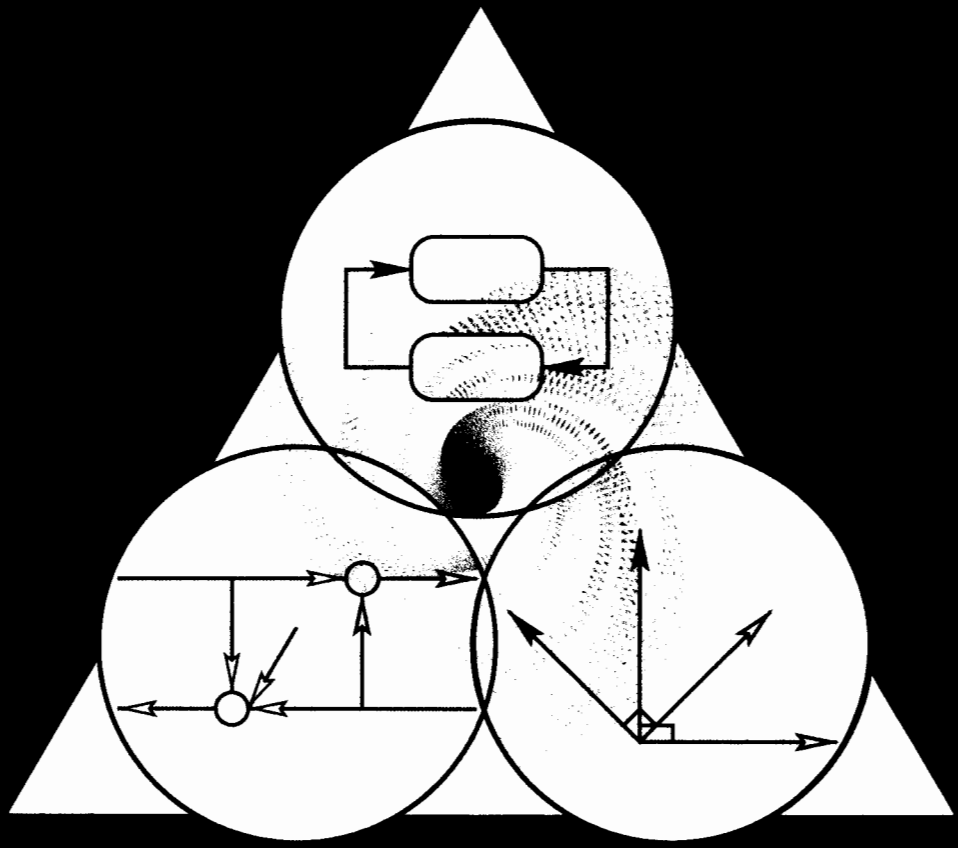
KAILATH
SAYED
HASSIBI

LINEAR ESTIMATION



THOMAS KAILATH ALI H. SAYED BABAK HASSIBI

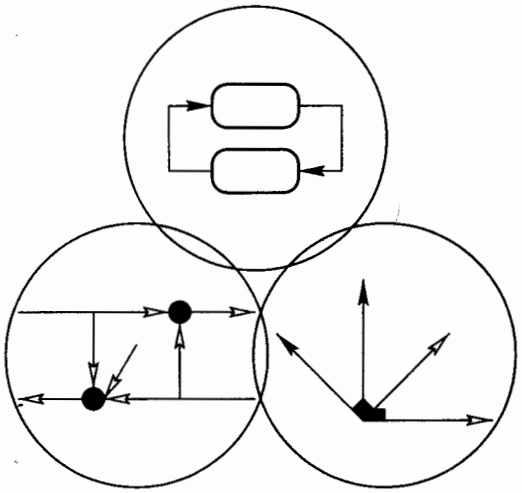
Prentice Hall Information and System Sciences Series

LINEAR ESTIMATION

ISBN 0-13-022464-2

9 780130 224644 90000

LINEAR ESTIMATION



T. Kailath
Stanford University

A. H. Sayed
UCLA

B. Hassibi
Bell Laboratories

PRENTICE HALL
Upper Saddle River, New Jersey 07458

RENTICE HALL INFORMATION AND SYSTEM SCIENCES SERIES

iomas Kailath, Editor

STROM & WITTENMARK	*Computer-Controlled Systems: Theory and Design, 2/E*
HATTACHARYA, CHAPELLAT & KEEL	*Robust Control: The Parametric Approach*
IEDLAND	*Advanced Control System Design*
ARDNER	*Statistical Spectral Analysis: A Nonprobabilistic Theory*
REWALL & ANDREWS	*Kalman Filtering: Theory and Practice*
AYKIN	*Adaptive Filter Theory, 3/E*
AYKIN, ED.	*Blind Deconvolution*
IN & CHELLAPA	*Fundamentals of Digital Image Processing, 2/E*
AILATH	*Linear Systems*
JNG	*VLSI Array Processors*
JNG, WHITEHOUSE & KAILATH, EDS.	*VLSI and Modern Signal Processing*
VAKERNAAK & SIVAN	*Signals and Systems*
UNG	*System Identification: Theory for the User*
UNG & GLAD	*Modeling of Dynamic Systems*
ACOVSKI	*Medical Imaging Systems*
OSCA	*Stochastic and Predictive Adaptive Control*
ARENDRA & ANNASWAMY	*Stable Adaptive Systems*
EKOOGAR & MORIARTY	*Digital Control Using Digital Signal Processing*
ORAT	*Digital Processing of Random Signals: Theory & Methods*
JGH	*Linear System Theory, 2/E*
ELIMAN & SRINATH	*Continuous and Discrete-Time Signal and Systems, 2/E*
ELO & KONG	*Adaptive Signal Processing Algorithms: Stability & Performance*
INATH, RAJASEKARAN, & VISWANATHAN	*Introduction to Statistical Signal Processing with Applications*
ELLS	*Applied Coding and Information Theory for Engineers*
LLIAMS	*Designing Digital Filters*

To our parents and families

of Congress Cataloging-in-Publication Data

Thomas.
ear estimation / T. Kailath, A.H. Sayed, B. Hassibi.
p. cm.
udes bibliographical references.
N 0-13-022464-2 (case)
stimation theory. 2. Least squares. I. Sayed, Ali H. II. Hassibi, Babak. III, Title.
3 .K33 1999
-dc21 99-047033

her: *Tom Robbins*
iate editor: *Alice Dworkin*
ction editor: *Audri Anna Bazlen*
resident and editorial director: *Marcia Horton*
tive managing editor: *Vince O'Brien*
ging editor: *David A. George*
resident of production and manufacturing: *David W. Riccardi*
rector: *Jayne Conte*
design: *Bruce Kenselaar*
facturing buyer: *Pat Brown*
eting manager: *Danny Hoyt*
osition: *PreTEX*

d in the United States of America
8 7 6 5 4 3 2 1

0-13-022464-2

ice-Hall International (UK) Limited, *London*
ice-Hall of Australia Pty. Limited, *Sydney*
ice-Hall Canada Inc., *Toronto*
ice-Hall Hispanoamericana, S.A., *Mexico*
ice-Hall of India Private Limited, *New Delhi*
ice-Hall of Japan, Inc., *Tokyo*
on Education Asia Pte. Ltd., *Singapore*
ra Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

Contents

Preface

The problem of estimating the values of a random (or stochastic) process given observations of a related random process is encountered in many areas of science and engineering, *e.g.*, communications, control, signal processing, geophysics, econometrics, and statistics. Although the topic has a rich history, and its formative stages can be attributed to illustrious investigators such as Laplace, Gauss, Legendre, and others, the current high interest in such problems began with the work of H. Wold, A. N. Kolmogorov, and N. Wiener in the late 1930s and early 1940s. N. Wiener in particular stressed the importance of modeling not just "noise" but also "signals" as random processes. His thought-provoking originally classified 1942 report, released for open publication in 1949 and now available in paperback form under the title *Time Series Analysis*, is still very worthwhile background reading.

As with all deep subjects, the extensions of these results have been very far-reaching as well. A particularly important development arose from the incorporation into the theory of multichannel state-space models. Though there were various earlier partial intimations and explorations, especially in the work of R. L. Stratonovich in the former Soviet Union, the chief credit for the explosion of activity in this direction goes to R. E. Kalman, who also made important related contributions to linear systems, optimal control, passive systems, stability theory, and network synthesis.

In fact, least-squares estimation is one of those happy subjects that is interesting not only in the richness and scope of its results, but also because of its mutually beneficial connections with a host of other (often apparently very different) subjects. Thus, beyond those already named, we may mention connections with radiative transfer and scattering theory, linear algebra, matrix and operator theory, orthogonal polynomials, moment problems, inverse scattering problems, interpolation theory, decoding of Reed–Solomon and BCH codes, polynomial factorization and root distribution problems, digital filtering, spectral analysis, signal detection, martingale theory, the so-called $\mathcal{H}_\infty$ theories of estimation and control, least-squares and adaptive filtering problems, and many others. We can surely apply to it the lines written by William Shakespeare about another (beautiful) subject:

"Age does not wither her, nor custom stale,
Her infinite variety."

Though we were originally tempted to cover a wider range, many reasons have led us to focus this volume largely on estimation problems for finite-dimensional linear systems with state-space models, covering most aspects of an area now generally known as Wiener and Kalman filtering theory. Three distinctive features of our treatment are the pervasive use of a geometric point of view, the emphasis on the numerically favored square-root/array forms of many algorithms, and the emphasis on equivalence and duality concepts for the solution of several related problems in adaptive filtering, estimation, and control. These features are generally absent in most prior treatments, ostensibly on the grounds that they are too abstract and complicated. It is our hope that these misconceptions will be dispelled by the presentation herein, and that the fundamental simplicity and power of these ideas will be more widely recognized and exploited.

The material presented in this book can be broadly categorized into the following topics:

- **Introduction and Foundations**
  - Chapter 1: Overview
  - Chapter 2: Deterministic Least-Squares Problems
  - Chapter 3: Stochastic Least-Squares Problems
  - Chapter 4: The Innovations Process
  - Chapter 5: State-Space Models
- **Estimation of Stationary Processes**
  - Chapter 6: Innovations for Stationary Processes
  - Chapter 7: Wiener Theory for Scalar Processes
  - Chapter 8: Recursive Wiener Filters
- **Estimation of Nonstationary Processes**
  - Chapter 9: The Kalman Filter
  - Chapter 10: Smoothed Estimators
- **Fast and Array Algorithms**
  - Chapter 11: Fast Algorithms
  - Chapter 12: Array Algorithms
  - Chapter 13: Fast Array Algorithms
- **Continuous-Time Estimation**
  - Chapter 16: Continuous-Time State-Space Estimation
- **Advanced Topics**
  - Chapter 14: Asymptotic Behavior
  - Chapter 15: Duality and Equivalence in Estimation and Control
  - Chapter 17: A Scattering Theory Approach

Being intended for a graduate-level course, the book assumes familiarity with basic concepts from matrix theory, linear algebra, linear system theory, and random processes. Four appendices at the end of the book provide the reader with background material in all these areas.

There is ample material in this book for the instructor to fashion a course to his or her needs and tastes. The authors have used portions of this book as the basis for one-quarter first-year graduate level courses at Stanford University, the University of California at Los Angeles, and the University of California at Santa Barbara; the students were expected to have had some exposure to discrete-time and state-space theory. A typical course would start with Secs. 1.1–1.2 as an overview (perhaps omitting the matrix derivations), with the rest of Ch. 1 left for a quick reading (and re-reading from time to time), most of Chs. 2 and 3 (focusing on the geometric approach) on the basic deterministic and stochastic least-squares problems, Ch. 4 on the innovations process, Secs. 6.4–6.5 and 7.3–7.7 on scalar Wiener filtering, Secs. 9.1–9.3, 9.5, and 9.7 on Kalman filtering, Secs. 10.1–10.2 as an introduction to smoothing, Secs. 12.1–12.5 and 13.1–13.4 on array algorithms, and Secs. 16.1–16.4 and 16.6 on continuous-time problems.

More advanced students and researchers would pursue selections of material from Sec. 2.8, Chs. 8, 11, 14, 15, and 17, and Apps. E and F. These cover, among other topics, least-squares problems with uncertain data, the problem of canonical spectral factorization, convergence of the Kalman filter, the algebraic Riccati equation, duality, backwards-time and complementary models, scattering, etc. Those wishing to go on to the more recent $\mathcal{H}_\infty$ theory can find a treatment closely related to the philosophy of the current book (*cf.* Sec. 1.6) in the research monograph of Hassibi, Sayed, and Kailath (1999).

A feature of the book is a collection of nearly 300 problems, several of which complement the text and present additional results and insights. However, there is little discussion of real applications or of the error and sensitivity analyses required for them. The main issue in applications is constructing an appropriate model, or actually a set of models, which are further analyzed and then refined by using the results and algorithms presented in this book. Developing good models and analyzing them effectively requires not only a good appreciation of the actual application, but also a good understanding of the theory, at both an analytical and intuitive level. It is the latter that we have tried to achieve here; examples of successful applications have to be sought in the literature, and some references are provided to this end.

Acknowledgments

The development of this textbook has spanned many years. So the material, as well as its presentation, has benefited greatly from the inputs of the many bright students who have worked with us on these topics: J. Omura, P. Frost, T. Duncan, R. Geesey, D. Duttweiler, H. Aasnaes, M. Gevers, H. Weinert, A. Segall, M. Morf, B. Dickinson, G. Sidhu, B. Friedlander, A. Vieira, S. Y. Kung, B. Levy, G. Verghese, D. Lee, J. Delosme, B. Porat, H. Lev-Ari, J. Cioffi, A. Bruckstein, T. Citron, Y. Bresler, R. Roy, J. Chun, D. Slock, D. Pal, G. Xu, R. Ackner, Y. Cho, P. Park, T. Boros, A. Erdogan, U. Forsell, B. Halder, H. Hindi, V. Nascimento, T. Pare, R. Merched, and our young friend Amir Ghazanfarian (*in memoriam*) from whom we had so much more to learn.

We are of course also deeply indebted to the many researchers and authors in this beautiful field. Partial acknowledgment is evident through the citations and references; while the list of the latter is quite long, we apologize for omissions and inadequacies arising from the limitations of our knowledge and our energy. Nevertheless, we would be remiss not to explicitly mention the inspiration and pleasure we have gained in studying the papers and books of N. Wiener, R. E. Kalman, and P. Whittle.

Major support for the many years of research that led to this book was provided by the Mathematics Divisions of the Air Force Office of Scientific Research and the Army Research Office, by the Joint Services Electronics Program, by the Defense Advanced Research Projects Agency, and by the National Science Foundation. Finally, we would like to thank Bernard Goodwin and Tom Robbins, as well as the staff of Prentice Hall, for their patience and other contributions to this project.

T. Kailath
Stanford, CA

A. H. Sayed
Westwood, CA

B. Hassibi
Murray Hill, NJ

Symbols

We collect here, for ease of reference, a list of the main symbols and signs used throughout the text.

$\mathbb{R}$	The set of real numbers.
$\mathbb{C}$	The set of complex numbers.
$.^T$	Matrix transposition.
$.^*$	Complex conjugation; Hermitian transposition.
$\diamond$	denotes the end of a theorem/lemma/proof/example/remark.
z^{-1}	denotes a unit-time delay.
$\alpha \in \mathcal{S}$	The element α belongs to the set $\mathcal{S}$.
$\mathbf{x}$	a boldface letter denotes a random variable.
x	a letter in normal font denotes a vector in Euclidean space.
$E\mathbf{x}$	denotes the expected value of a random variable $\mathbf{x}$.
$\langle \mathbf{x}, \mathbf{y} \rangle$	denotes $E\mathbf{x}\mathbf{y}^*$ for column random vectors $\mathbf{x}$ and $\mathbf{y}$.
$\|\mathbf{x}\|^2$	denotes $E\mathbf{x}\mathbf{x}^*$ for a zero-mean random variable $\mathbf{x}$.
$\mathbf{x} \perp \mathbf{y}$	denotes uncorrelated zero-mean random variables $\mathbf{x}$ and $\mathbf{y}$.

$\langle x, y \rangle$	denotes the inner product x^*y for column vectors x and y.
$\|x\|^2$	denotes x^*x for a column vector x.
$\|x\|$	denotes $\sqrt{x^*x}$ for a column vector x.
$x \perp y$	denotes orthogonal vectors x and y.
$\|A\|_2$	The 2-induced norm = the maximum singular value of A.
$\|A\|_F$	The Frobenius norm of A.
$a \triangleq b$	The quantity a is defined as b.
$a \propto b$	The quantity a is proportional to b.
$\text{col}\{a, b\}$	a column vector with entries a and b.
$\text{vec}\{A\}$	a column vector formed by stacking the columns of A.
$\text{diag}\{a, b\}$	a diagonal matrix with diagonal entries a and b.
$a \oplus b$	The same as $\text{diag}\{a, b\}$.
0	a zero scalar, vector, or matrix.
$\mathbf{I}_n$	The identify matrix of size $n \times n$.
$x(z)$ or $X(z) = \mathcal{Z}\{x_i\}$	denotes the bilateral z-transform of a sequence $\{x_i\}$.
$X(f) = \mathcal{F}\{x(t)\}$	denotes the Fourier transform of a function $x(t)$.
$X(s) = \mathcal{L}\{x(t)\}$	denotes the bilateral Laplace transform of $x(t)$.
$X(e^{j\omega})$	denotes the Discrete-Time Fourier Transform of $\{x_i\}$.
$\mathcal{L}\{\mathbf{x}_1, \mathbf{x}_2, ..\}$	denotes the linear span of the variables $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$.
$\hat{\mathbf{x}}_{i\|j}$	l.l.m.s. estimator of $\mathbf{x}_i$ given observations up to time j.
$\hat{\mathbf{x}}_i$	l.l.m.s. estimator of $\mathbf{x}_i$ given observations up to time $i-1$.
$\tilde{\mathbf{x}}_{i\|j}$	The estimation error $\mathbf{x}_i - \hat{\mathbf{x}}_{i\|j}$.
$\tilde{\mathbf{x}}_i$	The estimation error $\mathbf{x}_i - \hat{\mathbf{x}}_i$.
$\{\cdot\}_+$	Causal part of a transfer function.
$\{\cdot\}_-$	Anti-causal part of a transfer function.
$\{\cdot\}_{s.c.}$	Strictly causal part of a transfer function.

$P > 0$	a positive-definite (p.d.) matrix P.
$P \geq 0$	a positive-semidefinite (p.s.d.) matrix P.
$P^{1/2}$	a square-root factor of a matrix $P \geq 0$, usually triangular.
$A > B$	means that $A - B$ is positive-definite.
$A \geq B$	means that $A - B$ is positive-semidefinite.
$\det A$	Determinant of the matrix A.
$\text{trace } A$	Trace of the matrix A.
$O(n)$	A constant multiple of n, or of the order of n.
QR	The QR factorization of a matrix.
LDU	Lower-diagonal-upper decomposition of a matrix.
UDL	Upper-diagonal-lower decomposition of a matrix.
LDL^*	LDU decomposition of a Hermitian matrix.
UDU^*	UDL decomposition of a Hermitian matrix.
Thm.	"Theorem."
Cor.	"Corollary."
Def.	"Definition."
Fig.	"Figure."
LTI	"Linear time-invariant."
l.l.s.	"linear least-squares."
l.l.m.s.	"linear least-mean-squares."
l.l.m.s.e	"linear least-mean-squares estimation/estimator."
m.m.s.e.	"minimum mean-square error."
LS	"least-squares."
p.d.f.	"probability density function."
iff	"if and only if."
a.e.	"almost everywhere."
w.r.t.	"with respect to."
RHS	"right-hand side."
LHS	"left-hand side."

ROC	"Region of convergence."
ARE	"Algebraic Riccati equation."
DARE	"Discrete-time algebraic Riccati equation."
CARE	"Continuous-time algebraic Riccati equation."
LMI	"Linear Matrix Inequality."
AR	"Autoregressive model."
MA	"Moving average model."
ARMA	"Autoregressive moving average model."
FIR	"Finite impulse response filter."
IIR	"Infinite impulse response filter."
SNR	"Signal to noise ratio."
MAP	"Maximum a-posteriori."
EKF	"Extended Kalman filter."
SISO	"Single-input single-output."
MIMO	"Multiple-input multiple-output."

CHAPTER 1

Overview

Estimation problems arise in diverse fields, such as communications, control, econometrics, and signal processing. Underlying these, of course, are many general results in probability and statistics. What distinguishes the particular applications mentioned above is the fact that they have additional structure that can be used to further refine these general results. The proper exploration and exploitation of this structure lead to many new problems and results.

In this book we shall focus on a certain rather narrowly defined class of problems. This is essentially the study of linear least-squares estimation problems for signals with known finite-dimensional linear state-space models. [The terminology, for those unfamiliar with it, will be made clear very soon in Sec. 1.2.] However, despite its apparent narrowness, this is a rich subject with useful applications in fields such as quadratic control, adaptive filtering, $\mathcal{H}_\infty$-filtering and control, signal detection, and even matrix theory and linear algebra (see Sec. 1.6).

Since many readers will have had prior exposure to state-space theory, we begin with a brief review of what is called an *asymptotic observer* for determining the states, and then describe how to modify it in the presence of random disturbances (Sec. 1.2). The resulting *optimum transient observer* will be, in fact, a basic form of the celebrated *Kalman filter* for state-space estimation, and will raise some of the important issues to be studied in this book (see Secs. 1.3–1.5). There is at least one, small but important, fact that holds us back from plunging straight into a detailed study of the Kalman filter: in observer theory we assume a particular form for the solution and just have to optimize the choice of a particular parameter. How do we know that a different initial structure would not give a better solution? In fact, the assumed structure in Sec. 1.2 is "optimum," but it is very valuable to back off and take a more fundamental approach in which no particular structure, no matter how reasonable, is assumed a priori.

We shall start out on this more leisurely route, compared to the deliberately rapid pace of this review chapter, in Ch. 2 and only actually return to the Kalman filter itself in Ch. 9. This might appear to be an unnecessary delay to some readers. However our apparent detour will have several benefits, not only in solving problems for which no reasonable structure is evident a priori (*e.g.*, finding what are called *smoothed* estimators — see Sec. 1.3.1), but also in better understanding the properties of the Kalman filter

itself, *e.g.*, its asymptotic behavior (see Sec. 1.5). For this it will be necessary to bring in the covariance functions of the state and observation processes, which was the starting point of the earlier Wiener approach to the estimation problem. The two formulations — starting with a (state-space) model or starting with the covariance or power spectral (in the stationary case) information — complement each other nicely, as we shall already be able to illustrate in this introductory chapter (see Sec. 1.5).

At least at first glance, it would seem from the rapid survey in this overview chapter, that the subject involves a host of special formulas and massive algebraic manipulations. However, we shall show that a geometric formulation of the estimation problem — treating random variables as vectors with matrix-valued inner products — will bring both simplicity and order into the whole subject. The foundation for this treatment is developed essentially from first principles, in Chs. 2 to 4. Ch. 5 both motivates and surveys state-space descriptions. Ch. 6 reviews some basic facts from linear systems and their interaction with second-order random processes. Ch. 7 treats the classical Wiener theory for scalar-valued stationary processes. Then Ch. 8 shows how the use of state-space structure completes the theory for vector-valued processes and presages the Kalman filter for nonstationary and stationary processes. A glimpse of the material in later chapters will be provided in Sec. 1.3.

1.1 THE ASYMPTOTIC OBSERVER

We shall begin with a problem probably familiar to those who have studied linear system theory. For others, enough background is provided so that they can follow the discussion; those less accustomed to state-space language may wish to consult any of the numerous textbooks now available.

It should also be emphasized that the intent of this chapter is largely motivational; we are planning to explore a large territory, and this preliminary reconnaissance is undoubtedly going to be harder for many interested readers. Readers will be well advised merely to skim through this chapter, and not to be unduly concerned about the details — there will be ample opportunity later to explore the details in a more leisurely fashion.

In deterministic linear system theory, the so-called *observer design* problem is one of developing a realistic solution to the problem of determining the states of a linear system, given access to its inputs and outputs. Consider a linear system in state-space form,

$$\begin{cases} x_{i+1} = Fx_i + Gu_i, & i \geq 0, \\ y_i = Hx_i, \end{cases} \tag{1.1.1}$$

where $F \in \mathbb{C}^{n\times n}$, $G \in \mathbb{C}^{n\times m}$, $H \in \mathbb{C}^{p\times n}$ are known matrices, $\{u_i, i \geq 0\}$ is a *known* input sequence, and $\{y_i, i \geq 0\}$ is the observed system output. The problem is to determine the state-sequence $\{x_i, i \geq 0\}$. Here $\mathbb{C}$ denotes the field of complex numbers.

With this wealth of information, the problem would not seem to be hard, especially if we know the initial state, x_0, or at least have a good estimate of it, say $\hat{x}_0$. For then we can set up a "dummy" system, using the known F, G, $\{u_i\}$, and the initial estimate $\hat{x}_0$,

$$\hat{x}_{i+1} = F\hat{x}_i + Gu_i, \quad i \geq 0, \tag{1.1.2}$$

and just compute $\hat{x}_1, \hat{x}_2, \ldots$, and so on. Unfortunately, no matter how good the estimate $\hat{x}_0$ is, unless it is perfect, the above method will not work if the matrix F is unstable,

i.e., if F has eigenvalues that lie outside the unit disc. To see this, note that the error

$$x_i - \hat{x}_i \triangleq \tilde{x}_i, \quad \text{say,} \tag{1.1.3}$$

obeys the equation

$$\tilde{x}_{i+1} = F\tilde{x}_i, \quad \tilde{x}_0 = x_0 - \hat{x}_0,$$

so that

$$\tilde{x}_i = F^i\tilde{x}_0. \tag{1.1.4}$$

If F is unstable, $\|\tilde{x}_i\|$ will grow exponentially (unless $\tilde{x}_0$ is completely in the subspace spanned by the eigenvectors and generalized eigenvectors corresponding to the stable eigenvalues of F; since the error $\tilde{x}_0$ can be quite arbitrary, this condition will, in general, not be met). The notation $\|\cdot\|$ stands for the Euclidean norm of a vector. Even when F is stable, Eq. (1.1.4) shows that the rate of convergence of $\tilde{x}_i$ to zero is determined by the eigenvalues of the given matrix F, and it would therefore be desirable to have a mechanism that allows us to control this rate of convergence.

In making a postmortem analysis of this unsuccessful attempt, we may note that one problem with it is that it is "open-loop": it makes no use of the information in the available output sequence $\{y_i\}$. Thus

$$\tilde{y}_i \triangleq y_i - \hat{y}_i = H(x_i - \hat{x}_i) = H\tilde{x}_i, \quad i \geq 0, \tag{1.1.5}$$

will give us an indication of how $\tilde{x}_i$ is behaving (provided it is not always in the nullspace of H, again an unlikely event given that we may have no *a priori* information on $\tilde{x}_0$). From past experience with the beneficial effects of "feedback," we can consider using $\tilde{y}_i$ in a feedback mode so as to drive the error $\tilde{x}_i$ to zero.

Thus, suppose we also drive the dummy system (1.1.2) with a term $\bar{K}\tilde{y}_i$, where $\bar{K} \in \mathbb{C}^{n\times p}$ is to be suitably chosen,

$$\hat{x}_{i+1} = F\hat{x}_i + Gu_i + \bar{K}\tilde{y}_i. \tag{1.1.6}$$

The error will now obey the equation

$$\tilde{x}_{i+1} = F\tilde{x}_i - \bar{K}\tilde{y}_i = (F - \bar{K}H)\tilde{x}_i, \tag{1.1.7}$$

so that

$$\tilde{x}_i = (F - \bar{K}H)^i\tilde{x}_0. \tag{1.1.8}$$

This reduces, of course, to the previous expression when $\bar{K} = 0$. But clearly, if we can choose $\bar{K} \neq 0$ so that $F - \bar{K}H$ has any desired eigenvalues, then the error $\tilde{x}_i$ can be made to go to zero as i increases and also at any desired rate! In fact, it is a well-known result in linear system theory (see, *e.g.*, Kailath (1980), Callier and Desoer (1991) , Antsaklis and Michel (1997)) that under a certain condition, *viz.*, that $\{F, H\}$ is observable, we can always find $\bar{K} \in \mathbb{C}^{n\times p}$ to make $(F - \bar{K}H)$ have n arbitrary eigenvalues. Therefore, we can make the error go to zero as fast as we wish by suitably choosing the *feedback gain matrix*, $\bar{K}$. For obvious reasons, the system described by (1.1.6) is called an *asymptotic observer*.

The observability requirement can be defined and characterized in several different ways, all discussed in the books mentioned above (and others — see also App. C).

Here we may just note one of them:

$$\{F, H\} \text{ is observable} \iff \mathcal{O} \triangleq \begin{bmatrix} H^* & F^* & H^* & \dots & F^{*(n-1)}H^* \end{bmatrix}^* \text{ has full rank,}$$

where $\mathcal{O}$ is referred to as the *observability matrix* associated with $\{F, H\}$, and the $*$ denotes Hermitian conjugation (complex conjugation for scalars).

1.2 THE OPTIMUM TRANSIENT OBSERVER

There are other variations and refinements of the asymptotic observer described in the literature of linear system theory. Here we shall address the issue of how this solution behaves in the inevitable presence of noise in the measurements, or noise arising from the process of (digital) computation. In fact, the inevitable uncertainties in the model itself (*i.e.*, the approximations and assumptions involved in setting up the linear $\{F, G, H\}$ model) contribute to randomness in the state, the inputs, and the outputs. The proper modeling of such phenomena is a large and rather imprecise subject by itself, which we shall not explore in this book. For the moment, it will suffice to say that a lot of experiment and experience leads us to generalize the noiseless state-space model (1.1.1) to the following *noisy* state-space model:

$$\begin{cases} \mathbf{x}_{i+1} = F\mathbf{x}_i + G(u_i + \mathbf{u}_i), & i \geq 0, \\ \mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{1.2.1}$$

where the so-called process noise $\{\mathbf{u}_i\}$, the measurement noise $\{\mathbf{v}_i\}$, and the initial state $\mathbf{x}_0$ are all assumed to be random. This *primary* randomness in turn makes the states $\{\mathbf{x}_i\}$ and the outputs $\{\mathbf{y}_i\}$ themselves random. *Note that to distinguish between the deterministic quantities in (1.1.1) and the random quantities in (1.2.1), we have used boldface characters to denote random variables.* We shall adopt this convention throughout this book.

How shall we describe the random processes $\{\mathbf{u}_i\}$, $\{\mathbf{v}_i\}$, and the initial state, $\mathbf{x}_0$? Again, there are various possibilities and various circumstances, but the model we shall study is where all random variables have known mean values, which we can take without loss of generality to be zero (see Prob. 1.1), and known covariances, as follows. The sequences $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are (so-called) white-noise processes , with variance matrices Q and R, respectively, that is,

$$E\mathbf{u}_i\mathbf{u}_j^* = Q\delta_{ij} \quad \text{and} \quad E\mathbf{v}_i\mathbf{v}_j^* = R\delta_{ij},$$

where δ_{ij} is the Kronecker delta function (identically zero except when $i = j$). The initial condition, $\mathbf{x}_0$, is zero-mean, has variance Π_0, and is uncorrelated with the processes $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$, *i.e.*,

$$E\mathbf{x}_0\mathbf{x}_0^* = \Pi_0, \quad E\mathbf{u}_i\mathbf{x}_0^* = 0, \quad E\mathbf{v}_i\mathbf{x}_0^* = 0.$$

It is natural to ask about dependence between the processes $\{\mathbf{u}_i, \mathbf{v}_i\}$. A common assumption is that they are completely uncorrelated, because often they have different physical origins — the $\{\mathbf{u}_i\}$ arises from disturbances and uncertainties in the system, while the $\{\mathbf{v}_i\}$ arises from the measurement process. However, in situations where feedback is involved, *e.g.*, in control problems where the output $\{\mathbf{y}_i\}$ is used to modify

the state equation, it will be useful to consider models in which there is some dependence. This can be of different forms, but one of the most useful is to assume that they are correlated only at the same time instant, or perhaps one instant apart, *i.e.*, to assume that

$$E\mathbf{u}_i\mathbf{v}_i^* = S\delta_{ij} \quad \text{or} \quad E\mathbf{u}_{i+1}\mathbf{v}_i^* = S\delta_{ij},$$

Here, for definiteness, we shall make the first choice.

The above assumptions can be compactly summarized by the following equation, where the alert reader will notice that the zero-mean assumption has also been included:

$$E\begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix}\begin{bmatrix} \mathbf{u}_j^* & \mathbf{v}_j^* & \mathbf{x}_0^* & 1 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\delta_{ij} & \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \Pi_0 & 0 \end{bmatrix}. \tag{1.2.2}$$

The matrices $Q \in \mathbb{C}^{m\times m}$, $R \in \mathbb{C}^{p\times p}$, $S \in \mathbb{C}^{m\times p}$, $\Pi_0 \in \mathbb{C}^{n\times n}$ are assumed to be known. We also note that, by definition, the matrices $\{Q, R\}$ must be Hermitian and nonnegative-definite, *i.e.*,

$$Q = Q^* \geq 0, \qquad R = R^* \geq 0.$$

The matrix S need not be Hermitian, but it is not completely arbitrary, because it must be true that

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \geq 0,$$

since the above matrix denotes the variance matrix of the aggregate vector process $\text{col}\{\mathbf{u}_i, \mathbf{v}_i\}$.[1]

Sometimes a point of confusion, for beginning readers, is that though the sequences $\{\mathbf{u}_i, \mathbf{v}_i\}$ are white-noise processes, when $m > 1$ and $p > 1$, the covariance matrices $\{Q, R, S\}$ are not necessarily diagonal matrices. In other words, the components of the $m \times 1$ vector $\mathbf{u}_i$ may be correlated with each other, even though all of them are completely uncorrelated with the components of the vector $\mathbf{u}_j$, $j \neq i$. [We may also remark, for more advanced readers, that the assumptions in (1.2.2) ensure that the processes $\{\mathbf{x}_i\}$ and $\text{col}\{\mathbf{x}_i, \mathbf{y}_i\}$ are so-called wide-sense Markov processes, a property discussed in detail in Ch. 5.]

After this long introduction of the model, we now return to the observer problem. We first note that because of the randomness introduced into the model, the observer equation (1.1.6) will have to be written as

$$\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + Gu_i + \bar{K}(\mathbf{y}_i - \hat{\mathbf{y}}_i) \quad \text{with} \quad E\hat{\mathbf{x}}_0 = 0 = E\mathbf{x}_0, \tag{1.2.3}$$

where the predicted output $\hat{\mathbf{y}}_i$ is still taken as $H\hat{\mathbf{x}}_i$ because there is nothing we know at the moment to allow us to make any guess at what the random variable $\mathbf{v}_i$ may be; so also for $E\hat{\mathbf{x}}_0$, which is taken to be zero. Therefore now

$$\tilde{\mathbf{y}}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i = H\mathbf{x}_i + \mathbf{v}_i - H\hat{\mathbf{x}}_i = H(\mathbf{x}_i - \hat{\mathbf{x}}_i) + \mathbf{v}_i = H\tilde{\mathbf{x}}_i + \mathbf{v}_i, \tag{1.2.4}$$

[1] The notation $\text{col}\{a, b\}$ denotes a column vector whose entries are a and b.

and subtracting (1.2.3) from (1.2.1), we have, instead of the homogeneous equation (1.1.7),

$$\tilde{\mathbf{x}}_{i+1} = F\tilde{\mathbf{x}}_i + G\mathbf{u}_i - \bar{K}(H\tilde{\mathbf{x}}_i + \mathbf{v}_i) = (F - \bar{K}H)\tilde{\mathbf{x}}_i + G\mathbf{u}_i - \bar{K}\mathbf{v}_i. \tag{1.2.5}$$

This is a big difference, because with the nonzero *driving* terms $\{G\mathbf{u}_i, \bar{K}\mathbf{v}_i\}$, we cannot now assert that the stability of $F - \bar{K}H$ will ensure that $\tilde{\mathbf{x}}_i$ will tend to zero; rather $\{\tilde{\mathbf{x}}_i\}$ itself will now asymptotically be a (wide-sense) stationary random process. The mean value of the error will obey the same equation as (1.1.7), *viz.*,

$$E\tilde{\mathbf{x}}_{i+1} = (F - \bar{K}H)E\tilde{\mathbf{x}}_i \quad \text{with} \quad E\tilde{\mathbf{x}}_0 = 0,$$

which means that the mean value of the error will be identically zero (rather than asymptotically zero), no matter what $\bar{K}$ is. However, the actual error will fluctuate about the mean value and the point is to try to choose $\bar{K}$ to minimize the average (or expected) value of some function of the error. The proper choice of this function of the error is another topic for much debate (see the notes in Sec. 1.7).

1.2.1 The Mean-Square-Error Criterion

For various reasons, the most studied choice is the square function, so that the criterion is to minimize the mean-square value of the error (see the notes at the end of this chapter). However there are still some choices to be made. We could ask, for each i, to minimize the variance of each component of the error vector $\tilde{\mathbf{x}}_i$, or more generally, to minimize for each i the variance of an arbitrary linear combination of the components of the error vector, say $a^*\tilde{\mathbf{x}}_i$. This variance is a quadratic form in the column vector a, written as

$$\xi_{i+1}(a) = a^*(E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*)a, \quad \text{for} \quad i \geq 0. \tag{1.2.6}$$

It turns out that because of the special structure of $E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*$, which results from $\tilde{\mathbf{x}}_{i+1}$ being linear in the matrix $\bar{K}$, we can choose $\bar{K}$ so that it *simultaneously* minimizes $\xi_{i+1}(a) = a^*(E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*)a$ for *all* nonzero vectors $a \in \mathbb{C}^n$.

Whenever a quadratic form in a, say $a^*P(\bar{K})a$, can be minimized over a (matrix) parameter $\bar{K}$ for all $a \in \mathbb{C}^n$, where $P(\bar{K}) \in \mathbb{C}^{n\times n}$ is some positive-semi-definite matrix function of $\bar{K}$, it means that the optimum choice, say $\bar{K}_o$, has the property that $a^*P(\bar{K})a \geq a^*P(\bar{K}_o)a$, *i.e.*,

$$a^*[P(\bar{K}) - P(\bar{K}_o)]a \geq 0$$

for all $a \in \mathbb{C}^n$ and for all $\bar{K}$. This means that $P(\bar{K}) - P(\bar{K}_o)$ is nonnegative-definite for all $\bar{K}$, which is commonly denoted as $P(\bar{K}) \geq P(\bar{K}_o)$. In this sense we can speak of minimizing not only the quadratic form $\xi_{i+1}(a) = a^*(E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*)a$ for all $a \in \mathbb{C}^n$, but in fact of minimizing the matrix error variance, $E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*$, itself. *This is the terminology we shall use henceforth.*

A further point to be made is that one expects the choice of $\bar{K}$ to depend upon the instant i at which the minimum is sought. Therefore we shall allow $\bar{K}$ to vary with i and write the estimator equation as

$$\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + Gu_i + \bar{K}_i(\mathbf{y}_i - H\hat{\mathbf{x}}_i). \tag{1.2.7}$$

Finally, what about the initial condition? In the deterministic case, we could assign any value for $\hat{x}_0$, *e.g.*, $\hat{x}_0 = 0$. But when we have a random $\mathbf{x}_0$, in order to be consistent with (1.2.6), we should choose a deterministic initial value $\hat{x}_0$ such that

$$E(\mathbf{x}_0 - \hat{x}_0)(\mathbf{x}_0 - \hat{x}_0)^* = \text{ a minimum.}$$

A simple calculation shows (see also Prob. 1.3) that the best choice is $\hat{x}_0 = E\mathbf{x}_0$, which in our case is equal to zero, by assumption. So now we can write a complete error equation,

$$\begin{aligned}\tilde{\mathbf{x}}_{i+1} &= (F - \bar{K}_iH)\tilde{\mathbf{x}}_i + G\mathbf{u}_i - \bar{K}_i\mathbf{v}_i, \quad \tilde{\mathbf{x}}_0 = \mathbf{x}_0. \\ &= (F - \bar{K}_iH)\tilde{\mathbf{x}}_i + \begin{bmatrix} G & -\bar{K}_i \end{bmatrix}\begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \quad \tilde{\mathbf{x}}_0 = \mathbf{x}_0.\end{aligned} \tag{1.2.8}$$

Equation (1.2.8) shows that $\tilde{\mathbf{x}}_i$ can be expressed as a linear combination of the vector-valued random variables $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}, \mathbf{v}_0, \ldots, \mathbf{v}_{i-1}\}$. Since both $\mathbf{u}_i$ and $\mathbf{v}_i$ are uncorrelated with $\mathbf{x}_0$ and with past $\mathbf{u}_j$ and $\mathbf{v}_j$, it follows that

$$E\tilde{\mathbf{x}}_i\mathbf{u}_i^* = 0 \quad \text{and} \quad E\tilde{\mathbf{x}}_i\mathbf{v}_i^* = 0.$$

Using these facts, a direct calculation gives the recursion

$$E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^* = \tag{1.2.9}$$

$$(F - \bar{K}_iH)(E\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^*)(F - \bar{K}_iH)^* + \begin{bmatrix} G & -\bar{K}_i \end{bmatrix}\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\begin{bmatrix} G^* \\ -\bar{K}_i^* \end{bmatrix}.$$

For later convenience, we shall define

$$P_i \triangleq \text{ the covariance matrix of the error in the } \textit{optimum} \text{ estimator of } \mathbf{x}_i. \tag{1.2.10}$$

Since we have already noted that the optimum estimator of $\mathbf{x}_0$ is zero, we shall have $P_0 = \Pi_0$, the covariance matrix of $\mathbf{x}_0$ (*cf.* (1.2.2)).

1.2.2 Minimization via Completion of Squares

We shall now proceed to determine the optimum values of $\bar{K}_i$ (and hence the optimum estimators $\hat{\mathbf{x}}_i$) by induction. To this end, suppose we have found the optimum estimator of $\mathbf{x}_i$ and the corresponding error covariance matrix, P_i. Then the optimum choice of $\bar{K}_i$, denoted by $K_{p,i}$ (p for prediction — see Prob. 1.4) that will give us the minimum value of the term $E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*$ is given by the following lemma, where for various reasons, we have made the additional assumption that $R > 0$ (see though Remark 1 further ahead).

Lemma 1.2.1 (Optimum Prediction Gain Matrix) *Assume $R > 0$. Given P_i as in (1.2.10), the optimum choice for the gain matrix $\bar{K}_i$ in (1.2.7) that minimizes the matrix error variance $E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*$ in (1.2.9) is*

$$K_{p,i} = (FP_iH^* + GS)R_{e,i}^{-1} \qquad \textit{where} \qquad R_{e,i} \triangleq R + HP_iH^*. \tag{1.2.11}$$

The corresponding minimum value of the mean-square error at time $(i+1)$ *is given by the so-called discrete-time Riccati recursion*

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0 = \Pi_0. \tag{1.2.12}$$

Also $R_{e,i}$ *has the interpretation*

$$R_{e,i} = E\mathbf{e}_i\mathbf{e}_i^* \qquad \textit{where} \qquad \mathbf{e}_i \triangleq \mathbf{y}_i - H\hat{\mathbf{x}}_i = H\tilde{\mathbf{x}}_i + \mathbf{v}_i. \tag{1.2.13}$$

■

Proof: Expression (1.2.9) can be rearranged as

$$E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^* = \begin{bmatrix} I & \bar{K}_i \end{bmatrix} \begin{bmatrix} FP_iF^* + GQG^* & -(FP_iH^* + GS) \\ -(HP_iF^* + S^*G^*) & R + HP_iH^* \end{bmatrix} \begin{bmatrix} I \\ \bar{K}_i^* \end{bmatrix}.$$

Now using an *upper-lower* block triangular factorization (see App. A) we may write the above 2×2 block matrix as

$$\begin{bmatrix} FP_iF^* + GQG^* & -(FP_iH^* + GS) \\ -(HP_iF^* + S^*G^*) & R + HP_iH^* \end{bmatrix} = \tag{1.2.14}$$

$$\begin{bmatrix} I & -(FP_iH^* + GS)R_{e,i}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Delta & 0 \\ 0 & R_{e,i} \end{bmatrix} \begin{bmatrix} I & 0 \\ -R_{e,i}^{-1}(FP_iH^* + GS)^* & I \end{bmatrix},$$

where, for compactness, we have introduced the notation

$$\Delta \triangleq FP_iF^* + GQG^* - (FP_iH^* + GS)R_{e,i}^{-1}(FP_iH^* + GS)^*,$$

as well as $R_{e,i} = R + HP_iH^*$. Note that the above factorization is well defined since $R > 0$ (by assumption) and $P_i \geq 0$ (by definition) imply that $R_{e,i} = R + HP_iH^* > 0$. Therefore, the expression for $E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*$ readily becomes

$$E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^* = \Delta + (\bar{K}_i - K_{p,i})R_{e,i}(\bar{K}_i - K_{p,i})^*. \tag{1.2.15}$$

Therefore the error criterion (1.2.6) is equal to the following

$$\xi_{i+1}(a) = a^*\left[FP_iF^* + GQG^* - (FP_iH^* + GS)R_{e,i}^{-1}(FP_iH^* + GS)^*\right]a + a^*\left[(\bar{K}_i - K_{p,i})R_{e,i}(\bar{K}_i - K_{p,i})^*\right]a.$$

Since the first term in the above expression is independent of $\bar{K}_i$, and since $R_{e,i} > 0$ (so that the second term is always nonnegative-definite), we readily see that to make $\xi_{i+1}(a)$ as small as possible we must choose $a^*(\bar{K}_i - K_{p,i}) = 0$. In order for this choice of $\bar{K}_i$ to simultaneously minimize $\xi_{i+1}(a)$ for *all* $a \in \mathbb{C}^n$, we must set

$$\bar{K}_i = K_{p,i} = (FP_iH^* + GS)R_{e,i}^{-1}.$$

With this choice, which is independent of a, we immediately see that the minimum error covariance matrix P_{i+1} is given by

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^* = \Delta.$$

Finally, using $\mathbf{e}_i = H\tilde{\mathbf{x}}_i + \mathbf{v}_i$, a straightforward calculation will show that $E\mathbf{e}_i\mathbf{e}_i^* = R_{e,i}$. ◆

Comparing (1.2.9) and (1.2.15) leads to the characterization of the above proof as a *completion of squares* argument, used, as all readers will recall, to solve (scalar) quadratic equations. The algebra gets less obvious in the matrix case; however, the derivation via the matrix factorization formula makes the algebra a little more routine (*i.e.*, we may need less trial and error to get (1.2.15)).

Remark 1. The above derivation relied on the invertibility of $R_{e,i}$, which we guaranteed by our assumption that $R > 0$. Neither of these assumptions is essential (though both turn out to be very good to try to ensure by proper modeling of the physical problem). The general result is that we can use any solution $K_{p,i}$ of the equation[2]

$$K_{p,i}R_{e,i} = FP_iH^* + GS. \tag{1.2.16}$$

This can be proved by rewriting the block triangular factorization (1.2.14) used in the proof as

$$\begin{bmatrix} FP_iF^* + GQG^* & -(FP_iH^* + GS) \\ -(HP_iF^* + S^*G^*) & R + HP_iH^* \end{bmatrix} = \begin{bmatrix} I & -K_{p,i} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Delta & 0 \\ 0 & R_{e,i} \end{bmatrix} \begin{bmatrix} I & 0 \\ -K_{p,i}^* & I \end{bmatrix}$$

where

$$\Delta = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*,$$

and $K_{p,i}$ is any solution of (1.2.16). An alternative derivation is outlined in Prob. 1.2. ◆

1.2.3 The Optimum Transient Observer

The block diagram depiction in Fig. 1.1 enables us to summarize the above discussion of the optimum transient observer (where the blocks with the symbol z^{-1} denote unit-time delays).

Given a system that can be described by the state-space model depicted in the top-block, we wish to operate on the observable inputs $\{u_i\}$ and the observable outputs $\{\mathbf{y}_i\}$ to find the minimum mean-square-error estimators of the (not directly observable) states $\{\mathbf{x}_i\}$. For this purpose, we set up a model of the system as shown in the lower block of Fig. 1.1. However, instead of the unavailable random inputs $\{\mathbf{u}_i\}$ and the unknown random initial state $\mathbf{x}_0$, we drive our model with a "feedback" term proportional to the error $\mathbf{e}_i = \mathbf{y}_i - H_i\hat{\mathbf{x}}_i$ and with initial state $\hat{x}_0 = 0$ (the mean value of the random variable $\mathbf{x}_0$). By properly choosing the proportionality matrix $K_{p,i}$, we can minimize the mean-square error $E\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^*$ at each time instant i.

[2] This equation is always consistent, meaning that it always has a solution $K_{p,i}$. This is because, as will be shown in Prob. 3.22, equations of the form $KE\mathbf{y}\mathbf{y}^* = E\mathbf{x}\mathbf{y}^*$, for any random variables $\{\mathbf{x}, \mathbf{y}\}$, are always consistent. Now, it is argued in Prob. 1.5, and also in Sec. 9.2.2, assuming $u_i = 0$ for simplicity, that the term $FP_iH^* + GS$ in (1.2.16) is equal to $E\mathbf{x}_{i+1}\mathbf{e}_i^*$. Hence, the equation (1.2.16) has the special form $K_{p,i}E\mathbf{e}_i\mathbf{e}_i^* = E\mathbf{x}_{i+1}\mathbf{e}_i^*$, in terms of the variance and cross-covariance matrices of two random variables $\{\mathbf{x}_{i+1}, \mathbf{e}_i\}$. The existence of solutions $K_{p,i}$ will thus follow from Prob. 3.22.

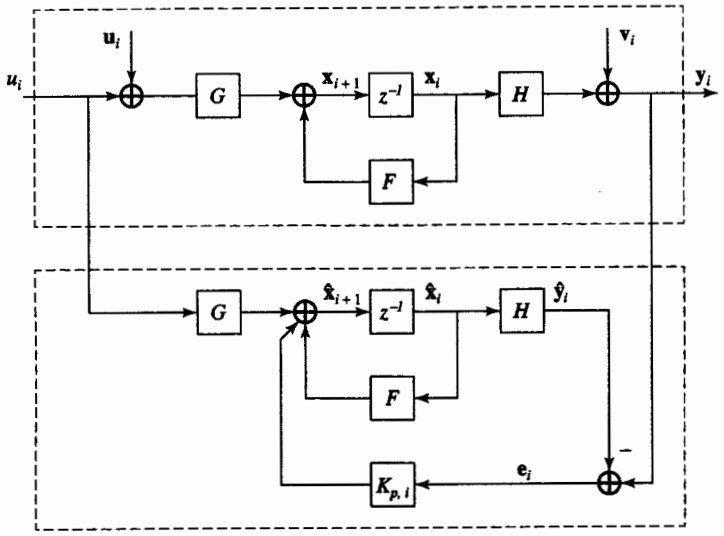


Figure 1.1 The optimum transient observer.

Moreover, under reasonable assumptions, it turns out that as $i \to \infty$, $K_{p,i}$ tends to a constant matrix K_p, which defines the optimum steady-state observer — see Sec. 1.5. We summarize the above discussions in the following theorem.

Theorem 1.2.1 (Optimum Transient Observer) *Consider a state-space system specified by (1.2.1)–(1.2.2), where $R > 0$, and a hypothesized observer structure*

$$\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + Gu_i + K_{p,i}(\mathbf{y}_i - H\hat{\mathbf{x}}_i), \quad \hat{\mathbf{x}}_0 = 0. \tag{1.2.17}$$

We can minimize the error covariance matrix, $E\tilde{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^*$, for $i \geq 0$, by choosing

$$K_{p,i} = (FP_iH^* + GS)R_{e,i}^{-1} \quad \text{where} \quad R_{e,i} = R + HP_iH^*, \tag{1.2.18}$$

and the $\{P_i\}$ can be found recursively via the discrete-time Riccati recursion

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0 = \Pi_0.$$

Moreover, $P_i = E\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^*$ is the minimum value of the error covariance matrix at time i, while $R_{e,i} = E\mathbf{e}_i\mathbf{e}_i^*$, where $\mathbf{e}_i = \mathbf{y}_i - H\hat{\mathbf{x}}_i$ is the (optimum) feedback signal. ■

1.2.4 The Kalman Filter

It is an interesting and important fact that the optimum transient observer is actually identical to the celebrated Kalman filter for state estimation. The only reason we have not yet dubbed it as such is that here we started by assuming an estimator of a particular

form, *viz.*, (1.2.7). The question is would we be able to get a better estimator, *i.e.*, one with a smaller mean-square error, by assuming a different structure? The answer is negative, but to establish this we must start without any a priori assumptions, except linearity, about the optimum structure.

And to do this well,[3] we shall have to back off a bit and begin with the more basic problem of finding the linear least-mean-squares estimator of one set of random variables given another such set. For fixed finite sets, this is quite straightforward and is done in Ch. 3 (starting from first principles; see also Prob. 1.3). The more interesting problems in many applications, and the main focus of this book, involve stochastic process estimation, where we have either or both infinite and/or growing sets of indexed random variables. This study is begun in Ch. 4, via the concept of what is called the *innovations process* (see also Sec. 1.4 in this chapter). After that, one could, if desired, proceed directly to a straightforward derivation of the Kalman filter, which we choose to delay to Ch. 9.

Our reasons for taking a less hurried route are several, especially the fact that a proper study of several other problems is thereby greatly aided. One such problem is that of finding the so-called smoothed estimators, which use both past and future data. Another is the following. A major attraction of the Kalman filter formulation is that, unlike the earlier Wiener filtering theory (to be studied here in Chs. 7 and 8), it starts with (the often physically available) state-space models for the process $\{\mathbf{y}_i\}$, rather than with the (often less directly available) covariance functions (or, in steady-state, power spectral densities) of the observed process $\{\mathbf{y}_i\}$. However, it turns out that to get a fuller understanding of the properties of the Kalman filter it is necessary to bring in the covariance/spectral descriptions; indications of this will be seen in Secs. 1.4 and 1.5 below.

Why the Name "Filter" for the Algorithms?. Wiener studied continuous-time problems and noted that his algorithm could be implemented by a linear circuit. In this sense, his solution extended the classical methods of circuit theory for designing "filters" for separating signals in different frequency ranges, as in telephony and in radios, to problems of "filtering signals out of noisy measurements." In fact, in Sec. 7.1.5 we shall compare the performance of the classical so-called "linear distortionless filter" (unit gain and linear phase over the passband) with the (optimum) Wiener filter for filtering out bandlimited signals from additive white noise. Kalman went beyond Wiener by assuming knowledge of a state-space model for the "signal process" $\{\mathbf{x}_i\}$ and his solution could then be readily extended to time-variant models. It could be naturally depicted in "block diagram" form (see Fig. 1.1), and so the early researchers continued to use Wiener's term "filter" for the new solution.

1.3 COMING ATTRACTIONS

In the remainder of this chapter, we shall briefly describe several of the major issues and results that we shall be exploring in later chapters.

[3] With hindsight and considerable calculation, we can also prove the optimality at this point — see Prob. 1.5. However, a more fundamental approach will be developed starting with Ch. 3.

1.3.1 Smoothed Estimators

The most common so-called smoothing problem is that of determining the linear least-mean-squares estimator $\hat{\mathbf{x}}_{i|N}$ of $\mathbf{x}_i$ given $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}, \mathbf{y}_i, \mathbf{y}_{i+1}, \ldots, \mathbf{y}_N\}$. No a priori reasonable structure is readily available for this problem, unlike the one of Fig. 1.1 for the predicted estimator. So in fact it took a while for solutions to appear, first in the Ph.D. dissertation of Rauch (1962). However, using the innovations process just mentioned in Sec. 1.2.4 (see also Sec. 1.4), we shall see quite readily that the smoothed estimators are completely determined by knowledge of the predicted estimators. Among the many possible smoothing formulas derived in Ch. 10, here we mention only the following:

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i + P_i \boldsymbol{\lambda}_{i|N}, \tag{1.3.1}$$

where the $\{\boldsymbol{\lambda}_{i|N}\}$ are computed via the backwards recursion

$$\boldsymbol{\lambda}_{i|N} = [F - K_{p,i}H]^* \boldsymbol{\lambda}_{i+1|N} + H^* R_{e,i}^{-1}[\mathbf{y}_i - H\hat{\mathbf{x}}_i], \qquad \boldsymbol{\lambda}_{N+1|N} = 0. \tag{1.3.2}$$

So after a forwards pass over the data to compute the predicted estimators $\{\hat{\mathbf{x}}_i, i = 0, 1, \ldots, N\}$, along with the error covariance matrices $\{P_i\}$, we perform a backwards pass to compute the $\{\boldsymbol{\lambda}_{i|N}, i = N, N-1, \ldots, 0\}$; then these are linearly combined as in (1.3.1) to obtain the smoothed estimators $\{\hat{\mathbf{x}}_{i|N}, i = 0, 1, \ldots, N\}$. We mention again that there are several variants of this formula, including one that only uses the $\{\mathbf{x}_i\}$ with no further need for the observations $\{\mathbf{y}_i\}$.

1.3.2 Extensions to Time-Variant Models

A notable feature of the results so far is that they can be carried over with only notational changes to time-variant system models:

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i(u_i + \mathbf{u}_i), \\ \mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{1.3.3}$$

with

$$E \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_j^* & \mathbf{v}_j^* & \mathbf{x}_0^* & 1 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} Q_i & S_i \\ S_i^* & R_i \end{bmatrix} \delta_{ij} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \begin{matrix} \Pi_0 & 0 \end{matrix} \end{bmatrix}, \tag{1.3.4}$$

and $R_i > 0$. The predicted estimators $\{\hat{\mathbf{x}}_i\}$ can be computed exactly as in Thm. 1.2.1, with the time-invariant parameters replaced by $\{F_i, G_i, H_i, Q_i, R_i, S_i\}$. [Of course, the same remark applies, perforce, to the calculation of the smoothed estimators.] So we have the striking fact that the extension to time-variant systems can be made just by the stroke of a pen, an illustration of the power of working with state-space models. On the other hand, some reflection (and some hindsight) will suggest that this might also conceal a weakness. In fact, while we would expect that some computational simplifications should be possible when the model is time-invariant, the Kalman filter cannot exploit this fact, as we shall explain.

The main computational burden in the Kalman filter is in updating the $n \times n$ matrices P_i via the Riccati recursion (1.2.12). Here the most expensive operation, when $n \geq \max\{m, p\}$ (the usual case), is the formation of the triple products FP_iF^*, or $F_iP_iF_i^*$ in the time-variant case, both of which take $O(n^3)$ elementary additions and multiplications or floating point operations (flops). In other words, the computational burden is the same for time-invariant or time-variant systems, though, of course, the storage requirements increase in the latter case because we have to store the whole sequence of matrices $\{F_i, G_i, H_i, Q_i, R_i, S_i\}$.

However, there is an alternative to the Riccati-based method used in the Kalman filter for computing the gains $\{K_{p,i}\}$ in the estimator recursion (1.2.17) that does involve fewer computations when the model is time-invariant.

1.3.3 Fast Algorithms for Time-Invariant Systems

For time-invariant systems $\{F, G, H, R, S, Q\}$, it will be shown in Ch. 11 that we can proceed as follows. First define $\delta P_0 = P_1 - P_0$, *i.e.*,

$$\delta P_0 \triangleq F\Pi_0 F^* + GQG^* - (F\Pi_0 H^* + GS)(R + H\Pi_0 H^*)^{-1}(F\Pi_0 H^* + GS)^* - \Pi_0,$$

and determine its rank, say $\alpha = \text{rank}(\delta P_0)$. Then factor δP_0 (nonuniquely) as $\delta P_0 = L_0 M_0 L_0^*$, where L_0 and M_0 are full rank $n \times \alpha$ and $\alpha \times \alpha$ matrices, respectively. Finally, define

$$K_i \triangleq FP_iH^* + GS \qquad \text{so that} \quad K_{p,i} \triangleq K_i R_{e,i}^{-1}.$$

Now we can propagate $\{K_i, R_{e,i}\}$ via recursions involving certain auxiliary sequences $\{L_i, R_{r,i}\}$:

$$K_{i+1} = K_i - FL_iR_{r,i}^{-1}L_i^*H^*, \qquad L_{i+1} = (F - K_iR_{e,i}^{-1}H)L_i,$$

$$R_{e,i+1} = R_{e,i} - HL_iR_{r,i}^{-1}L_i^*H^*, \qquad R_{r,i+1} = R_{r,i} - L_i^*H^*R_{e,i}^{-1}HL_i,$$

with initial conditions

$$K_0 = F\Pi_0 H^* + GS, \quad R_{e,0} = R + H\Pi_0 H^*, \quad R_{r,0} = -M_0^{-1}.$$

The $n \times n$ error variance matrices $\{P_i\}$ do not enter these recursions, but they can be computed, when desired, as

$$P_{i+1} = \Pi_0 - \sum_{j=0}^{i} L_j R_{r,j}^{-1} L_j^*.$$

Now the main computational burden is in forming the products FL_i, which requires only $O(n^2\alpha)$ flops, compared to $O(n^3)$ for the Riccati-based algorithm.

These so-called Chandrasekhar-Kailath-Morf-Sidhu (**CKMS**) recursions are of special interest when the parameter α is significantly smaller than n, which happens in several important cases. For example, when $\Pi_0 = 0$, $\delta P_0 = Q - SR^{-1}S^*$, and $\alpha \leq m$, the number of inputs (which is often much less than n, the number of states). Another important special case is when the processes $\{\mathbf{x}_i, \mathbf{y}_i\}$ are stationary, in which case $\alpha \leq p$, the number of outputs — see Ch. 11. Moreover, the results have important connections to the theory of structured matrices with *displacement structure*, and to a host of related problems (see App. F and also Kailath and Sayed (1995,1999)).

1.3.4 Numerical Issues

Returning to the Kalman filter, we should note that more is at stake than the amount of computation: round-off errors arising from finite-precision requirements can cause several problems.

One consequence of round-off error is that the computed P_i may be non-Hermitian.[4] This is sometimes compensated for by averaging the computed P_i and its Hermitian transpose. A better solution is to propagate only half the elements in P_i — say the ones on and below the main diagonal.

A more serious consequence arises from the fact that the P_i, being covariance matrices, have to be nonnegative-definite. But round-off errors in the computation might destroy this property. Moreover, this is not an easy property to check — a matrix may be indefinite even if all its diagonal entries are nonnegative. The diagonal entries are the mean-square errors in the estimators of each of the components of the state vector and, of course, the computation would be seriously off if these diagonal entries turned out to be negative.

Nevertheless, it has been observed that such situations need not always be catastrophic — it can happen that the computation recovers, and that some iterations later the P_i are nonnegative-definite; see the discussion in Sec. 1.5 below.

Despite these possibilities, it is desirable to try to ensure that P_i is always nonnegative-definite. It turns out that an important step in this direction is to propagate not P_i but a square-root factor, *i.e.*, a matrix A_i such that $P_i = A_i A_i^*$. There will be of course round-off errors in propagating A_i, just as for P_i, but the point is that the product of the computed factors, say $\hat{P}_i = \hat{A}_i \hat{A}_i^*$, is almost certainly nonnegative-definite. In theory, $\hat{A}_i \hat{A}_i^*$ always is nonnegative-definite, but of course again round-off effects may arise — however, they are much easier to control, and in fact, it is easy to see that the diagonal elements will never be negative. There are several forms of square-root algorithms, but the most popular have a so-called array form, explained next.

1.3.5 Array Algorithms

As noted above, a matrix A such that $P = AA^*$ is called a square-root factor of P. Such factors are not unique, since $A\Theta$, where $\Theta\Theta^* = I$, is clearly also a square-root factor. We can choose Θ to make the factor unique, *e.g.*, by making $A\Theta$ Hermitian, or (as we shall prefer) by making it lower triangular with positive diagonal elements. For notational convenience, we shall denote a square-root factor of a matrix P by $P^{1/2}$ and almost always understand it as the unique lower triangular square-root factor. We shall also write

$$P = \left(P^{1/2}\right)\left(P^{1/2}\right)^* = P^{1/2}P^{*/2},$$

and

$$P^{-1} = \left(P^{*/2}\right)^{-1}\left(P^{1/2}\right)^{-1} = P^{-*/2}P^{-1/2}.$$

[4] A square matrix X is said to be Hermitian if $X = X^*$.

With these notations, we can use the following algorithm for propagating the square-root factor $P_i^{1/2}$ in the special case $S = 0$. Form the pre-array

$$\mathcal{A}_1 = \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix},$$

and triangularize it by post-multiplication by any unitary matrix[5] Θ to obtain a post-array

$$\mathcal{A}_2 = \begin{bmatrix} X & 0 & 0 \\ Y & Z & 0 \end{bmatrix}.$$

Then it turns out that we can identify

$$Z = P_{i+1}^{1/2}, \quad X = R_{e,i}^{1/2}, \quad Y = F_i P_i H_i^* R_{e,i}^{-*/2} \triangleq \bar{K}_{p,i} \; ! \tag{1.3.5}$$

This striking (and perhaps surprising) claim can in fact readily be verified by noting that

$$\mathcal{A}_2\mathcal{A}_2^* = \mathcal{A}_1 \underbrace{\Theta\Theta^*}_{I} \mathcal{A}_1^* = \mathcal{A}_1\mathcal{A}_1^*,$$

which translates into

$$\begin{bmatrix} XX^* & XY^* \\ YX^* & YY^* + ZZ^* \end{bmatrix} = \begin{bmatrix} R_i + H_i P_i H_i^* & H_i P_i F_i^* \\ F_i P_i H_i^* & F_i P_i F_i^* + G_i Q_i G_i^* \end{bmatrix}.$$

Therefore, $XX^* = R_{e,i}$, so that X can be identified as $R_{e,i}^{1/2}$. Next $YX^* = F_i P_i H_i^*$ or $Y = F_i P_i H_i^* R_{e,i}^{-*/2}$. Finally

$$ZZ^* = F_i P_i F_i^* + G_i Q_i G_i^* - YY^* = P_{i+1}.$$

Hence, $Z = P_{i+1}^{1/2}$.

There are many ways of obtaining Θ — it is usually best to determine it as a succession of easily specified elementary unitary matrices. This is discussed in detail in Ch. 12. Here we only remark that no explicit equations are involved in the above algorithm — we just set up a certain pre-array of numbers, unitarily triangularize it, and read off the desired quantities from the post-array.

Such "array algorithms" have several other nice features, especially the fact that many algorithms involving complicated sets of equations can be replaced by much simpler arrays (see, *e.g.*, Sayed and Kailath (1994b)). Moreover, not surprisingly fast array algorithms can also be obtained for time-invariant systems (see Ch. 13).

1.3.6 Other Topics

Besides those mentioned so far, several other topics are also considered in this book.

The geometric point of view that we use in this book leads us to the study of linear vector space duality, which is shown in Ch. 15 to provide both new insights and new results. We mention especially the concept of complementary state-space models

[5] A square matrix Θ is unitary if, and only if, $\Theta\Theta^* = \Theta^*\Theta = I$.

and applications to the solution of several deterministic and stochastic quadratic control problems. Several of the results in the book are extended to continuous-time state-space models in Ch. 16.

Finally, a very different approach to the whole subject is described in Ch. 17, where a scattering/generalized transmission line/physical picture is used to study the whole estimation problem, ab initio. It may be mentioned that the Riccati equation/recursion, which is so important in state-space estimation theory, was already encountered in the 1940s in transmission line theory and the closely related radiative transfer theory.

1.4 THE INNOVATIONS PROCESS

It turns out that the error feedback process $\{\mathbf{e}_i = \mathbf{y}_i - H\hat{\mathbf{x}}_i\}$, which in Lemma 1.2.1 we introduced without comment, will be fundamental to our study of estimation problems for several reasons. Chief among these are that $\{\mathbf{e}_i\}$ is a white process and that it is related to the observations process $\{\mathbf{y}_i\}$ through a causal and causally invertible linear transformation. By this we mean that, at every i, the random variable $\mathbf{e}_i$ can be obtained as a linear combination of the observations up to that time instant, *viz.*, $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\}$ and that, in a similar vein, the observation $\mathbf{y}_i$ can be obtained as a linear combination of the $\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_i\}$. Furthermore, we shall see that the $\{\mathbf{e}_i\}$ are independent of the particular state-space model used to describe the process $\{\mathbf{y}_i\}$; that is, the $\{\mathbf{e}_i\}$ are uniquely determined by knowledge of the covariance function of the process $\{\mathbf{y}_i\}$.

> *A deeper consequence of this last remark is that while here we shall demonstrate all the previously mentioned properties by (detailed) calculations using the formulas of Lemma 1.2.1 (or Thm. 1.2.1), obtained by starting with the state-space model (1.2.1)–(1.2.2), the $\{\mathbf{e}_i\}$ can be directly defined in a way that makes the properties mentioned above obvious — see Ch. 4. As we shall see, the process $\{\mathbf{e}_i\}$ can then be used to obtain a very direct derivation of the Kalman filter, and also to solve several other problems, e.g., the smoothing problem of Sec. 1.3.1.*

Such an important process clearly deserves a distinct name and, in fact, there is a very appropriate one: the *innovations process* of $\{\mathbf{y}_i\}$. The reason is that, when $\hat{\mathbf{x}}_i$ is actually the optimum linear least-mean-squares estimator of $\mathbf{x}_i$ given $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$ (as will be studied in Ch. 3), with a similar interpretation of $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{y}}_i$, we can write

$$\mathbf{e}_i \triangleq \mathbf{y}_i - H\hat{\mathbf{x}}_i = \mathbf{y}_i - H\hat{\mathbf{x}}_i - \hat{\mathbf{v}}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i.$$

The last equality will follow by linearity and the one before it by the fact that $\hat{\mathbf{v}}_i = 0$ since $\mathbf{v}_i$ is uncorrelated with $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}, \mathbf{v}_0, \ldots, \mathbf{v}_{i-1}\}$ and therefore with all previous $\{\mathbf{y}_j = H\mathbf{x}_j + \mathbf{v}_j,\ j \le i-1\}$. So $\mathbf{e}_i$ can be regarded as the "new information" or the "innovation" in the observation $\mathbf{y}_i$ after we remove from it all we can say (in the linear least-mean-squares sense) about it from knowledge of past observations — see Ch. 4.

The proofs of the above claims are carried out in a number of subsections, which may be omitted on a first reading.

1.4.1 Whiteness of the Innovations Process

It makes little difference to the basic results if the model is time-invariant or not; so we shall assume the time-variant model (1.3.3)–(1.3.4). The innovations will then be given by

$$\mathbf{e}_i = \mathbf{y}_i - H_i\hat{\mathbf{x}}_i,$$

which it is also useful to rewrite as

$$\mathbf{e}_i = (H_i\mathbf{x}_i + \mathbf{v}_i) - H_i\hat{\mathbf{x}}_i = H_i\tilde{\mathbf{x}}_i + \mathbf{v}_i.$$

Recalling the extension of the estimator equation (1.2.17) to the time-variant case, *viz.*,

$$\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_i + G_iu_i + K_{p,i}(\mathbf{y}_i - H_i\hat{\mathbf{x}}_i), \quad \hat{\mathbf{x}}_0 = 0,$$

we can write a recursion for the error $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$ as

$$\begin{aligned}\tilde{\mathbf{x}}_{i+1} &= F_i\tilde{\mathbf{x}}_i + G_i\mathbf{u}_i - K_{p,i}(H_i\tilde{\mathbf{x}}_i + \mathbf{v}_i),\\ &= F_{p,i}\tilde{\mathbf{x}}_i + G_i\mathbf{u}_i - K_{p,i}\mathbf{v}_i, \quad \tilde{\mathbf{x}}_0 = \mathbf{x}_0,\end{aligned} \tag{1.4.1}$$

where

$$F_{p,i} = F_i - K_{p,i}H_i, \quad K_{p,i} = (F_iP_iH_i^* + G_iS_i)R_{e,i}^{-1}, \quad R_{e,i} = R_i + H_iP_iH_i^*. \tag{1.4.2}$$

The above recursion will be useful in computing the covariance function of the $\{\mathbf{e}_i\}$,

$$\begin{aligned}E\mathbf{e}_i\mathbf{e}_j^* &= E(H_i\tilde{\mathbf{x}}_i + \mathbf{v}_i)(H_j\tilde{\mathbf{x}}_j + \mathbf{v}_j)^*,\\ &= H_i(E\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_j^*)H_j^* + E\mathbf{v}_i\mathbf{v}_j^* + H_i(E\tilde{\mathbf{x}}_i\mathbf{v}_j^*) + (E\mathbf{v}_i\tilde{\mathbf{x}}_j^*)H_j^*.\end{aligned}$$

We shall show first that

$$E\mathbf{e}_i\mathbf{e}_j^* = 0, \quad i > j.$$

For this purpose, we note that by assumption $E\mathbf{v}_i\mathbf{v}_j^* = 0$ and also $E\mathbf{v}_i\tilde{\mathbf{x}}_j^* = 0$, since by (1.4.1) $\tilde{\mathbf{x}}_j$ is a linear combination of the variables $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{j-1}, \mathbf{v}_0, \ldots, \mathbf{v}_{j-1}\}$, each of which is uncorrelated with $\mathbf{v}_i$. To compute the other two terms, we must examine (1.4.1) more closely. A little algebra will show that we can write, for $i > j$,

$$\tilde{\mathbf{x}}_i = \Phi_p(i, j)\tilde{\mathbf{x}}_j + \xi_i, \tag{1.4.3}$$

where ξ_i is a random variable that is a linear combination of the random variables $\{\mathbf{u}_j, \ldots, \mathbf{u}_{i-1}, \mathbf{v}_j, \ldots, \mathbf{v}_{i-1}\}$, and

$$\Phi_p(i, j) = F_{p,i-1}F_{p,i-2}\ldots F_{p,j}, \quad \Phi_p(i, i) = I, \tag{1.4.4}$$

is the state transition matrix from time j to time i. More explicitly,

$$\xi_i = \sum_{k=j}^{i-1}\Phi_p(i, k+1)[G_k\mathbf{u}_k - K_{p,k}\mathbf{v}_k]. \tag{1.4.5}$$

It follows that the variable ξ_i is uncorrelated with $\tilde{\mathbf{x}}_j$ since the latter is a linear combination of the random variables $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{j-1}, \mathbf{v}_0, \ldots, \mathbf{v}_{j-1}\}$, all of which are uncor-

related with the random variables that define ξ_i. Hence, $E\xi_i\tilde{\mathbf{x}}_j^* = 0$. Using this fact, along with $P_j = E\tilde{\mathbf{x}}_j\tilde{\mathbf{x}}_j^*$, we conclude from (1.4.3) that

$$E\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_j^* = \Phi_p(i,j)P_j.$$

Therefore, we can write

$$\begin{aligned}
E\tilde{\mathbf{x}}_i\mathbf{v}_j^* &= E[\Phi_p(i,j)\tilde{\mathbf{x}}_j + \xi_i]\mathbf{v}_j^*, \\
&= \Phi_p(i,j)(E\tilde{\mathbf{x}}_j\mathbf{v}_j^*) + E\xi_i\mathbf{v}_j^*, \\
&= 0 + E\xi_i\mathbf{v}_j^*, \qquad \text{since } E\tilde{\mathbf{x}}_j\mathbf{v}_j^* = 0, \\
&= \Phi_p(i,j+1)E[G_j\mathbf{u}_j - K_{p,j}\mathbf{v}_j]\mathbf{v}_j^*, \qquad \text{by (1.4.5)}, \\
&= \Phi_p(i,j+1)[G_jS_j - K_{p,j}R_j].
\end{aligned}$$

Collecting all these results shows that, for $i > j$,

$$\begin{aligned}
E\mathbf{e}_i\mathbf{e}_j^* &= H_i\Phi_p(i,j)P_jH_j^* + 0 + H_i\Phi_p(i,j+1)[G_jS_j - K_{p,j}R_j] + 0, \\
&= H_i\Phi_p(i,j+1)[F_{p,j}P_jH_j^* + G_jS_j - K_{p,j}R_j], \\
&= H_i\Phi_p(i,j+1)[F_jP_jH_j^* - K_{p,j}H_jP_jH_j^* + G_jS_j - K_{p,j}R_j], \\
&= 0,
\end{aligned}$$

because, by the definition of $K_{p,j}$ in (1.4.2),

$$K_{p,j}[R_j + H_jP_jH_j^*] = F_jP_jH_j^* + G_jS_j. \tag{1.4.6}$$

Similarly, for $i < j$, we simply note that

$$E\mathbf{e}_i\mathbf{e}_j^* = (E\mathbf{e}_j\mathbf{e}_i^*)^* = 0.$$

It remains to evaluate $E\mathbf{e}_i\mathbf{e}_i^*$. But this follows directly from $\mathbf{e}_i = H_i\tilde{\mathbf{x}}_i + \mathbf{v}_i$, which implies that $E\mathbf{e}_i\mathbf{e}_i^* = R_i + H_iP_iH_i^*$. In summary, the above discussion establishes our earlier claim that the process $\{\mathbf{e}_i\}$ is white with

$$E\mathbf{e}_i\mathbf{e}_j^* = (R_i + H_iP_iH_i^*)\delta_{ij} \triangleq R_{e,i}\delta_{ij}.$$

Remark 2. The above calculations show that we can minimize the error variance (1.2.9) by choosing $K_{p,i}$ so as to make the process $\{\mathbf{e}_i = \mathbf{y}_i - H_i\hat{\mathbf{x}}_i\}$, with $\hat{\mathbf{x}}_i$ defined by (1.2.7), a white process. This property suggests a way of "tuning" candidate filters until they are optimum (see, *e.g.*, Mehra (1970)). ◆

1.4.2 Innovations Representations

By rewriting the Kalman filter equations of Thm. 1.2.1 as

$$\begin{cases} \hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + Gu_i + K_{p,i}\mathbf{e}_i, \quad \hat{\mathbf{x}}_0 = 0, \\ \mathbf{y}_i = H\hat{\mathbf{x}}_i + \mathbf{e}_i, \end{cases} \tag{1.4.7}$$

and using the whiteness of the innovations, $E\mathbf{e}_i\mathbf{e}_j^* = R_{e,i}\delta_{ij}$, we see that we have another state-space model for the process $\{\mathbf{y}_i\}$ as the output of a linear system (in state-space form) driven by a white-noise process. Moreover, this model is causally invertible, *i.e.*, we can calculate the $\mathbf{e}_i$ from $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$ as

$$\begin{aligned}
\mathbf{e}_i &= \mathbf{y}_i - H\hat{\mathbf{x}}_i, \\
\hat{\mathbf{x}}_{j+1} &= F\hat{\mathbf{x}}_j + Gu_j + K_{p,j}[\mathbf{y}_j - H\hat{\mathbf{x}}_j], \quad \hat{\mathbf{x}}_0 = 0, \quad j \le i-1.
\end{aligned}$$

That is, we first compute $\{\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_i\}$ using sequentially $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$, and then calculate $\mathbf{e}_i = \mathbf{y}_i - H\hat{\mathbf{x}}_i$. This causal invertibility is in sharp contrast to our initial model (1.2.1)–(1.2.2), where it is clearly impossible to determine the random variables $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}, \mathbf{v}_0, \ldots, \mathbf{v}_{i-1}\}$ from knowledge of the random variables $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$. The representation (1.4.7) is called an *innovations representation* of the process $\{\mathbf{y}_i\}$.

The fact that we have two different models for the process $\{\mathbf{y}_i\}$ (we can actually have many more — see Sec. 9.2.5) leads us to ask what is common to the different models. The only answer will be that both models must of course give processes $\{\mathbf{y}_i\}$ with the same second-order statistics: $\{E\mathbf{y}_i\}$ and $\{E\mathbf{y}_i\mathbf{y}_j^*\}$. This connection is worth a closer look.

1.4.3 Canonical Covariance Factorization

To simplify our discussions in this section, we shall assume that the deterministic inputs $\{u_i\}$ in our state-space model (1.3.3) are identically zero, so that the random variables $\{\mathbf{x}_i, \mathbf{y}_i\}$ in it will be zero-mean.

Now we can compute the covariance matrix of the output process $\{\mathbf{y}_i\}$ by using methods similar to those just used in Sec. 1.4.1 to compute $E\mathbf{e}_i\mathbf{e}_j^*$. Doing this for the model (1.3.3)–(1.3.4) will show that the entries of the covariance matrix

$$R_y \triangleq [E\mathbf{y}_i\mathbf{y}_j^*]_{i,j=0}^N = E\mathbf{y}\mathbf{y}^*, \qquad \mathbf{y} \triangleq \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\},$$

are given by (more details can be found in Sec. 5.3.4):

$$E\mathbf{y}_i\mathbf{y}_j^* = \begin{cases} H_i\Phi(i,j+1)N_j & i > j, \\ R_i + H_i\Pi_iH_i^* & i = j, \\ N_i^*\Phi^*(j,i+1)H_j^* & i < j, \end{cases} \tag{1.4.8}$$

where $N_i = F_i\Pi_iH_i^* + G_iS_i$,

$$\Phi(i,j) \triangleq F_{i-1}F_{i-2}\ldots F_j, \quad i > j, \quad \Phi(i,i) = I,$$

is the (so-called) state transition matrix from time j to time i, and $\Pi_i = E\mathbf{x}_i\mathbf{x}_i^*$ is the state variance matrix. It can be seen to satisfy the recursion

$$\Pi_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^*, \quad i \ge 0,$$

with initial condition Π_0.

An alternative expression can be obtained from the innovations model (1.4.7),

$$\begin{cases} \hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + K_{p,i}\mathbf{e}_i, \quad \hat{\mathbf{x}}_0 = 0, \\ \mathbf{y}_i = H\hat{\mathbf{x}}_i + \mathbf{e}_i. \end{cases} \tag{1.4.9}$$

Defining the column vector $\mathbf{e} = \text{col}\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_N\}$, we first note that we can relate $\mathbf{y}$ and $\mathbf{e}$ as $\mathbf{y} = L\mathbf{e}$, where

$$L \triangleq \begin{bmatrix} I & 0 & 0 & \ldots & 0 \\ H_1K_{p,0} & I & 0 & \ldots & 0 \\ H_2\Phi(2,1)K_{p,0} & H_2K_{p,1} & I & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_N\Phi(N,1)K_{p,0} & H_N\Phi(N,2)K_{p,1} & H_N\Phi(N,3)K_{p,2} & \ldots & I \end{bmatrix}. \tag{1.4.10}$$

The causality of the mapping from $\{\mathbf{e}_i\}$ to $\{\mathbf{y}_i\}$ is reflected in the fact that L is lower triangular; the unit diagonal implies that L^{-1} exists and is also lower triangular with unit diagonal entries, so that the mapping from $\{\mathbf{e}_i\}$ to $\{\mathbf{y}_i\}$ is causally invertible.

Now from (1.4.10) we obtain

$$R_y = E\mathbf{y}\mathbf{y}^* = L[E\mathbf{e}\mathbf{e}^*]L^* = LR_eL^*, \quad \text{say,} \tag{1.4.11}$$

where by the whiteness of the innovations,

$$R_e \triangleq \text{diag}\{R_{e,0}, R_{e,1}, \ldots, R_{e,N}\}. \tag{1.4.12}$$

It is also clear from (1.4.11) that the positive-definiteness of the $\{R_{e,i}\}$, which follows from the assumed positive-definiteness of the $\{R_i\}$ and from the fact that the $\{P_i\}$ are nonnegative-definite, guarantees the positive-definiteness of R_y.

Therefore, (1.4.11) gives us a so-called *canonical* triangular (or lower-diagonal-upper) factorization of the positive-definite covariance matrix. It is wellknown and easily proved (see App. A) that such factorizations are *unique*, which proves the uniqueness of the innovations process, a fact also evident from the construction in Sec. 1.4.2. In other words, while there can be many state-space models giving rise to a process $\{\mathbf{y}_i\}_{i=0}^N$ with a given covariance matrix R_y, all of them will have the same associated innovations process. [One can show that there is no real loss of generality in requiring all models to have the same $\{F_i, H_i\}$ matrices, so that the variety arises from different choices of $\{G_i, Q_i, S_i, \Pi_0\}$. For more on the range of such choices, see the discussion (of the so-called positive-real property) in Secs. 8.2 and 8.8.]

1.4.4 Exploiting State-Space Structure for Matrix Problems

The factorization result we have just obtained is also an example of an application of state-estimation theory to matrix algebra. The triangular factorization of an $N \times N$ positive-definite matrix takes $O(N^3)$ flops, but when we have a state-space structure, the effort goes down to $O(Nn^3)$ flops. Moreover, when the state-space model has time-invariant coefficients, evaluation of the $\{K_{p,i}\}$ by the fast CKMS algorithm will reduce the effort to $O(Nn^2)$ flops.

We note also that the inverse, R_y^{-1}, can also be found via the formula

$$R_y^{-1} = L^{-*}R_e^{-1}L^{-1},$$

where L^{-1} can be found as the impulse response matrix of the system mapping $\{\mathbf{y}_i\}$ to $\{\mathbf{e}_i\}$,

$$\hat{\mathbf{x}}_{i+1} = F_{p,i}\hat{\mathbf{x}}_i + K_{p,i}\mathbf{y}_i, \qquad \mathbf{e}_i = \mathbf{y}_i - H_i\hat{\mathbf{x}}_i.$$

More explicitly,

$$L^{-1} = \begin{bmatrix} I & 0 & 0 & \ldots & 0 \\ -H_1K_{p,0} & I & 0 & \ldots & 0 \\ -H_2\Phi_p(2,1)K_{p,0} & -H_2K_{p,1} & I & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ -H_N\Phi_p(N,1)K_{p,0} & -H_N\Phi_p(N,2)K_{p,1} & -H_N\Phi_p(N,3)K_{p,2} & \ldots & I \end{bmatrix}$$

where $\Phi_p(i,j)$ is given by (1.4.4).

These results, which follow so directly from the stochastic point of view, are not so easy to derive in a purely algebraic way (though this can be done, especially with hindsight — see Apps. 9.A and 13.A; the literature of the 1950s contains many complicated attempts in this direction, with clarity coming only with the explicit introduction and use of state-space structure.)

1.5 STEADY-STATE BEHAVIOR

We return now to time-invariant models as in Sec. 1.2, and examine the behavior, as $i \longrightarrow \infty$, of the optimum transient observer,

$$\begin{aligned} \hat{\mathbf{x}}_{i+1} &= F\hat{\mathbf{x}}_i + K_{p,i}(\mathbf{y}_i - H\hat{\mathbf{x}}_i), \quad \hat{\mathbf{x}}_0 = 0, \\ K_{p,i} &= (FP_iH^* + GS)R_{e,i}^{-1}, \quad R_{e,i} = R + HP_iH^*, \\ P_{i+1} &= FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0 = \Pi_0, \end{aligned} \tag{1.5.1}$$

where we continue to assume that the deterministic inputs $\{u_i\}$ in the state-space model (1.2.1) are identically zero so that, as in the previous section, the random variables $\{\mathbf{x}_i, \mathbf{y}_i\}$ are zero-mean.

Now we would expect that the variable P_i tends to a constant value, P, and hence of course, also $R_{e,i} \longrightarrow R_e$ and $K_{p,i} \longrightarrow K_p$, say. It is also natural to expect that the limiting value P will satisfy the equation

$$P = FPF^* + GQG^* - K_pR_eK_p^*, \tag{1.5.2}$$

with

$$K_p = (FPH^* + GS)R_e^{-1}, \qquad R_e = R + HPH^*. \tag{1.5.3}$$

Equation (1.5.2) is called a Discrete Algebraic Riccati Equation (or DARE). The corresponding steady-state observer equation is

$$\begin{aligned}\hat{\mathbf{x}}_{i+1} &= F\hat{\mathbf{x}}_i + K_p(\mathbf{y}_i - H\hat{\mathbf{x}}_i),\\ &= (F - K_pH)\hat{\mathbf{x}}_i + K_p\mathbf{y}_i, \quad \hat{\mathbf{x}}_0 = 0. \end{aligned} \tag{1.5.4}$$

All this is true, but only under certain, sometimes unexpected, conditions, as we shall describe in detail in Ch. 14. The analysis is undoubtedly the most complex in the book, and in the whole subject of linear state estimation. In this introductory chapter, we shall only indicate some of the issues that arise, how they are resolved, and why they are important.

1.5.1 Appropriate Solutions of the DARE

For example, it is natural to ask if the steady-state value P can be found by directly solving the DARE (1.5.2). This is a nontrivial question because (1.5.2) is a highly nonlinear equation and even in the scalar case ($n = 1$), solutions may or may not exist, and even if they do, they may not be unique.

Thus consider the simple case: $F = 1/2$, $G = H = Q = R = 1$, $S = 0$, for which the DARE is quadratic:

$$P = \frac{1}{4}P + 1 - \frac{P^2}{4(1+P)},$$

or

$$4P^2 - P - 4 = 0.$$

This equation has two solutions, $(1 \pm \sqrt{65})/8$. Which one should we choose? Fortunately, here one solution is negative and can be rejected because the error variances, P_i, are nonnegative and therefore so also their limiting value. Similarly, one might hope that in the general matrix case ($n > 1$), there would be only one nonnegative definite Hermitian solution. Unfortunately, this is only true under additional assumptions on the state-space model, and moreover it is not the only issue in choosing among the solutions.

Our discussion of the observer in Secs. 1.1 and 1.2 indicates that a more critical property than nonnegativity is that P must also be such that the closed-loop matrix, $F - K_pH$ is stable, *i.e.*, have all its eigenvalues of magnitude strictly less than unity; otherwise the variance of the error $\tilde{\mathbf{x}}_i$ would grow without bound. It turns out that, under reasonable assumptions, while there can be several nonnegative-definite solutions P, there can only be one stabilizing solution, *provided it exists*. One frequently encountered set of assumptions that ensures all this is that

(i) F is stable, and

(ii) $R > 0$, and when $S \neq 0$, that

(iii) the power spectral density function of the (stationary) process $\{\mathbf{y}_i\}$ is positive-definite on the unit circle. [When $S = 0$, the assumption (ii) suffices to ensure (iii).][6]

[6] It will be shown in Ch. 6 that this assumption is critical in ensuring that the steady-state innovations process is well defined. An equivalent characterization is that the pair of matrices $\{F - GSR^{-1}H, G(Q - SR^{-1}S)^{1/2}\}$ is unit-circle controllable, as discussed in App. D and in Lemma 8.3.1.

This result is a special case of a more general result established in App. E (Thms. E.5.1 and E.6.1 — see also Lemma 8.3.1). Here we only remark that once again (*cf.* Sec. 1.4.3), we have to go to the second-order statistics (power spectra and covariance functions) to understand the behavior of the state estimators.

So far we have only addressed the issue of appropriate solutions of the DARE without even discussing how they may be found if they exist; some methods for doing this are also described in App. E. Of course, one way of avoiding solution of the DARE is to hope that the clearly unique matrices P_i defined by the Riccati recursion (1.5.1) will converge to the unique stabilizing solution of the DARE (1.5.2), which suggests attention to fast ways of carrying out the recursion. Such so-called "doubling" algorithms do exist, though we only introduce them in Ch. 9 (Prob. 9.25) and Ch. 17 (Secs. 17.5 and 17.6.6). [They could have been introduced earlier, but the authors had many such choices to make.]

1.5.2 Wiener Filters

When F is stable,[7] and $i \longrightarrow \infty$, the state process $\{\mathbf{x}_i\}$ and the observation process $\{\mathbf{y}_i\}$ become zero-mean stationary random processes. Since the origin of time is irrelevant in studying such processes, we can replace the interval $(0, \infty)$ with $(-\infty, i)$ and regard the estimation problem as being one of finding an estimator, $\hat{\mathbf{x}}_i$, for $\mathbf{x}_i$ given all past observations, $\{\mathbf{y}_j,\ -\infty < j \leq i-1\}$ that minimizes $E\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^*$ (more details will be provided in Chs. 3, 8, and 9).

Such stationary stochastic process estimation problems were first studied from an engineering perspective by N. Wiener in 1941–42, who was happy to find that the (analogous continuous-time) estimator was determined by an integral equation he had shown how to solve in 1931! This was the famous Wiener-Hopf equation, which had been introduced in radiative transfer theory in the early 1900s; the name was bestowed after Wiener and Hopf (1931) gave a beautiful closed-form solution, long thought to have been impossible to obtain, based on a so-called *canonical factorization* of the *power spectrum* of the observed stationary stochastic process, $\{\mathbf{y}_i\}$. Such factorizations were relatively easy to obtain for scalar-valued processes with rational z-spectra (see Ch. 6), but there was no computationally satisfactory solution to problems with vector-valued observations, or to problems (as studied in Sec. 1.2) with only a finite but growing set of observations. After nearly a decade of effort (Wiener's 1942 wartime report was only declassified and published in 1949),[8] Kalman (1960a) showed a way through these difficulties by introducing state-space models for the processes involved, rather than by specifying their power spectra (or in the nonstationary case, their covariance functions). However, while a description via a state-space model is often more directly available than power spectra or covariance data, especially in the space guidance and navigation

[7] This subsection calls on some background in linear system theory (reviewed in Ch. 6) and on Wiener theory, studied in Chs. 7 and 8, and may be omitted on a first reading. It is introduced here to provide some perspective on the relation of the Kalman filter theory to the older Wiener filter theory.

[8] Wiener's 1949 book was reissued in paperback form under the title *Time Series Analysis* (MIT Press); it is well worth reading for its perspectives on many topics.

problems that appeared in the 1960s (see a fascinating review article by McGee and Schmidt (1985), who first seized upon, further developed, and implemented the Kalman filter), the difference in the Kalman and Wiener formulations is not as significant as it may at first appear. We have already noted that properties of the power spectra are important in understanding some aspects of the Kalman filter, and more examples will appear soon. The major point is that finding the Kalman filter is equivalent to finding the canonical spectral factorization that is at the heart of the Wiener solution; in the finite interval case, it is equivalent to finding the canonical triangular factorization of the covariance matrix R_y — see Sec. 1.4.3.

To see this, we first note that when F is stable and we are in steady state, the so-called z-spectrum of the process $\{\mathbf{y}_i\}$ described by the model

$$\begin{cases} \mathbf{x}_{i+1} = F\mathbf{x}_i + G\mathbf{u}_i, \\ \mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{1.5.5}$$

with

$$E\begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix}\begin{bmatrix} \mathbf{u}_j^* & \mathbf{v}_j^* & \mathbf{x}_0^* & 1 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\delta_{ij} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \begin{matrix} \Pi_0 & 0 \end{matrix} \end{bmatrix}, \tag{1.5.6}$$

is (*cf.* Ch. 6)

$$S_y(z) \triangleq \text{the } z\text{-transform of the covariance sequence } R_y(i) = E\mathbf{y}_j\mathbf{y}_{j-i}^*,$$

$$= \begin{bmatrix} H(zI - F)^{-1} & G \end{bmatrix}\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ G^* \end{bmatrix}.$$

Now the Wiener solution (Ch. 7) relies on the canonical spectral factorization

$$S_y(z) = L(z)R_eL^*(z^{-*}), \tag{1.5.7}$$

where R_e is the steady-state innovations variance and $L(z)$ is a so-called minimum-phase function, characterized by having all its poles and zeros strictly inside the unit circle. Then under the three assumptions (i)–(iii) mentioned in Sec. 1.5.1, it can be shown (see Ch. 8) that we can find $L(z)$ as[9]

$$L(z) = I + H(zI - F)^{-1}K_p, \tag{1.5.8}$$

where

$$K_p = (FPH^* + GS)R_e^{-1}, \qquad R_e = R + HPH^*, \tag{1.5.9}$$

and P is the unique stabilizing solution of the DARE (1.5.2). [We may remark that $S_y(z)$ and $L(z)$ are compact ways of representing the now-infinite matrices R_y and L in the triangular factorization $R_y = LDL^*$.]

[9] For time-invariant systems, the matrix L in (1.4.10) can be recognized as (part of) the impulse response matrix associated with the transfer function $L(z)$.

But (1.5.8) is exactly the transfer function of the steady-state innovations representation obtained by rewriting the *steady-state Kalman filter* as

$$\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + K_p\mathbf{e}_i, \qquad \mathbf{y}_i = H\hat{\mathbf{x}}_i + \mathbf{e}_i. \tag{1.5.10}$$

The point is that the Wiener and Kalman approaches can both be pursued to yield the same result, after we incorporate Kalman's crucial and fruitful insight as to introducing state-space descriptions. More on all this in later chapters (Chs. 6 to 8). The same equivalence also carries over to finite time nonstationary problems (see Ch. 9). And in some problems, *e.g.*, the steady-state solution to the smoothing problem discussed in Sec. 1.3.1, the Wiener solution is simpler (*cf.* Secs. 7.1 and 7.3.1).

More knowledgeable readers may complain that the Wiener formulation started with spectral data (in particular, a matrix of rational functions in z and z^{-1}) and not with a state-space model, as we have done here. This is an important point. We shall show in detail in Ch. 8 that the key to satisfactorily solving the Wiener problem for vector-valued processes is to recast the original rational power spectrum information into state-space form. Briefly this means that we should seek to write the (rational) z-spectrum as

$$S_y(z) = R_y(0) + H(zI - F)^{-1}\bar{N} + \bar{N}^*(z^{-1}I - F^*)^{-1}H^*, \tag{1.5.11}$$

where $\{H, F, \bar{N}\}$ are $p \times n$, $n \times n$, $n \times p$, matrices, respectively. Then some straightforward calculation (see Sec. 8.5) will show that the canonical factor can again be found as (it is unique)

$$L(z) = I + H(zI - F)^{-1}K_p,$$

but now K_p can be expressed entirely in terms of the covariance parameters $\{H, F, \bar{N}\}$:

$$K_p = (\bar{N} - F\bar{\Sigma}H^*)R_e^{-1}, \qquad R_e = R_y(0) - H\bar{\Sigma}H^*,$$

where $\bar{\Sigma}$ is the unique nonnegative-definite and stabilizing solution of the (slightly different) algebraic Riccati equation

$$\bar{\Sigma} = F\bar{\Sigma}F^* + K_pR_eK_p^*.$$

Again "stabilizing" means the $\bar{\Sigma}$ that makes $F - K_pH$ stable.

All this will be more evident as we proceed with our studies. Here we introduced these results to say that what is important is not whether we start with a model for the process $\{\mathbf{y}_i\}$ (the Kalman formulation) or with the z-spectrum of $\{\mathbf{y}_i\}$ (the Wiener formulation), but that we use state-space models for $\{\mathbf{y}_i\}$ and/or for $S_y(z)$ — this was (in retrospect) Kalman's key insight. Moreover as we have already noted (Sec. 1.3.2) once we introduce state-space descriptions, extensions to time-variant models (and nonstationary) processes can essentially be made by adding a few more subscripts.

But there is more. State-space models were first (re)introduced by Bellman and Kalman in the late 1950s to solve control problems, which led Kalman to realize that one could also study the steady-state behavior of the filter for models with unstable F matrices, which are a major issue in control. Now in the estimation problem, unstable F means that the processes $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$ will have unbounded variance as $i \longrightarrow \infty$ and the problem may appear to be meaningless. But, as we may guess from the fact that the model and the filter have the same F matrix, what can happen (under certain

further assumptions on the model) is that the estimators $\{\hat{\mathbf{x}}_i\}$ will become unbounded at the same rate as the $\{\mathbf{x}_i\}$, so that the error variances, P_i, can remain bounded. This is a nice result, and the assumptions needed are important properties of linear systems — controllability and observability, and their somewhat weaker counterparts, stabilizability and detectability (see App. D). However, as the above heuristic explanation already suggests, the result is somewhat fragile. Even a small discrepancy between the "true" F matrix and the one used in the filter equations will cause the variances of $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ to grow out at different rates, making the error variance diverge. [For more on this, and why things can work out in the pure control problem, see Secs. 14.1.3–14.1.4.]

1.5.3 Convergence Results

We mentioned before that one way of avoiding direct solution of the DARE is to hope that the clearly unique matrices P_i defined by the Riccati recursion (1.5.1) will converge to the unique stabilizing solution of the DARE (1.5.2). The general results are fairly involved, so we only note some important special cases and special features. A standing assumption will be that

$$R > 0.$$

The simplest result is that when F is stable and $S = 0$, the Riccati variable P_i tends to a constant matrix P, at an exponential rate, for all nonnegative-definite starting values of the recursion (1.5.1), *i.e.*, for all $P_0 \geq 0$. Moreover, this constant value is the unique stabilizing solution of the DARE (1.5.2).

More unexpected is the fact that a nonzero value of S can disturb the convergence even when F is stable. Now, as expected, from our discussion of the DARE (*cf.* Lemma 8.3.1), we have to bring in the assumption that[10]

$$S_y(z) \quad \text{is positive-definite on the unit circle.}$$

Under this additional assumption, we can characterize a rather complicated set of initial conditions for which convergence is guaranteed — see Thm. 14.5.1. But this set does not include *all* nonnegative definite matrices P_0! For that to hold, we have to bring in the stronger assumption that $\{F - GSR^{-1}H, G(Q - SR^{-1}S)^{1/2}\}$ is stabilizable[11] — see Thm. 14.5.3. We shall not pursue the details but do mention the interesting implications on the behavior of the innovations variances $\{R_{e,i}\}$ — they will always be nonsingular but they may fail to be positive-definite for a *finite* number of time instants (see Lemma 14.5.3)!

All the above results can be extended to the case of *unstable* F matrices by adding the assumption that $\{F, H\}$ is *detectable*,[12] which will guarantee that the error variance P_i is bounded — see Prob. 14.4.

[10] An equivalent condition is that $\{F - GSR^{-1}H, G(Q - SR^{-1}S)^{1/2}\}$ must be unit-circle controllable, as discussed in App. D and in Lemma 8.3.1.

[11] There are various characterizations of stabilizability (see App. C). For example, a pair $\{F, G\}$ is said to be stabilizable if, and only, if there exists a constant matrix K such that $F - GK$ is stable.

[12] Again, there are various characterizations of detectability (see App. C). For example, a pair $\{F, H\}$ is said to be detectable if, and only, if there exists a constant matrix K such that $F - KH$ is stable.

Finally, we note the surprising result that convergence can hold also for certain *indefinite*, and even *negative-semi-definite*, initial conditions Π_0 (provided they are bounded below by a certain negative-semi-definite matrix). This unexpected result has some important implications.

The first is numerical. Although no physical initial conditions can be indefinite, starting with $\Pi_0 \geq 0$, it is quite possible that due to numerical effects (say, round-off errors in the computations) the computed P_i, at some instant $i > 0$, may lose its required nonnegative-definite character (it is a variance matrix). The question therefore is will the algorithm then break down and give meaningless results thereafter? Or if we continue the recursion, will P_i eventually recover its nonnegative-definiteness and, more to the point, ultimately converge to the stabilizing solution, P? We will show in Ch. 14 that, as stated above, the answer is yes. This phenomenon has also been observed in practice.

The next issue is to consider why such results are possible. One explanation is that P_i is an "internal" quantity, dependent upon the state-space model we have chosen for the process $\{\mathbf{y}_i\}$. Such internal quantities might be "nonphysical" provided we do not violate the statistical properties of the process $\{\mathbf{y}_i\}$. In the observer equations of Thm. 1.2.1, the "external" quantity influenced by P_i is the innovations variance $R_{e,i} = R + HP_iH^*$, which, as explained in the previous section, is uniquely determined by the covariance matrix R_y of the observed process $\{\mathbf{y}_i\}$ (the $\{R_{e,i}\}$ are the diagonal entries in the unique LDL* factorization of R_y — see (1.4.12)). It is the positive-definiteness of the $\{R_{e,i}\}$ that should not be violated if we are to have a meaningful solution. Now since we assumed $R > 0$, it is clear that we can have $R_{e,i} = R + HP_iH^* > 0$ even if P_i is indefinite! This is in fact what happens, as we shall show in Lemmas 14.5.3 and 14.5.6 of Ch. 14.

In summary, however, the main message of this rapid survey is to indicate that the topic of convergence is a nontrivial one, which is discussed in detail in Ch. 14.

1.6 SEVERAL RELATED PROBLEMS

There are several other problems, in addition to the transient observer design of the earlier sections, that lend themselves rather directly to the methods discussed in this book. Here we briefly outline some of them.

This section can be omitted on a first reading. It is meant to whet the reader's appetite for (some of) the benefits to be gained by a study of least-squares estimation theory.

1.6.1 Adaptive RLS Filtering

Sayed and Kailath (1994b) showed that certain equivalence relations (*i.e.*, connections with the Kalman filter solution) exist that allow us to immediately solve several so-called deterministic recursive least-squares (RLS) problems by reducing them to equivalent state-space estimation problems.

For the sake of illustration, consider a set of $(N+1)$ data $\{h_i, y(i)\}_{i=0}^{N}$, where the $y(i)$ are given scalars (the desired output) and the h_i are given $1 \times M$ row vectors (the given input vectors). Consider also an $M \times M$ positive-definite matrix $\Pi_0 = \delta I$, with $\delta > 0$.

A problem of interest in RLS theory is the recursive estimation of the unknown vector w that minimizes the quadratic criterion:

$$\min_w \left[w^* \Pi_0^{-1} w + \sum_{i=0}^{N} |y(i) - h_i w|^2 \right].$$

It will be shown in Ch. 2 that the solution to the above problem can be recursively computed as follows: start with the initial condition $w_{-1} = 0$ and repeat for $i \geq 0$:

$$w_i = w_{i-1} + k_{p,i}[y(i) - h_i w_{i-1}],$$

where

$$k_{p,i} = P_{i-1} h_i^* r_e^{-1}(i), \quad r_e(i) = 1 + h_i P_{i-1} h_i^*,$$

and P_{i-1} satisfies the Riccati recursion

$$P_i = P_{i-1} - P_{i-1} h_i^* (1 + h_i P_{i-1} h_i^*)^{-1} h_i P_{i-1}, \quad P_{-1} = \Pi_0.$$

There is clearly a striking resemblance between the above solution of the RLS problem and the recursions of the optimum transient observer of Thm. 1.2.1. The ramifications of this resemblance can be pursued much further. Once such a connection is established, many other variants of the RLS equations, including array and fast versions, will follow rather transparently. Much of the now vast literature on adaptive filtering (see, *e.g.*, the widely used textbook of Haykin (1996)) can be unified in this way (see Sayed and Kailath (1994b)).

1.6.2 Linear Quadratic Control

Consider the time-variant state-space model

$$x_{i+1} = F_i x_i + G_i u_i, \quad 0 \leq i \leq N, \tag{1.6.1}$$

and some given linear combination of the states, say $s_i = H_i x_i$, where the $F_i \in \mathbb{C}^{n \times n}$, $G_i \in \mathbb{C}^{n \times m}$, and $H_i \in \mathbb{C}^{p \times n}$ are known matrices. The initial condition x_0 of (1.6.1) is known and we would like to choose a *control* signal $\{u_i\}$ so as to regulate the $\{s_i\}$ in a certain sense (*e.g.*, to keep it close to a zero trajectory). The conventional LQR (linear quadratic regulator) method proposes the following criterion for choosing the $\{u_i\}$:

$$\min_{\{u_i\}} \left[x_{N+1}^* P_{N+1}^c x_{N+1} + \sum_{i=0}^{N} u_i^* Q_i^c u_i + \sum_{i=0}^{N} s_i^* R_i^c s_i \right], \tag{1.6.2}$$

subject to the state-space constraint (1.6.1), where $P_{N+1}^c \geq 0$, $Q_i^c > 0$, and $R_i^c \geq 0$ are given matrices that penalize the final state, the inputs, and the intermediary states, respectively.

The solution (*i.e.*, the desired control signal) was shown, essentially first by Kalman and Koepcke (1958), to be given by the following state feedback law:

$$u_i = -K_i^{c*} x_i,$$

where the feedback gain matrix K_i^c is given by

$$K_i^{c*} = (Q_i^c + G_i^* P_{i+1}^c G_i)^{-1} G_i^* P_{i+1}^c F_i,$$

and where P_i^c satisfies the backwards-time Riccati recursion

$$P_i^c = F_i^* P_{i+1}^c F_i - F_i^* P_{i+1}^c G_i (Q_i^c + G_i^* P_{i+1}^c G_i)^{-1} G_i^* P_{i+1}^c F_i + H_i^* R_i^c H_i,$$

with the value at time $N+1$ equal to the given matrix P_{N+1}^c. Comparing with the statement of the Kalman filter in Thm. 1.2.1 (in the time-variant case), we see that if we replace the variables $\{F_i^*, H_i^*, G_i^*, R_i^c, Q_i^c\}$ by $\{F_i, G_i, H_i, Q_i, R_i\}$,[13] and if we reverse the time order of the LQR Riccati recursion (so as to propagate forwards in time, starting at time 0 with some initial covariance matrix Π_0), then the LQR recursion for P_i^c reduces to the Kalman recursion for P_i (assuming the cross-covariance matrix S_i is zero). Moreover, the state feedback matrix, K_i^{c*} becomes the Hermitian transpose of the Kalman gain, $K_{p,i}$.

Thus, it is not surprising that when Kalman (1960a) obtained his state-space estimation formulas a little later, he was able to declare that the control and estimation *solutions* could be regarded as duals of each other! However, rather than solving both problems independently (and by quite different methods), we shall show, in Ch. 15, that, by using the (geometrical) concept of duality, the two *problems* can be recognized as being dual to each other, so that the solution of the control problem can be immediately written down from that of the estimation problem. This duality concept is introduced in Ch. 15, and its application to control problems is studied in Sec. 15.3.5 and also in the monograph (Hassibi, Sayed, and Kailath (1999)).

1.6.3 $\mathcal{H}_\infty$ Estimation

The optimum transient observer of Thm. 1.2.1 requires a priori knowledge of the statistical properties of the random variables $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i\}$. In some applications, however, one is faced with model uncertainties and lack of statistical information on the (exogenous) signals. This had led in recent years to an interest in mini-max estimation, with the belief that the resulting so-called $\mathcal{H}_\infty$ algorithms will be more robust, and less sensitive, to disturbance variations and modeling assumptions.

Now it happens that $\mathcal{H}_\infty$ filters bear a striking resemblance to classical Kalman filters and involve a similar Riccati recursion. Despite these similarities, the now large literature on $\mathcal{H}_\infty$ theory has used techniques very different from those of the stochastic state-space estimation (Kalman filtering) theory; in these approaches, the similarities do not seem to have a simple/natural explanation. However, an extension of the usual stochastic formulation allows a unified approach to the least-mean-squares and the $\mathcal{H}_\infty$ theories, in which the similarities and the differences have natural explanations. The extension involves working with certain indefinite metric (Krein) spaces, rather than standard Hilbert spaces. Although Hilbert spaces and Krein spaces share many characteristics, they differ in special ways that mark the differences between the classical LQR and observer design theories and the more recent $\mathcal{H}_\infty$ theories. This approach is developed in detail in the monograph (Hassibi, Sayed, and Kailath (1999)). Here we shall only introduce the $\mathcal{H}_\infty$ filtering problem and present a solution.

[13] Strictly speaking, we should replace $\{F_i^*, H_i^*, G_i^*\}$ by $\{F_i, -G_i, H_i\}$, as will become clear later in our studies of dual models in Sec. 15.2.3.

Consider the time-variant state-space model

$$\begin{cases} x_{i+1} &= F_i x_i + G_i u_i, \\ y_i &= H_i x_i + v_i, \end{cases} \qquad i \geq 0, \tag{1.6.3}$$

where $F_i \in \mathbb{C}^{n\times n}$, $G_i \in \mathbb{C}^{n\times q}$, and $H_i \in \mathbb{C}^{p\times n}$ are known matrices, x_0, $\{u_i\}$, and $\{v_i\}$ are *unknown* quantities and y_i is the measured output. We can regard v_i as a measurement noise and u_i as a process noise or driving disturbance. We make no assumption on the nature of the disturbances (such as uncorrelated, normally distributed, etc).[14] Suppose we want to estimate some arbitrary linear combination of the states, say

$$s_i = L_i x_i, \quad \text{where } L_i \in \mathbb{C}^{m\times n} \text{ is known,} \tag{1.6.4}$$

using the observations $\{y_0, \ldots y_i\}$. Let $\hat{s}_{i|i} = \mathcal{F}_f(y_0, \ldots, y_i)$ denote an estimate of s_i given observations $\{y_j\}$ from time 0 up to and including time i, and let

$$e_{f,i} = \tilde{s}_{i|i} = s_i - \hat{s}_{i|i},$$

be the corresponding so-called *filtered* error. Note that since $\hat{z}_{i|i}$ is a *causal* function of the $\{y_j\}$ we shall call such an estimate the *a posteriori* or *filtered* estimate of s_i; a priori or predicted estimates use $\{y_0, \ldots, y_{i-1}\}$ to estimate s_i and can be studied in the same way.

The $\mathcal{H}_\infty$ formulation deals with the following problem. Given a scalar $\gamma_f > 0$, a positive-definite matrix Π_0, and an initial guess $\check{x}_0$, find, if possible, an estimator $\hat{s}_{j|j} = \mathcal{F}_f(y_0, \ldots, y_j)$ that achieves for all i

$$\sup_{x_0, u\in h_2, v\in h_2} \frac{\sum_{j=0}^{i} e_{f,j}^* e_{f,j}}{(x_0 - \check{x}_0)^* \Pi_0^{-1} (x_0 - \check{x}_0) + \sum_{j=0}^{i} u_j^* u_j + \sum_{j=0}^{i} v_j^* v_j} < \gamma_f^2. \tag{1.6.5}$$

[Here, h_2 denotes the space of square-summable sequences (*i.e.*, finite-energy sequences), and u and v denote the sequences $\{u_i\}$ and $\{v_i\}$, respectively.]

An interpretation of the above criterion is now in order. Note that the denominator of the ratio in (1.6.5) can be regarded as the energy of the unknown disturbances $\{\Pi_0^{-1/2}(x_0 - \check{x}_0, \{u_j\}, \{v_j\}\}$, and that the numerator can be regarded as the energy of the estimation error sequence $\{e_{f,j}\}$. Thus the ratio in (1.6.5) is just the energy gain from the disturbances to the estimation errors. Of course, a desirable estimator is one for which this energy gain is small (so that the disturbances are attenuated), and a nondesirable estimator is one for which the energy gain is large (since the disturbances are amplified). The $\mathcal{H}_\infty$ framework proposes to choose the estimator such that the worst-case energy gain is bounded by the prescribed value γ_f^2. In other words, $\mathcal{H}_\infty$ estimators yield an energy gain less than γ_f^2 for all (bounded energy) disturbances, no matter what they are. The robustness of $\mathcal{H}_\infty$ estimators arises from this latter fact. Note, moreover, that the smaller γ_f^2 is, the more robust the resulting estimator is. Of course, γ_f^2 cannot be made arbitrarily small and there is a certain value, denoted by $\gamma_{f,opt}^2$, beyond which it cannot be reduced.

[14] In particular, in the $\mathcal{H}_\infty$ framework the disturbances are not assumed to be random processes.

The solution of (1.6.5) can be shown to have the following form. If the matrices $\{F_j \;\; G_j\}$ have full rank, a solution exists (*i.e.*, it is possible to bound the worst-case energy gain by γ_f^2), if, and only if, the following positivity conditions hold:

$$P_j^{-1} + H_j^* H_j - \gamma_f^{-2} L_j^* L_j > 0, \qquad j = 0, \ldots, i, \tag{1.6.6}$$

where $P_0 = \Pi_0$ and P_i satisfies the Riccati recursion

$$P_{i+1} = F_i P_i F_i^* + G_i G_i^* - F_i P_i \begin{bmatrix} H_i^* & L_i^* \end{bmatrix} R_{e,i}^{-1} \begin{bmatrix} H_i \\ L_i \end{bmatrix} P_i F_i^*,$$

with

$$R_{e,i} = \begin{bmatrix} I & 0 \\ 0 & -\gamma_f^2 I \end{bmatrix} + \begin{bmatrix} H_i \\ L_i \end{bmatrix} P_i \begin{bmatrix} H_i^* & L_i^* \end{bmatrix}.$$

If this is the case, then one possible $\mathcal{H}_\infty$ estimator is given by $\hat{s}_{i|i} = L_i \hat{x}_{i|i}$, where

$$\hat{x}_{i+1|i+1} = F_i \hat{x}_{i|i} + K_{s,i}(y_{i+1} - H_{i+1} F_i \hat{x}_{i|i}), \tag{1.6.7}$$

with initial condition $\hat{x}_{-1|-1} = F_{-1}^{-1} \check{x}_0$, and

$$K_{s,i} = P_{i+1} H_{i+1}^* (I + H_{i+1} P_{i+1} H_{i+1}^*)^{-1}. \tag{1.6.8}$$

The point to highlight here is that the above recursions look very much like a Kalman filter solution (compare with Prob. 1.4), except that the Riccati recursion differs from that of the Kalman filter, since

1. We now have indefinite "*covariance*" matrices, $\begin{bmatrix} I & 0 \\ 0 & -\gamma_f^2 I \end{bmatrix}$.
2. The L_i (of the quantity to be estimated) enters the Riccati recursion.
3. We have an additional condition, (1.6.6), that must be satisfied for the filter to exist; in the Kalman filter problem the L_i would not appear, and the P_i would be positive-definite, so that (1.6.6) is immediate.

Despite these differences, we show in the monograph (Hassibi, Sayed, and Kailath (1999)) that the above $\mathcal{H}_\infty$ filter, and in fact other variants as well, can be obtained by developing a partial equivalence to a Kalman filtering problem, not in a Hilbert space, but in a certain indefinite vector space, called a Krein space. The indefinite "covariance" matrices and the appearance of L_i in the Riccati recursion are readily explained in this framework. The additional condition (1.6.6) will be seen to arise from the fact that in Krein space, unlike as in the usual Hilbert space context, quadratic forms need not always have minima or maxima, unless certain additional conditions are met. [A simple problem where these issues also arise will be treated in Sec. 2.7.]

1.6.4 $\mathcal{H}_\infty$ Adaptive Filtering

The $\mathcal{H}_\infty$-estimator (1.6.6)–(1.6.8) provides novel insights into the performance of some widely used adaptive filtering schemes, such as the normalized least-mean squares (NLMS) algorithm and the Gauss-Newton (GN) algorithm. We discuss this issue in more detail in the monograph (Hassibi, Sayed, and Kailath (1999)). Here we only wish to formulate the problems and to show that the NLMS and GN algorithms are indeed special cases of recursion (1.6.7). Similar claims can also be made for the famous Widrow-Hoff LMS algorithm (described in, *e.g.*, Widrow and Hoff (1960), Widrow and Stearns (1985), and Haykin (1996)), but the criterion is slightly different from (1.6.9) below, so a variant of (1.6.11)–(1.6.14) will be needed; space forbids further discussion (see Hassibi, Sayed, and Kailath (1996c,1999) and Sayed and Rupp (1997)).

We start with the special state-space model

$$\begin{cases} x_{i+1} &= x_i, \\ y(i) &= h_i x_i + v(i), \end{cases} \qquad i \geq 0.$$

Comparing with (1.6.3), we see that we have assumed $G_i = 0$, F_i equal to the identity matrix, and H_i a row vector that we indicate by h_i. We may also recognize that the above state-space model corresponds to the adaptive filtering problem of Sec. 1.6.1, where $y(i)$ is the given scalar output, h_i is the known input vector, and $x_i = x_0$, for all $i \geq 0$ (since the state equation is trivial), is the unknown weight vector. The unknown scalar sequence $v(i)$ can be regarded as measurement noise and is used to represent the discrepancies between the desired output $y(i)$ and the ideal model $h_i x_0$. We further assume that we want to estimate this uncorrupted output, which we denote by $s(i) = h_i x_i$. Comparing with (1.6.4), we see that this corresponds to choosing $L_i = h_i$. The resulting $\mathcal{H}_\infty$ problem is therefore to find, if possible, an estimator $\hat{s}(j|j) = \mathcal{F}_f(y(0), \ldots, y(j))$ that achieves for all i

$$\sup_{x_0, v \in h_2} \frac{\sum_{j=0}^{i} |e_f(j)|^2}{(x_0 - \check{x}_0)^* \Pi_0^{-1} (x_0 - \check{x}_0) + \sum_{j=0}^{i} |v(j)|^2} < \gamma_f^2, \tag{1.6.9}$$

where $e_f(j) = s(j) - \hat{s}(j|j)$.

In this case, the $\mathcal{H}_\infty$ estimator of the previous section reduces to the following form:

$$\hat{s}(i|i) = h_i \hat{x}_{i|i}, \tag{1.6.10}$$

$$\hat{x}_{i+1|i+1} = \hat{x}_{i|i} + K_{s,i}[y(i+1) - h_{i+1}\hat{x}_{i|i}], \quad \hat{x}_{-1|-1} = \check{x}_0, \tag{1.6.11}$$

$$K_{s,i} = P_{i+1} h_{i+1}^* (1 + h_{i+1} P_{i+1} h_{i+1}^*)^{-1}, \tag{1.6.12}$$

$$P_{i+1} = P_i - P_i \begin{bmatrix} h_i^* & h_i^* \end{bmatrix} R_{e,i}^{-1} \begin{bmatrix} h_i \\ h_i \end{bmatrix} P_i, \tag{1.6.13}$$

$$R_{e,i} = \begin{bmatrix} 1 & 0 \\ 0 & -\gamma_f^2 \end{bmatrix} + \begin{bmatrix} h_i \\ h_i \end{bmatrix} P_i \begin{bmatrix} h_i^* & h_i^* \end{bmatrix}. \tag{1.6.14}$$

Applying the matrix inversion lemma (given in App. A (vii)) to the Riccati recursion (1.6.13) yields

$$P_{i+1}^{-1} = P_i^{-1} + (1 - \gamma_f^{-2}) h_i^* h_i, \quad P_0^{-1} = \Pi_0^{-1}. \tag{1.6.15}$$

Solving the above recursion for P_j^{-1} yields

$$P_j^{-1} = \Pi_0^{-1} + (1 - \gamma_f^{-2}) \sum_{k=0}^{j-1} h_k^* h_k,$$

from which we conclude that the existence condition (1.6.6) becomes

$$P_j^{-1} + (1 - \gamma_f^{-2}) h_j^* h_j = \Pi_0^{-1} + (1 - \gamma_f^{-2}) \sum_{k=0}^{j} h_k^* h_k > 0, \qquad j = 0, \ldots, i, \tag{1.6.16}$$

which is clearly satisfied for any $\gamma_f \geq 1$, since $\Pi_0 > 0$.

The Normalized Least-Mean-Squares (NLMS) Algorithm. To further study the existence condition (1.6.16) let us assume that the input vectors $\{h_i\}$ are such that

$$\lim_{N \to \infty} \sum_{k=0}^{N} h_k h_k^* = \infty. \tag{1.6.17}$$

Note that this is a mild condition on the input vectors that states that they do not "die out too fast." When (1.6.17) is true, it is easy to see that for *any* $\gamma_f < 1$, and for sufficiently large j, at least one of the diagonal entries of the matrix on the LHS of (1.6.16) must become negative. This violates the existence condition so that we must have $\gamma_{f,opt} = 1$, which is the optimum value of the disturbance attenuation.

It is also interesting to study the structure of the optimal $\mathcal{H}_\infty$ adaptive filter corresponding to $\gamma_{f,opt} = 1$. Setting $\gamma_f = 1$ and $\Pi_0 = \mu I$, a positive multiple of the identity, the recursion (1.6.15) shows that $P_{i+1} = \mu I$ for all i. In this case (1.6.11) becomes:

$$\hat{x}_{i+1|i+1} = \hat{x}_{i|i} + \frac{\mu h_{i+1}^*}{1 + \mu h_{i+1} h_{i+1}^*} [y(i+1) - h_{i+1}\hat{x}_{i|i}], \quad \hat{x}_{-1|-1} = \check{x}_0, \tag{1.6.18}$$

which is the update expression for the so-called NLMS algorithm.

In other words, NLMS is an *optimal* $\mathcal{H}_\infty$ filter (see further Hassibi, Sayed, and Kailath (1996c)). To summarize this important fact, we can say that NLMS satisfies the following min-max optimization problem

$$\inf_{\mathcal{F}_f} \sup_{x_0, v \in h_2} \frac{\sum_{j=0}^{i} |e_f(j)|^2}{(x_0 - \check{x}_0)^* \Pi_0^{-1} (x_0 - \check{x}_0) + \sum_{j=0}^{i} |v(j)|^2}. \tag{1.6.19}$$

The Gauss-Newton (GN) Adaptive Filtering Algorithm. When condition (1.6.17) is not satisfied, the following infinite sum has a finite limit,

$$\lim_{N \to \infty} \sum_{k=0}^{N} h_k^* h_k \triangleq A < \infty I.$$

With this definition of A, the existence condition (1.6.16) for large enough j becomes

$$\frac{1}{\mu} I + (1 - \gamma_f^{-2})A > 0,$$

where we have used $\Pi_0 = \mu I$. For a $\gamma_f < 1$ the above is true if, and only if,

$$\frac{1}{\mu} > (\gamma_f^{-2} - 1)\bar{\sigma}(A),$$

where $\bar{\sigma}(A)$ denotes the maximum singular value of A. This latter inequality implies that a solution exists if, and only if,

$$\gamma_f^2 > \frac{\mu\bar{\sigma}(A)}{1 + \mu\bar{\sigma}(A)} \triangleq \gamma_{f,opt}^2.$$

Now for any $\gamma_f^2 \neq 1$, we can re-apply the matrix inversion lemma to (1.6.15) to obtain

$$P_{i+1} = P_i - \frac{P_i h_i h_i^* P_i}{(1 - \gamma_f^{-2})^{-1} + h_i P_i h_i^*}, \tag{1.6.20}$$

which is an alternative rewriting of the original Riccati recursion (1.6.13). When $\gamma_f^2 > 1$, so that $1 - \gamma_f^{-2} > 0$, expressions (1.6.11)–(1.6.12) and (1.6.20) constitute the a posteriori form of the so-called Gauss-Newton algorithm (see Sayed and Rupp (1997)).

1.6.5 $\mathcal{H}_\infty$ Control

In Sec. 1.6.2 we introduced the problem of linear quadratic control and mentioned that it is dual (in a certain sense) to the problem of least-mean-squares estimation. However, as we show in the monograph (Hassibi, Sayed and Kailath (1999)), this duality between estimation and control is deeper and is not confined to LQR control and least-mean-squares estimation. In fact, if one considers control problems with an $\mathcal{H}_\infty$ criterion, then it can be shown that the problem (and its solution) is again dual to the problem (and solution) of $\mathcal{H}_\infty$ estimation.

We only mention here a simple variant of the $\mathcal{H}_\infty$ control problem, namely the so-called *full information* version. So consider the time-variant state-space model

$$x_{i+1} = F_i x_i + G_{1,i} w_i + G_{2,i} u_i, \quad 0 \leq i \leq N, \tag{1.6.21}$$

where the $F_i \in \mathbb{C}^{n \times n}$, $G_{1,i} \in \mathbb{C}^{n \times m_1}$, and $G_{2,i} \in \mathbb{C}^{n \times m_2}$ are known matrices, $\{w_i\}$ is the exogenous input (or disturbance), and $\{u_i\}$ is the control signal. Assume, for simplicity, that the initial condition x_0 is zero. As mentioned in Sec. 1.6.2, the objective in control is to regulate some linear combination of the states, say $s_i = L_i x_i$, where $L_i \in \mathbb{C}^{p \times n}$ is known. This regulation is achieved by choosing the control signal $\{u_i\}$ which, in the full information problem, is allowed to be a linear causal function of the disturbances $\{w_i\}$, *viz.*,

$$u_i = \mathcal{F}_i(w_0, \ldots, w_i), \tag{1.6.22}$$

for some linear function $\mathcal{F}_i$. In the LQR control problem, the $\{u_i\}$ were chosen to minimize the quadratic cost in (1.6.2). However, because of the appearance of w_i in the state equation (1.6.21), this cost is now also a function of the unknown exogenous input $\{w_i\}$ and so it is not clear how to minimize (or even define minimization) over $\{u_i\}$, subject to the causality constraint (1.6.22).

In $\mathcal{H}_\infty$ control, we instead seek to choose $u_i = \mathcal{F}_i(w_0, \ldots, w_i)$ so as to bound the worst-case energy gain from the disturbances $\{w_i\}$ to the LQR cost by a prescribed value γ^2, *i.e.*, we would like to find $\mathcal{F}_i$ such that

$$\sup_{w \in h_2} \frac{x_{N+1}^* P_{N+1}^c x_{N+1} + \sum_{i=0}^N u_i^* Q_i^c u_i + \sum_{i=0}^N s_i^* R_i^c s_i}{\sum_{i=0}^N w_i^* Q_i^w w_i} < \gamma^2,$$

where $Q_i^w > 0$ is given and w denotes the sequence $\{w_i\}$.

The robustness of $\mathcal{H}_\infty$ controllers follows from the fact that they bound the energy gain by γ^2 for all (bounded energy) disturbances, no matter what they are. Clearly, the smaller γ^2 is, the more robust the resulting controller will be. However, γ^2 cannot be made arbitrarily small and there is a certain value, denoted by γ_{opt}^2, beyond which it cannot be further reduced.

In the monograph (Hassibi, Sayed and Kailath (1999))it is shown that by formulating an appropriate dual $\mathcal{H}_\infty$ filtering problem we obtain the following result. If the matrices $\left\{\begin{bmatrix} F_j^* & L_j^* \end{bmatrix}\right\}$ have full rank, a control solution exists (*i.e.*, it is possible to bound the worst-case energy gain by γ^2) if, and only if, the following positivity conditions hold:

$$(P_j^c)^{-1} + G_{2,j}^* (Q_j^c)^{-1} G_{2,j} - \gamma^{-2} G_{1,j}^* (Q_j^w)^{-1} G_{1,j} > 0, \qquad \text{for} \quad j = 0, \ldots, N,$$

where P_i^c satisfies the backwards Riccati recursion

$$P_i^c = F_i^* P_{i+1}^c F_i + L_i^* R_i^c L_i - F_i^* P_{i+1}^c \begin{bmatrix} G_{2,i} & G_{1,i} \end{bmatrix} (R_{e,i}^c)^{-1} \begin{bmatrix} G_{2,i}^* \\ G_{1,i}^* \end{bmatrix} P_{i+1}^c F_i,$$

with boundary condition P_{N+1}^c and where

$$R_{e,i}^c = \begin{bmatrix} Q_i^c + G_{2,i}^* P_{i+1}^c G_{2,i} & G_{2,i}^* P_{i+1}^c G_{1,i} \\ G_{1,i}^* P_{i+1}^c G_{2,i} & -\gamma^2 Q_i^w + G_{1,i}^* P_{i+1}^c G_{1,i} \end{bmatrix}.$$

If this is the case, then one possible $\mathcal{H}_\infty$ controller is given by the state feedback law

$$u_i = -K_i^{c*} x_{i+1} = -(Q_i^c + G_{2,i}^* P_{i+1}^c G_{2,i})^{-1} G_{2,i}^* P_{i+1}^c x_{i+1}.$$

Comparing with the $\mathcal{H}_\infty$ estimator of Sec. 1.6.3, we see that if we replace the variables $\{F_i^*, L_i^*, G_{2,i}^*, G_{1,i}^*, R_i^c, Q_i^c\}$ by $\{F_i, G_i, H_i, L_i, I_m, I_p\}$, and if we reverse the time direction of the $\mathcal{H}_\infty$ control Riccati recursion (so as to propagate forwards in time, rather than backwards in time), then the $\mathcal{H}_\infty$ control Riccati recursion for P_i^c collapses to the $\mathcal{H}_\infty$ filtering Riccati recursion for P_i. Moreover, the state feedback matrix, K_i^{c*}, becomes the Hermitian transpose of the observer gain, $K_{s,i}$.

1.6.6 Linear Algebra and Matrix Theory

This book uses a lot of results from linear algebra and matrix theory. On the other hand, several new results and new insights in matrix theory have arisen from the methods and results of least-squares estimation theory. Several examples of this interplay are scattered throughout the book (in fact, an important one has already been discussed

in Sec. 1.4.3), while several others are implicit in the sense that the matrix analogs (or consequences) have not been explicitly worked out.

The major matrix theory results are those related to the triangular and orthogonal factorizations of matrices, a key step in many linear algebra calculations. In estimation theory terms, this is the problem of finding the innovations process of a given second-order stochastic process (*cf.* Sec. 4.2). The contribution of estimation theory, and more generally of system theory, to this problem is the exploitation of structure in the process — stationarity (and more generally, displacement structure) and state-space structure in order to derive fast algorithms for such factorizations. For the linear algebra (and operator theory) implications of displacement structure, we may refer to the survey papers of Kailath, Vieira, and Morf (1978), Kailath (1987,1999), Kailath and Sayed (1995), and to the edited volume by Kailath and Sayed (1999). Our book of course focuses on the state-space structure, and among noteworthy results here we may cite those in Appendices 8.B, 9.A, 9.B, 13.A, and, for example, Secs. 8.3, 9.4, 9.6, 9.8.4, and 11.5.

1.7 COMPLEMENTS

This introductory chapter plunges immediately into a general state estimation problem. However, it is not necessary to follow the arguments in detail in order to get a feeling for the kinds of problems to be studied in this book. Beginning in Ch. 2, we take a more leisurely route, starting with more elementary results and in fact building up motivation for the important role of state-space models in linear estimation theory.

Sec. 1.1. The Asymptotic Observer. Åström (1970, pp. 157–158) states that "the idea of reconstructing the state of a dynamical system using a mathematical model [as in Secs. 1.1 and 1.2] is probably very old. It was, *e.g.*, encountered in discussions with John Bertram [of IBM] in 1961." The now widely used designation "observer" for the asymptotic estimator comes from Luenberger (1964, 1971).

Sec. 1.2.1. The Squared Error Criterion. There are several reasons for choosing a squared error criterion; some are presented here and some in the notes to Ch. 3. Some of the best discussions of this matter can be found in the classical works of Gauss and Legendre, all available in translation. In particular, we are fortunate to have a fine translation by Stewart (1995) of Gauss's final publications on the topic, including the text of three very clear general lectures. After a nice discussion of regular vs. random errors, Gauss goes on to formulate the least-squares error criterion, with remarks such as the following (Stewart (1995, pp. 9–11)):

"It is by no means self-evident how much loss should be assigned to a given observation error. On the contrary, the matter depends in some part on our own judgment. Clearly, we cannot set the loss equal to the error itself; for if positive errors were taken as losses, negative errors would have to represent gains. The size of the loss is better represented by a function that is naturally positive. Since the number of such functions is infinite, it would seem that we should choose the simplest function having this property. That function is unarguably the square, and the principle proposed above results from this adoption.

Laplace has also considered the problem in a similar manner, but he adopted the absolute value of the error as the measure of this loss. Now if I am not mistaken, this convention is no less arbitrary than mine. Should an error of double size be considered as tolerable as a single error twice repeated or worse? Is it better to assign only twice as much influence to a double error or more? The answers are not self-evident, and the problem cannot be resolved by mathematical proofs, but only by an arbitrary decision.

Moreover, it cannot be denied that Laplace's convention violates continuity and hence resists analytic treatment, while the results that my convention leads to are distinguished by their wonderful simplicity and generality."

This discussion relates to the so-called deterministic least-squares problem described in Sec. 1.6.1. Gauss studied these problems over many years. Among other results, he showed the equivalence to stochastic least-squares problems, where the criterion is now the minimum mean-square error as used in this chapter. For more on this equivalence, see Sec. 3.4.2 and App. 3.5.

Sec. 1.2. The Optimum Transient Observer. The formulation of the optimum transient observer problem and the recognition that the solution is identical to the Kalman filter are also quite old; the oldest reference we are aware of is Battin (1962).

■ PROBLEMS

In keeping with the motivational spirit of this introductory chapter, beginning students could well skip all these problems. Others may find them useful as a way of checking out some of the details omitted in the main text, and as a way of getting more familiar with operations involving vector-valued random variables.

1.1 **(Nonzero means)** Suppose that the random variables $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i\}$ have *known* non-zero means $\{E\mathbf{x}_0 = m_{x_0}, E\mathbf{u}_i = m_{u_i}, E\mathbf{v}_i = m_{v_i}\}$ so that (1.2.2) is replaced by

$$E\begin{bmatrix} \mathbf{u}_i - m_{u_i} \\ \mathbf{v}_i - m_{v_i} \\ \mathbf{x}_0 - m_{x_0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_j^* - m_{u_j}^* & \mathbf{v}_j^* - m_{v_j}^* & \mathbf{x}_0^* - m_{x_0}^* \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \delta_{ij} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \Pi_0 \end{bmatrix}.$$

(a) Suppose that we use the same structure for the estimator as was ysed in the zero-mean case (1.2.7), *viz.*, $\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + Gu_i + \bar{K}_i(\mathbf{y}_i - H\hat{\mathbf{x}}_i)$, $\hat{\mathbf{x}}_0 = 0$. Show that the mean value of the state estimation error $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$ will be nonzero for this choice of estimator, *i.e.*, $E\tilde{\mathbf{x}}_i \neq 0$.

(b) Show that to obtain a zero-mean value for the error $\tilde{\mathbf{x}}_i$ we must replace the (previous) estimator (1.2.3) by $\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + Gu_i + Gm_{u_i} + \bar{K}_i(\mathbf{y}_i - H\hat{\mathbf{x}}_i - m_{v_i})$, $\hat{\mathbf{x}}_0 = m_{x_0}$.

(c) Conclude that the recursion for the estimation error is now given by

$$\tilde{\mathbf{x}}_{i+1} = (F - K_{p,i}H)\tilde{\mathbf{x}}_i + G(\mathbf{u}_i - m_{u_i}) - K_{p,i}(\mathbf{v}_i - m_{v_i}), \quad \tilde{\mathbf{x}}_0 = \mathbf{x}_0 - m_{x_0},$$

and that the arguments leading to Thm. 1.2.1 can now all be applied to the *centered* random variables $\{\mathbf{x}_0 - m_{x_0}, \mathbf{u}_i - m_{u_i}, \mathbf{v}_i - m_{v_i}\}$.

Remark. The very important message is that it is best to first center the random variables in the problem before proceeding further. ◆

1.2 **(Transient observer via differentiation)** A second proof of the result of Lemma 1.2.1 can be obtained by using the notion of complex gradients, as discussed in App. A.6. Consider the quadratic form (1.2.6).

(a) Using (1.2.9), check that (1.2.6) can be rearranged as a quadratic form in $\bar{K}_i^* a$,

$$\xi_{i+1}(a) = a^*\bar{K}_i(R + HP_iH^*)\bar{K}_i^* a - a^*\bar{K}_i(HP_iF^* + S^*G^*)a \\ - a^*(FP_iH^* + GS)\bar{K}_i^* a + a^*(FP_iF^* + GQG^*)a.$$

(b) Verify that for the complex gradient $\partial\xi_{i+1}(a)/\partial(\bar{K}_i^* a)$ to be zero for any a, we must choose $\bar{K}_i = K_{p,i}$, where $K_{p,i}R_{e,i} = FP_iH^* + GS$ and $R_{e,i} = R + HP_iH^*$. Verify also that the Hessian matrix is given by

$$\frac{\partial^2\xi_{i+1}}{\partial(\bar{K}_i^* a)\partial(a^*\bar{K}_i)} = (R + HP_iH^*) \geq 0.$$

Conclude that the above choice for $\bar{K}_i$ is in fact a minimum, and that

$$\xi_{i+1,\min} \triangleq a^*P_{i+1}a = a^*\left[FP_iF^* + GQG^* - K_{p,i}(R + HP_iH^*)K_{p,i}^*\right]a,$$

which means that P_{i+1} is defined by the Riccati recursion (1.2.12).

1.3 **(Initial conditions)** Suppose $\mathbf{x}_0$ is a random variable with *known* mean, $E\mathbf{x}_0 = m_0$, and variance matrix $\Pi_0 = E(\mathbf{x}_0 - m_0)(\mathbf{x}_0 - m_0)^*$.

(a) Show that the minimum of $E(\mathbf{x}_0 - m)(\mathbf{x}_0 - m)^*$ over m is achieved for the choice $m = m_0$, and that the minimum value is Π_0.

(b) Suppose we have an observation $\mathbf{y}_0 = H\mathbf{x}_0 + \mathbf{v}_0$, with $E\mathbf{x}_0 = 0$, $E\mathbf{v}_0 = 0$, $E\mathbf{x}_0\mathbf{x}_0^* = \Pi_0$, $E\mathbf{v}_0\mathbf{v}_0^* = R > 0$, and $E\mathbf{x}_0\mathbf{v}_0^* = 0$. Show that the linear estimator $\hat{\mathbf{x}}_{0|0} \triangleq K_{f,o}\mathbf{y}_0$ that solves

$$\min_{K_f} E(\mathbf{x}_0 - K_f\mathbf{y}_0)(\mathbf{x}_0 - K_f\mathbf{y}_0)^*,$$

is obtained for the choice $K_{f,o} = \Pi_0H^*(R + H\Pi_0H^*)^{-1}$, and that the minimum value of the mean-square error is $\Pi_0 - \Pi_0H^*(R + H\Pi_0H^*)^{-1}H\Pi_0 \triangleq P_{0|0}$.

(c) Show, using the matrix inversion formula in App. A, that $P_{0|0}^{-1} = \Pi_0^{-1} + H^*R^{-1}H$.

1.4 **(Filtered estimators)** The discussion given in the text can be carried out under a slightly different set of assumptions. Consider once more the state-space model (1.2.1)–(1.2.2) where, for simplicity, we shall assume $S = 0$ and $u_i = 0$. Until now we used estimators of the state, $\hat{\mathbf{x}}_i$, based on the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$, which are called *predicted* estimates. In this problem we shall be interested in the so-called *filtered* estimators of the state, $\hat{\mathbf{x}}_{i|i}$, based on the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}$. To use this, it turns out we should hypothesize an observer structure of the form

$$\hat{\mathbf{x}}_{i+1|i+1} = F\hat{\mathbf{x}}_{i|i} + \bar{K}_{i+1}(\mathbf{y}_{i+1} - HF\hat{\mathbf{x}}_{i|i}), \quad \hat{\mathbf{x}}_{0|0} = \Pi_0H^*(R + H\Pi_0H^*)^{-1}\mathbf{y}_0.$$

(a) Show that the resulting (filtered) error $\tilde{\mathbf{x}}_{i|i} = \mathbf{x}_i - \hat{\mathbf{x}}_{i|i}$ satisfies the recursion

$$\tilde{\mathbf{x}}_{i+1|i+1} = (F - \bar{K}_{i+1}HF)\tilde{\mathbf{x}}_{i|i} + (G - \bar{K}_{i+1}HG)\mathbf{u}_i - \bar{K}_{i+1}\mathbf{v}_{i+1}.$$

(b) Suppose that we want to choose $\bar{K}_{i+1}$ so as to minimize the filtered error covariance matrix, $E\tilde{\mathbf{x}}_{i+1|i+1}\tilde{\mathbf{x}}_{i+1|i+1}^*$. Denoting this minimum by $P_{i+1|i+1}$ and the minimizing gain by $K_{f,i+1}$, show that $K_{f,i+1} = (FP_{i|i}F^* + GQG^*)H^*R_{f,i+1}^{-1}$, where $R_{f,i+1} = R + H(FP_{i|i}F^* + GQG^*)H^*$, and $P_{i|i}$ satisfies the recursion (assuming the inverses on the right-hand side exist — sufficient conditions for such inverses to exist are established in Lemma 9.5.1 and Prob. 9.17)

$$P_{i+1|i+1}^{-1} = (FP_{i|i}F^* + GQG^*)^{-1} + H^*R^{-1}H, \qquad P_{0|0}^{-1} = \Pi_0^{-1} + H^*R^{-1}H.$$

Remark. Starting with $\mathbf{y}_{i+1} = H\mathbf{x}_{i+1} + \mathbf{v}_{i+1} = HF\mathbf{x}_i + HG\mathbf{u}_i + \mathbf{v}_{i+1}$, it can be shown that the optimal estimator of $\mathbf{y}_{i+1}$ given $\{\mathbf{y}_j,\ j \leq i\}$ (in the linear least-mean-squares sense) is $\hat{\mathbf{y}}_{i+1|i} = HF\hat{\mathbf{x}}_{i|i}$. This motivates the feedback term $(\mathbf{y}_{i+1} - HF\hat{\mathbf{x}}_{i|i})$ that appears in the equation for the observer introduced above. It can also be shown, with some algebra, that $FP_{i|i}F^* + GQG^* = P_{i+1}$, so that we can write $R_{f,i+1} = R_{e,i+1}$, $K_{f,i+1} = P_{i+1}H^*R_{e,i+1}^{-1}$. The origin of these identities, and immediate proofs, will appear later in Ch. 9. ◆

1.5 **(Orthogonality properties)** Refer to Thm. 1.2.1 and define $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$. Assume further that $u_i = 0$ for all i so that the observations $\mathbf{y}_i$ are zero-mean.

(a) First show that $E\hat{\mathbf{x}}_{i+1}\tilde{\mathbf{x}}_{i+1}^* = F(E\hat{\mathbf{x}}_i\tilde{\mathbf{x}}_i^*)F_{p,i}^*$.

(b) Then show that $E\hat{\mathbf{x}}_j\tilde{\mathbf{x}}_i^* = 0$ for all $j \leq i$.

(c) Now show that $E\mathbf{y}_j\tilde{\mathbf{x}}_i^* = 0$ for all $j < i$.

(d) Verify that $E\mathbf{x}_{i+1}\mathbf{e}_i^* = FP_iH^* + GS$.

(e) Conclude that $E\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^* \leq E(\mathbf{x}_i - \mathbf{a})(\mathbf{x}_i - \mathbf{a})^*$, for all $\mathbf{a}$ that are linear functions of $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$, with equality only when $\mathbf{a} = \hat{\mathbf{x}}_i$.

(f) How would the results of parts (a)–(e) change if the u_i were nonzero?

Remark. We shall see in Chs. 3 and 4 that these somewhat painfully derived properties (as well as the results of Sec. 1.4) are immediate when $\hat{\mathbf{x}}_i$ is the linear least-mean-squares estimator of $\mathbf{x}_i$ given $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{i-1}\}$. ◆

1.6 **(Another array algorithm)** Here we consider a special case of an array algorithm that propagates a square-root factor of P_i^{-1} (rather than of P_i). So assume $F_i = I$, $G_i = 0$, and $R_i = I$. Form the pre-array $\mathcal{A}$ shown below and apply to it any unitary rotation matrix Θ that will introduce a (block) zero entry in the first (block) row of the post-array $\mathcal{B}$:

$$\underbrace{\begin{bmatrix} P_i^{-*/2} & H_i^* \\ \hat{\mathbf{x}}_i^* P_i^{-*/2} & \mathbf{y}_i^* \\ 0 & I \end{bmatrix}}_{\mathcal{A}} \Theta = \underbrace{\begin{bmatrix} X & 0 \\ \mathbf{a} & \mathbf{b} \\ Y & Z \end{bmatrix}}_{\mathcal{B}}.$$

By comparing entries on both sides of the equality $\mathcal{A}\mathcal{A}^* = \mathcal{B}\mathcal{B}^*$ show that we can make the identifications

$$X = P_{i+1}^{-*/2}, \quad \mathbf{a} = \hat{\mathbf{x}}_{i+1}^* P_{i+1}^{-*/2}, \quad \mathbf{b} = \mathbf{e}_i^* R_{e,i}^{-*/2}, \quad Y = H_i P_{i+1}^{1/2}, \quad Z = R_{e,i}^{-*/2}.$$

Remark. Note that the post-array propagates the quantities $P_{i+1}^{-1/2}$ and $P_{i+1}^{-1/2}\hat{\mathbf{x}}_{i+1}$. Hence, the state estimator can also be determined by solving the triangular system of linear equations $\mathbf{a} = \hat{\mathbf{x}}_{i+1}^* P_{i+1}^{-*/2}$, rather than via the usual recursion (1.2.17). This algorithm will be used in the discussion at the end of Sec. 2.6. ◆

CHAPTER 2

Deterministic Least-Squares Problems

The motivational Ch. 1 introduced a relatively recent and relatively advanced least-squares estimation problem. In this chapter, whose intent is also partly motivational, we introduce one of the earliest and in some ways a more elementary least-squares problem. Ch. 1 will have shown the importance of getting to be at ease with matrix manipulations. This chapter will reemphasize that fact, as well as show the need for a good grasp of some basic facts from linear algebra. Though the chapter is relatively self-contained, and though it does present much of the necessary linear algebra material (see App. 2.A), many readers may find it useful to also review the material on linear equations in one or more of the numerous linear algebra textbooks now available.

The key section to master here is Sec. 2.3 on the geometric formulation; Secs. 2.1 to 2.2 may well be skimmed rather rapidly and reread after the more leisurely discussions in Sec. 2.3. Sec. 2.6 should be read more closely, particularly to note some important analogies between the purely deterministic least-squares problem of this chapter and the stochastic least-squares problem of Ch. 1.

Sec. 2.7 on the less-studied downdating problem brings up the provocative idea that working in indefinite metric spaces rather than the usual Euclidean spaces can be helpful; this theme is related to the discussions in Secs. 1.6.3 and 1.6.4 on $\mathcal{H}_\infty$ problems (the monograph by Hassibi, Sayed, and Kailath (1999) gives a more detailed treatment). Sec. 2.8 describes some of the many directions in which further important work continues in this classical field.

Readers wishing to proceed more quickly to the Kalman filter can go directly to Ch. 3; others will find the material and several of the problems useful for studying adaptive filtering theory.

2.1 THE DETERMINISTIC LEAST-SQUARES CRITERION

There is a long history, going back to Legendre (1805) and Gauss (1809) at the beginning of the nineteenth century, of problems reducible to the solution of an *inconsistent* overdetermined set of linear equations

$$Hx \cong y, \tag{2.1.1}$$

where H is a given $N \times n$ matrix, $N \geq n$ (hence overdetermined), y is a given $N \times 1$ vector, and x is an unknown $n \times 1$ vector. Inconsistency (denoted by the symbol $\cong$) means that y is not a linear combination of the columns of H, *i.e.*, y is not an element of $\mathcal{R}(H)$, where $\mathcal{R}(H)$ is the column space of H (see App. 2.A). This is the reason we did not use an equality sign in (2.1.1). The fact that y does not belong to $\mathcal{R}(H)$ means that

$$y = Hx + v, \tag{2.1.2}$$

for some nonzero $N \times 1$ vector v, which is referred to as the *residual*. Since $y \notin \mathcal{R}(H)$, there is no choice of x that will make $v = 0$. A *least-squares* (LS) solution, $\hat{x}$, is one that minimizes the *length* of the residual vector, *i.e.*, it is one with the property that

$$\| y - H\hat{x} \|^2 \leq \| y - Hx \|^2, \tag{2.1.3}$$

for all $x \in \mathbb{C}^n$, where $\| \cdot \|^2$ denotes the squared Euclidean norm, *viz.*,

$$\|v\|^2 = v^*v = \sum_{i=1}^{N} |v(i)|^2. \tag{2.1.4}$$

Here, $v(i)$ stands for the i-th entry of the vector v. It can be shown (see Prob. 2.1) that if the equations (2.1.1) are consistent (*i.e.*, if $y \in \mathcal{R}(H)$), then the least-squares solution $\hat{x}$ is an exact solution; we say an rather than the because there will be many solutions $\hat{x}$ if H is not full rank.

Of course, we could use other criteria for choosing $\hat{x}$, *e.g.*, so as to minimize $\|y - Hx\|_1 = \|v\|_1 = \sum_{i=1}^{N} |v(i)|$; the resulting solutions are often known as l_1 estimators. While several facts are known about such estimates, from the time of Laplace and Gauss (see, *e.g.*, Plackett (1972), Sheynin (1977,1979), and Stewart (1995)), the least-squares theory is much richer and the only one we shall study here.

Numerous physical problems lead to overdetermined systems of linear equations of the form (2.1.1). Sometimes they arise directly, as in adaptive RLS filtering (see Sec. 2.6). However, they often arise in the course of solving other optimization problems, *e.g.*, in using the method of the *second variation* in optimal control (see Bryson and Ho (1969)); also in the increasingly popular interior-point methods for solving convex optimization problems (see, *e.g.*, Nesterov and Nemirovskii (1994) and Boyd et al. (1994)).

2.2 THE CLASSICAL SOLUTION

Since many readers may already have studied least-squares solutions in other contexts (*e.g.*, in elementary linear algebra), we shall first present the straightforward solution of the nonregularized least-squares problem $\min_x J(x)$, where the *cost function* $J(x)$ is defined as

$$J(x) \triangleq \|y - Hx\|^2 = x^*H^*Hx - x^*H^*y - y^*Hx + y^*y. \tag{2.2.1}$$

2.2.1 The Normal Equations

We shall show that the least-squares solutions can be obtained via certain so-called "normal" equations. One reason for the name will be seen in Sec. 2.3, where the equations are derived by using certain orthogonality arguments.

Lemma 2.2.1 (The Normal Equations) *A vector $\hat{x}$ is a minimizer of the cost function (2.2.1) if, and only if, it satisfies the (always consistent) so-called normal equations*

$$H^*H\hat{x} = H^*y. \tag{2.2.2}$$

The resulting minimum value of the cost function $J(x)$ can be written as

$$J(\hat{x}) \triangleq \|y - H\hat{x}\|^2 = \|y\|^2 - y^*H\hat{x} = \|y\|^2 - \|H\hat{x}\|^2, \tag{2.2.3}$$

which is suggestive of a geometric interpretation (see Sec. 2.3). ■

Proof: An immediate proof follows by differentiation (see App. A.6):

$$0 = \frac{\partial}{\partial x}(x^*H^*Hx - x^*H^*y - y^*Hx + y^*y)\Big|_{x=\hat{x}} = \hat{x}^*H^*H - y^*H,$$

which shows that every solution $\hat{x}$ must satisfy the normal equations $H^*H\hat{x} = H^*y$. That any such $\hat{x}$ will minimize the cost function $J(x)$ can be seen by noting that the Hessian matrix is positive-semi-definite:

$$\frac{\partial^2\|y - Hx\|^2}{\partial x^*\partial x} = H^*H \geq 0. \tag{2.2.4}$$

The value of $J(\cdot)$ at the minimum can be expressed as

$$\begin{aligned} J(\hat{x}) &= \|y - H\hat{x}\|^2 = (y - H\hat{x})^*(y - H\hat{x}), \\ &= y^*(y - H\hat{x}), \quad \text{since by (2.2.2)} \quad H^*(y - H\hat{x}) = 0, \\ &= \|y\|^2 - y^*H\hat{x} = \|y\|^2 - (H^*y)^*\hat{x} = \|y\|^2 - (H^*H\hat{x})^*\hat{x}, \\ &= \|y\|^2 - \|H\hat{x}\|^2. \end{aligned} \tag{2.2.5}$$

◆

When the matrix H has full rank n, the matrix H^*H will be nonsingular (see below) and we can explicitly solve for $\hat{x}$ and give a more explicit formula for the minimum cost.

Lemma 2.2.2 (Unique Solutions) *When H has full rank n, there is a unique $\hat{x}$ satisfying (2.2.3), which can be expressed as*

$$\hat{x} = (H^*H)^{-1}H^*y. \tag{2.2.6}$$

Moreover, the resulting minimum value of the cost can be written as

$$J(\hat{x}) = \| y - H\hat{x} \|^2 = y^*(I - H(H^*H)^{-1}H^*)y. \tag{2.2.7}$$

■

Proof: We first have to show that H being full rank implies that H^*H is nonsingular, or equivalently, that H^*H being singular implies that H does not have full rank. Indeed, if H^*H is singular then there must exist a nonzero vector c such that $H^*Hc = 0$, which implies that $c^*H^*Hc = \|Hc\|^2 = 0$, so that $Hc = 0$. This in turn means that the columns of H are linearly dependent. Hence H is not full rank, which is a contradiction. Once we know that H^*H is nonsingular, the formulas (2.2.6)–(2.2.7) follow easily from the results of Lemma 2.2.1. ◆

When H is not full rank, H^*H will be singular, and one needs to address the question of whether the normal equations have a solution, and if so, whether it is unique. It turns out that the normal equations will always have a solution, although the solution will be nonunique when H is not full rank. However, no matter which solution $\hat{x}$ is used, the quantity $H\hat{x}$ will be the same, and the minimum value of $\|y - H\hat{x}\|^2$ will always be the same. These facts will be fairly evident from the geometric formulation to be given shortly. An algebraic proof is described in App. 2.A. We just state the results here.

Lemma 2.2.3 (The General Case) *Consider the normal equations (2.2.2).*

(a) *When H is full rank, the unique solution is given by $\hat{x} = (H^*H)^{-1}H^*y$.*

(b) *When H is not full rank, the normal equations always have more than one solution, where any two solutions $\hat{x}_1$ and $\hat{x}_2$ differ by a vector in the nullspace of H, i.e., $H(\hat{x}_1 - \hat{x}_2) = 0$.*

(c) *The projection of y onto $\mathcal{R}(H)$ is unique and is defined by $\hat{y} \triangleq H\hat{x}$, where $\hat{x}$ is any solution to the normal equations; when H has full rank, we can write $\hat{y} = H(H^*H)^{-1}H^*y$.*

■

One way to obtain a unique solution to the normal equations is to insist on a solution that has the minimum Euclidean norm; it turns out that this is equivalent to solving the normal equations using the so-called Moore-Penrose pseudo-inverse (see App. A.4 for the definition of the pseudo-inverse and App. 2.A for a derivation of the minimum norm solution).

2.2.2 Weighted Least-Squares Problems

In many applications (*e.g.*, adaptive filtering), a weighted least-squares criterion is more appropriate,

$$J(x) = \|y - Hx\|_W^2 \triangleq (y - Hx)^* W (y - Hx), \tag{2.2.8}$$

where W is any Hermitian positive-definite matrix. It is not difficult to show that the earlier arguments generalize to give the following statement.

Lemma 2.2.4 (Weighted Least-Squares Solutions) *The weighted least-squares solutions, $\hat{x}_W$, of the inconsistent equations $Hx \cong y$, are those with the property*

$$\|y - H\hat{x}_W\|_W^2 \quad \leq \quad \|y - Hx\|_W^2, \tag{2.2.9}$$

for all $x \in \mathbb{C}^n$. They are given by any solution of the consistent (normal) system of equations

$$H^*WH\hat{x} = H^*Wy. \tag{2.2.10}$$

The corresponding minimum value of $\|y - Hx\|_W^2$ is

$$\|y - H\hat{x}\|_W^2 = y^*Wy - y^*WH\hat{x}, \tag{2.2.11}$$

which, in the case of a full rank H, can be written as

$$\|y - H\hat{x}\|_W^2 = y^*(W - WH(H^*WH)^{-1}H^*W)y. \tag{2.2.12}$$

■

2.2.3 Statistical Assumptions on the Noise

Such weighted least-squares problems arise not only in adaptive filtering, as mentioned earlier, but also in many problems where we have a model of the form $\mathbf{y} = Hx + \mathbf{v}$, where x is a deterministic but unknown vector while $\mathbf{v}$ is a random "noise" or "disturbance" vector, with known mean and variance, say $E\mathbf{v} = 0$ and $E\mathbf{v}\mathbf{v}^* = R_v$. In this case, the weighted least-squares estimator,

$$\hat{\mathbf{x}} = (H^*WH)^{-1}H^*W\mathbf{y},$$

will also be random with mean

$$E\hat{\mathbf{x}} = E(H^*WH)^{-1}H^*W\mathbf{y} = x,$$

since $E\mathbf{y} = Hx$. Such estimators are said to be *unbiased*. The covariance matrix can be computed as

$$E(\hat{\mathbf{x}} - x)(\hat{\mathbf{x}} - x)^* = (H^*WH)^{-1}H^*WR_vWH(H^*WH)^{-1}.$$

It can be shown (see Sec. 3.4.2) that the covariance matrix is smallest when we choose the weighting matrix W as $W = R_v^{-1}$ so that

$$E(\hat{\mathbf{x}} - x)(\hat{\mathbf{x}} - x)^* = (H^*WH)^{-1}.$$

In the statistical literature (see, *e.g.*, Rao (1973)), the corresponding $\hat{\mathbf{x}}$ is often called the Minimum Variance Unbiased Estimator (MVUE) or the Gauss-Markov estimator, and the above result in fact follows from the Gauss-Markov theorem of Sec. 3.4.2. While, for simplicity, we assumed the invertibility of H^*WH, a more general treatment is possible. We shall encounter these estimators again at the end of Sec. 3.4.1.

2.3 A GEOMETRIC FORMULATION: THE ORTHOGONALITY CONDITION

Our expression (2.2.3) for the minimum cost, *viz.*,

$$\|y - H\hat{x}\|^2 = \|y\|^2 - \|H\hat{x}\|^2,$$

suggests that we can regard the column matrices, $y - H\hat{x}$ and $H\hat{x}$, as orthogonal vectors in an Euclidean space — see Fig. 2.1. Pursuing this further we rewrite the normal equations (2.2.2) as

$$H^*(y - H\hat{x}) = 0,$$

which states that $y - H\hat{x}$ is orthogonal to the vectors defined by the columns of H. We pursue this interpretation in this section. We proceed rather slowly, because we shall soon be using the concepts in a nontypical setting, *viz.*, for vectors defined by random variables and with matrix-valued (rather than scalar-valued) inner products. Let $\underline{h}_i \in \mathbb{C}^N$ denote the columns of the matrix H, *i.e.*, $H = [\underline{h}_0 \ \dots \ \underline{h}_{n-1}]$. [We shall reserve the notation $\{h_0, h_1, \dots, h_{N-1}\}$, *i.e.*, without underbars, for the rows of H. The above definition for the columns of H is not very elegant but is made to reduce notational explosion.]

Also let the components of $x \in \mathbb{C}^n$ be denoted by $x(0), x(1), \dots x(n-1)$, *i.e.*, $x = \text{col}\{x(0), x(1), \dots, x(n-1)\}$. The least-squares problem is, as noted before, one of finding a linear combination of the columns of H, say $H\hat{x}$, such that

$$\|y - H\hat{x}\|^2 \triangleq \left\| y - \sum_{i=0}^{n-1} \underline{h}_i \hat{x}(i) \right\|^2 = \text{minimum}.$$

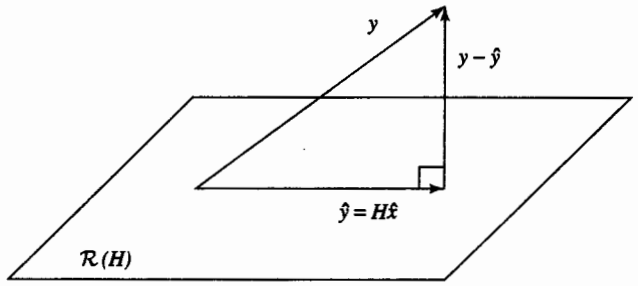


Figure 2.1 The least-squares solution is obtained when $(y - \hat{y}) \perp \mathcal{R}(H)$.

Now if $N = 3$ and $n = 2$, then the geometry of the 3-dimensional space (see Fig. 2.1) will tell us that $\hat{x}$ must be such that

$$y - \underline{h}_0\hat{x}(0) - \underline{h}_1\hat{x}(1) \perp \underline{h}_0, \quad y - \underline{h}_0\hat{x}(0) - \underline{h}_1\hat{x}(1) \perp \underline{h}_1.$$

That is,

$$\underline{h}_0^*\left[y - \underline{h}_0\hat{x}(0) - \underline{h}_1\hat{x}(1)\right] = 0 \quad \text{and} \quad \underline{h}_1^*\left[y - \underline{h}_0\hat{x}(0) - \underline{h}_1\hat{x}(1)\right] = 0,$$

or equivalently, $H^*y = H^*H\hat{x}$, which are exactly the normal equations for this problem!

Actually the power of matrix notation is such that the reader may have no hesitation in saying that the above argument can be used for arbitrary N-dimensional vectors, not just for 3-dimensional vectors. Therefore using a geometric formulation, the result of Lemma 2.2.1 is "obvious." Moreover, since the projection onto a linear space is *obviously* unique, so is the result of Lemma 2.2.3. Also, the result of Lemma 2.2.4 would follow by using a generalization of the notion of Euclidean length and Euclidean inner products, namely,

$$\|a\|_W^2 = a^*Wa, \quad \langle a, b\rangle_W \triangleq b^*Wa. \tag{2.3.1}$$

In other words, we could just interpret the geometric condition

$$(y - \underline{h}_0\hat{x}(0) - \underline{h}_1\hat{x}(1)) \perp \underline{h}_0$$

as

$$\underline{h}_0^*W\left[y - \underline{h}_0\hat{x}(0) - \underline{h}_1\hat{x}(1)\right] = 0 = \langle \underline{h}_0, y - \underline{h}_0\hat{x}(0) - \underline{h}_1\hat{x}(1)\rangle_W,$$

and so on.

But now, if not earlier, one might hesitate a little bit. Can one use any *weighting* matrix W in the above argument? Moreover, even if $W = I$, can we still be sure that our geometric knowledge from 3-dimensional spaces always extends to N-dimensional spaces, $N > 3$?

In fact, care is needed; it turns out that the weighting matrix W has to be positive-definite.[1] And we must remember that not all our geometric intuition from 3-dimensional spaces carries over to N-dimensional spaces. For example, readers with a background in information theory (not to speak of those with a background in N-dimensional geometry) may recall that the volume of an N-dimensional sphere is mostly concentrated in the outer shells of the sphere.

The reason our geometric intuition from 3-dimensional spaces is valid for our particular problem is that the properties of inner products in 3-dimensional spaces carry over to any linear vector space for which the concept of inner product is *properly defined*. In other words, we now appeal to the fact that mathematicians have developed a general formulation of linear vector spaces with inner products (so-called inner product spaces), where the *vectors* and *inner products* can be abstract objects obeying certain rules. For example, the vectors need not be elements of $\mathbb{C}^N$ and the inner products need not be real or complex numbers. We shall give appropriate definitions soon, but

[1] In the monograph (Hassibi, Sayed, and Kailath (1999)), we consider *indefinite* weighting matrices W. In that case, it turns out that projections may not always exist or be unique, and that they do not, in general, minimize the length of the residual vector.

to dispel the feeling that this is unnecessary generality, let us just mention that such a specific situation will be encountered in the next chapter, where the vectors can be scalar- (or even vector-) valued random variables and where inner products can be rectangular matrices.

2.3.1 The Projection Theorem in Inner Product Spaces

We assume we have a linear space, say $\mathcal{V}$, whose elements are called "vectors." We also have a space of "scalars," $\mathcal{S}$ say, such that

$$\alpha v \in \mathcal{V} \quad \text{for any} \quad v \in \mathcal{V}, \quad \text{any} \quad \alpha \in \mathcal{S}.$$

The space $\mathcal{S}$ is often the field of complex numbers, but it can be a more general algebraic object; in the next chapter $\mathcal{S}$ will be a ring of matrices with complex entries. We shall not repeat here the formal definition of a linear vector space over $\mathcal{S}$; an intuitive understanding will be adequate, though it will not hurt to look up the definition in any textbook on linear algebra. We shall not define "ring" either; for our application it will mean that the product of matrices is another matrix.

Now with each pair of vectors, say $v \in \mathcal{V}$, $u \in \mathcal{V}$, the *inner product* $\langle v, u\rangle$ is an element of $\mathcal{S}$ characterized by the following properties:

1. Linearity: $\langle \alpha_1 v_1 + \alpha_2 v_2, u\rangle = \alpha_1\langle v_1, u\rangle + \alpha_2\langle v_2, u\rangle$.
2. Reflexivity: $\langle u, v\rangle = \langle v, u\rangle^*$.
3. Nondegeneracy: $\|v\|^2 \triangleq \langle v, v\rangle$ is zero only when $v = 0$.

Here 0 denotes the zero element in $\mathcal{V}$. The operation $*$ depends on the space $\mathcal{S}$. When $\mathcal{S}$ is the field of complex numbers, $*$ denotes complex conjugate; when $\mathcal{S}$ is the ring of matrices, $*$ stands for the conjugate transpose (usually called Hermitian transpose).

Now in this chapter we are interested in the vector space, usually denoted $\mathbb{C}^N$ or $\mathbb{C}^{N\times 1}$, of N-dimensional column vectors (*i.e.*, $N \times 1$ matrices) over the field of complex numbers. The first definition we used above was

$$\langle a, b\rangle = \sum_{i=1}^{N} a(i)b^*(i) = b^*a, \quad \text{where} \quad a, b \in \mathbb{C}^N.$$

The reader should verify that this definition meets the three requirements listed above for being a legitimate inner product. The reader should also verify that the (weighted) inner products in (2.3.1) satisfy the axioms, and in particular that $W > 0$ is necessary to meet the nondegeneracy requirement.

Now in any inner-product space, the following facts about projections can be established. Let $\mathcal{L}$ be a linear subspace of $\mathcal{V}$ and let y be an arbitrary element of $\mathcal{V}$. The *projection* of y onto $\mathcal{L}$, denoted by $\hat{y}_{\mathcal{L}}$, or often just $\hat{y}$, is a unique element of $\mathcal{L}$ such that

$$\langle y - \hat{y}, a\rangle = 0, \quad \text{for all } a \in \mathcal{L}. \tag{2.3.2}$$

The proof that there must exist such an element is deferred to App. 4.A. But assuming the existence of such a unique projection we can readily prove the following.

Lemma 2.3.1 (Orthogonality and Approximation) *Let $\mathcal{L}$ be a subspace of a linear vector space $\mathcal{V}$ and let y be any element of $\mathcal{V}$. Then the projection, $\hat{y}_{\mathcal{L}}$, has the property that $\|y - \hat{y}_{\mathcal{L}}\|^2 \leq \|y - a\|^2$ for any $a \in \mathcal{L}$.* ∎

Proof: We can write

$$\begin{aligned}\|y - a\|^2 &= \|y - \hat{y}_{\mathcal{L}} + \hat{y}_{\mathcal{L}} - a\|^2, \\ &= \|y - \hat{y}_{\mathcal{L}}\|^2 + \|\hat{y}_{\mathcal{L}} - a\|^2 + \langle y - \hat{y}_{\mathcal{L}}, \hat{y}_{\mathcal{L}} - a\rangle + \langle \hat{y}_{\mathcal{L}} - a, y - \hat{y}_{\mathcal{L}}\rangle.\end{aligned}$$

But since $\hat{y}_{\mathcal{L}} \in \mathcal{L}$ and $a \in \mathcal{L}$, $(\hat{y}_{\mathcal{L}} - a) \in \mathcal{L}$ and, by definition, $(y - \hat{y}_{\mathcal{L}})$ is orthogonal to $(\hat{y}_{\mathcal{L}} - a)$. Therefore,

$$\|y - \hat{y}_{\mathcal{L}}\|^2 = \|y - a\|^2 - \|\hat{y}_{\mathcal{L}} - a\|^2 \leq \|y - a\|^2.$$

◆

The orthogonality condition (2.3.2) plays a fundamental role in least-squares theory.

2.3.2 Geometric Insights

Lemma 2.3.1 is what justifies the geometric argument used at the beginning of this section. The subspace $\mathcal{L}$ is now the space $\mathcal{R}(H)$ spanned by the columns of the matrix H. The least-squares solution, $\hat{x}$, is characterized by the fact that the residual vector $y - H\hat{x}$ is orthogonal to $\mathcal{R}(H)$, or equivalently, $\hat{y} \triangleq H\hat{x}$ is given by the unique projection of y onto $\mathcal{R}(H)$. This is depicted in Fig. 2.1.

Likewise, the weighted least-squares solution of Lemma 2.2.4 can be characterized by the fact that the residual vector $y - H\hat{x}$ is orthogonal to $\mathcal{R}(H)$. Now, however, the orthogonality is with respect to the weighted inner product defined by $\langle a, b\rangle_W = b^*Wa$ in (2.3.1). This yields $H^*W[y - H\hat{x}] = 0$, which gives us again the (weighted) normal equations (2.2.10), *viz.*, $H^*WH\hat{x} = H^*Wy$.

Another bonus of the geometric formulation is that it is immediate that the solution of the normal equations actually gives us a minimum; this fact needed a separate step in the algebraic approach. There is another (small) way in which the geometric picture gives us a better understanding of the problem that is immediately evident from the algebraic approach. It shows that the least-squares approach to the inconsistent linear equations $Hx \cong y$ is to first make them consistent by replacing y by $\hat{y}$, the unique projection of y onto the space $\mathcal{R}(H)$, spanned by the columns of H (see Fig. 2.1). The equations $Hx = \hat{y}$ are consistent and the solution is denoted by $\hat{x}$. Note of course that though $\hat{y}$ is uniquely defined by $\{H, y\}$, the same is not necessarily true of the solution of the consistent equations $Hx = \hat{y}$. If H has full rank, *i.e.*, if the vectors defined by the columns of H are linearly independent, then there is a unique solution $\hat{x}$, *i.e.*, a unique combination of the columns of H that will give us $\hat{y}$. However, if the rank of H is not full, *i.e.*, its columns are linearly dependent, then of course there can be many different combinations that will produce $\hat{y}$ so that $\hat{x}$ will not be unique.

2.3.3 Projection Matrices

When the matrix H has full rank we can be more explicit about projections onto $\mathcal{R}(H)$. In this case, we can write using (2.2.10) that the projection of y onto $\mathcal{R}(H)$ is

$$\hat{y} = H\hat{x} = H(H^*WH)^{-1}H^*Wy \triangleq \mathcal{P}_H y.$$

The matrix $\mathcal{P}_H = H(H^*WH)^{-1}H^*W$ is the so-called (weighted) *projector matrix* onto the space $\mathcal{R}(H)$, spanned by the columns of H. It has the useful and easily verified (algebraically and geometrically) properties of

(i) Symmetry: $(W\mathcal{P}_H)^* = W\mathcal{P}_H$ (ii) Idempotency: $\mathcal{P}_H = \mathcal{P}_H^2$,

which will look more familiar when $W = I$.

We note that $P_H^{\perp} \triangleq I - \mathcal{P}_H$ is also a (weighted) projector matrix, projecting vectors onto the linear space of all vectors orthogonal to $\mathcal{R}(H)$. This space is denoted by $\mathcal{R}^{\perp}(H)$ and is called the orthogonal complement of $\mathcal{R}(H)$. If H consists of two 3×1 column vectors, then $\mathcal{R}(H)$ is a plane, and $\mathcal{R}^{\perp}(H)$ will be the one-dimensional space spanned by all multiples of any vector orthogonal to the plane $\mathcal{R}(H)$.

2.3.4 An Application: Order-Recursive Least-Squares

In order to appreciate the power of the geometric approach, let us develop an order-recursive (unweighted, for notation simplicity) least-squares solution. More specifically, let $\hat{x}_{n,N}$ now denote the $n \times 1$ solution of the overdetermined system of linear equations $Hx \cong y$ where, as above, $H \in \mathbb{C}^{N \times n}$ has full column rank and $y \in \mathbb{C}^N$. The two subscripts n and N used in $\hat{x}_{n,N}$ indicate that it is a vector of dimension n and is based on data $\{H, y\}$ with N rows.

Assume that we add one *column* to H, and correspondingly one entry to x, to obtain the following overdetermined system:

$$\begin{bmatrix} H & \underline{h}_n \end{bmatrix} \begin{bmatrix} x \\ x(n) \end{bmatrix} \cong y,$$

where $\begin{bmatrix} H & \underline{h}_n \end{bmatrix}$ is also assumed to have full column rank. We denote the least-squares solution of this higher-order problem by $\hat{x}_{n+1,N}$, and it is now a vector of dimensions $(n+1) \times 1$. The order-recursive least-squares problem is to relate $\hat{x}_{n+1,N}$ to $\hat{x}_{n,N}$.

One can in principle proceed algebraically and derive the desired relation by relying on the defining formula

$$\hat{x}_{n+1,N} = \left(\begin{bmatrix} H^* \\ \underline{h}_n^* \end{bmatrix} \begin{bmatrix} H & \underline{h}_n \end{bmatrix} \right)^{-1} \begin{bmatrix} H^* \\ \underline{h}_n^* \end{bmatrix} y. \tag{2.3.3}$$

However, a more direct route to the solution is possible if we exploit the geometry of the least-squares problem. To see this, let $\hat{\underline{h}}_n$ denote the projection of $\underline{h}_n$ onto the column span of H, *viz.*,

$$\hat{\underline{h}}_n = \mathcal{P}_H \underline{h}_n = H(H^*H)^{-1}H^*\underline{h}_n = Ha, \quad \text{say.}$$

Note that a denotes the column vector that projects $\underline{h}_n$ onto $\mathcal{R}(H)$, *i.e.*, $a = (H^*H)^{-1}H^*\underline{h}_n$. Let also the resulting residual vector be denoted by

$$\tilde{\underline{h}}_n = \underline{h}_n - \hat{\underline{h}}_n = \underline{h}_n - \mathcal{P}_H \underline{h}_n = h_n - Ha.$$

Then the matrix $\begin{bmatrix} H & \tilde{\underline{h}}_n \end{bmatrix}$ provides a new basis for the space $\mathcal{R}\left\{\begin{bmatrix} H & \underline{h}_n \end{bmatrix}\right\}$, with the advantage of having a column $\tilde{\underline{h}}_n$ that is orthogonal to H. Moreover, the projection of y onto $\mathcal{R}\left\{\begin{bmatrix} H & \underline{h}_n \end{bmatrix}\right\}$ coincides with the projection of y onto $\mathcal{R}\left\{\begin{bmatrix} H & \tilde{\underline{h}}_n \end{bmatrix}\right\}$. If we denote this projection by $\hat{y}_{n+1}$, and using the orthogonality of $\tilde{\underline{h}}_n$ and H, we readily obtain

$$\hat{y}_{n+1} = \mathcal{P}_H y + \mathcal{P}_{\tilde{\underline{h}}_n} y = H\hat{x}_{n,N} + \mathcal{P}_{\tilde{\underline{h}}_n} y,$$

where $\mathcal{P}_{\tilde{\underline{h}}_n}$ is the projection matrix onto $\tilde{\underline{h}}_n$, and $H\hat{x}_{n,N}$ is the projection of y onto $\mathcal{R}(H)$. Since $\tilde{\underline{h}}_n$ is a single column,

$$\mathcal{P}_{\tilde{\underline{h}}_n} y = \frac{\tilde{\underline{h}}_n^* y}{\|\tilde{\underline{h}}_n\|^2}\,\tilde{\underline{h}}_n = \beta\tilde{\underline{h}}_n, \quad \text{say.}$$

Therefore,

$$\hat{y}_{n+1} = H\hat{x}_{n,N} + \beta\tilde{\underline{h}}_n = H\hat{x}_{n,N} + \beta[\underline{h}_n - Ha] = \begin{bmatrix} H & \underline{h}_n \end{bmatrix}\begin{bmatrix} \hat{x}_{n,N} - \beta a \\ \beta \end{bmatrix}.$$

But $\hat{y}_{n+1} = [H \;\; \underline{h}_n]\hat{x}_{n+1,N}$ and $\hat{x}_{n+1,N}$ is unique. Hence, we can identify

$$\hat{x}_{n+1,N} = \begin{bmatrix} \hat{x}_n - a\beta \\ \beta \end{bmatrix}, \tag{2.3.4}$$

which is the desired order-update relation.

As mentioned before, one can manipulate expression (2.3.3) and after some algebra arrive at the result (2.3.4) (*e.g.*, by using the block matrix inversion formula (A.1.7)). The geometric derivation, however, is conceptually much simpler, and the formulas have more significance.

Remark 1. We may mention that the simple ideas here can be carried much further — $\tilde{\underline{h}}_n$ can be regarded as the "new information" or "innovation" brought by the additional column $\underline{h}_n$; this concept will be developed much further in Ch. 4 and used extensively thereafter, though largely in a stochastic context. [In the deterministic context, pursuing this example will lead us to the so-called lattice filtering algorithms — see, *e.g.*, Sayed and Kailath (1994b) and Haykin (1996).] ◆

2.4 REGULARIZED LEAST-SQUARES PROBLEMS

A more general cost function that is often used instead of (2.2.9) is

$$J(x) = (x - x_0)^*\Pi_0^{-1}(x - x_0) + \|y - Hx\|_W^2. \tag{2.4.1}$$

This is still a quadratic cost function in the unknown vector x, but it includes the additional term $(x - x_0)^*\Pi_0^{-1}(x - x_0)$, where Π_0 is a given positive-definite (weighting) matrix and x_0 is also a given vector. Choosing $\Pi_0 = \infty I$ leads us back to the original expression (2.2.9), *viz.*,

$$\min_x \quad \|y - Hx\|_W^2. \tag{2.4.2}$$

One reason for using (2.4.1) is that we shall always get a unique least-squares solution $\hat{x}$, even when the matrix H is not full rank. And when H is full rank, including this extra term can improve the condition number of the matrix appearing in the normal equations, and thereby result in better numerical behavior.

Another point is that the availability of the extra parameters $\{\Pi_0, x_0\}$ allows us to incorporate additional a priori knowledge into the statement of the problem; different choices for Π_0 will indicate how confident we are about the closeness of the optimal solution $\hat{x}$ to a given vector x_0. Assume, for example, that we set $\Pi_0 = \epsilon I$, where ϵ is a very small positive number. Then the first term in the new cost function (2.4.1) becomes dominant, and we see that this will tend to force $\hat{x}$ to be close to x_0. In loose words, a "small" Π_0 reflects a high confidence that x_0 is a good guess for the optimal solution $\hat{x}$, while a "large" Π_0 indicates a high degree of uncertainty in the initial guess x_0.

To facilitate the solution of (2.4.1), we introduce the change of variables $x' = x - x_0$ and $y' = y - Hx_0$. Then our regularized least-squares problem can be written as

$$\min_{x'}\left[x'^*\Pi_0^{-1}x' + \|y' - Hx'\|_W^2\right], \tag{2.4.3}$$

which can be rewritten as

$$\min_{x'}\left\|\begin{bmatrix} 0 \\ y' \end{bmatrix} - \begin{bmatrix} \Pi_0^{-1/2} \\ H \end{bmatrix} x'\right\|_{I\oplus W}^2 \quad \text{where} \quad I \oplus W \triangleq \begin{bmatrix} I & 0 \\ 0 & W \end{bmatrix},$$

and $\Pi_0^{1/2}$ is a square-root of Π_0, *i.e.*, $\Pi_0 = \Pi_0^{1/2}\Pi_0^{*/2}$. This is now of the same form as our earlier minimization problem (2.4.2). Thus to obtain the least-squares solution $\widehat{x'}$ we need to project

$$\begin{bmatrix} 0 \\ y' \end{bmatrix} \quad \text{onto the column space of} \quad \begin{bmatrix} \Pi_0^{-1/2} \\ H \end{bmatrix}.$$

The orthogonality condition gives

$$\begin{bmatrix} \Pi_0^{-1/2} \\ H \end{bmatrix}^* \begin{bmatrix} I & 0 \\ 0 & W \end{bmatrix}\left(\begin{bmatrix} 0 \\ y' \end{bmatrix} - \begin{bmatrix} \Pi_0^{-1/2} \\ H \end{bmatrix}\widehat{x'}\right) = 0, \tag{2.4.4}$$

which reduces to the linear system of equations

$$\left[\Pi_0^{-1} + H^*WH\right](\hat{x} - x_0) = H^*W(y - Hx_0). \tag{2.4.5}$$

Table 2.1 Linear least-squares problems and solutions.

Optimization Problem	Solution
Given $\{H, y\}$, $H \in \mathbb{C}^{N\times n}$ full rank, $N \geq n$ solve $\min_x J(x)$, where $J(x) = \|y - Hx\|^2$	$\hat{x} = (H^*H)^{-1}H^*y$ minimum value is $J(\hat{x}) =$ $y^*\left[I - H(H^*H)^{-1}H^*\right]y = y^*\mathcal{P}_H^{\perp}y$
Given $\{x_0, y, H, \Pi_0, W\}$, with $\Pi_0 > 0$, $W > 0$, solve $\min_x J(x)$, where $J(x) =$ $\left[(x - x_0)^*\Pi_0^{-1}(x - x_0) + \|y - Hx\|_W^2\right]$	$\hat{x} = x_0 + \left[\Pi_0^{-1} + H^*WH\right]^{-1}H^*W\left[y - Hx_0\right]$ minimum value is $J(\hat{x}) =$ $(y - Hx_0)^*\left[W^{-1} + H\Pi_0H^*\right]^{-1}(y - Hx_0)$

As mentioned earlier, instead of requiring the invertibility of H^*WH, we now require the invertibility of the matrix $\left[\Pi_0^{-1} + H^*WH\right]$, which is guaranteed by the assumption $\Pi_0 > 0$.

The minimum value of (2.4.1) can be calculated to be

$$J(\hat{x}) = (y - Hx_0)^*\left[W^{-1} + H\Pi_0H^*\right]^{-1}(y - Hx_0), \tag{2.4.6}$$

and we can also write

$$\hat{x} = x_0 + \left[\Pi_0^{-1} + H^*WH\right]^{-1}H^*W\left[y - Hx_0\right], \tag{2.4.7}$$

and

$$J(\hat{x}) = (y - Hx_0)^*\left[W^{-1} + H\Pi_0H^*\right]^{-1}(y - Hx_0). \tag{2.4.8}$$

For ease of reference, we summarize the results of the previous discussions in Table 2.1.

2.5 AN ARRAY ALGORITHM: THE QR METHOD

The normal equations $(H^*H)\hat{x} = H^*y$ can be solved by the standard method of Gaussian elimination, but in our case since H^*H is Hermitian and positive-definite (when H is full rank), numerical analysts prefer to use its (unique) Cholesky decomposition, say

$$H^*H = \hat{R}^*\hat{R}, \tag{2.5.1}$$

where $\hat{R}$ is upper triangular with positive diagonal entries. Then we proceed as follows:

1. Solve the lower triangular system of equations $\hat{R}^*w = H^*y$ for w.
2. Solve the upper triangular system of equations $\hat{R}\hat{x} = w$ for $\hat{x}$.

However, this procedure can encounter numerical difficulties when implemented on a finite precision machine. This is because for ill-conditioned[2] data matrices H, numerical precision is lost when the matrix product H^*H is formed. Consider the simple example of a full rank matrix

$$H = \begin{bmatrix} 1 & 1 \\ 0 & \epsilon \\ 1 & 1 \end{bmatrix}, \tag{2.5.2}$$

where ϵ is a very small positive number that is of the same order of magnitude as the machine precision (this is essentially the smallest number that can be represented in finite precision arithmetic). But then

$$H^*H = \begin{bmatrix} 2 & 2 \\ 2 & 2+\epsilon^2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad \text{in finite precision,}$$

a singular matrix!

The best engineering approach to addressing the issue of ill-conditioned matrices is to reexamine the physical problem to see if through some reasonable changes (*e.g.*, using different variables, using different basis functions, using a different approximation) we can get a new set of better-conditioned equations. However, if we are (still) stuck with an ill-conditioned H, a safeguard is to avoid forming H^*H. But is it possible to solve the problem without doing this? Since, as just noted, solving the least-squares problem can be reduced to solving two triangular systems with coefficient matrices $\hat{R}$ and $\hat{R}^*$, *one may ask whether $\hat{R}$ cannot be directly obtained from H rather than via a Cholesky factorization of H^*H.* After all, there is no more information in H^*H than in the original matrix H, and in fact there may be less because (not to speak of finite precision effects) we cannot always recover H from H^*H. The fortunate answer to all this is that the key matrix $\hat{R}$ can be directly computed from H via a well-known decomposition in numerical analysis — the so-called reduced QR decomposition (see App. A):[3]

$$H = \hat{Q}\hat{R} \quad \text{where } H \in \mathbb{C}^{N\times n} \text{ and rank } H = n \leq N, \tag{2.5.3}$$

and $\hat{Q} \in \mathbb{C}^{N\times n}$ has orthonormal columns, and $\hat{R} \in \mathbb{C}^{n\times n}$ is upper triangular with positive diagonal entries. Now note that

$$H^*H = \hat{R}^*\hat{Q}^*\hat{Q}\hat{R} = \hat{R}^*\hat{R}, \tag{2.5.4}$$

since $\hat{Q}$ has orthonormal columns. But clearly this is also a Cholesky decomposition of H^*H. In other words, $\hat{R}$ can be found directly from H without forming H^*H!

[2] Ill-conditioned matrices are those that have a very large ratio of the largest to the smallest singular values. For further discussion, see any textbook on numerical linear algebra, *e.g.*, those cited at the end of App. A.

[3] The full QR decomposition is described in App. A (see also Prob. 2.19).

In fact, we gain a little more. Recall that we have to solve the triangular equations $\hat{R}\hat{x} = w$, where $\hat{R}^* w = H^* y$. But note that $w = \hat{R}^{-*}(\hat{Q}\hat{R})^* y = \hat{R}^{-*}\hat{R}^*\hat{Q}^* y = \hat{Q}^* y$, so that we can solve just the single equation

$$\hat{R}\hat{x} = \hat{Q}^* y. \tag{2.5.5}$$

Moreover, the minimum cost can be computed as

$$\|y - H\hat{x}\|^2 = \|y - \hat{Q}\hat{R}\hat{R}^{-1}\hat{Q}^* y\|^2 = \|(I - \hat{Q}\hat{Q}^*)y\|^2. \tag{2.5.6}$$

The above so-called QR method, first proposed by Householder (1958, pp. 72–73) , and the practical details worked out in Golub (1965) and Businger and Golub (1965), is now widely recommended. It has several nice features for both hardware and software implementations. However, it can be up to twice as slow as the normal equations method when $N \gg n$: the operation counts are (see, *e.g.*, Trefethen and Bau (1997, Ch. 11))

$$\left(Nn^3 + \frac{n^3}{3}\right) \text{ for the normal eqs method vs. } \left(2Nn^3 - \frac{2n^3}{3}\right) \text{ for the QR method.}$$

Therefore, when H is wellconditioned (which generally means that we have formulated a good model of the physical problem), the normal equations are quite adequate. As mentioned before, when H is ill-conditioned and not of full rank (or is close to being rank deficient), one should first try to see if reexamining the original problem and for example choosing different variables/basis functions will not lead to a better-conditioned system of equations. If this is not an option then it is recommended to use the QR method. Sometimes, it is recommended to use an even more stable, but significantly more expensive, method based on the singular value decomposition of H (see App. A). However, we shall not pursue such numerically oriented discussions further here — a vast literature is available on the numerical aspects of least-squares problems (see, *e.g.*, Björck (1996), Trefethen and Bau (1997, p. 84 and p. 143), Higham (1996), Stewart (1998), as well as the notes in Sec. 2.9).

Remark 2. As will be shown in detail in App. B, the recommended method for obtaining the matrix $\hat{R}$ is via a sequence of elementary unitary operations applied to the matrix H. In this sense, the QR method is an *array* algorithm of the type noted in Sec. 1.3.5. ◆

Remark 3. The normal equations approach arises naturally from application of the orthogonality condition for the optimum least-squares estimate. The underlying geometric picture also suggests a different so-called innovations approach, which will lead us directly to the QR method (see Sec. 4.3). ◆

Remark 4. For convenience we have used the terminology QR method, when in fact we have only used the reduced $\hat{Q}\hat{R}$ decomposition. The full QR decomposition can also be used — see Prob. 2.16, and in fact yields a little more information. We may note that the array algorithm of Sec. 1.3.5 has the form $\mathcal{A}_1\Theta = \mathcal{A}_2$, where Θ is unitary and $\mathcal{A}_2$ is lower triangular. By rewriting it as $\mathcal{A}_1^* = \Theta\mathcal{A}_2^*$, we see that the array algorithm of Sec. 1.3.5 can be interpreted as one of forming the QR decomposition of $\mathcal{A}_1^*$. ◆

Remark 5. The approach based on the given matrix H may be called a model-based (or sometimes, a first-order) approach, while the normal equations method may be called a covariance (Gramian) or second-order approach. The reasons for this terminology, and for this remark, will only be fully clear much later in this book, where in stochastic process estimation the normal equations approach will be seen to be analogous to the so-called Wiener filter methodology (based on power spectra/covariance functions), while the QR method is analogous to the (state-space) model-based Kalman filter approach. ◆

2.6 UPDATING LEAST-SQUARES SOLUTIONS: RLS ALGORITHMS

Since in least-squares problems the number of equations N may be much larger than the number of unknowns n and may actually increase as we gather more observations, the storage of the actual data may become a problem as well. These problems can be alleviated by using what are called recursive updating methods, which of course become especially useful (see the discussion below Eqs. (2.6.3)–(2.6.4)) when the data for the least-squares problem arises sequentially.

2.6.1 The RLS Algorithm

Thus suppose that, at step $i-1$, we have solved the least-squares problem for

$$H_{i-1}x \cong y_{i-1}, \tag{2.6.1}$$

where $H_{i-1} = \text{col}\{h_0, \ldots, h_{i-1}\} \in \mathcal{C}^{i\times n}$ and $y_{i-1} = \text{col}\{y(0), \ldots, y(i-1)\} \in \mathbb{C}^i$ are given.[4] [Note that the j-th row of H_{i-1} is denoted by h_j; earlier we used the notation $\{\underline{h}_j\}$ for the columns of H_{i-1}.] We shall consider the regularized form of the least-squares problem, though for notational simplicity we shall assume that $W = I$ and $x_0 = 0$, *i.e.*, we consider the problem

$$\min_x \left[x^*\Pi_0^{-1}x + \|y_{i-1} - H_{i-1}x\|^2\right]. \tag{2.6.2}$$

Now suppose that at step i we are given one extra row h_i and one extra scalar $y(i)$,[5] so that we must augment (2.6.1) with an additional equation, *i.e.*,

$$\underbrace{\begin{bmatrix} H_{i-1} \\ h_i \end{bmatrix}}_{H_i} x \cong \underbrace{\begin{bmatrix} y_{i-1} \\ y(i) \end{bmatrix}}_{y_i}, \tag{2.6.3}$$

and seek a minimizing solution under the new criterion

$$\min_x \left[x^*\Pi_0^{-1}x + \|y_i - H_i x\|^2\right]. \tag{2.6.4}$$

[4] Here, for clarity, we are using parentheses $(\cdot)$ to indicate the time dependency for scalar quantities, *e.g.*, $y(i-1)$. In the vector case we use subscripts, *e.g.*, y_{i-1}.

[5] The argument that follows is equally applicable for block updates where more than one row and more than one scalar entries are added to H_{i-1} and y_{i-1}, respectively.

At first sight it may seem that to find the new least-squares solution $\hat{x}_i$, we must re-solve the least-squares problem, which requires (i) storing all the previous data $\{h_j, y(j)\}_{j=0}^{i-1}$ and (ii) computing $(\Pi_0^{-1} + H_i^* H_i)^{-1} H_i^* y_i$. However, one may suspect that if we have already solved the least-squares problem (2.6.1) (whose solution is likewise denoted by $\hat{x}_{i-1}$), then we should be able to find $\hat{x}_i$ without needing to store the previous data and with much less computational effort. It turns out that this is the case, and not surprisingly this was already known to and used by Legendre (1805) and Gauss (1809). Pursuing this fact at this stage will have two useful consequences. First, it is a useful exercise in matrix manipulations; secondly, it will reveal an important connection between the deterministic least-squares problem and the optimum transient observer (or Kalman filter) problem studied in Ch. 1.

First, since $(\Pi_0^{-1} + H_{i-1}^* H_{i-1})$ is invertible, then so is

$$(\Pi_0^{-1} + H_i^* H_i) = (\Pi_0^{-1} + H_{i-1}^* H_{i-1}) + h_i^* h_i.$$

Then we can write

$$\begin{aligned} \hat{x}_i &= \left(\Pi_0^{-1} + H_i^* H_i\right)^{-1} H_i^* y_i, \\ &= \left(\Pi_0^{-1} + H_{i-1}^* H_{i-1} + h_i^* h_i\right)^{-1} [H_{i-1}^* y_{i-1} + h_i^* y(i)]. \end{aligned} \tag{2.6.5}$$

The above equation suggests that we define

$$P_i = (\Pi_0^{-1} + H_i^* H_i)^{-1}, \quad P_{-1} = \Pi_0, \tag{2.6.6}$$

so that

$$P_i^{-1} = P_{i-1}^{-1} + h_i^* h_i, \quad P_{-1}^{-1} = \Pi_0^{-1}. \tag{2.6.7}$$

Then using the (readily verified) matrix inversion identity (see App. A)

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}, \tag{2.6.8}$$

with $A = P_{i-1}^{-1}$, $B = h_i^*$, $C = 1$, and $D = h_i$, we obtain a recursive formula for P_i,

$$P_i = P_{i-1} - \frac{P_{i-1}h_i^* h_i P_{i-1}}{1 + h_i P_{i-1} h_i^*}, \quad P_{-1} = \Pi_0. \tag{2.6.9}$$

The point is that no matrix inversions are necessary. Each updating step requires only $O(n^2)$ computations (additions and multiplications), so that for N steps we need only $O(Nn^2)$ operations compared to $O(N^3)$ for a direct solution of the $N \times N$ normal equations. It is the special (recursive) structure of the P_i that has led to this reduction.

The recursion for P_i also gives one for updating the least-squares solutions:

$$\begin{aligned} \hat{x}_i &= \underbrace{\left(P_{i-1} - \frac{P_{i-1}h_i^* h_i P_{i-1}}{1 + h_i P_{i-1} h_i^*}\right)}_{P_i} [H_{i-1}^* y_{i-1} + h_i^* y(i)], \\ &= \underbrace{P_{i-1}H_{i-1}^* y_{i-1}}_{\hat{x}_{i-1}} - \frac{P_{i-1}h_i^*}{1 + h_i P_{i-1} h_i^*} h_i \underbrace{P_{i-1}H_{i-1}^* y_{i-1}}_{\hat{x}_{i-1}} + \\ &\quad P_{i-1}h_i^* \underbrace{\left(1 - \frac{h_i P_{i-1} h_i^*}{1 + h_i P_{i-1} h_i^*}\right)}_{\frac{1}{1+h_i P_{i-1}h_i^*}} y(i), \\ &= \hat{x}_{i-1} + \frac{P_{i-1}h_i^*}{1 + h_i P_{i-1} h_i^*}(y(i) - h_i \hat{x}_{i-1}). \end{aligned} \tag{2.6.10}$$

We summarize the discussion so far in the following lemma, where we have deliberately introduced some additional notation ($k_{p,i}$, $r_e(i)$) and terminology (Riccati). The resulting algorithm is a celebrated one, known as the *Recursive Least-Squares* (or RLS) algorithm.

Lemma 2.6.1 (Recursive Updating: The RLS Algorithm) *The solution $\hat{x}_i$ of problem (2.6.4) can be computed as*

$$\hat{x}_i = \hat{x}_{i-1} + k_{p,i}(y(i) - h_i \hat{x}_{i-1}), \quad \hat{x}_{-1} = 0, \tag{2.6.11}$$

where[6] $k_{p,i} = P_{i-1}h_i^* r_e^{-1}(i)$, $r_e(i) = 1 + h_i P_{i-1} h_i^*$, *and P_i satisfies the Riccati recursion*

$$P_i = P_{i-1} - P_{i-1}h_i^*(1 + h_i P_{i-1} h_i^*)^{-1} h_i P_{i-1}, \quad P_{-1} = \Pi_0, \tag{2.6.12}$$

and $\hat{x}_{i-1}$ is the regularized least-squares solution of (2.6.2). The effort required for one step of the recursion is $O(n^2)$ flops. ∎

Similarity to the Kalman Filter Recursions. The point of the additional notation is that we can now readily compare the above lemma with Thm. 1.2.1 on the Optimum Transient Observer. Doing so shows the striking fact that except *for the replacement of*

[6] Again for a variety of reasons, we shall often use small letters to denote vectors and scalars and capital letters to denote matrices. That is why we write here $k_{p,i}$ instead of $K_{p,i}$ and $r_e(i)$ instead of $R_{e,i}$.

random quantities by deterministic quantities, the solution given in Lemma 2.6.1 is *the same* as the (Kalman filter) solution for the special state-space model

$$\begin{cases} \mathbf{x}_{j+1} = \mathbf{x}_j, \quad \mathbf{x}_0 = \mathbf{x}, \\ \mathbf{y}(j) = h_j \mathbf{x}_j + \mathbf{v}(j), \end{cases} \tag{2.6.13}$$

with

$$E\mathbf{x}_0\mathbf{x}_0^* = \Pi_0 \text{ and } E\mathbf{v}(i)\mathbf{v}^*(j) = \delta_{ij}. \tag{2.6.14}$$

Note also that, if we define

$$\mathbf{y}_i \triangleq \text{col}\{\mathbf{y}(0), \mathbf{y}(1), \ldots, \mathbf{y}(i)\}, \quad \mathbf{v}_i \triangleq \text{col}\{\mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(i)\},$$

the state-space model (2.6.13) gives

$$\mathbf{y}_i = H_i \mathbf{x} + \mathbf{v}_i, \tag{2.6.15}$$

which is exactly the equation obeyed by the deterministic quantities y_i, x, and v_i — see (2.1.2).

We see that there is a close connection between Kalman filtering and the minimization of certain deterministic quadratic forms; there is a simple underlying reason for this equivalence, which we shall discuss in Sec. 3.5. We note also that this connection can be generalized and used to solve several deterministic control and adaptive filtering problems by showing their equivalence to certain stochastic state-space estimation problems.

But why not a direct approach to these deterministic problems? The point is that the matrix algebra used in the above derivation can rapidly get relatively involved and more complicated as we work with more general quadratic forms (corresponding for example to more general state equations than $\mathbf{x}_{i+1} = \mathbf{x}_i$). However, the equivalence between stochastic and deterministic minimization will allow us to obtain simpler derivations of general recursive algorithms and also various alternative forms, such as the fast algorithms and array algorithms mentioned in Sec. 1.3.

In fact, we shall illustrate this right now by showing how to directly obtain a recursive version of the QR approach of Sec. 2.5 (*cf.* Sayed and Kailath (1994b)) [A direct algebraic derivation of the resulting algorithm is pursued in Prob. 2.17.]

2.6.2 An Array Algorithm for RLS

Writing down the Kalman equations that correspond to the special state-space model (2.6.13)–(2.6.14) from Thm. 1.2.1, we obtain

$$\begin{aligned} \hat{\mathbf{x}}_{i+1} &= \hat{\mathbf{x}}_i + k_{p,i}[\mathbf{y}(i) - h_i\hat{\mathbf{x}}_i], \quad \hat{\mathbf{x}}_0 = 0, \\ k_{p,i} &= P_i h_i^* r_e^{-1}(i), \quad r_e(i) = 1 + h_i P_i h_i^*, \\ P_{i+1} &= P_i - P_i h_i^*(1 + h_i P_i h_i^*)^{-1} h_i P_i, \quad P_0 = \Pi_0. \end{aligned}$$

Table 2.2 Correspondence between RLS and Kalman variables.

Kalman problem	RLS problem
$\mathbf{x}$	x
$\mathbf{y}(i)$	$y(i)$
$\mathbf{e}(i)$	$e_a(i)$
h_i	h_i
Π_0	Π_0
P_i	P_{i-1}
$k_{p,i}$	$k_{p,i}$
$r_e(i)$	$r_e(i)$

Comparing these equations with the RLS equations in Lemma 2.6.1 we see that the only difference is in the time indexing of the variables $\{P_i, \hat{\mathbf{x}}_i, \hat{x}_i\}$. In the RLS solution, the time subscripts of $\{P_i, \hat{x}_i\}$ lag behind by one relative to the above Kalman equations. This is merely a consequence of our choice of notation. Had we written P_{i+1} instead of P_i in the definition (2.6.6), with initial condition P_0 rather than P_{-1}, and had we denoted the solution of (2.6.3) by $\hat{x}_{i+1}$ rather than $\hat{x}_i$, then both sets of equations (RLS and Kalman) would have been identical. The convention we adopted, however, is the common one in the literature. In any case, this comparison shows that we can translate a Kalman type equation to an RLS type equation by making the identifications shown in Table 2.2, where we further introduced the so-called a priori estimation error, $e_a(i) = y(i) - h_i\hat{x}_{i-1}$, for the RLS problem.

We can now invoke the special array algorithm derived in Prob. 1.6 for Kalman filtering and use the above correspondences to write for the RLS problem the following array equations

$$\begin{bmatrix} P_{i-1}^{-*/2} & h_i^* \\ \hat{x}_{i-1}^* P_{i-1}^{-*/2} & y^*(i) \\ 0 & 1 \end{bmatrix} \Theta = \begin{bmatrix} P_i^{-*/2} & 0 \\ \hat{x}_i^* P_i^{-*/2} & e_a^*(i) r_e^{-*/2}(i) \\ h_i P_i^{1/2} & r_e^{-*/2}(i) \end{bmatrix}, \tag{2.6.16}$$

with $P_{-1}^{-1/2} = \Pi_0^{-1/2}$ and $\hat{x}_{-1} = 0$. These equations are known as the QR algorithm in the adaptive filtering literature (see, *e.g.*, Haykin (1996) , Sayed and Kailath (1994b), and also Prob. 2.17).

2.7 DOWNDATING LEAST-SQUARES SOLUTIONS

We shall now introduce a new issue, *viz.*, that of *removing* the effect of earlier data, which is referred to as *downdating*. To this end, consider once more the possibly inconsistent system of linear equations

$$H_i x = \underbrace{\begin{bmatrix} h_0 \\ H_{1:i} \end{bmatrix}}_{H_{0:i}} x \cong \begin{bmatrix} y(0) \\ y_{1:i} \end{bmatrix} = y_i, \tag{2.7.1}$$

where $H_i \in \mathbb{C}^{(i+1)\times n}$ and $y_i \in \mathcal{C}^{(i+1)}$ are given. We have also denoted H_i by $H_{0:i}$ to indicate that it consists of the rows h_0 through h_i.

Suppose now that we wish to remove the effect of the *first* of the above equations, *i.e.*, we are interested in the regularized least-squares solution of the new system of equations

$$H_{1:i}x \cong y_{1:i}. \tag{2.7.2}$$

We shall show that if we have already determined $\hat{x}_{0:i}$ ($= \hat{x}_i$, the least-squares solution of (2.7.1)), then the least-squares solution to (2.7.2), denoted by $\hat{x}_{1:i}$, can be found with much less computational effort than by re-solving (2.7.2). Using Lemma 2.2.2 we have

$$\begin{aligned}\hat{x}_{1:i} &= [\Pi_0^{-1} + H_{1:i}^* H_{1:i}]^{-1} H_{1:i} y_{1:i},\\ &= [\Pi_0^{-1} + H_{0:i}^* H_{0:i} - h_0^* h_0]^{-1}[H_{0:i}^* y_i - h_0^* y(0)].\end{aligned}$$

This suggests the definitions

$$P_{0:i} \triangleq (\Pi_0^{-1} + H_{0:i}^* H_{0:i})^{-1}, \qquad P_{1:i} \triangleq (\Pi_0^{-1} + H_{1:i}^* H_{1:i})^{-1}. \tag{2.7.3}$$

From (2.7.3) we clearly have

$$P_{1:i}^{-1} = P_{0:i}^{-1} - h_0^* h_0, \tag{2.7.4}$$

so that using again the matrix inversion lemma (2.6.8) we can write

$$P_{1:i} = P_{0:i} - \frac{P_{0:i} h_0^* h_0 P_{0:i}}{-1 + h_0 P_{0:i} h_0^*}.$$

The term $(-1 + h_0 P_{0:i} h_0^*)$ that appears in the denominator can be shown to be strictly negative (and hence nonzero). Indeed, equation (2.7.4) shows that $P_{1:i}^{-1}$ can be interpreted as the Schur complement with respect to the rightmost diagonal entry of the following extended matrix

$$\begin{bmatrix} P_{0:i}^{-1} & h_0^* \\ h_0 & 1 \end{bmatrix}.$$

Since both $P_{0:i}$ and $P_{1:i}$ are positive-definite (by definition, see (2.7.3)), we see (using the last result in Sec. A.1) that the above extended matrix is also positive-definite. Hence, its Schur complement with respect to the leftmost diagonal block must also be positive, *i.e.*, $1 - h_0 P_{0:i}^{-1} h_0^* > 0$. Thus it holds that

$$-1 + h_0 P_{0:i}^{-1} h_0^* \;<\; 0, \tag{2.7.5}$$

which confirms our claim. [We should mention that this conclusion does not hold automatically in the nonregularized case, where $\Pi_0 = \infty I$. In this case, it must be imposed as a condition in order to guarantee an invertible matrix $H_{1:i}^* H_{1:i}$ — see Prob. 2.12.]

Returning to the expression for $\hat{x}_{1:i}$ we now find that

$$\begin{aligned}\hat{x}_{1:i} &= P_{1:i}[H_{0:i}^* y_i - h_0^* y(0)],\\ &= \left[P_{0:i} - \frac{P_{0:i} h_0^* h_0 P_{0:i}}{-1 + h_0 P_{0:i} h_0^*}\right][H_{0:i}^* y_i - h_0^* y(0)],\\ &= \underbrace{P_{0:i} H_{0:i}^* y_i}_{\hat{x}_{0:i}} - \frac{P_{0:i} h_0^*}{-1 + h_0 P_{0:i} h_0^*} h_0 \underbrace{P_{0:i} H_{0:i}^* y_i}_{\hat{x}_{0:i}} - P_{0:i} h_0^* \underbrace{\left[1 - \frac{h_0 P_{0:i} h_0^*}{-1 + h_0 P_{0:i} h_0^*}\right]}_{\frac{-1}{-1+h_0 P_{0:i} h_0^*}} y(0),\\ &= \hat{x}_{0:i} + \frac{P_{0:i} h_0^*}{-1 + h_0 P_{0:i} h_0^*}[y(0) - h_0 \hat{x}_{0:i}].\end{aligned}$$

We summarize our results in the following lemma, where we have again introduced some suggestive notation.

Lemma 2.7.1 (Recursive Downdating) *Suppose $\hat{x}_{1:i}$ is the regularized least-squares solution to the overdetermined system of linear equations $H_{1:i}x \cong y_{1:i}$, obtained by deleting the first equation in the system (2.7.1). Then*

$$\hat{x}_{1:i} = \hat{x}_{0:i} + k_{p,1:i}(y(0) - h_0 \hat{x}_{0:i}), \tag{2.7.6}$$

where $k_{p,1:i} = P_{0:i} h_0^* r_e^{-1}(1:i)$, $r_e(1:i) = -1 + h_0 P_{0:i} h_0^*$, and

$$P_{1:i} = P_{0:i} - \frac{P_{0:i} h_0^* h_0 P_{0:i}}{-1 + h_0 P_{0:i} h_0^*}. \tag{2.7.7}$$

■

As with updating, this solution turns out to be the same, apart from the property (2.7.5), as the (Kalman filter) solution of Thm. 1.2.1 for the state-space model

$$\begin{cases} \mathbf{x}_{j+1} = \mathbf{x}_j, \quad \mathbf{x}_0 = \mathbf{x}, \\ \mathbf{y}(j) = h_j \mathbf{x}_j + \mathbf{v}(j), \end{cases} \tag{2.7.8}$$

but with

$$E\mathbf{x}_0\mathbf{x}_0^* = \Pi_0 \;\text{ and }\; E\mathbf{v}(i)\mathbf{v}^*(j) = r(i)\delta_{ij}, \quad r(i) = -1. \tag{2.7.9}$$

This is intriguing. Since $r(i)$ here is negative, it cannot be thought of as the variance of some random variable $\mathbf{v}(i)$. However, due to the formal similarities between the solutions of Lemmas 2.6.1 and 2.7.1, one can speculate whether Kalman filtering and least-squares problems can be appropriately generalized. It turns out that this is indeed the case, and in the monograph (Hassibi, Sayed, and Kailath (1999)) we show how a parallel theory for estimation in indefinite metric spaces can be developed. This generalized theory shows how the results in this book can be extended fairly readily to the so-called $\mathcal{H}_\infty$ problems (*cf.* Secs. 1.6.3 and 1.6.5).

2.8 SOME VARIATIONS OF LEAST-SQUARES PROBLEMS

There are other variations of the least-squares criterion that have been proposed in the literature with the intent of addressing some of the deficiencies of least-squares-based designs.

Returning to the formulation of the least-squares problem in Sec. 2.1, we see from the measurement equation (2.1.2) that the observation vector y is assumed to be corrupted by additive noise while the data matrix H itself is assumed to be known exactly. Consequently, least-squares solutions can be senstive to errors in H. To see this, assume that we determine an optimal least-squares estimate $\hat{x}$ under the assumption that the given observation vector y has been generated by a full rank matrix H, say $y = Hx + v$. Then $\hat{x} = (H^*H)^{-1}H^*y$ and the resulting minimum residual norm will be $\|y - H\hat{x}\|$. But what if the observation vector y has not been generated by H but by a perturbed version of H, say by $y = (H + \Delta H)x + v$ for some unknown ΔH? In this case, if we continue to use the above construction for $\hat{x}$, then the corresponding residual norm will be $\|y - (H + \Delta H)\hat{x}\|$. This residual satisfies, in view of the triangle inequality of norms,

$$\|y - (H + \Delta H)\hat{x}\| \leq \underbrace{\|y - H\hat{x}\|}_{\text{LS residual}} + \underbrace{\|\Delta H\ \hat{x}\|}_{\text{additional term}} .$$

The upper bound on the new residual norm is tight only when the vector $\Delta H\hat{x}$ happens to be collinear with the original residual vector $y - H\hat{x}$. It is thus not difficult to envision perturbations ΔH that can degrade the performace of the least-squares solution. Examples to this effect can be found in (Sayed, Nascimento, and Chandrasekaran (1998)).

Perturbation errors in the data are common in practice and they can be due to several factors including the approximation of complex models by simpler ones, the occurrence of experimental errors when collecting data, or even the presence of unknown or unmodeled effects.

2.8.1 The Total Least-Squares Criterion

One criterion that has been devised to deal with data errors in H and y is the so-called *Total Least-Squares* (TLS) method, also known as *orthogonal regression* or *errors-in-variables* method in statistics and system identification. The method is well surveyed in the monograph of Van Huffel and Vandewalle (1991).

Given (H, y), the TLS problem seeks a matrix $\hat{H}$ and a vector $\hat{y}$ in its range space by solving

$$\min_{\hat{H}, \hat{y} \in \mathcal{R}(\hat{H})} \|[H\ \ y] - [\hat{H}\ \ \hat{y}]\|_F^2, \tag{2.8.1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of its argument (see App. A). The resulting optimal solution $\hat{H}_o$ is regarded as an approximation for H, which is then used along with the optimal $\hat{y}_o$ to determine $\hat{x}$ by solving the consistent linear system of equations $\hat{H}_o\hat{x} = \hat{y}_o$.

In the so-called nondegenerate case (more general cases are treated in Van Huffel and Vandewalle (1991)), the TLS solution turns out to admit an interpretation in terms of a regularized least-squares solution. More specifically, assume H is full rank with smallest singular value σ_n. Assume also $[H\ \ y]$ is full rank with smallest singular value $\bar{\sigma}_{n+1}$. When $\bar{\sigma}_{n+1} < \sigma_n$, a unique solution to the TLS problem exists and it can be expressed in the form

$$\hat{x} = (H^*H - \bar{\sigma}_{n+1}^2 I_n)^{-1}H^*y, \tag{2.8.2}$$

where I_n is the identity matrix of size $n \times n$.

Comparing with (2.4.7) we see that the TLS solution can be regarded as the solution of a *regularized* cost function but one with a *negative-definite* matrix Π_0 (compare with (2.4.1) where the positive-definite matrix Π_0^{-1} is now replaced by the negative-definite matrix $-\bar{\sigma}_{n+1}^2 I_n$):

$$\min_x \left[\ \|y - Hx\|^2 - \bar{\sigma}_{n+1}^2\|x\|^2\ \right].$$

This brings up connections with our earlier discussions on the $\mathcal{H}_\infty$ problem of Sec. 1.6.3 and on the simple downdating problem of Sec. 2.7, where we also encountered negative-definite weighting matrices (or coefficients) as well as certain solvability conditions. Some of these connections are discussed in the article (Sayed, Hassibi, and Kailath (1996)).

2.8.2 Criteria with Bounds on Data Uncertainties

The TLS approach does not impose any explicit bound on how far the approximation $\hat{H}_o$ can be from H. In fact, the spectral norm (defined in App. A) of the correction $(H - \hat{H}_o)$ turns out to be determined by the smallest singular value of $[H\ \ y]$. This singular value can be relatively large even when H is known almost precisely, *e.g.*, when y is sufficiently far from the column space of H. In this case, the TLS method may end up overcorrecting H.

In Chandrasekaran et al. (1998) and Sayed et al. (1998, 1999), new cost functions have been formulated that explicitly incorporate a priori bounds on perturbations to the data matrix H. In so doing, the resulting so-called BDU problems (with BDU standing for Bounded Data Uncertainties) guarantee that the correction to H is never beyond what is assumed by the given bounds and, consequently, the effect of the uncertainties will not be overemphasized. Let x be again a column vector of unknown parameters, y a vector of measurements, and H a known full rank matrix. The matrix H represents nominal data in the sense that the true matrix that relates y to x is not H itself but rather a perturbed version of H, say

$$y = (H + \Delta H)x + v. \tag{2.8.3}$$

The perturbation ΔH is not known. What is known is a bound on how far the true matrix $(H + \Delta H)$ can be from the assumed nominal value H, say $\|\Delta H\| \leq \eta$ (in terms of the spectral norm of ΔH, or equivalently, its maximum singular value).

The standard least-squares criterion (2.1.3) would seek to recover x from y by relying on the available nominal data H, and without taking into account the fact that the true data matrix may not be H itself but lies around H within a ball of size η. The TLS criterion, on the other hand, is sensitive to possible perturbations in H and therefore tries to replace it with an approximation $\hat{H}_o$ before seeking to estimate x. It, however, does not explicitly incorporate the a priori bound η into the solution. In this way, there is no guarantee that the approximation $\hat{H}_o$ will lie within the ball of size η and, consequently, H may end up being overcorrected.

There are several ways in which the a priori bound η can explicitly be incorporated into the problem formulation, as well as several ways for modeling data uncertainties. Some of these so-called BDU criteria are listed below.

The most basic problem is one that corresponds to worst-case BDU estimation. It requires that we solve

$$\min_x \max_{\|\Delta H\| \le \eta} \|y - (H + \Delta H)\ x\|. \tag{2.8.4}$$

That is, we seek a solution $\hat{x}$ that performs "best" in the worst possible scenario. This formulation can be regarded as a constrained two-player game problem, with the designer trying to pick an x that minimizes the residual norm while the opponent ΔH tries to maximize the residual norm. The game problem is constrained since it imposes a limit on how large (or how damaging) the opponent ΔH can be. We may remark that in related work, a formulation similar to (2.8.4), with bounds on the Frobenius norm of ΔH rather than its spectral norm, was also posed and solved by El-Ghaoui and Lebret (1997). Their solution is based on (more costly) convex optimization and LMI (linear matrix inequality) techniques.

It turns out that the solution of (2.8.4) can be found at essentially the same computational cost as standard least-squares solutions, thus making the BDU problem attractive for practical use. The solution was obtained algebraically in Chandrasekaran et al. (1997, 1998) and we shall comment on it later. Interestingly enough, a geometric formulation can also be pursued for such problems, along the lines of least-squares theory, and this is developed in Sayed et al. (1998). In particular, the orthogonality principle of least-squares problems (*cf.* Sec. 2.3) can be extended to the BDU context and can be shown to lead to useful insights about the nature of the solution.

We can also consider several other variations of the BDU problem. For example, in some applications, we might be uncertain only about part of the data matrix while the remaining data is known exactly. This leads to a formulation of a BDU problem with partial uncertainties, say

$$\min_x \left(\max_{\|\Delta H_2\| \le \eta_2} \left\| \begin{bmatrix} H_1 & H_2 + \Delta H_2 \end{bmatrix} x - y \right\| \right). \tag{2.8.5}$$

The BDU formulation can also handle situations with different levels of uncertainties in different parts of the data matrix. This can be done via the solution of a BDU problem

with multiple uncertainties (*e.g.*, Sayed et al. (1998) and Sayed and Chandrasekaran (1998, 2000)), say

$$\min_x \max_{\left(\begin{matrix} \|\Delta H_j\| \le \eta_j \\ 1 \le j \le K \end{matrix}\right)} \left\| \begin{bmatrix} H_1 + \Delta H_1 & \dots & H_K + \Delta H_K \end{bmatrix} x - y) \right\|. \tag{2.8.6}$$

Here, the $\{H_j\}$ denote submatrices (column-wise) of H. For control problems, it is useful to introduce the following BDU formulations (see Sayed et al. (1998) and Sayed and Nascimento (1999)):

$$\min_x \left(\max_{\|\Delta H\| \le \eta,\ \|\Delta y\| \le \beta} \|(H + \Delta H)x - (y + \Delta y)\|^2 \ + \ \rho \|x\|^2 \right), \tag{2.8.7}$$

and

$$\min_x \max_{\substack{\|\Delta H\| \le \eta \\ \|\Delta y\| \le \beta}} [(H + \Delta H)x - (y + \Delta y)]^* W [(H + \Delta H)x - (y + \Delta y)] + x^* Q x, \tag{2.8.8}$$

where we now allow for uncertainties in H and y, and also employ weighting factors W, Q, and ρ (with positive-definite matrices W and Q, and a positive scalar ρ).

We can also formulate cost functions that treat data uncertainties multiplicatively rather than additively,

$$\min_x \max_{\|\Delta H\| \le \eta} \|(I + \Delta H)Hx - y\|, \tag{2.8.9}$$

and also treat weight uncertainties,

$$\min_x \max_{\|\Delta W\| \le \eta_w} \|(W + \Delta W)(Hx - y)\|. \tag{2.8.10}$$

This latter cost is a variation of the weighted least-squares criterion in which the uncertainty is taken to be in the weight matrix itself.

Solutions of several of the above problems, along with examples and geometric formulations, can be found in the aforementioned references. Here we comment only on the solution of the worst-case BDU formulation,

$$\min_x \max_{\|\Delta H\| \le \eta} \|y - (H + \Delta H)\ x\|.$$

It can be shown that this problem is equivalent to solving

$$\min_x \ (\eta \|x\| \ + \ \|y - Hx\|). \tag{2.8.11}$$

This equivalent problem is deceptively similar, but significantly distinct, from a regularized least-squares problem of the form,

$$\min_x \ (\gamma\|x\|^2 \ + \ \|y - Hx\|^2),$$

for some positive scalar γ, and where the *squared* Euclidean norms $\{\|x\|^2, \|y - Hx\|^2\}$ are used rather than the *norms* themselves as in (2.8.11). For this reason, the solution of such (sum of distances) problems is not as immediate as regularized LS problems.

Still, it turns out that the solution of (2.8.11) is always unique, except in special degenerate cases. When $\eta \geq \|H^*y\|/\|y\|$, *i.e.*, when the uncertainties are relatively large, the solution is $\hat{x} = 0$. Otherwise, when $y \notin \mathcal{R}(H)$, the solution is given by

$$\hat{x} = \left(H^*H + \hat{\alpha} I_n\right)^{-1} H^*y, \tag{2.8.12}$$

where $\hat{\alpha}$ is the unique positive root of a certain nonlinear equation; $\hat{\alpha}$ can be determined, for example, by employing a bisection-type method requiring $O\left(n \log \frac{\hat{\alpha}}{\epsilon}\right)$ flops, where ϵ is the desired precision.

The above expression for $\hat{x}$ has the same form of a regularized solution (2.4.7), except that the regularization parameter $\hat{\alpha}$ is not given a priori but is determined by the algorithm. In this sense, we say that the new problem formulation performs *automatic* regularization. Further remarks are provided in the notes in the next section.

2.9 COMPLEMENTS

Sec. 2.1. The Deterministic Least-Squares Criterion. Here is how Gauss introduces the deterministic least-squares problem (Stewart (1995, pp. 31, 33)):

"Suppose a quantity that depends on another unknown quantity is estimated by an observation that is not absolutely precise. If the unknown is calculated from this observation, it will also be subject to error, and there will be no freedom in this estimate of it. But if *several* quantities depending on the same unknown have been determined by inexact observations, we can recover the unknown either from one of the observations or from any of an infinite number of combinations of the observations. Although the value of an unknown determined in this way is always subject to error, there will be less error in some combinations than in others.

A similar situations occurs when we observe several quantities depending on several unknowns. The number of observations may be equal to, less than, or greater than the number of unknowns. In the first case the problem is well determined; in the second it is indeterminate. In the third case the problem is (generally speaking) overdetermined, and the observations can be combined in an infinite number of ways to estimate the unknowns. One of the most important problems in the application of mathematics to the natural sciences is to choose the best of these many combinations, *i.e.*, the combination that yields values of the unknowns that are least subject to the errors."

The romantic story of how Gauss used his methods to predict where the planetoid Ceres would appear after having been lost to sight after 41 days of observation is nicely told in Hall (1970, pp. 61–63). The excitement was in the fact that the great philosopher

Hegel claimed to have proved by pure logic that there were exactly seven planets; then on Jan. 1, 1801, an Italian astronomer discovered a moving object in the constellation of Aries, and the question was whether the new heavenly body was a planet or a comet?

Sec. 2.3. A Geometric Formulation. The natural geometric formulation of the deterministic least-squares problem is one of the major reasons for including this chapter. In the next chapter this intuition is to be carried over to the somewhat less familiar geometric formulation of stochastic least-squares problems. More on vectors, linear vector spaces, and inner product spaces, can be found in standard textbooks on linear algebra, several of which are mentioned at the end of App. A.

In this book, we shall also use, without much comment, infinite dimensional vector spaces and inner product spaces — a textbook that discusses the finer issues that we ignore (*e.g.*, closure) is a classic book by Luenberger (1969).

Sec. 2.2. The Classical Solutions. The effective computational solution of systems of linear equations — either exactly or in the least-squares sense — is a major part of the field of Numerical Linear Algebra. There are numerous textbooks in this area, at various levels of sophistication, and several are cited at the end of App. A. We may mention that Trefethen and Bau (1997) follows a lecture format and is a good first introduction to numerical analysis. Higham (1996) has extensive historical notes (and apt quotations); the book by Björck (1996) is the most detailed as far as the least-squares problem goes.

Sec. 2.5. The QR Method: An Array Algorithm. The origin of the QR method goes back to Householder (1953, pp. 72–73), one of the pioneers of numerical analysis. However, it first gained attention through a paper by Golub (1965), which according to Björck (1996, p. 64) "started a new epoch in least-squares methods." As noted in the remarks in Sec. 2.5, in the engineering literature the QR method can be interpreted as using the innovations approach (see Sec. 4.3), and as a model-based rather than a second-order approach (*i.e.*, using H directly rather than going via H^*H). These remarks will become clearer as we progress.

Sec. 2.8. Some Variations of Least-Squares Problems. We have only presented a glimpse of an area that is under active development. Compared to regularized least-squares methods, the BDU formulations lead to automatic procedures for the selection of the regularization parameters. These formulations are particularly suitable when a priori bounds on the sizes of the uncertainties are available, in which case they can guarantee a robust performance for uncertain models that lie within a pre-specified domain. The article by Sayed, Nascimento, and Chandrasekaran (1998) provides a treatment of this area of research with emphasis on geometric formulations and with several application examples from the areas of image restoration, image separation, array signal processing, and control. Further details and alternative formulations can be found in Chandrasekaran et al. (1998), Sayed and Chandrasekaran (1998, 2000), Nascimento and Sayed (1999), and Sayed and Nascimento (1999). The latter reference introduces a very general design criterion and also studies the case of structured uncertainties in the data. The article Sayed (2000) develops a framwork for state-space estimation with uncertain models. The resulting recursions turn out to shsare many important features with the different variants of Kalman filtering algorithms studied in this book.

PROBLEMS

2.1 **(Consistent equations)** Show that if the overdetermined system (2.1.1) is consistent, *i.e.*, if $y \in \mathcal{R}(H)$, then the least-squares solution gives an exact solution.

2.2 **(Special least-squares problems)** Consider the least-squares problem

$$\min_x \ \|y - Hx\|^2.$$

Comment on the solution when $y \in \mathcal{N}(H)$, and when $y \in \mathcal{N}(H^*)$.

2.3 **(Recursive least-squares with nonzero x_0)** Verify, for example, by using the change of variables suggested prior to (2.4.3), that if we pose the optimization problem

$$\min_x \left[(x - x_0)^* \Pi_0^{-1} (x - x_0) + \|y_i - H_i x\|^2\right],$$

with a nonzero x_0, then the same recursive solution of Lemma 2.6.1 still holds except for the value of the initial guess $\hat{x}_{-1}$, which should now be taken as $\hat{x}_{-1} = x_0$.

2.4 **(An expression for the minimum cost)** Consider the regularized least-squares problem $\min_x (x-x_0)^*\Pi_0^{-1}(x-x_0) + \|y-Hx\|^2$. Denote its solution by $\hat{x}$ and the corresponding minimum cost by $J(\hat{x})$. Show that $J(\hat{x}) = (y - Hx_0)^*(y - H\hat{x})$. [This result provides an expression for the minimum cost that is not explicitly dependent on the regularization matrix Π_0 and, therefore, it also holds for the pure least-squares problem.]

2.5 **(Time update for the minimum cost)** Refer to the RLS algorithm of Sec. 2.6 and define the following so-called a priori and a posteriori estimation errors, $e_a(i) \triangleq y(i) - h_i\hat{x}_{i-1}$ and $e_p(i) \triangleq y(i) - h_i\hat{x}_i$, respectively.

(a) Show that $e_p(i) = r_e^{-1}(i)e_a(i)$, and also that $r_e^{-1}(i) = 1 - h_i P_i h_i^*$.

(b) Let $J(\hat{x}_i)$ denote the minimum value of $\min_x \left[x^*\Pi_0^{-1}x + \|y_i - H_i x\|^2\right]$. Show that

$$J(\hat{x}_i) = J(\hat{x}_{i-1}) + e_a^*(i)e_p(i) \ = \ J(\hat{x}_{i-1}) + r_e^{-1}(i)|e_a(i)|^2.$$

2.6 **(Exponentially weighted RLS)** Let $\hat{x}_i$ denote the solution of the regularized and exponentially-weighted least-squares problem

$$\min_x \left[(x - x_0)^* \left[\lambda^{-(i+1)}\Pi_0\right]^{-1} (x - x_0) + \sum_{j=0}^{i} \lambda^{i-j} |y(j) - h_j x|^2 \right],$$

where x_0 is given, Π_0 is a positive-definite matrix, and λ is a positive scalar, usually $0 \ll \lambda < 1$. Following the derivation of the RLS algorithm in Sec. 2.6.1, show that $\hat{x}_i$ can be updated recursively as follows:

$$\begin{aligned}
\hat{x}_i &= \hat{x}_{i-1} + k_{p,i}[y(i) - h_i\hat{x}_{i-1}], \quad \hat{x}_{-1} = x_0, \\
k_{p,i} &= \lambda^{-1}P_{i-1}h_i^* r_e^{-1}(i), \\
r_e(i) &= 1 + \lambda^{-1}h_i P_{i-1}h_i^*, \\
P_i &= \lambda^{-1}[P_{i-1} - k_{p,i}h_i P_{i-1}], \quad P_{-1} = \Pi_0.
\end{aligned}$$

Remark. The scalar λ is known as the forgetting factor in the adaptive filtering literature (see, *e.g.*, Haykin (1996)). Since $\lambda < 1$, the factor λ^{i-j} in the cost function weights past data (*i.e.*, those that occur at time instants j that are sufficiently far from i) less heavily than more recent data (*i.e.*, those that occur at time instants j relatively close to i). This feature enables an adaptive algorithm to respond to variations in data statistics by "forgetting" data from the remote past. ◆

2.7 **(The QR method)** For the matrix H in (2.5.2), explicitly find the Cholesky decomposition of H^*H and the $\hat{Q}\hat{R}$ decomposition of H. Show that though setting $\epsilon^2 = 0$ causes the normal equations approach to break down, the QR method still allows a solution.

2.8 **(A useful bound)** Let $\hat{x}$ denote the unique solution of the regularized least-squares problem

$$\min_x \left[\gamma\|x\|^2 \ + \ \|y - Hx\|^2\right],$$

where γ is a finite positive number, $H \in \mathbb{C}^{N\times n}$ is assumed to have full rank with $N > n$, and y does not lie in the range space of H. Assume further that $H^*y \neq 0$ so that $\hat{x}$ is nonzero. Define $\eta^2 = \gamma^2\|\hat{x}\|^2/\|y - H\hat{x}\|^2$, and show that $\eta^2 < \|H^*y\|^2/\|y\|^2$.

Remark. This bound can be used to show that, under the given conditions on the data $\{H, y\}$, the nonzero solution of a regularized least-squares problem is also the solution of a BDU estimation problem of the form (2.8.4) — see Sayed et al. (1998). ◆

2.9 **(Two matrix least-squares problems)** We write $\|A\|_F$ to denote the Frobenius norm of a matrix A (*cf.* App. A).

(a) Consider matrices $Y \in \mathbb{C}^{N\times m}$, $H \in \mathbb{C}^{N\times n}$, and $X \in \mathbb{C}^{n\times m}$. Assume H has full column rank. Show that the minimizer of $\|Y - HX\|_F$ is $\hat{X} = (H^*H)^{-1}H^*Y$.

(b) Consider matrices $H_i \in \mathbb{C}^{N\times m}$, $i = 1, 2, \ldots, p$, and introduce the scalars $a_{ij} = \mathrm{Tr}(H_i^*H_j)$ and $b(j) = \mathrm{Tr}(H_j^*Y)$, as well as the $p \times p$ matrix $A = [a_{ij}]$ and the $p \times 1$ column vector $b = \mathrm{col}\{b(1), \ldots, b(p)\}$. Show that the scalars $\hat{x}(i)$ that minimize $\left\|Y - \sum_{i=1}^p x(i)H_i\right\|_F$ are given by $\hat{x} = A^{-1}b$, where $\hat{x} = \mathrm{col}\{\hat{x}(1), \ldots, \hat{x}(p)\}$.

2.10 **(A system identification problem)** Suppose we are able to measure the state vectors x_k and the control input sequence u_k of a system governed by the linear time-invariant state equation $x_{k+1} = Fx_k + Gu_k + v_k$, $k \geq 0$, where F and G are unknown matrices, and v_k is an unknown disturbance. We propose to estimate F and G by solving the optimization problem

$$\min_{\{F,G\}} \sum_{k=0}^{N} \|x_{k+1} - Fx_k - Gu_k\|_2^2.$$

We denote the solution by $\{\hat{F}_N, \hat{G}_N\}$. Assuming that the matrix

$$\begin{bmatrix} x_0 & x_1 & x_2 & \ldots & x_N \\ u_0 & u_1 & u_2 & \ldots & u_N \end{bmatrix},$$

has full row rank, write down expressions for the estimates $\hat{F}_N$ and $\hat{G}_N$. Also show how $\{\hat{F}_{N+1}, \hat{G}_{N+1}\}$, which are the optimal solutions using data up to time $N+1$, are related to $\{\hat{F}_N, \hat{G}_N\}$. [*Hint.* Use the result of Prob. 2.9 and recall the derivation of the RLS algorithm in Sec. 2.6.]

2.11 **(Ordering of matrices)** Let A and B be two Hermitian positive-definite matrices satisfying $A \geq B > 0$, where by $A \geq B$ we mean that the difference $A - B$ is nonnegative-definite. Choose a positive scalar ϵ and define $A(\epsilon) = A + \epsilon I$.

(a) Express $A(\epsilon)$ in the form $A(\epsilon) = B + [A(\epsilon) - B]$ and show that

$$A^{-1}(\epsilon) - B^{-1} = -B^{-1}[B^{-1} + (A(\epsilon) - B)^{-1}]^{-1}B^{-1}.$$

(b) By taking the limit as $\epsilon \to 0$, conclude that $B^{-1} \geq A^{-1} > 0$.

(c) Now consider the recursive updating and downdating solutions of Lemmas 2.6.1 and 2.7.1. Use the result of part (b) to argue that $P_i \leq P_{i-1}$ and $P_{1:i} \geq P_{0:i}$. Provide an interpretation of these results.

2.12 **(Full rank after downdating)** Assume a full rank data matrix $H_{0:i}$ in (2.7.1), where h_0 is a row vector.

(a) Prove that $H_{1:i}$ is full rank if, and only if, $1 - h_0[H_{0:i}^* H_{0:i}]^{-1}h_0^* \neq 0$.
Hint. For $n \times 1$ column vectors a and b, $\det(I + ab^*) = 1 + b^*a$.]

(b) Prove that if $H_{1:i}$ is full rank, then $1 - h_0[H_{0:i}^* H_{0:i}]^{-1}h_0^* > 0$.

2.13 **(Simultaneous optimization (Björck (1996, p. 20)))** Assume that $H \in \mathbb{C}^{N \times n}$ is full rank with $N \geq n$. Show that the solution $\text{col}\{z, x\}$ of the nonsingular linear system of equations

$$\begin{bmatrix} I & H \\ H^* & 0 \end{bmatrix} \begin{bmatrix} z \\ x \end{bmatrix} = \begin{bmatrix} y \\ c \end{bmatrix},$$

for given $\{y, c\}$, is the solution of the so-called *primal* and *dual* least-squares problems

$$\min_x \; [\|y - Hx\|^2 + c^*x] \quad \text{and} \quad \min_z \; \|y - z\|^2 \text{ subject to } H^*z = c.$$

Now let $H = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$ denote the full QR decomposition of H. Show that z and x can be computed from

$$x = R^{-1}[d_1 - R^{-*}c], \quad z = Q \begin{bmatrix} R^{-*}c \\ d_2 \end{bmatrix}, \quad \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = Q^*y.$$

Remark. Note that by setting $y = 0$, the solution z becomes the minimum-norm solution of the underdetermined linear system $H^*z = x$ (*cf.* the discussion in App. A). Likewise, by setting $c = 0$, the resulting x becomes the solution of the LS problem $\min_x \|y - Hx\|$. ◆

2.14 **(Isometric matrices)** $N \times n$ matrices, with $n \leq N$, are called *isometric* if their columns are mutually orthogonal. The matrix $\hat{Q}$ in Sec. 2.5 is an isometric matrix.

(a) Show that $\hat{Q}^*\hat{Q} = I$, while $\hat{Q}\hat{Q}^*$ and $I - \hat{Q}\hat{Q}^*$ are projection matrices.

(b) Show that the minimum cost in the LS problem can be computed as $\|y - H\hat{x}\|^2 = y^*(I - \hat{Q}\hat{Q}^*)y$.

2.15 **(Alternative derivation of the QR solution)** Refer to the discussion in Sec. 2.5. Establish the identity $y - Hx = (I - \hat{Q}\hat{Q}^*)y - \hat{Q}(\hat{R}x - \hat{Q}^*y)$, and use it to show that

$$\|y - H\hat{x}\|^2 = \|\hat{R}x - \hat{Q}^*y\|^2 + \|(I - \hat{Q}\hat{Q}^*)y\|^2.$$

Conclude that the least-squares solution satisfies the equation $\hat{R}\hat{x} = \hat{Q}^*y$.

2.16 **(LS solution using the full QR decomposition)** Refer again to the discussion in Sec. 2.5. We now solve the least-squares problem by using the full QR decomposition of H rather than its reduced QR decomposition (2.5.3). So let

$$H = Q \begin{bmatrix} R \\ 0 \end{bmatrix},$$

where Q is now $N \times N$ and unitary. As explained in App. A, such a full QR factorization for H can be obtained from its reduced QR decomposition by appending an additional $N - n$ orthonormal columns to $\hat{Q}$ so that it becomes a unitary $N \times N$ matrix. Likewise, we append rows of zeros to R so that it becomes an $N \times n$ matrix.

(a) Verify that $\|y - Hx\|^2 = \|Q^*(y - Hx)\|^2 = \|z_1 - Rx\|^2 + \|z_2\|^2$, where we have partitioned Q^*y into $Q^*y = \text{col}\{z_1, z_2\}$ and z_1 is $n \times 1$.

(b) Conclude that the optimal least-squares solution is obtained by solving $z_1 = R\hat{x}$ and that the resulting minimum cost is $\|z_2\|^2$.

2.17 **(An array algorithm for updating the LS solution)** We use the notation of Sec. 2.6. Suppose that at step $i - 1$ we have solved the standard least-squares problem $\min_x \|y_{i-1} - H_{i-1}x\|^2$ via the full QR method of Prob. 2.16 and obtained $\hat{x}_{i-1}$. Let

$$Q_{i-1}H_{i-1} = \begin{bmatrix} R_{i-1} \\ 0 \end{bmatrix},$$

denote the QR factorization of H_{i-1}, and define $Q_{i-1}y_{i-1} \triangleq \text{col}\{z_{1,i-1}, z_{2,i-1}\}$. Now suppose that at step i we are given one extra row h_i and one extra scalar $y(i)$, so that we must now solve

$$\min_x \left\| \begin{bmatrix} y_{i-1} \\ y(i) \end{bmatrix} - \begin{bmatrix} H_{i-1} \\ h_i \end{bmatrix} x \right\|^2.$$

(a) Verify that

$$\left\| \begin{bmatrix} Q_{i-1} & 0 \\ 0 & 1 \end{bmatrix} (y_i - H_i x) \right\|^2 = \left\| \begin{bmatrix} z_{1,i-1} \\ y(i) \end{bmatrix} - \begin{bmatrix} R_{i-1} \\ h_i \end{bmatrix} \right\|^2 + \|z_{2,i-1}\|^2.$$

(b) Let Q be a unitary matrix that upper triangularizes $\begin{bmatrix} R_{i-1} \\ h_i \end{bmatrix}$,

$$Q \begin{bmatrix} R_{i-1} \\ h_i \end{bmatrix} = \begin{bmatrix} R_i \\ 0 \end{bmatrix},$$

and define

$$Q \begin{bmatrix} z_{1,i-1} \\ y(i) \end{bmatrix} \triangleq \begin{bmatrix} z_{1,i} \\ z_{2,i} \end{bmatrix}.$$

Show that $\hat{x}_i = R_i^{-1}z_{1,i}$ and the optimum cost $J(\hat{x}_i)$ at step i is given by $J(\hat{x}_i) = J(\hat{x}_{i-1}) + \|z_{2,i}\|^2$.

Remark. The update procedure can be combined into a single matrix equation as follows:

$$Q\begin{bmatrix} R_{i-1} & z_{1,i-1} \\ h_i & y(i) \end{bmatrix} = \begin{bmatrix} R_i & z_{1,i} \\ 0 & z_{2,i} \end{bmatrix}.$$

By considering the conjugate transpose we obtain

$$\begin{bmatrix} R_{i-1}^* & 0 \\ z_{1,i-1}^* & y^*(i) \end{bmatrix} Q^* = \begin{bmatrix} R_i^* & 0 \\ z_{1,i}^* & z_{2,i}^* \end{bmatrix}.$$

By comparing with (2.6.16) we see that these equations agree, apart from the different notation and from the zero initial condition $\Pi_0 = 0$, with the first two lines of the array algorithm (2.6.16). In fact, it can be easily verified that the quantities $\{R_i^*, z_{1,i}^*, z_{2,i}^*\}$ in the above equations coincide with the quantities $\{P_i^{-*/2}, \hat{x}_i P_i^{-*/2}, e_a^*(i) r_e^{-*/2}(i)\}$ in (2.6.16). ◆

2.18 (Constrained least-squares) In constrained least-squares problems one is required to minimize $\| y - Hx \|^2$ subject to the linear constraint $Ax = b$, where H and A are $N \times n$ ($N \geq n$) and $M \times n$ ($M \leq n$) full rank matrices, respectively. Solve this problem by using Lagrange multiplier techniques, *i.e.*, minimizing over x and stationarizing over λ the following *unconstrained* quadratic form:

$$J = \| y - Hx \|^2 + \frac{1}{2}\lambda^*(Ax - b) + \frac{1}{2}(Ax - b)^*\lambda.$$

Show also that the stationarizing λ is given by

$$\frac{1}{2}\hat{\lambda} = \left[A(H^*H)^{-1}A^*\right]^{-1}(A\hat{x}_{LS} - b),$$

where $\hat{x}_{LS}$ is the unconstrained least-squares solution. Conclude that the constrained solution is given by $\hat{x}_c = \hat{x}_{LS} - (H^*H)^{-1}A^*\left[A(H^*H)^{-1}A^*\right]^{-1}(A\hat{x}_{LS} - b)$, and that

$$\min_{Ax=b} \|y - Hx\|^2 = \|y - H\hat{x}_{LS}\|^2 + (A\hat{x}_{LS} - b)^*\left[A(H^*H)^{-1}A^*\right]^{-1}(A\hat{x}_{LS} - b).$$

2.19 (A QR method for constrained least-squares) Consider the constrained least-squares formulation of Prob. 2.18 and introduce the full QR decomposition of A^*, *viz.*, $A^* = Q\,\mathrm{col}\{R, 0\}$, where Q is $n \times n$ unitary and R is $M \times M$ upper triangular. Define $z \triangleq Q^*x$ and partition it into $z = \mathrm{col}\{z_1, z_2\}$ where z_1 is $M \times 1$. Hence, solving for x is equivalent to solving for z. Define also $\bar{H} = HQ$ and partition it as $\bar{H} = \begin{bmatrix} \bar{H}_1 & \bar{H}_2 \end{bmatrix}$, where $\bar{H}_1$ is $N \times M$. Let $\bar{y}_2 = y - \bar{H}_1 z_1$.

(a) Show that we must have $z_1 = R^{-*}b$.

(b) Show that z_2 should be determined by solving the unconstrained least-squares problem $\min_{z_2} \|\bar{y}_2 - \bar{H}_2 z_2\|^2$.

2.20 (A sliding window (or finite memory) adaptive filter) Consider the overdetermined system of linear equations

$$y_{i-L+1:i} = \begin{bmatrix} y(i-L+1) \\ \vdots \\ y(i) \end{bmatrix} \cong \begin{bmatrix} h_{i-L+1} \\ \vdots \\ h_i \end{bmatrix} x = H_{i-L+1:i}x,$$

where L is fixed and i varies with time, and denote its least-squares solution by $\hat{x}_{i-L+1:i}$. Since the above linear system consists of L equations at all time instants i, $\hat{x}_{i-L+1:i}$ is referred to as the *finite memory* or *sliding* window, of length L, least-squares solution. Find a recursion for going from $\hat{x}_{i-L:i-1}$ to $\hat{x}_{i-L+1:i}$ by combining the downdating recursion $\hat{x}_{i-L:i-1} \rightarrow \hat{x}_{i-L+1:i-1}$ and the updating recursion $\hat{x}_{i-L+1:i-1} \rightarrow \hat{x}_{i-L+1:i}$.
[*Hint.* Recall the results of Secs. 2.6 and 2.7 on updating and downdating LS solutions.]

Appendix for Chapter 2

2.A ON SYSTEMS OF LINEAR EQUATIONS

Many questions in estimation theory (and many other disciplines as well) ultimately reduce to the study of the system of linear equations

$$Hx = y, \tag{2.A.1}$$

where $H \in \mathbb{C}^{N\times n}$ and $y \in \mathbb{C}^{N\times 1}$ are known, and $x \in \mathbb{C}^{n\times 1}$ is unknown. Depending on the values of the matrix H and vector y, the equation (2.A.1) may have a unique solution, many solutions, or no solutions. Likewise, depending on whether $N > n$, $N < n$, or $N = n$, the system of equations is said to be overdetermined, underdetermined, or square, respectively.

The study of linear equations is a major topic in many books on linear algebra and matrix computation, which should be consulted for a detailed treatment. Our aim here is largely to outline some major points, mostly to introduce some terminology and some more geometric intuition.

We begin by recalling that Hx can be regarded as a linear combination of the columns of H, where each column is weighted by the corresponding component of x. It follows that (2.A.1) will have a solution if, and only if, the right-hand side y is some linear combination of the columns of H. In this case, the system (2.A.1) is referred to as *consistent*.

To formalize this observation, we define the *range space* of H, denoted by $\mathcal{R}(H)$, as the linear space spanned by the columns of H, *i.e.*,

$$\mathcal{R}(H) = \{Ha \mid a \in \mathbb{C}^n\}. \tag{2.A.2}$$

Note that $\mathcal{R}(H)$ is indeed a linear space since it satisfies the *closure* property, *i.e.*, if $z_1, z_2 \in \mathcal{R}(H)$, then for any complex scalars c_1, c_2, we have $c_1z_1 + c_2z_2 \in \mathcal{R}(H)$. We can now write the condition for (2.A.1) to be a consistent system of linear equations as $y \in \mathcal{R}(H)$. It is also clear that (2.A.1) will have *no solution* if y does not belong to $\mathcal{R}(H)$.

Now if the equations are consistent, is it possible to have more than one solution? To investigate this question, let us suppose that (2.A.1) has two solutions $x_1 \neq x_2$, so that $Hx_1 = y$ and $Hx_2 = y$. We readily see that $H(x_1 - x_2) = 0$, from which we conclude that we can have more than one solution if, and only if, $H(x_1 - x_2) = 0$ has a nontrivial solution, *i.e.*, a solution whose components are not all identically zero.

To formalize this result, we define the *nullspace* of a matrix H as

$$\mathcal{N}(H) = \{a \in \mathbb{C}^n \mid Ha = 0\}, \tag{2.A.3}$$

i.e., as the linear space of those vectors that after multiplication by H yield the zero vector. Note once more that $\mathcal{N}(H)$ is indeed a linear space since if $z_1, z_2 \in \mathcal{N}(H)$, then for any complex scalars c_1, c_2, we have $c_1z_1 + c_2z_2 \in \mathcal{N}(H)$. We therefore conclude that the system (2.A.1) will have *more than one* solution if, and only if, $y \in \mathcal{R}(H)$ and $\mathcal{N}(H) \neq \{0\}$. In this case any two solutions x_1 and x_2 will differ by a vector in the nullspace of H, *i.e.*, $x_1 - x_2 = c \in \mathcal{N}(H)$.

Now $\mathcal{N}(H) = \{0\}$ if, and only if, there is no nontrivial combination of the columns of H that can be zero. Any set of such columns is said to be a *linearly independent* set, and any matrix whose columns are independent is said to have *full column rank*. It is thus obvious that (2.A.1) will have *a unique* solution if, and only if, $y \in \mathcal{R}(H)$ and H has full column rank. The matrix H will be said to have column rank α if at most α columns of H are linearly independent. In such case, $\mathcal{N}(H)$ will have dimension $n - \alpha$.

So far we have only talked about the columns of H and the equation $Hx = y$. Clearly, similar statements can be made about the rows of H and the equation $xH = y$, where now $y \in \mathbb{C}^{1\times n}$ is known and $x \in \mathbb{C}^{1\times N}$ is unknown. We shall not repeat the arguments here; it will suffice to mention that the *row rank* of H is defined as the maximum number of independent rows of H. It is an important fact, not proven here, that the row rank of a matrix is equal to its column rank.

Square Matrices. The special case of square matrices ($N = n$) is of particular importance. The properties of determinants show that a square full column (or row) rank matrix will always have a nonzero determinant. Such matrices are referred to as *nonsingular*, and are always invertible, so that the inverse H^{-1}, with the property that $HH^{-1} = H^{-1}H = I$, always exists. In this case the equation $Hx = y$ is always consistent and the unique solution is given by $x = H^{-1}y$.

On the other hand if H is singular, it will not have an inverse, and the nullspace will be nonempty (*i.e.*, it will contain something besides the zero column). Whether we have a solution or not, now depends on whether y lies in $\mathcal{R}(H)$ or not.

Range Spaces and Null Spaces. We have just seen that the question of existence and uniqueness for systems of linear equations depends upon the range space $\mathcal{R}(H)$ and the null space $\mathcal{N}(H)$ of the matrix H. It will therefore be useful to study the properties of these two spaces in some more detail.

We first define $\mathcal{L}^\perp$, the *orthogonal complement space* of a linear space $\mathcal{L} \subseteq \mathbb{C}^n$, as

$$\mathcal{L}^\perp = \{x \in \mathbb{C}^n \mid x^*y = 0, \quad \text{for all} \quad y \in \mathcal{L}\}. \tag{2.A.4}$$

Lemma 2.A.1 (Range Spaces and Null Spaces) *We have the following properties:* $\mathcal{N}(H) = \mathcal{R}(H^*)^\perp$, $\mathcal{N}(H^*) = \mathcal{R}(H)^\perp$, $\mathcal{R}(H) = \mathcal{N}(H^*)^\perp$, *and* $\mathcal{R}(H^*) = \mathcal{N}(H)^\perp$. ∎

Proof: We prove the first property. The rest follow in a similar fashion. Now one way we have

$$\begin{aligned} x \in \mathcal{N}(H) \;\Rightarrow\; Hx = 0 \Rightarrow x^*H^*y = 0 \quad \text{for all} \quad y \in \mathbb{C}^N \;\Rightarrow\; x \perp \mathcal{R}(H^*) \\ \Rightarrow x \in \mathcal{R}(H^*)^\perp \;\Rightarrow\; \mathcal{N}(H) \subseteq \mathcal{R}(H^*)^\perp. \end{aligned} \tag{2.A.5}$$

And the other way

$$x \in \mathcal{R}(H^*)^\perp \Rightarrow x \perp \mathcal{R}(H^*) \Rightarrow x^*H^*y = 0 \quad \text{for all} \quad y \in \mathbb{C}^N$$
$$\Rightarrow y^*(Hx) = 0 \quad \text{for all} \quad y \in \mathbb{C}^N \Rightarrow Hx = 0$$
$$\Rightarrow x \in \mathcal{N}(H) \Rightarrow \mathcal{R}(H^*)^\perp \subseteq \mathcal{N}(H). \tag{2.A.6}$$

◆

Corollary 2.A.1. (Range Spaces and Null Spaces) *Suppose H is an $N \times n$ matrix with complex entries. Then we have*

$$\mathcal{N}(H) \oplus \mathcal{R}(H^*) = \mathbb{C}^n \quad \textit{and} \quad \mathcal{N}(H^*) \oplus \mathcal{R}(H) = \mathbb{C}^N. \tag{2.A.7}$$

◆

Remark. Note that since $\dim[\mathcal{R}(H^*)] = \dim[\mathcal{R}(H)] = \text{rank}(H) = \alpha$, we have that $\dim[\mathcal{N}(H)] = n - \alpha$ and $\dim[\mathcal{N}(H^*)] = N - \alpha$. ◆

Application to the Normal Equations. Recall from Sec. 2.2 that the solution to the least-squares problem $\min_x \| y - Hx \|^2$, with $H \in \mathbb{C}^{N\times n}$ and $y \in \mathbb{C}^{N\times 1}$, is given by the normal equations

$$H^*H\hat{x} = H^*y. \tag{2.A.8}$$

Moreover, $\hat{y} = H\hat{x}$ had the interpretation of being the projection of the vector y onto the linear space $\mathcal{R}(H)$.

When H has full rank (so that H^*H is nonsingular) the unique solution to the normal equations is given by $\hat{x} = (H^*H)^{-1}H^*y$. However, when H is not full rank (so that H^*H is singular) we need to further study (2.A.8) to determine whether a solution exists or not. It turns out that a solution to the normal equations always exists and that, although $\hat{x}$ is not unique, the projection, $\hat{y} = H\hat{x}$, is. The fact that a solution to the normal equations always exists is seen from the following lemma.

Lemma 2.A.2 (Range Spaces of H^* and H^*H) *Let H be an $N \times n$ matrix with complex entries. Then we have $\mathcal{R}(H^*H) = \mathcal{R}(H^*)$.* ■

Proof: In view of Lemma 2.A.1 we can equivalently show that $\mathcal{N}(H^*H) = \mathcal{N}(H)$. To this end, consider one direction:

$$x \in \mathcal{N}(H^*H) \Rightarrow H^*Hx = 0 \Rightarrow x^*H^*Hx = 0 \Rightarrow \| Hx \|^2 = 0$$
$$\Rightarrow Hx = 0 \Rightarrow x \in \mathcal{N}(H) \Rightarrow \mathcal{N}(H^*H) \subseteq \mathcal{N}(H).$$

And now the other way:

$$x \in \mathcal{N}(H) \Rightarrow Hx = 0 \Rightarrow H^*Hx = 0 \Rightarrow x \in \mathcal{N}(H^*H) \Rightarrow \mathcal{N}(H) \subseteq \mathcal{N}(H^*H).$$

Comparing the above two conclusions yields the desired result. ◆

We turn next to the uniqueness of the projection $\hat{y}$. For this purpose, note that all solutions of the normal equations (2.A.8) will differ by a vector in the nullspace of H^*H, but since $\mathcal{N}(H) = \mathcal{N}(H^*H)$, they differ by a vector in the nullspace of H. Thus for any two such solutions $\hat{x}_1$ and $\hat{x}_2$, we have $\hat{x}_1 - \hat{x}_2 \in \mathcal{N}(H)$ or, equivalently, $H(\hat{x}_1 - \hat{x}_2) = 0$. This implies $H\hat{x}_1 = H\hat{x}_2$, which means that the projection $\hat{y} = H\hat{x}$ is unique.

From the above discussion we further note that when H is rank deficient, the space of all possible solutions is given by the affine linear space

$$\hat{\mathcal{L}} = \{x_0 + c \mid H^*Hx_0 = H^*y, \quad c \in \mathcal{N}(H)\}.$$

One way to obtain a unique solution to the normal equations is to insist on a solution that has the minimum norm, *i.e.*, one with the property that

$$x_{min} \in \hat{\mathcal{L}} \text{ such that } \|x_{min}\|^2 < \|x\|^2, \quad \text{for all } x \neq x_{min} \in \hat{\mathcal{L}}.$$

It is clear that x_{min} is that solution to the normal equations that is perpendicular to $\mathcal{N}(H)$. In other words, $x_{min} \perp \mathcal{N}(H)$ and $x_{min} \in \hat{\mathcal{L}}$. But since $\mathcal{N}(H)^\perp = \mathcal{R}(H^*) = \mathcal{R}(H^*H)$, this may be equivalently written as $x_{min} = \hat{\mathcal{L}} \cap \mathcal{R}(H^*H)$. In other words, for some vector λ, we must have $H^*H\lambda = x_{min}$. Combining this equation with the normal equations for x_{min}, $H^*Hx_{min} = H^*y$, we obtain that λ satisfies $(H^*H)^2\lambda = H^*y$. Note that since $\mathcal{R}(H^*) = \mathcal{R}(H^*H) = \mathcal{R}((H^*H)^2)$ the above equation always has a solution for λ. However, since any two solutions will differ by a vector in the nullspace of $(H^*H)^2$, or equivalently in the nullspace of H^*H, we see that $x_{min} = H^*H\lambda$ is *unique*. We can thus find the unique minimum norm solution of the normal equations, x_{min}, as follows:

(i) Find any solution λ to the consistent system of linear equations $(H^*H)^2\lambda = H^*y$.

(ii) The unique minimum norm solution is given by $x_{min} = H^*H\lambda$, for any λ found in part (i).

We may note that an alternative method for finding x_{min} is based on the singular value decomposition of the matrix H, which is discussed in Sec. A.4.

CHAPTER 3

Stochastic Least-Squares Problems

In many senses, the basic development of the subject of this textbook begins in this chapter. The material in Sec. 3.2 is quite straightforward, but the simple results derived here can often be recognized in more complicated contexts and in fact underlie many later derivations. The geometric formulation of Sec. 3.3 is fundamental and will repay careful study and rereading from time to time until the material becomes second nature.

In Sec. 3.4, we specifically address the case of "linear" models, those in which there is a linear relationship between the observations and the quantities to be estimated. This additional structure can be exploited to develop some useful identities. In fact, these will allow us in Sec. 3.5 to draw up an important equivalence (not duality) relation between the deterministic and stochastic least-squares problems for linear models. We have chosen the word *equivalence* deliberately, because we are not referring to the widely quoted (for more advanced readers) *duality* between these two classes of problems — in fact, there are both equivalences and dualities, and both are useful. However, it is too early for our purposes to pursue this issue here — more eager readers can turn to Ch. 15.

An important terminological issue needs to be cleared up here. Strictly speaking, we should distinguish between the *stochastic* linear least-squares problem (often called the linear least-mean-squares problem) and the *deterministic* linear least-squares problem of Ch. 2. We shall often abuse the notation, however, on the (arguable) grounds of brevity and euphony, especially since we shall show that the results of one problem can be obtained from the results of the other, and vice versa. We note also that we shall often abbreviate "linear least-mean-squares" as "l.l.m.s." and "linear least-squares" as "l.l.s."

3.1 THE PROBLEM OF STOCHASTIC ESTIMATION

Consider two (scalar or column-vector) random variables $\mathbf{x}$ and $\mathbf{y}$ (possibly complex-valued) with joint probability density function $f_{\mathbf{x},\mathbf{y}}(\cdot,\cdot)$. If the random variables are independent, *i.e.*, if they assume values independently of each other, then there is little (if any) that can be said about the value assumed by one random variable when the value assumed by the other is known or measured. So we shall assume that the random variables are dependent and ask the following question: given that the variable $\mathbf{y}$ assumed the value y in a particular experiment, what can be said (or guessed) about the value assumed by the variable $\mathbf{x}$?

Such questions often arise when the quantity of interest is not directly observable or directly measurable while it is possible to monitor another related quantity. For example, we may have available only noisy measurements $\mathbf{y}$ of $\mathbf{x}$, say $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where the random variable $\mathbf{v}$ represents additive noise or disturbance. With a proper formulation, reasonable information about $\mathbf{x}$ can be extracted from the noisy measurement $\mathbf{y}$.

To tackle the general question, an estimate of the value assumed by $\mathbf{x}$, say $\hat{x}$, can be described as a function of the value assumed by $\mathbf{y}$, say

$$\hat{x} = h(y). \tag{3.1.1}$$

We shall refer to $\hat{x}$ as the *estimate*. Likewise, we shall refer to the random variable $\hat{\mathbf{x}}$ defined by

$$\hat{\mathbf{x}} = h(\mathbf{y}), \tag{3.1.2}$$

as the *estimator*; evaluating the estimator $\hat{\mathbf{x}}$ at a particular value for $\mathbf{y}$ results in an estimate $\hat{x}$. The challenge is to suitably choose the function $h(\cdot)$ to yield reasonable estimates $\hat{x}$. By reasonable we mean estimates that satisfy a desired optimality criterion. There are several criteria that can be used for estimation problems, but for signal processing, communications, and control one of the most important, at least in the sense of having had the most applications, is the least-mean-squares criterion (which is the stochastic analogue of the deterministic least-squares criterion of Ch. 2 — see also the notes on the squared error criterion in Chs. 1 and 2). With the mean-square-error criterion, it can be shown, and in fact quite readily (see App. 3.A), that the optimum estimator is

$$\hat{\mathbf{x}} = E[\mathbf{x}|\mathbf{y}],$$

the conditional expectation of $\mathbf{x}$ given $\mathbf{y}$. Calculating this expectation requires full knowledge of the joint probability density function of $\{\mathbf{x}, \mathbf{y}\}$, which is often hard to obtain. If however we deliberately restrict the estimator $h(\cdot)$ to be a *linear* function of the observations, then it turns out that all we shall need is knowledge of the first- and second-order statistics, $E\mathbf{x}$, $E\mathbf{y}$, $E\mathbf{x}\mathbf{x}^*$, $E\mathbf{y}\mathbf{y}^*$, and $E\mathbf{x}\mathbf{y}^*$. This is the condition we shall impose henceforth.

We may also remark that when $\{\mathbf{x}, \mathbf{y}\}$ are jointly Gaussian, an assumption that is often reasonable, then the unconstrained least-mean-squares estimator is linear, as is shown in Apps. 3.B–3.C.

Therefore, what we shall study is a linear least-mean-squares estimation problem, which, as we shall see, is a stochastic counterpart of the least-squares problem of Ch. 2.

3.2 LINEAR LEAST-MEAN-SQUARES ESTIMATORS

The problem of linear least-mean-squares estimation for a finite number of real-valued random variables is straightforward and the basic facts can be found in many introductory textbooks on probability. In many applications in communications and signal processing, however, we have to deal with complex-valued random variables.

For such complex-valued and zero-mean random variables, say $\mathbf{x}$ and $\mathbf{y}$, we shall define the covariance matrix as $\text{cov}(\mathbf{x}, \mathbf{y}) = E\mathbf{x}\mathbf{y}^*$, where $*$ denotes the complex conjugate for a scalar random variable and the complex conjugate transpose (the so-called Hermitian transpose) for a vector-valued random variable. The main reason for this definition is to ensure that $\text{var}(\mathbf{x}) = \text{cov}(\mathbf{x}, \mathbf{x}) = E\mathbf{x}\mathbf{x}^*$, is a nonnegative scalar when $\mathbf{x}$ is scalar, and a nonnegative-definite matrix when $\mathbf{x}$ is a vector random variable. One could ask about the significance of quantities such as $E\mathbf{x}\mathbf{y}^T$ or $E\mathbf{x}\mathbf{x}^T$, where T denotes transpose (not Hermitian transpose). This is not usually done because these quantities are not encountered in the conventional formulation of the l.l.m.s.e. problem, which is the one we shall begin with in Sec. 3.2.1. Then in Sec. 3.2.5 we shall re-examine the issue and show that under a so-called circularity assumption it is not required to consider quantities such as $E\mathbf{x}\mathbf{y}^T$.

3.2.1 The Fundamental Equations

We begin with one particular form of linear estimator. The solution will be seen to involve only first- and second-order statistics (*i.e.*, the mean values and variances and covariances) of the variables involved.

So consider a complex vector-valued zero-mean random variable $\mathbf{x}$ of dimension n and a related set of complex vector-valued zero-mean random variables $\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$, where each $\mathbf{y}_i$ has dimension p.

Our objective is to estimate the value assumed by the random variable $\mathbf{x}$ given that the random variables $\{\mathbf{y}_i\}$ assumed certain values $\{y_i\}$. We are interested in *linear* estimators for $\mathbf{x}$, *viz.*, estimators that operate linearly on the random variables $\{\mathbf{y}_i\}$.[1]

We shall write, as appropriate, $\hat{\mathbf{x}}_{|\mathbf{y}}$ or $\hat{\mathbf{x}}_{|N}$ or simply $\hat{\mathbf{x}}$, to denote a linear estimator for $\mathbf{x}$ given the N random variables $\{\mathbf{y}_i\}$. We assume $\hat{\mathbf{x}}$ is constructed as a linear combination of the form

$$\hat{\mathbf{x}} = K_o\mathbf{y}, \tag{3.2.1}$$

where $K_o \in \mathbb{C}^{n \times p(N+1)}$ is a coefficient matrix that we wish to determine so as to minimize the resulting error covariance matrix, *i.e.*,

$$P(K_o) \triangleq E[\mathbf{x} - \hat{\mathbf{x}}][\mathbf{x} - \hat{\mathbf{x}}]^* = \text{minimum}. \tag{3.2.2}$$

[1] We shall see in Sec. 3.2.5 that linear estimators for *complex-valued* random variables can be defined in two different ways; depending upon the kind of prior information that is available about the variables $\{\mathbf{x}, \mathbf{y}_i\}$. Each choice will lead, in general, to a different optimal estimator. They, however, will coincide in the case of real-valued random variables and in the case of *circular* (or spherically invariant) random variables (see App. 3.B for a brief discussion of circular random variables). These issues will be clarified in Sec. 3.2.5.

That is, we should find K_o such that for every $K \in \mathbb{C}^{n \times p(N+1)}$ we obtain

$$P(K) \triangleq E[\mathbf{x} - K\mathbf{y}][\mathbf{x} - K\mathbf{y}]^* \geq P(K_o).$$

We recall from Sec. 1.2 that this is equivalent to requiring that

$$aP(K)a^* \geq aP(K_o)a^*, \tag{3.2.3}$$

for every K and for every row vector a. The solution to the above problem is given by the following theorem.

Theorem 3.2.1 (Optimal Linear L.M.S. Estimators) *Given two complex zero mean random variables $\mathbf{x}$ and $\mathbf{y}$, the l.l.m.s. estimator of $\mathbf{x}$ given $\mathbf{y}$, defined by (3.2.1)–(3.2.2) is given by any solution K_o of the so-called normal equations*

$$K_oR_y = R_{xy}, \tag{3.2.4}$$

where $R_y = E\mathbf{y}\mathbf{y}^*$ and $R_{xy} = E\mathbf{x}\mathbf{y}^* = R_{yx}^*$. The corresponding minimum-mean-square-error matrix (or error covariance matrix) is

$$P(K_o) = R_x - K_oR_{yx} = R_x - R_{xy}K_o^*. \tag{3.2.5}$$

■

Proof: K_o is a solution of the optimization problem (3.2.1)–(3.2.2) if, and only if, for all vectors a, aK_o is a minimum of $aP(K)a^*$, where

$$aP(K)a^* = aE[(\mathbf{x} - K\mathbf{y})(\mathbf{x} - K\mathbf{y})^*]a^* = a\left[R_x - R_{xy}K^* - KR_{yx} + KR_yK^*\right]a^*.$$

Note that $aP(K)a^*$ is a *scalar* function of a complex-valued (row) vector quantity aK. Then (see App. A.6) differentiating $aP(K)a^*$ with respect to aK and setting the derivative equal to zero at $K = K_o$ leads to the equations $R_{xy} = K_oR_y$. The corresponding minimum-mean-square-error (or, m.m.s.e. for short) matrix is

$$\begin{aligned}\text{m.m.s.e.} \triangleq P(K_o) &= E(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^* = E(\mathbf{x} - \hat{\mathbf{x}})\mathbf{x}^* - E(\mathbf{x} - \hat{\mathbf{x}})\hat{\mathbf{x}}^* \\ &= E(\mathbf{x} - K_o\mathbf{y})\mathbf{x}^* - E(\mathbf{x} - K_o\mathbf{y})\mathbf{y}^*K_o^* \\ &= R_x - K_oR_{yx} - (R_{xy} - K_oR_y)K_o^* = R_x - K_oR_{yx}.\end{aligned}$$

◆

Two remarks are in order here. First, the result is clearly valid for real-valued random variables $\{\mathbf{x}, \mathbf{y}\}$ as well, with the complex-conjugate symbol $*$ replaced by the transpose symbol T. Secondly, it turns out that the optimum choice K_o also minimizes various other criteria (see Prob. 3.4). Note, in particular, that (3.2.3) implies that the solution K_o also minimizes the mean-square error in the estimator of each component of the vector $\mathbf{x}$.

Theorem 3.2.2 (Unique Solutions) *Assume that $R_y > 0$. Then the optimum choice K_o that minimizes $P(K) = E[\mathbf{x} - K\mathbf{y}][\mathbf{x} - K\mathbf{y}]^*$ is given by*

$$K_o = R_{xy}R_y^{-1}. \tag{3.2.6}$$

The m.m.s.e. (see (3.2.5)) can be written as

$$P(K_o) = R_x - R_{xy}R_y^{-1}R_{yx} \triangleq R_{\tilde{x}}. \tag{3.2.7}$$

■

Proof: This result of course immediately follows from Thm. 3.2.1. However, since $R_y > 0$, we can give a different proof (analogous to that in Lemma 1.2.1) based on the idea of *completion of squares*. Thus

$$\begin{aligned} P(K) &= E[\mathbf{x} - K\mathbf{y}][\mathbf{x} - K\mathbf{y}]^* = R_x - KR_{yx} - R_{xy}K^* + KR_yK^*, \\ &= \begin{bmatrix} I & -K \end{bmatrix} \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix} \begin{bmatrix} I \\ -K^* \end{bmatrix}. \end{aligned}$$

Since $R_y > 0$, we can use an upper-lower block triangular factorization (see App. A)

$$\begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix} = \begin{bmatrix} I & R_{xy}R_y^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} R_{\tilde{x}} & 0 \\ 0 & R_y \end{bmatrix} \begin{bmatrix} I & 0 \\ R_y^{-1}R_{yx} & I \end{bmatrix},$$

where we have defined $R_{\tilde{x}} = R_x - R_{xy}R_y^{-1}R_{yx}$, so that we can write

$$P(K) = R_x - R_{xy}R_y^{-1}R_{yx} + (R_{xy}R_y^{-1} - K)R_y(R_{xy}R_y^{-1} - K)^*.$$

Now the quadratic form involving K is always nonnegative (because $R_y > 0$), so the smallest value it can have is zero, which is achieved by choosing $K = R_{xy}R_y^{-1} = K_o$. This choice also immediately establishes the formula (3.2.7) for $P(K_o)$. ◆

3.2.2 Stochastic Interpretation of Triangular Factorization

It is important to note that the covariance matrix $R_{\tilde{x}}$ of the error in estimating $\mathbf{x}$ from $\mathbf{y}$, assuming $R_y > 0$, *i.e.*, $R_{\tilde{x}} = R_x - R_{xy}R_y^{-1}R_{yx}$, is the *Schur complement* of R_y in the joint covariance matrix

$$E\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{x}^* & \mathbf{y}^* \end{bmatrix} = \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}.$$

So also the other Schur complement, assuming $R_x > 0$, $R_y - R_{yx}R_x^{-1}R_{xy} \triangleq R_{\tilde{y}}$, is the covariance matrix of the error in estimating $\mathbf{y}$ from $\mathbf{x}$.

This is a good place to provide a stochastic derivation of the block LDL* and UDU* triangular decompositions of a joint covariance matrix (in fact, this is a convenient way of recalling the decompositions). When $R_y > 0$, the UDU* decomposition

$$\begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix} = \begin{bmatrix} I & R_{xy}R_y^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} R_x - R_{xy}R_y^{-1}R_{yx} & 0 \\ 0 & R_y \end{bmatrix} \begin{bmatrix} I & 0 \\ R_y^{-1}R_{yx} & I \end{bmatrix}$$

follows from the representation of the pair $\{\mathbf{x}, \mathbf{y}\}$ of correlated random variables in terms of the obviously uncorrelated pair $\{\tilde{\mathbf{x}}_{|\mathbf{y}}, \mathbf{y}\}$, where $\tilde{\mathbf{x}}_{|\mathbf{y}} = \mathbf{x} - R_{xy}R_y^{-1}\mathbf{y}$:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} I & R_{xy}R_y^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{|\mathbf{y}} \\ \mathbf{y} \end{bmatrix}.$$

Likewise, when $R_x > 0$, the LDL* decomposition

$$\begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix} = \begin{bmatrix} I & 0 \\ R_{yx}R_x^{-1} & I \end{bmatrix} \begin{bmatrix} R_x & 0 \\ 0 & R_y - R_{yx}R_x^{-1}R_{xy} \end{bmatrix} \begin{bmatrix} I & R_x^{-1}R_{xy} \\ 0 & I \end{bmatrix},$$

is obtained from the representation of $\{\mathbf{x}, \mathbf{y}\}$ in terms of the uncorrelated pair $\{\mathbf{x}, \tilde{\mathbf{y}}_{|\mathbf{x}}\}$, where now $\tilde{\mathbf{y}}_{|\mathbf{x}} = \mathbf{y} - R_{yx}R_x^{-1}\mathbf{x}$:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} I & 0 \\ R_{yx}R_x^{-1} & I \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \tilde{\mathbf{y}}_{|\mathbf{x}} \end{bmatrix}.$$

To continue in this way, consider the zero-mean vector-valued random variables $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$, where

$$E\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{bmatrix}^*\right) = \begin{bmatrix} R_x & R_{xy} & 0 \\ R_{yx} & R_y & R_{yz} \\ 0 & R_{zy} & R_z \end{bmatrix},$$

and verify that the covariance matrix of the error in estimating $\mathbf{y}$ given both $\mathbf{x}$ and $\mathbf{z}$ is given by

$$E\tilde{\mathbf{y}}_{|\mathbf{x},\mathbf{z}}\tilde{\mathbf{y}}_{|\mathbf{x},\mathbf{z}}^* \triangleq R_{\tilde{y}_{|x,z}} = R_y - R_{yx}R_x^{-1}R_{xy} - R_{yz}R_z^{-1}R_{zy} = R_{\tilde{y}_{|x}} + R_{\tilde{y}_{|z}} - R_y.$$

This result is useful for the problem of combining estimators that are based on different observations, which we consider in Sec. 3.4.3 and Prob. 3.23.

Remark 1. It is possible to extend the above stochastic interpretation to non-Hermitian matrices by introducing the notion of oblique projections — see, *e.g.*, Sayed and Kailath (1995) and also Prob. 3.25. ◆

3.2.3 Singular Data Covariance Matrices

It is usually desirable to ensure that $R_y > 0$. The reason is that the singularity of R_y means that for some nonzero column vector $c \in \mathbb{C}^{(N+1)p}$, $c^* R_y = 0$, so that

$$0 = c^* R_y c = c^*(E\mathbf{y}\mathbf{y}^*)c = E|c^*\mathbf{y}|^2 = 0.$$

Therefore $c^*\mathbf{y}$ is a zero-mean, zero-variance complex-valued random variable, which we can identify with zero.[2] To see this, let $\mathbf{a}$ and $\mathbf{b}$ be real-valued random variables that denote the real and imaginary parts of $c^*\mathbf{y}$,

$$c^*\mathbf{y} = \mathbf{a} + j\mathbf{b}, \qquad \text{with} \quad j \triangleq \sqrt{-1}.$$

Then

$$0 = E|c^*\mathbf{y}|^2 = (E\mathbf{a}^2 + E\mathbf{b}^2),$$

which requires $E|\mathbf{a}|^2 = 0 = E|\mathbf{b}|^2$. That is, $\mathbf{a}$ and $\mathbf{b}$ are zero-variance real-valued random variables, which we shall consider to be identically zero random variables (or, more precisely, zero almost surely). In other words, $\mathbf{a} = \mathbf{b} = 0$ and, hence, $c^*\mathbf{y} = 0$. This means that there is a linear dependence between the components of $\mathbf{y}$. Such situations are usually degenerate and in general one will be better served by re-examining the mathematical modeling of the physical problem.

However, if we persist with the case where R_y is singular, the first question that must be addressed is whether the normal equations, $K_o R_y = R_{xy}$, have a solution (*i.e.*, whether the equations are consistent), and if so, whether the solution is unique. It turns out that, as in the deterministic case studied in Ch. 2, the normal equations will always have a solution, although the solution will be nonunique when R_y is singular. However, no matter which solution is used, the l.l.m.s.e. $\hat{\mathbf{x}} = K_o\mathbf{y}$ will be unique, and so will the corresponding m.m.s.e., $P(K_o)$. Both these facts are readily suggested by the geometric formulation to be given shortly. A rigorous, though less intuitive, algebraic proof is suggested in Prob. 3.22. We just state the results here.

Theorem 3.2.3 (Nonunique Solutions) *Even if $R_y = E\mathbf{y}\mathbf{y}^*$ is singular, the normal equations $K_o R_y = R_{xy}$ will be consistent, and there will be many solutions. No matter which solution K_o is used, the corresponding l.l.m.s. estimator $\hat{\mathbf{x}} = K_o\mathbf{y}$ will, however, be unique, and so of course will $P(K_o)$.* ■

Proof: See Prob. 3.22. ◆

3.2.4 Nonzero-Mean Values and Centering

When the random variables $\{\mathbf{x}, \mathbf{y}\}$ have known nonzero means, the simplest (*and best*) way is to proceed by *centering* the random variables. That is, if $E\mathbf{x} = m_x$ and $E\mathbf{y} = m_y$, we define

$$\mathbf{x}^o \triangleq \mathbf{x} - m_x, \qquad \mathbf{y}^o \triangleq \mathbf{y} - m_y.$$

[2] More precisely, we should say that the random variable $c^*\mathbf{y}$ is zero almost surely.

The transformation between $\{\mathbf{x}, \mathbf{y}\}$ and $\{\mathbf{x}^o, \mathbf{y}^o\}$ is reversible, so there is no loss of information in making such a transformation. Note further that

$$\begin{aligned} E\mathbf{x}^o\mathbf{x}^{o*} &= E(\mathbf{x} - m_x)(\mathbf{x} - m_x)^* = E\mathbf{x}\mathbf{x}^* - m_x m_x^*, \\ &\triangleq R_x, \quad \text{the covariance matrix of } \mathbf{x}. \\ E\mathbf{x}^o\mathbf{y}^{o*} &= E(\mathbf{x} - m_x)(\mathbf{y} - m_y)^* = E\mathbf{x}\mathbf{y}^* - m_x m_y^*, \\ &\triangleq R_{xy}, \quad \text{the cross-covariance matrix of } \mathbf{x} \text{ and } \mathbf{y}. \end{aligned}$$

Now

$$\hat{\mathbf{x}}^o = (E\mathbf{x}^o\mathbf{y}^{o*})(E\mathbf{y}^o\mathbf{y}^{o*})^{-1}\mathbf{y}^o,$$

or, equivalently,

$$\begin{aligned} \hat{\mathbf{x}} &= m_x + R_{xy}R_y^{-1}(\mathbf{y} - m_y), \\ &= R_{xy}R_y^{-1}\mathbf{y} + (m_x - R_{xy}R_y^{-1}m_y). \end{aligned} \tag{3.2.8}$$

We see that, strictly speaking, the *linear* minimum-mean-square estimator of $\mathbf{x}$ given $\mathbf{y}$ is really an *affine* function of $\mathbf{y}$ rather than a linear function; however, one can easily be excused for continuing to call $\hat{\mathbf{x}}$ a linear function of $\mathbf{y}$.

To appreciate the simplicity of using the centering approach to handle the case of nonzero mean values, we urge the reader to work out Probs. 3.1 and 3.2 at this stage (even under the simplifying assumption that $\mathbf{x}$ and $\mathbf{y}$ are scalar random variables). One will soon appreciate the virtue of making our standing assumption that all random variables have mean value equal to zero, unless otherwise specified (or otherwise unlikely). Moreover, to avoid confusion with some common terminology in the statistics literature, it is useful to be aware of the issue raised in Prob. 3.3 on Minimum Variance Unbiased Estimators (MVUEs).

3.2.5 Estimators for Complex-Valued Random Variables

We remarked earlier in the footnote in Sec. 3.2.1 that some care is needed in defining the linear estimation problem for *complex-valued* random variables. We now examine this issue.

Given complex-valued zero-mean random variables $\mathbf{x}$ and $\mathbf{y}$, we constructed in Sec. 3.2 an optimal linear estimator for $\mathbf{x}$ as

$$\hat{\mathbf{x}} = K_o\mathbf{y}, \tag{3.2.9}$$

for any matrix K_o satisfying the normal equations

$$K_o R_y = R_{xy}. \tag{3.2.10}$$

This estimator is optimal in the sense that, among all estimators of the form $\hat{\mathbf{x}} = K\mathbf{y}$, it is the one that minimizes the error covariance matrix $E\tilde{\mathbf{x}}\tilde{\mathbf{x}}^*$, where $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$.

Now a complex-valued random variable is one whose real and imaginary parts are real-valued random variables. Correspondingly, expression (3.2.9) can be thought of as also defining an estimator for the real and imaginary parts of $\mathbf{x}$ in terms of the real and imaginary parts of $\mathbf{y}$. In order to highlight this point, let us denote the random real and imaginary parts of $\mathbf{x}$, $\mathbf{y}$, and $\hat{\mathbf{x}}$ by

$$\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I, \qquad \mathbf{y} = \mathbf{y}_R + j\mathbf{y}_I, \qquad \hat{\mathbf{x}} = \hat{\mathbf{x}}_R + j\hat{\mathbf{x}}_I,$$

respectively. We also let $K_{o,R}$ and $K_{o,I}$ denote the real and imaginary parts of the optimal coefficient matrix K_o in (3.2.10). It then follows from (3.2.9) that

$$\hat{\mathbf{x}}_R + j\hat{\mathbf{x}}_I = (K_{o,R} + jK_{o,I})(\mathbf{y}_R + j\mathbf{y}_I),$$

or, equivalently, in matrix notation,

$$\begin{bmatrix} \hat{\mathbf{x}}_R \\ \hat{\mathbf{x}}_I \end{bmatrix} = \underbrace{\begin{bmatrix} K_{o,R} & -K_{o,I} \\ K_{o,I} & K_{o,R} \end{bmatrix}}_{M_o} \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_I \end{bmatrix}, \tag{3.2.11}$$

where we have denoted the coefficient matrix multiplying $\mathrm{col}\{\mathbf{y}_R, \mathbf{y}_I\}$ by M_o. Note that M_o has a *special* structure; its (1, 1) and (2, 2) blocks are identical, whereas its (1, 2) block is the negative of its (2, 1) block. Moreover, its entries are completely determined by the entries of the covariance and cross-covariance matrices R_y and R_{xy}. These facts have important implications on the meaning of the optimality of the linear estimator defined by (3.2.9)–(3.2.10), as we now explain.

We first note that expression (3.2.11), which followed from (3.2.9), is not really the only way to construct a linear estimator for $(\mathbf{x}_R, \mathbf{x}_I)$. To explain why, observe that we can rewrite R_y and R_{xy} in terms of the covariance and cross-covariance matrices of the real and imaginary parts of $\mathbf{x}$ and $\mathbf{y}$ as follows:

$$R_y = (R_{y_R} + R_{y_I}) + j(R_{y_I,y_R} - R_{y_R,y_I}), \tag{3.2.12}$$

and

$$R_{xy} = (R_{x_R,y_R} + R_{x_I,y_I}) + j(R_{x_I,y_R} - R_{x_R,y_I}). \tag{3.2.13}$$

These expressions show that knowledge of the second-order statistics (R_y, R_{xy}) is only equivalent to knowledge of the sums and differences

$$(R_{y_R} + R_{y_I}), \quad (R_{x_R,y_R} + R_{x_I,y_I}), \quad (R_{y_I,y_R} - R_{y_R,y_I}), \quad (R_{x_I,y_R} - R_{x_R,y_I}). \tag{3.2.14}$$

These sums and differences do not completely characterize the covariance and cross-covariance matrices of the real and imaginary parts of $\mathbf{x}$ and $\mathbf{y}$. In other words, knowledge of the quantities in (3.2.14), or equivalently knowledge of $\{R_y, R_{xy}\}$, is not enough to determine the individual quantities

$$\{R_{y_R}, R_{y_I}, R_{y_I,y_R}, R_{x_R,y_R}, R_{x_R,y_I}, R_{x_I,y_R}, R_{x_I,y_I}\}. \tag{3.2.15}$$

For this reason, we say that the linear estimator (3.2.9) (or, equivalently, (3.2.11)) is only optimal with respect to the information available from the first- and second-order statistics of $\mathbf{x}$ and $\mathbf{y}$, *viz.*, $\{R_y, R_{xy}\}$ in (3.2.12) and (3.2.13). However, the estimator is not optimal had we been given instead all the first- and second-order statistics of the real and imaginary parts of $\mathbf{x}$ and $\mathbf{y}$, *viz.*, the more complete information in (3.2.15). The covariance and cross-covariance matrices in (3.2.15) provide more information about the random variables $\{\mathbf{x}, \mathbf{y}\}$ than what is contained in (3.2.14); knowledge of these matrices completely determines the sums and differences in (3.2.14), but the contrary is not true.

The possible lack of optimality of the estimator (3.2.11) can be seen from the special structure of the coefficient matrix M_o. If we assume that we are given the information in (3.2.15) rather than (3.2.14), then an alternative linear estimator can be defined for $\{\mathbf{x}_R, \mathbf{x}_I\}$ that can be shown to outperform the estimator defined by (3.2.11) in the sense that it leads to a smaller error covariance matrix $E\tilde{\mathbf{x}}\tilde{\mathbf{x}}^*$. It also leads to a different optimal coefficient matrix M_o without the restrictions in (3.2.11).

So assume that we are given the covariance and cross-covariance matrices (3.2.15). Then an optimal linear estimator for $\mathbf{x}$ can be defined as one that optimally estimates its real and imaginary parts as

$$\begin{bmatrix} \hat{\mathbf{x}}_R \\ \hat{\mathbf{x}}_I \end{bmatrix} = N_o \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_I \end{bmatrix}, \tag{3.2.16}$$

where the coefficient matrix N_o is determined such that the resulting error covariance matrix is minimized, *i.e.*,

$$P(N_o) \triangleq E\left(\begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{x}}_R \\ \hat{\mathbf{x}}_I \end{bmatrix}\right)\left(\begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{x}}_R \\ \hat{\mathbf{x}}_I \end{bmatrix}\right)^T = \text{minimum}. \tag{3.2.17}$$

As noted in part (b) of Prob. 3.4, N_o also results in the minimum value for $E\|\tilde{\mathbf{x}}_R\|^2 + E\|\tilde{\mathbf{x}}_I\|^2$ since

$$E\left(\begin{bmatrix} \tilde{\mathbf{x}}_R \\ \tilde{\mathbf{x}}_I \end{bmatrix}^T \begin{bmatrix} \tilde{\mathbf{x}}_R \\ \tilde{\mathbf{x}}_I \end{bmatrix}\right) = E\|\tilde{\mathbf{x}}_R\|^2 + E\|\tilde{\mathbf{x}}_I\|^2,$$

where we are denoting the real and imaginary parts of the error vector $\tilde{\mathbf{x}}$ by $\tilde{\mathbf{x}}_R$ and $\tilde{\mathbf{x}}_I$, respectively. Likewise, as mentioned prior to Thm. 3.2.2 and in part (b) of the same Prob. 3.4, K_o also minimizes $E\tilde{\mathbf{x}}^*\tilde{\mathbf{x}}$, which is easily seen to be equal to $E\|\tilde{\mathbf{x}}_R\|^2 + E\|\tilde{\mathbf{x}}_I\|^2$ since $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_R + j\tilde{\mathbf{x}}_I$. In other words, the linear estimators defined by (3.2.9) and (3.2.16) are minimizing the same cost function. The difference between them is that they use different prior statistical information; one uses (3.2.14) while the other uses (3.2.15).

By following the same arguments as in Sec. 3.2.1, it is easy to see that the fundamental equations that correspond to problem (3.2.16) are given by

$$N_o \begin{bmatrix} R_{y_R,y_R} & R_{y_R,y_I} \\ R_{y_I,y_R} & R_{y_I,y_I} \end{bmatrix} = \begin{bmatrix} R_{x_R,y_R} & R_{x_R,y_I} \\ R_{x_I,y_R} & R_{x_I,y_I} \end{bmatrix}. \tag{3.2.18}$$

The resulting matrix N_o is, in general, quite different from the matrix M_o in (3.2.11). Prob. 3.5 gives a simple scalar example that confirms this statement. However, since the linear estimator (3.2.9) imposes a special structure on the resulting matrix M_o, it is clear that the linear estimator (3.2.16) will always lead to a smaller mean-square error (assuming the same information (3.2.15) is available to both estimators).

One might then wonder if both estimators can ever lead to similar results. In fact, we can show that in the following cases both constructions lead to the same estimator:

(a) All random variables are real.

(b) The random variables $\{\mathbf{x}, \mathbf{y}\}$ are complex-valued but satisfy the requirement

$$E\mathbf{y}\mathbf{y}^T = 0 \quad \text{and} \quad E\mathbf{x}\mathbf{y}^T = 0. \tag{3.2.19}$$

A zero-mean complex-valued random variable $\mathbf{z}$ that satisfies $E\mathbf{z}\mathbf{z}^T = 0$ is said to be *circular*.[3] Hence, (3.2.19) requires the observation vector $\mathbf{y}$ to be circular, and the pair $\{\mathbf{x}, \mathbf{y}\}$ to be second-order circular, *i.e.*, $E\mathbf{x}\mathbf{y}^T = 0$.

The real case is a trivial (degenerate) situation and one can easily verify that M_o and N_o indeed coincide, especially since all imaginary parts become zero. The circular case guarantees, in view of (3.2.19), that the information contained in (3.2.14) is equivalent to the information contained in (3.2.15). To verify this claim, note that we can write

$$E\mathbf{y}\mathbf{y}^T = (R_{y_R} - R_{y_I}) + j(R_{y_I y_R} + R_{y_R y_I}),$$

and

$$E\mathbf{x}\mathbf{y}^T = (R_{x_R,y_R} - R_{x_I,y_I}) + j(R_{x_I y_R} + R_{x_R y_I}).$$

It then follows from (3.2.19) that we must necessarily have

$$R_{y_R} = R_{y_I}, \quad R_{y_I y_R} = -R_{y_R y_I}, \quad R_{x_R,y_R} = R_{x_I,y_I}, \quad R_{x_I y_R} = -R_{x_R y_I}.$$

These relations, along with the given sums and differences (3.2.14), allow us to determine all the quantities in (3.2.15). Hence, under the circular assumption, both estimators (the one with M_o and the one with N_o) have access to the same first- and second-order statistics and the solutions should therefore coincide. One can, in fact, go further and verify explicitly that $M_o = N_o$, which we leave as an exercise for the reader.

In summary, our discussion establishes the following conclusions:

(i) If all random variables are real or complex-valued and circular (*cf.* (3.2.19)), then the linear estimator defined in Sec. 3.2.1 is optimal in the sense that it leads to the minimum-mean-square error.

(ii) For general complex-valued random variables, if the prior information that is available to us is only $\{R_y, R_{xy}\}$ (which is equivalent to knowledge of the sums and differences in (3.2.14)), then the linear estimator defined in Sec. 3.2.1 is still optimal in the sense that it leads to the minimum mean-square error relative to the available information.

(iii) But if for general complex-valued random variables, the available prior information is more than simply $\{R_y, R_{xy}\}$, and it includes all the covariances and cross-covariances of the real and imaginary parts of the random variables involved (as in (3.2.15)), then the linear estimator defined in Sec. 3.2.1 is *not* optimal. We can obtain the optimal linear estimator by reducing the problem to the real context, as done in (3.2.16)–(3.2.17), and then proceeding with a linear estimator design just as in Sec. 3.2.1 but now working with the extended observation vector $\{\mathbf{y}_R, \mathbf{y}_I\}$ and with the normal equations (3.2.18).

[3] Note that in the definition of a circular random variable we have used transposes and not "conjugate" transposes; for us $E\mathbf{y}\mathbf{y}^* = 0$ always implies $\mathbf{y} = 0$. Also, a simple example of a circular random variable $\mathbf{z}$ is one whose real and imaginary parts are uncorrelated and have the same covariance matrices, *i.e.*, $E\mathbf{z}_R\mathbf{z}_I^T = 0$ and $R_{z_R} = R_{z_I}$.

Regardless of whether we reduce the problem to a real context, as in step (iii) above, or continue to work in the complex context, as in parts (i) and (ii), the same normal equations (and the same fundamental geometric insights of the next section) will still be applicable. The only difference is that we might be working either with the given vectors $\{\mathbf{x}, \mathbf{y}\}$ or with extended versions of them, $\{\mathbf{x}_R, \mathbf{x}_I, \mathbf{y}_R, \mathbf{y}_I\}$. For this reason, we shall continue to use, here and in all future chapters, and without loss of generality, the formulation introduced without comment in Sec. 3.2.1.

3.3 A GEOMETRIC FORMULATION

The results in Sec. 3.2, and especially the somewhat detailed arguments in Sec. 3.2.3, will become fairly obvious if we introduce a geometric formulation of the above problem. [The reader may find it useful to reread at this point the presentation in Sec. 2.3; note that there are certain differences in the details, especially in the definition of inner products.]

3.3.1 The Orthogonality Condition

We remarked above that the optimum estimator $\hat{\mathbf{x}} = K_o\mathbf{y}$ was determined by any solution K_o of the equation

$$K_o R_y = R_{xy}. \tag{3.3.1}$$

Now this equation is equivalent to $K_o E\mathbf{y}\mathbf{y}^* = E\mathbf{x}\mathbf{y}^*$, or

$$E(\mathbf{x} - K_o\mathbf{y})\mathbf{y}^* = 0. \tag{3.3.2}$$

This suggests that if we can regard the random variables $\mathbf{x}$ and $\mathbf{y}$ as vectors (*i.e.*, elements) in an inner product space, with inner product defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle \triangleq E\mathbf{x}\mathbf{y}^*, \tag{3.3.3}$$

then the above condition has the geometric meaning that $\langle \mathbf{x} - K_o\mathbf{y}, \mathbf{y} \rangle = 0$, *i.e.*, that $\mathbf{x} - K_o\mathbf{y}$ is orthogonal to $\mathbf{y}$, written as

$$\mathbf{x} - K_o\mathbf{y} \perp \mathbf{y}.$$

Note that this geometric interpretation holds even when $\mathbf{x}$ and $\mathbf{y}$ are *scalar* random variables.

All this is very suggestive, except that one might raise at least a couple of relevant issues:

1. In our applications, $\mathbf{x}$ and $\mathbf{y}$ will generally be vector-valued, *e.g.*, $\mathbf{x} \in \mathbb{C}^{n\times 1}$, $\mathbf{y} \in \mathbb{C}^{p\times 1}$, random variables, so that the inner product $\langle \mathbf{x}, \mathbf{y}\rangle \triangleq E\mathbf{x}\mathbf{y}^*$ will often be a (rectangular) matrix, and not a scalar as is usually assumed in the theory and textbooks on linear vector spaces. Is this a problem?
2. More fundamentally, we need to ask: in what vector space are the random variables defined?

The answer to the first question is easy. To be a well-defined inner product, it is only required that the following conditions be satisfied:

1. Linearity: $\langle a_1\mathbf{x}_1 + a_2\mathbf{x}_2, \mathbf{y}\rangle = a_1\langle \mathbf{x}_1, \mathbf{y}\rangle + a_2\langle \mathbf{x}_2, \mathbf{y}\rangle$, for any $a_1, a_2 \in \mathbb{C}$.
2. Reflexivity: $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle^*$.
3. Nondegeneracy: $\|\mathbf{x}\|^2 \triangleq \langle \mathbf{x}, \mathbf{x}\rangle$ is zero only when $\mathbf{x} = 0$.

These requirements are discussed in books on linear algebra. In most books, inner products are taken to be scalar-valued, but this is not essential. We discuss the issue further in App. 4.A, which the reader may skim at this stage with particular attention to the examples.

We can readily check that $\langle \mathbf{x}, \mathbf{y}\rangle \triangleq E\mathbf{x}\mathbf{y}^*$ satisfies the above conditions and is thus a legitimate inner product. Note that the properties of inner products imply that, for example,

$$\langle \mathbf{x}, K\mathbf{y}\rangle = \langle K\mathbf{y}, \mathbf{x}\rangle^* = (K\langle \mathbf{y}, \mathbf{x}\rangle)^* = \langle \mathbf{x}, \mathbf{y}\rangle K^*,$$

which is consistent with the calculation

$$\langle \mathbf{x}, K\mathbf{y}\rangle = E\mathbf{x}(K\mathbf{y})^* = (E\mathbf{x}\mathbf{y}^*)K^* = \langle \mathbf{x}, \mathbf{y}\rangle K^*.$$

But in what sense can we regard a random variable $\mathbf{x}$ as a vector? Here is where the formal definition of a random variable, as given in the first chapter of almost any book on probability theory, becomes useful. A scalar-valued random variable is a function from a probability space, Ω, to the space of complex numbers:

$$\mathbf{x} : \Omega \to \mathbb{C}.$$

[There is an additional condition called measurability that is needed to ensure well-behavedness of the definition; however we can ignore this issue here. Interested readers can find more on this matter in books on probability theory.]

For a very simple example, we can take Ω = the real line, $n = 1$, and define a scalar-valued random variable

$$\mathbf{x}(\omega) = 1, \quad \omega > 0; \quad \mathbf{x}(\omega) = -1, \quad \omega \le 0.$$

Now we can imagine a space, with one coordinate axis for *every* $\omega \in \Omega$, and think of the random variable $\mathbf{x}$ as a *vector* with value $\mathbf{x}(\omega)$ along the coordinate axis ω. Once we

accept this abstraction, it is not too much to go a bit further and consider vector-valued random variables, where a collection of say n complex variables is associated with each point $\omega \in \Omega$. Given the confusion that beginners often have with this concept, it is worth restating. When we think of a vector-valued random variable, $\mathbf{x}$, with say N components, we think of an $N \times 1$ column vector that can take lots of different values, a different one in each "trial" or each "experiment". All these possible different values are summarized in the single $N \times 1$ vector $\mathbf{x}(\omega)$, which now lives in the abstract space Ω: for each $\omega \in \Omega$, we get one of the different values the random vector can take. We suggest also that the reader might find it helpful, at least initially, to think of only scalar random variables, and to let the power of matrix notation carry the burden of the fact that they may be vector-valued.

Once we have a framework for regarding random variables as vectors, we can use the projection theorem of inner product spaces to obtain least-mean-squares estimators. This has been indirectly demonstrated above (using (3.3.2)–(3.3.3)), but let us, for emphasis, present the direct arguments.

Thus to estimate the random variable $\mathbf{x}$ given $\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$, we consider these variables as vectors in the space Ω with the inner product as defined above. Then to find a linear combination, $K\mathbf{y}$, that minimizes $E[\mathbf{x} - K\mathbf{y}][\mathbf{x} - K\mathbf{y}]^*$, we have to project $\mathbf{x}$ onto the linear space spanned by the $\{\mathbf{y}_i, i = 0, \ldots, N\}$, say $\mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$; see Fig. 3.1. Then the projection is defined by $\hat{\mathbf{x}} = K_o\mathbf{y}$, where K_o is such that

$$\mathbf{x} - K_o\mathbf{y} \perp \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\},$$

i.e., $\langle \mathbf{x}, \mathbf{y}_i\rangle = K_o\langle \mathbf{y}, \mathbf{y}_i\rangle$ for $i = 0, 1, \ldots N$, or $\langle \mathbf{x}, \mathbf{y}\rangle = K_o\langle \mathbf{y}, \mathbf{y}\rangle$, which can also be written as $R_{xy} = K_oR_y$. The properties of inner product spaces ensure[4] that the projection exists and is unique even if R_y is singular (see Prob. 3.22).

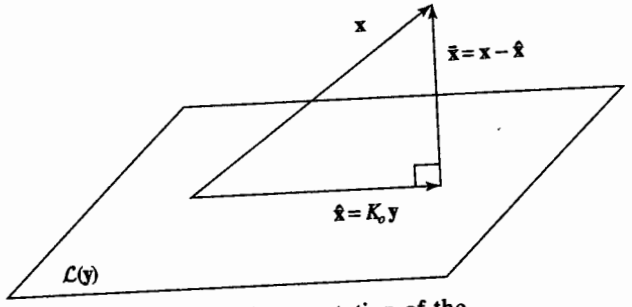


Figure 3.1 Geometric interpretation of the l.l.m.s. estimator.

[4] Here again we are ignoring some issues regarding what is called completeness of the inner product space of random variables: Adding to all finite linear combinations of the specific random variables, their limits in the norm gives a complete space called a Hilbert space. Our random variables are elements of this Hilbert space. Interested readers can find a reasonably elementary discussion in Picinbono (1993), or of course in books on Hilbert space theory.

To gain a little more familiarity with these arguments, note that we can readily prove that the projection $K_o\mathbf{y}$ is the element that minimizes the mean-square error. For if $K\mathbf{y}$ is some other element in $\mathcal{L}\{\mathbf{y}\}$, we can write

$$\begin{aligned}\|\mathbf{x} - K\mathbf{y}\|^2 &= \|\mathbf{x} - K_o\mathbf{y} + K_o\mathbf{y} - K\mathbf{y}\|^2,\\ &= \|\mathbf{x} - K_o\mathbf{y}\|^2 + \langle \mathbf{x} - K_o\mathbf{y}, K_o\mathbf{y} - K\mathbf{y}\rangle +\\ &\quad \langle K_o\mathbf{y} - K\mathbf{y}, \mathbf{x} - K_o\mathbf{y}\rangle + \|K_o\mathbf{y} - K\mathbf{y}\|^2.\end{aligned}$$

But the two cross-terms are zero, because $(K_o - K)\mathbf{y} \in \mathcal{L}\{\mathbf{y}\}$, and by definition of the projection, $\mathbf{x} - K_o\mathbf{y} \perp \mathcal{L}\{\mathbf{y}\}$. Hence

$$\|\mathbf{x} - K\mathbf{y}\|^2 = \|\mathbf{x} - K_o\mathbf{y}\|^2 + \|K_o\mathbf{y} - K\mathbf{y}\|^2,$$

which is clearly minimized by choosing $K = K_o$.

In the same vein, we can make use of the Pythagoras theorem to find various expressions for the error,

$$\begin{aligned}E\tilde{\mathbf{x}}\tilde{\mathbf{x}}^* &= \|\tilde{\mathbf{x}}\|^2 = \|\mathbf{x}\|^2 - \|\hat{\mathbf{x}}\|^2,\\ &= \langle \mathbf{x}, \mathbf{x}\rangle - \langle K_o\mathbf{y}, K_o\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{x}\rangle - K_o\langle \mathbf{y}, \mathbf{y}\rangle K_o^*,\\ &= R_x - K_oR_yK_o^* = R_x - R_{xy}K_o^* = R_x - K_oR_{yx}.\end{aligned}$$

Alternatively, we could also have used orthogonality directly,

$$\|\tilde{\mathbf{x}}\|^2 = \langle \mathbf{x} - K_o\mathbf{y}, \mathbf{x} - K_o\mathbf{y}\rangle = \langle \mathbf{x} - K_o\mathbf{y}, \mathbf{x}\rangle = R_x - K_oR_{yx}.$$

Recall that when R_y is invertible, we can also write

$$\|\tilde{\mathbf{x}}\|^2 = R_x - R_{xy}R_y^{-1}R_{yx}.$$

We shall compactly describe the use of the geometric formulation for finding least-mean-squares estimators as using orthogonality.

Lemma 3.3.1 (The Orthogonality Condition) *The linear least-mean-squares estimator (l.l.m.s.e.) of a random variable* $\mathbf{x}$ *given a set of other random variables* $\mathbf{y}$ *is characterized by the fact that the error* $\tilde{\mathbf{x}}$ *in the estimator is orthogonal to (i.e., uncorrelated with) each of the random variables used to form the estimator. Equivalently, the l.l.m.s.e. is the projection of* $\mathbf{x}$ *onto* $\mathcal{L}(\mathbf{y})$. ∎

Projection onto the linear space $\mathcal{L}(\mathbf{y})$ (which we denote here by $\hat{\mathbf{x}}_{|\mathbf{y}}$) has the important properties

$$\widehat{(\mathbf{x}_1 + \mathbf{x}_2)}_{|\mathbf{y}} = \hat{\mathbf{x}}_{1|\mathbf{y}} + \hat{\mathbf{x}}_{2|\mathbf{y}} \tag{3.3.4}$$

and

$$\hat{\mathbf{x}}_{|\mathbf{y}_1,\mathbf{y}_2} = \hat{\mathbf{x}}_{|\mathbf{y}_1} + \hat{\mathbf{x}}_{|\mathbf{y}_2} \text{ if, and only if, } \mathbf{y}_1 \perp \mathbf{y}_2. \tag{3.3.5}$$

These geometrically intuitive properties can be formally verified by using the explicit formula $\hat{\mathbf{x}}_{|\mathbf{y}} = \langle \mathbf{x}, \mathbf{y}\rangle \|\mathbf{y}\|^{-2}\mathbf{y}$.

Final Remarks. The abstractions introduced in this section may cause some understandable uneasiness for beginning readers, and this may be compounded when put together with the perhaps (more familiar) abstractions of the Euclidean vector spaces of Sec. 2.3. The facts are that different vector spaces can have different kinds of elements: (ordinary) complex column vectors or complex row vectors, or scalar random variables, or vector-valued (row or column) random variables. We can also have vector spaces in which the elements (vectors) are column vectors whose components are say $1 \times p$ row (or $p \times 1$ column) vectors themselves; when the probability space is finite-dimensional, say p-dimensional, then the space of vector-valued random variables is exactly of this type. The abstractions will get more familiar as we progress through the book (see also App. 4.A).

When tempted to worry too much about what is going on (*e.g.*, what exactly is the space Ω?), one might bring to mind an aphorism attributed to Albert Einstein: "Make things as simple as possible, but not simpler."

3.3.2 Examples

The geometric viewpoint and the orthogonality condition will be extensively used hereafter. Here are a few examples, the last of which shows the value of the stochastic/geometric point of view in some purely algebraic problems.

Example 3.3.1 Consider a zero-mean real-valued stationary process $\{\mathbf{y}(t)\}$ with autocovariance function $\langle \mathbf{y}(t), \mathbf{y}(t-\tau)\rangle = R_y(\tau)$. Find the linear least-mean squares estimator of its integral $\int_0^T \mathbf{y}(t)dt$ in terms of its endpoints $\mathbf{y}(0)$ and $\mathbf{y}(T)$.

Solution: Let $\mathbf{z} \triangleq \int_0^T \mathbf{y}(t)dt$ and $\hat{\mathbf{z}} = a\mathbf{y}(0) + b\mathbf{y}(T)$, for some constants a and b that we have to find. The orthogonality condition

$$\mathbf{z} - \hat{\mathbf{z}} = \int_0^T \mathbf{y}(t)dt - a\mathbf{y}(0) - b\mathbf{y}(T) \perp \{\mathbf{y}(0), \mathbf{y}(T)\}$$

yields the equations

$$\left\langle \int_0^T \mathbf{y}(t)dt - a\mathbf{y}(0) - b\mathbf{y}(T), \mathbf{y}(0)\right\rangle = \int_0^T R_y(t)dt - aR_y(0) - bR_y(T) = 0,$$

$$\left\langle \int_0^T \mathbf{y}(t)dt - a\mathbf{y}(0) - b\mathbf{y}(T), \mathbf{y}(T)\right\rangle = \underbrace{\int_0^T R_y(t-T)dt}_{\int_0^T R_y(t)dt} - a\underbrace{R_y(-T)}_{R_y(T)} - bR_y(0) = 0,$$

or in matrix form,

$$\begin{bmatrix} R_y(0) & R_y(T) \\ R_y(T) & R_y(0)\end{bmatrix}\begin{bmatrix} a \\ b\end{bmatrix} = \begin{bmatrix} 1 \\ 1\end{bmatrix}\int_0^T R_y(t)dt.$$

Solving the above system, we finally obtain

$$\hat{\mathbf{z}} = \frac{\int_0^T R_y(t)dt}{R_y(0) + R_y(T)}[\mathbf{y}(0) + \mathbf{y}(T)].$$

◆

Example 3.3.2 (Exponential Covariance) Consider a scalar zero-mean stationary random process $\{\mathbf{x}(t)\}$ with autocovariance function

$$\langle \mathbf{x}(t), \mathbf{x}(t-\tau)\rangle = R_x(\tau) = e^{-\alpha|\tau|}.$$

Find the linear least-mean squares estimator of $\mathbf{x}(T_3)$ using $\mathbf{x}(T_1)$ and $\mathbf{x}(T_2)$, assuming that $T_1 < T_2 < T_3$.

Solution: Define

$$\mathbf{z} = \mathbf{x}(T_3) \quad \text{and} \quad \mathbf{y} = \text{col}\{\mathbf{x}(T_1), \mathbf{x}(T_2)\}.$$

Now the desired estimator will be $\hat{\mathbf{z}} = k\mathbf{y}$ where, using the orthogonality principle, $k = \begin{bmatrix} k_1 & k_2 \end{bmatrix}$ satisfies

$$\begin{bmatrix} k_1 & k_2 \end{bmatrix} \begin{bmatrix} R_x(0) & R_x(T_1 - T_2) \\ R_x(T_2 - T_1) & R_x(0) \end{bmatrix} = \begin{bmatrix} R_x(T_3 - T_1) & R_x(T_3 - T_2) \end{bmatrix}.$$

Using the exponential autocovariance function, we have

$$\begin{bmatrix} k_1 & k_2 \end{bmatrix} \begin{bmatrix} 1 & e^{-\alpha(T_2 - T_1)} \\ e^{-\alpha(T_2 - T_1)} & 1 \end{bmatrix} = \begin{bmatrix} e^{-\alpha(T_2 - T_1)} & 1 \end{bmatrix} e^{-\alpha(T_3 - T_2)}.$$

The vector on the right-hand side is a multiple of the second row of R_y. Thus we readily see that

$$\begin{bmatrix} k_1 & k_2 \end{bmatrix} = \begin{bmatrix} 0 & e^{-\alpha(T_3 - T_2)} \end{bmatrix},$$

and therefore

$$\hat{\mathbf{z}} = \hat{\mathbf{x}}(T_3|T_2, T_1) = e^{-\alpha(T_3 - T_2)}\mathbf{x}(T_2) = \hat{\mathbf{x}}(T_3|T_2).$$

Note the surprising fact that the l.l.m.s.e. of $\mathbf{x}(T_3)$ using $\mathbf{x}(T_1)$ and $\mathbf{x}(T_2)$ depends only upon the most recent observation $\mathbf{x}(T_2)$. Note, moreover, that since our choice of $T_1 < T_2$ was arbitrary, we may guess that the l.l.m.s.e. of $\mathbf{x}(T_3)$ given $\{\mathbf{x}(\tau), -\infty < \tau \le T_2\}$ is equal to the l.l.m.s.e. of $\mathbf{x}(T_3)$ given $\mathbf{x}(T_2)$. This can be confirmed by checking that $\mathbf{x}(T_3) - e^{-\alpha(T_3 - T_2)}\mathbf{x}(T_2) \perp \mathbf{x}(\tau)$, $\tau < T_2$.

Stochastic processes that have this property are referred to as *wide-sense Markov* processes and will be studied in detail in Ch. 5. ◆

Example 3.3.3 (A Matrix Cauchy-Schwarz Inequality) Recall from Sec. 3.2.2 that, when R_x is nonsingular, the covariance matrix of the error in estimating $\mathbf{y}$ given $\mathbf{x}$ is equal to $R_y - R_{yx}R_x^{-1}R_{xy}$. Now since the covariance matrix of any vector-valued random variable is nonnegative-definite, we have $R_y - R_{yx}R_x^{-1}R_{xy} \ge 0$, or, equivalently,

$$\|\mathbf{y}\|^2 - \langle \mathbf{y}, \mathbf{x}\rangle \|\mathbf{x}\|^{-2} \langle \mathbf{x}, \mathbf{y}\rangle \ge 0. \tag{3.3.6}$$

This is a matrix generalization of the well-known *Cauchy-Schwarz inequality*. ◆

Example 3.3.4 (Stochastic Proof of a Matrix Identity) Provide a stochastic derivation of the identity

$$\begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}^{-1} = \tag{3.3.7}$$

$$\begin{bmatrix} R_x - R_{xy}R_y^{-1}R_{yx} & 0 \\ 0 & R_y - R_{yx}R_x^{-1}R_{xy} \end{bmatrix}^{-1} \begin{bmatrix} I & -R_{xy}R_y^{-1} \\ -R_{yx}R_x^{-1} & I \end{bmatrix}.$$

Solution: Note that we can project $\mathbf{x}$ onto $\mathcal{L}\{\mathbf{y}\}$ and form the error $\tilde{\mathbf{x}}_{|\mathbf{y}} = \mathbf{x} - R_{xy}R_y^{-1}\mathbf{y}$. Combining this with a similar expression for $\tilde{\mathbf{y}}_{|\mathbf{x}}$, we may thus write

$$\begin{bmatrix} \tilde{\mathbf{x}}_{|\mathbf{y}} \\ \tilde{\mathbf{y}}_{|\mathbf{x}} \end{bmatrix} = \begin{bmatrix} I & -R_{xy}R_y^{-1} \\ -R_{yx}R_x^{-1} & I \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}. \tag{3.3.8}$$

Since $\tilde{\mathbf{x}}_{|\mathbf{y}}$ is orthogonal to $\mathbf{y}$ and $\tilde{\mathbf{y}}_{|\mathbf{x}}$ is orthogonal to $\mathbf{x}$, the cross-covariance matrix of the random variables $\{\tilde{\mathbf{x}}_{|\mathbf{y}}, \tilde{\mathbf{y}}_{|\mathbf{x}}\}$ with $\{\mathbf{x}, \mathbf{y}\}$ is diagonal. Indeed,

$$\left\langle \begin{bmatrix} \tilde{\mathbf{x}}_{|\mathbf{y}} \\ \tilde{\mathbf{y}}_{|\mathbf{x}} \end{bmatrix}, \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \right\rangle = \begin{bmatrix} R_x - R_{xy}R_y^{-1}R_{yx} & 0 \\ 0 & R_y - R_{yx}R_x^{-1}R_{xy} \end{bmatrix}. \tag{3.3.9}$$

But we could have used (3.3.8) to directly write the cross-covariance matrix as

$$\left\langle \begin{bmatrix} \tilde{\mathbf{x}}_{|\mathbf{y}} \\ \tilde{\mathbf{y}}_{|\mathbf{x}} \end{bmatrix}, \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \right\rangle = \begin{bmatrix} I & -R_{xy}R_y^{-1} \\ -R_{yx}R_x^{-1} & I \end{bmatrix} \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}. \tag{3.3.10}$$

Combining (3.3.9) and (3.3.10) leads to (3.3.7). ◆

Remark 2. Several of the matrix identities in App. A have similarly elegant stochastic proofs, as we encourage readers to check for themselves. Though these identities are often easy to confirm algebraically, once the exact form is given, the stochastic derivations help to recall the identities and also to explore variations. ◆

3.4 LINEAR MODELS

An extremely important special case that will often arise in our analysis occurs when $\mathbf{y}$ and $\mathbf{x}$ are *linearly* related, say as

$$\mathbf{y} = H\mathbf{x} + \mathbf{v}, \tag{3.4.1}$$

where $H \in \mathbb{C}^{p \times n}$ is a known matrix and $\mathbf{v}$ is a zero-mean random-noise vector uncorrelated with $\mathbf{x}$. Assume that $R_x = \langle \mathbf{x}, \mathbf{x}\rangle$ and $R_v = \langle \mathbf{v}, \mathbf{v}\rangle$ are known and also that $R_y = HR_xH^* + R_v > 0$. Then the l.l.m.s.e. and the corresponding m.m.s.e. can be written as

$$\hat{\mathbf{x}} = K_o\mathbf{y}, \quad K_o = R_xH^*[HR_xH^* + R_v]^{-1}, \tag{3.4.2}$$

and

$$P_x \triangleq R_{\tilde{x}} = R_x - R_xH^*[R_v + HR_xH^*]^{-1}HR_x. \tag{3.4.3}$$

These formulas will be encountered in many different contexts in later chapters.

3.4.1 Information Forms When $R_x > 0$ and $R_v > 0$

In many problems it happens that we can also assume that $R_x > 0$ and $R_v > 0$, which of course also ensures that $R_y > 0$. Then we can express the formulas (3.4.2) and (3.4.3) in certain alternative forms that are often useful. We could state the appropriate forms directly and verify them, but it will be pedagogically useful to proceed a little more discursively.

When we have a matrix of the form $A + BCD$, it is often useful to see what happens by applying the matrix inversion lemma (App. A). In our case, this yields

$$R_y^{-1} = (R_v + HR_xH^*)^{-1} = R_v^{-1} - R_v^{-1}H(R_x^{-1} + H^*R_v^{-1}H)^{-1}H^*R_v^{-1}.$$

Then it follows that (see (3.4.2))

$$\begin{aligned} K_o &= R_xH^*R_v^{-1} - R_xH^*R_v^{-1}H(R_x^{-1} + H^*R_v^{-1}H)^{-1}H^*R_v^{-1}, \\ &= R_x\left[(R_x^{-1} + H^*R_v^{-1}H) - H^*R_v^{-1}H\right](R_x^{-1} + H^*R_v^{-1}H)^{-1}H^*R_v^{-1}, \\ &= (R_x^{-1} + H^*R_v^{-1}H)^{-1}H^*R_v^{-1}. \end{aligned} \tag{3.4.4}$$

In the same way, we can establish the identity

$$P_x = R_x - R_xH^*[R_v + HR_xH^*]^{-1}HR_x = (R_x^{-1} + H^*R_v^{-1}H)^{-1}. \tag{3.4.5}$$

We may remark that formulas using inverses of covariance matrices are sometimes called *Information Form* results, because loosely speaking the amount of information obtained by observing a random variable varies inversely as its variance.

An important consequence of (3.4.5) is the nice formula

$$P_x^{-1}\hat{\mathbf{x}} = H^*R_v^{-1}\mathbf{y}, \tag{3.4.6}$$

which reveals the interesting fact that the combination $P_x^{-1}\hat{\mathbf{x}}$ is independent of the covariance matrix, R_x, of $\mathbf{x}$. An important application of this result is described in Sec. 3.4.3.

Remark 3. In Ch. 15 we shall give a geometric interpretation of the identities (3.4.4)–(3.4.5) as describing projections in a so-called *dual space*. ◆

Remark 4. When $R_v > 0$ but not necessarily R_x, it is immediate to verify that the identities (3.4.4)–(3.4.5) can be replaced by

$$\begin{aligned} K_o &= R_xH^*(R_v + HR_xH^*)^{-1} = (I + R_xH^*R_v^{-1}H)^{-1}R_xH^*R_v^{-1}, \\ P(K_o) &= R_x - R_xH^*(R_v + HR_xH^*)^{-1}HR_x = (I + R_xH^*R_v^{-1}H)^{-1}R_x. \end{aligned}$$

◆

3.4.2 The Gauss-Markov Theorem

In certain problems, one assumes very little a priori knowledge of $\mathbf{x}$, in which case one often writes $R_x = \alpha R$, $\alpha \gg 1$. Then in the limit as $\alpha \to \infty$, we can write (assuming that H has full rank),

$$\hat{\mathbf{x}} \to \hat{\mathbf{x}}_\infty \triangleq (H^*R_v^{-1}H)^{-1}H^*R_v^{-1}\mathbf{y} \quad \text{and} \quad P_x \to P_\infty \triangleq (H^*R_v^{-1}H)^{-1}. \tag{3.4.7}$$

One situation in which the above assumptions are invoked is where it is felt that $\mathbf{x}$ is not random at all but is some unknown constant. We may remark that such problems bring us into the domain of statistics associated with the name of R. A. Fisher, so that $\hat{\mathbf{x}}_\infty$ is often referred to as the *Fisher* estimator of $\mathbf{x}$ given $\mathbf{y}$; the estimators for finite values of R_x are often called *Bayes* estimators. For more on the reasons for this terminology and on the pros and cons of the different points of view, the interested reader can consult almost any book on statistics (*e.g.*, Rao (1973)).

Note that if we write $R_v^{-1} = W$, the formulas (3.4.7) are the same as those encountered in the weighted least-squares problem studied in Sec. 2.2.2. While this similarity will be pursued in more detail in Sec. 3.5, we note here that it is a special case of a famous result in the statistical literature.

Theorem 3.4.1 (Gauss-Markov Theorem) *Consider a model $\mathbf{y} = Hx + \mathbf{v}$, where $\mathbf{v}$ is a zero-mean random variable with unit variance, $\langle \mathbf{v}, \mathbf{v}\rangle = I$, x is a deterministic vector, and H has full column rank. Then the estimator that is defined by $\hat{\mathbf{x}}_\infty = (H^*H)^{-1}H^*\mathbf{y}$ is the optimum unbiased linear least-mean-squares estimator of x.* ■

Proof: Assume $\hat{\mathbf{z}} = K\mathbf{y}$ is any other linear estimator for x. In order for $\hat{\mathbf{z}}$ to be unbiased we must require $KH = I$ since $E\hat{\mathbf{z}} = EK(Hx + \mathbf{v})^* = KHx$. In this case, the covariance matrix of $\hat{\mathbf{z}}$ becomes $\text{cov}(\hat{\mathbf{z}}) = \langle \hat{\mathbf{z}} - x, \hat{\mathbf{z}} - x\rangle = KK^*$. On the other hand, $\hat{\mathbf{x}}_\infty$ is also an unbiased estimator for x and its covariance matrix is given by

$$\text{cov}(\hat{\mathbf{x}}_\infty) = \langle \hat{\mathbf{x}}_\infty - x, \hat{\mathbf{x}}_\infty - x\rangle = (H^*H)^{-1},$$

which, using $KH = I$, can be rewritten as

$$\langle \hat{\mathbf{x}}_\infty - x, \hat{\mathbf{x}}_\infty - x\rangle = KH(H^*H)^{-1}H^*K^*.$$

Therefore,

$$\text{cov}(\hat{\mathbf{z}}) - \text{cov}(\hat{\mathbf{x}}_\infty) = K[I - H(H^*H)^{-1}H^*]K^*.$$

But $H(H^*H)^{-1}H^*$ is a projection matrix and we necessarily have $I - H(H^*H)^{-1}H^* \geq 0$. Consequently, $\text{cov}(\hat{\mathbf{x}}_\infty) \leq \text{cov}(\hat{\mathbf{z}})$, for any other unbiased linear estimator of x. ◆

When the variance of $\mathbf{v}$ is not unity, say $\langle \mathbf{v}, \mathbf{v}\rangle = R_v > 0$, we can scale the equation $\mathbf{y} = Hx + \mathbf{v}$ by the inverse of any matrix L satisfying $R_v = LL^*$. That is, we may write

$$L^{-1}\mathbf{y} = L^{-1}Hx + L^{-1}\mathbf{v} = L^{-1}Hx + \bar{\mathbf{v}}, \quad \langle \bar{\mathbf{v}}, \bar{\mathbf{v}}\rangle = I.$$

Then, according to the Gauss-Markov theorem, the optimal unbiased linear l.m.s. estimator of x given $L^{-1}\mathbf{y}$ is given by

$$\hat{\mathbf{x}}_\infty = (H^*L^{-*}L^{-1}H)^{-1}H^*L^{-*}L^{-1}\mathbf{y} = (H^*R_v^{-1}H)^{-1}H^*R_v^{-1}\mathbf{y}. \tag{3.4.8}$$

This same expression for $\hat{\mathbf{x}}_\infty$ can be obtained as a special case of a weighted least-squares problem, say (*cf.* Sec. 2.2.2)

$$\min_x \quad \|y - Hx\|_W^2. \tag{3.4.9}$$

The solution of (3.4.9) is given by $\hat{x} = (H^*WH)^{-1}H^*Wy$. We therefore see that the choice $W = R_v^{-1}$ leads to the expression for the minimum variance unbiased estimator (MVUE), a result that we presented earlier in Sec. 2.2.3.

Constrained Linear Estimation. We may finally remark that the estimator $\hat{\mathbf{x}}_\infty$ in (3.4.8) can also be regarded as the solution to a constrained estimation problem, as we now clarify. Such constrained formulations arise in various applications, including beamforming for interference attentuation and decision feedback channel equalization (see, *e.g.*, Prob. 3.26).

Let $\hat{\mathbf{x}}_\infty = K\mathbf{y}$ denote a linear estimator for the deterministic vector x given $\mathbf{y} = Hx + \mathbf{v}$. Assume now that we seek a K that satisfies two requirements:

1. The first requirement is that $\hat{\mathbf{x}}_\infty$ be an unbiased estimator for x, *i.e.*, $E\hat{\mathbf{x}}_\infty = x$ or, equivalently, $x = KE\mathbf{y} = KE(Hx + \mathbf{v}) = KHx$. So by requiring $KH = I$, the identity matrix, we can guarantee an unbiased estimator.
2. The second requirement is to minimize $E(\hat{\mathbf{x}}_\infty - x)(\hat{\mathbf{x}}_\infty - x)^*$, the covariance matrix of the unbiased random variable $\hat{\mathbf{x}}$. But since $KH = I$, we see that $\hat{\mathbf{x}}_\infty = K\mathbf{y} = K(Hx + \mathbf{v}) = x + K\mathbf{v}$. This shows that the covariance matrix of $\hat{\mathbf{x}}_\infty$ is equal to

$$E(\hat{\mathbf{x}} - x)(\hat{\mathbf{x}} - x)^* = K(E\mathbf{v}\mathbf{v}^*)K^* = KR_vK^*.$$

[Note that R_v is also equal to R_y, the covariance matrix of $\mathbf{y}$, since $E\mathbf{y} = Hx$ and $E(\mathbf{y} - Hx)(\mathbf{y} - Hx)^* = E\mathbf{v}\mathbf{v}^*$.]

We are therefore reduced to solving the *constrained minimization* problem

$$\min_K \; KR_vK^* \quad \text{subject to} \quad KH = I, \tag{3.4.10}$$

where $R_v > 0$ and H has full column rank. The solution (*cf.* the proof of Thm. 3.4.1) is

$$K_o = (H^*R_v^{-1}H)^{-1}H^*R_v^{-1}. \tag{3.4.11}$$

3.4.3 Combining Estimators

In this section we shall present an important result on how to combine estimators of two separate observations of a random variable.

Lemma 3.4.1 (Combining Estimators) *Let $\mathbf{y}_a$ and $\mathbf{y}_b$ be two separate observations of a zero-mean random variable $\mathbf{x}$, such that $\mathbf{y}_a = H_a\mathbf{x} + \mathbf{v}_a$ and $\mathbf{y}_b = H_b\mathbf{x} + \mathbf{v}_b$, where $\{\mathbf{v}_a, \mathbf{v}_b, \mathbf{x}\}$ are mutually uncorrelated zero-mean random variables with covariance matrices R_a, R_b, and R_x, respectively. Denote by $\hat{\mathbf{x}}_a$ and $\hat{\mathbf{x}}_b$ the l.l.m.s. estimators of $\mathbf{x}$ given $\mathbf{y}_a$ and $\mathbf{y}_b$, respectively, and likewise define the error covariance matrices, $P_a = \langle \mathbf{x} - \hat{\mathbf{x}}_a, \mathbf{x} - \hat{\mathbf{x}}_a \rangle$ and $P_b = \langle \mathbf{x} - \hat{\mathbf{x}}_b, \mathbf{x} - \hat{\mathbf{x}}_b \rangle$. Then $\hat{\mathbf{x}}$, the l.l.m.s. estimator of $\mathbf{x}$ given both $\mathbf{y}_a$ and $\mathbf{y}_b$, can be found as*

$$P^{-1}\hat{\mathbf{x}} = P_a^{-1}\hat{\mathbf{x}}_a + P_b^{-1}\hat{\mathbf{x}}_b, \tag{3.4.12}$$

where P, the corresponding error covariance matrix, is given by

$$P^{-1} = P_a^{-1} + P_b^{-1} - R_x^{-1}. \tag{3.4.13}$$

■

Remark 5. Before giving a proof, we note that (3.4.12) fits with our intuition. If P_a is small, the estimator is good and vice versa. Therefore it makes sense to combine the estimators inversely as their variances. The only unusual aspect is the appearance of the term R_x^{-1} in (3.4.13); it vanishes when we have no a priori information on $\mathbf{x}$, in which case $R_x \to \infty$. ◆

Proof: Recall from (3.4.6) that we can write

$$P_a^{-1}\hat{\mathbf{x}}_a = H_a^*R_a^{-1}\mathbf{y}_a \quad \text{and} \quad P_b^{-1}\hat{\mathbf{x}}_b = H_b^*R_b^{-1}\mathbf{y}_b, \tag{3.4.14}$$

which shows that $P_a^{-1}\hat{\mathbf{x}}_a$ and $P_b^{-1}\hat{\mathbf{x}}_b$ are independent of R_x. Now we can write the observations $\mathbf{y}_a$ and $\mathbf{y}_b$ as one aggregate observation

$$\underbrace{\begin{bmatrix}\mathbf{y}_a\\ \mathbf{y}_b\end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix}H_a\\ H_b\end{bmatrix}}_{H}\mathbf{x} + \underbrace{\begin{bmatrix}\mathbf{v}_a\\ \mathbf{v}_b\end{bmatrix}}_{\mathbf{v}}.$$

Therefore, writing the expression analogous to (3.4.14) for the above linear model, we have

$$\begin{aligned} P^{-1}\hat{\mathbf{x}} = H^*R^{-1}\mathbf{y} &= \begin{bmatrix}H_a^* & H_b^*\end{bmatrix}\begin{bmatrix}R_a^{-1} & 0\\ 0 & R_b^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{y}_a\\ \mathbf{y}_b\end{bmatrix},\\ &= H_a^*R_a^{-1}\mathbf{y}_a + H_b^*R_b^{-1}\mathbf{y}_b. \end{aligned}$$

Combining the last equation with (3.4.14), we obtain (3.4.12). Now using the aggregate linear model we can write the following expression for P:

$$\begin{aligned} P &= (R_x^{-1} + H^*R^{-1}H)^{-1},\\ &= \left(R_x^{-1} + \begin{bmatrix}H_a^* & H_b^*\end{bmatrix}\begin{bmatrix}R_a^{-1} & 0\\ 0 & R_b^{-1}\end{bmatrix}\begin{bmatrix}H_a\\ H_b\end{bmatrix}\right)^{-1},\\ &= (R_x^{-1} + H_a^*R_a^{-1}H_a + H_b^*R_b^{-1}H_b)^{-1}. \end{aligned}$$

Now using the expressions (3.4.5) for P_a and P_b, we obtain $P = (P_a^{-1} + P_b^{-1} - R_x^{-1})^{-1}$, as desired. ◆

3.5 EQUIVALENCE TO DETERMINISTIC LEAST-SQUARES

An important comparison can now be made between some results of Ch. 2 and those in Sec. 3.4. Recall from Sec. 2.4 (or Table 2.1) that the solution $\hat{x}$ to the optimization problem:

$$\min_x \left[(x - x_0)^*\Pi_0^{-1}(x - x_0) + \|y - Hx - v_0\|_W^2\right],$$

is given by

$$\hat{x} = x_0 + \left[\Pi_0^{-1} + H^*WH\right]^{-1}H^*W\left[y - Hx_0 - v_0\right], \tag{3.5.1}$$

where Π_0 and W are given positive-definite matrices and where we have included an additional term v_0 — this simply corresponds to replacing y in (2.4.1) by $(y - v_0)$. Also, the notation $\|a\|_W^2$ stands for a^*Wa.

Now, we just showed in the previous section that given a linear model of the form $\mathbf{y} = H\mathbf{x} + \mathbf{v}$, where $\mathbf{x}$ is a random variable with known mean m_x and covariance matrix R_x, and $\mathbf{v}$ is a random noise vector with mean m_v and covariance matrix $R_v = \langle v - m_v, v - m_v \rangle$, and which is uncorrelated with $\mathbf{x}$ in the sense $\langle \mathbf{x} - m_x, \mathbf{v} - m_v \rangle = 0$, then the linear least-mean-squares estimator of $\mathbf{x}$ given $\mathbf{y}$ is

$$\hat{\mathbf{x}} = m_x + \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1} [\mathbf{y} - Hm_x - m_v]. \qquad (3.5.2)$$

Comparing the expressions (3.5.1) and (3.5.2), we see that they can be related by making the associations

$$R_x \leftrightarrow \Pi_0, \quad m_x \leftrightarrow x_0, \quad R_v \leftrightarrow W^{-1}, \quad m_v \leftrightarrow v_0.$$

This equivalence relationship allows us to move back and forth between the deterministic and the stochastic frameworks, a very important result that can be exploited to great effect. In particular, it explains the similarity of the recursive least-squares solution of Sec. 2.6 (recall the remark after Lemma 2.6.1) and the Kalman filter of Thm. 1.2.1. We shall give other examples (from adaptive filtering and quadratic control) in later chapters. For more on such applications, see the article on adaptive filtering by Sayed and Kailath (1994b) and the monograph on indefinite quadratic forms in estimation and control by Hassibi, Sayed, and Kailath (1999).

Table 3.1 summarizes the relations between the variables in both frameworks. Moreover, it is worthwhile to restate the important conclusion that results from this discussion. If we pose a stochastic linear least-mean-squares problem, *viz.*, estimating a random variable $\mathbf{x}$ from a collection of random variables in $\mathbf{y}$,

$$\min_K \langle \mathbf{x} - m_x - K(\mathbf{y} - m_y), \mathbf{x} - m_x - K(\mathbf{y} - m_y) \rangle \quad \Longrightarrow \quad \hat{\mathbf{x}},$$

then the equivalence result tells us that we can formulate a corresponding deterministic problem as follows:

(i) exhibit the linear relation that exists between $\mathbf{y}$ and $\mathbf{x}$, say $\mathbf{y} = H\mathbf{x} + \mathbf{v}$, with the appropriate covariance matrices and mean values for $\mathbf{x}$ and $\mathbf{v}$, and then

(ii) write down the deterministic criterion,

$$\min_x \left[(x - m_x)^* R_x^{-1} (x - m_x) + (y - Hx - m_v)^* R_v^{-1} (y - Hx - m_v)\right] \quad \Longrightarrow \quad \hat{x}.$$

It is important to observe that *it is the formulas for the optimum gain matrix K_o that can be made identical by the mapping in Table 3.1; the expressions for the minimum costs $P(K_o)$ and $J(\hat{x})$ cannot be directly related to each other.* We remark that more insight into these facts will be presented in Ch. 15, where we shall also explain the difference between "equivalent" problems and "dual" problems.

Table 3.1 An equivalence between the stochastic and deterministic frameworks. The expressions for $\hat{\mathbf{x}}$ and $\hat{x}$ are related to each other by the transformations $R_x \leftrightarrow \Pi_0$, $m_x \leftrightarrow x_0$, $R_v \leftrightarrow W^{-1}$, $m_v \leftrightarrow v_0$.

Stochastic Framework	Deterministic Framework
$\mathbf{y} = H\mathbf{x} + \mathbf{v}$	$y = Hx + v$
$m_x = E\mathbf{x}$	initial guess of x: x_0
$E(\mathbf{x} - m_x)(\mathbf{x} - m_x)^* = R_x$	weighting matrix: Π_0
$m_v = E\mathbf{v}$	initial value: v_0
$E(\mathbf{v} - m_v)(\mathbf{v} - m_v)^* = R_v$	inverse of weighting matrix: W^{-1}
$m_y = E\mathbf{y} = Hm_x + m_v$	initial guess of y: $y_0 = Hx_0 + v_0$
$\hat{\mathbf{x}}$	$\hat{x}$
$\min_K P(K)$, where $P(K) = \|\mathbf{x} - m_x - K(\mathbf{y} - m_y)\|^2$	$\min_x J(x)$, where $J(x) = (x - x_0)^* \Pi_0^{-1} (x - x_0) + \|y - Hx - v_0\|_W^2$
$\hat{\mathbf{x}} = m_x + K_o[\mathbf{y} - Hm_x - m_v]$ where $K_o = R_x H^* (R_v + HR_x H^*)^{-1}$ $K_o = \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1} H^* R_v^{-1}$	$\hat{x} = x_0 + K_o[y - Hx_0 - v_0]$ where $K_o = \Pi_0 H^* (W^{-1} + H\Pi_0 H^*)^{-1}$ $K_o = \left[\Pi_0^{-1} + H^* W H\right]^{-1} H^* W$
$P(K_o) = R_x - R_x H^* (R_v + HR_x H^*)^{-1} HR_x = \left[R_x^{-1} + H^* R_v^{-1} H\right]^{-1}$	$J(\hat{x}) = \|y - Hx_0 - v_0\|^2_{\left[W^{-1} + H\Pi_0 H^*\right]^{-1}}$

3.6 COMPLEMENTS

Sec. 3.1. The Problem of Stochastic Estimation. In the notes to Sec. 1.2.1, we presented Gauss' comments on the physical reasonableness, mathematical tractability, and elegance of the mean-square-error criterion. However, there are several other reasons for interest in minimum mean-square error (stochastic least-squares) estimation problems. Here are some of them:

(i) The optimal estimator of a random variable given another collection of random variables can be identified as a conditional expectation (see, *e.g.*, App. 3.A or almost any textbook on probability theory). Though this is not always easy to calculate explicitly, conditional expectation sequences, as encountered in recursive

(increasing data) estimation problems, form a very important kind of stochastic process called a *martingale* process, which is very important in further studies of estimation theory. Unfortunately (or fortunately, depending upon the reader's point of view), we shall confine ourselves in this book to linear estimators, in which case the martingale property is not essential.

(ii) When all the random variables involved are jointly Gaussian, then the optimal mean-square estimator is linear (see App. 3.A). In this case, the linear estimator is also optimum for essentially any error criterion (see, *e.g.*, Zakai (1964) and the references therein).

(iii) The solutions of stochastic least-squares problems turn out to have many mutually illuminating connections with the ubiquitous deterministic problems of the solution of linear (matrix and integral) equations and the closely related problems of triangular (LDL*) and orthogonal (QR) matrix factorizations.

Sec. 3.2. Linear Least-Mean-Squares Estimators. The material in this section is quite standard, except for the issues raised by the use of complex-valued random variables. Some of the issues have been known for a long time, see, *e.g.*, the classical book of Doob (1953, pp. 74–75). However, in recent years, it was Picinbono (1993,1994) who re-emphasized that more care was needed in the usual discussions of linear least-mean-squares estimation for complex-valued random variables; see also Neeser and Massey (1993), Edelblute (1996), and van den Bos (1998).

Sec. 3.2.1. The Fundamental Equations. An unfortunate historical anomaly is the name "Wiener solution" sometimes given to the very simply obtained formulas (3.2.4) and (3.2.6), especially in much of the literature on adaptive filtering. This grew out of the early excitement associated with the introduction of the deterministic and stochastic least-squares criteria into the field. As we shall see in Sec. 4.1.3, and in detail in Chs. 6 and 7, Wiener solved the much deeper "causal" estimation problem.

Sec. 3.3. A Geometric Formulation. The idea of regarding random variables as elements of a linear vector space apparently goes back to Fréchet (1937). This formulation, which made it natural to interpret least-squares estimation as projection onto a linear subspace, was used by Wold (1938), and later more fully exploited by Kolmogorov (1939,1941a,1941b) to obtain fundamental results on the representation and prediction of general stationary stochastic processes (see Ch. 6).

It took many years for the geometric formulation to penetrate into the engineering literature, even after its use in the pioneering paper of Kalman (1960a), which explicitly cites and uses the basic ideas in Doob (1953); Yaglom (1962) also gave an engineering-oriented presentation emphasizing the value of the geometric formulation.

As the reader may expect, it is possible to include the different geometric formulations of Ch. 2 and Ch. 3 in a unified framework, which is the theory of linear spaces. As noted in the text, we have to be somewhat more general than conventional treatments because of our use of matrix-valued inner products. However, in order not to slow down the flow of ideas with a mathematical interlude, we defer that discussion to App. 4.A.

Sec. 3.4.3. Combining Estimators. Not all the results presented in this section are widely known; the special case $\Pi_0 = \infty$ is the one generally discussed in the literature. Though simple, the results are often very useful (see, *e.g.*, Sec. 10.4.3).

Sec. 3.5. Equivalence to Deterministic Least-Squares. The equivalence described here is very useful in expanding the scope of the stochastic theory to fields such as adaptive filtering and $\mathcal{H}_\infty$ estimation. We may remark that what we have called "equivalence" is sometimes called "duality" in the literature. However, it is very useful to distinguish between these two terms, for reasons that will be discussed in Ch. 15. We may also mention that the equivalence described in Table 3.1 can be regarded as a slight generalization of the Gauss-Markov theorem (Thm. 3.4.1) to accomodate Π_0 (or x_0).

PROBLEMS

3.1 **(Nonzero means)** Suppose that $\mathbf{x}$ and $\mathbf{y}$ are random variables such that $E\mathbf{x} = m_x$, $E\mathbf{y} = m_y$, $E(\mathbf{x}-m_x)(\mathbf{x}-m_x)^* = R_x$, and $E(\mathbf{x}-m_x)(\mathbf{y}-m_y)^* = R_{xy}$.

(a) Determine $K \in \mathbb{C}^{n\times p}$ and $l \in \mathbb{C}^{n\times 1}$ such that $\hat{\mathbf{x}} = K\mathbf{y}+l$ minimizes $E(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^*$. Also find an expression for the minimum. *Hint.* Form the quadratic form $a^*E(\mathbf{x}-K\mathbf{y}-l)(\mathbf{x}-K\mathbf{y}-l)^*a$, for any $a \in \mathbb{C}^{n\times 1}$, and differentiate w.r.t. both a^*K and a^*l.]

(b) Verify that the above result can also be obtained by using the conditions $\mathbf{x}-\hat{\mathbf{x}} \perp \mathbf{y}$ and $\mathbf{x}-\hat{\mathbf{x}} \perp 1$ (the constant random variable 1).

3.2 **(More on nonzero means)** Consider the same setting as in Prob. 3.1, but now write $\hat{\mathbf{x}} = K\mathbf{y}$. That is, we set $l = 0$.

(a) Find the value of K that minimizes $E(\mathbf{x}-K\mathbf{y})(\mathbf{x}-K\mathbf{y})^*$.

(b) Assume $m_y = 0$. Show that the minimum value achieved by the solution in part (a) is greater than the minimum value achieved by the solution in Prob. 3.1 (where we assumed $l \neq 0$).

(c) Repeat part (b) but now with $m_x = 0$.

(d) Repeat part (b) with $m_x = 0$ and $m_y = 0$.

3.3 **(Minimum-variance-unbiased estimators)** Consider again the setting of Prob. 3.1.

(a) When $m_x = 0$ and $m_y = 0$, show that the mean-square-error matrix $E\tilde{\mathbf{x}}\tilde{\mathbf{x}}^*$, $\tilde{\mathbf{x}} = \mathbf{x}-\hat{\mathbf{x}}$, is the same as the covariance matrix, $E\tilde{\mathbf{x}}\tilde{\mathbf{x}}^* - (E\tilde{\mathbf{x}})(E\tilde{\mathbf{x}})^*$.

(b) When $m_x \neq 0$ and $m_y \neq 0$, try to find $\{K, l\}$ so that $\hat{\mathbf{x}} = K\mathbf{y}+l$ minimizes the covariance matrix of $\mathbf{x}-K\mathbf{y}-l$ rather than the mean-square-error matrix. All attempts will fail. Why?

(c) Repeat (b) after making the additional assumption that $\hat{\mathbf{x}}$ is an unbiased linear estimator of $\mathbf{x}$, *i.e.*, that $m_x = E\mathbf{x} = E\hat{\mathbf{x}} = Km_y + l$. Show that the resulting so-called Linear Minimum Variance Unbiased Estimator (MVUE) coincides with the linear minimum mean-square-error estimator.

3.4 **(Other minimization criteria)** Show that the l.l.m.s. estimator $\hat{\mathbf{x}} = R_{xy}R_y^{-1}\mathbf{y}$ also minimizes the following criteria:

(a) $\mathrm{Tr}\{E(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^*\}$, where $\mathrm{Tr}(A) \triangleq$ the trace of the matrix A = the sum of the diagonal entries of A.

(b) $E(\mathbf{x}-\hat{\mathbf{x}})^*W(\mathbf{x}-\hat{\mathbf{x}})$, where W is any Hermitian nonnegative-definite matrix.

(c) $\det\{E(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^*\}$.

Hint. For (c) use the fact that $A \geq B \geq 0$ implies $\det A \geq \det B$.

3.5 **(Optimal linear estimators for complex-valued random variables)** Assume $\mathbf{x}$ and $\mathbf{y}$ are two scalar complex-valued random variables with $E\mathbf{y}_R^2 = \gamma_{11}$, $E\mathbf{y}_I^2 = \gamma_{22}$, $E\mathbf{y}_R\mathbf{y}_I = \gamma_{12}$, $E\mathbf{x}_R\mathbf{y}_R = \alpha_{11}$, $E\mathbf{x}_R\mathbf{y}_I = \alpha_{12}$, $E\mathbf{x}_I\mathbf{y}_R = \alpha_{21}$, and $E\mathbf{x}_I\mathbf{y}_I = \alpha_{22}$. Assume further that $\gamma_{11}\gamma_{22} \neq \gamma_{12}^2$ and $\gamma_{11} + \gamma_{22} \neq 0$. Verify that the matrices M_o and N_o that solve (3.2.11) and (3.2.18), respectively, are distinct.

3.6 **(A separation principle)** All variables are zero-mean. Consider the linear model $\mathbf{y} = H\mathbf{x} + \mathbf{v}$, where $\mathbf{v}$ and $\mathbf{x}$ are uncorrelated with variances R_v and R_x, respectively. Let $\hat{\mathbf{z}}_{|x}$ and $\hat{\mathbf{z}}_{|y}$ denote the l.l.m.s. estimators of a random variable $\mathbf{z}$ given $\mathbf{x}$ and given $\mathbf{y}$, respectively (both $\mathbf{z}$ and $\mathbf{v}$ are assumed uncorrelated). Let further $\widehat{\hat{\mathbf{z}}_{|x}}$ denote the l.l.m.s.e. of $\hat{\mathbf{z}}_{|x}$ given $\mathbf{y}$. Verify that $\hat{\mathbf{z}}_{|y}$ and $\widehat{\hat{\mathbf{z}}_{|x}}$ coincide.

3.7 **(Linear estimator of $\mathbf{x}^2$)** Consider $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where $\mathbf{x}$ and $\mathbf{v}$ are independent zero-mean Gaussian real- and scalar-valued random variables with variances σ_x^2 and σ_v^2, respectively. Find the l.l.m.s. estimator of the random variable $\mathbf{x}^2$ using $\{\mathbf{y}, \mathbf{y}^2\}$. [*Hint.* Recall that for a zero-mean real-valued Gaussian random variable $\mathbf{n}$, with variance σ^2, we have $E\mathbf{n}^3 = 0$ and $E\mathbf{n}^4 = 3\sigma^4$.]

3.8 **(Stochastic sampling theorem)** A continuous-time zero-mean scalar stationary process $\{\mathbf{y}(t), -\infty < t < \infty\}$ is called *band limited* if its power spectral density function, defined as the Fourier transform of its autocorrelation function,

$$S_y(j\omega) \triangleq \mathcal{F}\{R_y(\tau)\} = \mathcal{F}\{\langle \mathbf{y}(t), \mathbf{y}(t-\tau)\rangle\},$$

is band limited, *i.e.*, for some W, $S_y(j\omega) = 0$ for all $|w| > W$. The well-known deterministic sampling theorem shows that $R_y(t)$ can be recovered from its samples through the formula

$$R_y(t-a) = \sum_{n=-\infty}^{\infty} \operatorname{sinc}[W(t-nT)]\, R_y(nT-a), \quad \operatorname{sinc} x = \frac{\sin \pi x}{\pi x},$$

for any a, provided the sampling rate is such that $T < \pi/W$.

(a) Show that the l.l.m.s. estimator of $\mathbf{y}(t)$, t any real number, given the samples $\{\mathbf{y}(nT)\}_{n=-\infty}^{\infty}$, can be written as

$$\hat{\mathbf{y}}(t) = \sum_{n=-\infty}^{\infty} \operatorname{sinc}[W(t-nT)]\, \mathbf{y}(nT).$$

(b) Show that the m.m.s.e is zero. What is the interpretation of this result?

3.9 **(A stationary state-space model)** Consider scalar zero-mean random variables $\{\mathbf{x}(i), \mathbf{u}(i), \mathbf{v}(i), \mathbf{y}(i)\}$ that are related via the state-space equations

$$\mathbf{x}(i+1) = 0.5\mathbf{x}(i) + \mathbf{u}(i), \qquad \mathbf{y}(i) = \mathbf{x}(i) + \mathbf{v}(i), \quad i > -\infty,$$

where $\mathbf{u}(i)$ and $\mathbf{v}(i)$ are uncorrelated unit-variance white-noise processes. It is assumed that the system is operating in stationary mode and that $\mathbf{u}(i)$ and $\mathbf{v}(i)$ are further uncorrelated with $\mathbf{x}(i)$. Define the column vectors

$$\mathbf{a} \triangleq \operatorname{col}\{\mathbf{x}(1), \mathbf{x}(3), \mathbf{y}(2)\}, \qquad \mathbf{b} \triangleq \operatorname{col}\{\mathbf{x}(2), \mathbf{y}(1), \mathbf{y}(3)\}.$$

Determine the l.l.m.s. estimator of $\mathbf{a}$ given $\mathbf{b}$ and compute the corresponding m.m.s.e.

3.10 **(Interpolation)** Suppose that $\{\mathbf{y}(t),\ t \geq 0,\ \mathbf{y}(0) = 0\}$ is a continuous-time stochastic process with $E\mathbf{y}(t) = 0$, $E\mathbf{y}(t)\mathbf{y}^*(s) = \min(t, s)$. Find the l.l.m.s. estimator of $\mathbf{y}(\sigma)$ given $\mathbf{y}(t_0)$ and $\mathbf{y}(t_1)$, $t_0 < \sigma < t_1$.

3.11 **(Estimator for an integral)** Consider a zero-mean wide-sense stationary random scalar process $\mathbf{y}(t)$ with $E\mathbf{y}(t)\mathbf{y}^*(s) = 0.5^{|t-s|}$ for all $0 \leq t, s \leq 1$. Determine the l.l.m.s. estimator of $\mathbf{x} = \int_0^1 \mathbf{y}(t)dt$ given the observations $\mathbf{y}(0)$ and $\mathbf{y}(1)$. What is the variance of the resulting estimation error?

3.12 **(Stationary wide-sense Markov processes)** Suppose $\mathbf{x}(\cdot)$ is a stationary process with the so-called wide-sense Markov property that, for some k,

$$\begin{aligned}\hat{\mathbf{x}}(t+\lambda|t) &\triangleq \text{l.l.m.s.e. of } \mathbf{x}(t+\lambda) \text{ given } \{\mathbf{x}(s), -\infty < s \leq t\},\\ &= k\mathbf{x}(t), \quad \text{for all } t,\end{aligned}$$

and for some constant k. Show that $R_x(\tau) = \langle \mathbf{x}(t+\tau), \mathbf{x}(t)\rangle$ must have the form $R_x(\tau) = \beta e^{-\alpha|\tau|}$ for some constants α and β.

3.13 **(Linear l.m.s. estimator of $\mathbf{y}^2$)** Consider a random scalar variable $\mathbf{y}$ with moments $E\mathbf{y}^i = m_i$ for $i \geq 1$. Find the l.l.m.s. estimator of $\mathbf{y}^2$ given $\mathbf{y}$. What is the best nonlinear estimator (*cf.* App. 3.A)?

3.14 **(Uncertain models)** Consider the noisy measurement $\mathbf{y} = (h + \boldsymbol{\delta h})\mathbf{x} + \mathbf{v}$, where the scalar real-valued quantities $\{\mathbf{x}, \mathbf{v}, \boldsymbol{\delta h}\}$ are zero-mean independent random variables with variances $\{\sigma_x^2, \sigma_v^2, \sigma_h^2\}$. The h is a known real scalar coefficient and represents the uncertainty in h. Define the signal-to-noise ratio $\mathsf{SNR} = \sigma_x^2/\sigma_v^2$. Determine the l.l.m.s. estimator of $\mathbf{x}$ given $\mathbf{y}$ and the resulting m.m.s.e. Express your answers in terms of $\{h, \sigma_h^2, \mathsf{SNR}, \mathbf{y}\}$. Argue that the expression for $\hat{\mathbf{x}}$ has the same form as the solution of a regularized least-squares problem of the form

$$\min_x \left[\frac{1}{\pi_0}x^2 + |y - hx|^2\right].$$

What is the value of π_0?

3.15 **(Multiplicative noise)** Consider the noisy measurement $\mathbf{y} = (1 + \mathbf{v})\mathbf{x}$, where $\mathbf{x}$ and $\mathbf{v}$ are zero-mean independent random variables. Only the variance of the noise variable $\mathbf{v}$ is known, say σ_v^2. Determine the l.l.m.s. estimator of $\mathbf{x}$ given $\mathbf{y}$. Show that the m.m.s.e. is smaller than the variance of $\mathbf{x}$.

3.16 **(Defective measurement sensors)** Consider a zero-mean random variable $\mathbf{x}$ with variance Π_0 and two possible measurements for $\mathbf{x}$, say

$$\mathbf{y}_1 = H_1\mathbf{x} + \mathbf{v}_1, \qquad \mathbf{y}_2 = H_2\mathbf{x} + \mathbf{v}_2,$$

where $\{\mathbf{v}_1, \mathbf{v}_2\}$ are zero-mean uncorrelated sensor noise with variances R_1 and R_2, respectively. They are also uncorrelated with $\mathbf{x}$. One of the measurement sensors is defective and it is either sensor 1 with probability $1 - p$ or sensor 2 with probability p. The measurement that is used to estimate $\mathbf{x}$ is therefore either $\mathbf{y}_1$ with probability p or $\mathbf{y}_2$ with probability $(1 - p)$. Denote this measurement by $\mathbf{z}$. Find the l.l.m.s. estimator of $\mathbf{x}$ given $\mathbf{z}$ and the corresponding m.m.s.e. How would your answers change if the variables $\{\mathbf{v}_1, \mathbf{v}_2\}$ were correlated? Any comments on the special case $H_1 = H_2 \triangleq H$?

3.17 **(Parameter estimation problem)** The pressure in a valve is decreasing exponentially fast. Three noisy measurements of the pressure are available at time instants $\{t_0, t_1, t_2\}$, say

$$\mathbf{y}(i) = p_0 e^{-t_i} + \mathbf{v}(i), \quad i = 0, 1, 2,$$

where p_0 is the unknown initial pressure and the $\{\mathbf{v}(i)\}$ are uncorrelated zero-mean random variables with variances σ_i^2. Let $\hat{\mathbf{p}}_0$ denote the optimum unbiased l.l.m.s. estimator of p_0 given $\{\mathbf{y}(0), \mathbf{y}(1), \mathbf{y}(2)\}$. Determine $\hat{\mathbf{p}}_0$. [*Hint.* Recall the Gauss-Markov theorem (Thm. 3.4.1).]

3.18 **(Equalization of a communications channel)** All random variables in this problem are scalar and real-valued. A random sequence $\mathbf{x}(n)$ is generated by applying a white-noise zero-mean process $\mathbf{u}(n)$, of variance σ_u^2, to a first-order stable filter

$$H(z) = \frac{1}{1 + \alpha z^{-1}}, \quad |\alpha| < 1, \quad \alpha \in \mathbb{R}.$$

The sequence $\mathbf{x}(n)$ is then transmitted through a communications channel with transfer function

$$C(z) = \frac{1}{1 + \beta z^{-1}}, \quad |\beta| < 1, \quad \alpha \in \mathbb{R}.$$

The output of the channel, denoted by $\mathbf{r}(n)$, is corrupted by additive zero-mean noise $\mathbf{v}(n)$ of variance σ_v^2 and the signal at the receiver is therefore $\mathbf{y}(n) = \mathbf{r}(n) + \mathbf{v}(n)$. The white-noise processes $\{\mathbf{v}(n), \mathbf{u}(n)\}$ are uncorrelated. Design a two-tap transversal filter receiver in order to estimate the transmitted symbol in the linear least-mean squares sense. In other words, determine the optimal values of the weights w_0 and w_1 in Fig. 3.2. Determine also the resulting m.m.s.e., as well as the numerical values when $\alpha = 0.5$, $\beta = -0.6$, $\sigma_u^2 = 0.30$, and $\sigma_v^2 = 0.15$.

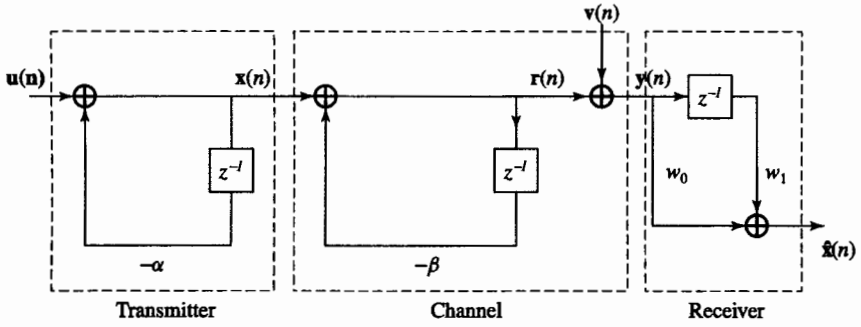


Figure 3.2 Design of an equalizer for a communications channel.

3.19 **(Block processing)** Consider two zero-mean jointly wide-sense stationary scalar-valued random processes $\{\mathbf{x}(n), \mathbf{y}(n)\}$. Let $\hat{\mathbf{x}}(n)$ denote the l.l.m.s. estimator of $\mathbf{x}(n)$ given the past 6 observations $\mathbf{y} = \text{col}\{\mathbf{y}(n), \mathbf{y}(n-1), \ldots, \mathbf{y}(n-5)\}$. That is, $\hat{\mathbf{x}}(n) = k_o \mathbf{y}$, where $k_o = R_{xy} R_y^{-1}$ is independent of n in view of the wide-sense stationarity assumption. We can therefore regard $\hat{\mathbf{x}}(n)$ as the output of an FIR filter with transfer function

$$k_o(z) = k_0 + k_1 z^{-1} + k_2 z^{-2} + k_3 z^{-3} + k_4 z^{-4} + k_5 z^{-5},$$

where the $\{k_i\}$ denote the individual entries of the row vector k_o. Alternatively, the estimators $\{\hat{\mathbf{x}}(n)\}$ can be evaluated on a block-by-block basis as follows. Introduce the column vectors

$$\hat{\mathbf{x}}_n \triangleq \begin{bmatrix} \hat{\mathbf{x}}(3n) \\ \hat{\mathbf{x}}(3n-1) \\ \hat{\mathbf{x}}(3n-2) \end{bmatrix}, \quad \mathbf{y}_n \triangleq \begin{bmatrix} \mathbf{y}(3n) \\ \mathbf{y}(3n-1) \\ \mathbf{y}(3n-2) \end{bmatrix},$$

and verify that the transfer *matrix* function that maps $\mathbf{y}_n$ to $\hat{\mathbf{x}}_n$ is given by

$$K_o(z) = \begin{bmatrix} E_0(z) & E_1(z) & E_2(z) \\ z^{-1}E_2(z) & E_0(z) & E_1(z) \\ z^{-1}E_1(z) & z^{-1}E_2(z) & E_0(z) \end{bmatrix},$$

where the $\{E_i(z)\}$ denote the (3rd order) polyphase components of (the so-called wideband filter) $k_o(z)$, *viz.*,

$$k_o(z) = E_0(z^3) + z^{-1}E_1(z^3) + z^{-2}E_2(z^3)$$

where

$$E_0(z) = k_0 + k_3 z^{-1}, \quad E_1(z) = k_1 + k_4 z^{-1}, \quad E_2(z) = k_2 + k_5 z^{-1}.$$

Remark. We say that the transfer matrix $K_o(z)$ has a pseudo-circulant structure. A pseudo-circulant matrix is essentially a circulant matrix with the exception that all entries below the main diagonal are further multiplied by the same factor z^{-1}. For implications of the pseudo-circulant property in digital filtering and in subband adaptive filtering, see Lin and Mitra (1996) and Merched and Sayed (1998), respectively. ◆

3.20 **(Estimation of the mean of a random variable)** Consider the scalar measurements $\mathbf{y}(n) = m + \mathbf{v}(n)$, where m is a constant nonrandom complex parameter to be estimated, and $\mathbf{v}(\cdot)$ is a zero-mean random noise process. Given a collection of measurements $\mathbf{y} = \text{col}\{\mathbf{y}(0), \mathbf{y}(1), \ldots, \mathbf{y}(N)\}$, we would like to estimate m linearly from the observations.

(a) Show that $\hat{\mathbf{m}} = c^* R_v^{-1} \mathbf{y} / c^* R_v^{-1} c$, where $c = \text{col}\{1, 1, \ldots, 1\}$ and $R_v = \langle \mathbf{v}, \mathbf{v} \rangle$, $\mathbf{v} = \text{col}\{\mathbf{v}(0), \ldots, \mathbf{v}(N)\}$. Verify that the resulting m.m.s.e. is $(c^* R_v^{-1} c)^{-1}$.

(b) Show that for the special case of a white-noise sequence $\mathbf{v}(n)$ with variance σ_v^2, the estimator becomes the sample mean

$$\hat{\mathbf{m}} = \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{y}(i),$$

with m.m.s.e. $= \sigma_v^2/(N+1)$. [This estimator is consistent in the sense that the m.m.s.e. tends to zero as $N \to \infty$.]

3.21 **(Linear models)** Given zero-mean random vectors $\{\mathbf{x}, \mathbf{y}\}$, assume that the covariance matrices

$$R_x, \quad R_y, \quad \text{and} \quad \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix},$$

are all invertible. Show that we can always find $\{H, R_v\}$, $R_v > 0$, and construct a linear model of the form

$$\mathbf{y} = H\mathbf{x} + \mathbf{v}, \quad \left\langle \begin{bmatrix} \mathbf{x} \\ \mathbf{v} \end{bmatrix}, \begin{bmatrix} \mathbf{x} \\ \mathbf{v} \end{bmatrix} \right\rangle = \begin{bmatrix} R_x & 0 \\ 0 & R_v \end{bmatrix},$$

such that

$$E \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}^* = \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}.$$

Remark. This result shows that there is no loss of generality in assuming a linear model of the form (3.4.1) as opposed to assuming arbitrarily related random variables $\{\mathbf{x}, \mathbf{y}\}$. ◆

3.22 (Consistency of normal equations) Here we invoke the material developed in App. 2.A to establish that the normal equations $K_o R_y = R_{xy}$ are always consistent and that for any two solutions $K_{o,1}$ and $K_{o,2}$ we always obtain $K_{o,1}\mathbf{y} = K_{o,2}\mathbf{y}$. That is, the l.l.m.s.e. of $\mathbf{x}$ given $\mathbf{y}$ is unique.

(a) The statement is obviously true when R_y is nonsingular. So assume R_y is singular. Argue that the normal equations are consistent if, and only if, $\mathcal{R}(R_{yx}) \subseteq \mathcal{R}(R_y)$.

(b) Show that $\mathcal{N}(R_{xy}) \oplus \mathcal{R}(R_{yx}) = \mathcal{N}(R_y) \oplus \mathcal{R}(R_y)$. Conclude that the requirement of part (a) becomes $\mathcal{N}(R_y) \subseteq \mathcal{N}(R_{xy})$.

(c) Show that $\mathcal{N}(R_y) \subseteq \mathcal{N}(R_{xy})$ always holds.

(d) Let $K_{o,1}$ and $K_{o,2}$ denote any two solutions to the normal equations, when multiple solutions exist. Show that $K_{o,1}\mathbf{y} = K_{o,2}\mathbf{y}$.

3.23 (More general combined estimators) Refer to Sec. 3.4.3. Let $\mathbf{y}_a$ and $\mathbf{y}_b$ be two separate observations of a zero-mean random variable $\mathbf{x}$, such that $\mathbf{y}_a = H_a\mathbf{x} + \mathbf{v}_a$ and $\mathbf{y}_b = H_b\mathbf{x} + \mathbf{v}_b$, where we assume that

$$\left\langle \begin{bmatrix} \mathbf{v}_a \\ \mathbf{x} \end{bmatrix}, \begin{bmatrix} \mathbf{v}_a \\ \mathbf{x} \end{bmatrix} \right\rangle = \begin{bmatrix} R_a & 0 \\ 0 & M_a \end{bmatrix}, \quad \left\langle \begin{bmatrix} \mathbf{v}_b \\ \mathbf{x} \end{bmatrix}, \begin{bmatrix} \mathbf{v}_b \\ \mathbf{x} \end{bmatrix} \right\rangle = \begin{bmatrix} R_b & 0 \\ 0 & M_b \end{bmatrix}.$$

That is, the covariance matrix of $\mathbf{x}$ is assumed different in both experiments. Let $\hat{\mathbf{x}}_a$ denote the l.l.m.s. estimator of $\mathbf{x}$ given $\mathbf{y}_a$ in the first case, and let $\hat{\mathbf{x}}_b$ denote the l.l.m.s. estimator of $\mathbf{x}$ given $\mathbf{y}_b$ in the second case. The corresponding error covariance matrices are denoted by P_a and P_b, respectively. Now, let $\hat{\mathbf{x}}$ denote the l.l.m.s. estimator of $\mathbf{x}$ given both $\mathbf{y}_a$ and $\mathbf{y}_b$ and assuming $\langle \mathbf{x}, \mathbf{x} \rangle = \Pi$. Show that

$$P^{-1}\hat{\mathbf{x}} = P_a^{-1}\hat{\mathbf{x}}_a + P_b^{-1}\hat{\mathbf{x}}_b,$$

and

$$P^{-1} = P_a^{-1} + P_b^{-1} + \Pi^{-1} - M_a^{-1} - M_b^{-1}.$$

Remark. The above result shows how to combine estimators for any choice of covariances M_a and M_b. A useful choice is $M_a = \Pi$ and $M_b = \infty I$, which yields a combination of Bayes and Fischer estimators:

$$P^{-1}\hat{\mathbf{x}} = P_a^{-1}\hat{\mathbf{x}}_a + P_\infty^{-1}\hat{\mathbf{x}}_\infty, \qquad P^{-1} = P_a^{-1} + P_\infty^{-1}.$$

Here we have introduced the notation $\hat{\mathbf{x}}_\infty$ for the Fisher estimator and P_∞ for the corresponding error covariance matrix. These results will be used in Ch. 10. ◆

3.24 (Separation of signal and structured noise) Consider the model

$$\mathbf{y} = Hx + S\theta + \mathbf{v},$$

where $\mathbf{v}$ is a zero-mean additive noise random vector process with unit variance, and $\{x, \theta\}$ are unknown constant vectors. The $H \in \mathbb{C}^{N \times n}$ and $S \in \mathbb{C}^{N \times m}$ are known matrices such that $[H \;\; S]$ is full rank and $N \geq m + n$. The term $S\theta$ can be interpreted as a structured perturbation that is known to lie in the column span of S. On the other hand, the term $s \triangleq Hx$ denotes the desired signal that is corrupted by $S\theta$ and $\mathbf{v}$. We wish to estimate Hx from $\mathbf{y}$ and, hence, separate Hx from $S\theta$.

(a) Define the column vector $z \triangleq \text{col}\{x, \theta\}$. Determine the optimum unbiased l.l.m.s. estimator $\hat{\mathbf{z}}$ of z given $\mathbf{y}$.

(b) Partition $\hat{\mathbf{z}}$ into $\hat{\mathbf{z}} = \text{col}\{\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}\}$. Let $\hat{\mathbf{s}} = H\hat{\mathbf{x}}$ denote the estimator of s. Show that $\hat{\mathbf{s}}$ can be written in the form $\hat{\mathbf{s}} = \mathcal{E}_{H,S}\,\mathbf{y}$, where $\mathcal{E}_{H,S}$ is given by either expression,

$$\mathcal{E}_{H,S} = \mathcal{P}_H \left[I - S(S^*\mathcal{P}_H^\perp S)^{-1} S^*\mathcal{P}_H^\perp \right] = H(H^*\mathcal{P}_S^\perp H)^{-1} H^*\mathcal{P}_S^\perp,$$

with $\mathcal{P}_H^\perp = I - \mathcal{P}_H$, $\mathcal{P}_S^\perp = I - \mathcal{P}_S$, and $\mathcal{P}_S$ ($\mathcal{P}_H$) denotes the orthogonal projector onto the column span of S (H).

(c) Conclude that $\mathcal{E}_{H,S}S = 0$ and provide a geometric interpretation for $\mathcal{E}_{H,S}$.

(d) Let $\tilde{\mathbf{s}} = s - \hat{\mathbf{s}}$. Show that the resulting mean-square-error, $E\tilde{\mathbf{s}}\tilde{\mathbf{s}}^*$, is equal to $\mathcal{E}_{H,S}\mathcal{E}_{H,S}^*$.

(e) Assume now that $\mathbf{x}$ is modelled as a zero-mean random variable and that we have a priori knowledge about its variance, say $\Pi_0 > 0$ and $\mathbf{y} = H\mathbf{x} + S\theta + \mathbf{v}$. Show that the l.l.m.s. estimator of $\mathbf{s} = H\mathbf{x}$ is now given by $\hat{\mathbf{s}} = \mathcal{E}_{H,S}^\pi \mathbf{y}$, where

$$\mathcal{E}_{H,S}^\pi = \mathcal{P}_H^\pi \left[I - S(S^*\mathcal{P}_H^{\pi,\perp} S)^{-1} S^*\mathcal{P}_H^{\pi,\perp} \right],$$

where $\mathcal{P}_H^{\pi,\perp} = I - \mathcal{P}_H^\pi$ and $\mathcal{P}_H^\pi = H(H^*H + \Pi_0)^{-1}H^*$. Verify that we still have $\mathcal{E}_{H,S}^\pi S = 0$. Compute the new m.m.s.e. and compare it with that of part (d).

3.25 (Stochastic oblique projection) Consider two sets of observations

$$\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\} \quad \text{and} \quad \mathbf{d} = \text{col}\{\mathbf{d}_0, \ldots, \mathbf{d}_N\},$$

and two random variables $\mathbf{x}$ and $\mathbf{z}$. All variables are zero-mean. The cross-correlation matrix of $\mathbf{y}$ and $\mathbf{d}$ is denoted by $R_{yd} = \langle \mathbf{y}, \mathbf{d} \rangle$ and it is assumed invertible. Likewise, we define $R_{zy} = \langle \mathbf{z}, \mathbf{y} \rangle$ and $R_{xd} = \langle \mathbf{x}, \mathbf{d} \rangle$. We wish to determine estimators for $\mathbf{x}$ and $\mathbf{z}$ that satisfy the following requirements:

- $\hat{\mathbf{x}}$ lies in the linear span of $\mathbf{y}$, *viz.*, $\hat{\mathbf{x}} = K_x^o\mathbf{y}$ for some K_x^o.
- The estimation error $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ is orthogonal to the linear span of $\mathbf{d}$ (and not $\mathbf{y}$). That is, $\tilde{\mathbf{x}} \perp \mathbf{d}_i$ for all i.
- $\hat{\mathbf{z}}$ lies in the linear span of $\mathbf{d}$, *viz.*, $\hat{\mathbf{z}} = K_z^o\mathbf{d}$ for some K_z^o.
- The estimation error $\tilde{\mathbf{z}} = \mathbf{z} - \hat{\mathbf{z}}$ is orthogonal to the linear span of $\mathbf{y}$ (and not $\mathbf{d}$). That is, $\tilde{\mathbf{z}} \perp \mathbf{y}_i$ for all i.

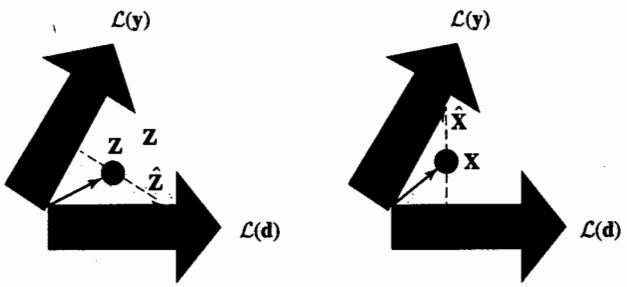


Figure 3.3 Geometric interpretation of oblique estimation.

We say that $\hat{\mathbf{z}}$ is the *oblique projection* of $\mathbf{z}$ onto the linear span of $\mathbf{d}$, $\mathcal{L}(\mathbf{d})$. This is because the resulting estimation error is not orthogonal to $\mathcal{L}(\mathbf{d})$ but rather to another space, $\mathcal{L}(\mathbf{y})$. Hence, we are projecting $\mathbf{z}$ onto $\mathbf{d}$ along a direction that is orthogonal to $\mathbf{y}$. Similarly, we say that $\hat{\mathbf{x}}$ is the oblique projection of $\mathbf{x}$ onto the linear span of $\mathbf{y}$. The situation is illustrated in Fig. 3.3.

(a) Show that $\hat{\mathbf{z}} = R_{zy}R_{dy}^{-1}\mathbf{d}$ and $\hat{\mathbf{x}} = R_{xd}R_{yd}^{-1}\mathbf{y}$. That is, $K_z^o = R_{zy}R_{dy}^{-1}$ and $K_x^o = R_{xd}R_{yd}^{-1}$.

(b) Show that (K_x^o, K_z^o) is a stationary point of the cost function $P(K_x, K_z) \triangleq \langle \tilde{\mathbf{x}}, \tilde{\mathbf{z}} \rangle$.

(c) Assume now that the $\{\mathbf{y}, \mathbf{d}, \mathbf{x}, \mathbf{z}\}$ are related via linear models of the form $\mathbf{y} = H\mathbf{x}+\mathbf{v}$ and $\mathbf{d} = A\mathbf{z} + \mathbf{w}$, where $\langle \mathbf{v}, \mathbf{w} \rangle = W$, $\langle \mathbf{z}, \mathbf{x} \rangle = \Pi$, $\langle \mathbf{v}, \mathbf{z} \rangle = 0$ and $\langle \mathbf{w}, \mathbf{x} \rangle = 0$. Determine $\hat{\mathbf{z}}$ and $\hat{\mathbf{x}}$ in terms of $\{H, A, \Pi, W, \mathbf{y}, \mathbf{d}\}$.

Remark. More details on the oblique estimation formulation of this problem can be found in Sayed and Kailath (1995). Such oblique projection problems arise in the study of instrumental variable methods in system identification (see, *e.g.*, Söderström and Stoica (1983) and Young (1984)). They also arise in some communications and array processing applications and in higher-order spectra (HOS) analysis methods (*e.g.*, Kayalar and Weinert (1989), Behrens and Scharf (1994)). A Kalman-type formulation of oblique estimation for state-space models is presented in Prob. 9.27. ◆

3.26 (Decision feedback equalization) The following is a simplified discrete time model for decision feedback equalization (DFE), which is used in practice to attentuate intersymbol interference (ISI) over communication channels. A sequence $\mathbf{x}(i), i > -\infty$, is transmitted over an FIR channel $C(z)$, say $C(z) = c_0 + c_1 z^{-1} + \ldots + c_M z^{-M}$. The channel output is corrupted by additive zero-mean noise $\mathbf{v}(i)$ that is uncorrelated with the input sequence $\{\mathbf{x}(\cdot)\}$.

The structure of the receiver includes a feedforward FIR filter $F(z)$, say $F(z) = f_0 + f_1 z^{-1} + \ldots + f_N z^{-N}$, and a nonlinear decision device that acts upon the signal $\mathbf{z}(i)$ and provides a decision regarding the transmitted signal $\mathbf{x}(i)$. The output of the decision device, denoted by $\check{\mathbf{x}}(i)$, is then fed back through an FIR feedback filter $B(z)$ of the form

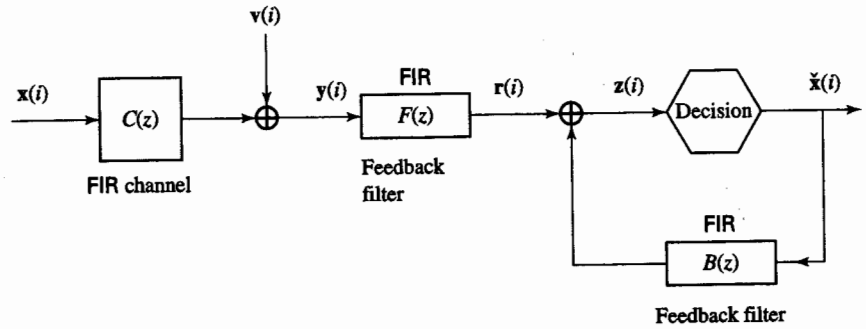


Figure 3.4 A decision feedback equalizer.

(with no direct path — see Fig. 3.4) $B(z) = -b_1 z^{-1} - b_2 z^{-2} - \ldots - b_Q z^{-Q}$. Decision feedback equalizers attempt to estimate the transmitted symbols $\{\mathbf{x}(\cdot)\}$ by employing previous decisions $\{\check{\mathbf{x}}(\cdot)\}$ in order to attenuate (or possibly eliminate) any trailing intersymbol interference. This is inherently a nonlinear scheme. In order to design $\{F(z), B(z)\}$, we make the following assumption (which is common in the literature on the analysis and design of DFE): *the decisions* $\{\check{\mathbf{x}}(i)\}$ *are correct and, hence, equal to* $\{\mathbf{x}(i)\}$.

Define the column vectors

$$\mathbf{y} \triangleq \text{col}\{\mathbf{y}(i), \ldots, \mathbf{y}(i-N)\}, \qquad \underline{\mathbf{x}} \triangleq \text{col}\{\mathbf{x}(i), \ldots, \mathbf{x}(i-M-N)\},$$
$$\mathbf{v} \triangleq \text{col}\{\mathbf{v}(i), \ldots, \mathbf{v}(i-N)\}, \qquad \mathbf{x} \triangleq \text{col}\{\mathbf{x}(i), \ldots, \mathbf{x}(i-Q)\}.$$

The covariance matrices of $\mathbf{x}$, $\underline{\mathbf{x}}$, and $\mathbf{v}$ will be denoted by R_x, $R_{\underline{x}}$, and R_v, respectively, and they are assumed known. Define also the row vectors

$$f = \begin{bmatrix} f_0 & f_1 & \ldots & f_N \end{bmatrix}, \quad b = \begin{bmatrix} 1 & b_1 & b_2 & \ldots & b_q \end{bmatrix}.$$

(a) Show that $\mathbf{y} = H\underline{\mathbf{x}}+\mathbf{v}$ for some H to be determined and find the covariance matrix R_y of $\mathbf{y}$ in terms of $\{H, R_{\underline{x}}, R_v\}$.

(b) Show that $\tilde{\mathbf{x}}(i) = b\mathbf{x} - f\mathbf{y}$, where $\tilde{\mathbf{x}}(i)$ denotes the difference $\mathbf{x}(i) - \mathbf{z}(i)$.

(c) Show that the optimal choices $\{f_{\text{opt}}, b_{\text{opt}}\}$ that minimize the error variance $\|\tilde{\mathbf{x}}(i)\|^2$ are given by

$$f_{\text{opt}} = b_{opt}R_{xy}R_y^{-1}, \quad b_{opt} = \frac{\text{first row of } R_\Delta^{-1}}{[R_\Delta^{-1}]_{00}},$$

where $R_\Delta = R_x - R_{xy}R_y^{-1}R_{yx}$, and $[R_\Delta^{-1}]_{00}$ denotes the first diagonal entry of R_Δ^{-1}.

3.27 (An optimal nonlinear estimator) Consider the observations $\mathbf{y}(i) = \mathbf{x}+\mathbf{v}(i)$, where $\mathbf{x}$ and $\mathbf{v}(i)$ are independent real-valued random variables, $\mathbf{v}(i)$ is a white-noise Gaussian process with zero-mean and unit variance, and $\mathbf{x}$ takes the values ± 1 with equal probability.

(a) Use the result of Thm. 3.A.1 to show that the l.m.s. estimate of $\mathbf{x}$ given N observations $\{y(0), \ldots, y(N-1)\}$ is

$$\hat{x}_N = \tanh\left(\sum_{i=0}^{N-1} y(i)\right).$$

(b) Assume $\mathbf{x}$ takes the value 1 with probability p and the value -1 with probability q $(p+q=1)$. Show that the l.m.s. estimate of $\mathbf{x}$ given $\{y(0), \ldots, y(N-1)\}$ is now given by

$$\hat{x}_N = \tanh\left[\frac{1}{2}\ln\left(\frac{p}{q}\right) + \sum_{i=0}^{N-1} y(i)\right].$$

3.28 **(Another nonlinear estimator)** Suppose we observe $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where $\mathbf{x}$ and $\mathbf{v}$ are independent real-valued random variables with exponential distribution with parameters λ_1 and λ_2 $(\lambda_1 \neq \lambda_2)$. Use again Thm. 3.A.1 to show that the least-mean-squares estimate of $\mathbf{x}$ given $\mathbf{y} = y$ is

$$\hat{x} = \frac{1}{\lambda_1 - \lambda_2} - \frac{e^{-\lambda_1 y}}{e^{-\lambda_2 y} - e^{-\lambda_1 y}}\, y.$$

Remark The exponential p.d.f. with parameter λ is given by $f_{\mathbf{x}}(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. ◆

Appendix for Chapter 3

3.A LEAST-MEAN-SQUARES ESTIMATION

In this appendix we consider the more general problem of determining a possibly nonlinear function $h(\cdot)$ that serves as an optimal estimator in the least-mean-squares sense of one random variable given observations of another (*cf.* (3.1.1) and (3.1.2)).

Returning to (3.1.1), we see that for each new observed value for $\mathbf{y}$ we obtain a new value for $\hat{\mathbf{x}}$. This defines an error quantity, which is also a random variable, $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - h(\mathbf{y})$. The least-mean-squares criterion minimizes the "variance" of the error variable. More explicitly, we seek a function $h(\cdot)$ that solves

$$\min_{h(\cdot)} \quad E(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^*). \tag{3.A.1}$$

We can proceed in several ways here in order to solve the optimization problem (3.A.1). One possibility is to invoke the following very useful result: for any two random variables $\mathbf{x}$ and $\mathbf{y}$ it holds that

$$E(\mathbf{x}) = E\{E\{\mathbf{x}|\mathbf{y}\}\}. \tag{3.A.2}$$

Therefore, for any (measurable, for experts) function of $\mathbf{y}$, say $g(\mathbf{y})$, we obtain

$$E\mathbf{x}g^*(\mathbf{y}) = E\left\{E\left\{\mathbf{x}g^*(\mathbf{y})|\mathbf{y}\right\}\right\} = E\left\{E\{\mathbf{x}|\mathbf{y}\}\, g^*(\mathbf{y})\right\}. \tag{3.A.3}$$

We now write the cost function as

$$E(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^* = E(\mathbf{x} - E(\mathbf{x}|\mathbf{y}) + E(\mathbf{x}|\mathbf{y}) - \hat{\mathbf{x}})(\mathbf{x} - E(\mathbf{x}|\mathbf{y}) + E(\mathbf{x}|\mathbf{y}) - \hat{\mathbf{x}})^*,$$

and take $g(\mathbf{y}) = E(\mathbf{x}|\mathbf{y}) - \hat{\mathbf{x}}$. Then

$$E(\mathbf{x} - E(\mathbf{x}|\mathbf{y}))g^*(\mathbf{y}) = E\mathbf{x}g^*(\mathbf{y}) - E\left\{E\{\mathbf{x}|\mathbf{y}\}\, g^*(\mathbf{y})\right\} = 0.$$

Therefore, the expression for the cost becomes

$$E(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^*) = E[\mathbf{x} - E(\mathbf{x}|\mathbf{y})][\mathbf{x} - E(\mathbf{x}|\mathbf{y})]^* + E[E(\mathbf{x}|\mathbf{y}) - \hat{\mathbf{x}}][E(\mathbf{x}|\mathbf{y}) - \hat{\mathbf{x}}]^*,$$

and the minimum is achieved for $\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y})$. In summary we have the following result.

Theorem 3.A.1 (The Optimal Least-Mean-Squares Estimator) *The optimal least-mean-squares (l.m.s.) estimator (cf. (3.A.1)) of a random variable* $\mathbf{x}$ *given the value of another random variable* $\mathbf{y}$ *is given by the conditional expectation*

$$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}).$$

In particular, if $\mathbf{x}$ *and* $\mathbf{y}$ *are independent random variables, then the optimal estimator of* $\mathbf{x}$ *is* $\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) = E(\mathbf{x})$. ■

The result of the theorem can also be easily extended to the case where several observable random variables $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ are available:

$$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n).$$

This holds also when $\mathbf{x}$ is itself a vector, in which case each component of $\hat{\mathbf{x}}$ is the l.m.s.e. of the corresponding component of $\mathbf{x}$.

A Geometric Interpretation. Least-mean-squares estimators also have a very important geometric interpretation in terms of an orthogonality principle. Indeed, let $g(\mathbf{y})$ denote any (measurable) function of the observable variable $\mathbf{y}$. Then, using (3.A.3) yields,

$$E\mathbf{x}g^*(\mathbf{y}) = E\hat{\mathbf{x}}g^*(\mathbf{y}),$$

which shows that $E\tilde{\mathbf{x}}g^*(\mathbf{y}) = 0$. That is, the error vector $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ is uncorrelated with (or orthogonal to) any function of the observations.

3.B GAUSSIAN RANDOM VARIABLES

The p.d.f. of a real-valued p-dimensional Gaussian random vector $\mathbf{x}$, with mean $\bar{x}$ and nonsingular covariance matrix R_x, has the form

$$f_{\mathbf{x}}(x) = \frac{1}{\sqrt{(2\pi)^p}} \frac{1}{\sqrt{\det R_x}} \exp\left\{-\tfrac{1}{2}(x-\bar{x})^T R_x^{-1}(x-\bar{x})\right\}. \tag{3.B.1}$$

The joint probability density function of two jointly Gaussian real-valued random vectors $(\mathbf{x}, \mathbf{y})$ is

$$f_{\mathbf{x},\mathbf{y}}(x, y) = \tag{3.B.2}$$

$$\frac{1}{\sqrt{(2\pi)^p}} \frac{1}{\sqrt{(2\pi)^q}} \frac{1}{\sqrt{\det R}} \exp\left\{-\tfrac{1}{2}\begin{bmatrix}(x-\bar{x})^T & (y-\bar{y})^T\end{bmatrix} R^{-1} \begin{bmatrix}x-\bar{x} \\ y-\bar{y}\end{bmatrix}\right\},$$

where R denotes the covariance matrix of the random vector $\text{col}\{\mathbf{x}, \mathbf{y}\}$, and $\text{col}\{\bar{x}, \bar{y}\}$ its mean,

$$R = \begin{bmatrix} E(\mathbf{x}-\bar{x})(\mathbf{x}-\bar{x})^T & E(\mathbf{x}-\bar{x})(\mathbf{y}-\bar{y})^T \\ E(\mathbf{y}-\bar{y})(\mathbf{x}-\bar{x})^T & E(\mathbf{y}-\bar{y})(\mathbf{y}-\bar{y})^T \end{bmatrix} = \begin{bmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{bmatrix}.$$

It is clear from (3.B.2) that the p.d.f. of two jointly Gaussian *real-valued* random variables is completely specified by the first- and second-order statistics of the random variables, *viz.*, their means $(\bar{x}, \bar{y})$ and their covariance and cross-covariance matrices (R_x, R_y, R_{xy}). While this might seem an obvious conclusion in this case, it does not necessarily hold for complex-valued Gaussian random variables unless the variables satisfy additional conditions. Let us first introduce complex-valued random variables, which are useful, for example, in studying the so-called narrowband random processes encountered in communications and radar problems.

Complex-Valued Gaussian Random Variables. A complex-valued random variable $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ is said to be Gaussian if its real and imaginary parts $(\mathbf{x}, \mathbf{y})$ are jointly Gaussian. Now assume that we are given the first- and second-order statistics of $\mathbf{z}$, *viz.*, its mean $\bar{z}$ and its covariance matrix R_z. The given mean completely specifies the values of $\bar{x}$ and $\bar{y}$ since $\bar{z} = \bar{x} + j\bar{y}$. However, the given covariance matrix R_z does not completely specify the covariance matrices (R_x, R_y, R_{xy}) that we need in order to write down the joint probability density function of $(\mathbf{x}, \mathbf{y})$. This is because

$$R_z \triangleq E(\mathbf{z}-\bar{z})(\mathbf{z}-\bar{z})^* = (R_x + R_y) + j(R_{yx} - R_{xy}), \tag{3.B.3}$$

where

$$R_x = E(\mathbf{x}-\bar{x})(\mathbf{x}-\bar{x})^T, \quad R_{xy} = E(\mathbf{x}-\bar{x})(\mathbf{y}-\bar{y})^T,$$

and similarly for R_y. We therefore see from (3.B.3) that knowledge of R_z alone provides us with the values of the sum $(R_x + R_y)$ and the difference $(R_{yx} - R_{xy})$. This information is not enough to uniquely determine all covariances (R_x, R_y, R_{xy}).

In order to be able to uniquely determine (R_x, R_y, R_{xy}) we need additional information, which can be obtained by also assuming knowledge of $R_{z,2} = E(\mathbf{z}-\bar{z})(\mathbf{z}-\bar{z})^T$. It is easy to verify that

$$R_{z,2} = E(\mathbf{z}-\bar{z})(\mathbf{z}-\bar{z})^T = (R_x - R_y) + j(R_{yx} + R_{xy}). \tag{3.B.4}$$

By using (3.B.3) and (3.B.4), we can uniquely determine (R_x, R_y, R_{xy}) and therefore completely specify the joint p.d.f. of the random variables $(\mathbf{x}, \mathbf{y})$ that define $\mathbf{z}$. In other words, a complex-valued Gaussian random variable is completely specified in terms of its mean and in terms of the covariance matrices $\{R_z, R_{z,2}\}$.

It is *generally* assumed that

$$R_{z,2} = 0, \tag{3.B.5}$$

in which case we can solve for (R_x, R_y, R_{xy}) from R_z alone. Indeed, $R_{z,2} = 0$ implies that $R_x = R_y$ and $R_{yx} = -R_{xy}$. Consequently,

$$R_x = \frac{1}{2}\,\text{Real}(R_z) = R_y \quad \text{and} \quad R_{xy} = -R_{yx} = -\frac{1}{2}\,\text{Imag}(R_z). \tag{3.B.6}$$

Complex-valued Gaussian random variables $\mathbf{z}$ that satisfy (3.B.5) are said to be *circular* (or spherically invariant). In this case, the joint p.d.f. of the real and imaginary parts of $\mathbf{z}$ is given by (3.B.2) with $p = q$ and with the above expressions (3.B.6), *i.e.*, with

$$R = \frac{1}{2}\begin{bmatrix}\text{Real}(R_z) & -\text{Imag}(R_z) \\ \text{Imag}(R_z) & \text{Real}(R_z)\end{bmatrix} = \begin{bmatrix} R_x & R_{xy} \\ -R_{xy} & R_x \end{bmatrix}. \tag{3.B.7}$$

Now, the joint p.d.f. of $(\mathbf{x}, \mathbf{y})$ can be re-expressed in terms of $\mathbf{z}$ and R_z in the following form, which should be compared with (3.B.1); the proof is omitted.

Lemma 3.B.1 (Circular Gaussian Random Variables) *The p.d.f. of a complex valued circular (or spherically invariant) Gaussian random variable* $\mathbf{z}$ *of dimension* p *is given by*

$$f_{\mathbf{z}}(z) = \frac{1}{\pi^p}\frac{1}{\det R_z}\exp\{-(z-\bar{z})^* R_z^{-1}(z-\bar{z})\}. \tag{3.B.8}$$

■

When (3.B.8) holds, we can check that uncorrelated jointly Gaussian random variables will also be independent; this is one important reason for the assumption of circularity (*cf.* Doob (1953, Sec. II.3)).

3.C OPTIMAL ESTIMATION FOR GAUSSIAN VARIABLES

We now verify that the optimal nonlinear least-mean-squares estimator of Thm. 3.A.1 collapses to a linear estimator when $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian zero-mean random vector variables, say with (a nonsingular) covariance matrix

$$R \triangleq \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}.$$

Here, $R_x = E\mathbf{x}\mathbf{x}^*$, $R_y = E\mathbf{y}\mathbf{y}^*$, and $R_{xy} = E\mathbf{x}\mathbf{y}^* = R_{yx}^*$. Different dimensions are assumed for $\mathbf{x}$ and $\mathbf{y}$ for the sake of generality: $\mathbf{x}$ is $p \times 1$ and $\mathbf{y}$ is $q \times 1$. The random variables are also assumed to be circular, and hence, $E\mathbf{y}\mathbf{y}^T = 0$, $E\mathbf{x}\mathbf{x}^T = 0$, and $E\mathbf{x}\mathbf{y}^T = 0$. With these assumptions, we can write

$$f_{\mathbf{x}}(x) = \frac{1}{\pi^p}\frac{1}{\det R_x}\exp\{-x^* R_x^{-1}x\}, \qquad f_{\mathbf{y}}(x) = \frac{1}{\pi^q}\frac{1}{\det R_y}\exp\{-y^* R_y^{-1}y\},$$

and, consequently, the joint p.d.f. of $\{\mathbf{x}, \mathbf{y}\}$ is proportional to

$$f_{\mathbf{x},\mathbf{y}}(x, y) \propto \exp\left\{-\begin{bmatrix} x^* & y^* \end{bmatrix} R^{-1}\begin{bmatrix} x \\ y \end{bmatrix}\right\}.$$

The l.m.s.e. of $\mathbf{x}$ requires the determination of the conditional probability density function

$$f_{\mathbf{x}|\mathbf{y}}(x|y) = \frac{f_{\mathbf{x},\mathbf{y}}(x, y)}{f_{\mathbf{y}}(y)} \propto \exp\left\{-\begin{bmatrix} x^* & y^* \end{bmatrix} R^{-1}\begin{bmatrix} x \\ y \end{bmatrix}\right\} \exp\{y^* R_y^{-1} y\}.$$

The above expression can be simplified by invoking the following factorization (see App. A),

$$R^{-1} = \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -R_y^{-1}R_{yx} & I \end{bmatrix}\begin{bmatrix} \Delta^{-1} & 0 \\ 0 & R_y^{-1} \end{bmatrix}\begin{bmatrix} I & -R_{xy}R_y^{-1} \\ 0 & I \end{bmatrix},$$

where $\Delta = R_x - R_{xy}R_y^{-1}R_{yx}$ is the Schur complement of R_y in R. It follows that

$$\begin{bmatrix} x^* & y^* \end{bmatrix} R^{-1}\begin{bmatrix} x \\ y \end{bmatrix} = (x - R_{xy}R_y^{-1}y)^*\Delta^{-1}(x - R_{xy}R_y^{-1}y) + y^*R_y^{-1}y,$$

which allows us to simplify $f_{\mathbf{x}|\mathbf{y}}(x|y)$ to

$$f_{\mathbf{x}|\mathbf{y}}(x|y) \propto \exp\left\{-(x-R_{xy}R_y^{-1}y)^*\Delta^{-1}(x-R_{xy}R_y^{-1}y)\right\}.$$

This has the form of the p.d.f. of a circular Gaussian variable with covariance matrix Δ and mean $R_{xy}R_y^{-1}y$. Consequently,

$$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) = R_{xy}R_y^{-1}\mathbf{y}.$$

Moreover, the corresponding minimum mean-square-error (m.m.s.e) matrix is readily seen to be

$$\text{m.m.s.e.} = E[\mathbf{x} - \hat{\mathbf{x}}][\mathbf{x} - \hat{\mathbf{x}}]^* = E\mathbf{x}\mathbf{x}^* - E\mathbf{x}\hat{\mathbf{x}}^* = R_x - R_{xy}R_y^{-1}R_{yx} = \Delta.$$

This quantity can be evaluated a priori, which provides a mechanism for checking whether the l.m.s. estimator will be an acceptable solution.

We may note that if the complex-valued Gaussian variables $\mathbf{x}$ and $\mathbf{y}$ were not circular, then we could have proceeded by first replacing each of them by its real and imaginary parts, say

$$\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I, \quad \mathbf{y} = \mathbf{y}_R + j\mathbf{y}_I,$$

and then posing the problem of estimating $(\mathbf{x}_R, \mathbf{x}_I)$ from $(\mathbf{y}_R, \mathbf{y}_I)$. By working with the joint probability density functions of the now real-valued Gaussian random variables $(\mathbf{x}_R, \mathbf{x}_I, \mathbf{y}_R, \mathbf{y}_I)$, we can repeat the analysis of this section to conclude that the optimal least mean-square estimator of $(\mathbf{x}_R, \mathbf{x}_I)$ given $(\mathbf{y}_R, \mathbf{y}_I)$ is still linear and given by the estimator defined by (3.2.16) and (3.2.17).

CHAPTER 4

The Innovations Process

So far in the book we have considered the problem of estimating one random variable, $\mathbf{x}$, given another random variable, $\mathbf{y}$. In many applications these random variables have additional structure and, for example, they may arise from random processes as in the problem studied in Ch. 1. When there is structure, new issues and new results arise, and we shall begin to explore them in this chapter and in the rest of this book.

In particular, when the random variables come from an indexed family, *i.e.*, they are stochastic processes, the number of observed random variables can be very large. In such problems both computational and data storage considerations lead us to search for additional structure in the stochastic processes that will allow a reduction in the very large effort needed to solve the normal equations when this structure is exploited.

In this chapter, we shall show that the geometric formulation suggests that the introduction of a so-called *innovations* or *new information* process will help us in trying to exploit the presence of structure. The basic, but simple, observation is that projections on a linear subspace are particularly easy to compute when we have an orthogonal basis for the subspace (see Sec. 4.2). The value of this point of view will be illustrated in Secs. 4.2.5 and 4.3. However, the number of computations, whether we use the innovations or not, will be the same, and very high, unless we can understand how to identify and exploit special structure in the problem. We begin to illustrate this in a very simple problem in Sec. 4.4. Further study of this problem will lead us to introduce state-space models in Ch. 5. These models will prove very convenient in solving estimation problems for vector-valued and/or nonstationary stochastic processes.

However, before launching into the discussions described above, we should emphasize that while most of our discussion will be in the context of random variables and stochastic processes, the geometric formulation can also be applied to ordinary vectors (in $\mathbb{R}^n$ or $\mathbb{C}^n$), as we showed in Ch. 2. Therefore to facilitate a broader applicability of our results, we shall often (without being overly pedantic about it) use the terms *vectors* and *Gramians* instead of *random variables* and *covariance matrices*. This is explained in App. 4.A. There we also address the issue of freely applying the geometric language to vector-valued random variables, which will require the (relatively unfamiliar) concept of inner products that are not just complex numbers, but that can be complex matrices.

As an illustration of the fact that our geometric insights apply equally to random variables and to Euclidean vectors, we shall show in Sec. 4.3 that applying the innovations technique to the deterministic least-squares problem of Ch. 2 will lead directly to the array algorithm of Sec. 2.5.

4.1 ESTIMATION OF STOCHASTIC PROCESSES

We shall now assume that we are dealing with indexed sets of random variables, *i.e.*, with stochastic processes. In particular, we shall focus on discrete-time zero-mean random processes $\{\mathbf{s}_i\}$ and $\{\mathbf{y}_i\}$ with known cross-covariance and covariance sequences, $R_{sy}(i,l) = E\mathbf{s}_i\mathbf{y}_l^*$ and $R_y(i,l) = E\mathbf{y}_i\mathbf{y}_l^*$. The process $\{\mathbf{s}_i\}$ will be referred to as the *signal* process and is not assumed observable, while $\{\mathbf{y}_i\}$ is the observable *measurement* process. The goal will be to use the measurement process, $\{\mathbf{y}_i\}$, to estimate the signal process, $\{\mathbf{s}_i\}$.

Three types of problems can be envisioned.

(a) Smoothing: For each i, and for fixed $N \geq i$, estimate $\mathbf{s}_i$ given the observations $\{\mathbf{y}_j, 0 \leq j \leq N\}$, say

$$\hat{\mathbf{s}}_{i|N} = \sum_{j=0}^{N} k_{s,ij}\mathbf{y}_j. \tag{4.1.1}$$

In other words, determine the set of coefficients $\{k_{s,ij}\}$ such that the error variance, $\|\mathbf{s}_i - \hat{\mathbf{s}}_{i|N}\|^2$, is minimized. Such a problem is referred to as a *smoothing problem*, and the corresponding estimators as *smoothed estimators* since the estimator $\hat{\mathbf{s}}_{i|N}$ depends on future, as well as current and past, observations of the process $\{\mathbf{y}_i\}$. For this reason, we shall often say that $\hat{\mathbf{s}}_{i|N}$ is a noncausal function of the observations.

(b) Causal Filtering: For each i, estimate $\mathbf{s}_i$ given only the past and present observations $\{\mathbf{y}_j\}_{j=0}^{i}$, say

$$\hat{\mathbf{s}}_{i|i} = \sum_{j=0}^{i} k_{f,ij}\mathbf{y}_j. \tag{4.1.2}$$

In other words, determine some (in general, different) set of coefficients $\{k_{f,ij}\}$, satisfying $k_{f,ij} = 0$ for $j > i$, that minimize the error variance $\|\mathbf{s}_i - \hat{\mathbf{s}}_{i|i}\|^2$. This problem is referred to as a *filtering problem*, and the corresponding estimators as *filtered estimators*.

(c) Prediction and Filtering: For each i, and for a fixed integer λ, estimate $\mathbf{s}_{i+\lambda}$ given the observations $\{\mathbf{y}_j\}_{j=0}^{i}$, say

$$\hat{\mathbf{s}}_{i+\lambda|i} = \sum_{j=0}^{i} k_{\lambda,ij}\mathbf{y}_j. \tag{4.1.3}$$

In other words, determine the set of coefficients $\{k_{\lambda,ij}\}$ (satisfying $k_{\lambda,ij} = 0$ for $j > i$) that minimize the error variance $\|\mathbf{s}_{i+\lambda} - \hat{\mathbf{s}}_{i+\lambda|i}\|^2$. When $\lambda > 0$, this is referred to as a *prediction problem*, since the goal is to predict the future values of the signal using current and past observations.

As we shall presently see, the first of the above problems can be solved by a straightforward application of the results of the previous chapter. The second problem, however, turns out to be more challenging since the estimators are restricted to be causal functions of the observations; a solution using (a simple version of) a famous so-called Wiener-Hopf technique will be presented in Sec. 4.1.3 (and via the so-called innovations method in Sec. 4.2.5).

We describe the solution in Sec. 4.1.3, along with some remarks on relations between the smoothing and filtering problems. We shall not study the general prediction problem here, though some special cases will be encountered in the problems (see Probs. 4.9–4.11).

4.1.1 The Fixed Interval Smoothing Problem

The key to estimating one stochastic process, $\{\mathbf{s}_j\}$, from another stochastic process, $\{\mathbf{y}_j\}$, is to estimate the random variable $\mathbf{s}_i$ from the observations, for *each* value of i. Therefore we begin by choosing a certain time instant i, and attempting to estimate $\mathbf{s}_i$ given $\{\mathbf{y}_j\}_{j=0}^N$, as $\hat{\mathbf{s}}_{i|N} = \sum_{j=0}^N k_{s,ij}\mathbf{y}_j$. The orthogonality principle says that for optimality we must have

$$\left(\mathbf{s}_i - \sum_{j=0}^{N} k_{s,ij}\mathbf{y}_j\right) \perp \mathbf{y}_l, \qquad \text{for } l = 0, \ldots N, \tag{4.1.4}$$

so that computing the inner product of the above estimation error with $\mathbf{y}_l$ yields

$$R_{sy}(i,l) = \sum_{j=0}^{N} k_{s,ij} R_y(j,l), \quad \text{for } l = 0, \ldots N. \tag{4.1.5}$$

Collecting these equations together gives the matrix equation

$$\begin{bmatrix} R_{sy}(i,0) & \ldots & R_{sy}(i,N) \end{bmatrix} = \begin{bmatrix} k_{s,i0} & \ldots & k_{s,iN} \end{bmatrix} R_y, \tag{4.1.6}$$

where $R_y = \left[R_y(i,j)\right]_{i,j=0}^N$.

We have to solve a similar equation for all $i = 0, 1, \ldots, N$. However, notice that the coefficient matrix R_y is the same for all i, so that we can write the solution of the smoothing problem even more compactly as follows. Introduce the matrices

$$R_{sy} \triangleq \left[R_{sy}(i,j)\right]_{i,j=0}^N, \qquad K_s \triangleq \left[k_{s,ij}\right]_{i,j=0}^N,$$

and let

$$\mathbf{s} = \text{col}\{\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_N\}, \quad \hat{\mathbf{s}}_s = \text{col}\{\hat{\mathbf{s}}_{0|N}, \hat{\mathbf{s}}_{1|N}, \ldots, \hat{\mathbf{s}}_{N|N}\}, \quad \mathbf{y} = \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}.$$

[The subscript s in $\hat{\mathbf{s}}_s$ denotes smoothing.] Then note that $R_y = \langle \mathbf{y}, \mathbf{y} \rangle$, $R_{sy} = \langle \mathbf{s}, \mathbf{y} \rangle$, and that collecting Eqs. (4.1.1) and (4.1.6) for each $i = 0, \ldots, N$ gives the equations

$$\hat{\mathbf{s}}_s = K_s\mathbf{y} \quad \text{and} \quad K_s R_y = R_{sy}, \tag{4.1.7}$$

or when $R_y > 0$,

$$\hat{\mathbf{s}}_s = R_{sy}R_y^{-1}\mathbf{y} = \langle \mathbf{s}, \mathbf{y} \rangle \langle \mathbf{y}, \mathbf{y} \rangle^{-1} \mathbf{y}.$$

That is, we just have the problem of estimating the vector $\mathbf{s}$ given the vector $\mathbf{y}$. Therefore the solution to the smoothing problem is no different from the solution to the standard estimation problem: we must solve the normal equations. Often, however, the fact that the components of $\mathbf{y}$ come from a random process, $\{\mathbf{y}_i\}$, introduces additional structure that allows one to solve these equations more efficiently. In fact, introducing and exploiting special structures will be the major theme of this book.

4.1.2 The Causal Filtering Problem

In the filtering problem, we seek as explained above (*cf.* (4.1.2)), coefficients $\{k_{f,ij}\}$ such that, for each i and l,

$$k_{f,ij} = 0 \quad \text{for } j > i \quad \text{and} \quad \left(\mathbf{s}_i - \sum_{j=0}^{i} k_{f,ij}\mathbf{y}_j\right) \perp \mathbf{y}_l, \quad \text{for } l = 0, \ldots i. \tag{4.1.8}$$

In other words, the estimation error is only orthogonal to past and current observations, since those are the only observations that can be used to determine the estimators. Forming the inner products in (4.1.8) leads to the equations

$$R_{sy}(i,l) = \sum_{j=0}^{i} k_{f,ij} R_y(j,l), \quad l = 0, \ldots i. \tag{4.1.9}$$

Collecting these equations for each i will give us a matrix equation different from (4.1.7). To write it down, we need to first introduce some matrix notation. Let K_f be the $(N+1) \times (N+1)$ *lower triangular* matrix

$$K_f \triangleq \left[k_{f,ij}\right]_{i,j=0}^N, \qquad k_{f,ij} = 0 \quad \text{for} \quad j > i. \tag{4.1.10}$$

Then the reader should check that we can write (4.1.2) as

$$\hat{\mathbf{s}}_f \triangleq \begin{bmatrix} \hat{\mathbf{s}}_{0|0} \\ \hat{\mathbf{s}}_{1|1} \\ \vdots \\ \hat{\mathbf{s}}_{N|N} \end{bmatrix} = \begin{bmatrix} k_{f,00} & & & \\ k_{f,10} & k_{f,11} & & \\ \vdots & & \ddots & \\ k_{f,N0} & k_{f,N1} & \ldots & k_{f,NN} \end{bmatrix} \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = K_f\mathbf{y}. \tag{4.1.11}$$

[The subscript f in $\hat{\mathbf{s}}_f$ denotes filtering.] However, to write (4.1.9) in matrix form needs more notation. To see this, note that with the definition (4.1.10), the right-hand side of Eq. (4.1.9) is simply the i-th row of the matrix product $K_f R_y$, $R_y = \langle \mathbf{y}, \mathbf{y} \rangle$, and the left-hand side is the i-th row of $R_{sy} = \langle \mathbf{s}, \mathbf{y} \rangle$. However, since the equality in (4.1.9) only holds for $l \leq i$, the entries of the matrices R_{sy} and $K_f R_y$ are only equal at the *lower triangular* entries (*i.e.*, those for which $l \leq i$). To capture this fact, let us introduce an operator that retains the lower triangular entries of a matrix,

$$\left[\{A\}_{\text{lower}}\right]_{ij} = \begin{cases} 0 & i < j, \\ A_{ij} & i \geq j. \end{cases} \tag{4.1.12}$$

Then collecting Eqs. (4.1.9), for $i = 0, 1, \ldots, N$, leads to the unconventional matrix equation in K_f

$$\{R_{sy} - K_f R_y\}_{\text{lower}} = 0. \tag{4.1.13}$$

Note the difference from the ordinary normal equation (4.1.7) where we have equality for all entries of the matrix on the left-hand side, and not just the lower triangular part.

In fact, since the equality in (4.1.13) holds only for the lower triangular part, and since the solution K_f is restricted to be lower triangular, solving Eq. (4.1.13) seems to be a formidable task. However, it can be solved by invoking an ingenious technique that was originally introduced by Wiener and Hopf (1931) for the solution of what is now known as the Wiener-Hopf integral equation (see Eq. (7.2.2) — see also Wiener (1942)). The interested reader can get a simple preview of that ingenious technique here.

4.1.3 The Wiener-Hopf Technique

To begin to solve Eq. (4.1.13), we rewrite the equation as

$$K_f R_y - R_{sy} = U^+, \tag{4.1.14}$$

where U^+ is a yet-unknown *strictly* upper triangular matrix, *i.e.*, $\{U^+\}_{\text{lower}} = 0$. Now the key (unmotivated for the moment — but see Sec. 4.2.5) to determining K_f from (4.1.13) is to introduce the *unique lower-diagonal-upper triangular* factorization of R_y,

$$R_y = L R_e L^*, \tag{4.1.15}$$

where L is lower triangular with unit diagonal and R_e is diagonal.[1] Then we may write

$$K_f L R_e L^* - R_{sy} = U^+, \tag{4.1.16}$$

and from there

$$K_f L = R_{sy} L^{-*} R_e^{-1} + U^+ L^{-*} R_e^{-1}. \tag{4.1.17}$$

Since the product of two lower triangular matrices is lower triangular, and since the product of a strictly upper triangular matrix with an upper triangular matrix is strictly upper triangular, we can therefore identify the structure of the matrices in (4.1.17) as follows:

$$\underbrace{K_f L}_{\text{lower}} = \underbrace{R_{sy} L^{-*} R_e^{-1}}_{\text{mixed}} + \underbrace{U^+ L^{-*} R_e^{-1}}_{\text{strictly upper}}. \tag{4.1.18}$$

Clearly, it must be true that

$$K_f L = \{R_{sy} L^{-*} R_e^{-1}\}_{\text{lower}}, \tag{4.1.19}$$

so that

$$K_f = \{R_{sy} L^{-*} R_e^{-1}\}_{\text{lower}} L^{-1}. \tag{4.1.20}$$

A neat idea, indeed! (The interesting story of how the name Wiener-Hopf is attached to equations of the form (4.1.13) is told in Sec. 7.2). An important ingredient (the *deus ex machin*) is really the triangular factorization (4.1.15), which is unmotivated at this point. However, we shall show in Sec. 4.2.5 that by pursuing the physical meaning of the LR_eL^* factorization of R_y we can obtain the result (4.1.20) in a less mysterious way. We summarize the discussion in the following statement for ease of reference.

[1] Sufficient conditions for such a factorization to exist and be unique are that $R_y > 0$ or that R_y be strongly regular (all leading minors nonzero) — see App. A.

Lemma 4.1.1 (Optimal Causal Filter) *The optimal coefficient matrix K_f in (4.1.11) that solves the causal filtering problem (4.1.2) (or, equivalently, (4.1.8)) is given by*

$$K_f = \{R_{sy} L^{-*} R_e^{-1}\}_{\text{lower}} L^{-1}.$$

■

The formula (4.1.20) can be simplified in the following important special case.

Signals in Additive White Noise. Consider the linear model $\mathbf{y} = \mathbf{s} + \mathbf{v}$, with

$$\left\langle \begin{bmatrix} \mathbf{s} \\ \mathbf{v} \end{bmatrix}, \begin{bmatrix} \mathbf{s} \\ \mathbf{v} \end{bmatrix} \right\rangle = \begin{bmatrix} R_s & 0 \\ 0 & R_v \end{bmatrix}, \quad \text{with } R_v \text{ diagonal and nonsingular.} \tag{4.1.21}$$

Since $R_{sy} = R_s$ and $R_y = R_s + R_v$, expression (4.1.20) reduces to

$$\begin{aligned} K_f &= \{(R_y - R_v) L^{-*} R_e^{-1}\}_{\text{lower}} L^{-1}, \\ &= \{L - R_v L^{-*} R_e^{-1}\}_{\text{lower}} L^{-1}, \\ &= I - \{R_v L^{-*} R_e^{-1}\}_{\text{lower}} L^{-1}. \end{aligned} \tag{4.1.22}$$

Now since L^{-*} is upper triangular, and $\{R_v, R_e\}$ are diagonal, we obtain

$$\{R_v L^{-*} R_e^{-1}\}_{\text{lower}} = R_v R_e^{-1} \quad \text{and therefore} \quad K_f = I - R_v R_e^{-1} L^{-1}. \tag{4.1.23}$$

For the smoothing problem, we shall have

$$K_s = R_s (R_s + R_v)^{-1} = I - R_v R_y^{-1} = I - R_v R_y^{-1} = I - R_v L^{-*} R_e^{-1} L^{-1}. \tag{4.1.24}$$

Relations between the Causal Filtering and Smoothing Problems. The optimal coefficient matrices $\{K_s, K_f\}$ in (4.1.7) and (4.1.20) can be related to each other as follows. Let $\{\cdot\}_{\text{s.upper}}$ denote the strictly upper triangular part of its argument. Now assuming $R_y > 0$, we have $R_y^{-1} = L^{-*} R_e^{-1} L^{-1}$ so that

$$\begin{aligned} K_s = R_{sy} R_y^{-1} &= R_{sy} L^{-*} R_e^{-1} L^{-1}, \\ &= \left[\{R_{sy} L^{-*} R_e^{-1}\}_{\text{lower}} + \{R_{sy} L^{-*} R_e^{-1}\}_{\text{s.upper}} \right] L^{-1}, \\ &= K_f + \{R_{sy} L^{-*} R_e^{-1}\}_{\text{s.upper}} L^{-1}. \end{aligned} \tag{4.1.25}$$

Observe that the correction term that is added to K_f in order to obtain K_s is of mixed nature (it is neither lower triangular nor upper triangular).

Likewise, we can relate the cost functions that are minimized by K_s and K_f. Recall that K_s is the optimal coefficient for estimating s from y in the linear least-mean-squares sense. Therefore, K_s minimizes the error variance matrix, *i.e.*, it solves

$$\min_K \quad \|\mathbf{s} - K\mathbf{y}\|^2 \quad \Longrightarrow \quad \hat{\mathbf{s}} = K_s\,\mathbf{y}.$$

The matrix K_f in (4.1.20), on the other hand, is the optimal solution of the following optimization problem

$$\min_{\text{lower trian. } K} \quad \text{Tr } \|\mathbf{s} - K\mathbf{y}\|^2 \quad \Longrightarrow \quad \hat{\mathbf{s}} = K_f\mathbf{y}. \tag{4.1.26}$$

That is, K_f minimizes the *trace* of the error variance matrix over all *lower* triangular matrices K. This is because each individual row of K_f is, by definition (see (4.1.11)), the optimal vector that minimizes the error variance in estimating the corresponding entry of s. For example, the nonzero part of the i-th row of K_f is the solution to the following optimization problem:

$$\min_{k_i} \quad E\,\|\,\mathbf{s}_i - k_i\,\mathbf{y}_{0:i}\,\|^2, \qquad \mathbf{y}_{0:i} \triangleq \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\},$$

so that K_f solves (see also Prob. 4.3):

$$\min_{k_0,k_1,\ldots,k_m} \left(\sum_{i=0}^{m} E\,\|\,\mathbf{s}_i - k_i\,\mathbf{y}_{0:i}\,\|^2\right) = \min_{\text{lower trian. } K} \text{Tr } \|\mathbf{s} - K\mathbf{y}\|^2.$$

Computational Issues. We have seen that the key to the solution of the filtering problem was the triangular factorization of the covariance matrix R_y. In the conceptually easier smoothing problem, all we needed was inversion of the matrix R_y or, at least, the solution of a certain set of linear equations with coefficient matrix R_y. Actually, triangular factorization, matrix inversion, solution of linear equations all take essentially the same number of computations: $O(N^3)$ flops. Moreover, triangular factorization also turns out to be a useful step for matrix inversion and linear equation solution. The reasons why this is true will be made clear in the next sections, where we shall also begin to explore the issue of how the number of computations required can be reduced by exploiting additional structure in the problem.

4.1.4 A Note on Terminology — Vectors and Gramians

However, before proceeding further, it is useful to introduce the terminology of vectors and Gramians in order to unify our treatment of both the deterministic and stochastic least-squares problems. Recall that for the apparently very different deterministic least-squares problem of Ch. 2 and the stochastic least-squares problem of Ch. 3, the respective solutions were obtained by projecting one vector onto a space spanned by some other vector(s). In Ch. 2, the vectors were elements in an Euclidean space, while in Ch. 3 they were vectors in a probability space (a more abstract space, in the sense that one need not be too concerned with exactly specifying the space Ω, or the probability measure on it).

The essential unity of the two problems is masked to some extent by using the language of random variables, variances, and covariances. While, for a variety of reasons, most of the discussion in the rest of the book will deal with the stochastic case, it will be helpful to keep the geometric picture clearly in view by adopting a more neutral language. Thus, while it will be overly pedantic to insist on always replacing the term *random variable* by *vector*, by and large we shall abandon the notations $R_x = E\mathbf{x}\mathbf{x}^*$ and $R_{xy} = E\mathbf{x}\mathbf{y}^*$ in favor of $R_x = \langle \mathbf{x}, \mathbf{x}\rangle$ and $R_{xy} = \langle \mathbf{x}, \mathbf{y}\rangle$. Moreover, we shall generally refer to these quantities not as covariances and cross-covariances, but rather as Gramians and cross-Gramians. For convenience, a brief review of abstract linear vector spaces is given in App. 4.A, along with several specific examples relevant to our discussions in this book. However, the study of this appendix is not essential at this point.

Here we proceed to pursue the geometric notion of projection in order to introduce the fundamental concept of innovations processes as an aid in exploiting special structure.

4.2 THE INNOVATIONS PROCESS

In Chs. 2 and 3 and Sec. 4.1 we saw that the solution of deterministic and stochastic least-squares problems reduced to the solution of certain linear equations. Since the solution of linear equations is a much studied problem, it would seem that there is not much more to be said, except to refer to some books on the subject. However, there are at least two features of the problem that should give us some pause:

(a) It takes proportional N^3 operations (an operation may be taken as the multiplication or addition of two real numbers) to solve an $N \times N$ set of linear equations. This can be a substantial amount of work when N is large: N could be of the order 10–100 in several aerospace problems and 500–2000–4000–10,000 in many environmental, geodetic, power-system, econometric, and image processing problems.

(b) For large N, there may be a problem of data storage, especially since in many applications the data comes in sequentially, so that we have to solve the estimation problem for sequentially increasing values of N. The storage problem could be ameliorated if we could develop a *sequential* or *recursive* method of solving the equations: it would be nice if the new datum could be used to update the previous estimate, and then discarded, so that no data storage is necessary. Note that recursive solutions can be useful whenever N is large, whether or not it is growing.

While general methods are known for the recursive solution of linear equations, the problem must have some special structure if the number of computations (and the amount of storage) is to be significantly reduced, to say $O(N^2)$ or even $O(N)$ from $O(N^3)$. The exploration of structure can be carried out by algebraic or geometric methods, and in several different ways.

4.2.1 A Geometric Approach

Recall that we are not interested in linear equations as such, but in those that arise from the problem of computing the projection of a vector, say x, onto the linear space spanned by another set of vectors (or random variables) $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}$. As we have

seen in Sec. 3.2.1, this problem reduces to the solution of a simultaneous set of linear equations, say $K_o R_y = R_{xy}$, where

$$R_y = \left[\langle \mathbf{y}_i, \mathbf{y}_j \rangle\right]_{i,j=0:N} \quad \text{and} \quad R_{xy} = \left[\langle \mathbf{x}, \mathbf{y}_i \rangle\right]_{i=0:N}.$$

Now it is a pretty obvious remark that these equations would be easy to solve if R_y were a diagonal matrix, or equivalently if the $\{\mathbf{y}_i\}$ were orthogonal to each other, in which case the projection would reduce to just the sum of the projections of $\mathbf{x}$ onto each orthogonal vector. Of course, in most problems, R_y would not be diagonal; in fact, it is the nature of the dependence between the vectors $\{\mathbf{y}_i\}$ that distinguishes various physical problems from each other. Now, as mentioned before, from now on we shall always assume that the variables $\{\mathbf{y}_i\}$ are not an arbitrary collection, but belong to an indexed or ordered set, in the sense that $\mathbf{y}_{i+1}$ follows $\mathbf{y}_i$. In other words, we assume that the $\{\mathbf{y}_i\}$ constitute a *stochastic process*, where the index i will be assumed, for definiteness, to be a time index, though it could also be a space index if desired.

The fact that the generally nonorthogonal vectors $\{\mathbf{y}_i\}$ arise from an indexed set may immediately remind us of the obvious (in retrospect) *recursive* Gram-Schmidt procedure for replacing a set of indexed vectors by an *equivalent* orthogonal set of vectors. Thus assume that we have transformed $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ to an equivalent set of orthogonal vectors $\{\mathbf{e}_0, \ldots, \mathbf{e}_N\}$, equivalent in the sense that they span the same linear (sub)space, written

$$\mathcal{L}\{\mathbf{e}_0, \ldots, \mathbf{e}_N\} = \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\} \triangleq \mathcal{L}_N, \quad \text{say.} \tag{4.2.1}$$

If now we have an additional vector, $\mathbf{y}_{N+1}$, a natural way of proceeding is by projecting $\mathbf{y}_{N+1}$ onto $\mathcal{L}_N$ to get

$$\mathbf{e}_{N+1} = \mathbf{y}_{N+1} - \text{Proj.}\{\mathbf{y}_{N+1}|\mathcal{L}_N\}. \tag{4.2.2}$$

Moreover, finding the above projection is aided by property (4.2.1), which allows us to find the projection by separately projecting onto each of the previously found orthogonal vectors $\{\mathbf{e}_i\}$,

$$\text{Proj.}\{\mathbf{y}_{N+1}|\mathcal{L}_N\} = \sum_{j=0}^{N} \langle \mathbf{y}_{N+1}, \mathbf{e}_j \rangle \|\mathbf{e}_j\|^{-2} \mathbf{e}_j.$$

This then leads to the recursive formula

$$\mathbf{e}_{N+1} = \mathbf{y}_{N+1} - \sum_{j=0}^{N} \langle \mathbf{y}_{N+1}, \mathbf{e}_j \rangle \|\mathbf{e}_j\|^{-2} \mathbf{e}_j, \tag{4.2.3}$$

which can be begun with $\mathbf{e}_0 = \mathbf{y}_0$. This is the *Gram-Schmidt orthogonalization procedure* (see also App. A, where this procedure is described for vectors in Euclidean space).

When the $\{\mathbf{y}_i\}$ are random variables, a suggestive terminology can be associated with the orthogonal variables $\{\mathbf{e}_i\}$. Thus recall that in the stochastic case,

$$\text{Proj.}\{\mathbf{y}_{N+1}|\mathcal{L}_N\} = \text{the l.l.m.s. estimator of } \mathbf{y}_{N+1} \text{ given } \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$$

$$\triangleq \hat{\mathbf{y}}_{N+1}, \quad \text{say.}$$

This is the part of the random variable $\mathbf{y}_{N+1}$ that is determined by knowledge of the previous random variables $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$. The remainder is the random variable

$$\mathbf{e}_{N+1} \triangleq \mathbf{y}_{N+1} - \hat{\mathbf{y}}_{N+1}, \tag{4.2.4}$$

which we can regard as the "new information" or the "innovation" in $\mathbf{y}_{N+1}$ given $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$. Therefore we shall call

$$\{\mathbf{e}_i\} = \text{the innovations process associated with } \{\mathbf{y}_i\}.$$

As befits the name, each vector $\mathbf{e}_i$ brings *new* information, since $\mathbf{e}_i$ is uncorrelated with all other vectors $\{\mathbf{e}_j\}_{j \neq i}$; in other words, the innovations process is a white-noise process. However, the white-noise property by itself is not enough to characterize the innovations. It is important that there be a causal relationship between the indexed collections $\{\mathbf{y}_i\}$ and $\{\mathbf{e}_i\}$: for every $i \geq 0$,

$$\mathbf{e}_i \in \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}, \tag{4.2.5}$$

and

$$\mathbf{y}_i \in \mathcal{L}\{\mathbf{e}_0, \ldots, \mathbf{e}_i\}. \tag{4.2.6}$$

In other words, the processes $\{\mathbf{y}_i\}$ and $\{\mathbf{e}_i\}$ are related by a causal and causally invertible linear transformation. This causality restriction makes the white-noise process $\{\mathbf{e}_i\}$ unique (apart from scaling); if we relax the causality requirement, there are many other white-noise processes that can be obtained by linear operations on the process $\{\mathbf{y}_i\}$. This fact is most easily seen from an algebraic argument.

4.2.2 An Algebraic Approach

We return to the remark at the beginning of Sec. 4.2.1 that the fundamental linear equations $K_o R_y = R_{xy}$ would be easy to solve if R_y were a diagonal matrix, or equivalently if the associated random variables $\{\mathbf{y}_i\}$ were uncorrelated with each other. Since this will rarely be true, we can seek to achieve it by some linear operations. What we need to find is a square matrix A such that the variables

$$\boldsymbol{\epsilon} \triangleq A\mathbf{y}, \quad \boldsymbol{\epsilon} = \text{col}\{\boldsymbol{\epsilon}_0, \ldots, \boldsymbol{\epsilon}_N\}, \tag{4.2.7}$$

where $\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$, are uncorrelated with each other, *i.e.*, such that

$$R_\epsilon = \langle \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \rangle = \langle A\mathbf{y}, A\mathbf{y} \rangle = AR_yA^* = \quad \text{a diagonal matrix.} \tag{4.2.8}$$

Of course, we should require that A be nonsingular so that no *information* is lost in the transformation from the $\{\mathbf{y}_i\}$ to the $\{\boldsymbol{\epsilon}_i\}$. Thus to find such an A we need to factor R_y as

$$R_y = A^{-1} R_\epsilon A^{-*}, \quad R_\epsilon \quad \text{diagonal.} \tag{4.2.9}$$

Such factorizations are highly nonunique because there are N^2 free parameters in A and only $N(N+1)/2$ constraints imposed by (4.2.8). One choice is to make the eigenvector-eigenvalue decomposition of the Hermitian matrix R_y. In this case the matrix A will be unitary. But note that unless the transformation matrix is lower triangular, each of the resulting variables $\{\boldsymbol{\epsilon}_i\}$ will in general be linear combinations of *all* the original variables $\{\mathbf{y}_i\}$. To have a causal relationship as in (4.2.5)–(4.2.6), the

transformation matrix A must be lower triangular. That is, we must seek factorizations of the form

$$R_y = LDL^*, \tag{4.2.10}$$

where D is diagonal, and L is lower triangular. So again we encounter the triangular factorization that was key to the solution of the causal filtering problem in Sec. 4.1.3. Here we note that as mentioned in Sec. 4.1.3 (and proven in App. A), if we require that L be lower triangular with unit diagonal, and that all leading minors of R_y be nonzero, then such factorizations always exist and are also unique.

We shall denote the white-noise process obtained from the factorization (4.2.10), as

$$\mathbf{e} = L^{-1}\mathbf{y}\ ,\ \ R_e = L^{-1}R_yL^{-*} = D. \tag{4.2.11}$$

The choice of the same symbol, $\mathbf{e}$, as for the innovations is deliberate because the variables $L^{-1}\mathbf{y}$ coincide with those obtained via the Gram-Schmidt (GS) procedure. One way to see this fact is to note that the GS procedure also yields a triangular matrix factorization of R_y, and as just stated above, such factorizations must be unique. More specifically, by rearranging the earlier formula (4.2.3) for $\mathbf{e}_i$, we get

$$\mathbf{y}_i = \mathbf{e}_i + \sum_{j=0}^{i-1} \langle \mathbf{y}_i, \mathbf{e}_j \rangle \|\mathbf{e}_j\|^{-2}\mathbf{e}_j, \tag{4.2.12}$$

which of course could also have been written down directly as the projection of $\mathbf{y}_i$ on $\mathcal{L}_i = \mathcal{L}\{\mathbf{e}_0, \ldots, \mathbf{e}_i\}$. Collecting these expressions for $i = 0, 1, \ldots, N$ gives us

$$\mathbf{y} = L_1\mathbf{e}, \quad \text{say} \tag{4.2.13}$$

where $L_1 =$

$$\begin{bmatrix} I & & & & \\ \langle \mathbf{y}_1, \mathbf{e}_0\rangle \|\mathbf{e}_0\|^{-2} & I & & & \\ \langle \mathbf{y}_2, \mathbf{e}_0\rangle \|\mathbf{e}_0\|^{-2} & \langle \mathbf{y}_2, \mathbf{e}_1\rangle \|\mathbf{e}_1\|^{-2} & I & & \\ \vdots & \vdots & \vdots & \ddots & \\ \langle \mathbf{y}_N, \mathbf{e}_0\rangle \|\mathbf{e}_0\|^{-2} & \langle \mathbf{y}_N, \mathbf{e}_1\rangle \|\mathbf{e}_1\|^{-2} & \langle \mathbf{y}_N, \mathbf{e}_2\rangle \|\mathbf{e}_2\|^{-2} & \cdots & I \end{bmatrix}. \tag{4.2.14}$$

Note that an expression for L_1^{-1} can be deduced from (4.2.14), but it will be more complicated, at least in algebraic form. It follows from (4.2.13) that

$$R_y = L_1D_1L_1^*, \qquad D_1 = \text{diag}\{\ \|\mathbf{e}_i\|^2\ \}. \tag{4.2.15}$$

But, by uniqueness, the factorizations (4.2.10) and (4.2.15) must coincide, so that we must have $L_1 = L$ and $D_1 = D$.

Therefore, the processes $L^{-1}\mathbf{y}$ and $L_1^{-1}\mathbf{y}$ coincide and give the unique innovations process. Of course, for the white-noise innovations process, we mean unique up to scaling of the variables. One specific choice is to normalize the process to have unit variance:

$$\bar{\mathbf{e}} = D^{-1/2}\mathbf{e}, \qquad R_{\bar{e}} = I. \tag{4.2.16}$$

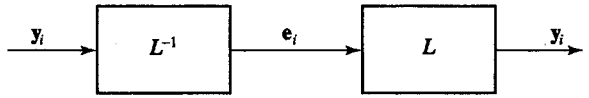


Figure 4.1 Causal and causally invertible transformations relating $\{\mathbf{y}_i\}$ and $\{\mathbf{e}_i\}$.

Correspondingly, we can write

$$R_y = \bar{L}\bar{L}^*, \qquad \bar{L} = LD^{1/2}, \tag{4.2.17}$$

known as the Cholesky decomposition of R_y.

We can summarize the above discussion in block diagram form (see Fig. 4.1). The fact that L and L^{-1} are both triangular reflects the causal equivalence between the processes $\{\mathbf{y}_i\}$ and $\{\mathbf{e}_i\}$. We should remark that L is often called the canonical modeling filter, and L^{-1} the canonical whitening filter. The representation $\mathbf{y} = L\mathbf{e}$ is often called the *innovations representation* of the process $\{\mathbf{y}_i\}$. We shall encounter these names in much less obvious contexts in Ch. 6 (Sec. 6.4) and in Ch. 9.

Remark. The alert reader will have noticed that the matrix in (4.2.14) is written as a *block* matrix, which would be the case when the random variables $\{\mathbf{y}_i\}$ are vector-valued. However, our discussion of triangular factorization was phrased as if the entries of R_y were real or complex scalars. For simplicity, on a first reading, one could assume that all the random variables are scalar-valued. Then a second reading could assume vector-valued random variables, in which case the matrices $\{L, D\}$ in the triangular factorization (4.2.10) would have to be interpreted as being block triangular and block diagonal matrices. These extended interpretations will not often cause any problems. They will really only need to be made explicit occasionally — such are the blessings of matrix notation. ◆

4.2.3 The Modified Gram-Schmidt Procedure

While the innovations process $\{\mathbf{e}_i\}$, or equivalently the triangular matrix L, is unique, this does not mean that there is only one way of constructing them. Here we describe one interesting alternative — the so-called modified Gram-Schmidt (MGS) procedure:

(a) Set $\mathbf{e}_0 = \mathbf{y}_0$.

(b) Form $\tilde{\mathbf{y}}_{i|0} = \mathbf{y}_i - \langle \mathbf{y}_i, \mathbf{e}_0\rangle \|\mathbf{e}_0\|^{-2}\mathbf{e}_0$, and then set $\mathbf{e}_1 = \tilde{\mathbf{y}}_{1|0}$.

(c) Form $\tilde{\mathbf{y}}_{i|1} = \tilde{\mathbf{y}}_{i|0} - \langle \tilde{\mathbf{y}}_{i|0}, \mathbf{e}_1\rangle \|\mathbf{e}_1\|^{-2}\mathbf{e}_1$, and then set $\mathbf{e}_2 = \tilde{\mathbf{y}}_{2|1}$,

and so on. The partial residuals $\{\mathbf{y}_i, \tilde{\mathbf{y}}_{i|0}, \tilde{\mathbf{y}}_{i|1}, \ldots\}$ can be rearranged in a triangular array, the diagonal entries of which are the innovations $\{\mathbf{e}_i\}$:

$$\begin{array}{ccccc} \mathbf{y}_0 & & & & \\ \mathbf{y}_1 & \tilde{\mathbf{y}}_{1|0} & & & \\ \mathbf{y}_2 & \tilde{\mathbf{y}}_{2|0} & \tilde{\mathbf{y}}_{2|1} & & \\ \vdots & \vdots & \vdots & \ddots & \\ \mathbf{y}_N & \tilde{\mathbf{y}}_{N|0} & \tilde{\mathbf{y}}_{N|1} & \cdots & \tilde{\mathbf{y}}_{N|N-1}. \end{array}$$

Now it can be shown that the i-th column of L $(i = 0, 1, \ldots, N)$ is obtained by taking the inner product of the i-th column in the above table with $\|\mathbf{e}_i\|^{-2}\mathbf{e}_i$ (the normalized top entry in that column). Indeed, in Prob. 4.6 we show that if R_y denotes the Gramian matrix of the first column of the above array, and if $R_{y,1}$ denotes the Gramian matrix of the second column of the same array, then $R_{y,1}$ is the Schur complement of R_y with respect to its top leftmost entry. Moreover, in App. A we show that the first columns of the successive Schur complements of a matrix, when properly normalized, are the columns of its Cholesky factor. Now the first column of $R_{y,1}$ is easily seen to be the inner product of the second column of the above array with its top entry, which is equal to $\mathbf{e}_1$. Normalizing this inner product by $\|\mathbf{e}_1\|^{-2}$ yields the second column of L. The argument applies to the other columns of L as well.

4.2.4 Estimation Given the Innovations Process

We recall that the reason for seeking to determine the innovations is that we can then replace the problem of estimation given the process $\{\mathbf{y}_i, i \le k\}$, with the simpler one of estimation given the orthogonal innovations process $\{\mathbf{e}_i, i \le k\}$. Thus

$$\hat{\mathbf{x}}_{|N} \triangleq \text{ the l.l.m.s. estimator of } \mathbf{x} \text{ given } \{\mathbf{y}_0, \ldots, \mathbf{y}_N\},$$

can also be expressed as

$$\hat{\mathbf{x}}_{|N} = \text{ the l.l.m.s. estimator of } \mathbf{x} \text{ given } \{\mathbf{e}_0, \ldots, \mathbf{e}_N\},$$

which, due to the orthogonality of the $\{\mathbf{e}_j\}$, is given by

$$\hat{\mathbf{x}}_{|N} = \sum_{j=0}^{N} \langle \mathbf{x}, \mathbf{e}_j \rangle \, \|\mathbf{e}_j\|^{-2}\mathbf{e}_j. \tag{4.2.18}$$

Moreover, if we now have an additional observation $\mathbf{y}_{N+1}$, then the estimator $\hat{\mathbf{x}}_{|N}$ can readily be updated by using the innovation $\mathbf{e}_{N+1}$,

$$\begin{aligned} \hat{\mathbf{x}}_{|N+1} &= \hat{\mathbf{x}}_{|N} + (\text{l.l.m.s.e. of } \mathbf{x} \text{ given } \mathbf{e}_{N+1}), \\ &= \hat{\mathbf{x}}_{|N} + \langle \mathbf{x}, \mathbf{e}_{N+1} \rangle \|\mathbf{e}_{N+1}\|^{-2}\mathbf{e}_{N+1}, \quad \hat{\mathbf{x}}_{|-1} = 0, \end{aligned} \tag{4.2.19}$$

where

$$\mathbf{e}_{N+1} = \mathbf{y}_{N+1} - \hat{\mathbf{y}}_{N+1|N} = \mathbf{y}_{N+1} - \sum_{j=0}^{N} \langle \mathbf{y}_{N+1}, \mathbf{e}_j \rangle \|\mathbf{e}_j\|^{-2}\mathbf{e}_j, \quad \mathbf{e}_0 = \mathbf{y}_0. \tag{4.2.20}$$

The simple formulas (4.2.18), (4.2.19), and (4.2.20) are the key to many results in linear least-squares estimation theory.

A first illustration of this fact is obtained by applying the innovations method to the filtering problem of Sec. 4.1.2, which we solved in Sec. 4.1.3 via the ingenious, but unmotivated, Wiener-Hopf technique.

4.2.5 The Filtering Problem via the Innovations Approach

As described in Sec. 4.2.4, in the innovations approach, the estimation problem is broken into two parts: (i) finding the innovations $\{\mathbf{e}_i\}$ from the observations $\{\mathbf{y}_i\}$, and (ii) finding the estimators $\{\hat{\mathbf{s}}_{i|i}\}$ from the innovations. The second step is easy because the innovations are a white noise process. So, once the innovations are found via $\mathbf{e} = L^{-1}\mathbf{y}$, we can seek to compute

$$\hat{\mathbf{s}}_{i|i} = \sum_{j=0}^{i} g_{f,ij}\mathbf{e}_j, \tag{4.2.21}$$

for some coefficients $\{g_{f,ij}\}$. The orthogonality condition gives

$$\langle \mathbf{s}_i - \hat{\mathbf{s}}_{i|i}, \mathbf{e}_l \rangle = 0 \quad \text{for} \quad 0 \le l \le i, \tag{4.2.22}$$

which leads to the now trivial Wiener-Hopf equation

$$R_{se}(i, l) = \sum_{j=0}^{i} g_{f,ij}\|\mathbf{e}_j\|^2\delta_{jl} = g_{f,il}\|\mathbf{e}_l\|^2, \quad \text{for} \quad 0 \le l \le i. \tag{4.2.23}$$

In other words, the desired coefficients $\{g_{f,il}\}$ are given by

$$g_{f,il} = \begin{cases} R_{se}(i, l)\|\mathbf{e}_l\|^{-2} & \text{for } 0 \le l \le i, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2.24}$$

However, we are given the $\{R_{sy}(i, l)\}$ and not the $\{R_{se}(i, l)\}$. The latter can be readily computed by using the linear relationship $\mathbf{e} = L^{-1}\mathbf{y}$. Thus, in matrix notation, we have

$$R_{se} = [R_{se}(i, l)]_{il}, \qquad R_{se} = \langle \mathbf{s}, \mathbf{e} \rangle = \langle \mathbf{s}, \mathbf{y} \rangle \, L^{-*} = R_{sy}L^{-*}.$$

Therefore, in view of (4.2.24), the lower triangular matrix $G_f = [g_{f,il}]_{0 \le l \le i}$ is given by

$$G_f = \{R_{sy}L^{-*}R_e^{-1}\}_{\text{lower}}. \tag{4.2.25}$$

Finally, the mapping from the original observations $\{\mathbf{y}_i\}$ to the estimators $\{\hat{\mathbf{s}}_{i|i}\}$ is given by

$$\hat{\mathbf{s}}_f = G_f\mathbf{e} = G_fL^{-1}\mathbf{y} = [R_{sy}L^{-*}R_e^{-1}]_{\text{lower}}L^{-1}\mathbf{y} = K_f\mathbf{y},$$

which agrees with our earlier expression (4.1.20)!

Note that once we have understood the significance of the matrix L in the triangular factorization of R_y, we no longer need the clever Wiener-Hopf trick of Eqs. (4.1.14) and (4.1.16).

4.2.6 Computational Issues

The various methods for determining the innovations all have special features of interest, but for the moment the point we wish to make is that they all take essentially the same order of *elementary computations*, $O(N^3)$, to yield the factors A (and L) and thereby the innovations. This is in fact the same order of complexity as for inverting R_y and thus directly solving the estimation problem without first finding the innovations! So what have we gained?

The point is, as often mentioned already, that in applications we often have special structures, *e.g.*, *stationarity* of the process or the availability of *state-space* or *difference equation* models for it, that enable *fast* ways of obtaining the innovations, or equivalently, fast ways of *factoring* the associated covariance matrix. Thus, for example, it is known that the stationarity structure can be exploited to find the innovations and thereby to factor (and invert) R_y with $O(N^2)$ operations. In fact, a now-classical illustration of this fact is given in Prob. 4.11 on the Levinson-Durbin algorithm for the prediction of stationary stochastic processes. These results have also been extended to nonstationary processes with *displacement* structure by using the so-called generalized Schur algorithm; if the so-called displacement rank of the process is r, it can be shown that it takes only $O(rN^2)$ operations to find the innovations, and hence to factor and invert R_y (see, *e.g.*, App. F and the survey articles Kailath (1999) and Kailath and Sayed (1995)). Certain so-called doubling methods allow even further reductions to $O(Nr^2 \log N \log N)$ computations (see, *e.g.*, Chun and Kailath (1991)).

In this book our focus shall be on state-space structure. If we have an n-dimensional state-space model for the process $\{\mathbf{y}_i\}$, then it turns out that the innovations can be found with $O(Nn^3)$ operations (see Ch. 9), which can be very much less than $O(N^3)$ if $n \ll N$. Processes with *constant parameter* state-space models do have displacement structure, and displacement rank $r \leq n$; for such processes we shall only need $O(Nn^2r)$ operations (see Ch. 11). We shall slowly build up to these results. We shall also illustrate many of the points of this section by studying a simple example in Sec. 4.4.

4.3 INNOVATIONS APPROACH TO DETERMINISTIC LEAST-SQUARES PROBLEMS

Most of our discussions so far have been in terms of random variables regarded as vectors. However as noted several times already, the discussion can often be applied to vectors in any linear space. So, as an example, here we shall consider vectors in N-dimensional Euclidean space as arise in the problems studied in Ch. 2:

$$\min_x \ \|y - Hx\|^2, \quad H \quad \text{full rank},$$

where $H \in \mathbb{C}^{N\times n}$, the space of possibly complex-valued $N \times n$ matrices, $n \leq N$. The solution was seen to be (Sec. 2.2)

$$\hat{x} = (H^*H)^{-1}H^*y, \tag{4.3.1}$$

which was obtained by projecting y onto the space spanned by the columns $\{\underline{h}_i\}$ of H,

$$H = \begin{bmatrix} \underline{h}_0 & \underline{h}_1 & \cdots & \underline{h}_{n-1} \end{bmatrix}.$$

Now in the innovations approach, we should first replace this nonorthogonal set of vectors by an equivalent orthogonal set. This can be done by the Gram-Schmidt procedure (*cf.* App. A), which in matrix form can be expressed as

$$\begin{bmatrix} \underline{e}_0 & \underline{e}_1 & \cdots & \underline{e}_{n-1} \end{bmatrix} = \begin{bmatrix} \underline{h}_0 & \underline{h}_1 & \cdots & \underline{h}_{n-1} \end{bmatrix} \begin{bmatrix} \times & \times & \cdots & \times \\ & \times & & \times \\ & & \ddots & \vdots \\ & & & \times \end{bmatrix},$$

or $E = HU^{-1}$, say, where U is a unit-diagonal $n \times n$ upper triangular matrix and the columns of E are orthogonal to each other so that E^*E is a diagonal matrix (and trivial to invert). Once E and U are found, the solution $\hat{x}$ can be determined by using the condition

$$\underline{e}_i^*(y - H\hat{x}) = 0, \quad i = 0, 1, \ldots, n-1,$$

or

$$E^*y = E^*H\hat{x} = (E^*E)U\hat{x},$$

and finally

$$U\hat{x} = (E^*E)^{-1}E^*y. \tag{4.3.2}$$

By now, the alert reader may have noticed that what we have done here is essentially to rederive in a very direct way the array method of Sec. 2.5 for solving the deterministic least-squares problem. The only difference is that there we used a normalized form of the decomposition $E = HU^{-1}$, or $H = EU$. If we normalize the $\{\underline{e}_i\}$ to have unit length, then the decomposition will take the form $H = \hat{Q}\hat{R}$, used in Sec. 2.5 (and in App. A). More explicitly, we can write

$$H = EU = \underbrace{E(E^*E)^{-*/2}}_{\hat{Q}} \underbrace{(E^*E)^{*/2}U}_{\hat{R}}.$$

Then (4.3.2) can be written as

$$(E^*E)^{*/2}U\hat{x} = \hat{R}\hat{x} = (E^*E)^{-1/2}E^*y = \hat{Q}^*y,$$

which is exactly the solution (2.5.5) given by the QR method.

Finally, note that replacing E by HU^{-1} in (4.3.2) gives us back the usual formula (4.3.1). At least in retrospect, one may ask if we are given H, then why not work directly with H rather than form H^*H and then factor it? This is exactly what we do in the innovations approach! The potential numerical benefits of this approach when H is ill-conditioned were explained in Sec. 2.5. We should also emphasize, as stated there, that the decomposition $H = \hat{Q}\hat{R}$ is not usually carried out via the numerically unreliable GS method — we use instead a sequence of elementary unitary operations, as described in App. B.

4.4 THE EXPONENTIALLY CORRELATED PROCESS

In this section we shall explore some of the procedures suggested in Sec. 4.2 by studying a simple example. Thus suppose $\{\mathbf{y}_i, 0 \le i \le N\}$ is a segment of a scalar zero-mean wide-sense stationary process with

$$\langle \mathbf{y}_i, \mathbf{y}_j \rangle = E\mathbf{y}_i\mathbf{y}_j^* = a^{|i-j|}, \quad 0 \le i, j \le N, \tag{4.4.1}$$

for some a such that $0 < a < 1$. [Note that any such a can be written as $e^{-\rho}$, explaining the name *exponentially correlated.*] Such processes have been used as models for several physical phenomena and are known by various names in different contexts (*e.g.*, RC noise, Ornstein-Uhlenbeck process). We shall find that their special structure allows the innovations of $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ to be found with $O(N)$ elementary computations, compared to the $O(N^3)$ required for a general process.

4.4.1 Triangular Factorization of R_y

We shall first consider two standard algebraic methods for factoring the covariance matrix as

$$R_y = \bar{L}\bar{L}^*, \quad \text{(Cholesky factorization)}$$

or as

$$R_y = LDL^*, \quad \text{(LDU or lower-diagonal-upper factorization),}$$

where $\bar{L}$ is lower triangular, L is lower triangular with unit diagonal, and D is diagonal.

For simplicity, we shall assume that $N = 2$, but the extension for general N will be apparent. When $N = 2$, R_y is the 3×3 matrix

$$R_y = \left\langle \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \right\rangle = \begin{bmatrix} 1 & a & a^2 \\ a & 1 & a \\ a^2 & a & 1 \end{bmatrix}, \tag{4.4.2}$$

which is a Toeplitz matrix (constant along diagonals).

Cholesky Method. We set

$$R_y = \bar{L}\bar{L}^*, \qquad \bar{L} = \begin{bmatrix} \bar{l}_{00} & 0 & 0 \\ \bar{l}_{10} & \bar{l}_{11} & 0 \\ \bar{l}_{20} & \bar{l}_{21} & \bar{l}_{22} \end{bmatrix}, \tag{4.4.3}$$

and compare coefficients to find

$$\bar{l}_{00} = 1, \quad \bar{l}_{10}^* = a, \quad \bar{l}_{11} = \sqrt{1-a^2},$$
$$\bar{l}_{20}^* = a^2, \quad \bar{l}_{21} = a\sqrt{1-a^2}, \quad \bar{l}_{22} = \sqrt{1-a^2}.$$

Therefore,

$$\bar{L} = \begin{bmatrix} 1 & 0 & 0 \\ a & \sqrt{1-a^2} & 0 \\ a^2 & a\sqrt{1-a^2} & \sqrt{1-a^2} \end{bmatrix}, \tag{4.4.4}$$

from which we can readily deduce that

$$L = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ a^2 & a & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-a^2 & 0 \\ 0 & 0 & 1-a^2 \end{bmatrix}. \tag{4.4.5}$$

The above method for finding $\bar{L}$ is often called the Cholesky method.

Symmetric Gaussian Elimination. Here we find L and D by the well-known Gaussian elimination procedure. For this purpose, we carry out elementary operations as shown. First,

$$\begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ -a^2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & a & a^2 \\ a & 1 & a \\ a^2 & a & 1 \end{bmatrix} = \begin{bmatrix} 1 & a & a^2 \\ 0 & 1-a^2 & a-a^3 \\ 0 & a-a^3 & 1-a^4 \end{bmatrix}, \tag{4.4.6}$$

and then

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -a & 1 \end{bmatrix} \begin{bmatrix} 1 & a & a^2 \\ 0 & 1-a^2 & a-a^3 \\ 0 & a-a^3 & 1-a^4 \end{bmatrix} = \begin{bmatrix} 1 & a & a^2 \\ 0 & 1-a^2 & a-a^3 \\ 0 & 0 & 1-a^2 \end{bmatrix} \tag{4.4.7}$$
$$\triangleq DU, \quad \text{say}$$

where

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-a^2 & 0 \\ 0 & 0 & 1-a^2 \end{bmatrix}, \qquad U = \begin{bmatrix} 1 & a & a^2 \\ 0 & 1 & a \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.4.8}$$

Writing the product of the two elementary matrices as L^{-1} gives

$$L^{-1}R_y = DU \text{ or } R_y = LDU, \tag{4.4.9}$$

where, by the fact that R_y is Hermitian, we can identify, as found before, that

$$L = U^* = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ a^2 & a & 1 \end{bmatrix}. \tag{4.4.10}$$

Note that this Gaussian elimination procedure effectively determines L by columns (rather than by rows as in the first method).

We may also remark that this procedure is equivalent to the so-called Schur reduction procedure described in App. A, which effectively operates on R_y from both the left and the right. Therefore instead of (4.4.6), we can note by inspection that

$$\begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ -a^2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & a & a^2 \\ a & 1 & a \\ a^2 & a & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ -a^2 & 0 & 1 \end{bmatrix}^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-a^2 & a-a^3 \\ 0 & a-a^3 & 1-a^4 \end{bmatrix},$$

or, equivalently and preferably, (as written in App. A),

$$\begin{bmatrix} 1 & a & a^2 \\ a & 1 & a \\ a^2 & a & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ a \\ a^2 \end{bmatrix} \begin{bmatrix} 1 & a & a^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1-a^2 & a-a^3 \\ 0 & a-a^3 & 1-a^4 \end{bmatrix}.$$

The 2×2 nonzero submatrix matrix on the right-hand side of the above equation is the *Schur complement* of the top leftmost entry of R_y. Now we can proceed similarly on the Schur complement by subtracting the outer product of its first row and first column, normalized by the inverse of the top leftmost entry. And so on.

As shown in detail in App. A, this procedure yields the (column-wise) LDL* factorization of R_y.

4.4.2 Finding L^{-1} and the Innovations

At this point we should recall that our major goal was actually not so much to find the triangular factors L and $\bar{L}$, but, as pointed out in Sec. 4.2, to find the innovations process $\{\mathbf{e}_i\}$. The triangular factorization of R_y presents one method of finding the innovations. Indeed from (4.2.11) we have

$$\mathbf{e} = L^{-1}\mathbf{y} \quad \text{or} \quad \bar{\mathbf{e}} = \bar{L}^{-1}\mathbf{y}.$$

Now neither $\bar{L}$ in (4.4.4) nor L in (4.4.10) is hard to invert, but it is worth noting that once we have performed Gaussian elimination as in (4.4.6) and (4.4.7), we can find L^{-1} quite easily as the product of the elementary matrices, which yields

$$L^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & -a & 1 \end{bmatrix}. \tag{4.4.11}$$

Therefore,

$$\mathbf{e} = L^{-1}\mathbf{y} = L^{-1} \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \tag{4.4.12}$$

so that

$$\mathbf{e}_0 = \mathbf{y}_0, \quad \mathbf{e}_1 = \mathbf{y}_1 - a\mathbf{y}_0, \quad \mathbf{e}_2 = \mathbf{y}_2 - a\mathbf{y}_1. \tag{4.4.13}$$

By now, the formulas for general N should be pretty obvious. In fact, we have

$$\mathbf{e}_k = \mathbf{y}_k - a\mathbf{y}_{k-1}, \quad \text{with } \|\mathbf{e}_k\|^2 = (1-a^2), \tag{4.4.14}$$

which can be directly verified, if one wishes, by checking that $\mathbf{e}_k$ satisfies the orthogonality property,

$$\begin{aligned} \text{for } j \le k-1, \quad \langle \mathbf{e}_k, \mathbf{y}_j \rangle &= \langle \mathbf{y}_k, \mathbf{y}_j \rangle - a\langle \mathbf{y}_{k-1}, \mathbf{y}_j \rangle \\ &= a^{k-j} - a \cdot a^{k-1-j} = 0. \end{aligned}$$

We see that it takes only two elementary computations to find each innovation, so that the innovations of $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ can be found with $O(N)$ computations, as claimed earlier.

4.4.3 Innovations via the Gram-Schmidt Procedures

Instead of the algebraic approach, let us consider the geometric Gram-Schmidt procedures. Recall from Sec. 4.2.1 that in the (classical) Gram-Schmidt procedure, we successively compute

$$\mathbf{e}_0 = \mathbf{y}_0, \quad \text{with } \|\mathbf{e}_0\|^2 = 1,$$

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{y}_1 - \langle \mathbf{y}_1, \mathbf{e}_0 \rangle \|\mathbf{e}_0\|^{-2}\mathbf{e}_0, \\ &= \mathbf{y}_1 - a\mathbf{y}_0, \quad \text{with } \|\mathbf{e}_1\|^2 = (1-a^2), \end{aligned}$$

$$\begin{aligned} \mathbf{e}_2 &= \mathbf{y}_2 - \sum_{j=0}^{1} \langle \mathbf{y}_2, \mathbf{e}_j \rangle \|\mathbf{e}_j\|^{-2}\mathbf{e}_j, \\ &= \mathbf{y}_2 - a^2\mathbf{y}_0 - (a-a^3)(1-a^2)^{-1}(\mathbf{y}_1 - a\mathbf{y}_0), \\ &= \mathbf{y}_2 - a\mathbf{y}_1, \quad \text{with } \|\mathbf{e}_2\|^2 = (1-a^2), \end{aligned}$$

exactly as in (but more simply than via) the algebraic route. Organizing these results in matrix form immediately gives the algebraic result (4.4.11)–(4.4.12):

$$\begin{bmatrix} \mathbf{e}_0 \\ \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & -a & 1 \end{bmatrix} \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = L^{-1}\mathbf{y}.$$

Moreover, note that now the triangular factor, L, can be obtained, if desired, by simply *rearranging* the above formulas as (see (4.2.12)–(4.2.14))

$$\begin{aligned} \mathbf{y}_0 &= \mathbf{e}_0, \\ \mathbf{y}_1 &= a\mathbf{y}_0 + \mathbf{e}_1 = a\mathbf{e}_0 + \mathbf{e}_1, \\ \mathbf{y}_2 &= a\mathbf{y}_1 + \mathbf{e}_2 = a^2\mathbf{e}_0 + a\mathbf{e}_1 + \mathbf{e}_2, \end{aligned}$$

which leads to the same L obtained earlier after some algebra in (4.4.5).

In fact, for determining L, the modified Gram-Schmidt (MGS) procedure is more direct. Recall from Sec. 4.2.3 that in the MGS procedure we must form the triangular array

$$\begin{array}{lll} \mathbf{y}_0 & & \\ \mathbf{y}_1 & \tilde{\mathbf{y}}_{1|0} & \\ \mathbf{y}_2 & \tilde{\mathbf{y}}_{2|0} & \tilde{\mathbf{y}}_{2|1} \end{array} \tag{4.4.15}$$

and that the diagonal elements of the above array will be the innovations $\{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2\}$. Now using the expressions given in Sec. 4.2.3, the second column of the above array is given by

$$\tilde{\mathbf{y}}_{1|0} = \mathbf{y}_1 - \langle \mathbf{y}_1, \mathbf{y}_0 \rangle \|\mathbf{y}_0\|^{-2} \mathbf{y}_0 = \mathbf{y}_1 - a\mathbf{y}_0 \triangleq \mathbf{e}_1,$$

$$\tilde{\mathbf{y}}_{2|0} = \mathbf{y}_2 - \langle \mathbf{y}_2, \mathbf{y}_0 \rangle \|\mathbf{y}_0\|^{-2} \mathbf{y}_0 = \mathbf{y}_2 - a^2\mathbf{y}_0.$$

Moreover, the third column of the array is given by

$$\begin{aligned} \tilde{\mathbf{y}}_{2|1} &= \tilde{\mathbf{y}}_{2|0} - \langle \tilde{\mathbf{y}}_{2|0}, \mathbf{e}_1 \rangle \|\mathbf{e}_1\|^{-2} \mathbf{e}_1, \\ &= \mathbf{y}_2 - a^2\mathbf{y}_0 - \langle \mathbf{y}_2 - a^2\mathbf{y}_0, \mathbf{y}_1 - a\mathbf{y}_0 \rangle \|\mathbf{y}_1 - a\mathbf{y}_0\|^{-2} (\mathbf{y}_1 - a\mathbf{y}_0), \\ &= \mathbf{y}_2 - a^2\mathbf{y}_0 - (a - a^3 - a^3 + a^3)(1 - a^2)^{-1} (\mathbf{y}_1 - a\mathbf{y}_0), \\ &= \mathbf{y}_2 - a\mathbf{y}_1. \end{aligned}$$

In other words, the triangular array (4.4.15) is just (note that the innovations appear along the diagonal)

$$\begin{array}{lll} \mathbf{y}_0 & & \\ \mathbf{y}_1 & \mathbf{y}_1 - a\mathbf{y}_0 & \\ \mathbf{y}_2 & \mathbf{y}_2 - a^2\mathbf{y}_0 & \mathbf{y}_2 - a\mathbf{y}_1. \end{array}$$

As mentioned before, to obtain the i-th column of L we just form the inner product of the i-th column in the above array with $\|\mathbf{e}_i\|^{-2}\mathbf{e}_i$. The reader should verify that this will indeed lead to

$$L = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ a^2 & a & 1 \end{bmatrix}.$$

We should note the striking simplicity of the expressions for L, L^{-1}, and for the innovations in this case, and speculate whether there is a deeper reason for this simplicity and whether there is a more direct route to finding the innovations. We shall take up this issue in Ch. 5 where we will show that the direct route is to consider a *model* for the process $\{\mathbf{y}_i\}$. This then allows one to construct the innovations (and hence the triangular factor) directly from the model, rather than through factorization of the covariance matrix, which can become quite complicated as we go beyond the above simple example.

However, let us close this section by noting that we introduced the triangular factorization of R_y in connection with the filtering problem of Sec. 4.1.2. The smoothing problem of Sec. 4.1.1 was conceptually simpler, requiring only the inversion of R_y. However, from a numerical point of view, often the best way of finding R_y^{-1} is via the triangular factorization, $R_y = LR_eL^*$. That is, we compute R_y^{-1} as $R_y^{-1} = L^{-*}R_e^{-1}L^{-1}$. In the present simple problem, we may note that this yields the simple explicit formula

$$\frac{1}{1-a^2} \begin{bmatrix} 1 & -a & 0 \\ -a & 1+a^2 & -a \\ 0 & -a & 1 \end{bmatrix},$$

whose tri-diagonal structure is evident; apart from a scaling by $1/(1-a^2)$, the off-diagonal entries are all equal to $-a$ while the diagonal entries are $\{1, 1+a^2, 1\}$.

Of course, what we really need in the smoothing problem is not to invert R_y but to solve the linear equations (4.1.7). We would not form R_y^{-1} explicitly, but would find K_s by solving two triangular systems of equations. More specifically, we first solve for the intermediate matrix X in $XR_eL^* = R_{sy}$ and then solve for K_s in $K_sL = X$. Issues such as these, and several related ones, are well discussed in books on numerical linear algebra (*e.g.*, those listed in App. A).

4.5 COMPLEMENTS

This chapter introduced the fundamental concept of the (uncorrelated) innovations process associated with a stochastic process or, equivalently, a given ordered set of (correlated) vectors. The concept is usually applied to random variables but it is important to note that it is useful for deterministic vectors as well, as we illustrated in Sec. 4.3.

Sec. 4.1. Estimation of Stochastic Processes. The first general studies of estimation problems for stochastic processes were made by A. N. Kolmogorov in 1939–1941 and by N. Wiener in 1942; as acknowledged by Kolmogorov , he built on the important contributions made by H. Wold in his remarkable Ph.D. dissertation [Wold(1938)]. Kolmogorov considered the problem of k-step prediction of a general discrete-time stationary process and obtained an elegant formula (*cf.* Sec. 7.7.1) for the associated minimum-mean-square error, clearly identifying also the cases of perfect prediction. He made no effort to explicitly determine the form of the predictor. Such formulas are obviously necessary for applications and in fact, in an effort to make his own contribution to the war effort, Wiener in 1941 was led to formulate the antiaircraft fire control problem as a continuous-time linear prediction problem; the optimum predictor turned out to be the solution of an integral equation very familiar to him — the Wiener-Hopf equation. Eq. 4.1.13 is a finite dimensional version of this equation, and it can be solved by using the same fundamental idea (see Sec. 4.1.3 and also App. 7.A).

Wiener also studied the problem of filtering signals out of noise and noted that similar approaches could be used for other problems in control theory and circuit theory. In fact, though Wiener's solution was not effective for the anti-aircraft problem, his emphasis on the statistical nature of all communication problems and on seeking solutions that met specified optimization criteria greatly influenced the development

of mathematical system theory — see, *e.g.*, Kailath (1997) and the references therein for a review of Wiener's work and notable influence on the development of what might be called mathematical engineering. Chs. 7 and 8 will go into more detail on Wiener's results.

Sec. 4.2. The Innovations Process. The innovations concept goes back to Wold (1938), who picked up a suggestion by Fréchet (1937) that random variables could be regarded as elements of a linear vector space. Wold then noted that it would be convenient to represent a correlated set of random variables by an uncorrelated set, and that the Gram-Schmidt method provided a way of doing this. Kolmogorov (1939,1941a,b) developed the ideas much further for the case of discrete-time stationary processes, as we shall (partly) describe in Ch. 7. The term innovations itself was perhaps first used by Wiener and Masani in the mid-fifties — see also Cramér (1960); it was reintroduced in connection with (continuous-time) nonstationary linear and nonlinear process estimation by Kailath (1968, 1969a, 1972a) — see Sec. 16.4.1.

PROBLEMS

4.1 **(Whitening filters are nonunique)** Given $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}$, form linear combinations $\epsilon_i = \sum_{j=0}^{N} a_{ij}\mathbf{y}_j$, $i = 0, 1, \ldots, N$. Let $A = [a_{ij}]$ and $R_y = [\langle \mathbf{y}_i, \mathbf{y}_j \rangle]$, with R_y nonsingular.

(a) Determine conditions on A so that the $\{\epsilon_i\}$ are mutually uncorrelated and have unit variance.

(b) Consider the modal decomposition $R_y = U\Lambda U^*$, where U is a unitary matrix, whose columns are the eigenvectors of R_y, and where Λ is a diagonal matrix of the eigenvalues of R_y. Show that all such A can be written as $Q\Lambda^{-\frac{1}{2}}U^*$ for some arbitrary unitary matrix Q. ($\Lambda^{\frac{1}{2}}$ is the diagonal matrix whose entries are the positive square roots of the eigenvalues of R_y.)

(c) A can be made unique by imposing additional constraints, *e.g.*, such that it be Hermitian or triangular with positive diagonal entries. Show how to find such a Hermitian A.

4.2 **(A simple moving average process)** Let $\mathbf{y}_k = \mathbf{v}_k + \mathbf{v}_{k-1}$, $k \geq 0$, where $\{\mathbf{v}_j, j \geq -1\}$ is a zero-mean stationary white-noise scalar process with unit variance. Show that

$$\hat{\mathbf{y}}_{k+1|k} = \frac{k+1}{k+2}(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}),$$

and that the error variance is given by

$$\|\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}\|^2 = \frac{k+2}{k+1}.$$

4.3 **(A minimization criterion for causal estimation)** Refer to the discussion in Sec. 4.1.2. Show that K_f also minimizes the cost function

$$E(\mathbf{s} - K\mathbf{y})^* W (\mathbf{s} - K\mathbf{y}),$$

over all lower triangular matrices K, and for any nonnegative-definite matrix W.

4.4 **(A separation principle)** All variables are zero-mean. Consider the linear model $\mathbf{y} = H\mathbf{x} + \mathbf{v}$, where $\mathbf{v}$ and $\mathbf{x}$ are uncorrelated with variances R_v and R_x, respectively. Assume R_x is diagonal and consider a random variable $\mathbf{z}$ that is also uncorrelated with $\mathbf{v}$, but whose correlation with $\mathbf{x}$ is lower triangular, *i.e.*, R_{zx} is lower triangular. Let $\hat{\mathbf{z}}_{|x}$ and $\hat{\mathbf{z}}_{|y}$ denote the *causal* l.l.m.s. estimators of $\mathbf{z}$ given $\mathbf{x}$ and given $\mathbf{y}$, respectively. Let further $\widehat{\hat{\mathbf{z}}_{|x}}$ denote the *causal* l.l.m.s.e. of $\hat{\mathbf{z}}_{|x}$ given $\mathbf{y}$. Verify that $\hat{\mathbf{z}}_{|y}$ and $\widehat{\hat{\mathbf{z}}_{|x}}$ coincide.

4.5 **(Matched filters)** All quantities in this problem are real-valued. Consider scalar measurements $\mathbf{y}(i) = \alpha m(i) + \mathbf{v}(i)$, for $i = 0, 1, \ldots, N$, where $\{\mathbf{v}(i)\}$ is a unit-variance zero-mean white-noise process and the $\{m(i)\}$ is a known deterministic sequence. In order to recover $m(i)$, it is suggested that we form the linear combination $\sum_{i=0}^{N} h(i)\mathbf{y}(i)$ for some $\{h(i)\}$ that are chosen in order to maximize the signal-to-noise ratio (SNR) defined as

$$\text{SNR} \triangleq \frac{\left|\sum_{i=0}^{N} h(i)m(i)\right|^2}{E\left|\sum_{i=1}^{N} h(i)\mathbf{v}(i)\right|^2},$$

is a maximum.

(a) Show that an optimum choice of $h(i)$ is $h(i) = m(i)$. Are there other choices?

(b) Suppose now that $\mathbf{v}(\cdot)$ is colored noise with $E\mathbf{v}(i)\mathbf{v}^T(j) = r_{ij}$. Show that an optimum choice of the $\{h(i)\}$ is now the solution of the equation $hR = m$, where $m^T = \text{col}\{m(0), \ldots, m(N)\}$, $h^T = \text{col}\{h(0), \ldots, h(N)\}$, and $R = [r_{ij}]_{i,j=0}^{N}$.

(c) How would your results change if the quantities were complex-valued?

Remark. The resulting filters $\{h(\cdot)\}$ are known as matched filters. ◆

4.6 **(MGS procedure)** Refer to the description of the modified Gram-Schmidt procedure in Sec. 4.2.3. Let R_y denote the Gramian matrix of $\mathbf{y} = \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}$. Let also $R_{y,1}$ denote the Gramian matrix of $\tilde{\mathbf{y}} = \text{col}\{\tilde{\mathbf{y}}_{1|0}, \tilde{\mathbf{y}}_{2|0}, \ldots, \tilde{\mathbf{y}}_{N|0}\}$. Show that $R_{y,1}$ is the Schur complement of the top leftmost entry of R_y. That is, if we partition R_y as

$$R_y = \begin{bmatrix} \|\mathbf{y}_0\|^2 & a^* \\ a & M \end{bmatrix},$$

then $R_{y,1} = M - a\|\mathbf{y}_0\|^{-2}a^*$.

4.7 **(A useful relation)** Refer to the discussion in Sec. 4.2.2 and recall that $\mathbf{y} = L\mathbf{e}$. Show that

$$\mathbf{y}^* R_y^{-1} \mathbf{y} = \sum_{i=0}^{N} \mathbf{e}_i^* \, \|\mathbf{e}_i\|^{-2} \, \mathbf{e}_i.$$

Remark. Some readers may recognize the left-hand side as a log likelihood function for Gaussian processes. The result is useful, for example, when studying the connections between Kalman filtering and adaptive filtering problems (see, *e.g.*, Sayed and Kailath (1994b) — see also Sec. 10.7). ◆

4.8 **(Correlated signal and noise processes)** Consider the model $\mathbf{y}_i = \mathbf{z}_i + \mathbf{v}_i$, with $\|\mathbf{v}_i\|^2 = R_i\delta_{ij}$, $\langle \mathbf{v}_i, \mathbf{z}_j \rangle = 0$ for $i > j$, and $\langle \mathbf{v}_i, \mathbf{z}_i \rangle = D_i$. All random variables are zero-mean. Let $R_{e,i}$ denote the variance of the innovations of the process $\{\mathbf{y}_i\}$. Let also $\hat{\mathbf{z}}_{i|i}$ denote the l.l.m.s.e. of $\mathbf{z}_i$ given $\{\mathbf{y}_j, 0 \leq j \leq i\}$. Show that we can write

(a) $\hat{\mathbf{z}}_{i|i} = \mathbf{y}_i - (D_i + R_i)R_{e,i}^{-1}\mathbf{e}_i$.

(b) $\|\tilde{\mathbf{z}}_{i|i}\|^2 = R_i - (D_i + R_i)R_{e,i}^{-1}(D_i + R_i)^*$.

4.9 (Order-recursive prediction) Consider a zero-mean discrete-time random process $\{\mathbf{y}_t\}$. Let $\mathcal{U}$ denote the linear subspace spanned by $\{\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-N}\}$, written as

$$\mathcal{U} = \mathcal{L}\{\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-N}\}.$$

Let also $\hat{\mathbf{y}}_{t|\mathcal{U}}$ denote the l.l.m.s. estimator of $\mathbf{y}_t$ given the observations in $\mathcal{U}$. Define

$$\mathbf{e}_{N,t} \triangleq \mathbf{y}_t - \hat{\mathbf{y}}_{t|\mathcal{U}} \triangleq \text{ the } N\text{−th order } \textit{forward residual} \text{ at time } t,$$

$$\mathbf{r}_{N,t-1} \triangleq \mathbf{y}_{t-N-1} - \hat{\mathbf{y}}_{t-N-1|\mathcal{U}} \triangleq \text{ the } N\text{−th order } \textit{backward residual} \text{ at time } t-1.$$

Note that $\mathbf{e}_{t,t} = \mathbf{y}_t - \hat{\mathbf{y}}_{t|\{\mathbf{y}_{t-1},\ldots,\mathbf{y}_0\}}$ is the innovations at time t. Denote

$$\|\mathbf{e}_{N,t}\|^2 = R_{N,t}^e, \quad \|\mathbf{r}_{N,t-1}\|^2 = R_{N,t-1}^r, \quad \Delta_{N,t} = \langle \mathbf{e}_{N,t}, \mathbf{r}_{N,t-1}\rangle = \langle \mathbf{r}_{N,t-1}, \mathbf{e}_{N,t}\rangle^*.$$

(a) Show that we can write recursions, for fixed t and increasing N, as follows:

$$\begin{aligned}
\mathbf{e}_{N+1,t} &= \mathbf{e}_{N,t} - \Delta_{N,t}R_{N,t-1}^{-r}\mathbf{r}_{N,t-1},\\
\mathbf{r}_{N+1,t} &= \mathbf{r}_{N,t-1} - \Delta_{N,t}^*R_{N,t}^{-e}\mathbf{e}_{N,t},\\
R_{N+1,t}^e &= R_{N,t}^e - \Delta_{N,t}R_{N,t-1}^{-r}\Delta_{N,t}^*,\\
R_{N+1,t}^r &= R_{N,t-1}^r - \Delta_{N,t}^*R_{N,t}^{-e}\Delta_{N,t}.
\end{aligned}$$

(b) Define the normalized variables $\bar{\mathbf{e}}_{N,t} = \|\mathbf{e}_{N,t}\|^{-1}\mathbf{e}_{N,t}$, $\bar{\mathbf{r}}_{N,t} = \|\mathbf{r}_{N,t}\|^{-1}\mathbf{r}_{N,t}$ and the so-called *reflection* coefficient $\gamma_{N+1,t} = \langle \bar{\mathbf{e}}_{N,t}, \bar{\mathbf{r}}_{N,t-1}\rangle$. Show that

$$\|\gamma_{N+1,t}\|^2 \leq \|\bar{\mathbf{e}}_{N,t}\|^2\,\|\bar{\mathbf{r}}_{N,t-1}\|^2 = 1.$$

Remark. For reasons that we shall not pursue here, $\gamma_{N+1,t}$ can be interpreted as the partial correlation coefficient of the variables $\{\mathbf{y}_t, \mathbf{y}_{t-N-1}\}$ conditioned on knowledge of $\{\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-N}\}$. ◆

(c) Assume $\|\gamma_{N,t}\| < 1$ for all t, N. Show that we can write

$$\begin{bmatrix}\bar{\mathbf{e}}_{N+1,t}\\ \bar{\mathbf{r}}_{N+1,t}\end{bmatrix} = \Theta(\gamma_{N+1,t})\begin{bmatrix}\bar{\mathbf{e}}_{N,t}\\ \bar{\mathbf{r}}_{N,t-1}\end{bmatrix},$$

where

$$\Theta(\gamma_{N,t}) = \begin{bmatrix}(I - \gamma_{N,t}\gamma_{N,t}^*)^{-1/2} & 0\\ 0 & (I - \gamma_{N,t}^*\gamma_{N,t})^{-1/2}\end{bmatrix}\begin{bmatrix}I & -\gamma_{N,t}\\ -\gamma_{N,t}^* & I\end{bmatrix}.$$

(d) Show that $\Theta(\gamma_{N,t})$ is J-unitary, $J = (I \oplus -I)$, *i.e.*, that

$$\Theta(\gamma_{N,t})\begin{bmatrix}I & 0\\ 0 & -I\end{bmatrix}\Theta(\gamma_{N,t})^* = \begin{bmatrix}I & 0\\ 0 & -I\end{bmatrix}.$$

4.10 (The stationary case) We continue with the setting of Prob. 4.9. When the process $\{\mathbf{y}_t, t \geq 0\}$ is wide-sense stationary, *i.e.*, when

$$\langle \mathbf{y}_t, \mathbf{y}_s\rangle = \langle \mathbf{y}_{t-j}, \mathbf{y}_{s-j}\rangle, \quad j = 1, 2, \ldots,$$

show that $\{R_{N,t}^e, R_{N,t-1}^r, \gamma_{N+1,t}\}$ become independent of t, say $\{R_N^e, R_N^r, \gamma_{N+1}\}$. [Strictly speaking, $\gamma_{N,t} = 0$ for $t < N$ and is constant for $t \geq N$.]

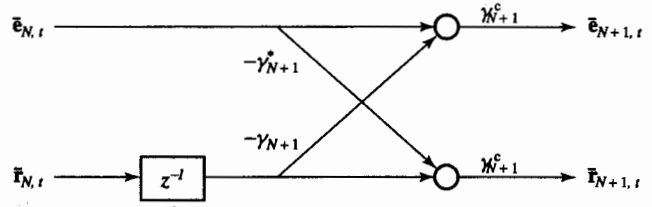


Figure 4.2 A section of a lattice filter for the forward and backward residuals.

Show further that we can build a so-called lattice filter, as shown in Fig. 4.2, to generate the residuals $\{\bar{\mathbf{e}}_{N,t}, \bar{\mathbf{r}}_{N,t-1}\}$. For simplicity assume that the random variables $\{\mathbf{y}_t\}$ and the reflection coefficients $\{\gamma_N\}$ are scalar-valued. The coefficient γ_{N+1}^c shown in the figure stands for

$$\gamma_{N+1}^c = \frac{1}{\sqrt{1 - |\gamma_{N+1}|^2}},$$

while z^{-1} denotes a unit-time delay.

4.11 (The (scalar) Levinson-Durbin algorithm) Consider again the same setting of Probs. 4.9 and 4.10 with $\{\mathbf{y}_t\}$ now assumed, for simplicity, a scalar-valued stationary process. Let

$$\mathbf{e}_{N,t} = \mathbf{y}_t + a_{N,1}\mathbf{y}_{t-1} + \ldots + a_{N,N}\mathbf{y}_{t-N},$$

for some coefficients $\{a_{N,j}, j = 1, \ldots, N\}$.

(a) Show that

$$\mathbf{r}_{N,t-1} = \mathbf{y}_{t-N-1} + a_{N,N}^*\mathbf{y}_{t-1} + \ldots + a_{N,1}^*\mathbf{y}_{t-N}.$$

(b) Introduce the row vector $a_N \triangleq [\,a_{N,N} \;\ldots\; a_{N,1}\;1\,]$. Show that a_N satisfies the so-called Yule-Walker equations

$$a_N T_N = [0 \;\ldots\; 0 \;\sigma_N^2],$$

where T_N is a Toeplitz matrix whose first column is $\text{col}\{c_0, c_1, \ldots, c_N\}$, with $c_i = \langle \mathbf{y}_t, \mathbf{y}_{t-i}\rangle$ and $\sigma_N^2 = \|\mathbf{e}_{N,t}\|^2$.

(c) Use the results of Prob. 4.9 to deduce the following so-called Levinson-Durbin recursions:

$$\begin{bmatrix}a_{N+1}\\ a_{N+1}^\#\end{bmatrix} = \begin{bmatrix}1 & -\gamma_{N+1}\\ -\gamma_{N+1}^* & 1\end{bmatrix}\begin{bmatrix}0 & a_N\\ a_N^\# & 0\end{bmatrix}, \quad \begin{bmatrix}a_0\\ a_0^\#\end{bmatrix} = \begin{bmatrix}1\\ 1\end{bmatrix},$$

where $a_N^\# \triangleq [\,1\; a_{N,1}^* \;\ldots\; a_{N,N}^*\,]$.

(d) Show also that

$$\sigma_{N+1}^2 = \sigma_N^2(1 - |\gamma_{N+1}|^2),$$

and that

$$\gamma_{N+1} = \frac{a_{N,N}c_1 + \ldots + a_{N,1}c_N + c_{N+1}}{\sigma_N^2}.$$

(e) Show that the number of elementary operations for going from a_k to a_{k+1} is $O(k)$, so that to finally obtain a_N (and $e_{N,t}$) requires $O(N^2)$ elementary operations.

Remark. The linear equation in part (b) relating $\{a_N, T_N, \sigma_N^2\}$ was first introduced by Yule (1927) in the problem of fitting an autoregressive model to sunspot data. The efficient algorithm of parts (c) and (d) for solving the Yule-Walker equations was derived by Durbin (1960) and earlier (in a more general context for a general right-hand side) by Levinson (1947). Whittle (1963) and Wiggins and Robinson (1965) extended these results to the block case. The usual derivations are quite algebraic. The stochastic/geometric arguments in Probs. 4.9 and 4.10 are more motivated and, as noted, apply to vector and nonstationary processes; they are taken from Lev-Ari, Kailath, and Cioffi (1984). Note also that the formation of γ_{N+1} requires an inner product evaluation, which is not fully parallelizable. An alternative fully parallelizable approach to the efficient solution of the Yule-Walker equations uses the generalized Schur algorithm — see App. F. ◆

4.12 (Causal estimation) Note first that for any square matrix A we can write

$$A = \{A\}_{m\text{-lower}} + \{A\}_{m\text{-upper}},$$

where we define

$$\left[\{A\}_{m\text{-upper}}\right]_{ij} \triangleq \begin{cases} A_{ij} & i \le j + m, \\ 0 & i > j + m, \end{cases} \qquad \left[\{A\}_{m\text{-lower}}\right]_{ij} \triangleq \begin{cases} 0 & i \le j + m, \\ A_{ij} & i > j + m. \end{cases}$$

Now consider the finite-horizon Wiener filtering problem of Sec. 4.1.2 where we are now interested in finding $\hat{\mathbf{s}}_{i|i-m}$, the l.l.m.s. estimator of $\mathbf{s}_i$ using the observations $\{\mathbf{y}_j\}_{j=0}^{i-m}$, for a given $m \ge 0$.

(a) Show that if we define $\hat{\mathbf{s}}_m \triangleq \text{col}\{\hat{\mathbf{s}}_{0|-m}, \hat{\mathbf{s}}_{1|1-m}, \ldots, \hat{\mathbf{s}}_{N|N-m}\} = K_m\mathbf{y}$, then K_m must be a lower triangular matrix with zeros on its diagonal and first $(m-1)$ subdiagonals.

(b) Show that $K_m = \left\{R_{sy}L^{-*}R_e^{-1}\right\}_{m\text{-lower}} L^{-1}$.

(c) Define $U^+ \triangleq K_m R_y - R_{sy}$. Use an argument similar to the Wiener-Hopf technique to show that

$$U^+ = -\left\{R_{sy}L^{-*}R_e^{-1}\right\}_{m\text{-upper}} R_e L^*.$$

(d) Show that $K_m = R_{sy}R_y^{-1} + U^+R_y^{-1} = (R_{sy} + U^+)R_y^{-1}$, where $\mathbf{s} = \text{col}\{\mathbf{s}_0, \ldots, \mathbf{s}_N\}$.

Remark. The above expression for K_m shows the deviation of the *causal* estimation gain from the *smoothing* estimation gain; in the smoothing problem $U^+ = 0$ and the gain becomes $R_{sy}R_y^{-1}$. ◆

(e) Show that the covariance matrix of the causal estimation error can be written as

$$E\|\mathbf{s} - \hat{\mathbf{s}}_m\|^2 = R_s - R_{sy}R_y^{-1}R_{ys} + U^+R_y^{-1}U^{+*},$$

which displays the increase in the error covariance arising from the use of a causal estimator.

4.13 (Additive noise) Assume $\mathbf{y} = \mathbf{s} + \mathbf{v}$, where

$$E\begin{bmatrix}\mathbf{s}\\ \mathbf{v}\end{bmatrix}\begin{bmatrix}\mathbf{s}\\ \mathbf{v}\end{bmatrix}^* = \begin{bmatrix}R_s & 0\\ 0 & R_v\end{bmatrix}, \quad \text{with } R_v \text{ diagonal.}$$

(a) Show from Prob. 4.12 that $U^+ = \left\{R_v L^{-*}R_e^{-1} - L\right\}_{m\text{-upper}} R_e L^*$.

(b) When $m = 1$ (strict prediction), we denote K_m by K_p and $\hat{\mathbf{s}}_m$ by $\hat{\mathbf{s}}_p$. Show that $K_p = I - L^{-1}$ and $U^+ = R_v - R_e L^*$.

(c) Relate $\hat{\mathbf{s}}_p$ and $\hat{\mathbf{s}}_f$ (which corresponds to $m = 0$). Relate also $\hat{\mathbf{s}}_{i|i}$ and $\hat{\mathbf{s}}_{i|i-1}$.

4.14 (Noise cancellation) Two sets of noisy measurements of an unknown scalar-valued deterministic trajectory $m(i)$ are available, say

$$\mathbf{y}_1(i) = m(i) + \mathbf{v}_1(i), \quad \mathbf{y}_2(i) = m(i) + \mathbf{v}_2(i), \quad 0 \le i \le N,$$

where $\{\mathbf{v}_1(\cdot), \mathbf{v}_2(\cdot)\}$ are zero-mean noise processes that are uncorrelated with each other and such that

$$\langle \mathbf{v}_1(i), \mathbf{v}_1(j)\rangle \triangleq r_1(i, j), \qquad \langle \mathbf{v}_2(i), \mathbf{v}_2(j)\rangle \triangleq r(i)\delta_{ij}.$$

That is, only the random process $\{\mathbf{v}_2(\cdot)\}$ is white while $\{\mathbf{v}_1(\cdot)\}$ is correlated.

It is suggested that the 2-input 1-output structure of Fig. 4.3 be used to estimate the trajectory $m(i)$. The output of the structure is denoted by $m(i) + \mathbf{w}(i)$, where $\mathbf{w}(i)$ denotes

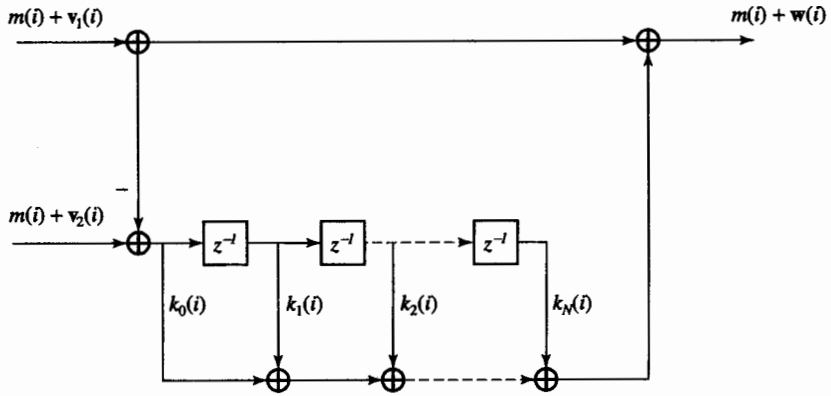


Figure 4.3 A structure for noise cancellation from two signal measurements.

the perturbation relative to the desired $m(i)$. The block in the lower branch indicates a finite-impulse response filter of length $(N+1)$ that is initially at rest. Determine the optimal coefficients of the filter in order to result in a perturbation $\mathbf{w}(i)$ of smallest variance. Solve the problem first for general $\{N, r(i), r_1(i, j)\}$ and then specialize the result to the case

$$N = 2, \quad r(i) = \frac{1}{2}, \quad r_1(i, j) = \left(\frac{1}{2}\right)^{|i-j|+1}, \quad 0 \le i, j \le 2.$$

[*Hint.* The coefficients of the filter can be time-variant.]

Appendix for Chapter 4

4.A LINEAR SPACES, MODULES, AND GRAMIANS

We have referred freely to linear vector spaces of vectors in $\mathbb{C}^n$ in App. 2.A. We shall often need to use more general/abstract vectors and more general concepts of inner products than common even in the mathematical literature. So this appendix gives a few formal definitions and, especially, several examples.

Linear Vector Spaces. Consider a linear space, say $\mathcal{V}$, whose elements are called *vectors*. This means that there exists some ring of *scalars*, say $\mathcal{S}$, an operation called multiplication by a scalar such that $\alpha x \in \mathcal{V}$ for all $x \in \mathcal{V}$ and $\alpha \in \mathcal{S}$, and an operation called addition of vectors, $x + y$, defined for every $x, y \in \mathcal{V}$ with values in $\mathcal{V}$. These operations are characterized by the properties

(i) $x + y = y + x$
(ii) $(x + y) + z = x + (y + z)$
(iii) $\alpha(x + y) = \alpha x + \alpha y$
(iv) $(\alpha + \beta)x = \alpha x + \beta x$
(v) $(\alpha\beta)x = \alpha(\beta x)$
(vi) $0 \cdot x = 0, \quad 1 \cdot x = x$

The ring $\mathcal{S}$ is usually the field of complex numbers but can be a more general algebraic object; in particular, we shall often take $\mathcal{S}$ as the ring of square complex matrices. For our discussions it will suffice to know that the product of any two elements in the ring $\mathcal{S}$ is also an element of $\mathcal{S}$.

Now corresponding to any pair of vectors, say $x \in \mathcal{V}$, $y \in \mathcal{V}$, the *inner product* $\langle x, y \rangle$ is defined as an element of $\mathcal{S}$ characterized by the following properties:

1. Linearity: $\langle \alpha_1 x_1 + \alpha_2 x_2, y \rangle = \alpha_1 \langle x_1, y \rangle + \alpha_2 \langle x_2, y \rangle$.
2. Reflexivity: $\langle y, x \rangle = \langle x, y \rangle^*$.
3. Nondegeneracy: $\|x\|^2 \triangleq \langle x, x \rangle$ is zero only when $x = 0$.

The operation $*$ depends on the space $\mathcal{S}$ and is referred to as an *involution* over $\mathcal{S}$. It has the property that, for all real λ and $\alpha, \beta \in \mathcal{S}$,

$$(\alpha^*)^* = \alpha, \quad (\alpha\beta)^* = \beta^*\alpha^*, \quad (\lambda\alpha)^* = \lambda\alpha^*. \tag{4.A.1}$$

When $\mathcal{S}$ is the field of complex numbers, $*$ is just the operation of taking the complex conjugate; when $\mathcal{S}$ is the ring of square matrices, $*$ stands for the conjugate transpose (usually called Hermitian transpose).

A triple $\{\mathcal{V}, \mathcal{S}, \langle , \rangle\}$, with the aforementioned properties, is called a *module*; when $\mathcal{S}$ is $\mathbb{C}$, or more generally when $\mathcal{S}$ is any field, then the triple $\{\mathcal{V}, \mathcal{S}, \langle , \rangle\}$ is called an inner product space. There are not many discussions of matrix-valued inner products in the literature. The original references (*e.g.,* Goldstine and Horwitz (1966)) go far beyond our needs; a more recent and more readable source is the appendix in Ben-Artzi and Gohberg (1994).

Gramian Matrices. In the usual mathematical terminology, the *Gramian* of a collection of vectors $\{a_0, \ldots, a_N\}$ (with $a_i \in \mathcal{V}$) is the matrix

$$G = \left[\langle a_i, a_j\rangle\right]_{i,j=0}^{N}. \tag{4.A.2}$$

By the reflexivity property of the inner product, we can readily see that the Gramian is Hermitian, $G^* = G$. Indeed,

$$[G^*]_{ij} = (G_{ji})^* = \langle a_j, a_i\rangle^* = \langle a_i, a_j\rangle = G_{ij},$$

as desired.

Remark. It is also possible to have the scalars $\alpha \in \mathcal{S}$ be such that they multiply the vectors $x \in \mathcal{V}$ from the *right*, *i.e.*, $x\alpha \in \mathcal{V}$ for all $x \in \mathcal{V}$ and $\alpha \in \mathcal{S}$. In this case, all of the above discussions go through, provided we modify (iii)–(iv) accordingly and rewrite the linearity condition for inner products as

$$\langle x_1\alpha_1 + x_2\alpha_2, y\rangle = \langle x_1, y\rangle\alpha_1 + \langle x_2, y\rangle\alpha_2.$$

We shall encounter such a case in Example 4.A.2. ◆

Example 4.A.1 [n-Dimensional Column Vectors] Suppose the vectors x are $n \times 1$ columns of complex entries, say $x = \underline{h} \in \mathbb{C}^n = \mathcal{V}$, with $\mathcal{S} = \mathbb{C}$, and the usual inner product $\langle \underline{h}_1, \underline{h}_2\rangle = \underline{h}_2^*\underline{h}_1 \in \mathbb{C}$. It is straightforward to see that $\{\mathbb{C}^n, \mathbb{C}, \langle\cdot,\cdot\rangle\}$ has the properties mentioned above and is a module. It is usually referred to as an n-dimensional Euclidean space. Moreover, if we define

$$H = \left[\underline{h}_1 \ \ldots \ \underline{h}_N\right], \tag{4.A.3}$$

then the Gramian of the collection $\{\underline{h}_1, \ldots, \underline{h}_N\}$ is the matrix

$$G = \left[\langle \underline{h}_i, \underline{h}_j\rangle\right] = \left[\underline{h}_j^*\underline{h}_i\right] = H^*H. \tag{4.A.4}$$

◆

Example 4.A.2 [$n \times N$ Matrices] Let us now suppose that the vectors x are $n \times N$ matrices, *i.e.*, that $\mathcal{V} = \mathbb{C}^{n\times N}$. Moreover, let us take $\mathcal{S} = \mathbb{C}^{N\times N}$, *i.e.*, the ring of square $N \times N$ matrices. If we consider standard matrix addition and multiplication (here the scalars are multiplied from the right), then it is straightforward to verify that $\mathcal{V}$ and $\mathcal{S}$ have the properties (i)–(vi) and that $\mathcal{V}$ is a linear vector space over $\mathcal{S}$.

To obtain an inner product space (or module) we define the inner product as

$$\langle x, y\rangle = y^*x \in \mathbb{C}^{N\times N} = \mathcal{S}, \quad \text{for every} \quad x, y \in \mathbb{C}^{n\times N} = \mathcal{V}. \tag{4.A.5}$$

Once more we can readily check that the above definition for inner product satisfies the three required properties of linearity, reflexivity, and nondegeneracy.

We can now remark that if we write an element $H \in \mathcal{V}$ in terms of its columns

$$H = \left[\underline{h}_1 \ \ldots \ \underline{h}_N\right],$$

then the squared norm of H is just $\langle H, H\rangle = H^*H$, which is simply the Gramian corresponding to the column vectors $\{\underline{h}_1, \ldots, \underline{h}_N\}$ of Example 4.A.1.

Of course, if we have a collection of elements $\{H_1, \ldots, H_m\}$ in $\mathcal{V} = \mathbb{C}^{n\times N}$, then their Gramian is the block matrix $G = \left[\langle H_i, H_j\rangle\right] = \left[H_j^*H_i\right]$. If we define the $n \times Nm$ matrix $A = \left[H_1 \ \ldots \ H_m\right]$, then note that G now is the $Nm \times Nm$ matrix $G = A^*A$. ◆

Example 4.A.3 [Scalar Random Variables] Suppose the vectors are scalar random variables, say

$$x = \mathbf{y}(\omega) \ : \ \Omega \to \mathbb{C}. \tag{4.A.6}$$

Take $\mathcal{S} = \mathbb{C}$, the ring of complex scalars, and define the inner product as

$$\langle \mathbf{y}_1(\omega), \mathbf{y}_2(\omega)\rangle = E\mathbf{y}_1(\omega)\mathbf{y}_2^*(\omega) \in \mathbb{C} = \mathcal{S}. \tag{4.A.7}$$

We can verify that the above space and inner product satisfy conditions (i)–(vi) and (1)–(3), respectively. If we now consider a collection of scalar random variables, say $\{\mathbf{y}_1(\omega), \ldots, \mathbf{y}_N(\omega)\}$, then their Gramian is given by

$$G = \left[\langle \mathbf{y}_i(\omega), \mathbf{y}_j(\omega)\rangle\right] = \left[E\mathbf{y}_i(\omega)\mathbf{y}_j^*(\omega)\right] = R_{\mathbf{y}}. \tag{4.A.8}$$

◆

Example 4.A.4 [Vector-Valued Random Variables] Consider now the space of n-dimensional column-vector-valued random variables. Therefore, the elements of the space are

$$x = \mathbf{y}(\omega) = \text{col}\{\mathbf{y}_1(\omega), \ldots, \mathbf{y}_N(\omega)\}, \qquad \text{where} \quad \mathbf{y}_i(\omega) \ : \ \Omega \to \mathbb{C}, \ \ i = 1, \ldots, N.$$

This space is obviously linear over the ring of square $N \times N$ matrices, $\mathcal{S} = \mathbb{C}^{N\times N}$. If we define the inner product as

$$\langle \mathbf{y}(\omega), \mathbf{z}(\omega)\rangle = E\mathbf{y}(\omega)\mathbf{z}^*(\omega) \in \mathbb{C}^{N\times N} = \mathcal{S}, \tag{4.A.9}$$

then we readily see that the required conditions of linearity, reflexivity, and nondegeneracy are satisfied.

Moreover, note that the inner product

$$\langle \mathbf{y}(\omega), \mathbf{y}(\omega)\rangle = E\mathbf{y}(\omega)\mathbf{y}^*(\omega)$$

yields the Gramian of the elements $\{\mathbf{y}_1(\omega), \ldots, \mathbf{y}_N(\omega)\}$ in Example 4.A.3. ◆

The random variables that will be of interest to us in this book will generally be vector-valued, so that their Gramian will be a block matrix (which is of course a matrix itself). This is usually not the case in the deterministic least-squares problems that are usually studied, but they very well could be (*e.g.*, in multichannel adaptive filtering).

An important notion in linear vector spaces is that of linear independence.

Definition 4.A.1. (Linear Independence) *A set of vectors $x_i \in \mathcal{V}$, $i = 0, \ldots, N$ is linearly independent if, and only if, there exists no set of scalars $\alpha_i \in \mathcal{S}$, $i = 0, \ldots, N$, not all identically zero, such that $\sum_{i=0}^{N} \alpha_i x_i = 0$.* ◆

Gramians were first introduced to provide a test for linear independence.

Lemma 4.A.1 (A Gramian Test) *The vectors $\{a_i \in \mathcal{V},\ i = 0, 1, \ldots, N\}$ are linearly independent if, and only if, their corresponding Gramian matrix is nonsingular.* ■

Proof: To prove the "only if" direction of the lemma, suppose that the $\{a_i\}$ are linearly dependent. Then there exists a nonzero linear combination of the $\{a_i\}$ that yields the zero vector, *i.e.*, $\sum_{i=0}^{N} c_i a_i = 0$ for some $c_i \in \mathcal{S}$ that are not all identically zero. Taking the inner product of the above expression with a_j implies $\sum_{i=0}^{N} c_i \langle a_i, a_j \rangle = 0$, for $j = 0, 1, \ldots N$, or $cG = 0$, where $c = \begin{bmatrix} c_0 & \ldots & c_N \end{bmatrix}$ is nonzero and G is the Gramian of the $\{a_i\}$. But this last expression shows that G is singular.

Conversely, suppose that the Gramian is singular. Then there must exist a nonzero row vector c such that $0 = cG$. Then

$$0 = cGc^* = c\left[\langle a_i, a_j \rangle\right] c^* = \langle \sum_{i=0}^{N} c_i a_i, \sum_{j=0}^{N} c_j a_j \rangle = \left\| \sum_{i=0}^{N} c_i a_i \right\|^2 .$$

But by the nondegeneracy of inner products the only vector with zero length is the zero vector. Thus we must have $\sum_{i=0}^{N} c_i a_i = 0$, meaning that the $\{a_i\}$ are linearly dependent. ◆

The above lemma explains the assumption, often made in the stochastic problem, that the *observed* random variables $\{\mathbf{y}_i\}$ are such that $R_y > 0$, *i.e.*, that the Gramian is positive-definite. This appears to be stronger than the requirement that the Gramian be nonsingular, but actually it is not because of the fact that Gramians are always nonnegative-definite, so that being nonsingular is equivalent to being positive-definite.

In any case, note that in both the deterministic and stochastic problems that we studied in Chs. 2 and 3, the Gramians are the coefficient matrices in the basic linear equations defining the solutions, $(H^*H)\hat{x} = H^*y$ and $K_o R_y = R_{xy}$.

We should also recall that these equations follow immediately from the orthogonality conditions that characterize projections,

$$\langle y - H\hat{x}, \underline{h}_i \rangle = 0, \quad i = 0, 1, \ldots N, \quad \text{or} \quad \langle \mathbf{x} - K_o \mathbf{y}, \mathbf{y} \rangle = 0.$$

Uniqueness of Projections. Let $\mathcal{L}$ be a linear subspace of an inner-product space $\mathcal{V}$ and let y be an arbitrary element of $\mathcal{V}$. Let $\mathcal{S}$ denote the corresponding space of "scalars". We shall be dealing with finite-dimensional linear vector spaces, say $\mathcal{V}$ is N-dimensional and $\mathcal{L}$ is M-dimensional with $M \leq N$. This in turn implies the existence of a collection of N orthonormal vectors $\{v_i\}_{i=1}^{N}$, also called *basis vectors*, such that

1. The $\{v_i\}$ are orthogonal to each other and have unit norm, *i.e.*, $\langle v_i, v_j \rangle = 1$ for $i = j$ and zero otherwise.
2. Every $y \in \mathcal{V}$ can be uniquely expressed as a linear combination of the $\{v_i\}$, say $y = \sum_{i=1}^{N} \alpha_i v_i$, $\alpha_i \in \mathcal{S}$, where $\alpha_i = \langle y, v_i \rangle$.

We may assume, without loss of generality, that the first M basis vectors, $\{v_i\}_{i=1}^{M}$, also form an orthonormal basis for the subspace $\mathcal{L}$.

Lemma 4.A.2 (Uniqueness of Projections) *Given $y \in \mathcal{V}$, there exists a unique element of $\mathcal{L}$, denoted by $\hat{y}$, such that*

$$\langle y - \hat{y}, a \rangle = 0, \quad \textit{for all} \quad a \in \mathcal{L}.$$

■

Proof: We start by expressing y as a linear combination of the basis vectors $\{v_i\}_{i=1}^{N}$, say $y = \sum_{i=1}^{N} \alpha_i v_i$. Then we claim that $\hat{y} = \sum_{i=1}^{M} \alpha_i v_i$. Indeed, it is obvious that $\hat{y} \in \mathcal{L}$ since, by assumption, the first M basis vectors span the subspace $\mathcal{L}$. Moreover, it is easy to see that, for any $a \in \mathcal{L}$, $\langle y - \hat{y}, a \rangle = 0$, since a is a linear combination of the first M basis vectors, while $(y - \hat{y})$ is a linear combination of the remaining basis vectors and, as defined above, the basis vectors are orthogonal to each other.

To establish uniqueness, we now assume that there exist two elements in $\mathcal{L}$, say $\hat{y}_1$ and $\hat{y}_2$, such that

$$\langle y - \hat{y}_1, a \rangle = 0 \quad \text{and} \quad \langle y - \hat{y}_2, a \rangle = 0, \quad \text{for all } a \in \mathcal{L}.$$

But since $\hat{y}_1 \in \mathcal{L}$ and $\hat{y}_2 \in \mathcal{L}$, $(\hat{y}_1 - \hat{y}_2) \in \mathcal{L}$ and we obtain

$$\begin{aligned} \langle \hat{y}_1 - \hat{y}_2, \hat{y}_1 - \hat{y}_2 \rangle &= \langle \hat{y}_1 - \hat{y}_2, (\hat{y}_1 - y) + (y - \hat{y}_2) \rangle, \\ &= \langle \hat{y}_1 - \hat{y}_2, \hat{y}_1 - y \rangle + \langle \hat{y}_1 - \hat{y}_2, y - \hat{y}_2 \rangle \ = 0 + 0 = 0. \end{aligned}$$

Therefore, $\|\hat{y}_1 - \hat{y}_2\|^2 = 0$ and we conclude that we must have $\hat{y}_1 = \hat{y}_2$. ◆

CHAPTER 5

State-Space Models

Much of the discussion in this chapter may be superfluous for readers already familiar with state-space models and their power. We have included a more motivated and more leisurely approach for the benefit of readers from other fields (*e.g.*, communications and signal processing) where the use of state-space models is not yet widespread. However, all may still find some value in Sec. 5.4, where we shall present one of the deeper reasons for the success of state-space descriptions — the so-called wide-sense Markov property of the state-vector process and the so-called Markovian representations that it yields for the (generally non-Markov) output process. Among other things, this discussion will lead us to introduce the concept of reverse-time, or backwards, Markovian models, which will be useful in the study of smoothing problems in Ch. 10 and in the study of duality in Ch. 15).

5.1 THE EXPONENTIALLY CORRELATED PROCESS

As noted in Sec. 4.4, the surprising simplicity of the formulas for the innovations of the process with the exponential covariance function invites a closer examination. In this section we shall show that we can also describe the process by a simple time-domain model, which will allow us to find the innovations much more directly than when starting with the covariance function. Moreover, this time-domain model will naturally suggest various nontrivial extensions.

A standard way of obtaining a model for a scalar stationary process is by studying its so-called z-spectrum, which is the bilateral z-transform of the covariance function. While more details on z-spectra are provided in Ch. 6, here it suffices to note that for the exponentially correlated process $\{\mathbf{y}_i\}$ with $\langle \mathbf{y}_i, \mathbf{y}_j \rangle = E\mathbf{y}_i\mathbf{y}_j^* = a^{|i-j|}$, and $0 < a < 1$,

the z-spectrum is readily evaluated as

$$S_y(z) \triangleq \sum_{j=-\infty}^{\infty} a^{|j|}z^{-j} = \sum_{j=-\infty}^{-1} a^{-j}z^{-j} + \sum_{j=0}^{\infty} a^j z^{-j},$$

$$= \frac{az}{1-az} + \frac{1}{1-az^{-1}}, \quad \text{for} \quad a < |z| < \frac{1}{a},$$

$$= \frac{(1-a^2)}{(1-az^{-1})(1-az)}. \tag{5.1.1}$$

Now it is a well-known (and easily verified) fact from the elementary theory of random processes that we can model the process $\{\mathbf{y}_i\}$ as the output of a linear filter, say $H(z)$, driven by a white-noise process (see also Secs. 6.3.2 and 6.4). If we denote the white-noise process by $\{\mathbf{u}_i\}$, with variance $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = (1-a^2)\delta_{ij}$, then $H(z)$ can be taken as any stable and causal transfer function (*i.e.*, with poles strictly inside the unit disc) that satisfies

$$S_y(z) = (1-a^2)H(z)\left[H\left(\frac{1}{z^*}\right)\right]^*.$$

In view of the expression (5.1.1) for $S_y(z)$ we see that we can take

$$H(z) = \frac{z^{-m}}{1-az^{-1}}, \quad \text{for any integer} \quad m \geq 0.$$

A natural choice for $H(z)$ is $H(z) = L(z) \triangleq 1/(1-az^{-1})$, which corresponds to $m = 0$. [The resulting function $L(z)$ is known as the canonical spectral factor of $S_y(z)$ — see Sec. 6.3.] However, for reasons explained below, we shall choose the model with $m = 1$,

$$H(z) = \frac{z^{-1}}{1-az^{-1}} = z^{-1}\left[1 + \sum_{j=1}^{\infty} a^j z^{-j}\right]. \tag{5.1.2}$$

This transfer function correponds to the input-output equation,

$$\mathbf{y}(z) = H(z)\mathbf{u}(z) = \frac{z^{-1}}{1-az^{-1}}\mathbf{u}(z) = \frac{1}{z-a}\mathbf{u}(z),$$

or

$$z\mathbf{y}(z) - a\mathbf{y}(z) = \mathbf{u}(z), \tag{5.1.3}$$

where

$$\mathbf{y}(z) \triangleq \sum_{i=-\infty}^{\infty} \mathbf{y}_i z^{-i}, \qquad \mathbf{u}(z) \triangleq \sum_{i=-\infty}^{\infty} \mathbf{u}_i z^{-i}.$$

Comparing the coefficients of the $\{z^{-i}\}$ on both sides of (5.1.3) gives the difference equation

$$\mathbf{y}_{i+1} - a\mathbf{y}_i = \mathbf{u}_i, \quad i > -\infty, \tag{5.1.4}$$

where, as befits the stationarity assumption, we assume that all processes begin in the remote past. From (5.1.4), we can write

$$\mathbf{y}_i = \sum_{j=0}^{\infty} a^j \mathbf{u}_{i-j-1},$$

from which we immediately obtain the important property that the process $\{\mathbf{u}_i\}$ is uncorrelated with present and past output, *viz.*,

$$\langle \mathbf{u}_i, \mathbf{y}_j \rangle = 0, \quad i \geq j, \tag{5.1.5}$$

a property whose significance will appear shortly.

We can now explain that the choice $m = 1$, which leads to (5.1.4), was made[1] because (5.1.4) can be regarded as a special case of the state-space model (1.2.1) assumed earlier in Ch. 1: with $\mathbf{x}_i \triangleq \mathbf{y}_i$, the process $\mathbf{y}_i$ in (5.1.4) can be rewritten in state-space form as

$$\mathbf{x}_{i+1} = a\mathbf{x}_i + \mathbf{u}_i, \qquad \mathbf{y}_i = \mathbf{x}_i.$$

5.1.1 Finite Interval Problems; Initial Conditions for Stationarity

When the exponentially correlated process $\{\mathbf{y}_i\}$ was introduced in Sec. 4.4, we considered it starting at the finite time instant $i = 0$. So a natural question is whether representations such as (5.1.4) can still apply. They can, provided we properly define the statistics of the initial conditions. Thus let us write

$$\mathbf{y}_{i+1} - a\mathbf{y}_i = \mathbf{u}_i, \quad i \geq 0, \qquad \langle \mathbf{u}_i, \mathbf{u}_j \rangle = (1 - a^2)\delta_{ij}. \tag{5.1.6}$$

The difference from (5.1.4) is now that at $i = 0$, we have an initial condition $\mathbf{y}_0$. What should we assume about its statistics? It is natural to assume that its variance is exactly what it would be for the stationary process, and that (*cf.* (5.1.5)) it is uncorrelated with $\{\mathbf{u}_i, i \geq 0\}$, *i.e.*,

$$\|\mathbf{y}_0\|^2 = 1, \quad \langle \mathbf{u}_i, \mathbf{y}_0 \rangle = 0, \quad i \geq 0. \tag{5.1.7}$$

In fact, with these assumptions it turns out that the process defined by (5.1.6)–(5.1.7) is stationary over $[0, \infty)$ with

$$\langle \mathbf{y}_i, \mathbf{y}_j \rangle = a^{|i-j|}, \quad i, j \geq 0. \tag{5.1.8}$$

A formal verification is useful and will be given in Sec. 5.2; here let us first present our major reason for introducing the model (5.1.6)–(5.1.7).

[1] Had we chosen the model $H(z) = L(z) = 1/(1 - az^{-1})$, the resulting difference equation would have been $\mathbf{y}_i = a\mathbf{y}_{i-1} + \mathbf{u}_i$. By introducing the state vector $\mathbf{x}_i = \mathbf{y}_{i-1}$ we then obtain the alternative state-space realization $\mathbf{x}_{i+1} = a\mathbf{x}_i + \mathbf{u}_i$ and $\mathbf{y}_i = a\mathbf{x}_i + \mathbf{u}_i$, which does not satisfy (5.1.5). However, of course, it can still be handled by the methods of this book — see Prob. 5.1.

5.1.2 Innovations from the Process Model

In Sec. 4.4 we obtained the innovations of the exponentially correlated process by using the given covariance function. Suppose however that we happen to have been directly given the model (5.1.6)–(5.1.7). We could of course compute the covariances (5.1.8) and then proceed as in Sec. 4.4.

However, since all the statistical information about the process $\{\mathbf{y}_i\}$ is already present in the model (5.1.6)–(5.1.7), we might ask if it is possible to obtain the innovations directly from the model without first computing the covariances. The answer is in fact yes. In Ch. 9, this fact will be used to derive the Kalman filter recursions. The reader may also recognize parallels with the discussion of the so-called QR *method in Sec. 2.5.*

Given the model (5.1.6)–(5.1.7), it is natural to proceed as follows. Recall that

$$\begin{aligned} \hat{\mathbf{y}}_{i+1|i} &= \text{the l.lm.s. estimator of } \mathbf{y}_{i+1} \text{ given } \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}, \\ &= \text{the projection of } \mathbf{y}_{i+1} \text{ on } \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}. \end{aligned}$$

Now by the *linearity* property of l.l.m.s. estimators (*i.e.*, that the projection of a sum is the sum of the projections), we can write immediately from the model equations (5.1.6)–(5.1.7) that

$$\hat{\mathbf{y}}_{i+1|i} - a\hat{\mathbf{y}}_{i|i} = \hat{\mathbf{u}}_{i|i}, \quad i \geq 0,$$

where "$\cdot|_{-1}$" means given no observations. Now, for $i \geq 0$, we clearly have

$$\hat{\mathbf{y}}_{i|i} \triangleq \text{the l.l.m.s. estimator of } \mathbf{y}_i \text{ given } \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_i\} = \mathbf{y}_i. \tag{5.1.9}$$

Moreover, from the defining equation (5.1.6) we see that $\mathbf{y}_j \in \mathcal{L}\{\mathbf{y}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{j-1}\}$, *i.e.*, $\mathbf{y}_j$ depends only upon past $\{\mathbf{u}_k, k < j\}$. Therefore, since $\mathbf{u}_i \perp \mathbf{u}_k$ for $i > k$, it holds that $\mathbf{u}_i \perp \mathbf{y}_j$, $i \geq j$, and so

$$\hat{\mathbf{u}}_{i|i} = \text{the l.l.m.s. estimator of } \mathbf{u}_i \text{ given } \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_i\} = 0, \quad i \geq 0. \tag{5.1.10}$$

Hence, it follows that

$$\hat{\mathbf{y}}_{i+1|i} = a\mathbf{y}_i, \quad i \geq 0, \tag{5.1.11}$$

and

$$\mathbf{e}_{i+1} = \mathbf{y}_{i+1} - \hat{\mathbf{y}}_{i+1|i} = \mathbf{y}_{i+1} - a\mathbf{y}_i, \quad i \geq 0, \qquad \mathbf{e}_0 = \mathbf{y}_0. \tag{5.1.12}$$

This is exactly the same expression obtained in Sec. 4.4 (*cf.* (4.4.14)) by starting with the known (exponential) covariance function and making several computations. On the other hand, starting with the model allows a much more direct approach.

5.2 GOING BEYOND THE STATIONARY CASE

The above arguments show the convenience of working with a model for the random process. There is yet another advantage in adopting such model-based arguments. While we have assumed so far that the underlying process is stationary, starting with

the model (5.1.4) allows us to show that the restriction to a stationary process is not at all essential. To explore this, let us start afresh by considering the process $\{\mathbf{y}_i, i \geq 0\}$ described by the linear stochastic difference equation (5.1.4), *viz.*,

$$\mathbf{y}_{i+1} - a\mathbf{y}_i = \mathbf{u}_i, \quad i \geq 0, \tag{5.2.1}$$

with an initial condition $\mathbf{y}_0$ and $\mathbf{u}_i$ such that

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = Q_i \delta_{ij}, \qquad \|\mathbf{y}_0\|^2 = \Pi_0. \tag{5.2.2}$$

Note that we are allowing a *nonstationary* input process since $\|\mathbf{u}_i\|^2$ is not assumed to be constant. We shall see presently that with certain additional assumptions, the innovations can still be computed as in Sec. 5.1.2.

The first additional assumption is that the initial condition $\mathbf{y}_0$ is uncorrelated with the input sequence $\{\mathbf{u}_i, i \geq 0\}$,

$$\langle \mathbf{u}_i, \mathbf{y}_0 \rangle = 0, \quad i \geq 0. \tag{5.2.3}$$

Then, as earlier in Sec. 5.1.2, it follows that

$$\langle \mathbf{u}_i, \mathbf{y}_j \rangle = 0, \quad j \leq i. \tag{5.2.4}$$

In any case, let us now give a first illustration of the value of the assumption (5.2.3), and its consequence (5.2.4), by computing the variances of the variables $\{\mathbf{y}_i, i \geq 0\}$. Let us denote

$$\|\mathbf{y}_i\|^2 = \Pi_i, \quad i \geq 0. \tag{5.2.5}$$

Then it follows easily from (5.2.1)–(5.2.3) that

$$\Pi_{i+1} = a^2 \Pi_i + Q_i, \quad i \geq 0, \tag{5.2.6}$$

from which the values $\{\Pi_i\}$ can be successively computed, starting with Π_0 when $i = 0$.

5.2.1 Stationary Processes

Since the values of the $\{\Pi_i\}$ change with i, we shall not have a stationary process in general, though this can be arranged by proper choice of the a priori values $\{a, \Pi_0\}$. For stationarity, it is reasonable to assume that the input process $\{\mathbf{u}_i\}$ is stationary, so that $Q_i = Q$, say. We can now see that choosing $\Pi_0 = Q/(1 - a^2)$, will give

$$\Pi_1 = \frac{a^2}{1 - a^2} Q + Q = \frac{Q}{1 - a^2},$$

and so also for all Π_i, $i \geq 0$. Of course it is necessary that $|a| < 1$, in order to have positive variances. In other words, when the system described by (5.2.6) is stable and we start with the initial condition $\Pi_0 = Q/(1 - a^2)$, then the process variance will be constant, $\Pi_i = Q/(1 - a^2)$ for $i \geq 0$. Of course, for stationarity, we also have to check that $\langle \mathbf{y}_i, \mathbf{y}_j \rangle$ depends only upon $|i - j|$, and we shall leave it to the reader to show that this holds when we choose $\Pi_0 = Q/(1 - a^2)$ — see also Sec. 5.3.

5.2.2 Nonstationary Processes

The important point is that stationarity is only achieved by a special choice of the variance of the initial condition, and that in general the model (5.2.1)–(5.2.3) will describe a *nonstationary* process. Moreover, a review of Sec. 5.1.2 will show that the result and derivation of (5.1.12) for the innovations will still continue to hold even if $|a| \geq 1$, *i.e., the model is unstable. We do not even need constancy of the model parameters.* That is, we could let a be dependent on time, and consider

$$\mathbf{y}_{i+1} - a_i \mathbf{y}_i = \mathbf{u}_i, \quad i \geq 0, \tag{5.2.7}$$

where $\mathbf{u}_i$ and $\mathbf{y}_0$ are zero-mean and

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = Q_i \delta_{ij}, \quad \|\mathbf{y}_0\|^2 = \Pi_0, \quad \langle \mathbf{u}_i, \mathbf{y}_0 \rangle = 0, \tag{5.2.8}$$

and still obtain, for known $\{a_i, Q_i\}$,

$$\hat{\mathbf{y}}_{i+1|i} = a_i \mathbf{y}_i, \quad i \geq 0. \tag{5.2.9}$$

However, while the formula for the innovations is the same, whether $|a| > 1$ or $|a| < 1$, we might wonder about the innovations variance. Clearly, the earlier formula (4.4.14), *viz.*, $\|\mathbf{e}_i\|^2 = (1 - a^2)$, will never be meaningful if $|a| > 1$. Therefore we return to the model equations (5.2.1)–(5.2.3) and the formula (5.1.12), which show that

$$\mathbf{e}_{i+1} = \mathbf{u}_i, \quad \|\mathbf{e}_{i+1}\|^2 = \|\mathbf{u}_i\|^2 = Q_i, \quad i \geq 0. \tag{5.2.10}$$

The point is that when $|a| > 1$, the nonstationary process has a growing variance, but the variance of the predicted estimator $\hat{\mathbf{y}}_{i+1|i}$ also grows (*cf.* (5.2.9)), so that the error variance remains finite: the estimator still tracks the process.

5.3 HIGHER-ORDER PROCESSES AND STATE-SPACE MODELS

To generalize the several results derived in Sec. 5.2, it is natural to consider processes $\{\mathbf{y}_i\}$ generated by models of the form

$$\mathbf{y}_{i+1} = a_{0,i}\mathbf{y}_i + \ldots + a_{n-1,i}\mathbf{y}_{i-n+1} + b_{0,i}\mathbf{u}_i + \ldots + b_{m,i}\mathbf{u}_{i-m}. \tag{5.3.1}$$

Such higher-order models are often encountered in time-series analysis, system identification, control theory, digital signal processing, etc. However, the efforts to compute some of the things we did previously (*e.g.*, expressions for variances and covariances, or determining initial condition statistics to ensure stationarity for constant parameter models) can very rapidly bog down in notational complexity.

The insight from linear system theory is that the method of choice for alleviating these notational burdens (especially in the general time-variant case) is to rewrite (5.3.1) in a so-called *state-space form*. We shall illustrate this point by first considering an important special case of (5.3.1).

5.3.1 Autoregressive Processes

We shall define an n-th order autoregressive (AR) process $\{\mathbf{y}_i\}$ as one that satisfies the stochastic difference equation

$$\mathbf{y}_{i+1} = a_{0,i}\mathbf{y}_i + \ldots + a_{n-1,i}\mathbf{y}_{i-n+1} + \mathbf{u}_i, \quad i \geq 0, \tag{5.3.2}$$

where $\{\mathbf{u}_i\}$ is a zero-mean white-noise process with

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = Q_i \delta_{ij}, \tag{5.3.3}$$

and the initial values $\{\mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-n+1}\}$ are zero-mean random variables with some known $n \times n$ covariance *matrix*, say Π_0,

$$\Pi_0 = \left[\langle \mathbf{y}_{-j}, \mathbf{y}_{-k} \rangle\right]_{j,k \in \{0,\ldots,n-1\}}.$$

We also assume that

$$\langle \mathbf{u}_i, \mathbf{y}_j \rangle = 0 \quad \text{for} \quad -n+1 \leq j \leq 0 \quad \text{and} \quad i \geq 0, \tag{5.3.4}$$

which will ensure that (show this)

$$\langle \mathbf{u}_i, \mathbf{y}_j \rangle = 0, \quad i \geq j \geq 0. \tag{5.3.5}$$

We shall show that many of the results of the earlier sections can readily be extended to such processes.

5.3.2 Handling Initial Conditions

So assume first that the coefficients $\{a_{k,i}\}$ are independent of i, say $\{a_k\}$, as are the $Q_i = \|\mathbf{u}_i\|^2$ in (5.3.3), and that we would like to choose Π_0 so that the process $\{\mathbf{y}_i, i \geq -n\}$ is stationary. The stationarity implies that Π_0 must be a Toeplitz matrix, *i.e.*, one with constant values along each diagonal. It also implies that all sequences

$$\mathbf{x}_i \triangleq \text{col}\{\mathbf{y}_i, \mathbf{y}_{i-1}, \ldots, \mathbf{y}_{i-n+1}\}, \quad i \geq 0, \tag{5.3.6}$$

must have the same covariance matrix, *i.e.*, in our case $\|\mathbf{x}_i\|^2 = \Pi_0, i \geq 0$. There are n^2 unknowns in Π_0 (actually n since Π_0 is Toeplitz) and we need at least that many equations to determine them. Now note, for example, that

$$\mathbf{x}_{i+1} \triangleq \begin{bmatrix} \mathbf{y}_{i+1} \\ \mathbf{y}_i \\ \mathbf{y}_{i-1} \\ \vdots \\ \mathbf{y}_{i-n+2} \end{bmatrix} = \underbrace{\begin{bmatrix} a_0 & a_1 & \ldots & a_{n-1} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \quad 0 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} \mathbf{y}_i \\ \mathbf{y}_{i-1} \\ \mathbf{y}_{i-2} \\ \vdots \\ \mathbf{y}_{i-n+1} \end{bmatrix}}_{\mathbf{x}_i} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{G} \mathbf{u}_i$$

and, hence, we obtain the state-space representation

$$\mathbf{x}_{i+1} = F\mathbf{x}_i + G\mathbf{u}_i, \quad \text{for the above } F \text{ and } G, \tag{5.3.7}$$

$$\mathbf{y}_i = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix} \mathbf{x}_i. \tag{5.3.8}$$

The matrix F is a so-called *companion* matrix associated with the characteristic polynomial of F,

$$\det(zI - F) = z^n - a_0 z^{n-1} - \ldots - a_{n-1} = a(z), \quad \text{say.} \tag{5.3.9}$$

We say that F is *stable* if all its eigenvalues are less than one in magnitude, or equivalently, if all the roots of $a(z)$ lie strictly within the unit disc in the complex z-plane.

Now the equality of the covariances of the vectors $\mathbf{x}_{i+1}$ and $\mathbf{x}_i$ yields the linear equation

$$\Pi_0 = F\Pi_0 F^* + GQG^*, \quad Q = \|\mathbf{u}_i\|^2. \tag{5.3.10}$$

This is in fact a famous (so-called Lyapunov) matrix equation in system theory, much studied in exploring the stability of discrete-time systems by the so-called method of Lyapunov, (see App. D). In our problem, it can be shown that if F is stable, then there is a unique, Hermitian, nonnegative-definite solution Π_0 to (5.3.10). Solutions of the equation (5.3.10) are not necessarily Toeplitz, but it turns out that in fact they are for all matrices $\{F, G\}$ of the form (5.3.7) — see Prob 5.16.

The reader may wish to check the power of the above matrix-based approach to handling the initial conditions by trying even the case $n = 3$ by direct calculation. There are many other computational and conceptual advantages of introducing state-space models. The above problem led us, almost involuntarily, to the use of matrix notation and state equations. Actually one can directly introduce state-space descriptions for autoregressive processes and also ARMA processes. For AR processes, Eq. (5.3.7) suggests a direct way of getting such a representation, even in the time-variant case.

5.3.3 State-Space Descriptions

Thus if we define a state vector

$$\mathbf{x}_i \triangleq \text{col}\{\mathbf{y}_i, \mathbf{y}_{i-1}, \ldots, \mathbf{y}_{i-n+1}\}, \quad i \geq 0, \tag{5.3.11}$$

we can write the autoregressive process (5.3.2) in state-space form as

$$\begin{cases} \mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i \\ \mathbf{y}_i = H_i\mathbf{x}_i \end{cases} \tag{5.3.12}$$

where F_i is a companion matrix as in (5.3.7), but with top row $[a_{0,i} \ldots a_{n-1,i}]$, $G_i = \text{col}\{1, 0, \ldots, 0\}$, $H_i = [1, 0, \ldots, 0]$, and $\mathbf{x}_0 = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_{-n+1}\}$. The process $\{\mathbf{u}_i\}$ is zero-mean and white with $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = Q_i\delta_{ij}$. Let us introduce the state covariance matrix $\Pi_i = \langle \mathbf{x}_i, \mathbf{x}_i \rangle$. Then it is easy to check that we have the recursion

$$\Pi_{i+1} = F_i\Pi_i F_i^* + G_i Q_i G_i, \quad i \geq 0, \tag{5.3.13}$$

and that when the system is time-invariant and F is stable, the equation (5.3.10) is the steady state form of (5.3.13) — see App. D.

ARMA Processes. State-space models for ARMA processes (those with $m > 0$ in (5.3.1)) can be obtained in several different ways. Here is one so-called *controller* canonical form (see, *e.g.*, Kailath (1980), Ch. 2) which we construct for the time-invariant ARMA model

$$\mathbf{y}_{i+1} = a_0\mathbf{y}_i + \ldots + a_{n-1}\mathbf{y}_{i-n+1} + b_0\mathbf{u}_i + \ldots + b_{n-1}\mathbf{u}_{i-n+1}.$$

We begin by choosing the state vector as $\mathbf{x}_i = \text{col}\{\boldsymbol{\xi}_i, \boldsymbol{\xi}_{i-1}, \ldots, \boldsymbol{\xi}_{i-n+1}\}$, where $\boldsymbol{\xi}_i$ satisfies the difference equation

$$\boldsymbol{\xi}_{i+1} = a_0\boldsymbol{\xi}_i + \ldots + a_{n-1}\boldsymbol{\xi}_{i-n+1} + \mathbf{u}_i. \tag{5.3.14}$$

That is, $\boldsymbol{\xi}_{i+1}$ is the output of an AR model driven by $\mathbf{u}_i$. We already know that this model can be written in state-space form as follows:

$$\begin{bmatrix} \boldsymbol{\xi}_{i+1} \\ \boldsymbol{\xi}_i \\ \boldsymbol{\xi}_{i-1} \\ \vdots \\ \boldsymbol{\xi}_{i-n+2} \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & \ldots & a_{n-1} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \quad 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_i \\ \boldsymbol{\xi}_{i-1} \\ \boldsymbol{\xi}_{i-2} \\ \vdots \\ \boldsymbol{\xi}_{i-n+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \mathbf{u}_i. \tag{5.3.15}$$

Now, by linearity and time-invariance, the output of the ARMA model may readily be written as a linear combination of the outputs of the AR model (5.3.14) as follows:

$$\mathbf{y}_{i+1} = b_0\boldsymbol{\xi}_{i+1} + \ldots + b_{n-1}\boldsymbol{\xi}_{i-n+2}. \tag{5.3.16}$$

Combining (5.3.15) and (5.3.16), we arrive at the state-space model

$$\begin{cases} \mathbf{x}_{i+1} = \begin{bmatrix} a_0 & a_1 & \ldots & a_{n-1} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \quad 0 \end{bmatrix} \mathbf{x}_i + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \mathbf{u}_i, \\ \mathbf{y}_i = \begin{bmatrix} b_0 & b_1 & \ldots & b_{n-1} \end{bmatrix} \mathbf{x}_i. \end{cases}$$

There are of course many other methods of obtaining state-space models, either from descriptions of the form (5.3.1) or directly from the physical system (by using Newton's laws, Kirchoff's laws, etc). This material is widely available in textbooks on linear systems, see, *e.g.*, Kailath (1980) and Antsaklis and Michel (1997). So we shall not pursue the issue any further here.

An important reason for the convenience of state-space models is that the state-vector process $\{\mathbf{x}_i\}$ is just a vector form of the scalar exponentially correlated process, and its nonstationary and time-variant generalizations, studied in the beginning of this chapter. Many of the striking results obtained there (*e.g.*, the simple construction of the innovations) go over to the vector case. We now introduce what we shall call the

standard state-space model. It will be the starting point of most of our future discussions. A closer study of the properties of such state-vector processes will be pursued in Sec. 5.4.

5.3.4 The Standard State-Space Model

The standard state-space model that we shall most often employ in this book takes the form

$$\begin{cases} \mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i, & i \geq 0, \\ \mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{5.3.17}$$

where $F_i \in \mathbb{C}^{n\times n}$, $G_i \in \mathbb{C}^{n\times m}$, and $H_i \in \mathbb{C}^{p\times n}$ are known matrices, while $\{\mathbf{u}_i\}$, $\{\mathbf{v}_i\}$, and $\{\mathbf{x}_0\}$ are variables obeying

$$\left\langle \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_j \\ \mathbf{v}_j \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} \Pi_0 & 0 & 0 & 0 \\ 0 & Q_i\delta_{ij} & S_i\delta_{ij} & 0 \\ 0 & S_i^*\delta_{ij} & R_i\delta_{ij} & 0 \end{bmatrix}. \tag{5.3.18}$$

Here we have made explicit our standing assumption that all variables have zero-mean. Moreover, for generality, the driving sequences $\{\mathbf{u}_i, \mathbf{v}_i\}$ are allowed to be distinct. In many applications, the $\{\mathbf{v}_i\}$ can be interpreted as output disturbance or measurement noise, while the $\{\mathbf{u}_i\}$ can be interpreted as process noise.

We collect here various important orthogonality and covariance properties of the model (5.3.17)–(5.3.18). We begin with some simple, but very useful properties, that the reader is advised to understand thoroughly. Note especially that no explicit calculations (*i.e.*, using explicit formulas for $\mathbf{x}_i$ in terms of $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}\}$) are needed in the proofs.

Lemma 5.3.1 (Basic Orthogonality Properties) *In the standard state-space model (5.3.17)–(5.3.18) we can assert that*

(i) For $i \geq j$, $\langle \mathbf{u}_i, \mathbf{x}_j \rangle = 0$ *and* $\langle \mathbf{v}_i, \mathbf{x}_j \rangle = 0$.

(ii) For $i > j$, $\langle \mathbf{u}_i, \mathbf{y}_j \rangle = 0$ *and* $\langle \mathbf{v}_i, \mathbf{y}_j \rangle = 0$.

(iii) For $i = j$, $\langle \mathbf{u}_i, \mathbf{y}_i \rangle = S_i$ *and* $\langle \mathbf{v}_i, \mathbf{y}_i \rangle = R_i$.

(iv) When the $\{F_i\}$ *are nonsingular,* $\langle \mathbf{u}_i, \mathbf{x}_0 \rangle = 0$ *if, and only if,* $\langle \mathbf{u}_i, \mathbf{x}_i \rangle = 0$, $i \geq 0$.

■

Proof: The arguments are very similar, so not all of them will be given. We should mention that (iv) will be proved later (Lemma 5.4.3). Here is a typical proof: from the state equation, we note that $\mathbf{x}_j \in \mathcal{L}\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{j-1}\}$ and that $\mathbf{u}_i$ is orthogonal to $\mathcal{L}\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{j-1}\}$ for $i \geq j$. Hence the first of the equations in (i).

Similarly, for $i \geq j$,

$$\langle \mathbf{v}_i, \mathbf{y}_j \rangle = \langle \mathbf{v}_i, \mathbf{v}_j \rangle + \langle \mathbf{v}_i, \mathbf{x}_j \rangle H_j^* = R_i\delta_{ij} + 0,$$

because $\mathbf{x}_j \in \mathcal{L}\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{j-1}\}$ and $\mathbf{v}_i$ is orthogonal to $\mathcal{L}\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{j-1}\}$ for $i \geq j$.

◆

Lemma 5.3.2 (Covariance Expressions) *Consider the standard model (5.3.17)–(5.3.18) and denote the state covariance matrix by* $\langle \mathbf{x}_i, \mathbf{x}_i \rangle = \|\mathbf{x}_i\|^2 = \Pi_i$. *Then* Π_i *satisfies the recursion*

$$\Pi_{i+1} = F_i \Pi_i F_i^* + G_i Q_i G_i^*, \quad i \geq 0.$$

Moreover, if we define the state transition matrix

$$\Phi(i, j) = F_{i-1} F_{i-2} \ldots F_j, \quad i > j, \quad \Phi(i, i) = I, \tag{5.3.19}$$

then the covariances of the state variables can be computed via

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \begin{cases} \Phi(i, j)\Pi_j & i \geq j, \\ \Pi_i \Phi^*(j, i) & i \leq j, \end{cases}$$

and the covariances of the output process $\{\mathbf{y}_i\}$ *are given by*

$$\langle \mathbf{y}_i, \mathbf{y}_j \rangle = \begin{cases} H_i \Phi(i, j+1) N_j & i > j, \\ R_i + H_i \Pi_i H_i^* & i = j, \\ N_i^* \Phi^*(j, i+1) H_j^* & i < j, \end{cases} \tag{5.3.20}$$

where $N_i = F_i \Pi_i H_i^* + G_i S_i$. ∎

Proof: The recursion for Π_i follows by computing the variance of both sides of the state equation and using the property $\langle \mathbf{x}_i, \mathbf{u}_i \rangle = 0$.

To compute the covariances of the state variables, note that from the state equation we can write

$$\mathbf{x}_i = \Phi(i, j)\mathbf{x}_j + \text{ some linear combination of } \mathcal{L}\{\mathbf{u}_j, \ldots, \mathbf{u}_{i-1}\}, \quad i \geq j.$$

Therefore, $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \Phi(i, j)\langle \mathbf{x}_j, \mathbf{x}_j \rangle + 0$ for $i \geq j$. Next note that for $i < j$, $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \langle \mathbf{x}_j, \mathbf{x}_i \rangle^* = \Pi_i \Phi^*(j, i)$. Finally, note that

$$\langle \mathbf{y}_i, \mathbf{y}_j \rangle = H_i \langle \mathbf{x}_i, \mathbf{x}_j \rangle H_i^* + H_i \langle \mathbf{x}_i, \mathbf{v}_j \rangle + \langle \mathbf{v}_i, \mathbf{x}_j \rangle H_j^* + \langle \mathbf{v}_i, \mathbf{v}_j \rangle.$$

Now for $i > j$, the last two inner products on the right-hand side are zero, while the second inner product is $\langle \mathbf{x}_i, \mathbf{v}_j \rangle = \Phi(i, j+1) G_j \langle \mathbf{u}_i, \mathbf{v}_j \rangle = \Phi(i, j+1) G_j S_j$, and the first $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \Phi(i, j)\Pi_j = \Phi(i, j+1) F_j \Pi_j$. Collecting these results gives the first of the desired expressions for $\langle \mathbf{y}_i, \mathbf{y}_j \rangle$. The remaining two (for $i = j$ and $i < j$) follow in a similar fashion. ◆

The covariances (5.3.20) completely specify the entries of the Gramian matrix R_y of the output process $\{\mathbf{y}_i\}$ that is generated by the standard state-space model. For example, assuming three observation vectors $\{\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2\}$, we obtain

$$R_y = \begin{bmatrix} R_0 + H_0 \Pi_0 H_0^* & N_0^* H_1^* & N_0^* F_1^* H_2^* \\ H_1 N_0 & R_1 + H_1 \Pi_1 H_1^* & N_1^* H_2^* \\ H_2 F_1 N_0 & H_2 N_1 & R_2 + H_2 \Pi_2 H_2^* \end{bmatrix}. \tag{5.3.21}$$

5.3.5 Examples of Other State-Space Models

There are of course many other forms of state-space models, arising in different contexts. For example, in Sec. 5.3, while discussing state-space models for AR and ARMA processes we were led to a state-space representation of the form

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i, & i \geq 0, \\ \mathbf{y}_i = H_i \mathbf{x}_i, \end{cases} \tag{5.3.22}$$

with zero-mean random variables $\{\mathbf{x}_0, \mathbf{u}_i\}$ that satisfy

$$\left\langle \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_i \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_j \end{bmatrix} \right\rangle = \begin{bmatrix} \Pi_0 & 0 \\ 0 & Q_i \delta_{ij} \end{bmatrix}.$$

This is a special case of the standard model (5.3.17)–(5.3.18) with $R_i = 0$.

Another example of a modified state-space model is the following:

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_{i+1}, & i \geq 0, \\ \mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{5.3.23}$$

with random variables $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i\}$ that satisfy

$$\left\langle \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_j \\ \mathbf{v}_j \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} \Pi_0 & 0 & 0 & 0 \\ 0 & Q_i \delta_{ij} & S_i \delta_{ij} & 0 \\ 0 & S_i^* \delta_{ij} & R_i \delta_{ij} & 0 \end{bmatrix}.$$

Observe now that the time indices of the driving processes $\{\mathbf{u}_{i+1}, \mathbf{v}_i\}$ differ by one unit-time delay. Prob. 5.6 exhibits the appropriate orthogonality conditions for such models. It may also happen that the state equation is driven by a combination of $\mathbf{u}_i$ and $\mathbf{u}_{i+1}$, say (see Prob. 9.10)

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i + B_i \mathbf{u}_{i+1}, & i \geq 0, \\ \mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i. \end{cases} \tag{5.3.24}$$

Yet another example of a modified state-space model is the case of *nonwhite* (*i.e.*, colored) measurement noise such as

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i, & i \geq 0, \\ \mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{n}_i, \\ \mathbf{n}_{i+1} = A_i \mathbf{n}_i + \mathbf{v}_i, \end{cases} \tag{5.3.25}$$

where

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \\ \mathbf{n}_0 \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \\ \mathbf{n}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & 0 & 0 & 0 \\ 0 & R_i \delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \\ 0 & 0 & 0 & \Pi_0^{(n)} \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{5.3.26}$$

Here, the process $\{\mathbf{n}_i\}$ that drives the output equation is itself generated by a state equation. Prob. 5.8 provides more details and exhibits certain orthogonality conditions, which are used later in Prob. 9.15 to determine the innovations of the output process $\{\mathbf{y}_i\}$.

5.4 WIDE-SENSE MARKOV PROCESSES[2]

Recall that our discussion of state-space models was motivated in part by the observation that the exponentially correlated process (and in fact all *first-order* autoregressive processes) $\{\mathbf{y}_i\}$ had the striking property that

$$\begin{aligned}\hat{\mathbf{y}}_{i|i-1} &= \text{the l.l.m.s.e. of } \mathbf{y}_i \text{ given } \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\},\\ &= \text{the l.l.m.s.e. of } \mathbf{y}_i \text{ given } \mathcal{L}\{\mathbf{y}_{i-1}\}. \qquad (5.4.1)\end{aligned}$$

This result is reminiscent of the fact that for "strict-sense" Markov stochastic processes, we have the following identity for the conditional probability density functions,

$$f_{\mathbf{y}_i|\mathbf{y}_j,\mathbf{y}_k}(y_i|y_j, y_k) = f_{\mathbf{y}_i|\mathbf{y}_j}(y_i|y_j) \text{ if } i > j > k.$$

For linear least-squares estimation, we only use the first- and second-order statistical information, and so we may say, following Doob (1953), that property (5.4.1) characterizes a so-called "wide-sense" Markov process. This is a very useful concept, which underlies the proper determination of state-space models for second-order processes.

Definition 5.4.1. (Wide-Sense Markov Process) *For any integer k, consider indices $i_0, i_1, \ldots, i_k$ ordered so that $N \geq i_k > i_{k-1} > \ldots > i_0 \geq 0$. Then a (possibly vector-valued) process $\{\mathbf{x}_i, i = 0, 1, \ldots, N\}$ is wide-sense Markov (WSM) if the l.l.m.s.e. of $\mathbf{x}_{i_k}$ given $\mathcal{L}\{\mathbf{x}_{i_0}, \ldots, \mathbf{x}_{i_{k-1}}\}$ equals the l.l.m.s.e. of $\mathbf{x}_{i_k}$ given $\mathcal{L}\{\mathbf{x}_{i_{k-1}}\}$. That is, the l.l.m.s. estimator depends only on the most recent observation.* ◆

There is a simple covariance test for a process being WSM.

Lemma 5.4.1 (Covariance Test for WSM Processes) *Assume that $\|\mathbf{x}_i\|^2 > 0$, $i \geq 0$. Then the process $\{\mathbf{x}_i, i \geq 0\}$ is* WSM *if, and only if, for any $i > j > k$,*

$$\langle \mathbf{x}_i, \mathbf{x}_k \rangle = \langle \mathbf{x}_i, \mathbf{x}_j \rangle \|\mathbf{x}_j\|^{-2} \langle \mathbf{x}_j, \mathbf{x}_k \rangle. \qquad (5.4.2)$$

■

Proof: By definition, $\{\mathbf{x}_i\}$ is WSM if, and only if, $\tilde{\mathbf{x}}_{i|\mathbf{x}_j} \perp \mathbf{x}_k$ for $i > j > k$. This is because, by the WSM property, $\hat{\mathbf{x}}_{i|\mathbf{x}_k,\mathbf{x}_j} = \hat{\mathbf{x}}_{i|\mathbf{x}_j}$, for any $i > j > k$. Hence,

$$\begin{aligned}0 = \langle \mathbf{x}_i - \hat{\mathbf{x}}_{i|\mathbf{x}_j}, \mathbf{x}_k \rangle &= \langle \mathbf{x}_i - \langle \mathbf{x}_i, \mathbf{x}_j \rangle \|\mathbf{x}_j\|^{-2} \mathbf{x}_j, \mathbf{x}_k \rangle,\\ &= \langle \mathbf{x}_i, \mathbf{x}_k \rangle - \langle \mathbf{x}_i, \mathbf{x}_j \rangle \|\mathbf{x}_j\|^{-2} \langle \mathbf{x}_j, \mathbf{x}_k \rangle,\end{aligned}$$

as desired. ◆

[2]This section can be omitted on a first reading; the material will not be used till Sec. 9.8.

5.4.1 Forwards Markovian Models

Note the assumption in the above lemma that $\|\mathbf{x}_i\|^2 = \Pi_i > 0$. This is a very reasonable assumption, but it is not essential. One way of seeing this is to give an alternative characterization of the WSM property in terms of so-called Markovian state-space models. For this purpose, we return to Def. 5.4.1 and recall that the simple exponentially correlated process studied in Secs. 4.4 and 5.1 has precisely the WSM property that the l.l.m.s. estimator depends only on the most recent observation. Now knowing the result for the first-order (scalar) case, we may expect that a vector process with a state-space model of the form

$$\mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i, \quad i \geq 0, \qquad (5.4.3)$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & 0 \\ 0 & \Pi_0 \end{bmatrix}, \qquad (5.4.4)$$

will be WSM. Indeed, note that since $\mathbf{x}_i \in \mathcal{L}\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}\}$, it follows from the assumptions (5.4.3)–(5.4.4) that

$$\mathbf{u}_i \perp \mathbf{x}_j, \quad i \geq j \geq 0. \qquad (5.4.5)$$

We should emphasize that it is the assumption $\langle \mathbf{u}_j, \mathbf{x}_0 \rangle = 0$ for $j \geq 0$ that enables the conclusion (5.4.5).

Thus projecting (5.4.3) onto $\mathcal{L}\{\mathbf{x}_0, \ldots, \mathbf{x}_i\}$, and using the linearity property of the projection, yields

$$\begin{aligned}\hat{\mathbf{x}}_{i+1|\{\mathbf{x}_0,\ldots,\mathbf{x}_i\}} &= F_i \hat{\mathbf{x}}_{i|\{\mathbf{x}_0,\ldots,\mathbf{x}_i\}} + G_i \hat{\mathbf{u}}_{i|\{\mathbf{x}_0,\ldots,\mathbf{x}_i\}},\\ &= F_i \mathbf{x}_i + 0 = F_i \hat{\mathbf{x}}_{i|\mathbf{x}_i} = \hat{\mathbf{x}}_{i+1|\mathbf{x}_i},\end{aligned}$$

thus showing that *a process $\{\mathbf{x}_i, i \geq 0\}$ with state-space model of the form (5.4.3)–(5.4.4) is a wide-sense Markov process.*

It is a very interesting fact that the converse also holds. To clarify this, assume we are given a wide-sense Markov process $\{\mathbf{x}_i, i \geq 0\}$. Then its innovations process can be found as

$$\mathbf{e}_0 = \mathbf{x}_0, \quad \mathbf{e}_{i+1} = \mathbf{x}_{i+1} - K_{o,i} \mathbf{x}_i, \quad i \geq 0,$$

where $K_{o,i}$ is *any* solution of the normal equation $K_{o,i} \|\mathbf{x}_i\|^2 = \langle \mathbf{x}_{i+1}, \mathbf{x}_i \rangle$. The reason for saying *any solution* is that, as shown in Sec. 3.2, a solution to the above equation exists even when $\|\mathbf{x}_i\|^2$ is singular. Now define $\mathbf{u}_i = \mathbf{e}_{i+1}$ and $F_i = K_{o,i}$ for $i \geq 0$. Then we can write

$$\mathbf{x}_{i+1} = F_i \mathbf{x}_i + \mathbf{u}_i, \quad i \geq 0 \qquad (5.4.6)$$

and $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \langle \mathbf{e}_{i+1}, \mathbf{e}_{j+1} \rangle = 0$, $i \neq j$, by the orthogonality of the innovations. Moreover,

$$\|\mathbf{u}_i\|^2 = \|\mathbf{e}_{i+1}\|^2 = \Pi_{i+1} - F_i \Pi_i F_i^* \triangleq Q_i,$$

where we have defined the state covariance matrix as $\Pi_i \triangleq \|\mathbf{x}_i\|^2$. Finally, notice that the state-space model so defined satisfies the critical property that

$$\mathbf{u}_i = \mathbf{e}_{i+1} \perp \mathbf{e}_0 = \mathbf{x}_0, \quad i \geq 0.$$

Hence,

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & 0 \\ 0 & \Pi_0 \end{bmatrix}, \tag{5.4.7}$$

and the process $\{\mathbf{x}_i\}$ has the desired state-space representation (5.4.3)–(5.4.4). We summarize the above discussion in the following theorem, which gives a stronger result than Lemma 5.4.1 (because it is not assumed that $\|\mathbf{x}_i\|^2 > 0$).

Theorem 5.4.1 (Characterization of WSM Processes) *A process $\{\mathbf{x}_i, i \geq 0\}$ is* WSM *if, and only if, it has a so-called forwards Markovian state-space representation of the form (5.4.3)–(5.4.4).* ■

Proof: We have already seen that if $\{\mathbf{x}_i, i \geq 0\}$ has a state-space representation of the form (5.4.3)–(5.4.4), then it is WSM. Conversely, if $\{\mathbf{x}_i, i \geq 0\}$ is WSM, then the discussion prior to the statement of the theorem establishes that $\{\mathbf{x}_i\}$ admits a state-space representation of the form (5.4.6)–(5.4.7). ◆

The assumption that $\{\mathbf{u}_i, i \geq 0\}$ is uncorrelated with the initial condition $\mathbf{x}_0$ is very important in simplifying the analyses to be carried out in this book. To emphasize this fact we have called state-space models with this particular property, *Markovian*. The point is that one could of course have state-space models that do not have this property.

An immediate example arises when we consider reverse-time (or backwards) state-space models.

5.4.2 Backwards Markovian Models

Consider a WSM process with a forwards *Markovian* representation of the form (5.4.3)–(5.4.4). Assume further that the F_i are invertible. We can then reverse time and write

$$\mathbf{x}_i = F_i^{-1}\mathbf{x}_{i+1} - F_i^{-1}G_i\mathbf{u}_i, \quad i \leq N. \tag{5.4.8}$$

Now though $\{\mathbf{u}_i\}$ is still white, we do not have the critical backward time property $\langle \mathbf{u}_i, \mathbf{x}_{N+1} \rangle = 0, i \leq N$. Therefore (5.4.8) is *not* a backwards-time Markovian model.

On the other hand, the process $\{\mathbf{x}_i, 0 \leq i \leq N\}$ is still a WSM process: the process is not changed because we examine the random variables in a different order.

Lemma 5.4.2 (Independence from Time Direction) *Let $\{\mathbf{x}_i, i \geq 0\}$ be a* WSM *process. Then for $i > j > k$, the l.l.m.s.e. of $\mathbf{x}_k$ given $\mathcal{L}\{\mathbf{x}_i, \mathbf{x}_j\}$ is equal to the l.l.m.s.e. of $\mathbf{x}_k$ given $\mathcal{L}\{\mathbf{x}_j\}$.* ■

Proof: A simple but useful computation for the reader. ◆

The point is that when we do not have the property $\langle \mathbf{u}_i, \mathbf{x}_0 \rangle = 0, i \geq 0$ (or $\langle \mathbf{u}_i, \mathbf{x}_{N+1} \rangle = 0, i \leq N$ for backwards models), the WSM property of the $\{\mathbf{x}_i\}$ is less evident and will have to be verified by further calculation. However, the above result

suggests that, even though the simple model (5.4.8) did not work, we should be able to find a *backwards-time* Markovian state-space model for a WSM process $\mathbf{x}_i$. This is indeed possible (see Sec. 5.4.3), but let us first formally define what we mean by a backwards Markovian model.

Definition 5.4.2. (Backwards Markovian Models) *A representation of a process $\{\mathbf{x}_i\}$ in the following form is called a backwards Markovian state-space model:*

$$\mathbf{x}_i = F_{i+1}^b \mathbf{x}_{i+1} + G_{i+1}^b \mathbf{u}_{i+1}^b, \quad N \geq i \geq 0, \tag{5.4.9}$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i^b \\ \mathbf{x}_{N+1} \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j^b \\ \mathbf{x}_{N+1} \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i^b \delta_{ij} & 0 \\ 0 & \Pi_{N+1} \end{bmatrix}. \tag{5.4.10}$$

◆

Note that time flows in the reverse direction in (5.4.9): starting from $i = N$ and moving down to $i = 0$. In this respect, the state at time $(N+1)$ is now considered the initial state. Hence, note that the input process $\mathbf{u}_i^b$ is now required to be uncorrelated with $\mathbf{x}_{N+1}$.

We have shown earlier how to construct a forwards Markovian representation (5.4.6) for a WSM process $\mathbf{x}_i$. We now follow a similar argument in order to derive a backwards Markovian model for the same process $\mathbf{x}_i$.

For this purpose, we now define the *backwards* innovations of $\{\mathbf{x}_i\}$ as $\mathbf{e}_{N+1}^b = \mathbf{x}_{N+1}$, and, for $i \geq 0$,

$$\begin{aligned} \mathbf{e}_i^b &= \mathbf{x}_i - \hat{\mathbf{x}}_{i|\{\mathbf{x}_{i+1},\ldots,\mathbf{x}_{N+1}\}} = \mathbf{x}_i - \hat{\mathbf{x}}_{i|\mathbf{x}_{i+1}} \quad \text{(using Lemma 5.4.2)}, \\ &= \mathbf{x}_i - K_{o,i+1}^b \mathbf{x}_{i+1}, \end{aligned}$$

where $K_{o,i+1}^b$ is any solution to the equation $K_{o,i+1}^b \|\mathbf{x}_{i+1}\|^2 = \langle \mathbf{x}_i, \mathbf{x}_{i+1} \rangle$. Now if we define $F_{i+1}^b = K_{o,i+1}^b$, $G_{i+1}^b = I$, and $\mathbf{u}_{i+1}^b = \mathbf{e}_i^b$, $i \leq N$, we can write

$$\mathbf{x}_i = F_{i+1}^b \mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b, \quad i \geq 0.$$

Moreover, the whiteness of the $\{\mathbf{u}_i^b\}$, and the orthogonality of $\mathbf{x}_{N+1}$ with $\{\mathbf{u}_i^b\}$, both follow from the whiteness of the backwards innovations $\{\mathbf{e}_i^b\}$. Therefore, all that remains to be shown is the expression for Q_{i+1}^b,

$$\begin{aligned} Q_{i+1}^b \triangleq \|\mathbf{u}_{i+1}^b\|^2 = \|\mathbf{e}_i^b\|^2 &= \langle \mathbf{x}_i - F_{i+1}^b \mathbf{x}_{i+1}, \mathbf{x}_i - F_{i+1}^b \mathbf{x}_{i+1} \rangle \\ &= \Pi_i - \underbrace{\langle \mathbf{x}_i, \mathbf{x}_{i+1} \rangle}_{F_{i+1}^b \Pi_{i+1}} F_{i+1}^{*b} - F_{i+1}^b \underbrace{\langle \mathbf{x}_{i+1}, \mathbf{x}_i \rangle}_{\Pi_{i+1} F_{i+1}^{*b}} + F_{i+1}^b \Pi_{i+1} F_{i+1}^{*b} \\ &= \Pi_i - F_{i+1}^b \Pi_{i+1} F_{i+1}^{*b}, \end{aligned}$$

as desired.

We summarize the above discussion, as well as the one relevant to the forwards model, in the following theorem.

Theorem 5.4.2 (Forwards and Backwards Markovian Models) *Consider a* WSM *process* $\{\mathbf{x}_i,\ 0 \le i \le N\}$ *and denote its covariance matrix by* $\Pi_i = \|\mathbf{x}_i\|^2$. *Then* $\{\mathbf{x}_i\}$ *admits both forwards and backwards Markovian state-space models that are constructed as follows. Let* F_i *and* F_{i+1}^b *be any solutions to the normal equations*

$$F_i\|\mathbf{x}_i\|^2 = \langle \mathbf{x}_{i+1}, \mathbf{x}_i\rangle, \qquad F_{i+1}^b\|\mathbf{x}_{i+1}\|^2 = \langle \mathbf{x}_i, \mathbf{x}_{i+1}\rangle,$$

and introduce the covariance matrices

$$Q_i = \Pi_{i+1} - F_i\Pi_i F_i^*, \qquad Q_{i+1}^b = \Pi_i - F_{i+1}^b\Pi_{i+1}F_{i+1}^{b*}.$$

Then $\mathbf{x}_i$ *has the following forwards and backwards Markovian models:*

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + \mathbf{u}_i, \qquad \mathbf{x}_i = F_{i+1}^b\mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b,$$

with

$$\left\langle \begin{bmatrix}\mathbf{u}_i\\ \mathbf{x}_0\end{bmatrix}, \begin{bmatrix}\mathbf{u}_j\\ \mathbf{x}_0\end{bmatrix}\right\rangle = \begin{bmatrix}Q_i\delta_{ij} & 0\\ 0 & \Pi_0\end{bmatrix},$$

and

$$\left\langle \begin{bmatrix}\mathbf{u}_i^b\\ \mathbf{x}_{N+1}\end{bmatrix}, \begin{bmatrix}\mathbf{u}_j^b\\ \mathbf{x}_{N+1}\end{bmatrix}\right\rangle = \begin{bmatrix}Q_i^b\delta_{ij} & 0\\ 0 & \Pi_{N+1}\end{bmatrix}.$$

■

We also note here a useful fact available when the $\{F_i\}$ are invertible.

Lemma 5.4.3 (Markovian Models for Invertible F_i) *Consider the state-space equation* $\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i$ *for* $i \ge 0$. *When the* $\{F_i\}$ *are invertible, the following two conditions are equivalent for this model to be a forwards Markovian model:*

(i) $\langle \mathbf{u}_i, \mathbf{x}_0\rangle = 0,\ \ \langle \mathbf{u}_i, \mathbf{u}_j\rangle = Q_i\delta_{ij},\ \ i, j \ge 0.$

(ii) $\langle \mathbf{u}_i, \mathbf{x}_i\rangle = 0,\ \ \langle \mathbf{u}_i, \mathbf{u}_j\rangle = Q_i\delta_{ij},\ \ i, j \ge 0.$

Likewise, consider the state-space equation $\mathbf{x}_i = F_{i+1}^b\mathbf{x}_{i+1} + G_{i+1}^b\mathbf{u}_{i+1}^b$, *for* $i \le N$. *When the* $\{F_i^b\}$ *are invertible, the following two conditions are equivalent for this model to be a backwards Markovian model:*

(i) $\langle \mathbf{u}_i^b, \mathbf{x}_{N+1}\rangle = 0,\ \ \langle \mathbf{u}_i^b, \mathbf{u}_j^b\rangle = Q_i^b\delta_{ij},\ \ i, j \le N+1.$

(ii) $\langle \mathbf{u}_i^b, \mathbf{x}_i\rangle = 0,\ \ \langle \mathbf{u}_i^b, \mathbf{u}_j^b\rangle = Q_i^b\delta_{ij},\ \ i, j \le N+1.$

■

Proof: We shall prove the result for forwards Markovian models. The proof for backwards representations is similar. Proving the direction (i) $\Rightarrow$ (ii) was done before and follows easily by noting that $\mathbf{x}_i \in \mathcal{L}\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}\}$. For the converse, note that we can use the state-space equations to write

$$\mathbf{x}_i = F_{i-1}\ldots F_0\mathbf{x}_0 + \text{a linear combination of } \{\mathbf{u}_0, \ldots, \mathbf{u}_{i-1}\}.$$

Therefore, it is straightforward to see that $\langle \mathbf{u}_i, \mathbf{x}_i\rangle = \langle \mathbf{u}_i, \mathbf{x}_0\rangle F_0^*\ldots F_{i-1}^*$, so that if the $\{F_i\}$ are invertible, $\langle \mathbf{u}_i, \mathbf{x}_i\rangle = 0$ implies $\langle \mathbf{u}_i, \mathbf{x}_0\rangle = 0$. ◆

5.4.3 Backwards Models from Forwards Models

It is sometimes useful to be able to obtain a backwards Markovian model from such a forwards model. Recall that we unsuccessfully attempted this earlier in (5.4.8) by simply inverting the matrix F_i. But now we know more about the properties of such models and can successfully solve the problem, and in fact, in a couple of useful ways.

First, note that, by using Thm. 5.4.2, we can write

$$F_{i+1}^b\|\mathbf{x}_{i+1}\|^2 = \langle \mathbf{x}_i, \mathbf{x}_{i+1}\rangle = \langle \mathbf{x}_{i+1}, \mathbf{x}_i\rangle^* = (F_i\|\mathbf{x}_i\|^2)^* = \|\mathbf{x}_i\|^2F_i^*.$$

Therefore if the matrices $\|\mathbf{x}_i\|^2 \triangleq \Pi_i$ are invertible, we can obtain F_{i+1}^b from F_i as $F_{i+1}^b = \Pi_iF_i^*\Pi_{i+1}^{-1}$. Then it is not hard to find Q_{i+1}^b. This is shown in the following lemma without assuming the invertibility of the $\{\Pi_i\}$.

Lemma 5.4.4 (Backwards Models from Forwards Models) *Given a forwards Markovian model of the form*

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i, \quad 0 \le i \le N, \tag{5.4.11}$$

with zero-mean uncorrelated random variables $\{\mathbf{u}_i, \mathbf{x}_0\}$ *and such that* $\langle \mathbf{u}_i, \mathbf{u}_j\rangle = Q_i\delta_{ij}$, $\langle \mathbf{x}_0, \mathbf{x}_0\rangle = \Pi_0$, *we can obtain a backwards Markovian model of the process* $\{\mathbf{x}_i\}$ *as follows:*

$$\mathbf{x}_i = F_{i+1}^b\mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b, \quad 0 \le i \le N,$$

with

$$\left\langle \begin{bmatrix}\mathbf{u}_i^b\\ \mathbf{x}_{N+1}\end{bmatrix}, \begin{bmatrix}\mathbf{u}_j^b\\ \mathbf{x}_{N+1}\end{bmatrix}\right\rangle = \begin{bmatrix}Q_i^b\delta_{ij} & 0\\ 0 & \Pi_{N+1}\end{bmatrix},$$

where F_{i+1}^b *is any solution of the equation* $F_{i+1}^b\Pi_{i+1} = \Pi_iF_i^*$ *and* $Q_{i+1}^b = \Pi_i - F_{i+1}^b\Pi_{i+1}F_{i+1}^{b*}$. *Here,* $\Pi_i = \|\mathbf{x}_i\|^2$ *and satisfies* $\Pi_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^*$. ■

Proof: All we need to show is that the equation $F_{i+1}^b\Pi_{i+1} = \Pi_iF_i^*$ always has a solution. But this is the case, since from (5.4.11) the above equation is just the normal equation for estimating $\mathbf{x}_i$ given $\mathbf{x}_{i+1}$,

$$F_{i+1}^b\|\mathbf{x}_{i+1}\|^2 = \langle \mathbf{x}_i, \mathbf{x}_{i+1}\rangle = \Pi_iF_i^*.$$

The expression for Q_{i+1}^b now follows immediately. ◆

As noted earlier, the backwards Markovian model of the above lemma can be made more explicit if we assume that the $\{\Pi_i\}_{i=0}^N$ are nonsingular, in which case we can write

$$F_{i+1}^b = \Pi_iF_i^*\Pi_{i+1}^{-1}, \qquad Q_{i+1}^b = \Pi_i - \Pi_iF_i^*\Pi_{i+1}^{-1}F_i\Pi_i.$$

A necessary and sufficient condition for this is to require that (*cf.* Prob. 5.15)

$$\Pi_0 > 0 \text{ and } \begin{bmatrix}F_i & G_iQ_i^{\frac{1}{2}}\end{bmatrix} \text{ have full rank for all } i = 0, 1, \ldots, N. \tag{5.4.12}$$

Perhaps more useful sufficient conditions are either (see also Prob. 5.15)

(a) $\Pi_0 > 0$ and F_i nonsingular for all $i = 0, 1, \ldots, N$, or

(b) $\Pi_0 > 0$ and the pair $\{F_i, G_i Q_i^{\frac{1}{2}}\}$ controllable (see App. C) for all $i = 0, 1, \ldots, N$.

Backwards Models via Time Reversal. A second solution to the problem arises by returning to the non-Markovian reversed time model (5.4.8), assuming F_i invertible,

$$\mathbf{x}_i = F_i^{-1}\mathbf{x}_{i+1} - F_i^{-1}G_i\mathbf{u}_i, \quad 0 \le i \le N, \tag{5.4.13}$$

and seeing if it cannot be somehow converted to a Markovian model. In fact, this can be done, as we shall show by using the method of Verghese and Kailath (1979). The reason the reversed-time model (5.4.13) is not (backwards) Markovian is because (see Lemma 5.4.3 part (ii)) $\langle \mathbf{u}_i, \mathbf{x}_{i+1}\rangle \neq 0$. However, we can obviously make it so by replacing $\mathbf{u}_i$ with

$$\begin{aligned}\tilde{\mathbf{u}}_i &= \mathbf{u}_i - \hat{\mathbf{u}}_{i|\mathbf{x}_{i+1}} = \mathbf{u}_i - \langle \mathbf{u}_i, \mathbf{x}_{i+1}\rangle \|\mathbf{x}_{i+1}\|^{-2}\mathbf{x}_{i+1}, \\ &= \mathbf{u}_i - Q_i G_i^* \Pi_{i+1}^{-1}\mathbf{x}_{i+1}. \end{aligned} \tag{5.4.14}$$

Substituting the above expression into the reversed-time state-space equation (5.4.13) yields

$$\begin{aligned}\mathbf{x}_i &= (F_i^{-1} - F_i^{-1}G_iQ_iG_i^*\Pi_{i+1}^{-1})\mathbf{x}_{i+1} - F_i^{-1}G_i\tilde{\mathbf{u}}_i, \quad i \le N, \\ &\triangleq F_{i+1}^b\mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b, \quad \text{say.}\end{aligned} \tag{5.4.15}$$

Now by our construction of $\mathbf{u}_i^b$, we have $\langle \mathbf{u}_i^b, \mathbf{x}_i\rangle = 0$; showing that $\mathbf{u}_i^b$ is white requires more work and is left as an exercise (see Prob. 5.12). Now these two properties imply that the model (5.4.15) is backwards Markovian, with

$$\langle \mathbf{u}_{i+1}^b, \mathbf{u}_{i+1}^b\rangle = Q_{i+1}^b = F_i^{-1}G_i(Q_i - Q_iG_i^*\Pi_{i+1}^{-1}G_iQ_i)G_i^*F_i^{*}.$$

We have used the same notation $\{F_i^b, Q_i^b\}$ as for the backwards Markovian model of Thm. 5.4.2 because they are, in fact, the same model:

$$\begin{aligned}F_{i+1}^b &= F_i^{-1} - F_i^{-1}G_iQ_iG_i^*\Pi_{i+1}^{-1} = F_i^{-1}\left[\Pi_{i+1} - (\Pi_{i+1} - F_i\Pi_iF_i^*)\Pi_{i+1}^{-1}\right], \\ &= \Pi_iF_i^*\Pi_{i+1}^{-1}, \quad \text{as before,}\end{aligned}$$

while

$$\begin{aligned}Q_{i+1}^b &= F_i^{-1}G_iQ_iG_i^*\left[-\Pi_{i+1}^{-1}(\Pi_{i+1} - F_i\Pi_iF_i^*)\right]F_i^{-*}, \\ &= F_i^{-1}G_iQ_iG_i^*\Pi_{i+1}^{-1}F_i\Pi_i, \\ &= F_i^{-1}(\Pi_{i+1} - F_i\Pi_iF_i^*)\Pi_{i+1}^{-1}F_i\Pi_i, \\ &= \Pi_i - \Pi_iF_i^*\Pi_{i+1}^{-1}F_i\Pi_i, \quad \text{as before.}\end{aligned}$$

In summary, we have the following result.

Lemma 5.4.5 (Backwards Model via Time-Reversal) *Consider the forwards Markovian state-space model*

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i, \quad i \ge 0,$$

with invertible F_i, $\Pi_0 > 0$, zero-mean uncorrelated random variables $\{\mathbf{u}_i, \mathbf{x}_0\}$, and such that $\langle \mathbf{u}_i, \mathbf{u}_j\rangle = Q_i\delta_{ij}$, $\langle \mathbf{x}_0, \mathbf{x}_0\rangle = \Pi_0$. Then the following is a backwards Markovian state-space model for $\{\mathbf{x}_i\}$,

$$\mathbf{x}_i = F_{i+1}^b\mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b,$$

with $F_{i+1}^b = \Pi_iF_i^*\Pi_{i+1}^{-1}$, $Q_{i+1}^b = \Pi_i - \Pi_iF_i^*\Pi_{i+1}^{-1}F_i\Pi_i$, and

$$\left\langle \begin{bmatrix}\mathbf{u}_i^b \\ \mathbf{x}_{N+1}\end{bmatrix}, \begin{bmatrix}\mathbf{u}_j^b \\ \mathbf{x}_{N+1}\end{bmatrix}\right\rangle = \begin{bmatrix}Q_i^b\delta_{ij} & 0 \\ 0 & \Pi_{N+1}\end{bmatrix}.$$

■

These results on backwards models will be useful in the study of smoothing problems (Ch. 10), as well as in the study of dual control and estimation problems.

5.4.4 Markovian Representations and the Standard Model

We now return to the standard state-space model (5.3.17)–(5.3.18) that was introduced earlier in Sec. 5.3.4. As shown in Sec. 5.4, the assumption $\langle \mathbf{u}_i, \mathbf{x}_0\rangle = 0$, $i \ge 0$ makes $\{\mathbf{x}_i\}$ a WSM process. Unfortunately, while $\{\mathbf{x}_i\}$ is WSM and, of course, so is $\{\mathbf{v}_i\}$, it turns out that $\{H_i\mathbf{x}_i\}$ is not, unless $H_i = I$, and even in that case, $\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i$ is not WSM. The problem is that the sum of WSM processes is generally not WSM, as can be readily checked from the definition. Processes such as $\{\mathbf{y}_i\}$ in (5.3.17) are sometimes called *projections* of Markov processes — see also Lemma 5.4.6 below.

Nevertheless, as we might expect, the fact that by using the state-space description, the output process $\{\mathbf{y}_i\}$ can be directly related to a WSM process, makes the state-space description very useful. For example, one useful consequence is the fact that given a forwards Markovian model, we can always obtain a backwards Markovian model (a result that will be used later in Sec. 9.8 while deriving the so-called backwards Kalman filter).

So consider again the standard state-space model (5.3.17)–(5.3.18) and assume, for simplicity, that $S_i = 0$. Using the result of Lemma 5.4.4, we can associate the following backwards Markovian model with the state-vector process $\{\mathbf{x}_i\}$,

$$\mathbf{x}_i = F_{i+1}^b\mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b, \quad 0 \le i \le N,$$

with

$$\left\langle \begin{bmatrix}\mathbf{u}_i^b \\ \mathbf{x}_{N+1}\end{bmatrix}, \begin{bmatrix}\mathbf{u}_j^b \\ \mathbf{x}_{N+1}\end{bmatrix}\right\rangle = \begin{bmatrix}Q_i^b\delta_{ij} & 0 \\ 0 & \Pi_{N+1}\end{bmatrix},$$

and where F_{i+1}^b is any solution to the equation $F_{i+1}^b\Pi_{i+1} = \Pi_iF_i^*$, with $Q_{i+1}^b = \Pi_i - F_{i+1}^b\Pi_{i+1}F_{i+1}^{b*}$.

This suggests that we can replace the standard state-space representation for $\{\mathbf{y}_i\}$ by the following so-called backwards state-space representation:

$$\begin{cases} \mathbf{x}_i &= F_{i+1}^b \mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b, \\ \mathbf{y}_i &= H_i \mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{5.4.16}$$

or, equivalently,

$$\begin{cases} \mathbf{x}_i &= F_{i+1}^b \mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b, \\ \mathbf{y}_i &= H_i F_{i+1}^b \mathbf{x}_{i+1} + H_i \mathbf{u}_{i+1}^b + \mathbf{v}_i \triangleq H_i F_{i+1}^b \mathbf{x}_{i+1} + \mathbf{v}_{i+1}^b, \end{cases} \tag{5.4.17}$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i^b \\ \mathbf{v}_i^b \\ \mathbf{x}_{N+1} \\ 1 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j^b \\ \mathbf{v}_j^b \\ \mathbf{x}_{N+1} \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i^b \delta_{ij} & Q_i^b H_{i-1}^* \delta_{ij} & 0 \\ H_{i-1} Q_i^b \delta_{ij} & (R_{i-1} + H_{i-1} Q_i^b H_{i-1}^*) \delta_{ij} & 0 \\ 0 & 0 & \Pi_{N+1} \\ 0 & 0 & 0 \end{bmatrix}, \tag{5.4.18}$$

and where we have defined

$$\mathbf{v}_{i+1}^b \triangleq H_i \mathbf{u}_{i+1}^b + \mathbf{v}_i.$$

We conclude with the remark that though the process $\{\mathbf{y}_i\}$ is not WSM, the aggregate process obtained by considering $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$ together is indeed WSM — this explains why $\{\mathbf{y}_i\}$ is often called the projection of a WSM process.

Lemma 5.4.6 ($\mathrm{col}\{\mathbf{x}_i, \mathbf{y}_i\}$ is WSM) *Consider the model (5.3.17)–(5.3.18) (with $\langle \mathbf{u}_i, \mathbf{v}_i \rangle = S_i = 0$). It follows that the aggregate vector process $\mathrm{col}\{\mathbf{x}_i, \mathbf{y}_i\}$ is* WSM. *Moreover, the output process $\{\mathbf{y}_i\}$ admits the backwards state-space representation (5.4.17)–(5.4.18).* ■

Proof: The argument prior to the statement of the lemma establishes the validity of the backwards state-space model. We only need to show that $\mathrm{col}\{\mathbf{x}_i, \mathbf{y}_i\}$ is indeed WSM. For this purpose, note that we can write

$$\mathbf{y}_{i+1} = H_{i+1}\mathbf{x}_{i+1} + \mathbf{v}_{i+1} = H_{i+1}F_i\mathbf{x}_i + H_{i+1}G_i\mathbf{u}_i + \mathbf{v}_{i+1}.$$

Hence,

$$\begin{bmatrix} \mathbf{x}_{i+1} \\ \mathbf{y}_{i+1} \end{bmatrix} = \begin{bmatrix} F_i & 0 \\ H_{i+1}F_i & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{bmatrix} + \begin{bmatrix} G_i & 0 \\ H_{i+1}G_i & I \end{bmatrix} \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_{i+1} \end{bmatrix}.$$

Moreover, note that

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_{i+1} \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix} \right\rangle = 0 \quad \text{for all} \quad i \geq 0.$$

This fact follows from the conditions $\langle \mathbf{u}_i, \mathbf{x}_0 \rangle = 0$, $\langle \mathbf{v}_i, \mathbf{x}_0 \rangle = 0$, and $\langle \mathbf{u}_i, \mathbf{v}_j \rangle = 0$ for all i, j. Now using Thm. 5.4.1 we conclude that $\mathrm{col}\{\mathbf{x}_i, \mathbf{y}_i\}$ is WSM. ◆

In fact, it also holds that the vector process $\mathrm{col}\{\mathbf{x}_i, \mathbf{y}_i\}$ is WSM even if $S_i \neq 0$ (see Prob. 5.7).

5.5 COMPLEMENTS

Sec. 5.3.2. Handling Initial Conditions. It was realized by early workers in estimation theory that special attention had to be given to the handling of "initial conditions", and that state-space descriptions were helpful in doing so, especially for time-variant systems. See, for example, the remark in the 1956 textbook of Laning and Battin (p. 340): "In effect, the essential difficulties associated with the point $t = 0$ are absorbed in the calculations needed to reduce the N-th order equation to N first-order equations."

Sec. 5.3.3. State-Space Descriptions. The use of what (control) engineers call state-space descriptions/models/representations of systems described by high-order differential or difference equations is wellknown in mathematical textbooks on differential equations. Physicists used these representations in studying Markov processes (see, *e.g.*, Wang and Uhlenbeck (1945), which is reprinted in the collection edited by N. Wax (1954)). Doob (1944) made a detailed study of stationary Gaussian processes having the Markov property in a long paper that, for some reason, was not cited in his famous 1953 textbook (Doob (1953)). Had he done so, development of state-space estimation results may have occurred much earlier; in fact, a lot of the material in Sec. 5.4 on wide-sense Markov processes can be found in that paper. R. E. Bellman began in the mid-fifties to reemphasize the value of state-space descriptions in control problems; similar inputs were appearing from switching circuits and automata theory and even classical circuit theory (see Bashkow (1957)). And specifically in estimation theory, their use was also, as noted earlier, suggested by Laning and Battin (1956).

At about the same time, R. L. Stratonovich in the (then) USSR was strongly urging that attention turn from Gaussian processes (and linear filters) to Markov processes (and except in the Gaussian case, nonlinear filters) — see, *e.g.*, Stratonovich (1959, 1960a, 1960b) and his book, Stratonovich (1966). For example, in Stratonovich (1960a) he remarks, referring to the difficulties in studying general non-Gaussian processes, that "we may conveniently use to eliminate those difficulties a Markoff process apparatus, involving Markoff processes and their corresponding equations. Markoff processes may be continuous or discontinuous, may have one or many components, may correspond to continuous time or form discrete sequences. . . ." More remarks on Stratonovich's notable but somewhat overlooked contributions can be found in the notes to Ch 16. [It is intriguing to note that Doob's 1944 paper was added by A. Yaglom to the Russian translation of Doob's 1953 textbook.]

However, it was undoubtedly R. E. Kalman, through his outstanding research on a wide range of fundamental problems, who brought the state-space point of view into center stage in system theory (see, *e.g.*, Kalman (1960b, 1960c, 1963c) and the chapters by Kalman in Kalman, Falb, and Arbib (1969)); see also the papers in the tribute edited by Antoulas (1991).

Sec. 5.4.3. Backwards Markovian Models. In connection with smoothing problems, researchers had found it useful to consider reversing the time direction, by writing $\mathbf{x}_i = F_i^{-1}\mathbf{x}_{i+1} - F_i^{-1}G_i\mathbf{u}_i$, $i \le N$. Of course $\langle \mathbf{x}_{N+1}, \mathbf{u}_i \rangle$ will not generally be zero, so the above is not a Markovian model. However to obtain the correct smoothing formulas, early authors somewhat arbitrarily introduced the assumption, $\|\mathbf{x}_{N+1}\|^2 = \infty$, which can be regarded as implying that the random variable $\mathbf{x}_{N+1}$ is so random that it is uncorrelated with all other random variables, and in particular with the $\{\mathbf{u}_i,\ i \le N\}$, so this assumption avoids the issue of proper modeling of the reverse-time process. The first insight into the problem came through the study of certain scattering/transmission models for the estimation problem (see Ch. 17). In transmission lines, we have waves going in both directions, and so it is natural to explore both forwards and backwards evolution. This idea led to the result of Thm. 5.4.2 (in continuous time) — see Ljung and Kailath (1976b), where a direct proof was given once the result itself had been obtained via the scattering picture; later other proofs were also presented (*e.g.*, Sidhu and Desai (1976) and Lainiotis (1976a)). However, these proofs only showed that the forwards and backwards models gave processes $\{\mathbf{x}_i, \mathbf{x}_i^b\}$ with the same second-order statistics. Verghese and Kailath (1979) gave the more physical argument described at the end of Sec. 5.4.3, which showed in fact that the sample paths of $\{\mathbf{x}_i, \mathbf{x}_i^b\}$ were the same.

PROBLEMS

5.1 **(An alternative AR model)** Refer to the discussion in Sec. 5.1 where we mentioned the alternative state-space realization

$$\mathbf{x}_{i+1} = a\mathbf{x}_i + \mathbf{u}_i, \quad \mathbf{y}_i = a\mathbf{x}_i + \mathbf{u}_i, \quad i > -\infty,$$

for the zero-mean stationary exponentially correlated process $\{\mathbf{y}_i\}$. Using this model, show that we again obtain $\hat{\mathbf{y}}_{i|i-1} = a\mathbf{y}_{i-1}$.

5.2 **(Innovations of two random processes)** Let $\{\mathbf{y}_i\}$ be a scalar real-valued random process defined by the equations

$$\mathbf{y}_i + a_1\mathbf{y}_{i-1}^2 + a_2|\mathbf{y}_{i-2}| = \mathbf{w}_i, \quad i \ge 0,$$

$$\mathbf{w}_i = \mathbf{u}_i + \mathbf{u}_{i-1}, \quad i \ge 0,$$

with $\mathbf{y}_{-1} = \mathbf{y}_{-2} = 0$, $\mathbf{u}_{-1} = 0$, and $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij}$. Show that the innovations of $\{\mathbf{y}_i\}$ are the same as that of $\{\mathbf{w}_i\}$ or, more specifically, that

$$\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_{i|\mathbf{y}_0,\ldots,\mathbf{y}_{i-1}} = \mathbf{w}_i - \hat{\mathbf{w}}_{i|\mathbf{w}_0,\ldots,\mathbf{w}_{i-1}}.$$

5.3 **(A second-order model)** Consider a zero-mean stationary scalar-valued random process $\{\mathbf{y}_i\}$ that is generated by the difference equation

$$\mathbf{y}_{i+1} = a_0\mathbf{y}_i + a_1\mathbf{y}_{i-1} + \mathbf{u}_i + b\mathbf{u}_{i-1}, \quad i > -\infty,$$

where $\{\mathbf{u}_i\}$ is a stationary white-noise sequence with $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = Q\delta_{ij}$ and $\langle \mathbf{u}_i, \mathbf{y}_j \rangle = 0$ for $j \le i$. The roots of the characteristic polynomial $z^2 - a_0 z - a_1$ are further assumed to be strictly inside the unit circle. Extend the derivation of Sec. 5.1.2 to show that the innovations process can be found recursively as follows:

$$\mathbf{e}_{i+1} = \begin{cases} -b\mathbf{e}_i + \mathbf{y}_{i+1} - a_0\mathbf{y}_i - a_1\mathbf{y}_{i-1} & \text{if } |b| < 1 \\ -\frac{1}{b}\mathbf{e}_i + \mathbf{y}_{i+1} - a_0\mathbf{y}_i - a_1\mathbf{y}_{i-1} & \text{if } |b| > 1. \end{cases}$$

Remark. When $|b| < 1$, the transfer function from $\mathbf{e}_i$ to $\mathbf{y}_i$ can be seen to be

$$\frac{\mathbf{y}(z)}{\mathbf{e}(z)} = \frac{z(z+b)}{z^2 - a_0 z - a_1}.$$

Now note that the transfer function from $\mathbf{u}_i$ to $\mathbf{y}_i$ is given by

$$\frac{\mathbf{y}(z)}{\mathbf{u}(z)} = \frac{z+b}{z^2 - a_0 z - a_1}.$$

That is, the above transfer functions differ only by a factor of z. Later, in Sec. 6.4, when we study canonical spectral factorizations, we shall see that this fact has a natural explanation (see Ex. 6.5.3). ◆

5.4 **(Higher-order predictors via innovations)** Consider again the second-order stationary process $\{\mathbf{y}_i\}$ of Prob. 5.3.

(a) Verify that the following relations hold:

$$\hat{\mathbf{y}}_{i+1|i-2} = a_0\hat{\mathbf{y}}_{i|i-2} + a_1\hat{\mathbf{y}}_{i-1|i-2},$$

$$\hat{\mathbf{y}}_{i|i-2} = a_0\hat{\mathbf{y}}_{i-1|i-2} + a_1\mathbf{y}_{i-2},$$

$$\hat{\mathbf{y}}_{i-1|i-2} = a_0\mathbf{y}_{i-2} + a_1\mathbf{y}_{i-3} + bQr_e^{-1}\mathbf{e}_{i-2}.$$

(b) Determine an expression for $\hat{\mathbf{y}}_{i+3|i}$ in terms of $\{\mathbf{y}_i, \mathbf{y}_{i-1}, \mathbf{e}_i\}$ only. Determine also the transfer function from $\mathbf{y}_i$ to $\hat{\mathbf{y}}_{i+3|i}$. Consider both cases: $|b| < 1$ and $|b| > 1$.

Remark. For stationary random processes, we shall develop in Ch. 7 the Wiener-Hopf technique, which will allow us to obtain the same results by working with the so-called z-spectrum of the output process — see Prob. 7.13. In Ch. 8 we shall rederive (see Prob. 8.2) the same results using the state-space formalism of Sec. 5.3. ◆

5.5 **(The polynomial approach)** Consider the setting of Prob. 5.4. When $|b| < 1$, we can rederive the transfer function of part (b) in an alternative (more algebraic) manner as follows. We first use the model to express $\mathbf{y}(z)$ in terms of $\mathbf{u}(z)$,

$$\mathbf{y}(z) = \frac{z+b}{z^2 - a_0 z - a_1}\,\mathbf{u}(z).$$

Now since we are interested in predicting 3 steps ahead in time, we multiply both sides by z^3 and use polynomial division to write

$$z^3\mathbf{y}(z) = E(z)\mathbf{u}(z) + \frac{F(z)}{z^2 - a_0 z - a_1}\,\mathbf{u}(z) = E(z)\mathbf{u}(z) + \frac{F(z)}{z+b}\,\mathbf{y}(z),$$

for some polynomials $\{E(z), F(z)\}$, with the degree of $F(z)$ strictly less than the degree of $z^2 - a_0 z - a_1$. Note that the transfer function $F(z)/(z+b)$ is stable because $|b| < 1$. The above relation expresses $z^3\mathbf{y}(z)$ as the sum of two terms. The first term, $E(z)\mathbf{u}(z)$, when translated to the time domain will include input noise terms that are uncorrelated with all current and past observations, $\{\mathbf{y}_j, j \le i\}$. The second term, on the other hand, is only dependent on these observations. Hence, the transfer function from $\mathbf{y}_i$ to $\hat{\mathbf{y}}_{i+3|i}$ should be equal to $F(z)/(z+b)$.

(a) Carry out the above calculations and verify that you obtain the same result as in part (b) of Prob. 5.4.

(b) Let m_j denote the coefficients of $E(z)$, say $E(z) = m_0 + m_1 z + \ldots + m_p z^p$. Find the $\{m_j\}$ and the order p. Show that the resulting minimum mean-square error is given by

$$E|\mathbf{y}_{i+3} - \hat{\mathbf{y}}_{i+3|i}|^2 = Q\sum_{i=0}^{p}|m_i|^2.$$

Remark. The method described in this problem is sometimes refered to as the polynomial approach — see, *e.g.*, Ch. 7 of Söderström (1994). The approach handles prediction and estimation problems for stationary processes that are described by minimum-phase models (*i.e.*, by systems whose poles and zeros are inside the unit circle). For such systems, as it turns out, the determination of the transfer function from the innovations $\{\mathbf{e}_i\}$ to the observations $\{\mathbf{y}_i\}$ is actually trivial. This is because it is equal, apart from a power of z, to the transfer function of the given system, *i.e.*, from $\{\mathbf{u}_i\}$ to $\{\mathbf{y}_i\}$, as explained in the remark to Prob. 5.3 and in greater detail in Ex. 6.5.3 of Ch. 6. In Ch. 7 we shall develop a more general Wiener-Hopf technique for stationary processes that can handle both minimum- and nonminimum-phase systems (see Ex. 7.13). In the minimum-phase case, the polynomial approach and the Wiener-Hopf technique become equivalent. This latter claim is established in Prob. 7.14. For nonstationary processes, the use of state-space models is almost essential. ◆

5.6 **(A nonstandard state-space model)** Consider the state-space model

$$\begin{cases}\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_{i+1}, & i \ge 0,\\ \mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i,\end{cases}$$

with random variables $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i\}$ that continue to satisfy (5.3.18). Show that

(a) $\langle \mathbf{u}_i, \mathbf{x}_i\rangle = Q_i G_{i-1}^*$ and, for $i > j$, $\langle \mathbf{u}_i, \mathbf{x}_j\rangle = 0$ and $\langle \mathbf{v}_i, \mathbf{x}_j\rangle = 0$.

(b) $\langle \mathbf{u}_i, \mathbf{y}_i\rangle = Q_i G_{i-1}^* H_i^* + S_i$ and, for $i > j$, $\langle \mathbf{u}_i, \mathbf{y}_j\rangle = 0$ and $\langle \mathbf{v}_i, \mathbf{y}_j\rangle = 0$.

5.7 **(WSM property of col$\{\mathbf{x}_i, \mathbf{y}_i\}$)** Consider the standard state-space model (5.3.17)–(5.3.18) with $S_i \neq 0$. Show that

$$\begin{bmatrix}\widehat{\mathbf{x}_{i+1}}\\ \mathbf{y}_{i+1}\end{bmatrix}_{|\{\mathbf{x}_0,\ldots,\mathbf{x}_i,\mathbf{y}_0,\ldots,\mathbf{y}_i\}} = \begin{bmatrix}\widehat{\mathbf{x}_{i+1}}\\ \mathbf{y}_{i+1}\end{bmatrix}_{|\{\mathbf{x}_i,\mathbf{y}_i\}}$$

and conclude that col$\{\mathbf{x}_i, \mathbf{y}_i\}$ is WSM.

5.8 **(Colored measurement noise)** Consider the state-space model (5.3.25)–(5.3.26).

(a) Show that, for $i \ge j$, $\langle \mathbf{u}_i, \mathbf{x}_j\rangle = 0$, $\langle \mathbf{v}_i, \mathbf{x}_j\rangle = 0$, and $\langle \mathbf{n}_i, \mathbf{x}_j\rangle = 0$.

(b) Find recursions for $\|\mathbf{x}_i\|^2$ and $\|\mathbf{n}_i\|^2$.

(c) Compute $\langle \mathbf{n}_i, \mathbf{y}_i\rangle$, $\langle \mathbf{n}_i, \mathbf{y}_{i+1}\rangle$, and $\langle \mathbf{n}_i, \mathbf{y}_{i-1}\rangle$.

5.9 **(A simple test for the WSM property)** Let $\{\mathbf{x}_i\}_{i=0}^N$ be WSM and define $\mathbf{x} = \text{col}\{\mathbf{x}_i\}$ and $\tilde{\mathbf{x}} = \text{col}\{\tilde{\mathbf{x}}_i\}$, where $\tilde{\mathbf{x}}_i = \mathbf{x} - \hat{\mathbf{x}}_i$, $\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_{i|i-1,i-2,\ldots 0}$.

(a) Show that the (lower triangular) matrix M defined via $\tilde{\mathbf{x}} = M\mathbf{x}$ is block bidiagonal.

(b) Show that the matrix $R_{\tilde{x}} = \langle \tilde{\mathbf{x}}, \tilde{\mathbf{x}}\rangle$ is block diagonal.

(c) Use parts (a) and (b) to prove that a process $\{\mathbf{x}_i\}_{i=0}^N$ is WSM if, and only if, $R_x^{-1} = \langle \mathbf{x}, \mathbf{x}\rangle^{-1}$ is tri-diagonal.

Remark. See Ackner and Kailath (1989a, 1989b). ◆

5.10 **(The smoothing error is WSM)** Consider the linear model $\mathbf{y} = H\mathbf{x} + \mathbf{v}$, where $\mathbf{x}$ and $\mathbf{v}$ are zero-mean uncorrelated random variables with covariances R_x and R_v, respectively. Assume $\mathbf{x} = \text{col}\{\mathbf{x}_0, \ldots, \mathbf{x}_N\}$, $\mathbf{v} = \text{col}\{\mathbf{v}_0, \ldots, \mathbf{v}_N\}$, $\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$, and $H = \text{diag}(H_0, \ldots, H_N)$, $R_x = \text{diag}(R_{x_0}, \ldots, R_{x_N})$, $R_v = \text{diag}(R_{v_0}, \ldots, R_{v_N})$.

(a) Prove that the smoothing error $\left(\tilde{\mathbf{x}}_{i|\{\mathbf{y}_0,\ldots,\mathbf{y}_N\}}\right)_{i=0}^N$ is WSM. [*Hint.* Use the result of Prob. 5.9.]

(b) Show that the result of part (a) is true even if $\{\mathbf{v}_i\}$ is not white, but a WSM process itself.

5.11 **(Stationary WSM processes)** Let $\{\mathbf{x}_i\}$ be a WSM process.

(a) Show that we can write $r_{ik} = T_{ij} r_{jk}$, for some doubly indexed sequence $\{T_{ij}\}$, where we have defined $r_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j\rangle$.

(b) We shall say that $\{\mathbf{x}_i\}$ is *stationary* if, and only if, $r_{ij} = r_{i-j}$, for all i, j. Show that this implies $T_{ij} = W_{i-j}$, for some sequence $\{W_i\}$.

(c) Deduce from part (b) that $T_{ij} = F^{i-j}$, $i \ge j$.

(d) Prove that a stationary process is WSM if, and only if, $r_{ij} = F^{i-j}\bar{\Pi}$, for $i \ge j$, where $\bar{\Pi} = r_{ii}$ is the stationary variance of $\mathbf{x}_i$.

5.12 **(Whiteness of $\mathbf{u}_i^b$)** Show that $\mathbf{u}_{i+1}^b = -F_i^{-1}G_i\left(\mathbf{u}_i - Q_i G_i^* \Pi_{i+1}^{-1}\mathbf{x}_{i+1}\right)$ in (5.4.15) is a white-noise process.

5.13 **(Proof of Lemma 5.A.2)** Use expression (5.3.20) to verify the formula (5.A.6) for the output Gramian R_y in Lemma 5.A.2. Then use (5.A.4) and (5.A.5) to prove the second formula (5.A.7) for R_y.

5.14 **(Properties of Lyapunov recursions)** Consider the Lyapunov recursion

$$\Pi_{i+1} = F\Pi_i F^* + GQG^*, \quad \text{with initial condition} = \Pi_0,$$

where F is stable (eigenvalues strictly inside the unit disc).

(a) Define $\tilde{\Pi}_{i+1} = \Pi_{i+1} - \Pi_i$ and show that $\lim_{i\to\infty} \tilde{\Pi}_i = 0$.

(b) Show that if F is stable and $\{F, GQ^{\frac{1}{2}}\}$ is controllable (*cf.* App. C), then there exists a unique positive-definite solution to the Lyapunov equation $\Pi = F\Pi F^* + GQG^*$.

(c) Prove that in the Lyapunov recursion $\lim_{i\to\infty} \Pi_i = \Pi$, for all Π_0.

Remark. More discussion of Lyapunov equations can be found in App. D. ◆

5.15 **(Invertibility of the state covariance matrix)** Let $\{\mathbf{x}_i\}_{i=0}^N$ be a WSM process with Markovian representation (5.4.3)–(5.4.4) and define $\Pi_i = \|\mathbf{x}_i\|^2$, $i = 0, \ldots, N$.

(a) Show that the $\{\Pi_i\}_{i=0}^N$ are invertible if, and only if,

$$\Pi_0 > 0 \text{ and } \begin{bmatrix} F_i & G_i Q_i^{\frac{1}{2}} \end{bmatrix} \text{ has full rank for all } i = 0, 1, \ldots, N.$$

(b) Show that the following two conditions are each sufficient to guarantee the invertibility of the $\{\Pi_i\}_{i=0}^N$:

(i) $\Pi_0 > 0$ and F_i nonsingular for all $i = 0, 1, \ldots, N$.

(ii) $\Pi_0 > 0$ and the pair $\{F_i, G_i Q_i^{\frac{1}{2}}\}$ controllable (*cf.* App. C) for all $i = 0, 1, \ldots, N$.

5.16 **(A special equation)** Show that every solution of the Lyapunov equation $\Pi - C\Pi C^* = gqg^*$, where C is a *top-companion* matrix as in (5.3.7), q is a scalar, and $g^* = [1\ 0\ \ldots\ 0]$, is a Toeplitz matrix. [*Hint.* Let e_i denote the row unit vector with 1 in the i-th position and zero elsewhere. Note first that $e_i C = e_{i-1}$, $i \geq 1$, and then examine the relations obtained by pre-multiplying the Lyapunov equation by e_i and post-multiplying it by e_j, for $i, j \geq 1$.]

Appendix for Chapter 5

5.A SOME GLOBAL FORMULAS

Here we shall rewrite the state-space equations (5.3.17) in a global form and use them to obtain some useful formulas for the Gramian/covariance matrix R_y. We start by defining the column vectors

$$\mathbf{y} \triangleq \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}, \quad \mathbf{u} \triangleq \text{col}\{\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_N\}, \quad \mathbf{v} \triangleq \text{col}\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_N\}.$$

Define also $\text{zcol}\{\mathbf{x}_0, \mathbf{u}\}$. Then we can rewrite the standard equations (5.3.17) as

$$\mathbf{y} = A\mathbf{z} + \mathbf{v}, \tag{5.A.1}$$

where

$$A = \begin{bmatrix} H_0 & & & & \\ H_1\Phi(1,0) & H_1G_0 & & & \\ H_2\Phi(2,0) & H_2\Phi(2,1)G_0 & H_2G_1 & & \\ \vdots & & & \ddots & \\ H_N\Phi(N,0) & H_N\Phi(N,1)G_0 & H_N\Phi(N,2)G_1 & \ldots & H_NG_{N-1}\ 0 \end{bmatrix}.$$

[Note that the last (block) column of A is zero, since $\mathbf{u}_N$ does not affect $\mathbf{y}_N$.] We thus have a linear model $\mathbf{y} = A\mathbf{z} + \mathbf{v}$ of the same form studied earlier in Sec. 3.4. But here we see that the matrix A has further structure: it is (block) lower triangular and its entries are completely specified in terms of the given $\{F_i, G_i, H_i\}$. Also, we now have

$$< \mathbf{v}, \mathbf{v} > = \text{diag}\{R_0, R_1, \ldots, R_N\} \quad \text{and} \quad < \mathbf{z}, \mathbf{z} > = \text{diag}\{\Pi_0, Q_0, \ldots, Q_N\}.$$

When $\mathbf{u}_i = 0$ and $\mathbf{v}_i = 0$, for $i \geq 0$, the output $\{\mathbf{y}_i\}$ will be given by $\mathbf{y}_i = H_i\Phi(i,0)\mathbf{x}_0$. This suggests defining the *observability map*

$$\mathcal{O} \triangleq \mathcal{O}(N,0) = \begin{bmatrix} H_0\Phi(0,0) \\ H_1\Phi(1,0) \\ H_2\Phi(2,0) \\ \vdots \\ H_N\Phi(N,0) \end{bmatrix} = \begin{bmatrix} H_0 \\ H_1F_0 \\ H_2F_1F_0 \\ \vdots \\ H_NF_{N-1}\ldots F_0 \end{bmatrix}, \tag{5.A.2}$$

which maps the initial state $\mathbf{x}_0$ to the output vector $\mathbf{y}$ (when $\mathbf{u}_i = 0, \mathbf{v}_i = 0$), $\mathbf{y} = \mathcal{O}\mathbf{x}_0$. Likewise, when both $\mathbf{x}_0$ and $\mathbf{v}$ are zero, the output will be

$$\mathbf{y}_i = \sum_{j=0}^{i-1} H_i\Phi(i, j+1)G_j\mathbf{u}_j = \sum_{j=0}^{i-1} \Gamma_{ij}\mathbf{u}_j,$$

where

$$\Gamma_{ij} \triangleq H_i\Phi(i, j+1)G_j = H_iF_{i-1}\dots F_{j+1}G_j$$

is the response at time i to an impulse at time $j < i$. This suggests that we introduce the (strictly lower triangular) *impulse response* matrix

$$\Gamma = \begin{bmatrix} 0 & & & \\ \Gamma_{10} & 0 & & \\ \Gamma_{20} & \Gamma_{21} & 0 & \\ \Gamma_{30} & \Gamma_{31} & \Gamma_{32} & 0 \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

It is now straightforward to write the following global expression relating the quantities $\{\mathbf{x}_0, \mathbf{u}, \mathbf{v}\}$,

$$\mathbf{y} = \mathcal{O}\mathbf{x}_0 + \Gamma\mathbf{u} + \mathbf{v} = \begin{bmatrix}\mathcal{O} & \Gamma\end{bmatrix}\begin{bmatrix}\mathbf{x}_0 \\ \mathbf{u}\end{bmatrix} + \mathbf{v}. \tag{5.A.3}$$

Comparing (5.A.3) and (5.A.1) shows that $A = \begin{bmatrix}\mathcal{O} & \Gamma\end{bmatrix}$.

We can now give some global formulas for the Gramian matrix R_y of the output process $\mathbf{y}$.

Lemma 5.A.1 (Output Gramian Matrix) *Consider the standard state-space model (5.3.17)–(5.3.18). Then*

$$R_y = \mathcal{O}\Pi_0\mathcal{O}^* + \begin{bmatrix}\Gamma & I\end{bmatrix}\begin{bmatrix}\mathcal{Q} & \mathcal{S} \\ \mathcal{S}^* & \mathcal{R}\end{bmatrix}\begin{bmatrix}\Gamma^* \\ I\end{bmatrix}, \tag{5.A.4}$$

where

$$\mathcal{Q} = diag(Q_0, \dots, Q_N), \quad \mathcal{R} = diag(R_0, \dots, R_N), \quad \mathcal{S} = diag(S_0, \dots, S_N).$$

In particular, if $S_i = 0$, $i = 0, \dots, N$, then we have $R_y = \mathcal{O}\Pi_0\mathcal{O}^* + \Gamma\mathcal{Q}\Gamma^* + \mathcal{R}$. ■

Proof: Follows immediately from (5.A.3). ◆

Alternative expressions can be obtained by exploiting the fact that the entries of $\mathcal{O}$ and Γ inherit the assumed state-space structure. Indeed, let us denote by

$$\mathcal{Z} \triangleq \begin{bmatrix} 0 & & & \\ I & 0 & & \\ & I & 0 & \\ & & \ddots & \ddots \\ & & & I \quad 0 \end{bmatrix},$$

the lower triangular shift matrix (with identity on the first lower block subdiagonal and zeros elsewhere); $\mathcal{Z}$ has the property that it shifts column vectors one block element downwards: if $a = \text{col}\{a_0, \dots, a_N\}$, then $\mathcal{Z}a = \text{col}\{0, a_0, \dots, a_{N-1}\}$. Using the shift

matrix $\mathcal{Z}$, we can rewrite the i-th block column of the impulse response matrix as (refer to (5.A.1) where the structure of the columns of Γ is shown, and where the first block column of Γ corresponds to the index $i = 0$ below):

$$i\text{-th column of } \Gamma = \mathcal{Z}^i \begin{bmatrix} 0 \\ H_{i+1}\Phi(i+1, i+1) \\ H_{i+2}\Phi(i+2, i+1) \\ \vdots \\ H_N\Phi(N, i+1) \end{bmatrix} G_i. \tag{5.A.5}$$

Consequently, the following expressions for R_y also result.

Lemma 5.A.2 (Two Expressions for the Output Gramian) *The Gramian R_y corresponding to the state-space model (5.3.17)–(5.3.18) can be written in either of the following two forms:*

$$R_y = \tag{5.A.6}$$

$$\sum_{i=0}^{N} \mathcal{Z}^i \begin{bmatrix} 0 & I \\ H_{i+1}\Phi(i+1, i+1) & 0 \\ H_{i+2}\Phi(i+2, i+1) & 0 \\ \vdots & \vdots \\ H_N\Phi(N, i+1) & 0 \end{bmatrix} \begin{bmatrix} 0 & N_i \\ N_i^* & R_i + H_i\Pi_iH_i^* \end{bmatrix} \begin{bmatrix} 0 & I \\ H_{i+1}\Phi(i+1, i+1) & 0 \\ H_{i+2}\Phi(i+2, i+1) & 0 \\ \vdots & \vdots \\ H_N\Phi(N, i+1) & 0 \end{bmatrix}^* \mathcal{Z}^{i*}$$

or

$$R_y = \mathcal{O}\Pi_0\mathcal{O}^* + \tag{5.A.7}$$

$$\sum_{i=0}^{N} \mathcal{Z}^i \begin{bmatrix} 0 & I \\ H_{i+1}\Phi(i+1, i+1) & 0 \\ H_{i+2}\Phi(i+2, i+1) & 0 \\ \vdots & \vdots \\ H_N\Phi(N, i+1) & 0 \end{bmatrix} \begin{bmatrix} G_iQ_iG_i^* & S_i \\ S_i^* & R_i \end{bmatrix} \begin{bmatrix} 0 & I \\ H_{i+1}\Phi(i+1, i+1) & 0 \\ H_{i+2}\Phi(i+2, i+1) & 0 \\ \vdots & \vdots \\ H_N\Phi(N, i+1) & 0 \end{bmatrix}^* \mathcal{Z}^{i*},$$

where $N_i = F_i\Pi_iH_i^* + G_iS_i$ and Π_i satisfies the recursion

$$\Pi_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^*, \quad i \geq 0,$$

with initial condition Π_0. ■

Proof: The proof of (i) follows from inspection of the expressions (5.3.20) for the elements of R_y. Likewise (ii) follows by inspecting (5.A.4) and using (5.A.5) — see Prob. 5.13. ◆

[The equivalence of the two representations (5.A.6) and (5.A.7) becomes even more evident when we later consider (see Sec. 8.1) the important special case of constant-parameter systems.]

Remark. The state-space model $\{F_i, G_i, H_i, I\}$ for the process $\{\mathbf{y}_i\}$ reflects itself into the covariance matrix by the fact that R_y is completely determined by the set of (small) matrices $\{F_i, H_i, N_i, I\}$. Therefore, it should not be surprising that there exist algorithms for factoring R_y and R_y^{-1} that require only $O(Npn^3)$ computations rather than the $O(N^3p^3)$ required for a general $(N+1)p \times (N+1)p$ matrix. However, such fast algorithms are not readily evident. Although it is possible to take a linear algebraic approach, such as the GS or MGS procedures, to find the triangular factors of R_y (or R_y^{-1}), as noted at the beginning of Sec. 5.3, the calculations are somewhat cumbersome, though of course they can be done. In App. 9.A, we shall do them by using the MGS procedure and in App. 9.B we shall describe a method that uses the two representations of R_y in Lemma 5.A.2. However, as was the case in Sec. 5.3, it is easier to work with the state-space model itself, rather than with the covariance matrix, as we shall show in Ch. 9. ◆

CHAPTER 6

Innovations for Stationary Processes

In this chapter we consider the extension of the conceptually straightforward triangular factorization for finite covariance matrices of Ch. 4 to doubly infinite, but Toeplitz, covariance matrices that arise when we have scalar-valued stationary processes. We shall show in Sec. 6.1 that the doubly infinite Toeplitz matrix factorization problem can be translated to a problem of canonical spectral factorization. Background material required for the study of such factorizations, on discrete-time signals and systems, and on the z-spectra of discrete-time stationary processes, is reviewed in Secs. 6.2 and 6.3. The canonical spectral factorization problem is introduced in Sec. 6.4, with the important case of scalar rational spectra studied in Sec. 6.5. The spectral factorization problem for vector-valued processes is introduced in Sec. 6.6.

6.1 INNOVATIONS VIA SPECTRAL FACTORIZATION

In Sec. 4.2 we showed that the evaluation of the innovations $\{\mathbf{e}_i\}$ of a finite number of observations, $\{\mathbf{y}_i, 0 \le i \le N\}$, is equivalent to the triangular factorization of the corresponding finite covariance matrix, $R_y = [\langle \mathbf{y}_i, \mathbf{y}_j \rangle]_{i,j=0}^N$; our standing assumption is that $R_y > 0$ for all N. Then the factorization $R_y = LDL^*$, where L is lower triangular with unit diagonal entries and D is diagonal, allows $\{\mathbf{y}_i\}$ and $\{\mathbf{e}_i\}$ to determine each other uniquely via the relations

$$\mathbf{y} = L\mathbf{e} \quad \text{or} \quad \mathbf{e} = L^{-1}\mathbf{y} \triangleq W\mathbf{y}, \quad \text{say,} \tag{6.1.1}$$

where $\mathbf{y} = \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}$, $\mathbf{e} = \text{col}\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_N\}$, and $\langle \mathbf{e}, \mathbf{e} \rangle = D$. Then, for each i, we can write

$$\mathbf{e}_i = w_{i0}\mathbf{y}_0 + w_{i1}\mathbf{y}_1 + w_{i2}\mathbf{y}_2 + \ldots + w_{i,i-1}\mathbf{y}_{i-1} + \mathbf{y}_i,$$

where the $\{w_{ij}\}$ denote the entries of the i-th (block) row of L^{-1} (or W), with $w_{ii} = 1$. An important fact to note is that the coefficients $\{w_{ij}\}$ generally change with the time instant i and, therefore, the above equation describes a *time-variant* mapping from the observations $\{\mathbf{y}_j\}$ to the innovations $\{\mathbf{e}_i\}$.

However, when $\{\mathbf{y}_i\}$ is a stationary process, observed from $i = -\infty$, then the mapping from $\{\mathbf{e}_i\}$ to $\{\mathbf{y}_i\}$, and vice versa, turns out to be time-invariant. This was the setting of the classical studies of Wold (1938,1954), Kolmogorov (1939,1941a), and Wiener (1942,1949), and we shall introduce some of their results in this chapter.

It turns out that there are significant differences in the analysis between *scalar-valued* and *vector-valued* processes. For this reason, we shall focus mostly in this chapter on *scalar-valued* stationary processes; the difficulties in the vector case are described in Sec. 6.6, and resolved in Ch. 8 using state-space models.

6.1.1 Stationary Processes

Consider a zero-mean scalar stationary process $\{\mathbf{y}_i, -\infty < i < \infty\}$, with covariance function $R_y(i) = \langle \mathbf{y}_j, \mathbf{y}_{j-i} \rangle$. We shall show that the computation of the corresponding innovations process

$$\{\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_{i|i-1} = \mathbf{y}_i - \hat{\mathbf{y}}_i\} \qquad \text{for } -\infty < i < \infty,$$

can be reduced to the equivalent problem of computing the so-called canonical factorization of the so-called z-spectrum of $\{\mathbf{y}_i\}$ (to be defined shortly).

To begin with, note that when $\{\mathbf{y}_i\}$ is a stationary stochastic process defined for $-\infty < i < \infty$, the corresponding covariance matrix will be a doubly infinite matrix,

$$R_y = \begin{bmatrix} \ddots & \ddots & \ddots & \cdots & & \\ \cdots & R_y(1) & R_y(0) & R_y^*(1) & \cdots & \\ & \cdots & R_y(1) & \boxed{R_y(0)} & R_y^*(1) & \cdots \\ & & \cdots & R_y(1) & R_y(0) & R_y^*(1) & \cdots \\ & & & \cdots & \ddots & \ddots & \ddots \end{bmatrix}, \qquad R_y(i) = \langle \mathbf{y}_j, \mathbf{y}_{j-i} \rangle.$$

Note further that because of the stationarity assumption, R_y has a special *Toeplitz* (or constant-along-diagonals) structure. Now we would like to determine, if possible, a triangular factorization for R_y, say

$$R_y = LDL^*, \tag{6.1.2}$$

with the factors $\{L, D\}$ having a *similar* Toeplitz structure,[1]

$$L = \begin{bmatrix} \ddots & \ddots & \ddots & & & \bigcirc \\ \cdots & l_2 & l_1 & 1 & & \\ & \cdots & l_2 & l_1 & \boxed{1} & \\ & & \cdots & l_2 & l_1 & 1 \\ & & & \cdots & \ddots & \ddots & \ddots \end{bmatrix}, \quad D = \begin{bmatrix} \ddots & & & \bigcirc \\ & r_e & & \\ & & \boxed{r_e} & \\ & & & r_e \\ \bigcirc & & & \ddots \end{bmatrix}.$$

The central $(0, 0)$ elements in these doubly infinite matrices have been highlighted by rectangular boxes.

Three of the issues that arise when dealing with doubly infinite processes are now evident. First, for the the case of a finite matrix R_y, we saw that performing its LDL* decomposition was straightforward (using, say, the GS or MGS procedures). When the matrix is doubly infinite, however, it is not immediately obvious how to go about finding the elements $\{l_i, r_e\}$.

The second issue, although not as obvious at first sight, is more fundamental. Apart from being an uncorrelated (white) process, the innovations process, $\{\mathbf{e}_i\}$, must be related to the observations process, $\{\mathbf{y}_i\}$, by *causal* and *causally invertible* operations. This, of course, means that the doubly infinite Toeplitz matrix L that we seek should be invertible, so that we can obtain the innovations from the observations via

$$\text{col}\{\ldots, \mathbf{e}_{i-1}, \boxed{\mathbf{e}_i}, \mathbf{e}_{i+1}, \ldots\} = L^{-1} \text{ col}\{\ldots, \mathbf{y}_{i-1}, \boxed{\mathbf{y}_i}, \mathbf{y}_{i+1}, \ldots\}, \tag{6.1.3}$$

and, more importantly, that L^{-1} should also be lower triangular (and Toeplitz) so that the inverse mapping is causal (and time-invariant) as well. In the finite case this is not a significant issue since any lower triangular matrix with unit diagonal is invertible and has an inverse that is also lower triangular. In the case of doubly infinite matrices, although it is still possible to assert the invertibility of any lower triangular Toeplitz matrix with unit diagonal, it is not necessarily true that the inverse is *lower triangular* as well. Indeed, Prob. 6.1 gives an example of a doubly infinite lower triangular Toeplitz matrix whose inverse is *upper triangular*!

The third issue is that the resulting time-invariant mappings between the sequences $\{\mathbf{y}_i, \mathbf{e}_i\}$ should guarantee that a wide-sense stationary signal $\{\mathbf{e}_i\}$ is mapped into a similar wide-sense stationary signal $\{\mathbf{y}_i\}$ and vice versa. Now since

$$\mathbf{y}_k = \mathbf{e}_k + \sum_{i=1}^{\infty} l_i \mathbf{e}_{k-i}, \qquad \mathbf{e}_k = \mathbf{y}_k + \sum_{i=1}^{\infty} w_i \mathbf{y}_{k-i}, \tag{6.1.4}$$

[1] While the Toeplitz requirement on $\{L, D\}$ might seem strong, our arguments will show that this is always possible for rational z-spectra that are strictly positive on the unit circle, while a certain Paley-Wiener condition is required for general z-spectra. One motivation for seeking Toeplitz matrices L and D is that one expects the innovations process $\{\mathbf{e}_i\}$ of a stationary process $\{\mathbf{y}_i\}$ to be stationary as well.

where we are using $\{w_i\}$ to denote the entries in any particular row of (a now Toeplitz) L^{-1}, the above condition is met by requiring

$$\sum_{i=1}^{\infty} |l_i|^2 \;<\; \infty, \qquad \sum_{i=1}^{\infty} |w_i|^2 \;<\; \infty. \tag{6.1.5}$$

Remark 1. These conditions guarantee that the series in (6.1.4) converge in the mean-square sense to wide-sense stationary processes (see, *e.g.*, Picinbono (1993, pp. 72–73)). A sequence $\{\mathbf{s}_i\}$ of random variables is said to be mean-square convergent to a random variable $\mathbf{s}$ if $\|\mathbf{s}_i - \mathbf{s}\|^2 \to 0$ as $i \to \infty$, *i.e.*, if the variance of $(\mathbf{s}_i - \mathbf{s})$ tends to zero. An equivalent Cauchy criterion states that $\{\mathbf{s}_i\}$ is mean-square convergent if, and only if, $\|\mathbf{s}_i - \mathbf{s}_j\|^2 \to 0$ as $i, j \to \infty$ independently. To show that both series in (6.1.4) are mean-square convergent under (6.1.5) we proceed as follows. Consider, for example, the first series in (6.1.4) and introduce the partial sums

$$\mathbf{S}_p \triangleq \mathbf{e}_k + \sum_{i=1}^{p} l_i \mathbf{e}_{k-i},$$

for all integers $p \geq 1$. Now verify that the sequence $\{\mathbf{S}_p\}$ satisfies the Cauchy criterion. ◆

6.1.2 Generating Functions and z-Spectra

With the above remarks in mind,[2] we see that to find the innovations of a doubly infinite process $\{\mathbf{y}_i\}$, we must perform the LDL* factorization in (6.1.2) in such a way that the matrix L has a well-defined lower triangular Toeplitz inverse. This seems to be a formidable task. However, a strategy that is often useful when working with doubly infinite Toeplitz matrices is to introduce "generating functions," which allows us to reduce the doubly infinite covariance matrix factorization problem to the "more tractable" one of suitably factoring a generating function.

The generating function corresponding to the doubly infinite Toeplitz matrix

$$T = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & & \\ \cdots & t_1 & t_0 & t_{-1} & \cdots & \\ & \cdots & t_1 & \boxed{t_0} & t_{-1} & \cdots \\ & & \cdots & t_1 & t_0 & t_{-1} & \cdots \\ & & & \cdots & \ddots & \ddots & \ddots \end{bmatrix},$$

is defined as

$$T(z) \triangleq \sum_{i=-\infty}^{\infty} t_i z^{-i}, \tag{6.1.6}$$

[2] The following discussion is largely for motivational purposes, and we do not pursue all the mathematical issues, nor the most general case.

where z is a "place holder" (or indeterminate). It will be convenient to introduce the doubly infinite row vector

$$\begin{bmatrix} \cdots & z^2 & z & \boxed{1} & z^{-1} & z^{-2} & \cdots \end{bmatrix} \triangleq \Lambda(z).$$

Premultiplying T by $\Lambda(z)$ yields

$$\Lambda(z)T = T(z)\begin{bmatrix} \cdots & z^2 & z & \boxed{1} & z^{-1} & z^{-2} & \cdots \end{bmatrix} = T(z)\Lambda(z).$$

In other words, $\Lambda(z)$ is a left eigenvector of T with eigenvalue $T(z)$.

We can now use this result to check that pre-multiplying both sides of the desired matrix factorization (6.1.2) by $\Lambda(z)$ gives the compact form

$$S_y(z) \;=\; L(z) r_e L^*(z^{-*}), \tag{6.1.7}$$

where we introduced the generating functions of R_y, L, and $D = \text{diag}\{r_e\}$,

$$S_y(z) \triangleq \sum_{i=-\infty}^{\infty} R_y(i) z^{-i}, \tag{6.1.8}$$

and

$$L(z) \triangleq 1 + \sum_{i=1}^{\infty} l_i z^{-i}, \qquad D(z) \;=\; r_e. \tag{6.1.9}$$

The notation $L^*(z^{-*})$ that appears in (6.1.7) will be used throughout this chapter and it stands for

$$L^*(z^{-*}) \;\triangleq\; \left[L\left(\frac{1}{z^*}\right)\right]^*.$$

That is, we replace z by $1/z^*$ and then evaluate the complex conjugate of the resulting expression. For example, if $L(z) = 1 + az^{-1}$, then $L^*(z^{-*}) = 1 + a^* z$. $L^*(z^{-*})$ is known as the *para-Hermitian conjugate* of $L(z)$; when $|z| = 1$, it is the usual Hermitian conjugate.

The corresponding function version of the innovations representation (6.1.3) is

$$\mathbf{e}(z) \;=\; L^{-1}(z)\mathbf{y}(z), \tag{6.1.10}$$

where we are writing $\mathbf{e}(z)$ and $\mathbf{y}(z)$ to denote the (formal) series

$$\mathbf{e}(z) \triangleq \Lambda(z)\ \text{col}\{\ldots, \mathbf{e}_{i-1}, \boxed{\mathbf{e}_i}, \mathbf{e}_{i+1}, \ldots\} \;=\; \sum_{i=-\infty}^{\infty} \mathbf{e}_i z^{-i}$$

$$\mathbf{y}(z) \triangleq \Lambda(z)\ \text{col}\{\ldots, \mathbf{y}_{i-1}, \boxed{\mathbf{y}_i}, \mathbf{y}_{i+1}, \ldots\} \;=\; \sum_{i=-\infty}^{\infty} \mathbf{y}_i z^{-i}.$$

If we now regard z as a complex variable (and assuming that the series (6.1.8) in $\{z, z^{-1}\}$ converge in the sense described in Sec. 6.2.1 below), then $S_y(z)$ is what we shall call the *z-spectrum* of the stationary process $\{\mathbf{y}_i\}$; the reason is that when $z = e^{j\omega}$, $S_y(e^{j\omega})$ is the power spectral density function of the stationary process (here, $j = \sqrt{-1}$). The series $L(z)$ is the z-transform of the sequence $\{l_i\}$ and can be regarded as the *transfer function* of a linear system with impulse response $\{l_i\}$.

We now have to translate the requirement that the infinite Toeplitz matrices L and L^{-1} must both be lower triangular into properties of the transfer function $L(z)$. We begin to explain how to do this in the next two sections. In particular, the discussion that follows is aimed largely at answering the following questions:

Q1. Under what conditions is the power series (6.1.8) well defined?

A1. We shall restrict our discussion in this chapter to exponentially bounded covariance sequences $\{R_y(i)\}$, *viz.*, sequences that satisfy

$$|R_y(i)| < K\alpha^{|i|}, \tag{6.1.11}$$

for some $K > 0$ and $0 < \alpha < 1$. In this case, the series (6.1.8) will be well defined and will converge absolutely for all values of z satisfying $\alpha < |z| < \frac{1}{\alpha}$. This domain includes the unit circle, $|z| = 1$, and hence, the power spectral density function $S(e^{j\omega})$ will also be well defined.[3]

Q2. What conditions should the factor $L(z)$ in (6.1.7) and (6.1.9) satisfy in order for its coefficients $\{l_i\}$ to correspond to a triangular factorization (6.1.2) with L and L^{-1} both lower triangular and Toeplitz with square summable rows?

A2. We shall see that, for rational z-spectra, $L(z)$ will be rational with its poles and zeros strictly inside the unit circle; such a transfer function is said to be *minimum-phase*.

Q3. When does a factorization of the form (6.1.7) with the minimum-phase condition on $L(z)$ exist? How do we find it?

A3. We shall see, in a fairly obvious manner, that the factorization always exists and is unique for *rational* z-spectra that are strictly positive on the unit circle. For general z-spectra, a certain so-called Paley-Wiener condition should be satisfied (*cf.* (6.4.2)). We shall also note several methods for computing $\{L(z), r_\epsilon\}$ from knowledge of $S_y(z)$.

The above questions and answers are treated in the sequel, especially in Secs. 6.4 and 6.5.

[3] The exponential boundedness condition on $\{R_y(i)\}$ can be relaxed, in which case one should also allow for power spectral density functions $S(e^{j\omega})$ that converge in a weaker sense, *e.g.*, in a mean-square sense or in a distributional sense.

6.2 SIGNALS AND SYSTEMS

In this section, we briefly review some results from the theory of discrete-time linear time-invariant (LTI) systems. More detailed expositions can be found in textbooks such as Mitra (1998) and Oppenheim and Schafer (1998).

6.2.1 The z-Transform

The bilateral z-transform of a doubly infinite sequence, $\{u_i\}_{i=-\infty}^{\infty}$, is denoted by $u(z)$ and defined as

$$u(z) \triangleq \sum_{i=-\infty}^{\infty} u_i z^{-i}. \tag{6.2.1}$$

Although $u(z)$ may be considered as a formal sum, the definition of the z-transform is most useful when z takes values in the complex plane, $\mathbb{C}$, in which case the definition only makes sense for those values of z for which the series (6.2.1) *converges* in an appropriate sense. The values of z for which we have convergence are said to form the *region of convergence* (ROC) of the z-transform. We shall restrict ourselves in this book to sequences that are *exponentially bounded*, *i.e.*, for all i, there exist α and K such that

$$|u_i| < K\alpha^{|i|}, \quad 0 < \alpha < 1, \ K > 0. \tag{6.2.2}$$

In this case, convergence can be taken as *absolute convergence*, *i.e.*, that

$$\sum_{i=-\infty}^{\infty} |u_i| \cdot |z^{-i}| < \infty$$

for all z in the ROC. Under the assumption (6.2.2), the ROC can be verified to be the *annulus* $\alpha < |z| < \alpha^{-1}$, which contains the unit circle since $0 < \alpha < 1$.

The original sequence $\{u_i\}$ can then be recovered from its z-transform by the *inverse z-transform*,

$$u_i = \frac{1}{j2\pi} \oint_C z^{i-1} u(z) dz, \tag{6.2.3}$$

where C is a counterclockwise closed contour in the ROC that encircles the origin, $z = 0$ and $j = \sqrt{-1}$. [Of course, in various special cases, the inverse transform can be found more directly, *e.g.*, from a table of transforms or by partial fraction expansions in the rational case.] It is important to note that given any function, $u(z)$, its inverse z-transform, $\{u_i\}$, will depend on the ROC that we choose (if one is not explicitly specified).

For example, if we are given $u(z) = 1/(z - 0.5)$, we can expand it as

$$u(z) = \frac{1}{z - 0.5} = \frac{z^{-1}}{1 - 0.5z^{-1}} = \sum_{i=1}^{\infty} (0.5)^{i-1} z^{-i}, \quad |z| > 0.5, \tag{6.2.4}$$

which, as indicated, converges for values of z outside the disc of radius 0.5, $|z| > 0.5$. The inverse z-transform is (*cf.* the definition (6.2.1))

$$u_i = \begin{cases} 0.5^{i-1} & i > 0, \\ 0 & i \le 0. \end{cases} \tag{6.2.5}$$

On the other hand, we could also expand $u(z)$ as

$$u(z) = \frac{1}{z - 0.5} = \frac{-2}{1 - 2z} = -2\left(\sum_{i=0}^{\infty} 2^i z^i\right) = -\sum_{i=-\infty}^{0} 2^{-i+1} z^{-i}, \quad |z| < 0.5. \quad (6.2.6)$$

Here, the ROC is $|z| < 0.5$ and the inverse z-transform now is

$$u_i = \begin{cases} 0 & i > 0, \\ -2^{-i+1} & i \leq 0. \end{cases} \quad (6.2.7)$$

The first sequence (6.2.5) is a *causal* sequence; the second (6.2.7) is *anticausal*.[4] The difference arises from the difference in the ROCs — in the first (causal) case, the ROC is the outside of a circle; in the second (anticausal) case it is the inside of the circle. As stated earlier, we shall only be dealing with sequences that are *exponentially bounded*, in which case the ROC will be an annulus containing the unit circle. *All our z-transforms will be regarded as having such a region of convergence*, so that in the above example the *only* time sequence that can be associated with $u(z) = (z - 0.5)^{-1}$ is the exponentially decaying and causal sequence (6.2.5).

Now consider $v(z) = -1/(z - 2)$. If the ROC is to contain the unit circle, the only associated time sequence is obtained from the expansion

$$v(z) = \frac{0.5}{1 - 0.5z} = \frac{1}{2}\left(1 + \frac{z}{2} + \frac{z^2}{4} + \ldots\right), \quad \text{for} \quad |z| < 2,$$

corresponding to the exponentially bounded anticausal sequence

$$v_i = \begin{cases} 2^{-(i+1)} & i \leq 0, \\ 0 & i > 0. \end{cases}$$

[The function $v(z)$ has a pole at 2, which control engineers tend to think of as corresponding to an unbounded growing sequence — this is true if $v(z)$ is associated with a causal sequence, or equivalently if the ROC is taken as the region $|z| > 2$. In other words, associating poles outside the unit circle with unbounded sequences is only true if our interest is always in causal sequences, as is (often implicitly) assumed in presentations of transform theory for control engineers.]

In any case, the above discussion suggests the following general conclusion (which can be established by working with the contour integral formula (6.2.3)):

(i) Exponentially bounded and *causal* sequences have transforms that are analytic (*i.e.*, have no singularities) in a region of the form $|z| > \alpha$, for some $0 < \alpha < 1$ (*i.e.*, outside a circular region that includes the unit circle).

[4] In general, a causal sequence $\{u_i\}$ is one that satisfies $u_i = 0$ for all $j < i_0$ for some i_0. That is, the sequence exists to the right of a time instant i_0. Likewise, an anticausal sequence $\{u_i\}$ is one that satisfies $u_i = 0$ for all $j > i_0$ for some i_0. That is, the sequence exists to the left of a time instant i_0. In this chapter, we shall assume that $i_0 = 0$, unless otherwise specified. We note also that $\{u_i\}$ will be called *strictly* causal if $u_i = 0$, $i \leq 0$, and strictly anticausal if $u_i = 0$, $i \geq 0$.

(ii) Exponentially bounded and *anticausal* sequences have transforms that are analytic (*i.e.*, have no singularities) in a region of the form $|z| < \beta$, for some $\beta > 1$ (*i.e.*, inside a circular region that includes the unit circle).

The terminology *causal* and *anticausal* is not very meaningful for sequences as such and actually arises from the study of linear systems. However, before studying such systems, let us note an important consequence of having the ROC contain the unit circle, $|z| = 1$. In this case, the *discrete-time Fourier transform* (DTFT)

$$u(e^{j\omega}) \triangleq \sum_{i=-\infty}^{\infty} u_i e^{-j\omega i}, \qquad j \triangleq \sqrt{-1},$$

converges (absolutely) for all values of $\omega \in [-\pi, \pi]$. Conversely, if we have absolute convergence of the discrete-time Fourier transform, then the unit circle must be in the ROC of the corresponding z-transform. [If we do not have absolute convergence, the Fourier transform may still exist, *e.g.*, in a distributional sense; in such cases the ROC will generally not be an annulus but only the unit circle; examples are the unit step sequence or sample functions of stationary random processes.]

6.2.2 Linear Time-Invariant Systems

A discrete-time linear time-invariant (LTI) system maps an input sequence $\{u_i\}$ to an output sequence $\{y_i\}$ according to a convolution rule,

$$y_i = \sum_{k=-\infty}^{\infty} h_{i-k} u_k = \sum_{k=-\infty}^{\infty} h_k u_{i-k}, \quad (6.2.8)$$

where the sequence $\{h_i\}$ is the output of the LTI system when the input is a unit sample (or Kronecker delta), $u_i = \delta_0$, which leads to $\{h_i\}$ being called the *impulse response* of the system.

Another characterization of an LTI system is via its *transfer function*, defined as the (bilateral) z-transform of its impulse response,

$$H(z) \triangleq \sum_{i=-\infty}^{\infty} h_i z^{-i}.$$

An important reason for introducing transfer functions is that the convolution (6.2.8) can be replaced by multiplication in the z-transform domain,

$$y(z) = H(z)u(z). \quad (6.2.9)$$

In this case, the ROC of $y(z)$ will be the intersection of the ROCs of $H(z)$ and $u(z)$, *i.e.*, it is their common annulus of convergence. By our standing assumption that both the sequences $\{h_i\}$ and $\{u_i\}$ are exponentially bounded, the ROC of their z-transforms contains the unit circle, and therefore the same is true of the ROC of $y(z)$. In fact, it is also true that $\{y_i\}$ is exponentially bounded.

We may mention that our main interest will be in rational transfer functions, namely those having the form

$$H(z) = \frac{b_0 z^m + b_1 z^{m-1} + \ldots + b_m}{a_0 z^n + a_1 z^{n-1} + \ldots + a_n} \triangleq \frac{b(z)}{a(z)},$$

where we further assume that the polynomials $\{b(z), a(z)\}$ are coprime, *i.e.*, they have no factors in common. In this case, the roots of the denominator and numerator polynomials are called the *poles* and *zeros* of the system, respectively.

Another important concept regarding LTI systems is *stability*. There are many different notions of stability. For present purposes, it is enough to consider (bounded input bounded output) BIBO stability. It can be shown that an LTI system is BIBO stable if, and only if, its impulse response sequence is absolutely summable, *viz.*,

$$\sum_{i=-\infty}^{\infty} |h_i| < \infty, \tag{6.2.10}$$

which in turn is equivalent to the requirement that the ROC of the transfer function $H(z)$ should include the unit circle, $|z| = 1$.

6.2.3 Causal, Anticausal, and Minimum-Phase Systems

In our discussion, it is going to be relevant whether signals and stable systems are causal or not, and therefore we shall further elaborate on this point here.

As mentioned before, a sequence $\{u_i\}$ will be said to be *causal* if it is zero for negative time, *i.e.*, if

$$u_i = 0, \quad \text{for } i < 0.$$

It will be said to be *strictly causal*if it is zero for $i \leq 0$. Our standing assumption of exponential boundedness means that the z-transform

$$u(z) = \sum_{i=0}^{\infty} u_i z^{-i}.$$

of a causal sequence is analytic in a region of the form, $|z| > \alpha, 0 < \alpha < 1$. When $u(z)$ is rational, this analyticity is equivalent to $u(z)$ having all its poles strictly within $|z| \leq \alpha$ (and hence, strictly within the unit circle).

Likewise, a sequence $\{u_i\}$ is *anticausal* if it is zero for positive time, *i.e.*, if

$$u_i = 0, \quad \text{for } i > 0.$$

The sequence is *strictly anticausal* if it is zero for $i \geq 0$. Our assumption on the exponential boundedness of $\{u_i\}$ now means that the z-transform must be analytic in a region of the form $|z| < \alpha, \alpha > 1$. When $u(z)$ is rational, this analyticity is equivalent to $u(z)$ having all its poles strictly within $|z| \geq \alpha$ (and, hence, strictly outside the unit circle).

A linear system is called (strictly) causal if its impulse response, $\{h_i\}$, is (strictly) causal. Note that the input-output relation (6.2.8) implies that causal systems map causal inputs to causal outputs. Moreover, in this case we can be more specific and, for $i \geq 0$, rewrite (6.2.8) as

$$y_i = \sum_{k=0}^{i} h_{i-k} u_k = \sum_{k=0}^{i} h_k u_{i-k}.$$

From the above discussions we conclude that a stable linear system is causal if its transfer function, $H(z)$, is analytic outside a circular region that includes the unit circle (recall that the BIBO stability of $H(z)$ is equivalent to the absolute summability of its impulse response sequence $\{h_i\}$, which in turn implies that the ROC of $H(z)$ should include the unit circle). When $H(z)$ is rational, this means that all the poles of the transfer function lie strictly within the unit circle.

Similarly, a (strictly) anticausal system is one for which the impulse response, $\{h_i\}$, is (strictly) anticausal. An alternative characterization for stable anticausal systems is that their transfer function, $H(z)$, be analytic inside a circular region that includes the unit circle, which in the rational case means that $H(z)$ has all its poles strictly outside the unit circle.

The *inverse* of an LTI system is one that maps the output of the original system to the input of the original system. Thus if the original system is defined by the input-output relation $y(z) = H(z)u(z)$, then the inverse system must obey $u(z) = H^{-1}(z)y(z)$. In other words, the transfer function of the inverse of an LTI system is the inverse of the transfer function of the original system.

Finally, a very important definition. A LTI system $H(z)$ is called *minimum-phase* if both it and its inverse, $H^{-1}(z)$, are stable and causal. Using the characterizations described above, this means that both $H(z)$ *and* $H^{-1}(z)$ must be analytic outside circular regions that include the unit circle. When $H(z)$ is rational, since the poles of $H^{-1}(z)$ are the zeros of $H(z)$, $H(z)$ will be minimum phase if, and only if, all its poles and zeros lie strictly inside the unit circle, $|z| = 1$. This is further equivalent to saying that $H(z)$ *and* $H^{-1}(z)$ are analytic on and outside the unit circle.

6.3 STATIONARY RANDOM PROCESSES

In this section, we very briefly review some elementary facts from the theory of stationary random processes.[5] For more extensive, and more general treatments, see *e.g.*, the textbook of Doob (1953), and the more engineering-oriented presentations in Caines (1988), Picinbono (1993), and Porat (1994).

The *covariance sequence* of a scalar zero-mean discrete-time stationary process, $\{\mathbf{y}_i\}$, is the sequence

$$R_y(i) = \langle \mathbf{y}_k, \mathbf{y}_{k-i} \rangle = E\mathbf{y}_k\mathbf{y}_{k-i}^* = R_y^*(-i), \quad -\infty < i < \infty.$$

Note that, by stationarity, the covariance sequence is a function only of i (and not of k and $k - i$).

[5] We often omit the adjectives "second-order" or "wide-sense" that should be used because we only assume knowledge of the first- and second-order statistics of the process, *i.e.*, of the mean-value and covariance functions.

Conversely, any para-Hermitian sequence $\{R_y(i) = R_y^*(-i)\}_{i=-\infty}^{\infty}$ that satisfies the nonnegativity condition

$$\begin{bmatrix} R_y(0) & R_y(-1) & \dots & R_y(-i) \\ R_y(1) & R_y(0) & \dots & R_y(-i+1) \\ \vdots & \vdots & \ddots & \vdots \\ R_y(i) & R_y(i-1) & \dots & R_y(0) \end{bmatrix} \geq 0, \tag{6.3.1}$$

for all $i = 0, 1, 2, \dots$, is called a covariance sequence since it can be regarded as being generated by a scalar zero-mean discrete-time stationary process $\{\mathbf{y}_i\}$.

6.3.1 Properties of the z-Spectrum

Our standing assumption that the sequence $R_y(i)$ is exponentially bounded (*cf.* (6.1.11)) means that the z-transform,

$$S_y(z) \triangleq \sum_{i=-\infty}^{\infty} R_y(i) z^{-i} = [S_y(z^{-*})]^* \triangleq S_y^*(z^{-*}),$$

converges absolutely and is analytic in an annulus containing the unit circle, *viz.*, $\alpha < |z| < 1/\alpha$. As noted in Sec. 6.1, the function $S_y(z)$ will be called the *z-spectrum* of the process $\{\mathbf{y}_i\}$. Since the ROC of $S_y(z)$ contains the unit circle, the discrete-time Fourier transform (DTFT)

$$S_y(e^{j\omega}) \triangleq \sum_{i=-\infty}^{\infty} R_y(i) e^{-j\omega i}, \qquad j \triangleq \sqrt{-1},$$

converges for all $\omega \in [0, 2\pi]$ and is also real-valued. The function $S_y(e^{j\omega})$ is called the *power spectral density function* or, more briefly, the *power spectrum* of $\{\mathbf{y}_i\}$. It has the important properties of:

(i) Hermitian symmetry:

$$S_y(e^{j\omega}) = S_y^*(e^{j\omega}). \tag{6.3.2}$$

(ii) Nonnegativity:

$$S_y(e^{j\omega}) \geq 0, \qquad 0 \leq \omega \leq 2\pi. \tag{6.3.3}$$

The first property (i) is easily checked:

$$S_y^*(e^{j\omega}) = \sum_{i=-\infty}^{\infty} R_y^*(i) e^{j\omega i} = \sum_{i=-\infty}^{\infty} R_y(-i) e^{j\omega i} = \sum_{i=-\infty}^{\infty} R_y(i) e^{-j\omega i} = S_y(e^{j\omega}).$$

The other claim, which is a special case of the famous Wiener-Khinchin theorem, can be readily justified under our strong assumption (6.1.11) — see Prob. 6.7. Moreover, the converse is also true: a function $S_y(z)$ obeying (i) and (ii) above must be the z-transform of a covariance sequence.

We should also mention that the covariance sequence $\{R_y(i)\}$ can be recovered from the power spectrum $S_y(e^{j\omega})$ via the inverse DTFT formula

$$R_y(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_y(e^{j\omega}) e^{jk\omega} d\omega. \tag{6.3.4}$$

Moreover, our assumption of an exponentially bounded covariance sequence $\{R_y(i)\}$ implies that $R_y(0)$ is bounded so that

$$R_y(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(e^{j\omega}) d\omega < \infty, \tag{6.3.5}$$

and $\{\mathbf{y}_i\}$ is a finite power process.

6.3.2 Linear Operations on Stationary Stochastic Processes

In many applications, random signals are processed by stable linear time-invariant systems. In these situations, the statistical properties of the input and output random processes can be related via the transfer function of the LTI system. First, however, given two stationary random processes $\{\mathbf{y}_i\}$ and $\{\mathbf{u}_i\}$, we shall define their z-cross-spectrum as

$$S_{yu}(z) \triangleq \sum_{i=-\infty}^{\infty} R_{yu}(i) z^{-i} = S_{yu}^*(z^{-*}). \tag{6.3.6}$$

Lemma 6.3.1 (Filtering of Stationary Processes) *Let $\{\mathbf{y}_i\}$ be the stationary process that is obtained by passing a zero-mean stationary process, $\{\mathbf{u}_i\}$, through a stable linear time-invariant system with transfer function $H(z)$. Then the following relations hold:*

$$S_y(z) = H(z) S_u(z) H^*(z^{-*}) \quad \textit{and} \quad S_{yu}(z) = H(z) S_u(z),$$

and

$$S_y(e^{j\omega}) = H(e^{j\omega}) S_u(e^{j\omega}) H^*(e^{j\omega}) \quad \textit{and} \quad S_{yu}(e^{j\omega}) = H(e^{j\omega}) S_u(e^{j\omega}).$$

Finally, if $\{\mathbf{x}_i\}$ is jointly stationary with $\{\mathbf{y}_i, \mathbf{u}_i\}$ as just defined, then

$$S_{xy}(z) = S_{xu}(z) H^*(z^{-*}).$$

Moreover, these relations also hold for vector-valued processes (where now $*$ denotes the Hermitian transpose). ∎

Proof: We provide two proofs. The first one is based on traditional arguments. The second is simpler and more useful in treating systems interconnected in various ways; it is based on taking transforms of random sequences.

First proof. The input-output relation can be written in the time domain as

$$\mathbf{y}_i = \sum_{p=-\infty}^{\infty} h_{i-p}\mathbf{u}_p = \sum_{p=-\infty}^{\infty} h_p \mathbf{u}_{i-p}, \tag{6.3.7}$$

where $\{h_k\}$ denotes the impulse response sequence of $H(z) \triangleq \sum_{i=-\infty}^{\infty} h_i z^{-i}$. It then follows that $E\mathbf{y}_i = \sum_{p=-\infty}^{\infty} h_p E\mathbf{u}_{i-p} = 0$, meaning that the output process is also zero-mean.

Next, we evaluate the output covariance function and write

$$\langle \mathbf{y}_i, \mathbf{y}_p \rangle = \langle \sum_{m=-\infty}^{\infty} h_m \mathbf{u}_{i-m}, \sum_{k=-\infty}^{\infty} h_k \mathbf{u}_{p-k} \rangle = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} h_m R_u(i-p-m+k) h_k^*$$

$$\triangleq R_y(i-p),$$

which shows that the output process is also wide-sense stationary. Now, $S_y(z)$ is given by

$$\begin{aligned} S_y(z) &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} h_m R_u(n-m+k) h_k^* z^{-n}, \\ &= \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} h_m R_u(l) h_k^* z^{-(l+m-k)}, \\ &= \left(\sum_{m=-\infty}^{\infty} h_m z^{-m} \right) \left(\sum_{l=-\infty}^{\infty} R_u(l) z^{-l} \right) \left(\sum_{k=-\infty}^{\infty} h_k^* z^k \right), \\ &= H(z) S_u(z) \left[H\left(\frac{1}{z^*}\right) \right]^*, \\ &= H(z) S_u(z) H^*(z^{-*}). \end{aligned} \tag{6.3.8}$$

The other formulas follow similarly.

Second proof. We now give an alternative (more direct) derivation of these results by using the following observation. Given two zero-mean jointly stationary random processes $\{\mathbf{a}_i\}$ and $\{\mathbf{b}_i\}$, the z-cross-spectrum, $S_{ab}(z)$, can be written as

$$S_{ab}(z) = E\mathbf{a}(z)\mathbf{b}_0^*, \tag{6.3.9}$$

since

$$E\mathbf{a}(z)\mathbf{b}_0^* = E\left(\sum_{i=-\infty}^{\infty} \mathbf{a}_i z^{-i} \mathbf{b}_0^* \right) = \sum_{i=-\infty}^{\infty} (E\mathbf{a}_i \mathbf{b}_0^*) z^{-i} = \sum_{i=-\infty}^{\infty} R_{ab}(i) z^{-i} = S_{ab}(z).$$

Now using (6.3.9) with $\mathbf{a}(z) = \mathbf{y}(z) = H(z)\mathbf{u}(z)$ and $\mathbf{b}(z) = \mathbf{u}(z)$, we obtain

$$S_{yu}(z) = E\mathbf{y}(z)\mathbf{u}_0^* = E[H(z)\mathbf{u}(z)\mathbf{u}_0^*] = H(z)E\mathbf{u}(z)\mathbf{u}_0^* = H(z)S_u(z),$$

which is the second relation. Likewise, or by taking conguate transposes, we can obtain $S_{uy}(z) = S_u(z)H^*(z^{-*})$. Finally, if we use (6.3.9) with $\mathbf{a}(z) = \mathbf{b}(z) = \mathbf{y}(z) = H(z)\mathbf{u}(z)$, we obtain

$$S_y(z) = E\mathbf{y}(z)\mathbf{y}_0^* = E[H(z)\mathbf{u}(z)\mathbf{y}_0^*] = H(z)S_{uy}(z) = H(z)S_u(z)H^*(z^{-*}),$$

which establishes the first relation. Finally, note that

$$S_{yx}(z) = E\mathbf{y}(z)\mathbf{x}_0^* = H(z)S_{ux}(z),$$

so that

$$S_{xy}(z) = S_{yx}^*(z^{-*}) = S_{xu}(z)H^*(z^{-*}).$$

◆

Remark 2. The simple argument in the second proof is often avoided because of concerns about the convergence of the series $\sum_i \mathbf{a}_i z^{-i}$. These concerns can be addressed (see, *e.g.*, Doob (1953) and Caines (1988, Sec. 1.4)), but the reader may wish to regard it as a convenient mnemonic device for recalling the input-output relations of Lemma 6.3.1. ◆

6.4 CANONICAL SPECTRAL FACTORIZATION

We now return to the problem originally posed in Sec. 6.1, *viz.*, given a doubly infinite stationary sequence $\{\mathbf{y}_i\}$ with covariance matrix R_y, we would like to determine a factorization of the form (6.1.2). The factorization should be such that L is a doubly infinite lower triangular Toeplitz matrix with a lower triangular Toeplitz inverse. Moreover, any particular row of L or L^{-1} should be absolutely summable.

Using the generating function description (6.1.6), we argued in that section that the triangular factorization (6.1.2) of R_y is equivalent to the factorization (6.1.8) of the z-spectrum, $S_y(z)$, where the function $L(z)$ was defined by (6.1.9). We are now in a position to translate the above requirements on the triangular factor L into requirements on $L(z)$. We do this in steps:

(A–1). The function $L(z)$ as defined by (6.1.9) is the z-transform of a causal sequence since the coefficients l_i exist only for $i \geq 1$. This means that $L(z)$ must be the z-transform of a causal sequence and, hence, its ROC must be the outside of a circular region (*i.e.*, of the form $|z| > \alpha$ for some $\alpha \geq 0$).

(A–2). The square summability of the sequence $\{l_i\}$ of any row of L (*cf.* (6.1.5)) implies that $\{l_i\}$ is a bounded sequence, say $|l_i| < M$ for some finite M. In this case, $L(z)$ will be analytic in $|z| > 1$ since

$$\sum_{i=1}^{\infty} |l_i z^{-i}| \leq M \sum_{i=1}^{\infty} |z^{-i}| = \frac{M}{|z|-1} \quad \text{for all} \quad |z| > 1.$$

(A–3). Since the process $\{\mathbf{y}_i\}$ has finite power (by assumption), and its z-spectrum is rational (also by assumption), there can be no poles of $S_y(z)$, and hence of $L(z)$, on the unit circle $|z| = 1$. This is because unit-circle poles of $S_y(z)$ would make the integral in (6.3.5) diverge. Combining this fact with (A–1) and (A–2) above, we see that $L(z)$ is analytic in $|z| \geq 1$ which is equivalent to saying that $L(z)$ must have all its poles strictly inside the unit circle. This means that $L(z)$ should be a BIBO stable system so that its impulse response sequences is in fact absolutely summable,

$$\sum_{i=1}^{\infty} |l_i| \ < \ \infty.$$

[This condition is stronger than the square summability condition in (6.1.5).]

(B–1). Since we also require the inverse of L to be lower triangular and Toeplitz, this means, just like $L(z)$, that $L^{-1}(z)$ should be the z-transform of a causal sequence. Therefore, its ROC must also be the outside of a circular region (*i.e.*, of the form $|z| > \beta$ for some $\beta \geq 0$).

(B–2). The square-summability of the sequence $\{w_i\}$ of any row of L^{-1} implies that $L^{-1}(z)$ is analytic in $|z| > 1$.

(B–3). Combining (B–1) and (B–2) we see that $L^{-1}(z)$ must be analytic in $|z| > 1$, which means that $L(z)$ must have all its zeros in $|z| \leq 1$.

(B–4). We shall however impose the *stronger* condition that $L^{-1}(z)$ be analytic in $|z| \geq 1$ so that it is a BIBO stable system as well and, hence, all zeros of $L(z)$ will be strictly inside the unit circle. This guarantees that

$$\sum_{i=1}^{\infty} |w_i| \ < \ \infty,$$

again a stronger assumption than in (6.1.5).

In summary, conditions (A) and (B) above show that $L(z)$ must correspond to a minimum-phase system. We should note that the assumption that $L^{-1}(z)$ is analytic on the unit circle rules out the possibility of having any zeros of $S_y(z)$ on the unit circle. This means that $S_y(e^{j\omega})$ must be strictly positive, *i.e.*,

$$S_y(e^{j\omega}) > 0, \quad -\pi \leq \omega \leq \pi.$$

Definition 6.4.1. (Canonical Spectral Factorization) *Let $S_y(z)$ be a rational z-spectrum of a finite power process and assume that $S_y(e^{j\omega})$ is strictly positive. The canonical spectral factorization of $S_y(z)$ is*

$$S_y(z) = L(z) r_e L^*(z^{-*}), \tag{6.4.1}$$

where $L(z)$ is minimum-phase (i.e., the zeros and poles of $L(z)$ are strictly inside the unit circle), $r_e > 0$, and $L(\infty) = 1$. ◆

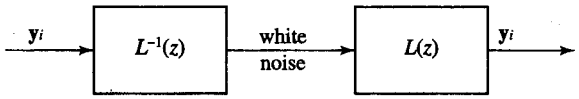


Figure 6.1 The whitening and modeling filters.

Modeling and Whitening Filters. Comparison of the formula (6.4.1) and expression (i) in Lemma 6.3.1 for the z-spectrum of the output of an LTI system driven by a stationary input yields the following important physical interpretations of the canonical spectral factor $L(z)$ and its inverse $L^{-1}(z)$:

(i) $L(z)$ can be regarded as a *modeling* filter since it shows how to construct the process $\{\mathbf{y}_i\}$ by passing a white-noise process (the innovations process) with variance r_e through a stable causal filter.

(ii) When $S_y(z)$ has no unit-circle zeros (as we have assumed), $L(z)$ will also have no unit-circle zeros, we can invert the modeling filter to obtain the stable *causal* filter $L^{-1}(z)$. This filter can be regarded as a (causal) *whitening* filter since it shows how to obtain a white-noise process (the innovations process) with variance r_e by passing the original process $\{\mathbf{y}_i\}$ through a stable causal filter (see Fig. 6.1).

For these reasons, the filters $\{L(z), L^{-1}(z)\}$ are often denoted, more mnemonically, by the symbols $M(z)$ and $W(z)$. Following Bode and Shannon (1950) and Zadeh and Ragazzini (1950), the above interpretations will be used in the next chapter to give a "physical" approach to the solution of the causal filtering problem.

Remark 3. Of course, there are several possible modeling filters, since we can always replace $L(z)$ by $L(z)U(z)$, where $U(z)U^*(z^{-*}) = 1$; for example with $U(z) = e^{j\theta}$. However, the canonical modeling filter is *uniquely* characterized by the two further properties: $L(\infty) = I$ and its $L^{-1}(z)$ is also causal. ◆

Remark 4. It will be seen in Sec. 6.5 that the existence and uniqueness of the canonical spectral factor $L(z)$ is fairly obvious for rational z-spectra. It can happen however that a rational $S_y(z)$ has zeros on the unit circle, a case we have excluded. However, a generalized definition of canonical factorization is possible in which $L(z)$ is allowed to have unit-circle zeros, which means that $L^{-1}(z)$ can now only be analytic in a region of the form $|z| > 1$. Recovering the process $\{\mathbf{y}_i\}$ from the innovations process $\{\mathbf{e}_i\}$ is now not straightforward, though it can be done but not in a numerically satisfactory way — see Prob. 6.5 and Hannan (1970). ◆

Remark 5. For general (not necessarily rational) z-spectra $S_y(z)$, the following result can be proved (see, *e.g.*, Doob (1953) and Grenander and Rosenblatt (1957)). There exists a unique function $L(z)$ satisfying $S_y(z) = L(z)r_eL^*(z^{-*})$ with the following properties:

(i) $L(z)$ and $L^{-1}(z)$ are analytic in $|z| > 1$, and

(ii) $\sum_{i=1}^{\infty} |l_i|^2 < \infty$ (*i.e.*, the impulse response of $L(z)$ is square-summable),

if, and only if, $S_y(z)$ is the z-spectrum of a finite power process (*cf.* (6.3.5)) and satisfies the so-called Paley-Wiener condition (see Probs. 6.8 and 6.14):

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \left[S_y(e^{j\omega}) \right] d\omega \ > \ -\infty. \tag{6.4.2}$$

In this case, it turns out that there exists an elegant so-called Szegö formulafor r_e (see Prob. 6.12),

$$r_e = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left[S_y(e^{j\omega})\right] d\omega\right]. \tag{6.4.3}$$

Moreover, because of the finite-power condition (6.3.5), the above Paley-Wiener condition is also equivalent to the absolute integrability of $\ln[S_y(e^{j\omega})]$, *viz.*,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\ln\left[S_y(e^{j\omega})\right]\right| \, d\omega \; < \; \infty. \tag{6.4.4}$$

Note that the finite-power assumption (6.3.5) rules out the possibility of poles of $L(z)$ on the unit circle. Still, $L(z)$ can have an essential singularity on $|z| = 1$ so that it can only be guaranteed to be analytic in $|z| > 1$. In the rational case, essential singularities cannot occur and, hence, $L(z)$ will be analytic in $|z| \geq 1$ and all its poles will be strictly inside the unit circle. Note further that while (ii) guarantees the existence of a canonical factor $L(z)$ with a square-summable impulse response, nothing is said about the square-summability of the impulse response of $L^{-1}(z)$. In fact, this latter condition cannot be guaranteed in general, *e.g.*, because $L(z)$ may have isolated unit-circle zeros (see Prob. 6.5).

However, $S_y(z)$ cannot have zeros over a set of nonzero measure on the unit circle. Thus, for example, random processes with band-limited power spectra will not qualify. ◆

Remark 6 [Computing the Canonical Factorization]. So far we have only discussed existence issues. How can we find the factors? There are general formulas for doing this (see, *e.g.*, Doob (1953), Grenander and Rosenblatt (1957), and also Wiener (1949, esp. App. 2)). We shall not pursue them here, except for an elegant method of Kolmogorov (under a stronger condition than (6.4.2)) described in Prob. 6.12. ◆

6.5 SCALAR RATIONAL z-SPECTRA

When the z-spectrum is rational it means that, up to a constant scaling factor, $S_y(z)$ is uniquely determined by its poles and zeros. However, since z-spectra have the properties of para-Hermitian symmetry, $S_y(z) = S_y^*(z^{-*})$, and nonnegativity, the poles and zeros have a certain structure:

(i) Since $S_y(z) = S_y^*(z^{-*})$, for every pole (or zero) at a point $z = \alpha$, there must be a pole (respectively a zero) at $z = \alpha^{-*}$.

(ii) Since the process has finite power (by assumption), there can be no poles on the unit circle $|z| = 1$. This is because unit-circle poles of $S_y(z)$ would make the integral in (6.3.5) diverge.

(iii) Since $S_y(z)$ is nonnegative on $|z| = 1$, any zeros on the unit circle must be of even multiplicity. However, our standing assumption that the z-spectrum is positive on $|z| = 1$ rules out the possibility of any unit-circle zeros.

(iv) The constant scaling factor appearing in the z-spectrum must be positive because $S_y(z)$ is positive on the unit circle.

It follows from the properties (i)–(iv) above that any rational z-spectrum $S_y(z)$ that is strictly positive on the unit circle can be written as

$$S_y(z) = r_e \cdot \frac{\prod_{i=1}^{m}(z - \alpha_i)(z^{-1} - \alpha_i^*)}{\prod_{i=1}^{n}(z - \beta_i)(z^{-1} - \beta_i^*)}, \quad \text{with } |\alpha_i| < 1, \; |\beta_i| < 1, \; r_e > 0. \tag{6.5.1}$$

Now it is fairly obvious how to construct modeling filters and in particular the canonical modeling filter. First note that we can (in several ways) factor $S_y(z)$ as

$$S_y(z) = H(z)H^*(z^{-*}). \tag{6.5.2}$$

Clearly if $H(z)$ has a pole (respectively zero) at $z = \alpha$, then $H^*(z^{-*})$ will have a pole (respectively zero) at $z = \alpha^{-*}$. Therefore, comparing the above expression with that of Eq. (6.5.1) we see that for each pair of poles (respectively zeros) of $S_y(z)$ at $z = \alpha$ and $z = \alpha^{-*}$ we can *arbitrarily* assign one to $H(z)$ and one to $H^*(z^{-*})$. Therefore, there are many different causal transfer functions $H(z)$ that can generate the z-spectrum $S_y(z)$, in the sense that if a unit-variance white-noise process is applied to the causal $H(z)$ the output has z-spectrum $S_y(z)$. We just have to put all the stable poles into $H(z)$.

However, to get the canonical factorization, we must further put all the stable zeros into $H(z)$. But this is not enough. We also have to meet the normalization condition that $L(\infty) = 1$. This can be achieved by choosing

$$L(z) = z^{n-m}\frac{\prod_{i=1}^{m}(z - \alpha_i)}{\prod_{i=1}^{n}(z - \beta_i)}, \quad |\alpha_i| < 1\,, \; |\beta_i| < 1. \tag{6.5.3}$$

In summary, to form $L(z)$ we retain the stable poles and zeros of $S_y(z)$ and then multiply by the factor z^{n-m} so that the normalization constraint $L(\infty) = 1$ is satisfied.

This task is easy to perform for low-order numerator and denominator (Laurent) polynomials in $S_y(z)$. However, other methods may be preferred when $\{n, m\}$ are large. The literature describes several methods of separately factoring the numerator and denominator polynomials of $S_y(z)$, *e.g.*, Bauer's method (Bauer (1955,1956)), and the use of the Levinson-Durbin algorithm (see Prob. 4.11) and of the Schur algorithm (see App. F). It turns out that all these methods can be better understood once we have obtained algorithms for state-space estimation, so we shall not pursue them here; we may note that once the state-space connection is made, several other methods (*e.g.*, array algorithms, doubling algorithms, CKMS algorithms) become evident. A somewhat different method was also proposed by Wilson (1969) — see Goodman et al. (1997) For interest, we only describe Bauer's method, using a simple example,

$$S_y(z) = 2z^{-1} + 5 + 2z,$$

Bauer (1955,1956) suggested forming a (semi-infinite) tri-diagonal Toeplitz matrix with 5 on the diagonal and 2 on the first sub- and superdiagonals, as shown below, and

determining its LDU factorization

$$\begin{bmatrix} 5 & 2 & & & \\ 2 & 5 & 2 & & \\ & 2 & 5 & 2 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} l_{00} & & & \\ l_{10} & l_{11} & & \\ & l_{21} & l_{22} & \\ & & \ddots & \ddots \end{bmatrix} D \begin{bmatrix} l_{00} & & & \\ l_{10} & l_{11} & & \\ & l_{21} & l_{22} & \\ & & \ddots & \ddots \end{bmatrix}^*$$

and $D = \text{diag}\{d_0, d_1, d_2, \ldots\}$. Then the reader can verify that as $i \to \infty$,

$$d_i \to r_e = 4, \quad l_{ii} + l_{i,i-1}z^{-1} \to 1 + \frac{1}{2}z^{-1}.$$

We might mention that if the covariance coefficients $\{R_y(i)\}$ are available, then even when $S_y(z)$ is not rational, the Levinson-Durbin (*cf.* Prob 4.11) and Schur algorithms (*cf.* App. F) can be used to obtain approximate spectral factors (as suggested, for example, by Whittle (1963, p. 37, p. 102)) even in the vector case — see also Dewilde and Dym (1981a,1981b). The article by Sayed and Kailath (2000) provides a unified treatment of several methods for spectral factorization.

Example 6.5.1 Suppose $S_y(z) = (2z+1)(2z^{-1}+1)$. To find r_e and $L(z)$, we first convert $S_y(z)$ to the standard form (6.5.1):

$$S_y(z) = 4\left(z + \frac{1}{2}\right)\left(z^{-1} + \frac{1}{2}\right),$$

from which we conclude that $r_e = 4$. Moreover, since $m = 1$ and $n = 0$, expression (6.5.3) gives

$$L(z) = z^{-1}\left(z + \frac{1}{2}\right) = \frac{\left(z + \frac{1}{2}\right)}{z} = 1 + \frac{1}{2}z^{-1},$$

so that $L(\infty) = 1$ and $L(z)$ has all poles and zeros inside the unit circle. ◆

Example 6.5.2 (Exponentially Correlated Process) Suppose $0 < a < 1$ and

$$S_y(z) = \sum_{i=-\infty}^{\infty} a^{|i|}z^{-i} = \frac{1-a^2}{(1-az^{-1})(1-az)}.$$

Then we have

$$L(z) = \frac{1}{1-az^{-1}}, \quad r_e = 1 - a^2.$$

◆

Example 6.5.3 (Second-Order (Minimum-Phase) ARMA Process) Consider again the second-order stationary process $\{\mathbf{y}_i\}$ of Prob. 5.3, *viz.*,

$$\mathbf{y}_{i+1} = a_0\mathbf{y}_i + a_1\mathbf{y}_{i-1} + \mathbf{u}_i + b\mathbf{u}_{i-1}, \quad i > -\infty,$$

with the same assumptions in that problem and with $|b| < 1$. The transfer function from $\mathbf{u}_i$ to $\mathbf{y}_i$ is

$$\frac{\mathbf{y}(z)}{\mathbf{u}(z)} = \frac{z+b}{z^2 - a_0z - a_1}.$$

Since $S_u(z) = Q$, the z-spectrum is

$$S_y(z) = Q\left(\frac{z+b}{z^2 - a_0z - a_1}\right)\left(\frac{z^{-1}+b}{z^{-2} - a_0z^{-1} - a_1}\right).$$

When $|b| < 1$, the canonical factorization is

$$r_e = Q, \quad L(z) = \frac{z(z+b)}{z^2 - a_0z - a_1},$$

a result obtained differently in Prob. 5.3. However, when $|b| > 1$, the reader should check that the canonical factorization is determined by

$$r_e = b^2Q, \quad L(z) = \frac{z\left(z + \frac{1}{b}\right)}{z^2 - a_0z - a_1},$$

as also obtained differently in Prob. 5.3. ◆

6.6 VECTOR-VALUED STATIONARY PROCESSES

As mentioned earlier, the spectral factorization problem is considerably more difficult for general vector-valued processes $\{\mathbf{y}_i\}$. To see this, let us examine the definition of z-spectra for vector-valued processes as well as the notion of canonical spectral factorization in this context.

Let $\{\mathbf{y}_i\}$ denote a p-dimensional zero-mean stationary random process with a matrix-valued covariance sequence

$$R_y(i) = E\mathbf{y}_j\mathbf{y}_{j-i}^*, \quad -\infty < i < \infty. \tag{6.6.1}$$

Then, of course, *i.e.*,

$$R_y(i) = R_y^*(-i), \quad -\infty < i < \infty, \tag{6.6.2}$$

and it is not hard to check that the matrices

$$\begin{bmatrix} R_y(0) & R_y(-1) & \ldots & R_y(-i) \\ R_y(1) & R_y(0) & \ldots & R_y(-i+1) \\ \vdots & \vdots & \ddots & \vdots \\ R_y(i) & R_y(i-1) & \ldots & R_y(0) \end{bmatrix}$$

are nonnegative-definite for all $i \geq 0$.

If we also make our (stronger than necessary) standard assumption that the covariance sequence is exponentially bounded, *i.e.*, that there exists a positive-definite constant *matrix* $K > 0$, and a positive scalar $\alpha < 1$, such that

$$R_y(i) < K\alpha^{|i|}, \quad -\infty < i < \infty,$$

then the power series defining the z-spectrum of $\{\mathbf{y}_i\}$,

$$S_y(z) \triangleq \sum_{i=-\infty}^{\infty} R_y(i)z^{-i} = S_y^*(z^{-*}),$$

converges absolutely in an annulus containing the unit circle, $\alpha < |z| < \alpha^{-1}$. In particular, the power spectral density matrix function

$$S_y(e^{j\omega}) = \sum_{i=-\infty}^{\infty} R_y(i)e^{-j\omega i}, \quad j \triangleq \sqrt{-1},$$

will exist for all $\omega \in [-\pi, \pi]$ and moreover,

$$S_y(e^{j\omega}) \geq 0, \quad -\pi \leq \omega \leq \pi, \tag{6.6.3}$$

where now nonnegativity is meant in the sense of matrices, *i.e.*, that $S_y(e^{j\omega})$ is a positive-semi-definite matrix. Moreover, the converse is also true: a matrix function $S_y(z)$ obeying these two properties must be the z-transform of a matrix-valued covariance sequence.

Although the matrix function $S_y(e^{j\omega})$ is nonnegative-definite for all $\omega \in [-\pi, \pi]$, this does not mean that it has constant rank for all frequencies ω. In fact, its rank on the unit circle will in general depend upon the value of ω and, in particular, on whether a given $e^{j\omega}$ is a unit-circle (transmission) zero of $S_y(z)$.[6] To see why, recall first that the *normal rank* of a matrix function $S_y(z)$ is defined as the maximum rank of $S_y(z)$ over all $z \in \mathbb{C}$. For example, the following matrix function

$$S_y(z) = \begin{bmatrix} (z + 0.5)(z^{-1} + 0.5) & 0 \\ 0 & 1 \end{bmatrix} \tag{6.6.4}$$

has normal rank 2 (although its rank is unity at $z = -0.5$). Now a complex number β is said to be a transmission zero of a $p \times p$ matrix function $S_y(z)$ with *full* normal rank (*i.e.*, with normal rank equal to p) if there exists a nonzero vector $w \in \mathbb{C}^p$, called the *left zero direction*, such that

$$wS_y(\beta) = 0. \tag{6.6.5}$$

In other words, the *transmission zeros* of $S_y(z)$ are those values of z for which the rank of $S_y(z)$ drops below p. It is important to note that, for matrix functions, transmission zeros must be identified together with their zero directions. The $S_y(z)$ in (6.6.4) has full normal rank and a zero at $\beta = -0.5$ along the direction $w = [1\ 0]$.

Fortunately, in the rational case that we are considering, the situation is relatively simple. In this case, it turns out that an $S_y(z)$ with full normal rank will have constant rank almost everywhere on the unit circle, *viz.*,

$$\text{rank}[S_y(z)] = p, \qquad \text{a.e. on the unit circle.}$$

The reason for this is simple: $S_y(e^{j\omega})$ can only drop rank at those values of ω for which $S_y(z)$ has an isolated unit-circle zero — since $S_y(z)$ is rational there are only finitely many such zeros, and so the rank is constant almost everywhere and given by the normal rank (assumed equal to p).

[6] There are a lot of (surprising) issues in a general treatment, which we make no attempt to cover here. The interested reader may refer to the following references for more detailed studies — Gohberg and Krein (1958), Popov (1964,1973), Rozanov (1967), Hannan (1970); Caines (1988) and Hannan and Deistler (1988) are somewhat more oriented towards an engineering audience.

In fact, as in the scalar case, we shall further assume that $S_y(z)$ has no unit-circle zeros, which is equivalent to $S_y(z)$ having constant rank p *everywhere* on the unit circle, *i.e.*,

$$\text{rank}[S_y(e^{j\omega})] = p, \qquad -\pi \leq \omega \leq \pi.$$

This is guaranteed by the assumption that $S_y(z)$ is positive-definite on the unit circle, *i.e.*, when

$$S_y(e^{j\omega}) > 0, \quad -\pi \leq \omega \leq \pi.$$

When this holds, *i.e.*, when $S_y(z)$ is rational and has maximal normal rank everywhere on the unit circle, one can show that it is always possible to perform the following canonical spectral factorization:

$$S_y(z) = L(z)\ R_e\ L^*(z^{-*}), \tag{6.6.6}$$

where

(i) $L(z)$ is a $p \times p$ rational matrix function that is analytic on and outside the unit circle ($|z| \geq 1$).
(ii) $L^{-1}(z)$ is analytic on and outside the unit circle ($|z| \geq 1$).
(iii) $L(\infty) = I_p$.
(iv) $R_e > 0$.

We should also mention that the normalization $L(\infty) = I_p$ makes the factorization (6.6.6) unique.[7] We shall provide a constructive proof in Sec. 8.3.3 for rational z-spectra described in state-space form — see Thm. 8.3.2.

The reader will note that this is simply the matrix analog of the canonical spectral factorization first introduced in (6.4.1). The causal and causally invertible transfer matrix $L(z)$ is referred to as the modeling filter and allows us to compute the innovations process $\{\mathbf{e}_i\}$ from the original process $\{\mathbf{y}_i\}$ via

$$\mathbf{e}(z) = L^{-1}(z)\mathbf{y}(z). \tag{6.6.7}$$

Finally, we should mention that if we do not rule out the possibility of unit circle zeros, then item (ii) in the above canonical factorization must be replaced by:[8]

(ii) $L^{-1}(z)$ is analytic outside the unit circle ($|z| > 1$).

[7] For nonrational finite-power spectra that have full normal rank a.e. on the unit circle, the following generalized Paley-Wiener condition will guarantee the existence of a canonical factor $L(z)$ such that it and its inverse are analytic in $|z| > 1$:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln \det\left[S_y(e^{j\omega})\right] d\omega > -\infty.$$

More details can be found in Rozanov (1967, Thm. 6.1).

[8] We are continuing to assume that $S_y(z)$ still has full normal rank a.e. on the unit circle. If this is not the case, then we need to replace (iv) by $R_e \geq 0$ as well. What this means is that in the rank deficient case, say

$$\text{rank}[S_y(e^{j\omega})] = m < p, \qquad \text{a.e. on the unit circle,}$$

there exists a $p \times m$ function $\bar{L}(z)$ analytic on and outside the unit circle ($|z| \geq 1$) such that (see, *e.g.*, Rozanov (1967))

$$S_y(z) = \bar{L}(z)\bar{L}^*(z^{-*}).$$

Computing the Spectral Factors. A natural question is how to actually compute the canonical factor $L(z)$ of (6.6.6) in the vector-valued case. In the scalar case this was in principle straightforward: compute the poles and zeros of $S_y(z)$ and retain the stable ones in the canonical factor. Unfortunately, generalizing this method to the matrix case is not straightforward. We have already had an inkling of why this is so — unlike the scalar case, zeros and poles alone *do not* determine $S_y(z)$ (one needs, in addition, to consider the zero directions).

The first complete approach for computing the canonical factorization in the rational matrix-valued case is perhaps due to Youla (1961) , whose method is based on the so-called Smith-McMillan canonical form of rational matrices. Other methods were described by Yaglom (1960), Rozanov (1960), Davis (1963); see also Prob. 6.16. Moreover, these methods are all algebraically very cumbersome, and one tends to "miss the forest for the trees"; in part because of this, they can all present various numerical difficulties. We may also mention here the use of iterative methods as noted at the end of Sec. 6.5 (*e.g.*, Whittle (1963, p. 102), Wilson (1972), and Rissanen and Kailath (1972)).

In Ch. 8 we shall show that by using state-space models for the process $\{\mathbf{y}_i\}$, the canonical spectral factorization can be computed by solving a discrete-time algebraic Riccati equation, which is essentially a certain system of *nonlinear* algebraic equations. Although this may not at first appear to be a simplification of the problem, various effective methods have been proposed for solving the Riccati equation (see App. E), so that it appears to be the currently most computationally satisfactory way of computing the canonical spectral factorization in the rational matrix-valued case. Of course, this method can also be effective in the scalar case.

6.7 COMPLEMENTS

The problem of finding innovations for stationary discrete-time scalar-valued processes was first studied in a famous dissertation by H. Wold in 1938 (published as Wold (1954)), and then in great generality by A. N. Kolmogorov (1939 and 1941). We gave a simplified presentation here, making stronger assumptions than needed — general treatments can now be found in any of several textbooks on stochastic processes, *e.g.*, Doob (1953), Grenander and Rosenblatt (1957). The vector case is discussed in Rozanov (1967) and Hannan (1970).

PROBLEMS

6.1 **(A doubly infinite triangular Toeplitz matrix)** Consider a lower triangular Toeplitz matrix A_N of dimensions $(2N+1) \times (2N+1)$, and whose first column is $\text{col}\{1, (a-b), a(a-b), \ldots, a^{2N-1}(a-b)\}$. Assume that $0 < a < 1$ and $b > 1$.

(a) Show that for every finite N, $B_N \triangleq A_N^{-1}$ is also a lower triangular Toeplitz matrix whose first column is $\text{col}\{1, (b-a), b(b-a), \ldots, b^{2N-1}(b-a)\}$.

(b) Show that, as $N \to \infty$, some of the elements of B_N become unbounded, whereas the elements of A_N all remain bounded. Is it then possible to conclude that

$$\left[\lim_{N\to\infty} A_N\right]^{-1} = \lim_{N\to\infty} B_N \, ?$$

(c) Consider now the $(2N+1) \times (2N+1)$ upper triangular Toeplitz matrix C_N whose last column is

$$\text{col}\{(a-b)/b^{2N-1}, (a-b)/b^{2N-2}, \ldots, (a-b)/b^2, a/b\},$$

and define the matrix W_N as $[W_{N,ij}]_{i,j=-N}^{N} = W_N = A_N C_N$. Note that we have numbered the rows and columns of W_N from $-N$ to N. Use the facts that

$$A_{N,ij} = \begin{Bmatrix} 0 & i < j \\ 1 & i = j \\ a^{i-j-1}(a-b) & i > j \end{Bmatrix} \text{ and } C_{N,ij} = \begin{Bmatrix} \frac{1}{b^{j-i+1}}(a-b) & i < j \\ \frac{a}{b} & i = j \\ 0 & i > j \end{Bmatrix}$$

to show that the entries of W_N are given by

$$W_{N,ij} = \begin{cases} \dfrac{a-b}{b^{j-i+1}} + \dfrac{(a-b)^2}{ab} \cdot \dfrac{1}{b^{j-i}} \cdot \dfrac{(\frac{a}{b})^{i+N-1}}{1-\frac{b}{a}} & i < j, \\ \dfrac{a}{b} + \dfrac{(a-b)^2}{ab} \cdot \dfrac{(\frac{a}{b})^{i+N-1}}{1-\frac{b}{a}} & i = j, \\ \dfrac{a^{i-j}(a-b)}{b} + \dfrac{(a-b)^2 a^{i-j}}{ab} \cdot \dfrac{(\frac{a}{b})^{j+N-1}}{1-\frac{b}{a}} & i > j. \end{cases}$$

(d) Show that as $N \to \infty$, we have

$$W_{N,ij} \to \begin{cases} 0 & i < j, \\ 1 & i = j, \\ 0 & i > j. \end{cases}$$

(e) Show that the entries of C_N remain bounded as $N \to \infty$, and conclude that

$$\left[\lim_{N\to\infty} A_N\right]^{-1} = \lim_{N\to\infty} C_N.$$

(f) This means that the inverse of the doubly infinite lower triangular Toeplitz matrix $(\lim_{N\to\infty} A_N)$ is *upper triangular*! Note also one more surprising fact: the matrix $\lim_{N\to\infty} A_N$ has unit diagonal, whereas its inverse has $\frac{a}{b}$ as its diagonal entries. Explain the above results using the generating function $(z-b)/(z-a)$ with $0 < a < 1$ and $b > 1$.

6.2 **(A harmonic proces)** Let $\mathbf{a}$ be a Rayleigh distributed random variable, *i.e.*, its probability density function is given by

$$f_A(a) = \frac{a}{\sigma_a^2} \exp\left(\frac{-a^2}{2\sigma_a^2}\right), \quad a \geq 0.$$

Now assume a random sequence $\{\mathbf{x}_i\}$ is generated via $\mathbf{x}_i = \mathbf{a}\cos(\omega_0 i + \boldsymbol{\phi})$, where ω_0 is a known frequency in the range $[-\pi, \pi]$ and $\boldsymbol{\phi}$ is a uniformly distributed random variable in $[-\pi, \pi]$. Both $\boldsymbol{\phi}$ and $\mathbf{a}$ are assumed independent.

(a) Find $E\mathbf{a}$, $E\mathbf{a}^2$, and $E(\mathbf{a} - E\mathbf{a})^2$.

(b) Show that $\{\mathbf{x}_i\}$ is wide-sense stationary with zero mean, and covariance sequence

$$R_x(n) \triangleq \langle \mathbf{x}_k, \mathbf{x}_{k-n} \rangle = \sigma_a^2 \cos(\omega_0 n).$$

[That is, both the process and its covariance sequence have the same frequency.] Verify that $S_x(e^{j\omega}) = \sigma_a^2 \pi [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]$, where $\delta(\cdot)$ is the delta function.

(c) Show that the innovations of the process $\{\mathbf{x}_i\}$ are identically zero.

6.3 **(A discrete spectrum)** Consider a covariance sequence that is the sum of two complex sinusoidal signals, say $R_y(k) = a_0 e^{jk\omega_0} + a_1 e^{jk\omega_1}$, for some positive constants $\{a_0, a_1\}$ and with $0 \le \omega_0, \omega_1 < 2\pi$, $\omega_1 \ne \omega_2$.

(a) Verify that the corresponding power spectral density function is discrete and given by

$$S(e^{j\omega}) = 2\pi [a_0 \delta(\omega - \omega_0) + a_1 \delta(\omega - \omega_1)].$$

(b) Does $S(e^{j\omega})$ satisfy the Paley-Wiener condition (6.4.2)?

(c) Let T_i be the Hermitian Toeplitz covariance matrix whose first column is given by $\text{col}\{R_y(0), R_y(1), \ldots, R_y(i)\}$. Is T_i positive-definite for all i?

6.4 **(A band-limited spectrum)** Consider a continuous-time random process $\mathbf{x}(\cdot)$ whose power spectrum is constant and equal to $N_0/2$ for all frequencies f in the range $[-B, B]$, and zero otherwise. Show that the autocorrelation function of $\mathbf{x}(\cdot)$ is given by

$$R_x(\tau) = \frac{N_0 \sin(2\pi B\tau)}{2\pi\tau}.$$

Consider the discrete-time process $\mathbf{y}_i = \mathbf{x}\left(t = \frac{i}{2B}\right)$. That is, $\{\mathbf{y}_i\}$ is obtained by sampling $\mathbf{x}(t)$ at multiples of $1/2B$. Determine the innovations of $\{\mathbf{y}_i\}$.

6.5 **(Unit-circle zeros)** Consider a zero-mean wide-sense stationary random process $\{\mathbf{y}_i\}$ with the rational z-spectrum $S_y(z) = (1 - z)(1 - z^{-1})$.

(a) Show that $S_y(z)$ satisfies the Paley-Wiener condition (6.4.2). Show also that the process has finite power (*cf.* (6.3.5)).

(b) Find the unique canonical spectral factor $L(z)$, and its inverse $L^{-1}(z)$. Verify that they are both analytic in $|z| > 1$.

(c) Show that the impulse response of $L(z)$ is square-summable, while that of $L^{-1}(z)$ is not.

(d) Does the series

$$\mathbf{e}_i = \mathbf{y}_i + \sum_{k=1}^{\infty} w_k \mathbf{y}_{i-k},$$

in (6.1.4) converge in the mean-square sense? Can it be used to describe the innovations of $\{\mathbf{y}_i\}$?

(e) Let $\hat{\mathbf{y}}_{i+1|i,i-1,\ldots,i-n}$ denote the predictor of $\mathbf{y}_{i+1}$ given the past $(n+1)$ observations $\{\mathbf{y}_i, \mathbf{y}_{i-1}, \ldots, \mathbf{y}_{i-n}\}$, say

$$\hat{\mathbf{y}}_{i+1|i,i-1,\ldots,i-n} = \sum_{j=0}^{n} \alpha_j \mathbf{y}_{i-j},$$

for some coefficients $\{\alpha_j\}$. Show that $\alpha_j = -(n - j + 1)/(n + 2)$ and $\|\mathbf{y}_{i+1} - \hat{\mathbf{y}}_{i+1|i,\ldots,i-n}\|^2 = (n+2)/(n+1)$. Conclude that

$$\hat{\mathbf{y}}_{i+1} = \lim_{n\to\infty} \left[-\sum_{j=0}^{n} \frac{n - j + 1}{n + 2} \mathbf{y}_{i-j} \right].$$

Remark. This problem shows the difficulty in evaluating the innovations of processes with unit-circle spectral zeros. The limit in part (e) is referred to as a *first Cesàro sum* and it converges in the mean-square sense. We should mention that in practice, due to round-off errors, Cesàro sums cannot be computed and generally result in limits with infinite variance. Thus, the above discussions are mostly of theoretical interest and have little practical value. In fact, one may claim that processes with unit-circle spectral zeros cannot be found in practice, as even the slightest perturbation can result in the zeros moving away from the unit circle. ◆

6.6 **(Covariance sequences)** Let $S_y(z)$ denote a function that obeys the two properties of (para-)Hermitian symmetry, $S_y(z) = S_y^*(z^{-*})$, and nonnegativity on the unit circle, $S_y(e^{j\omega}) \ge 0$. Define the inverse discrete-time Fourier transform (DTFT) sequence

$$R_y(i) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} S_y(e^{j\omega}) e^{j\omega i} d\omega, \qquad j \triangleq \sqrt{-1}, \qquad -\infty < i < \infty.$$

Show that $\{R_y(i)\}$ is a covariance sequence.

[*Hint.* For any sequence of complex numbers $\{a_i\}$, and for any N, let $A_N(e^{j\omega}) = \sum_{i=0}^{N} a_i e^{-j\omega i}$ and verify that

$$\sum_{i=0}^{N} \sum_{k=0}^{N} a_i^* a_k R_y(i - k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |A_N(e^{j\omega})|^2 S_y(e^{j\omega}) d\omega.]$$

Remark. Let T_i be the Hermitian Toeplitz covariance matrix whose first column is given by $\text{col}\{R_y(0), R_y(1), \ldots, R_y(i)\}$. The problem thus shows that $S(e^{j\omega}) \ge 0$ implies $T_i \ge 0$ for all i. ◆

6.7 **(Nonnegativity of the spectral density function)** Consider a covariance sequence $R_y(i)$ that decays exponentially to zero, say $|R_y(i)| \le K\alpha^{|i|}$ for some constant K and real number $\alpha \in (0, 1)$. Let T_N be the Hermitian Toeplitz covariance matrix whose first column is $\text{col}\{R_y(0), R_y(1), \ldots, R_y(N)\}$.

Pick any finite scalar a and define $b = \text{col}\{a, ae^{-j\omega}, ae^{-j2\omega}, \ldots, ae^{-jN\omega}\}$, where $j = \sqrt{-1}$. Establish that

$$0 \le \frac{1}{N} b^* T_N b = a^* \left[\sum_{i=-N}^{N} R_y(i) e^{-j\omega i} \right] a - \frac{1}{N} \sum_{i=-N}^{N} |i| a^* R_y(i) a e^{-j\omega i}.$$

Remark. Now we can take the limit as $N \to \infty$ to conclude that $S_y(e^{j\omega}) \ge 0$. This problem thus shows that $T_i \ge 0$ for all i implies $S(e^{j\omega}) \ge 0$. The assumption of an exponentially bounded sequence $\{R_y(i)\}$ can be relaxed to establish that $T_i \ge 0$ for all i implies $S(e^{j\omega}) \ge 0$ almost everywhere (see, *e.g.*, Doob (1953) or Grenander and Szegö (1958)). ◆

6.8 **(Paley-Wiener condition)** Let $S_y(z)$ be a z-spectrum, possibly nonrational, that satisfies the Paley-Wiener condition (6.4.2). Let T_i be the Hermitian Toeplitz covariance matrix whose first column is $\text{col}\{R_y(0), R_y(1), \ldots, R_y(i)\}$. We want to show that the Paley-Wiener condition guarantees $T_i > 0$ for all i.

(a) Assume T_i is singular for some i, say $T_i a = 0$ for some nonzero column vector a with entries $\{a_k\}$. Show that this implies

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} S_y(e^{j\omega})\left|\sum_{k=0}^{i} a_k e^{-jk\omega}\right|^2 d\omega \;=\; 0, \qquad j=\sqrt{-1}.$$

(b) Conclude that the above implies $S_y(e^{j\omega}) = 0$ almost everywhere, which contradicts the Paley-Wiener condition (6.4.2).

Remark. The argument further shows that if one starts with a $z-$spectrum $S_y(z)$ that is assumed to be strictly positive on the unit circle, then it will also follow that $T_i > 0$ for all i. This is because if T_i were singular for some i, then by part (b) above we would be able to conclude that $S_y(e^{j\omega}) = 0$ a. e., a contradiction. ◆

6.9 **(Eigenvalues of covariance matrices)** Let $\{R_y(i)\}$ denote the covariance sequence of a doubly infinite stationary process $\{\mathbf{y}_i\}$. Let T_i denote a Hermitian Toeplitz matrix whose first column is $\text{col}\{R_y(0), R_y(1), \ldots, R_y(i)\}$. Show that, for all i,

$$\lambda_{\max}(T_{i+1}) \geq \lambda_{\max}(T_i) \qquad \text{and} \qquad \lambda_{\min}(T_{i+1}) \leq \lambda_{\min}(T_i),$$

where $\lambda_{\max}$ and $\lambda_{\min}$ denote the maximum and minimum eigenvalues of their arguments. [*Hint.* For any matrix A, it holds that

$$\lambda_{\max}(A) = \sup_{\|x\|=1} x^*Ax, \qquad \lambda_{\min}(A) = \inf_{\|x\|=1} x^*Ax.]$$

6.10 **(Strong positivity)** Consider a z-spectrum $S_y(z)$ that is strongly positive on the unit circle, say $S_y(e^{j\omega}) > \epsilon > 0$ for all $\omega \in [-\pi, \pi]$ and for some positive ϵ. Let T_i denote a Hermitian Toeplitz matrix whose first column is $\text{col}\{R_y(0), R_y(1), \ldots, R_y(i)\}$. We know from Prob. 6.6 that $T_i \geq 0$ for all i. We want to show that, under strong positivity, it holds that $T_i > \epsilon I$ for all i.

(a) For any i, let a be a unit-norm eigenvector that corresponds to the smallest eigenvalue of T_i, say $T_i a = \lambda_{\min} a$ with $\|a\|^2 = 1$. Let $\{a_k\}$ denote the entries of a. Use (6.3.4) to show that

$$\lambda_{\min} \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} S_y(e^{j\omega})\left|\sum_{k=0}^{i} a_k e^{-jk\omega}\right|^2 d\omega, \qquad j=\sqrt{-1}.$$

(b) Use Parseval's theorem, *viz.*,

$$\sum_{k=0}^{i}|h_k|^2 \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi}|H(e^{j\omega})|^2 d\omega \qquad \text{where} \qquad H(e^{j\omega}) \triangleq \sum_{k=0}^{i} h_k e^{-j\omega k},$$

to conclude that $\lambda_{\min} > \epsilon$.

6.11 **(Extreme values of the power spectrum)** Refer to Prob. 6.7.

(a) Argue that $S_y(e^{j\omega})$ is a continuous function of ω in $[-\pi, \pi]$.

(b) Using $\lambda_{\max}(T_N) \geq b^*T_N b/\|b\|^2$, and the equality established in Prob. 6.7, show that for all N,

$$\lambda_{\max}(T_N) \;\geq\; \max_{\omega\in[-\pi,\pi]} S_y(e^{j\omega}).$$

Use a similar argument to establish that

$$\lambda_{\min}(T_N) \;\leq\; \min_{\omega\in[-\pi,\pi]} S_y(e^{j\omega}).$$

(c) Show that

$$\lim_{N\to\infty}\lambda_{\max}(T_N) \;=\; \max_{\omega\in[-\pi,\pi]} S_y(e^{j\omega}), \qquad \lim_{N\to\infty}\lambda_{\min}(T_N) \;=\; \min_{\omega\in[-\pi,\pi]} S_y(e^{j\omega}).$$

6.12 **(Kolmogorov's exp-log method)** In this problem we deduce the canonical factorization and the Szegö formula (6.4.3) under the assumption that $\ln\left[S_y(z)\right]$ is analytic in an annulus that includes the unit circle, $|z| = 1$, so that it can be expanded in a Laurent series, say

$$\ln\left[S_y(z)\right] = \sum_{j=-\infty}^{\infty}\gamma_j z^{-j}.$$

(a) Show that the unique canonical factorization (6.4.1) of $S_y(z)$ is given by

$$L(z) = \exp\left[\sum_{j=1}^{\infty}\gamma_j z^{-j}\right] \qquad \text{and} \qquad r_e = \exp[\gamma_0].$$

(b) Show that

$$\ln r_e = \gamma_0 \triangleq \left[\frac{1}{j2\pi}\oint_C \ln\left[S_y(z)\right]\frac{dz}{z}\right],$$

to deduce the formula (6.4.3) for r_e.

Remark. This method, perhaps first given by Kolmogorov (1941a), is often called the *cepstral* method. ◆

6.13 **(Minimum-delay property)** Consider two stable and causal transfer functions (*i.e.*, analytic outside a circular region that includes the unit circle),

$$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + \ldots \quad \text{and} \quad G(z) = g_0 + g_1 z^{-1} + g_2 z^{-2} + \ldots$$

such that $H(z)H^*(z^{-*}) = G(z)G^*(z^{-*})$. In particular, this means that $|H(e^{j\omega})|^2 = |G(e^{j\omega})|^2$ so that both systems have the same magnitude response. It also means, in view of Parseval's theorem, *viz.*,

$$\sum_{i=0}^{\infty}|h_i|^2 \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi}|H(e^{j\omega})|^2 d\omega,$$

that both impulse response sequences, $\{h_i\}$ and $\{g_i\}$, have the same total energies. Now assume that $H(z)$ is minimum-phase, *i.e.*, that $H^{-1}(z)$ is also analytic outside a circular

region that includes the unit circle. We want to establish that, for any $G(z)$ as above, it holds that

$$\sum_{i=0}^{N} |g_i|^2 \leq \sum_{i=0}^{N} |h_i|^2 \quad \text{for all } N.$$

We can interpret this to mean that among all systems with the same magnitude response, the minimum-phase system is the one that yields the most energy first. For this reason, minimum-phase systems were called minimum-delay systems by E. A. Robinson (1963), who first discovered this property.

(a) Show that $\Delta(z) = H^{-1}(z)G(z)$ is a lossless or all-pass system , *i.e.,* that $\Delta(z)$ is stable, causal, and satisfies $\Delta(z)\Delta^*(z^{-*}) = 1$. Let $\{\delta_i,\ 0 \leq i < \infty\}$ denote its impulse response. Conclude that $\sum_{i=0}^{\infty} |\delta_i|^2 = 1$.

(b) For any integer N, define the partial transfer functions

$$G_N(z) = \sum_{i=0}^{N} g_i z^{-i}, \quad H_N(z) = \sum_{i=0}^{N} h_i z^{-i}, \quad \Delta_N(z) = \sum_{i=0}^{N} \delta_i z^{-i}.$$

Using the relation $G(z) = H(z)\Delta(z)$, and the fact that $\{G(z), H(z), \Delta(z)\}$ are all causal functions, argue that $G_N(z) = H_N(z)\Delta_N(z)$.

(c) Using Parseval's theorem, and the triangle inequality of norms, verify that the following holds:

$$\sum_{i=0}^{N} |g_i|^2 \leq \|H_N(z)\|^2 \cdot \|\Delta_N(z)\|^2 \leq \sum_{i=0}^{N} |h_i|^2,$$

where we are defining the squared norm of a stable function $K(z)$ by

$$\|K(z)\|^2 \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} |K(e^{j\omega})|^2 d\omega.$$

6.14 (Rational z-spectra) As shown in Sec. 6.5, a rational power spectrum $S_y(e^{j\omega})$ of a finite power process $\{\mathbf{y}_i\}$ can be written as

$$S_y(e^{j\omega}) = r_e \frac{|L_p(e^{j\omega})|^2}{|L_q(e^{j\omega})|^2},$$

where $L_p(z)$ and $L_q(z)$ are polynomials in z^{-1} with all their roots strictly inside the unit circle. Moreover, $L_p(\infty) = L_q(\infty) = 1$.

(a) Show that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |L_p(e^{j\omega})|^2 \, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |L_q(e^{j\omega})|^2 \, d\omega = 0.$$

(b) Conclude that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[S_y(e^{j\omega})] d\omega = \ln r_e,$$

and, hence, that every rational z-spectrum satisfies the Paley-Wiener condition (6.4.2).

6.15 (An optimization problem) Consider a Laurent polynomial $P(z)$, of degree m,

$$P(z) \triangleq \sum_{i=-m}^{m} p_i z^{-i}, \qquad p_i = p_{-i}^*,$$

and such that it is nonnegative on the unit circle, $P(e^{j\omega}) \geq 0$. Let $P(z) = L_p(z) r_p L^*(z^{-*})$ denote its canonical spectral factorization with $r_p > 0$, and $L_p(z)$ having all its roots are in $|z| \leq 1$. Let $g(z)$ denote any polynomial of degree at most m in z^{-1}, *viz.,*

$$g(z) = 1 + g_1 z^{-1} + g_2 z^{-2} + \ldots + g_k z^{-k}, \qquad k \leq m.$$

Show that the solution of the optimization problem:

$$\min_{g(z)} \int_{-\pi}^{\pi} \frac{|g(e^{j\omega})|^2}{P(e^{j\omega})} \, d\omega,$$

is given by the canonical spectral factor $L_p(z)$ of $P(z)$.

6.16 (Spectral factorization in the vector case) Consider a $p \times p$ rational z-spectrum $S_y(z)$ of a finite-power stationary process $\{\mathbf{y}_i\}$. Assume that $S_y(e^{j\omega}) > 0$ for all $\omega \in [-\pi, \pi]$. This problem follows an argument in Hassibi, Sayed, and Kailath (1999) to establish the existence of canonical spectral factors for such z-spectra.

(a) Suppose first that $p = 2$ and partition $S_y(z)$ as

$$S_y(z) = \begin{bmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ \Sigma_{21}(z) & \Sigma_{22}(z) \end{bmatrix}.$$

Introduce the Schur complement $\Sigma_\Delta(z) = \Sigma_{22}(z) - \Sigma_{21}(z)\Sigma_{11}^{-1}(z)\Sigma_{12}(z)$. Show that $0 < \Sigma_\Delta(e^{j\omega}) < \infty$. That is, $\Sigma_\Delta(z)$ is bounded and strictly positive on the unit circle.

(b) Introduce the canonical spectral factorizations of $\Sigma_{11}(z)$ and $\Sigma_\Delta(z)$, say

$$\Sigma_{11}(z) = r_1 \frac{a_1(z)}{b_1(z)} \frac{a_1^*(z^{-*})}{b_1^*(z^{-*})}, \qquad \Sigma_\Delta(z) = r_\delta \frac{a_\delta(z)}{b_\delta(z)} \frac{a_\delta^*(z^{-*})}{b_\delta^*(z^{-*})},$$

where $r_1 > 0$, $r_\delta > 0$, and the ratios $a_1(z)/b_1(z)$ and $a_\delta(z)/b_\delta(z)$ are minimum-phase (*i.e.,* both ratios and their inverses are analytic in $|z| \geq 1$). Conclude that $S_y(z)$ can be factored as $S_y(z) = B(z)B^*(z^{-*})$, where

$$B(z) = \begin{bmatrix} \sqrt{r_1} \frac{a_1(z)}{b_1(z)} & 0 \\ 0 & \sqrt{r_\delta} \frac{a_\delta(z)}{b_\delta(z)} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ g(z) & 1 \end{bmatrix},$$

and $g(z) \triangleq \frac{1}{\sqrt{r_\delta}} \frac{b_\delta(z)}{a_\delta(z)} \Sigma_{21}(z) \frac{1}{\sqrt{r_1}} \frac{b_1^*(z^{-*})}{a_1^*(z^{-*})}$. Is $B(z)$ minimum-phase?

(c) Argue that $g(z)$ is analytic on $|z| = 1$ so that it is a well-defined z-transform, say

$$g(z) = \sum_{i=-\infty}^{\infty} g_i z^{-i},$$

for some coefficients $\{g_i\}$ that are absolutely summable. Let $T(z)$ denote the strictly anticausal part of $g(z)$, written as

$$T(z) \triangleq \{g(z)\}_- = \sum_{i=-\infty}^{-1} g_i z^{-i}.$$

Hence, $T(z)$ is analytic in $|z| \leq 1$. Show that the following matrix and its inverse,

$$\begin{bmatrix} \sqrt{r_1}\frac{a_1(z)}{b_1(z)} & 0 \\ 0 & \sqrt{r_\delta}\frac{a_\delta(z)}{b_\delta(z)} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ g(z) - T(z) & 1 \end{bmatrix},$$

are analytic in $|z| \geq 1$.

(d) Let

$$\Theta(z) = \begin{bmatrix} \Theta_{11}(z) & \Theta_{12}(z) \\ \Theta_{21}(z) & \Theta_{22}(z) \end{bmatrix}$$

be a 2×2 rational matrix function such that both

$$\begin{bmatrix} 1 & 0 \\ T(z) & 1 \end{bmatrix} \Theta(z) \quad \text{and} \quad \Theta^*(z^{-*}) \begin{bmatrix} 1 & 0 \\ -T(z) & 1 \end{bmatrix},$$

are analytic in $|z| \geq 1$. Show that $\{\Theta_{11}(z), \Theta_{12}(z)\}$ must be analytic in $|z| \geq 1$, while $\{\Theta_{21}(z), \Theta_{22}(z)\}$ must be analytic in $|z| \leq 1$. Show further that the following relations must hold:

$$\begin{aligned} \Theta_{11}(z) - \{T^*(z^{-*})\Theta_{21}(z)\}_+ &= d_{11}, \\ \Theta_{21}(z) + \{T(z)\Theta_{11}(z)\}_{a.c} &= d_{21}, \\ \Theta_{12}(z) - \{T^*(z^{-*})\Theta_{22}(z)\}_+ &= d_{12}, \\ \Theta_{22}(z) + \{T(z)\Theta_{12}(z)\}_{a.c.} &= d_{22}, \end{aligned}$$

for some scalars $\{d_{11}, d_{12}, d_{21}, d_{22}\}$, where the notations $\{\cdot\}_+$ and $\{\cdot\}_{a.c.}$ represent the causal and anticausal parts of their arguments, *viz.*,

$$\{g(z)\}_+ = \sum_{i=0}^{\infty} g_i z^{-i}, \quad \{g(z)\}_{a.c.} = \sum_{i=-\infty}^{0} g_i z^{-i}.$$

(e) Show that, for any $\{d_{11}, d_{21}, d_{12}, d_{22}\}$, a matrix function $\Theta(z)$ that satisfies the above conditions can always be found. Show further that $\Theta^*(z^{-*})\Theta(z) = A$, a constant nonegative-definite matrix for all z.

(f) Show that with an appropriate choice of $\{d_{11}, d_{12}, d_{21}, d_{22}\}$, the matrix A can be made positive-definite. Conclude that the matrix function

$$\bar{L}(z) \triangleq \begin{bmatrix} \sqrt{r_1}\frac{a_1(z)}{b_1(z)} & 0 \\ 0 & \sqrt{r_\delta}\frac{a_\delta(z)}{b_\delta(z)} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ g(z) - T(z) & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ T(z) & 1 \end{bmatrix} \Theta(z) A^{-*/2},$$

is minimum-phase and satisfies $S_y(z) = \bar{L}(z)\bar{L}^*(z^{-*})$.

(g) Extend the argument by induction to $p > 2$.

(h) Extend the argument to continuous-time spectra $S_y(s)$ that are positive-definite on the imaginary axis.

Remark. This method may be numerically feasible (though this requires study), but it is algebraically and conceptually involved. We shall see in Ch. 8 how the introduction of state-space models at the least reduces the conceptual burden. ◆

Appendix for Chapter 6

i.A CONTINUOUS-TIME SYSTEMS AND PROCESSES

In this appendix we briefly note analogs for continuous time of the main results reviewed in the text for discrete-time linear systems and stationary processes.

The Laplace Transform. The *biltateral Laplace transform* of the function $u(\cdot)$ is defined as

$$U(s) \triangleq \int_{-\infty}^{\infty} u(t)e^{-st}dt, \quad s = \sigma + j\omega. \tag{6.A.1}$$

As in discrete time, the region of convergence (ROC) of the above Laplace transform is defined as those values of the complex variable s for which the above integral converges *absolutely*. In this appendix we shall assume that the time functions under consideration are exponentially bounded, *i.e.*, that there exist α and K such that for all t

$$|u(t)| < Ke^{-\alpha|t|}, \quad \alpha > 0, \quad K > 0. \tag{6.A.2}$$

Under the above assumption the ROC will be (a strip in the complex plane containing the imaginary axis) of the form $-\alpha < \mathrm{Re}(s) < \alpha$.

The original function can always be recovered from its Laplace transform using the *inverse Laplace transform*,

$$u(t) = \frac{1}{j2\pi}\int_{\sigma-j\infty}^{\sigma+j\infty} U(s)e^{st}ds, \quad j = \sqrt{-1}, \tag{6.A.3}$$

where the line $\mathrm{Re}(s) = \sigma$ belongs to the ROC of $U(s)$. [Of course, in various special cases, the inverse transform can be found more directly, *e.g.*, from a table of transforms or by partial fraction expansions in the rational case.] However, it is important to note that given any function, $U(s)$, its inverse Laplace transform, $u(t)$, will depend on the ROC that we choose. Under the exponential boundedness assumption, the ROC must include the imaginary axis, so that we may write

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} U(j\omega)e^{j\omega t}d\omega. \tag{6.A.4}$$

Linear Time-Invariant Systems. A continuous-time linear time-invariant (LTI) system maps an input function $u(\cdot)$ to an output function $y(\cdot)$ according to the convolution rule

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)u(\tau)d\tau = \int_{-\infty}^{\infty} h(\tau)u(t-\tau)d\tau, \tag{6.A.5}$$

where the function $h(\cdot)$ is known as the *impulse response* of the system.

Another characterization of an LTI system is through its transfer function, $H(s)$, defined as the bilateral Laplace transform of $h(\cdot)$, *i.e.*,

$$H(s) = \int_{-\infty}^{\infty} h(t)e^{-st}dt.$$

An important reason for introducing transfer functions is that the convolution (6.A.5) can be replaced by multiplication in the Laplace transform domain,

$$Y(s) = H(s)U(s). \tag{6.A.6}$$

The ROC of $Y(s)$ will be the intersection of the ROCs of $H(s)$ and $U(s)$, *i.e.*, it too will be a strip containing the imaginary axis.

Our main interest will be in rational transfer functions, namely those having the form

$$H(s) = \frac{b_0s^m + b_1s^{m-1} + \ldots + b_m}{a_0s^n + a_1s^{n-1} + \ldots + a_n} \triangleq \frac{b(s)}{a(s)},$$

where we further assume that the polynomials $\{b(s), a(s)\}$ are coprime, *i.e.*, they have no factors in common. In this case, the roots of the denominator and numerator polynomials are called the poles and the zeros of the system, respectively.

Another important concept is *stability*. An LTI system is said to be a BIBO stable system if it maps bounded input signals into bounded output signals. It can be shown that BIBO stability is equivalent to the absolute integrability of the impulse response,

$$\int_{-\infty}^{\infty} |h(t)|dt < \infty,$$

which in turn is equivalent to the requirement that the ROC of the transfer function $H(s)$ should include the imaginary axis, $\mathrm{Re}(s) = 0$.

Causal, Anticausal, and Minimum-Phase Systems. A time function, $u(\cdot)$, is called *causal* if it is zero for negative time, *i.e.*, if $u(t) = 0$ for $t < 0$. It is referred to as *strictly causal* if it is zero for $t \leq 0$. Our standing assumption of exponential boundedness means that the Laplace transform of a causal function is analytic in a domain of the form $\mathrm{Re}(s) > -\alpha$, which lies in the right-half plane (RHP) and also includes the imaginary axis. When $U(s)$ is rational, causality is equivalent to $U(s)$ having all its poles strictly within the left-half-plane (LHP). Similar statements apply to *anticausal* time functions, with $t > 0$ replaced by $t < 0$ and the RHP replaced by the LHP.

A linear system is called (strictly) causal if its impulse response, $h(\cdot)$, is (strictly) causal. Note that the input-output relation (6.A.5) implies that causal systems map causal inputs to causal outputs; more explicitly, we may write,

$$y(t) = \int_0^t h(t-\tau)u(\tau)d\tau = \int_0^t h(\tau)u(t-\tau)d\tau, \quad t \geq 0.$$

Moreover, a stable linear system is causal if its transfer function, $H(s)$, is analytic in a domain of the form $\mathrm{Re}(s) > -\alpha$, which includes the RHP. When $H(s)$ is rational, this means that all the poles of the transfer function lie strictly within the LHP.

The *inverse* of an LTI system is one that maps the output of the original system to the input of the original system; its transfer function is the inverse of the transfer function of the original system.

Finally, an LTI system is called *minimum-phase* if both it and its inverse are stable and causal. This means that both $H(s)$ *and* $H^{-1}(s)$ must be analytic in a domain of the form $\mathrm{Re}(s) > -\alpha$, which includes the RHP. When $H(s)$ is rational, all the poles and zeros of $H(s)$ must lie strictly in the LHP. This is further equivalent to saying that $H(s)$ *and* $H^{-1}(s)$ are analytic in the closed RHP (which includes the imaginary axis).

Stochastic Processes. The covariance function of a continuous-time zero-mean stochastic process $\mathbf{y}(\cdot)$ is defined as

$$\langle \mathbf{y}(t), \mathbf{y}(\tau) \rangle = R_y(t,\tau) = R_y^*(\tau,t), \quad -\infty < t,\tau < \infty.$$

When $\mathbf{y}(\cdot)$ is a zero-mean (wide-sense) stationary process, its covariance function depends only on the time difference $t-\tau$, *i.e.*,

$$\langle \mathbf{y}(t), \mathbf{y}(\tau) \rangle = R_y(t-\tau), \quad -\infty < t,\tau < \infty.$$

The s-spectrum of $\mathbf{y}(\cdot)$ is defined as the Laplace transform of the covariance function,

$$S_y(s) \triangleq \int_{-\infty}^{\infty} R_y(t) e^{-st} dt = S_y^*(-s^*), \tag{6.A.7}$$

which is analytic in some strip containing the imaginary axis (since $R_y(\cdot)$ is assumed to be exponentially bounded). This, in particular, means that the continuous-time Fourier transform $S_y(j\omega)$ exists,

$$S_y(j\omega) \triangleq \int_{-\infty}^{\infty} R_y(t) e^{-j\omega t} dt.$$

One sometimes also writes $S_y(f)$, where $\omega = 2\pi f$. The function $S_y(j\omega)$ (or $S_y(f)$) is known as the *power spectral density function* or, more briefly, the *power spectrum*, and has the characterizing properties of Hermitian symmetry ($S_y(j\omega) = S_y^*(j\omega)$), and nonnegativity ($S_y(j\omega) \geq 0$ for all $\infty < \omega < \infty$).

Linear Operations on Stationary Stochastic Processes. Given two continuous-time (zero-mean) jointly stationary random processes $\mathbf{y}(\cdot)$ and $\mathbf{u}(\cdot)$, their s-cross-spectrum is defined as

$$S_{yu}(s) = \int_{-\infty}^{\infty} R_{yu}(t) e^{-st} dt = S_{uy}^*(-s^*). \tag{6.A.8}$$

If, furthermore, $\mathbf{y}(\cdot)$ and $\mathbf{u}(\cdot)$ are related by a linear system with transfer function $H(s)$, *i.e.*, if $Y(s) = H(s)U(s)$, then we may write

$$S_y(s) = H(s)S_u(s)H^*(-s^*) \quad \text{and} \quad S_{yu}(s) = H(s)S_u(s). \tag{6.A.9}$$

Also, if $\mathbf{x}(\cdot)$ is jointly stationary with $\mathbf{u}(\cdot)$ and $\mathbf{y}(\cdot)$ as just defined, then

$$S_{xy}(s) = S_{xu}(s)H^*(-s^*). \tag{6.A.10}$$

A simple proof of the above formulae uses the following simple fact: given two zero-mean stationary stochastic processes $\mathbf{a}(\cdot)$ and $\mathbf{b}(\cdot)$, we have

$$\begin{aligned} E\mathbf{A}(s)\mathbf{b}^*(0) &= E\left[\int_{-\infty}^{\infty} \mathbf{a}(t)e^{-st}dt\right]\mathbf{b}^*(0) \\ &= \int_{-\infty}^{\infty} E\mathbf{a}(t)\mathbf{b}^*(0)e^{-st}dt = \int_{-\infty}^{\infty} R_{ab}(t)e^{-st}dt = S_{ab}(s). \end{aligned}$$

Thus, taking $\mathbf{A}(s) = \mathbf{Y}(s) = H(s)\mathbf{U}(s)$ and $\mathbf{b}(\cdot) = \mathbf{u}(\cdot)$ in the above identity, we have

$$S_{yu}(s) = EH(s)\mathbf{U}(s)\mathbf{u}^*(0) = H(s)E\mathbf{U}(s)\mathbf{u}^*(0) = H(s)S_u(s).$$

The first equality in (6.A.9) follows similarly. For (6.A.10), note that

$$S_{yx}(s) = E\mathbf{Y}(s)\mathbf{x}^*(0) = H(s)[E\mathbf{U}(s)\mathbf{x}^*(0)] = H(s)S_{ux}(s),$$

and therefore

$$S_{xy}(s) = S_{yx}^*(-s^*) = S_{ux}^*(-s^*)H^*(-s^*) = S_{xu}(s)H^*(-s^*).$$

We may remark, as we did in the discrete-time case, that engineers often hesitate to use transforms of (stationary) random processes — their use is just as legitimate as the use of delta functions and can be regularized by working under an integral sign (see, *e.g.*, Doob (1953)).

Canonical Spectral Factorization When $S_y(\infty) > 0$. For the applications in this book, we shall further always assume that $S_y(s)$ is strictly positive on the imaginary axis, $S_y(j\omega) > 0$, and that $S_y(s)$ is rational and proper (i.e., the degree of its numerator is equal to the degree of its denominator). The canonical spectral factorization of $S_y(s)$ is then one of the form

$$S_y(s) = L(s)RL^*(-s^*), \tag{6.A.11}$$

where $R > 0$, $L(\infty) = 1$, and $L(s)$ is a minimum-phase system (*i.e.*, all its zeros and poles are strictly inside the left-half plane).

The spectral factor $L(s)$ can be completely determined from the poles and zeros of $S_y(s)$. Indeed, the relation $S_y(s) = S_y^*(-s^*)$ shows that for every pole (zero) at α (β) we must also have a pole (zero) at $-\alpha^*$ ($-\beta^*$). Moreover, a finite-power assumption on the process $\mathbf{y}(\cdot)$ rules out the possibility of imaginary poles, whereas a strict positivity assumption on $S_y(j\omega)$ rules out the possibility of imaginary zeros. Under these conditions,

$$S_y(s) = R \cdot \frac{\prod_{i=1}^{n}(s-\beta_i)(-s-\beta_i^*)}{\prod_{j=1}^{n}(s-\alpha_i)(-s-\alpha_i^*)}, \quad \mathrm{Re}(\alpha_i) < 0, \quad \mathrm{Re}(\beta_i) < 0,$$

where clearly $R > 0$. Then

$$L(s) = \prod_{i=1}^{n} \frac{(s-\beta_i)}{(s-\alpha_i)}, \quad \mathrm{Re}(\alpha_i) < 0, \quad \mathrm{Re}(\beta_i) < 0.$$

Remark [The White-Noise Assumption]. The assumption that $S_y(\cdot)$ is proper means that the process $\mathbf{y}(\cdot)$ has a white-noise component with covariance $R\delta(t-\tau)$. [This assumption allows us to write the continuous-time innovations as $\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{y}}(t|t-)$, as we shall discuss in some detail in Ch. 16.] We may remark that the assumption of an additive white-noise component is not necessary in discrete-time.

Of course there are rational spectra with numerator of lower degree than the denominator. For example, when $R_y(t) = e^{-\alpha|t|}$, we have $S_y(s) = 2\alpha/(\alpha^2 - s^2)$. We can still define a canonical factorization $S_y(s) = L(s)RL^*(-s^*)$, but now we cannot normalize $L(s)$ such that $L(\infty) = 1$. Now $L^{-1}(s)$ will not be proper and the innovations process will involve $\mathbf{y}(\cdot)$ and its derivatives. Such cases can be studied but are of limited interest because in practice one is reluctant to differentiate the observed process $\mathbf{y}(\cdot)$. ◆

Remark. We may add that, just as in discrete time, canonical spectral factorization can also be defined for *nonrational* finite-power s-spectra, $S_y(s)$, say $S_y(s) = L(s)L^*(-s^*)$, with $L(s)$ and its inverse analytic in the open right-half plane. [Unlike discrete-time, the factor $L(s)$ cannot always be normalized to $L(\infty) = 1$.] The existence of $L(s)$ requires that a certain Paley-Wiener condition be satisfied, which in continuous time is of the form (compare with (6.4.4)):

$$\int_{-\infty}^{\infty} \frac{|\ln S_y(f)|}{1+f^2}\, df \;<\; \infty.$$

◆

Vector-Valued Processes. Finally, for vector-valued processes with proper rational z-spectra that are positive-definite on the imaginary axis, there always exists a unique canonical spectral factor $L(s)$ such that

$$S_y(s) \;=\; L(s)RL^*(-s^*),$$

with $R > 0$, $L(\infty) = I$, and $L(s)$ and $L^{-1}(s)$ analytic in $\mathrm{Re}(s) \geq 0$. That is, all the poles and zeros of $L(s)$ are in the open left-half plane.

CHAPTER 7

Wiener Theory for Scalar Processes

Thus far in the book, we have only considered estimation problems involving a finite number of random variables. In this chapter we shall study problems involving an infinite (in fact, sometimes an uncountably infinite) collection of random variables. In this way, we shall extend the discussions of Sec. 4.1 to continuous-time random processes. However, we shall see that the geometric interpretation of random variables as vectors allows us to proceed fairly directly, at least if certain technicalities are ignored — see the footnote for Eq. (7.1.1).

We may remark that estimation problems for stochastic processes observed over infinite and semi-infinite intervals were first considered by Norbert Wiener around 1940 in now famous studies that are often regarded as launching the Statistical Theory of Communications and Mathematical System Theory.

7.1 CONTINUOUS-TIME WIENER SMOOTHING

Consider two *continuous-time* jointly stationary random processes: a *signal* process, $\mathbf{s}(\cdot)$, that is not directly observable, and another observable *measurement* random process, $\mathbf{y}(\cdot)$. We assume that we have available the first- and second-order statistics, *i.e.*, the mean values which are *assumed*, as always, to be zero, and the covariance and cross-covariance functions of the processes $\{\mathbf{s}(\cdot), \mathbf{y}(\cdot)\}$. We wish to estimate, in the sense of minimum mean-square error, the values of the signal process by means of linear operations on the process $\mathbf{y}(\cdot)$, say

$$\hat{\mathbf{s}}(t) \;=\; \int_{-\infty}^{\infty} w(t,\tau)\mathbf{y}(\tau)\,d\tau. \tag{7.1.1}$$

In systems language, we can interpret this as saying that the process $\mathbf{y}(\cdot)$ is operated upon by a linear (time-variant) *filter*, $w(\cdot, \cdot)$, to get the *estimated* process, $\hat{\mathbf{s}}(\cdot)$. The filter $w(\cdot, \cdot)$ is time-variant because in general we may expect that the *weights*, $w(t, \tau)$, applied to the $\mathbf{y}(\tau)$ might depend on the time t at which the value of the signal process is being estimated. However, as we shall see, the assumption that the processes $\mathbf{s}(\cdot)$ and $\mathbf{y}(\cdot)$ are jointly stationary will imply that the filtering operations are time-invariant, *i.e.*, that the values $w(t, \tau)$ depend only upon the difference $(t - \tau)$ and not on t and τ separately.[1]

As mentioned earlier, we shall use the geometric approach of Sec. 3.3 to solve the above problem. For this, first note that a random process $\mathbf{y}(\cdot)$ can be thought of either as a collection of sample functions (of time), or as an *indexed* (infinite) collection of random variables $\{\mathbf{y}(t), -\infty < t < \infty\}$. It is the second point of view that will be more useful for us, because then for every t, the random variable $\mathbf{y}(t)$ can be regarded as a *vector* in an abstract probability space. The relations between the vectors comprising the whole random process $\mathbf{y}(\cdot)$ are given by the inner products

$$\langle \mathbf{y}(t), \mathbf{y}(\tau) \rangle = E\mathbf{y}(t)\mathbf{y}^*(\tau) \triangleq R_y(t - \tau), \qquad \infty < t, \tau < \infty,$$

where $R_y(\cdot)$ is the covariance function of the stationary process $\mathbf{y}(\cdot)$.

The signal process $\mathbf{s}(\cdot)$ similarly defines a collection of vectors $\{\mathbf{s}(t), -\infty < t < \infty\}$ in the same space, and it and its relation to the collection $\{\mathbf{y}(t), -\infty < t < \infty\}$ are given by the inner products:

$$\langle \mathbf{s}(t), \mathbf{s}(\tau) \rangle = E\mathbf{s}(t)\mathbf{s}^*(\tau) = R_s(t - \tau), \quad -\infty < t, \tau < \infty,$$

and

$$\langle \mathbf{s}(t), \mathbf{y}(\tau) \rangle = E\mathbf{s}(t)\mathbf{y}^*(\tau) = R_{sy}(t - \tau), \quad -\infty < t, \tau < \infty.$$

We can now put the problem of estimating $\mathbf{s}(\cdot)$ given observations of the process $\mathbf{y}(\cdot)$ into our framework as follows. Fix t and form $\hat{\mathbf{s}}(t)$ as a linear combination of the infinite collection of (observable) random variables $\{\mathbf{y}(t), -\infty < t < \infty\}$, say

$$\hat{\mathbf{s}}(t) = \int_{-\infty}^{\infty} w(t, \tau)\mathbf{y}(\tau)d\tau, \tag{7.1.2}$$

[1] Some technicalities arise, not only in defining integrals of the form (7.1.1), but also in deciding whether such expressions suffice to describe a general linear functional of the process $\mathbf{y}(\cdot)$. Most readers can ignore this issue, but for the more adventurous we make the following comments. As to the first question, if the process $\mathbf{y}(\cdot)$ has finite variance, then the integral can be defined as a limit in mean-square of Riemann sums (see, *e.g.*, Doob (1953) and Davis (1977)) provided the function $w(t, \cdot)$ is square-integrable,

$$\int_{-\infty}^{\infty} |w(t, \tau)|^2 d\tau < \infty.$$

However, if the process $\mathbf{y}(\cdot)$ is differentiable, then a general linear functional could be expected to include values of the derivative at certain points, *e.g.*, $\dot{\mathbf{y}}(t_1)$. This implies that such $w(t, \cdot)$ should contain the derivative of an impulse function, written $\delta^{(1)}(t - t_1)$, which is not square-integrable (or even a function in the ordinary sense). So for smooth processes, differentiable to various orders, or even to all orders (as bandlimited processes are), a general linear functional cannot be written as in (7.1.1), because it is not easy to say what $w(t, \cdot)$ should be. Fortunately, when $\mathbf{y}(\cdot)$ contains pure (continuous-time) white noise, then it suffices to assume the square-integrability of $w(t, \cdot)$; this useful assumption will be made for all the discussions in Ch. 8.

where $w(t, \cdot)$ is to be chosen so that

$$E\left|\mathbf{s}(t) - \int_{-\infty}^{\infty} w(t, \tau)\mathbf{y}(\tau)d\tau\right|^2 = \text{minimum}. \tag{7.1.3}$$

Notice that in the estimate $\hat{\mathbf{s}}(t)$ we are using *all* values of $\mathbf{y}(\cdot)$, those before t and those after t as well. This will mean that the linear filter $w(\cdot, \cdot)$ is *noncausal* (or *anticipative*) and cannot be implemented in *real time* or *on-line*. It is only appropriate for *off-line* situations where all the observations $\mathbf{y}(\cdot)$ have already been recorded. One application is what is often called *post flight analysis*, carried out to discover reasons for crashes or unusual events. Of course, there is no fundamental reason to restrict t to being a "time" index — it could be spatial, as in the analysis of beams, or multidimensional, as in image processing applications (in medicine, seismology, etc.). So the term noncausal filtering is somewhat prejudicial and one often uses the terms *smoothing filter* or just *smoother*. When only data prior to the index t is used in the estimator, we speak of *causal filtering* or usually just *filtering*.

7.1.1 The Geometric Formulation

Now returning to our problem (7.1.3), the geometric picture tells us that to achieve the minimum, we must choose $w(t, \cdot)$ so that the orthogonality condition is satisfied, *i.e.*,

$$\left(\mathbf{s}(t) - \int_{-\infty}^{\infty} w(t, \tau)\mathbf{y}(\tau)d\tau\right) \perp \mathbf{y}(\sigma), \quad -\infty < \sigma < \infty,$$

or

$$\langle \mathbf{s}(t), \mathbf{y}(\sigma) \rangle = \langle \int_{-\infty}^{\infty} w(t, \tau)\mathbf{y}(\tau)d\tau, \mathbf{y}(\sigma) \rangle, \quad -\infty < \sigma < \infty,$$

or

$$R_{sy}(t - \sigma) = \int_{-\infty}^{\infty} w(t, \tau)R_y(\tau - \sigma)d\tau, \quad -\infty < \sigma < \infty. \tag{7.1.4}$$

From (7.1.4) we see that instead of linear *algebraic* equations, for continuous-time stochastic process observations we have a linear *integral* equation, or rather a *set of linear integral equations*, one for each value of t. There are only a few cases in which we can write down explicit formulas for the solution of such equations, but fortunately our assumptions in the present problem will allow such a solution.

To see this we first note that the stationarity assumptions (which make $R_{sy}(\cdot)$ and $R_y(\cdot)$ functions of a single variable) imply that so is $w(t, \tau)$. We can show this by making two changes of variables: let $\tau - \sigma = \tau'$ (to eliminate τ) and then define $t' = t - \sigma$ (to eliminate t), after which the equation (7.1.4) will take the form

$$R_{sy}(t') = \int_{-\infty}^{\infty} w(t' + \sigma, \tau' + \sigma)R_y(\tau')d\tau', \quad -\infty < t', \sigma < \infty. \tag{7.1.5}$$

But the left-hand side is independent of σ, and therefore the right-hand side must be the same for all values of σ. For this to hold, it must be true that $w(\cdot, \cdot)$ has the property

$$w(t' + \sigma, \tau' + \sigma) = w(t', \tau') \quad \text{for any } \sigma,$$

which further shows that $w(\cdot,\cdot)$ should be a function of the difference of its arguments, say

$$w(t',\tau') = k(t'-\tau') \tag{7.1.6}$$

for some function $k(\cdot)$. So expression (7.1.5) can be rewritten as

$$R_{sy}(t') = \int_{-\infty}^{\infty} k(t'-\tau')R_y(\tau')d\tau', \quad -\infty < t' < \infty,$$

or by redefining the variables (t',τ') as (t,τ),

$$R_{sy}(t) = \int_{-\infty}^{\infty} k(t-\tau)R_y(\tau)d\tau, \quad -\infty < t < \infty. \tag{7.1.7}$$

Another important consequence is that we can write the smoothing filter (7.1.2) as the convolution integral

$$\hat{\mathbf{s}}(t) = \int_{-\infty}^{\infty} k(t-\tau)\mathbf{y}(\tau)d\tau = \int_{-\infty}^{\infty} k(\tau)\mathbf{y}(t-\tau)d\tau. \tag{7.1.8}$$

In retrospect, of course, such a result is not unexpected, since because of the stationarity assumptions, the form of the solution for finding $\hat{\mathbf{s}}(t)$ is independent of the particular value of t. Therefore instead of the set of linear integral equations (7.1.4), we have just one equation.

7.1.2 Solution via Fourier Transforms

Since the right-hand side of the integral equation (7.1.7) has the form of a convolution, the equation is readily solved by taking Fourier transforms to get $S_{sy}(f) = K(f)S_y(f)$. Then[2]

$$K(f) = \frac{S_{sy}(f)}{S_y(f)}, \tag{7.1.9}$$

where

$$S_{sy}(f) \triangleq \int_{-\infty}^{\infty} R_{sy}(t)e^{-j2\pi ft}dt, \qquad j \triangleq \sqrt{-1},$$

is the cross-spectral density function of $\{\mathbf{s}(\cdot), \mathbf{y}(\cdot)\}$,

$$S_y(f) \triangleq \int_{-\infty}^{\infty} R_y(t)e^{-j2\pi ft}dt$$

is the power spectral density function of $\mathbf{y}(\cdot)$, and

$$K(f) \triangleq \int_{-\infty}^{\infty} k(t)e^{-j2\pi ft}dt.$$

Recalling the basic formula for estimating a random variable, say $\mathbf{s}$, from an observed random variable, say $\mathbf{y}$,

$$\hat{\mathbf{s}} = \langle \mathbf{s}, \mathbf{y}\rangle \, \|\mathbf{y}\|^{-2} \, \mathbf{y} = R_{sy}R_y^{-1}\mathbf{y},$$

[2] We need to assume that $S_y(f) \neq 0$ for almost all f, which is certainly true when $S_y(f)$ is rational.

the formula (7.1.9) implies that our problem can be regarded as one of estimating each frequency component of the process $\mathbf{s}(\cdot)$ from the corresponding frequency component of the process $\mathbf{y}(\cdot)$, *i.e.*, each frequency can be treated independently of every other frequency. This independence is a consequence of stationarity and the fact that the observations of $\mathbf{y}(t)$ are over the whole range, $-\infty < t < \infty$; it would not be true (as we shall soon see in Sec. 7.4) if the observations are over only a finite, or even semi-infinite, range.

We should note that the filter with transfer function $K(f)$ is *noncausal* because to find the estimate at each time t, it uses all the $\{\mathbf{y}(\tau), -\infty < \tau < \infty\}$. If τ is physical time, then such a filter is not physically realizable, though it can be approximated by introducing delay into the system.

7.1.3 The Minimum Mean-Square Error

The variance of the error, $E|\tilde{\mathbf{s}}(t)|^2 = E|\mathbf{s}(t) - \hat{\mathbf{s}}(t)|^2$, can be computed in several ways. However, in all these ways it is useful to go back to the formula (7.1.8), which shows that by letting t range over $(-\infty, \infty)$ we can regard the estimates $\hat{\mathbf{s}}(t)$ as defining a stochastic process, and that $\hat{\mathbf{s}}(t)$ can be obtained by passing the observed stochastic process $\mathbf{y}(\cdot)$ through a linear time-invariant filter with impulse response $k(\cdot)$. Therefore, since $\mathbf{y}(\cdot)$ is stationary, so will be the process $\hat{\mathbf{s}}(\cdot)$,[3] and in fact so will be the error process defined by $\tilde{\mathbf{s}}(\cdot) = \mathbf{s}(\cdot) - \hat{\mathbf{s}}(\cdot)$. With this in mind, we can write

$$E|\tilde{\mathbf{s}}(t)|^2 = E|\mathbf{s}(t)|^2 - E|\hat{\mathbf{s}}(t)|^2 = R_s(0) - R_{\hat{s}}(0) = \int_{-\infty}^{\infty} S_s(f)df - \int_{-\infty}^{\infty} S_{\hat{s}}(f)df.$$

Now using the stationarity of all the processes involved, and the time invariance of the optimal filter, it follows that[4]

$$S_{\hat{s}}(f) = |K(f)|^2 S_y(f) = \frac{|S_{sy}(f)|^2}{S_y(f)}.$$

Therefore,

$$E|\tilde{\mathbf{s}}(t)|^2 = \int_{-\infty}^{\infty} \left[S_s(f) - \frac{|S_{sy}(f)|^2}{S_y(f)} \right] df, \tag{7.1.10}$$

which again may be compared with that for the error of estimating a random variable, $\mathbf{s}$, given another random variable, $\mathbf{y}$:

$$\text{m.m.s.e.} = \|\mathbf{s}\|^2 - \langle \mathbf{s}, \mathbf{y}\rangle \|\mathbf{y}\|^{-2} \langle \mathbf{y}, \mathbf{s}\rangle = R_s - R_{sy}R_y^{-1}R_{ys}.$$

[3] The stationarity of the output process $\hat{\mathbf{s}}(\cdot)$ can be guaranteed by requiring the filter $k(\cdot)$ to be square-integrable (as remarked in the first footnote in this chapter — see also Picinbono (1993, pp. 181–182)).

[4] A familiarity with the interaction of stationary processes and time-invariant linear systems is assumed for the rest of the discussion (see, App. 6.A or standard textbooks, *e.g.*, Davenport and Root (1958)).

7.1.4 Filtering Signals out of Noisy Measurements

A very important special case is when the processes $\mathbf{s}(\cdot)$ and $\mathbf{y}(\cdot)$ are related in an additive way,

$$\mathbf{y}(t) = \mathbf{s}(t) + \mathbf{v}(t), \tag{7.1.11}$$

where $\mathbf{v}(\cdot)$ is a stationary *noise* process, independent of $\mathbf{s}(\cdot)$, with power spectral density function $S_v(f)$. In this case it is easy to see that

$$K(f) = \frac{S_s(f)}{S_s(f) + S_v(f)} = 1 - \frac{S_v(f)}{S_y(f)}, \tag{7.1.12}$$

and that the

$$\text{m.m.s.e.} = \int_{-\infty}^{\infty} \frac{S_s(f)S_v(f)}{S_s(f) + S_v(f)} df. \tag{7.1.13}$$

Note that the m.m.s.e. is zero whenever the signal and noise spectra are disjoint — we would be quite disturbed if this were not so. In fact, it will be interesting to take a closer look at the significance of the Wiener smoothing solution by comparing it to earlier approaches to the problem.

7.1.5 Comparison with the Ideal Filter

Until Wiener introduced the concept of a statistical optimization criterion, it was felt that the best one could do was design a device (a filter) that would pass the signal through without distortion ("high fidelity") while admitting as little noise as possible. Since the interesting signals (voice, music) generally had a smaller bandwidth than the noise, the best device under this criterion was a so-called ideal lowpass filter, *i.e.*, a filter with unit gain in the signal bandwidth and zero elsewhere, say

$$I(f) = \begin{cases} 1 & |f| \le \Omega, \\ 0 & |f| > \Omega, \end{cases}$$

where the signal frequencies are assumed to lie in the range $(-\Omega, \Omega)$, or equivalently the signal is assumed to have bandwidth 2Ω.

If we now turn to the optimal Wiener solution, based on the least-mean-squares criterion, we see from (7.1.12) that it too rejects all inputs outside the frequency range of the signal. However, *unlike the ideal lowpass filter, it does not equally weight the frequencies within the signal range*: it gives more weight to frequencies where $S_s(f)$ is high, *i.e.*, where the expected signal energy is high. To be more specific, consider the common case where the noise has a much wider bandwidth than the signal so that it is modeled as a white-noise process,

$$S_v(f) = N, \quad -\infty < f < \infty,$$

so that the optimum Wiener filter is given by

$$K(f) = \frac{S_s(f)}{S_s(f) + N}.$$

The optimum solution is generally of most value when the noise is high relative to the signal (small signal-to-noise ratio), in which case we have

$$K(f) \to \frac{S_s(f)}{N} \quad \text{as} \quad \frac{S_s(f)}{N} \to 0.$$

We see that the frequency weighting is proportional to the expected power in each frequency range — so regions of lesser "signal-to-noise" are deemphasized in favor of regions with a better signal-to-noise ratio. This is certainly a strategy that makes intuitive sense; it was implicitly used in the so-called "square-law-detector" in radio engineering, and in the well-known "matched filter" for signal detection.

When the noise is very low compared to the signal (high signal-to-noise ratio), the Wiener smoother equally weights all frequencies in the signal range

$$K(f) \to 1 \quad \text{as} \quad \frac{S_s(f)}{N} \to \infty.$$

This is the same as the ideal lowpass filter, which is clearly the optimal solution (under all criteria) when there is no noise. So while the optimal filter is easy to specify in the two limiting cases, it is in more ambiguous situations that this theory is useful. The Wiener formulas give us an explicit procedure for all situations, including where the noise is not additive — see Prob. 7.3. The literature contains many other examples of the application of the Wiener smoothing filter.

7.2 THE CONTINUOUS-TIME WIENER-HOPF EQUATION

So far we have assumed that we have access to *all* the observations $\mathbf{y}(\tau)$ for estimating $\mathbf{s}(t)$. If we require that only *past* and *present* values of $\mathbf{y}(\tau)$ be used in estimating $\mathbf{s}(t)$, we are led to a more difficult equation to solve. Let us assume without proof (see Sec. 7.3.2 and Prob. 7.4 for the discrete-time case) that the optimum filter is time-invariant, so that the estimator has the form

$$\hat{\mathbf{s}}(t|t) = \int_{-\infty}^{t} k(t-\tau)\mathbf{y}(\tau)d\tau = \int_{0}^{\infty} k(\tau)\mathbf{y}(t-\tau)d\tau. \tag{7.2.1}$$

Assuming the form (7.2.1), we proceed by invoking the orthogonality condition that the error should be orthogonal to all past observations $\{\mathbf{y}(\sigma), -\infty < \sigma \le t\}$, *i.e.*,

$$\left(\mathbf{s}(t) - \int_0^{\infty} k(\tau)\mathbf{y}(t-\tau)d\tau\right) \perp \mathbf{y}(\sigma), \quad -\infty < \sigma \le t,$$

or

$$R_{sy}(t-\sigma) = \int_0^{\infty} k(\tau)R_y(t-\tau-\sigma)d\tau, \quad -\infty < \sigma \le t,$$

and after changing variables $t - \sigma \to t$,

$$R_{sy}(t) = \int_0^{\infty} k(\tau)R_y(t-\tau)d\tau, \quad t \ge 0. \tag{7.2.2}$$

We may replace the lower limit by $-\infty$, say

$$R_{sy}(t) = \int_{-\infty}^{\infty} k(\tau) R_y(t-\tau)d\tau, \quad t \geq 0, \tag{7.2.3}$$

if we also assume that $k(\cdot)$ is causal, *i.e.*, $k(t) = 0$ for $t < 0$.

This is the celebrated Wiener-Hopf integral equation. It looks deceptively like Eq. (7.1.7) for the Wiener smoother except that the equality holds only for $t \geq 0$ and that we have the constraint that $k(\cdot)$ be causal (*i.e.*, $k(t) = 0$ for $t < 0$). Therefore one cannot just equate the Fourier transforms of the functions on both sides of the equality in (7.2.2). In other words, although $R_{sy}(t)$, $k(t)$, $R_y(t)$ and $\int_0^\infty k(\tau)R_y(t-\tau)d\tau$ are defined for all instants t, and each of these functions has Fourier transforms (*viz.*, $S_{sy}(f)$, $K(f)$, $S_y(f)$, and $K(f)S_y(f)$), the above equality holds only for the $t \geq 0$ portions of the functions $R_{sy}(\cdot)$ and $\int_0^\infty k(\tau)R_y(t-\tau)d\tau$.[5] This is a nontrivial difficulty, which baffled the astrophysicists who introduced such equations: first O. D. Hvol'son (1894, Leningrad) while studying the scattering of light by milk glass, and later (ca. 1920), E. A. Milne, K. Schwarzschild, and others while studying problems in astrophysics (these references can be found, for example, in Sobolev (1963)).

Since explicit analytical solutions were long thought to be impossible, Wiener and Hopf (1931) won instant acclaim when they solved (7.2.2) by a beautiful method, now called the Wiener-Hopf factorization technique. In fact, so striking was their solution that not only the technique, but the equation itself, came to be known by the name "Wiener-Hopf." The tools used in the solution were largely developed by Wiener himself and involve the study of Fourier transforms extended to the complex domain, which are now called bilateral Laplace transforms. Since the major emphasis in this book will be on discrete-time processes, we shall explain the Wiener-Hopf technique in that context and postpone the discussion of the continuous-time Wiener-Hopf technique to App. 7.A.

7.3 DISCRETE-TIME PROBLEMS

In this and the remaining sections of this chapter, we shall be considering discrete-time zero-mean jointly wide-sense stationary random processes $\{\mathbf{s}_i\}$ and $\{\mathbf{y}_i\}$ with known covariance and cross-covariance functions $R_s(i)$, $R_y(i)$, and $R_{sy}(i)$.

7.3.1 The Discrete-Time Wiener Smoother

To find the l.l.m.s. estimator of $\mathbf{s}_i$ using all the observations $\{\mathbf{y}_m\}_{m=-\infty}^{\infty}$, let us write[6]

$$\hat{\mathbf{s}}_i = \sum_{m=-\infty}^{\infty} w_{im}\mathbf{y}_m, \tag{7.3.1}$$

[5] The reader at this point may wish to compare these statements with those given after Eq. (4.1.13) on discrete-time finite-horizon causal filtering.

[6] As with Eq. (7.1.1), we also have the issue of whether the most general linear functional of an infinite collection of random variables can be represented as in (7.3.1) — in general, the answer is no. The representation will be justified if the $\{\mathbf{y}_i\}$ are uncorrelated, or equivalently, if they can be transformed to such a set, which will be the situation of interest to us.

for some set of filter weights, $\{w_{im}\}$, that need to be determined. The orthogonality condition yields

$$\left(\mathbf{s}_i - \sum_{m=-\infty}^{\infty} w_{im}\mathbf{y}_m\right) \perp \mathbf{y}_l, \quad -\infty < l < \infty,$$

or,

$$R_{sy}(i-l) = \sum_{m=-\infty}^{\infty} w_{im}R_y(m-l), \quad -\infty < l < \infty. \tag{7.3.2}$$

Note that (7.3.2) is an infinite set of linear equations, one for each l. However, because of our stationarity assumptions, we can reduce (7.3.2) to an equation that can be solved by Fourier transform techniques. If we employ the change of variables ($m - l = m'$ and $i - l = i'$), we can write

$$R_{sy}(i') = \sum_{m'=-\infty}^{\infty} w_{i'+l,m'+l}R_y(m'), \quad -\infty < l < \infty, \tag{7.3.3}$$

which allows us to conclude that we must have

$$w_{i'+l,m'+l} = w_{i',m'} = k_{i'-m'}, \quad -\infty < l < \infty,$$

for some sequence $\{k_i\}$, since the left-hand side of (7.3.3) is independent of l. Consequently, we obtain that

$$R_{sy}(i) = \sum_{m=-\infty}^{\infty} k_{i-m}R_y(m), \tag{7.3.4}$$

and the smoothed estimator becomes

$$\hat{\mathbf{s}}_i = \sum_{m=-\infty}^{\infty} k_{i-m}\mathbf{y}_m. \tag{7.3.5}$$

As in continuous time, the solution to (7.3.4) can be obtained by taking discrete-time Fourier transforms, *e.g.*,

$$K(e^{j\omega}) \triangleq \sum_{l=-\infty}^{\infty} k_l e^{-j\omega l},$$

and correspondingly for $S_{sy}(e^{j\omega})$ and $S_y(e^{j\omega})$.

Theorem 7.3.1 (The Optimal Discrete-Time Smoothing Filter) *Given two discrete-time zero-mean jointly stationary random processes $\{\mathbf{s}_i, \mathbf{y}_i\}$, the linear least mean-squares smoother of the process $\{\mathbf{s}_i\}$ given all the observations of the process $\{\mathbf{y}_i\}$ is the time-invariant filter*

$$K(e^{j\omega}) = \frac{S_{sy}(e^{j\omega})}{S_y(e^{j\omega})}, \tag{7.3.6}$$

with the corresponding minimum mean-square error

$$E|\tilde{\mathbf{s}}_i|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(S_s(e^{j\omega}) - \frac{|S_{sy}(e^{j\omega})|^2}{S_y(e^{j\omega})}\right)d\omega. \tag{7.3.7}$$

■

7.3.2 The Discrete-Time Wiener-Hopf Equation

In the Wiener *filtering* problem we wish to estimate $\mathbf{s}_i$ given the past and present observations $\{\mathbf{y}_m\}_{m=-\infty}^{i}$ as

$$\hat{\mathbf{s}}_{i|i} = \sum_{m=-\infty}^{i} w_{im}\mathbf{y}_m, \tag{7.3.8}$$

for some set of coefficients $\{w_{im}\}$ that are to be determined. Using the orthogonality condition we require

$$\left(\mathbf{s}_i - \sum_{m=-\infty}^{i} w_{im}\mathbf{y}_m\right) \perp \mathbf{y}_l \quad \text{for} \quad -\infty < l \le i,$$

or, equivalently,

$$R_{sy}(i-l) = \sum_{m=-\infty}^{i} w_{im}R_y(m-l) \quad \text{for} \quad -\infty < l \le i. \tag{7.3.9}$$

The change of variable $i' = i - l$ yields

$$R_{sy}(i') = \sum_{m=-\infty}^{i'+l} w_{i'+l,m}R_y(m-l) \quad \text{for} \quad i' \ge 0.$$

The further change of variable $m' = m - l$ leads to the equation

$$R_{sy}(i') = \sum_{m'=-\infty}^{i'} w_{i'+l,m'+l}R_y(m') \quad \text{for} \quad i' \ge 0, \tag{7.3.10}$$

which should hold for all l. But since the left-hand-side of (7.3.10) is independent of l, the coefficients $\{w_{im}\}$ must have the property that

$$w_{i'+l,m'+l} = w_{i'm'} = k_{i'-m'}, \tag{7.3.11}$$

for some sequence $\{k_i\}$ (see Prob. 7.4). This implies that we can rewrite (7.3.8) as

$$\hat{\mathbf{s}}_{i|i} = \sum_{m=-\infty}^{i} k_{i-m}\mathbf{y}_m = \sum_{m=0}^{\infty} k_m\mathbf{y}_{i-m}. \tag{7.3.12}$$

In other words, because of the joint stationarity of the processes $\{\mathbf{s}_i, \mathbf{y}_i\}$, the form of the solution for finding $\hat{\mathbf{s}}_{i|i}$ is independent of the particular value of i. Moreover,

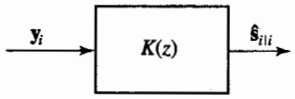


Figure 7.1 The estimator $\hat{\mathbf{s}}_{i|i}$ is obtained by passing the process $\{\mathbf{y}_i\}$ through a linear time-invariant (LTI) filter $K(z)$.

note that $\{\hat{\mathbf{s}}_{i|i}\}$ is also a stationary process since it can be regarded as the output of a linear time-invariant filter $K(z)$, with impulse response $\{k_i\}$,[7]

$$K(z) = \sum_{i=0}^{\infty} k_i z^{-i},$$

driven by the stationary process $\{\mathbf{y}_i\}$, as depicted in Fig. 7.1.

Then using (7.3.11), we can also rewrite (7.3.10) as

$$R_{sy}(i) = \sum_{m=-\infty}^{i} k_{i-m}R_y(m) = \sum_{m=0}^{\infty} k_m R_y(i-m), \quad i \ge 0. \tag{7.3.13}$$

We can also write the equation as

$$R_{sy}(i) = \sum_{m=-\infty}^{\infty} k_{i-m}R_y(m) = \sum_{m=-\infty}^{\infty} k_m R_y(i-m), \quad i \ge 0 \tag{7.3.14}$$

if we impose the constraint that

$$k_m = 0, \quad m < 0. \tag{7.3.15}$$

This is the discrete-time version of the Wiener-Hopf equation (7.2.2). It does not immediately lend itself to Fourier transform techniques since the equality in (7.3.14) is only valid for $i \ge 0$. That is, although $R_{sy}(i)$ and $\sum_{m=0}^{\infty} k_m R_y(i-m)$ are defined for all values of i, they are only equal for their $i \ge 0$ portions.

We shall present the discrete-time version of the ingenious Wiener-Hopf technique for solving the above equation in the next section. For this purpose we shall need to use some properties of discrete-time stationary processes and their spectral factorizations, which we presented in Secs. 6.2–6.5. These results should be reviewed at this point, because they will also be used later in Sec. 7.7 to present the more physically transparent innovations solution of the estimation problem (and thereby of the Wiener-Hopf equation).

7.4 THE DISCRETE-TIME WIENER-HOPF TECHNIQUE

In the Wiener-Hopf equation (7.3.14), the cross-correlation and autocorrelation functions, $R_{sy}(\cdot)$ and $R_y(\cdot)$, are given and it is assumed that their z-transforms are well-defined *rational* functions in an annulus containing the unit circle. Here, unlike the

[7] Again, the stationarity of the output process $\{\hat{\mathbf{s}}_{i|i}\}$ can be guaranteed by requiring the sequence $\{k_i\}$ to be square-summable.

smoothing problem, we shall also need to assume that[8]

$$S_y(z) > 0 \quad \text{on} \quad |z| = 1 \; - \; S_y(z) \text{ has no unit-circle zeros.}$$

[This will allow us to invoke the canonical factorization results of Ch. 6, which will soon be needed here.]

Eq. (7.3.14) is hard to solve because the equality holds only for $i \geq 0$; otherwise, taking z-transforms would give the solution in a simple way. To overcome this problem, Wiener and Hopf used the following clever technique.[9] Define the sequence

$$g_i \triangleq R_{sy}(i) - \sum_{m=0}^{\infty} k_m R_y(i-m), \quad -\infty < i < \infty, \tag{7.4.1}$$

which by (7.3.14) we see is strictly anticausal, *i.e.*,

$$g_i = 0 \quad \text{for} \quad i \geq 0. \tag{7.4.2}$$

Since (7.4.1) is defined for all i over $(-\infty, \infty)$, we can now take z-transforms to get

$$G(z) = S_{sy}(z) - K(z)S_y(z). \tag{7.4.3}$$

Now comes the critical insight of Wiener and Hopf. Recall from Sec. 6.4 the definition of the canonical factorization of the z-spectrum $S_y(z)$,

$$S_y(z) = L(z)r_e L^*(z^{-*}), \tag{7.4.4}$$

where (i) $L(z)$ is analytic on and outside the unit circle, (ii) $L^{-1}(z)$ is analytic on and outside the unit circle, and (iii) $L(\infty) = 1$. Properties (i) and (ii) are equivalent to saying that all the zeros and poles of $L(z)$ are strictly inside the unit circle, *i.e.*, that $L(z)$ is minimum phase.

Now dividing both sides of the above equation by $r_e L^*(z^{-*})$ we get

$$\frac{G(z)}{r_e L^*(z^{-*})} = \frac{S_{sy}(z)}{r_e L^*(z^{-*})} - K(z)L(z). \tag{7.4.5}$$

Since $L^{-1}(z)$ has a causal inverse transform, $L^{-*}(z^{-*})$ will have an anticausal inverse transform.[10] Moreover, by (7.4.2), $G(z)$ is the z-transform of a strictly anticausal sequence, and therefore $G(z)r_e^{-1}L^{-*}(z^{-*})$ is the z-transform of a strictly anticausal sequence as well. This is because the convolution of an anticausal sequence and a strictly anticausal sequence is strictly anticausal. On the other hand, since $K(z)$ and $L(z)$ are

[8] We are making stronger assumptions than needed in the general Wiener-Hopf theory. Interested readers can refer to Krein (1958), Gohberg and Krein (1958), and Gohberg and Fel'dman (1974) for general treatments.

[9] The reader may wish to compare the following argument with the finite-time argument in Sec. 4.1.3.

[10] This is easy to check since $L^{-1}(z)$ having a causal inverse transform means that we can write

$$L^{-1}(z) = 1 + \sum_{j=1}^{\infty} w_j z^{-j} \quad \text{and} \quad L^{-*}(z^{-*}) = 1 + \sum_{j=1}^{\infty} w_j^* z^j,$$

so that $L^{-*}(z^{-*})$ has an anticausal inverse transform.

analytic on and outside the unit circle, $K(z)L(z)$ has a causal inverse transform. Hence we have

$$\underbrace{\frac{G(z)}{r_e L^*(z^{-*})}}_{\text{strictly anticausal IT}} = \frac{S_{sy}(z)}{r_e L^*(z^{-*})} - \underbrace{K(z)L(z)}_{\text{causal IT}}, \tag{7.4.6}$$

where IT denotes "inverse tranform." For equality to hold, the causal and anticausal parts of the inverse transform of the mixed function $S_{sy}(z)/r_e L^*(z^{-*})$ must match properly. That is, if we take inverse z-transforms of both sides of the above equation, the inverse z-transform of $K(z)L(z)$ must be equal to the causal portion of the inverse z-transform of

$$\frac{S_{sy}(z)}{r_e L^*(z^{-*})}.$$

This is really the end of the story. [Note that $G(z)$ can also be found by a similar argument.] However, to obtain a nice-looking formula, it will be useful to introduce some additional notation as follows.

Let $F(z)$ be an analytic function in some annulus that contains the unit circle, with Laurent expansion

$$F(z) = \sum_{i=-\infty}^{\infty} f_i z^{-i}. \tag{7.4.7}$$

We introduce an operator $\{\cdot\}_+$ that yields the "causal part" of the function to which it is applied:

$$\{F\}_+ = \sum_{i=0}^{\infty} f_i z^{-i}. \tag{7.4.8}$$

We can now apply the $\{\cdot\}_+$ operator to both sides of (7.4.6) to obtain

$$\left\{\frac{G(z)}{r_e L^*(z^{-*})}\right\}_+ = \left\{\frac{S_{sy}(z)}{r_e L^*(z^{-*})}\right\}_+ - \{K(z)L(z)\}_+ . \tag{7.4.9}$$

But since the left-hand side has a *strictly* anticausal inverse transform, we have

$$\left\{\frac{G(z)}{r_e L^*(z^{-*})}\right\}_+ = 0. \tag{7.4.10}$$

Similarly, since the second term on the right-hand side of (7.4.9) has a causal inverse transform,

$$\{K(z)L(z)\}_+ = K(z)L(z). \tag{7.4.11}$$

Finally, substitution of (7.4.10)–(7.4.11) in (7.4.9) gives the formula[11]

$$K(z) = \left\{\frac{S_{sy}(z)}{r_e L^*(z^{-*})}\right\}_+ \frac{1}{L(z)}. \tag{7.4.12}$$

Eq. (7.4.12) is the form of solution to the Wiener-Hopf equation that we shall most often use.

Note in particular that the $K(z)$ obtained via (7.4.12) will be BIBO stable (*i.e.*, all its poles will be strictly inside the unit circle) since the ROC of the function on the right-hand side of (7.4.12) can be shown to include the unit circle. To see this, recall that the ROC of $S_{sy}(z)$ is an annulus containing the unit circle, while $L^{-*}(z^{-*})$ is analytic in $|z| \leq 1$. Hence, the ROC of the ratio $S_{sy}(z)/r_e L^*(z^{-*})$ is also an annulus containing the unit circle. It then follows that the ROC of the causal part

$$\left\{\frac{S_{sy}(z)}{r_e L^*(z^{-*})}\right\}_+$$

includes the unit circle, so that $K(z)$ is indeed BIBO stable.

The results obtained so far, and some easy consequences, are summarized in the following statement, where we also use the definition

$$\{F(z)\}_- \triangleq F(z) - \{F(z)\}_+ \quad \text{for any } F(z). \tag{7.4.13}$$

Theorem 7.4.1 (The Wiener Filter) *Consider two zero-mean jointly stationary scalar random processes $\{\mathbf{s}_i\}$ and $\{\mathbf{y}_i\}$ with known rational z-spectra and z-cross-spectra $S_s(z)$, $S_y(z)$, and $S_{sy}(z)$, respectively. Assume further that $S_y(z)$ has no unit circle zeros. Then the l.l.m.s. estimator of $\mathbf{s}_i$ given $\{\mathbf{y}_m, -\infty < m \leq i\}$ is given by the filter $K(z)$ in (7.4.12). The corresponding minimum mean-square error is $E|\mathbf{s}_i - \hat{\mathbf{s}}_{i|i}|^2 =$*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[S_s(e^{j\omega}) - \frac{|S_{sy}(e^{j\omega})|^2}{S_y(e^{j\omega})}\right]d\omega + \frac{1}{2\pi r_e}\int_{-\pi}^{\pi}\left|\left\{\frac{S_{sy}(e^{j\omega})}{L^*(e^{-j\omega})}\right\}_-\right|^2 d\omega. \tag{7.4.14}$$

■

Proof: The only result that has not been derived so far is the expression for the minimum mean-square error, which we leave as an exercise for active readers. ◆

It is the beauty and elegance of this closed-form solution to an equation that long seemed impossible of such resolution that led to the equation itself, and various later generalizations, being called the Wiener-Hopf equation. The appearance of the canonical spectral factorization in this solution will be illuminated in the alternative innovations approach which we shall present in Sec. 7.7. First, however, let us describe some simplifications that are available in the rational case and then work out some examples.

[11] We can also define $\{F\}_+$ by using the inverse z-transform formula (6.2.3); Eq. (7.4.12) is rewritten in this way in Prob. 7.6.

7.5 CAUSAL PARTS VIA PARTIAL FRACTIONS

We explained in Sec. 6.5 how the canonical spectral factorization of a rational z-spectrum, $S_y(z)$, can be performed. We shall not repeat the details here and instead remark that once the spectral factorization has been performed, the Wiener filtering problem still requires additional evaluations, *e.g.*, of the operator $\{\cdot\}_+$ as in Eq. (7.4.12).

When $F(z)$ is rational, it is easiest to compute $\{F(z)\}_+$ by first performing its unique partial fraction expansion, say

$$F(z) = r_0 + \sum_{i=1}^{m}\sum_{k=1}^{l_i}\frac{r_{ik}}{(z-p_i)^k}, \tag{7.5.1}$$

where we have assumed that $F(z)$ is proper (*i.e.*, the degree of the numerator is less than or equal to the degree of the denominator). Note that in the above expansion p_i is a pole of $F(z)$ with degree (or multiplicity) l_i. Since we can write

$$\{F(z)\}_+ = \{r_0\}_+ + \sum_{i=1}^{m}\sum_{k=1}^{l_i}\left\{\frac{r_{ik}}{(z-p_i)^k}\right\}_+, \tag{7.5.2}$$

we can obtain $\{F(z)\}_+$ for any rational function using the following simple rules given below. We remind the reader that, as explained above, the formula (7.4.12) requires that we evaluate the causal part of the function $F(z) = S_{sy}(z)/r_e L^*(z^{-*})$, whose ROC includes the unit circle. Hence, in the rules below we assume that the regions of convergence of all transforms include the unit circle.

Three Simple Rules.

(a) $\{c\}_+ = c$ for any constant c.

(b)

$$\left\{\frac{1}{z+\alpha}\right\}_+ = \begin{cases}\dfrac{1}{z+\alpha} & \text{for } |\alpha| < 1,\\[2ex] \dfrac{1}{\alpha} & \text{for } |\alpha| > 1.\end{cases} \tag{7.5.3}$$

Proof: The ROC of a function of the form $1/(z+\alpha)$ is either $|z| > \alpha$ or $|z| < \alpha$. First suppose $|\alpha| < 1$. Then the ROC has to be $|z| > |\alpha|$ in order to include the unit circle. In this case, the inverse transform of $1/(z+\alpha)$ will be a causal sequence so that

$$\left\{\frac{1}{z+\alpha}\right\}_+ = \frac{1}{z+\alpha}.$$

Now suppose $|\alpha| > 1$. Then the ROC has to be $|z| < |\alpha|$ in order to include the unit circle. In this case, the inverse transform of $1/(z+\alpha)$ will be an anticausal sequence so that

$$\left\{\frac{1}{z+\alpha}\right\}_+ = \frac{1}{\alpha}.$$

The ratio $1/\alpha$ is the value of the anticausal sequence at the origin and it can be found by simply evaluating the $z-$transform, $1/(z+\alpha)$, at $z=0$. ◆

(c)

$$\left\{\frac{1}{(z+\alpha)^i}\right\}_+ = \begin{cases} \dfrac{1}{(z+\alpha)^i} & \text{for } |\alpha| < 1, \\ \dfrac{1}{\alpha^i} & \text{for } |\alpha| > 1. \end{cases} \tag{7.5.4}$$

Proof: Similar to part (b). ◆

Three simple examples will illustrate how these rules can be used.

EXAMPLE 7.5.1 Compute

$$\left\{\frac{4z+3}{z^2+\frac{7}{3}z+\frac{2}{3}}\right\}_+.$$

Solution: We write

$$\frac{4z+3}{z^2+\frac{7}{3}z+\frac{2}{3}} = \frac{4z+3}{(z+2)(z+\frac{1}{3})} = \frac{3}{z+2}+\frac{1}{z+\frac{1}{3}}.$$

Therefore,

$$\left\{\frac{4z+3}{z^2+\frac{7}{3}z+\frac{2}{3}}\right\}_+ = \left\{\frac{3}{z+2}+\frac{1}{z+\frac{1}{3}}\right\}_+ = \left\{\frac{3}{z+2}\right\}_+ + \left\{\frac{1}{z+\frac{1}{3}}\right\}_+$$

$$= 3\cdot\frac{1}{2}+\frac{1}{z+\frac{1}{3}} = \frac{\frac{3}{2}z+\frac{3}{2}}{z+\frac{1}{3}}.$$

◆

EXAMPLE 7.5.2 Compute

$$\left\{\frac{(z+\frac{1}{3})(z^{-1}+\frac{1}{3})}{(z+\frac{1}{2})(z^{-1}+\frac{1}{2})}\right\}_+.$$

Solution: The first stage in the calculation is to eliminate the powers of z^{-1}. This can be done by multiplying both the numerator and the denominator by an appropriate power of z^n (in our case $n=1$). Moreover, since the function is not strictly proper (the degree of the denominator is not greater than that of the numerator) we first separate out the strictly proper part and then use partial fractions. In particular,

$$\frac{(z+\frac{1}{3})(z^{-1}+\frac{1}{3})}{(z+\frac{1}{2})(z^{-1}+\frac{1}{2})} = \frac{(z+\frac{1}{3})(1+\frac{1}{3}z)}{(z+\frac{1}{2})(1+\frac{1}{2}z)} = \frac{\frac{1}{3}z^2+\frac{4}{3}z+\frac{1}{3}}{\frac{1}{2}z^2+\frac{5}{4}z+\frac{1}{2}},$$

$$= \frac{2}{3}+\frac{z}{(z+\frac{1}{2})(z+2)} = \frac{2}{3}-\frac{\frac{1}{3}}{z+\frac{1}{2}}+\frac{\frac{4}{3}}{z+2},$$

or

$$\left\{\frac{(z+\frac{1}{3})(z^{-1}+\frac{1}{3})}{(z+\frac{1}{2})(z^{-1}+\frac{1}{2})}\right\}_+ = \frac{2}{3}-\frac{\frac{1}{3}}{z+\frac{1}{2}}+\frac{4}{3}\cdot\frac{1}{2} = -\frac{\frac{1}{3}}{z+\frac{1}{2}}.$$

◆

EXAMPLE 7.5.3 Compute

$$\left\{\frac{2z+3}{(z+\frac{1}{2})^2(z+2)}\right\}_+.$$

Solution: We can write

$$\frac{2z+3}{(z+\frac{1}{2})^2(z+2)} = \frac{A}{z+\frac{1}{2}}+\frac{B}{(z+\frac{1}{2})^2}+\frac{C}{z+2},$$

and evaluate the unknowns using

$$B = \left.\frac{2z+3}{z+2}\right|_{z=-\frac{1}{2}}, \quad C = \left.\frac{2z+3}{(z+\frac{1}{2})^2}\right|_{z=-2}, \quad A = \left.\frac{d}{dz}\cdot\frac{2z+3}{z+2}\right|_{z=-\frac{1}{2}},$$

to obtain $A=4/9$, $B=4/3$, and $C=-4/9$. Therefore,

$$\left\{\frac{2z+3}{(z+\frac{1}{2})^2(z+2)}\right\}_+ = \frac{\frac{4}{9}}{z+\frac{1}{2}}+\frac{\frac{4}{3}}{(z+\frac{1}{2})^2}-\frac{4}{9}\cdot\frac{1}{2} = \frac{-\frac{2}{9}z^2+\frac{2}{9}z+\frac{3}{2}}{(z+\frac{1}{2})^2}.$$

◆

7.6 IMPORTANT SPECIAL CASES AND EXAMPLES

There are often-encountered special cases in which the above results simplify even more. The corresponding formulas are well worth remembering. They will also show the way to alternative and ultimately more powerful approaches to the Wiener filtering problem.

7.6.1 Pure Prediction

Consider a zero-mean scalar stationary process $\{\mathbf{y}_i\}$ with z-spectrum $S_y(z)$. We would like to construct linear least-mean-squares *predictors* of $\mathbf{y}_{i+\lambda}$ using the past observations $\{\mathbf{y}_m\}_{m=-\infty}^{i}$, for some given $\lambda > 0$.

We can solve this problem directly, or by using the results of Thm. 7.4.1 with the assumption that

$$\mathbf{s}_i = \mathbf{y}_{i+\lambda}. \tag{7.6.1}$$

Then $\mathbf{s}(z) = z^\lambda \mathbf{y}(z)$ and $S_{sy}(z) = z^\lambda S_y(z)$. Therefore, using (7.4.12), the formula for the predictor becomes

$$K_\lambda(z) = \left\{\frac{z^\lambda S_y(z)}{r_e L^*(z^{-*})}\right\}_+ \frac{1}{L(z)} = \{z^\lambda L(z)\}_+ \frac{1}{L(z)}. \tag{7.6.2}$$

We can further simplify the above expression if we write the modeling filter as $L(z) = 1 + \sum_{m=1}^{\infty} l_m z^{-m}$, so that

$$\{z^{\lambda}L(z)\}_+ = \sum_{m=\lambda}^{\infty} l_m z^{-m+\lambda} = z^{\lambda}L(z) - z^{\lambda}\left[1 + \sum_{m=1}^{\lambda-1} l_m z^{-m}\right], \tag{7.6.3}$$

and the expression for the predictor becomes

$$K_{\lambda}(z) = z^{\lambda}\left(1 - \frac{1 + \sum_{m=1}^{\lambda-1} l_m z^{-m}}{L(z)}\right), \tag{7.6.4}$$

which is our desired result. It can also be shown that the m.m.s.e. is given by (see Prob. 7.15)

$$\text{m.m.s.e.} = r_e\left[1 + \sum_{m=1}^{\lambda-1} |l_m|^2\right]. \tag{7.6.5}$$

The special case of $\lambda = 1$ is instructive. In this case (7.6.4) reduces to

$$K_1(z) = z\left(1 - \frac{1}{L(z)}\right), \tag{7.6.6}$$

which is known as the *pure prediction filter* for estimating $\mathbf{y}_{i+1}$ given $\{\mathbf{y}_m, m \le i\}$.

For later use, we should note that the one-step predictor is usually expressed in terms of estimating $\mathbf{y}_i$ (not $\mathbf{y}_{i+1}$) given $\{\mathbf{y}_m, m \le i-1\}$ and the filter for doing this is

$$K_p(z) = 1 - \frac{1}{L(z)}. \tag{7.6.7}$$

It is worthwhile to dwell a bit on why this follows from (7.6.6). The point is that in the transfer function domain, we define

$$\mathbf{y}(z) = \sum_{i=-\infty}^{\infty} \mathbf{y}_i z^{-i}, \quad \hat{\mathbf{y}}(z) = \sum_{i=-\infty}^{\infty} \hat{\mathbf{y}}_i z^{-i}.$$

But what the filter $K_1(z)$ gives is (*cf.* (7.6.1))

$$K_1(z)\mathbf{y}(z) = \sum_{i=-\infty}^{\infty} \hat{\mathbf{s}}_i z^{-i} = \sum_{i=-\infty}^{\infty} \hat{\mathbf{y}}_{i+1} z^{-i} = z\sum_{i=-\infty}^{\infty} \hat{\mathbf{y}}_i z^{-i} = z\hat{\mathbf{y}}(z).$$

Therefore if we write $\hat{\mathbf{y}}(z) = K_p(z)\mathbf{y}(z)$, we see that

$$K_p(z) = z^{-1}K_1(z) = 1 - \frac{1}{L(z)}, \tag{7.6.8}$$

as claimed above. [More insight into the difference between (7.6.7) and (7.6.8) comes from the innovations approach to be discussed in Sec. 7.7.2.]

EXAMPLE 7.6.1 (Exponentially Correlated Processes) Consider a zero-mean scalar process $\{\mathbf{y}_i\}$ with

$$\langle \mathbf{y}_i, \mathbf{y}_j \rangle = a^{|i-j|} \text{ or } S_y(z) = \frac{1 - a^2}{(1 - az^{-1})(1 - az)}, \quad |a| < 1, \quad |a| < |z| < \frac{1}{|a|}.$$

The canonical factor is

$$L(z) = \frac{1}{1 - az^{-1}} = 1 + \sum_{i=1}^{\infty} a^i z^{-i}, \tag{7.6.9}$$

so that

$$\{z^{\lambda}L(z)\}_+ = \left\{z^{\lambda} + \sum_{i=1}^{\infty} a^i z^{(\lambda-i)}\right\}_+ = a^{\lambda}(1 + az^{-1} + a^2z^{-2} + \ldots)$$

$$= \frac{a^{\lambda}}{1 - az^{-1}}. \tag{7.6.10}$$

Then (7.6.2) yields

$$K_{\lambda}(z) = \frac{a^{\lambda}}{1 - az^{-1}}(1 - az^{-1}) = a^{\lambda},$$

or, in the time domain,

$$k_{\lambda,i} = \begin{cases} a^{\lambda} & \text{for } i = 0, \\ 0 & \text{for } i \neq 0. \end{cases}$$

The predicted estimator then is

$$\hat{\mathbf{y}}_{i+\lambda|i} = \sum_{m=0}^{\infty} k_{\lambda,m}\mathbf{y}_{i-m} = a^{\lambda}\mathbf{y}_i,$$

a striking result, for which one might expect a good physical reason and consequently a more direct derivation. In fact, the reader may recall that the above result was actually more directly derived in Ex. 3.3.2 and was revisited in Sec. 4.4; we pursued it much further in Sec. 5.1, a discussion to which we shall return at the end of this section. ◆

EXAMPLE 7.6.2 For another illustrative (and related) example, suppose now that $r_e = 1$, while

$$L(z) = \frac{1 - bz^{-1}}{1 - az^{-1}}, \quad |a| < 1, \quad |b| < 1.$$

Then

$$\{z^{\lambda}L(z)\}_+ = \left\{\frac{z^{\lambda}}{1 - az^{-1}} - \frac{bz^{\lambda-1}}{1 - az^{-1}}\right\}_+ = \frac{a^{\lambda}}{1 - az^{-1}} - \frac{ba^{\lambda-1}}{1 - az^{-1}} = \frac{a^{\lambda-1}(a - b)}{1 - az^{-1}},$$

where in the second step we used our calculation of $\{z^\lambda L(z)\}_+$ in (7.6.10). The desired filter is now given by

$$K_\lambda(z) = \{z^\lambda L(z)\}_+ \frac{1}{L(z)} = a^{\lambda-1}\frac{a-b}{1-bz^{-1}},$$

or in the time domain

$$k_{\lambda,i} = \begin{cases} a^{\lambda-1}(a-b)b^i & \text{for } i \geq 0, \\ 0 & \text{for } i < 0. \end{cases}$$

Therefore,

$$\hat{\mathbf{y}}_{i+\lambda|i} = a^{\lambda-1}(a-b)\sum_{m=0}^{\infty} b^m \mathbf{y}_{i-m}.$$

◆

Remark 1. [Recursive Form] From the rational function form of $K_\lambda(z)$, we can readily deduce a recursive form for the estimators. Thus, in Ex. 7.6.2, we can write

$$\hat{\mathbf{y}}_{i+\lambda|i} - b\hat{\mathbf{y}}_{i+\lambda-1|i-1} = a^{\lambda-1}(a-b)\mathbf{y}_i,$$

which could be initiated with $\hat{\mathbf{y}}_{-N+\lambda|-N} = 0$ or, in fact any arbitrary value, because the effect of the initial condition should die out for very large N.

This possibility was explicitly pointed out by Whittle (1963) who, unaware at the time of the Kalman filter, wrote that "such relations are often convenient for the recursive calculation of forecasts as time advances, and new data become available!"

Of course, going from a rational *matrix* function to a simple time-domain recursion is much less obvious, generally requiring the use of coprime matrix fraction descriptions (see, *e.g.*, Kailath (1980, Ch. 6)). ◆

7.6.2 Additive White Noise

In many problems, the processes $\{\mathbf{s}_i\}$ and $\{\mathbf{y}_i\}$ are related in an additive way,

$$\mathbf{y}_i = \mathbf{s}_i + \mathbf{v}_i, \tag{7.6.11}$$

where $\{\mathbf{v}_i\}$ is another zero-mean stationary noise process, in general jointly stationary with $\{\mathbf{s}_i\}$. In this case, obtaining $\{\hat{\mathbf{s}}_{i|i}\}$ is often spoken of as *filtering* the process $\{\mathbf{s}_i\}$ out of the noisy observations $\{\mathbf{y}_j,\ j \leq i\}$ and the estimators are called filtered estimators. When $\{\mathbf{v}_i\}$ is a white-noise process completely uncorrelated with $\{\mathbf{s}_i\}$, *i.e.*,

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = r\delta_{ij} \quad \text{and} \quad \langle \mathbf{v}_i, \mathbf{s}_j \rangle = 0, \tag{7.6.12}$$

or $S_v(z) = r$ and $S_{vs}(z) = 0$, the expression for the optimum filter can be further simplified. First, note that now

$$S_{sy}(z) = S_s(z) \quad \text{and} \quad S_y(z) = S_s(z) + S_v(z) = S_s(z) + r. \tag{7.6.13}$$

Then (7.4.12) reduces to

$$\begin{aligned} K(z) &= \left\{\frac{S_{sy}(z)}{L^*(z^{-*})}\right\}_+ \frac{1}{r_e L(z)}, \\ &= \left\{\frac{S_y(z) - r}{L^*(z^{-*})}\right\}_+ \frac{1}{r_e L(z)}, \\ &= \left\{L(z)r_e - \frac{r}{L^*(z^{-*})}\right\}_+ \frac{1}{r_e L(z)}, \\ &= \left[\{L(z)\}_+ - \left\{\frac{1}{L^*(z^{-*})}\right\}_+ \frac{r}{r_e}\right]\frac{1}{L(z)}. \end{aligned} \tag{7.6.14}$$

Since $L(z)$ is causal, we have $\{L(z)\}_+ = L(z)$, and since $L^{-*}(z^{-*})$ is anticausal and monic ($L^*(0) = 1$), we have

$$\left\{\frac{1}{L^*(z^{-*})}\right\}_+ = 1. \tag{7.6.15}$$

Using these last two expressions in (7.6.14) yields

$$K_f(z) = 1 - \frac{r}{r_e}L^{-1}(z), \tag{7.6.16}$$

so that finally we have the nice result

$$\hat{\mathbf{s}}_f(z) = \left[1 - \frac{r}{r_e}\frac{1}{L(z)}\right]\mathbf{y}(z). \tag{7.6.17}$$

Here, $\hat{\mathbf{s}}_f(z)$ denotes the z-transform of the sequence $\{\hat{\mathbf{s}}_{i|i}\}$. In other words, knowledge of the canonical spectral factor immediately defines the optimal filter. The subscripts f and p in (7.6.16) and (7.6.7) are used to indicate that they give the filtered and (one-step) predicted estimates, $\{\hat{\mathbf{s}}_{i|i}\}$ and $\{\hat{\mathbf{s}}_{i|i-1}\}$, respectively.

The simple form of the results (7.6.16) and (7.6.7) suggests that they may be more directly obtained. This is in fact true, as we shall show in Sec. 7.7.2 by using the concept of the innovations process.

One might wonder about the filter that yields the predicted signal estimators

$$\hat{\mathbf{s}}_i \triangleq \text{the l.l.m.s.e. of } \mathbf{s}_i \text{ given } \{\mathbf{y}_k,\ k \leq i-1\}.$$

From (7.6.11)–(7.6.12), we see that $\hat{\mathbf{y}}_i = \hat{\mathbf{s}}_i + \hat{\mathbf{v}}_i = \hat{\mathbf{s}}_i + 0$, so that finding $\hat{\mathbf{s}}_i$ is equivalent to finding the one-step predicted estimators $\{\hat{\mathbf{y}}_i\}$. But we studied this problem in Sec. 7.6.1 where we showed that (*cf.* (7.6.8))

$$K_p(z) = 1 - L^{-1}(z). \tag{7.6.18}$$

Remark 2. The reader may wish to ponder the reasons for the differences in (7.6.16) and (7.6.18). The formulas will become much more evident when we adopt the innovations approach — see Sec. 7.7.2. ◆

EXAMPLE 7.6.3 (Exponentially Correlated Process) Assume that a is real-valued and

$$S_s(z) = \frac{\sigma^2}{(1 - az^{-1})(1 - az)}, \quad |a| < 1, \quad S_v(z) = 1.$$

Then

$$S_y(z) = \frac{\sigma^2}{(1 - az^{-1})(1 - az)} + 1 = \frac{1 + \sigma^2 + a^2 - az - az^{-1}}{(1 - az^{-1})(1 - az)}.$$

We need to factor $S_y(z)$ in the form

$$S_y(z) = L(z) r_e L^*(z^{-*}) = r_e \frac{1 - \alpha z}{1 - az} \frac{1 - \alpha z^{-1}}{1 - az^{-1}}, \quad |\alpha| < 1, \quad r_e > 0.$$

This requires that we determine (r_e, α) such that

$$r_e[1 + \alpha^2 - \alpha z - \alpha z^{-1}] = 1 + \sigma^2 + a^2 - az - az^{-1},$$

which shows that we should have

$$r_e(1 + \alpha^2) = 1 + \sigma^2 + a^2 \quad \text{and} \quad r_e \alpha = a.$$

Solving for r_e we obtain the quadratic equation $r_e^2 - (1 + \sigma^2 + a^2) r_e + a^2 = 0$. This quadratic equation has two positive roots for r_e (since $a^2 > 0$ and $1 + \sigma^2 + a^2 > 0$). We must choose that root which results in $|\alpha| < 1$. Since $\alpha = a/r_e$, and the product of the two roots is a^2, this means that we must choose the larger root which is

$$r_e = \frac{1 + \sigma^2 + a^2 + \sqrt{(1 + \sigma^2 + a^2)^2 - 4a^2}}{2}.$$

The resulting α is

$$\alpha = \frac{2a}{1 + \sigma^2 + a^2 + \sqrt{(1 + \sigma^2 + a^2)^2 - 4a^2}}.$$

Now we have all the information needed to compute $K_f(z)$ and $K_p(z)$. ◆

Remark 3. The important thing about this example is that even in this almost the simplest filtering problem, the expressions for the Wiener filter are already quite complicated. Indeed for higher-order signal processes the algebra rapidly gets very much more cumbersome and closed form solutions are rarely feasible. This complexity was a major barrier to the widespread use of the theory and to extensions to problems with vector data, finite data, nonstationary signals, etc. This suggests that one should take a different approach to the problem. The innovations approach, which shows the significance of having (albeit a very special) a model for the process, is a step in this direction and we turn to it now. [The deeper message, reinforcing what we learned in Ch. 5, is that a way to (conceptually) tame this notational complexity is to use state-space descriptions. We shall demonstrate this in Ch. 8.] ◆

7.7 INNOVATIONS APPROACH TO THE WIENER FILTER

As first described in Sec. 4.2.4, in this approach, the estimation problem is broken into two parts: (i) finding the innovations $\{\mathbf{e}_i\}$ from the observations $\{\mathbf{y}_i\}$, and (ii) finding the estimators $\{\hat{\mathbf{s}}_{i|i}\}$ (of a process $\{\mathbf{s}_i\}$) from the innovations, as depicted in Fig. 7.2. As we may expect, the estimation problem given the innovations will be easy because the innovations are white. Thus, suppose that we wish to find

$$\hat{\mathbf{s}}_{i|i} = \sum_{k=-\infty}^{i} g_{i-k} \mathbf{e}_k. \tag{7.7.1}$$

The orthogonality condition gives

$$\langle \mathbf{s}_i - \hat{\mathbf{s}}_{i|i}, \mathbf{e}_l \rangle = 0 \quad \text{for} \quad -\infty < l \le i. \tag{7.7.2}$$

Now, by stationarity,

$$\langle \mathbf{s}_i, \mathbf{e}_l \rangle \triangleq R_{se}(i - l), \quad \langle \mathbf{e}_k, \mathbf{e}_l \rangle = r_e \delta_{kl},$$

so that (7.7.2) becomes the (trivial) Wiener-Hopf equation

$$R_{se}(i - l) = \sum_{k=-\infty}^{i} g_{i-k} r_e \delta_{kl} = g_{i-l} r_e, \quad \text{for} \quad l \le i.$$

In other words, the desired coefficients $\{g_i\}$ in (7.7.1) are given by

$$g_i = \begin{cases} R_{se}(i) r_e^{-1} & \text{for } i \ge 0, \\ 0 & \text{for } i < 0. \end{cases} \tag{7.7.3}$$

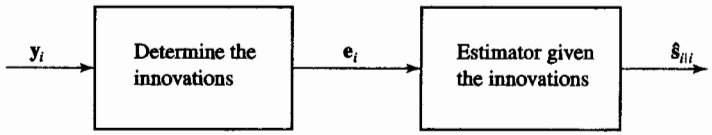


Figure 7.2 Estimation using the innovations approach.

Of course, this result is expected, since we could readily have written down the following expression for $\hat{\mathbf{s}}_{i|i}$ by projecting (component-wise) onto the orthogonal basis $\{\mathbf{e}_k\}_{k=0}^{i}$,

$$\hat{\mathbf{s}}_{i|i} = \sum_{k=-\infty}^{i} \langle \mathbf{s}_i, \mathbf{e}_k \rangle \|\mathbf{e}_k\|^{-2} \mathbf{e}_k = \sum_{k=-\infty}^{i} R_{se}(i-k) r_e^{-1} \mathbf{e}_k. \tag{7.7.4}$$

However, $R_{se}(i)$ is not part of the original statistical information, but it can be readily computed from the fact that $\{\mathbf{e}_i\}$ is the response of the whitening filter $L^{-1}(z)$ to the input $\{\mathbf{y}_i\}$. Therefore (see Lemma 6.3.1), since in the transform domain we have $\mathbf{e}(z) = L^{-1}(z)\mathbf{y}(z)$, it follows that the cross-spectrum $S_{se}(z)$ can be expressed in terms of the cross-spectrum $S_{sy}(z)$ as follows:

$$S_{se}(z) = \frac{S_{sy}(z)}{L^*(z^{-*})}, \tag{7.7.5}$$

where, of course,

$$S_{se}(z) \triangleq \sum_{k=-\infty}^{\infty} R_{se}(k) z^{-k}.$$

Using (7.7.3), which expresses $\{R_{se}(k)\}$ in terms of the desired coefficients $\{g_k\}$, we conclude that we must have

$$G(z) \triangleq \sum_{i=0}^{\infty} g_i z^{-i} = \sum_{i=0}^{\infty} R_{se}(i) r_e^{-1} z^{-i} = r_e^{-1} \{S_{se}(z)\}_+.$$

Comparing this expression for $G(z)$ with the equality (7.7.5), we see that we can make the identification

$$G(z) = \frac{1}{r_e} \left\{ \frac{S_{sy}(z)}{L^*(z^{-*})} \right\}_+.$$

It then follows that the transfer function of the (overall) filter for estimating $\mathbf{s}_i$ from $\{\mathbf{y}_m\}_{m=-\infty}^{i}$ can be found as the cascade (see Fig. 7.2)

$$K(z) = G(z) \cdot \frac{1}{L(z)} = \frac{1}{L(z)} \left\{ \frac{S_{sy}(z)}{r_e L^*(z^{-*})} \right\}_+. \tag{7.7.6}$$

Of course, we should note that the optimum filter will not generally be built as a cascade as in Fig. 7.2, because cancellations between $L^{-1}(z)$ and the other term in (7.7.6) will in general lead to a simpler (lower-order) filter — see the examples in Sec. 7.6.

It is nice, but not surprising, to check that (7.7.6) is exactly the solution found in Sec. 7.4 by using the Wiener-Hopf technique. Recall that that ingenious technique was based on the introduction (by fiat) of the canonical spectral factorization, followed by clever use of the analyticity properties of one-sided functions. Here the reasoning is much more physical: the canonical factorization provides the innovations, explaining the appearance of $L^{-1}(z)$ in the formula (7.7.6), while the remaining quantities we now realize to be just the solution of the (trivial) Wiener-Hopf equation for estimation given the white-noise innovations process.

These insights were first achieved independently by Bode and Shannon (1950) and Zadeh and Ragazzini (1950), while attempting to better understand Wiener's results on prediction and filtering of continuous-time stationary processes. Since these authors worked on the more difficult continuous-time problem, they were apparently not aware of the fact that the innovations in discrete time can be directly defined as (*cf.* Sec. 4.2.1) $\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i$, as done in fact by Wold (1938) and Kolmogorov (1939) in their elegant solutions of the pure prediction problem. We shall demonstrate the power of using this simpler definition in the next two subsections.

7.7.1 The Pure Prediction Problem

In Sec. 7.6.1, we noted that when $\mathbf{s}_i = \mathbf{y}_{i+1}$, the general formula (7.7.6) reduced to the simple form

$$K_1(z) = z\left(1 - \frac{1}{L(z)}\right).$$

In fact, this follows immediately from the expression

$$\hat{\mathbf{y}}_{i+1|i} = \mathbf{y}_{i+1} - \mathbf{e}_{i+1}, \quad -\infty < i < \infty,$$

by taking z-transforms:

$$\mathcal{Z}\{\hat{\mathbf{y}}_{i+1|i}\} = z\mathbf{y}(z) - z\mathbf{e}(z) = z\left(1 - \frac{1}{L(z)}\right)\mathbf{y}(z) = K_1(z)\mathbf{y}(z).$$

Moreover, the one-step prediction error is just (by stationarity)

$$\|\mathbf{y}_{i+1} - \hat{\mathbf{y}}_{i+1|i}\|^2 = \|\mathbf{e}_{i+1}\|^2 = r_e.$$

There are also simple expressions for multistep prediction. Thus note that, by definition, $\mathbf{y}(z) = L(z)\mathbf{e}(z)$, or

$$\mathbf{y}_i = \mathbf{e}_i + \sum_{k=1}^{\infty} l_k \mathbf{e}_{i-k},$$

from which we can write

$$\mathbf{y}_{i+\lambda} = \mathbf{e}_{i+\lambda} + \sum_{k=1}^{\infty} l_k \mathbf{e}_{i+\lambda-k}, \quad \lambda > 0,$$

so that

$$\hat{\mathbf{y}}_{i+\lambda|i} = \sum_{k=\lambda}^{\infty} l_k \mathbf{e}_{i+\lambda-k}.$$

Therefore,

$$\mathbf{y}_{i+\lambda} - \hat{\mathbf{y}}_{i+\lambda|i} = \mathbf{e}_{i+\lambda} + l_1 \mathbf{e}_{i+\lambda-1} + \ldots + l_{\lambda-1} \mathbf{e}_{i+1}, \tag{7.7.7}$$

and (since the innovations are mutually orthogonal)

$$\|\mathbf{y}_{i+\lambda} - \hat{\mathbf{y}}_{i+\lambda|i}\|^2 = r_e \left[1 + \sum_{k=1}^{\lambda-1} |l_k|^2\right], \tag{7.7.8}$$

while the transfer function of the multistep predictor is

$$K_\lambda(z) = z^\lambda \left(1 - \frac{1 + \sum_{k=1}^{\lambda-1} l_k z^{-k}}{L(z)}\right). \tag{7.7.9}$$

7.7.2 Additive White-Noise Problems

When

$$\mathbf{y}_i = \mathbf{s}_i + \mathbf{v}_i, \quad \langle \mathbf{v}_i, \mathbf{v}_j \rangle = r\delta_{ij}, \quad \langle \mathbf{v}_i, \mathbf{s}_j \rangle = 0, \tag{7.7.10}$$

then in an obvious notation

$$S_{sy}(z) = S_s(z) \ , \ S_y(z) = S_s(z) + S_v(z) = S_s(z) + r,$$

and in Sec. 7.6.2 we showed that

$$\hat{\mathbf{s}}_f(z) = \left[1 - \frac{r}{r_e} \cdot \frac{1}{L(z)}\right] \mathbf{y}(z), \tag{7.7.11}$$

or in the time domain

$$\hat{\mathbf{s}}_{i|i} = \mathbf{y}_i - \frac{r}{r_e}\mathbf{e}_i. \tag{7.7.12}$$

However, this expression can easily be obtained by simple geometric arguments. Note that $\hat{\mathbf{s}}_{i|i}$ is the projection of $\mathbf{s}_i$ onto the linear space $\mathcal{L}\{\mathbf{y}_j, j \le i\}$. But from (7.7.10) we can write

$$\hat{\mathbf{y}}_{i|i} = \hat{\mathbf{s}}_{i|i} + \hat{\mathbf{v}}_{i|i}, \tag{7.7.13}$$

where $\hat{\mathbf{y}}_{i|i}$ and $\hat{\mathbf{v}}_{i|i}$ are the obvious projections onto $\mathcal{L}\{\mathbf{y}_j, j \le i\}$. Now clearly $\hat{\mathbf{y}}_{i|i} = \mathbf{y}_i$, while we can compute $\hat{\mathbf{v}}_{i|i}$ via the innovations as

$$\hat{\mathbf{v}}_{i|i} = \sum_{k=-\infty}^{i} \langle \mathbf{v}_i, \mathbf{e}_k \rangle \|\mathbf{e}_k\|^{-2} \mathbf{e}_k = \langle \mathbf{v}_i, \mathbf{e}_i \rangle \ r_e^{-1} \ \mathbf{e}_i,$$

where the fact that $\langle \mathbf{v}_i, \mathbf{e}_j \rangle = 0$, for all $j < i$, follows easily from the assumptions in (7.7.10) and the fact that $\mathbf{e}_i \in \mathcal{L}\{\mathbf{y}_k, k \le i\}$. Now we can compute $\langle \mathbf{v}_i, \mathbf{e}_i \rangle$ as follows:

$$\langle \mathbf{v}_i, \mathbf{e}_i \rangle = \langle \mathbf{v}_i, \mathbf{y}_i - \hat{\mathbf{y}}_i \rangle = \langle \mathbf{v}_i, \mathbf{y}_i \rangle - \underbrace{\langle \mathbf{v}_i, \hat{\mathbf{y}}_i \rangle}_{=0} = \langle \mathbf{v}_i, \mathbf{s}_i + \mathbf{v}_i \rangle = \underbrace{\langle \mathbf{v}_i, \mathbf{s}_i \rangle}_{=0} + \langle \mathbf{v}_i, \mathbf{v}_i \rangle = r.$$

Collecting these expressions gives

$$\mathbf{y}_i = \hat{\mathbf{s}}_{i|i} + \frac{r}{r_e}\mathbf{e}_i,$$

which is the same as (7.7.12)!

A bonus of this derivation is a simple expression for the m.m.s.e. Thus

$$\mathbf{s}_i - \hat{\mathbf{s}}_{i|i} = \frac{r}{r_e}\mathbf{e}_i - \mathbf{v}_i,$$

so that

$$\|\mathbf{s}_i - \hat{\mathbf{s}}_{i|i}\|^2 = \frac{r^2}{r_e} - \frac{2r}{r_e}\langle \mathbf{e}_i, \mathbf{v}_i \rangle + r = \frac{r^2}{r_e} - \frac{2r^2}{r_e} + r = r\left(1 - \frac{r}{r_e}\right). \tag{7.7.14}$$

The important fact to remember from this example is that in the additive noise problem, knowledge of the canonical spectral factor $L(z)$, or rather $L^{-1}(z)$, immediately determines the solution — see (7.7.11).

7.8 VECTOR PROCESSES

There is no formal difficulty in extending many of the earlier results to the case of vector-valued random processes with rational z-spectra (defined in Sec. 6.6), as we now indicate.

For example, in an obvious notation, we can see that the optimum Wiener smoother is given by

$$K(z) = S_{sy}(z)S_y^{-1}(z).$$

In the filtering problem, the matrix Wiener-Hopf equation is

$$R_{sy}(i) = \sum_{m=0}^{\infty} K_m R_y(i - m) \quad \text{for} \quad i \ge 0, \tag{7.8.1}$$

where now $\{K_i\}$ is a sequence of $n \times p$ complex matrices. We shall assume that the rational z-spectrum $S_y(z)$ has maximal normal rank *everywhere* on the unit circle, *i.e.*, (*cf.* Sec. 6.6)

$$S_y(e^{j\omega}) > 0, \quad -\pi \le \omega \le \pi.$$

[In our further discussions, this condition will often be ensured by assuming that the observations $\{\mathbf{y}_i\}$ have a white-noise component that is completely uncorrelated with the process we wish to estimate.]

Then according to the discussion in Sec. 6.6, these assumptions guarantee the existence of a positive-definite matrix R_e, and of a unique $p \times p$ rational canonical spectral factor $L(z)$ such that

$$S_y(z) = L(z)R_eL^*(z^{-*}), \tag{7.8.2}$$

where $L(z)$ and $L^{-1}(z)$ are analytic on and outside the unit circle ($|z| \ge 1$), and $L(\infty) = I_p$. Such a factorization allows us to extend the Wiener-Hopf solution to the filtering problem in the vector case and to obtain

$$K(z) = \left\{S_{sy}(z)L^{-*}(z^{-*})\right\}_+ R_e^{-1}L^{-1}(z).$$

We consider a special case in the following statement.

Theorem 7.8.1 (Additive White Noise) *When $\mathbf{y}_i = \mathbf{s}_i + \mathbf{v}_i$, with known rational z-spectra and z-cross-spectra, $S_s(z)$, $S_v(z) = R$, and $S_{sv}(z) = 0$, then the l.l.m.s.e. of $\mathbf{s}_{i+\lambda}$ (for some integer $\lambda > 0$) given $\{\mathbf{y}_m, -\infty < m \le i\}$ is given by*

$$\hat{\mathbf{s}}_{i+\lambda|i} = \sum_{m=-\infty}^{i} K_{i-m}\mathbf{y}_m = \sum_{m=0}^{\infty} K_m \mathbf{y}_{i-m},$$

where $K(z)$, the z-transform of $\{K_i\}$, is given by

$$K(z) = \left\{z^{\lambda}[L(z) - RL^{-*}(z^{-*})R_e^{-1}]\right\}_+ L^{-1}(z),$$

and $L(z)$ and R_e are found from the canonical factorization

$$S_y(z) = S_s(z) + R = L(z)R_eL^*(z^{-*}).$$

Moreover, for $\lambda = 1$, the filter $K(z)$ specializes to $K_p(z) = z[I - L^{-1}(z)]$ and for $\lambda = 0$, we have $K_f(z) = I - RR_e^{-1}L^{-1}(z)$. ∎

Remark 4. As noted earlier, a major issue with the use of the Wiener formulas is the computation of the canonical spectral factor. It turns out that the explicit introduction of state-space structure helps not only in this regard, but also in addressing the issue of algebraic complexity noted in Remark 3. It is not that the algebra mysteriously disappears — the point is that the state-space language helps to push the complexities to a lower level, allowing the conceptual structure to be clearer. We shall see all this in Ch. 8 for stationary processes and then in Ch. 9 for nonstationary processes. First, however, we shall briefly review some of the developments between the understanding and dissemination of Wiener's work in the early 1950s and the introduction of state-space ideas at the end of that decade. ◆

7.9 EXTENSIONS OF WIENER FILTERING

Early extensions to the theory of Wiener filtering were to accommodate power and saturation constraints of various sorts (see, *e.g.*, Jaffe and Rechtin (1955) and Newton, Gould, and Kaiser (1957)); a generic example is presented in Prob. 7.10. There were also extensions, beginning with Zadeh and Ragazzini (1950), to estimation over finite (fixed or growing) observation intervals and for stationary and nonstationary processes. With some exceptions, most of these results were for continuous-time problems, as studied here in App. 7.A. The corresponding generalization of the classical Wiener-Hopf integral equation (7.2.3) is

$$R_{sy}(t, \tau) = \int_0^t k(t, \sigma)R_y(\sigma, \tau)d\sigma, \qquad 0 \le \tau \le t, \tag{7.9.1}$$

corresponding to the estimator

$$\hat{\mathbf{s}}(t) = \int_0^t k(t, \sigma)\mathbf{y}(\sigma)d\sigma.$$

The classical Wiener-Hopf spectral factorization technique can be extended to such equations by now seeking a canonical *covariance* factorization of the form

$$R_y(t, \tau) = \int_0^T \bar{L}(t, \sigma)\bar{L}^*(\tau, \sigma)d\sigma, \qquad 0 \le t, \tau \le T, \tag{7.9.2}$$

where $\bar{L}(.,.)$ is causal, *i.e.*, $L(t, \sigma) = 0$ for $\sigma > t$.

The problem is that, especially when $R_y(.,.)$ is well behaved, the "factor" $\bar{L}(.,.)$ is very difficult to obtain or even characterize. So direct covariance factorization was abandoned. Instead a host of special methods was developed[12] (see, *e.g.*, Zadeh and Ragazzini (1950), Yaglom(1955,1962), Laning and Battin (1956), Darlington (1959), Hajek (1962), Pisarenko and Rozanov (1963), Whittle (1963), Helstrom (1965), and Slepian and Kadota (1969)) for solving (7.9.1) and several variants.

It turns out that the most useful results are obtained by assuming that the observations contain a white noise component, so that $R_y(.,.)$ has the form

$$R_y(t, \tau) = R(t)\delta(t - \tau) + K(t, \tau), \tag{7.9.3}$$

in which case the canonical covariance factor will have the form

$$\bar{L}(t, \tau) = R^{1/2}(t)\delta(t - \tau) + l(t, \tau), \tag{7.9.4}$$

and $l(.,.)$ is now as smooth as $K(.,.)$. Furthermore, all the above authors assumed the stationary case in which $R_y(t)$ was a sum of terms of the form $t^k e^{-\alpha_k|t|}$ (*i.e.*, $\mathbf{y}(.)$ had a rational power spectral density function). The natural generalization to the nonstationary case was to the so-called "semi-separable" kernels of the form

$$R_y(t, \tau) = \begin{cases} \sum_{i=1}^n a_i(t)b_i(\tau), & 0 \le t \le \tau, \\ \sum_{i=1}^n b_i(t)a_i(\tau), & 0 \le \tau \le t, \end{cases} \tag{7.9.5}$$

which were studied by Shinbrot (1957). [Of course this form includes the stationary case; for example, choosing $n = 1$, $a(t) = e^{\alpha t}$, $b(t) = e^{\alpha t}$ implies that $R_y(t - \tau) = e^{-\alpha|t-\tau|}$.] It was later realized that processes with finite-dimensional, possibly time-variant, state-space models had "semi-separable" covariance functions.

However, while it was true that closed-form solutions could be obtained under the above assumptions, they were algebraically quite complicated. Nevertheless they were used to study various missile tracking, intercept, navigation, and guidance problems (see, *e.g.*, the monograph of Peterson (1961), especially Chs. 6 and 8). We should mention that Peterson explicitly introduced linear and nonlinear state-space process models and also discussed the realization of the optimum impulse response in analog-computer form. [As noted in Sec. 5.5, such descriptions had also been used by Laning and Battin (1956) (and perhaps others).] Nevertheless, the concept of a recursive solution was not

[12] So much of the literature dealt with minor variations and special cases that already in 1958 Elias felt compelled to editorialize in the *IEEE Transactions on Information Theory* that it was time to stop writing "two famous papers." One was "The optimum linear mean-square filter for separating sinusoidally-modulated triangular signals from randomly-sampled stationary Caussian noise, with applications to a problem in radar." (The other was "Information theory, photosynthesis, and religion," a title suggested by D. Huffman.)

explicit and, as stated before, the integral equation solutions were complicated both in form and derivation.

Complete and elegant solutions awaited the work in discrete time of Kalman (1960a), and in continuous time of Carlton and Follin (1956), Bucy (1959), Kalman and Bucy (1961), and Stratonovich (1959,1960a,1960b). These results will be discussed at length beginning in Ch. 9 (see Ch. 16 for the continuous-time case). Though these results were all differently obtained, in retrospect the key fact is that the assumption of a state-space model enables computationally efficient algorithms for spectral/covariance factorization. We shall illustrate this point in the next chapter and show how it offers several useful perspectives on the relation between the Wiener and Kalman approaches.

7.10 COMPLEMENTS

As mentioned on several occasions already, Wiener's solution of the problems of prediction and filtering had a huge impact on many fields of engineering and mathematics (even though it fell short of its original goal of solving the anti-aircraft-gun control problem). For example, Doob (1953) devoted the last chapter of his famous textbook to it. So also the first sentence of Kalman (1963b) is "There is no doubt that Wiener's theory of statistical prediction and filtering is one of the great contributions to engineering science."

On the mathematical side, Wiener's development even in the scalar case was not as general as the independent work of Kolmogorov (1939,1941a,b). Wiener focused on the case where explicit expressions could be obtained for the optimum predictor. [A footnote in Wiener (1949, p. 59) well describes the relationships of his work to Kolmogorov's.]

Among engineering-oriented textbooks that treat the discrete-time Wiener-Hopf theory we cite specifically the remarkable little monograph of Whittle (1963), and its updated second edition, Whittle (1983); see also Yaglom (1962), which focuses on the rational case. Treatments exclusively in continuous time are given by Lee (1960) and Van Trees (1968), both of which have large collections of problems. We should also of course mention Wiener's original monograph (1942, 1949); in fact, most of the explicit examples appearing in later books were first worked out by Wiener. Moreover, Wiener's monograph is well worth browsing in for many reasons, including its clear emphasis (in 1942) on the statistical nature of the communications problem and its illustration of the ways in which practical constraints can be explored within the optimal theory (see, *e.g.*, Kailath (1997) and the references therein).

On the Wiener-Hopf equation itself there is a huge literature. Here we may only mention two long papers by Krein (1958) and Gohberg and Krein (1958), and also the book of Gohberg and Fel'dman (1974), which present the theory in the most general case (*e.g.*, without the assumption that $S_y(z) > 0$ on $|z| = 1$).

Finally, we may remark that the material in this chapter has generally been ignored in most textbooks on state-space estimation on the grounds that the Kalman filter theory subsumes the Wiener theory. For many reasons, this has turned out to be a rather short-sighted point of view. For one thing, the simple and physically meaningful Wiener formulas for smoothing are readily derived (see Secs. 7.1 and 7.3), while they are more difficult to obtain as asymptotic cases of the various state-space smoothing formulas of Ch. 10. Moreover, the Wiener-Hopf equation arises in many fields and has a long history; this can enable fruitful insights to be transferred across fields. A direct example is the application to estimation theory of insights and results first obtained in radiative transfer theory (see Ch. 17). In fact, it was here in the work of Ambartsumian (1943) that the Riccati equation was perhaps first introduced to solve a Wiener-Hopf equation. And this led to the work of Chandrasekhar (1947a,1947b), and many years later to the results in Chs. 11, 13, and 17. We may also mention here the occurrence of the Wiener-Hopf equations in the theory of queuing systems and computer networks (see, *e.g.*, Kleinrock (1975)); cross-couplings with state-space estimation have not yet been explored.

■ PROBLEMS

7.1 **(An alternative derivation)** Consider the continuous-time zero-mean stationary random processes $\mathbf{s}(\cdot)$ and $\mathbf{y}(\cdot)$ with given power spectra $S_s(f)$ and $S_y(f)$, and cross-power spectrum $S_{sy}(f)$. We would like to estimate the signal process $\mathbf{s}(\cdot)$ from the observations process $\mathbf{y}(\cdot)$ in the least-mean-squares sense using a linear *time-invariant* filter $K(f)$. In other words, let $\hat{\mathbf{s}}(f)$ denote the Fourier transform of the estimated signal and let $\mathbf{y}(f)$ denote the Fourier transform of the observations process. We seek a filter $K(f)$ so that $\hat{\mathbf{s}}(f) = K(f)\mathbf{y}(f)$.

(a) Show that $S_{\tilde{s}}(f)$, the power spectrum of the estimation error $\tilde{\mathbf{s}}(\cdot) = \mathbf{s}(\cdot) - \hat{\mathbf{s}}(\cdot)$, is given by

$$S_{\tilde{s}}(f) = \begin{bmatrix} 1 & -K(f) \end{bmatrix} \begin{bmatrix} S_s(f) & S_{sy}(f) \\ S_{sy}^*(f) & S_y(f) \end{bmatrix} \begin{bmatrix} 1 \\ -K^*(f) \end{bmatrix}.$$

(b) Show that we can write

$$S_{\tilde{s}}(f) = S_s(f) - \frac{|S_{sy}(f)|^2}{S_y(f)} - \left[K(f) - \frac{S_{sy}(f)}{S_y(f)} \right] S_y(f) \left[K(f) - \frac{S_{sy}(f)}{S_y(f)} \right]^*,$$

and deduce that the m.s.e. $E|\tilde{\mathbf{s}}|^2$ is minimized by the choice $K(f) = S_{sy}(f)/S_y(f)$.

Remark. The above derivation, first given by Bode and Shannon (1950), assumes a priori that the optimum filter is time-invariant; this is not necessary, as shown in Sec. 7.3.1. ◆

7.2 **(Comparison with the ideal filter)** Consider the Wiener smoothing problem for the additive model $\mathbf{y}(t) = \mathbf{s}(t) + \mathbf{v}(t)$, where the signal and noise processes $\mathbf{s}(\cdot)$ and $\mathbf{v}(\cdot)$ are zero-mean, uncorrelated, and have power spectra $S_v(f) = N_0/2$ for $-\infty < f < \infty$ and

$$S_s(f) = \begin{cases} A\left(a - \frac{|f|}{f_c}\right) & \text{for } |f| < f_c, \\ 0 & \text{elsewhere,} \end{cases}$$

with $A \geq 1$ (make a sketch of $S_s(f)$).

(a) Calculate and sketch the transfer function of the optimum smoothing filter.

(b) Show that the

$$\text{m.m.s.e.} = N_0 f_c \left[1 - \frac{N_0}{2A} \log \frac{a + N_0/(2A)}{a - 1 + N_0/(2A)} \right].$$

(c) Study the asymptotic behavior in the two cases (A fixed, $N_0 \to \infty$) and ($A \to \infty$, N_0 fixed), and compare with the performance of the ideal lowpass filter

$$I(f) = \begin{cases} 1 & \text{for } |f| < f_c, \\ 0 & \text{elsewhere.} \end{cases}$$

7.3 **(Multiplicative noise)** Consider the model $\mathbf{y}(t) = \mathbf{n}(t)\mathbf{s}(t)$ for $-\infty < t < \infty$, where the signal process $\mathbf{s}(\cdot)$ is stationary and zero-mean with power spectral density function $S_s(f)$, and the noise process $\mathbf{n}(\cdot)$ is independent of $\mathbf{s}(\cdot)$ and has mean $\bar{n}$ and spectrum $S_n(f)$.

(a) Show that the optimum Wiener smoother for estimating $\mathbf{s}(t)$ using $\{\mathbf{y}(\tau),\ -\infty < \tau < \infty\}$ is given by

$$K(f) = \frac{\bar{n} S_s(f)}{\bar{n}^2 S_s(f) + S_n(f) \star S_s(f)},$$

where $\star$ denotes convolution.

(b) Determine $K(f)$ when the autocorrelation functions are $R_n(\tau) = e^{-\alpha^2 \tau^2}$ and $R_s(\tau) = e^{-\beta^2 \tau^2}$.

7.4 **(Time invariance)** Here we demonstrate that Eq. (7.3.10), *viz.*,

$$R_{sy}(i) = \sum_{m=-\infty}^{i} w_{i+l,m+l} R_y(m) \quad \text{for } i \geq 0 \text{ and for all } l,$$

implies that we must have $w_{i+l,m+l} = w_{im} = k_{i-m}$, for some sequence $\{k_i\}$.

(a) Show that $R_{sy}(i) = \sum_{m=-\infty}^{i} w_{0,m-i} R_y(m)$ for $i \geq 0$.

(b) Conclude that $\{w_{0,i}\}$ is the unique solution to the Wiener-Hopf equation

$$R_{sy}(i) = \sum_{m=-\infty}^{i} k_{i-m} R_y(m) \quad \text{for } i \geq 0.$$

(c) Likewise show that the sequence $\{w_{l,i+l}\}$ is the unique solution to the Wiener-Hopf equation for any l.

(d) Conclude that $\{w_{l,i+l}\} = \{w_{0,i}\} = \{k_i\}$.

7.5 **(Filtered estimator)** Consider the model $\mathbf{y}_i = \mathbf{s}_i + \mathbf{v}_i + b\mathbf{v}_{i-1}$, for $i > -\infty$, where $\mathbf{s}_i$ and $\mathbf{v}_i$ are uncorrelated zero-mean scalar stationary processes with known z-spectra, $S_s(z)$ and $S_v(z) = r$. Assume that we have carried out the canonical spectral factorization of $S_y(z)$, so that we know the modeling and whitening filters,

$$L(z) = 1 + \sum_{k=1}^{\infty} l_k z^{-k}, \qquad W(z) = 1 + \sum_{k=1}^{\infty} w_k z^{-k},$$

respectively, and the innovations variance r_e. Show that

$$\hat{\mathbf{s}}_{i|i} = \mathbf{y}_i - \frac{r}{r_e}\left(1 + bw_1^* + |b|^2\right)\mathbf{e}_i - \frac{br}{r_e}\mathbf{e}_{i-1}.$$

7.6 **(The Wiener solution)** Let $\{\alpha_i, -\infty < i < \infty\}$ denote the inverse transform sequence of $S_{sy}(z)/r_e L^*(z^{-*})$. Show that

$$K(z)L(z) = \sum_{i=0}^{\infty} \alpha_i z^{-i} = \sum_{i=0}^{\infty} \left[\frac{1}{j2\pi} \oint_C \frac{S_{sy}(w)}{r_e L^*(w^{-*})} w^{i-1} dw \right] z^{-i},$$

where C is a closed contour outside and encircling the unit circle.

7.7 **(Multistep prediction and filtering)** Consider the zero-mean jointly stationary scalar random processes $\{\mathbf{s}_i\}$ and $\{\mathbf{y}_i\}$, and suppose we would like to estimate $\mathbf{s}_{i+\lambda}$, where λ is a fixed positive or negative integer, using the observations $\{\mathbf{y}_j\}_{j=-\infty}^{i}$. Thm. 7.4.1 considers the special case $\lambda = 0$.

(a) Let $\hat{s}_\lambda(z)$ denote the z-transform of the estimate $\hat{\mathbf{s}}_{i+\lambda|i}$. Show that

$$\hat{\mathbf{s}}_\lambda(z) = \left\{ \frac{z^\lambda S_{sy}(z)}{L^*(z^{-*})} \right\}_+ \frac{1}{r_e L(z)} \mathbf{y}(z),$$

where $L(z)$ and r_e are found from the unique canonical spectral factorization of $S_y(z)$.

(b) Show that the m.m.s.e. is given by $E|\mathbf{s}_i - \hat{\mathbf{s}}_{i+\lambda|i}|^2 =$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left[S_s(e^{j\omega}) - \frac{|S_{sy}(e^{j\omega})|^2}{S_y(e^{j\omega})} \right] d\omega + \frac{1}{2\pi r_e} \int_{-\pi}^{\pi} \left| \left\{ \frac{e^{j\lambda\omega} S_{sy}(e^{j\omega})}{L^*(e^{-j\omega})} \right\}_- \right|^2 d\omega.$$

7.8 **(A simple example)** Find the l.l.m.s. predictor of $\mathbf{y}_{i+\lambda}$, $\lambda \geq 0$, for a process with autocorrelation function

$$R_y(i) = \frac{1}{2}\delta(i+1) + \frac{5}{4}\delta(i) + \frac{1}{2}\delta(i-1), \qquad -\infty < i < \infty.$$

7.9 **(An interpolation problem (Whittle (1983)))** Consider a zero-mean stationary continuous-time process $\mathbf{y}(t)$ and suppose that it is observed at discrete instants of time, say at multiples of a sampling period T, $\mathbf{y}_i = \mathbf{y}(t)|_{t=iT}$, for all integer i. The autocorrelation function of the continuous-time process $\{\mathbf{y}(\cdot)\}$ is $R_y(\tau) = \frac{\sigma^2}{2\alpha} e^{-\alpha|\tau|}$. Given all the discrete-time obsevations $\{\mathbf{y}_i, -\infty < i < \infty\}$, we would like to determine the l.l.m.s. estimator of $\mathbf{y}(iT + \nu)$ for some real value $\nu \in (0, T)$. Let $\rho = e^{-\alpha T}$ and $\lambda = e^{-\alpha \nu}$. Show that the desired estimator is

$$\hat{\mathbf{y}}(iT + \nu) = \frac{\lambda - \lambda^{-1}\rho^2}{1 - \rho^2}\, \mathbf{y}(iT) \;+\; \frac{\rho(\lambda^{-1} - \lambda)}{1 - \rho^2}\, \mathbf{y}((i+1)T).$$

7.10 **(Regularized Wiener filtering)** Consider two zero-mean scalar-valued stationary random processes $\{\mathbf{s}_i, \mathbf{y}_i\}$ with z-spectra and cross-spectra $S_y(z)$ and $S_{sy}(z)$. Let

$$\hat{\mathbf{s}}_{i|i} = \sum_{j=-\infty}^{i} k_{i-j}\mathbf{y}_j \triangleq k \star \mathbf{y},$$

and define the filtered sequence

$$\hat{\mathbf{s}}^f_{i|i} = \sum_{j=-\infty}^{i} g_{i-j}\hat{\mathbf{s}}_{j|j} \triangleq g \star \hat{\mathbf{s}},$$

where $G(z)$ is a given stable and causal transfer function with impulse response $\{g_j\}$. Consider the problem

$$\min_{\{k_i\}} \; E\left[|\mathbf{s}_i - \hat{\mathbf{s}}_{i|i}|^2 \;+\; |\hat{\mathbf{s}}^f_{i|i}|^2\right].$$

(a) Verify that the cost function can be rewritten in the equivalent form

$$\min_{k_i} \; E(\mathbf{s}'_{i|i} - \hat{\mathbf{s}}'_{i|i})(\mathbf{s}'_{i|i} - \hat{\mathbf{s}}'_{i|i})^*,$$

where $\mathbf{s}'_i = [\mathbf{s}_i \;\; 0]$ and $\hat{\mathbf{s}}'_{i|i} = k \star [\mathbf{y} \;\; g \star \mathbf{y}]$.

(b) Conclude that

$$K(z) = \left\{\frac{S_{sy}(z)}{\bar{r}_e \bar{L}^*(z^{-*})}\right\}_+ \frac{1}{\bar{L}(z)},$$

where $\bar{L}(z)\bar{r}_e\bar{L}^*(z^{-*})$ is the canonical spectral factorization of the z-spectrum $S_y(z) + G(z)S_y(z)G^*(z^{-*})$.

Remark. The additional term $g \star \hat{\mathbf{s}}$ allows us to incorporate weighting into the frequency domain. For example, by choosing g as a low- or highpass filter we can attribute less or more significance to low or high frequencies in the estimator $\hat{\mathbf{s}}_{i|i}$. Newton, Gould, and Kaiser (1957) introduced the idea of Wiener filtering with constraints (which can be reduced to the above form by using Lagrange multipliers) to study control problems with (for example) saturation constraints. ◆

7.11 **(Additive white noise)** Consider the model $\mathbf{y}_i = \mathbf{s}_i + \mathbf{v}_i$, where the zero-mean stationary process $\{\mathbf{s}_i\}$ is completely uncorrelated with the zero-mean white noise $\{\mathbf{v}_i\}$ of intensity r. Show that the m.m.s.e. in estimating $\mathbf{s}_i$ given $\{\mathbf{y}_m, m \le i\}$ is equal to $E\,|\mathbf{s}_i - \hat{\mathbf{s}}_i|^2 = r(1 - \frac{r}{r_e})$.

7.12 **(Unit-circle zeros)** Let $W(z)$ be a rational transfer function that is analytic in $|z| > 1$ (and, hence, its poles are inside the closed unit circle).

(a) Assume first that $W(z) = 1/(z - \alpha)$, where $|\alpha| = 1$. Find the impulse response sequence $\{w_i\}$ of $W(z)$ and show that it is not square-summable.

(b) More generally, by expanding $W(z)$ in partial fractions, show that if $W(z)$ has a pole on the unit circle, then its impulse response sequence cannot be square-summable.

7.13 **(Higher-order prediction for ARMA process)** Consider a zero-mean stationary process $\{\mathbf{y}_i\}$ generated by a second-order ARMA model of the form (*cf.* Prob. 5.3),

$$\mathbf{y}_{i+1} = a_0\mathbf{y}_i + a_1\mathbf{y}_{i-1} + \mathbf{u}_i + b\mathbf{u}_{i-1}, \quad i > -\infty,$$

where $\{\mathbf{u}_i\}$ denotes a zero-mean white-noise stationary sequence with variance Q. The signal $\mathbf{u}_i$ is assumed to be uncorrelated with current and past observations, $\{\mathbf{y}_j, j \le i\}$. The roots of the characteristic polynomial $z^2 - a_0 z - a_1$ are assumed to be strictly inside the unit circle. We would like to construct a linear least-mean-squares *predictor* of $\mathbf{y}_{i+3}$ using the past observations $\{\mathbf{y}_m\}_{m=-\infty}^{i}$. Using (7.4.12), the formula for the predictor is given by

$$K_3(z) = \left\{z^3 L(z)\right\}_+ \frac{1}{L(z)},$$

where $L(z)$ denotes the canonical spectral factor of the z-spectrum of the process $\{\mathbf{y}_i\}$ (recall Ex. 6.5.3). Determine $K_3(z)$ when $|b| < 1$ and when $|b| > 1$. Compare the results with those of Prob. 5.4.

7.14 **(Polynomial approach to higher-order prediction)** We introduced earlier in Prob. 5.5 a polynomial approach to the solution of a higher-order prediction problem for an ARMA process. We treated this same situation above in Prob. 7.13 by using the Wiener-Hopf technique. We now establish that both methods, the polynomial method and the Wiener-Hopf method, are in fact identical in this situation where the characteristic polynomial is assumed stable. So consider the same setting as in Prob. 7.13. We know using (7.4.12) that the desired predictor is given by

$$K_3(z) = \left\{z^3 L(z)\right\}_+ \frac{1}{L(z)},$$

where, according to the result of Ex. 6.5.3, $L(z) = z(z + b)/(z^2 - a_0 z - a_1)$. Show that $K_3(z) = F(z)/(z + b)$, where $F(z)$ is such that

$$\frac{z^3(z + b)}{z^2 - a_0 z - a_1} \;=\; E(z) + \frac{F(z)}{z^2 - a_0 z - a_1},$$

with the degree of $F(z)$ strictly less than that of $z^2 - a_0 z - a_1$.

7.15 **(Pure prediction)** Refer to Ex. 7.6.1 and consider a zero-mean stationary process $\{\mathbf{y}_i\}$, for which we would like to predict $\mathbf{y}_{i+\lambda}$ ($\lambda > 0$) using the past observations $\{\mathbf{y}_m\}_{m=-\infty}^{i}$.

(a) Define the error signal as $\boldsymbol{\epsilon}_i = \mathbf{y}_{i+\lambda} - \hat{\mathbf{y}}_{i+\lambda|i}$, and show that

$$S_\epsilon(z) = r_e\left[1 + \sum_{k=1}^{\lambda-1} l_k z^{-k}\right]\left[1 + \sum_{k=1}^{\lambda-1} l_k^* z^k\right].$$

(b) Deduce that the corresponding m.m.s.e. is

$$\text{m.m.s.e.} = r_e\left[1 + \sum_{k=1}^{\lambda-1} |l_k|^2\right].$$

[*Hint.* One method is to find the coefficient of z^0 in $S_{\tilde{s}}(z)$.]

7.16 (Mean-square error) Consider the first problem in Sec. 7.6.

(a) Show that the z-transform of the prediction error $\mathbf{y}_{i+\lambda} - \hat{\mathbf{y}}_{i+\lambda|i}$ is given by $\tilde{\mathbf{y}}(z) = (z^\lambda - a^\lambda)\mathbf{y}(z)$.

(b) Now deduce that the m.m.s.e. is given by $E|\tilde{\mathbf{y}}_{i+\lambda|i}|^2 = 1 - a^{2\lambda}$. [*Hint.* Find the coefficient of z^0 in $S_{\tilde{y}}(z)$.]

7.17 (An interpolation problem) Consider the additive noise model $\mathbf{y}_i = \mathbf{s}_i + \mathbf{v}_i$ for $-\infty < i < \infty$, where $\mathbf{s}_i$ and $\mathbf{v}_i$ are uncorrelated zero-mean scalar stationary random processes with z-spectra $S_s(z)$ and $S_v(z)$, respectively. Suppose we want to estimate $\mathbf{s}_0$ given all $\{\mathbf{y}_i\}$ *except* $\mathbf{y}_0$, *i.e.*, to find

$$\hat{\mathbf{s}}_{0|i\neq 0} = \sum_{i\neq 0} k_i \mathbf{y}_i, \qquad (k_0 = 0).$$

(a) Use the orthogonality principle to show that

$$R_s(i) = \sum_{m\neq 0} k_m R_y(i-m) \qquad \text{for all } i \neq 0.$$

(b) Use the Wiener-Hopf idea to show that $S_s(z) - K(z)S_y(z) = g_0$, a constant, where $K(z)$ is the bilateral z-transform of $\{k_i\}$.

(c) Show that g_0 is determined by the equation

$$\left\{\frac{S_s(z) - g_0}{S_y(z)}\right\}_0 = 0,$$

where the notation $\{f(z)\}_0$ means the coefficient of z^0 in the expansion $f(z) = \sum_{i=-\infty}^{\infty} f_i z^{-i}$.

(d) If $S_v(z) = 1$, show that the m.m.s.e. is given by g_0.

(e) Find $K(z)$ when $S_v(z) = 1$ and

$$S_s(z) = \frac{1 - \frac{|\alpha|^2}{2}}{(1 - \alpha^* z)(1 - \alpha z^{-1})}, \qquad |\alpha| < 1.$$

Show that the m.m.s.e. is $1 - \frac{|\alpha|^2}{2}$.

7.18 (Smoothing using even-indexed data) Consider scalar zero-mean random variables $\{\mathbf{x}(i), \mathbf{u}(i), \mathbf{v}(i), \mathbf{y}(i)\}$ that are related via the state-space equations

$$\mathbf{x}(i+1) = 0.5\mathbf{x}(i) + \mathbf{u}(i), \qquad \mathbf{y}(i) = \mathbf{x}(i) + \mathbf{v}(i), \quad i > -\infty,$$

where $\mathbf{u}(\cdot)$ and $\mathbf{v}(\cdot)$ are uncorrelated unit-variance white-noise processes. It is assumed that the system is operating in stationary mode and that $\mathbf{u}(i)$ and $\mathbf{v}(i)$ are further uncorrelated with $\mathbf{x}_i$. Find the l.l.m.s. filter for estimating $\mathbf{x}(2N_o + 1)$ given $\{\mathbf{y}(2i), -\infty < i < \infty\}$ for a fixed value N_o. What is the z-spectrum of the estimator, $S_{\hat{x}}(z)$?

7.19 (A general estimation problem) Fig. 7.3 shows two zero-mean unit-variance independent white-noise processes $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$, and two known LTI systems, $H(z)$ and $L(z)$. We would like to design a filter $K(z)$ in order to obtain optimal linear l.m.s. estimators for $\mathbf{s}_i$.

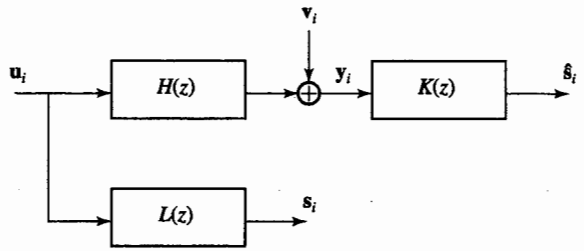


Figure 7.3 A general estimation problem.

(a) Assume $\hat{\mathbf{s}}_i$ in the figure denotes the l.l.m.s. estimator of $\mathbf{s}_i$ given $\{\mathbf{y}_m, -\infty < m < \infty\}$. Show that the optimum Wiener smoother is given by

$$K(z) = \frac{L(z)H^*(z^{-*})}{1 + H(z)H^*(z^{-*})}.$$

(b) Assume now $\hat{\mathbf{s}}_i$ in the figure denotes the l.l.m.s. estimator of $\mathbf{s}_i$ given $\{\mathbf{y}_m, -\infty < m \le i\}$. Find the optimum *causal* filter $K(z)$.

(c) Show that, in either case, the transfer matrix from the disturbances $\mathbf{u}(z)$ and $\mathbf{v}(z)$ to the estimation error $\tilde{\mathbf{s}}_i = \mathbf{s}_i - \hat{\mathbf{s}}_i$ is given by

$$T(z) = L(z) - K(z)H(z) \quad -K(z).$$

(d) The 2-norm of the transfer matrix $T(z)$ is defined as

$$\|T(z)\|_2 \triangleq \left(\frac{1}{2\pi}\text{trace}\int_{-\pi}^{\pi} T^*(e^{j\omega})T(e^{j\omega})d\omega\right)^{1/2}.$$

Show that

$$\|T(z)\|_2^2 = \frac{1}{2\pi}\int_0^{2\pi}\left[\left|K(e^{j\omega})L(e^{j\omega}) - \frac{L(e^{j\omega})H^*(e^{j\omega})}{L^*(e^{j\omega})}\right|^2 + \frac{L(e^{j\omega})L^*(e^{j\omega})}{1 + H(e^{j\omega})H^*(e^{j\omega})}\right]d\omega.$$

(e) Deduce that the optimum Wiener smoother minimizes $\|T(z)\|_2^2$ over all $K(z)$.

(f) Deduce that the optimum Wiener filter minimizes $\|T(z)\|_2^2$ over all causal $K(z)$.
Hint. Note that for any function $a(z)$, we have

$$|a(e^{j\omega})|^2 = \left|\{a(e^{j\omega})\}_+\right|^2 + \left|\{a(e^{j\omega})\}_-\right|^2.$$

7.20 (Decision feedback equalizers) Consider the general model shown in Fig. 7.4 for a decision feedback equalizer. The sequence $\{\mathbf{b}_i\}$ represents the transmitted information and it consists of independent and identically distributed data with values 1 or -1, each with probability 1/2. The transfer function $H(z)$ represents a stable causal LTI communications channel, while $K_1(z)$ and $K_2(z)$ represent the transfer functions of causal stable filters that we wish to design in order to estimate $\mathbf{b}_{i-d}$, for some given delay $d \ge 0$. The sequence

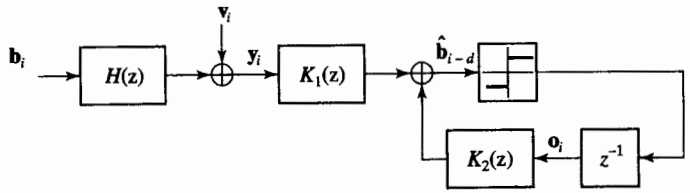


Figure 7.4 A decision feedback equalizer.

$\{\mathbf{v}_i\}$ is white noise with variance σ_v^2. The $\{\mathbf{y}_i\}$ represent the received noisy sequence at the input of the feedforward filter $K_1(z)$, while the $\{\mathbf{o}_i\}$ represent the sequence at the input of the feedback filter $K_2(z)$. The estimator $\hat{\mathbf{b}}_{i-d}$ is the sum of the outputs of $K_1(z)$ and $K_2(z)$. The nonlinear decision device has output 1 if a realization of $\hat{\mathbf{b}}_{i-d}$ is nonnegative and 0 otherwise. In order to simplify the design of the filters $\{K_1(z), K_2(z)\}$, it is assumed that past decisions are always correct, *i.e.*, $\mathbf{o}_i = \mathbf{b}_{i-d-1}$.

(a) Show that the configuration in Fig. 7.4 is equivalent to the one in Fig. 7.5 where $K(z)$ is a stable causal filter that is designed to estimate $\mathbf{s}_i$. That is, determine $\{K(z), H_e(z), L(z), \mathbf{u}_i, \mathbf{w}_i, \mathbf{s}_i\}$ in terms of $\{H(z), K_1(z), K_2(z), \mathbf{y}_i, \mathbf{b}_i, \mathbf{v}_i, d\}$.

(b) Show that the z-spectrum of the signal $\mathbf{u}_i$ can be factored as $S_u(z) = P(z)P^*(z^{-*})$, where

$$P(z) = \begin{bmatrix} \sigma_v & H(z)z^{d+1} \\ 0 & 1 \end{bmatrix}.$$

(c) Let $S_u(z) = M(z)M^*(z^{-*})$ denote a factorization for $S_u(z)$ with $M(z)$ causal and causally invertible (*i.e.*, $M(z)$ is a minimum phase system). Argue that we can express $M(z)$ in terms of $P(z)$ above via $M(z) = P(z)\Theta(z)$, where $\Theta(z)$ satisfies $\Theta(z)\Theta^*(z^{-*}) = I$. Partition the 2×2 matrix $\Theta(z)$ into

$$\Theta(z) = \begin{bmatrix} \Theta_{11}(z) & \Theta_{12}(z) \\ \Theta_{21}(z) & \Theta_{22}(z) \end{bmatrix}.$$

Show that the transfer functions $\{\Theta_{21}(z), \Theta_{22}(z)\}$ must correspond to stable causal systems while the transfer functions $\{\Theta_{11}(z), \Theta_{12}(z)\}$ must correspond to stable anticausal systems (refer to the discussion in Sec. 6.2.3 for a definition of causal and anticausal systems). Show further that each of the transfer functions $\{\Theta_{ij}(z)\}$ has at most order $(d+1)$.

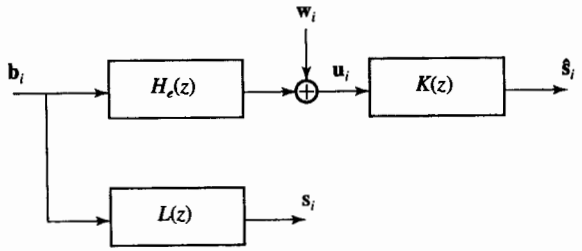


Figure 7.5 An equivalent configuration when past decisions are assumed correct.

(d) If $K(z)$ is allowed to be noncausal, what would be the resulting smoothing error?

(e) Find the optimal linear causal filter $K(z)$ by using the Wiener-Hopf theory. Let $S_e(z)$ denote the z-spectrum of the resulting error sequence $\{\mathbf{s}_i - \hat{\mathbf{s}}_i\}$. Show that $S_e(z)$ is a constant. More specifically, show that

$$S_e(z) = |\Theta_{21}(\infty)|^2 + |\Theta_{22}(\infty)|^2.$$

7.21 **(A multichannel smoothing problem)** Fig. 7.6 shows a zero-mean scalar stationary process $\{\mathbf{x}_i\}$ with variance function $R_x(i) = \sigma_x^2\delta_i$ and an LTI system $H(z)$. The process $\{\mathbf{v}_i\}$ is white noise (independent of $\{\mathbf{x}_i\}$) with variance function $R_v(i) = \sigma_v^2\delta_i$.

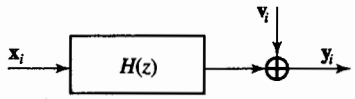


Figure 7.6 Single-channel transmission.

(a) Show that the l.l.m.s. estimator of $\mathbf{x}_i$ given $\{\mathbf{y}_k\}_{k=-\infty}^{\infty}$, denoted by $\hat{\mathbf{x}}_i^s$, can be found as shown in Fig. 7.7. It is useful to note that $H^*(z^{-*})$ is often, especially in communications applications, referred to as the *matched filter*.

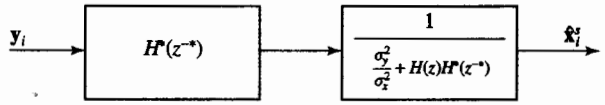


Figure 7.7 Optimal Wiener smoother.

(b) Now suppose, as shown in Fig. 7.8, that $\mathbf{x}_i$ passes through m transfer functions $H_1(z), \ldots, H_m(z)$ and that the outputs are corrupted by independent white-noise sequences $\{\mathbf{v}_i^1\}, \ldots, \{\mathbf{v}_i^m\}$ with $\langle \mathbf{v}_i^k, \mathbf{v}_j^l \rangle = \sigma^2\delta_{ij}\delta_{kl}$. It is claimed that $\hat{\mathbf{x}}_i^s$, the l.l.m.s. estimator of $\mathbf{x}_i$ given $\left\{\{\mathbf{y}_k^1\}_{k=-\infty}^{\infty}, \ldots, \{\mathbf{y}_k^m\}_{k=-\infty}^{\infty}\right\}$, can be found by using the structure of Fig. 7.9, for some appropriate transfer function $G(z)$. Show that this is true and find $G(z)$.

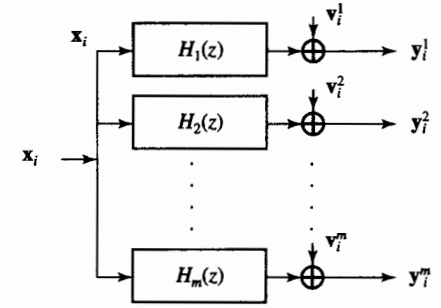


Figure 7.8 A multichannel transmission.

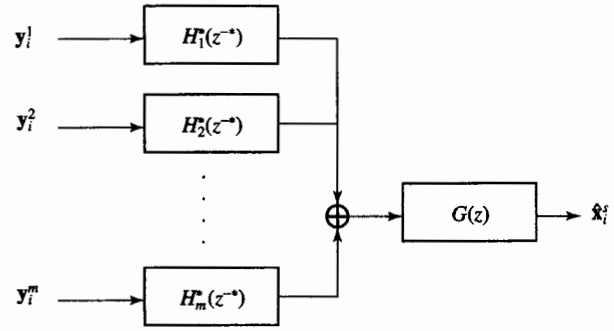


Figure 7.9 Multichannel optimal smoothing.

7.22 **(Continuous-time prediction)** Use the results of App. 7.A to find the transfer function of the l.l.m.s. predictor of $\mathbf{x}(t+\lambda)$ $(\lambda > 0)$ for a stationary process $\mathbf{x}(\cdot)$ with s-spectrum $S_x(s) = (s^2 - \beta^2)/(s^2 - \alpha^2)$, where $\alpha > 0$ and $\beta > 0$.

7.23 **(Continuous-time filtering and smoothing)** Let $\mathbf{y}(t) = \sqrt{P}\mathbf{x}(t) + \mathbf{v}(t)$, where $\mathbf{x}(\cdot)$ and $\mathbf{v}(\cdot)$ are uncorrelated zero-mean stationary processes with s-spectra $S_x(s) = 2\alpha/(\alpha^2 - s^2)$, $S_v(s) = R$, respectively, and $\alpha > 0$, $P > 0$.

(a) Compute $S_y(s)$ and $S_{xy}(s)$ and express the results in terms of $\Delta \triangleq 2P/\alpha R$.

(b) Using the results of App. 7.A, determine the transfer function of the causal filter $K(s)$ for the following three cases:

1. $\lambda = 0$. This corresponds to filtering with zero delay. What happens as $\Delta \to 0$ and as $\Delta \to \infty$? Interpret your results.
2. λ negative. This corresponds to filtering with delay. What happens as $|\lambda|$ increases?
3. λ positive. This corresponds to prediction. How does the filters with zero delay and with prediction relate? Interpret your results.

7.24 **(Noise cancellation via Wiener filtering)** Two sets of noisy measurements of a deterministic signal $m(\cdot)$ are available, say

$$\mathbf{y}_1(t) = m(t) + \mathbf{v}_1(t), \quad \mathbf{y}_2(t) = m(t) + \mathbf{v}_2(t), \quad t > -\infty,$$

where $\{\mathbf{v}_1(\cdot), \mathbf{v}_2(\cdot)\}$ are zero-mean uncorrelated stationary noise processes with s-spectra

$$S_{v_1}(s) = \frac{1}{3 - s^2}, \quad S_{v_2}(s) = \frac{-s^2}{4 - s^2}.$$

It is suggested that the 2-input 1-output structure of Fig. 7.10 be used to estimate the signal $m(\cdot)$. The output of the structure is denoted by $m(\cdot) + \mathbf{w}(\cdot)$, where $\mathbf{w}(\cdot)$ denotes the perturbation relative to the desired signal $m(\cdot)$. The block in the lower branch indicates a causal linear filter with transfer function $K(s)$. Determine the $K(s)$ that yields a perturbation $\mathbf{w}(\cdot)$ with the smallest variance.

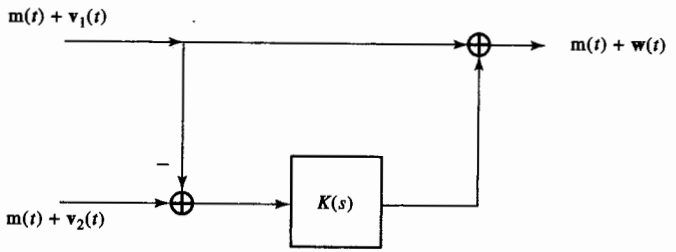


Figure 7.10 A structure for noise cancellation from two signal measurements.

7.25 **(Pre-emphasis filter)** A stationary process $\mathbf{x}(\cdot)$ passes through a LTI communications channel with transfer function $P(s)$, before being corrupted by additive white noise $\mathbf{v}(\cdot)$ with intensity $N_o/2$ (*i.e.*, $S_v(s) = N_o/2$). Assume $\mathbf{x}(\cdot)$ and $\mathbf{v}(\cdot)$ are uncorrelated and define $\mathbf{y}(t) = \mathbf{x}(t) \star p(t) + \mathbf{v}(t)$. Here, $\star$ denotes linear convolution and $\mathbf{x}(t) \star p(t)$ denotes the output of the communications channel (whose impulse response we denote by $p(t)$).

(a) Find the l.l.m.s. smoothing filter for estimating $\mathbf{x}(\cdot)$ given $\{\mathbf{y}(\cdot), -\infty < t < \infty\}$.

(b) Assume that

$$P(s) = \left(\frac{\alpha}{\beta}\right)^{1/2} \frac{s + \beta}{s + \alpha}, \quad S_s(s) = \frac{2\beta}{\beta^2 - s^2}, \quad \alpha > 0, \quad \beta > 0.$$

Find the l.l.m.s. causal filter for estimating $\mathbf{x}(\cdot)$ given all past $\mathbf{y}(\cdot)$.

(c) Find an expression for the m.m.s. error and use it to find the value of α that minimizes this error.

(d) How will the results change if

$$P(s) = \left(\frac{\alpha}{\beta}\right)^{1/2} \frac{s - \beta}{s + \alpha}.$$

Appendix for Chapter 7

7.A THE CONTINUOUS-TIME WIENER-HOPF TECHNIQUE

In this appendix we briefly present the Wiener-Hopf technique for solving the equation

$$R_{sy}(t+\lambda) = \int_0^\infty k(\tau) R_y(t-\tau)d\tau, \quad t \geq 0, \qquad \lambda > 0, \tag{7.A.1}$$

where $k(\tau) = 0$ for $\tau < 0$. We assume that

$$S_y(f) \triangleq \mathcal{F}\{R_y(\tau)\} > 0, \quad -\infty < f < \infty,$$

and that the bilateral Laplace transform of $R_y(\cdot)$,[13]

$$S_y(s) \triangleq \mathcal{L}\{R_y(\tau)\},$$

which is referred to as the s-spectrum, exists in a strip containing the imaginary axis. We also assume that $S_y(s)$ is a proper rational function that is strictly positive on the unit circle. Then $S_y(s)$ admits the following canonical factorization (*cf.* App. 6.A):

$$S_y(s) = L(s)RL^*(-s^*), \tag{7.A.2}$$

for some $R > 0$ and where $L(s)$ and $L^{-1}(s)$ are analytic in the closed right-half plane. We can now outline the Wiener-Hopf technique. We first extend Eq. (7.A.1) to the whole line by introducing the function

$$g(t) = R_{sy}(t+\lambda) - \int_0^\infty k(\tau)R_y(t-\tau)d\tau, \quad -\infty < t < \infty. \tag{7.A.3}$$

By the fact that Eq. (7.A.1) holds for $t \geq 0$, we know that $g(\cdot)$ is one-sided, but *anticausal*,

$$g(t) = 0, \quad t \geq 0,$$

unlike the unknown causal $k(\cdot)$. Of course now we have two unknown functions $g(\cdot)$ and $k(\cdot)$, but we shall nonetheless persist. Taking bilateral Laplace transforms yields

$$G(s) = S_{sy}(s)e^{s\lambda} - K(s)S_y(s). \tag{7.A.4}$$

Now using the canonical factorization (7.A.2) of $S_y(s)$ we may write

$$\frac{G(s)}{RL^*(-s^*)} = \frac{S_{sy}(s)e^{s\lambda}}{RL^*(-s^*)} - K(s)L(s). \tag{7.A.5}$$

[13] The letter s in normal font used in $L(s)$ denotes the complex argument, $s = \sigma + j\omega$, of the Laplace transform $L(s)$. It is clear from the context that it is distinct from the boldface letter $\mathbf{s}$ that we use to denote the random variable (or process) that we wish to estimate.

By construction, the time function obtained by the inverse Laplace transform of the term $G(s)/RL^*(-s^*)$ will be zero for $t \geq 0$ (strictly anticausal), while the time function corresponding to $K(s)L(s)$ will be zero for $t < 0$ (causal). We may also represent this fact as

$$\underbrace{\frac{G(s)}{RL^*(-s^*)}}_{\text{strictly anticausal IT}} = \frac{S_{sy}(s)e^{s\lambda}}{RL^*(-s^*)} - \underbrace{K(s)L(s)}_{\text{causal IT}}.$$

This means that the latter function must be equal to the $t \geq 0$ portion of the inverse Laplace transform of $S_{sy}(s)/RL^*(-s^*)$, leading to the famous formula for the solution to the Wiener-Hopf equation (7.A.1),

$$K(s) = \frac{1}{L(s)}\int_0^\infty \left[\frac{1}{j2\pi}\int_{Br}\frac{S_{sy}(p)e^{p\lambda}}{RL^*(-p^*)}e^{pt}dp\right]e^{-st}dt, \tag{7.A.6}$$

where the contour of integration (denoted by Br) should lie in the region of convergence of $S_{sy}(s)/RL^*(-s^*)$. We can also write this expression in the form

$$K(s) = \left\{\frac{S_{sy}(s)e^{s\lambda}}{RL^*(-s^*)}\right\}_+ \frac{1}{L(s)}, \tag{7.A.7}$$

where, as in the discrete-time case studied in the body of the chapter, the notation $\{\cdot\}_+$ denotes an operator that extracts the causal part of the function to which it is applied. More specifically, if $f(t)$ is a time function with bilateral Laplace transform $F(s) = \mathcal{L}[f(t)]$, and if $1(t)$ denotes the unit step (Heaviside) function that is zero for $t < 0$ and unity otherwise, then $\{F(s)\}_+ = \mathcal{L}[f(t)1(t)]$.

Of course, the innovations approach (*cf.* Sec. 7.7) can also be used to obtain this solution, as was first done by Bode and Shannon (1950) and by Zadeh and Ragazzini (1950). We leave the solution by this method as an exercise for active readers.

Example 7.A.1 (Signal in Additive Noise) Consider a stationary random process $\mathbf{y}(t) = \mathbf{s}(t) + \mathbf{v}(t)$, where $\mathbf{s}(\cdot)$ and $\mathbf{v}(\cdot)$ are zero-mean and uncorrelated, $\mathbf{v}(\cdot)$ is white with unit variance, and the power spectral density function of $\mathbf{s}(\cdot)$ is

$$S_s(f) = \frac{2\alpha}{\alpha^2 + 4\pi^2 f^2} = \mathcal{F}\{e^{-\alpha|t|}\}.$$

The $s-$spectrum of $\mathbf{s}(\cdot)$ is $S_s(s) = \frac{2\alpha}{\alpha^2 - s^2}$ and, consequently,

$$S_{sy}(s) = S_s(s) = \frac{2\alpha}{\alpha^2 - s^2}, \qquad S_y(s) = S_s(s) + 1 = \frac{s^2 - \alpha^2 - 2\alpha}{s^2 - \alpha^2} = L(s)RL^*(-s^*),$$

with

$$L(s) = \frac{s + \sqrt{\alpha^2 + 2\alpha}}{s + \alpha}, \qquad R = 1.$$

Using (7.A.7), the Wiener filter for estimating $\mathbf{s}(\cdot)$ from $\mathbf{y}(\cdot)$ is given by

$$K(s) = \frac{s+\alpha}{s+\sqrt{\alpha^2+2\alpha}} \left\{ \frac{s-\alpha}{s-\sqrt{\alpha^2+2\alpha}} \cdot \frac{2\alpha}{\alpha^2-s^2} \right\}_+$$

$$= \frac{s+\alpha}{s+\sqrt{\alpha^2+2\alpha}} \left\{ \frac{-2\alpha}{(s-\sqrt{\alpha^2+2\alpha})(s+\alpha)} \right\}_+$$

$$= \frac{s+\alpha}{s+\sqrt{\alpha^2+2\alpha}} \left\{ \frac{\frac{-2\alpha}{\alpha+\sqrt{\alpha^2+2\alpha}}}{s-\sqrt{\alpha^2+2\alpha}} + \frac{\frac{2\alpha}{\alpha+\sqrt{\alpha^2+2\alpha}}}{s+\alpha} \right\}_+$$

$$= \frac{s+\alpha}{s+\sqrt{\alpha^2+2\alpha}} \frac{\frac{2\alpha}{\alpha+\sqrt{\alpha^2+2\alpha}}}{s+\alpha} = \frac{\frac{2\alpha}{\alpha+\sqrt{\alpha^2+2\alpha}}}{s+\sqrt{\alpha^2+2\alpha}} = \frac{\sqrt{\alpha^2+2\alpha}-\alpha}{s+\sqrt{\alpha^2+2\alpha}}.$$

◆

EXAMPLE 7.A.2 (More General Setting) Let $\mathbf{y}(t) = \mathbf{s}(t) + \mathbf{v}(t)$, where

$$E\mathbf{s}(t)\mathbf{v}^*(\tau) = 0, \quad E\mathbf{v}(t)\mathbf{v}^*(\tau) = R\delta(t-\tau).$$

Let us show that $K(s) = 1 - L^{-1}(s)$. Indeed, we have

$$S_y(s) = S_s(s) + R \triangleq L(s)RL^*(-s^*),$$

and $S_{sy}(s) = S_s(s) = S_y(s) - R$. Then

$$K(s) = \frac{1}{L(s)} \left\{ \frac{S_y(s)-R}{RL^*(-s^*)} \right\}_+ = \frac{1}{L(s)} \{L(s)\}_+ - \frac{1}{L(s)} \left\{ \frac{1}{L^*(-s^*)} \right\}_+ = 1 - \frac{1}{L(s)} \cdot 1 = 1 - L^{-1}(s).$$

The reader can check that this formula simplifies the solution of the previous example. ◆

CHAPTER 8

Recursive Wiener Filtering

In Ch. 7 we showed that canonical spectral factorization was the key to the solution of the estimation problem for stationary random processes over infinite and semi-infinite intervals. For scalar processes, and especially when the power spectral density function is rational, determining the canonical factorization is relatively straightforward. However, with vector stationary processes, even when the resulting matrix-valued spectral densities are rational, computing the canonical factorization is considerably more difficult and was, for a long time (see the discussion at the end of Sec. 6.6), a major stumbling block to the useful application of the theory.

The introduction of state-space structure into the problem by R. E. Kalman in 1960 helped to cross this barrier. In this chapter, we shall exploit this critical insight in a different way than usual in the literature. For time-invariant state-space models driven by stationary random processes, we first introduce an equivalence class of all the input covariances that give rise to the same power spectral density. The flexibility so obtained will suggest a method for computing the spectral factorization by finding a particular so-called stabilizing solution of an algebraic Riccati equation (ARE). Moreover, the factorization will be found in state-space form, immediately leading to a recursive solution to the estimation problem, which will turn out to be a forerunner of the Kalman filter for estimating nonstationary processes (see Ch. 9). The usual approach reverses this route, and goes from the finite-time estimation results to their steady-state versions (*cf.* Sec. 1.5.2 and Ch. 14).

8.1 TIME-INVARIANT STATE-SPACE MODELS

A process $\{\mathbf{y}_i\}$ is said to have a *time-invariant* (or *constant parameter*) state-space model if we can write

$$\begin{cases} \mathbf{x}_{i+1} = F\mathbf{x}_i + G\mathbf{u}_i, & i \geq 0, \\ \mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{8.1.1}$$

where $F \in \mathbb{C}^{n\times n}$, $G \in \mathbb{C}^{n\times m}$, and $H \in \mathbb{C}^{p\times n}$ are known time-independent (or constant) matrices, and the $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are zero-mean jointly stationary vector random variables that, along with the zero-mean random variable $\mathbf{x}_0$, satisfy the conditions

$$\left\langle \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_j \\ \mathbf{v}_j \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \delta_{ij} & 0 \end{bmatrix}. \tag{8.1.2}$$

As mentioned earlier in Sec. 5.3.4, we shall refer to the above as the *standard* time-invariant state-space model.

8.1.1 Covariance Functions for Time-Invariant Models

The first result is simply a restatement of Lemma 5.3.2 when the state-space model is taken to be time-invariant.

Lemma 8.1.1 (Covariance Functions) *Consider the time-invariant state-space model (8.1.1)–(8.1.2) and denote the state covariance matrix by* $\langle \mathbf{x}_i, \mathbf{x}_i \rangle = \|\mathbf{x}_i\|^2 = \Pi_i$. *Then* Π_i *satisfies*

$$\Pi_{i+1} = F\Pi_i F^* + GQG^*, \quad i \geq 0. \tag{8.1.3}$$

The covariances of the state variables can be written as

$$R_x(i,j) \triangleq \langle \mathbf{x}_i, \mathbf{x}_j \rangle = \begin{cases} F^{i-j}\Pi_j & i \geq j, \\ \Pi_i F^{(j-i)*} & i \leq j, \end{cases} \tag{8.1.4}$$

and the covariances of the output process $\{\mathbf{y}_i\}$ *as*

$$R_y(i,j) \triangleq \langle \mathbf{y}_i, \mathbf{y}_j \rangle = \begin{cases} HF^{i-j-1}N_j & i > j, \\ R + H\Pi_i H^* & i = j, \\ N_i^* F^{(j-i-1)*} H^* & i < j, \end{cases} \tag{8.1.5}$$

where $N_i \triangleq F\Pi_i H^* + GS$. ∎

Proof: By specializing the results of Lemma 5.3.2, or better, by straightforward direct calculations. ◆

8.1.2 The Special Case of Stationary Processes

An important consequence of Lemma 8.1.1 is that although the underlying state-space model is time-invariant, and although the input disturbances $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are stationary, when we start at some finite time, $i = 0$ say, neither the state process $\{\mathbf{x}_i, i \geq 0\}$ nor the output process $\{\mathbf{y}_i, i \geq 0\}$ is in general stationary. The reason is that the state variance, Π_i, is generally time-dependent.

However, the processes $\{\mathbf{x}_i, i \geq 0\}$ and $\{\mathbf{y}_i, i \geq 0\}$ can be stationary in special circumstances. To this end, suppose that F is a *stable* matrix, *i.e.*, all its eigenvalues are strictly inside the unit circle, and let $\bar{\Pi}$ be a constant matrix obeying the equation

$$\bar{\Pi} = F\bar{\Pi}F^* + GQG^*. \tag{8.1.6}$$

This is the famous discrete-time Lyapunov (sometimes known as Stein) equation, for which a celebrated theorem asserts that when F is stable, there is a unique (Hermitian) solution; moreover, when $Q \geq 0$ this solution is positive-semi-definite (see App. D). Subtracting (8.1.6) from (8.1.3) we obtain

$$\Pi_{i+1} - \bar{\Pi} = F(\Pi_i - \bar{\Pi})F^*,$$

so that after repeated applications of this equality, it follows that $\Pi_{i+1} - \bar{\Pi} = F^i(\Pi_0 - \bar{\Pi})F^{i*}$. When F is stable, we have $F^i \to 0$ as $i \to \infty$, so that for any initial value $\Pi_0 \geq 0$, $\Pi_i \to \bar{\Pi}$ as $i \to \infty$, *i.e.*, the processes $\{\mathbf{x}_i, \mathbf{y}_i, i \geq 0\}$ will be *asymptotically* stationary.

Now assuming that F is stable, *if* we choose the initial covariance matrix Π_0 as the unique solution of (8.1.6), *i.e.*, $\Pi_0 = \bar{\Pi}$, then it follows from (8.1.3) that

$$\Pi_i = \bar{\Pi} \quad \text{for} \quad i \geq 0, \tag{8.1.7}$$

and, consequently, that N_i is time-invariant, $N_i = F\bar{\Pi}H^* + GS \triangleq \bar{N}$, say. Therefore, the process $\{\mathbf{y}_i, i \geq 0\}$ will be stationary, *i.e.*, $R_y(i,j)$ will be a function only of $|i-j|$:

$$R_y(i,j) = R_y(i-j) = \langle \mathbf{y}_i, \mathbf{y}_j \rangle = \begin{cases} HF^{i-j-1}\bar{N} & i > j, \\ R + H\bar{\Pi}H^* & i = j, \\ \bar{N}^*(F^*)^{j-i-1}H^* & i < j. \end{cases} \tag{8.1.8}$$

So a stationary covariance function is completely determined by the triple $\{H, F, \bar{N}\}$. We summarize this discussion in the following Lemma.

Lemma 8.1.2 (Covariance Functions in the Stationary Case) *Consider the time-invariant state-space model (8.1.1)–(8.1.2) and suppose that F is stable and that* $\Pi_0 = \bar{\Pi}$, *where* $\bar{\Pi}$ *is the unique solution of the Lyapunov equation (8.1.6). Then the processes* $\{\mathbf{x}_i, \mathbf{y}_i\}$ *are both stationary with covariance sequences*

$$R_x(i-j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle = \begin{cases} F^{i-j}\bar{\Pi} & i \geq j, \\ \bar{\Pi}F^{*(i-j)} & i \leq j, \end{cases} \tag{8.1.9}$$

and $R_y(i-j)$ *as in (8.1.8) where* $\bar{N} \triangleq F\bar{\Pi}H^* + GS$. ∎

8.1.3 Expressions for the z-Spectrum

The above formulas show that when the process $\{\mathbf{y}_i\}$ is stationary, the covariance function is exponentially decaying. Therefore, as in Sec. 6.6, we can define its *z-spectrum*,

$$S_y(z) = \sum_{k=-\infty}^{\infty} R_y(k) z^{-k}, \tag{8.1.10}$$

which will be well defined in an annulus in the complex z-plane that includes the unit circle, $z = e^{j\omega}$.

The state-space model (8.1.1)–(8.1.2) allows us to give two useful explicit formulas for $S_y(z)$. One expression follows by using (8.1.8) and (8.1.10) to get

$$S_y(z) = R_y(0) + \sum_{k=1}^{\infty} HF^{k-1}\bar{N}z^{-k} + \sum_{-\infty}^{k=-1} \bar{N}^*(F^*)^{-k-1}H^*z^{-k}, \tag{8.1.11}$$

$$= (R + H\bar{\Pi}H^*) + H(zI - F)^{-1}\bar{N} + \bar{N}^*(z^{-1}I - F^*)^{-1}H^*, \tag{8.1.12}$$

which can be rearranged in matrix form as

$$S_y(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} 0 & \bar{N} \\ \bar{N}^* & R + H\bar{\Pi}H^* \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}. \tag{8.1.13}$$

The ROC of $S_y(z)$ is seen from (8.1.12) to be — see Prob. 8.6,

$$|\lambda_{\max}(F)| < |z| < \frac{1}{|\lambda_{\max}(F)|}, \tag{8.1.14}$$

which is nonempty since F is stable and thus $|\lambda_{\max}(F)| < 1$.

An alternative expression for $S_y(z)$ can be obtained by using the well-known formulas for expressing the z-spectrum in terms of the transfer function of the linear filter relating the (input) white-noise processes $\{\mathbf{u}_i, \mathbf{v}_i\}$ to the output process $\{\mathbf{y}_i\}$ (see Sec. 6.3.2). Thus note that taking z-transforms of the state equations (8.1.1), rewritten as

$$\begin{cases} \mathbf{x}_{i+1} = F\mathbf{x}_i + \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} G\mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \\ \mathbf{y}_i = H\mathbf{x}_i + \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} G\mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \end{cases} \tag{8.1.15}$$

we get

$$\mathbf{y}(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} G\mathbf{u}(z) \\ \mathbf{v}(z) \end{bmatrix}. \tag{8.1.16}$$

In Sec. 6.3.2, we showed that if an $m \times 1$ stationary process $\{\mathbf{r}_i\}$ with z-spectrum $S_r(z)$ is applied to a $p \times m$ stable linear system with transfer matrix $H(z)$ to yield an output $\{\mathbf{o}_i\}$, the z-spectrum of the output is given by

$$S_o(z) = H(z)S_r(z)H^*(z^{-*}).$$

Therefore (8.1.16) shows that the z-spectrum of $\{\mathbf{y}_i\}$ can be written as

$$S_y(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}. \tag{8.1.17}$$

Comparing (8.1.13) with (8.1.17) we see that the only difference between these two *representations* of the output z-spectrum is in the matrix appearing in the center of these equations. In the case of (8.1.17), this matrix is the variance matrix of $\{G\mathbf{u}_i, \mathbf{v}_i\}$,

$$\begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix} \geq 0. \tag{8.1.18}$$

This fact immediately shows that when $z = e^{j\omega}$, $S_y(e^{j\omega})$ is nonnegative-definite, as a power spectral density function should be. However note that in the alternative representation (8.1.13), the center matrix

$$\begin{bmatrix} 0 & \bar{N} \\ \bar{N}^* & R + H\bar{\Pi}H^* \end{bmatrix} \quad \text{is indefinite.} \tag{8.1.19}$$

But of course since $R_y(k)$ is a covariance sequence, it still must be true that $S_y(e^{j\omega}) \geq 0$, even though the center matrix in (8.1.19) is not nonnegative-definite and therefore cannot be thought of as the variance matrix of some random variables, say $\{\mathbf{u}_i^{(1)}, \mathbf{v}_i^{(1)}\}$; were this to be so, $\mathbf{u}_i^{(1)}$ would need to have zero variance but nonzero cross-variance with $\mathbf{v}_i^{(1)}$!

This seems puzzling. However, if we broaden our domain of discourse, and instead of random variables, consider vectors that belong to an abstract *indefinite* (so-called Krein) space, then the matrix (8.1.19) can be considered as the "covariance" of such an abstract process $\{\mathbf{u}_i^{(1)}, \mathbf{v}_i^{(1)}\}$. The question is whether anything useful can be gained from such a generalization. Indeed there is, as we shall show in App. 8.A. However, in order to move more quickly towards the final goal, for the moment we shall proceed in a less motivated and purely algebraic way.

8.2 AN EQUIVALENCE CLASS FOR INPUT GRAMIANS

The fact that two different central matrices as in (8.1.18)–(8.1.19) can lead to the same z-spectrum leads us to ask whether there are other such matrices. In fact, there is an infinity of them (for convenience of designation, we shall call them *input Gramians*, as described next).

[Before proceeding we make a *temporary* assumption in order to reduce the notational burden: this is just that $G = I$. There is clearly no loss of generality here; just replace Q by GQG^* and S by GS, if desired.]

Lemma 8.2.1 (Equivalence Class for Input Gramians) *Consider the model (8.1.1)–(8.1.2) with $G = I$ and stable F. The following facts hold.*

(a) *For any Hermitian matrix Z, the output z-spectrum of the process $\{\mathbf{y}_i\}$,*

$$S_y(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}, \tag{8.2.1}$$

is invariant under the input Gramian transformation[1]

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \to \begin{bmatrix} Q - Z + FZF^* & S + FZH^* \\ S^* + HZF^* & R + HZH^* \end{bmatrix}. \tag{8.2.2}$$

(b) *If for an observable system $\{F, H\}$,*

$$\begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q_1 & S_1 \\ S_1^* & R_1 \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix} =$$

$$\begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q_2 & S_2 \\ S_2^* & R_2 \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix},$$

then there exists a unique Hermitian matrix Z such that

$$\begin{bmatrix} Q_1 & S_1 \\ S_1^* & R_1 \end{bmatrix} = \begin{bmatrix} Q_2 - Z + FZF^* & S_2 + FZH^* \\ S_2^* + HZF^* & R_2 + HZH^* \end{bmatrix}, \tag{8.2.3}$$

$$= \begin{bmatrix} Q_2 & S_2 \\ S_2^* & R_2 \end{bmatrix} + \begin{bmatrix} -Z + FZF^* & FZH^* \\ HZF^* & HZH^* \end{bmatrix}.$$

■

(Algebraic) Proof: Part (a) follows directly by showing via a calculation that (8.2.2) is true for any Hermitian matrix Z — just check that

$$0 = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} -Z + FZF^* & FZH^* \\ HZF^* & HZH^* \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix} \tag{8.2.4}$$

is true for any $Z = Z^*$. For part (b), let us assume that

$$\begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q_1 - Q_2 & S_1 - S_2 \\ S_1^* - S_2^* & R_1 - R_2 \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix} = 0. \tag{8.2.5}$$

[1] See App. 8.A for an explanation of the form of this transformation.

Now introduce Z as the unique Hermitian solution[2] to the Lyapunov equation

$$Z = FZF^* + Q_2 - Q_1. \tag{8.2.6}$$

We can now apply the input Gramian transformation of part (a) with the matrix $-Z$. Thus (8.2.5) becomes

$$\begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Z - FZF^* + Q_1 - Q_2 & S_1 - S_2 - FZH^* \\ S_1^* - S_2^* - HZF^* & R_1 - R_2 - HZH^* \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix} = 0$$

The $(1, 1)$ block entry of the central matrix is zero due to (8.2.6). Hence, this last equation can be expanded as

$$H(zI - F)^{-1}(S_1 - S_2 - FZH^*) + R_1 - R_2 - HZH^* + (S_1^* - S_2^* - HZF^*)(z^{-1}I - F^*)^{-1}H^* = 0.$$

Now note that the assumption of a stable matrix F allows us to expand $H(zI - F)^{-1}$ in the power series

$$H(zI - F)^{-1} = z^{-1}H + z^{-2}HF + z^{-3}HF^2 + \ldots,$$

so that the earlier equality can be written as

$$\sum_{i=1}^{\infty} HF^{i-1}(S_1 - S_2 - FZH^*)z^{-i} + R_1 - R_2 - HZH^* + \sum_{i=1}^{\infty}(S_1^* - S_2^* - HZF^*)F^{(i-1)*}H^*z^i = 0.$$

The above equality shows that all the coefficients of z^i, $-\infty < i < \infty$, must be zero. Thus, for $i = 0$,

$$R_1 - R_2 = HZH^*, \tag{8.2.7}$$

and for $i > 0$, $HF^{i-1}(S_1 - S_2 - FZH^*) = 0$. These equations can be written in matrix form as

$$\begin{bmatrix} H \\ HF \\ HF^2 \\ \vdots \end{bmatrix} (S_1 - S_2 - FZH^*) = 0.$$

Since $\{F, H\}$ is observable, the matrix on the left-hand side has full rank and therefore

$$S_1 - S_2 = FZH^*. \tag{8.2.8}$$

Eqs. (8.2.6), (8.2.7), and (8.2.8) now establish part (b). ◆

The reader should check (see Prob. 8.4) that the choice $Z = \bar{\Pi}$, where $\bar{\Pi}$ is as in (8.1.6), relates the input covariances (8.1.18) and (8.1.19). Moreover, although we have required F to be stable in the statement of Lemma 8.2.1, the algebraic argument (8.2.4) clearly shows that the result of part (a) holds without any assumptions on $\{F, Q, R, S\}$. We can now proceed fairly directly to obtain spectral factorizations.

[2] The stability of F ensures that the Lyapunov equation has a unique Hermitian solution — see App. D.

Remark 1 [The Kalman-Yakubovich-Popov (KYP) and Related Lemmas]. The results of Lemma 8.2.1 do not use the fact that the process $\{\mathbf{y}_i\}$ is a true stochastic process, *i.e.*, that its z-spectrum $S_y(z)$ is nonnegative-definite on the unit circle. When that is true, stronger results are possible, *e.g.*, the so-called Kalman-Yakubovich-Popov (KYP) lemma. Though this is an elegant and important result, we shall not need to use it in this book. For our purposes, the comparatively trivial Lemma 8.2.1 is sufficient. However, the KYP Lemma, and the closely related Positive Real and Bounded Real lemmas, are very useful in many applications. Since much of the background material has already been developed here, we elaborate on these important lemmas in Apps. 8.C and 8.D. ◆

8.3 CANONICAL SPECTRAL FACTORIZATION

Recall from Sec. 6.6 that in order for $S_y(z)$ to admit a canonical spectral factorization as defined in that section, it must have full normal rank everywhere on the unit circle or, equivalently, it must have no unit-circle zeros. In other words, that

$$S_y(e^{j\omega}) > 0 \qquad \text{for all} \quad -\pi \le \omega \le \pi. \tag{8.3.1}$$

8.3.1 Unit-Circle Controllability Condition

A natural question is to determine the requirements on the matrices $\{F, G, H, Q, S, R\}$ such (8.3.1) is true.[3]

Lemma 8.3.1 (Positive-Definite Spectra) *Consider the z-spectrum (8.1.17), where F is stable and*

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \ge 0 \quad \textit{and} \quad R > 0. \tag{8.3.2}$$

Introduce the matrices

$$F^s \triangleq F - GSR^{-1}H \quad \textit{and} \quad Q^s \triangleq Q - SR^{-1}S^*.$$

Then $S_y(e^{j\omega}) > 0$ for all $-\pi \le \omega \le \pi$ if, and only if, the matrix pair $\{F^s, GQ^{s/2}\}$ is unit-circle uncontrollable, or equivalently if, and only if, there exists no unit-circle eigenvalue of F^s, with corresponding left eigenvector x (i.e., $xF^s = \lambda x$, $|\lambda| = 1$), such that $xGQ^{s/2} = 0$. ■

Proof: Note that since $R > 0$ we can perform the "upper-diagonal-lower" factorization (*cf.* App. A)

$$\begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix} = \begin{bmatrix} I & GSR^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} GQ^sG^* & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} I & 0 \\ R^{-1}S^*G^* & I \end{bmatrix},$$

[3] The reader may find it useful to refer to App. C for a review of pertinent results from linear system theory that will be encountered in the next few sections.

of the center matrix in the z-spectrum of (8.1.17). This allows us to rewrite (8.1.17) as

$$S_y(z) = H(zI - F)^{-1}GQ^sG^*(z^{-1} - F^*)^{-1}H^* +$$

$$\left[H(zI - F)^{-1}GSR^{-1} + I\right] R \left[H(z^{-*}I - F)^{-1}GSR^{-1} + I\right]^*.$$

Now since $R > 0$ and $GQ^sG^* \ge 0$, clearly $S_y(z) \ge 0$, for all $|z| = 1$. However, it will drop rank at some point on the unit circle if, and only if, there exists a nonzero vector w and a scalar λ, with $|\lambda| = 1$, such that

$$\begin{cases} wH(\lambda I - F)^{-1}GQ^{s/2} = 0, \\ w\left[H(\lambda I - F)^{-1}GSR^{-1} + I\right] = 0. \end{cases} \tag{8.3.3}$$

We shall now show that (8.3.3) is equivalent to the unit circle noncontrollability of the pair $\{F^s, GQ^{s/2}\}$.

To prove one direction, assume that (8.3.3) holds. Then it turns out that $x = wH(\lambda I - F)^{-1}$ is an uncontrollable left eigenvector of $\{F^s, GQ^{s/2}\}$. Indeed $xGQ^{s/2} = wH(\lambda I - F)^{-1}GQ^{s/2} = 0$, and

$$\begin{aligned} xF^s &= wH(\lambda I - F)^{-1}(F - GSR^{-1}H), \\ &= wH(\lambda I - F)^{-1}F - wH(\lambda I - F)^{-1}GSR^{-1}H, \\ &= wH(\lambda I - F)^{-1}F - (-wH) \qquad \text{using (8.3.3)}, \\ &= wH(\lambda I - F)^{-1}(F + \lambda I - F) = wH(\lambda I - F)^{-1}\lambda = \lambda x. \end{aligned}$$

This shows that x is a left eigenvector of F^s that is orthogonal to $GQ^{s/2}$ so that an uncontrollable unit-circle eigenvalue exists.

To prove the other direction, assume that the pair $\{F^s, GQ^{s/2}\}$ has a unit circle uncontrollable mode at $|\lambda| = 1$, say $xF^s = \lambda x$ and $xGQ^{s/2} = 0$ for some nonzero x. Then, using the definition of F^s, this implies that $x(F - GSR^{-1}H) = \lambda x$, which can be rewritten as

$$x(\lambda I - F) = -xGSR^{-1}H \triangleq wH, \tag{8.3.4}$$

where we defined the vector $w = -xGSR^{-1}$. Note that w is necessarily nonzero, since otherwise it would follow that $xF = \lambda x$, which contradicts the fact that all the eigenvalues of F are strictly inside the unit circle.

Using (8.3.4) and the above definition for w we can write

$$x = wH(\lambda I - F)^{-1} \qquad \text{and} \qquad xGSR^{-1} + w = 0,$$

so that what we have shown is that there exists a nonzero vector w such that

$$wH(\lambda I - F)^{-1}GQ^{s/2} = 0 \quad \text{and} \quad w\left[H(\lambda I - F)^{-1}GSR^{-1} + I\right] = 0.$$

These are the two equalities in (8.3.3). ◆

Remark 2. The unit-circle controllability assumption arises often in system theory. We note however that it is automatically met when F is stable *and* $S = 0$. Moreover, since many different state-space models can give rise to the same z-spectrum $S_y(z)$, it is striking that unit-circle controllability is not model-dependent. ◆

Remark 3 [A Weaker Condition]. The proof of Lemma 8.3.1 also shows that the stability assumption on F can be replaced by the weaker condition that F has no unit-circle eigenvalues — see Prob. 8.10 and also Prob. 8.11. In this case, $S_y(z)$ will not be a true $z-$spectrum; it will be called a Popov function and will be studied in Apps. 8.A and 8.C. ◆

8.3.2 An Inertia Property

As a further step toward canonical spectral factorization, we shall use the result of part (a) of Lemma 8.2.1 to equivalently write

$$S_y(z) = \tag{8.3.5}$$

$$\begin{bmatrix} H(zI-F)^{-1} & I \end{bmatrix} \begin{bmatrix} -Z + FZF^* + GQG^* & FZH^* + GS \\ HZF^* + S^*G^* & R + HZH^* \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix},$$

for any arbitrary Hermitian matrix Z. Now although we cannot make any statements about the positivity of the center matrix appearing in (8.3.5) for an arbitrary Z, we can assert the following.

Lemma 8.3.2 (An Inertia Property) *Consider the z-spectrum (8.1.17) and suppose that $S_y(e^{j\omega}) > 0$ for all $-\pi \le \omega \le \pi$. Assume further that F does not have unit-circle eigenvalues. Then the central matrix*

$$T = \begin{bmatrix} -Z + FZF^* + GQG^* & FZH^* + GS \\ HZF^* + S^*G^* & R + HZH^* \end{bmatrix}, \tag{8.3.6}$$

must have at least p positive eigenvalues for any Hermitian matrix Z (recall that p is the dimension of the vectors $\mathbf{y}_i$). ■

Proof: The fact that F does not have unit-circle eigenvalues guarantees the existence of the inverse $(zI-F)^{-1}$ for all $|z| = 1$. Now note that for each $z = e^{j\omega}$, the expression (8.3.5) is the product of a $p \times (n+p)$, an $(n+p) \times (n+p)$, and an $(n+p) \times p$ matrix, which is strictly positive-definite and therefore has p positive eigenvalues. Therefore, the central $(n+p) \times (n+p)$ matrix must have at least p positive eigenvalues. Indeed, if the center matrix has $q < p$ positive eigenvalues, then we can express it as (by using the modal decomposition of T)

$$T = L_1 L_1^* - L_2 L_2^*,$$

where L_1 is $(n+p) \times q$ and has full rank q, while L_2 is $(n+p) \times (n+p-q)$. It follows that for any given $z = e^{j\omega}$, we can write

$$S_y(e^{j\omega}) = A - B,$$

where $A \ge 0$, $B \ge 0$, A is $p \times q$, B is $p \times (n+p-q)$, and A has rank q (at most). But since $q < p$, this means that A does not have full row rank, so that there exists some $1 \times p$ vector, a, such that $aA = 0$. But this implies $aS_y(e^{j\omega})a^* = -aBa^* \le 0$, which contradicts the assumption that $S_y(e^{j\omega}) > 0$. ◆

8.3.3 Algebraic Riccati Equations and Spectral Factorization

Now that we have shown that if (the $p \times p$ matrix) $S_y(e^{j\omega}) > 0$, the matrix T has *at least* p positive eigenvalues for any Hermitian Z, it is interesting to ask whether Z can be chosen so that T has *only* p positive eigenvalues and no negative eigenvalues, *i.e.*, if Z can be chosen so that T has minimal rank p.

Suppose for now that we can make such a choice, which we designate by $Z = P$. *Also, assume temporarily that for this choice $R + HPH^*$ is nonsingular.* Then we can perform the following block "upper-diagonal-lower" triangular factorization of T:

$$T = \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \begin{bmatrix} \Delta & 0 \\ 0 & R + HPH^* \end{bmatrix} \begin{bmatrix} I & 0 \\ X^* & I \end{bmatrix}, \tag{8.3.7}$$

where $X = (FPH^* + GS)(R + HPH^*)^{-1}$, and Δ is the Schur complement of $R + HPH^*$ in T, *i.e.*,

$$\Delta = -P + FPF^* + GQG^* - (FPH^* + GS)(R + HPH^*)^{-1}(HPF^* + S^*G^*).$$

Now we can see from Eq. (8.3.7) that if we choose P to make Δ equal to zero, then T reduces to

$$T = \begin{bmatrix} (FPH^*+GS)(R+HPH^*)^{-1} \\ I \end{bmatrix} (R+HPH^*) \begin{bmatrix} (FPH^*+GS)(R+HPH^*)^{-1} \\ I \end{bmatrix}^*$$

showing that T becomes of minimal rank p. For reasons to become clear later (Sec. 8.3.5), we shall define

$$R_e \triangleq R + HPH^* \quad \text{and} \quad K_p \triangleq (FPH^* + GS)R_e^{-1}. \tag{8.3.8}$$

Then we can write

$$\begin{aligned} S_y(z) &= \begin{bmatrix} H(zI-F)^{-1} & I \end{bmatrix} \begin{bmatrix} K_p \\ I \end{bmatrix} R_e \begin{bmatrix} K_p^* & I \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix} \\ &= \begin{bmatrix} H(zI-F)^{-1}K_p + I \end{bmatrix} R_e \begin{bmatrix} H(z^{-*}I - F)^{-1}K_p + I \end{bmatrix}^*. \end{aligned}$$

In other words, if we define

$$L(z) \triangleq H(zI-F)^{-1}K_p + I, \tag{8.3.9}$$

then we have a factorization of $S_y(z)$,

$$S_y(z) = L(z)R_eL^*(z^{-*}), \tag{8.3.10}$$

and there will be a different factorization for every solution P of the nonlinear equation

$$\Delta = -P + FPF^* + GQG^* - K_pR_eK_p^* = 0. \tag{8.3.11}$$

Now since F is stable, every factor $L(z)$ will have all its poles strictly inside the unit circle, so that all the $L(z)$ will be stable and causal (*cf.* Sec. 6.2). If we choose a solution P for which the inverse of $L(z)$ is also stable and causal, this particular $L(z)$ will define the canonical factor of $S_y(z)$. Now by an application of the matrix inversion lemma, we can write

$$[L(z)]^{-1} = \left[I + H(zI - F)^{-1}K_p\right]^{-1} = I - H(zI - F + K_pH)^{-1}K_p. \quad (8.3.12)$$

Therefore, if the solution P of (8.3.11) that we use is such that *the matrix $F - K_pH$ is stable*, then $L(z)$ will be causally invertible as well. Such a P will be called a *stabilizing* solution. The issue is whether such P exist. This requires a closer examination of (8.3.11).

8.3.4 Appropriate Solutions of the DARE

Equation (8.3.11) will be called the Discrete-Time Algebraic Riccati Equation (DARE). It will be convenient to rewrite it using (8.3.8) as

$$P = FPF^* + GQG^* - K_pR_eK_p^*, \quad (8.3.13)$$

which is the form we shall henceforth use.

To see if the DARE has a positive-semi-definite solution such that $F - K_pH$ is stable and $R + HPH^*$ is invertible, we must carefully study the properties of (8.3.13), a *nonlinear* algebraic set of equations. There may exist many solutions; for example, in the scalar case, the reader may verify that the DARE (8.3.13) is just a quadratic equation in the scalar unknown, P. And perhaps none of these solutions may be stabilizing; or they may all be such that $R + HPH^*$ is singular even though $R > 0$.

The DARE is studied at length in App. E, from which we quote the following general result (Thm. E.5.1): *The* DARE *(14.1.2) will have a stabilizing solution if, and only if, $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle. Moreover, any such stabilizing solution is unique and in fact also positive semi-definite (so that automatically $R + HPH^* > 0$).*

The condition that $\{F, H\}$ be detectable makes sense, since otherwise $F - KH$ will be unstable for *all* K (see App. C); this condition is automatically satisfied when F is stable. The above result (*viz.*, Thm. E.5.1) states that the stabilizing solution (when it exists) is unique and positive-semi-definite. This may encourage us to think that finding any positive-semi-definite solution of the DARE will be good enough to give us a stable closed-loop matrix, $F_p = F - K_pH$. Unfortunately, the answer is no. Under the detectability and unit circle controllability assumptions, it turns out that there can exist positive-semi-definite solutions of the DARE that are not stabilizing. To rule out this possibility, we must make a further assumption.

The following general result is established in App. E (see Thm. E.6.1). Assume that $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle. Then the DARE (8.3.13) will have only one positive-semi-definite solution if, and only if, $\{F^s, GQ^{s/2}\}$ is *stabilizable*. Moreover, the unique positive-semi-definite solution of the DARE also defines its stabilizing solution.

In other words, to guarantee that there be only one positive-semi-definite solution, we need the additional condition that $\{F^s, GQ^{s/2}\}$ be stabilizable, *i.e.*, that it be controllable on and outside the unit circle (and not just on the unit circle, which was the condition for the existence of a stabilizing solution to the DARE).[4] We summarize the above conclusions, which justify the assumptions we made at the beginning of Sec. 8.3.3 to carry out the algebra leading to the canonical factorization (8.3.10).

Theorem 8.3.1 (The DARE) *Assume that F is stable (or otherwise, that $\{F, H\}$ is detectable), $\{F^s, GQ^{s/2}\}$ controllable on the unit circle,*

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \geq 0 \quad \text{and} \quad R > 0.$$

*Under these conditions, the discrete-time algebraic Riccati equation (*DARE*),*

$$P = FPF^* + GQG^* - K_pR_eK_p^*, \quad (8.3.14)$$

where K_p and R_e are given by (8.3.8), has a unique solution P such that $F - K_pH$ is stable. Moreover, this so-called stabilizing solution is positive-semi-definite and results in a positive-definite $R_e = R + HPH^*$.

If, in addition, $\{F^s, GQ^{s/2}\}$ is stabilizable, then the stabilizing solution P is also the unique positive-semi-definite solution of the DARE *(8.3.14).* ■

Remark 4 [Generalizations] We may note that we can establish the existence of a stabilizing solution by using only the facts that $S_y(z)$ is a rational z-spectrum that is positive-definite on the unit circle — a brief overview of results in this direction is given in App. 8.C. ◆

8.3.5 Canonical Spectral Factorization and Innovations Models

We can summarize the discussions of the earlier sections in the following fundamental theorem.

Theorem 8.3.2 (Canonical Spectral Factorization) *Consider a spectrum $S_y(z)$ of the form (8.1.17), with $\{F, G, H, Q, S, R\}$ satisfying the conditions stated in Thm. 8.3.1. Then its canonical spectral factorization can be obtained as*

$$S_y(z) = L(z)R_eL^*(z^{-*}), \quad L(\infty) = I, \quad R_e > 0,$$

where

$$L(z) = I + H(zI - F)^{-1}K_p, \qquad L^{-1}(z) = I - H(zI - F + K_pH)^{-1}K_p,$$

$$K_p = (FPH^* + GS)R_e^{-1}, \qquad R_e = R + HPH^*,$$

and P is the unique positive-semi-definite solution to the DARE

$$P = FPF^* + GQG^* - K_pR_eK_p^*.$$

Moreover, $F - K_pH$ is stable, which, in addition to the stability of F, will guarantee that $L(z)$ is minimum-phase (i.e., both it and its inverse are analytic in $|z| \geq 1$). ■

[4] An example that shows the existence of more than one positive-semi-definite solution to the DARE, in the absence of the stabilizability assumption, is given in Sec. 14.2.

An important bonus of starting with state-space descriptions for $S_y(z)$ is that the canonical modeling filter, $L(z)$, and the canonical whitening filter, $L^{-1}(z)$, also display state-space structure. Thus recall that the modeling filter $L(z)$ relates the input innovations process $\{\mathbf{e}_i\}$ to the output process $\{\mathbf{y}_i\}$, or in z-transform notation

$$\mathbf{y}(z) = L(z)\mathbf{e}(z) = \big(H(zI - F)^{-1}K_p + I\big)\,\mathbf{e}(z).$$

Let us define

$$\boldsymbol{\theta}(z) \triangleq (zI - F)^{-1}K_p\mathbf{e}(z), \tag{8.3.15}$$

so that $\mathbf{y}(z) = H\boldsymbol{\theta}(z) + \mathbf{e}(z)$. The time-domain equivalent is

$$\begin{cases} \boldsymbol{\theta}_{i+1} &= F\boldsymbol{\theta}_i + K_p\mathbf{e}_i, \\ \mathbf{y}_i &= H\boldsymbol{\theta}_i + \mathbf{e}_i, \end{cases} \tag{8.3.16}$$

which is therefore a state-space description of the canonical modeling filter $L(z)$, with a stationary state vector process $\{\boldsymbol{\theta}_i\}$.

Of course this filter has a causal inverse, which in fact follows easily from (8.3.16). A simple rearrangement gives

$$\begin{cases} \boldsymbol{\theta}_{i+1} &= F\boldsymbol{\theta}_i + K_p(\mathbf{y}_i - H\boldsymbol{\theta}_i) = (F - K_pH)\boldsymbol{\theta}_i + K_p\mathbf{y}_i, \\ \mathbf{e}_i &= \mathbf{y}_i - H\boldsymbol{\theta}_i, \end{cases} \tag{8.3.17}$$

or in the transform domain

$$\mathbf{e}(z) = \big[I - H(zI - F + K_pH)^{-1}K_p\big]\mathbf{y}(z), \tag{8.3.18}$$

consistent with the formula (8.3.12) derived earlier for $L^{-1}(z)$. In summary, we obtain the following statement.

Theorem 8.3.3 (Time-Domain Canonical Models) *The canonical modeling filter $L(z)$ has a realization*

$$\begin{cases} \boldsymbol{\theta}_{i+1} &= F\boldsymbol{\theta}_i + K_p\mathbf{e}_i, \\ \mathbf{y}_i &= H\boldsymbol{\theta}_i + \mathbf{e}_i, \end{cases}$$

where

$$\langle \mathbf{e}_i, \mathbf{e}_j\rangle = R_e\delta_{ij}, \quad \langle \mathbf{e}_i, \boldsymbol{\theta}_j\rangle = 0, \quad j \le i,$$

and $\{K_p, R_e\}$ are as in Thm. 8.3.2. So also $L^{-1}(z)$ has a realization

$$\begin{cases} \boldsymbol{\theta}_{i+1} &= (F - K_pH)\boldsymbol{\theta}_i + K_p\mathbf{y}_i, \\ \mathbf{e}_i &= \mathbf{y}_i - H\boldsymbol{\theta}_i, \end{cases}$$

■

Remark 5 [Singular R_e] Although the matrix R_e in the above statements is always positive-definite, there are situations where we need to define the DARE for possibly singular R_e (*e.g.*, when $S_y(z)$ does not have full normal rank on the unit circle). These cases can be handled by working instead with a system of Riccati equations (SDARE) of the form,

$$\begin{cases} P &= FPF^* + GQG^* - K_pR_eK_p^* \\ K_pR_e &= FPH^* + GS \\ R_e &= R + HPH^*. \end{cases} \tag{8.3.19}$$

It is straightforward to check that for the above $\{P, K_p, R_e\}$, if we define $L(z)$ as in (8.3.9),then it still holds that $S_y(z) = L(z)R_eL^*(z^{-*})$ — see Prob. 8.5. The SDARE is further studied in Probs. 14.16 and 14.17. ◆

Remark 6 [Different Models] The results for the model (8.3.16) should be compared with those for the original state-space model

$$\begin{cases} \mathbf{x}_{i+1} &= F\mathbf{x}_i + \begin{bmatrix} G & 0 \end{bmatrix}\begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \\ \mathbf{y}_i &= H\mathbf{x}_i + \begin{bmatrix} 0 & I \end{bmatrix}\begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}. \end{cases} \tag{8.3.20}$$

We have written it in the above form to emphasize that model (8.3.20) for the process $\{\mathbf{y}_i\}$ is causal but not invertible — we cannot recover the inputs $\{\mathbf{u}_i, \mathbf{v}_i\}$ from the $\{\mathbf{y}_i\}$, unlike the canonical model which is both causal and causally invertible. ◆

8.3.6 A Digression: A Criterion for Positivity

We can combine Thms. 8.3.1 and 8.3.2 to obtain the following result. We mention it here because it is a special case of some important system theory results that go by the names of the Bounded Real lemma, the Positive-Real lemma, and the KYP lemma. These lemmas hold under weaker conditions than those stated below, and they are discussed in App. 8.C.

Lemma 8.3.3 (A Positivity Criterion) *Assume that F is stable, $R > 0$, and*

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \ge 0,$$

and denote

$$S_y(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix}\begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix}\begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}.$$

Then the following statements are equivalent:

(i) $S_y(e^{j\omega}) > 0$ *for all* $\omega \in [-\pi, \pi]$.

(ii) There exists a unique nonnegative definite solution P *of the* DARE,

$$P = FPF^* + GQG^* - K_p R_e K_p^*,$$

such that $F - K_p H$ *is stable, where* $K_p = (FPH^* + GS)R_e^{-1}$ *and* $R_e = R + HPH^* > 0$.

■

Proof: We prove both directions.

$(i) \Rightarrow (ii)$. Assume $S_y(e^{j\omega}) > 0$. Then by Lemma 8.3.1 it must hold that $\{F^s, GQ^{s/2}\}$ is unit-circle controllable where

$$F^s \triangleq F - GSR^{-1}H \quad \text{and} \quad Q^s \triangleq Q - SR^{-1}S^*.$$

Now the stability of F and the unit-circle controllability of $\{F^s, GQ^{s/2}\}$ imply by Thm. 8.3.1 that a unique nonnegative definite solution P of the DARE exists with the desired properties.

$(ii) \Rightarrow (i)$. Now assume that the DARE has a solution P with the stated properties. Then the factorization in Thm. 8.3.2 exists from which we conclude that $S_y(e^{j\omega}) = L(e^{j\omega})R_e L^*(e^{j\omega}) > 0$. ◆

8.4 RECURSIVE ESTIMATION GIVEN STATE-SPACE MODELS

Since the modeling and whitening filters have state-space structure, it is reasonable that the Wiener predictors, filters, and smoothers for stationary processes with state-space structure can also be written in state-space form.

8.4.1 Recursive Predictors

As often noted in Ch. 7, knowledge of the canonical modeling filter $L(z)$ immediately solves linear estimation problems. Thus recall from the results of Sec. 7.6.2 that the filter for estimating $\mathbf{y}_i$ from prior observations $\{\mathbf{y}_j, 0 \le j < i\}$ is given by (*cf.* (7.6.7))

$$\hat{\mathbf{y}}(z) = \left[I - L^{-1}(z)\right]\mathbf{y}(z). \tag{8.4.1}$$

Now when we have the standard state-space model (8.3.20) for $\mathbf{y}_i$, we can write more explicitly

$$\hat{\mathbf{y}}(z) = H(zI - F + K_p H)^{-1}K_p \mathbf{y}(z) = H\boldsymbol{\theta}(z).$$

Therefore, in the time-domain, we can obtain the Wiener predictor recursively via the equations

$$\hat{\mathbf{y}}_i = H\boldsymbol{\theta}_i, \tag{8.4.2}$$

$$\boldsymbol{\theta}_{i+1} = (F - K_p H)\boldsymbol{\theta}_i + K_p \mathbf{y}_i, \quad i > -\infty. \tag{8.4.3}$$

Remark 7. This result shows how the introduction of state-space structure nicely completes the discussion initiated in Remark 1 at the end of Sec. 7.6.1. ◆

Remark 8. Of course, since $\mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i$, we also have (by linearity) that $\hat{\mathbf{y}}_i = H\hat{\mathbf{x}}_i$, where $\hat{\mathbf{x}}_i$ denotes the l.l.m.s. estimator of $\mathbf{x}_i$ given $\{\mathbf{y}_j, -\infty < j < i\}$. Comparing this with (8.4.2) suggests that we may identify $\boldsymbol{\theta}_i = \hat{\mathbf{x}}_i$, which not only makes the canonical model (8.3.16) even more explicit, but also immediately gives us a recursive solution for the predicted state estimators of the original model. However, since H is generally not full rank, the above argument does not prove that $\boldsymbol{\theta}_i = \hat{\mathbf{x}}_i$. The result is true however, and the proof requires some more effort, as described below. ◆

8.4.2 Recursive State Predictors

Returning to the standard state-space model (8.1.1)–(8.1.2), let us denote the predicted estimator of the state in the z-domain by $\hat{\mathbf{x}}(z) = \mathcal{Z}\{\hat{\mathbf{x}}_i\}$. Now in terms of the innovations $\mathbf{e}(z)$, and since $S_e(z) = R_e$, the estimator $\hat{\mathbf{x}}(z)$ is given by the Wiener-Hopf formula (7.4.12), *viz.*,

$$\hat{\mathbf{x}}(z) = \{S_{xe}(z)\}_{\text{S.C.}} R_e^{-1}\mathbf{e}(z), \tag{8.4.4}$$

where $\{S_{xe}(z)\}_{\text{S.C.}}$ represents the *strictly* causal part of $S_{xe}(z)$.[5]

In Prob. 8.7 it is shown that $S_{xe}(z)$ can be expressed as the sum of three terms

$$S_{xe}(z) = (zI - F)^{-1}(FPH^* + GS) + PH^* + PF_p^*(z^{-1} - F_p^*)^{-1}H^*, \tag{8.4.5}$$

which allows us to readily identify the strictly causal part of $S_{xe}(z)$ as

$$\{S_{xe}(z)\}_{\text{S.C.}} = (zI - F)^{-1}(FPH^* + GS) = (zI - F)^{-1}K_p R_e.$$

This is because the stability of the matrices F and F_p implies that the term

$$(zI - F)^{-1}(FPH^* + GS) = \sum_{j=1}^{\infty} z^{-j} F^{j-1}(FPH^* + GS)$$

is stricly causal, while the term

$$PF_p^*(z^{-1}I - F_p^*)^{-1}H^* = \sum_{j=1}^{\infty} z^j P F_p^{j*} H^*$$

is strictly anticausal. Substituting into (8.4.4) we find that

$$\hat{\mathbf{x}}(z) = (zI - F)^{-1}K_p\mathbf{e}(z). \tag{8.4.6}$$

Comparing this expression with (8.3.15) we conclude that $\hat{\mathbf{x}}(z) = \boldsymbol{\theta}(z)$, *i.e.*, that the state in the modeling filter is the predicted estimator of the state in the original model, $\hat{\mathbf{x}}_i = \boldsymbol{\theta}_i$, so that the modeling filter (8.3.16) can be written as

$$\begin{cases} \hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + K_p\mathbf{e}_i, \\ \mathbf{y}_i = H\hat{\mathbf{x}}_i + \mathbf{e}_i, \quad i > -\infty. \end{cases} \tag{8.4.7}$$

[5] Note that here, unlike (7.4.12), we have used the strictly causal part of $S_{xe}(z)$, rather than its causal part $\{S_{xe}(z)\}_+$, since $\hat{\mathbf{x}}_i$ is a predicted estimator and depends only on *past* (and not current) values of $\mathbf{e}_j$. In other words, $\hat{\mathbf{x}}_i$ is a strictly causal function of the $\{\mathbf{e}_j\}$.

8.4.3 Recursive Smoothed Estimators

Let $\hat{\mathbf{x}}_{i|\infty}$ denote the smoothed estimator of the state $\mathbf{x}_i$ of (8.1.1) given all observations $\{\mathbf{y}_j, -\infty < j < \infty\}$. It will be denoted in the z-domain by $\hat{\mathbf{x}}_{|\infty}(z)$ and is given by (see Sec. 7.3.1)

$$\hat{\mathbf{x}}_{|\infty}(z) = S_{xy}(z)S_y^{-1}(z)\mathbf{y}(z) = S_{xe}(z)R_e^{-1}\mathbf{e}(z). \tag{8.4.8}$$

We shall now seek a recursive time-domain version of this formula. One way to proceed is to recall the decomposition (8.4.5)

$$\hat{\mathbf{x}}_{|\infty} = \{S_{xe}(z)\}_{\text{s.c.}} R_e^{-1}\mathbf{e}(z) + \{S_{xe}(z)\}_{\text{a.c.}} R_e^{-1}\mathbf{e}(z), \tag{8.4.9}$$

where

$$\{S_{xe}(z)\}_{\text{a.c.}} \triangleq PH^* + PF_p^*(z^{-1}I - F_p^*)^{-1}H^* = z^{-1}P(z^{-1}I - F_p^*)^{-1}H^*,$$

is the anticausal part of $S_{xe}(z)$. But this suggests, along with (8.4.4), that (8.4.9) can be rewritten as

$$\hat{\mathbf{x}}_{|\infty}(z) = \hat{\mathbf{x}}(z) + P\boldsymbol{\lambda}(z), \tag{8.4.10}$$

where we defined

$$\boldsymbol{\lambda}(z) \triangleq z^{-1}(z^{-1}I - F_p^*)^{-1}H^*\mathbf{e}(z) = (I - zF_p^*)^{-1}H^*\mathbf{e}(z). \tag{8.4.11}$$

Let $\boldsymbol{\lambda}_{i|\infty}$ be the time series associated with $\boldsymbol{\lambda}(z)$. It follows that $\boldsymbol{\lambda}_{i|\infty}$ is an anticausal function of the innovations $\{\mathbf{e}_i\}$ and that it satisfies the backwards-time model

$$\boldsymbol{\lambda}_{i|\infty} = F_p^*\boldsymbol{\lambda}_{i+1|\infty} + H^*R_e^{-1}\mathbf{e}_i. \tag{8.4.12}$$

This can be readily checked by taking z-transforms and verifying that $\boldsymbol{\lambda}(z)$ is related to $\mathbf{e}(z)$ as in (8.4.11). These equations still do not give us much insight into the physical meaning of the variables $\{\boldsymbol{\lambda}_{i|\infty}\}$ — for this, we shall have to wait until Ch. 10, where we will be able to recognize (8.4.10) as a steady-state Bryson-Frazier formula.

We summarize the discussions presented in this section in the following theorem.

Theorem 8.4.1 (Recursive Estimators) *Consider the time-invariant state-space model (8.1.1)–(8.1.2). The one-step process predictors $\{\hat{\mathbf{y}}_i\}$ are obtained as*

$$\hat{\mathbf{y}}_i = H\hat{\mathbf{x}}_i, \tag{8.4.13}$$

where $\{\hat{\mathbf{x}}_i\}$, the l.l.m.s. predictors of the states $\{\mathbf{x}_i\}$ using the observations $\{\mathbf{y}_j\}_{j=-\infty}^{i-1}$, satisfy the forwards-time recursion

$$\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + K_p(\mathbf{y}_i - H\hat{\mathbf{x}}_i), \qquad i > -\infty, \tag{8.4.14}$$

where P is the unique positive-semi-definite solution to the DARE *(8.3.14) that stabilizes $F_p = F - K_pH$, with $R_e = R + HPH^*$ and $K_p = (FPH^* + GS)R_e^{-1}$.*

Moreover, $\hat{\mathbf{x}}_{i|\infty}$, the l.l.m.s.e. of $\mathbf{x}_i$ given all the observations $\{\mathbf{y}_j\}_{j=-\infty}^{\infty}$, can be found as

$$\hat{\mathbf{x}}_{i|\infty} = \hat{\mathbf{x}}_i + P\boldsymbol{\lambda}_{i|\infty}, \qquad i > -\infty, \tag{8.4.15}$$

where $\boldsymbol{\lambda}_{i|\infty}$ satisfies the backwards-time recursion (8.4.12). ∎

8.5 FACTORIZATION GIVEN COVARIANCE DATA: RECURSIVE WIENER FILTERS

The alert reader may have noted a conceptual discontinuity in our discussions in this chapter. The Wiener theory is based on access to the second-order statistics (covariances and power spectra), whereas we have assumed more in this chapter — namely the availability of a state-space model for the process $\{\mathbf{y}_i\}$, from which we computed the canonical factorization.

However, since the modeling filter $L(z)$ is uniquely determined by the z-spectrum $S_y(z)$, it should come as no surprise that we should be able to obtain equivalent factorization results by working directly with the covariance function, as we shall show in this section.

Thus let us assume that instead of a known state-space model $\{F, G, H, Q, R, S\}$, we have the covariance sequence $\{R_y(i)\}$, or the power spectral density function

$$S_y(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_y(k)e^{-j\omega k}.$$

Let us further assume that the $\{R_y(i)\}$ arise from some finite-dimensional time-invariant linear system of order n. Then by using what are called minimal realization techniques from linear system theory (see, *e.g.*, Ho and Kalman (1965) or Kailath (1980, Ch. 5)), we can find a triple $\{H, F, \bar{N}\}$ such that F is stable and

$$R_y(i) = HF^{i-1}\bar{N}, \quad i > 0. \tag{8.5.1}$$

Then the (infinite) covariance sequence is completely determined by knowledge of the parameters $\{H, F, \bar{N}, R_y(0)\}$.

Now, for motivation, suppose that

$$\begin{cases} \mathbf{x}_{i+1} = F\mathbf{x}_i + G\mathbf{u}_i, \\ \mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{8.5.2}$$

with

$$\left\langle \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_j \\ \mathbf{v}_j \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} \bar{\Pi} & 0 & 0 \\ 0 & \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\delta_{ij} & 0 \end{bmatrix}, \tag{8.5.3}$$

where $\bar{\Pi}$ is the unique nonnegative-definite solution of the Lyapunov equation

$$\bar{\Pi} = F\bar{\Pi}F^* + GQG^*, \tag{8.5.4}$$

is some state-space model that leads to the same covariance values $\{R_y(0), R_y(i)\}$, where $R_y(i)$ is as in (8.5.1). In other words, we must have (*cf.* (8.1.8))

$$\bar{N} = F\bar{\Pi}H^* + GS. \tag{8.5.5}$$

Now we showed in Sec. 8.4.1 that the modeling filter $L(z)$ for such a process $\{\mathbf{y}_i\}$ could be written as in (8.4.7) with

$$K_p = (FPH^* + GS)R_e^{-1}, \quad R_e = R + HPH^*,$$

and P is the unique stabilizing solution of the DARE

$$P = FPF^* + GQG^* - K_pR_eK_p^*. \tag{8.5.6}$$

But since the modeling filter must be completely determined by the covariance function, *i.e.*, by knowledge of $\{H, F, \bar{N}, R_y(0)\}$, we must be able to rewrite $\{K_p, R_e\}$ in terms of these parameters. This is not hard to do. Define

$$\bar{\Sigma} \triangleq \|\hat{\mathbf{x}}_i\|^2, \tag{8.5.7}$$

and recall that

$$\bar{\Pi} = \|\mathbf{x}_i\|^2, \qquad P = \|\tilde{\mathbf{x}}_i\|^2, \tag{8.5.8}$$

and that

$$\bar{\Pi} = P + \bar{\Sigma}. \tag{8.5.9}$$

Therefore, we can write

$$R_e = R + HPH^* = R + H(\bar{\Pi} - \bar{\Sigma})H^* = R_y(0) - H\bar{\Sigma}H^*.$$

Next we can rewrite K_pR_e as

$$K_pR_e = FPH^* + GS = F(\bar{\Pi} - \bar{\Sigma})H^* + GS = \bar{N} - F\bar{\Sigma}H^*.$$

Therefore, $\{K_p, R_e\}$ are expressed in terms of the covariance parameters and $\bar{\Sigma}$. Fortunately, $\bar{\Sigma}$ is also determined by these parameters since

$$\bar{\Sigma} = \|\hat{\mathbf{x}}_i\|^2 = F\bar{\Sigma}F^* + K_pR_eK_p^*. \tag{8.5.10}$$

This leads to the following result.

Theorem 8.5.1 (Covariance-based Formulas) *Consider a stationary process $\{\mathbf{y}_i\}$ with covariance function expressed as in (8.1.8). Then the canonical modeling filter for $\{\mathbf{y}_i\}$ can be written as*

$$\begin{cases} \boldsymbol{\theta}_{i+1} = F\boldsymbol{\theta}_i + K_p\mathbf{e}_i, \\ \mathbf{y}_i = H\boldsymbol{\theta}_i + \mathbf{e}_i, \end{cases} \tag{8.5.11}$$

where

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = R_e\delta_{ij}, \qquad R_e = R_y(0) - H\bar{\Sigma}H^*, \qquad K_p = [\bar{N} - F\bar{\Sigma}H^*]R_e^{-1}, \tag{8.5.12}$$

and $\bar{\Sigma} = \|\hat{\mathbf{x}}_i\|^2$ is the unique nonnegative-definite solution of the algebraic Riccati equation

$$\bar{\Sigma} = F\bar{\Sigma}F^* + K_pR_eK_p^*, \tag{8.5.13}$$

that yields a stable $F - K_pH$. Note that $\bar{\Sigma} = \|\boldsymbol{\theta}_i\|^2$. Also that the Wiener predictor can be obtained as

$$\hat{\mathbf{y}}_i = H\boldsymbol{\theta}_i. \tag{8.5.14}$$

■

Lemma 8.5.1 (Relation to Underlying Model) *The state $\boldsymbol{\theta}_i$ in (8.5.11) can be identified as the l.l.m.s.e., $\hat{\mathbf{x}}_i$, of the state $\mathbf{x}_i$ of any state-space model for $\{\mathbf{y}_i\}$ with the same $\{F, H\}$ pair, so that of course $\bar{\Sigma} = \|\hat{\mathbf{x}}_i\|^2$.* ■

Proof of Thm. 8.5.1 and Lemma 8.5.1: Everything has been proved in the discussion prior to the theorem except that $\bar{\Sigma}$ stabilizes $F - K_pH$. But this is obvious because K_p is the same whether we specify it via P (*i.e.*, as $(FPH^* + GS)R_e^{-1}$) or via $\bar{\Sigma}$ (*i.e.*, as $(\bar{N} - F\bar{\Sigma}H^*)R_e^{-1}$). Therefore, $F - K_pH$ is stable. ◆

Remark 9. Thm. 8.5.1 can also be proved directly without invoking a process model such as (8.5.2). For this we can follow the method used in Sec. 8.2 but starting with (8.1.13) rather than (8.1.17). We leave the details to interested/active readers. ◆

8.6 EXTENSION TO TIME-VARIANT MODELS

We noted earlier that one advantage of state-space models over transfer function descriptions was that extensions to time-variant models were in many cases easily made — just replace constant parameters by time-variant ones. This is trivially evident in the case of the standard model — just replace $\{F, G, H, Q, R, S\}$ in the time-invariant model by $\{F_i, G_i, H_i, Q_i, R_i, S_i\}$. The powerful fact is that the same substitution works for (many) derived results, such as the estimator equations of Thms. 8.4.1 and 8.5.1. For example, (8.4.14) would now become

$$\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_i + K_{p,i}(\mathbf{y}_i - H\hat{\mathbf{x}}_i), \qquad \hat{\mathbf{x}}_0 = 0, \tag{8.6.1}$$

where

$$K_{p,i} = (F_iP_iH_i^* + G_iS_i)R_{e,i}^{-1}, \qquad R_{e,i} = R_i + H_iP_iH_i^*, \tag{8.6.2}$$

and the DARE for P is replaced by a Riccati recursion for P_i,

$$P_{i+1} = F_iP_iF_i^* + G_iQ_iG_i^* - K_{p,i}R_{e,i}K_{p,i}^*, \qquad P_0 = \Pi_0. \tag{8.6.3}$$

These are just the Kalman filter equations for the predicted state estimator (*cf.* Sec. 1.3.2 and Sec. 9.3). We could obtain a similar time-variant version of the covariance-based recursive Wiener estimators of Thm. 8.5.1.

These claims can be established by following the same line of argument as used earlier in this chapter, except that we should work with covariance matrices rather than z-spectra. Moreover, in the time-variant case, restriction to a finite observation interval is natural and avoids the problem of dealing with infinite matrices; note that such extensions also cover the case of time-invariant systems with observations starting at a finite time (and not in the remote past), *e.g.*, as in the transient observer discussed in Ch. 1. This approach will be presented in the appendices of the next chapter.

The reason for not doing this in the main text is that if we start with the assumption of a state-space model, then we can obtain the desired results by working directly with the model (as indicated by the dotted line in Fig. 8.1), instead of going from the model to the covariances, solving the problem, and then translating the solution back to state-space form. The latter is the route we used in this chapter for stationary processes, where for historical reasons (and motivated by communications applications) Wiener's

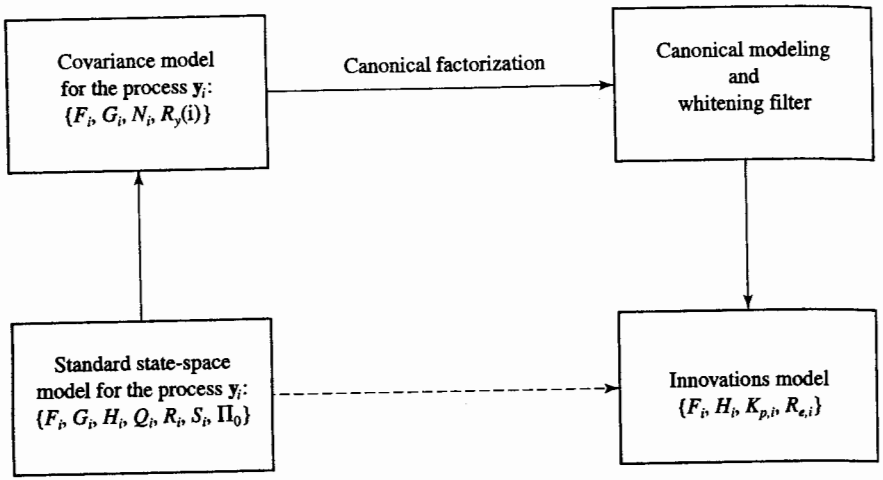


Figure 8.1 We shall show how to go along the dotted line in the next chapter.

approach via power spectral density functions was the paradigm. An early example of a more direct approach was in our discussion of the exponentially correlated process in Sec. 4.4. This direct route will be the one emphasized in the following chapters — see Fig. 8.1.

8.7 THE APPENDICES

In App. 8.A, we pursue in detail the remarks made at the end of Sec. 8.1. That is, we shall provide a more motivated approach to the results stated in Lemma 8.2.1.

As we described at length in the text, from this lemma we can proceed fairly quickly to obtain the canonical factorization of processes with state-space structure. An alternative way of exploiting the state-space description of the z-spectrum has been developed by Gohberg and Kaashoek (see, *e.g.*, Gohberg and Kaashoek (1986)). We describe this approach in App. 8.B, and reconcile the results with those in the main text.

Finally, as mentioned at the end of Sec. 8.2, in Apps. 8.C and 8.D we shall describe and discuss some important results in system theory — the KYP Lemma and the closely related Positive Real and Bounded Real lemmas.

8.8 COMPLEMENTS

The general theory for vector-valued stationary process estimation is quite elaborate and well outside our scope (see, *e.g.*, the early papers of Wiener and Masani (1957,1958)). Useful textbook references are Rozanov (1967) and Hannan (1970).

Processes with rational spectral density matrices are considerably easier to handle, but the early methods were generally rather involved and not easy to handle numerically. The breakthrough came with the introduction of state-space models and the work of Kalman in the early 1960s. With these models, much of the analysis is indifferent to whether the processes are scalar- or vector-valued. Also, in fact, to whether the model

coefficients are time-invariant or not, especially for the case of finite (even though growing) observation intervals; as noted in Sec. 7.9 (these were the problems of interest in the 1950s). The Kalman filter gave an elegant solution to these finite-time problems (see Ch. 9). Moreover, challenging and interesting problems arose in studying the limiting behavior of the Kalman filter, which were effectively solved by recognizing the duality to the solution of a quadratic regulator control problem; in other words, Lyapunov stability techniques were used to study the convergence. Though of course there are indications in the literature (see, *e.g.*, Willems (1971)), we would welcome references to other *direct* solutions of the Wiener problems for stationary processes with state-space models.

Sec. 8.2. An Equivalence Class for Input Gramians. In the (dual) controls context, continuous-time results equivalent to Lemma 8.2.1 (*cf.* App. 8.D) can be found in Willems (1971) and Molinari (1977). The interpretation of the results via Krein-space variables, as described in App. 8.A, is due to B. Hassibi.

Sec. 8.3. Canonical Spectral Factorization. The result of Lemma 8.3.1 can be found in Whittle (1990, p. 48). and no doubt elsewhere in the literature on algebraic Riccati equations. The fact (noted in Remark 2) that unit-circle controllability is model independent is notable. It reinforces the value of studying covariances and spectra along with the model. An interesting example of how spectral properties explain some apparently surprising phenomena (predictable, degenerate, and invariant directions) associated with discrete-time Riccati equations can be found in Gevers and Kailath (1973).

Sec. 8.5. Factorization Given Covariance Data: Recursive Wiener Filters. Actually, covariance-based results were first obtained in the finite-interval time-variant model problem — see Kailath and Geesey (1971) and also Sec. 16.2 (Thm. 16.2.3). They are especially relevant for stationary processes where state-space descriptions of the covariance function/z-spectrum can be obtained via realization techniques from linear system theory.

■ PROBLEMS

The system theory concepts of observability, controllability, detectability, stabilizability, and minimality, which are used in some of the problems below, are explained in App. C.

8.1 **(Innovations for a vector process)** A unit-variance zero-mean white-noise process $\{\mathbf{u}_i\}$ is applied to the stable and causal LTI system $1/(z-0.5)$. The output is denoted by $\mathbf{s}_i$ and is corrupted by additive zero-mean noise $\mathbf{v}_i$ to yield $\mathbf{y}_i = \mathbf{s}_i + \mathbf{v}_i$. The $\mathbf{v}_i$ is independent from $\mathbf{u}_i$ and is also white with variance r.

(a) Show that the whitening filter, $W(z)$, that yields the innovations $\{\mathbf{e}_i\}$ from the observations $\{\mathbf{y}_i\}$ is given by $W(z) = (z-0.5)/(z-\alpha)$, where α is the stable ($|\alpha| < 1$) solution to the quadratic equation $\alpha^2 - \frac{5r+4}{2r}\alpha + 1 = 0$.

(b) Now suppose that the same input $\mathbf{u}_i$ is applied to two *identical* LTI systems transfer function $1/(z-0.5)$ each, but that the outputs are corrupted by independent zero-mean white additive disturbances $\{\mathbf{v}_{i,1}, \mathbf{v}_{i,2}\}$ that have the same variance $r = 24/5$, and are uncorrelated with $\{\mathbf{u}_i\}$. The noisy outputs are denoted by $\mathbf{y}_{i,j} = \mathbf{s}_{i,j} + \mathbf{v}_{i,j}$, for $j = 1, 2$. Find a *first order* state-space model for the vector-valued process $\{\mathbf{y}_i = \text{col}\{\mathbf{y}_{i,1}, \mathbf{y}_{i,2}\}\}$, and compute the corresponding matrix-valued z-spectrum, $S_y(z)$.

(c) Clearly, if we apply the filter $W(z)$, obtained in part (a), to $\{\mathbf{y}_{i,1}\}$ and $\{\mathbf{y}_{i,2}\}$, the resulting process $\{\mathbf{e}_{i,1}\}$ will be the innovations for $\{\mathbf{y}_{i,1}\}$ and the resulting process $\{\mathbf{e}_{i,1}\}$ will be the innovations for $\{\mathbf{y}_{i,2}\}$. However, is it true that the vector-valued process $\{\text{col}\{\mathbf{e}_{i,1}, \mathbf{e}_{i,2}\}\}$ is the innovations for the vector-valued process $\{\mathbf{y}_i\}$? If yes, explain why. If not, find the true whitening filter for $\{\mathbf{y}_i\}$.

8.2 (An ARMA prediction problem) Consider again the second-order stationary process $\{\mathbf{y}_i\}$ of Prob. 5.3, *viz.*,

$$\mathbf{y}_{i+1} = a_0\mathbf{y}_i + a_1\mathbf{y}_{i-1} + \mathbf{u}_i + b\mathbf{u}_{i-1}, \quad i > -\infty,$$

with the same assumptions in that problem. We also assume that $|b| < 1$ and that the polynomials $z + b$ and $z^2 - a_0 z - a_1$ are coprime.

(a) Verify the validity of the following two-dimensional state-space model for the process $\{\mathbf{y}_i\}$:

$$\mathbf{x}_{i+1} = \underbrace{\begin{bmatrix} a_0 & a_1 \\ 1 & 0 \end{bmatrix}}_{F}\mathbf{x}_i + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{G}\mathbf{u}_i, \qquad \mathbf{y}_i = \underbrace{\begin{bmatrix} 1 & b \end{bmatrix}}_{H}\mathbf{x}_i.$$

Show that $\{F, G\}$ is controllable and $\{F, H\}$ is observable.

(b) The DARE that corresponds to the model in part (a) is given by

$$P = FPF^* - FPH^*(HPH^*)^{-1}HPF^* + GQG^*.$$

Compared with the DARE in (8.3.13), we see that the above equation is a special case with $R = 0$. Verify that $P = GQG^*$ is a positive-semi-definite solution of the above DARE. Verify also that $K_p = FG$ and that the resulting matrix $F - K_pH$ is stable.

(c) Show that the canonical spectral factor $L(z)$ is given by

$$L(z) = \frac{z(z+b)}{z^2 - a_0 z - a_1}.$$

Compare with the discussion in the remark to Prob. 5.3 and with the result of Ex. 6.5.3.

8.3 (Estimability) Consider the model (8.1.1)–(8.1.2), with stable F. Let $\bar{\Pi} = \|\mathbf{x}_i\|^2$.

(a) Show that the steady-state error variance, $P = \|\tilde{\mathbf{x}}_i\|^2$, is always less than or equal to $\bar{\Pi}$ ($P \le \bar{\Pi}$). Give a one-sentence interpretation of this result.

(b) We call the system *estimable* if P is strictly less than $\bar{\Pi}$ ($P < \bar{\Pi}$). Give a one-sentence interpretation of estimability.

(c) Show that the given linear system is not estimable if, and only if, there exists some row vector a such that $a\hat{\mathbf{x}}_i = 0$ for all i.

(d) Show that the given linear system is not estimable if, and only if, there exists a row vector a such that $a\mathbf{x}_i \perp \mathcal{L}\{\mathbf{y}_j,\ j < i-1\}$.

(e) Show that the linear system is estimable if, and only if, the matrix

$$[\bar{N} \quad F\bar{N} \quad F^2\bar{N} \quad \ldots]$$

has full rank, where $\bar{N} = F\bar{\Pi}H^* + GS$.

Remark. For more on estimability (and the dual concept of regulability), see Baram and Kailath (1988). ◆

8.4 (Choosing $Z = \bar{\Pi}$) Show that the choice $Z = \tilde{\Pi}$, where $\tilde{\Pi} = F\tilde{\Pi}F^* + GQG^*$, transforms the central matrix in (8.1.17) to the one in (8.1.13).

8.5 (Singular R_e) For any $\{F, G, H, R, Q, S\}$, let $\{P, K_p, R_e\}$ denote a solution, when it exists, of the system of equations (8.3.19), where R_e is possibly singular. Define $L(z) = H(zI - F)^{-1}K_p + I$ and verify that the following identify still holds:

$$L(z)R_eL^*(z^{-*}) = \begin{bmatrix} H(zI-F)^{-1} & I \end{bmatrix}\begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix}\begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}.$$

8.6 (Series expansion) Let F be a stable matrix and denote its maximum eigenvalue by $\lambda_{\max}(F)$. Show that the series

$$z^{-1}[I + z^{-1}F + z^{-2}F^2 + z^{-3}F^3 + \ldots]$$

converges uniformly to $(zI - F)^{-1}$ for all values of z in $|z| > |\lambda_{\max}(F)|$. Use this result to establish that the ROC of $S_y(z)$ is given by (8.1.14).

8.7 (Formula for $S_{\mathbf{xe}}(z)$) Start with the relation $\mathbf{e}(z) = L^{-1}(z)\mathbf{y}(z)$.

(a) Verify that $\mathbf{e}(z) = H(zI - F_p)^{-1}G\mathbf{u}(z) + [I - H(zI - F_p)^{-1}K_p]\mathbf{v}(z)$.

(b) Show further that $S_{xe}(z) = (zI - F)^{-1}GS_{ue}(z)$ and conclude that

$$S_{xe}(z) = (zI - F)^{-1}[GQG^* - GSK_p^*](z^{-1}I - F_p^*)^{-1}H^* + (zI - F)^{-1}GS.$$

(c) Now show that the DARE (8.3.13) can be re-expressed as $P = FPF_p^* + GQG^* - GSK_p^*$, and use this relation to verify that the expression for $S_{xe}(z)$ in part (b) can also be rewritten as

$$S_{xe}(z) = (zI - F)^{-1}(FPH^* + GS) + PH^* + PF_p^*(z^{-1} - F_p^*)^{-1}H^*.$$

8.8 (The additive decomposition of $S_{\mathbf{xe}}(z)$) We provide another derivation for the additive decomposition of $S_{xe}(z)$ in part (c) of Prob. 8.7.

(a) Show that any z-cross-spectrum of the form

$$S_{ab}(z) \triangleq \begin{bmatrix} H_a(zI - F_a)^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} (z^{-1}I - F_b^*)^{-1}H_b^* \\ I \end{bmatrix},$$

is invariant under the input cross-Gramian transformation

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \to \begin{bmatrix} A - Z + F_aZF_b^* & B + F_aZH_b^* \\ C + H_aZF_b^* & D + H_aZH_b^* \end{bmatrix},$$

for any matrix Z of appropriate dimensions.

(b) Now note that the z-cross-spectrum $S_{xe}(z)$, obtained in Prob.8.7 part (b), can be rearranged as follows:

$$S_{xe}(z) = \begin{bmatrix}(zI - F)^{-1} & I\end{bmatrix}\begin{bmatrix}GQG^* - GSK_p^* & GS \\ 0 & 0\end{bmatrix}\begin{bmatrix}(z^{-1}I - F_p^*)^{-1}H^* \\ I\end{bmatrix}.$$

Using the choice $Z = P$, where P is described in part (c) of Prob.8.7, verify that $S_{xe}(z)$ can also be written as

$$\begin{aligned} S_{xe}(z) &= \begin{bmatrix}(zI - F)^{-1} & I\end{bmatrix}\begin{bmatrix}0 & FPH^* + GS \\ PF_p^* & PH^*\end{bmatrix}\begin{bmatrix}(z^{-1}I - F_p^*)^{-1}H^* \\ I\end{bmatrix} \\ &= (zI - F)^{-1}(FPH^* + GS) + PH^* + PF_p^*(z^{-1} - F_p^*)^{-1}H^*. \end{aligned}$$

Remark. As with Lemma 8.2.1, there is also a Krein space interpretation for the result of part (a) — see Hassibi, Sayed, and Kailath (1999). ◆

8.9 (Strict positivity of Popov functions) Assume $\{F, H\}$ is detectable, $R > 0$,

$$\begin{bmatrix}Q & S \\ S^* & R\end{bmatrix} \geq 0,$$

and consider the Popov function

$$S_y(z) = \begin{bmatrix}H(zI - F)^{-1} & I\end{bmatrix}\begin{bmatrix}Q & S \\ S^* & R\end{bmatrix}\begin{bmatrix}(z^{-1}I - F^*)^{-1}H^* \\ I\end{bmatrix}.$$

Assume further that F has no unit-circle eigenvalues.

(a) Show that under these conditions, the statements (i) and (ii) of Lemma 8.3.3 are still equivalent.

(b) Show that the statements continue to be equivalent if we replace the condition of no unit-circle eigenvalues of F by the requirement that $\{F^s, Q^{s/2}\}$ be unit-circle controllable, where $F^s = F - SR^{-1}H$ and $Q^s = Q - SR^{-1}S^*$.

Hint. Use Thm. E.5.1. ◆

8.10 (Unit-circle zeros of Popov functions) Assume R is nonsingular, R and Q Hermitian, and that F has no unit circle eigenvalues. Consider the same Popov function of Prob. 8.9, where the center matrix may now be indefinite. Show that if $S_y(e^{j\omega})$ is nonsingular for all $\omega \in [-\pi, \pi]$ then $\{F^s, Q^s\}$ is unit-circle controllable. Is the converse true?

8.11 (Unit-circle zeros and a matrix pencil) Assume R is nonsingular, R and Q Hermitian, and that F has no unit circle eigenvalues. Consider the Popov function

$$S_y(z) = H(zI - F)^{-1}Q(z^{-1}I - F^*)^{-1}H^* + R,$$

where $\{R, Q\}$ can be indefinite. Let $\{\lambda, x\}$ denote a generalized eigenvalue-eigenvector pair of the matrix pencil shown below,

$$\begin{bmatrix}I & -Q \\ 0 & F^*\end{bmatrix}x = \lambda\begin{bmatrix}F & 0 \\ H^*R^{-1}H & I\end{bmatrix}x.$$

We want to show that $S_y(e^{j\omega})$ is nonsingular for all $\omega \in [-\pi, \pi]$ if, and only if, the above matrix pencil does not have any unit circle eigenvalues.

(a) Assume first that $|\lambda| = 1$ and partition x into $x = \text{col}\{x_1, x_2\}$. Show that $S_y(\lambda^*)x_1 = 0$.

(b) Now assume that $S_y(z)$ drops rank at some unit magnitude number, say $z = \lambda^*$ with $|\lambda| = 1$, and hence $S_y(\lambda^*)\bar{x}_1 = 0$ for some nonzero vector $\bar{x}_1$. Show that there exists a nonzero vector x such that $\{\lambda, x\}$ is an eigenvalue-eigenvector pair of the matrix pencil given above.

(c) In the special case $Q \geq 0$ and $R > 0$, show that the matrix pencil has no unit-circle eigenvalues if, and only if, $\{F, Q^{1/2}\}$ is unit-circle controllable. When Q and R are possibly indefinite matrices, is the unit-circle controllability of $\{F, Q\}$ sufficient?

Remark. The matrix pencil constructed in this problem is fundamental in the study of Riccati equations and is related to the so-called Hamiltonian matrix. (See App. E.7 and Probs. 14.16–14.17.) ◆

Appendices for Chapter 8

8.A THE POPOV FUNCTION

The *Popov function* is defined as

$$S_y(z) \triangleq \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}, \tag{8.A.1}$$

where Q and R are Hermitian matrices. It differs from the z-spectrum (8.2.1) in two important ways:

(i) The center matrix $\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}$ is not necessarily positive-semi-definite.

(ii) The system matrix F is not necessarily stable.

Note that unlike the z-spectrum, the Popov function can no longer be also defined via the Laurent series (8.1.11) since the ROC (8.1.14) will be empty when $|\lambda_{\max}(F)| > 1$. However, it can still be defined via (8.A.1) [for those values of z or z^{-1} that are not eigenvalues of F or F^*].

As a first application, we shall use this language to provide an interpretation of the invariance of the input Gramians result proved algebraically in Lemma 8.2.1. For this purpose, it will be useful to introduce time-invariant state-space models whose inputs lie in an indefinite metric (a so-called Krein) space.

Indefinite Metric Spaces. By an indefinite metric space, we shall mean a linear vector space $\mathcal{K}$ (see App. 4.A) where corresponding to any pair of elements or vectors, say $x \in \mathcal{K}$, $y \in \mathcal{K}$, the *inner product* $\langle x, y\rangle_{\mathcal{K}}$ is defined as an element of a ring $\mathcal{S}$ characterized by the following properties:

1. Linearity: $\langle \alpha_1 x_1 + \alpha_2 x_2, y\rangle_{\mathcal{K}} = \alpha_1\langle x_1, y\rangle_{\mathcal{K}} + \alpha_2\langle x_2, y\rangle_{\mathcal{K}}$.
2. Reflexivity: $\langle y, x\rangle_{\mathcal{K}} = \langle x, y\rangle_{\mathcal{K}}^*$.

However, the condition $\|x\|^2 \triangleq \langle x, x\rangle_{\mathcal{K}} = 0$ does not imply $x = 0$, so that nonzero vectors $x \in \mathcal{K}$ can have zero Gramian in this space. Moreover, Gramian matrices can be indefinite.

The ring $\mathcal{S}$ is usually the field of complex numbers but can be a more general algebraic object; in particular, for our discussions in this appendix it will be taken as the ring of square complex matrices. Likewise, the operation $*$ depends on the space $\mathcal{S}$. When $\mathcal{S}$ is the field of complex numbers, $*$ is just the operation of taking the complex conjugate; when $\mathcal{S}$ is the ring of square matrices, $*$ stands for the conjugate transpose.

A simple example of an indefinite metrix space is the following. Let $J = (1 \oplus -1)$ and consider the space of two-dimensional column vectors with $\langle \cdot, \cdot\rangle_{\mathcal{K}}$ defined by

$$\langle x, y\rangle_{\mathcal{K}} \triangleq x^* J y.$$

In this case, the operation $\langle \cdot, \cdot\rangle_{\mathcal{K}}$ associates a scalar with any two elements of $\mathbb{C}^2$. However, as was the case with random variables, the indefinite metric space used in this appendix is such that the operation $\langle \cdot, \cdot\rangle_{\mathcal{K}}$ is matrix-valued.

An Interpretation of Part (a) of Lemma 8.2.1. Now consider processes $\{\mathbf{x}_i, \mathbf{y}_i, \mathbf{u}_i, \mathbf{v}_i\}$ in an indefinite metric space $\mathcal{K}$ that satisfy a time-invariant state-space model of the form

$$\begin{cases} \mathbf{x}_{i+1} = F\mathbf{x}_i + \mathbf{u}_i, & i > -\infty, \\ \mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i, \end{cases} \tag{8.A.2}$$

with

$$\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \end{bmatrix} \rangle_{\mathcal{K}} = \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \delta_{ij}, \tag{8.A.3}$$

and where $Q = Q^*$, $R = R^*$ are Hermitian, but otherwise arbitrary, as is S. It is important to note that, since we are now considering the $\{\mathbf{u}_i, \mathbf{v}_i\}$ to lie in an indefinite metric space $\mathcal{K}$, the matrix appearing in (8.A.3) may be *indefinite*.

We define the generalized z-spectra of the white processes $\{\mathbf{u}_i, \mathbf{v}_i\}$ as the Popov functions $S_u(z) = Q$ and $S_v(z) = R$, respectively. Moreover, using the z-transform notation, we see that the output process $\{\mathbf{y}_i\}$ satisfies

$$\mathbf{y}(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} \mathbf{u}(z) \\ \mathbf{v}(z) \end{bmatrix},$$

and we define its generalized $z-$spectrum to be the Popov function (8.A.1).[6] The main difference is that now generalized z-spectra can be indefinite on the unit circle, $|z| = 1$.

Now suppose that we intend to add white and stationary disturbances $\{\bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$ (orthogonal in the indefinite metric space $\mathcal{K}$ to the original $\{\mathbf{u}_i, \mathbf{v}_i\}$) to the state-space model (8.A.2) such that the output z-spectrum $S_y(z)$ remains unchanged. In other words, the output of the modified state-space model

$$\begin{cases} \mathbf{x}_{i+1} + \bar{\mathbf{x}}_{i+1} = F(\mathbf{x}_i + \bar{\mathbf{x}}_i) + \mathbf{u}_i + \bar{\mathbf{u}}_i, \\ \mathbf{y}_i + \bar{\mathbf{y}}_i = H(\mathbf{x}_i + \bar{\mathbf{x}}_i) + \mathbf{v}_i + \bar{\mathbf{v}}_i, \end{cases} \tag{8.A.4}$$

should still have Popov function equal to $S_y(z)$, given in (8.A.1).

The Gramian matrix of $\{\bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$ is denoted by

$$\langle \begin{bmatrix} \bar{\mathbf{u}}_i \\ \bar{\mathbf{v}}_i \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{u}}_j \\ \bar{\mathbf{v}}_j \end{bmatrix} \rangle_{\mathcal{K}} = \begin{bmatrix} \bar{Q} & \bar{S} \\ \bar{S}^* & \bar{R} \end{bmatrix} \delta_{ij},$$

[6] We do not explicitly define $S_y(z)$ as the z-transform of the covariance function $R_y(i-j) = \langle \mathbf{y}_i, \mathbf{y}_j\rangle_{\mathcal{K}}$, say

$$S_y(z) = \sum_{i=-\infty}^{\infty} R_y(i) z^{-i},$$

because the above series will not converge when F is unstable — its ROC will be an empty set.

so that the Gramian matrix of $\{\mathbf{u}_i + \bar{\mathbf{u}}_i, \mathbf{v}_i + \bar{\mathbf{v}}_i\}$ is now given by

$$\begin{bmatrix} Q + \bar{Q} & S + \bar{S} \\ S^* + \bar{S}^* & R + \bar{R} \end{bmatrix},$$

and the new output generalized z-spectrum by

$$S_{y+\bar{y}}(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q + \bar{Q} & S + \bar{S} \\ S^* + \bar{S}^* & R + \bar{R} \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}.$$

Now, by linearity, $S_{y+\bar{y}}(z) = S_y(z) + S_{\bar{y}}(z)$. Therefore if $S_y(z)$ is to be unchanged, this implies that $S_{\bar{y}}(z)$, the generalized z-spectrum of the process $\{\bar{\mathbf{y}}_i\}$ defined by

$$\begin{cases} \bar{\mathbf{x}}_{i+1} = F\bar{\mathbf{x}}_i + \bar{\mathbf{u}}_i, \\ \bar{\mathbf{y}}_i = H\bar{\mathbf{x}}_i + \bar{\mathbf{v}}_i, \end{cases} \tag{8.A.5}$$

must be *zero, i.e.,*

$$S_{\bar{y}}(z) \triangleq \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} \bar{Q} & \bar{S} \\ \bar{S}^* & \bar{R} \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix} = 0.$$

In the sequel we shall first determine covariance matrices $\{\bar{Q}, \bar{R}, \bar{S}\}$ that result in an output process $\{\bar{\mathbf{y}}_i\}$ with a zero *covariance function.* Once this is done, we shall then show that the resulting $\{\bar{Q}, \bar{R}, \bar{S}\}$ yield a zero Popov function $S_{\bar{y}}(z)$, as desired.

To begin with, a simple calculation shows that

$$\langle \bar{\mathbf{y}}_i, \bar{\mathbf{y}}_i \rangle_{\mathcal{K}} = \bar{R} + H\langle \bar{\mathbf{x}}_i, \bar{\mathbf{x}}_i \rangle_{\mathcal{K}} H^*,$$

so that if we define the Hermitian matrix $Z = -\langle \bar{\mathbf{x}}_i, \bar{\mathbf{x}}_i \rangle_{\mathcal{K}}$, we may write[7]

$$\langle \bar{\mathbf{y}}_i, \bar{\mathbf{y}}_i \rangle_{\mathcal{K}} = \bar{R} - HZH^* = 0, \tag{8.A.6}$$

or $\bar{R} = HZH^*$. Likewise, a similar computation for $i > j$, shows that

$$\langle \bar{\mathbf{y}}_i, \bar{\mathbf{y}}_j \rangle_{\mathcal{K}} = HF^{i-j-1}(F\langle \bar{\mathbf{x}}_i, \bar{\mathbf{x}}_i \rangle_{\mathcal{K}} H^* + \bar{S}) = HF^{i-j-1}(-FZH^* + \bar{S}).$$

Thus choosing

$$\bar{S} = FZH^*, \tag{8.A.7}$$

we see that $\langle \bar{\mathbf{y}}_i, \bar{\mathbf{y}}_j \rangle_{\mathcal{K}} = 0$. Finally, using the state equation in (8.A.5) we may write

$$-Z = -FZF^* + \bar{Q}. \tag{8.A.8}$$

Combining (8.A.6), (8.A.7), and (8.A.8) shows that the variables $\{\bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$ can have as Gramian matrix

$$\begin{bmatrix} \bar{Q} & \bar{S} \\ \bar{S}^* & \bar{R} \end{bmatrix} = \begin{bmatrix} -Z + FZF^* & FZH^* \\ HZF^* & HZH^* \end{bmatrix}, \tag{8.A.9}$$

[7] Note that since the variables in (8.A.5) belong to an indefinite metric space, Z is in general indefinite.

for some Hermitian Z (which is the negative of the state Gramian matrix of the stationary process $\bar{\mathbf{x}}_i$). The above choice for $\{\bar{Q}, \bar{R}, \bar{S}\}$ thus results in a zero output covariance function. Now recall from the equality (8.2.4) that, for any Hermitian Z, the right-hand side of (8.A.9) results in a zero Popov function $S_{\bar{y}}(z)$, *i.e.,*

$$S_{\bar{y}}(z) \triangleq \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} -Z + FZF^* & FZH^* \\ HZF^* & HZH^* \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix} = 0,$$

so that the above choice for $\{\bar{Q}, \bar{R}, \bar{S}\}$ also annihilates $S_{\bar{y}}(z)$, as desired.

In summary, we provided an explanation for the invariance property of part (a) of Lemma 8.2.1 and in fact we did this for the general case of possibly unstable F and arbitrary matrices $\{Q, R, S\}$. The Hermitian matrix Z in (8.2.2) has the interpretation of being the Gramian matrix of the state vector in a state-space model with an identically zero Popov function.

8.B SYSTEM THEORY APPROACH TO RATIONAL SPECTRAL FACTORIZATION

There are several alternative approaches to rational matrix function factorizations. In our discussions in this chapter, we studied in some detail the problem of computing the canonical spectral factor of a z–spectrum $S_y(z)$ by exploiting the nonuniqueness of the input Gramian matrices (*cf.* Sec. 8.3). More specifically, we showed in Sec. 8.3.3 that a particular choice for the arbitrary matrix Z in (8.3.5) (*viz.*, the choice $Z = P$, where P is the unique stabilizing positive-semi-definite solution of the ARE (8.3.13) — see Thm. 8.3.2) leads to a rank deficient central matrix in (8.3.5), with minimal rank, and therefore to the canonical spectral factor $L(z)$.

We now explore a different route, often used in system theory (see, *e.g.,* Gohberg and Kaashoek (1986) and Bart, Gohberg, and Kaashoek (1979)) that is based on factoring a given rational transfer matrix function into a cascade of two or more rational transfer matrix functions. Their result is more abstractly stated (in terms of a certain projection operator (see Gohberg and Kaashoek (1986, p. 2)). Here we pursue the analysis more explicitly so as to again bring in the nonuniqueness of the input Gramians and then the DARE.

A Rational Transfer Matrix Representation. We start with a Popov function (assuming, without loss of generality, $G = I$),

$$S_y(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}, \tag{8.B.1}$$

and assume in this appendix that[8]

F is nonsingular.

[8] The case of singular F can be handled by employing singular transfer function representations, *viz.*, transfer functions of the form $G(z) = D + C(zE - A)^{-1}B$ with a possibly singular E.

We now try to express $S_y(z)$ in the form of a rational transfer matrix function, say

$$E(z) = D + [C_1 \;\; C_2]\left(\begin{bmatrix} zI & 0 \\ 0 & zI \end{bmatrix} - \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\right)^{-1}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}. \tag{8.B.2}$$

In other words, our first task is to determine matrices $\{A_{ij}, B_i, C_i, D\}$, for $i, j = 1, 2$, such that $S_y(z)$ is equal to $E(z)$.

To begin with, note that a block diagonal choice for A, say $A = \text{diag}\{A_{11}, A_{22}\}$, will not be sufficient since $(zI - A)^{-1}$ will not lead to the cross-terms that are present in (8.B.1). The next easiest choice is to assume a block triangular form for A, say

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}.$$

Now since $E(z)$ and $S_y(z)$ must have the same poles, which should coincide with the eigenvalues of F and F^{-*}, we further select A_{11} and A_{22} to be equal to F^{-*} and F, respectively. Hence,

$$A = \begin{bmatrix} F^{-*} & 0 \\ A_{21} & F \end{bmatrix},$$

and we obtain

$$(zI - A) = \begin{bmatrix} (zI - F^{-*}) & 0 \\ -A_{21} & (zI - F) \end{bmatrix},$$

and

$$(zI - A)^{-1} = \begin{bmatrix} (zI - F^{-*})^{-1} & 0 \\ (zI - F)^{-1}A_{21}(zI - F^{-*})^{-1} & (zI - F)^{-1} \end{bmatrix},$$

so that the expression (8.B.2) for $E(z)$ becomes $E(z)=$

$$D + C_1(zI - F^{-*})^{-1}B_1 + C_2(zI - F)^{-1}A_{21}(zI - F^{-*})^{-1}B_1 + C_2(zI - F)^{-1}B_2. \tag{8.B.3}$$

In order to facilitate a comparison of (8.B.3) with (8.B.1) we replace $(zI - F^{-*})^{-1}$ by

$$(zI - F^{-*})^{-1} = -F^* - F^*(z^{-1}I - F^*)^{-1}F^*,$$

to get

$$\begin{aligned} E(z) = {} & D - C_1F^*B_1 + C_2(zI - F)^{-1}(B_2 - A_{21}F^*B_1) - \\ & - C_1F^*(z^{-1}I - F^*)^{-1}F^*B_1 - \\ & - C_2(zI - F)^{-1}A_{21}F^*(z^{-1}I - F^*)^{-1}F^*B_1. \end{aligned} \tag{8.B.4}$$

Now comparing (8.B.4) with (8.B.1) we see that both expressions will coincide if the following equalities are satisfied:

$$\begin{gathered} D - C_1F^*B_1 = R, \quad C_2 = H, \quad B_2 - A_{21}F^*B_1 = S, \\ -C_1F^* = S^*, \quad F^*B_1 = H^*, \quad -A_{21}F^* = Q. \end{gathered}$$

These equations can be solved to yield

$$\begin{gathered} A_{21} = -QF^{-*}, \quad B_1 = F^{-*}H^*, \quad B_2 = S - QF^{-*}H^*, \\ C_1 = -S^*F^{-*}, \quad C_2 = H, \quad D = R - S^*F^{-*}H^*. \end{gathered}$$

In summary, we conclude that expression (8.B.1) can be matched with (8.B.2) by choosing

$$A = \begin{bmatrix} F^{-*} & 0 \\ -QF^{-*} & F \end{bmatrix}, \quad B = \begin{bmatrix} F^{-*}H^* \\ S - QF^{-*}H^* \end{bmatrix} = \begin{bmatrix} 0 \\ S \end{bmatrix} + A\begin{bmatrix} H^* \\ 0 \end{bmatrix}, \tag{8.B.5}$$

$$C = \begin{bmatrix} -S^*F^{-*} & H \end{bmatrix} = \begin{bmatrix} 0 & H \end{bmatrix} - \begin{bmatrix} S^* & 0 \end{bmatrix}A, \tag{8.B.6}$$

$$D = R - S^*F^{-*}H^*. \tag{8.B.7}$$

In particular, when $S = 0$ we obtain

$$A = \begin{bmatrix} F^{-*} & 0 \\ -QF^{-*} & F \end{bmatrix}, \quad B = \begin{bmatrix} F^{-*}H^* \\ QF^{-*}H^* \end{bmatrix}, \quad C = \begin{bmatrix} 0 & H \end{bmatrix}, \quad D = R.$$

Remark. Similar arguments apply had we started with a block upper triangular matrix A of the form

$$A = \begin{bmatrix} F & A_{12} \\ 0 & F^{-*} \end{bmatrix}.$$

◆

Nonunique Representations. The matrices $\{A, B, C, D\}$ just obtained are in fact highly nonunique. In particular, any similarity transformation will provide new matrices $\{\bar{A}, \bar{B}, \bar{C}, \bar{D}\}$ with the same transfer matrix representation $E(z)$. Now if we insist that the matrix $\bar{A}$ be of the same form

$$\bar{A} = \begin{bmatrix} F^{-*} & 0 \\ X & F \end{bmatrix},$$

with $\{F^{-*}, F\}$ on the diagonal, but possibly a different entry in the $(2, 1)$ block entry, then the only similarity transformations that we can use are those of the form

$$T = \begin{bmatrix} I & 0 \\ Z & I \end{bmatrix}, \tag{8.B.8}$$

for an arbitrary Z. Indeed, applying this similarity transformation to $\{A, B, C, D\}$ leads to the new matrices

$$\{\bar{A}, \bar{B}, \bar{C}, \bar{D}\} = \{TAT^{-1}, TB, CT^{-1}, D\},$$

and, hence,

$$\bar{A} = \begin{bmatrix} I & 0 \\ Z & I \end{bmatrix} \begin{bmatrix} F^{-*} & 0 \\ -QF^{-*} & F \end{bmatrix} \begin{bmatrix} I & \\ -Z & I \end{bmatrix} = \begin{bmatrix} F^{-*} & 0 \\ -\bar{Q}F^{-*} & F \end{bmatrix},$$

where we have defined $\bar{Q} \triangleq Q - Z + FZF^*$. Likewise,

$$\bar{B} = \begin{bmatrix} I & 0 \\ Z & I \end{bmatrix} \begin{bmatrix} F^{-*}H^* \\ S - QF^{-*}H^* \end{bmatrix} = \begin{bmatrix} F^{-*}H^* \\ \bar{S} - \bar{Q}F^{-*}H^* \end{bmatrix},$$

where we have defined $\bar{S} \triangleq S + FZH^*$. Moreover,

$$\bar{C} = \begin{bmatrix} -S^*F^{-*} & H \end{bmatrix} \begin{bmatrix} I & \\ -Z & I \end{bmatrix} = \begin{bmatrix} -\bar{S}F^{-*} & H \end{bmatrix},$$

and

$$\bar{D} = D = R - S^*F^{-*}H^* = \bar{R} - \bar{S}F^{-*}H^*,$$

where we have also defined $\bar{R} \triangleq R + HZH^*$. In summary, any similarity transformation of the form (8.B.8) yields a new realization $\{\bar{A}, \bar{B}, \bar{C}, \bar{D}\}$ whose parameters are given by

$$\bar{A} = \begin{bmatrix} F^{-*} & 0 \\ -\bar{Q}F^{-*} & F \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} F^{-*}H^* \\ \bar{S} - \bar{Q}F^{-*}H^* \end{bmatrix}, \tag{8.B.9}$$

$$\bar{C} = \begin{bmatrix} -\bar{S}^*F^{-*} & H \end{bmatrix}, \quad \bar{D} = \bar{R} - \bar{S}^*F^{-*}H^*, \tag{8.B.10}$$

where

$$\begin{bmatrix} \bar{Q} & \bar{S} \\ \bar{S}^* & \bar{R} \end{bmatrix} = \begin{bmatrix} -Z + FZF^* + Q & FZH^* + S \\ HZH^* + S^* & HZH^* + R \end{bmatrix}. \tag{8.B.11}$$

That is, given the earlier realization (8.B.5)–(8.B.7), in terms of the original matrices $\{Q, R, S\}$, we can always replace these matrices by new matrices $\{\bar{Q}, \bar{R}, \bar{S}\}$ that are constructed via (8.B.11), and still get an $E(z)$ in (8.B.2) that is equal to $S_y(z)$ in (8.B.1). Observe that (8.B.11) is the same parameterization we encountered earlier in Lemma 8.2.1 while studying the equivalence classes for input Gramian matrices (except that the present derivation requires an invertible F).

Cascade Factorization. Now that we have some freedom in selecting the parameters of the transfer matrix representation (8.B.2), let us examine how it can be exploited in order to determine a canonical spectral factorization of $E(z)$, *i.e.*, a factorization of the form

$$E(z) = L(z)R_eL^*(z^{-*}), \tag{8.B.12}$$

with $L(z)$ minimum phase and $L(\infty) = I$. The minimum phase requirement is equivalent to saying that $L(z)$ and its inverse should be analytic in $|z| \geq 1$.

For this purpose, we first note the following useful fact about the state-space realization of a cascade of two linear state-space models. Consider two transfer matrix functions

$$E_1(z) = D_1 + C_1(zI - A_1)^{-1}B_1, \quad E_2(z) = D_2 + C_2(zI - A_2)^{-1}B_2,$$

with state-space realizations

$$E_2(z): \begin{cases} \mathbf{x}_{i+1} = A_2\mathbf{x}_i + B_2\mathbf{u}_i \\ \mathbf{y}_i = C_2\mathbf{x}_i + D_2\mathbf{u}_i \end{cases} \qquad E_1(z): \begin{cases} \boldsymbol{\xi}_{i+1} = A_1\boldsymbol{\xi}_i + B_1\mathbf{v}_i \\ \boldsymbol{\eta}_i = C_1\boldsymbol{\xi}_i + D_1\mathbf{v}_i. \end{cases}$$

[The matrices $\{B_1, B_2, C_1, C_2\}$ are not related to the same matrices used earlier to denote the parameters of $E(z)$. They are simply used here for convenience of notation.]

Now, given the above realizations for $E_1(z)$ and $E_2(z)$, a state-space realization for the cascade $E_1(z)RE_2(z)$, with input $\mathbf{u}_i$ and output $\boldsymbol{\eta}_i$, and for some matrix R, can be verified to be

$$E_1(z)RE_2(z): \begin{cases} \begin{bmatrix} \mathbf{x}_{i+1} \\ \boldsymbol{\xi}_{i+1} \end{bmatrix} = \begin{bmatrix} A_2 & 0 \\ B_1RC_2 & A_1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_i \\ \boldsymbol{\xi}_i \end{bmatrix} + \begin{bmatrix} B_2 \\ B_1RD_2 \end{bmatrix} \mathbf{u}_i \\ \boldsymbol{\eta}_i = \begin{bmatrix} D_1RC_2 & C_1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_i \\ \boldsymbol{\xi}_i \end{bmatrix} + D_1RD_2\mathbf{u}_i. \end{cases}$$

In other words, the system matrices of the cascade system $E_1(z)RE_2(z)$ are

$$\left\{ \begin{bmatrix} A_2 & 0 \\ B_1RC_2 & A_1 \end{bmatrix}, \begin{bmatrix} B_2 \\ B_1RD_2 \end{bmatrix}, \begin{bmatrix} D_1RC_2 & C_1 \end{bmatrix}, D_1RD_2 \right\}. \tag{8.B.13}$$

Returning to (8.B.12), assume we represent the spectral factor $L(z)$ in state-space form as follows:

$$L(z) = I + C(zI - F)^{-1}B, \tag{8.B.14}$$

for some matrices (C, B) to be determined. It then follows that

$$\begin{aligned} L^*(z^{-*}) &= I + B^*(z^{-1}I - F^*)^{-1}C^*, \\ &= I - B^*[F^{-*} + F^{-*}(zI - F^{-*})^{-1}F^{-*}]C^*, \\ &= I - B^*F^{-*}C^* - B^*F^{-*}(zI - F^{-*})^{-1}F^{-*}C^*. \end{aligned} \tag{8.B.15}$$

Expressions (8.B.14) and (8.B.15) provide state-space descriptions for $L(z)$ and $L^*(z^{-*})$. Hence, in view of the result (8.B.13) for the cascade of two systems, we conclude that a realization for $E(z) = L(z)R_eL^*(z^{-*})$ in terms of the above $\{F, B, C\}$ matrices is given by

$$\begin{bmatrix} F^{-*} \\ -BR_eB^*F^{-*} \quad F \end{bmatrix}, \quad \begin{bmatrix} F^{-*}C^* \\ BR_e(I - B^*F^{-*}C^*) \end{bmatrix},$$
$$\begin{bmatrix} -R_eB^*F^{-*} \quad C \end{bmatrix}, \quad R_e(I - B^*F^{-*}C^*). \tag{8.B.16}$$

The question now is whether we can find a Z in (8.B.11), or equivalently a similarity transformation T in (8.B.8), that makes (8.B.9)–(8.B.10) agree with (8.B.16)

Comparing (8.B.16) with (8.B.9)–(8.B.10), we see that Z should be chosen such that

$$\bar{Q} = BR_eB^* = Q - Z + FZF^*, \qquad H = C,$$
$$\bar{S} = BR_e = S + FZH^*, \qquad \bar{R} = R_e = R + HZH^*.$$

These expressions establish that this particular choice for Z must satisfy

$$\begin{bmatrix} Q - Z + FZF^* & S + FZH^* \\ S^* + HZH^* & R + HZH^* \end{bmatrix} = \begin{bmatrix} B \\ I \end{bmatrix} R_e \begin{bmatrix} B^* & I \end{bmatrix},$$

and, hence, the matrix on the left-hand side of the above equality must be rank deficient. If $(R + HZH^*)$ is invertible, this means that Z should satisfy the DARE

$$Z = FZF^* + (FZH^* + S)(R + HZH^*)^{-1}(FZH^* + S)^* + Q,$$

so that $B = (FZH^* + S)R_e^{-1} = K_p$, as in Thm. 8.3.2.

8.C THE KYP AND RELATED LEMMAS

Lemma 8.3.3 made an important connection between z-spectra that are positive-definite on the unit circle and the existence of stabilizing solutions of DAREs. The lemma is in fact a special case of a deeper result in system theory that holds under considerably weaker assumptions. If we relax the requirements of a stable F and a nonnegative-definite center matrix in (8.1.17), $S_y(z)$ cannot be interpreted as a true z-spectrum anymore but is an object called a Popov function (*cf.* App. 8.A). It turns out that the nonnegativity on the unit circle of such Popov functions is again equivalent to the existence of stabilizing (or marginally stabilizing) solutions of DAREs (or SDAREs). This equivalence manifests itself in many other important results in system and control theories. In particular, it turns out to be crucial in characterizing so-called positive-real systems as well as bounded systems, which often arise in the context of adaptive and $\mathcal{H}_\infty$ control designs, among other fields.

For this reason, and for ease of reference, we state here some of these results. Proofs are omitted for brevity (see though the monograph (Hassibi, Sayed, and Kailath (1999)). The continuous-time counterparts are noted in App. 8.D.

We start with the Kalman-Yakubovich-Popov (KYP) Lemma.

Lemma 8.C.1 (The KYP Lemma) *Consider a detectable pair $\{F, H\}$ and assume F does not have unit-circle eigenvalues, with $F \in \mathbb{C}^{n\times n}$ and $H \in \mathbb{C}^{p\times n}$. Consider also arbitrary matrices $\{Q, S, R\}$ of appropriate dimensions and define the Popov function*

$$S_y(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix},$$

where the central matrix is Hermitian but may be indefinite. The following three statements are equivalent:

1. *$S_y(e^{j\omega}) \geq 0$ for all $\omega \in [-\pi, \pi]$.*
2. *There exists a Hermitian matrix P such that*
$$\begin{bmatrix} Q - P + FPF^* & S + FPH^* \\ S^* + HPF^* & R + HPH^* \end{bmatrix} \geq 0.$$
3. *There exist an $n \times n$ Hermitian matrix P, a $p \times p$ matrix $R_e \geq 0$, and an $n \times p$ matrix K_p, such that*
$$\begin{bmatrix} Q - P + FPF^* & S + FPH^* \\ S^* + HPF^* & R + HPH^* \end{bmatrix} = \begin{bmatrix} K_p \\ I \end{bmatrix} R_e \begin{bmatrix} K_p^* & I \end{bmatrix}.$$

The following two statements are also equivalent:

(a) *$S_y(e^{j\omega}) > 0$ for all $\omega \in [-\pi, \pi]$.*

(b) *There exists a unique Hermitian solution P of the* SDARE

$$P = FPF^* + Q - K_pR_eK_p^*, \qquad K_pR_e = (FPH^* + S), \qquad R_e = R + HPH^*,$$

such that $F - K_pH$ is stable and $R_e > 0$.

■

Proof: The directions (3)⇒(2)⇒(1) and (b)⇒(a) are immediate to verify. The other directions require more effort and they would follow by establishing first that the nonnegativity of $S_y(e^{j\omega})$ guarantees the existence of a Hermitian solution to the SDARE.

◆

This relationship between Popov functions and DAREs or SDAREs is also useful in characterizing rational positive-real systems. The reason is that positive-real systems are intimately related to Popov functions that are nonnegative-definite on the unit circle. Indeed, a rational positive-real system is a stable system with a square transfer matrix function, say $G(z)$, that satisfies

$$G(e^{j\omega}) + G^*(e^{j\omega}) \geq 0 \qquad \text{for all } \omega \in [-\pi, \pi]. \tag{8.C.1}$$

Clearly, if $G(z)$ is positive-real then $G(z) + G^*(z^{-*})$ is a Popov function that is nonnegative-definite on the unit circle. Thus by applying the KYP lemma to $G(z)+G^*(z^{-*})$, we obtain the following result.

Lemma 8.C.2 (Positive-Real Lemma) *Let $G(z) = D+H(zI-F)^{-1}\bar{N}$ be a given state-space realization of a stable rational matrix function, with $D \in \mathbb{C}^{p\times p}$ and $F \in \mathbb{C}^{n\times n}$. The following three statements are equivalent.*

1. *$G(z)$ is positive-real.*
2. *There exists an $n \times n$ matrix $\bar{\Sigma} \geq 0$ such that*

$$\begin{bmatrix} \bar{\Sigma} - F\bar{\Sigma}F^* & \bar{N} - F\bar{\Sigma}H^* \\ \bar{N}^* - H\bar{\Sigma}F^* & D + D^* - H\bar{\Sigma}H^* \end{bmatrix} \geq 0.$$

3. *There exist an $n \times n$ matrix $\bar{\Sigma} \geq 0$, a $p \times p$ matrix $R_e \geq 0$, and an $n \times p$ matrix K_p, such that*

$$\begin{bmatrix} \bar{\Sigma} - F\bar{\Sigma}F^* & \bar{N} - F\bar{\Sigma}H^* \\ \bar{N}^* - H\bar{\Sigma}F^* & D + D^* - H\bar{\Sigma}H^* \end{bmatrix} = \begin{bmatrix} K_p \\ I \end{bmatrix} R_e \begin{bmatrix} K_p^* & I \end{bmatrix}.$$

The following two statements are also equivalent:

(a) *$G(z)$ is strictly positive-real, i.e., $G(e^{j\omega}) + G^*(e^{j\omega}) > 0$ for all $\omega \in [-\pi, \pi]$.*

(b) *There exists a unique nonnegative-definite solution $\bar{\Sigma}$ of the* DARE

$$\bar{\Sigma} = F\bar{\Sigma}F^* + K_pR_eK_p^*, \quad R_e = D + D^* - H\bar{\Sigma}H^*, \quad K_pR_e = \bar{N} - F\bar{\Sigma}H^*,$$

such that $F - K_pH$ is stable and $R_e > 0$.

■

Proof: Once more, the directions (3)⇒(2)⇒(1) and (b)⇒(a) are immediate. The other directions follow from the KYP lemma. ◆

The positive-real lemma was first given by Yakubovic (1962) and Kalman (1963a) for scalar continuous-time systems and extended to the vector case by Popov (1964). Positive-real (p.r.) functions are also known in network theory as impedance functions since they can be regarded as the impedance functions of passive circuits (as first shown by Brune (1931)). The interplay between z-spectra and p.r. functions is much older and can be traced back to the early 1900s in works on the so-called trigonometric moment problem by Toeplitz (1907), Carathéodory (1907), Pick (1916), Schur (1917), and Nevanlinna (1919). The book by Akhiezer (1961) provides an overview of these early results on moment problems.

Another useful consequence of the KYP lemma is a characterization of bounded rational functions; such functions are known in circuit theory as scattering functions. These are stable (possibly nonsquare) systems with matrix functions $G(z)$ that satisfy

$$\sup_{\omega\in[-\pi,\pi]} \|G(e^{j\omega})\| \leq \gamma, \tag{8.C.2}$$

for some positive γ and for some matrix norm $\|\cdot\|$, say the induced 2-norm (or maximum singular value of its argument). Here again, if (8.C.2) holds, then $\gamma^2 I - G(z)G^*(z^{-*})$ is a Popov function that is nonnegative on the unit circle. In fact, bounded functions

provide yet another example of a situation where indefinite center matrices can arise. Indeed, note that

$$\gamma^2 I - G(z)G^*(z^{-*}) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} -BB^* & -BD^* \\ -DB^* & \gamma^2 I - DD^* \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix},$$

which exhibits an indefinite center matrix.

Now by applying the KYP lemma to $\gamma^2 I - G(z)G^*(z^{-*})$ (or to $\gamma^2 I - G^*(z^{-*})G(z)$), we obtain the following result.

Lemma 8.C.3 (Bounded Real Lemma) *Let $G(z) = D + H(zI - F)^{-1}B$ denote a stable rational matrix function with $D \in \mathbb{C}^{p\times p}$ and $F \in \mathbb{C}^{n\times n}$. The following three statements are equivalent:*

1. *$\sup_{\omega\in[-\pi,\pi]} \|G(e^{j\omega})\| \leq \gamma$.*
2. *There exists an $n \times n$ matrix $\bar{\Sigma} \geq 0$ such that*

$$\begin{bmatrix} \bar{\Sigma} - F\bar{\Sigma}F^* - BB^* & -BD^* - F\bar{\Sigma}H^* \\ -DB^* - H\bar{\Sigma}F^* & \gamma^2 I - DD^* - H\bar{\Sigma}H^* \end{bmatrix} \geq 0.$$

3. *There exist an $n \times n$ matrix $\bar{\Sigma} \geq 0$, a $p \times p$ matrix $R_e \geq 0$, and an $n \times p$ matrix K_p, such that*

$$\begin{bmatrix} \bar{\Sigma} - F\bar{\Sigma}F^* - BB^* & -BD^* - F\bar{\Sigma}H^* \\ -DB^* - H\bar{\Sigma}F^* & \gamma^2 I - DD^* - H\bar{\Sigma}H^* \end{bmatrix} = \begin{bmatrix} K_p \\ I \end{bmatrix} R_e \begin{bmatrix} K_p^* & I \end{bmatrix}.$$

The following two statements are also equivalent:

(a) *$\sup_{\omega\in[-\pi,\pi]} \|G(e^{j\omega})\| < \gamma$.*

(b) *There exists a unique nonnegative-definite solution $\bar{\Sigma}$ of the* DARE

$$\bar{\Sigma} = F\bar{\Sigma}F^* + BB^* + K_pR_eK_p^*, \quad R_e = \gamma^2 I - DD^* - H\bar{\Sigma}H^*, \quad K_pR_e = -BD^* - F\bar{\Sigma}H^*,$$

such that $F - K_pH$ is stable and $R_e > 0$.

■

8.D VECTOR SPECTRAL FACTORIZATION IN CONTINUOUS TIME

We shall very briefly present the continuous-time analogs of the results of this chapter. In particular, we shall show that finding the canonical spectral factor in the rational vector case can be reduced to the solution of a continuous-time algebraic Riccati equation (CARE). We shall also state the continuous-time counterparts of the KYP lemma, the positive-real lemma, and the bounded real lemma.

The model now is[9]

$$\dot{\mathbf{x}}(t) = F\mathbf{x}(t) + G\mathbf{u}(t), \tag{8.D.1}$$

$$\mathbf{y}(t) = H\mathbf{x}(t) + \mathbf{v}(t), \qquad t > -\infty, \tag{8.D.2}$$

where $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$ are zero-mean white-noise processes such that

$$\left\langle \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{v}(t) \end{bmatrix}, \begin{bmatrix} \mathbf{u}(s) \\ \mathbf{v}(s) \end{bmatrix} \right\rangle = \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \delta(t-s). \tag{8.D.3}$$

We assume that

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \geq 0, \qquad R > 0,$$

and F is a stable matrix (*i.e.*, all its eigenvalues are in the open left-half plane) so that the processes $\{\mathbf{x}(\cdot), \mathbf{y}(\cdot)\}$ are wide-sense stationary.

Let $\bar{\Pi} = \|\mathbf{x}(t)\|^2$ denote the state covariance matrix. It can be shown that $\bar{\Pi}$ is the unique solution of the Lyapunov equation (see, *e.g.*, Eq. (16.4.15)):

$$0 = F\bar{\Pi} + \bar{\Pi}F^* + GQG^*. \tag{8.D.4}$$

Introduce further the matrix $\bar{N} = \bar{\Pi}H^* + GS$. Then the covariance function of the output process $\{\mathbf{y}(\cdot)\}$ can also be shown to be given by

$$\langle \mathbf{y}(t), \mathbf{y}(s) \rangle = R\delta(t-s) + \begin{cases} He^{F(t-s)}\bar{N} & t \geq s, \\ \bar{N}^* e^{F^*(s-t)}H^* & t < s, \end{cases} \tag{8.D.5}$$

where the matrix exponential matrix e^{Ft} is defined in App. C; it is the unique solution of the matrix differential equation $\dot{X}(t) = FX(t)$, $X(0) = I$.

Using the Laplace transform notation, it is immediate to verify from the state-space model (8.D.1)–(8.D.2) that

$$\mathbf{Y}(s) = \begin{bmatrix} H(sI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} G\mathbf{U}(s) \\ \mathbf{V}(s) \end{bmatrix},$$

so that the s-spectrum of the output process $\{\mathbf{y}(\cdot)\}$ is given by (*cf.* (6.A.9))

$$S_y(s) = \begin{bmatrix} H(sI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix} \begin{bmatrix} (-sI - F^*)^{-1}H^* \\ I \end{bmatrix}. \tag{8.D.6}$$

A second expression for $S_y(s)$ follows by evaluating the Laplace transform of the covariance function $R_y(\cdot)$ in (8.D.5). Doing so leads to

$$S_y(s) = \begin{bmatrix} H(sI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} 0 & \bar{N} \\ \bar{N}^* & R \end{bmatrix} \begin{bmatrix} (-sI - F^*)^{-1}H^* \\ I \end{bmatrix}. \tag{8.D.7}$$

[9] Such continuous-time models are discussed in greater detail in Ch. 16.

We thus see from (8.D.6) and (8.D.7) that two different center matrices (one is non-negative-definite while the other is indefinite) lead to the same s–spectrum. So as in discrete-time, we have nonuniqueness in the "center" matrix, and by very similar arguments it can be checked that, for *any* Hermitian matrix Z, any center matrix of the form

$$T = \begin{bmatrix} FZ + ZF^* + GQG^* & ZH^* + GS \\ HZ + S^*G^* & R \end{bmatrix} \tag{8.D.8}$$

will lead to the same s-spectrum, $S_y(s)$. [Krein space arguments as in App. 8.A can be used to explain the particular form of the matrix T.]

We can now proceed fairly directly to obtain spectral factorizations. Recall from App. 6.A that in order for $S_y(s)$ to admit a canonical spectral factorization as defined in that section, it must be positive-definite on the imaginary axis,

$$S_y(j\omega) > 0 \qquad \text{for all} \qquad -\infty < \omega < \infty. \tag{8.D.9}$$

An argument similar to that of Lemma 8.3.1 will show that (8.D.9) holds if, and only if, the matrix pair $\{F^s, GQ^{s/2}\}$ has no uncontrollable modes on the imaginary axis, where

$$F^s \triangleq F - GSR^{-1}H \quad \text{and} \quad Q^s \triangleq Q - SR^{-1}S^*.$$

This means that there should exist no eigenvalue of F^s on the imaginary axis, with left eigenvector x (*i.e.*, $xF^s = \lambda x$, $\mathrm{Re}(\lambda) = 0$), such that $xGQ^{s/2} = 0$.

With this assumption, we now write $S_y(s)$ as

$$S_y(s) = \tag{8.D.10}$$

$$\begin{bmatrix} H(sI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} FZ + ZF^* + GQG^* & ZH^* + GS \\ HZ + S^*G^* & R \end{bmatrix} \begin{bmatrix} (-sI - F^*)^{-1}H^* \\ I \end{bmatrix},$$

for any arbitrary Hermitian matrix Z. Then, as in Lemma 8.3.2, we can argue that the central matrix T must have at least p positive eigenvalues for any Z (here p is the dimension of the output vector process). It is thus interesting to ask whether Z can be chosen so that T has *only* p positive eigenvalues and no negative eigenvalues, *i.e.*, if Z can be chosen so that T has minimal rank p.

Suppose that we can make such a choice, which we designate by $Z = P$. By repeating the argument of Sec. 8.3.3, it can be easily verified that P can be chosen as the solution of the CARE

$$0 = PF + PF^* + GQG^* - KRK^* = 0, \qquad K = (PH^* + GS)R^{-1}, \tag{8.D.11}$$

that results in a stable closed-loop matrix, $F_{cl} = F - KH$. [We shall argue further below in Thm. 8.D.1 that such a solution P always exists and that in fact it will be positive-semi-definite and unique.] In this case, the expression (8.D.6) for $S_y(s)$ collapses to

$$S_y(s) = \begin{bmatrix} H(zI - F)^{-1}K + I \end{bmatrix} R \begin{bmatrix} H(-sI - F)^{-1}K + I \end{bmatrix}^*,$$

so that if we define

$$L(s) \triangleq H(sI - F)^{-1}K + I, \tag{8.D.12}$$

then we have a factorization of $S_y(s)$,

$$S_y(s) = L(s)RL^*(-s^*), \tag{8.D.13}$$

and there will be a different factorization for every solution P of the Riccati equation (8.D.11). Now since F is stable, every transfer matrix $L(s)$ will have all its poles inside the open left-half plane, so that $L(s)$ will be stable and causal. Moreover, since $F - KH$ is stable, and

$$[L(s)]^{-1} = I - H(sI - F + KH)^{-1}K, \tag{8.D.14}$$

then the inverse of $L(s)$ is also stable and causal. This means that the above $L(s)$ defines the canonical factor of $S_y(z)$.

The equation (8.D.11) is the Continuous-Time Algebraic Riccati Equation (CARE). This equation is studied in App. E.9, and the following statement is a special case of the general Thm. E.9.2.

Theorem 8.D.1 (The CARE) *Assume that F is stable, $\{F^s, GQ^{s/2}\}$ is controllable on the imaginary axis, $R > 0$, and*

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \geq 0.$$

Under these conditions, the continuous-time algebraic Riccati equation (CARE),

$$0 = FP + PF^* + GQG^* - KRK^*, \tag{8.D.15}$$

where $K = (PH^* + GS)R^{-1}$, has a unique solution P such that $F - KH$ is stable. Moreover, this so-called stabilizing solution is positive-semi-definite. ∎

In the above theorem, the stability of F is required since we are dealing with stationary processes, whereas the assumption of controllability on the imaginary axis and the condition on the variances $\{R, Q, S\}$ are required to guarantee the positivity of the s-spectrum. In Thm. E.9.2 we show that the stability requirement on F can be relaxed (and replaced by the weaker assumption that $\{F, H\}$ is detectable), but we do not need this generality here.

We can now state a major result of this appendix, showing how the introduction of state-space structure gives a specific procedure for (canonical) spectral factorization in the vector case.

Theorem 8.D.2 (Spectral Factorization) *Consider the s-spectrum $S_y(s)$ (8.D.6), with $\{F, G, H, Q, S, R\}$ satisfying the conditions stated in Thm. 8.D.1. Then its canonical spectral factorization,*

$$S_y(s) = L(s)RL^*(-s^*), \quad L(\infty) = I, \quad R > 0,$$

is given by $L(s)$ in (8.D.12), and where P is the unique positive-semi-definite solution to the CARE (8.D.15). Moreover, $F - KH$ is stable, which in addition to the stability of F, will guarantee that $L(s)$ is minimum phase (i.e., both it and its inverse will be analytic in $Re(s) \geq 0$). ∎

The following result is now an immediate consequence of the discussions so far in the appendix.

Lemma 8.D.1 (Spectral Factorization and Riccati Equations) *Assume F is stable, $R > 0$,*

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \geq 0,$$

and consider the s-spectrum $S_y(s)$ defined below,

$$S_y(z) = \begin{bmatrix} H(sI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix} \begin{bmatrix} (-sI - F^*)^{-1}H^* \\ I \end{bmatrix}.$$

Then the following two statements are equivalent:

(i) $S_y(j\omega) > 0$ for all $\omega \in (-\infty, \infty)$.

(ii) There exists a unique nonnegative-definite solution P of the CARE,

$$0 = FP + PF^* + GQG^* - KRK^*,$$

such that $F - KH$ is stable, where $K = (PH^* + GS)R^{-1}$. ∎

Proof: The argument is similar to that of Lemma 8.3.3. ◆

Remark. Let us remark here that although we have taken F *stable* as our standing assumption in these discussions (otherwise we cannot speak of the output process of (8.D.1)–(8.D.2) as being stationary), many of the results do not require the stability of F; it can be replaced by the condition that $\{F, H\}$ is detectable. Moreover, in the general case, $S_y(s)$ as defined above is not a true s-spectrum but can be called a Popov function (as was explained in the discrete-time case in App. 8.A). ◆

Lemma 8.D.1 highlights an important connection between s-spectra that are positive definite on the imaginary axis and the existence of stabilizing solutions of CAREs. As in the discrete-time case, the lemma is also a special case of a more fundamental result in system theory that holds under considerably weaker assumptions.

Lemma 8.D.2 (The KYP Lemma) *Consider a detectable pair $\{F, H\}$ and assume F does not have eigenvalues on the imaginary axis, with $F \in \mathbb{C}^{n\times n}$ and $H \in \mathbb{C}^{p\times n}$. Consider also arbitrary matrices $\{Q, S, R\}$ of appropriate dimensions and define the Popov function*

$$S_y(s) = \begin{bmatrix} H(sI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} (-sI - F^*)^{-1}H^* \\ I \end{bmatrix},$$

where the central matrix is Hermitian but may be indefinite. The following three statements are equivalent:

1. *$S_y(j\omega) \geq 0$ for all $\omega \in (-\infty, \infty)$.*
2. *There exists a Hermitian matrix P such that*

$$\begin{bmatrix} Q + FP + PF^* & S + PH^* \\ S^* + HP & R \end{bmatrix} \geq 0.$$

3. *There exist an $n \times n$ Hermitian matrix P and an $n \times p$ matrix K such that*

$$\begin{bmatrix} Q + FP + PF^* & S + PH^* \\ S^* + HP & R \end{bmatrix} = \begin{bmatrix} K \\ I \end{bmatrix} R \begin{bmatrix} K^* & I \end{bmatrix}.$$

The following two statements are also equivalent:

(a) $S_y(j\omega) > 0$ *for all* $\omega \in (-\infty, \infty)$.

(b) $R > 0$ *and there exists a unique Hermitian solution P of the* CARE

$$0 = FP + PF^* + Q - KRK^*, \quad K = (PH^* + S)R^{-1},$$

such that $F - KH$ is stable.

■

Proof: The directions (3)⇒(2)⇒(1) and (b)⇒(a) are immediate to verify. The other directions require more effort and they would follow by first establishing that the nonnegativity of $S_y(j\omega)$ guarantees the existence of a Hermitian solution to a CARE. ◆

The equivalence between Popov functions and CAREs is also useful in characterizing rational positive-real systems. These are stable systems with square transfer matrix functions, say $G(s)$, that satisfy

$$G(j\omega) + G^*(j\omega) \geq 0 \quad \text{for all } \omega \in (-\infty, \infty). \tag{8.D.16}$$

Clearly, if $G(s)$ is positive-real, then $G(s)+G^*(-s^*)$ is a Popov function that is nonnegative definite on the imaginary axis. Thus by applying the KYP lemma to $G(s)+G^*(-s^*)$, we obtain the following result.

Lemma 8.D.3 (Positive-Real Lemma) *Let $G(s) = D+H(sI-F)^{-1}\bar{N}$ be a given state-space realization of a stable rational matrix function, with $D \in \mathbb{C}^{p\times p}$ and $F \in \mathbb{C}^{n\times n}$. The following three statements are equivalent:*

1. *$G(s)$ is positive-real.*
2. *There exists an $n \times n$ matrix $\bar{\Sigma} \geq 0$ such that*

$$\begin{bmatrix} -F\bar{\Sigma} - \bar{\Sigma}F^* & \bar{N} - \bar{\Sigma}H^* \\ \bar{N}^* - H\bar{\Sigma} & D + D^* \end{bmatrix} \geq 0.$$

3. *There exist an $n \times n$ matrix $\bar{\Sigma} \geq 0$, a $p \times p$ matrix $R \geq 0$, and an $n \times p$ matrix K, such that*

$$\begin{bmatrix} -F\bar{\Sigma} - \bar{\Sigma}F^* & \bar{N} - \bar{\Sigma}H^* \\ \bar{N}^* - H\bar{\Sigma} & D + D^* \end{bmatrix} = \begin{bmatrix} K \\ I \end{bmatrix} R \begin{bmatrix} K^* & I \end{bmatrix}.$$

The following two statements are also equivalent:

(a) *$G(s)$ is strictly positive-real, i.e., $G(j\omega) + G^*(j\omega) > 0$ for all $\omega \in (-\infty, \infty)$.*

(b) *$R = D + D^* > 0$ and there exists a unique nonnegative-definite solution $\bar{\Sigma}$ of the* CARE,

$$0 = F\bar{\Sigma} + \bar{\Sigma}F^* + KRK^*, \quad K = (\bar{N} - \bar{\Sigma}H^*)R^{-1},$$

such that $F - KH$ is stable.

■

Proof: Once more, the directions (3)⇒(2)⇒(1) and (b)⇒(a) are immediate. The other directions follow from the KYP lemma. ◆

Another useful consequence of the KYP lemma is the Bounded-Real lemma, which characterizes rational bounded functions. These are stable (possibly nonsquare) systems with matrix functions $G(s)$ that satisfy

$$\sup_{\omega\in(-\infty,\infty)} \|G(j\omega)\| \leq \gamma, \tag{8.D.17}$$

for some positive γ and for some matrix norm $\|\cdot\|$, say the induced 2-norm (or maximum singular value of its argument). Here again, if (8.D.17) holds, then $\gamma^2 I - G(s)G^*(-s^*)$ is a Popov function that is nonnegative on the imaginary axis.

Now by applying the KYP lemma to $\gamma^2 I - G(s)G^*(-s^*)$ (or to $\gamma^2 I - G^*(-s^*)G(s)$), we obtain the following result.

Lemma 8.D.4 (Bounded Real Lemma) *Let $G(s) = D + H(sI - F)^{-1}B$ denote a stable rational matrix function with $D \in \mathbb{C}^{p\times p}$ and $F \in \mathbb{C}^{n\times n}$. The following three statements are equivalent:*

1. *$\sup_{\omega\in(-\infty,\infty)} \|G(j\omega)\| \leq \gamma$.*
2. *There exists an $n \times n$ matrix $\bar{\Sigma} \geq 0$ such that*

$$\begin{bmatrix} -F\bar{\Sigma} - \bar{\Sigma}F^* - BB^* & -BD^* - \bar{\Sigma}H^* \\ -DB^* - H\bar{\Sigma} & \gamma^2 I - DD^* \end{bmatrix} \geq 0.$$

3. *There exist an $n \times n$ matrix $\bar{\Sigma} \geq 0$, a $p \times p$ matrix $R \geq 0$, and an $n \times p$ matrix K, such that*

$$\begin{bmatrix} -F\bar{\Sigma} - \bar{\Sigma}F^* - BB^* & -BD^* - \bar{\Sigma}H^* \\ -DB^* - H\bar{\Sigma} & \gamma^2 I - DD^* \end{bmatrix} = \begin{bmatrix} K \\ I \end{bmatrix} R \begin{bmatrix} K^* & I \end{bmatrix}.$$

The following two statements are also equivalent:

(a) *$\sup_{\omega\in(-\infty,\infty)} \|G(j\omega)\| < \gamma$.*

(b) *$R = \gamma^2 I - DD^* > 0$ and there exists a unique nonnegative-definite solution $\bar{\Sigma}$ of the* CARE,

$$0 = F\bar{\Sigma} + \bar{\Sigma}F^* + BB^* + KRK^*, \quad K = -(BD^* + \bar{\Sigma}H^*)R^{-1},$$

such that $F - KH$ is stable.

■

CHAPTER 9

THE KALMAN FILTER

In this chapter we shall show how the assumption of a finite-dimensional state-space model for a process allows the innovations to be recursively and efficiently computed, with $O(Nn^3)$ computations as opposed to $O(N^3)$, where n is the state dimension and N is the number of observations. There are many problems, especially in aerospace applications, where the state variables have a direct physical significance and where estimates of the state variables, or of some linear combinations of these variables, are needed. As noted earlier, once we have the innovations, the estimation of related quantities (states, inputs, and linear combinations thereof) is straightforward.

9.1 THE STANDARD STATE-SPACE MODEL

As often mentioned earlier, since the early sixties, much effort has been devoted to modeling processes $\{\mathbf{y}_i\}$ in state-space form, *i.e.*,

$$\mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i, \quad i \geq 0, \tag{9.1.1}$$

where the $n \times 1$ state-vector $\mathbf{x}_i$ obeys the recursion

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i, \quad i \geq 0. \tag{9.1.2}$$

The processes $\mathbf{v}_i$ and $\mathbf{u}_i$ are assumed to be $p \times 1$ and $m \times 1$ vector-valued zero-mean white-noise processes, with[1]

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i & S_i \\ S_i^* & R_i \end{bmatrix} \delta_{ij}, \tag{9.1.3}$$

[1] In this chapter, we shall consistently use the inner product notation $E\mathbf{a}\mathbf{b}^* \triangleq \langle \mathbf{a}, \mathbf{b} \rangle$ (see Chs. 3 and 4).

while the initial state $\mathbf{x}_0$ is assumed to have zero mean, covariance matrix Π_0, and to be uncorrelated with the $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$. These assumptions can be compactly stated as

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i\delta_{ij} & S_i\delta_{ij} & 0 & 0 \\ S_i^*\delta_{ij} & R_i\delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}. \tag{9.1.4}$$

It is also assumed that the matrices F_i (of dimension $n \times n$), G_i $(n \times m)$, H_i $(p \times n)$, Q_i $(m \times m)$, R_i $(p \times p)$, S_i $(m \times p)$, and Π_0 $(n \times n)$ are *known a priori*. The process $\mathbf{v}_i$ is often called *measurement noise* and the process $\mathbf{u}_i$, *plant or process noise*. They are often uncorrelated (*i.e.*, $S_i = 0$), but the more general assumption is necessary to handle problems where there may be feedback from the output to the states.

We shall not discuss here how the state equations have been obtained. In many situations, the definitions of the state variables are naturally suggested by the physical problem; linearization may often have to be used to actually obtain linear equations as in (9.1.1)–(9.1.2). As a result, the state-space model can be set up in slightly different forms, *e.g.*, with different assumptions on the correlation between $\{\mathbf{u}_i, \mathbf{v}_i\}$. These models can be analyzed in ways quite similar to the ones we are going to describe here, and therefore some of these variations will be left to the exercises. The model specified above will be henceforth called the *standard* model.

The following properties of the model (9.1.1)–(9.1.2) were derived in Ch. 5, but they are readily established anew, as we encourage readers to do; the geometric interpretation of random variables will be extensively and effectively used in this chapter, so Sec. 3.3 may also profitably be reviewed at this point.

For the moment it will be enough to be thoroughly familiar with the following covariance properties of the model.

1. **Uncorrelatedness Properties.** The assumptions (9.1.1)–(9.1.4) imply that $\{\mathbf{u}_i, \mathbf{v}_i\}$ are uncorrelated with all *past* or *present states*, *i.e.*,

$$\langle \mathbf{u}_i, \mathbf{x}_j \rangle = 0, \quad \langle \mathbf{v}_i, \mathbf{x}_j \rangle = 0, \quad j \leq i,$$

which we can (and often shall) write as

$$\mathbf{u}_i \perp \mathbf{x}_j, \quad \mathbf{v}_i \perp \mathbf{x}_j, \quad j \leq i. \tag{9.1.5}$$

The reason is that, according to (9.1.2), $\mathbf{x}_j$ depends linearly only upon the random variables $\{\mathbf{x}_0, \mathbf{u}_k, k \leq j-1\}$ (or briefly, $\mathbf{x}_j \in \mathcal{L}\{\mathbf{x}_0, \mathbf{u}_k, k \leq j-1\}$), and by assumption $\mathbf{u}_i$ is uncorrelated with or orthogonal to these random variables, and so is $\mathbf{v}_i$.[2] For the same reason, we can see that $\{\mathbf{u}_i, \mathbf{v}_i\}$ are orthogonal to *past outputs*, *i.e.*,

$$\mathbf{u}_i \perp \mathbf{y}_j, \quad \mathbf{v}_i \perp \mathbf{y}_j, \quad j \leq i-1. \tag{9.1.6}$$

[2] The notation $\mathbf{x}_j \in \mathcal{L}\{\mathbf{x}_0, \mathbf{u}_k, k \leq i-1\}$ means that $\mathbf{x}_j$ can be expressed as a linear combination of the variables $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}\}$. Now since $\mathbf{x}_0$ and the $\{\mathbf{u}_k\}$ generally have different dimensions, the coefficients of the linear combination are to be understood as matrices of appropriate dimensions. Ultimately, $\mathcal{L}\{\mathbf{x}_0, \mathbf{u}_k, k \leq i-1\}$ is the linear space formed by all the scalar random variables in $\mathbf{x}_0$ and $\{\mathbf{u}_k, k \leq i-1\}$.

However, for the *present output*, we have

$$\langle \mathbf{u}_i, \mathbf{y}_i \rangle = \langle \mathbf{u}_i, \mathbf{x}_i \rangle H_i^* + \langle \mathbf{u}_i, \mathbf{v}_i \rangle = 0 + S_i, \tag{9.1.7}$$

$$\langle \mathbf{v}_i, \mathbf{y}_i \rangle = \langle \mathbf{v}_i, \mathbf{x}_i \rangle H_i^* + \langle \mathbf{v}_i, \mathbf{v}_i \rangle = 0 + R_i. \tag{9.1.8}$$

2. Covariance Recursion. Let us define

$$\Pi_i \triangleq \langle \mathbf{x}_i, \mathbf{x}_i \rangle, \quad \text{the state covariance matrix.}$$

Then it is easy to check that

$$\begin{aligned} \Pi_{i+1} &= F_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle F_i^* + F_i \langle \mathbf{x}_i, \mathbf{u}_i \rangle G_i^* + G_i \langle \mathbf{u}_i, \mathbf{x}_i \rangle F_i^* + G_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle G_i^* \\ &= F_i \Pi_i F_i^* + G_i Q_i G_i^*, \quad i \geq 0, \end{aligned} \tag{9.1.9}$$

with initial value Π_0.

9.2 THE KALMAN FILTER RECURSIONS FOR THE INNOVATIONS

Now we go on to the problem of whether we can conveniently find the innovations,

$$\mathbf{e}_i \triangleq \mathbf{y}_i - \hat{\mathbf{y}}_{i|i-1},$$

when the $\{\mathbf{y}_i\}$ have state-space structure. It turns out that the recursive construction of the innovations combines nicely with the recursive evolution of the state variables to give a recursion for the innovations in terms of the parameters of the model and a pair of other matrices $\{K_{p,i}, R_{e,i}\}$. These can be computed in different ways, one of which we shall present in this section; others will be described in Chs. 11, 12, and 13.

9.2.1 Recursions for the Innovations

Starting with $\mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i$, and projecting onto the linear subspace spanned by $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$ yields

$$\hat{\mathbf{y}}_{i|i-1} = H_i \hat{\mathbf{x}}_{i|i-1} + \hat{\mathbf{v}}_{i|i-1}. \tag{9.2.1}$$

Our standard notational convention is that

$$\hat{\mathbf{x}}_{i|j} = \text{the projection of } \mathbf{x}_i \text{ on the linear subspace spanned by } \{\mathbf{y}_0, \ldots, \mathbf{y}_j\},$$

where the linear subspace is often denoted by $\mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_j\}$. Now as we noted at the end of Section 9.1, the assumptions on our state-space model imply that $\mathbf{v}_i \perp \mathbf{y}_j$ for $j \leq i-1$, so that $\hat{\mathbf{v}}_{i|i-1} = 0$ and

$$\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_{i|i-1} = \mathbf{y}_i - H_i \hat{\mathbf{x}}_{i|i-1}. \tag{9.2.2}$$

Therefore, we see that the problem of finding the innovations reduces to one of finding a convenient way of determining the one-step predictions of the state vector. For this purpose, we can try to use the basic formula for estimation given the (uncorrelated) innovations process,

$$\hat{\mathbf{x}}_{i+1|i} = \sum_{j=0}^{i} \langle \mathbf{x}_{i+1}, \mathbf{e}_j \rangle R_{e,j}^{-1} \mathbf{e}_j. \tag{9.2.3}$$

The major (invertibility) assumption made here is that $R_{e,i} > 0$, which corresponds to a *nondegeneracy assumption* on the process $\{\mathbf{y}_i\}$, *viz.*, that no variable $\mathbf{y}_i$ can be estimated without error by some linear combination of earlier variables. This assumption is independent of whether the matrices $R_i = \langle \mathbf{v}_i, \mathbf{v}_i \rangle$ are nonsingular or not; in fact, we could have $R_{e,i} > 0$ even if $R_i = 0$. However, it is usually a good modeling decision to assume that $R_i > 0$, which will ensure $R_{e,i} > 0$. In addition, several other simplifications occur in the $R_i > 0$ case and these are discussed at some length in Sec. 9.5.

Now given that the $\{R_{e,i}\}$ can be assumed invertible, the formula (9.2.3) seems puzzling (in fact, circular), because so far we have only defined the innovations $\{\mathbf{e}_i\}$ in terms of the one-step predictions, which are the variables we are trying to estimate. The reason (9.2.3) can make sense is that on the right-hand-side we have the quantities

$$\mathbf{e}_j = \mathbf{y}_j - H_j \hat{\mathbf{x}}_{j|j-1} \quad \text{for} \quad j \leq i,$$

so that in trying to find $\hat{\mathbf{x}}_{i+1|i}$ from (9.2.2), we are only using *earlier* one-step predictions $\{\hat{\mathbf{x}}_{j|j-1}, j \leq i\}$. This suggests that what we should try to find is a *recursive* solution, with the present value $\hat{\mathbf{x}}_{i+1|i}$ being computed from the most recent past value $\hat{\mathbf{x}}_{i|i-1}$ and the new information $\mathbf{e}_i = \mathbf{y}_i - H_i \hat{\mathbf{x}}_{i|i-1}$.

To see if this is possible, we first rewrite (9.2.3) in a form more indicative of a recursion

$$\begin{aligned} \hat{\mathbf{x}}_{i+1|i} &= \left(\sum_{j=0}^{i-1} \langle \mathbf{x}_{i+1}, \mathbf{e}_j \rangle R_{e,j}^{-1} \mathbf{e}_j \right) + \langle \mathbf{x}_{i+1}, \mathbf{e}_i \rangle R_{e,i}^{-1} \mathbf{e}_i, \\ &= \hat{\mathbf{x}}_{i+1|i-1} + \langle \mathbf{x}_{i+1}, \mathbf{e}_i \rangle R_{e,i}^{-1} (\mathbf{y}_i - H_i \hat{\mathbf{x}}_{i|i-1}). \end{aligned} \tag{9.2.4}$$

This is almost in the desired form, and would be exactly so if the term $\hat{\mathbf{x}}_{i+1|i-1}$ could be expressed in terms of just $\hat{\mathbf{x}}_{i|i-1}$ and $\mathbf{e}_i$. At this point, no more general statements can be made; to go further we must have more information about the way the states change with time.

In our problem we know that $\mathbf{x}_{i+1}$ obeys the state equation $\mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i$. But then projecting onto the linear subspace spanned by $\{\mathbf{y}_j, j \leq i-1\}$ shows that

$$\hat{\mathbf{x}}_{i+1|i-1} = F_i \hat{\mathbf{x}}_{i|i-1} + G_i \hat{\mathbf{u}}_{i|i-1} = F_i \hat{\mathbf{x}}_{i|i-1} + 0, \tag{9.2.5}$$

since by the assumptions on our model, $\mathbf{u}_i \perp \mathbf{y}_j$, $j \leq i-1$. But a relation as in (9.2.5) is exactly what we were seeking. In other words, by combining Eqs. (9.2.2)–(9.2.5) we have the following *recursive* set of equations for determining the innovations:

$$\mathbf{e}_i = \mathbf{y}_i - H_i \hat{\mathbf{x}}_{i|i-1}, \qquad \mathbf{e}_0 = \mathbf{y}_0, \tag{9.2.6}$$

$$\hat{\mathbf{x}}_{i+1|i} = F_i \hat{\mathbf{x}}_{i|i-1} + K_{p,i} \mathbf{e}_i, \quad i \geq 0, \tag{9.2.7}$$

where

$$K_{p,i} \triangleq \langle \mathbf{x}_{i+1}, \mathbf{e}_i \rangle R_{e,i}^{-1}. \tag{9.2.8}$$

The subscript "p" indicates that $K_{p,i}$ is used to update a *predicted* estimator.

The $\{K_{p,i}, R_{e,i}\}$ are *nonrandom* quantities that should be completely determinable from our knowledge of the means and covariances of the model, and in fact we shall show in the next section how this can be done.

We can combine (9.2.6) and (9.2.7) as

$$\hat{\mathbf{x}}_{i+1|i} = F_i\hat{\mathbf{x}}_{i|i-1} + K_{p,i}(\mathbf{y}_i - H_i\hat{\mathbf{x}}_{i|i-1}), \quad \hat{\mathbf{x}}_{0|-1} = 0, \ i \geq 0,$$

$$= F_{p,i}\hat{\mathbf{x}}_{i|i-1} + K_{p,i}\mathbf{y}_i, \quad F_{p,i} \triangleq F_i - K_{p,i}H_i, \tag{9.2.9}$$

which emphasizes that in finding the innovations, we actually also have a complete recursion for the state estimators $\{\hat{\mathbf{x}}_{i|i-1}\}$. Later we shall see that with the innovations we can readily determine other estimators such as $\hat{\mathbf{x}}_{i|i}$, $\hat{\mathbf{x}}_{i|i+1}$, $\hat{\mathbf{u}}_{i|i}$, etc.

In fact, we should reiterate that once we have the innovations $\{\mathbf{e}_i\}$ we can recursively estimate any other random variable, say $\mathbf{z}$, by the formula

$$\hat{\mathbf{z}}_{|i} = \hat{\mathbf{z}}_{|i-1} + \langle \mathbf{z}, \mathbf{e}_i\rangle R_{e,i}^{-1}\mathbf{e}_i, \tag{9.2.10}$$

where $\hat{\mathbf{z}}_{|i}$ denotes the l.l.m.s. estimator of $\mathbf{z}$ given the observations $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\}$.

Of course $\mathbf{z}$ can depend on time as well, so that it is a random process itself, say $\{\mathbf{z}_i\}$. But unless $\{\mathbf{z}_i\}$ varies in some known and suitable way with i, we cannot hope to get convenient recursions for the estimators $\{\hat{\mathbf{z}}_{i|i-1}\}$. The nicest case of course is when $\mathbf{z}_i$ is the state vector $\mathbf{x}_i$, where the preceding formula (9.2.9) can be applied directly. The next best situation is when the desired variables $\{\mathbf{z}_i\}$ are *known* linear combinations of the $\{\mathbf{x}_i\}$, say

$$\mathbf{z}_i = L_i\mathbf{x}_i, \quad \text{where } L_i \text{ are known matrices.} \tag{9.2.11}$$

By noting that

$$\hat{\mathbf{z}}_{i|i-1} = L_i\hat{\mathbf{x}}_{i|i-1}, \tag{9.2.12}$$

we can use the recursions (9.2.9) for the state estimators and then obtain the $\{\hat{\mathbf{z}}_{i|i-1}\}$ by multiplication by the known matrices $\{L_i\}$.

9.2.2 $R_{e,i}$ and $K_{p,i}$ in Terms of P_i

To complete the computation of the innovations, let us describe one way of computing the coefficients $\{K_{p,i}, R_{e,i}\}$ needed for the basic recursions (9.2.6)–(9.2.7). The formulas we shall present here were first explicitly given by Kalman (1960). Some important alternative methods (the so-called array and fast array methods) for computing $\{K_{p,i}, R_{e,i}\}$ will be presented in Chs. 11, 12, and 13.

Kalman began by introducing the quantity

$$P_{i|i-1} \triangleq \langle \tilde{\mathbf{x}}_{i|i-1}, \tilde{\mathbf{x}}_{i|i-1}\rangle, \quad \tilde{\mathbf{x}}_{i|i-1} \triangleq \mathbf{x}_i - \hat{\mathbf{x}}_{i|i-1}, \tag{9.2.13}$$

which is of course of independent interest as the variance matrix of the error in the predicted state estimator, and noting that the quantities $\{K_{p,i}, R_{e,i}\}$ in the basic recursions (9.2.6)–(9.2.8) could be expressed in terms of the $\{P_{i|i-1}\}$. It remains only to specify the $\{P_{i|i-1}\}$ in terms of the model parameters, and he showed that they could be computed via a discrete-time Riccati recursion,

$$P_{i+1|i} = F_iP_{i|i-1}F_i^* + G_iQ_iG_i^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad i \geq 0, \tag{9.2.14}$$

with initial condition

$$P_{0|-1} \triangleq \langle \tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_0\rangle = \langle (\mathbf{x}_0 - \hat{\mathbf{x}}_0), (\mathbf{x}_0 - \hat{\mathbf{x}}_0)\rangle = \langle \mathbf{x}_0, \mathbf{x}_0\rangle = \Pi_0. \tag{9.2.15}$$

The recursion was so named as the discrete-time analog of a famous quadratic nonlinear differential equation attributed to Jacopo Francesco, Count Riccati (ca. 1700), and first ingeniously exploited in the calculus of variations by A. M. Legendre (ca. 1786). It was reintroduced into control theory by Bellman (1957), and then in general matrix form by Kalman (1960a).

Important Remark on Notation. Since one-step predicted quantities will be encountered often, we shall use the following briefer notations (except when necessary for emphasis, or for some other special reason):

$$\hat{\mathbf{x}}_i \triangleq \hat{\mathbf{x}}_{i|i-1}, \quad \tilde{\mathbf{x}}_i \triangleq \tilde{\mathbf{x}}_{i|i-1}, \quad P_i \triangleq P_{i|i-1}. \tag{9.2.16}$$

◆

Now to see how P_i enters into the computation of $\{K_{p,i}, R_{e,i}\}$, note first that since

$$\mathbf{e}_i = \mathbf{y}_i - H_i\hat{\mathbf{x}}_i = H_i\mathbf{x}_i - H_i\hat{\mathbf{x}}_i + \mathbf{v}_i = H_i\tilde{\mathbf{x}}_i + \mathbf{v}_i, \tag{9.2.17}$$

the uncorrelatedness property noted in Sec. 9.1, $\mathbf{v}_i \perp \tilde{\mathbf{x}}_i$, shows that we can express the covariance matrix of $\mathbf{e}_i$ in terms of P_i,

$$R_{e,i} \triangleq \langle \mathbf{e}_i, \mathbf{e}_i\rangle = H_iP_iH_i^* + R_i. \tag{9.2.18}$$

It turns out that this is also true of $K_{p,i}$. For we have

$$\langle \mathbf{x}_{i+1}, \mathbf{e}_i\rangle = F_i\langle \mathbf{x}_i, \mathbf{e}_i\rangle + G_i\langle \mathbf{u}_i, \mathbf{e}_i\rangle. \tag{9.2.19}$$

Now

$$\langle \mathbf{x}_i, \mathbf{e}_i\rangle = \langle \mathbf{x}_i, \tilde{\mathbf{x}}_i\rangle H_i^* + \langle \mathbf{x}_i, \mathbf{v}_i\rangle = P_iH_i^* + 0, \tag{9.2.20}$$

where we have used the fact that $\langle \mathbf{x}_i, \tilde{\mathbf{x}}_i\rangle = \langle \hat{\mathbf{x}}_i + \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i\rangle = 0 + P_i$. Next, we compute

$$\langle \mathbf{u}_i, \mathbf{e}_i\rangle = \langle \mathbf{u}_i, \tilde{\mathbf{x}}_i\rangle H_i^* + \langle \mathbf{u}_i, \mathbf{v}_i\rangle = 0 + S_i, \tag{9.2.21}$$

where we have used the fact that

$$\tilde{\mathbf{x}}_i \in \mathcal{L}\{\mathbf{x}_i; \mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\} \subset \mathcal{L}\{\mathbf{x}_0; \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}; \mathbf{v}_0, \ldots, \mathbf{v}_{i-1}\} \perp \mathbf{u}_i.$$

Therefore

$$K_{p,i} \triangleq \langle \mathbf{x}_{i+1}, \mathbf{e}_i\rangle R_{e,i}^{-1} = (F_iP_iH_i^* + G_iS_i)R_{e,i}^{-1}, \tag{9.2.22}$$

so we see that $\{K_{p,i}, R_{e,i}\}$ can be determined once we have the error covariance matrices $\{P_i\}$. These, we shall show soon, can be computed successively via the previously mentioned *discrete-time Riccati recursion* (9.2.14).

Remark 1. It is important to note that the quantities $\{P_i, K_{p,i}, R_{e,i}\}$ depend only upon the prior assumptions on the model and not on the actual observations $\{\mathbf{y}_i\}$; therefore, these quantities can be pre-computed (or computed *off-line*) and stored for use in the actual prediction calculations. However, the above formulas do allow these quantities to be updated as needed (in real time), thus eliminating the need for extensive storage. ◆

9.2.3 Recursion for P_i

We can derive the formula (9.2.14) in several different ways, of which we shall present two here.

Using the State-Estimation Error. The most direct method is to seek a recursion for $\tilde{\mathbf{x}}_{i+1}$ and then form P_{i+1}. In fact, from the model equations $\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i$, $\mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i$, and the estimator equation

$$\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_i + K_{p,i}\mathbf{e}_i = F_i\hat{\mathbf{x}}_i + K_{p,i}(H_i\tilde{\mathbf{x}}_i + \mathbf{v}_i),$$

we can write

$$\begin{aligned}\tilde{\mathbf{x}}_{i+1} &= F_i\tilde{\mathbf{x}}_i + G_i\mathbf{u}_i - K_{p,i}H_i\tilde{\mathbf{x}}_i - K_{p,i}\mathbf{v}_i,\\ &= F_{p,i}\tilde{\mathbf{x}}_i + \begin{bmatrix}G_i & -K_{p,i}\end{bmatrix}\begin{bmatrix}\mathbf{u}_i\\ \mathbf{v}_i\end{bmatrix}, \quad \text{say,}\end{aligned} \tag{9.2.23}$$

where we have defined $F_{p,i} = F_i - K_{p,i}H_i$. Now it is easy to calculate from (9.2.23) that the covariance matrix obeys the recursion

$$P_{i+1} = F_{p,i}P_iF_{p,i}^* + \begin{bmatrix}G_i & -K_{p,i}\end{bmatrix}\begin{bmatrix}Q_i & S_i\\ S_i^* & R_i\end{bmatrix}\begin{bmatrix}G_i^*\\ -K_{p,i}^*\end{bmatrix}, \tag{9.2.24}$$

which the reader should check reduces, after some algebra, to the equation (9.2.14). The initial condition (9.2.15) follows from the fact that $\tilde{\mathbf{x}}_0 = \mathbf{x}_0 - \hat{\mathbf{x}}_0 = \mathbf{x}_0$, so that $P_0 = \langle\mathbf{x}_0, \mathbf{x}_0\rangle = \Pi_0$.

The active reader will see that there are some *redundant* calculations in this derivation, and we might wonder if a more direct approach is possible. In fact, the geometric viewpoint helps to get a nice proof.

Using the Difference of State and Estimator Covariance Matrices. The covariance matrix of the state vector of a white-noise driven process, $\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i$, obeys the (easily derived) recursion (*cf.* (9.1.9))

$$\Pi_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^*, \quad \Pi_i \triangleq \langle\mathbf{x}_i, \mathbf{x}_i\rangle. \tag{9.2.25}$$

Now we note that the estimator equation is also one driven by a white-noise process, namely the innovations:

$$\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_i + K_{p,i}\mathbf{e}_i, \quad \langle\mathbf{e}_i, \mathbf{e}_j\rangle \triangleq R_{e,i}\delta_{ij}.$$

Therefore, if we define the covariance matrix of the state estimators as

$$\Sigma_i \triangleq \langle\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_i\rangle, \tag{9.2.26}$$

then (as for Π_i) we can write

$$\Sigma_{i+1} = F_i\Sigma_iF_i^* + K_{p,i}R_{e,i}K_{p,i}^*, \tag{9.2.27}$$

with initial condition (recall $\hat{\mathbf{x}}_{0|-1} = 0$)

$$\Sigma_0 \triangleq \Sigma_{0|-1} = 0. \tag{9.2.28}$$

But the orthogonal decomposition $\mathbf{x}_i = \hat{\mathbf{x}}_i + \tilde{\mathbf{x}}_i$, $\hat{\mathbf{x}}_i \perp \tilde{\mathbf{x}}_i$, shows that

$$\Pi_i = \Sigma_i + P_i. \tag{9.2.29}$$

It is now immediate that

$$P_{i+1} = \Pi_{i+1} - \Sigma_{i+1} = F_i(\Pi_i - \Sigma_i)F_i^* + G_iQ_iG_i^* - K_{p,i}R_{e,i}K_{p,i}^*,$$

which is indeed the Riccati equation (9.2.14).

This is perhaps the most direct route to the Riccati recursion. The advantage of the first method of Sec. 9.2.3 is that it allows us to evaluate the error covariance matrix even when $K_{p,i}$ is not the optimal gain vector; the method of this section would not work in this case because then (9.2.29) would not hold.

9.2.4 The Kalman Filter Recursions for the Innovations

Once P_i is computed then $K_{p,i}$ and $R_{e,i}$ can be readily obtained, and thus the recursions for $\{\mathbf{e}_i\}$ completely specified. This collection of equations gives one (so-called covariance) form of the celebrated Kalman filter, which for convenience we present in the form of a Theorem 9.2.1 below.[3] As noted earlier, once we have the innovations, other quantities are readily estimated. In this way we shall obtain the Kalman filter recursions for the predicted and filtered state estimators, and later in Ch. 10 for the smoothed state estimators.

Theorem 9.2.1 (The Innovations Recursions) *Consider the state-space equations*

$$\begin{cases}\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i,\\ \mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i,\end{cases} \quad i \geq 0, \tag{9.2.30}$$

where the $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$ *are* $m \times 1$, $p \times 1$, *and* $n \times 1$*-dimensional random variables such that*

$$\left\langle\begin{bmatrix}\mathbf{u}_i\\ \mathbf{v}_i\\ \mathbf{x}_0\end{bmatrix}, \begin{bmatrix}\mathbf{u}_j\\ \mathbf{v}_j\\ \mathbf{x}_0\\ 1\end{bmatrix}\right\rangle = \begin{bmatrix}Q_i\delta_{ij} & S_i\delta_{ij} & 0 & 0\\ S_i^*\delta_{ij} & R_i\delta_{ij} & 0 & 0\\ 0 & 0 & \Pi_0 & 0\end{bmatrix}. \tag{9.2.31}$$

The matrices $\{F_i, G_i, H_i, \Pi_0, Q_i, S_i, R_i\}$ *are assumed known. Then the innovations of the process* $\{\mathbf{y}_i\}$ *can be recursively computed via the equations:*

$$\mathbf{e}_i = \mathbf{y}_i - H_i\hat{\mathbf{x}}_i, \quad \hat{\mathbf{x}}_0 = 0, \quad \mathbf{e}_0 = \mathbf{y}_0, \tag{9.2.32}$$

$$\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_i + K_{p,i}\mathbf{e}_i, \quad i \geq 0, \tag{9.2.33}$$

[3] The simple modifications for the nonzero-mean case are noted in Prob. 9.3.

where $K_{p,i} = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}$, $R_{e,i} = R_i + H_i P_i H_i^*$, *and*

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \qquad P_0 = \Pi_0.$$

We also note that $P_i \triangleq \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i \rangle$ *where* $\tilde{\mathbf{x}}_i \triangleq \mathbf{x}_i - \hat{\mathbf{x}}_i$. *When* $m \ll n$, $p \ll n$, *the number of computations required for going from* $\mathbf{e}_i$ *to* $\mathbf{e}_{i+1}$ *is* $O(n^3)$. ■

Proof: Simply collect together equations (9.2.6)–(9.2.8), (9.2.18), and (9.2.22)–(9.2.15). The number of computations is easy to check since the most expensive step is the computation of the triple product $F_i P_i F_i^*$, of $n \times n$ matrices. Now multiplying an $n \times n$ matrix by an $n \times 1$ vector takes $O(n^2)$ flops, so that multiplying two $n \times n$ matrices together takes $O(n^3)$ flops. Of course this assumes that $m \ll n$ and $p \ll n$. More exact counts in the general case can be obtained — see Tab. 12.2. ◆

9.2.5 Innovations Models for the Output Process

It is important to note that finding the innovations leads to a causal and causally invertible state-space model for the process $\{\mathbf{y}_i\}$.

Theorem 9.2.2 (An Innovations Model) *Knowing* $\{K_{p,i}, R_{e,i}\}$, *we can obtain a causal and causally invertible model for the process* $\{\mathbf{y}_i\}$ *as follows:*

$$\begin{cases} \hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_i + K_{p,i} \mathbf{e}_i, \quad \hat{\mathbf{x}}_0 = 0, \\ \mathbf{y}_i = H_i \hat{\mathbf{x}}_i + \mathbf{e}_i. \end{cases} \tag{9.2.34}$$

This is in contrast to the original causal, but not in general causally invertible — from $\{\mathbf{y}_i\}$ *to* $\{\mathbf{x}_0, \{\mathbf{u}_i\}, \{\mathbf{v}_i\}\}$ *— model (9.2.30) of Thm. 9.2.1.* ■

Proof: (9.2.34) follows from an obvious rearrangement of (9.2.33). The model is obviously causal, and to obtain $\mathbf{e}_i$ from $\mathbf{y}_i$ we just use the recursions

$$\begin{cases} \hat{\mathbf{x}}_{i+1} = F_{p,i} \hat{\mathbf{x}}_i + K_{p,i} \mathbf{y}_i, \quad \hat{\mathbf{x}}_0 = 0, \\ \mathbf{e}_i = -H_i \hat{\mathbf{x}}_i + \mathbf{y}_i. \end{cases}$$

◆

The equations (9.2.34) are said to define an *innovations representation* for the process $\{\mathbf{y}_i\}$; we say "an" because there are others – see, *e.g.*, Lemma 9.3.5 and also Prob. 9.26. Innovations models have many advantages, *e.g.*, fewer parameters are needed to specify (9.2.34) as compared to (9.2.30). Therefore if we had to identify models from input-output data, *e.g.*, from the impulse response, it is convenient to try to identify the parameters in the innovations representation — see *e.g.*, the textbook of Ljung (1987) on system identification. Other features of the innovations representation will be pursued later, especially the fact that they immediately yield efficient (*i.e.*, order $O(Nn^3)$ vs. $O(N^3)$) triangular factorizations of the covariance matrix of the observations and of its inverse – see Sec. 9.4.

9.3 RECURSIONS FOR PREDICTED AND FILTERED STATE ESTIMATORS

As noted before, one advantage of first computing the innovations is that this simplifies the process of finding estimators of other random variables given the observations $\{\mathbf{y}_i\}$.

9.3.1 The Predicted Estimators

A now trivial result that we can see immediately from the above discussion is that computing the innovations $\{\mathbf{e}_i\}$ is equivalent to computing the predicted state estimators $\hat{\mathbf{x}}_i$; all we need is a simple rearrangement of the above recursions.

Lemma 9.3.1 (The Kalman Filter Recursions for Predicted Estimators) *For the standard state-space model (see Thm. 9.2.1), the one-step predicted state estimators* $\{\hat{\mathbf{x}}_i\}$ *can be obtained via the recursions*

$$\hat{\mathbf{x}}_{i+1} = F_{p,i} \hat{\mathbf{x}}_i + K_{p,i} \mathbf{y}_i, \quad \hat{\mathbf{x}}_0 = 0, \quad F_{p,i} \triangleq F_i - K_{p,i} H_i, \tag{9.3.1}$$

where $\{K_{p,i}, R_{e,i}\}$ *are obtained as in Thm. 9.2.1. The Riccati recursion for* P_i *can also be expressed in terms of* $F_{p,i}$ *(cf. (9.2.24)),*

$$P_{i+1} = F_{p,i} P_i F_{p,i}^* + \begin{bmatrix} G_i & -K_{p,i} \end{bmatrix} \begin{bmatrix} Q_i & S_i \\ S_i^* & R_i \end{bmatrix} \begin{bmatrix} G_i^* \\ -K_{p,i}^* \end{bmatrix}. \tag{9.3.2}$$

■

Remark 2. As promised long ago in Ch. 1, this is the result that establishes that the observer structure (9.3.1) postulated in Ch. 1 is in fact *an* optimal structure. [Another (somewhat less obvious) optimal structure is (9.3.9) – *cf.* Prob. 1.4.] We remark also that we have now justified the hope expressed in Sec. 8.6 that the optimum steady-state recursion (8.4.14) would carry over to the time-variant case, largely just by adding time indices. ◆

9.3.2 Schmidt's Modification: Measurement and Time Updates

There are applications where the measurements are made at irregular, often widely spaced, intervals, *e.g.*, in tracking satellites using data from stations around the world. To address this issue, Dr. S. F. Schmidt[4] at the NASA Ames Research Center developed (in the late 1960s) a decomposition of the original problem into a *measurement-update* problem of going from the predicted estimator $\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_{i|i-1}$ to the so-called *filtered* estimator, $\hat{\mathbf{x}}_{i|i}$, and a separate *time-update* problem of going from $\hat{\mathbf{x}}_{i|i}$ to $\hat{\mathbf{x}}_{i+1}$. These problems are very easy to solve using the innovations approach, and in fact provide an alternative way of obtaining the results of Lemma 9.3.1 and Thm. 9.2.1.

Lemma 9.3.2 (Measurement Updates) *Consider the standard state-space model of Thm. 9.2.1, and suppose that we have computed* $\hat{\mathbf{x}}_{i|i-1}$, *the l.l.m.s.e. of* $\mathbf{x}_i$ *given*

[4] As mentioned in the notes in Sec. 9.9, Schmidt and his colleagues were the first to recognize the potential applicability of Kalman's results. Carrying this project through required several modifications and extensions (especially the extended Kalman filter noted in Sec. 9.7.2).

$\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$ and now get an additional measurement $\mathbf{y}_i$. We can update the estimator $\hat{\mathbf{x}}_i$ and its error P_i via the formulas

$$\hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_i + K_{f,i}\mathbf{e}_i, \tag{9.3.3}$$

$$\|\mathbf{x}_i - \hat{\mathbf{x}}_{i|i}\|^2 \triangleq P_{i|i} = P_i - K_{f,i}R_{e,i}K_{f,i}^* = P_i - P_iH_i^*R_{e,i}^{-1}H_iP_i, \tag{9.3.4}$$

where $K_{f,i} \triangleq P_iH_i^*R_{e,i}^{-1}$. ■

Proof: These formulas can be derived in several ways. The most straightforward is to start with the basic formula

$$\hat{\mathbf{x}}_{i|i} = \sum_{j=0}^{i}\langle \mathbf{x}_i, \mathbf{e}_j\rangle R_{e,j}^{-1}\mathbf{e}_j = \hat{\mathbf{x}}_{i|i-1} + \langle \mathbf{x}_i, \mathbf{e}_i\rangle R_{e,i}^{-1}\mathbf{e}_i.$$

In other words, we find the new information $\mathbf{e}_i$ in the new measurement $\mathbf{y}_i$ and use it to improve the earlier estimator, $\hat{\mathbf{x}}_i$; note that $\hat{\mathbf{x}}_{i|i} \neq \hat{\mathbf{x}}_{i|i-1} + \langle \mathbf{x}_i, \mathbf{y}_i\rangle \|\mathbf{y}_i\|^{-2}\mathbf{y}_i$.

Now

$$\langle \mathbf{x}_i, \mathbf{e}_i\rangle = \langle \mathbf{x}_i, H_i\tilde{\mathbf{x}}_i + \mathbf{v}_i\rangle = \langle \mathbf{x}_i, \tilde{\mathbf{x}}_i\rangle H_i^* + \langle \mathbf{x}_i, \mathbf{v}_i\rangle = P_iH_i^*.$$

Then defining $K_{f,i} = P_iH_i^*R_{e,i}^{-1}$ gives us the formula (9.3.3). Next, note that $\tilde{\mathbf{x}}_{i|i} = \mathbf{x}_i - \hat{\mathbf{x}}_i - K_{f,i}\mathbf{e}_i = \tilde{\mathbf{x}}_i - K_{f,i}\mathbf{e}_i$. Therefore,

$$\|\tilde{\mathbf{x}}_{i|i}\|^2 = \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i\rangle - K_{f,i}\langle \mathbf{e}_i, \tilde{\mathbf{x}}_i\rangle - \langle \tilde{\mathbf{x}}_i, \mathbf{e}_i\rangle K_{f,i}^* + K_{f,i}\langle \mathbf{e}_i, \mathbf{e}_i\rangle K_{f,i}^*.$$

Now

$$\langle \mathbf{e}_i, \tilde{\mathbf{x}}_i\rangle = H_i\langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i\rangle + \langle \mathbf{v}_i, \tilde{\mathbf{x}}_i\rangle = H_iP_i + 0, \tag{9.3.5}$$

because $\tilde{\mathbf{x}}_i \in \mathcal{L}\{\mathbf{x}_0; \mathbf{u}_0, \ldots \mathbf{u}_{i-1}; \mathbf{v}_0, \ldots, \mathbf{v}_{i-1}\}$ and $\mathbf{v}_i$ is clearly orthogonal to this linear subspace. Therefore

$$K_{f,i}\langle \mathbf{e}_i, \tilde{\mathbf{x}}_i\rangle = P_iH_i^*R_{e,i}^{-1}H_iP_i = \langle \tilde{\mathbf{x}}_i, \mathbf{e}_i\rangle K_{f,i}^*.$$

Substituting into (9.3.5) gives the formulas (9.3.4). Alternatively, and in fact, somewhat more directly:

$$\begin{aligned}\|\tilde{\mathbf{x}}_{i|i}\|^2 &= \langle \tilde{\mathbf{x}}_{i|i}, \mathbf{x}_i - \hat{\mathbf{x}}_{i|i}\rangle = \langle \tilde{\mathbf{x}}_{i|i}, \mathbf{x}_i\rangle,\\ &= \langle \tilde{\mathbf{x}}_i, \mathbf{x}_i\rangle - K_{f,i}\langle \mathbf{e}_i, \mathbf{x}_i\rangle,\\ &= P_i - K_{f,i}[H_i\langle \tilde{\mathbf{x}}_i, \mathbf{x}_i\rangle + 0],\\ &= P_i - K_{f,i}H_iP_i = P_i - P_iH_i^*R_{e,i}^{-1}H_iP_i.\end{aligned}$$

◆

Remark 3. It is often useful to note that $P_{i|i} = P_i - P_iH_i^*R_{e,i}^{-1}H_iP_i$ is the Schur complement of $R_{e,i}$ in the matrix (see Prob. 9.16 and App. A),

$$\begin{bmatrix} R_{e,i} & H_iP_i \\ P_iH_i^* & P_i \end{bmatrix}. \tag{9.3.6}$$

◆

Lemma 9.3.3 (Time Updates) *Consider the state-space model of Thm. 9.2.1, and suppose that we have computed $\{\hat{\mathbf{x}}_{i|i}, P_i\}$ and without any further measurements wish to find $\hat{\mathbf{x}}_{i+1}$ and P_{i+1}. This can be done via the formulas*

$$\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_{i|i} + G_i\hat{\mathbf{u}}_{i|i}, \qquad \hat{\mathbf{u}}_{i|i} = S_iR_{e,i}^{-1}\mathbf{e}_i, \tag{9.3.7}$$

$$P_{i+1} = F_iP_{i|i}F_i^* + G_i(Q_i - S_iR_{e,i}^{-1}S_i^*)G_i^* - F_iK_{f,i}S_i^*G_i^* - G_iS_iK_{f,i}^*F_i^*. \tag{9.3.8}$$

[Note that the formulas are much simpler when $S_i = 0$.] ■

Proof: The state equation $\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i$, allows us to write $\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_{i|i} + G_i\hat{\mathbf{u}}_{i|i}$. Now, we again use the basic innovations formula to find $\hat{\mathbf{u}}_{i|i}$,

$$\hat{\mathbf{u}}_{i|i} = \sum_{j=0}^{i}\langle \mathbf{u}_i, \mathbf{e}_j\rangle R_{e,j}^{-1}\mathbf{e}_j = 0 + \langle \mathbf{u}_i, \mathbf{e}_i\rangle R_{e,i}^{-1}\mathbf{e}_i,$$

because $\mathbf{e}_j \in \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_j\}$ and $\mathbf{u}_i$ is orthogonal to all such subspaces as long as $j \leq i-1$. Now

$$\langle \mathbf{u}_i, \mathbf{e}_i\rangle = \langle \mathbf{u}_i, \mathbf{v}_i + H_i\tilde{\mathbf{x}}_i\rangle = S_i + 0,$$

since $\tilde{\mathbf{x}}_i \in \mathcal{L}\{\mathbf{x}_0; \mathbf{u}_0, \ldots, \mathbf{u}_{i-1}; \mathbf{v}_0, \ldots, \mathbf{v}_{i-1}\}$ and $\mathbf{u}_i$ is orthogonal to this subspace. The formula for P_{i+1} follows by a straightforward calculation. ◆

Measurement and Time-Update Forms of the Kalman Filter. These results enable us to sequentially compute $\{\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_{i|i}, P_i\}$ as

$$0 = \hat{\mathbf{x}}_{0|-1} \xrightarrow{m.u.} \hat{\mathbf{x}}_{0|0} \xrightarrow{t.u.} \hat{\mathbf{x}}_1 \xrightarrow{m.u.} \hat{\mathbf{x}}_{1|1} \xrightarrow{t.u.} \hat{\mathbf{x}}_2 \xrightarrow{m.u.} \hat{\mathbf{x}}_{2|2} \xrightarrow{t.u.} \hat{\mathbf{x}}_3 \cdots,$$

and

$$\Pi_0 = P_{0|-1} \xrightarrow{m.u.} P_{0|0} \xrightarrow{t.u.} P_{1|0} \xrightarrow{m.u.} P_{1|1} \cdots,$$

where the abbreviations *m.u.* and *t.u.* stand for measurement and time updates, respectively.

As mentioned earlier, this *two-step* (measurement and time-update) procedure makes it clear how to proceed if we have a variable time between measurements or if, for some reason, certain measurements are lost. Therefore, most digital computer implementations of the Kalman filter tend to be of this form. Analog (or hybrid) computer realizations usually use the prediction estimator equation, which can in fact be now obtained as follows:

$$\begin{aligned}\hat{\mathbf{x}}_{i+1} &= F_i\underbrace{(\hat{\mathbf{x}}_i + P_iH_i^*R_{e,i}^{-1}\mathbf{e}_i)}_{\hat{\mathbf{x}}_{i|i}} + G_iS_iR_{e,i}^{-1}\mathbf{e}_i\\ &= F_i\hat{\mathbf{x}}_i + K_{p,i}\mathbf{e}_i, \quad K_{p,i} = (F_iP_iH_i^* + G_iS_i)R_{e,i}^{-1}.\end{aligned}$$

9.3.3 Recursions for Filtered Estimators

It is also possible to combine the measurement and time-update formulas in reverse order, *i.e.*, first apply a time update and then a measurement update, to obtain a recursion for the filtered estimator $\hat{\mathbf{x}}_{i|i}$. Indeed

$$\begin{aligned}\hat{\mathbf{x}}_{i+1|i+1} &= \hat{\mathbf{x}}_{i+1} + K_{f,i+1}\mathbf{e}_{i+1},\\ &= \hat{\mathbf{x}}_{i+1} + K_{f,i+1}(\mathbf{y}_{i+1} - H_{i+1}\hat{\mathbf{x}}_{i+1}),\\ &= (I - K_{f,i+1}H_{i+1})\hat{\mathbf{x}}_{i+1} + K_{f,i+1}\mathbf{y}_{i+1},\\ &= (I - K_{f,i+1}H_{i+1})(F_i\hat{\mathbf{x}}_{i|i} + G_iS_iR_{e,i}^{-1}\mathbf{e}_i) + K_{f,i+1}\mathbf{y}_{i+1},\\ &= (I - K_{f,i+1}H_{i+1})F_i\hat{\mathbf{x}}_{i|i} + K_{f,i+1}\mathbf{y}_{i+1} + (I - K_{f,i+1}H_{i+1})G_iS_iR_{e,i}^{-1}\mathbf{e}_i.\end{aligned}$$

The recursion for $\hat{\mathbf{x}}_{i|i}$ is rather complicated since it involves both $\mathbf{y}_{i+1}$ and $\mathbf{e}_i$ (or equivalently, both $\mathbf{e}_{i+1}$ and $\mathbf{e}_i$). However, a significant simplification occurs if we assume that $S_i = 0$, in which case the above expression becomes

$$\hat{\mathbf{x}}_{i+1|i+1} = (I - K_{f,i+1}H_{i+1})F_i\hat{\mathbf{x}}_{i|i} + K_{f,i+1}\mathbf{y}_{i+1}. \tag{9.3.9}$$

[*In Sec. 9.5.1 we shall show how, under the assumption that $R_i > 0$, problems with nonzero S_i can always be converted to ones with $S_i = 0$.*]

Let us also find a recursion for the error variance of the filtered state estimators, $P_{i|i}$. Note that using (9.3.9), we can write

$$\hat{\mathbf{x}}_{i+1|i+1} = F_i\hat{\mathbf{x}}_{i|i} + K_{f,i+1}\mathbf{e}_{i+1},$$

from which it follows easily that

$$\Sigma_{i+1|i+1} = F_i\Sigma_{i|i}F_i^* + K_{f,i+1}R_{e,i+1}K_{f,i+1}^*,$$

where we have defined $\Sigma_{i|i} \triangleq \|\hat{\mathbf{x}}_{i|i}\|^2$. Subtracting the above equation from the recursion for the variance of the state, $\Pi_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^*$, we obtain

$$P_{i+1|i+1} = F_iP_{i|i}F_i^* + G_iQ_iG_i^* - K_{f,i+1}R_{e,i+1}K_{f,i+1}^*. \tag{9.3.10}$$

This calculation does not need the assumption $S_i = 0$.

However, we also need to express $K_{f,i+1}$ and $R_{e,i+1}$ in terms of $P_{i|i}$. Now we assume $S_i = 0$, so that (9.3.8) simplifies to $P_{i+1} = F_iP_{i|i}F_i^* + G_iQ_iG_i^*$. Then we can write

$$K_{f,i+1} = (F_iP_{i|i}F_i^* + G_iQ_iG_i^*)H_{i+1}^*R_{e,i+1}^{-1}, \tag{9.3.11}$$

and

$$R_{e,i+1} = R_{i+1} + H_{i+1}(F_iP_{i|i}F_i^* + G_iQ_iG_i^*)H_{i+1}^*. \tag{9.3.12}$$

Finally, we need to find the initial conditions $\hat{\mathbf{x}}_{0|0}$ and $P_{0|0}$. But $\hat{\mathbf{x}}_{0|0}$ follows from the first step of the measurement-update equation (*cf.* (9.3.3))

$$\hat{\mathbf{x}}_{0|0} = \hat{\mathbf{x}}_0 + \Pi_0H_0^*R_{e,0}^{-1}(\mathbf{y}_0 - H_0\hat{\mathbf{x}}_0) = \Pi_0H_0^*R_{e,0}^{-1}\mathbf{y}_0,$$

and $P_{0|0}$ follows from the first step of the corresponding error Gramian calculation (*cf.* (9.3.4))

$$P_{0|0} = \Pi_0 - \Pi_0H_0^*R_{e,0}^{-1}H_0\Pi_0.$$

We organize the above discussion in the following lemma.

Lemma 9.3.4 (Recursions for Filtered Estimators when $S_i = 0$) *For the standard model of Thm. 9.2.1, with $S_i = 0$, the filtered state estimators $\{\hat{\mathbf{x}}_{i|i}\}$ can be obtained via the recursions*

$$\hat{\mathbf{x}}_{i+1|i+1} = F_i\hat{\mathbf{x}}_{i|i} + K_{f,i+1}(\mathbf{y}_{i+1} - H_{i+1}F_i\hat{\mathbf{x}}_{i|i}), \quad \hat{\mathbf{x}}_{0|0} = \Pi_0H_0^*R_{e,0}^{-1}\mathbf{y}_0, \tag{9.3.13}$$

where $K_{f,i+1}$ and $R_{e,i+1}$ are given by (9.3.11) and (9.3.12), and where the error variance $P_{i|i} = \|\tilde{\mathbf{x}}_{i|i}\|^2$ satisfies the recursion (9.3.10) with $P_{0|0} = \Pi_0 - \Pi_0H_0^*R_{e,0}^{-1}H_0\Pi_0$. ■

9.3.4 An Alternative Innovations Model

The above recursions give us an alternative (to Thm. 9.2.34) causal and causally invertible model for the process $\{\mathbf{y}_i\}$ with the filtered estimators as the state variables.

Lemma 9.3.5 (An Innovations Model when $S_i = 0$) *When $S_i = 0$, the following is a causal and causally invertible model for the process $\{\mathbf{y}_i\}$; an alternative to the model (9.2.34) in Thm. 9.2.2:*

$$\hat{\mathbf{x}}_{i+1|i+1} = F_i\hat{\mathbf{x}}_{i|i} + K_{f,i+1}\mathbf{e}_{i+1}, \tag{9.3.14}$$

$$\mathbf{y}_{i+1} = H_{i+1}F_i\hat{\mathbf{x}}_{i|i} + \mathbf{e}_{i+1}, \quad i \geq 0, \tag{9.3.15}$$

with initial conditions $\mathbf{y}_0 = \mathbf{e}_0$, $\hat{\mathbf{x}}_{0|0} = \Pi_0H_0^*R_{e,0}^{-1}\mathbf{y}_0$, and of course $\langle \mathbf{e}_i, \mathbf{e}_j\rangle = R_{e,i}\delta_{ij}$. ■

Proof: When $S_i = 0$, $\hat{\mathbf{x}}_{i+1|i} = F_i\hat{\mathbf{x}}_{i|i}$, so that $\mathbf{y}_{i+1} - H_{i+1}F_i\hat{\mathbf{x}}_{i|i} = \mathbf{y}_{i+1} - H_{i+1}\hat{\mathbf{x}}_{i+1|i} = \mathbf{e}_{i+1}$. ◆

Remark 4 [Uniqueness]. A natural question: are not "canonical" (causal and causally invertible) process models unique? The answer is yes, in the sense that they will have the same impulse response sequences, but of course not necessarily the same state-space model! We leave to active readers the verification of this claim. ◆

9.4 TRIANGULAR FACTORIZATIONS OF R_y AND R_y^{-1}

Quite early, in Sec. 4.2.2, we noted that the innovations $\{\mathbf{e}_i\}$ of a process $\{\mathbf{y}_i\}$ could always be found by factoring the covariance/Gramian matrix R_y of $\mathbf{y} = \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}$ as $R_y = LDL^*$, leading to the formulas $\mathbf{y} = L\mathbf{e}$ and $\mathbf{e} = L^{-1}\mathbf{y}$, where $\mathbf{e} = \text{col}\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_N\}$. The point was to use further structure in $\{\mathbf{y}_i\}$ and/or R_y to reduce the computational burden of these calculations from the $O(N^3)$ required for an arbitrary Gramian matrix. As an example, we noted in Sec. 5.1.2 that having a first-order state-space model for $\{\mathbf{y}_i\}$ led directly to a method for computing the $\{\mathbf{e}_i\}$ and thence immediately to formulas for L^{-1}, L, and R_y^{-1}. We can now show the same for general state-space models by using the innovations representation described in Thm. 9.2.2.

In fact, Eq. (9.2.34) allows us to write the following global expression relating $\mathbf{e}$ and $\mathbf{y}$:

$$\mathbf{y} = \begin{bmatrix} I & 0 & 0 & \dots & 0 \\ H_1 K_{p,0} & I & 0 & \dots & 0 \\ H_2 \Phi(2,1) K_{p,0} & H_2 K_{p,1} & I & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ H_N \Phi(N,1) K_{p,0} & H_N \Phi(N,2) K_{p,1} & H_N \Phi(N,3) K_{p,2} & \dots & I \end{bmatrix} \mathbf{e}, \tag{9.4.1}$$

where we have defined the state transition matrix

$$\Phi(i,j) \triangleq F_{i-1} F_{i-2} \dots F_j \quad \text{for } i > j \quad \text{and} \quad \Phi(i,i) = I.$$

If we denote the nonsingular block *lower triangular matrix* in (9.4.1) as L, it follows that L yields the canonical factorization of the output Gramian,

$$R_y = \langle \mathbf{y}, \mathbf{y} \rangle = L R_e L^*. \tag{9.4.2}$$

Finally, we should note that it is not in general easy to invert L directly to obtain a nice expression for L^{-1} (see also the discussion in Sec. 4.4.2). However, with a model for L, the situation is much easier. Thus note that it is easy to invert the state-space equations (9.2.34) to obtain

$$\begin{cases} \hat{\mathbf{x}}_{i+1} = F_{p,i} \hat{\mathbf{x}}_i + K_{p,i} \mathbf{y}_i, & \hat{\mathbf{x}}_0 = 0, \\ \mathbf{e}_i = -H_i \hat{\mathbf{x}}_i + \mathbf{y}_i, \end{cases}$$

from which we can immediately write

$$L^{-1} = \tag{9.4.3}$$

$$\begin{bmatrix} I & 0 & 0 & \dots & 0 \\ -H_1 K_{p,0} & I & 0 & \dots & 0 \\ -H_2 \Phi_p(2,1) K_{p,0} & -H_2 K_{p,1} & I & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ -H_N \Phi_p(N,1) K_{p,0} & -H_N \Phi_p(N,2) K_{p,1} & -H_N \Phi_p(N,3) K_{p,2} & \dots & I \end{bmatrix},$$

where we have defined the *closed-loop* state transition matrix

$$\Phi_p(i,j) \triangleq F_{p,i-1} F_{p,i-2} \dots F_{p,j} \quad \text{for } i > j, \quad \Phi_p(i,i) = I.$$

Remark 5 [Direct Factorization of R_y]. As is evident from (9.4.1)–(9.4.2), or as noted earlier in Sec. 5.A, the Gramian R_y of course exhibits structure when it arises from a state-space model. And it should not be surprising that this special structure can be exploited to efficiently factor R_y and deduce the formula (9.4.1). In fact, we can do this in several different ways, of which we shall present two. In App. 9.A we use the modified Gram-Schmidt procedure for finding L (see Sec. 4.2.3), while in App. 9.B we generalize the method described for time-invariant models in Ch. 8. While both methods have value, the reader may note again that, when possible, working directly with the model (rather than its Gramian) is algebraically much simpler (*cf.* the discussion in Sec. 8.6). ◆

9.5 AN IMPORTANT SPECIAL ASSUMPTION: $R_i > 0$

As explained earlier in Sec. 9.2, the major assumption made so far in the derivation of the Kalman equations is that the $\{R_{e,i}\}$ are positive-definite and, hence, invertible. A good modeling decision that guarantees $R_{e,i} > 0$ is to assume that $R_i > 0$. In fact, one would like to have R_i as close to an identity matrix as possible. One reason is that then the different measurements (the components of the vectors $\{\mathbf{y}_i\}$) are all roughly of the same *quality*, and one might expect that such situations are easier to handle numerically than more *unbalanced* ones.

Moreover, the assumption that $R_i > 0$ enables various useful simplifications and modifications of the earlier results — several of these are now discussed.

9.5.1 Simplifications for Correlated Noise Processes

Recall that the time-update formula (9.3.8) for P_{i+1} simplifies substantially when $S_i = 0$:

$$P_{i+1} = F_i P_{i|i} F_i^* + G_i Q_i G_i^*, \quad \text{when } S_i = 0. \tag{9.5.1}$$

When $R_i > 0$, the general formula (9.3.8) can be rewritten more compactly as

$$P_{i+1} = F_i^s P_{i|i} F_i^{s*} + G_i Q_i^s G_i^*, \tag{9.5.2}$$

where

$$F_i^s \triangleq F_i - G_i S_i R_i^{-1} H_i, \quad Q_i^s = Q_i - S_i R_i^{-1} S_i^*. \tag{9.5.3}$$

In other words, when $R_i > 0$, we can reduce the time-update error covariance update in the $S_i \neq 0$ case to the form of the equation that arises when $S_i = 0$.

More generally, we shall now establish that when $S_i \neq 0$, we can in fact rewrite *all* the Kalman filter formulas in a form similar to those for the special case $S_i = 0$ by making the replacements

$$F_i \to F_i^s, \quad Q_i \to Q_i^s \triangleq Q_i - S_i R_i^{-1} S_i^*, \tag{9.5.4}$$

$$K_{p,i} \to K_{p,i}^s \triangleq F_i^s P_i H_i^* R_{e,i}^{-1} \triangleq K_i^s R_{e,i}^{-1}. \tag{9.5.5}$$

Thus we can write instead of the Riccati recursion,

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1} (F_i P_i H_i^* + G_i S_i)^*, \tag{9.5.6}$$

the following expression

$$P_{i+1} = F_i^s P_i F_i^{s*} + G_i Q_i^s G_i^* - F_i^s P_i H_i^* R_{e,i}^{-1} H_i P_i F_i^{s*}. \tag{9.5.7}$$

However, to offset this simplification, the estimator equations have an extra *bias* term: if we wish to use F_i^s, we must write the recursion (9.2.33) as

$$\hat{\mathbf{x}}_{i+1} = F_i^s \hat{\mathbf{x}}_i + K_{p,i}^s \mathbf{e}_i + G_i S_i R_i^{-1} \mathbf{y}_i. \tag{9.5.8}$$

Note that this can also be written as

$$\hat{\mathbf{x}}_{i+1} = (F_i^s - F_i^s P_i H_i^* R_{e,i}^{-1} H_i)\hat{\mathbf{x}}_i + (F_i^s P_i H_i^* R_{e,i}^{-1} + G_i S_i R_i^{-1})\mathbf{y}_i. \tag{9.5.9}$$

Comparing this with the usual form

$$\hat{\mathbf{x}}_{i+1} = F_{p,i}\hat{\mathbf{x}}_i + K_{p,i}\mathbf{y}_i, \tag{9.5.10}$$

shows that we can identify

$$F_{p,i} = F_i^s(I - P_i H_i^* R_{e,i}^{-1} H_i) = F_i^s(I + P_i H_i^* R_i^{-1} H_i)^{-1}, \tag{9.5.11}$$

where the second equality follows by use of the matrix inversion formula. These formulae will be useful later. Note also that comparing (9.5.8) with (9.5.10) yields

$$K_{p,i} = F_i^s P_i H_i^* R_{e,i}^{-1} + G_i S_i R_i^{-1} = K_{p,i}^s + G_i S_i R_i^{-1}, \tag{9.5.12}$$

which can, in fact, be directly verified (as also (9.5.11)). Starting with this formula and substituting it into the general Kalman filter formulas (*cf.* Thm. 9.2.1) will establish the claimed formulas (9.5.8), (9.5.7), and (9.5.3).

However, rather than going through this nonintuitive and algebraic route, we shall pursue a more basic approach to the problem. The point is that we wish somehow to reduce the problem with correlated $\{\mathbf{u}_i, \mathbf{v}_i\}$ to some equivalent problem with uncorrelated variables. A standard way of achieving this is by a Gram-Schmidt procedure, which would transform, as can readily be checked,

$$\{\mathbf{v}_i, \mathbf{u}_i\} \to \{\mathbf{v}_i, \mathbf{u}_i^s\}, \quad \text{where } \mathbf{u}_i^s = \mathbf{u}_i - \langle \mathbf{u}_i, \mathbf{v}_i\rangle \|\mathbf{v}_i\|^{-2}\mathbf{v}_i = \mathbf{u}_i - S_i R_i^{-1}\mathbf{v}_i. \tag{9.5.13}$$

Note that

$$\langle \mathbf{u}_i^s, \mathbf{v}_i\rangle = 0, \quad \langle \mathbf{u}_i^s, \mathbf{u}_i^s\rangle = Q_i - S_i R_i^{-1} S_i^* \triangleq Q_i^s. \tag{9.5.14}$$

Now, using (9.5.13), we can rewrite the given state-space model as

$$\mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i(\mathbf{u}_i^s + S_i R_i^{-1}\mathbf{v}_i), \tag{9.5.15}$$

$$\mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i, \tag{9.5.16}$$

which, by writing $\mathbf{v}_i = \mathbf{y}_i - H_i\mathbf{x}_i$ in the state equation, can be further reduced to

$$\mathbf{x}_{i+1} = F_i^s \mathbf{x}_i + G_i \mathbf{u}_i^s + G_i S_i R_i^{-1}\mathbf{y}_i, \tag{9.5.17}$$

$$\mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i. \tag{9.5.18}$$

We now have a state estimation problem with uncorrelated noises $\{\mathbf{u}_i^s, \mathbf{v}_i\}$, but with an additional driving term in the state equation. However, this driving term is a *known* function of the observations $\{\mathbf{y}_i\}$ and, therefore, its effect is easy to take into account in the estimation. For example, we can argue that

$$\begin{aligned}
\hat{\mathbf{x}}_{i+1|i} &= F_i^s \hat{\mathbf{x}}_{i|i} + G_i \hat{\mathbf{u}}_{i|i}^s + G_i S_i R_i^{-1}\hat{\mathbf{y}}_{i|i}, \\
&= F_i^s \hat{\mathbf{x}}_{i|i} + 0 + G_i S_i R_i^{-1}\mathbf{y}_i, && (9.5.19) \\
&= F_i^s\left[\hat{\mathbf{x}}_{i|i-1} + \langle \mathbf{x}_i, \mathbf{e}_i\rangle R_{e,i}^{-1}\mathbf{e}_i\right] + G_i S_i R_i^{-1}\mathbf{y}_i, \\
&= F_i^s \hat{\mathbf{x}}_{i|i-1} + F_i^s P_i H_i^* R_{e,i}^{-1}\mathbf{e}_i + G_i S_i R_i^{-1}\mathbf{y}_i. && (9.5.20)
\end{aligned}$$

We note here that, of course, the innovations $\{\mathbf{e}_i\}$ depend only upon the $\{\mathbf{y}_i\}$ and are unaffected by the rewriting of $\{\mathbf{u}_i, \mathbf{v}_i\}$ as $\{\mathbf{u}_i^s, \mathbf{v}_i\}$; that is why the measurement-update equations are the same whether $S_i = 0$ or not. The various recursions for $\{P_i, P_{i|i}\}$ now follow in a routine manner.

[Curious readers might wonder what would have happened if we had replaced $\{\mathbf{u}_i, \mathbf{v}_i\}$ not by $\{\mathbf{v}_i, \mathbf{u}_i^s\}$ but by $\{\mathbf{u}_i, \mathbf{v}_i^s = \mathbf{v}_i - S^* Q^{-1}\mathbf{u}_i\}$, assuming $Q > 0$. They should check that this replacement does not simplify the original formulas.]

To complete the derivations, note that the errors obey

$$\tilde{\mathbf{x}}_{i+1} = F_i^s \tilde{\mathbf{x}}_{i|i} + G_i \tilde{\mathbf{u}}_{i|i}^s = F_i^s \tilde{\mathbf{x}}_i - K_{p,i}^s \mathbf{e}_i + G_i \tilde{\mathbf{u}}_{i|i}^s,$$

from which the formulas (9.5.2) and (9.5.7) follow easily.

For later use, it will be useful to note here, using formula (9.5.27) below for $P_{i|i}$, the following alternative forms:

$$\begin{aligned}
P_{i+1} &= F_i^s(P_i^{-1} + H_i^* R_i^{-1} H_i)^{-1} F_i^{s*} + G_i Q_i^s G_i^*, && (9.5.21) \\
&= F_i^s P_i(I + H_i^* R_i^{-1} H_i P_i)^{-1} F_i^{s*} + G_i Q_i^s G_i^*, && (9.5.22) \\
&= F_i^s(I + P_i H_i^* R_i^{-1} H_i)^{-1} P_i F_i^{s*} + G_i Q_i^s G_i^*, && (9.5.23)
\end{aligned}$$

which arise naturally in a so-called Redheffer scattering theory model for the estimation problem as we shall discuss in Ch. 17.

9.5.2 Measurement Updates in Information Form

Another useful implication of the assumption $R_i > 0$ is an alternative expression for the measurement-update equations of Lemma 9.3.2, which we rewrite below:

$$P_{i|i} = P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i = (I - K_{f,i} H_i) P_i, \tag{9.5.24}$$

$$\hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_i + K_{f,i}\mathbf{e}_i, \tag{9.5.25}$$

$$K_{f,i} = P_i H_i^* R_{e,i}^{-1}, \quad R_{e,i} = R_i + H_i P_i H_i^*. \tag{9.5.26}$$

When $R_i > 0$, and assuming that P_i^{-1} exists, we can write

$$P_{i|i}^{-1} = P_i^{-1} + H_i^* R_i^{-1} H_i, \tag{9.5.27}$$

and

$$P_{i|i}^{-1}\hat{\mathbf{x}}_{i|i} = \left[P_i^{-1}\hat{\mathbf{x}}_i + H_i^* R_i^{-1}\mathbf{y}_i\right]. \tag{9.5.28}$$

Moreover, while $K_{f,i}$ does not appear explicitly in these formulas, it will be useful to note that it can be rewritten as

$$K_{f,i} = P_{i|i}H_i^*R_i^{-1} = (P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}H_i^*R_i^{-1}. \tag{9.5.29}$$

Remark 6. Since the inverse of the variance of a parameter is a (rough) measure of the *information* in the parameter, *i.e.*, large variance means high uncertainty or less information, the formulas (9.5.27)–(9.5.28) are often described as *Information-Form* measurement update formulas. Correspondingly, the formulas in Thm. 9.2.1 are often called the *Covariance Forms*. ◆

Derivation. By applying the matrix inversion formula, we can re-express (9.3.4) as

$$P_{i|i} = P_i - P_iH_i^*(R_i + H_iP_iH_i^*)^{-1}H_iP_i = (P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1},$$

which is (9.5.27). Then we can write, by again using the matrix inversion lemma and (9.5.27),

$$\begin{aligned} K_{f,i} &= P_iH_i^*(R_i + H_iP_iH_i^*)^{-1}, \\ &= P_iH_i^*\left[R_i^{-1} - R_i^{-1}H_i(P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}H_i^*R_i^{-1}\right], \\ &= P_i\left[I - H_i^*R_i^{-1}H_iP_{i|i}\right]H_i^*R_i^{-1}, \\ &= P_i\left[I - (P_{i|i}^{-1} - P_i^{-1})P_{i|i}\right]H_i^*R_i^{-1} = P_{i|i}H_i^*R_i^{-1}, \end{aligned}$$

which is (9.5.29). Finally with (9.5.29) we can write (9.5.24)–(9.5.25) as

$$P_{i|i}^{-1}\hat{\mathbf{x}}_{i|i} = P_{i|i}^{-1}\left[(I - K_{f,i}H_i)\hat{\mathbf{x}}_i + K_{f,i}\mathbf{y}_i\right] = P_i^{-1}\hat{\mathbf{x}}_i + H_i^*R_i^{-1}\mathbf{y}_i,$$

which is (9.5.28).

Interpretation as Combination of Estimators. The expressions (9.5.27)–(9.5.28) may be recognized as the formula for combining the prior information, $\hat{\mathbf{x}}_{i|i-1}$, on $\mathbf{x}_i$ and the new (noisy) observation, $\mathbf{y}_i$, of $\mathbf{x}_i$ (see Sec. 3.4.3 and Prob. 3.23). Recall that we can think of obtaining $\hat{\mathbf{x}}_{i|i}$ as an appropriately weighted combination of the estimators that are separately based on $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$ and $\mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i$. The first estimator is simply $\hat{\mathbf{x}}_i$ with covariance matrix P_i. The second estimator is (recall expression (3.4.4))

$$\begin{aligned} \hat{\mathbf{x}}_{i|\{\mathbf{y}_i\}} &= \langle \mathbf{x}_i, \mathbf{y}_i \rangle \langle \mathbf{y}_i, \mathbf{y}_i \rangle^{-1} \mathbf{y}_i = \Pi_iH_i^*(R_i + H_i\Pi_iH_i^*)^{-1}\mathbf{y}_i, \\ &= (\Pi_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}H_i^*R_i^{-1}\mathbf{y}_i, \end{aligned}$$

with covariance matrix

$$P_{x_i|y_i} = \Pi_i - \Pi_iH_i^*(R_i + H_i\Pi_iH_i^*)^{-1}H_i\Pi_i = (\Pi_i^{-1} + H_i^*R_i^{-1}H_i).$$

We therefore conclude that (*cf.* Eq. (3.4.12))

$$P_{i|i}^{-1}\hat{\mathbf{x}}_{i|i} = \left[P_i^{-1}\hat{\mathbf{x}}_i + P_{x_i|y_i}^{-1}(\Pi_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}H_i^*R_i^{-1}\mathbf{y}_i\right] = \left[P_i^{-1}\hat{\mathbf{x}}_i + H_iR_i^{-1}\mathbf{y}_i\right],$$

which is the result (9.5.28).

9.5.3 Existence of P_i^{-1}

Of course, the formulas (9.5.27)–(9.5.28) presume the existence of P_i^{-1} and $P_{i|i}^{-1}$. Indeed, it may happen that for some reason P_i is very large. For example, when $i = 0$, a very large value of the initial covariance $P_0 = \Pi_0$ will reflect the fact that our knowledge of the initial state $\mathbf{x}_0$ is very limited. In this case, the computation of $P_{i|i}$ and $\hat{\mathbf{x}}_{i|i}$ by (9.5.24)–(9.5.25) will be numerically very difficult. On the other hand, (9.5.27)–(9.5.28) will serve very well in such situations. In any case, this discussion shows that we would not tend to use the formulas (9.5.27)–(9.5.28) when P_i is *small, i.e.*, likely to be singular. Moreover, the singularity of P_i will imply that certain (combinations of) components of the state vector can be determined without error, and this will correspond to a poorly modeled physical problem.

Lemma 9.5.1 (A Sufficient Condition for Existence of P_i^{-1}) *Assume $S_i = 0$, $R_i > 0$, $\Pi_0 > 0$, and F_i invertible. Then $P_i > 0$ and, hence, is invertible.* ■

Proof: A simple proof is by induction. The claim is certainly valid for $i = 0$ since $P_0 = \Pi_0$ and $\Pi_0 > 0$. So assume it is valid up to time i. In view of $P_i > 0$, we can rearrange the Riccati recursion (9.2.14) as

$$P_{i+1} = G_iQ_iG_i^* + F_i(P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}F_i^*,$$

which shows that $P_{i+1} \geq F_i(P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}F_i^* > 0$, where the last inequality depends upon the nonsingularity of F_i. ◆

For the correlated noise case ($S_i \neq 0$), we obtain a similar conclusion by requiring instead the *invertibility* of the matrices F_i^s that were defined in (9.5.4), *viz.*, $F_i^s = F_i - G_iS_iR_i^{-1}H_i$. Another condition for the invertibility of the $\{P_i\}$ is established in Prob. 9.17.

9.5.4 Sequential Processing

A very useful application of Eqs. (9.5.27)–(9.5.29) is to reduce the problem of vector measurements (*i.e.*, $\mathbf{y}_i$ a $p \times 1$ vector, $p > 1$) to that of a sequence of scalar measurements. Doing this would reduce computations because inversion of the $p \times p$ matrices $R_{e,i}$ would now be trivialized.

The first step is to arrange that the entries of the output noise vector be uncorrelated. So factor R_i as $R_i = L_iD_iL_i^*$ and scale the output equation $\mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i$ by L_i^{-1}, *i.e.*,

$$L_i^{-1}\mathbf{y}_i = L_i^{-1}H_i\mathbf{x}_i + L_i^{-1}\mathbf{v}_i.$$

Then the new noise sequence $\bar{\mathbf{v}}_i = L_i^{-1}\mathbf{v}_i$ is such that

$$E\bar{\mathbf{v}}_i\bar{\mathbf{v}}_j^* = D_i\delta_{ij}, \quad D_i = \text{diag}\{d_i^1, d_i^2, \ldots, d_i^p\},$$

for some positive numbers $\{d_i^j\}$.

We further partition the entries of the scaled output vector $L_i^{-1}\mathbf{y}_i$, and of the scaled matrix $L_i^{-1}H_i$, as follows:

$$L_i^{-1}\mathbf{y}_i \triangleq \text{col}\{\mathbf{y}^1(i), \mathbf{y}^2(i), \ldots, \mathbf{y}^p(i)\}, \qquad L_i^{-1}H_i \triangleq \text{col}\{h_i^1, h_i^2, \ldots, h_i^p\},$$

where $\{\mathbf{y}^k(i)\}$ are scalars and $\{h_i^k\}$ are row vectors.

Now the p measurement processes $\{\mathbf{y}^1(i), \ldots, \mathbf{y}^p(i)\}$ will be mutually uncorrelated and we should be able to incorporate them one at a time, essentially by making a series of measurement updates, first with $\mathbf{y}^1(i)$, then with $\mathbf{y}^2(i)$, ..., and finally with $\mathbf{y}^p(i)$.

To do this, first refer back to the formula (9.5.24) for a measurement update, which suggests that we successively compute a sequence of matrices

$$\begin{aligned} P_i^1 &= (I - K_{f,i}^1 h_i^1)P_i & K_{f,i}^1 &= P_i h_i^{1*}[h_i^1 P_i h_i^{1*} + d_i^1]^{-1}, \\ P_i^2 &= (I - K_{f,i}^2 h_i^2)P_i^1 & K_{f,i}^2 &= P_i^1 h_i^{2*}[h_i^2 P_i^1 h_i^{2*} + d_i^2]^{-1}, \\ &\vdots & &\vdots \\ P_i^p &= (I - K_{f,p}^p h_i^p)P_i^{p-1} & K_{f,i}^p &= P_i^{p-1} h_i^{p*}[h_i^p P_i^{p-1} h_i^{p*} + d_i^p]^{-1}. \end{aligned} \tag{9.5.30}$$

Then P_i^p will be the updated covariance matrix $P_{i|i}$ based on all the measurements. Note that all the inversions required here are trivial, *i.e.,* scalar.

The correctness of this scheme is really almost self-evident, using the basic formula for measurement updating, but a formal proof can also readily be given. Apply the matrix inversion formula (App. A) to formula (9.5.30) to obtain

$$(P_i^k)^{-1} = (P_i^{k-1})^{-1} + h_i^{k*}(d_i^k)^{-1}h_i^k, \quad k = 1, \ldots, p.$$

Then starting with $P_i^0 = P_i$, successive substitution yields

$$(P_i^p)^{-1} = P_i^{-1} + H_i^* D_i^{-1} H_i = P_i^{-1} + H_i^* R_i^{-1} H_i,$$

and comparing this with Eq. (9.5.27) shows that $P_i^p = P_{i|i}$.

As far as the estimators go, sequential incorporation of the new information in the components $\{\mathbf{y}^1(i), \ldots, \mathbf{y}^p(i)\}$ will lead to the equations $\hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_i^p$, where, by the basic measurement update formula,

$$\hat{\mathbf{x}}_i^k = \hat{\mathbf{x}}_i^{k-1} + K_{f,i}^k[\mathbf{y}^k(i) - h_i^k \hat{\mathbf{x}}_i^{k-1}], \quad k = 1, \ldots, p, \quad \hat{\mathbf{x}}_i^0 = \hat{\mathbf{x}}_i,$$

and

$$K_{f,i}^k = \langle \mathbf{x}_i, \mathbf{e}^k(i) \rangle \, \|\mathbf{e}^k(i)\|^{-2}, \quad \mathbf{e}^k(i) = \mathbf{y}^k(i) - h_i^k \hat{\mathbf{x}}_i^{k-1}.$$

Now if we define

$$P_i^k = \|\tilde{\mathbf{x}}_i^k\|^2, \quad \tilde{\mathbf{x}}_i^k = \mathbf{x}_i - \hat{\mathbf{x}}_i^k,$$

then we can readily see that

$$K_{f,i}^k = P_i^k h_i^{k*}[h_i^k P_i^k h_i^{k*} + d_i^k]^{-1}.$$

It should be noted that this is the same as the expression for $K_{f,i}^k$ in formula (9.5.30), where we did not make explicit the stochastic meaning of P_i^k and $K_{f,i}^k$.

9.5.5 Time Updates in Information Form ($Q_i > 0$)

To complete the discussion in Sec. 9.5.2, we need to obtain information form time updates for $P_{i|i}^{-1}$. However, examining the relevant direct formula (9.3.10), shows that this is likely to be complicated unless $S_i = 0$ (*and F_i is invertible*). So we shall start with the assumption that $S_i = 0$ and then apply the transformations of Sec. 9.5.1 to show how to handle the general case (when $R_i > 0$).

So we start with

$$P_{i+1} = F_i P_{i|i} F_i^* + G_i Q_i G_i^*,$$

and apply the matrix inversion formula, assuming Q_i invertible, to obtain

$$P_{i+1}^{-1} = F_i^{-*} P_{i|i}^{-1} F_i^{-1} - F_i^{-*} P_{i|i}^{-1} F_i^{-1} G_i [Q_i^{-1} + G_i^* F_i^{-*} P_{i|i}^{-1} F_i^{-1} G_i]^{-1} G_i^* F_i^{-*} P_{i|i}^{-1} F_i^{-1}.$$

If we further recall (9.5.27), *viz.*,

$$P_{i+1|i+1}^{-1} = P_{i+1}^{-1} + H_{i+1}^* R_{i+1}^{-1} H_{i+1}, \tag{9.5.31}$$

then we obtain a recursion for the inverse of the filtered-error covariance matrix, $P_{i|i}^{-1}$, rather than P_i^{-1}. This recursion can be written in compact form as follows. Introduce the quantities

$$K_{p,i}^d \triangleq F_i^{-*} P_{i|i}^{-1} F_i^{-1} G_i R_{e,i}^{-d},$$

and

$$R_{e,i}^d \triangleq Q_i^{-1} + G_i^* F_i^{-*} P_{i|i}^{-1} F_i^{-1} G_i.$$

Then the above recursions for P_{i+1}^{-1} and $P_{i+1|i+1}^{-1}$ lead to

$$P_{i+1|i+1}^{-1} = F_i^{-*} P_{i|i}^{-1} F_i^{-1} + H_{i+1}^* R_{i+1}^{-1} H_{i+1} - K_{p,i}^d R_{e,i}^d K_{p,i}^{d*}. \tag{9.5.32}$$

The interesting and important fact is that this recursion has the same structure as the standard Riccati recursion for P_i; this suggests a certain duality which we explore further in Sec. 12.8.4 — see Table 12.1.

Moreover, note from (9.5.28) that

$$\begin{aligned} P_{i+1|i+1}^{-1}\hat{\mathbf{x}}_{i+1|i+1} &= P_{i+1}^{-1}\hat{\mathbf{x}}_{i+1} + H_{i+1}^* R_{i+1}^{-1}\mathbf{y}_{i+1}, \\ &= [F_i^{-*} - K_{p,i}^d G_i^* F_i^{-*}] P_{i|i}^{-1} F_i^{-1}[F_i \hat{\mathbf{x}}_{i|i}] + H_{i+1}^* R_{i+1}^{-1}\mathbf{y}_{i+1}, \\ &= [F_i^{-*} - K_{p,i}^d G_i^* F_i^{-*}] P_{i|i}^{-1}\hat{\mathbf{x}}_{i|i} + H_{i+1}^* R_{i+1}^{-1}\mathbf{y}_{i+1}. \end{aligned} \tag{9.5.33}$$

Expressions (9.5.32)–(9.5.33) constitute the desired information form recursions for the filtered state estimator when $S_i = 0$. As noted earlier, we can obtain the formulas

for the correlated noise case by invoking the substitutions (9.5.4)–(9.5.5). Thus, for example,

$$K_{p,i}^d = F_i^{-s*} P_{i|i}^{-1} F_i^{-s} G_i R_{e,i}^{-d}, \qquad R_{e,i}^d = Q_i^{-s} + G_i^* F_i^{-s*} P_{i|i}^{-1} F_i^{-s} G_i,$$

and

$$P_{i+1|i+1}^{-1} = F_i^{-s*} P_{i|i}^{-1} F_i^{-s} + H_{i+1}^* R_{i+1}^{-1} H_{i+1} - K_{p,i}^d R_{e,i}^d K_{p,i}^{d*} \quad (9.5.34)$$

$$P_{i+1|i+1}^{-1} \hat{\mathbf{x}}_{i+1|i+1} = [F_i^{-s*} - K_{p,i}^d G_i^* F_i^{-s*}] P_{i|i}^{-1} \hat{\mathbf{x}}_{i|i} + H_{i+1}^* R_{i+1}^{-1} \mathbf{y}_{i+1} \quad (9.5.35)$$

The above expressions require the invertibility of Q_i^s rather than Q_i.

9.5.6 A Recursion for P_i^{-1}

The information form of the measurement-update equations (9.5.27)–(9.5.28) requires knowledge of P_i^{-1}, a recursion for which can be obtained by combining the time-update and filtered information forms. More specifically, using (9.5.27) in (9.5.34) leads to the equality

$$P_{i+1}^{-1} + H_{i+1}^* R_{i+1}^{-1} H_{i+1} = F_i^{-s*} [P_i^{-1} + H_i^* R_i^{-1} H_i] F_i^{-s} + H_{i+1}^* R_{i+1}^{-1} H_{i+1} - K_{p,i}^d R_{e,i}^d K_{p,i}^{d*},$$

which is equivalent to

$$P_{i+1}^{-1} = F_i^{-s*} P_i^{-1} F_i^{-s} + F_i^{-s*} H_i^* R_i^{-1} H_i F_i^{-s} - K_{p,i}^d R_{e,i}^d K_{p,i}^{d*}, \quad P_0^{-1} = \Pi_0^{-1}. \quad (9.5.36)$$

9.5.7 Summary of Results under Invertibility Conditions

It will be useful to summarize the major results of the earlier sections in a theorem.

Theorem 9.5.1 (Standard and Information Forms) *Consider the state-space equations (9.2.30)–(9.2.31), and assume that $R_i > 0$, $\Pi_0 > 0$ and that $\{F_i^s, Q_i^s\}$ are nonsingular whenever the inverses are needed, where*

$$F_i^s = F_i - G_i S_i R_i^{-1} H_i, \qquad Q_i^s = Q_i - S_i R_i^{-1} S_i^*.$$

Define further

$$K_{p,i} \triangleq (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}, \qquad R_{e,i} \triangleq R_i + H_i P_i H_i^*,$$

$$K_{p,i}^s \triangleq F_i^s P_i H_i^* R_{e,i}^{-1} = K_{p,i} - G_i S_i R_i^{-1},$$

$$K_{p,i}^d \triangleq F_i^{-s*} P_{i|i}^{-1} F_i^{-s} G_i R_{e,i}^{-d}, \qquad R_{e,i}^d \triangleq Q_i^{-s} + G_i^* F_i^{-s*} P_{i|i}^{-1} F_i^{-s} G_i,$$

$$F_{p,i} \triangleq F_i - K_{p,i} H_i = F_i^s - K_{p,i}^s H = F_i^s (I + P_i H_i R_i^{-1} H_i^*)^{-1}.$$

Then it follows that $P_i > 0$, and that the following relations hold:

- ***Predictor Updates:***

$$\begin{aligned}
\hat{\mathbf{x}}_{i+1} &= F_i \hat{\mathbf{x}}_i + K_{p,i}(\mathbf{y}_i - H_i \hat{\mathbf{x}}_i), \\
&= F_i^s \hat{\mathbf{x}}_i + K_{p,i}^s \mathbf{e}_i + G_i S_i R_i^{-1} \mathbf{y}_i, \\
P_{i+1} &= F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \\
&= F_i^s P_i F_i^{s*} + G_i Q_i^s G_i^* - K_{p,i}^s R_{e,i} K_{p,i}^{s*}, \\
&= F_i^s (I + P_i H_i^* R_i^{-1} H_i)^{-1} P_i F_i^{s*} + G_i Q_i^s G_i^*, \quad P_0 = \Pi_0.
\end{aligned}$$

- ***Measurement Updates:***

$$\begin{aligned}
\hat{\mathbf{x}}_{i|i} &= \hat{\mathbf{x}}_i + K_{f,i} \mathbf{e}_i, \\
K_{f,i} &= P_{i|i} H_i^* R_i^{-1} = P_i H_i^* R_{e,i}^{-1}, \\
P_{i|i} &= (I - K_{f,i} H_i) P_i.
\end{aligned}$$

- ***Measurement Updates in Information Form:***

$$\begin{aligned}
P_{i|i}^{-1} \hat{\mathbf{x}}_{i|i} &= P_i^{-1} \hat{\mathbf{x}}_i + H_i^* R_i^{-1} \mathbf{y}_i, \\
P_{i|i}^{-1} &= P_i^{-1} + H_i^* R_i^{-1} H_i.
\end{aligned}$$

- ***Time Updates:***

$$\begin{aligned}
\hat{\mathbf{x}}_{i+1} &= F_i^s \hat{\mathbf{x}}_{i|i} + G_i S_i R_i^{-1} \mathbf{y}_i, \\
P_{i+1} &= F_i^s P_{i|i} F_i^{s*} + G_i Q_i^s G_i^*.
\end{aligned}$$

- ***Time Updates in Information Form:***

$$P_{i+1}^{-1} = F_i^{-s*} P_{i|i}^{-1} F_i^{-s} - K_{p,i}^d R_{e,i}^d K_{p,i}^{d*}, \quad P_0^{-1} = \Pi_0^{-1}.$$

- ***Information Form for the Filtered Estimators:***

$$\begin{aligned}
P_{i+1|i+1}^{-1} \hat{\mathbf{x}}_{i+1|i+1} &= [F_i^{-s*} - K_{p,i}^d G_i^* F_i^{-s*}] P_{i|i}^{-1} \hat{\mathbf{x}}_{i|i} + H_{i+1}^* R_{i+1}^{-1} \mathbf{y}_{i+1}, \\
P_{i+1|i+1}^{-1} &= F_i^{-s*} P_{i|i}^{-1} F_i^{-s} + H_{i+1}^* R_{i+1}^{-1} H_{i+1} - K_{p,i}^d R_{e,i}^d K_{p,i}^{d*}.
\end{aligned}$$

- ***Recursion for the Inverse Riccati Variable:***

$$P_{i+1}^{-1} = F_i^{-s*} P_i^{-1} F_i^{-s} + F_i^{-s*} H_i^* R_i^{-1} H_i F_i^{-s} - K_{p,i}^d R_{e,i}^d K_{p,i}^{d*}, \quad P_0^{-1} = \Pi_0^{-1}.$$

■

9.6 COVARIANCE–BASED FILTERS

It is sometimes thought that the reason the Kalman filter goes beyond the Wiener filter is that the Kalman filter starts with a model for the process, rather than with the spectral/covariance data, thus apparently obviating the need for the difficult spectral/covariance factorization step. This is not correct. As noted before (Sec. 9.4), the Kalman filter recursions in effect carry out the factorization step in the process of finding the innovations. Another way of understanding this is to recall that the Kalman

filter recursions are equivalent to finding an innovations model for the observations (see Thm. 9.2.2), and of course this model must be completely determined by the spectral/covariance data and not by any particular model (*e.g.*, (9.1.2)) for the process $\{\mathbf{y}_i\}$. We shall demonstrate this explicitly below, after we write down expressions for the covariance of $\mathbf{y}_i$ in *state-space form* (this is the key ingredient). We shall then also show how to express the Kalman filter in terms of the covariance data.

To be specific, consider again the state-space model (9.2.30)–(9.2.31). For such a model, the covariance function of the output process $\{\mathbf{y}_i\}$ can readily be calculated (see, *e.g.*, Sec. 5.3.4) as

$$R_y(i,j) = \langle \mathbf{y}_i, \mathbf{y}_j \rangle = \begin{cases} H_i \Phi(i, j+1) N_j & i > j, \\ H_i \Pi_i H_i^* + R_i & i = j, \\ N_i^* \Phi^*(j, i+1) H_j^* & i < j, \end{cases} \tag{9.6.1}$$

where

$$\Phi(i,j) = F_{i-1} F_{i-2} \dots F_j, \quad i > j, \qquad \Phi(i,i) = I,$$

$$\langle \mathbf{x}_{i+1}, \mathbf{x}_{i+1} \rangle \triangleq \Pi_{i+1} = F_i \Pi_i F_i^* + G_i Q_i G_i^*,$$

$$N_i = F_i \Pi_i H_i^* + G_i S_i.$$

The problem we pose is to determine the innovations $\{\mathbf{e}_i\}$ by using only knowledge of the covariance parameters $\{F_i, H_i, N_i, R_y(i,i)\}$.

Now note that if we know the original model parameters, we can write $\mathbf{e}_i = \mathbf{y}_i - H_i \hat{\mathbf{x}}_i$, where

$$\hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_i + K_{p,i}(\mathbf{y}_i - H_i \hat{\mathbf{x}}_i), \quad \hat{\mathbf{x}}_0 = 0, \tag{9.6.2}$$

and

$$K_{p,i} = F_i P_i H_i^* R_{e,i}^{-1}, \qquad R_{e,i} = H_i P_i H_i^* + R_i, \tag{9.6.3}$$

with P_i given by the Riccati recursion

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \quad P_0 = \Pi_0.$$

So our problem is clearly to re-express $\{K_{p,i}, R_{e,i}\}$ in terms of the covariance data. To do this, we shall start by defining $\Sigma_i = \langle \hat{\mathbf{x}}_i, \hat{\mathbf{x}}_i \rangle$. Then the fact that $\{\mathbf{y}_i - H_i \hat{\mathbf{x}}_i = \mathbf{e}_i\}$ is a white process shows that from Eq. (9.6.2) we can write

$$\Sigma_{i+1} = F_i \Sigma_i F_i^* + K_{p,i} R_{e,i} K_{p,i}^*, \quad \Sigma_0 = 0. \tag{9.6.4}$$

Now recalling that $P_i \triangleq \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i \rangle$, $\tilde{\mathbf{x}}_i \perp \hat{\mathbf{x}}_i$, we see that $\Pi_i = \Sigma_i + P_i$. Hence, we can write

$$R_{e,i} = R_i + H_i P_i H_i^* = R_i + H_i(\Pi_i - \Sigma_i) H_i^* = R_y(i,i) - H_i \Sigma_i H_i^*,$$

$$K_{p,i} = F_i P_i H_i^* R_{e,i}^{-1} = [N_i - F_i \Sigma_i H_i^*] R_{e,i}^{-1},$$

where Σ_i is determined by the recursion (9.6.4). With this discussion in hand, the proof of the following theorem is immediate.

Theorem 9.6.1 (Canonical Modeling and Whitening Filters) *Consider an additive model process $\mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i$, where $\mathbf{v}_i$ is a zero-mean process, uncorrelated with $\{\mathbf{x}_j\}_{j=0}^i$, and such that $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = R_i \delta_{ij}$. Assume knowledge of the covariance function of $\{\mathbf{y}_i\}$, expressed in the state-space form (9.6.1). Then the canonical modeling filter for $\{\mathbf{y}_i\}$ is given by a state-space model $\{F_i, K_{p,i}, H_i, I\}$, say*

$$\begin{cases} \boldsymbol{\theta}_{i+1} = F_i \boldsymbol{\theta}_i + K_{p,i} \mathbf{e}_i, & \boldsymbol{\theta}_0 = 0, \\ \mathbf{y}_i = H_i \boldsymbol{\theta}_i + \mathbf{e}_i, \end{cases}$$

where $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = R_{e,i} \delta_{ij}$,

$$K_{p,i} = [N_i - F_i \Sigma_i H_i^*] R_{e,i}^{-1}, \qquad R_{e,i} = R_y(i,i) - H_i \Sigma_i H_i^*,$$

and

$$\Sigma_{i+1} = F_i \Sigma_i F_i^* + K_{p,i} R_{e,i} K_{p,i}^*, \quad \Sigma_0 = 0.$$

The canonical whitening filter for $\{\mathbf{y}_i\}$ is further given by

$$\begin{cases} \boldsymbol{\theta}_{i+1} = (F_i - K_{p,i} H_i) \boldsymbol{\theta}_i + K_{p,i} \mathbf{y}_i, & \boldsymbol{\theta}_0 = 0, \\ \mathbf{e}_i = H_i \boldsymbol{\theta}_i - \mathbf{y}_i, \end{cases}$$

i.e., by a state-space model with system matrices $\{F_i - K_{p,i} H_i, K_{p,i}, H_i, -I\}$. ■

Remark 7. We have denoted the state variable by $\boldsymbol{\theta}_i$, but the discussion before the theorem shows that $\boldsymbol{\theta}_i$ is the l.l.m.s.e. of the state of *any* state-space model (with system parameters $\{F_i, H_i, G_i\}$) whose output has the given covariance function (9.6.1). ◆

Remark 8 [The Stationary Case]. These formulas are most useful when the underlying state-space model is time-invariant, in which case

$$R_y(i,j) = \begin{cases} H F^{i-j-1} N_j & i > j, \\ H \Pi_i H^* + R & i = j, \\ N_i^* F^{*(j-i-1)} H^* & i < j. \end{cases}$$

They become even simpler when F is stable and we can assume that the process $\{\mathbf{y}_i\}$ is in steady state. In this case, the covariance function can be written as

$$R_y(i) = \begin{cases} H F^{i-1} \bar{N} & i > 0, \\ H \bar{\Pi} H^* + R & i = 0, \\ \bar{N}^* F^{*(-i-1)} H^* & i < 0, \end{cases}$$

where $\bar{N} = F \bar{\Pi} H^* + GS$ and $\bar{\Pi}$ is the unique positive-semi-definite solution of the Lyapunov equation

$$\bar{\Pi} = F \bar{\Pi} F^* + GQG^*.$$

Now the so-called minimal and partial realization procedures from linear system theory can be used to determine a triplet $\{H, F, \bar{N}\}$ (and $R_y(0)$) from a given covariance sequence $\{R_y(i)\}$. ◆

Remark 9 [Polynomial Factorization]. A nice application of the above results is to the problem of factoring a symmetric (positive) Laurent polynomial, which we encountered in Sec. 6.5. Any such Laurent polynomial can be regarded as the covariance function of a stationary moving-average process, whose z-spectrum has the form

$$S_y(z) = R_y(-n)z^n + \ldots + R_y(-1)z + R_y(0) + R_y(1)z^{-1} + \ldots + R_y(n)z^{-n}$$

for some finite integer n. It is easy to check that the choice

$$F = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & 0 & & \\ & & \ddots & \ddots & \\ & & & 1 & 0 \end{bmatrix}, \quad H = [0 \ \ldots \ 0 \ 1], \quad \bar{N} = \begin{bmatrix} R_y(n) \\ R_y(n-1) \\ \vdots \\ R_y(2) \\ R_y(1) \end{bmatrix},$$

where F is $n \times n$, H is $1 \times n$, and $\bar{N}$ is $n \times 1$, ensures that

$$R_y(i) = HF^{i-1}\bar{N}, \quad 0 < i \leq n.$$

We can then write a canonical modeling filter as in Thm. 9.6.1 for R_y, and its impulse response matrix will be the canonical factor L in the triangular factorization $R_y = LDL^*$. Now since the matrix F is stable, we might expect the Σ_i and $K_{p,i}$ to converge to constant values $\bar{\Sigma}$ and K_p, respectively.[5] This means that the $i-$th row of L will tend, as $i \to \infty$, to a constant row; which will in fact have the form (*cf.* (9.4.1))

$$[0 \ \ldots \ 0 \ HF^{n-1}K_p \ \ldots \ HFK_p \ HK_p \ 1]. \tag{9.6.5}$$

Moreover, the z-transform of this row, say

$$\begin{aligned} L(z) &= 1 + HK_pz^{-1} + HFK_pz^{-2} + \ldots + HF^{n-1}K_pz^{-n}, \\ &= 1 + H(zI - F)^{-1}K_p, \end{aligned}$$

is in fact the canonical spectral factor of $S_y(z)$, *i.e.*,

$$S_y(z) = L(z)R_eL^*(z^{-*}),$$

where $R_e = R_y(0) - H\bar{\Sigma}H^*$.

In fact, the reader can check that this is exactly the Bauer method for polynomial factorization that we briefly noted in Sec. 6.5. We see now that a Riccati recursion underlies Bauer's algorithm. Moreover, we see that the method applies also to nonscalar ($p > 1$) problems as well (with obvious notational changes, *e.g.*, the $1's$ in the F matrix will be $I_p's$, etc.)

[5] A rigorous proof will be given in Ch. 14, where it will be seen that Σ_i converges to the unique nonnegative definite solution of

$$\bar{\Sigma} = F\bar{\Sigma}F^* + K_pR_eK_p^*$$

that yields a stable closed-loop matrix $F - K_pH$, with

$$K_p = [\bar{N} - F\bar{\Sigma}H^*]R_e^{-1}, \quad R_e = R_y(0) - H\bar{\Sigma}H^*.$$

More importantly, for constant parameter models (including the stationary case), there are fast algorithms (*cf.* Chs. 11 and 13), and especially the doubling algorithm of Ch. 17, that can be used to obtain the limiting value of $K_{p,i}$ with fewer computations: $O(n^2)$ rather than $O(n^3)$ flops per iteration. ◆

9.7 APPROXIMATE NONLINEAR FILTERING

Most practical systems are nonlinear, but sometimes an idealized linear model suffices to describe the system. However, this is often not the case. Examples are nonlinear plant dynamics in control problems, perhaps due to actuator saturation or to a nonlinear measurement process. Here is an example from communication theory.

Consider the case of frequency modulation (FM) where the message $\lambda(t)$ has a first-order Butterworth spectrum, being modeled as the output of a first-order, time-invariant linear system with one real pole driven by continuous-time white noise. This message is passed through an integrator to yield $\theta(t) = \int_0^t \lambda(\tau)d\tau$, which then is used to phase modulate a carrier signal. The model state equations can be written as[6]

$$\begin{bmatrix} \dot{\lambda}(t) \\ \dot{\theta}(t) \end{bmatrix} = \begin{bmatrix} -1/\beta & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda(t) \\ \theta(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mathbf{u}(t), \tag{9.7.1}$$

$$\mathbf{y}(t) = \sqrt{2}\sin[\omega_c t + \theta(t)] + \mathbf{v}(t), \tag{9.7.2}$$

for some noise disturbances $\mathbf{v}(t)$ and $\mathbf{u}(t)$ and some $\beta > 0$. The equation for the state is linear, but the measurement equation is nonlinear.

A more general nonlinear state-space model in continuous time has the form

$$\dot{\mathbf{x}}(t) = f_t[\mathbf{x}(t)] + g_t[\mathbf{x}(t)]\mathbf{u}(t), \tag{9.7.3}$$

$$\mathbf{y}(t) = h_t[\mathbf{x}(t)] + \mathbf{v}(t), \tag{9.7.4}$$

where $\{f_t(\cdot), g_t(\cdot), h_t(\cdot)\}$ are time-variant nonlinear functions. Regardless of the model, the least-mean-squares estimator of the state vector $\mathbf{x}(t)$, at any particular time instant t, is given by the conditional mean (recall the discussion in App. 3.B)

$$E[\mathbf{x}(t)|\mathbf{Y}(t)], \quad \mathbf{Y}(t) = \{\mathbf{y}(\sigma), \ 0 < \sigma < t\}.$$

In general, this is too complicated to calculate, or to implement. While a lot of effort has been expended on the nonlinear problem, and a lot of interesting mathematical results obtained, for practical applications one resorts to certain ad hoc schemes, especially one first suggested by S. F. Schmidt and his colleagues at NASA Ames Research Center in their work on the feasibility studies for navigation and control of the Apollo space capsule.

[6] Applications of extended Kalman filtering to the frequency modulation example (9.7.2) can be found in Anderson and Moore (1979, pp. 200–203), Polk and Gupta (1973), and McBride (1973).

The first step was to discretize the continuous system, thus leading to a nonlinear discrete-time model of the general form (see Prob. 9.19 for one possibility)

$$\mathbf{x}_{i+1} = f_i(\mathbf{x}_i) + g_i(\mathbf{x}_i)\mathbf{u}_i, \tag{9.7.5}$$

$$\mathbf{y}_i = h_i(\mathbf{x}_i) + \mathbf{v}_i, \tag{9.7.6}$$

where the quantities $F_i\mathbf{x}_i$, $H_i\mathbf{x}_i$, and G_i of the linear model (9.2.30) are replaced by $f_i(\mathbf{x}_i)$, $h_i(\mathbf{x}_i)$, and $g_i(\mathbf{x}_i)$, with $f_i(\cdot)$, $h_i(\cdot)$ nonlinear (in general) and $g_i(\cdot)$ nonconstant (in general); $\mathbf{u}_i$, $\mathbf{v}_i$ are zero-mean, white processes, and $\mathbf{x}_0$ is a random variable with mean $\bar{x}_0$. We shall assume $\{\mathbf{u}_i\}$, $\{\mathbf{v}_i\}$, and $\mathbf{x}_0$ are mutually uncorrelated, and that

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 - \bar{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 - \bar{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i\delta_{ij} & 0 & 0 \\ 0 & R_i\delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \end{bmatrix}. \tag{9.7.7}$$

9.7.1 A Linearized Kalman Filter

The first approach was to linearize the state-space equations (9.7.5)–(9.7.6) around a known nominal trajectory x_i^{nom}. A common choice is the unforced solution

$$x_{i+1}^{\text{nom}} = f_i(x_i^{\text{nom}}), \quad x_0^{\text{nom}} = \bar{x}_0. \tag{9.7.8}$$

This defines a deterministic sequence and we can write

$$\mathbf{x}_i = x_i^{\text{nom}} + \Delta\mathbf{x}_i, \tag{9.7.9}$$

where $\Delta\mathbf{x}_i$ measures the perturbation away from the nominal trajectory and is a random variable.

Assuming the functions $\{f_i, g_i, h_i\}$ are smooth enough, and making a first-order Taylor expansion, we obtain

$$f_i(\mathbf{x}_i) \approx f_i(x_i^{\text{nom}}) + F_i\Delta\mathbf{x}_i, \quad h_i(\mathbf{x}_i) \approx h_i(x_i^{\text{nom}}) + H_i\Delta\mathbf{x}_i, \tag{9.7.10}$$

where the matrices F_i and H_i are defined by

$$F_i = \left.\frac{\partial f_i(x)}{\partial x}\right|_{x=x_i^{\text{nom}}}, \quad H_i = \left.\frac{\partial h_i(x)}{\partial x}\right|_{x=x_i^{\text{nom}}}.$$

That is, the (k, j)-th component of F_i is the partial derivative of the k-th component of $f_i(\cdot)$ with respect to the j-th component of x, and similarly for H_i, each derivative being evaluated at x_i^{nom}.

Likewise, taking a zero-th order expansion leads to

$$g_i(\mathbf{x}_i) \approx g_i(x_i^{\text{nom}}) \triangleq G_i. \tag{9.7.11}$$

Then (9.7.5)–(9.7.6) can be approximated as

$$x_{i+1}^{\text{nom}} + \Delta\mathbf{x}_{i+1} = f_i(x_i^{\text{nom}}) + F_i\Delta\mathbf{x}_i + g_i(x_i^{\text{nom}})\mathbf{u}_i,$$

so that

$$\Delta\mathbf{x}_{i+1} = F_i\Delta\mathbf{x}_i + G_i\mathbf{u}_i, \tag{9.7.12}$$

$$\mathbf{y}_i - h_i(x_i^{\text{nom}}) = H_i\Delta\mathbf{x}_i + \mathbf{v}_i. \tag{9.7.13}$$

If x_i^{nom} is known, (9.7.13) is a linear state space model in $\Delta\mathbf{x}_i$ and a standard Kalman filter solution can be readily applied to estimate $\Delta\mathbf{x}_i$. We write it in terms of time- and measurement-update relations:

$$\widehat{\Delta\mathbf{x}}_{i+1|i} = F_i\widehat{\Delta\mathbf{x}}_{i|i}, \quad \widehat{\Delta\mathbf{x}}_0 = 0, \quad P_0 = \Pi_0 \tag{9.7.14}$$

$$\widehat{\Delta\mathbf{x}}_{i|i} = \widehat{\Delta\mathbf{x}}_{i|i-1} + K_{f,i}[\mathbf{y}_i - h_i(x_i^{\text{nom}}) - H_i\widehat{\Delta\mathbf{x}}_{i|i-1}], \tag{9.7.15}$$

$$K_{f,i} = P_{i|i-1}H_i^*(H_iP_{i|i-1}H_i^* + R_i)^{-1}, \tag{9.7.16}$$

$$P_{i+1|i} = F_iP_{i|i}F_i^* + G_iQ_iG_i^*, \quad P_{i|i} = (I - K_{f,i}H_i)P_{i|i-1}.$$

An estimator $\hat{\mathbf{x}}_i$ for the state of the nonlinear model (9.7.5)–(9.7.6) can be found by conditioning both sides of (9.7.9) on the measurements:

$$\hat{\mathbf{x}}_{i|i} = x_i^{\text{nom}} + \widehat{\Delta\mathbf{x}}_{i|i}, \quad \hat{\mathbf{x}}_{i+1|i} = x_{i+1}^{\text{nom}} + \widehat{\Delta\mathbf{x}}_{i+1|i},$$

since x_i^{nom} is known (nonrandom). Therefore, adding (9.7.8) to (9.7.14) and x_i^{nom} to both sides of (9.7.15) results in the following algorithm.

Algorithm 9.7.1 (A Linearized Kalman Filter) *Consider the state-space model (9.7.5)–(9.7.6) with conditions (9.7.7). Consider also a nominal trajectory x_i^{nom} that is determined by solving (9.7.8). An approximate estimator for the state $\mathbf{x}_i$ can be recursively computed as follows. Start with $\hat{\mathbf{x}}_{0|-1} = \bar{x}_0$, $P_{0|-1} = \Pi_0$ and repeat:*

$$\hat{\mathbf{x}}_{i+1|i} = F_i(\hat{\mathbf{x}}_{i|i} - x_i^{\text{nom}}) + f_i(x_i^{\text{nom}}),$$

$$\hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_{i|i-1} + K_{f,i}[\mathbf{y}_i - h_i(x_i^{\text{nom}}) - H_i\hat{\mathbf{x}}_{i|i-1} + H_ix_i^{\text{nom}}],$$

$$K_{f,i} = P_{i|i-1}H_i^*(H_iP_{i|i-1}H_i^* + R_i)^{-1},$$

$$P_{i|i} = (I - K_{f,i}H_i)P_{i|i-1},$$

$$P_{i+1|i} = F_iP_{i|i}F_i^* + G_iQ_iG_i^*.$$

■

The performance of the linearized filter is clearly dependent on the quality of the approximation in (9.7.10)–(9.7.11). For small i, or small $\|g(\mathbf{x}_i)\mathbf{u}_i\|$, the nominal solution may be close to the true trajectory. However, with time the two will depart, often resulting in a breakdown of (9.7.10)–(9.7.11).

9.7.2 Schmidt Extended Kalman Filter (EKF)

S. F. Schmidt then suggested that relinearization about the current estimate might lead to smaller errors than the above (open-loop) linearization with respect to a nominal trajectory. This was a happy idea and, with some variations, is the most widely used nonlinear state-space estimator today. It is known as the Extended Kalman Filter (or

EKF), though the name Schmidt Extended Kalman Filter (Schmidt EKF) would be more accurate.

For this method, we define

$$f_i(\mathbf{x}_i) \approx f_i(\hat{x}_{i|i}) + F_i(\mathbf{x}_i - \hat{x}_{i|i}), \tag{9.7.17}$$

$$h_i(\mathbf{x}_i) \approx h_i(\hat{x}_{i|i-1}) + H_i(\mathbf{x}_i - \hat{x}_{i|i-1}), \tag{9.7.18}$$

$$g_i(\mathbf{x}_i) \approx g_i(\hat{x}_{i|i}) \triangleq G_i, \tag{9.7.19}$$

$$F_i = \left.\frac{\partial f_i(x)}{\partial x}\right|_{x=\hat{x}_{i|i}}, \quad H_i = \left.\frac{\partial h_i(x)}{\partial x}\right|_{x=\hat{x}_{i|i-1}}. \tag{9.7.20}$$

[Recall that $\hat{x}_{i|i}$ denotes the estimate while the boldface notation $\hat{\mathbf{x}}_{i|i}$ denotes the estimator.] Then (9.7.5)–(9.7.6) can be approximated as

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + \underbrace{(f_i(\hat{x}_{i|i}) - F_i\hat{x}_{i|i})}_{\text{known at time } i} + G_i\mathbf{u}_i,$$

$$\mathbf{y}_i - \underbrace{(h_i(\hat{x}_{i|i-1}) - H_i\hat{x}_{i|i-1})}_{\text{known at time } i-1} = H_i\mathbf{x}_i + \mathbf{v}_i,$$

which is a linear state-space model for $\mathbf{x}_i$. Therefore, we can apply the Kalman filter equations as in the case of the linearized Kalman filter to obtain the following equations for the estimators:

$$\hat{\mathbf{x}}_{i+1|i} = F_i\hat{\mathbf{x}}_{i|i} + f_i(\hat{\mathbf{x}}_{i|i}) - F_i\hat{\mathbf{x}}_{i|i} = f_i(\hat{\mathbf{x}}_{i|i}), \tag{9.7.21}$$

$$\begin{aligned}\hat{\mathbf{x}}_{i|i} &= \hat{\mathbf{x}}_{i|i-1} + K_{f,i}[\mathbf{y}_i - h_i(\hat{\mathbf{x}}_{i|i-1}) + H_i\hat{\mathbf{x}}_{i|i-1} - H_i\hat{\mathbf{x}}_{i|i-1}],\\ &= \hat{\mathbf{x}}_{i|i-1} + K_{f,i}[\mathbf{y}_i - h_i(\hat{\mathbf{x}}_{i|i-1})].\end{aligned} \tag{9.7.22}$$

The covariance and gain equations are the same as those of the linearized Kalman filter.

Algorithm 9.7.2 (The Extended Kalman Filter) *Consider the model (9.7.5)–(9.7.6) with conditions (9.7.7). An approximate estimator for the state $\mathbf{x}_i$ can be recursively computed as follows. Start with $\hat{\mathbf{x}}_{0|-1} = \bar{x}_0$, $P_{0|-1} = \Pi_0$ and repeat:*

$$\begin{aligned}\hat{\mathbf{x}}_{i+1|i} &= f_i(\hat{\mathbf{x}}_{i|i}),\\ \hat{\mathbf{x}}_{i|i} &= \hat{\mathbf{x}}_{i|i-1} + K_{f,i}[\mathbf{y}_i - h_i(\hat{\mathbf{x}}_{i|i-1})],\\ K_{f,i} &= P_{i|i-1}H_i^*(H_iP_{i|i-1}H_i^* + R_i)^{-1},\\ P_{i|i} &= (I - K_{f,i}H_i)P_{i|i-1},\\ P_{i+1|i} &= F_iP_{i|i}F_i^* + G_iQ_iG_i^*.\end{aligned}$$

■

Unlike the linearized Kalman filter, observe now that the matrices $\{F_i, H_i, G_i\}$ depend on the measurements and, therefore, the quantities $\{P_i, K_{f,i}\}$ cannot be precomputed. This represents an increased computational load. Moreover, while the linearized Kalman filter depended linearly on the $\{\mathbf{y}_i\}$, this is not the case any more for the extended Kalman filter since $K_{f,i}$ also depends nonlinearly on prior measurements.

9.7.3 The Iterated Schmidt EKF

A third (of many) variation is to employ an iterated version of the Schmidt EKF procedure, where an intermediate iterative procedure is used to compute the gain matrix $K_{f,i}$.

More specifically, let L denote a relatively small positive integer. Then for every time instant i, we compute $K_{f,i}$ as the matrix that is obtained at the end of L iterations of the following form. Let

$$H_i^{(0)} = \left.\frac{\partial h_i(x)}{\partial x}\right|_{x=\hat{x}_{i|i-1}},$$

and repeat for $j = 0$ to L:

$$\begin{aligned}K_{f,i}^{(j)} &= P_{i|i-1}H_i^{(j)*}\left[H_i^{(j)}P_{i|i-1}H_i^{(j)*} + R_i\right]^{-1},\\ \hat{\mathbf{x}}_{i|i}^{(j)} &= \hat{\mathbf{x}}_{i|i-1} + K_{f,i}^{(j)}\left[\mathbf{y}_i - h_i(\hat{\mathbf{x}}_{i|i-1})\right],\\ H_i^{(j+1)} &= \left.\frac{\partial h_i(x)}{\partial x}\right|_{x=\hat{x}_{i|i}^{(j)}}.\end{aligned}$$

At the end of the L steps we compute

$$\hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_{i|i}^{(L)}, \quad \hat{\mathbf{x}}_{i+1|i} = f_i(\hat{\mathbf{x}}_{i|i}),$$

$$P_{i|i} = \left(I - K_{f,i}^{(L)}H_i^{(L)}\right)P_{i|i-1}, \quad P_{i+1|i} = F_iP_{i|i}F_i^* + G_iQ_iG_i^*.$$

9.7.4 Performance of the Approximate Filters

There are almost no useful analytical results on the performance of the EKF. A considerable amount of experimentation and "tuning" is needed to get a reasonable filter. However, this has been done in numerous different applications (see, *e.g.*, those described in the IEEE Press reprint volume edited by Sorenson (1985)).

One technique for seeing whether an EKF is working satisfactorily, is to check the whiteness of the residuals $\mathbf{y}_i - h_i(\hat{\mathbf{x}}_{i|i-1})$, and to compare their actual (sample) covariance with the computed $R_{e,i} = R_i + H_iP_{i|i-1}H_i^*$. Just observing P_i is not enough; P_i may be finite or even decrease to 0, while the true error becomes unbounded.

9.7.5 Other Schemes

Higher-order filters can be designed by retaining more terms in the Taylor series. However, they are not necessarily better than an EKF. Also, more sophisticated filters can be developed that are based on Gaussian sum approximations (see, *e.g.,* Söderström (1994)), statistical linearization (see, *e.g.,* Gelb (1974)), spline approximations, etc.

9.8 BACKWARDS KALMAN RECURSIONS

The Kalman recursions in the earlier sections provide predicted and filtered estimators for the state vector $\mathbf{x}_i$ that are based on the observation data $\{\mathbf{y}_j\}$ up to and including time i. In some cases however, *e.g.,* in the smoothing problems of Ch. 10 (Sec. 10.4), it is of interest to compute *strictly noncausal* estimators of $\mathbf{x}_i$ that are based on future data vectors; in particular,

$$\hat{\mathbf{x}}_i^b \triangleq \text{the l.l.m.s.e. of } \mathbf{x}_i \text{ given } \{\mathbf{y}_{i+1}, \mathbf{y}_{i+2}, \ldots, \mathbf{y}_N\} \triangleq \hat{\mathbf{x}}_{i|i+1}^b, \tag{9.8.1}$$

$$\hat{\mathbf{x}}_{i|i}^b \triangleq \text{the l.l.m.s.e. of } \mathbf{x}_i \text{ given } \{\mathbf{y}_i, \mathbf{y}_{i+1}, \mathbf{y}_{i+2}, \ldots, \mathbf{y}_N\}. \tag{9.8.2}$$

We can readily obtain recursive formulas for computing these so-called backward predicted and filtered estimators by using the backwards Markovian state-space models introduced in Sec. 5.4.2. These results are useful in smoothing problems — see Ch. 10.

9.8.1 Backwards Markovian Representations of $\{\mathbf{y}_i\}$

We showed earlier in Thm. 5.4.2 that a forwards Markovian representation of a WSM process $\{\mathbf{x}_i\}$ of the form

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i, \quad 0 \le i \le N,$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i\delta_{ij} & 0 \\ 0 & \Pi_0 \end{bmatrix},$$

can be associated with a backwards Markovian representation, *viz.,*

$$\mathbf{x}_i = F_{i+1}^b\mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b, \quad 0 \le i \le N,$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i^b \\ \mathbf{x}_{N+1} \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j^b \\ \mathbf{x}_{N+1} \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i^b\delta_{ij} & 0 \\ 0 & \Pi_{N+1} \end{bmatrix}.$$

The matrix F_{i+1}^b is any solution to the equation $F_{i+1}^b\Pi_{i+1} = \Pi_i F_i^*$, with $Q_{i+1}^b = \Pi_i - F_{i+1}^b\Pi_{i+1}F_{i+1}^{b*}$. This fact was used in Sec. 5.4.4 to derive a backwards form of the standard model (9.2.30)–(9.2.31), when $S_i = 0$:

$$\begin{cases} \mathbf{x}_i &= F_{i+1}^b\mathbf{x}_{i+1} + \mathbf{u}_{i+1}^b, \\ \mathbf{y}_i &= H_iF_{i+1}^b\mathbf{x}_{i+1} + H_i\mathbf{u}_{i+1}^b + \mathbf{v}_i \triangleq H_iF_{i+1}^b\mathbf{x}_{i+1} + \mathbf{v}_{i+1}^b, \end{cases} \tag{9.8.3}$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i^b \\ \mathbf{v}_i^b \\ \mathbf{x}_{N+1} \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j^b \\ \mathbf{v}_j^b \\ \mathbf{x}_{N+1} \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i^b\delta_{ij} & Q_i^bH_{i-1}^*\delta_{ij} & 0 & 0 \\ H_{i-1}Q_i^b\delta_{ij} & (R_{i-1} + H_{i-1}Q_i^bH_{i-1}^*)\delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_{N+1} & 0 \end{bmatrix},$$

and where we have defined

$$\mathbf{v}_{i+1}^b \triangleq H_i\mathbf{u}_{i+1}^b + \mathbf{v}_i. \tag{9.8.4}$$

Note that the $\{\mathbf{x}_i\}$ are the same as in the forwards Markovian model. Now starting with the backwards representation (9.8.3), we can find the backwards innovations of $\{\mathbf{y}_i\}$.

9.8.2 Recursions for the Backwards Innovations Process

The arguments we shall use here exactly parallel those used for the forwards innovations recursions in Sec. 9.2, so we shall be brief. Let

$$\hat{\mathbf{y}}_i^b \triangleq \text{the l.l.m.s. estimator of } \mathbf{y}_i \text{ given } \{\mathbf{y}_{i+1}, \ldots, \mathbf{y}_N\},$$

and define backwards innovations as

$$\mathbf{e}_i^b = \mathbf{y}_i - \hat{\mathbf{y}}_i^b = \mathbf{y}_i - H_i\hat{\mathbf{x}}_i^b, \quad 0 \le i \le N, \tag{9.8.5}$$

with $\mathbf{e}_{N+1}^b = \mathbf{y}_{N+1}$, $\hat{\mathbf{x}}_{N+1}^b = 0$, and $\mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i$. Let further $\tilde{\mathbf{x}}_i^b = \mathbf{x}_i - \hat{\mathbf{x}}_i^b$ so that $\mathbf{e}_i^b = H_i\tilde{\mathbf{x}}_i^b + \mathbf{v}_i$. Then

$$R_{e,i}^b \triangleq \|\mathbf{e}_i^b\|^2 = R_i + H_iP_i^bH_i^*, \qquad P_i^b \triangleq \|\tilde{\mathbf{x}}_i^b\|^2.$$

We now expand $\hat{\mathbf{x}}_i^b$ in terms of the innovations $\{\mathbf{e}_j^b\}_{j=0}^{i+1}$,

$$\begin{aligned} \hat{\mathbf{x}}_{i|i+1}^b &= \sum_{j=i+1}^{N} \langle \mathbf{x}_i, \mathbf{e}_j^b\rangle R_{e,j}^{-b}\mathbf{e}_j^b = \sum_{j=i+2}^{N} \langle \mathbf{x}_i, \mathbf{e}_j^b\rangle R_{e,j}^{-b}\mathbf{e}_j^b + \langle \mathbf{x}_i, \mathbf{e}_{i+1}^b\rangle R_{e,i+1}^{-b}\mathbf{e}_{i+1}^b, \\ &= \hat{\mathbf{x}}_{i|i+2}^b + K_{l,i+1}^b\mathbf{e}_{i+1}^b, \quad \text{say,} \end{aligned} \tag{9.8.6}$$

where

$$K_{l,i+1}^b \triangleq \langle \mathbf{x}_i, \mathbf{e}_{i+1}^b\rangle R_{e,i+1}^{-b} = F_{i+1}^bP_{i+1}^bH_{i+1}^*R_{e,i+1}^{-b}. \tag{9.8.7}$$

Moreover, by projecting the state equation (9.8.3) onto $\mathcal{L}\{\mathbf{y}_{i+2}, \ldots, \mathbf{y}_N\}$ we obtain

$$\hat{\mathbf{x}}^b_{i|i+2} = F^b_{i+1}\hat{\mathbf{x}}^b_{i+1|i+2} + \underbrace{\hat{\mathbf{u}}^b_{i+1|i+2}}_{=0},$$

so we can conclude that

$$\hat{\mathbf{x}}^b_i = F^b_{i+1}\hat{\mathbf{x}}^b_{i+1} + K^b_{l,i+1}\mathbf{e}^b_{i+1}, \qquad \hat{\mathbf{x}}^b_{N+1} = 0, \tag{9.8.8}$$

All that remains is to obtain a recursion for P^b_i, which can be done by any of the methods described for the forwards filter in Sec. 9.2.3.

Theorem 9.8.1 (The Backwards Kalman Recursions) *Consider the state-space model (9.2.30)–(9.2.31) with $\langle \mathbf{u}_i, \mathbf{v}_j \rangle = 0$ for any i, j. Let F^b_{i+1} be any solution of $F^b_{i+1}\Pi_{i+1} = \Pi_i F^*_i$ and define $Q^b_{i+1} \triangleq \Pi_i - F^b_{i+1}\Pi_{i+1}F^{b*}_{i+1}$, where Π_i is the state-covariance matrix that is recursively constructed as follows:*

$$\Pi_{i+1} = F_i\Pi_i F^*_i + G_i Q_i G^*_i, \qquad \textit{with initial value} \quad \Pi_0.$$

The backwards innovations (9.8.5) can be computed via $\mathbf{e}^b_i = \mathbf{y}_i - H_i\hat{\mathbf{x}}^b_i$, where the backwards prediction estimator $\hat{\mathbf{x}}^b_i$ of (9.8.1) satisfies the backwards recursion (9.8.8) with

$$K^b_{l,i} = F^b_i P^b_i H^*_i R^{-b}_{e,i}, \qquad R^b_{e,i} = R_i + H_i P^b_i H^*_i,$$

and where P^b_i satisfies the backwards Riccati recursion

$$P^b_i = F^b_{i+1}P^b_{i+1}F^{b*}_{i+1} + Q^b_{i+1} - K^b_{l,i+1}R^b_{e,i+1}K^{b*}_{l,i+1}, \quad P^b_{N|N+1} = \Pi_N. \tag{9.8.9}$$

■

9.8.3 The Filtered Version of the Backwards Kalman Recursions

A reader may object that in the previous discussion we used $\mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i$ rather than $\mathbf{y}_i = H_i F^b_{i+1}\mathbf{x}_{i+1} + \mathbf{v}^b_{i+1}$ as in (9.8.3). We could start with this as well:

$$\mathbf{e}^b_i = \mathbf{y}_i - \hat{\mathbf{y}}^b_i = \mathbf{y}_i - H_i F^b_{i+1}\hat{\mathbf{x}}^b_{i+1|i+1} = H_i F^b_{i+1}\tilde{\mathbf{x}}^b_{i+1|i+1} + \mathbf{v}^b_{i+1},$$

where $\tilde{\mathbf{x}}^b_{i+1|i+1} \triangleq \mathbf{x}_{i+1} - \hat{\mathbf{x}}^b_{i+1|i+1}$. Now the covariance matrix of the backwards innovations is

$$R^b_{e,i} = \|\mathbf{e}^b_i\|^2 = (R_i + H_i Q^b_{i+1}H^*_i) + H_i F^b_{i+1}P^b_{i+1|i+1}F^{b*}_{i+1}H^*_i,$$

where $P^b_{i+1|i+1} \triangleq \langle \tilde{\mathbf{x}}^b_{i+1|i+1}, \tilde{\mathbf{x}}^b_{i+1|i+1} \rangle$.

We thus see that in order to compute the innovations we now need a recursion for $\hat{\mathbf{x}}^b_{i+1|i+1}$. But since (9.8.3)–(9.8.4) is the backwards time version of our standard state-space model, the desired recursion for $\hat{\mathbf{x}}^b_{i+1|i+1}$ is just the backwards time version of the standard Kalman filter,

$$\hat{\mathbf{x}}^b_{i|i} = F^b_{i+1}\hat{\mathbf{x}}^b_{i+1|i+1} + K^b_{p,i}\mathbf{e}^b_i,$$

where

$$K^b_{p,i} = (F^b_{i+1}P^b_{i+1|i+1}H^{b*}_{i+1} + G^b_{i+1}S^b_{i+1})R^{-b}_{e,i}.$$

By inspecting the backwards model (9.8.3)–(9.8.4), we can identify

$$H^b_{i+1} = H_i F^b_{i+1}, \quad G^b_{i+1} = I, \quad \text{and} \quad S^b_{i+1} = Q^b_{i+1}H^*_i.$$

Therefore,

$$K^b_{p,i} = (F^b_{i+1}P^b_{i+1|i+1}F^{b*}_{i+1}H^*_i + Q^b_{i+1}H^*_i)R^{-b}_{e,i}.$$

The recursion for $P^b_{i+1|i+1}$ is then

$$P^b_{i|i} = F^b_{i+1}P^b_{i+1|i+1}F^{b*}_{i+1} + Q^b_{i+1} - K^b_{p,i}R^b_{e,i}K^{b*}_{p,i}, \qquad P^b_{N+1|N+1} = \Pi_{N+1}.$$

These equations lead to the filtered version of the backwards Kalman filter of Thm. 9.8.1.

Theorem 9.8.2 (Filtered Version of the Backwards Recursions) *Consider the same setting as Thm. 9.8.1. The backwards innovations can be computed via $\mathbf{e}^b_i = \mathbf{y}_i - H_i F^b_{i+1}\hat{\mathbf{x}}^b_{i+1|i+1}$, where $\hat{\mathbf{x}}^b_{i+1|i+1}$ satisfies the backwards recursion*

$$\hat{\mathbf{x}}^b_{i|i} = F^b_{i+1}\hat{\mathbf{x}}^b_{i+1|i+1} + K^b_{p,i}\mathbf{e}^b_i, \quad \hat{\mathbf{x}}^b_{N+1|N+1} = 0,$$

with

$$K^b_{p,i} = (F^b_{i+1}P^b_{i+1|i+1}F^{b*}_{i+1}H^*_i + Q^b_{i+1}H^*_i)R^{-b}_{e,i},$$

$$R^b_{e,i} = R_i + H_i Q^b_{i+1}H^*_i + H_i F^b_{i+1}P^b_{i+1|i+1}F^{b*}_{i+1}H^*_i,$$

and where $P^b_{i|i}$ satisfies the backwards Riccati recursion

$$P^b_{i|i} = F^b_{i+1}P^b_{i+1|i+1}F^{b*}_{i+1} + Q^b_{i+1} - K^b_{p,i}R^b_{e,i}K^{b*}_{p,i}, \qquad P^b_{N+1|N+1} = \Pi_{N+1}. \tag{9.8.10}$$

■

The connection between the filters of Thms. 9.8.1 and 9.8.2 is essentially the connection between the filtered and predicted forms of the usual Kalman filter. Here it suffices to mention that this connection is established via the relations

$$\hat{\mathbf{x}}^b_{i|i} = F^b_i\hat{\mathbf{x}}^b_i \text{ and } P^b_i = F^b_{i+1}P^b_{i+1|i+1}F^{b*}_{i+1} + Q^b_{i+1}.$$

9.8.4 UDU^* Factorization of R_y

As with the forwards Kalman filter, the backwards recursions also yield a triangular factorization of R_y. However, this time the factorization has the form UDU^* with U unit upper triangular, rather than the LDL^* form described in Sec. 9.4. The details are left to Prob. 9.26.

9.9 COMPLEMENTS

Recursive solutions to least-squares problems are not of recent origin. Gauss was forced to invent them to handle the vast calculations he undertook in order to help astronomers locate the asteroid Ceres. His work dealt with the discrete-time model (9.1.1)–(9.1.2), where, however, the state $\mathbf{x}_i$ was constant (*i.e.*, F_i was the identity matrix and G_i was zero). Given hindsight one can generalize this work to handle dynamics and, for example, Rosenbrock (1965) has done so in an interesting note.

The Fundamental Paper of Kalman (1960a). The Kalman filter recursions in Thm. 9.2.1 were clearly and elegantly presented in Kalman (1960a). Kalman used a geometric formulation, and arguments very close to ours: in particular, the fact that (in our notation) $\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i$ is orthogonal to $\mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$ (Kalman (1960a, p. 38)). He also noted that the process $\{\mathbf{e}_i\}$ is white, but (for some reason) he did not note the *causal equivalence* of $\{\mathbf{e}_i\}$ to the process $\{\mathbf{y}_i\}$, which would have identified it as the *innovations process* of $\{\mathbf{y}_i\}$ and allowed a solution like ours, which as stated earlier goes back conceptually to Bode and Shannon (1950), Zadeh and Ragazzini (1950), and the earlier works of Wold (1938) and Kolmogorov (1939,1941a,1941b). [Interestingly, the Bode-Shannon paper is cited in Kalman (1960a) but for a different reason — see below.] Had the innovations approach been taken, a solution of the smoothing problem would actually have been well in hand (see Ch. 10), and not avoided as in Kalman (1960a, p. 39) with the remark that (in our notation) while $\hat{\mathbf{u}}_{i+s|i}$ is clearly zero when $s > 0$, "if $s < 0$, considerable complications result in evaluating this term. We shall only consider the case $s \geq 0$." So also in Kalman (1963b, p. 305): "The solution of the filtering problem is given in a convenient form only if $t_1 \geq t$." In fact, it took a few years after Kalman (1960a) for the first smoothing results to appear, with fairly complicated derivations.

Perhaps one reason for not using the innovations was that the corresponding concept for nonstationary continuous-time processes is less obvious, and relies heavily on the assumption of additive white noise in the observation process (see Sec. 16.9). So the paper (Kalman and Bucy (1961)) , where the continuous-time results are derived, uses a different approach (via the covariance functions and the Wiener-Hopf equation) from the one in Kalman (1960a).

Given the passage of time, it may be of interest to cite here what its author regarded as the highlights of Kalman (1960a):

(i) the use of orthogonal projection as a way of characterizing the optimal estimators.

Unfortunately Kalman's lead in this matter was not followed by most of the early researchers in this area, despite the prevalence of the geometric point of view on second-order stochastic processes, not only in mathematical textbooks such as Doob (1953) and Loève (1963), but also in the more engineering-oriented book of Yaglom (1962).

(ii) the use of models for random processes. We quote:

Following, in particular, Bode and Shannon [3], arbitrary random signals are represented (up to second order average statistical properties) as the output of a linear dynamic system excited by independent or uncorrelated random signals ("white noise"). This is a standard trick in the engineering applications of the Wiener theory [2–7]. The approach taken here differs from the conventional one only in the way in which linear dynamic systems are described. We shall emphasize the concepts of *state* and *state transition*; in other words, linear systems will be specified by systems of first-order difference (or differential) equations. This point of view is natural...

In fact, Kalman fortunately went well beyond Bode and Shannon, who only used the causal and causally invertible transfer function model defined by the canonical factorization of the power spectral density function; *i.e.*, this is the filter that defines the innovations process, the use of which makes the rest of the estimation problem

easy. What Kalman did was to start (not with the power spectral density, but) with a causal model for the process; in fact, had the model also been causally invertible, the state estimation problem would have been trivial. As a matter of fact (see Sec. 9.2.5), the Kalman filter solution is really a way of converting a given causal model to one that is also causally invertible. The key contribution was to start with a process model, and moreover, to describe it in state-space form; explicit formation of the covariance functions was avoided in the solution of the problem (see Fig. 8.1).

(iii) "with the state-transition method, a single derivation covers a large variety of problems: growing and infinite memory filters, stationary and nonstationary statistics, etc."

This is certainly a strength of the state-space model. On the other hand, it is in a different sense a weakness because we would expect simpler/faster algorithms when the system is time-invariant. In fact, this is true — see Ch. 11; in fact, as noted there, such fast algorithms were first introduced (ca 1947) as a way of reducing the computational burden of the conventional way of solving the Wiener-Hopf equation.

(iv) the fact "that the Wiener problem is the *dual* of the noise-free optimal regulator problem, which has been solved previously by the author, using the state-transition method to great advantage. The mathematical background of the two problems is identical – this had been suspected all along, but until now the analogies have never been made explicit."

In Kalman (1960a) and in almost all the literature, the possibility of a dual association is noted after both (the estimation and the control) problems have been independently solved. In Ch. 15, we shall first introduce duality in the general linear space context, and then use this framework to, among other results, solve a variety of control problems.

(v) Applications: "The power of the new method is most apparent in theoretical investigations and in numerical answers to complex practical problems. In the latter case, it is best to resort to machine computation. Examples of this type will be discussed later."

As for theory, the duality was used to study the asymptotic behavior of the Kalman filter, an issue studied in detail in Ch. 14. On the practical side, a nice historical review by McGee and Schmidt (1985) describes how "Kalman's work was introduced at a near-perfect time and why this near-ideal solution to the midcourse navigation problem might have gone undiscovered except for a fortunate meeting between Dr. Schmidt (one of the present authors) and Kalman." This review gives a fascinating account of the navigation and guidance problems being studied at NASA Ames Research Laboratory in Mountain View, California, in connection with President John Kennedy's proposed Apollo moon-landing project. Schmidt and his colleagues linearized the relevant equations of motion and introduced several clever ideas (*e.g.*, the decomposition of the original Kalman filter formulas into separate measurement-update and time-update steps, the use of array algorithms, the concept of the extended Kalman filter). As Kalman has often said, he himself had no such applications in mind. For example, in Kalman (1960b, p. 490)

he writes: "We must bear in mind that the [state-space] model is a mathematical fiction: it is merely a representation of (presumably empirically obtained) auto- and cross-correlation functions of signal and noise. This representation is standard in the engineering theory of filtering and prediction." For more recent perspectives, see Kalman (1965,1978).

The Contributions of Swerling. In this connection, it is important to mention the work of Swerling (1959,1968,1971), who proposed deterministic nonlinear least-squares as a "stagewise differential correction procedure for satellite tracking and prediction." Of course, linearization around the most recent estimate is necessary to obtain a tractable procedure and so Swerling in effect also developed the "Extended Kalman Filter" of Sec. 9.7.2. State-space models and stochastic descriptions were not mentioned in Swerling's first papers. However, given the equivalence (*cf.* Sec. 3.5) between deterministic and stochastic problems, one can recover the Kalman results from Swerling's (as explained in Swerling (1971)). Still, while Swerling's contributions were timely, and relevant and important (we recall that Kalman's investigation was a pure theoretical exercise), his paper was less elegant and harder to appreciate. Moreover, he did not have the good fortune to meet with Dr. Schmidt at just the right time (*cf.* Ecclesiastes, Ch. 9, v. 11).[7] In his editorial comments, Sorenson (1985, pp. 13–14) gives a balanced commentary on the two papers, which were both reprinted in that volume. We may also mention Sorenson (1970) as a nice review of the progression from Gauss' work on recursive least-squares to Kalman's papers.

Markovian Representations. The fact that our state-space model has the properties that make it Markovian in the sense described in Sec. 5.4 is critical in the derivations. If, for example, $\mathbf{x}_0$ were correlated with $\{\mathbf{u}_i, i \geq 0\}$, or if $\{\mathbf{v}_i\}$ were not a white process, our arguments would break down; in such cases, we should reformulate the model (*e.g.*, by introducing additional variables) in order to get a Markovian model. The Markov property is not mentioned explicitly in Kalman (1960a) and Kalman and Bucy (1961); however, it is explicit in Kalman (1963b, which appeared as a RIAS report in 1961). As already noted in Sec. 5.5, the first to emphasize the importance of Markov processes in signal estimation (and related problems, such as signal detection) was R. L. Stratonovich in the former USSR. Most of Stratonovich's work on filtering was on continuous-time problems, which we shall discuss in the notes to Ch. 16.

Other Derivations. The timeliness of the work (Kalman (1960a)) from the point of view of applications, as well as the relative unfamiliarity of the geometric approach, led to numerous other derivations, based on concepts/ideas more familiar in one context or the other, *e.g.*, via dynamic programming, invariant embedding, maximum likelihood, etc. Jazwinski (1970) is a convenient single source for most of these alternative derivations; it suffices perhaps only to add Duncan and Horn (1972) for a method based on reduction to deterministic least-squares. We should also mention here studies on

[7] "I returned and saw under the sun, that the race is not for the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favor to men of skill; but time and chance happeneth to them all." [Ecclesiastes, Ch. 9, v. 11.]

more general so-called descriptor state-space models — see Nikoukhah, Campbell, and Delebecque (1999).

Sec. 9.6. The Kalman Filter Given Covariance Data. In this section, it was noted that since the innovations representation of a process is unique, the innovations representation (and, hence, also the one-step predictor recursion) can be directly written down from the covariance specifications in state-space form, without having first to determine a state-space model. Thus both specifications, in terms of covariances or in terms of state-space models, are seen to be equivalent, not just in that they give the same final answer (because, of course, they must), but in that their solutions involve comparable amounts of work. The choice between them lies purely in whether state-space models or covariance specifications are more readily at hand. This fact is not widely appreciated and the literature contains many discussions of attempts to "identify" state-space models from covariance data so as to be able to use a Kalman filter.

Nevertheless, "modeling" is, as in all subjects, a thorny problem that we do not study in this book. However, we should say a few more words about it here. State-space models are often at hand in aerospace problems, where we may have enough information to write down the equations of motion, whether they be time-invariant or time-variant. But there are many ways of doing this and many assumptions and choices to be made. For example, the choice of the proper number of states to model a given problem adequately is not always an easy one. Now, in many problems of industrial process control and communications, it is generally impossible to write down state equations (as is clear if we try to do so for a large power grid, or chemical plant, or a telephone-line channel) and recourse has to be to terminal measurements, in order to estimate the covariance function or power spectrum of the channel output. Of course, covariance and especially spectral estimation is again a vast subject, which we shall not enter into here. But even if we assume that good estimates are somehow available, $R_y(i, j)$ will be available only as a numerical function of i and j and not in the factored form (9.6.1); therefore getting the matrices $\{H_i, \Phi(i, j+1), N_i, \Pi, R_i\}$ involves a further step of approximation.

The conceptual value of the covariance-based formulas is in rebuilding the ties to the Wiener-Hopf theory and the importance of canonical factorization. One immediate application is to the factorization problem for symmetric polynomials, as we showed in the text (Remark 9). Another was the fact that later (than 1931) results on the Wiener-Hopf equation of Ambartsumian (1943) and Chandrasekhar (1947a,b) could be nicely combined with the state-space model, as we shall see in Chs. 11, 13, 16, and 17. Another application is to adaptive filtering — see Sayed and Kailath (1994b).

Sec. 9.7. Approximate Nonlinear Filtering. The idea of the extended Kalman filter was originally proposed by S. F. Schmidt (see, *e.g.*, S. F. Schmidt (1970), G. T. Schmidt (1976) and also Bellantoni and Dodge (1967)). As mentioned in the text, higher-order filters can be obtained by retaining more terms in the Taylor series (see, *e.g.*, Schweppe (1973), Gelb (1974), and Maybeck (1979, 1982)). However, these filters are not necessarily better than an EKF. Also, more sophisticated filters can be developed that are based on Gaussian sum approximations, statistical linearization, spline approximations, and other variations (see, *e.g.*, Jazwinski (1970), Sage and Melsa (1971), Bucy and Senne

(1971), de Figueiredo and Jan (1971), Sorenson and Alspach (1971), Alspach and Sorenson (1972), Willsky (1974), and others). Bucy and Senne (1971) attempted to directly propagate the conditional density functions via Bayes' rule, but even with a sparse grid the computational burden was too high. Very recently, the use of modern sampling techniques (rejection sampling, sequential importance sampling, weighted resampling) has made this method more feasible. We may mention the papers of Smith and Gelfand (1992), Gordon, Salmond, and Smith (1993), Gordon, Salmond, and Ewing (1995), and Chen and Liu (2000).

Some Practical Issues. As mentioned several times already, the Kalman filter recursions have been very widely applied. Of course, several practical issues and difficulties arise in actual applications. There is a vast literature on these aspects, which is very difficult to adequately capture in a textbook. One of the most recent efforts is in the book of Grewal and Andrews (1993); among earlier efforts we especially mention the book edited by Gelb (1974), and the volume edited by Sorenson (1985). These references, and the journal literature (*e.g.*, fine surveys by Athans (1971), Hutchinson (1984), Gelb (1986), and Faurre (1991)), should be consulted by practitioners — see also Probs. 9.23–9.24.

One source of difficulty is uncertainty as to the values of the variance parameters $\{\Pi_0, Q_i, R_i\}$. Fortunately, it turns out that one solution is to use values that are upper bounds on these quantities – then it is easy to show that the actual state error variances will be less than that predicted by the solution of the Riccati recursion for the assumed (higher) variance parameters.

If we desire less conservative solutions, then we can attempt to "tune" the filter on-line by suitably estimating the parameters $\{Q, R\}$, now assumed to be constant (time-variant parameters are generally harder to estimate from data). One of the most successful techniques is based on checking how close the pseudo-innovations sequence obtained by using a trial $\{Q, R\}$ pair is to being white. However further reflection shows that when $\{Q, R\}$ are unknown, it will be more efficient to directly estimate the Kalman gain from the given data — in other words, instead of identifying a model with known $\{F, H\}$ but unknown $\{G, Q, R\}$, we estimate the innovations model (*cf.* Sec. 9.2.5), which has fewer parameters, in fact only K_p. This topic is discussed in Mehra (1970).

More complicated to study is the issue of uncertainties in the system coefficients $\{F, G, H\}$ — one of the few general conclusions is that for constant parameter unstable systems, such uncertainties can often lead to unbounded error variances (see Sec. 14.1.4 and also Prob. 9.24).

■ PROBLEMS

9.1 **(A simple moving average process)** Reconsider Prob. 4.2. Use a state-space model to re-establish the result

$$\hat{\mathbf{y}}_{k+1|k} = \frac{k+1}{k+2}(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}), \quad k \geq 0.$$

9.2 **(Estimation with prior information)** Consider a scalar-valued random variable $\mathbf{x}$ with variance σ_x^2 and mean $\bar{x}$, and $(N+1)$ noisy observations

$$\mathbf{y}(i) = \mathbf{x} + \mathbf{v}(i), \qquad i = 0, 1, 2, \ldots, N,$$

where $\mathbf{v}(i)$ is a zero-mean white-noise sequence that is uncorrelated with $\mathbf{x}$ and has variance σ_v^2. Show that the l.l.m.s. estimator of $\mathbf{x}$ given the $\{\mathbf{y}(i), 0 \leq i \leq N\}$ is a weighted combination of the form

$$\hat{\mathbf{x}}_{|N} = \alpha(N)\left[\frac{1}{N+1}\sum_{j=0}^{N}\mathbf{y}(i)\right] + \beta(N)\bar{x},$$

for some scalars $\alpha(N)$ and $\beta(N)$ that are dependent on N. Determine $\{\alpha(N), \beta(N)\}$. What happens when $N \to \infty$? What about the case $\sigma_x^2/\sigma_v^2 \to \infty$ (*i.e.*, the signal-to-noise ratio becomes significantly large with N finite)? [*Hint.* Use a state-space model.]

9.3 **(Nonzero-mean processes)** Consider the state-space model (9.2.30) but with the $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$ random variables such that

$$E\left(\begin{bmatrix}\mathbf{u}_i - \bar{u}_i \\ \mathbf{v}_i - \bar{v}_i \\ \mathbf{x}_0 - \bar{x}_0 \\ 1\end{bmatrix}\begin{bmatrix}\mathbf{u}_j - \bar{u}_j \\ \mathbf{v}_j - \bar{v}_j \\ \mathbf{x}_0 - \bar{x}_0\end{bmatrix}^*\right) = \begin{bmatrix}Q_i\delta_{ij} & S_i\delta_{ij} & 0 \\ S_i^*\delta_{ij} & R_i\delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \\ 0 & 0 & 0\end{bmatrix}.$$

Show that the innovations $\{\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i\}$ can be recursively computed as follows: start with $\hat{\mathbf{x}}_0 = \bar{x}_0$ and repeat for $i \geq 0$,

$$\mathbf{e}_i = \hat{\mathbf{y}}_i - H_i\hat{\mathbf{x}}_i - \bar{v}_i, \quad \hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_i + K_iR_{e,i}^{-1}\mathbf{e}_i + G_i\bar{u}_i,$$

where $R_{e,i} = H_iP_iH_i^* + R_i$, $K_i = F_iP_iH_i^* + G_iS_i$, and P_i satisfies

$$P_{i+1} = F_iP_iF_i^* - K_iR_{e,i}^{-1}K_i^* + G_iQ_iG_i^*, \qquad P_0 = \Pi_0.$$

What is $E\tilde{\mathbf{x}}_i$?

9.4 **(Ill-conditioned Riccati recursion)** Consider the standard state-space model of Thm. 9.2.1 and assume $S_i = 0$ and $R_i = 0$. Show that $P_{i|i}$ is singular.

9.5 **(Tracking a ramp trajectory)** Consider noisy measurements of a ramp trajectory, $\mathbf{y}(i) = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 i + \mathbf{v}(i)$, $i \geq 0$, where the initial value $\boldsymbol{\alpha}_0$ and the slope $\boldsymbol{\alpha}_1$ are independent zero-mean random variables with variances σ_0^2 and σ_1^2, respectively. The measurement noise $\{\mathbf{v}(i)\}$ is a zero-mean random process with variance σ_v^2 and uncorrelated with both $\{\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1\}$.

(a) Write down a state-space model in standard form for the process $\mathbf{y}(i)$.

(b) Derive a recursive algorithm for estimating $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ from measurements $\{\mathbf{y}(j)\}$ for $j = 0, 1, 2, \ldots$. That is, compute $\hat{\boldsymbol{\alpha}}_{0|i}$ and $\hat{\boldsymbol{\alpha}}_{1|i}$ recursively.

(c) Determine the resulting m.m.s.e. for $\hat{\boldsymbol{\alpha}}_{0|i}$ and $\hat{\boldsymbol{\alpha}}_{1|i}$, *viz.*, $\|\tilde{\boldsymbol{\alpha}}_{0|i}\|^2$ and $\|\tilde{\boldsymbol{\alpha}}_{1|i}\|^2$.

9.6 **(Estimating a bias term)** Consider the state-space model

$$\mathbf{x}_{i+1} = F\mathbf{x}_i + G\mathbf{u}_i, \quad \mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i + b,$$

where $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i\}$ are zero-mean uncorrelated random variables with variances Π_0, Q_i, and R_i, respectively. Moreover, the processes $\{\mathbf{u}_i, \mathbf{v}_i\}$ are white. The vector b is deterministic and represents an unknown bias term. Develop a recursive algorithm for estimating both the state vector and b, $\{\mathbf{x}_i, b\}$, given the observations $\{\mathbf{y}_j,\ j = 0, 1, \ldots, i-1\}$. [*Hint.* Introduce the extended state vector $\text{col}\{\mathbf{x}_i, b\}$. A way for carrying out the calculations efficiently was suggested by Friedland (1969).]

9.7 **(Defective sensor measurements in state space)** Consider the state equation $\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i$ and assume we have L sensor measurements for each time instant i of the form

$$\mathbf{y}_{i,k} = H_i\mathbf{x}_i + \mathbf{v}_{i,k}, \quad k = 1, 2, \ldots, L.$$

All random processes are zero-mean. Moreover, the processes $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_{i,k}\}$ are mutually uncorrelated with variances $\{\Pi_0, Q_i, R_{i,k}\}$, respectively, and the $\{\mathbf{u}_i, \mathbf{v}_{i,k}\}$ are white. At each time instant i, only one of the sensor measurements is retained with probability p_k. All other measurements are discarded. The retained measurement is denoted by $\mathbf{z}_i$. After $N+1$ time instants, we collect $(N+1)$ such measurements, $\{\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_N\}$. Derive a recursive Kalman-type algorithm for estimating $\mathbf{x}_i$ given the $\{\mathbf{z}_j, 0 \le j \le i-1\}$. How will the answer change if the sensor noises $\{\mathbf{v}_{i,k}\}$ were still white but correlated among each other? [*Hint.* Recall Prob. 3.16.]

9.8 **(Singular output noise)** Consider a standard state-space model as in Thm. 9.2.1 and assume it is time-invariant with $S = 0$ and $p \le n$ (*i.e.*, at most as much outputs as states). We further assume that R is singular, say with rank $r < p$ (where $p \times p$ is the size of R). Let U be a $p \times p$ unitary matrix that reduces R to the form $URU^* = (\bar{R} \oplus 0)$, where $\bar{R}$ is $r \times r$ and nonsingular. We want to show that the state vector can be estimated by using a Kalman filter of order r rather than n.

(a) Use U as a similarity transformation and apply it to the standard state-space model. Denote the transformed signals by $\{\bar{\mathbf{x}}_i, \bar{\mathbf{y}}_i, \bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$, where $\bar{\mathbf{x}}_i = U\mathbf{x}_i$. Similarly for the other variables. Verify that $\bar{\mathbf{y}}_i$ can be decomposed into two parts, $\bar{\mathbf{y}}_i = \text{col}\{\bar{\mathbf{y}}_{i,1}, \bar{\mathbf{y}}_{i,2}\}$, where $\bar{\mathbf{y}}_{i,2}$ is noise-free. That is, verify that we can write

$$\begin{bmatrix} \bar{\mathbf{y}}_{i,1} \\ \bar{\mathbf{y}}_{i,2} \end{bmatrix} = \begin{bmatrix} \bar{H}_1 \\ \bar{H}_2 \end{bmatrix} \bar{\mathbf{x}}_i + \begin{bmatrix} \bar{\mathbf{v}}_{i,1} \\ 0 \end{bmatrix},$$

for some $\{\bar{H}_1, \bar{H}_2, \bar{\mathbf{v}}_{i,1}\}$. What is the variance of $\bar{\mathbf{v}}_{i,1}$?

(b) Assume $\bar{H}_2$ is full rank and construct a matrix C such that

$$\begin{bmatrix} C \\ \bar{H}_2 \end{bmatrix} \quad \text{is square and invertible.}$$

Show that $\hat{\mathbf{x}}_i$ can be determined by using an r-th order Kalman filter.

Remark. One should be careful in using such reductions since the assumption of singular R is a very delicate one. ◆

9.9 **(A different state-space model)** In some applications, discretization leads to the model

$$\mathbf{x}_{i+1} = F\mathbf{x}_i + G\mathbf{u}_i, \quad \mathbf{y}_i = H\mathbf{x}_i + H_1\mathbf{x}_{i-1} + \mathbf{v}_i, \quad i \ge 0,$$

where $\{\mathbf{u}_i, \mathbf{v}_i\}$ satisfy the conditions of the standard state-space model of Thm. 9.2.1. Let $\mathbf{z}_0 = \text{col}\{\mathbf{x}_{-1}, \mathbf{x}_0\}$ denote the initial state vector with variance matrix Π_0 and assume that it is uncorrelated with all $\{\mathbf{u}_i, \mathbf{v}_i\}$. Find recursions for computing the innovations.

9.10 **(A modified state-space model)** Consider the modified model

$$\mathbf{x}_{i+1} = F\mathbf{x}_i + G_1\mathbf{u}_i + G_2\mathbf{u}_{i+1}, \quad i \ge 0,$$
$$\mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i,$$

with zero-mean uncorrelated random variables $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i\}$ such that $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = Q_i\delta_{ij}$, $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = R_i\delta_{ij}$, $\langle \mathbf{x}_0, \mathbf{x}_0 \rangle = \Pi_0$. Find recursive equations for $\hat{\mathbf{x}}_{i|i-1}$ and $P_{i|i-1}$.

9.11 **(Nonsingular Π_0)** Consider the standard state-space model (9.2.30)–(9.2.31). Assume $\Pi_0 > 0$. Show that $P_{i|i} > 0$ for all $i \ge 0$. What about P_i?

9.12 **(Two-step prediction)** Consider the standard state-space model (9.2.30)–(9.2.31) with $S_i = 0$.

(a) Show that $P_{i+1|i-1} = F_iP_iF_i^* + G_iQ_iG_i^*$.

(b) By working out and using a relation between $\tilde{\mathbf{x}}_{i+1|i-1}$ and $\tilde{\mathbf{x}}_{i+1}$, show also that

$$P_{i+1|i-1} = P_{i+1} + F_iP_iH_i^*[R_i + H_iP_iH_i^*]^{-1}H_iP_iF_i^*.$$

9.13 **(Filtered residuals)** Consider a process $\mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i$, where $\{\mathbf{v}_i\}$ is a white-noise zero-mean process with covariance matrix R_i and uncorrelated with the zero-mean process $\{\mathbf{x}_i\}$. The filtered residuals of $\{\mathbf{y}_i\}$ are defined as $\nu_i = \mathbf{y}_i - H_i\hat{\mathbf{x}}_{i|i}$. Show that the $\{\nu_i\}$ form a white-noise sequence with covariance matrix $R_{\nu,i} = R_i - H_iP_{i|i}H_i^*$.

9.14 **(A backwards Kalman filter)** Consider a state-space model of the form

$$\mathbf{x}_i^d = F_i^*\mathbf{x}_{i+1}^d + H_i^*\mathbf{u}_i^d, \qquad \mathbf{z}_i^d = G_i^*\mathbf{x}_{i+1}^d + \mathbf{v}_i^d,$$

where $\{\mathbf{u}_i^d, \mathbf{v}_i^d\}$ are uncorrelated white-noise sequences with variances $\{R_i^d \ge 0,\ Q_i^d > 0\}$; moreover, both are uncorrelated with $\mathbf{x}_{N+1}^d$, whose variance we denote by $P_{N+1}^d \ge 0$. All variables are zero-mean. Let $\{\hat{\mathbf{x}}_{i+1|i+1}^d, \hat{\mathbf{z}}_i^d\}$ denote the l.l.m.s.e. of $\{\mathbf{x}_{i+1}^d, \mathbf{z}_i^d\}$ given $\{\mathbf{z}_{i+1}^d, \ldots, \mathbf{z}_N^d\}$. Show that $\hat{\mathbf{x}}_{i|i}^d$ can be computed recursively as follows:

$$\hat{\mathbf{x}}_{i|i}^d = F_i^*\hat{\mathbf{x}}_{i+1|i+1}^d + K_i^dR_{e,i}^{-d}\mathbf{e}_i^d, \quad \hat{\mathbf{x}}_{N+1|N+1}^d = 0,$$
$$\mathbf{e}_i^d = \mathbf{z}_i^d - \hat{\mathbf{z}}_i^d = \mathbf{z}_i^d - G_i^*\hat{\mathbf{x}}_{i+1|i+1}^d,$$

where

$$R_{e,i}^d \triangleq \|\mathbf{e}_i^d\|^2 = G_i^*P_{i+1|i+1}^dG_i + Q_i^d, \quad K_i^d \triangleq \langle \mathbf{x}_i^d, \mathbf{e}_i^d \rangle = F_i^*P_{i+1|i+1}^dG_i,$$

and $P_{i|i}^d = \|\tilde{\mathbf{x}}_{i|i}^d\|^2$ satisfies the backwards Riccati recursion

$$P_{i|i}^d = F_i^*P_{i+1|i+1}^dF_i + H_i^*R_i^dH_i - K_i^dR_{e,i}^{-d}K_i^{d*}, \quad P_{N+1|N+1}^d = P_{N+1}^d.$$

9.15 **(Colored noise)** Consider the usual state-space model except that the measurement noise is nonwhite:

$$\begin{aligned} \mathbf{x}_{i+1} &= F\mathbf{x}_i + G\mathbf{u}_i, \quad i \geq 0, \\ \mathbf{y}_i &= H\mathbf{x}_i + \mathbf{n}_i, \\ \mathbf{n}_{i+1} &= A\mathbf{n}_i + \mathbf{v}_i, \end{aligned}$$

where $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0, \mathbf{n}_0\}$ are zero-mean uncorrelated random variables with $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = Q_i \delta_{ij}$, $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = I\delta_{ij}$, $\langle \mathbf{x}_0, \mathbf{x}_0 \rangle = \Pi_0$, $\langle \mathbf{n}_0, \mathbf{n}_0 \rangle = I$.

(a) Show that we can write the innovations of $\mathbf{y}_{i+1}$ as

$$\mathbf{e}_{i+1} = (\mathbf{y}_{i+1} - A\mathbf{y}_i) - \bar{H}\hat{\mathbf{x}}_{i|i} = \bar{H}\tilde{\mathbf{x}}_{i|i} + HG\mathbf{u}_i + \mathbf{v}_i,$$

where $\bar{H} \triangleq HF - AH$.

(b) Find expressions for $< \mathbf{e}_i, \mathbf{e}_i >$ and $< \mathbf{x}_{i+1}, \mathbf{e}_i >$ in terms of $P_{i-1|i-1}$ and given constants. Use this to obtain a recursion for the predicted state estimators $\hat{\mathbf{x}}_{i+1|i}$.

9.16 **(A formula for $\det P_{i|i}$)** Show that $\det P_{i|i} = (\det P_i)(\det R_i)/\det R_{e,i}$. [*Hint.* See (9.3.6).]

9.17 **(Invertible error covariance matrix)** Refer to the discussion that led to Lemma 9.5.1 and assume a time-invariant state-space model $\{F, G, H, R, Q, S\}$ with $S = 0$ and F singular. Show that the error covariance matrix P_i will be invertible if $\Pi_0 > 0$, $R > 0$, and the pair $\{F, GQ^{\frac{1}{2}}\}$ is controllable.

9.18 **(Fixed-point smoothing)** Consider the model (9.2.30)–(9.2.31) with $S_i = 0$, and let i_0 be a *fixed* time instant. Also let $\hat{\mathbf{x}}_{i_0|i}$ denote the l.l.m.s. estimator of the state $\mathbf{x}_{i_0}$ given the observations $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\}$ with $i > i_0$. We are interested in finding a recursive update that relates $\hat{\mathbf{x}}_{i_0|i}$ and $\hat{\mathbf{x}}_{i_0|i+1}$, as well as a recursive formula that relates $P_{i_0|i}$ and $P_{i_0|i+1}$, where $P_{i_0|i}$ is the covariance matrix of the error $(\mathbf{x}_{i_0} - \hat{\mathbf{x}}_{i_0|i})$. This is known as a *fixed-point smoothing problem.*

Argue that the answer can be obtained by using the following augmented state-space model: define the variable $\mathbf{z}_{i+1} = \mathbf{x}_{i_0}$ and write

$$\begin{bmatrix} \mathbf{x}_{i+1} \\ \mathbf{z}_{i+1} \end{bmatrix} = \begin{bmatrix} F_i & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{bmatrix} + \begin{bmatrix} G_i \\ 0 \end{bmatrix} \mathbf{u}_i,$$

$$\mathbf{y}_i = \begin{bmatrix} H_i & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{bmatrix} + \mathbf{v}_i, \quad \text{for} \quad i \geq i_0.$$

Use the above model to derive the desired recursions and specify the necessary initial conditions.

Remark. This approach is due to Zachrisson (1969) and Willman (1969). A more direct innovations solution is in fact simpler — see Ch. 10; however the formulation here will be used in Sec. 17.4.4.] ◆

9.19 **(Discretization of a continuous-time model)** Consider the nonlinear state-space model (9.7.3)–(9.7.4), where $\mathbf{u}(t)$, $\mathbf{v}(t)$ are white-noise uncorrelated signals with covariance matrices $Q(t)$ and $R(t)$, respectively.

We subdivide the time axis t into small intervals of width T each, and define $\mathbf{x}_0 = \mathbf{x}(0)$,

$$\mathbf{x}_i = \mathbf{x}(iT), \quad \mathbf{y}_i = \mathbf{y}(iT), \quad f_i = f_{(t=iT)}, \quad h_i = h_{(t=iT)}, \quad g_i = g_{(t=iT)}.$$

Also, $R_i = R(iT)$ and $Q_i = Q(iT)$. We further approximate $\dot{\mathbf{x}}(t)$ by

$$\dot{\mathbf{x}}(t) \approx \frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{T},$$

and define (see Sec. 16.1.2 for further motivation)

$$\mathbf{u}_i \triangleq \frac{1}{T} \int_{iT}^{(i+1)T} \mathbf{u}(t)dt.$$

Likewise for $\mathbf{v}_i$. Verify that these definitions lead to the discretized model

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i + T \cdot f_i(\mathbf{x}_i) + T \cdot g_i(\mathbf{x}_i)\mathbf{u}_i \\ \mathbf{y}_i &= h_i(\mathbf{x}_i) + \mathbf{v}_i, \end{aligned}$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 - \bar{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 - \bar{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} \frac{1}{T}Q_i\delta_{ij} & 0 & 0 \\ 0 & \frac{1}{T}R_i\delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \end{bmatrix}.$$

9.20 **(Wiener and Kalman filtering)** A zero-mean unit-variance white-noise stationary process $\{\mathbf{u}(i)\}$ is applied to the system $H(z) = 1/(z+0.5)$. We denote the output by $\{\mathbf{x}(i)\}$. Noisy measurements of $\mathbf{x}(i)$ are available, say $\mathbf{y}(i) = \mathbf{x}(i) + \mathbf{v}(i)$, where $\{\mathbf{v}(i)\}$ is also a zero-mean unit-variance white-noise stationary process that is further uncorrelated with $\{\mathbf{u}(i)\}$.

(a) By formal use of the Wiener formulas of Ch. 7, find the transfer function of the optimum linear filter for estimating $\mathbf{x}(k)$ given $\{\mathbf{y}(j), -\infty < j \leq k\}$.

(b) Repeat part (a) for estimating $\mathbf{x}(k+1)$ given the same observations.

(c) Write a state-space model for the system.

(d) Write down the Kalman filter equations for recursively computing $\hat{\mathbf{x}}(k+1|k+1)$ and $\hat{\mathbf{x}}(k+1|k)$ given the finite data $\{\mathbf{y}(j),\ 0 \leq j \leq k+1\}$ and $\{\mathbf{y}(j),\ 0 \leq j \leq k\}$, respectively.

(e) Do the limiting values of $P_{k+1|k}$ and $P_{k+1|k+1}$ exist as $k \to \infty$? If so, find them and compute the transfer function from $\mathbf{y}(k)$ to $\hat{\mathbf{x}}(k+1|k)$ and $\hat{\mathbf{x}}(k|k)$. Compare with the results of parts (a) and (b).

9.21 **(Comparing EKF with the optimal filter)** All variables in this problem are real and scalar-valued. Consider the state-space model

$$\mathbf{x}(i+1) = \mathbf{x}^2(i), \quad \mathbf{y}(i) = \mathbf{x}(i) + \mathbf{v}(i), \quad i \geq 0,$$

where $\mathbf{x}(0)$ is uniformly distributed over $[-1, 1]$ while $\mathbf{v}(i)$ is uniformly distributed over $[-0.5, 0.5]$. Moreover, $\{\mathbf{v}(i)\}$ is a white-noise sequence that is uncorrelated with $\mathbf{x}(0)$. We wish to estimate $\mathbf{x}(1)$ given $\mathbf{y}(0)$.

(a) Verify that the EKF equations lead to the estimator $\hat{\mathbf{x}}_{\text{ekf}}(1|0) = \frac{16}{25}\mathbf{y}^2(0)$.

(b) Show that the optimal estimator $\hat{\mathbf{x}}_{\text{opt}}(1|0) \triangleq E[\mathbf{x}(1)|\mathbf{y}_0]$ (*cf.* App. 3.A) is given by

$$\hat{\mathbf{x}}_{\text{opt}}(1|0) = \begin{cases} \frac{(\mathbf{y}+0.5)^2-(\mathbf{y}+0.5)+1}{3} & \text{if } -1.5 \leq \mathbf{y} \leq -0.5, \\ \mathbf{y}^2 + \frac{1}{12} & \text{if } -0.5 \leq \mathbf{y} \leq 0.5, \\ \frac{(\mathbf{y}-0.5)^2+(\mathbf{y}-0.5)+1}{3} & \text{if } 0.5 \leq \mathbf{y} \leq 1.5. \end{cases}$$

[*Hint.* Let $\mathbf{w} \triangleq \mathbf{x}(0)$ and $\mathbf{s} \triangleq \mathbf{y}(0)$. You need to determine the p.d.f. $f_{\mathbf{w}|\mathbf{s}}(w|s)$ and then evaluate $\int w^2 f_{\mathbf{w}|\mathbf{s}}(w|s)dw$ over appropriate intervals.]

(c) Sketch and compare the EKF and optimal estimators of $\mathbf{x}(1)$.

Remark. More details on this example can be found in Söderström (1994). ♦

9.22 **(An EKF-based frequency tracker)** There is a long history of the application of EKF techniques in frequency estimation and tracking (see, *e.g.*, Snyder (1969)). In this problem we consider the following nonlinear state-space model, studied in La Scala and Bitmead (1996),

$$\begin{bmatrix} \mathbf{x}_1(n+1) \\ \mathbf{x}_2(n+1) \\ \mathbf{x}_3(n+1) \end{bmatrix} = \begin{bmatrix} \cos \mathbf{x}_3(n) & -\sin \mathbf{x}_3(n) & 0 \\ \sin \mathbf{x}_3(n) & \cos \mathbf{x}_3(n) & 0 \\ 0 & 0 & 1-\alpha \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(n) \\ \mathbf{x}_2(n) \\ \mathbf{x}_3(n) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \mathbf{w}(n) \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{y}_1(n) \\ \mathbf{y}_2(n) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(n) \\ \mathbf{x}_2(n) \\ \mathbf{x}_3(n) \end{bmatrix} + \mathbf{v}(n),$$

where the parameter $\alpha \in (0, 1)$ determines the rate of time variation of $\mathbf{x}_3$ and is chosen so that the frequency varies slowly enough that the signal appears periodic over several cycles. Moreover, the sequences $\{\mathbf{v}(\cdot), \mathbf{w}(\cdot)\}$ are zero-mean uncorrelated random noise processes with variances R and σ_w^2, respectively. We wish to recover the frequency of the signal (*viz.*, $\mathbf{x}_3$). The $\{\mathbf{x}_1, \mathbf{x}_2\}$ are the in-phase and quadrature signals, respectively, and they are noiseless transformations of the unknown time-variant frequency $\mathbf{x}_3$. Let Π_0 denote the variance of the initial state vector $\text{col}\{\mathbf{x}_1(0), \mathbf{x}_2(0), \mathbf{x}_3(0)\}$. Write down the EKF equations for the above model and simulate the filter for different values of $\{R, \Pi_0, \sigma_w^2\}$.

9.23 **(Modeling errors)** Consider the standard state-space model (9.2.30)–(9.2.31) with $S_i = 0$. In this problem we study the effect of inaccuracies in the matrices $\{F_i, G_i, H_i, \Pi_0, Q_i, R_i\}$ on the performance of the Kalman filter. Thus assume that the *actual* system is described by matrices $\{\bar{F}_i, \bar{G}_i, \bar{H}_i, \bar{\Pi}_0, \bar{Q}_i, \bar{R}_i\}$ so that

$$\mathbf{x}_{i+1}^{ac} = \bar{F}_i \mathbf{x}_i^{ac} + \bar{G}_i \mathbf{u}_i^{ac}, \quad \mathbf{y}_i^{ac} = \bar{H}_i \mathbf{x}_i^{ac} + \mathbf{v}_i^{ac}, \quad i \geq 0,$$

where $\{\mathbf{x}_i^{ac}, \mathbf{y}_i^{ac}, \mathbf{u}_i^{ac}, \mathbf{v}_i^{ac}, \mathbf{x}_0^{ac}\}$ denote the state variable, the output measurement, the process noise, the output noise, and the unknown initial state, respectively, of the actual system. We shall further assume that $\{\mathbf{u}_i^{ac}, \mathbf{v}_i^{ac}, \mathbf{x}_0^{ac}\}$ are zero-mean uncorrelated random variables such that

$$\langle \mathbf{u}_i^{ac}, \mathbf{u}_j^{ac} \rangle = \bar{Q}_i \delta_{ij}, \quad \langle \mathbf{v}_i^{ac}, \mathbf{v}_j^{ac} \rangle = \bar{R}_i \delta_{ij}, \quad \langle \mathbf{x}_0^{ac}, \mathbf{x}_0^{ac} \rangle = \bar{\Pi}_0.$$

Let us further write $\bar{F}_i = F_i + \Delta F_i$, $\bar{H}_i = H_i + \Delta H_i$, and $\bar{G}_i = G_i + \Delta G_i$, for some $\{\Delta F_i, \Delta H_i, \Delta G_i\}$. In practice, these discrepancies between the model and the actual system affect the Kalman filter implementation in the following way. The measurements that we end up using in (9.2.33) are the $\{\mathbf{y}_i^{ac}\}$, so that the filter that is actually implemented is given by

$$\hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_i + K_{p,i} \mathbf{e}_i^{ac}, \qquad \hat{\mathbf{x}}_0 = 0,$$

where $\mathbf{e}_i^{ac} = \mathbf{y}_i^{ac} - H_i \hat{\mathbf{x}}_i$, and where $\{K_{p,i}, R_{e,i}, P_i\}$ are as in Thm. 9.2.1. Introduce the state-error vector $\tilde{\mathbf{x}}_i^{ac} = \mathbf{x}_i^{ac} - \hat{\mathbf{x}}_i$, and define $P_i^{ac} = \|\tilde{\mathbf{x}}_i^{ac}\|^2$, $\Pi_i^{ac} = \|\mathbf{x}_i^{ac}\|^2$, and $P_i^c =< \tilde{\mathbf{x}}_i^{ac}, \mathbf{x}_i^{ac} >$.

(a) Show that P_i^{ac} satisfies the recursion

$$P_{i+1}^{ac} = F_{p,i} P_i^{ac} F_{p,i}^* + \bar{G}_i \bar{Q}_i \bar{G}_i^* + K_{p,i} \bar{R}_i K_{p,i}^* + \Delta F_{p,i} \Pi_i^{ac} \Delta F_{p,i}^* + F_{p,i} P_i^c \Delta F_{p,i}^* + \Delta F_{p,i} P_i^{c*} F_{p,i}^*, \qquad P_0^{ac} = \bar{\Pi}_0,$$

where $\Delta F_{p,i} \triangleq \Delta F_i - K_{p,i} \Delta H_i$. Show also that

$$\Pi_{i+1}^{ac} = \bar{F}_i \Pi_i^{ac} \bar{F}_i^* + \bar{G}_i \bar{Q}_i \bar{G}_i^*, \qquad \Pi_0^{ac} = \bar{\Pi}_0,$$

$$P_{i+1}^c = F_{p,i} P_i^c \bar{F}_i^* + \Delta F_{p,i} \Pi_i^{ac} \bar{F}_i^* + \bar{G}_i \bar{Q}_i \bar{G}_i^*, \qquad P_0^c = \bar{\Pi}_0.$$

Remark. Assume a time-invariant realization. Then Π_i^{ac} will satisfy

$$\Pi_{i+1}^{ac} = \bar{F} \Pi_i^{ac} \bar{F}^* + \bar{G} \bar{Q} \bar{G}^*, \quad \Pi_0^{ac} = \bar{\Pi}_0,$$

which shows that Π_i^{ac}, and consequently P_i^{ac}, can become unbounded when $\bar{F}$ is unstable. This discussion suggests that at least for time-invariant systems, if the actual realization is unstable, then model innacuracies can result in an unbounded growth of the state estimation error. ♦

(b) Assume that the only discrepancy is in the value of Q_i, while all other model parameters are assumed exact. Show that in this case

$$P_{i+1} - P_{i+1}^{ac} = F_{p,i}(P_i - P_i^{ac}) F_{p,i}^* + G_i(Q_i - \bar{Q}_i) G_i^*, \qquad P_0 - P_0^{ac} = 0.$$

Show further that if $Q_i \geq \bar{Q}_i$, then $P_{i+1}^{ac} \leq P_{i+1}$.

Remark. This result shows that if the designer does not know the actual value of a noise covariance, say $\bar{Q}_i$, but has an upper bound for it, say Q_i, then by designing a Kalman filter that is based on Q_i, the designer guarantees that the actual error covariance matrix will be smaller than what is expected by the Kalman filter. A similar conclusion holds for R_i. ♦

9.24 (Effect of perturbations) Consider the model

$$\mathbf{x}_{i+1} = f\mathbf{x}_i, \quad \mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i,$$

with $f = 0.95$, $< \mathbf{x}_0, \mathbf{x}_0 >= 1$, and $< \mathbf{v}_i, \mathbf{v}_j >= 1/f^2\delta_{ij}$.

(a) Verify that the Riccati recursion of the Kalman filter that is associated with this state-space model can be evaluated by either recursion (in covariance and information forms):

$$p_{i+1} = \frac{f^2 p_i}{1 + f^2 p_i} \quad \text{or} \quad p_{i+1}^{-1} = \frac{1}{f^2} p_i^{-1} + 1,$$

with initial condition $p_0 = 1$.

(b) Assume that at a time instant i_o a perturbation is introduced into p_{i_o}, say due to numerical errors. No further errors are introduced at other time instants. Let $\bar{p}_i$ denote the Riccati variable that results for $i \geq i_o$ by applying

$$\bar{p}_{i+1} = \frac{f^2 \bar{p}_i}{1 + f^2 \bar{p}_i}, \quad i \geq i_o.$$

Let also $\hat{p}_i^{-1}$ denote the inverse Riccati variable that results for $i \geq i_o$ by applying

$$\hat{p}_{i+1}^{-1} = \frac{1}{f^2}\hat{p}_i^{-1} + 1, \quad i \geq i_o.$$

Derive the relations

$$\hat{p}_{i+1}^{-1} - p_{i+1}^{-1} = \frac{1}{f^2}[\hat{p}_i^{-1} - p_i],$$

and

$$\bar{p}_{i+1} - p_{i+1} = \frac{f^2}{1 + f^2 p_i + f^2 \bar{p}_i + f^4 \bar{p}_i p_i} (\bar{p}_i - p_i).$$

(c) Show that $\bar{p}_{i+1} - p_{i+1} \leq f^2(\bar{p}_i - p_i)$. Conclude that the recursion for $\bar{p}_i$ recovers from the perturbation and $\bar{p}_i$ eventually converges to p_i. Conclude also that the recursion for $\hat{p}_i$ does not recover from the perturbation and that the error $(\hat{p}_i - p_i)$ will grow unbounded.

Remark. This problem shows how different (but mathematically equivalent) filter implementations can behave quite differently under perturbations of the model. ♦

9.25 (A doubling algorithm) In this problem we assume all inverses exist whenever needed. Consider the Riccati recursion

$$P_{i+1} = FPF_i^* + GQG^* - K_{p,i} R_{e,i} K_{p,i}^*,$$

with zero initial condition, $P_0 = 0$, and where $K_{p,i} = FP_iH^*R_{e,i}^{-1}$ and $R_{e,i} = R + HP_iH^*$. We assume F is invertible and seek an algorithm that computes $P_1, P_2, P_4, P_8, \ldots, P_{2^k}, \ldots$ for $k \geq 0$.

(a) Show that we can express P_{i+1} in the form $P_{i+1} = (C + P_iD)(A + BP_i)^{-1}$, and determine $\{A, B, C, D\}$. Compare $\{A, B, C, D\}$ with the entries of the symplectic matrix M in (E.7.2) of App. E, and verify that

$$M^* = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \underbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_{\triangleq N} \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}.$$

Show that N is a simplectic matrix.

(b) Consider the homogeneous linear equation

$$\begin{bmatrix} X_{i+1} \\ Y_{i+1} \end{bmatrix} = N \begin{bmatrix} X_i \\ Y_i \end{bmatrix}, \quad \{X_i, Y_i\} \quad \text{are square matrices,}$$

and observe that if $Y_iX_i^{-1} = P_i$, then $Y_{i+1}X_{i+1}^{-1} = P_{i+1}$. Therefore, evaluating P_{2^i} requires that we evaluate N^{2^i}. Verify that the entries of N can be expressed in the form

$$N = \begin{bmatrix} \alpha_1 & \alpha_1^{-1}\beta \\ \gamma_1\alpha_1^{-1} & \alpha_1^* + \gamma_1\alpha_1^{-1}\beta_1 \end{bmatrix},$$

where $\alpha_1 = F^*$, $\beta_1 = H^*R^{-1}H$, and $\gamma_1 = GQG^*$. Use the fact that N is symplectic to conclude that N^{2^i} has a similar form, *viz.*,

$$N^{2^i} = \begin{bmatrix} \alpha_i & \alpha_i^{-1}\beta \\ \gamma_i\alpha_i^{-1} & \alpha_i^* + \gamma_i\alpha_i^{-1}\beta_i \end{bmatrix},$$

for some $\{\alpha_i, \beta_i, \gamma_i\}$.

(c) Use the relation $N^{2^{i+1}} = N^{2^i}N^{2^i}$ to establish the recursions

$$\alpha_{i+1} = \alpha_i(I + \beta_i\gamma_i)^{-1}\alpha_i,$$
$$\beta_{i+1} = \beta_i + \alpha_i(I + \beta_i\gamma_i)^{-1}\beta_i\alpha_i^*,$$
$$\gamma_{i+1} = \gamma_i + \alpha_i^*\gamma_i(I + \beta_i\gamma_i)^{-1}\alpha_i.$$

(d) Since $X_0 = I$ and $Y_0 = 0$, conclude that $P_{2^i} = \gamma_i$.

Remark. In Sec. 17.6.6 we shall rederive this doubling procedure by using a scattering formulation of state-space estimation theory, which will provide a natural physical explanation for these formulas. [Compare with recursions (17.6.37)–(17.6.39) and make the identifications $\alpha_i = \Phi_{2^i}^{o*}$, $\gamma_i = \mathcal{P}_{2^i}^o$, $\beta_i = \mathcal{O}_{2^i}^o$ (and also note that $\mathcal{O}_{2^i}^o$ is Hermitian).] The derivation given in this problem follows the one suggested in Anderson and Moore (1979). ♦

9.26 (***UDU**** **factorization of the Gramian matrix**) Consider the backwards recursions of Thm. 9.8.1 and introduce the column vectors $\mathbf{e}^b = \text{col}\{\mathbf{e}_0^b, \ldots, \mathbf{e}_N^b\}$, $\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$.

(a) Show that $\mathbf{e}^b$ and $\mathbf{y}$ are related via an upper triangular matrix with unit diagonal, say $\mathbf{y} = U\mathbf{e}^b$.

(b) The backwards Kalman filter induces the following backwards state-space model from $\mathbf{e}_i^b$ to $\mathbf{y}_i$:

$$\hat{\mathbf{x}}_i^b = F_{i+1}^b \hat{\mathbf{x}}_{i+1}^b + K_{l,i+1}^b \mathbf{e}_{i+1}^b, \quad \hat{\mathbf{x}}_{N+1}^b = 0,$$

$$\mathbf{y}_i = H_i \hat{\mathbf{x}}_i^b + \mathbf{e}_i^b.$$

Conclude that the backwards Kalman filter leads to an upper-diagonal-lower factorization of the Gramian matrix $R_y = \langle \mathbf{y}, \mathbf{y} \rangle$, *viz.*, $R_y = U R_e^b U^*$, with $R_e^b = \text{diag}\{R_{e,0}^b, R_{e,1}^b, \ldots, R_{e,N}^b\}$ and where the nonzero entries of the i-th (block) row of U are given by

$$\begin{bmatrix} I & H_i K_{l,i+1}^b & H_i F_{i+1}^b K_{l,i+2}^b & \cdots & H_i F_{i+1}^b \cdots F_{N-1}^b K_{l,N}^b \end{bmatrix}.$$

(c) Note that the nonzero entries of the i-th block column of U^* are

$$\text{col}\{I, K_{l,i+1}^{b*} H_i^*, K_{l,i+2}^{b*} F_{i+1}^{b*} H_i^*, \ldots, K_{l,N}^{b*} F_{N-1}^{b*} \cdots F_{i+1}^{b*} H_i^*\}.$$

By comparing with the discussion in Sec. 9.4, conclude that the entries of U^* can be generated from the (forwards) state-space model:

$$\mathbf{a}_{i+1} = F_i^{b*} \mathbf{a}_i + H_i^* \mathbf{b}_i, \quad \mathbf{a}_0 = 0,$$

$$\mathbf{s}_i = K_{l,i}^{b*} \mathbf{a}_i + \mathbf{b}_i.$$

That is, the map from $\text{col}\{\mathbf{b}_0, \ldots, \mathbf{b}_N\}$ to $\text{col}\{\mathbf{s}_0, \ldots, \mathbf{s}_N\}$ is U^*.

Remark. This result shows that the triangular factor U, which maps $\mathbf{e}^b$ to $\mathbf{y}$, admits a backwards state-space representation of the form shown in part (b), while its conjugate transpose admits a forwards state-space model of the form shown in part (c). ◆

9.27 **(Oblique Kalman filters)** Refer to Prob. 3.25 where we defined the notion of oblique projection (estimation). Now consider two state-space models

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i \\ \mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i \end{cases} \quad \text{and} \quad \begin{cases} \mathbf{z}_{i+1} = A_i \mathbf{z}_i + B_i \mathbf{r}_i \\ \mathbf{d}_i = C_i \mathbf{z}_i + \mathbf{w}_i \end{cases}$$

where all variables are zero-mean and satisfy

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{r}_j \\ \mathbf{w}_j \\ \mathbf{z}_0 \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & 0 & 0 & 0 \\ 0 & R_i \delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}.$$

The oblique innovations of $\mathbf{y}_i$ and $\mathbf{d}_i$ are defined by $\mathbf{e}_i \triangleq \mathbf{y}_i - \hat{\mathbf{y}}_i$ and $\boldsymbol{\nu}_i \triangleq \mathbf{d}_i - \hat{\mathbf{d}}_i$, where $\hat{\mathbf{y}}_i$ denotes the oblique estimator of $\mathbf{y}_i$ given $\{\mathbf{y}_j, j < i\}$ (*i.e.*, the estimation error is orthogonal to the $\{\mathbf{d}_j, j < i\}$). Likewise, $\hat{\mathbf{d}}_i$ denotes the oblique estimator of $\mathbf{d}_i$ given $\{\mathbf{d}_j, j < i\}$ (the estimation error is now orthogonal to the $\{\mathbf{y}_j, j < i\}$).

(a) Argue that the oblique innovations satisfy $\langle \mathbf{e}_i, \boldsymbol{\nu}_i \rangle = R_{ev,i}\, \delta_{ij}$, for some matrix $R_{ev,i}$. That is, show that they are uncorrelated. This procedure therefore uncorrelates the outputs of two state-space models.

(b) Let $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{z}}_i$ denote the oblique estimators of $\mathbf{x}_i$ and $\mathbf{z}_i$ given $\{\mathbf{y}_j, j < i\}$ and $\{\mathbf{d}_j, j < i\}$, respectively. Follow the derivation of the Kalman filter that is given in the text to show that the oblique innovations can be recursively evaluated as follows:

$$\hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_i + K_{v,i} R_{ev,i}^{-1} \mathbf{e}_i, \quad \hat{\mathbf{z}}_{i+1} = A_i \hat{\mathbf{z}}_i + K_{e,i} R_{ev,i}^{-*} \boldsymbol{\nu}_i,$$

$$\mathbf{e}_i = \mathbf{y}_i - H_i \hat{\mathbf{x}}_i, \quad \boldsymbol{\nu}_i = \mathbf{d}_i - C_i \hat{\mathbf{z}}_i,$$

$$K_{v,i} = F_i P_i C_i^*, \quad K_{e,i} = A_i P_i^* H_i^*, \quad R_{ev,i} = R_i + H_i P_i C_i^*,$$

$$P_{i+1} = F_i P_i A_i^* + G_i Q_i B_i^* + K_{v,i} R_{ev,i}^{-1} K_{e,i}^*, \quad P_0 = \Pi_0.$$

Remark. More details can be found in Sayed and Kailath (1995). ◆

Appendices for Chapter 9

The first two appendices present two different methods for canonical covariance factorization by exploiting the state-space structure of the covariance matrix R_y. As expected, Riccati equations are encountered in the process. Moreover, as in Ch. 8, we can then use the state-space structure of the resulting canonical factor to obtain the Kalman filter recursions for the innovations. However, while this (Wiener) approach is theoretically satisfying, it is less direct than proceeding directly from the state-space model, which is what we did in the main chapter (*cf.* the discussion relating to Fig. 8.1 in Sec. 8.6.)

Another aspect of the results in these appendices is that they provide fast algorithms for the factorization of what are called semi-separable matrices. Our expressions for L, D, and L^{-1} also enable determination of R_y^{-1} and also of the so-called orthogonal (or QR) decomposition of R_y. It is true that these results only apply to Hermitian nonnegative-definite semi-separable matrices. These are the ones most often encountered, but results for non-Hermitian matrices can also be obtained (*e.g.*, by pursuing the results of Prob. 9.27).

9.A FACTORIZATION OF R_y USING THE MGS PROCEDURE

In Sec. 4.2.3, we showed how the modified Gram-Schmidt (MGS) procedure led to a method of sequentially obtaining the columns of the triangular matrix L in the canonical factorization of the Gramian of a stochastic process, $\langle \mathbf{y}, \mathbf{y} \rangle = R_y = LDL^*$. Then $\mathbf{y} = L\mathbf{e}$, where $\mathbf{e} = L^{-1}\mathbf{y}$ is the associated innovations process. For a general $N \times N$ matrix, this takes $O(N^3)$ flops. In the main text, we used the Kalman filter recursions to obtain a formula for L in terms of the Riccati variable P_i that requires only $O(Nn^3)$ flops. In this appendix we shall show how the fact that the $\{\mathbf{y}_i\}$ have a state-space model can be introduced into the MGS procedure itself to obtain the same result by purely matrix arguments. The reader should re-read the description of this procedure in Sec. 4.2.3 to recall the notations used below, *e.g.*, $\tilde{\mathbf{y}}_{i|0} = \mathbf{y}_i - \hat{\mathbf{y}}_{i|\mathbf{y}_0}$.

Recall that the first step of the modified Gram-Schmidt (MGS) procedure provides the variables $\{\mathbf{e}_0, \tilde{\mathbf{y}}_{1|0}, \ldots, \tilde{\mathbf{y}}_{N|0}\}$. If we denote by $R_{y,1}$ the Gramian matrix of the variables $\{\tilde{\mathbf{y}}_{1|0}, \ldots, \tilde{\mathbf{y}}_{N|0}\}$ obtained in this first step, then it is easy to verify that R_y and $R_{y,1}$ are related via the Schur complementation step (*cf.* Prob. 4.6):

$$\begin{bmatrix} 0 & 0 \\ 0 & R_{y,1} \end{bmatrix} = R_y - \begin{bmatrix} < \mathbf{y}_0, \mathbf{y}_0 > \\ < \mathbf{y}_1, \mathbf{y}_0 > \\ \vdots \\ < \mathbf{y}_N, \mathbf{y}_0 > \end{bmatrix} < \mathbf{y}_0, \mathbf{y}_0 >^{-1} \begin{bmatrix} < \mathbf{y}_0, \mathbf{y}_0 > \\ < \mathbf{y}_1, \mathbf{y}_0 > \\ \vdots \\ < \mathbf{y}_N, \mathbf{y}_0 > \end{bmatrix}^* . \quad (9.A.1)$$

That is, $R_{y,1}$ is the *Schur complement* of the $(0,0)$ block entry of R_y. [The successive steps of the MGS procedure would correspond to successive Schur complementations of R_y (see App. A).]

The above facts hold regardless of any special structure. But if the $\{\mathbf{y}_i\}$ are known to be generated by a (standard) state-space model, then more can be said and done. More specifically, the Schur complementation steps on the $N \times N$ (block) matrix R_y can be implemented by using $n \times n$ matrices $\{P_i\}$ that obey a Riccati recursion, as we now verify. So consider again the standard state-space model (9.2.30)–(9.2.31) and let $\Pi_i = \langle \mathbf{x}_i, \mathbf{x}_i \rangle$. Note that

$$\Pi_{i+1} = F_i \Pi_i F_i^* + G_i Q_i G_i^*, \quad i \geq 0.$$

Also define the state-transition matrix $\Phi(j,i) = F_{j-1}F_{j-2}\ldots F_i$, for $j > i$ and $\Phi(j,j) = I$, as well as the quantities $R_{e,0} = \langle \mathbf{e}_0, \mathbf{e}_0 \rangle$ and $K_0 = \langle \mathbf{x}_1, \mathbf{e}_0 \rangle$. Using these quantities, we can now completely identify the entries of the first (block) column of R_y in terms of the model parameters as follows.

The state-space model (9.2.30)–(9.2.31) implies that

$$R_{e,0} = H_0 \Pi_0 H_0^* + R_0, \qquad K_0 = F_0 \Pi_0 H_0^* + G_0 S_0.$$

Now let $\bar{l}_0$ denote the first block column of R_y, *viz.*, $\bar{l}_0 = \text{col}\{\langle \mathbf{y}_0, \mathbf{y}_0 \rangle, \ldots, \langle \mathbf{y}_N, \mathbf{y}_0 \rangle\}$. It is immediate to use the state-space equations (9.2.30) to evaluate $\bar{l}_0$ in terms of $\{K_0, R_{e,0}\}$, which leads to

$$\bar{l}_0 = \text{col}\{R_{e,0}, H_1 K_0, H_2 F_1 K_0, \ldots, H_N \Phi(N,1) K_0\}.$$

Given the above expression for $\bar{l}_0$, we can now write the Schur complementation step (9.A.1) as

$$\begin{bmatrix} 0 & 0 \\ 0 & R_{y,1} \end{bmatrix} = R_y - \begin{bmatrix} R_{e,0} \\ H_1 K_0 \\ H_2 F_1 K_0 \\ \vdots \\ H_N \Phi(N,1) K_0 \end{bmatrix} R_{e,0}^{-1} \begin{bmatrix} R_{e,0} \\ H_1 K_0 \\ H_2 F_1 K_0 \\ \vdots \\ H_N \Phi(N,1) K_0 \end{bmatrix}^* .$$

This is simply a rewriting of (9.A.1) by using the explicit entries of $\bar{l}_0$. The fact that the process $\{\mathbf{y}_i\}$ arises from a state-space model manifests itself in at least two respects. First, note that all the entries of $\bar{l}_0$, except for the leading entry, are determined by the same matrix K_0 and by time-variant matrices $\{H_i\}$. Also, products of matrices $\{F_j\}$ appear between the H_i and K_0. Secondly, it turns out that this structure is preserved during successive Schur complementations. Thus if we let $\bar{l}_1$ denote the first (block) column of $R_{y,1}$, then all its entries (except for the first one) will also be determined by a single matrix K_1 and by the time-variant matrices $\{H_i\}$. Similar products of matrices $\{F_j\}$ will also appear between the H_i and K_1.

To see this, note that the leftmost top entry of $R_{y,1}$ is equal to $< \tilde{\mathbf{y}}_{1|0}, \tilde{\mathbf{y}}_{1|0} >$. It then follows from the above Schur complementation step that

$$< \tilde{\mathbf{y}}_{1|0}, \tilde{\mathbf{y}}_{1|0} > = < \mathbf{y}_1, \mathbf{y}_1 > - H_1 K_0 R_{e,0}^{-1} K_0^* H_1^*.$$

But $< \mathbf{y}_1, \mathbf{y}_1 >$ is the $(2,2)$ entry of R_y and it is equal to

$$< \mathbf{y}_1, \mathbf{y}_1 > = H_1 F_0 \Pi_0 F_0^* H_1^* + H_1 G_0 Q_0 G_0^* H_1^* + R_1.$$

Therefore, we have that

$$< \tilde{\mathbf{y}}_{1|0}, \tilde{\mathbf{y}}_{1|0} > \;=\; H_1 \left[F_0 \Pi_0 F_0^* + G_0 Q_0 G_0^* - K_0 R_{e,0}^{-1} K_0^* \right] H_1^* + R_1.$$

This suggests that we *introduce* the (nonnegative definite) quantity between brackets as

$$P_1 \triangleq F_0 \Pi_0 F_0^* + G_0 Q_0 G_0^* - K_0 R_{e,0}^{-1} K_0^*,$$

and rewrite the top-left corner entry of $R_{y,1}$, in the compact form

$$< \tilde{\mathbf{y}}_{1|0}, \tilde{\mathbf{y}}_{1|0} > \;=\; H_1 P_1 H_1^* + R_1 \triangleq R_{e,1}.$$

This is similar to the expression for the top-left corner entry of R_y, $< \mathbf{y}_0, \mathbf{y}_0 > = H_0 \Pi_0 H_0^* + R_0$, except that Π_0 is replaced by P_1 and the time indexes of H and R are updated, so that the first (block) column of $R_{y,1}$ is given by

$$\bar{l}_1 = \text{col}\{R_{e,1}, H_2 K_1, H_3 F_2 K_1, \ldots, H_N \Phi(N,2) K_1\},$$

where we defined $K_1 = F_1 P_1 H_1^* + G_1 S_1$.

This discussion suggests that $R_{y,1}$ can be regarded as the output covariance matrix of a state-space model that starts at time 1 with an initial state-covariance matrix P_1. We can now proceed with the argument as follows. Let $\mathbf{e}_1 = \tilde{\mathbf{y}}_{1|0}$ and note that

$$R_{e,1} \;=\; \langle \mathbf{e}_1, \mathbf{e}_1 \rangle \;=\; H_1 P_1 H_1^* + R_1, \qquad K_1 \;=\; \langle \mathbf{x}_2, \mathbf{e}_1 \rangle \;=\; F_1 P_1 H_1^* + G_1 S_1.$$

Now subtract from $R_{y,1}$ the outer product $\bar{l}_1 R_{e,1}^{-1} \bar{l}_1^*$. This would then lead us to introduce P_2 via the expression

$$P_2 \;=\; F_1 P_1 F_1^* + G_1 Q_1 G_1^* - K_1 R_{e,1}^{-1} K_1^*.$$

For an arbitrary time index i, we obtain $R_{e,i} = R_i + H_i P_i H_i^*$, $K_i = F_i P_i H_i^* + G_i S_i$, and

$$P_{i+1} \;=\; F_i P_i F_i^* - K_i R_{e,i}^{-1} K_i^* + G_i Q_i G_i^*, \qquad P_0 \;=\; \Pi_0. \tag{9.A.2}$$

In other words, the $\{P_i\}$ obey the Riccati recursions of the Kalman filter. Proceeding in this way, we obtain (*cf.* (9.4.1))

$$L \;=\; \begin{bmatrix} I & 0 & 0 & \ldots & 0 \\ H_1 K_{p,0} & I & 0 & \ldots & 0 \\ H_2 \Phi(2,1) K_{p,0} & H_2 K_{p,1} & I & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ H_N \Phi(N,1) K_{p,0} & H_N \Phi(N,2) K_{p,1} & H_N \Phi(N,3) K_{p,2} & \ldots & I \end{bmatrix},$$

with $K_{p,i} = K_i R_{e,i}^{-1}$. The terms in L are obtained via the Riccati recursion, which takes $O(n^3)$ operations per iteration, making for $O(Nn^3)$ operations overall as compared to $O(N^3)$ for factoring R_y without exploiting the state-space structure.

We now consider the canonical (causal and causally invertible) model $\mathbf{e} = L\mathbf{y}$ and use the state-space structure of L to rewrite this in state-space form,

$$\boldsymbol{\theta}_{i+1} \;=\; F_i \boldsymbol{\theta}_i + K_{p,i} \mathbf{e}_i, \qquad \mathbf{y}_i \;=\; H_i \boldsymbol{\theta}_i + \mathbf{e}_i,$$

where $\boldsymbol{\theta}_0 = 0$ and $K_{p,i} = K_i R_{e,i}^{-1}$. These are just the formulas for the innovations model of Thm. 9.2.2, which we derived by working directly with the state-space model (9.2.30)–(9.2.31) for $\{\mathbf{y}_i\}$. As noted before, that approach is more direct, unless we were directly provided with an expression for R_y in state-space form. Note that working with the model (9.2.30)–(9.2.31) also shows that the state variable $\boldsymbol{\theta}_i$ is in fact the predicted state estimator $\hat{\mathbf{x}}_i$. Note also that the above state-space model for $\{\mathbf{y}_i\}$ can easily be inverted to give recursions for $\{\mathbf{e}_i\}$ from which we can, if desired, write down an explicit expression for L^{-1}, which is not obvious directly from the matrix formula for L.

Remark [Displacement Structure]. We mentioned earlier that stationarity, and its generalization — displacement structure — can also be exploited to speed up matrix triangularization problems. It is shown in App. 13.A how this can be achieved by incorporating displacement structure into the Schur reduction procedure. ◆

9.B FACTORIZATION VIA GRAMIAN EQUIVALENCE CLASSES

In this appendix we show how to extend to the time-variant case the factorization technique method used to obtain the results of Ch. 8. As mentioned at the end of that chapter, the extension to the time-variant case is fairly straightforward (mostly adding subscripts in many places), as we shall now demonstrate. A review of Secs. 8.2 and 8.3 can be useful at this point.

We start with the standard time-variant state-space model (9.2.30)–(9.2.31), and collect the output measurements into a column vector, $\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$, and define the associated Gramian matrix $R_y = \langle \mathbf{y}, \mathbf{y} \rangle$.

An Equivalence Class for Input Covariances. We now show how to construct equivalent classes for the input covariance matrices (9.2.31) that will yield the same output covariance matrix R_y.

For this purpose, let us add disturbances $\{\mathbf{x}_0^0, \mathbf{u}_i^0, \mathbf{v}_i^0\}$ (orthogonal to the original $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i\}$) to the state-space model (9.2.30)–(9.2.31), such that

$$\left\langle \begin{bmatrix} \mathbf{x}_0^0 \\ \mathbf{u}_i^0 \\ \mathbf{v}_i^0 \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0^0 \\ \mathbf{u}_j^0 \\ \mathbf{v}_j^0 \end{bmatrix} \right\rangle \triangleq \begin{bmatrix} \Pi_0^0 & 0 \\ 0 & \begin{bmatrix} Q_i^0 & S_i^0 \\ S_i^{0*} & R_i^0 \end{bmatrix} \delta_{ij} \end{bmatrix} \quad \text{and} \quad \left\langle \begin{bmatrix} \mathbf{x}_0^0 \\ \mathbf{u}_i^0 \\ \mathbf{v}_i^0 \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_j \\ \mathbf{v}_j \end{bmatrix} \right\rangle = 0.$$

No definiteness properties are posed on the $\{\Pi_0^0, Q_i^0, S_i^0, R_i^0\}$. They are to be chosen so that the output process $\{\mathbf{y}_i + \mathbf{y}_i^0\}$ (defined below) has the same Gramian as $\{\mathbf{y}_i\}$. [The question is how we can define such variables $\{\mathbf{x}_0^0, \mathbf{u}_i^0, \mathbf{v}_i^0\}$ in our framework. This can be done, but here we proceed formally — the final results can be checked algebraically.]

With the augmented inputs to (9.2.30), we can, in an obvious notation, write

$$\begin{cases} \mathbf{x}_{i+1} + \mathbf{x}^0_{i+1} &= F_i(\mathbf{x}_i + \mathbf{x}^0_i) + G_i\mathbf{u}_i + \mathbf{u}^0_i, \\ \mathbf{y}_i + \mathbf{y}^0_i &= H_i(\mathbf{x}_i + \mathbf{x}^0_i) + \mathbf{v}_i + \mathbf{v}^0_i. \end{cases} \tag{9.B.1}$$

The covariance matrix of the new disturbances $(G_i\mathbf{u}_i + \mathbf{u}^0_i, \mathbf{v}_i + \mathbf{v}^0_i)$ is

$$\begin{bmatrix} G_iQ_iG_i^* + Q_i^0 & G_iS_i + S_i^0 \\ S_i^*G_i^* + S_i^{0*} & R_i + R_i^0 \end{bmatrix},$$

and the output covariance matrix is $R_{y+y^0} = R_y + R_{y^0}$. Therefore, if R_y is to be unchanged, this implies that R_{y^0}, the covariance matrix of the process $\{\mathbf{y}^0_i\}$, defined by

$$\begin{cases} \mathbf{x}^0_{i+1} &= F_i\mathbf{x}^0_i + \mathbf{u}^0_i, \\ \mathbf{y}^0_i &= H_i\mathbf{x}^0_i + \mathbf{v}^0_i, \end{cases} \tag{9.B.2}$$

must be *zero*, a condition that we now enforce by choosing $\{Q_i^0, R_i^0, S_i^0\}$.

From the first of the state equations in (9.B.2) we obtain

$$\langle \mathbf{x}^0_{i+1}, \mathbf{x}^0_{i+1}\rangle = F_i\langle \mathbf{x}^0_i, \mathbf{x}^0_i\rangle F_i^* + \langle \mathbf{u}^0_i, \mathbf{u}^0_i\rangle = F_i\langle \mathbf{x}^0_i, \mathbf{x}^0_i\rangle F_i^* + Q_i^0,$$

so that if we define $Z_i \triangleq -\langle \mathbf{x}^0_i, \mathbf{x}^0_i\rangle$, we can write

$$Q_i^0 = -Z_{i+1} + F_iZ_iF_i^*. \tag{9.B.3}$$

Moreover, we want

$$\langle \mathbf{y}^0_i, \mathbf{y}^0_i\rangle = H_i\langle \mathbf{x}^0_i, \mathbf{x}^0_i\rangle H_i^* + R_i^0 = -H_iZ_iH_i^* + R_i^0 = 0, \tag{9.B.4}$$

and, also for $j > 0$,

$$\langle \mathbf{y}^0_{i+j}, \mathbf{y}^0_i\rangle = H_{i+j}F_{i+j-1}\cdots F_{i+1}(-F_iZ_iH_i^* + S_i^0) = 0.$$

It is clear that the above equalities can be guaranteed by the choices

$$S_i^0 = F_iZ_iH_i^*, \quad R_i^0 = H_iZ_iH_i^*. \tag{9.B.5}$$

Combining (9.B.3), (9.B.4), and (9.B.5) yields

$$\Pi_0^0 = -Z_0 \quad \text{and} \quad \begin{bmatrix} Q_i^0 & S_i^0 \\ S_i^{0*} & R_i^0 \end{bmatrix} = \begin{bmatrix} -Z_{i+1} + F_iZ_iF_i^* & F_iZ_iH_i^* \\ H_iZ_iF_i^* & H_iZ_iH_i^* \end{bmatrix}.$$

This leads to the following lemma, the exact analog of Lemma 8.2.1, except for the (obvious) subscripts.

Lemma 9.B.1 (Equivalence Class for Input Covariances) *Consider the standard state-space model (9.2.30)–(9.2.31). Then $R_y \triangleq \langle \mathbf{y}, \mathbf{y}\rangle$, with $\mathbf{y} \triangleq col\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$, is invariant under the input Gramian transformations*

$$\Pi_0 \to \Pi_0 - Z_0 \tag{9.B.6}$$

and

$$\begin{bmatrix} G_iQ_iG_i^* & G_iS_i \\ S_i^*G_i^* & R_i \end{bmatrix} \to \begin{bmatrix} -Z_{i+1} + F_iZ_iF_i^* + G_iQ_iG_i^* & F_iZ_iH_i^* + G_iS_i \\ H_iZ_iF_i^* + S_i^*G_i^* & H_iZ_iH_i^* + R_i \end{bmatrix},$$

for any sequence of Hermitian matrices $\{Z_i\}$. ■

Proof: An algebraic proof can be obtained as follows. Note that the (block) diagonal entries of the original Gramian matrix $\langle \mathbf{y}_i, \mathbf{y}_i\rangle$ are given by

$$\langle \mathbf{y}_i, \mathbf{y}_i\rangle = H_i\Pi_iH_i^* + R_i, \tag{9.B.7}$$

where $\Pi_i = \langle \mathbf{x}_i, \mathbf{x}_i\rangle$ satisfies the recursion

$$\Pi_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^*, \quad \Pi_0 = \text{initial condition.}$$

Now we perform the transformations in the lemma, *i.e.*, we replace $G_iQ_iG_i^*$ by $G_iQ_iG_i^* - Z_{i+1} + F_iZ_iF_i^*$, R_i by $R_i + H_iZ_iH_i^*$, and Π_0 by $\Pi_0 - Z_0$. Then the Gramian of the "new" output measurement at time i becomes

$$H_iW_iH_i^* + (R_i + H_iZ_iH_i^*), \tag{9.B.8}$$

where W_i is the Gramian matrix of the "new" state variable and it satisfies the recursion (which now replaces the update for Π_i)

$$W_{i+1} = F_iW_iF_i^* + (G_iQ_iG_i^* - Z_{i+1} + F_iZ_iF_i^*), \quad W_0 = \Pi_0 - Z_0. \tag{9.B.9}$$

We want to show that (9.B.7) coincides with (9.B.8). That is, we want to show that the (block) diagonal entries of the output Gramian remain invariant. For this purpose, we define $M_{i+1} = W_{i+1} + Z_{i+1}$ and note from (9.B.9) that M_i satisfies

$$M_{i+1} = F_iM_iF_i^* + G_iQ_iG_i^*, \quad M_0 = \Pi_0.$$

It therefore follows that M_i satisfies the same recursion as Π_i and starts with the same initial condition; hence, $M_i = \Pi_i$. If we now replace $(W_i + Z_i)$ in (9.B.8) by Π_i we see that it collapses to (9.B.7), as desired.

The same argument can be used to verify that the off-diagonal entries, $\langle \mathbf{y}_j, \mathbf{y}_i\rangle$ for $j \neq i$, are also invariant under the transformations given in the statement of the lemma. ◆

Application to Factorization. Lemma 9.B.1 can now be used to provide a block triangular factorization of R_y. To begin with, note that it is not difficult to verify that the output Gramian R_y of $\mathbf{y} = col\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$, can be written as (recall Lemma 5.A.2)

$$R_y = \mathcal{O}(0)\Pi_0\mathcal{O}^*(0) + \sum_{j=0}^{N} \mathcal{Z}^j \begin{bmatrix} 0 & I \\ \mathcal{O}(j+1) & 0 \end{bmatrix} \begin{bmatrix} G_jQ_jG_j^* & G_jS_j \\ S_j^*G_j^* & R_j \end{bmatrix} \begin{bmatrix} 0 & I \\ \mathcal{O}(j+1) & 0 \end{bmatrix}^* \mathcal{Z}^{*j}$$

where $\mathcal{O}(j)$, the observability map at time j, and $\mathcal{Z}$, the block lower triangular shift matrix, are defined as

$$\mathcal{O}(j) \triangleq \begin{bmatrix} H_j \\ H_{j+1}F_j \\ H_{j+2}F_{j+1}F_j \\ \vdots \\ H_N F_{N-1}F_{N-2}\cdots F_j \end{bmatrix} \text{ and } \mathcal{Z} \triangleq \begin{bmatrix} 0 & & & \\ I & 0 & & \\ & I & 0 & \\ & & \ddots & \ddots \\ & & & I \quad 0 \end{bmatrix}.$$

Now Lemma 9.B.1 implies that for any sequence of Hermitian matrices $\{Z_j\}_{j=0}^N$ we can equivalently write

$$R_y = \mathcal{O}(0)(\Pi_0 - Z_0)\mathcal{O}^*(0) + \sum_{j=0}^{N} \mathcal{Z}^j \begin{bmatrix} 0 & I \\ \mathcal{O}(j+1) & 0 \end{bmatrix} T_j \begin{bmatrix} 0 & I \\ \mathcal{O}(j+1) & 0 \end{bmatrix}^* \mathcal{Z}^{*j}, \tag{9.B.10}$$

where

$$T_j = \begin{bmatrix} -Z_{j+1} + F_j Z_j F_j^* + G_j Q_j G_j^* & F_j Z_j H_j^* + G_j S_j \\ H_j Z_j F_j^* + S_j^* G_j^* & H_j Z_j H_j^* + R_j \end{bmatrix}.$$

We would now like to choose the $\{Z_j\}_{j=0}^N$ in such a manner that the block triangular factorization of R_y is obtained. In view of (9.B.10), this would be the case if $\Pi_0 - Z_0 = 0$ and if all the $(n+p)\times(n+p)$ matrices $\{T_j\}$ appearing in the center of (9.B.10) have rank p. In other words, if $\Pi_0 = Z_0$, and the T_j have the form

$$T_j = \begin{bmatrix} M_{2,j} \\ M_{1,j} \end{bmatrix} \begin{bmatrix} M_{2,j}^* & M_{1,j}^* \end{bmatrix},$$

for some $M_{1,j} \in \mathbb{C}^{p\times p}$ and $M_{2,j} \in \mathbb{C}^{n\times p}$.

Assume that $R_j + H_j Z_j H_j^*$ is nonsingular. Then we can perform the following block "upper-lower" triangular factorization of T_j (*cf.* App. A),

$$T_j = \begin{bmatrix} I & X_j \\ 0 & I \end{bmatrix} \begin{bmatrix} \Delta_j & 0 \\ 0 & R_j + H_j Z_j H_j^* \end{bmatrix} \begin{bmatrix} I & 0 \\ X_j^* & I \end{bmatrix}, \tag{9.B.11}$$

where

$$X_j \triangleq (F_j Z_j H_j^* + G_j S_j)(R_j + H_j Z_j H_j^*)^{-1},$$

and Δ_j is the Schur complement

$$\Delta_j = -Z_{j+1} + F_j Z_j F_j^* + G_j Q_j G_j^* - X_j(R_j + H_j Z_j H_j^*)X_j^*.$$

Eq. (9.B.11) shows that if we choose Z_j so as to make Δ_j zero, *i.e.*,

$$-Z_{j+1} + F_j Z_j F_j^* + G_j Q_j G_j^* - X_j(R_j + H_j Z_j H_j^*)X_j^* = 0,$$

we obtain

$$T_j = \begin{bmatrix} F_j Z_j H_j^* + G_j S_j \\ R_j + H_j Z_j H_j^* \end{bmatrix} (R_j + H_j Z_j H_j^*)^{-1} \begin{bmatrix} F_j Z_j H_j^* + G_j S_j \\ R_j + H_j Z_j H_j^* \end{bmatrix}^*,$$

meaning that T_j becomes of minimal rank p. The recursion for Z_j is simply the Kalman filter Riccati recursion for P_j! We can now identify

$$M_{1,j} = R_{e,j}^{1/2}, \qquad M_{2,j} = K_{p,j} R_{e,j}^{1/2},$$

so that we have the following result.

Lemma 9.B.2 (Triangular Factorization of R_y) *If R_y is positive-definite, its block triangular factorization is given by*

$$R_y = \sum_{j=0}^{N} \mathcal{Z}^j \begin{bmatrix} I \\ \mathcal{O}(j+1)K_{p,j} \end{bmatrix} R_{e,j} \begin{bmatrix} I & K_{p,j}^* \mathcal{O}^*(j+1) \end{bmatrix} \mathcal{Z}^{*j}. \tag{9.B.12}$$

■

Note that the above lemma implies that L and R_e in the canonical factorization of R_y are given by

$$L = \sum_{j=0}^{N} \mathcal{Z}^j \begin{bmatrix} I \\ \mathcal{O}(j+1)K_{p,j} \end{bmatrix} \text{ and } R_e = \operatorname{diag}\{R_{e,0}, \ldots, R_{e,N}\}.$$

More explicitly, L can be written as

$$L = \begin{bmatrix} I & 0 & 0 \ldots \\ H_1 K_{p,0} & I & 0 \ldots \\ H_2 F_1 K_{p,0} & H_2 K_{p,1} & I \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

We can now proceed as in Sec. 8.3.5 to obtain state-space models for the modeling filter corresponding to L, the whitening filter corresponding to L^{-1}, and so on. However, for brevity we shall not do so here. It should by now be clear to the reader how all the discussions of Ch. 8 can be carried over to the time-variant case.

CHAPTER 10

Smoothed Estimators

As we have seen in some detail, the Kalman filter and its variants give us recursive algorithms for computing the predicted and filtered state estimators, $\hat{\mathbf{x}}_{i|i-1}$ and $\hat{\mathbf{x}}_{i|i}$. It is not hard to compute *higher-order* predicted estimators $\hat{\mathbf{x}}_{i+m|i}$, $m > 0$. In fact

$$\hat{\mathbf{x}}_{i+m|i} = F_{i+m-1} \cdots F_i \hat{\mathbf{x}}_{i|i}, \quad m > 0.$$

However the determination of smoothed estimators, say $\hat{\mathbf{x}}_{i|N}$ for $i < N$, was not as evident by the methods used in the original papers of Kalman (1960a) and Kalman and Bucy (1961). The first solutions were obtained in quite different ways by Rauch (1962) and by Bryson and Frazier (1963). The latter used equivalence to actually solve a (continuous-time) deterministic problem (by arguments similar to those described in Sec. 10.6).

In this chapter, we shall use the innovations approach to first derive a general formula (not restricted to state-space models) showing how smoothed estimators are completely determined by knowledge of the predicted estimators; when specialized to state-space models, this general formula will lead us very directly to the major state-space smoothing algorithms (see Kailath and Frost (1968)).

There are slightly different formulations of the smoothing problem, but the most important and most widely used is the so-called *fixed-interval* smoothing problem of determining

$$\hat{\mathbf{x}}_{i|N} = \text{ the linear least-mean-squares estimator of } \mathbf{x}_i \text{ given } \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}.$$

If we keep the point i fixed, say at $i = 0$, and let N increase, we have a *fixed-point* smoothing problem (already studied in Prob. 9.18). If i varies, but so does N, according to the formula $N = i + L$, $L > 0$ fixed, then we have a *fixed-lag* smoothing problem. We shall discuss the fixed-interval problem, and then indicate in the problems how the results for fixed-point and fixed-lag estimators follow fairly easily (see Probs. 10.1 and 10.2).

10.1 GENERAL SMOOTHING FORMULAS

We begin with a general linear model

$$\mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i, \quad 0 \le i \le N, \tag{10.1.1}$$

with

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = R_i \delta_{ij}, \quad \langle \mathbf{v}_i, \mathbf{x}_j \rangle = 0, \quad \text{for } j \le i. \tag{10.1.2}$$

For the moment no further assumptions are made on the process $\{\mathbf{x}_i\}$, *e.g.*, that it has a state-space model.

Once again, the basic principle we shall use is the one laid out in Ch. 4: to find $\hat{\mathbf{x}}_{i|N}$, we should first replace $\{\mathbf{y}_j, 0 \le j \le N\}$ by the innovations and then estimate the desired quantities. For $\{\mathbf{y}_i\}$ as in (10.1.1)–(10.1.2), we recall that

$$\mathbf{e}_j = \mathbf{y}_j - \hat{\mathbf{y}}_j = \mathbf{y}_j - H_j \hat{\mathbf{x}}_j = H_j \tilde{\mathbf{x}}_j + \mathbf{v}_j, \quad 0 \le j \le N, \tag{10.1.3}$$

where $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$ and

$$R_{e,j} \triangleq \|\mathbf{e}_j\|^2 = H_j \langle \tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_j \rangle H_j^* + \langle \mathbf{v}_j, \mathbf{v}_j \rangle = H_j P_j H_j^* + R_j.$$

Let us introduce the covariance matrix of the $\{\tilde{\mathbf{x}}_i\}$,

$$P_{i,j} \triangleq \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle \quad \text{with} \quad P_{i,i} \triangleq P_i = \|\tilde{\mathbf{x}}_i\|^2.$$

Now, by the orthogonality of the innovations, we can write

$$\hat{\mathbf{x}}_{i|N} = \sum_{j=0}^{N} \langle \mathbf{x}_i, \mathbf{e}_j \rangle R_{e,j}^{-1} \mathbf{e}_j. \tag{10.1.4}$$

Note that setting $N = i - 1$ in (10.1.4) gives the predicted estimator $\hat{\mathbf{x}}_i$. This suggests that we break the sum into two parts to obtain

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i + \sum_{j=i}^{N} \langle \mathbf{x}_i, \mathbf{e}_j \rangle R_{e,j}^{-1} \mathbf{e}_j. \tag{10.1.5}$$

Moreover, it turns out that for $j \ge i$ we can be more explicit about the inner products $\langle \mathbf{x}_i, \mathbf{e}_j \rangle$:

$$\begin{aligned} \langle \mathbf{x}_i, \mathbf{e}_j \rangle &= \langle \mathbf{x}_i, H_j \tilde{\mathbf{x}}_j \rangle + \langle \mathbf{x}_i, \mathbf{v}_j \rangle, \\ &= \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle H_j^* + \langle \hat{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle H_j^* + \langle \mathbf{x}_i, \mathbf{v}_j \rangle. \end{aligned}$$

But for $j \ge i$, $\mathbf{v}_j \perp \mathbf{x}_i$ by assumption, and $\tilde{\mathbf{x}}_j \perp \mathcal{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_j\}$; the linear subspace spanned by the observations. This space contains $\hat{\mathbf{x}}_i$ and, hence,

$$\langle \mathbf{x}_i, \mathbf{e}_j \rangle = \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle H_j^* = P_{i,j} H_j^*, \quad j \ge i, \tag{10.1.6}$$

so that (10.1.5) reduces to

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i + \sum_{j=i}^{N} P_{i,j} H_j^* R_{e,j}^{-1} \mathbf{e}_j, \quad 0 \le i \le N. \tag{10.1.7}$$

The formulas (10.1.4) and (10.1.7) show that the smoothed estimators $\{\hat{\mathbf{x}}_{i|N}\}$ are nicely determined by knowledge of the prediction estimators $\{\hat{\mathbf{x}}_i\}$![1] Therefore, conceptually, the solution of the smoothing problem is straightforward once the predicted estimators are determined, though, of course, certain additional computations are required. Moreover, all the knowledge we have gained about computing $\hat{\mathbf{x}}_i$ in different problems can be applied to obtaining the smoothed estimators $\hat{\mathbf{x}}_{i|N}$ in those problems as well. We shall illustrate this very soon.

Here, however, we note that in the above, we could have obtained a similar decomposition to (10.1.5) using filtered estimators $\{\hat{\mathbf{x}}_{i|i}\}$ rather than predicted estimators $\{\hat{\mathbf{x}}_i\}$. We can also show how to compute $P_{i|N} = \|\tilde{\mathbf{x}}_{i|N}\|^2$ from knowledge of $P_{i,j} = \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle$. For convenience the results are summarized in a theorem.

Theorem 10.1.1 (General Smoothing Formulas) *Given (10.1.1)–(10.1.2), we can write*

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i + \sum_{j=i}^{N} P_{i,j} H_j^* R_{e,j}^{-1} \mathbf{e}_j, \tag{10.1.8}$$

where $\mathbf{e}_j = \mathbf{y}_j - H_j \hat{\mathbf{x}}_j$, $R_{e,j} = \|\mathbf{e}_j\|^2 = R_j + H_j P_j H_j^*$, $P_{i,j} = \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle$, *and* $P_i = P_{i,i}$. *The error covariance matrix for the smoothed estimators can be expressed as*

$$\langle \tilde{\mathbf{x}}_{i|N}, \tilde{\mathbf{x}}_{i|N} \rangle \triangleq P_{i|N} = P_i - \sum_{j=i}^{N} P_{i,j} H_j^* R_{e,j}^{-1} H_j P_{i,j}. \tag{10.1.9}$$

We could have also used filtered estimators to obtain

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_{i|i} + \sum_{j=i+1}^{N} P_{i,j} H_j^* R_{e,j}^{-1} \mathbf{e}_j, \tag{10.1.10}$$

$$P_{i|N} = P_{i|i} - \sum_{j=i+1}^{N} P_{i,j} H_j^* R_{e,j}^{-1} H_j P_{i,j}. \tag{10.1.11}$$

■

Proof: The formula (10.1.8) was already proved above; (10.1.10) follows analogously. For the error covariance matrices, note that

$$\tilde{\mathbf{x}}_{i|N} = \mathbf{x}_i - \hat{\mathbf{x}}_{i|N} = \tilde{\mathbf{x}}_i - \sum_{j=i}^{N} P_{i,j} H_j^* R_{e,j}^{-1} \mathbf{e}_j,$$

and that, as shown earlier (see (10.1.6)), $\langle \tilde{\mathbf{x}}_i, \mathbf{e}_j \rangle = P_{i,j} H_j^*$. Then forming $\langle \tilde{\mathbf{x}}_{i|N}, \tilde{\mathbf{x}}_{i|N} \rangle$ leads easily to the expression in (10.1.9). Analogously for (10.1.11). ♦

[1] Given the history of the smoothing problem, this fact was regarded as surprising. It is, however, an immediate consequence of the innovations approach to the estimation problem, because the innovations are determined by the predicted estimators.

10.2 EXPLOITING STATE-SPACE STRUCTURE

The natural question now is to explore what simplifications may ensue when the $\{\mathbf{x}_i\}$ process has more structure, and in particular a (standard) state-space model,

$$\begin{aligned} \mathbf{x}_{i+1} &= F_i \mathbf{x}_i + G_i \mathbf{u}_i, \\ \mathbf{y}_i &= H_i \mathbf{x}_i + \mathbf{v}_i, \end{aligned} \tag{10.2.1}$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & S_i \delta_{ij} & 0 \\ S_i^* \delta_{ij} & R_i \delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{10.2.2}$$

We shall see that this allows us to evaluate the $P_{i,j}$ in terms of the Kalman filter variables $\{P_i, K_{p,i}, R_{e,i}, F_i\}$.

Lemma 10.2.1 (State-Space Formula for $P_{i,j}$) *For the standard state-space model (10.2.1)–(10.2.2), it holds that* $P_{i,j} = P_i \Phi_p^*(j,i)$ *for* $j \geq i$, *where*

$$\Phi_p(j,i) = \begin{cases} F_{p,j-1} F_{p,j-2} \cdots F_{p,i} & \text{for} \quad j > i, \\ I & \text{for} \quad j = i, \end{cases} \tag{10.2.3}$$

and $F_{p,i} = F_i - K_{p,i} H_i$.

■

Proof: Recalling the Kalman filter equation $\hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_i + K_{p,i} \mathbf{e}_i$, we readily see (or recall from Eq. (9.2.23)) that

$$\tilde{\mathbf{x}}_{i+1} = F_{p,i} \tilde{\mathbf{x}}_i + G_i \mathbf{u}_i - K_{p,i} \mathbf{v}_i. \tag{10.2.4}$$

Therefore, for $j \geq i$, we shall have

$$\tilde{\mathbf{x}}_j = \Phi_p(j,i) \tilde{\mathbf{x}}_i + \sum_{k=i}^{j-1} \Phi_p(j,k+1) \left[G_k \mathbf{u}_k - K_{p,k} \mathbf{v}_k \right], \quad j \geq i.$$

Now taking inner products with $\tilde{\mathbf{x}}_i$ gives

$$P_{i,j} = \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle = \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i \rangle \Phi_p^*(j,i) + 0 = P_i \Phi_p^*(j,i), \quad j \geq i.$$

♦

10.2.1 The Bryson-Frazier (BF) Formulas

With the standard state-space model (10.2.1)–(10.2.2), we can combine the general results of Thm. 10.1.1 with the formula in Lemma 10.2.1 to write

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i + P_i \boldsymbol{\lambda}_{i|N}, \quad 0 \leq i \leq N, \tag{10.2.5}$$

where we defined

$$\boldsymbol{\lambda}_{i|N} \triangleq \sum_{j=i}^{N} \Phi_p^*(j,i) H_j^* R_{e,j}^{-1} \mathbf{e}_j. \tag{10.2.6}$$

But then the formula (10.2.3) for $\Phi_p(j,i)$ immediately suggests that $\lambda_{i|N}$ can be computed via the backwards-time recursion

$$\lambda_{i|N} = F_{p,i}^* \lambda_{i+1|N} + H_i^* R_{e,i}^{-1} e_i, \quad \lambda_{N+1|N} = 0. \tag{10.2.7}$$

Quantities obeying recursions of the form (10.2.5) and (10.2.7) are encountered in the calculus of variations, in particular, in the minimization of quadratic forms involving state variables — see Sec. 10.6 for further details. In fact, the state-space smoothing problem was first solved, in continuous-time, by formulating it as just such a quadratic minimization problem (Bryson and Frazier (1963)). The innovations approach yields this result quite directly, as we have just seen; it will also give us a stochastic interpretation of the so-called adjoint variables $\{\lambda_{i|N}\}$ (see Eq. (10.2.14) in Sec. 10.2.2). First, however, let us summarize the preceding results, and minor extensions thereof, in a theorem.

Theorem 10.2.1 (The Bryson-Frazier Formulas) *Given the model (10.2.1)–(10.2.2), we can find the smoothed estimators $\hat{x}_{i|N}$ via (10.2.5) where $\lambda_{i|N}$ is found via the backwards recursion (10.2.7). The corresponding error-covariance matrix can be found as*

$$P_{i|N} = P_i - P_i \Lambda_{i|N} P_i, \tag{10.2.8}$$

where

$$\Lambda_{i|N} \triangleq \|\lambda_{i|N}\|^2 = F_{p,i}^* \Lambda_{i+1|N} F_{p,i} + H_i^* R_{e,i}^{-1} H_i, \quad \Lambda_{N+1|N} = 0. \tag{10.2.9}$$

Alternative expressions in terms of filtered quantities are

$$\hat{x}_{i|N} = \hat{x}_{i|i} + P_i F_{p,i}^* \lambda_{i+1|N}, \tag{10.2.10}$$

with

$$P_{i|N} = P_{i|i} - P_i F_{p,i}^* \Lambda_{i+1|N} F_{p,i} P_i. \tag{10.2.11}$$

When $S_i = 0$, we can modify the formula (10.2.11) by replacing $P_i F_{p,i}^*$ by the equivalent quantity $P_{i|i} F_i^*$ (cf. (10.3.7) further ahead). The quantities

$$\{\hat{x}_i, \hat{x}_{i|i}, e_i, R_{e,i}, F_{p,i}, P_i, P_{i|i}\}$$

are as in the Kalman filter formulas of Thms. 9.2.1 and 9.5.1. ∎

Proof: Expression (10.2.5) was proved in the discussion preceding the theorem. Eq. (10.2.8) follows by noting that therefore $\tilde{x}_{i|N} = \tilde{x}_i - P_i \lambda_{i|N}$. We can proceed by a direct calculation, but some experience will show that it is better to rewrite the above as

$$\tilde{x}_i = \tilde{x}_{i|N} + P_i \lambda_{i|N}, \quad \text{since} \quad \lambda_{i|N} \perp \tilde{x}_{i|N}.$$

The reason is that $\tilde{x}_{i|N} \perp \{e_0, e_1, \ldots, e_N\}$ while $\lambda_{i|N} \in \mathcal{L}\{e_i, e_{i+1}, \ldots, e_N\}$. Therefore, $P_i = P_{i|N} + P_i \Lambda_{i|N} P_i^*$, from which the desired result (10.2.8) follows. Alternatively, we can use the evident orthogonality of $\hat{x}_i$ and $\lambda_{i|N}$ in (10.2.5), to write

$$\Sigma_{i|N} \triangleq \langle \hat{x}_{i|N}, \hat{x}_{i|N} \rangle = \|\hat{x}_i\|^2 + P_i \Lambda_{i|N} P_i^*,$$

and now the result follows from noting that $P_{i|N} = \Pi_i - \Sigma_{i|N}$ where $\Pi_i \triangleq \|x_i\|^2$. The proofs of the remaining results are quite straightforward. ◆

The BF formulas give us a "two-pass" algorithm. On a *forwards* pass, we compute the innovations and the predicted and filtered state estimators; then a *backwards* pass uses the innovations to compute the adjoint variables. Finally, an appropriate combination gives the smoothed estimators.

10.2.2 Stochastic Interpretation of the Adjoint Variable

It turns out that, at least when $S_i = 0$, the adjoint variable $\lambda_{i|N}$ is related to a smoothed estimator of the input, $\hat{u}_{i|N}$. (A more complicated relation holds in the general case — see Prob. 10.3.)

To see this note that, again calling on the innovations, we can write

$$\hat{u}_{i|N} = \sum_{j=0}^{N} \langle u_i, e_j \rangle \; R_{e,j}^{-1} e_j.$$

But since $S_i = 0$, we have $u_i \perp e_j$ for $j \le i$, so that

$$\hat{u}_{i|N} = \sum_{j=i+1}^{N} \langle u_i, e_j \rangle \; R_{e,j}^{-1} e_j.$$

Now, for $j > i$, $\langle e_j, u_i \rangle = H_j \langle \tilde{x}_j, u_i \rangle$, while via (10.2.4)

$$\langle \tilde{x}_j, u_i \rangle = \Phi_p(j, i+1) G_i Q_i, \quad j > i, \tag{10.2.12}$$

so that

$$\langle e_j, u_i \rangle = H_j \Phi_p(j, i+1) G_i Q_i = \langle u_i, e_j \rangle^*. \tag{10.2.13}$$

Therefore, finally recalling the definition of $\lambda_{i+1|N}$ in (10.2.6), we see that

$$\hat{u}_{i|N} = \sum_{j=i+1}^{N} Q_i G_i^* \Phi_p^*(j, i+1) H_j^* R_{e,j}^{-1} e_j = Q_i G_i^* \lambda_{i+1|N}, \tag{10.2.14}$$

a useful formula. For example, from the state equation we can write

$$\hat{x}_{i+1|N} = F_i \hat{x}_{i|N} + G_i \hat{u}_{i|N} = F_i \hat{x}_{i|N} + G_i Q_i G_i^* \lambda_{i+1|N}, \tag{10.2.15}$$

which as we shall see in the next section is essentially a so-called RTS smoothing formula.

10.3 THE RAUCH-TUNG-STRIEBEL (RTS) RECURSIONS

There are of course several other ways of exploiting state-space structure. The fact that $\{x_i, \hat{x}_i, \lambda_{i|N}\}$ all obey recursive equations suggests that so should the $\{\hat{x}_{i|N}\}$. Pursuing this remark leads to (different forms of) what have been called the RTS recursions.

10.3.1 First Form of RTS Recursions

The result (10.2.15) just presented leads directly to a recursion for the smoothed estimator if we add the three assumptions that

(a) F_i is invertible.

(b) $S_i = 0$.

(c) $P_i > 0$ for $i \geq 0$. [Recall from Lemma 9.5.1 that a sufficient condition for $P_i > 0$, when $S_i = 0$, is to assume $R_i > 0$, $\Pi_0 > 0$, and F_i invertible.]

Under these assumptions, we can combine (10.2.15) with the BF formula (10.2.5) to obtain

$$\hat{\mathbf{x}}_{i+1|N} = F_i\hat{\mathbf{x}}_{i|N} + G_iQ_iG_i^*P_{i+1}^{-1}(\hat{\mathbf{x}}_{i+1|N} - \hat{\mathbf{x}}_{i+1}),$$

or, equivalently,

$$\hat{\mathbf{x}}_{i|N} = F_{s,i}\hat{\mathbf{x}}_{i+1|N} + F_i^{-1}G_iQ_iG_i^*P_{i+1}^{-1}\hat{\mathbf{x}}_{i+1}, \quad i \leq N, \tag{10.3.1}$$

where we have defined

$$F_{s,i} \triangleq F_i^{-1}(I - G_iQ_iG_i^*P_{i+1}^{-1}), \tag{10.3.2}$$

$$= \text{the closed-loop state matrix of the smoothed estimator.}$$

Useful alternative expressions for $F_{s,i}$, that do not depend on F_i^{-1}, are given below in Thm. 10.3.1. Here we note that (10.3.1) is also a two-pass algorithm. On a forward pass we compute the predicted estimators $\{\hat{\mathbf{x}}_i\}$ and the error covariance matrices $\{P_i\}$, and on a backward pass, starting with the now available boundary condition $\hat{\mathbf{x}}_{N+1|N}$, we obtain the smoothed estimators $\{\hat{\mathbf{x}}_{N|N}, \hat{\mathbf{x}}_{N-1|N}, \ldots, \hat{\mathbf{x}}_{0|N}\}$. This is somewhat more direct than with the BF formulas, but on the other hand we need to have available the inverse quantities $\{F_i^{-1}, P_{i+1}^{-1}\}$. We should note that the inverse P_{i+1}^{-1} can be propagated recursively as explained in Sec. 9.5.6.

The formula (10.3.1) is the discrete-time equivalent of a continuous-time smoothing formula first given by Rauch, Tung, and Striebel (1965). Actually, these authors obtained the continuous-time formulas as a limit of certain other discrete-time formulas given in Thm. 10.3.2 a little later. First however let us complete the discussion of the form (10.3.1).

Theorem 10.3.1 (RTS Recursions) *Given the standard model (10.2.1)–(10.2.2) with $S_i = 0$, F_i invertible, and $P_i > 0$, the smoothed estimator can be computed via*

$$\hat{\mathbf{x}}_{i|N} = F_{s,i}\hat{\mathbf{x}}_{i+1|N} + F_i^{-1}G_iQ_iG_i^*P_{i+1}^{-1}\hat{\mathbf{x}}_{i+1}, \quad i \leq N, \tag{10.3.3}$$

with

$$F_{s,i} \triangleq F_i^{-1}(I - G_iQ_iG_i^*P_{i+1}^{-1}) = P_iF_{p,i}^*P_{i+1}^{-1} = P_{i|i}F_i^*P_{i+1}^{-1}. \tag{10.3.4}$$

The error covariance matrix obeys

$$P_{i|N} = F_{s,i}P_{i+1|N}F_{s,i}^* + F_i^{-1}G_iQ_i^rG_i^*F_i^{-*}, \tag{10.3.5}$$

where

$$Q_i^r \triangleq Q_i - Q_iG_i^*P_{i+1}^{-1}G_iQ_i. \tag{10.3.6}$$

■

Proof: The alternative formulae for $F_{s,i}$ can be proved in several ways. For example, starting with the measurement update formula $P_{i+1} = F_iP_{i|i}F_i^* + G_iQ_iG_i^*$, we have $I - G_iQ_iG_i^*P_{i+1}^{-1} = F_iP_{i|i}F_i^*P_{i+1}^{-1}$, which gives one of the expressions in (10.3.4). For the other expression, recall that (*cf.* (9.3.4))

$$P_{i|i}F_i^* = (P_i - P_iH_i^*R_{e,i}^{-1}H_iP_i)F_i^* = P_iF_{p,i}^*. \tag{10.3.7}$$

A direct derivation of the error covariance equation involves considerable algebra. Discovering the answer in some such way and thinking about the nature of the solution leads to the following (ultimately more far-reaching) approach. Given

$$\hat{\mathbf{x}}_{i|N} = \underbrace{F_i^{-1}(I - G_iQ_iG_i^*P_{i+1}^{-1})}_{F_{s,i}}\hat{\mathbf{x}}_{i+1|N} + F_i^{-1}G_iQ_iG_i^*P_{i+1}^{-1}\hat{\mathbf{x}}_{i+1},$$

and $\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i$, we seek an equation for $\tilde{\mathbf{x}}_{i+1|N}$. For this purpose, we start by rewriting the state equation as

$$\mathbf{x}_i = F_i^{-1}\mathbf{x}_{i+1} - F_i^{-1}G_i\mathbf{u}_i,$$

$$= F_i^{-1}(I - G_iQ_iG_i^*P_{i+1}^{-1})\mathbf{x}_{i+1} - F_i^{-1}G_i\mathbf{u}_i + F_i^{-1}G_iQ_iG_i^*P_{i+1}^{-1}\mathbf{x}_{i+1}.$$

Subtraction of the equations for $\mathbf{x}_i$ and $\hat{\mathbf{x}}_{i|N}$ yields

$$\tilde{\mathbf{x}}_{i|N} = F_{s,i}\tilde{\mathbf{x}}_{i+1|N} - F_i^{-1}G_i\mathbf{u}_i^r, \quad \text{say,} \tag{10.3.8}$$

where we introduced

$$\mathbf{u}_i^r \triangleq \mathbf{u}_i - Q_iG_i^*P_{i+1}^{-1}\tilde{\mathbf{x}}_{i+1}. \tag{10.3.9}$$

Now, looking at (10.3.5), it would seem that this recursion would follow immediately as the variance recursion for the state-equation (10.3.8) if $\{\mathbf{u}_i^r\}$ were a white-noise process with

$$\langle\mathbf{u}_i^r, \mathbf{u}_j^r\rangle = (Q_i - Q_iG_i^*P_{i+1}^{-1}G_iQ_i)\delta_{ij} \triangleq Q_i^r\delta_{ij},$$

and such that

$$\langle\tilde{\mathbf{x}}_{i+1|N}, \mathbf{u}_i^r\rangle = 0.$$

The fact is that these properties are true, as can be verified by a direct calculation — see Prob. 10.5!

◆

10.3.2 The Smoothing Errors are Backwards Markov

Of course what the above calculations show is that the recursion (10.3.8) for $\tilde{\mathbf{x}}_{i|N}$ is exactly a backwards-time Markovian process, as defined and studied in Sec. 5.4.2.

At first glance the Markov property is surprising because unlike the error in the predicted estimator $\tilde{\mathbf{x}}_i$, the smoothing error depends upon both past and future data. The history of this result is discussed in Bello, Willsky, and Levy (1989), along with several different proofs and interesting applications to problems where multiple observations are taken of the same set of variables, *e.g.*, in successive passes of an airplane over a given terrain. Here we present only one proof, a very direct one motivated by the approach of Verghese and Kailath (1979) used to obtain backwards-time Markovian

models in Sec. 5.4.3. We may also note that the result is in fact a special case of a more general result that does not even assume state-space structure — see Prob. 5.10.

A Direct Proof by Time Reversal. We follow the argument we employed earlier in Sec. 5.4.3 (prior to Lemma 5.4.5), when we studied the problem of generating a backwards Markovian model from a forwards model via time reversal.

To see this, let us first determine a suitable forwards model for the smoothing error. Starting with the state equation $\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i$, we obtain $\hat{\mathbf{x}}_{i+1|N} = F_i\hat{\mathbf{x}}_{i|N} + G_i\hat{\mathbf{u}}_{i|N}$, which leads to the desired forwards-time error equation $\tilde{\mathbf{x}}_{i+1|N} = F_i\tilde{\mathbf{x}}_{i|N} + G_i\tilde{\mathbf{u}}_{i|N}$. We would like to reverse this equation in time. Now since F_i is assumed invertible, we obtain

$$\tilde{\mathbf{x}}_{i|N} = F_i^{-1}\tilde{\mathbf{x}}_{i+1|N} - F_i^{-1}G_i\tilde{\mathbf{u}}_{i|N}. \tag{10.3.10}$$

This model is still not backwards Markovian because in general $\langle\tilde{\mathbf{u}}_{i|N}, \tilde{\mathbf{x}}_{i+1|N}\rangle \neq 0$. However, we can make it so by rewriting $\tilde{\mathbf{u}}_{i|N}$ as

$$\tilde{\mathbf{u}}_{i|N} = \mathbf{u}_i^r + Q_iG_i^*P_{i+1}^{-1}\tilde{\mathbf{x}}_{i+1|N}, \tag{10.3.11}$$

as we shall explain. But first note that the above equality follows from (10.2.5) and (10.2.14) since

$$\hat{\mathbf{u}}_{i|N} = Q_iG_i^*\boldsymbol{\lambda}_{i+1|N} = Q_iG_i^*P_{i+1}^{-1}[\tilde{\mathbf{x}}_{i+1} - \tilde{\mathbf{x}}_{i+1|N}],$$

and, hence,

$$\tilde{\mathbf{u}}_{i|N} = \mathbf{u}_i - Q_iG_i^*P_{i+1}^{-1}[\tilde{\mathbf{x}}_{i+1} - \tilde{\mathbf{x}}_{i+1|N}] = \mathbf{u}_i^r + Q_iG_i^*P_{i+1}^{-1}\tilde{\mathbf{x}}_{i+1|N}. \tag{10.3.12}$$

The important fact about this decomposition of $\tilde{\mathbf{u}}_{i|N}$ into $\mathbf{u}_i^r$ and $\tilde{\mathbf{x}}_{i+1|N}$ is that $\{\mathbf{u}_i^r\}$ is a white sequence that is orthogonal to $\tilde{\mathbf{x}}_{i+1|N}$ (as established in Probs. 10.4 and 10.5). We exploit this by substituting expression (10.3.12) into the reversed-time state-space equation (10.3.10) to obtain

$$\tilde{\mathbf{x}}_{i|N} = F_{s,i}\tilde{\mathbf{x}}_{i+1|N} - F_i^{-1}G_i\mathbf{u}_i^r,$$

which shows (again) that $\tilde{\mathbf{x}}_{i|N}$ is a WSM process.

10.3.3 The Original Rauch-Tung-Striebel (RTS) Formulas

As mentioned in Sec. 10.3.3, an alternative set of discrete-time formulas is what generally goes by the name RTS formulas. They are slightly more general than the set in Thm. 10.3.1 in that they do not require the invertibility of the F_i. To explore what is possible, we write the BF formula (10.2.5) for the time instant $i+1$ and use it to note that

$$F_{p,i}^*P_{i+1}^{-1}\hat{\mathbf{x}}_{i+1|N} = F_{p,i}^*P_{i+1}^{-1}(F_i\hat{\mathbf{x}}_i + K_{p,i}\mathbf{e}_i) - H_i^*R_{e,i}^{-1}\mathbf{e}_i + P_i^{-1}(\hat{\mathbf{x}}_{i|N} - \hat{\mathbf{x}}_i),$$

or better, since we are assuming $S_i = 0$, that

$$P_iF_{p,i}^*P_{i+1}^{-1}\hat{\mathbf{x}}_{i+1|N} = \hat{\mathbf{x}}_{i|N} + (P_iF_{p,i}^*P_{i+1}^{-1}F_i - I)\hat{\mathbf{x}}_i + (P_iF_{p,i}^*P_{i+1}^{-1}F_i - I)P_iH_i^*R_{e,i}^{-1}\mathbf{e}_i.$$

If we now denote $P_iF_{p,i}^*P_{i+1}^{-1}$ by $F_{s,i}$ (which is consistent with (10.3.4)), we obtain the backwards recursion

$$\begin{aligned}\hat{\mathbf{x}}_{i|N} &= F_{s,i}\hat{\mathbf{x}}_{i+1|N} + (I - F_{s,i}F_i)(\hat{\mathbf{x}}_i + P_iH_i^*R_{e,i}^{-1}\mathbf{e}_i),\\ &= F_{s,i}\hat{\mathbf{x}}_{i+1|N} + (I - F_{s,i}F_i)\hat{\mathbf{x}}_{i|i}.\end{aligned} \tag{10.3.13}$$

This is in fact one of the formulas obtained, using a very different argument, by Rauch, Tung, and Striebel (1965).

Theorem 10.3.2 (The Original RTS Formulas) *Consider the model (10.2.1)–(10.2.2), and assume that $S_i = 0$ and $P_i > 0$, for $0 \le i \le N$. Define $F_{s,i} = P_iF_{p,i}^*P_{i+1}^{-1}$. Then we can write*

$$\hat{\mathbf{x}}_{i|N} = F_{s,i}\hat{\mathbf{x}}_{i+1|N} + (\hat{\mathbf{x}}_{i|i} - F_{s,i}\hat{\mathbf{x}}_{i+1}), \tag{10.3.14}$$

and

$$P_{i|N} = F_{s,i}P_{i+1|N}F_{s,i}^* + P_{i|i} - F_{s,i}P_{i+1}F_{s,i}^*, \tag{10.3.15}$$

where the boundary conditions $\hat{\mathbf{x}}_{N+1|N}$ and $P_{N+1|N}$ can be obtained by applying the appropriate Kalman filter recursions to the data $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}$ (cf. Thms. 9.2.1 and 9.5.1). ■

Proof: Eq. (10.3.14) was deduced in the prior discussion since $\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_{i|i}$. But knowing the result, a more direct proof is possible. Starting with the BF formulas (10.2.5) and (10.2.10) we write

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_{i|i} + P_iF_{p,i}^*\boldsymbol{\lambda}_{i+1|N} = \hat{\mathbf{x}}_{i|i} + P_iF_{p,i}^*P_{i+1}^{-1}(\hat{\mathbf{x}}_{i+1|N} - \hat{\mathbf{x}}_{i+1}).$$

Now the desired result follows by using the definition of $F_{s,i}$ and the fact that $\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_{i|i}$ when $S_i = 0$.

For the error covariance formula, one method is to compute

$$\|\hat{\mathbf{x}}_{i|N}\|^2 = \|\hat{\mathbf{x}}_{i|i}\|^2 + P_iF_{p,i}^*\Lambda_{i+1|N}F_{p,i}P_i,$$

and then use the definition of $F_{s,i}$ and the BF formula (10.2.8) in the form

$$P_{i+1}^{-1}(P_{i+1} - P_{i+1|N})P_{i+1}^{-1} = \Lambda_{i+1|N},$$

to conclude that

$$\|\hat{\mathbf{x}}_{i|N}\|^2 = \|\hat{\mathbf{x}}_{i|i}\|^2 + F_{s,i}(P_{i+1} - P_{i+1|N})F_{s,i}^*.$$

Finally, use $P_{i|N} = \|\mathbf{x}_i\|^2 - \|\hat{\mathbf{x}}_{i|N}\|^2$ to obtain

$$\begin{aligned}P_{i|N} &= \|\mathbf{x}_i\|^2 - \|\hat{\mathbf{x}}_{i|i}\|^2 - F_{s,i}(P_{i+1} - P_{i+1|N})F_{s,i}^*,\\ &= P_{i|i} + F_{s,i}P_{i+1|N}F_{s,i}^* - F_{s,i}P_{i+1}F_{s,i}^*.\end{aligned}$$

◆

Remark 1. The RTS algorithm is also a two-pass algorithm, with all smoothed estimators being directly obtained at the end of the backwards pass; note that we need only the estimators $\{\hat{\mathbf{x}}_i\}$ and $\{\hat{\mathbf{x}}_{i|i}\}$ for the second pass, the original data $\{\mathbf{y}_i\}$ and even the innovations $\{\mathbf{e}_i\}$ need not be retained. The differences between the BF and RTS algorithms are small and much will depend on the actual codes and machines on which the algorithms are run. ◆

10.4 TWO-FILTER FORMULAS

The smoothing formulas in Secs. 10.1–10.3 were based on the formula (10.1.4), which expresses $\hat{\mathbf{x}}_{i|N}$ in terms of the forwards innovations. But for fixed-interval smoothing problems, the direction of time is not important, and we should be able to also process the data backwards starting with $\mathbf{y}_N$ and ending with $\mathbf{y}_0$. In fact, analogs of all the earlier algorithms can be obtained in this way, and some of these will be presented in the problems.

Moreover, by suitably combining the forwards and backwards expressions, we can obtain a set of so-called two-filter formulas, which yield smoothed estimators as appropriate combinations of forward and backwards estimators. The first set of such formulas was independently obtained, by very different arguments, by Mayne (1966), Fraser (1967), and Fraser and Potter (1969). Their results stimulated various questions that led to a more general set of formulas, which we shall present first. [As will be explained below, it was the search for a better understanding of the Mayne-Fraser-Potter formulas that prompted the first studies of backwards-time Markovian models (see Ljung and Kailath (1976b) and Verghese and Kailath (1979)), which ultimately led to Thm. 10.4.1.]

10.4.1 General Two Filter Formulas

It turns out that (in discrete time), we have two different choices. One is to combine forwards *predicted* estimators,

$$\hat{\mathbf{x}}_i \;=\; \text{the l.l.m.s. estimator of } \mathbf{x}_i \text{ given } \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{i-1}\},$$

and backwards *filtered* estimators,[2]

$$\hat{\mathbf{x}}^b_{i|i} \;=\; \text{the l.l.m.s. estimator of } \mathbf{x}_i \text{ given } \{\mathbf{y}_i, \mathbf{y}_{i+1}, \ldots, \mathbf{y}_N\}.$$

We could also combine forwards *filtered* estimators $\hat{\mathbf{x}}_{i|i}$ and backwards *predicted* estimators,

$$\hat{\mathbf{x}}^b_i \;=\; \text{the l.l.m.s. estimator of } \mathbf{x}_i \text{ given } \{\mathbf{y}_{i+1}, \mathbf{y}_{i+2}, \ldots, \mathbf{y}_N\}.$$

Theorem 10.4.1 (General Two-Filter Smoothing Formulas) *Given the state-space model (10.2.1)–(10.2.2) with F_i invertible and $S_i = 0$, we can write*

$$\hat{\mathbf{x}}_{i|N} \;=\; P_{i|N}\left(P_{i|i}^{-1}\hat{\mathbf{x}}_{i|i} + P_i^{-b}\hat{\mathbf{x}}^b_i\right), \tag{10.4.1}$$

[2] Note that $\hat{\mathbf{x}}^b_{i|i} \neq \sum_{j=i}^{N}\langle \mathbf{x}^b_i, \mathbf{e}_j\rangle R^{-1}_{e,j}\mathbf{e}_j$, since the innovations $\{\mathbf{e}_i, \ldots, \mathbf{e}_N\}$ do not span the same space as the observations $\{\mathbf{y}_i, \ldots, \mathbf{y}_N\}$!

with

$$P_{i|N} \;=\; \left(P_{i|i}^{-1} + P_i^{-b} - \Pi_i^{-1}\right)^{-1}, \tag{10.4.2}$$

or, alternatively,

$$\hat{\mathbf{x}}_{i|N} \;=\; P_{i|N}\left(P_i^{-1}\hat{\mathbf{x}}_i + P_{i|i}^{-b}\hat{\mathbf{x}}^b_{i|i}\right), \tag{10.4.3}$$

with

$$P_{i|N} \;=\; \left(P_i^{-1} + P_{i|i}^{-b} - \Pi_i^{-1}\right)^{-1}. \tag{10.4.4}$$

■

Proof: One proof will be given in Sec. 10.4.3; another is pursued in the problems at the end of this chapter (Probs. 10.13–10.14); a third proof based on the notion of dual bases is given in Sec. 15.7.6. [We may again note (see Prob. 5.15) that the invertibility of Π_i can be guaranteed by assuming, for example, $\Pi_0 > 0$ and F_i invertible.] ◆

Both formulas use forwards sweeps from $\mathbf{y}_0$ to $\mathbf{y}_i$ (or $\mathbf{y}_{i-1}$) to compute $\hat{\mathbf{x}}_{i|i}$ (or $\hat{\mathbf{x}}_i$), and backwards sweeps from $\mathbf{y}_N$ to $\mathbf{y}_{i+1}$ (or $\mathbf{y}_i$) to compute $\mathbf{x}^b_i$ (or $\hat{\mathbf{x}}^b_{i|i}$). The "backwards" quantities $\{\hat{\mathbf{x}}^b_i, \hat{\mathbf{x}}^b_{i|i}, P^b_i, P^b_{i|i}\}$ are computed as in Thms. 9.8.1 and 9.8.2, while the "forwards" quantities $\{\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_{i|i}, P_i, P_{i|i}\}$ are computed as in Thms. 9.2.1 and 9.5.1.

10.4.2 The Mayne and Fraser-Potter Formulas

It turns out that we have a useful degree of freedom in computing the backwards quantities in the above formulas. In particular, as we shall see, the recursions would be much simpler if we could assume (rather arbitrarily for now) that $\Pi_i = \infty \cdot I$, so that all the terms in Π_i^{-1} could be dropped; this would give the original two-filter formulas derived by Mayne (1966) and Fraser and Potter (1969). Some notation needs to be introduced (following the backward Kalman filter recursions of Thms. 9.8.1 and 9.8.2).

We thus assume that F_i is invertible and $S_i = 0$. Now introduce the quantities

$$R^b_{e,i,\infty} \;\triangleq\; R_i + H_i F_i^{-1}\left[P^b_{i+1,\infty|i+1} + G_i Q_i G_i^*\right] F_i^{-*} H_i^*, \tag{10.4.5}$$

$$K^b_{p,i,\infty} \;\triangleq\; F_i^{-1}\left[P^b_{i+1,\infty|i+1} + G_i Q_i G_i^*\right] F_i^{-*} H_i^* R^{-b}_{e,i,\infty}, \tag{10.4.6}$$

where $P^b_{i,\infty|i}$ satisfies the backwards recursion

$$P^b_{i,\infty|i} \;=\; F_i^{-1} P^b_{i+1,\infty|i+1} F_i^{-*} + F_i^{-1} G_i Q_i G_i^* F_i^{-*} - K^b_{p,i,\infty} R^b_{e,i,\infty} K^{b*}_{p,i,\infty}, \tag{10.4.7}$$

with boundary condition $P^b_{N+1,\infty|N+1} = \infty \cdot I$. [The subscript ∞ is used to indicate that these quantities are generated with a boundary condition that is equal to ∞.] Prob. 10.22 then shows, starting with the recursions of Thm. 9.8.2 and assuming $\Pi_0 > 0$, that the variables $\{R^b_{e,i,\infty}, K^b_{p,i,\infty}, P^b_{i,\infty|i}\}$ can be used to determine a filtered backward state estimator, which we shall denote by $\hat{\mathbf{x}}^b_{i,\infty|i}$.

Likewise, let

$$R^b_{e,i,\infty} \triangleq R_i + H_i P^b_{i,\infty} H_i^*, \; K^b_{l,i,\infty} \triangleq F_{i-1}^{-1} P^b_{i,\infty} H_i^* R^{-b}_{e,i,\infty}, \tag{10.4.8}$$

where $P^b_{i,\infty}$ satisfies the backwards recursion

$$P^b_{i,\infty} = F_i^{-1} P^b_{i+1,\infty} F_i^{-*} + F_i^{-1} G_i Q_i G_i^* F_i^{-*} - K^b_{l,i+1,\infty} R^b_{e,i,\infty} K^{b*}_{l,i,\infty}, \tag{10.4.9}$$

with boundary condition $P^b_{N+1,\infty} = \infty \cdot I$. An argument similar to Prob. 10.22, and starting now with the recursions of Thm. 9.8.1, will also show that the variables $\{R^b_{e,i,\infty}, K^b_{l,i,\infty}, P^b_{i,\infty}\}$ can be used to determine a predicted backward state estimator, which we shall denote by $\hat{\mathbf{x}}^b_{i,\infty}$.

In terms of these various quantities, we can now state the following variant of the two-filter formulas of Thm. 10.4.1.

Theorem 10.4.2 (The Fraser-Potter Formulas) *Given (10.2.1)–(10.2.2) with F_i invertible, $\Pi_0 > 0$, and $S_i = 0$, we can write*

$$\hat{\mathbf{x}}_{i|N} = P_{i|N}\left(P_{i|i}^{-1}\hat{\mathbf{x}}_{i|i} + P^{-b}_{i,\infty}\hat{\mathbf{x}}^b_{i,\infty}\right), \qquad P^{-1}_{i|N} = \left(P^{-1}_{i|i} + P^{-b}_{i,\infty}\right)^{-1}.$$

Likewise,

$$\hat{\mathbf{x}}_{i|N} = P_{i|N}\left(P_i^{-1}\hat{\mathbf{x}}_i + P^{-b}_{i,\infty|i}\hat{\mathbf{x}}^b_{i,\infty|i}\right), \qquad P^{-1}_{i|N} = \left(P^{-1}_i + P^{-b}_{i,\infty|i}\right)^{-1}. \tag{10.4.10}$$

The error covariance matrices $\{P^b_{i,\infty}, P^b_{i,\infty|i}\}$ are recursively computed via (10.4.7) and (10.4.9). Correspondingly, the state estimators $\{\hat{\mathbf{x}}^b_{i,\infty}, \hat{\mathbf{x}}^b_{i,\infty}\}$ are obtained as follows (cf. Thms. 9.8.1 and 9.8.2):

$$\hat{\mathbf{x}}^b_{i,\infty} = F_i^{-1}\hat{\mathbf{x}}^b_{i+1,\infty} + K^b_{l,i+1,\infty}[\mathbf{y}_{i+1} - H_{i+1}\hat{\mathbf{x}}^b_{i+1,\infty}], \tag{10.4.11}$$

$$\hat{\mathbf{x}}^b_{i,\infty|i} = F_i^{-1}\hat{\mathbf{x}}^b_{i+1,\infty|i+1} + K^b_{p,i,\infty}[\mathbf{y}_i - H_i F_i^{-1}\hat{\mathbf{x}}^b_{i+1,\infty|i+1}], \tag{10.4.12}$$

with boundary conditions $\hat{\mathbf{x}}^b_{N+1,\infty} = \hat{\mathbf{x}}^b_{N+1,\infty|N+1} = 0$. ∎

The above expressions are the discrete-time version of formulas first given by Fraser and Potter (1969), who interpreted the equation for $\hat{\mathbf{x}}^b_{i,\infty}$, for example, as the Kalman filter equation for processing the data $\{\mathbf{y}_{i+1}, \ldots, \mathbf{y}_N\}$ backwards starting with $\mathbf{y}_N$. However, this interpretation is only valid under the somewhat arbitrary assumption that $P^b_{N+1} = \infty \cdot I$ (*cf.* (10.4.9)). Some heuristic arguments were provided but they are incomplete (a nice discussion is provided in Wall, Willsky, and Sandell (1981)). A full justification requires a proper definition of backwards time (Markovian) state-space models, which cannot be obtained just by reversing the direction of time in the forwards model (see Sec. 5.4.2). When this is done, one is led to the general formulas of Thm. 10.4.1 with Π_i^{-1}. One proof is presented in Sec. 10.4.3.

Here we only remark that Fraser and Potter were aware that direct use of the formulas in Thm. 10.4.2 was impractical and proposed a slightly different set of equations, which as Fraser and Potter (1969) noted were just those originally obtained by Mayne (1966). To avoid the infinities, we directly propagate $P^{-b}_{i,\infty|i}$ and $P^{-b}_{i,\infty|i}\hat{\mathbf{x}}^b_{i,\infty|i}$.

Theorem 10.4.3 (The Mayne Formulas) *Given (10.2.1)–(10.2.2) with F_i invertible, $\Pi_0 > 0$, and $S_i = 0$, the smoothed estimator can be computed as*

$$\hat{\mathbf{x}}_{i|N} = P_{i|N}(P_i^{-1}\hat{\mathbf{x}}_i + \mathbf{z}^b_i), \tag{10.4.13}$$

$$\mathbf{z}^b_i = F_i^*(I + L^b_{i+1}G_iQ_iG_i^*)^{-1}\mathbf{z}^b_{i+1} + H_i^*R_i^{-1}\mathbf{y}_i, \quad \mathbf{z}^b_{N+1} = 0, \tag{10.4.14}$$

where

$$P^{-1}_{i|N} = P_i^{-1} + L_i^{-b},$$

$$L^b_i = F_i^*(I + L^b_{i+1}G_iQ_iG_i^*)^{-1}L^b_{i+1}F_i + H_i^*R_i^{-1}H_i, \tag{10.4.15}$$

with boundary condition $L^b_{N+1} = 0$. Moreover, by comparing directly with the recursions of Thm. 10.4.2, it follows that

$$\mathbf{z}^b_i = P^{-b}_{i,\infty|i}\hat{\mathbf{x}}^b_{i,\infty|i}, \qquad L^b_i = P^{-b}_{i,\infty|i}.$$

∎

10.4.3 Combined Estimators Derivation

The two-filter formulas can be derived in several ways. But one of the simplest (once we know about backwards Markovian models) arises from noting the similarity, which may already have occurred to the alert reader, between the two-filter formulas and the simple results for combining estimators presented in Sec. 3.4.3. We shall describe this now following Kailath and Wax (1984).

We first repeat here, for ease of reference, the conclusion of Lemma 3.4.1 concerning the procedure for combining estimators. More specifically, let $\mathbf{y}_a$ and $\mathbf{y}_b$ be two separate observations of a zero-mean random variable $\mathbf{x}$, such that

$$\mathbf{y}_a = H_a\mathbf{x} + \mathbf{v}_a, \qquad \mathbf{y}_b = H_b\mathbf{x} + \mathbf{v}_b,$$

where $\{\mathbf{v}_a, \mathbf{v}_b, \mathbf{x}\}$ are mutually uncorrelated zero-mean random variables with covariance matrices R_a, R_b, and R_x, respectively. Denote by $\hat{\mathbf{x}}_a$ and $\hat{\mathbf{x}}_b$ the l.l.m.s. estimators of $\mathbf{x}$ given $\mathbf{y}_a$ and $\mathbf{y}_b$, respectively, and likewise define the error covariance matrices, $P_a = \|\mathbf{x} - \hat{\mathbf{x}}_a\|^2$ and $P_b = \|\mathbf{x} - \hat{\mathbf{x}}_b\|^2$. Then $\hat{\mathbf{x}}$, the l.l.m.s. estimator of $\mathbf{x}$ given both $\mathbf{y}_a$ and $\mathbf{y}_b$, can be found as

$$P^{-1}\hat{\mathbf{x}} = P_a^{-1}\hat{\mathbf{x}}_a + P_b^{-1}\hat{\mathbf{x}}_b, \tag{10.4.16}$$

where P, the corresponding error covariance matrix, is given by

$$P^{-1} = P_a^{-1} + P_b^{-1} - R_x^{-1}. \tag{10.4.17}$$

To apply this result in the context of state-space smoothing, we call upon the forwards and backwards Markovian models of $\{\mathbf{x}_i, \mathbf{y}_i\}$, *viz.*, the standard state-space model

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i, \qquad \mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i,$$

and (*cf.* Sec. 9.8.1)

$$\mathbf{x}_i = F^b_{i+1}\mathbf{x}_{i+1} + \mathbf{u}^b_{i+1}, \qquad \mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i.$$

It is interesting to note that in order to allow the application of (10.4.16)–(10.4.17) we use the backwards model to express the *past* observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$ in terms of $\mathbf{x}_i$, while the forwards model is used to express the *present and future* observations $\{\mathbf{y}_i, \ldots, \mathbf{y}_N\}$ in terms of the same state $\mathbf{x}_i$:

$$\underbrace{\begin{bmatrix} \mathbf{y}_0 \\ \vdots \\ \mathbf{y}_{i-2} \\ \mathbf{y}_{i-1} \end{bmatrix}}_{\mathbf{y}_a} = \underbrace{\begin{bmatrix} H_0\Phi^b(0,i) \\ \vdots \\ H_{i-2}\Phi^b(i-2,i) \\ H_{i-1}\Phi^b(i-1,i) \end{bmatrix}}_{H_a}\mathbf{x}_i + \underbrace{\begin{bmatrix} \sum_{n=1}^{i} H_0\Phi^b(0,n-1)\mathbf{u}^b_n + \mathbf{v}_0 \\ \vdots \\ H_{i-2}F^b_{i-1}\mathbf{u}^b_i + H_{i-2}\mathbf{u}^b_{i-1} + \mathbf{v}_{i-2} \\ H_{i-1}\mathbf{u}^b_i + \mathbf{v}_{i-1} \end{bmatrix}}_{\mathbf{v}_a}, \tag{10.4.18}$$

and

$$\underbrace{\begin{bmatrix} \mathbf{y}_i \\ \mathbf{y}_{i+1} \\ \vdots \\ \mathbf{y}_N \end{bmatrix}}_{\mathbf{y}_b} \underbrace{\begin{bmatrix} H_i \\ H_{i+1}\Phi(i+1,i) \\ \vdots \\ H_N\Phi(N,i) \end{bmatrix}}_{H_b}\mathbf{x}_i + \underbrace{\begin{bmatrix} \mathbf{v}_i \\ H_{i+1}G_i\mathbf{u}_i + \mathbf{v}_{i+1} \\ \vdots \\ \sum_{n=i}^{N-1} H_N\Phi(N,n+1)G_n\mathbf{u}_n + \mathbf{v}_N \end{bmatrix}}_{\mathbf{v}_b}, \tag{10.4.19}$$

where

$$\Phi(k,i) = F_{k-1}F_{k-2}\ldots F_i, \qquad \Phi^b(k,i) = F^b_{k+1}F^b_{k+2}\ldots F^b_i. \tag{10.4.20}$$

Now since the $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are uncorrelated with each other, and since

$$\langle \mathbf{u}^b_k, \mathbf{x}_i\rangle = \langle \mathbf{v}_k, \mathbf{x}_i\rangle = 0 \quad \text{for } k \le i \quad \text{and} \quad \langle \mathbf{u}_k, \mathbf{x}_i\rangle = \langle \mathbf{v}_k, \mathbf{x}_i\rangle = 0 \quad \text{for } k > i,$$

it follows that

$$\langle \mathbf{v}_a, \mathbf{v}_b\rangle = 0, \qquad \langle \mathbf{v}_a, \mathbf{x}_i\rangle = 0, \qquad \langle \mathbf{v}_b, \mathbf{x}_i\rangle = 0. \tag{10.4.21}$$

So we are exactly in the situation envisaged in Lemma 3.4.1, and the rest is easy. If we identify

$$\mathbf{x} = \mathbf{x}_i, \qquad \hat{\mathbf{x}}_a = \hat{\mathbf{x}}_i, \qquad \hat{\mathbf{x}}_b = \hat{\mathbf{x}}^b_{i|i},$$

then (10.4.16)–(10.4.17) become exactly the second set of formulas, (10.4.3)–(10.4.4) in Thm. 10.4.1.

To obtain the first set of formulas (10.4.1)–(10.4.2) we again start from (10.4.18)–(10.4.19), only this time we group the *present* term with the *past* terms, and repeat the same steps to obtain

$$\hat{\mathbf{x}}_{i|N} = P_{i|N}\left(P^{-1}_{i|i}\hat{\mathbf{x}}_{i|i} + P_i^{-b}\hat{\mathbf{x}}^b_i\right), \tag{10.4.22}$$

$$P_{i|N} = \left(P^{-1}_{i|i} + P_i^{-b} - \Pi_i^{-1}\right)^{-1}. \tag{10.4.23}$$

If instead of the l.l.m.s. (Bayesian) estimator we use the Fisher (or deterministic) estimator while estimating $\mathbf{x}_i$ from $\mathbf{y}_b$, we shall obtain the Fraser-Potter formulas of Thm. 10.4.2.

10.5 THE HAMILTONIAN EQUATIONS ($R_i > 0$)

In studying the Kalman filtering problem, we noted (see Sec. 9.5) that the assumption $R_i > 0$ was generally a condition to seek in setting up our state-space model. Since the predicted and filtered estimators essentially determine the smoothed estimators, the condition $R_i > 0$ is helpful in smoothing problems too. Moreover, under this assumption, we can get a simpler form for the adjoint variable recursions and also introduce certain so-called Hamiltonian equations for the smoothing problem. These equations have several interesting features.

We shall assume for simplicity that $S_i = 0$ (see Prob. 10.15 for $S_i \neq 0$). Then we note that after some algebraic manipulation of the BF formulas of Sec. 10.2.1, we can rewrite the recursion (10.2.7) for $\boldsymbol{\lambda}_{i|N}$ as (another proof is suggested in Prob. 10.6),

$$\boldsymbol{\lambda}_{i|N} = F_i^*\boldsymbol{\lambda}_{i+1|N} - H_i^*R_i^{-1}H_i\hat{\mathbf{x}}_{i|N} + H_i^*R_i^{-1}\mathbf{y}_i. \tag{10.5.1}$$

We also derived earlier in (10.2.15) the RTS formula

$$\hat{\mathbf{x}}_{i+1|N} = F_i\hat{\mathbf{x}}_{i|N} + G_iQ_iG_i^*\boldsymbol{\lambda}_{i+1|N}. \tag{10.5.2}$$

[An alternative derivation is suggested in Prob. 10.7.] Combining these equations in matrix form we get

$$\begin{bmatrix} \hat{\mathbf{x}}_{i+1|N} \\ \boldsymbol{\lambda}_{i|N} \end{bmatrix} = \begin{bmatrix} F_i & G_iQ_iG_i^* \\ -H_i^*R_i^{-1}H_i & F_i^* \end{bmatrix}\begin{bmatrix} \hat{\mathbf{x}}_{i|N} \\ \boldsymbol{\lambda}_{i+1|N} \end{bmatrix} + \begin{bmatrix} 0 \\ H_i^*R_i^{-1} \end{bmatrix}\mathbf{y}_i, \tag{10.5.3}$$

where from (10.2.7) and (10.2.5) the boundary conditions are found to be

$$\hat{\mathbf{x}}_{0|N} = \Pi_0\boldsymbol{\lambda}_{0|N}, \qquad \boldsymbol{\lambda}_{N+1|N} = 0. \tag{10.5.4}$$

These are the Hamiltonian equations, so-called because equations of this type were encountered in certain classical (deterministic) variational problems (associated with famous names such as L. Euler, J. L. Lagrange, W. R. Hamilton) — see Secs. 10.6 and 10.7. As we shall see (in Ch. 17),[3] these equations are quite valuable in estimation theory as well. However an alert reader will note that the "recursions" in (10.5.3)–(10.5.4) are different from those we have seen so far, because the boundary conditions (10.5.4) are "mixed": one variable is specified at $i = N+1$ ($\boldsymbol{\lambda}_{N+1|N} = 0$), but the other one only at $i = 0$ (and that too only implicitly, $\hat{\mathbf{x}}_{0|N} = \Pi_0\boldsymbol{\lambda}_{0|N}$). Clearly, such so-called Two-Point Boundary Value Problems (TPBVP) cannot be solved by iteration!

There are of course several methods of solution (see, *e.g.*, Bryson and Ho (1969)). For example, certain so-called "sweep" or "Riccati transformation" methods will lead us back to formulas we already know — the BF and RTS formulas. This should not surprise the reader, because we actually derived the Hamiltonian equations by using these

[3] In Sec. 15.7.3, we shall derive the equations (10.5.3) directly without, as here, assuming knowledge of various filtering and smoothing formulas.

formulas. For us another interesting method arises from exploiting the equivalence between stochastic and deterministic estimation problems (*cf.* Sec. 3.5) as is discussed in Sec. 10.7 below.

Sweep Methods (Triangularization of the Hamiltonian). First we comment briefly on the "sweep" or "Riccati transformation" methods, which can also be interpreted as methods for lower or upper triangularizing the Hamiltonian equations. That is, they try to find a transformation from the variables $\{\hat{\mathbf{x}}_{i|N}, \boldsymbol{\lambda}_{i|N}\}$ to another pair $\{\mathbf{z}_{F,i}, \mathbf{z}_{B,i}\}$ such that for the new variables the equations have triangular form, say

$$\begin{bmatrix} \mathbf{z}_{F,i+1} \\ \mathbf{z}_{B,i} \end{bmatrix} = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix} \begin{bmatrix} \mathbf{z}_{F,i} \\ \mathbf{z}_{B,i+1} \end{bmatrix} + \begin{bmatrix} * \\ * \end{bmatrix} \mathbf{y}_i, \tag{10.5.5}$$

with boundary conditions $\{\mathbf{z}_{F,N+1}, \mathbf{z}_{B,0}\}$. Now using (10.5.5) we can first solve for the $\{\mathbf{z}_{F,i}\}$, then for the $\{\mathbf{z}_{B,i}\}$, and then transform back to the $\{\hat{\mathbf{x}}_{i|N}, \boldsymbol{\lambda}_{i|N}\}$.

In fact, with hindsight (or after some algebra), one can find that an appropriate transformation that achieves the lower-triangularization (10.5.5) is of the form

$$\mathbf{z}_{F,i} = \hat{\mathbf{x}}_{i|N} - P_{F,i}\boldsymbol{\lambda}_{i|N}, \qquad \mathbf{z}_{B,i} = \hat{\mathbf{x}}_{i|N}.$$

Carrying out this transformation, where $P_{F,i}$ is to be determined, one will find, after some algebra, that we can identify $\mathbf{z}_{F,i} = \hat{\mathbf{x}}_i$ and $P_{F,i} = P_i$. In other words, our transformation is the BF formula in reverse.

We repeat that this is not surprising; after all we derived the Hamiltonian equations from the BF formulas (or rather from a consequence of the BF formulas — the RTS formulas). It looks as if we are just spinning our wheels! But before rushing on to more fruitful ventures, let us raise the following question: what if we upper triangularize the Hamiltonian equations, rather than lower triangularize them? And in fact, what if we do both, and thereby diagonalize the Hamiltonian equations? Say,

$$\begin{bmatrix} \mathbf{z}_{F,i+1} \\ \mathbf{z}_{B,i} \end{bmatrix} = \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} \begin{bmatrix} \mathbf{z}_{F,i} \\ \mathbf{z}_{B,i+1} \end{bmatrix} + \begin{bmatrix} * \\ * \end{bmatrix} \mathbf{y}_i,$$

in the upper triangularized case or

$$\begin{bmatrix} \mathbf{z}_{F,i+1} \\ \mathbf{z}_{B,i} \end{bmatrix} = \begin{bmatrix} * & 0 \\ 0 & * \end{bmatrix} \begin{bmatrix} \mathbf{z}_{F,i} \\ \mathbf{z}_{B,i+1} \end{bmatrix} + \begin{bmatrix} * \\ * \end{bmatrix} \mathbf{y}_i,$$

in the diagonalized case, with boundary conditions $\{\mathbf{z}_{F,N+1}, \mathbf{z}_{B,0}\}$. Will we get some new results? The safe answer is yes, and this is correct. It turns out that upper triangularization effectively leads to the "backwards-time" BF and RTS formulas (described in Probs. 10.11–10.12); again this is not surprising, at least in retrospect. Finally, diagonalization leads to a solution that should appropriately combine forwards- and backwards-time BF/RTS formulas. This is true, and in fact the resulting so-called two-filter formulas are the Mayne and Fraser-Potter formulas. We shall not go through the details here — the (simpler) continuous-time calculations were presented in Kailath and Ljung (1982).

Extended Hamiltonian Equations. In cases where the matrices $\{R_i\}$ are ill-conditioned, it is desirable to see if the explicit presence of $\{R_i^{-1}\}$ can be avoided. For this purpose, define $\boldsymbol{\mu}_{i|N}$ via

$$R_i\boldsymbol{\mu}_{i|N} \triangleq \mathbf{y}_i - H_i\hat{\mathbf{x}}_{i|N}.$$

Then the equations (10.5.3) can be rearragned as

$$\begin{bmatrix} \hat{\mathbf{x}}_{i+1|N} \\ 0 \\ \boldsymbol{\lambda}_{i|N} \end{bmatrix} = \begin{bmatrix} G_iQ_iG_i^* & 0 & F_i \\ 0 & R_i & H_i \\ F_i^* & H_i^* & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{i+1|N} \\ \boldsymbol{\mu}_{i|N} \\ \hat{\mathbf{x}}_{i|N} \end{bmatrix} + \begin{bmatrix} 0 \\ -I \\ 0 \end{bmatrix} \mathbf{y}_i. \tag{10.5.6}$$

This form was introduced and used by Van Dooren (1981) and by Whittle (1983,1990) — see Sec. 10.7.6 and also the remark at the end of Sec. E.7.

10.6 VARIATIONAL ORIGIN OF HAMILTONIAN EQUATIONS

We mentioned in Sec. 10.5 that the Hamiltonian equations were originally encountered in certain classical deterministic variational problems. We present one of them here (see also Whittle (1990, App. 4) for the tie into classical physics). Consider a collection of $(N+1)$ deterministic data $\{y_i, x_i\}_{i=0}^N$, where the y_i and x_i are column vectors with the x_i satisfying the state equation

$$x_{i+1} = F_ix_i + G_iu_i. \tag{10.6.1}$$

We are also given matrices H_i, a positive-definite matrix Π_0, and positive-definite matrices $\{Q_i, R_i\}$. Then we pose the following deterministic least-squares problem: subject to the state-equation constraint (10.6.1), solve

$$\min_{\{x_0,u_0,\ldots,u_N\}} \left[x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^N (y_i - H_ix_i)^*R_i^{-1}(y_i - H_ix_i) + \sum_{i=0}^N u_i^*Q_i^{-1}u_i \right]. \tag{10.6.2}$$

We shall denote the resulting solution variables by $\{\hat{x}_{0|N}, \hat{u}_{j|N}\}$.

To solve (10.6.2), we employ a Lagrange multiplier argument and define the extended cost function

$$J_\lambda \triangleq \left[x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^N (y_i - H_ix_i)^*R_i^{-1}(y_i - H_ix_i) + \sum_{i=0}^N u_i^*Q_i^{-1}u_i \right] +$$

$$\sum_{i=0}^N \mathrm{Re}\left(\lambda_{i+1|N}^*[x_{i+1} - F_ix_i - G_iu_i]\right),$$

where $\boldsymbol{\lambda}_{i+1|N}$ is a column vector of Lagrange multipliers. Differentiating J_λ with respect to x_i and setting the derivative equal to zero leads to

$$\left.\frac{\partial J_\lambda}{\partial x_i}\right|_{\hat{x}_{i|N}} = -(y_i - H_i\hat{x}_{i|N})^*R_i^{-1}H_i + \lambda_{i|N}^* - \lambda_{i+1|N}^*F_i = 0, \qquad \text{for } 0 < i \le N,$$

where we are denoting the resulting stationary solution by $\hat{x}_{i|N}$. The above equation is equivalent to (*cf.* (10.5.1))

$$\lambda_{i|N} = F_i^* \lambda_{i+1|N} - H_i^* R_i^{-1} H_i \hat{x}_{i|N} + H_i^* R_i^{-1} y_i. \tag{10.6.3}$$

At the particular time instant $i = 0$ we need to consider the extra term $x_0^* \Pi_0^{-1} x_0$; thus leading to

$$\left.\frac{\partial J_\lambda}{\partial x_0}\right|_{\hat{x}_{0|N}} = \hat{x}_{0|N}^* \Pi_0^{-1} - (y_0 - H_0 \hat{x}_{0|N})^* R_0^{-1} H_0 - \lambda_{1|N}^* F_0 = 0,$$

which, in view of (10.6.3), implies that we must have $\hat{x}_{0|N}^* \Pi_0^{-1} - \lambda_{0|N} = 0$, or, equivalently, $\hat{x}_{0|N} = \Pi_0 \lambda_{0|N}$. Note also that

$$\left.\frac{\partial J_\lambda}{\partial x_{N+1}}\right|_{\hat{x}_{N+1|N}} = 0,$$

and, hence, $\lambda_{N+1|N} = 0$.

Similarly, differentiating J_λ with respect to u_i and setting the derivative equal to zero we obtain

$$\left.\frac{\partial J_\lambda}{\partial u_i}\right|_{\hat{u}_{i|N}} = \hat{u}_{i|N}^* Q_i^{-1} - \lambda_{i+1|N}^* G_i = 0, \quad \text{for } 0 \le i \le N,$$

where we have denoted the stationary solution by $\hat{u}_{i|N}$. This leads to the equality

$$\hat{u}_{i|N} = Q_i G_i^* \lambda_{i+1|N}. \tag{10.6.4}$$

Using (10.6.1) we get our first relation (*cf.* (10.5.2)),

$$\hat{x}_{i+1|N} = F_i \hat{x}_{i|N} + G_i Q_i G_i^* \lambda_{i+1|N}. \tag{10.6.5}$$

Combining (10.6.3) and (10.6.5), we are again led to the Hamiltonian equations (10.5.3) (but now for purely deterministic quantities):

$$\begin{bmatrix} \hat{x}_{i+1|N} \\ \lambda_{i|N} \end{bmatrix} = \begin{bmatrix} F_i & G_i Q_i G_i^* \\ -H_i^* R_i^{-1} H_i & F_i^* \end{bmatrix} \begin{bmatrix} \hat{x}_{i|N} \\ \lambda_{i+1|N} \end{bmatrix} + \begin{bmatrix} 0 \\ H_i^* R_i^{-1} \end{bmatrix} y_i,$$

with

$$\hat{x}_{0|N} = \Pi_0 \lambda_{0|N}, \qquad \lambda_{N+1|N} = 0. \tag{10.6.6}$$

Solving these equations by say the sweep methods noted in Sec. 10.5, we can determine the quantities $\{\hat{x}_{i|N}, \lambda_{i|N}, i = 0, 1, \ldots, N\}$ and then via (10.6.4) and (10.6.6) the desired solution $\{\hat{x}_{0|N}, \hat{u}_{i|N}, i = 0, 1, \ldots, N\}$. Of course, this is only a stationary solution. But the fact that the cost function in (10.6.2) is convex in the unknowns, along with the linear state-space constraints (10.6.1), can be used to verify that the solution $\{\hat{x}_{0|N}, \hat{u}_{i|N}\}$ actually yields a minimum.

Remark 2 [Extended Hamiltonian Equations]. The reader may refer to Prob. 10.20, which shows a similar derivation of the expanded Hamiltonian equations (10.5.6) by using two Lagrange multipliers. ♦

10.7 APPLICATIONS OF EQUIVALENCE

In this section we solve the deterministic minimization problem (10.6.2) by setting up an equivalent stochastic problem (as described in Sec. 3.5), which is readily solved using the innovations method. In fact, this approach will also lead to an alternative set of recursions for determining the desired variables $\{\hat{x}_{0|N}, \hat{u}_{j|N}\}$; it will also allow us to explain how the cost function (10.6.2) can be used to provide a deterministic interpretation for the Kalman filter.

10.7.1 The Equivalent Stochastic Problem

To solve (10.6.2) by equivalence, we shall first have to set up an equivalent stochastic problem.

For this purpose, we introduce the column vectors

$$z = \text{col}\{x_0, u_0, \ldots, u_N\}, \qquad y = \text{col}\{y_0, y_1, \ldots, y_N\},$$

and define the block-diagonal matrices

$$R_v \triangleq (R_0 \oplus R_1 \oplus \ldots \oplus R_N), \qquad R_z \triangleq (\Pi_0 \oplus Q_0 \oplus \ldots \oplus Q_N). \tag{10.7.1}$$

Then using the state equation (10.6.1) to express each term $H_i x_i$ in terms of combinations of the entries of z, we can rewrite the original cost function (10.6.2) as

$$\min_z \left[z^* R_z^{-1} z + (y - Az)^* R_v^{-1} (y - Az) \right], \tag{10.7.2}$$

where

$$A \triangleq \begin{bmatrix} H_0 & & & & \\ H_1\Phi(1,0) & H_1 G_0 & & & \\ H_2\Phi(2,0) & H_2\Phi(2,1)G_0 & H_2 G_1 & & \\ \vdots & & & \ddots & \\ H_N\Phi(N,0) & H_N\Phi(N,1)G_0 & H_N\Phi(N,2)G_1 & \ldots & H_N G_{N-1} \quad 0 \end{bmatrix}, \tag{10.7.3}$$

and

$$\Phi(i,j) = F_{i-1} F_{i-2} \ldots F_j, \quad i > j, \quad \Phi(i,i) = I.$$

Let $\hat{z}_{|N}$ denote the solution to the minimization problem (10.7.2), *i.e.*, $\hat{z}_{|N}$ is a column vector that contains the desired entries $\{\hat{x}_{0|N}, \hat{u}_{0|N}, \ldots, \hat{u}_{N|N}\}$. Now in view of the equivalence result of Sec. 3.5, we know that $\hat{z}_{|N}$ can be obtained by solving an equivalent stochastic estimation problem that is determined as follows.

We introduce random vectors $\{\mathbf{z}, \mathbf{y}\}$, with the same dimensions and partitionings as the above $\{z, y\}$, and assume that they are related via a linear model of the form

$$\mathbf{y} = A\mathbf{z} + \mathbf{v}, \tag{10.7.4}$$

where A is the same matrix as in the cost (10.7.2), and where $\mathbf{v}$ denotes an additive noise vector that is partitioned as

$$\mathbf{v} = \text{col}\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_N\}.$$

The dimensions of the $\{\mathbf{v}_i\}$ are compatible with those of $\{\mathbf{y}_i\}$. The variables $\{\mathbf{z}, \mathbf{v}\}$ are further assumed to be zero mean and such that

$$\langle \mathbf{z}, \mathbf{z} \rangle = R_z, \qquad \langle \mathbf{v}, \mathbf{v} \rangle = R_v, \qquad \langle \mathbf{z}, \mathbf{v} \rangle = 0. \tag{10.7.5}$$

Let $\hat{\mathbf{z}}_{|N}$ denote the l.l.m.s. estimator of $\mathbf{z}$ given the entries $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ in $\mathbf{y}$. We further partition $\mathbf{z}$ as $\mathbf{z} = \text{col}\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_N\}$.

Then the equivalence result of Sec. 3.5 states that the expression for $\hat{\mathbf{z}}_{|N}$ in terms of $\mathbf{y}$ in the stochastic problem (10.7.4) is identical to the expression for $\hat{z}_{|N}$ in terms of y in the deterministic problem (10.7.2).

10.7.2 Solving the Stochastic Problem

The linear model (10.7.4), coupled with the definitions of $\{R_z, R_v, A\}$ in (10.7.1), (10.7.3), and (10.7.5), show that the variables $\{\mathbf{y}_i, \mathbf{v}_i, \mathbf{x}_0, \mathbf{u}_i\}$ can be related via the following standard state-space model

$$\mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i, \qquad \mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i, \tag{10.7.6}$$

with

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & 0 & 0 \\ 0 & R_i \delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{10.7.7}$$

Therefore, the stochastic solution is readily obtained from the BF recursions for $\{\hat{\mathbf{x}}_{0|N}, \hat{\mathbf{u}}_{j|N}\}$ from Sec. 10.2. More specifically, using Thm. 10.2.1 and Eq. (10.2.14) we can write

$$\begin{cases} \hat{\mathbf{x}}_{0|N} = \Pi_0 \boldsymbol{\lambda}_{0|N}, \\ \hat{\mathbf{x}}_{j+1} = F_{p,j} \hat{\mathbf{x}}_j + K_{p,j} \mathbf{y}_j, \quad \hat{\mathbf{x}}_0 = 0, \\ \mathbf{e}_j = -H_j \hat{\mathbf{x}}_j + \mathbf{y}_j, \\ \hat{\mathbf{u}}_{j|N} = Q_j G_j^* \boldsymbol{\lambda}_{j+1|N}, \\ \boldsymbol{\lambda}_{j|N} = F_{p,j}^* \boldsymbol{\lambda}_{j+1|N} + H_i^* R_{e,j}^{-1} \mathbf{e}_j, \quad \boldsymbol{\lambda}_{N+1|N} = 0. \end{cases} \tag{10.7.8}$$

10.7.3 Solving the Deterministic Problem

We know by equivalence that the mappings from $\{\mathbf{y}_j\}$ to $\{\hat{\mathbf{x}}_{0|N}, \hat{\mathbf{u}}_{j|N}\}$ in the stochastic problem (10.7.4) coincide with the mappings from $\{y_j\}$ to $\{\hat{x}_{0|N}, \hat{u}_{j|N}\}$ in the deterministic problem (10.7.2). We are thus led to the following statement.

Theorem 10.7.1 (Solution of Deterministic Problem) *The solution of (10.6.2) can be recursively computed as follows:*

$$\begin{cases} \hat{x}_{0|N} = \Pi_0 \lambda_{0|N}, \\ \hat{x}_{i+1} = F_{p,i} \hat{x}_i + K_{p,i} y_i, \quad \hat{x}_0 = 0, \\ e_i = -H_i \hat{x}_i + y_i, \\ \hat{u}_{i|N} = Q_i G_i^* \lambda_{i+1|N}, \\ \lambda_{i|N} = F_{p,i}^* \lambda_{i+1|N} + H_i^* R_{e,i}^{-1} e_i, \quad \lambda_{N+1|N} = 0, \end{cases} \tag{10.7.9}$$

where the quantities $\{K_{p,i}, R_{e,i}\}$ are obtained from the Kalman-type filter recursions

$$\begin{cases} K_{p,i} = F_i P_i H_i^* R_{e,i}^{-1}, \quad R_{e,i} = R_i + H_i P_i H_i^*, \\ P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \quad P_0 = \Pi_0. \end{cases} \tag{10.7.10}$$

Moreover, the minimum value of (10.6.2) is equal to $\sum_{i=0}^{N} e_i^* R_{e,i}^{-1} e_i$. ■

Proof: The only result that remains to be proved is the expression for the minimum value of the cost (10.6.2). For this purpose, recall from Sec. 3.5 that the minimum value of (10.7.2) is given by

$$\text{minimum cost} = y^* [R_v + A R_z A^*]^{-1} y.$$

The center matrix $R_v + A R_z A^*$ is readily seen from (10.7.4) to coincide with the covariance matrix R_y of $\mathbf{y}$. Hence, the above minimum cost is also equal to $y^* R_y^{-1} y$. Now the quantities $\{e_i, y_i\}$ in the above statement are related through $y = Le$, where $y = \text{col}\{y_0, \ldots, y_N\}$, $e = \text{col}\{e_0, \ldots, e_N\}$, and where L is the lower triangular matrix with unit diagonal entries satisfying $R_y = L R_e L^*$, with $R_e = \text{diag}\{R_{e,0}, \ldots, R_{e,N}\}$ (*cf.* Sec. 9.4). We thus conclude that

$$y^* R_y^{-1} y = y^* L^{-*} R_e^{-1} L^{-1} y = e^* R_e^{-1} e,$$

as desired. ◆

10.7.4 An Alternative Direct Solution

One can also solve the stochastic problem (10.7.4) directly without assuming knowledge of the BF recursions as follows.

Thus let $\hat{\mathbf{z}}_{|i}$ denote the l.l.m.s. estimator of $\mathbf{z}$ given the top entries $\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}$ in $\mathbf{y}$. To determine $\hat{\mathbf{z}}_{|i}$, and ultimately $\hat{\mathbf{z}}_{|N}$, we can proceed recursively by employing the innovations $\{\mathbf{e}_i\}$ of the observations $\{\mathbf{y}_i\}$. We thus write

$$\begin{aligned} \hat{\mathbf{z}}_{|i} &= \hat{\mathbf{z}}_{|i-1} + \langle \mathbf{z}, \mathbf{e}_i \rangle \, R_{e,i}^{-1} \, \mathbf{e}_i, \\ &= \hat{\mathbf{z}}_{|i-1} + \underbrace{\langle \mathbf{z}, \tilde{\mathbf{x}}_i \rangle}_{K_{z,i}} H_i^* R_{e,i}^{-1} \, \mathbf{e}_i, \quad \hat{\mathbf{z}}_{|-1} = 0, \end{aligned} \tag{10.7.11}$$

where we have used in the last equality the innovations equation $\mathbf{e}_i = H_i \tilde{\mathbf{x}}_i + \mathbf{v}_i$, and the fact that $< \mathbf{x}_0, \mathbf{v}_i >= 0$ and $< \mathbf{u}_j, \mathbf{v}_i >= 0$ for all j. We have also introduced a new

gain matrix defined by $K_{z,i} = < \mathbf{z}, \tilde{\mathbf{x}}_i >$. It is easy to see that the entries of $\hat{\mathbf{z}}_{|i}$ have the following structure:

$$\hat{\mathbf{z}}_{|i} = \text{col}\{\hat{\mathbf{x}}_{0|i}, \hat{\mathbf{u}}_{0|i}, \hat{\mathbf{u}}_{1|i}, \ldots, \hat{\mathbf{u}}_{i-1|i}, 0, 0, \ldots, 0\}.$$

That is, the trailing entries of $\hat{\mathbf{z}}_{|i}$ are zero since $\hat{\mathbf{u}}_{j|i} = 0$ for $j \geq i$.

The above recursive construction would be complete, and hence provide the desired quantity $\hat{\mathbf{z}}_{|N}$, once we show how to evaluate the gain matrix $K_{z,i}$. For this, recall from Eq. (9.2.23) that

$$\tilde{\mathbf{x}}_{i+1} = F_{p,i}\tilde{\mathbf{x}}_i + G_i\mathbf{u}_i - K_{p,i}\mathbf{v}_i,$$

where $F_{p,i} = F_i - K_{p,i}H_i$. Using this recursion, one can easily verify that

$$K_{z,i+1} \triangleq < \mathbf{z}, \tilde{\mathbf{x}}_{i+1} >= K_{z,i}\left[F_i - K_{p,i}H_i\right]^* + \begin{bmatrix}0\\ \hline 0\\ I\\ 0\end{bmatrix} Q_i G_i^*, \tag{10.7.12}$$

with

$$K_{z,0} = \begin{bmatrix}\Pi_0\\ 0\end{bmatrix}. \tag{10.7.13}$$

The identity matrix that appears in the second term of the recursion for $K_{z,i+1}$ occurs at the position that corresponds to the entry $\mathbf{u}_i$ in the vector $\mathbf{z}$. Substituting into (10.7.11) we find that the following recursions hold:

$$\begin{cases}\hat{\mathbf{x}}_{0|i} = \hat{\mathbf{x}}_{0|i-1} + \Pi_0\Phi_p^*(i,0)H_i^*R_{e,i}^{-1}\mathbf{e}_i, & \hat{\mathbf{x}}_{0|-1} = 0,\\ \hat{\mathbf{u}}_{j|i} = \hat{\mathbf{u}}_{j|i-1} + Q_jG_j^*\Phi_p^*(i,j+1)H_i^*R_{e,i}^{-1}\mathbf{e}_i, & j < i,\\ \hat{\mathbf{u}}_{j|i} = 0, \quad j \geq i,\end{cases} \tag{10.7.14}$$

where

$$\Phi_p(i,j) \triangleq \begin{cases}F_{p,i-1}F_{p,i-2}\cdots F_{p,j} & i > j,\\ I & i = j,\end{cases}$$

and $F_{p,i} = F_i - K_{p,i}H_i$.

The above recursions are seen to provide the estimators $\{\hat{\mathbf{x}}_{0|i}, \hat{\mathbf{u}}_{j|i}\}$ for successive values of i, and not only for $i = N$ as in the BF recursions (10.7.8). [In Prob. 10.8 we show how the BF recursions for $\{\hat{\mathbf{x}}_{j|N}, \boldsymbol{\lambda}_{j|N}\}$ follow from those in (10.7.14).] A bonus is that the estimators $\{\hat{\mathbf{x}}_{0|i}, \hat{\mathbf{u}}_{j|i}\}$ can be related to solutions of nested optimization problems. Indeed, by equivalence, the expressions that provide the solutions $\{\hat{\mathbf{x}}_{0|i}, \hat{\mathbf{u}}_{j|i}\}$ in (10.7.14) should coincide with those that provide the solutions $\{\hat{x}_{0|i}, \hat{u}_{j|i}\}$ for the following deterministic problem, with data up to time i (rather than N as in (10.6.2)):

$$\min_{x_0,u_0,\ldots,u_i}\left[x_0^*\Pi_0^{-1}x_0 + \sum_{j=0}^{i}(y_j - H_jx_j)^*R_j^{-1}(y_j - H_jx_j) + \sum_{j=0}^{i}u_j^*Q_j^{-1}u_j\right]. \tag{10.7.15}$$

We thus conclude that the following recursions also hold for each $i \geq 0$:

$$\begin{cases}\hat{x}_{0|i} = \hat{x}_{0|i-1} + \Pi_0\Phi_p^*(i,0)H_i^*R_{e,i}^{-1}e_i, & \hat{x}_{0|-1} = 0,\\ \hat{u}_{j|i} = \hat{u}_{j|i-1} + Q_jG_j^*\Phi_p^*(i,j+1)H_i^*R_{e,i}^{-1}e_i, & j < i,\\ \hat{u}_{j|i} = 0, \quad j \geq i.\end{cases} \tag{10.7.16}$$

10.7.5 MAP Estimation and a Deterministic Interpretation for the Kalman Filter

We mentioned earlier that the cost function (10.6.2) can be motivated in terms of maximum a posteriori (MAP) estimation for Gaussian random variables. We expand briefly on this remark here and show also how the deterministic quadratic form (10.6.2) arises in the context of Kalman filtering.

Thus assume we have a state-space model of the form (10.7.6)–(10.7.7), which in turn induces the linear relation (10.7.4). We shall further assume that the variables $\{\mathbf{z}, \mathbf{v}\}$ are circular Gaussian and independent.

We now pose the problem of estimating the state variables $\{\mathbf{x}_0, \ldots, \mathbf{x}_{N+1}\}$ given the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$. That is, we wish to determine the (smoothed) estimators $\{\hat{\mathbf{x}}_{0|N}, \hat{\mathbf{x}}_{1|N}, \ldots, \hat{\mathbf{x}}_{N+1|N}\}$. We mentioned earlier, following Eq. (10.1.7), that the smoothed estimators $\{\hat{\mathbf{x}}_{i|N}\}$ can be determined by knowledge of the prediction estimators $\{\hat{\mathbf{x}}_i\}$, so that the solution of the smoothing problem is straightforward once the predicted estimators are determined via the Kalman filter. We shall encounter the same situation here.

First note that in view of the state equation (10.7.6), which can be used to express each $\mathbf{x}_i$ as a linear combination of the variables $\{\mathbf{x}_0, \mathbf{u}_j, j < i\}$, we can equivalently consider the problem of estimating $\{\mathbf{x}_0, \mathbf{u}_0, \ldots, \mathbf{u}_N\}$ from the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ — see Prob. 10.9. In other words, we can consider the problem of estimating the variable $\mathbf{z}$ from the vector $\mathbf{y}$ in the linear model (10.7.4).

The maximum a posteriori (MAP) approach to the estimation of $\mathbf{z}$ seeks an estimator that maximizes the conditional probability density function $f_{\mathbf{z}|\mathbf{y}}(z|y)$. Using Bayes' rule we can write

$$f_{\mathbf{z}|\mathbf{y}}(z|y) = \frac{f_{\mathbf{y}|\mathbf{z}}(y|z)f_{\mathbf{z}}(z)}{f_{\mathbf{y}}(y)}. \tag{10.7.17}$$

From the equation $\mathbf{y} = A\mathbf{z} + \mathbf{v}$ it follows that

$$f_{\mathbf{y}|\mathbf{z}}(y|z) = f_{\mathbf{v}}(y - Az) \propto \exp^{-(y-Az)^*R_v^{-1}(y-Az)},$$

where the symbol $\propto$ stands for "proportional to". Likewise, by assumption,

$$f_{\mathbf{z}}(z) \propto \exp^{-(z^*R_z^{-1}z)}.$$

The denominator in (10.7.17) is independent of $\mathbf{z}$. We therefore see that maximizing $f_{\mathbf{z}|\mathbf{y}}(z|y)$ over z is equivalent to minimizing the following cost function over z,

$$\min_z\left[z^*R_z^{-1}z + (y - Az)^*R_v^{-1}(y - Az)\right],$$

which is identical to (10.7.2) and, hence, to (10.6.2). Moreover, the recursions (10.7.8) provide the desired quantities $\{\hat{\mathbf{x}}_{0|N}, \hat{\mathbf{u}}_{j|N}\}$. Note in particular how all the quantities that are required in these recursions, such as $\{K_{p,i}, R_{e,i}, \mathbf{e}_i\}$, are completely determined by the Kalman filtering recursions applied to the state-space model (10.7.6)–(10.7.7).

Remark 3. The fact that the Kalman filtering recursions can be used to minimize the cost function (10.6.2) is wellknown in the literature. However, some of the arguments used to justify this result are not as transparent as the one given above, which was based on equivalence. We explain this issue more closely in Prob. 10.9. For example, the argument in Jazwinski (1969, pp. 87–88) requires an invertible $G_i Q_i G_i^*$, which is generally not true; therefore he proposes certain signal transformations that would replace $G_i \mathbf{u}_i$ by $\bar{G}_i \bar{\mathbf{u}}_i$ such that $\bar{G}_i \bar{Q}_i \bar{G}_i^*$ is invertible. In Anderson and Moore (1979, p. 135) it is suggested that pseudo-inverses be used. The equivalence argument that we employed above avoids these artifices. ◆

10.7.6 The Deterministic Approach of Whittle

In his very interesting books, Whittle (1983,1990) also studied stochastic estimation problems via equivalence, using the expanded Hamiltonian equations (10.5.6) — see Probs. 10.20 and 10.21. Here we shall specialize his approach to the traditional Hamiltonian equations (10.6.6), so as to better compare it with our approach.

Whittle works with time-invariant models and assumes that steady-state has been reached. The first step is to rewrite the Hamiltonian equations (10.6.6) so as to make the coefficient matrix Hermitian:

$$\begin{bmatrix} \hat{x}_{i+1|N} \\ \lambda_{i|N} \end{bmatrix} = \begin{bmatrix} GQG^* & F \\ F^* & -H^*R^{-1}H \end{bmatrix} \begin{bmatrix} \lambda_{i+1|N} \\ \hat{x}_{i|N} \end{bmatrix} + \begin{bmatrix} 0 \\ H^*R^{-1} \end{bmatrix} y_i. \tag{10.7.18}$$

This is simply achieved by switching the order of the variables $\{\lambda_{i+1|N}, \hat{x}_{i|N}\}$ on the right-hand side. Now let $\{\hat{x}(z), \lambda(z), y(z)\}$ denote the z-transforms of the sequences $\{\hat{x}_{i|N}, \lambda_{i+1|N}, y_i\}$, respectively — these are the sequences that appear on the right-hand side of (10.7.18). Then, in the $z-$transform domain, we can rewrite (10.7.18) as

$$\begin{bmatrix} GQG^* & -zI + F \\ -z^{-1}I + F^* & -H^*R^{-1}H \end{bmatrix} \begin{bmatrix} \lambda(z) \\ \hat{x}(z) \end{bmatrix} + \begin{bmatrix} 0 \\ H^*R^{-1} \end{bmatrix} y(z) = 0. \tag{10.7.19}$$

We shall denote the para-Hermitian coefficient matrix that multiplies $\text{col}\{\lambda(x), \hat{x}(z)\}$ by

$$\Psi(z) \triangleq \begin{bmatrix} GQG^* & -zI + F \\ -z^{-1}I + F^* & -H^*R^{-1}H \end{bmatrix}. \tag{10.7.20}$$

Observe that $\Psi(z)$ has the simple expansion

$$\Psi(z) = \Psi_{-1}z + \Psi_0 + \Psi_1 z^{-1}, \tag{10.7.21}$$

where

$$\Psi_{-1} = \begin{bmatrix} 0 & -I \\ 0 & 0 \end{bmatrix}, \quad \Psi_0 = \begin{bmatrix} GQG^* & F \\ F^* & -H^*R^{-1}H \end{bmatrix}, \quad \Psi_1 = \begin{bmatrix} 0 & 0 \\ -I & 0 \end{bmatrix}.$$

Since Ψ_0 can be seen to be indefinite (its (1, 1) block entry is nonnegative-definite while its (2, 2) block entry is nonpositive-definite), $\Psi(z)$ is not a z-spectrum. Nevertheless, $\Psi(z)$ can be factored as

$$\Psi(z) = A(z)A_0\left[A\left(\frac{1}{z^*}\right)\right]^*, \tag{10.7.22}$$

where $A(z)$ is the *rational* matrix function

$$A(z) = \begin{bmatrix} -zP & zI - F_p \\ I & 0 \end{bmatrix}, \tag{10.7.23}$$

where once again P is the unique stabilizing solution of the DARE (recall the discussion in Sec. 8.3.4, where conditions for the existence of such a P were given)

$$P = FPF^* + GQG^* - K_p R_e K_p^*, \quad K_p = FPH^*R_e^{-1}, \quad R_e = R + HPH^*,$$

and F_p is the stable closed-loop matrix, $F_p = F - K_pH$. Furthermore, A_0 is also determined from knowledge of P,

$$A_0 = \begin{bmatrix} -H^*R^{-1}H & -(I + H^*R^{-1}HP) \\ -(I + PH^*R^{-1}H) & -P(I + H^*R^{-1}HP) \end{bmatrix}.$$

[Of course, (10.7.22) can be verified by multiplication of the given expressions. Whittle (1990, p. 149) gives a method for finding factorizations of rational matrix functions $\Psi(z)$ of the general form (10.7.21).] Note further that[4]

$$A^{-1}(z) = \begin{bmatrix} 0 & I \\ (zI - F_p)^{-1} & (I - z^{-1}F_p)^{-1}P \end{bmatrix},$$

$$A_0^{-1} = \begin{bmatrix} P & -I \\ -I & H^*R^{-1}H(I + PH^*R^{-1}H)^{-1} \end{bmatrix}.$$

Having found the factorization, we can now proceed to solve the Hamiltonian equations. Thus from (10.7.19) and (10.7.22) we have

$$\left[A\left(\frac{1}{z^*}\right)\right]^* \begin{bmatrix} \lambda(z) \\ \hat{x}(z) \end{bmatrix} = A_0^{-1}A^{-1}(z) \begin{bmatrix} 0 \\ -H^*R^{-1} \end{bmatrix} y(z).$$

Using the above expressions for $\{A_0^{-1}, A^{-1}(z)\}$ we can find, *after some algebra*, that

$$\begin{bmatrix} -z^{-1}P & I \\ z^{-1}I - F_p^* & 0 \end{bmatrix} \begin{bmatrix} \lambda(z) \\ \hat{x}(z) \end{bmatrix} = \begin{bmatrix} (zI - F_p)^{-1}K_p y(z) \\ H^*R_e^{-1}e(z) \end{bmatrix}, \tag{10.7.24}$$

[4] We may mention that $A^{-1}(z)$ is analytic in $|z| \geq 1$ since all eigenvalues of F_p are strictly inside the unit circle. Moreover, $A(z)$ is defined everywhere except at ∞. Hence, both $A(z)$ and its inverse are analytic (well-defined) in $1 \leq |z| < \infty$.

where

$$e(z) = \left[I - H(zI - F_p)^{-1}K_p\right] y(z)$$

is the z-transform of a sequence $\{e_i\}$. The second equality in (10.7.24) then gives

$$(z^{-1}I - F_p^*)\lambda(z) = H^*R_e^{-1}e(z),$$

which in the time domain corresponds to the recursion (compare with the steady-state version of the last recursion in (10.7.9))

$$\lambda_{i|N} = F_p^*\lambda_{i+1|N} + H^*R_e^{-1}e_i. \tag{10.7.25}$$

Likewise, the first equality in (10.7.24) gives

$$-z^{-1}P\lambda(z) + \hat{x}(z) = (zI - F_p)^{-1}K_p y(z),$$

which is equivalent to

$$-z^{-1}(zI - F_p)P\lambda(z) + (zI - F_p)\hat{x}(z) = K_p y(z).$$

In the time domain we obtain

$$-P\lambda_{i+1|N} + F_pP\lambda_{i|N} + \hat{x}_{i+1|N} = F_p\hat{x}_{i|N} + K_p y_i. \tag{10.7.26}$$

Finally, we shall introduce the variable (consistent with the Bryson-Frazier relation (10.2.5))

$$\hat{x}_i \triangleq \hat{x}_{i|N} - P_i\lambda_{i|N}.$$

Then (10.7.26) collapses to the steady-state Kalman filter recursion (compare with the steady-state version of the second equation in (10.7.9))

$$\hat{x}_{i+1} = F_p\hat{x}_i + K_p y_i. \tag{10.7.27}$$

In this way, we again obtain the basic filtering and smoothing recursions. This is certainly an interesting approach, which has some links to the methods described in Sec. 10.5.

We should say, however, that to us it seems considerably easier to solve stochastic problems directly, using the geometric recursive projection/innovations approach, rather than to go to an equivalent deterministic problem. In fact, as in the earlier subsections, it seems preferable to us rather to go in the other direction, and to solve deterministic problems by reduction to equivalent (or dual, *cf.* Ch. 15) stochastic problems. We have successfully applied this point of view to the solution of adaptive filtering problems in Sayed and Kailath (1994b) and to the solution of $\mathcal{H}_\infty$ problems in Hassibi, Sayed, and Kailath (1999).

However, a feature stressed by Whittle (1990, Ch. 12) is that this approach allows us to more readily study models given in higher-order (AR or ARMA) forms. The point is that while we could convert these to state-space form, the resulting forms have special structure (*cf.* Sec. 5.3) that is not fully exploited by our state-space recursions. These aspects certainly deserve closer examination.

10.8 COMPLEMENTS

The fact that smoothing problems were left unsolved in the original papers of Kalman (1960a,1963b) and Kalman and Bucy (1961) led to a flurry of efforts, beyond those already noted in the text. A nice survey of the literature up to 1973 was given by Meditch (1973) with a supplement by Kailath (1975). However, there has been a continuing interest, with approaches based on stochastic realization theory (*e.g.,* Faurre, Clerget, and Germain (1979) and Badawi, Lindquist, and Pavon (1979)), on the notion of complementary models (Weinert and Desai (1981), Desai, Weinert, and Yusypchuk (1983), Bello et al. (1986), Ackner and Kailath (1989a,1989b); see also Sec. 15.7.6), on scattering theory (Ljung and Kailath (1976a) and also Sec. 17.4.4), triangularization and diagonalization of the Hamiltonian equations (Kailath and Ljung (1982); also Sec. 10.5), and others.

As noted earlier, it is ironic that smoothing problems are easier to address in the Wiener theory (*cf.* Secs. 7.3.1 and 7.3.2). One reason is that recursive implementations of the Wiener smoother were not pursued. Doing so would have required not only the introduction of state-space models (see Sec. 8.4), but also the need for properly defining backwards-time Markov models. The early confusion in this regard was reflected in some of the interpretations and derivations of the Fraser-Potter and Mayne formulas of Sec. 10.4; the early literature is examined in detail in Wall, Willsky, and Sandell (1981).

As described in the notes on Ch. 5, clear insight into these issues (and several new results) first came from the scattering approach to estimation theory (see Ch. 17). This approach begins with the Hamiltonian equations (10.5.3), but now derived from first principles using the structure of the linear space generated by the fundamental random variables of the standard state-space model.

Finally, we should also mention that there have been several detailed studies of the properties and implementation of the fixed-lag smoother, for potential use in communications problems (see Anderson and Moore (1979) and the references therein).

■ PROBLEMS

10.1 **(Fixed-point smoothing)** Consider the model (10.2.1)–(10.2.2). Fix a time instant i_o and let N increase. Let $\hat{\mathbf{x}}_{i_o|N}$ denote the l.l.m.s. estimator of $\mathbf{x}_{i_0}$ given the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$, and denote the error covariance matrix at time i_o by $P_{i_o|N}$. Likewise, let $\hat{\mathbf{x}}_{i_o|N+1}$ denote the l.l.m.s. estimator of $\mathbf{x}_{i_0}$ given $\{\mathbf{y}_0, \ldots, \mathbf{y}_N, \mathbf{y}_{N+1}\}$ and let $P_{i_o|N+1}$ denote the corresponding error covariance matrix. Show that

$$\hat{\mathbf{x}}_{i_o|N+1} = \hat{\mathbf{x}}_{i_o|N} + P_{i_o}\Phi_p^*(N+1, i_o)H_{N+1}^*R_{e,N+1}^{-1}\mathbf{e}_{N+1},$$

$$P_{i_o|N+1} = P_{i_o|N} - P_{i_o}\Phi_p^*(N+1, i_o)H_{N+1}^*R_{e,N+1}^{-1}H_{N+1}\Phi_p(N+1, i_o)P_{i_o}.$$

Remark. Recall that we solved the fixed-point smoothing problem earlier in Prob. 9.18 by using a method due to Zachrisson (1969) and Willman (1969). ◆

10.2 **(Fixed-lag smoothing)** Consider again the state-space model (10.2.1)–(10.2.2) and assume $S_i = 0$ for simplicity. Now choose a positive integer L and let i increase. Let $\hat{\mathbf{x}}_{i|i+L}$ denote the l.l.m.s. estimator of $\mathbf{x}_i$ given the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i+L}\}$. We want to determine a recursion relating $\hat{\mathbf{x}}_{i|i+L}$ and $\hat{\mathbf{x}}_{i-1|i-1+L}$.

(a) Using (10.1.6), the formula for $P_{i,j}$ in Lemma 10.2.1, and the innovations $\{\mathbf{e}_j\}_{j=0}^{i+L}$ show that

$$\hat{\mathbf{x}}_{i|i+L} = \hat{\mathbf{x}}_{i|i+L-1} + P_i\Phi_p^*(i+L,i)H_{i+L}^* R_{e,i+L}^{-1}\mathbf{e}_{i+L}.$$

(b) Using the state equation $\mathbf{x}_i = F_{i-1}\mathbf{x}_{i-1} + G_{i-1}\mathbf{u}_{i-1}$, the relation (10.2.13), and the innovations $\{\mathbf{e}_j\}_{j=0}^{i+L}$ show also that

$$\hat{\mathbf{x}}_{i|i+L-1} = F_{i-1}\hat{\mathbf{x}}_{i-1|i-1+L} + G_{i-1}Q_{i-1}G_{i-1}^*\boldsymbol{\lambda}_{i|i+L-1},$$

where we defined

$$\boldsymbol{\lambda}_{i|i+L-1} \triangleq \sum_{j=i}^{i+L-1}\Phi_p^*(j,i)H_j^*R_{e,j}^{-1}\mathbf{e}_j.$$

(c) Using the above definition of $\boldsymbol{\lambda}_{i|i+L-1}$ and the relation $\Phi_p^*(j,i) = F_{p,i}^*\Phi_p^*(j,i+1)$, conclude that

$$\hat{\mathbf{x}}_{i|i+L} = F_{i-1}\hat{\mathbf{x}}_{i-1|i-1+L} + G_{i-1}Q_{i-1}G_{i-1}^*\boldsymbol{\lambda}_{i|i+L-1} + P_i\Phi_p^*(i+L,i)H_{i+L}^*R_{e,i+L}^{-1}\mathbf{e}_{i+L},$$

$$\boldsymbol{\lambda}_{i|i+L-1} = F_{p,i}^*\boldsymbol{\lambda}_{i+1|i+L} + H_i^*R_{e,i}^{-1}\mathbf{e}_i - \Phi_p^*(i+L,i)H_{i+L}^*R_{e,i+L}^{-1}\mathbf{e}_{i+L}.$$

10.3 **($\hat{\mathbf{u}}_{i|N}$ when $S_i \neq 0$)** Refer to Sec. 10.2.2 but now assume $S_i \neq 0$. That is, $\langle \mathbf{u}_j, \mathbf{v}_i\rangle = S_i\delta_{ij}$.

(a) Verify that $\hat{\mathbf{u}}_{i|N} = \sum_{j=i}^N \langle \mathbf{u}_i, \mathbf{e}_j\rangle R_{e,j}^{-1}\mathbf{e}_j$.

(b) Show that for $j > i$, $\langle \tilde{\mathbf{x}}_j, \mathbf{u}_i\rangle = \Phi_p(j,i+1)[G_iQ_i - K_{p,i}S_i^*]$.

(c) Verify also that

$$\langle \mathbf{e}_j, \mathbf{u}_i\rangle = \begin{cases} H_j\Phi_p(j,i+1)[G_iQ_i - K_{p,i}S_i^*] & j > i \\ S_i^* & j = i. \end{cases}$$

(d) Conclude that $\hat{\mathbf{u}}_{i|N} = S_iR_{e,i}^{-1}\mathbf{e}_i + [Q_iG_i^* - S_iK_{p,i}^*]\boldsymbol{\lambda}_{i+1|N}$.

10.4 **(Orthogonality of $\mathbf{u}_i^r$ and $\tilde{\mathbf{x}}_{i+1|N}$)** Refer to Eq. (10.3.8) and the definition of $\mathbf{u}_i^r$ (recall that we assumed $S_i = 0$).

(a) Use (10.1.5), (10.2.12), and (10.2.13) to verify that

$$\langle \tilde{\mathbf{x}}_{i+1|N}, \mathbf{u}_i\rangle = G_iQ_i - \sum_{j=i+1}^N P_{i+1,j}H_j^*R_{e,j}^{-1}H_j\Phi_p(j,i+1)G_iQ_i.$$

(b) Show that for $j \geq i+1$, $\langle \mathbf{e}_j, \tilde{\mathbf{x}}_{i+1}\rangle = H_j\Phi_p(j,i+1)P_{i+1}$.

(c) Use (10.1.5) and part (b) to verify that

$$\langle \tilde{\mathbf{x}}_{i+1|N}, \tilde{\mathbf{x}}_{i+1}\rangle = P_{i+1} - \sum_{j=i+1}^N P_{i+1,j}H_j^*R_{e,j}^{-1}H_j\Phi_p(j,i+1)P_{i+1}.$$

(d) Use the results of parts (a) and (c), and the definition of $\mathbf{u}_i^r$, to conclude that $\langle \tilde{\mathbf{x}}_{i+1|N}, \mathbf{u}_i^r\rangle = 0$.

10.5 **(Whiteness of $\mathbf{u}_i^r$)** Refer again to the definition of $\mathbf{u}_i^r$ in (10.3.9), where we assumed F_i invertible, $S_i = 0$, and $P_i > 0$ for all i. Recall also that $\mathbf{u}_i$ itself is a white sequence with covariance matrix Q_i.

(a) Verify the following relations

$$\langle \tilde{\mathbf{x}}_{j+1}, \mathbf{u}_i\rangle = \begin{cases} G_iQ_i & j = i \\ 0 & j < i \\ \Phi_p(j+1,i+1)G_iQ_i & j > i. \end{cases}$$

$$\langle \tilde{\mathbf{x}}_{j+1}, \tilde{\mathbf{x}}_{i+1}\rangle = \begin{cases} P_{i+1} & j = i \\ P_{j+1}\Phi_p^*(i+1,j+1) & j < i \\ \Phi_p(j+1,i+1)P_{i+1} & j > i. \end{cases}$$

(b) Use part (a) to conclude that $\mathbf{u}_i^r$ is a white sequence, $\langle \mathbf{u}_j^r, \mathbf{u}_i^r\rangle = Q_i^r\delta_{ij} = [Q_i - Q_iG_i^*P_{i+1}^{-1}G_iQ_i]\delta_{ij}$.

10.6 **(Alternative recursion for $\boldsymbol{\lambda}_{i|N}$ when $R_i > 0$)** Assume $R_i > 0$, $S_i = 0$, and $P_i > 0$ in the Kalman recursions of Thms. 9.2.1 and 9.5.1.

(a) Use (10.2.10) to show that $P_{i|i}^{-1}\hat{\mathbf{x}}_{i|N} = P_{i|i}^{-1}\hat{\mathbf{x}}_{i|i} + F_i^*\boldsymbol{\lambda}_{i+1|N}$.

(b) Then use the measurement update formulas to show that

$$\boldsymbol{\lambda}_{i|N} = F_i^*\boldsymbol{\lambda}_{i+1|N} - H_i^*R_i^{-1}H_i\hat{\mathbf{x}}_{i|N} + H_i^*R_i^{-1}\mathbf{y}_i.$$

10.7 **(Alternative proof of (10.5.2))** Use (10.2.5) and (10.2.7) to verify that

$$\hat{\mathbf{x}}_{i+1|N} = \hat{\mathbf{x}}_{i+1} + P_{i+1}\boldsymbol{\lambda}_{i+1|N} = F_i\hat{\mathbf{x}}_i + K_{p,i}\mathbf{e}_i + P_{i+1}\boldsymbol{\lambda}_{i+1|N},$$

and

$$F_i\hat{\mathbf{x}}_{i|N} = F_i\hat{\mathbf{x}}_i + F_iP_i\boldsymbol{\lambda}_{i|N} = F_i\hat{\mathbf{x}}_i + F_iP_iF_{p,i}^*\boldsymbol{\lambda}_{i+1|N} + F_iP_iH_i^*R_{e,i}^{-1}\mathbf{e}_i.$$

Conclude that

$$\hat{\mathbf{x}}_{i+1|N} = F_i\hat{\mathbf{x}}_{i|N} + G_iQ_iG_i^*\boldsymbol{\lambda}_{i+1|N}.$$

10.8 **(BF recursions by equivalence)** Consider the model (10.7.6)–(10.7.7) and the recursions (10.7.14) for $\{\hat{\mathbf{x}}_{0|i}, \hat{\mathbf{u}}_{j|i}\}$.

(a) Verify that

$$\hat{\mathbf{x}}_{0|N} = \hat{\mathbf{x}}_{0|i-1} + \sum_{j=i}^N \Pi_0\Phi_p^*(j,0)H_j^*R_{e,j}^{-1}\mathbf{e}_j,$$

$$\hat{\mathbf{u}}_{j|N} = \hat{\mathbf{u}}_{j|i-1} + \sum_{k=i}^N Q_jG_j^*\Phi_p^*(k,j+1)H_k^*R_{e,k}^{-1}\mathbf{e}_k.$$

(b) Use the state-space model (10.7.6) to show that

$$\hat{\mathbf{x}}_{i|N} = \Phi(i,0)\hat{\mathbf{x}}_{0|N} + \sum_{j=0}^{i-1} \Phi(i,j+1)G_j\hat{\mathbf{u}}_{j|N},$$

where $\Phi(i,j) = F_{i-1}F_{i-2}\cdots F_j$ for $i > j$ and $\Phi(i,i) = I$. Find a similar expression for $\hat{\mathbf{x}}_i$ in terms of $\{\hat{\mathbf{x}}_{0|i-1}, \hat{\mathbf{u}}_{j|i-1}\}$.

(c) Introduce the quantity

$$\boldsymbol{\lambda}_{i|N} \triangleq \sum_{j=i}^{N} \Phi_p^*(j,i)H_j^* R_{e,j}^{-1}\mathbf{e}_j.$$

Now substitute the expression for $\hat{\mathbf{x}}_{0|N}$ from part (a) into the expression for $\hat{\mathbf{x}}_{i|N}$ in part (b), and use the transition property $\Phi_p(i,j) = \Phi_p(i,k)\Phi_p(k,j)$, for all $i \geq k \geq j$, to show that $\hat{\mathbf{x}}_{i|N}$ satisfies

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i + \left[\Phi(i,0)\Pi_0\Phi_p^*(i,0) + \sum_{j=0}^{i-1}\Phi(i,j+1)G_jQ_jG_j^*\Phi_p^*(i,j+1)\right]\boldsymbol{\lambda}_{i|N}.$$

(d) Show that the quantity between brackets in the above expression for $\hat{\mathbf{x}}_{i|N}$ is equal to P_i.

10.9 (Deterministic interpretation) Consider the model (10.7.6)–(10.7.7) and assume that $\Pi_0 > 0$, $R_i > 0$, $Q_i > 0$, and $G_i = I$ (so that the product $G_iQ_iG_i^*$ is invertible). Assume further that all random variables $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i\}$ are circular Gaussian.

(a) Define the aggregate vector $\mathbf{x} = \{\mathbf{x}_0, \ldots, \mathbf{x}_{N+1}\}$ and consider the problem of estimating $\mathbf{x}$ from the observations $\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ using the maximum a posteriori (MAP) approach. Show that the conditional probability density function of $\mathbf{y}$ given $\mathbf{x}$ is proportional to

$$f_{\mathbf{y}|\mathbf{x}}(y|x) \propto \exp^{-\left(\sum_{i=0}^{N}(y_i - H_ix_i)^*R_i^{-1}(y_i-H_ix_i)\right)}.$$

(b) Show further that the probability density function of $\mathbf{x}$ is proportional to

$$f_{\mathbf{x}}(x) = \propto \exp^{\left(-x_0^*\Pi_0^{-1}x_0\right)}\exp^{-\left(\sum_{i=0}^{N}(x_{i+1}-F_ix_i)^*Q_i^{-1}(x_{i+1}-F_ix_i)\right)}.$$

(c) Argue that maximizing $f_{\mathbf{x}|\mathbf{y}}(x|y)$ over x is equivalent to minimizing the following cost over $\{x_0, x_1, \ldots, x_{N+1}\}$:

$$x_0^*\Pi_0^{-1}x_0 + \sum_{i=0}^{N}(y_i - H_ix_i)^*R_i^{-1}(y_i - H_ix_i) + \sum_{i=0}^{N}u_i^*Q_i^{-1}u_i.$$

Remark. If $G_i \neq I$, then in part (b) above the matrix Q_i^{-1} would need to be replaced by $(G_iQ_iG_i^*)^{-1}$. However, the product $G_iQ_iG_i^*$ is in general rank deficient and, hence, not invertible. This poses a difficulty in justifying the deterministic cost function (10.6.2), which appears in part (c) as well. As mentioned in the remark at the end of Sec. 10.7, although this issue has been addressed in different ways in the literature, the equivalence argument that we employed in that section is more transparent and avoids these difficulties. ♦

10.10 (Fixed-interval smoothing via backwards innovations) Consider the standard state-space model (10.2.1)–(10.2.2), with $S_i = 0$, and define $\{\mathbf{x}_i^b, \mathbf{x}_{i|i}^b, \mathbf{e}_i^b\}$ as in Sec. 9.8. By expressing the smoothed estimator $\hat{\mathbf{x}}_{i|N}$ in terms of the backwards innovations $\{\mathbf{e}_i^b\}$, show that

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i^b + \sum_{j=0}^{i} P_{i,b}^b H_j^* R_{e,j}^{-b}\mathbf{e}_j^b = \hat{\mathbf{x}}_{i|i}^b + \sum_{j=0}^{i-1} P_{i,j}^b H_j^* R_{e,j}^{-b}\mathbf{e}_j^b,$$

where, for $j < i$, $P_{i,j}^b \triangleq \langle\tilde{\mathbf{x}}_i^b, \tilde{\mathbf{x}}_j^b\rangle = P_i^b\Phi_i^{b*}(j,i)$, $\Phi_i^b(j,i) = F_{i,j+1}^b\cdots F_{i,i}^b$, and $F_{i,i}^b = F_i^b - K_{i,i}^bH_i$ is the closed-loop state matrix of the backwards Kalman filter of Thm. 9.8.1.

10.11 (BF formulas using the backwards innovations) Consider the setting of Prob. 10.10. Derive the following analogs of the Bryson-Frazier formulas, where the recursion for $\boldsymbol{\lambda}_i^b$ is now forwards in time:

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i^b + P_i^b\boldsymbol{\lambda}_i^b, \qquad \boldsymbol{\lambda}_i^b = F_{i,i}^{b*}\boldsymbol{\lambda}_{i-1}^b + H_i^*R_{e,i}^{-b}\mathbf{e}_i^b, \quad \boldsymbol{\lambda}_{-1} = 0.$$

10.12 (RTS formulas using the backwards innovations) Consider the setting of Prob. 10.10.

(a) Derive the following forwards-time RTS recursion:

$$\hat{\mathbf{x}}_{i|N} = F_{s,i}^b\hat{\mathbf{x}}_{i-1|N} + (\hat{\mathbf{x}}_{i|i}^b - F_{s,i}^b\hat{\mathbf{x}}_{i-1}^b), \qquad F_{s,i}^b = P_i^bF_{i,i}^{b*}P_{i-1}^{-b}.$$

(b) Use the result of Prob. 10.10 and the recursion for the predicted state estimator in the backwards Kalman filter of Thm. 9.8.1 to show that $\hat{\mathbf{x}}_{i|N} = F_{s,i}^b\hat{\mathbf{x}}_{i-1|N} + F_i^{-b}Q_i^bP_{i-1}^{-b}\hat{\mathbf{x}}_{i-1}^b$.

10.13 (Two-filter formulas using backwards BF and RTS) Consider the setting of Prob. 10.10 and assume F_i is invertible. The two-filter formulas of Thm. 10.4.1 can be derived by combining the forwards RTS formula of Thm. 10.3.1 and the backwards RTS formula of Prob. 10.12.

(a) Consider the forwards RTS smoother (10.3.1) and replace $\hat{\mathbf{x}}_{i+1|N}$ by its expression from part (b) of Prob. 10.12 to conclude that

$$(I - F_{s,i}F_{s,i+1}^b)\hat{\mathbf{x}}_{i|N} = F_{s,i}F_{i+1}^{-b}Q_{i+1}^bP_i^{-b}\hat{\mathbf{x}}_i^b + F_i^{-1}G_iQ_iG_i^*P_{i+1}^{-1}\hat{\mathbf{x}}_{i+1}.$$

(b) Use the alternative expressions for $F_{s,i}$ in Thm. 10.3.1 to shows its equivalence to the usual formula.

10.14 (Another two-filter formula from the RTS formulas) Consider the same setting of Prob. 10.10 and assume F_i is invertible.

(a) Start with the expression $\hat{\mathbf{x}}_{i|N} = F_{s,i}^b\hat{\mathbf{x}}_{i-1|N} + F_i^{-b}Q_i^bP_{i-1}^{-b}\hat{\mathbf{x}}_{i-1}^b$, and replace $\hat{\mathbf{x}}_{i-1|N}$ by its expression from Thm. 10.3.1 to show that the following relation holds:

$$(I - F_{s,i}^bF_{s,i-1})\hat{\mathbf{x}}_{i|N} = F_{s,i}^bF_{i-1}^{-1}G_{i-1}Q_{i-1}G_{i-1}^*P_i^{-1}\hat{\mathbf{x}}_i + F_i^{-b}Q_i^bP_{i-1}^{-b}\hat{\mathbf{x}}_{i-1}^b.$$

(b) Show its equivalence to the usual formula.

10.15 (Hamiltonian equations when $S_i \neq 0$) Consider again the state-space model (10.2.1)–(10.2.2) with $S_i \neq 0$ and $R_i > 0$. Show that

$$\begin{bmatrix}\hat{\mathbf{x}}_{i+1|N}\\ \boldsymbol{\lambda}_{i|N}\end{bmatrix} = \begin{bmatrix}F_i^s & G_iQ_i^sG_i^*\\ -H_i^*R_i^{-1}H_i & F_i^{s*}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}_{i|N}\\ \boldsymbol{\lambda}_{i+1|N}\end{bmatrix} + \begin{bmatrix}G_iS_iR_i^{-1}\\ H_i^*R_i^{-1}\end{bmatrix}\mathbf{y}_i,$$

with boundary conditions $\hat{\mathbf{x}}_{0|N} = \Pi_0\boldsymbol{\lambda}_{0|N}$ and $\boldsymbol{\lambda}_{N+1|N} = 0$, where $F_i^s = F_i - G_iS_iR_i^{-1}H_i$ and $Q_i^s = Q_i - S_iR_i^{-1}S_i^*$.

10.16 **(Backwards-time models)** Consider the backwards-time state-space model of Prob. 9.14, *viz.*,

$$\mathbf{x}_i^d = F_i^* \mathbf{x}_{i+1}^d + H_i^* \mathbf{u}_i^d, \qquad \mathbf{z}_i^d = G_i^* \mathbf{x}_{i+1}^d + \mathbf{v}_i^d,$$

where $\{\mathbf{u}_i^d, \mathbf{v}_i^d\}$ are uncorrelated white-noise sequences with variances $\{R_i^d \geq 0, Q_i^d > 0\}$; moreover, both are uncorrelated with $\mathbf{x}_{N+1}^d$, whose variance we denote by $P_{N+1}^d \geq 0$. All variables are zero-mean. With the same definitions as in Prob. 9.14, follow arguments similar to those in Sec. 10.2 and derive the following BF-type recursions:

$$\begin{cases} \hat{\mathbf{x}}_{i|0}^d = \hat{\mathbf{x}}_{i|i}^d + P_{i|i}^d \boldsymbol{\lambda}_{i|0}^d, \\ \boldsymbol{\lambda}_{i+1|0}^d = F_{d,i}^* \boldsymbol{\lambda}_{i|0}^d + G_i R_{e,i}^{-d} \mathbf{e}_i^d, \quad \boldsymbol{\lambda}_{0|0}^d = 0, \end{cases}$$

where $F_{d,i} = (F_i - G_i R_{e,i}^{-d} K_i^{d*})^*$,

$$\boldsymbol{\lambda}_{i|0}^d \triangleq P_{i|i}^d \left(\sum_{j=0}^{i-1} \Psi_d^*(j+1, i) G_j R_{e,j}^{-b} \mathbf{e}_j^d \right),$$

$\Psi_d(i, i) = I$, and $\Psi_d(i, j) = F_{d,i} F_{d,i-1} \ldots F_{d,j-1}$ for $j > i$. Verify further, by repeating the argument of Sec. 10.2.2, that $\hat{\mathbf{u}}_{i|0}^d = R_i^d H_i \boldsymbol{\lambda}_{i|0}^d$. In the above, $\{\hat{\mathbf{x}}_{i|0}^d, \hat{\mathbf{u}}_{i|0}^d\}$ are the l.l.m.s. estimators of $\{\mathbf{x}_i^d, \mathbf{u}_i^d\}$ given the observations $\{\mathbf{z}_0^d, \ldots, \mathbf{z}_N^d\}$.

10.17 **(Backwards-time smoothing formulas)** Consider the same setting of Prob. 10.16. Follow arguments similar to those that led to (10.7.8) and derive the following recursions:

$$\begin{cases} \hat{\mathbf{x}}_{N+1|i}^d = \hat{\mathbf{x}}_{N+1|i+1}^d + P_{N+1}^d \Psi_d^*(i+1, N+1) G_i R_{e,i}^{-d} \mathbf{e}_i^d, \quad \hat{\mathbf{x}}_{N+1|N+1}^d = 0, \\ \hat{\mathbf{u}}_{j|i}^d = \hat{\mathbf{u}}_{j|i+1}^d + R_j^d H_j \Psi_d^*(i+1, j) G_i R_{e,i}^{-d} \mathbf{e}_i^d, \quad j > i, \\ \hat{\mathbf{u}}_{j|i}^d = 0, \quad j \leq i. \end{cases}$$

Here $\{\hat{\mathbf{x}}_{N+1|i}^d, \hat{\mathbf{u}}_{j|i}^d\}$ denote the l.l.m.s. estimators of $\{\mathbf{x}_{N+1}^d, \mathbf{u}_i^d\}$ given the observations $\{\mathbf{z}_i^d, \ldots, \mathbf{z}_N^d\}$.

10.18 **(Steady-state RTS and Wiener filters)** Consider a constant parameter state-space model (10.2.1)–(10.2.2) with a stable invertible matrix F and $S = 0$. Assume the initial covariance matrix Π_0 is the unique solution of

$$\bar{\Pi} = F \bar{\Pi} F^* + GQG^*,$$

so that the zero-mean random processes $\{\mathbf{x}_i, \mathbf{y}_i\}$ are wide-sense stationary. Define $\mathbf{s}_i = H\mathbf{x}_i$.

(a) Let $\hat{\mathbf{s}}_{i|N}$ denote the smoothed estimator of $\mathbf{s}_i$ given the observations $\{\mathbf{y}_j, 0 \leq j \leq N\}$. Using the steady-state form of the RTS equation (10.3.1), *i.e.*, with $N \to \infty$, show that the transfer function from $\mathbf{y}_i$ to $\hat{\mathbf{s}}_{i|\infty}$ is given by (with P assumed invertible — see the discussion in Sec. 9.5.3)

$$W(z) = H(z^{-1}I - F_s)^{-1} F^{-1} GQG^* P^{-1} (zI - F + K_p H)^{-1} K_p,$$

where $F_s = F^{-1}(I - GQG^* P^{-1})$.

(b) Show that the transfer function of part (b) collapses to $I - RS_y^{-1}(z)$ and is therefore equal to the (noncausal) Wiener solution, *i.e.*, $W(z) = S_s(z) S_y^{-1}(z)$.

10.19 **(Steady-state smoothing via the Hamiltonian equations)** Consider the same setting of Prob. 10.18.

(a) Use the Hamiltonian equations (10.5.3) to show that the transfer function $W(z)$ from $\mathbf{y}_i$ to $\hat{\mathbf{s}}_{i|\infty}$ is also given by the expression $W(z) =$

$$[\mathrm{I} + \mathrm{H}(\mathrm{zI} - \mathrm{F})^{-1}\mathrm{GQG}^*(\mathrm{I} - \mathrm{zF}^*)^{-1}\mathrm{H}^*\mathrm{R}^{-1}]^{-1}\mathrm{H}(\mathrm{zI} - \mathrm{F})^{-1}\mathrm{GQG}^*(\mathrm{I} - \mathrm{zF}^*)^{-1}\mathrm{H}^*\mathrm{R}^{-1}.$$

(b) Show again that $W(z) = I - RS_y^{-1}(z)$.

10.20 **(Extented Hamiltonian equations)** Consider the optimization problem

$$\min_{\substack{\{x_0, u_0, \ldots, u_N \\ v_0, v_1, \ldots, v_N\}}} \left[x_0^* \Pi_0^{-1} x_0 + \sum_{i=0}^{N} v_i^* R_i^{-1} v_i + \sum_{i=0}^{N} u_i^* Q_i^{-1} u_i \right]$$

subject to the state-space constraints

$$x_{i+1} = F_i x_i + G_i u_i, \qquad y_i = H_i x_i + v_i.$$

Introduce two sequences of Lagrange multipliers, $\{\lambda_{i+1|N}, \mu_{i|N}\}$, and follow the derivation of Sec. 10.6 to arrive at the following expanded Hamiltonian equations:

$$\begin{bmatrix} \hat{x}_{i+1|N} \\ 0 \\ \lambda_{i|N} \end{bmatrix} = \begin{bmatrix} GQG^* & 0 & F \\ 0 & R & H \\ F^* & H^* & 0 \end{bmatrix} \begin{bmatrix} \lambda_{i+1|N} \\ \mu_{i|N} \\ \hat{x}_{i|N} \end{bmatrix} + \begin{bmatrix} 0 \\ -I \\ 0 \end{bmatrix} y_i,$$

with boundary conditions $\lambda_{N+1|N} = 0$ and $\Pi_0^{-1} x_0 = \lambda_{0|N}$.

10.21 **(Whittle's deterministic approach)** Refer to the Hamiltonian equations of Prob. 10.20, which can be rewritten in the z-transform domain as

$$\begin{bmatrix} GQG^* & 0 & -zI + F \\ 0 & R & H \\ -z^{-1}I + F^* & H^* & 0 \end{bmatrix} \begin{bmatrix} \lambda(z) \\ \mu(z) \\ \hat{x}(z) \end{bmatrix} + \begin{bmatrix} 0 \\ -y(z) \\ 0 \end{bmatrix} = 0.$$

Denote the para-Hermitian coefficient matrix in the above equation by $\Psi(z)$.

(a) Verify that $\Psi(z)$ can be factored as $\Psi(z) = A(z) A_0 [A(1/z^*)]^*$, where

$$A(z) = \begin{bmatrix} -zP & K_p & -zI + F \\ 0 & I & H \\ I & 0 & 0 \end{bmatrix}, \qquad A_0 = \begin{bmatrix} 0 & 0 & I \\ 0 & R_e & 0 \\ I & 0 & -P \end{bmatrix},$$

where P is the unique stabilizing solution of the DARE

$$P = FPF^* + GQG^* - K_p R_e K_p^*, \qquad K_p = FPH^* R_e^{-1}, \qquad R_e = R + HPH^*.$$

(b) Verify that

$$A^{-1}(z) = \begin{bmatrix} I & 0 & 0 \\ 0 & I & -H \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} 0 & 0 & I \\ 0 & I & 0 \\ (-zI + F_p)^{-1} & (zI - F_p)^{-1}K_p & (-I + z^{-1}F_p)^{-1}P \end{bmatrix}$$

$$A_0^{-1} = \begin{bmatrix} P & 0 & I \\ 0 & R_e^{-1} & 0 \\ I & 0 & 0 \end{bmatrix}.$$

(c) Show that

$$A_0^{-1}A^{-1}(z) \begin{bmatrix} 0 \\ y(z) \\ 0 \end{bmatrix} = \begin{bmatrix} (zI - F_p)^{-1}K_p y(z) \\ R_e^{-1}e(z) \\ 0 \end{bmatrix},$$

where $F_p = F - K_pH$ and $e(z) = \big[I - H(zI - F_p)^{-1}K_p\big]y(z)$.

(d) Conclude that

$$(z^{-1}I - F_p^*)\lambda(z) = H^*R_e^{-1}e(z),$$

$$-z^{-1}P\lambda(z) + \hat{x}(z) = (zI - F_p)^{-1}K_p y(z),$$

and therefore establish the recursions

$$\lambda_{i|N} = F_p^*\lambda_{i+1|N} + H^*R_e^{-1}e_i, \qquad \hat{x}_{i+1} = F_p\hat{x}_i + K_p y_i,$$

where

$$\hat{x}_i \triangleq \hat{x}_{i|N} - P_i\lambda_{i|N}.$$

10.22 **(A backward recursion for $P^b_{i,\infty|i}$)** Thm. 9.8.2 provides a statement of the filtered version of the backwards Kalman filter, which we shall now apply to the state-space model (10.2.1)–(10.2.2) with $\Pi_0 > 0$, $S_i = 0$, and under the additional assumption that F_i is invertible. The result of Prob. 5.15 then guarantees that the Π_i generated via $\Pi_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^*$ are invertible.

Let F^b_{i+1} be any solution to the equation $F^b_{i+1}\Pi_{i+1} = \Pi_iF_i^*$ and define $\{Q^b_{i+1}, R^b_{e,i}, K^b_{p,i}\}$ as in Thm. 9.8.2, where it is shown that the Riccati variable $P^b_{i|i}$ satisfies the backwards recursion

$$P^b_{i|i} = F^b_{i+1}P^b_{i+1|i+1}F^{b*}_{i+1} + Q^b_{i+1} - K^b_{p,i}R^b_{e,i}K^{b*}_{p,i},$$

with initial condition $P^b_{N+1|N+1} = \Pi_{N+1}$. In this problem we study what happens to the above recursion when we choose the boundary condition as $\Pi_{N+1} = \infty.I$

(a) Show that we can write, for every i,

$$F^b_{i+1} = \Pi_iF_i^*\Pi_{i+1}^{-1} = \big[F_i^{-1}\Pi_{i+1} - F_i^{-1}G_iQ_iG_i^*\big]\Pi_{i+1}^{-1}.$$

Conclude, by induction and starting with $\Pi_{N+1} = \infty.I$, that we can take $F^b_{i+1} = F_i^{-1}$.

(b) Use the relation $F^b_{i+1} = \Pi_iF_i^*\Pi_{i+1}^{-1}$ to write $Q^b_{i+1} = \Pi_i - \Pi_iF_i^*\Pi_{i+1}^{-1}F_i\Pi_i$. Verify that

$$F_iQ^b_{i+1}F_i^* = F_i\Pi_iF_i^*\big[I - \Pi_{i+1}^{-1}F_i\Pi_iF_i^*\big] = [\Pi_{i+1} - G_iQ_iG_i^*]\big[\Pi_{i+1}^{-1}G_iQ_iG_i^*\big],$$

and conclude, again by induction (*i.e.*, using the fact that $\Pi_i = \infty \cdot I$), that $F_iQ^b_{i+1}F_i^* = G_iQ_iG_i^*$ and, consequently, $Q^b_{i+1} = F_i^{-1}G_iQ_iG_i^*F_i^{-*}$.

(c) Let $P^b_{i,\infty|i}$ denote the resulting backward Riccati variable when the boundary condition is taken as $\Pi_{N+1} = \infty \cdot I$. Let also $K^b_{p,i,\infty}$ and $R^b_{e,i,\infty}$ denote the corresponding quantities $\{K^b_{p,i}$ and $R^b_{e,i}\}$ that result for this particular choice of the boundary condition. Verify, with the above expressions for Q^b_{i+1} and F^b_{i+1}, that the backwards recursions now take the forms (10.4.5) and (10.4.6). Show that the backwards Riccati variable $P^b_{i,\infty|i}$ satisfies the backwards recursion (10.4.7).

(d) Repeat the analysis for the recursions of Thm. 9.8.1, *i.e.*, let $\{P^b_{i,\infty}, K^b_{l,i,\infty}, R^b_{e,i,\infty}\}$ denote the quantities that result when the boundary condition is taken as $\Pi_{N+1} = \infty \cdot I$. Deduce the results in (10.4.8)–(10.4.9).

CHAPTER 11

Fast Algorithms

As has been noted before, the computational requirements of the Riccati equation-based Kalman filter of Ch. 9 are indifferent to whether the coefficient matrices $\{F_i, G_i, H_i, Q_i, R_i, S_i\}$ are constant (time-invariant) or not. In particular, it takes $O(n^3)$ operations (additions and multiplications of real numbers) to update P_i to P_{i+1} via the Riccati recursion (9.2.14), whether the matrices $\{F_i, G_i, H_i, Q_i, R_i, S_i\}$ are constant or not. This is a striking advantage of the state-space formulation — the difference between the time-invariant and time-variant models and solutions is just a matter of a few additional subscripts. Paradoxically, however, this very strength can be a weakness: one would expect that in some way constant-parameter problems should be easier to handle than similar time-variant problems.

It turns out that estimation for a constant-parameter state-space model can be achieved by replacing the Riccati recursion used in the Kalman filter by a different set of recursions. Apart from the fact that a new solution method for a much studied problem is of intrinsic interest, we shall see that in important cases the new equations can be solved with significantly less effort than those of the Riccati-equation based Kalman filter: $O(n^2)$ rather than $O(n^3)$ flops for each update, a great reduction for large n. However, the fast algorithms can be extended to cover certain forms of time variation encountered in several applications (especially adaptive filtering).

11.1 THE FAST (CKMS) RECURSIONS

The standard state-space model, with constant parameters, is

$$\begin{cases} \mathbf{x}_{i+1} = F\mathbf{x}_i + G\mathbf{u}_i, \\ \mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i, \quad i \geq 0, \end{cases} \tag{11.1.1}$$

with $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$ random variables such that

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q\delta_{ij} & S\delta_{ij} & 0 & 0 \\ S^*\delta_{ij} & R\delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}, \tag{11.1.2}$$

where the (time-invariant) matrices $\{F, G, H, \Pi_0, Q, S, R\}$ are assumed known, and $\{\mathbf{x}_i, \mathbf{y}_i, \mathbf{u}_i\}$ are vectors of dimensions $\{n, p, m\}$, respectively. Though not essential, we shall also make the desirable (see Sec. 9.5) assumption that $R > 0$.

The Kalman filter recursions yield the innovations as[1]

$$\mathbf{e}_i = \mathbf{y}_i - H\hat{\mathbf{x}}_i, \qquad \hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + K_i R_{e,i}^{-1}\mathbf{e}_i, \qquad \hat{\mathbf{x}}_0 = 0, \tag{11.1.3}$$

where K_i and $R_{e,i}$ are computed as

$$K_i = FP_iH^* + GS, \quad R_{e,i} = R + HP_iH^*, \tag{11.1.4}$$

with P_i found through the Riccati difference equation

$$P_{i+1} = FP_iF^* + GQG^* - K_iR_{e,i}^{-1}K_i^*, \qquad P_0 = \Pi_0. \tag{11.1.5}$$

Now whether or not the model parameters are constant, forming the triple matrix product $F_iP_iF_i^*$ takes $O(n^3)$ operations. Therefore, to exploit the constancy of the state-space model we have to find a way of computing $\{K_i, R_{e,i}\}$ in (11.1.3) that does not require the computation of P_i.

An indication of how to achieve this came from studies of the Wiener-Hopf equation in radiative transfer theory, as elaborated later (see Remark 5). This suggested that one approach to exploiting constancy is via the following result on the increments of the Riccati variable P_i:

$$\delta P_i \triangleq P_{i+1} - P_i, \quad i \geq 0.$$

Theorem 11.1.1 (A Generalized Stokes Identity) *The increments δP_i obey the identity*

$$\delta P_{i+1} = F_{p,i}[\delta P_i - \delta P_i H^* R_{e,i+1}^{-1} H \delta P_i] F_{p,i}^*, \tag{11.1.6}$$

where $F_{p,i} = F - K_iR_{e,i}^{-1}H$. ■

Proof: One proof involves straightforward algebra and is deferred to Lemma 11.1.3 further ahead. ◆

Remark 1 [An Alternative Form]. Application of the matrix inversion formula to the matrix $(I + H^*R_{e,i}^{-1}H\delta P_i)$ yields the following alternative identity:

$$\delta P_{i+1} = F_{p,i}\delta P_i(I + H^*R_{e,i}^{-1}H\delta P_i)^{-1}F_{p,i}^*, \tag{11.1.7}$$

which will arise in our discussions on a scattering theoretic/transmission-line model approach to estimation theory described in Ch. 17 (see Eq. (17.6.35)). The discussion in that chapter will

[1] In earlier chapters, we used $K_{p,i}$ rather than $K_iR_{e,i}^{-1}$; the separation of K_i and $R_{e,i}$ simplifies the formulas to be given in this chapter.

further reveal that the apparently quite algebraic identity (11.1.6) in fact has a nice physical interpretation that shows why it is intimately tied to the constancy of $\{F, G, H, Q, R, S\}$ (and will also explain the name). ◆

An immediate consequence of the relation (11.1.6) is that the rank of δP_i will never exceed the rank of δP_0. Hence, *though the P_i may have full rank (generally $P_i > 0$), the δP_i can have low rank* and this fact is the key to developing the fast recursions, as we now proceed to explain.

Since P_i is Hermitian, so are δP_i and δP_0; therefore we can always write δP_0 as

$$\delta P_0 = L_0 M_0 L_0^*, \tag{11.1.8}$$

where M_0 is $\alpha \times \alpha$, L_0 is $n \times \alpha$, and α is the rank of δP_0. That is,

$$\alpha \triangleq \text{rank of } (F\Pi_0 F^* + GQG^* - K_0 R_{e,0}^{-1} K_0^* - \Pi_0). \tag{11.1.9}$$

The factorization (11.1.8) can be obtained in many ways, using $O(n^2\alpha)$ flops, *e.g.*, by the so-called Bunch-Kaufman algorithm (*e.g.*, Higham (1996)). The factorization is also not unique, but here let us assume that we have chosen a particular one.

Example 11.1.1 (Zero Initial Condition) When $\Pi_0 = 0$,

$$\alpha = \text{rank}\,(G(Q - SR^{-1}S^*)G^*) \leq m, \quad \text{the number of inputs.}$$

See Sec. 11.2 for further discussion and another important example. ◆

Lemma 11.1.1 (Factorization of δP_i) *Assume that we have a factored representation of δP_0 as $\delta P_0 = L_0 M_0 L_0^*$, where M_0 is Hermitian, nonsingular, and of size $\alpha \times \alpha$. Then we can also write δP_i in factored form as*

$$\delta P_i = L_i M_i L_i^*, \quad i \geq 0, \tag{11.1.10}$$

where M_i is Hermitian, nonsingular, and also of size $\alpha \times \alpha$. It can be defined, along with L_i, by the recursions

$$L_{i+1} = (F - K_i R_{e,i}^{-1} H) L_i = F_{p,i} L_i, \qquad i \geq 0, \tag{11.1.11}$$

$$M_{i+1} = M_i - M_i L_i^* H^* R_{e,i+1}^{-1} H L_i M_i, \qquad i \geq 0. \tag{11.1.12}$$

■

Proof: By induction. If (11.1.10) holds, then substitution into (11.1.6) gives

$$\begin{aligned}
\delta P_{i+1} &= F_{p,i}\left[L_i M_i L_i^* - L_i M_i L_i^* H^* R_{e,i+1}^{-1} H L_i M_i L_i^*\right] F_{p,i}^*, \\
&= F_{p,i} L_i [M_i - M_i L_i^* H^* R_{e,i+1}^{-1} H L_i M_i] L_i^* F_{p,i}^*,
\end{aligned}$$

which we can rewrite as $\delta P_{i+1} = L_{i+1} M_{i+1} L_{i+1}^*$ by defining $\{L_{i+1}, M_{i+1}\}$ as in (11.1.11)-(11.1.12). The matrix M_{i+1} is clearly Hermitian. For the nonsingularity of M_{i+1}, note

that if M_i is nonsingular, then so is M_{i+1} since, in view of the matrix inversion formula,

$$\begin{aligned}
M_{i+1}^{-1} &= M_i^{-1} + L_i^* H^* (R_{e,i+1} - H L_i M_i L_i^* H^*)^{-1} H L_i, \\
&= M_i^{-1} + L_i^* H^* (R + H P_{i+1} H^* - H(P_{i+1} - P_i) H^*)^{-1} H L_i, \\
&= M_i^{-1} + L_i^* H^* R_{e,i}^{-1} H L_i,
\end{aligned} \tag{11.1.13}$$

which shows that the inverse of M_{i+1} is well defined in terms of the inverses of $\{M_i, R_{e,i}\}$. ◆

Remark 2 [Rank $L_i \leq$ Rank L_0]. While the $\{M_i\}$ are nonsingular, the matrices $\{L_i\}$ that are generated recursively via (11.1.11) can drop rank when any of the $\{F_{p,i}\}$ become singular.

To see this, consider the contrived example

$$\mathbf{x}_{i+1} = \mathbf{u}_i, \quad \mathbf{y}_i = \mathbf{v}_i,$$

with $F = 0$, $H = 0$, $G = I$. Assume that $\delta P_0 = Q - SR^{-1}S^* - \Pi_0$ has rank α. Now since $H = 0$, we have $F_{p,i} = F = 0$ and, hence, $M_i = M_0$ and $L_i = 0$ for all $i \geq 1$. That is, $\delta P_i = 0$ for all $i \geq 1$. This is a consequence of the fact that for this example, the matrices $\{P_i\}$ are constant and equal to

$$P_i = Q - SR^{-1}S^*, \quad \text{for all } i \geq 1.$$

Therefore, the factorization (11.1.10) shows that while the rank of δP_i can never exceed α, it may drop below α. ◆

The representation (11.1.10) leads immediately to the following *fast* algorithm, which for reasons described further ahead we call the Chandrasekhar-Kailath-Morf-Sidhu (CKMS) algorithm.

Theorem 11.1.2 (The Fast (CKMS) Recursions) *The $\{K_i, R_{e,i}\}$ in (11.1.3) can be recursively computed by the following set of coupled recursions, for $i \geq 0$:*

$$K_{i+1} = K_i - F L_i R_{r,i}^{-1} L_i^* H^*, \tag{11.1.14}$$

$$L_{i+1} = F L_i - K_i R_{e,i}^{-1} H L_i, \tag{11.1.15}$$

$$R_{e,i+1} = R_{e,i} - H L_i R_{r,i}^{-1} L_i^* H^*, \tag{11.1.16}$$

$$R_{r,i+1} = R_{r,i} - L_i^* H^* R_{e,i}^{-1} H L_i. \tag{11.1.17}$$

The state error variance matrices can also be computed, if desired, as

$$P_{i+1} = \Pi_0 - \sum_{j=0}^{i} L_j R_{r,j}^{-1} L_j^*. \tag{11.1.18}$$

The initial conditions are computed as follows. Clearly, $K_0 = F\Pi_0 H^* + GS$ and $R_{e,0} = R + H\Pi_0 H^*$. Now factor (nonuniquely)

$$\delta P_0 \triangleq F\Pi_0 F^* + GQG^* - K_0 R_{e,0}^{-1} K_0^* - \Pi_0 \tag{11.1.19}$$

as $\delta P_0 = -L_0 R_{r,0}^{-1} L_0^*$, where L_0 has size $n \times \alpha$, and $R_{r,0}$ is Hermitian, nonsingular, and of size $\alpha \times \alpha$, to obtain the initial conditions $\{L_0, R_{r,0}\}$. ■

Proof: First define

$$R_{r,i} \triangleq -M_i^{-1}. \tag{11.1.20}$$

Then (11.1.13) is exactly (11.1.17) in Thm. 11.1.2. Likewise note that

$$\begin{aligned} R_{e,i+1} &\triangleq R + HP_{i+1}H^* = R + HP_iH^* + H\delta P_iH^*, \\ &= R_{e,i} + HL_iM_iL_i^*H^* = R_{e,i} - HL_iR_{r,i}^{-1}L_i^*H^*, \end{aligned}$$

which is exactly (11.1.16) of Thm. 11.1.2. Moreover, combining (11.1.25) below and (11.1.10) gives

$$K_{i+1} = K_i + FL_iM_iL_i^*H^* = K_i - FL_iR_{r,i}^{-1}L_i^*H^*,$$

which is just (11.1.14). Next note that (11.1.11) is just (11.1.15). Finally, to calculate the error variance matrices, just note that

$$P_{i+1} = P_0 + \sum_{j=0}^{i} \delta P_j = \Pi_0 - \sum_{j=0}^{i} L_jR_{r,j}^{-1}L_j^*.$$

♦

Remark 3 [Number of Computations]. We see that in place of propagating the $n \times n$ matrices $\{P_i\}$, we now propagate matrices of generally smaller dimensions, assuming $m \leq n$, $p \leq n$: the $\{K_i\}$ are $n \times p$, the $\{L_i\}$ are $n \times \alpha$, the $\{R_{e,i}\}$ are $p \times p$, and the $\{R_{r,i}\}$ are $\alpha \times \alpha$. So instead of $O(n^3)$ flops per update form i to $i+1$, the dependence on n is now $O(n^2)$ and on α is $O(\alpha^3)$. So whenever $\alpha \ll n$, we can have a significant computational benefit by using the CKMS recursions. Two important situations where this always holds are discussed in Sec. 11.2. ♦

Remark 4. The recursions of Thm. 11.1.2 can be recast in several slightly different forms (see Kailath, Morf, and Sidhu (1973) and Morf, Sidhu, and Kailath (1974)); the form (11.1.14)–(11.1.17) used above arises naturally in the stationary case (see below) and also in a Redheffer scattering formulation of the state-space estimation problem (to be discussed in Ch. 17). ♦

Remark 5 [Terminology]. The recursions (11.1.14)–(11.1.17) are the discrete time analogs of certain nonlinear differential equations developed by Kailath (1972b,1973) for the continuous time estimation problem, which were generalizations of equations introduced by Chandrasekhar (1947a,1947b,1950) to solve the finite-time Wiener-Hopf equations arising in certain radiative transfer problems (see the discussion and the concluding notes in Ch. 16). The discrete-time versions were first presented in Kailath, Morf, and Sidhu (1973), and in more detail in Morf, Sidhu, and Kailath (1974). Though their resemblance to the original Chandrasekhar equations is even more remote than in the continuous-time case, these authors named them as the Chandrasekhar recursions. Whittle (1983,1990) called them the Kailath-Chandrasekhar equations. Since, apart from earlier applications (*e.g.*, Desalu et al. (1974), Saint et al. (1985), Mahalanabis and Xue (1987)), these equations are now being increasingly encountered (especially in adaptive filtering – see, *e.g.*, Houacine and Demoment (1986), Slock (1989), Houacine (1991), Sayed (1992), Sayed and Kailath (1994a,1994b), Merched and Sayed (1999)), the present authors decided to use a more accurate designation. ♦

More on the $\{R_{r,i}\}$. Despite the notation (the reasons for which lie in a different innovations derivation of the CKMS recursions (Sidhu and Kailath (1974))), the $\{R_{r,i}\}$ are not covariance matrices, unlike the innovations variances $\{R_{e,i}\}$. They are nonsingular, but in general indefinite. However, they do have an important property — their inertia is constant over time; this property will be re-established and exploited in Ch. 13 to derive a so-called array version of the fast recursions.

Lemma 11.1.2 (The $R_{r,i}$ have constant inertia) *Assume $P_0 \geq 0$. The matrices $R_{r,i}$ defined by the recursions (11.1.16)–(11.1.17) have the same inertia for all i. More specifically,*

$$In\{R_{r,i}\} = In\{\delta P_0\}, \tag{11.1.21}$$

where $\delta P_0 = P_1 - \Pi_0 = F\Pi_0F^* + GQG^* - K_{p,0}R_{e,0}K_{p,0}^* - \Pi_0$. ■

Proof: Consider the block matrix

$$\begin{bmatrix} R_{e,i} & HL_i \\ L_i^*H^* & R_{r,i} \end{bmatrix}$$

and note that $R_{r,i+1}$ is its Schur complement with respect to $R_{e,i}$, while $R_{e,i+1}$ is its Schur complement with respect to $R_{r,i}$. Hence, the lower-upper and upper-lower block triangular factorizations of the above matrix would lead to the following equalities (*cf.* App. A):

$$\begin{bmatrix} R_{e,i} & HL_i \\ L_i^*H^* & R_{r,i} \end{bmatrix} =$$

$$\begin{bmatrix} I & 0 \\ L_i^*H^*R_{e,i}^{-1} & I \end{bmatrix} \begin{bmatrix} R_{e,i} & 0 \\ 0 & R_{r,i+1} \end{bmatrix} \begin{bmatrix} I & 0 \\ L_i^*H^*R_{e,i}^{-1} & I \end{bmatrix}^*, \tag{11.1.22}$$

$$\begin{bmatrix} I & HL_iR_{r,i}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} R_{e,i+1} & 0 \\ 0 & R_{r,i} \end{bmatrix} \begin{bmatrix} I & HL_iR_{r,i}^{-1} \\ 0 & I \end{bmatrix}^*. \tag{11.1.23}$$

It then follows from Sylvester's law of inertia that the matrices $\text{diag}\{R_{e,i}, R_{r,i+1}\}$ and $\text{diag}\{R_{e,i+1}, R_{r,i}\}$ have the same inertia. But since both $R_{e,i}$ and $R_{e,i+1}$ are positive-definite (in view of the assumptions $R > 0$ and $P_0 \geq 0$), we conclude that $R_{r,i+1}$ and $R_{r,i}$ should have the same inertia. ♦

Remark 6. In Sec. 14.8, when we study the asymptotic behavior of the fast recursions, we shall allow for possibly indefinite initial conditions P_0. In this case, the matrices $\{R_{e,i}\}$ need not be positive-definite for all i and, consequently, the $\{R_{r,i}\}$ will generally not have constant inertia for all i — see Prob. 14.7. ♦

To complete our derivation of the CKMS algorithm, we still need to establish the identity of Thm. 11.1.1. We do so in the following statement, which also establishes several other useful increment relations.

Lemma 11.1.3 (Increment Relations) *The increments of $R_{e,i}$, $K_{p,i}$, $F_{p,i}$, and P_i can be written in the form*

$$R_{e,i+1} - R_{e,i} = H\delta P_i H^*, \tag{11.1.24}$$

$$K_{i+1} - K_i = F\delta P_i H^*, \tag{11.1.25}$$

$$K_{p,i+1} - K_{p,i} = F_{p,i}\delta P_i H^* R_{e,i+1}^{-1}, \tag{11.1.26}$$

$$F_{p,i+1} = F_{p,i}(I - \delta P_i H^* R_{e,i+1}^{-1} H), \tag{11.1.27}$$

$$\delta P_{i+1} = F_{p,i}[\delta P_i - \delta P_i H^* R_{e,i+1}^{-1} H \delta P_i] F_{p,i}^*. \tag{11.1.28}$$

■

Proof: The first two equalities are immediate:

$$R_{e,i+1} \triangleq R + HP_{i+1}H^* = R + HP_iH^* + H\delta P_iH^* = R_{e,i} + H\delta P_iH^*,$$

$$K_{i+1} \triangleq FP_{i+1}H^* + GS = FP_iH^* + GS + F\delta P_iH^* = K_i + F\delta P_iH^*.$$

The third takes a little more work:

$$\begin{aligned} K_{p,i+1} &\triangleq (FP_{i+1}H^* + GS)R_{e,i+1}^{-1} = (FP_iH^* + GS + F\delta P_iH^*)R_{e,i+1}^{-1}, \\ &= K_{p,i}R_{e,i}R_{e,i+1}^{-1} + F\delta P_iH^*R_{e,i+1}^{-1}, \\ &= K_{p,i}(R_{e,i+1} - H\delta P_iH^*)R_{e,i+1}^{-1} + F\delta P_iH^*R_{e,i+1}^{-1}, \\ &= K_{p,i} + (F - K_{p,i}H)\delta P_iH^*R_{e,i+1}^{-1}, \\ &= K_{p,i} + F_{p,i}\delta P_iH^*R_{e,i+1}^{-1}. \end{aligned}$$

The increment for $F_{p,i}$ follows from the identities

$$\begin{aligned} F_{p,i+1} &\triangleq F - K_{p,i+1}H = F - K_{p,i}H - F_{p,i}\delta P_iH^*R_{e,i+1}^{-1}H, \\ &= F_{p,i}[I - \delta P_iH^*R_{e,i+1}^{-1}H]. \end{aligned}$$

Finally, from the Riccati recursion (11.1.5) we have

$$\delta P_{i+1} = P_{i+2} - P_{i+1} = F\delta P_iF^* - K_{p,i+1}R_{e,i+1}K_{p,i+1}^* + K_{p,i}R_{e,i}K_{p,i}^*.$$

Now using the just-derived increment formulas (and a rearrangement of the formula $F_{p,i} = F - K_{p,i}H$), we can express this as

$$\begin{aligned} \delta P_{i+1} = {} & (F_{p,i} + K_{p,i}H)\delta P_i(F_{p,i}^* + H^*K_{p,i}^*) \\ & - (K_{p,i} + F_{p,i}\delta P_iH^*R_{e,i+1}^{-1})R_{e,i+1}(K_{p,i}^* + R_{e,i+1}^{-1}H\delta P_iF_{p,i}^*) \\ & + K_{p,i}(R_{e,i+1} - H\delta P_iH^*)K_{p,i}^*, \end{aligned}$$

which, thanks to nice cancellations, gives (11.1.6). ♦

11.2 TWO IMPORTANT CASES

The significance of the CKMS recursions can now be seen by considering two special cases, where $\alpha \leq m$, the number of inputs, and $\alpha \leq p$, the number of outputs. Since we often have $m \ll n$ and $p \ll n$, considerable computational reductions can be obtained in these special cases.

11.2.1 Zero Initial Conditions

When $\Pi_0 = 0$ we obtain

$$\delta P_0 = G(Q - SR^{-1}S^*)G^* = GQ^sG^*, \quad Q^s \triangleq Q - SR^{-1}S^*. \tag{11.2.1}$$

When G and Q^s have full rank m, we will have

$$\alpha = \text{rank}(\delta P_0) = m,$$

and we can take

$$L_0 = G, \qquad R_{r,0} = -Q^{-s}.$$

Furthermore, the Riccati variable can be computed, when desired, as

$$P_i = -\sum_{j=0}^{i-1} L_j R_{r,j}^{-1} L_j^*. \tag{11.2.2}$$

When G is a column (*i.e.*, when we have a state-space model with a single input, $m = 1$), we obtain $\alpha = 1$ and the $\{R_{r,i}\}$ become scalars while the $\{L_i\}$ become column vectors.

Remark 7 [$\{P_i\}$ is now Monotone Nondecreasing]. From (11.2.2) we would expect that $R_{r,i} \leq 0$. In fact, since $R_{r,0} = -Q^{-s} < 0$, by Lemma 11.1.2 it follows that the $\{R_{r,i}\}$ will be negative-definite for all $i > 0$. It also follows from (11.2.2) that $P_{i+1} \geq P_i$. ◆

11.2.2 Stationary Processes

When F in the standard state-space model (11.1.1) is *stable*, then we recall from Ch. 8 (Sec. 8.1.2) that there is a unique nonnegative definite matrix $\bar{\Pi}$ obeying the Lyapunov equation

$$\bar{\Pi} = F\bar{\Pi}F^* + GQG^*, \tag{11.2.3}$$

and that when we choose $\Pi_0 = \bar{\Pi}$, the processes $\{\mathbf{x}_i, i \geq 0\}$ and $\{\mathbf{y}_i, i \geq 0\}$ will be stationary, *i.e.*, $< \mathbf{x}_i, \mathbf{x}_j >$ and $< \mathbf{y}_i, \mathbf{y}_j >$ will depend only on $|i - j|$.

For this special initial condition $\Pi_0 = \bar{\Pi}$, δP_0 simplifies to

$$\delta P_0 = -K_0R_{e,0}^{-1}K_0^* = -(F\bar{\Pi}H^* + GS)(R + H\bar{\Pi}H^*)^{-1}(F\bar{\Pi}H^* + GS)^*. \tag{11.2.4}$$

Then, assuming K_0 has full rank p (which requires $n \geq p$) and since $R + H\bar{\Pi}H^* > 0$, we have

$$\alpha \triangleq \text{rank}\,(\delta P_0) = p, \tag{11.2.5}$$

and we can take

$$L_0 = F\bar{\Pi}H^* + GS, \qquad R_{r,0} = R + H\bar{\Pi}H^* = R_{e,0}. \tag{11.2.6}$$

The error covariance matrix can be found, when desired, as

$$P_i = \bar{\Pi} - \sum_{j=0}^{i-1} L_j R_{r,j}^{-1} L_j^*. \tag{11.2.7}$$

When H is a row, *i.e.*, when we have a scalar output process, then $\alpha = 1$ and the $\{L_i\}$ become column vectors.

Remark 8 [$\{P_i\}$ is now Monotone Nonincreasing]. From (11.2.6) we know that $R_{r,0} > 0$ and therefore, by Lemma 11.1.2, we shall also have $R_{r,i} > 0$ for all i. Using this fact and (11.2.7), we note the useful fact that in the stationary case the $\{P_i\}$ are monotone nonincreasing

$$P_{i+1} \leq P_i, \qquad \text{when } \Pi_0 = \bar{\Pi} \geq 0.$$

◆

Remark 9. The recursions in this special (stationary) case were noted by Kailath, Morf, and Sidhu (1973) to be related to the so-called Levinson-Whittle-Wiggins-Robinson (LWWR) algorithm for the prediction of general stationary processes, *i.e.*, not necessarily with a finite-dimensional (state-space) model (see, *e.g.*, Wiggins and Robinson (1965) and the remarks to Prob. 4.11). The recursions for the stationary case were also obtained by Lindquist (1974) by a method (rediscovering some of the results of Wiggins and Robinson (1965)) that was unique to the case of stationary processes. It was later discovered that the deeper relationship is to the generalized Schur algorithms, which apply to both stationary and nonstationary processes (see App. 13.A). ◆

11.3 STRUCTURED TIME-VARIANT SYSTEMS

The earlier discussions were limited to the case of time-invariant system matrices $\{F, G, H\}$. While there have been some efforts over the years to obtain extensions to time-variant state-space models, the most useful progress in this area has come about through a particular application in recursive least-squares problems (Sayed and Kailath (1994b)). This application motivated an extension by Sayed and Kailath (1992,1994a) of the CKMS filter to a class of time-variant state-space models where the time variation occurs in a certain structured manner. We shall refer to this variant as the extended CKMS filter.

Thus consider a general time-variant state-space model of the form

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i, \\ \mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i, \quad i \geq 0, \end{cases} \tag{11.3.1}$$

with $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$ random variables such that

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & S_i \delta_{ij} & 0 & 0 \\ S_i^* \delta_{ij} & R_i \delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}. \tag{11.3.2}$$

The time variation in the model (11.3.1) is said to be *structured* if there exist $n \times n$ matrices Ψ_i such that the matrices F_i, H_i, and G_i vary according to the rules (see Prob. 11.7 for one example):

$$H_i = H_{i+1}\Psi_i, \qquad F_{i+1}\Psi_i = \Psi_{i+1}F_i, \qquad G_{i+1} = \Psi_{i+1}G_i. \tag{11.3.3}$$

In other words, the matrices $\{F_i, G_i, H_i\}$ are allowed to vary in time but in such a way that the time variations are tied together via *known* time-variant matrices Ψ_i. In this section we shall further assume that the covariance matrices R_i, S_i, and Q_i are constant for all i ($R_i = R$, $Q_i = Q$, $S_i = S$); the results however can be extended to time-variant covariance matrices (by following the discussion in Morf and Kailath (1975) — see, *e.g.*, Sayed (1992)).

Constant-parameter systems clearly satisfy (11.3.3) with $\Psi_i = I$. An even more general case is mentioned at the end of Sec. 13.4 and in some of the problems, but here we stay with (11.3.3) in order to convey the main ideas. Further ahead we shall provide a physical interpretation for the constraints (11.3.3).

The savings in computation are now achieved by considering generalized difference matrices of the form

$$\delta_\Psi P_i \triangleq P_{i+1} - \Psi_i P_i \Psi_i^*,$$

and by factoring them (nonuniquely) as

$$P_{i+1} - \Psi_i P_i \Psi_i^* = -L_i R_{r,i}^{-1} L_i^*, \qquad i \geq 0,$$

for some $n \times \alpha$ matrix L_i and $\alpha \times \alpha$ Hermitian and nonsingular matrix $R_{r,i}$. Following the same arguments as in Sec. 11.1, we can verify that α can be chosen as

$$\begin{aligned} \alpha &= \text{rank } (P_1 - \Psi_0 P_0 \Psi_0^*), \\ &= \text{rank } (F_0 \Pi_0 F_0^* + G_0 Q G_0^* - K_0 R_{e,0}^{-1} K_0^* - \Psi_0 \Pi_0 \Psi_0^*), \end{aligned}$$

and that the following recursions now replace those of Thm. 11.1.2 — see Prob. 11.6.

Theorem 11.3.1 (Extended Fast Recursions) *The K_i and $R_{e,i}$ needed in the estimator equation*

$$\hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_i + K_i R_{e,i}^{-1} (\mathbf{y}_i - H_i \hat{\mathbf{x}}_i), \quad \hat{\mathbf{x}}_0 = 0,$$

for the structured state-space model (11.3.1)–(11.3.3) can be recursively computed via the recursions, for $i \geq 0$,

$$K_{i+1} = \Psi_{i+1} K_i - F_{i+1} L_i R_{r,i}^{-1} L_i^* H_{i+1}^*, \tag{11.3.4}$$

$$L_{i+1} = F_{i+1} L_i - \Psi_{i+1} K_i R_{e,i}^{-1} H_{i+1} L_i, \tag{11.3.5}$$

$$R_{e,i+1} = R_{e,i} - H_{i+1} L_i R_{r,i}^{-1} L_i^* H_{i+1}^*, \tag{11.3.6}$$

$$R_{r,i+1} = R_{r,i} - L_i^* H_{i+1}^* R_{e,i}^{-1} H_{i+1} L_i. \tag{11.3.7}$$

The initial conditions are computed as follows. First $K_0 = F_0 \Pi_0 H_0^* + G_0 S$ and $R_{e,0} = R + H_0 \Pi_0 H_0^*$, while $\{L_0, R_{r,0}\}$ are defined via the (nonunique) factorization $P_1 - \Psi_0 P_0 \Psi_0^* = -L_0 R_{r,0}^{-1} L_0^*$. ■

Physical Interpretation. The structured rules (11.3.3) have an interesting physical interpretation that shows that time invariance is actually hidden in the conditions (11.3.3): the system is *internally* time-variant but *externally* time-invariant. Note first that if both $\mathbf{x}_0$ and the noise sequence $\{\mathbf{v}_i\}$ were zero in the model (11.3.1), then the output of the state-space model will be given by

$$\mathbf{y}_i = \sum_{j=0}^{i-1} H_i \Phi(i, j+1) G_j \mathbf{u}_j = \sum_{j=0}^{i-1} \Gamma_{ij} \mathbf{u}_j,$$

where $\Phi(i, j)$ is the state-transition matrix, defined by

$$\Phi(i, j) = F_{i-1} F_{i-2} \dots F_j, \quad i > j, \qquad \Phi(i, i) = I,$$

and the Γ_{ij} are the so-called impulse response or *Markov parameters*,

$$\Gamma_{ij} = H_i \Phi(i, j+1) G_j = H_i F_{i-1} \dots F_{j+1} G_j.$$

It is easy to see that Γ_{ij} is equal to the output of the model (11.3.1) at time i in response to an impulse at time $j < i$. The following conclusion is now a consequence of the structured time variation specified by (11.3.3).

Lemma 11.3.1 (Markov Parameters) *If conditions (11.3.3) hold, then the system Markov parameters are constant, i.e., the system has a time-invariant impulse response function so that Γ_{ij} is a function of the time increment $(i - j)$, written as $\Gamma_{ij} = \Gamma_{i-j}$.* ∎

Proof: This follows from the following sequence of easily verifiable equalities (for $i > j$):

$$\begin{aligned}
\Gamma_{ij} &= H_i F_{i-1} F_{i-2} \dots F_{j+1} G_j, \\
&= H_{i+1} \Psi_i F_{i-1} F_{i-2} \dots F_{j+1} G_j, \\
&= H_{i+1} F_i \Psi_{i-1} F_{i-2} \dots F_{j+1} G_j, \\
&= H_{i+1} F_i F_{i-1} \dots F_{j+2} \Psi_{j+1} G_j, \\
&= H_{i+1} F_i F_{i-1} \dots F_{j+2} G_{j+1} \quad = \quad \Gamma_{i+1, j+1}.
\end{aligned}$$

◆

In fact, a converse statement also holds under certain controllability and observability assumptions. The details are pursued in Prob. 11.10.

11.4 CKMS RECURSIONS GIVEN COVARIANCE DATA

Since the quantities $\{R_{e,i}, K_i\}$ determine the innovations representation (see Sec. 9.2.5), they can be determined from knowledge of the covariance data rather than from an explicit state-space model for the output process $\{\mathbf{y}_i\}$. This was shown in Sec. 9.6, using a Riccati recursion (9.6.4) that involved only the covariance parameters. Here we shall show how the CKMS recursions can be expressed using only covariance information.

Recall from Sec. 9.6 that for the state-space model (11.1.1)–(11.1.2), the covariance function of the output process $\{\mathbf{y}_i\}$ is given by

$$R_y(i, j) = \langle \mathbf{y}_i, \mathbf{y}_j \rangle = \begin{cases} H F^{i-j-1} N_j & \text{if } i > j, \\ H \Pi_i H^* + R & \text{if } i = j, \\ N_i^* F^{*(j-i-1)} H^* & \text{if } i < j, \end{cases} \tag{11.4.1}$$

where

$$\Pi_{i+1} = F \Pi_i F^* + G Q G^*, \qquad N_i = F \Pi_i H^* + G S. \tag{11.4.2}$$

So now assume that instead of the model parameters $\{F, G, H, R, Q, S\}$ for the process $\{\mathbf{y}_i\}$, we are given the covariance matrices $\{F, H, N_i\}$ and the values $R_y(i, i)$. Now the initial conditions K_0 and $R_{e,0}$ that are needed for the CKMS recursions of Thm. 11.1.2 are simply

$$R_{e,0} = R_y(0, 0), \qquad K_0 = N_0.$$

However, to obtain $\{L_0, R_{r,0}\}$ we need to factor

$$P_1 - P_0 = \Pi_1 - N_0 R_y^{-1}(0, 0) N_0^* - \Pi_0. \tag{11.4.3}$$

This means that we still need to identify the difference matrix $\Pi_1 - \Pi_0$ from the given covariance data. For this purpose, we proceed as follows. Introduce the covariance matrices

$$\mathcal{R}_0 = \left[\langle \mathbf{y}_i, \mathbf{y}_j \rangle\right]_{i,j=0}^{n-1}, \qquad \mathcal{R}_1 = \left[\langle \mathbf{y}_i, \mathbf{y}_j \rangle\right]_{i,j=1}^{n}.$$

Then it is easy to verify by direct calculations, and using (11.4.1)–(11.4.2), that

$$\mathcal{R}_1 - \mathcal{R}_0 = \mathcal{O}(\Pi_1 - \Pi_0)\mathcal{O}^*,$$

where

$$\mathcal{O} \triangleq \text{col}\{H, HF, HF^2, \dots, HF^{n-1}\},$$

is the observability matrix of the pair $\{F, H\}$. If we assume that this matrix is full rank, *i.e.*, that $\{F, H\}$ is observable, then we can find $(\Pi_1 - \Pi_0)$ as

$$\Pi_1 - \Pi_0 = \mathcal{O}^\dagger (\mathcal{R}_1 - \mathcal{R}_0) \mathcal{O}^{\dagger *}, \tag{11.4.4}$$

where $\mathcal{O}^\dagger \triangleq (\mathcal{O}^* \mathcal{O})^{-1} \mathcal{O}^*$. Once $(\Pi_1 - \Pi_0)$ is known, $P_1 - P_0$ is determined and can be factored as $P_1 - P_0 = -L_0 R_{r,0}^{-1} L_0^*$. With the initial conditions $\{K_0, L_0, R_{e,0}, R_{r,0}\}$ so determined, we can proceed with the recursions of Thm. 11.1.2.

Remark 10. The above argument further shows that, when $\{F, H\}$ *is observable*, the rank of $(P_1 - P_0)$ and, hence, the value of the parameter α, is model-independent; its value can be determined from the covariance data alone. In fact, we shall soon make a more direct connection with the so-called displacement rank of the covariance matrix R_y. ◆

Remark 11 [Stationary Models]. When the process model is stationary, $\Pi_1 = \Pi_0 = \bar{\Pi}$, so that (11.4.3) becomes $P_1 - P_0 = -N_0 R_y^{-1}(0, 0) N_0^*$, which is directly specified by knowledge of the covariance function. ◆

11.5 RELATION TO DISPLACEMENT RANK

The computational advantages of the CKMS recursions depend upon the value of the parameter

$$\alpha = \text{rank of } (F\Pi_0 F^* + GQG^* - K_0 R_{e,0}^{-1} K_0^* - \Pi_0),$$

a rather complex expression determined by the model parameters $\{F, G, Q, R, S, \Pi_0\}$. Once again, a deeper insight into its significance comes from examining the covariance function, $R_y(\cdot,\cdot)$, of the output process $\{\mathbf{y}_i\}$, and in particular the displacement rank, say r, of the covariance matrix R_y. We shall show that $r \leq \alpha + p$, where p is the dimension of the vector output process $\{\mathbf{y}_i\}$, with equality except in degenerate cases. The concept of displacement structure has been briefly mentioned before (see Sec. 4.2.6) and it is discussed in more detail in Apps. F and 13.A.

Thus let R_y denote the covariance matrix of the output process $\{\mathbf{y}_i\}$ that is described by the model (11.1.1)–(11.1.2); note that R_y has $p \times p$ block entries that correspond to cross-covariance matrices $\langle \mathbf{y}_i, \mathbf{y}_j \rangle$. The displacement of R_y is defined as (see Kailath and Sayed (1995,1999) for overviews of displacement structure theory):

$$\nabla_{Z^p} R_y \triangleq R_y - Z^p R_y [Z^p]^*, \tag{11.5.1}$$

where Z denotes the semi-infinite lower triangular shift matrix with ones of the first subdiagonal and zeros elsewhere,

$$Z = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & 1 & 0 & \\ & & \ddots & \ddots \end{bmatrix}, \tag{11.5.2}$$

and Z^p is then the lower triangular shift matrix with ones on the p-th subdiagonal and zeros elsewhere. The displacement rank of R_y is defined as

$$r \triangleq \text{rank } \left(\nabla_{Z^p} R_y\right).$$

[A classic example is the case of a Toeplitz matrix T with scalar entries, in which case the reader can check that the displacement is a matrix with the same first row and column as T, and zeros everywhere else. The displacement rank of T is clearly less than or equal to 2. When the entries are $p \times p$ blocks, the displacement rank is less than or equal to $2p$.]

Now let us consider the matrix R_y defined by the $R_y(i, j)$ as specified in (11.4.1)–(11.4.2),

$$\langle \mathbf{y}_i, \mathbf{y}_j \rangle = \begin{cases} HF^{i-j-1}N_j & \text{if } i > j, \\ H\Pi_i H^* + R & \text{if } i = j, \\ N_i^* F^{*(j-i-1)} H^* & \text{if } i < j, \end{cases} \tag{11.5.3}$$

where

$$\Pi_{i+1} = F\Pi_i F^* + GQG^*, \qquad N_i = F\Pi_i H^* + GS. \tag{11.5.4}$$

It then follows that

$$\Pi_{i+1} - \Pi_i = F^i \Delta F^{*i}, \quad \Delta \triangleq \Pi_1 - \Pi_0,$$

and

$$\langle \mathbf{y}_i, \mathbf{y}_i \rangle - \langle \mathbf{y}_{i-1}, \mathbf{y}_{i-1} \rangle = HF^{(i-1)} \Delta F^{*(i-1)} H^*,$$
$$\langle \mathbf{y}_i, \mathbf{y}_{i+1} \rangle - \langle \mathbf{y}_{i-1}, \mathbf{y}_i \rangle = HF^{(i-1)} \Delta F^{*i} H^*.$$

Using these identities, we can verify that

$$\nabla_{Z^p} R_y = \begin{bmatrix} R_{e,0} & K_0^* H^* & K_0^* F^* H^* & K_0^* F^{*2} H^* & \cdots \\ HK_0 & H\Delta H^* & H\Delta F^* H^* & H\Delta F^{*2} H^* & \\ HFK_0 & HF\Delta H^* & HF\Delta F^* H^* & HF\Delta F^{*2} H^* & \\ HF^2 K_0 & HF^2 \Delta H^* & HF^2 \Delta F^* H^* & HF^2 \Delta F^{*2} H^* & \\ \vdots & & & & \ddots \end{bmatrix} \tag{11.5.5}$$

where

$$R_{e,0} = R + H\Pi_0 H^*, \qquad K_0 = N_0 = F\Pi_0 H^* + GS.$$

We thus see that the displacement rank of R_y is determined by $\Delta = \Pi_1 - \Pi_0$. Note that in the stationary case (when $\Pi_0 = \bar{\Pi}$), $\Delta = 0$ and $\nabla_{Z^p} R_y$ will have rank $\leq 2p$.

Note also that even in the general case there is significant redundancy in the elements of $\nabla_{Z^p} R_y$ since successive rows differ only by multiples of F. One therefore suspects that the rank of $\nabla_{Z^p} R_y$ is still low, and this can be verified by going through the first few (in fact, two) steps of Schur reduction. Let us begin with the Schur complement of the (0, 0) block entry of $\nabla_{Z^p} R_y$, which can be seen to be equal to

$$\nabla_{Z^p} R_y - \begin{bmatrix} R_{e,0} \\ HK_0 \\ HFK_0 \\ \vdots \end{bmatrix} R_{e,0}^{-1} \begin{bmatrix} R_{e,0} & K_0^* H^* & K_0^* F^* H^* & \cdots \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ 0 & H\delta P_0 H^* & H\delta P_0 F^* H^* & H\delta P_0 F^{*2} H^* & \\ 0 & HF\delta P_0 H^* & HF\delta P_0 F^* H^* & HF\delta P_0 F^{*2} H^* & \\ 0 & HF^2 \delta P_0 H^* & HF^2 \delta P_0 F^* H^* & HF^2 \delta P_0 F^{*2} H^* & \\ \vdots & & & & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ H \\ HF \\ \vdots \end{bmatrix} \delta P_0 \begin{bmatrix} 0 & H^* & F^* H^* & \cdots \end{bmatrix},$$

where we used the easily verifiable relation

$$H\Delta H^* - HK_0R_{e,0}^{-1}K_0^*H^* = H(P_1 - P_0)H^* \triangleq H\delta P_0 H^*,$$

and where P_1 is the Riccati variable that satisfies the recursion

$$P_1 = FP_0F^* + GQG^* - K_0R_{e,0}^{-1}K_0^*.$$

It now follows easily that

$$\nabla_{Z^p}R_y = \begin{bmatrix} R_{e,0} \\ HK_0 \\ HFK_0 \\ \vdots \end{bmatrix} R_{e,0}^{-1} \begin{bmatrix} R_{e,0} \\ HK_0 \\ HFK_0 \\ \vdots \end{bmatrix}^* + \begin{bmatrix} 0 \\ H \\ HF \\ \vdots \end{bmatrix} \delta P_0 \begin{bmatrix} 0 \\ H \\ HF \\ \vdots \end{bmatrix}^*,$$

which we can further factor as follows. Assuming δP_0 has rank α, we factor δP_0 as $\delta P_0 = -L_0R_{r,0}^{-1}L_0^*$, where L_0 is $n \times \alpha$ and $R_{r,0}$ is $\alpha \times \alpha$ and nonsingular. If we also introduce any decomposition (*e.g.*, the eigendecomposition) of the Hermitian matrix $-R_{r,0}^{-1}$, say

$$-R_{r,0}^{-1} = U_0JU_0^*,$$

where J is an $\alpha \times \alpha$ signature matrix with as many $\pm 1's$ as the number of positive and negative eigenvalues of $R_{r,0}$, then we can express δP_0 as

$$\delta P_0 = \bar{L}_0J\bar{L}_0^*, \qquad \bar{L}_0 = L_0U_0.$$

Then we can rewrite $\nabla_{Z^p}R_y$ as

$$\nabla_{Z^p}R_y = \mathcal{G}\begin{bmatrix} I_p & 0 \\ 0 & J \end{bmatrix}\mathcal{G}^*, \tag{11.5.6}$$

where $\mathcal{G}$ still has two (block) columns with entries

$$\mathcal{G} = \begin{bmatrix} R_{e,0}^{1/2} & 0 \\ H\bar{K}_{p,0} & H\bar{L}_0 \\ HF\bar{K}_{p,0} & HF\bar{L}_0 \\ \vdots & \vdots \end{bmatrix}, \qquad \bar{K}_{p,0} = K_0R_{e,0}^{-*/2}.$$

Now the displacement rank, *i.e.*, $r = \text{rank}(\nabla_{Z^p}R_y)$, is clearly equal to the rank of $\mathcal{G}$. The first block column has rank p, since $R_{e,0} > 0$, while the rank of the second block column is less than or equal to α, the number of columns in $\bar{L}_0$. Therefore, $r \leq p + \alpha$, as claimed earlier; the generic situation is one of equality, *i.e.*, $r = p + \alpha$.

In App. 13.A we shall show how the displacement structure representation (11.5.6) can be used to derive an array form of the CKMS recursions by means of a generalized Schur algorithm presented in App. F.

11.6 COMPLEMENTS

Sec. 11.1. The Fast (CKMS) Recursions. The CKMS recursions were first presented in Kailath, Morf, and Sidhu (1973), as an extension to discrete time of results first obtained in continuous time (Kailath (1972b,1973)). The extension was actually quite difficult, for an interesting reason. The key fact is the generalized Stokes identity (11.1.6), which in continuous time takes the differential form (see Sec. 16.6)

$$\dot{P}(t) = \Psi(t,0)\dot{P}(0)\Psi^*(t,0),$$

where $\Psi(t,0)$ is the solution of

$$\frac{d\Psi(t,0)}{dt} = [F - K(t)H]\Psi(t,0), \qquad \Psi(0,0) = I.$$

The discrete-time version might therefore be expected to be

$$\delta P_{i+1} = F_{p,i}\delta P_i F_{p,i}^*, \qquad F_{p,i} = (F - K_{p,i}H),$$

whereas it is (*cf.* (11.1.6))

$$\delta P_{i+1} = F_{p,i}[\delta P_i - \delta P_i H^* R_{e,i+1}^{-1} H\delta P_i]F_{p,i}^*.$$

Briefly stated, the reason is that the "quadratic" term in δP_i vanishes in the continuous limit (because it goes to zero faster than δP_i). There are other simplifications in continuous time that are discussed in Sec. 16.6, *e.g.*, the $\{R_{e,i}, R_{r,i}\}$ recursions do not appear.

The origins of these fast algorithms in the work of Ambartsumian (1943) and Chandrasekhar (1947a,1947b) on solving a class of Wiener-Hopf equations (encountered in radiative transfer theory) by reduction to nonlinear (partial) differential equations of Riccati-type, will be discussed in Ch. 16.

Sec. 11.2. Two Important Cases. The continuous-time stationary case is actually the one that was studied in radiative transfer theory. The recursions for discrete-time stationary processes can be related to the now famous Szegö-Levinson-Durbin algorithms for solving Toeplitz systems of linear equations, and to their generalization by Wiggins and Robinson (1965) to the block Toeplitz case (*i.e.*, vector-valued processes $\{\mathbf{y}_i\}$) — see the remarks to Prob. 4.11 and also Friedlander et al. (1978) and Kailath (1974). Lindquist (1974,1976) used Wiggins-Robinson-type arguments to independently obtain the results for the stationary case. However, such arguments are difficult to extend to the general case, though this can be done (see Sidhu and Kailath (1974)).

Actually, the more appropriate (for handling both stationary and nonstationary processes) connection is to the problem of factoring Toeplitz matrices (rather than their inverses, as above); this direct factorization problem was studied by Rissanen (1973), Morf (1974), and others, essentially rediscovering an algorithm due to Schur (1917); see also Chang and Georgiou (1992). The important fact is that the Schur algorithm can be nicely generalized to non-Toeplitz matrices by introducing the notion of displacement structure. As one application, it will be shown in App. 13.A that when state-space structure is explicitly incorporated, a generalized Schur algorithm reduces to the CKMS recursions of Thm. 11.1.2 — see also the discussions in Sayed, Lev-Ari, and Kailath (1992), and Sayed, Kailath, and Lev-Ari (1994).

Sec. 11.3. Structured Time-Variant Systems. The extended fast recursions of this section were derived by Sayed and Kailath (1992,1994a). They include as special cases some equations that appeared in Houacine and Demoment (1986), Slock (1989), and Houacine (1991). These earlier works did not study or introduce time-variant state-space systems as in (11.3.3) for general time-variant matrices $\{F_i, G_i, H_i, \Psi_i\}$. Their results can be obtained as a special case by setting $G_i = 0$, $F_i = \alpha I$, and $\Psi_i = Z$ (the lower triangular shift matrix). One particular application where such structured state-space models are useful is adaptive filtering (*cf.* Sayed and Kailath (1994b) and Merched and Sayed (1999)).

■ PROBLEMS

11.1 **(Formulas for δP_i)** Show that we can rewrite the Riccati recursion (11.1.5) as $P_{i+1} = F_{p,i} P_i F^* + GQG^*$. Use this form to derive the increment formulas of Lemma 11.1.1.

11.2 **(A formula for $R_{r,i}$)** Let $\Phi_p(i, 0)$ be the closed-loop state transition matrix defined by

$$\Phi_p(i+1, 0) \;=\; [F - K_{p,i}H]\Phi_p(i, 0).$$

Introduce the observability Gramian matrix

$$\mathcal{O}_i \triangleq \sum_{j=0}^{i} \Phi_p^*(j, 0) H^* R_{e,j}^{-1} H \Phi_p(j, 0), \qquad \mathcal{O}_{-1} = 0.$$

(a) Verify that $\mathcal{O}_i = \mathcal{O}_{i-1} + \Phi_p^*(i, 0) H^* R_{e,i}^{-1} H \Phi_p(i, 0)$.

(b) Use (11.1.17) to show that $R_{r,i+1} = R_{r,i} - L_0^*(\mathcal{O}_i - \mathcal{O}_{i-1})L_0$. Conclude that $R_{r,i} = R_{r,0} - L_0^* \mathcal{O}_{i-1} L_0$.

11.3 **(A modified fast algorithm)** In Sec. 11.1 we derived the fast recursions by propagating the factors $\{L_i, R_{r,i}\}$ in the factorization $P_{i+1} - P_i = -L_i R_{r,i}^{-1} L_i^*$. In this problem we derive a modified algorithm by propagating the factors $\{L_{1,i}, L_{2,i}, R_{z,i}\}$ of the following alternative factorization:

$$\begin{bmatrix} H \\ F \end{bmatrix} (P_{i+1} - P_i) \begin{bmatrix} H \\ F \end{bmatrix}^* \triangleq - \begin{bmatrix} L_{1,i} \\ L_{2,i} \end{bmatrix} R_{z,i}^{-1} \begin{bmatrix} L_{1,i} \\ L_{2,i} \end{bmatrix}^*,$$

for some invertible matrix $R_{z,i}$, say of size $\lambda \times \lambda$.

(a) Verify that $\{R_{e,i}, K_i\}$ can be propagated as follows

$$R_{e,i+1} = R_{e,i} - L_{1,i} R_{z,i}^{-1} L_{1,i}^*, \qquad K_{i+1} = K_i - L_{2,i} R_{z,i}^{-1} L_{1,i}^*.$$

(b) Using the expression (11.1.6) for δP_{i+1} show that $\{L_{1,i+1}, L_{2,i+1}, R_{z,i+1}\}$ can be updated as follows:

$$\begin{bmatrix} L_{1,i+1} \\ L_{2,i+1} \end{bmatrix} = \begin{bmatrix} H \\ F \end{bmatrix} (L_{2,i} - K_i R_{e,i}^{-1} L_{1,i}),$$

$$R_{z,i+1} = R_{z,i} - L_{1,i}^* R_{e,i}^{-1} L_{1,i}.$$

(c) Show also that $R_{z,i+1}$ is invertible, and that the initial conditions can be obtained by a factorization of the form

$$\begin{bmatrix} H \\ F \end{bmatrix} (P_1 - P_0) \begin{bmatrix} H \\ F \end{bmatrix}^* \triangleq - \begin{bmatrix} L_{1,0} \\ L_{2,0} \end{bmatrix} R_{z,0}^{-1} \begin{bmatrix} L_{1,0} \\ L_{2,0} \end{bmatrix}^*,$$

for some full rank $\lambda \times \lambda$ matrix $R_{z,0}$.

Remark. The value of λ clearly does not exceed the rank α of $P_1 - P_0$, so that this algorithm is at least as efficient as the CKMS recursions of Thm. 11.1.2. The value of λ can however drop below α when the multiplication by $\text{col}\{H, F\}$ reduces the rank due to rank deficiencies, in which case the algorithm becomes more efficient. Also, an initial condition P_0 that does not result in a low rank difference $P_1 - P_0$ can still lead to a small λ after multiplication by $\text{col}\{H, F\}$. ◆

11.4 **(Fast recursions for the adjoint variable)** Consider the adjoint state variable $\lambda_{i|N}$ that was introduced in the solution of the smoothing problem in Sec. 10.2 (see Eq. (10.2.5)). Recall from (10.2.7) that $\lambda_{i|N}$ satisfies the backwards recursion

$$\lambda_{i|N} = F_{p,i}^* \lambda_{i+1|N} + H^* R_{e,i}^{-1} \mathbf{e}_i, \qquad \lambda_{N+1|N} = 0,$$

where $F_{p,i} = F - K_{p,i}H$ and $K_{p,i} = K_i R_{e,i}^{-1}$. Fix a time instant i_0 and introduce the quantities $W_j = H\Phi_p(j, i_0)$ and $Z_j = L_{i_0}^* \mathcal{O}_{j-1,i_0}$, where $\mathcal{O}_{j-1,i_0}$ is the observability Gramian defined by

$$\mathcal{O}_{j-1,i_0} \triangleq \sum_{k=i_0}^{j-1} \Phi_p^*(k, i_0) H^* R_{e,k}^{-1} H \Phi_p(k, i_0),$$

and

$$\Phi_p(j, i_0) = F_{p,j-1} F_{p,j-2} \ldots F_{p,i_0} \quad \text{for} \quad j > i_0 \quad \text{and} \quad \Phi_p(i_0, i_0) = I.$$

Moreover, L_{i_0} arises from the factorization $P_{i_0+1} - P_{i_0} = -L_{i_0} R_{r,i_0}^{-1} L_{i_0}^*$.

(a) Use the change-in-initial-conditions formula in part (c) of Prob. 14.10 to show that

$$W_j = W_{j-1}[I - L_{i_0} R_{r,i_0}^{-1} Z_{j-1}]^{-1} F_{p,i_0}, \qquad W_{i_0} = H.$$

(b) Show further that $Z_j = Z_{j-1} + L_{i_0}^* W_{j-1}^* R_{e,j-1}^{-1} W_{j-1}$ with $Z_{i_0} = 0$.

(c) Show also that $\lambda_{i_0|N}$ can evaluated via

$$\lambda_{i_0|N} = \sum_{j=i}^{N} W_j^* R_{e,j}^{-1} \mathbf{e}_j.$$

Remark. The variables $\{L_{i_0}, R_{r,i_0}, R_{e,j}\}$ that are needed in the above solution can of course be computed via the fast recursions of Thm. 11.1.2. We have thus derived a fast procedure for evaluating the adjoint variable $\lambda_{i_0|N}$ using the auxiliary quantities $\{W_j, Z_j\}$. The resulting recursions are the discrete-time analogs of a fast algorithm derived in Ljung and Kailath (1977). Their algorithm was motivated by results in a scattering theory approach to estimation (see Ch. 17). ◆

11.5 **(Fast recursions in information form)** Assume that the matrices

$$F^s \triangleq F - GSR^{-1}H \quad \text{and} \quad Q^s \triangleq Q - SR^{-1}S^*$$

are invertible. Then recall that we derived in Sec. 9.5.5 the following information form recursions for the filtering problem, written here for the constant-parameter case:

$$P_{i+1|i+1}^{-1} = F^{-s*}P_{i|i}^{-1}F^{-s} + H^*R^{-1}H - K_{p,i}^d R_{e,i}^d K_{p,i}^{d*},$$

where

$$K_{p,i}^d = K_i^d R_{e,i}^{-d}, \quad K_i^d = F^{-s*}P_{i|i}^{-1}F^{-s}G, \quad R_{e,i}^d = Q^{-s} + G^*F^{-s*}P_{i|i}^{-1}F^{-s}G.$$

Introduce the factorization $P_{1|1}^{-1} - P_{0|0}^{-1} = -L_0^d R_{r,0}^{-d} L_0^{d*}$, where $R_{r,0}^d$ is invertible. Show that the quantities $\{K_i^d, R_{e,i}^d\}$ can be propagated recursively via the following CKMS-type recursions:

$$\begin{aligned}
K_{i+1}^d &= K_i^d - F^{-s*}L_i^d R_{r,i}^{-d} L_i^{d*} F^{-s}G, \\
L_{i+1}^d &= F^{-s*}L_i^d - K_i^d R_{e,i}^{-d} G^*F^{-s*}L_i^d, \\
R_{e,i+1}^d &= R_{e,i}^d - G^*F^{-s*}L_i^d R_{r,i}^{-d} L_i^{d*} F^{-s}G, \\
R_{r,i+1}^d &= R_{r,i}^d - L_i^{d*}F^{-s}GR_{e,i}^{-d}G^*F^{-s*}L_i^d,
\end{aligned}$$

with initial conditions $K_0^d = F^{-s*}P_{0|0}^{-1}F^{-s}G$ and $R_{e,0}^d = Q^{-s} + G^*F^{-s*}P_{0|0}^{-1}F^{-s}G$.

Hint. Consider the auxiliary state-space model

$$\mathbf{x}_{i+1}^d = F^{-s*}\mathbf{x}_i^d + H^*\mathbf{v}_i^d, \qquad \mathbf{y}_i^d = G^*F^{-s*}\mathbf{x}_i^d + \mathbf{u}_i^d,$$

where $\{\mathbf{v}_i^d, \mathbf{u}_i^d, \mathbf{x}_0^d\}$ are zero-mean random variables with covariance matrices

$$\left\langle \begin{bmatrix} \mathbf{v}_i^d \\ \mathbf{x}_0^d \\ \mathbf{u}_i^d \end{bmatrix}, \begin{bmatrix} \mathbf{v}_j^d \\ \mathbf{x}_0^d \\ \mathbf{u}_j^d \end{bmatrix} \right\rangle = \begin{bmatrix} R^{-1}\delta_{ij} & 0 & 0 \\ 0 & P_{0|0}^{-1} & 0 \\ 0 & 0 & Q^{-s}\delta_{ij} \end{bmatrix}.$$

11.6 **(Increment formulas for structured systems)** Consider the structured state-space model (11.3.1)–(11.3.3). Define $F_{p,i}^{\Psi} = F_{i+1} - \Psi_{i+1}K_{p,i}H_{i+1}$.

(a) Show that $R_{e,i+1} = R_{e,i} + H_{i+1}\delta_\Psi P_i H_{i+1}^*$ and $K_{i+1} = \Psi_{i+1}K_i + F_{i+1}\delta_\Psi P_i H_{i+1}^*$.

(b) Show that $K_{p,i+1} = \Psi_{i+1}K_{p,i} + F_{p,i}^{\Psi}\delta_\Psi P_i H_{i+1}^* R_{e,i+1}^{-1}$.

(c) Using the relations in (a) and (b) derive the identity

$$\delta_\Psi P_{i+1} = F_{p,i}^{\Psi}\left[\delta_\Psi P_i - \delta_\Psi P_i H_{i+1}^* R_{e,i+1}^{-1} H_{i+1}\delta_\Psi P_i\right] F_{p,i}^{\Psi *}.$$

(d) Assume $\delta_\Psi P_i$ is factored as $\delta_\Psi P_i = -L_i R_{r,i}^{-1} L_i^*$, where $R_{r,i}$ is $\alpha \times \alpha$ and L_i is $n \times \alpha$. Show that $\delta_\Psi P_{i+1}$ can be factored as

$$\delta_\Psi P_{i+1} = -L_{i+1}R_{r,i+1}^{-1}L_{i+1}^*,$$

where $L_{i+1} = F_{p,i}^{\Psi}L_i$ and $R_{r,i+1}$ is invertible and given by

$$R_{r,i+1} = R_{r,i} - L_i^*H_{i+1}^*R_{e,i}^{-1}H_{i+1}L_i.$$

11.7 **(A useful structured model)** Consider the state-space model (11.3.1) with $F_i = \lambda I$, a constant multiple of the identity, and where $\{H_i, G_i\}$ are defined as follows. The $\{H_i\}$ are row vectors such that starting with an initial vector $\{H_0\}$, the successive $\{H_i, i \geq 1\}$ are generated by shifting the entries of H_0 circularly to the right. For example, for $n = 3$, we would have

$$H_0 = \begin{bmatrix} a & b & c \end{bmatrix}, \quad H_1 = \begin{bmatrix} c & a & b \end{bmatrix}, \quad H_2 = \begin{bmatrix} b & c & a \end{bmatrix}, \quad H_3 = \begin{bmatrix} a & b & c \end{bmatrix}, \quad \ldots$$

Likewise, the $\{G_i\}$ are column vectors that are also generated by circularly shifting the entries of G_0 downwards. Show that the $\{F, G_i, H_i\}$ so defined satisfy a relation of the form (11.3.3) for some Ψ_i to be determined.

11.8 **(A structured covariance matrix)** Consider the structured state-space model (11.3.1)–(11.3.3) with constant $\{R, S, Q\}$ and let R_y denote the covariance matrix of the output process $\{\mathbf{y}_i\}$. Follow the discussion in Sec. 11.5 to show that the displacement of R_y is given by

$$\nabla_{Zp}R_y = \mathcal{G}\begin{bmatrix} I_p & 0 \\ 0 & J \end{bmatrix}\mathcal{G}^*,$$

where the signature matrix J is found from the factorization

$$P_1 - \Psi_0 P_0 \Psi_0 = \bar{L}_0 J \bar{L}_0^*,$$

and

$$\mathcal{G} = \begin{bmatrix} R_{e,0}^{1/2} & 0 \\ H_1\bar{K}_{p,0} & H_1\bar{L}_0 \\ H_2\Phi(2,1)\bar{K}_{p,0} & H_2\Phi(2,1)\bar{L}_0 \\ H_3\Phi(3,1)\bar{K}_{p,0} & H_3\Phi(3,1)\bar{L}_0 \\ \vdots & \vdots \end{bmatrix},$$

with

$$\Phi(i,1) \triangleq F_{i-1}F_{i-2}\ldots F_1, \quad i > 1.$$

11.9 **(Displacement ranks)** Consider the state-space model (11.1.1)–(11.1.2).

(a) Assume first that F is stable and Π_0 is the unique solution of $\bar{\Pi} = F\bar{\Pi}F^* + GQG^*$. Assume further that $H = 0$, $m < p \leq n$, and $\text{rank}(GS) = m$. Verify that $\text{rank}(P_1 - P_0) = m$ while $\text{rank}(\nabla_{Zp}R_y) = p$. [Recall that p and m are the sizes of $\mathbf{v}_i$ and $\mathbf{u}_i$, respectively. Moreover, R_y and $\nabla_{Zp}R_y$ are defined in Sec. 11.5.]

(b) Assume now $F = 0$, $H = 0$, and choose $P_0 \geq 0$ such that $\text{rank}(GQ^sG^* - P_0) = m$. Verify again that $\text{rank}(P_1 - P_0) = m$ while $\text{rank}(\nabla_{Zp}R_y) = p$. Recall that $Q^s = Q - SR^{-1}S^*$.

Remark. These examples show that in some degenerate cases, α and the displacement rank, r, of R_y can be totally independent of each other. Note however that we still have $r \leq \alpha + p$. ◆

11.10 (Constant Markov parameters) Refer to the notation in Sec. 11.3 and define the observability matrix

$$\mathcal{O}(i, i+n-1) \triangleq \begin{bmatrix} H_i\Phi(i,i) \\ H_{i+1}\Phi(i+1,i) \\ \vdots \\ H_{i+n-1}\Phi(i+n-1,i) \end{bmatrix},$$

as well as the controllability matrix

$$\mathcal{C}(i-1, i-n) \triangleq \begin{bmatrix} \Phi(i,i)G_{i-1} & \Phi(i,i-1)G_{i-2} & \dots & \Phi(i,i-n+1)G_{i-n} \end{bmatrix}.$$

(a) Assume that the Markov parameters Γ_{ij} are only functions of the time increment, say Γ_{i-j}. Verify that $\mathcal{O}(i,i+n-1)\mathcal{C}(i-1,i-n) = \mathcal{O}(i+1,i+n)\mathcal{C}(i,i-n+1)$.

(b) Assume that $\mathcal{C}(i-1,i-n)$ and $\mathcal{O}(i,i+n-1)$ are full rank for all i and define

$$\Psi_i = \mathcal{C}(i,i-n+1)\mathcal{C}^*(i-1,i-n)\left[\mathcal{C}(i-1,i-n)\mathcal{C}^*(i-1,i-n)\right]^{-1}.$$

Conclude from part (a) that $\mathcal{O}(i,i+n-1) = \mathcal{O}(i+1,i+n)\Psi_i$.

(c) From the first (block) row of the equality in part (b) conclude that $H_i = H_{i+1}\Psi_i$ for all i. Establish in a similar manner, by examining the other block rows, that it also holds that $\mathcal{O}(i+2,i+n+1)\Psi_{i+1}F_i = \mathcal{O}(i+2,i+n+1)F_{i+1}\Psi_i$. Conclude that we must have $\Psi_{i+1}F_i = F_{i+1}\Psi_i$.

(d) Repeat the above argument, but now working with the controllability matrix, to show that $\Psi_{i+1}G_i = G_{i+1}$.

(e) Parts (a)–(d) establish that for constant Markov parameters, and under certain full rank conditions, there should exist a Ψ_i that satisfies relations (11.3.3). Show that Ψ_i is unique.

CHAPTER 12

ARRAY ALGORITHMS

We derived in Sec. 9.2 the state estimator recursion of (9.2.7) in terms of the nonrandom functions $K_{p,i}$ and $R_{e,i}$ and of the given parameters $\{F_i, G_i, H_i, Q_i, R_i, S_i\}$. Then we showed that *one* way of calculating $\{K_{p,i}, R_{e,i}\}$ was in terms of the $n \times n$ prediction-error covariance matrix P_i, as described in Thm. 9.2.1.

In this chapter we shall consider another way of calculating the $\{K_{p,i}, R_{e,i}\}$, namely the so-called *array* methods. These methods were originally introduced as a way of alleviating some computational problems that are associated with the Riccati recursion of Thm. 9.2.1. However, as we shall see, such algorithms have several other advantages, in particular, reduced dynamic range for fixed-point implementations. As mentioned earlier, the largest amount of computation in the Kalman filter recursions arises in propagating the error covariance matrix P_i. However, more is at stake than the amount of computation. One consequence of round-off error is that the computed P_i may be non-Hermitian. This is sometimes compensated for by averaging the computed P_i and its Hermitian transpose. A better solution is to propagate only half the elements in P_i — say the ones on and below the main diagonal.

A more difficult consequence of round-off error arises from the fact that the P_i, being covariance matrices, have to be nonnegative definite. But round-off errors in the computation might destroy this property. Moreover, this is not an easy property to check — a matrix may be indefinite even if all its diagonal entries are nonnegative. The diagonal entries are the mean-square errors in the estimators of each of the components of the state vector and, of course, the computation would be seriously off if these diagonal entries turned out to be negative, which can also happen because of the build up of numerical errors.

It is therefore desirable to try to ensure that P_i is always nonnegative-definite. It turns out that an important step in this direction is to propagate not P_i but a *square-root* factor, *i.e.*, a matrix A_i such that $P_i = A_i A_i^*$. [The term square root is reserved for the case where A_i is Hermitian, *i.e.*, $A_i = A_i^*$, in which case $P_i = A_i^2$. In estimation theory, one generally uses *triangular* square-root factors.] There will be of course round-off errors in propagating A_i, just as for P_i, but the point is that the product of the computed factors, say $\hat{P}_i = \hat{A}_i \hat{A}_i^*$, is almost certainly nonnegative-definite. In theory, $\hat{A}_i \hat{A}_i^*$ is always nonnegative-definite, but of course again round-off effects may arise — however, they are much easier to control, and in fact, it is easy to see that the diagonal elements will never be negative.

Except in a single special case (see Sec. 12.2), the square-root factors can be propagated by algorithms of the following form:

1. We form a certain *pre-array* of numbers based on the given data at time i.
2. This array is reduced to a specified form (often triangular) by *a sequence of elementary unitary operations* (rotations or reflections).
3. The desired quantities at time $i+1$ can be immediately read off from the resulting so-called *post-array*.
4. No explicit equations are necessary.

Such array algorithms are often much simpler to describe and implement (in software or hardware) than explicit sets of equations; in fact, they are more like algorithms in the computer science sense of the word. They are becoming the method of choice in many applications. In this chapter, the main quantities being propagated are matrix square roots of covariance matrices and, in this context, the array algorithms were called square-root algorithms (see, *e.g.*, Dyer and McReynolds (1969), Kaminski, Bryson, and Schmidt (1971), and Morf and Kailath (1975)). In the next chapter, we shall present array forms of the fast algorithms of Ch. 11, and here the quantities of interest are (generalized) square-roots of increments of covariance matrices. There are of course also many array algorithms that propagate noncovariance related objects, *e.g.*, the Schur algorithm (see Kailath and Sayed (1995) and also App. 13.A). Actually, in Sec. 12.7, we present an algorithm that propagates something like a square-root factor of a square-root factor. Moreover, in Sec. 2.5, we have already presented an array algorithm for solving the classical deterministic least-squares problem. Therefore in this book, we have used the general term, array algorithms, rather than the earlier denomination as square-root algorithms.

Before proceeding, it will be useful to establish our notation and to review, for ease of reference, the relevant formulas of Ch. 9 for the filtering problem. Array algorithms for smoothing are presented in Sec. 12.9.

12.1 REVIEW AND NOTATIONS

Consider again the state-space equations

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{u}_i, \\ \mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i, \quad i \geq 0, \end{cases} \tag{12.1.1}$$

with $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$ zero-mean random variables such that

$$\left\langle \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_j \\ \mathbf{v}_j \\ \mathbf{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & S_i \delta_{ij} & 0 \\ S_i^* \delta_{ij} & R_i \delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \end{bmatrix},$$

and where the matrices $\{F_i, G_i, H_i, \Pi_0, Q_i, S_i, R_i\}$ are assumed known. Then the one-step predicted state estimator of $\mathbf{x}_i$, given $\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}$, $\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_{i|i-1}$, can be recursively computed via the equations: $\hat{\mathbf{x}}_0 = 0$,

$$\hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_i + K_{p,i} \mathbf{e}_i = F_{p,i} \hat{\mathbf{x}}_i + K_{p,i} \mathbf{y}_i, \quad i \geq 0, \tag{12.1.2}$$

where $\mathbf{e}_i = \mathbf{y}_i - H_i \hat{\mathbf{x}}_i$, $F_{p,i} = F_i - K_{p,i} H_i$,

$$K_{p,i} = K_i R_{e,i}^{-1}, \quad K_i = F_i P_i H_i^* + G_i S_i, \quad R_{e,i} = R_i + H_i P_i H_i^*, \tag{12.1.3}$$

and P_i could itself be computed via the Riccati difference equation (9.2.14),

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \quad P_0 = \Pi_0 \geq 0, \quad i \geq 0. \tag{12.1.4}$$

The nonnegative definiteness of the initial condition, $\Pi_0 \geq 0$, guarantees that all successive P_i will also be nonnegative definite.

Equivalently, the state-estimator $\hat{\mathbf{x}}_i$ can be updated in an alternative useful way that involves both time and measurement updates,

$$\hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_i + K_{f,i} \mathbf{e}_i, \tag{12.1.5}$$

$$\hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_{i|i} + G_i S_i R_{e,i}^{-1} \mathbf{e}_i, \tag{12.1.6}$$

where

$$K_{f,i} = P_i H_i^* R_{e,i}^{-1}.$$

The corresponding covariance matrices are updated via

$$P_{i|i} = P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i, \tag{12.1.7}$$

$$P_{i+1} = F_i P_{i|i} F_i^* + G_i (Q_i - S_i R_{e,i}^{-1} S_i^*) G_i^* - F_i K_{f,i} S_i^* G_i^* - G_i S_i K_{f,i}^* F_i^*. \tag{12.1.8}$$

We postpone discussion of the information form recursions to Sec. 12.8.

12.1.1 Notation

A "square-root factor" of a positive-definite matrix P is defined to be any $n \times n$ matrix, say A, such that

$$P = AA^*.$$

Square-root factors are not unique because if Θ is any unitary matrix, *i.e.*, if $\Theta\Theta^* = \Theta^*\Theta = I$, then clearly $A\Theta$ will also be a square-root factor of P. However, we need not worry about this because the nonuniqueness will disappear whenever we form the desired quantity P by "squaring."

Square-root factors can be made unique by imposing additional constraints, *e.g.*, that they be triangular (with positive diagonal entries) or Hermitian — see Prob. 12.1. In fact, the Hermitian factor is a "true" square root, because then

$$P = AA^* = A^2, \quad \text{and we can write } A = P^{1/2}.$$

In most applications, it is convenient to choose the square-root factor to be (lower) *triangular*, rather than Hermitian. However to reduce the notational and conceptual burden somewhat, we shall abuse the notation and use $P^{1/2}$ to denote *any square-root factor, not just the Hermitian one.* In this chapter, we shall always assume, unless otherwise stated, that $P^{1/2}$ refers to the unique lower triangular factor. We shall also employ the following notational conventions

$$P^{*/2} \triangleq (P^{1/2})^*, \quad P^{-1/2} \triangleq (P^{1/2})^{-1}, \quad P^{-*/2} \triangleq (P^{-1/2})^*,$$

so that we can write

$$P = P^{1/2}P^{*/2}, \qquad P^{-1} = P^{-*/2}P^{-1/2}.$$

Remark 1. If the matrix P is nonnegative-definite and, hence, rank-deficient, say of rank $\beta < n$, then a square-root factor of P is defined as any full rank $n \times \beta$ matrix A such that $P = AA^*$. In Prob. 12.3 we establish that any such nonnegative-definite matrix P also admits a unique factorization of the form $P = AA^*$, where A is $n \times n$ lower triangular with β positive diagonal entries and with $n - \beta$ identically zero columns, as demonstrated by the following example:

$$0 \le P = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 2 & 0 & \\ 0 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & & \\ 2 & 0 & \\ 0 & 0 & \sqrt{2} \end{bmatrix}^*.$$

In these rank-deficient cases, we shall continue to use $P^{1/2}$ to refer to this unique $n \times n$ lower triangular square-root factor. ♦

12.1.2 Normalizations

The availability of square-root factors enables various normalizations. For example, we can take the innovations process to be a unit variance process,

$$\bar{\mathbf{e}}_i \triangleq R_{e,i}^{-1/2}\mathbf{e}_i.$$

Correspondingly, the gain vectors in the Kalman filter recursions may be normalized. For example, the estimator measurement update (12.1.5) can be rewritten as

$$\hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_i + \bar{K}_{f,i}\bar{\mathbf{e}}_i,$$

where

$$\bar{K}_{f,i} = P_i H_i^* R_{e,i}^{-1} R_{e,i}^{1/2} = P_i H_i^* R_{e,i}^{-*/2}. \tag{12.1.9}$$

So also for the predictor update equation (12.1.2), we can write

$$\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_i + \bar{K}_{p,i}\bar{\mathbf{e}}_i,$$

where

$$\bar{K}_{p,i} = K_{p,i}R_{e,i}^{1/2} = F_i P_i H_i^* R_{e,i}^{-*/2}.$$

12.1.3 A Demonstration of Round-Off Error Effects

In order to further motivate the discussion in this chapter, consider the following contrived example:

$$F_i = 1, \quad H_i = 1, \quad G_i = 0, \quad R_i = 1, \quad S_i = 0.$$

In this case, the variable P_i is a scalar and the Riccati recursion (12.1.4) collapses to

$$P_{i+1} = P_i - \frac{P_i^2}{1 + P_i}, \tag{12.1.10}$$

which is also equivalent to

$$P_{i+1} = \frac{P_i}{1 + P_i}. \tag{12.1.11}$$

Now assume that at a particular time instant i, the value of P_i is sufficiently large that, in finite precision, $1 + P_i = P_i$. Assume further that the term $P_i^2/(1 + P_i)$ in (12.1.10) is implemented as follows:

$$\frac{P_i^2}{1 + P_i} = P_i \cdot \frac{P_i}{1 + P_i}.$$

Then the ratio $P_i/(1 + P_i)$ will evaluate to 1 and hence, if we compute P_{i+1} using (12.1.10) we obtain

$$P_{i+1} = P_i - P_i \cdot 1 = 0.$$

On the other hand, recursion (12.1.11) leads to $P_{i+1} = 1$.

The values for P_{i+1} are obviously different and the second one is in fact the correct value. We thus see that two different implementations of the same equation can behave very differently in the presence of round-off errors (or finite-precision effects). In the case of (12.1.10), the nonnegative quantity P_{i+1} is evaluated as the difference of two nearly equal large positive numbers, and this procedure leads in this example to an undesired cancellation. Such cancellations are often, but not always, bad phenomena in finite-precision implementations (see, *e.g.*, Higham (1996, p. 10)).

This example shows that there is merit in looking for alternative implementations of the Kalman filtering equations. In particular, for the above example we shall suggest, among others, the following array method for computing $\sqrt{P_{i+1}}$ from $\sqrt{P_i}$ (*cf.* (12.3.10)). We form the 2×2 pre-array of numbers

$$\mathcal{A} = \begin{bmatrix} 1 & \sqrt{P_i} \\ 0 & \sqrt{P_i} \end{bmatrix},$$

and then choose an elementary Givens rotation of the form (as explained in Sec. B.2 and also in Sec. 12.4 further ahead):

$$\Theta = \frac{1}{\sqrt{1 + \rho^2}} \begin{bmatrix} 1 & -\rho \\ \rho & 1 \end{bmatrix}, \qquad \rho = \sqrt{P_i}.$$

Since we are assuming in this example that P_i is large enough that $1 + P_i = P_i$ in finite precision, then $\sqrt{1+\rho^2}$ evaluates to $\sqrt{P_i}$. Moreover, the effect of Θ, when multiplied by $\mathcal{A}$ from the left, will be to null the (1, 2) entry of $\mathcal{A}$, *viz.*,

$$\begin{bmatrix} 1 & \sqrt{P_i} \\ 0 & \sqrt{P_i} \end{bmatrix} \Theta = \begin{bmatrix} \sqrt{P_i} & 0 \\ \sqrt{P_i} & 1 \end{bmatrix}.$$

It will be shown in (12.3.10) that the (2, 2) entry of the post-array above is equal to $\sqrt{P_{i+1}}$ and, hence, this method of implementation leads again to $P_{i+1} = 1$. [Note that the array method works with square roots of P_i rather than P_i itself, which is clearly an advantage when P_i assumes large values.]

We now move to a closer study of different array methods for filtering and smoothing.

12.2 POTTER'S EXPLICIT ALGORITHM FOR SCALAR MEASUREMENT UPDATE

We start by noting that it was actually for the measurement-update problem that the concept of a "square-root" algorithm was first introduced by Potter (see Potter and Stern (1963) and Battin (1964)). Potter worked directly with the measurement update equation (12.1.7), but for the special case $p = 1$, *i.e., scalar observations*. To emphasize this, we shall use the special notations

$$H_i \triangleq h_i, \text{ a row vector}, \quad R_i \triangleq r(i) \text{ and } R_{e,i} \triangleq r_e(i), \quad \text{both scalars.}$$

Then we can write (12.1.7) as

$$P_{i|i} = P_i^{1/2}(I - r_e^{-1}(i)a_i a_i^*)P_i^{*/2}, \tag{12.2.1}$$

where

$$a_i = P_i^{*/2}h_i^* \quad \text{and} \quad r_e(i) = r(i) + a_i^* a_i, \text{ a scalar.} \tag{12.2.2}$$

Now Potter noted that we can factor

$$I - r_e^{-1}(i)a_i a_i^* = (I - \gamma(i)a_i a_i^*)(I - \gamma(i)a_i a_i^*), \tag{12.2.3}$$

by choosing $\gamma(i)$ such that

$$-r_e^{-1}(i) = -2\gamma(i) + \gamma^2(i)(a_i^* a_i). \tag{12.2.4}$$

Then (12.2.1) will yield the updating formula

$$P_{i|i}^{1/2} = P_i^{1/2}(I - \gamma(i)a_i a_i^*). \tag{12.2.5}$$

In fact, Potter went a little further, by solving the quadratic equation (12.2.4) to get

$$\gamma(i) = \frac{1 \pm \sqrt{1 - r_e^{-1}(i)(a_i^* a_i)}}{(a_i^* a_i)} = \frac{1 \pm \sqrt{r(i)r_e^{-1}(i)}}{r_e(i) - r(i)} = \frac{1}{\sqrt{r_e(i)}(\sqrt{r_e(i)} \pm \sqrt{r(i)})}.$$

Potter chose the + sign so as to divide by a larger quantity, thus finally obtaining the recursion

$$P_{i|i}^{1/2} = P_i^{1/2}\left[I - \frac{a_i a_i^*}{\sqrt{r_e(i)}(\sqrt{r_e(i)} + \sqrt{r(i)})}\right], \tag{12.2.6}$$

where $r_e(i)$ and a_i are as defined in (12.2.2).

This nice formula also has an extension to the case of vector observations (*i.e.*, $p > 1$) — see Prob. 12.5, but it has two limitations:

(i) There is no simple analogous formula for the *time-update* step of going from $P_{i|i}^{1/2}$ to $P_{i+1}^{1/2}$. [Dyer and McReynolds (1969) do give a formula for doing this, but it requires knowledge of $P_{i|i}^{-1/2}$ — see Prob. 12.6.]

(ii) Even if $P_i^{1/2}$ is triangular or Hermitian, the same will not be generally true of $P_{i|i}^{1/2}$ as computed via (12.2.6). (Why? See Prob. 12.4). So we will need more storage and more computation in the next step. Carlson (1973) presented a variation of Potter's update formula (12.2.6) that gave the square-root factor $P_{i|i}^{1/2}$ in a triangular form. He did so by requiring $P_i^{1/2}$ to be in triangular form and by computing a triangular square root for the matrix $I - r_e^{-1}(i)a_i a_i^*$ in (12.2.3), rather than using the factor $(I - \gamma(i)a_i a_i^*)$.[1]

It turns out (see Sec. 12.3.3) that there are more direct updating methods for going from a triangular $P_i^{1/2}$ to a triangular $P_{i|i}^{1/2}$, even in the vector case ($p > 1$), but these cannot be expressed in as explicit a form as (12.2.6). Once we free ourselves from demanding explicit equations, we can also solve the time-updating problem.

12.3 SEVERAL ARRAY ALGORITHMS

In fact, the time-update problem is simpler for this purpose, so we consider it first. Then we shall go to the somewhat less obvious case of measurement update, followed by updates for the predicted and filtered estimators.

12.3.1 A Standing Assumption

In this chapter we also make, unless otherwise specified, the standing assumption that

$$S_i = 0 \quad \text{and} \quad R_i > 0.$$

We recall from Sec. 9.5.1 that when $R_i > 0$, a circumstance to be favored in setting up the state-space model, nonzero S_i can be accommodated by replacing $\{F_i, Q_i\}$ by

$$F_i^s = F_i - G_i S_i R_i^{-1} H_i \quad \text{and} \quad Q_i^s = Q_i - S_i R_i^{-1} S_i^*.$$

We also refer the reader to Prob. 12.11 and to Secs. 12.6 and 12.8.4.

[1] Let W denote a triangular square-root factor of $I - r_e^{-1}(i)a_i a_i^*$. The entries of W were determined in Carlson (1973) by comparing the entries on both sides of the identity $I - r_e^{-1}(i)a_i a_i^* = WW^*$ — see Grewal and Andrews (1993, pp. 237–239). The formulas are somewhat involved and have been dropped in favor of the array algorithms.

12.3.2 Time Updates

The time-update equation (12.1.8), assuming $S_i = 0$, is

$$P_{i+1} = F_i P_{i|i} F_i^* + G_i Q_i G_i^*, \quad i \geq 0. \tag{12.3.1}$$

Here the furthest we can get with Potter's line of argument is the expression

$$P_{i+1} = \left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]\left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]^*.$$

This gives a factorization of P_{i+1}, but unfortunately the dimensions of

$$\left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]$$

are too large, $n \times (n+m)$ rather than $n \times n$. However, here we could take advantage of the nonuniqueness of square-root factors and introduce a unitary matrix Θ,

$$P_{i+1} = \left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]\Theta\Theta^*\left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]^*,$$

and try to choose Θ so that

$$\left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]\Theta = \left[X \;\; 0_{n\times m}\right], \tag{12.3.2}$$

where $0_{n\times m}$ denotes an $n \times m$ matrix of all zero elements and X denotes a presently undetermined $n \times n$ matrix. If we can find such a Θ, then it must hold by *squaring* that

$$\left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]\underbrace{\Theta\Theta^*}_{I}\left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]^* = \left[X \;\; 0_{n\times m}\right]\left[X \;\; 0_{n\times m}\right]^*,$$

and, hence,

$$F_i P_{i|i} F_i^* + G_i Q_i G_i^* = XX^*.$$

But since the left-hand side is equal to P_{i+1}, X can be identified as $P_{i+1}^{1/2}$, a square-root factor of P_{i+1}. So we have the following algorithm. Form a so-called *pre-array*

$$\mathcal{A}_1 = \left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right], \tag{12.3.3}$$

and unitarily (block) triangularize it to yield a *post-array* of the form

$$\mathcal{A}_1\Theta = \left[X \;\; 0_{n\times m}\right], \quad \Theta\Theta^* = I_{n+m} = \Theta^*\Theta.$$

We can identify X as a square root of P_{i+1}. Uniqueness could be ensured by assuming that X is, say, lower triangular with nonnegative diagonal entries.

In summary, the array algorithm for the time-update problem (12.1.8) takes the following form (assuming $S_i = 0$):

$$\left[F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2}\right]\Theta = \left[P_{i+1}^{1/2} \;\; 0\right], \tag{12.3.4}$$

where Θ is any unitary matrix that (lower) triangularizes the pre-array.[2]

This solves the time-updating problem, provided we can show how to find an appropriate Θ. In fact, Θ can be found in several ways via well-known methods in numerical linear algebra. In fact, it follows from (12.3.3) and (12.3.4) that

$$\mathcal{A}_1^* = \Theta\begin{bmatrix} P_{i+1}^{*/2} \\ 0 \end{bmatrix},$$

which shows that the triangularization (12.3.4) can be accomplished by performing the so-called QR decomposition (Q unitary, R upper triangular) of the matrix $\mathcal{A}_1^*$ (*cf.* App. A). This is best done by using a sequence of elementary unitary transformations as described in App. B (see also the examples in Sec. 12.4). The point to note here is that each of the elementary transformations is such that no explicit formulas are required. While each operation is also easy to describe explicitly, getting an explicit formula for the resulting triangular factor $P_{i+1}^{1/2}$ would be cumbersome and unrewarding. We may also note that many software packages already include ready-to-use routines that perform QR decompositions so that, for all practical purposes, transformations of the form (12.3.4) are straightforward to implement. [We further remark that, as explained earlier, if P_{i+1} is rank-deficient, then some of the rows of the upper triangular factor $P_{i+1}^{*/2}$ will be identically zero.]

12.3.3 Measurement Updates

We wish to go from $P_i^{1/2}$ to $P_{i|i}^{1/2}$ in accordance with the measurement-update equation

$$P_{i|i} = P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i. \tag{12.3.5}$$

If there were a plus sign in (12.3.5) rather than a minus sign, we could readily use the technique of the previous subsection. Because of the minus sign, we would need to employ a so-called J-unitary transformation (see Prob. 12.10 and also App. B). However, it turns out that in this particular case, there is an alternative formulation, first suggested by Kaminski, Bryson, and Schmidt (1971), that uses only unitary transformations. We just present it here, deferring the motivation to Sec. 12.5.2.

[2] The matrix Θ is of course time-dependent and, for completeness, we should have written Θ_i in (12.3.4) in order to emphasize the fact that in general different Θ's are needed for different time instants. We shall however continue to write Θ for simplicity of notation. The algorithm in (12.3.4) was first described by Schmidt (1970), who used the Gram-Schmidt construction for triangularizing the pre-array; in those early days it was not immediate that any other unitary triangularization method would do as well.

The method is to start with the pre-array

$$\mathcal{A}_2 = \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix},$$

and then to triangularize it via a unitary transformation Θ:

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}. \tag{12.3.6}$$

The entries $\{X, Y, Z\}$ in the post-array can be identified by squaring both sides:

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix} \underbrace{\Theta\Theta^*}_{I} \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix}^* = \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix} \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}^*, \tag{12.3.7}$$

to obtain the equalities

$$\begin{aligned} XX^* &= R_i^{1/2} R_i^{*/2} + H_i P_i^{1/2} P_i^{*/2} H_i^* = R_i + H_i P_i H_i^* = R_{e,i}, \\ YX^* &= P_i^{1/2} P_i^{*/2} H_i^* = P_i H_i^*, \\ ZZ^* &= P_i^{1/2} P_i^{*/2} - YY^* = P_i - P_i H_i^* (X^{-*} X^{-1}) H_i P_i, \\ &= P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i = P_{i|i}. \end{aligned}$$

Therefore, as claimed, we can identify Z as a square-root factor of $P_{i|i}$, and similarly identify X and Y as $X = R_{e,i}^{1/2}$ and $Y = \bar{K}_{f,i} = P_i H_i^* R_{e,i}^{-*/2}$. Hence, the array algorithm for the measurement-update problem (12.1.7) takes the following form:

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ \bar{K}_{f,i} & P_{i|i}^{1/2} \end{bmatrix}, \tag{12.3.8}$$

where Θ is any unitary matrix that (lower) triangularizes the pre-array. As we shall show via several explicit numerical examples in Sec. 12.4, there are many possible choices of Θ.

We may also remark that, even in the scalar case, where we have Potter's and Carlson's explicit equations, the present algorithm appears to be at least conceptually simpler. Moreover, by insisting on a full (rather than block) triangularization of the pre-array, we can achieve a triangular $P_{i|i}^{1/2}$ for both the scalar and vector cases, which is not true of Potter's approach (12.2.6) — see also Prob. 12.4.

Relation to Potter's Explicit Square-Root Update. Of course, one would expect that a particular choice of Θ in (12.3.8) would give us Potter's equation (12.2.6). In fact, some algebra, first carried out by Bierman (1973a), shows that we can rearrange Potter's formulas as

$$\begin{bmatrix} \sqrt{r(i)} & h_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} \sqrt{r_e(i)} & 0 \\ \bar{K}_{f,i} & P_{i|i}^{1/2} \end{bmatrix},$$

where Θ is the so-called Householder matrix, $\Theta = I - 2w(w^*w)^{-1}w^*$, with

$$w^* = \begin{bmatrix} (\sqrt{r_e(i)} - \sqrt{r(i)}) & -P_i^{*/2} h_i^* \end{bmatrix}.$$

The (Householder) matrix Θ has the property that it rotates the first row of the pre-array to lie along the direction $\begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}$. This block triangularization of the pre-array does not of course imply that the (2, 2) block in the post-array is triangular; to achieve this, we would have to apply further unitary operations. Clearly, as mentioned above, the specification of these operations would be algebraically rather complex, which is why explicit expressions are not given for obtaining triangular $P_{i|i}^{1/2}$.

12.3.4 Predicted Estimators

By *combining* the measurement- and time-update steps of the earlier sections (as indicated in Prob. 12.7, which the reader is well advised to attempt), or by other more direct methods (see Sec. 12.5.3), we can obtain the following algorithm (when $S_i = 0$).

Form the pre-array

$$\mathcal{A}_3 = \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix},$$

and triangularize it via a unitary transformation Θ to get

$$\mathcal{A}_3 \Theta = \begin{bmatrix} X & 0 & 0 \\ Y & Z & 0 \end{bmatrix}, \quad \text{say.} \tag{12.3.9}$$

We can now identify the $\{X, Y, Z\}$ by "squaring up." However, while this procedure is quite simple and straightforward, there is a slightly different alternative approach that will be quite helpful on some later occasions (*e.g.*, Sec. 12.8.5).

For this, note that (12.3.9) may be regarded as saying that the rotation Θ sets up a conformal (*i.e.*, a norm- and angle-preserving) mapping between the (block) rows of the pre-array and the (block) rows of the post-array. So for example,

$$\langle \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \end{bmatrix}, \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \end{bmatrix} \rangle = \langle \begin{bmatrix} X & 0 & 0 \end{bmatrix}, \begin{bmatrix} X & 0 & 0 \end{bmatrix} \rangle,$$

$$\langle \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \end{bmatrix}, \begin{bmatrix} 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \rangle = \langle \begin{bmatrix} X & 0 & 0 \end{bmatrix}, \begin{bmatrix} Y & Z & 0 \end{bmatrix} \rangle,$$

where, for two row vectors a and b, the notation $\langle a, b \rangle$ stands for the inner product ab^*.

From this we can see that X can be identified by equating the norms of the first rows on both sides of (12.3.9):

$$R_i^{1/2} R_i^{*/2} + H_i P_i^{1/2} P_i^{*/2} H_i^* = R_{e,i} = XX^*,$$

so that $X = R_{e,i}^{1/2}$. Then Y can be found by equating the inner products of the first and second rows:

$$R_i^{1/2} \cdot 0 + H_i P_i^{1/2} \cdot P_i^{*/2} F_i^* + 0 \cdot Q_i^{*/2} G_i^* = K_i^* = XY^*.$$

Hence, $Y = K_i R_{e,i}^{-*/2} \triangleq \bar{K}_{p,i}$. Finally, Z can be identified by equating the norms of the second rows,

$$ZZ^* = F_i P_i F_i^* + G_i Q_i G_i^* - YY^* = P_{i+1},$$

which implies that $Z = P_{i+1}^{1/2}$.

Therefore, the array algorithm for the update of the prediction problem (12.1.3)–(12.1.4), assuming $S_i = 0$, takes the following form:

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \end{bmatrix}, \tag{12.3.10}$$

where Θ is any unitary matrix that (lower) triangularizes the pre-array.

12.3.5 Filtered Estimators

By combining the time and measurement steps (see Prob. 12.8), or by other more direct methods, we can also obtain the following algorithm.

Form the pre-array

$$\mathcal{A}_3 = \begin{bmatrix} R_{i+1}^{1/2} & H_{i+1} F_i P_{i|i}^{1/2} & H_{i+1} G_i Q_i^{1/2} \\ 0 & F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix},$$

and triangularize it via a unitary transformation Θ to get

$$\mathcal{A}_3 \Theta = \begin{bmatrix} X & 0 & 0 \\ Y & Z & 0 \end{bmatrix}, \quad \text{say.} \tag{12.3.11}$$

The entries in the post-array can be identified as above, *e.g.*, by "squaring" both sides of (12.3.11) or by forming appropriate norms and inner products. This leads to the following array form for the filtering problem (when $S_i = 0$):

$$\begin{bmatrix} R_{i+1}^{1/2} & H_{i+1} F_i P_{i|i}^{1/2} & H_{i+1} G_i Q_i^{1/2} \\ 0 & F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 & 0 \\ \bar{K}_{f,i+1} & P_{i+1|i+1}^{1/2} & 0 \end{bmatrix}.$$

12.3.6 Estimator Update

Returning to the array algorithm (12.3.10), we see that though numerically reliable unitary operations are used to propagate $P_{i+1}^{1/2}$, and to compute the $R_{e,i}^{1/2}$ and $\bar{K}_{p,i}$, the state estimators are still to be obtained as (*cf.* (12.1.2))

$$\hat{\mathbf{x}}_{i+1} = (F_i)(\hat{\mathbf{x}}_i) + (\bar{K}_{p,i})(R_{e,i}^{1/2})^{-1}(\mathbf{y}_i - H_i \hat{\mathbf{x}}_i). \tag{12.3.12}$$

An alternative solution to this problem is to extend the pre-array by an additional row and to apply the same unitary rotation as in (12.3.10) to obtain the following array form:

$$\left[\begin{array}{ccc} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \\ \hline -\mathbf{y}_i^* R_i^{-*/2} & \hat{\mathbf{x}}_i^* P_i^{-*/2} & 0 \end{array}\right] \Theta = \left[\begin{array}{ccc} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \\ \hline \boldsymbol{\alpha} & \boldsymbol{\beta} & \boldsymbol{\gamma} \end{array}\right]. \tag{12.3.13}$$

We can identity $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ as follows. First equate the inner products of the first and second block rows of the pre- and post-arrays in (12.3.13) to get

$$R_i^{1/2}(-R_i^{-1/2}\mathbf{y}_i) + H_i P_i^{1/2} P_i^{-1/2} \hat{\mathbf{x}}_i + 0 = R_{e,i}^{1/2} \boldsymbol{\alpha}^*,$$

or, equivalently,

$$\boldsymbol{\alpha}^* = -R_{e,i}^{-1/2}(\mathbf{y}_i - H_i \hat{\mathbf{x}}_i) = -R_{e,i}^{-1/2} \mathbf{e}_i = -\bar{\mathbf{e}}_i.$$

Now equate the inner products of the second and third block rows to get

$$0 + F_i P_i^{1/2} P_i^{-1/2} \hat{\mathbf{x}}_i + 0 = \bar{K}_{p,i} \boldsymbol{\alpha}^* + P_{i+1}^{1/2} \boldsymbol{\beta}^*,$$

or

$$\boldsymbol{\beta}^* = P_{i+1}^{-1/2}(F_i \hat{\mathbf{x}}_i + \bar{K}_{p,i} \bar{\mathbf{e}}_i) = P_{i+1}^{-1/2} \hat{\mathbf{x}}_{i+1}.$$

An expression for $\boldsymbol{\gamma}$ can be found by equating the norms of the third block rows in (12.3.13), which yields after some algebra

$$\boldsymbol{\gamma} = -\hat{\mathbf{x}}_{i+1}^* K_{b,i}^* Q_i^{-*r/2},$$

with

$$K_{b,i} = Q_i G_i P_{i+1}^{-1} \quad \text{and} \quad Q_i^r = Q_i - Q_i G_i^* P_{i+1}^{-1} G_i Q_i. \tag{12.3.14}$$

The apparently complicated quantities $\{\boldsymbol{\gamma}, K_{b,i}, Q_i^r\}$ will be useful, later in this chapter, in finding array forms for the so-called Rauch-Tung-Striebel (RTS) smoothing algorithm (see Sec. 12.9), and also in the derivation of an array form for the information filter of the time-update problem (*cf.* (12.8.7)). However, the point of the algorithm in (12.3.13) is that the estimator $\hat{\mathbf{x}}_{i+1}$ can now be constructed as

$$\hat{\mathbf{x}}_{i+1} = (P_{i+1}^{1/2}) \boldsymbol{\beta}^*.$$

Since the state estimators are now found as products of quantities that are available from the post-array, this algorithm is more amenable to parallel implementation than the array algorithm (12.3.10), where the state estimator is further evaluated via (12.3.12). [Note also that in (12.3.12) we have a $p \times p$ matrix inversion, albeit of a triangular matrix $R_{e,i}^{1/2}$ (so that the inversion step can be replaced by solving linear equations by backwards substitution).] We may also note that if our major interest is in updating the estimators, then some computation may be reduced in the extended arrays by not computing the entry $\bar{K}_{p,i}$ in the post-array.

Remark 2. The importance of proper modeling in which we seek to properly scale the R_i to be as near identity matrices as possible is also seen from the above algorithm, where the only inverses required are of the matrices $R_i^{1/2}$. ◆

Of course, we have left an air of mystery around all these striking results/arrays by not fully describing their origins. This will be done in Secs. 12.5 and 12.6.

12.3.7 Operation Counts and Condensed Forms

The number of operations needed in going from step i to step $(i+1)$ in the estimator update (12.3.13) is $O(n^3)$, the same order as the Riccati-based algorithm. In general, though, the actual number of computations in the array method would tend to be somewhat larger than in the direct Riccati equation method. However, there are of course important compensatory numerical advantages. Moreover, it appears (see Bierman and Thornton (1977)), that with proper programming, the computational efforts can be made essentially the same. Verhaegen and Van Dooren (1986) have also noted that it can be useful to first transform the given model parameters $\{F_i, G_i, H_i\}$ by unitary operations to so-called condensed forms, which help reduce the operations count further – see App. 12.B.

12.4 NUMERICAL EXAMPLES

Consider a 2-state 1-input system with the following values for some quantities of interest at time i:

$$F_i = \begin{bmatrix} 0.8 & 0.3 \\ 0.5 & 0.7 \end{bmatrix}, \quad G_i = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}, \quad Q_i = 1.0, \quad P_{i|i}^{1/2} = \begin{bmatrix} 1 & 0 \\ 0.25 & 0.5 \end{bmatrix}.$$

Then P_{i+1} can be computed directly from the formula

$$P_{i+1} = F_i P_{i|i}^{1/2} P_{i|i}^{*/2} F_i^* + G_i Q_i G_i^* = \begin{bmatrix} 1.7881 & 1.1431 \\ 1.1431 & 0.8281 \end{bmatrix}.$$

Alternatively, we can employ an array algorithm to compute $P_{i+1}^{1/2}$. For this purpose, we form the pre-array,

$$\begin{bmatrix} F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix} = \begin{bmatrix} 0.875 & 0.15 & 1.0 \\ 0.675 & 0.35 & 0.5 \end{bmatrix} \tag{12.4.1}$$

and proceed to triangularize it. This will be illustrated here by employing Givens rotations, Householder transformations, and square-root free rotations. We urge the reader to review their descriptions as given in App. B, because the following explanations are somewhat abbreviated.

12.4.1 Triangularization via Givens Rotations

We start by annihilating the $(1,3)$ entry of the pre-array (12.4.1) by pivoting with its $(1,1)$ entry. According to expression (B.2.2), the orthogonal transformation Θ_1 that achieves this result is given by

$$\Theta_1 = \begin{bmatrix} 0.6585 & -0.7526 \\ 0.7526 & 0.6585 \end{bmatrix}, \quad \text{where we used} \quad \rho_1 = \frac{1}{0.875}.$$

Applying Θ_1 to the pre-array (12.4.1) leads to (recall that we are only operating on the first and third columns, leaving the second column unchanged)

$$\begin{bmatrix} 0.875 & 0.15 & 1 \\ 0.675 & 0.35 & 0.5 \end{bmatrix} \begin{bmatrix} 0.6585 & 0 & -0.7526 \\ 0 & 1 & 0 \\ 0.7526 & 0 & 0.6585 \end{bmatrix} = \begin{bmatrix} 1.3288 & 0.1500 & 0.0000 \\ 0.8208 & 0.3500 & -0.1788 \end{bmatrix}.$$

We now annihilate the $(1,2)$ entry of the resulting matrix in the above equation by pivoting with its $(1,1)$ entry. This requires that we choose

$$\Theta_2 = \begin{bmatrix} 0.9937 & -0.1122 \\ 0.1122 & 0.9937 \end{bmatrix}, \quad \text{where we used} \quad \rho_2 = \frac{0.1500}{1.3288}.$$

Applying Θ_2 to the post-array of the first step leads to (now we leave the third column unchanged)

$$\begin{bmatrix} 1.3288 & 0.1500 & 0.0000 \\ 0.8208 & 0.3500 & 0.1788 \end{bmatrix} \begin{bmatrix} 0.9937 & -0.1122 & 0 \\ 0.1122 & 0.9937 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.3373 & 0.0000 & 0.0000 \\ 0.8549 & 0.2557 & 0.1788 \end{bmatrix}.$$

We finally annihilate the $(2,3)$ entry of the resulting post-array by pivoting with its $(2,2)$ entry. In principle this requires that we choose

$$\Theta_3 = \begin{bmatrix} 0.8195 & 0.5731 \\ -0.5731 & 0.8195 \end{bmatrix}, \quad \text{using} \quad \rho_3 = \frac{0.1788}{-0.2557},$$

and apply it to the post-array, which would then lead to

$$\begin{bmatrix} 1.3373 & 0.0000 & 0.0000 \\ 0.8549 & -0.2557 & 0.1788 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.8195 & 0.5731 \\ 0 & -0.5731 & 0.8195 \end{bmatrix} = \begin{bmatrix} 1.3373 & 0.0000 & 0.0000 \\ 0.8549 & -0.3120 & 0.0000 \end{bmatrix}.$$

Alternatively, this last step can be implemented without explicitly forming Θ_3. We simply replace the row vector $[-0.2557 \;\; 0.1788]$, which contains the $(2,2)$ and $(2,3)$ entries of the pre-array in the above equation, by the row vector

$$[\pm\sqrt{(-0.2557)^2 + (0.1788)^2} \;\; 0.0000] = [\pm 0.3120 \;\; 0.0000].$$

We choose the positive sign in order to conform with our earlier convention that the diagonal entries of triangular square-roots factors should be taken to be positive. The resulting post-array is therefore

$$\begin{bmatrix} 1.3373 & 0.0000 & 0.0000 \\ 0.8549 & 0.3120 & 0.0000 \end{bmatrix}.$$

We have thus exhibited a sequence of elementary orthogonal transformations that triangularizes the pre-array of numbers (12.4.1). In view of (12.3.4), we see that we can identify

$$P_{i+1}^{1/2} = \begin{bmatrix} 1.3373 & 0.0000 \\ 0.8549 & 0.3120 \end{bmatrix}. \tag{12.4.2}$$

Remark 3. We may also remark that the combined effect of the sequence of transformations $\{\Theta_1, \Theta_2, \Theta_3\}$ corresponds to the orthogonal rotation Θ required in (12.3.4). However, we do not need to know or to form $\Theta = \Theta_1\Theta_2\Theta_3$. In fact, we can avoid explicit introduction of the Givens matrices Θ_i and speak only of appropriate column operations with coefficients

$$\left\{ \frac{1}{\sqrt{1+\rho_i^2}}, \pm\frac{\rho_i}{\sqrt{1+\rho_i^2}} \right\}.$$

This is a somewhat pedantic point when using Givens rotations. It is much less so when we use Householder transformations. ♦

12.4.2 Triangularization via Householder Transformations

Again the reader will be well served by reviewing App. B. We again start with the pre-array of numbers (12.4.1) and annihilate the (1, 2) and (1, 3) entries by using the (1, 1) entry as a pivot. According to expression (B.1.4), the 3×3 Householder transformation Θ_1 that achieves this for real-valued data is given by

$$\Theta_1 = I_3 - 2\,\frac{g_1^T g_1}{g_1 g_1^T} \qquad \text{where} \qquad g_1 = x_1 \pm \|x_1\| \begin{bmatrix} 1 & 0 & 0 \end{bmatrix},$$

and x_1 is the first row of the array (12.4.1), $x_1 = \begin{bmatrix} 0.875 & 0.15 & 1.0 \end{bmatrix}$. The Euclidean norm of x_1 is $\|x_1\| = 1.3372$. We choose the sign in the expression for g_1 to be the same as the sign of the leading entry of x_1, which is positive. Therefore,

$$g_1 = \begin{bmatrix} 0.875 & 0.15 & 1.0 \end{bmatrix} + \begin{bmatrix} 1.3372 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2.2122 & 0.1500 & 1.0000 \end{bmatrix}.$$

Note that we do not need to explicitly form Θ_1. Instead, we can just operate on the pre-array (12.4.1) row by row, say

$$x_i\Theta = x_i - 2\frac{x_i g_1^T}{g_1 g_1^T} g_1,$$

for the i-th row x_i. Only the first row x_1, and any proportional to it, will be replaced by one of the form $\begin{bmatrix} \bullet & 0 & 0 \end{bmatrix}$.

Operating on both rows of the pre-array (12.4.1) we obtain

$$\begin{bmatrix} 0.875 & 0.15 & 1.0 \\ 0.675 & 0.35 & 0.5 \end{bmatrix} \Theta_1 = \begin{bmatrix} -1.3372 & 0.0000 & 0.0000 \\ 0.8549 & 0.2463 & -0.1916 \end{bmatrix}. \tag{12.4.3}$$

We now annihilate the (2, 3) entry of the matrix appearing on the right-hand side of (12.4.3), by pivoting with its (2, 2) entry. In principle, the 2×2 Householder transformation that achieves this is given by

$$\Theta_2 = I_2 - 2\,\frac{g_2^T g_2}{g_2 g_2^T} \qquad \text{where} \qquad g_2 = x_2 \pm \|x_2\| \begin{bmatrix} 1 & 0 \end{bmatrix},$$

where x_2 and g_2 are given by

$$x_2 = \begin{bmatrix} 0.2463 & -0.1916 \end{bmatrix},$$

$$g_2 = \begin{bmatrix} 0.2463 & -0.1916 \end{bmatrix} + \begin{bmatrix} 0.3120 & 0 \end{bmatrix} = \begin{bmatrix} 0.5583 & -0.1916 \end{bmatrix}.$$

We can then operate on the rows of the matrix on the right-hand side of (12.4.3) with Θ_2 to obtain

$$\begin{bmatrix} -1.3372 & 0.0000 & 0.0000 \\ 0.8549 & 0.2463 & -0.1916 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \Theta_2 \end{bmatrix} = \begin{bmatrix} -1.3372 & 0.0000 & 0.0000 \\ -0.8549 & -0.3120 & 0.0000 \end{bmatrix}.$$

Alternatively, this last step can be implemented without explicitly forming Θ_2. We simply replace the row vector $\begin{bmatrix} 0.2463 & -0.1916 \end{bmatrix}$, which contains the (2, 2) and (2, 3) entries of the pre-array in the above equation, by the row vector

$$\begin{bmatrix} \pm\sqrt{(0.2463)^2 + (-0.1916)^2} & 0.0000 \end{bmatrix} = \begin{bmatrix} \pm 0.31200 & 0.0000 \end{bmatrix}.$$

We again choose the positive sign in order to conform with our earlier convention that the diagonal entries of triangular square-root factors should be taken to be positive. The resulting post-array is therefore

$$\begin{bmatrix} -1.3372 & 0.0000 & 0.0000 \\ -0.8549 & 0.3120 & 0.0000 \end{bmatrix}.$$

We may also take the sign of the (1, 1) entry of the above square-root factor to be positive. Then, in view of (12.3.4), we see that we can identify

$$P_{i+1}^{1/2} = \begin{bmatrix} 1.3373 & 0.0000 \\ -0.8549 & 0.3120 \end{bmatrix}. \tag{12.4.4}$$

12.4.3 Triangularization via Square-Root Free Rotations

The reader should review the material on modified Givens transformations in App. B at this point. We start with the pre-array of numbers (12.4.1) and annihilate its (1, 3) entry by pivoting with its (1, 1) entry.

We have $p_1 = 0.8750$, $p_2 = 1$ and start by choosing $D_{p1} = 1$ and $D_{p2} = 1$. Hence, using (B.3.3) and (B.3.4),

$$D_{q1} = 0.7656 + 1 = 1.7656, \qquad D_{q2} = 1/1.7656 = 0.5664.$$

The transformation matrix (B.3.5) becomes

$$D_p^{1/2}\Theta D_q^{-1/2} = \begin{bmatrix} 0.4956 & -1 \\ 0.5664 & 0.8750 \end{bmatrix}.$$

Applying this transformation to the pre-array (12.4.1) leads to

$$\begin{bmatrix} 0.8750 & 0.15 & 1.0 \\ 0.6750 & 0.35 & 0.5 \end{bmatrix} \begin{bmatrix} 0.4956 & 0 & -1 \\ 0 & 1 & 0 \\ 0.5664 & 0 & 0.8750 \end{bmatrix} = \begin{bmatrix} 1.0000 & 0.15 & 0.0000 \\ 0.6177 & 0.35 & -0.2375 \end{bmatrix}.$$

We now proceed to annihilate the $(1, 2)$ entry of the resulting postarray by pivoting with its $(1, 1)$ entry. In this case we have $p_1 = 1.000$, $p_2 = 0.15$ and we choose $D_{p1} = 1.7656$ and $D_{p2} = 1$. The choice for D_{p1} is the value obtained earlier for D_{q1}, which is the weight associated with the entry that we are now employing as p_1. Hence,

$$D_{q1} = 1.7656 + (0.15)^2 = 1.7881, \qquad D_{q2} = 1.7656/1.7881 = 0.9874.$$

Applying the transformation (B.3.5) to the above postarray leads to

$$\begin{bmatrix} 1.0000 & 0.15 & 0.0000 \\ 0.6177 & 0.35 & -0.2375 \end{bmatrix} \begin{bmatrix} 0.9874 & -0.15 & 0 \\ 0.0839 & 1.00 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.6393 & 0.2573 & -0.2375 \end{bmatrix}.$$

We finally annihilate the $(2, 3)$ entry of the postarray in the above equation by pivoting with its $(2, 2)$ entry. We now have $p_1 = 0.2573$, $p_2 = -0.2375$ and choose $D_{p1} = 0.9874$ and $D_{p2} = 0.5664$. The values for D_{p1} and D_{p2} are the weights associated with the entries that we are now employing as p_1 and p_2. Hence, $D_{q1} = 0.0973$ and $D_{q2} = 5.7478$.

Applying the transformation (B.3.5) leads to

$$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.6393 & 0.2573 & -0.2375 \end{bmatrix} \begin{bmatrix} 1 & 0.0000 & 0.0000 \\ 0 & 2.6111 & 0.2375 \\ 0 & -1.3825 & 0.2573 \end{bmatrix} = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.6393 & 1.0000 & 0.0000 \end{bmatrix}.$$

The resulting right-hand side is the desired normalized post-array. Its columns are respectively weighted by the numbers $(1.7881)^{1/2}$, $(0.0973)^{1/2}$, $(5.7478)^{1/2}$. This implies that we can identify

$$P_{i+1}^{1/2} = \begin{bmatrix} 1.0000 & 0.0000 \\ 0.6393 & 1.0000 \end{bmatrix} \begin{bmatrix} (1.7881)^{1/2} & 0 \\ 0 & (0.0973)^{1/2} \end{bmatrix} = \begin{bmatrix} 1.3772 & 0.0000 \\ 0.8549 & 0.3119 \end{bmatrix}.$$

12.5 DERIVATIONS OF THE ARRAY ALGORITHMS

We have indicated several array algorithms in Sec. 12.3. While it was easy to deduce the entries of the post-array once the pre-array was specified, the question of course is how one would know the form of the pre-array. Some insight into different ways of deducing this will be presented in this section.

A key result in this regard is Lemma A.5.1 (in App. A), which states that for two $n \times m$ $(n \leq m)$ matrices A and B, the equality $AA^* = BB^*$ holds if, and only if, there exists an $m \times m$ unitary matrix Θ $(\Theta\Theta^* = I = \Theta^*\Theta)$ such that $A = B\Theta$.

12.5.1 The Time-Update Algorithm

We return to the time-update equation (12.1.8) for the error covariance matrix, with $S_i = 0$,

$$P_{i+1} = F_i P_{i|i} F_i^* + G_i Q_i G_i^*, \tag{12.5.1}$$

and note that we can factor both the left- and the right-hand sides as follows:

$$P_{i+1} = P_{i+1}^{1/2} P_{i+1}^{*/2} = \begin{bmatrix} P_{i+1}^{1/2} & 0 \end{bmatrix} \begin{bmatrix} P_{i+1}^{1/2} & 0 \end{bmatrix}^*,$$

$$F_i P_{i|i} F_i^* + G_i Q_i G_i^* = \begin{bmatrix} F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \begin{bmatrix} F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix}^*,$$

where $P_{i+1}^{1/2}$ and $P_{i|i}$ are $n \times n$ square-root factors of P_{i+1} and $P_{i|i}$, respectively. It then follows from the equality (12.5.1) that

$$\begin{bmatrix} F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \begin{bmatrix} F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix}^* = \begin{bmatrix} P_{i+1}^{1/2} & 0 \end{bmatrix} \begin{bmatrix} P_{i+1}^{1/2} & 0 \end{bmatrix}^*.$$

In view of Lemma A.5.1, we conclude that there should exist a unitary rotation Θ relating the arrays shown below

$$\begin{bmatrix} F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} P_{i+1}^{1/2} & 0 \end{bmatrix}, \tag{12.5.2}$$

which explains the origin of the array equations (12.3.4). A similar explanation holds for the array algorithm in the measurement-update case, as we further elaborate.

12.5.2 The Measurement-Update Algorithm

Consider again the measurement-update relation (12.1.7),

$$P_{i|i} = P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i.$$

Because of the minus sign, this is not as trivial to put into an array form as in the time-update case (12.5.2). [The reader is referred to Prob. 12.10, which exhibits an array form for propagating the square-root factor of $P_{i|i}$ by employing *hyperbolic rotations*.]

Here we proceed by using an observation that is helpful in several situations where one has expressions that can be identified as Schur complements of simpler block matrices. Thus note that $P_{i|i}$ can be regarded as the Schur complement of $R_{e,i}$ in the block matrix

$$\begin{bmatrix} R_{e,i} & H_i P_i \\ P_i H_i^* & P_i \end{bmatrix}. \tag{12.5.3}$$

We can now use the block matrix formulas of App. A to obtain the upper-diagonal-lower and lower-diagonal-upper factorizations of (12.5.3) and, hence, the equality

$$\begin{bmatrix} I & H_i \\ 0 & I \end{bmatrix}\begin{bmatrix} R_i & 0 \\ 0 & P_i \end{bmatrix}\begin{bmatrix} I & 0 \\ H_i^* & I \end{bmatrix} =$$

$$\begin{bmatrix} I & 0 \\ P_i H_i^* R_{e,i}^{-1} & I \end{bmatrix}\begin{bmatrix} R_{e,i} & 0 \\ 0 & P_{i|i} \end{bmatrix}\begin{bmatrix} I & R_{e,i}^{-1} H_i P_i \\ 0 & I \end{bmatrix}.$$

The nonnegtive-definiteness of the diagonal terms $\{R_i, P_i, P_{i|i}, R_{e,i}\}$ allows us to incorporate their square-root factors $\{R_i^{1/2}, P_i^{1/2}, P_{i|i}^{1/2}, R_{e,i}^{1/2}\}$ into the triangular blocks, thus leading to the equality

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix}\begin{bmatrix} R_i^{*/2} & 0 \\ P_i^{*/2} H_i^* & P_i^{*/2} \end{bmatrix} = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ P_i H_i^* R_{e,i}^{-*/2} & P_{i|i}^{1/2} \end{bmatrix}\begin{bmatrix} R_{e,i}^{*/2} & R_{e,i}^{-1/2} H_i P_i \\ 0 & P_{i|i}^{*/2} \end{bmatrix}.$$

This now implies, by virtue of Lemma A.5.1, that there exists a unitary matrix Θ that relates the two arrays shown below:

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix}\Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ P_i H_i^* R_{e,i}^{-*/2} & P_{i|i}^{1/2} \end{bmatrix}, \tag{12.5.4}$$

which explains the array equations (12.3.8).

12.5.3 Algorithm for the State Predictors

We again invoke the result of Lemma A.5.1 to justify the array equations of Sec. 12.3.4. To begin with, we write the Riccati recursion as

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_i R_{e,i}^{-1} K_i^*, \tag{12.5.5}$$

where

$$R_{e,i} = R_i + H_i P_i H_i^*, \quad K_i = F_i P_i H_i^*. \tag{12.5.6}$$

Here we have assumed $S_i = 0$ and also $R_i > 0$.

We note from (12.5.5) that P_{i+1} can be obtained as the Schur complement of $R_{e,i}$ in the block matrix

$$\begin{bmatrix} R_{e,i} & K_i^* \\ K_i & F_i P_i F_i^* + G_i Q_i G_i^* \end{bmatrix}. \tag{12.5.7}$$

We now invoke the block matrix formulas of App. A to obtain the lower-diagonal-upper factorization of (12.5.7),

$$\begin{bmatrix} R_{e,i} & K_i^* \\ K_i & F_i P_i F_i^* + G_i Q_i G_i^* \end{bmatrix} = \begin{bmatrix} I & 0 \\ K_i R_{e,i}^{-1} & I \end{bmatrix}\begin{bmatrix} R_{e,i} & 0 \\ 0 & P_{i+1} \end{bmatrix}\begin{bmatrix} I & R_{e,i}^{-1} K_i^* \\ 0 & I \end{bmatrix}.$$

The nonnegative-definiteness of the diagonal terms $\{R_{e,i}, P_{i+1}\}$ again allows us to incorporate their square-root factors $\{R_{e,i}^{1/2}, P_{i+1}^{1/2}\}$ into the triangular blocks, thus leading to the equality

$$\begin{bmatrix} R_{e,i} & K_i^* \\ K_i & F_i P_i F_i^* + G_i Q_i G_i^* \end{bmatrix} = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ K_i R_{e,i}^{-*/2} & P_{i+1}^{1/2} \end{bmatrix}\begin{bmatrix} R_{e,i}^{*/2} & R_{e,i}^{-1/2} K_i^* \\ 0 & P_{i+1}^{*/2} \end{bmatrix}.$$

The upper-diagonal-lower factorization of (12.5.7), on the other hand, is complicated (*e.g.*, we need the inverse of $F_i P_i F_i^* + G_i Q_i G_i^*$). However, note that by going to "larger" factors we can write

$$\begin{bmatrix} R_{e,i} & K_i^* \\ K_i & F_i P_i F_i^* + G_i Q_i G_i^* \end{bmatrix} = \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix}\begin{bmatrix} R_i^{*/2} & 0 \\ P_i^{*/2} H_i^* & P_i^{*/2} F_i^* \\ 0 & Q_i^{*/2} G_i^* \end{bmatrix}.$$

By equating the last two factorizations we obtain

$$\underbrace{\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} R_i^{*/2} & 0 \\ P_i^{*/2} H_i^* & P_i^{*/2} F_i^* \\ 0 & Q_i^{*/2} G_i^* \end{bmatrix}}_{A^*} =$$

$$\underbrace{\begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ K_i R_{e,i}^{-*/2} & P_{i+1}^{1/2} & 0 \end{bmatrix}}_{B}\underbrace{\begin{bmatrix} R_{e,i}^{*/2} & R_{e,i}^{-1/2} K_i^* \\ 0 & P_{i+1}^{*/2} \\ 0 & 0 \end{bmatrix}}_{B^*}.$$

This last equality fits into the statement of Lemma A.5.1. We thus conclude that there should exist a unitary matrix Θ that relates the arrays A and B,

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix}\Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ K_i R_{e,i}^{-*/2} & P_{i+1}^{1/2} & 0 \end{bmatrix}, \tag{12.5.8}$$

which leads us to the array form (12.3.10). [The case $S_i \neq 0$ is treated in Prob. 12.11.]

12.6 A GEOMETRIC DERIVATION OF THE ARRAYS

While the algebraic method presented in the preceding section is generally the simplest way of deducing array algorithms, a geometric interpretation can lend further insight into the origin of the array algorithms. The fact that the pre- and post-arrays are connected by a unitary transformation is a clue that they correspond to representations of certain vectors in terms of two different sets of basis vectors, which are related by a rotation. We shall now demonstrate this fact.

Assume that at time i we have processed all the earlier observations and obtained the innovations $\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_{i-1}\}$. Now consider the linear space[3]

$$\mathcal{L}\left\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \quad \boxed{\mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_{i+1}}\right\}, \tag{12.6.1}$$

where we have added the new observation $\mathbf{y}_i$, the current and future state vectors $\{\mathbf{x}_i, \mathbf{x}_{i+1}\}$, and the disturbances $\{\mathbf{u}_i, \mathbf{v}_i\}$. These variables are added because by projecting onto the span of the earlier innovations we obtain the newest innovation, as well as the filtered and predicted estimation errors as we now explain.

More specifically, we perform two projections (or, equivalently, two steps of MGS computations). The first step projects all the new variables $\{\mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_{i+1}\}$ onto the space of the earlier innovations and keeps the resulting estimation errors. This leads to the equivalent space

$$\mathcal{L}\left\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \quad \boxed{\mathbf{e}_i, \tilde{\mathbf{x}}_{i|i-1}, \mathbf{u}_i, \mathbf{v}_i, \tilde{\mathbf{x}}_{i+1|i-1}}\right\}. \tag{12.6.2}$$

Note that we used the fact that $\hat{\mathbf{u}}_{i|i-1} = 0$ and $\hat{\mathbf{v}}_{i|i-1} = 0$ so that $\tilde{\mathbf{u}}_{i|i-1} = \mathbf{u}_i$ and $\tilde{\mathbf{v}}_{i|i-1} = \mathbf{v}_i$. The above space contains the predicted estimation error $\tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_{i|i-1}$ as well as the new innovations variable $\mathbf{e}_i$. Since we are also interested in filtered errors, we perform a second projection. This time we project the last four variables $\{\tilde{\mathbf{x}}_i, \mathbf{u}_i, \mathbf{v}_i, \tilde{\mathbf{x}}_{i+1|i-1}\}$ onto the enlarged span of the innovations $\{\mathbf{e}_0, \ldots, \mathbf{e}_{i-1}, \mathbf{e}_i\}$, and keep the resulting estimation errors. This step leads to the following equivalent space:

$$\mathcal{L}\left\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \quad \boxed{\mathbf{e}_i, \tilde{\mathbf{x}}_{i|i}, \tilde{\mathbf{u}}_{i|i}, \tilde{\mathbf{v}}_{i|i}, \tilde{\mathbf{x}}_{i+1|i}}\right\}. \tag{12.6.3}$$

It turns out that the variables in the first step (12.6.2) determine the pre-arrays while the variables in the second step (12.6.3) determine the post-arrays.

12.6.1 Predicted Form of the Arrays

To justify the above claim we start by representing $\{\mathbf{y}_i, \mathbf{x}_{i+1}\}$ in terms of variables in the first step (12.6.2) as

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_{i+1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i+1|i-1} \end{bmatrix} + \begin{bmatrix} I & H_i & 0 \\ 0 & F_i & G_i \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \tilde{\mathbf{x}}_{i|i-1} \\ \mathbf{u}_i \end{bmatrix}. \tag{12.6.4}$$

Returning to the original space (12.6.1), the above representation has the following interpretation. It expresses $\{\mathbf{y}_i, \mathbf{x}_{i+1}\}$ in terms of the earlier innovations up to time $i-1$ and, hence, the terms $\{\hat{\mathbf{y}}_{i|i-1}, \hat{\mathbf{x}}_{i+1|i-1}\}$. The resulting estimation errors are then expressed in terms of the quantities $\{\mathbf{x}_i, \mathbf{u}_i, \mathbf{v}_i\}$. This is possible since these variables fully characterize $\{\mathbf{y}_i, \mathbf{x}_{i+1}\}$ and should therefore provide the additional information about the $\{\mathbf{y}_i, \mathbf{x}_{i+1}\}$ that cannot be extracted from the earlier innovations. However, in (12.6.4), rather than use the $\{\mathbf{x}_i, \mathbf{u}_i, \mathbf{v}_i\}$ directly, we use their estimation errors after

[3] As often noted before, by this notation we mean the linear space spanned by all the scalar random variables in the vectors $\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_{i+1}\}$.

removing the information that is already present in the earlier innovations. In other words, we employ the variables $\{\tilde{\mathbf{x}}_{i|i-1}, \mathbf{u}_i, \mathbf{v}_i\}$ that we get from the first step (12.6.2).

As we shall see, the equations (12.6.4) (when normalized) will lead to the pre-array. For the post-array, we use the fact that the newest observation $\{\mathbf{y}_i\}$ is already available and, hence, we proceed to represent $\{\mathbf{y}_i, \mathbf{x}_{i+1}\}$ in terms of the observations up to time i. This of course includes $\mathbf{y}_i$ and leads to a rather trivial representation for $\mathbf{y}_i$ itself. It nevertheless leads to a useful representation for $\mathbf{x}_{i+1}$ since we now write

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_{i+1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i+1|i-1} \end{bmatrix} + \begin{bmatrix} I & 0 \\ K_{p,i} & I \end{bmatrix} \begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i+1|i} \end{bmatrix},$$

where we defined the matrix $K_{p,i} = \langle \mathbf{x}_{i+1}, \mathbf{e}_i \rangle \|\mathbf{e}_i\|^{-2}$. We see that we now use the variables $\{\mathbf{e}_i, \tilde{\mathbf{x}}_{i+1|i}\}$ from the second step (12.6.3).

In order to be able to compare this representation with the earlier one in (12.6.4) we introduce an auxiliary variable $\boldsymbol{\mu}_i$ such that

$$\boldsymbol{\mu}_i \perp \mathcal{L}\{\mathbf{e}_i, \tilde{\mathbf{x}}_{i+1|i}\} \quad \text{and} \quad \mathcal{L}\{\mathbf{e}_i, \tilde{\mathbf{x}}_{i+1|i}, \boldsymbol{\mu}_i\} = \mathcal{L}\{\mathbf{v}_i, \tilde{\mathbf{x}}_{i|i-1}, \mathbf{u}_i\}. \tag{12.6.5}$$

In this case, we can rewrite the above representation as

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_{i+1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i+1|i-1} \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ K_{p,i} & I & 0 \end{bmatrix} \begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i+1|i} \\ \boldsymbol{\mu}_i \end{bmatrix}. \tag{12.6.6}$$

A Digression: An Alternative Derivation of the Riccati Recursion. Note that the expressions (12.6.4) and (12.6.6) have been obtained without explicitly computing $K_{p,i}$ (in terms of P_i or anything else). Now (12.6.4) and (12.6.6) show that

$$\begin{bmatrix} I & H_i & 0 \\ 0 & F_i & G_i \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \tilde{\mathbf{x}}_{i|i-1} \\ \mathbf{u}_i \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ K_{p,i} & I & 0 \end{bmatrix} \begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i+1|i} \\ \boldsymbol{\mu}_i \end{bmatrix}. \tag{12.6.7}$$

Then with the definitions (as before), $R_{e,i} = \langle \mathbf{e}_i, \mathbf{e}_i \rangle$ and $P_i = \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i \rangle$, forming the covariance matrices of the quantities on both sides of the equality (12.6.7) gives (after minor algebra)

$$\begin{bmatrix} R_i + H_i P_i H_i^* & H_i P_i F_i^* + S_i^* G_i^* \\ F_i P_i H_i^* + G_i S_i & F_i P_i F_i^* + G_i Q_i G_i^* \end{bmatrix} = \begin{bmatrix} R_{e,i} & K_i^* \\ K_i & P_{i+1} + K_i R_{e,i}^{-1} K_i^* \end{bmatrix}. \tag{12.6.8}$$

Now equating corresponding elements simultaneously yields the relations separately derived in Sec. 9.2.2,

$$R_{e,i} = R_i + H_i P_i H_i^*, \quad K_i = F_i P_i H_i^* + G_i S_i, \quad P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_i R_{e,i}^{-1} K_i^*.$$

Normalized Bases. After this digression, we return to our main theme: deriving the array algorithms. [Note that for generality, in the above derivation of the Riccati recursion we modified our standing assumption that $S_i = 0$. In fact, the geometric approach allows us to handle the case $S_i \neq 0$ fairly directly, as we shall see below. So in the following we shall assume $S_i \neq 0$.]

Now reviewing the derivations of (12.6.4) and (12.6.6), we note that $\{\mathbf{v}_i, \mathbf{u}_i, \tilde{\mathbf{x}}_{i|i-1}\}$ and $\{\mathbf{e}_i, \tilde{\mathbf{x}}_{i+1|i}, \boldsymbol{\mu}_i\}$ span the same $(n+m+p)$-dimensional space of variables. In fact, since the variables in each set are linearly independent, these two sets form two different *bases* of the same space of random variables. If these variables were also orthonormalized so as to be independent of each other, and of unit variance, then any relation between the normalized bases can only be a "rotation", *i.e.*, the two normalized bases can only be related by a unitary transformation. The set $\{\mathbf{e}_i, \tilde{\mathbf{x}}_{i+1|i}, \boldsymbol{\mu}_i\}$ in the post-array is very easy to handle, because the quantities are mutually orthogonal, and so we can normalize them individually to get the normalized variables

$$\bar{\mathbf{e}}_i = R_{e,i}^{-1/2}\mathbf{e}_i, \quad \bar{\tilde{\mathbf{x}}}_{i+1|i} = P_{i+1}^{-1/2}\tilde{\mathbf{x}}_{i+1|i}, \quad \bar{\boldsymbol{\mu}}_i = \|\boldsymbol{\mu}_i\|^{-1}\boldsymbol{\mu}_i.$$

In terms of these normalized variables, the post-array (12.6.6) can be written as

$$\begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i+1|i-1} \end{bmatrix} = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ K_i R_{e,i}^{-*/2} & P_{i+1}^{1/2} & 0 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{e}}_i \\ \bar{\tilde{\mathbf{x}}}_{i+1|i} \\ \bar{\boldsymbol{\mu}}_i \end{bmatrix}. \tag{12.6.9}$$

For the variables $\{\mathbf{v}_i, \mathbf{u}_i, \tilde{\mathbf{x}}_{i|i-1}\}$ in the pre-array, we can use a Gram-Schmidt technique for sequential orthonormalization:

$$\bar{\mathbf{v}}_i = R_i^{-1/2}\mathbf{v}_i,$$

$$\bar{\mathbf{u}}_i = \|\mathbf{u}_i - \langle \mathbf{u}_i, \bar{\mathbf{v}}_i\rangle \bar{\mathbf{v}}_i\|^{-1}(\mathbf{u}_i - \langle \mathbf{u}_i, \bar{\mathbf{v}}_i\rangle \bar{\mathbf{v}}_i) = Q_i^{-s/2}(\mathbf{u}_i - S_i R_i^{-*/2}\bar{\mathbf{v}}_i),$$

where we have defined

$$Q_i^s \triangleq Q_i - S_i^* R_i^{-1} S_i.$$

Finally, since $\tilde{\mathbf{x}}_{i|i-1} \perp \{\mathbf{u}_i, \mathbf{v}_i\}$, we have

$$\bar{\tilde{\mathbf{x}}}_{i|i-1} = P_i^{-1/2}\tilde{\mathbf{x}}_{i|i-1}. \tag{12.6.10}$$

We therefore obtain

$$\begin{bmatrix} \mathbf{v}_i \\ \tilde{\mathbf{x}}_{i|i-1} \\ \mathbf{u}_i \end{bmatrix} = \begin{bmatrix} R_i^{1/2} & 0 & 0 \\ 0 & P_i^{1/2} & 0 \\ S_i R_i^{-*/2} & 0 & Q_i^{s/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{v}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i-1} \\ \bar{\mathbf{u}}_i \end{bmatrix}, \tag{12.6.11}$$

which leads to the normalized pre-array (*cf.* (12.6.4))

$$\begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i+1|i-1} \end{bmatrix} = \begin{bmatrix} I & H_i & 0 \\ 0 & F_i & G_i \end{bmatrix} \begin{bmatrix} R_i^{1/2} & 0 & 0 \\ 0 & P_i^{1/2} & 0 \\ S_i R_i^{-*/2} & 0 & Q_i^{s/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{v}}_i \\ \bar{\mathbf{u}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i-1} \end{bmatrix},$$

$$= \begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ G_i S_i R_i^{-*/2} & F_i P_i^{1/2} & G_i Q_i^{s/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{v}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i-1} \\ \bar{\mathbf{u}}_i \end{bmatrix}. \tag{12.6.12}$$

Now since the two orthonormal sets of variables on the right-hand sides of (12.6.9) and (12.6.12) are related by a unitary transformation, the same must be true of the coefficient arrays in (12.6.9) and (12.6.12), namely, there should exist a unitary matrix Θ such that

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ G_i S_i R_i^{-*/2} & F_i P_i^{1/2} & G_i Q_i^{s/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ K_i R_{e,i}^{-*/2} & P_{i+1}^{1/2} & 0 \end{bmatrix}. \tag{12.6.13}$$

For $S_i = 0$, the above array equations collapse to the earlier equations (12.3.10) — see also Prob. 12.11.

12.6.2 Measurement Updates

To further reinforce the geometric ideas, let us go through them in the simpler problems of measurement and time updates.

We now represent $\{\mathbf{y}_i, \mathbf{x}_i\}$ in terms of variables from the first and second steps, (12.6.2) and (12.6.3), respectively. This leads to

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_i \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i|i-1} \end{bmatrix} + \begin{bmatrix} I & H_i \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \tilde{\mathbf{x}}_{i|i-1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i|i-1} \end{bmatrix} + \begin{bmatrix} I & 0 \\ K_{f,i} & I \end{bmatrix} \begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i|i} \end{bmatrix},$$

where we defined $K_{f,i}' = \langle \mathbf{x}_i, \mathbf{e}_i\rangle R_{e,i}^{-1}$. This shows that we must have

$$\begin{bmatrix} I & H_i \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \tilde{\mathbf{x}}_{i|i-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ K_{f,i} & I \end{bmatrix} \begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i|i} \end{bmatrix}. \tag{12.6.14}$$

Then, with

$$R_i = \langle \mathbf{v}_i, \mathbf{v}_i\rangle, \quad P_i = \langle \tilde{\mathbf{x}}_{i|i-1}, \tilde{\mathbf{x}}_{i|i-1}\rangle, \quad P_{i|i} = \langle \tilde{\mathbf{x}}_{i|i}, \tilde{\mathbf{x}}_{i|i}\rangle, \quad R_{e,i} = \langle \mathbf{e}_i, \mathbf{e}_i\rangle,$$

we can further normalize the variables $\{\mathbf{v}_i, \mathbf{e}_i, \tilde{\mathbf{x}}_{i|i-1}, \tilde{\mathbf{x}}_{i|i}\}$ in order to have unit variance, say

$$\begin{bmatrix} \mathbf{v}_i \\ \tilde{\mathbf{x}}_{i|i-1} \end{bmatrix} = \begin{bmatrix} R_i^{1/2} & 0 \\ 0 & P_i^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{v}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i-1} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i|i} \end{bmatrix} = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ 0 & P_{i|i}^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{e}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i} \end{bmatrix}.$$

It then follows from (12.6.14) that

$$\begin{bmatrix} I & H_i \\ 0 & I \end{bmatrix} \begin{bmatrix} R_i^{1/2} & 0 \\ 0 & P_i^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{v}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ K_{f,i} & I \end{bmatrix} \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ 0 & P_{i|i}^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{e}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i} \end{bmatrix},$$

or, equivalently,

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{v}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i-1} \end{bmatrix} = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ P_i H_i^* R_{e,i}^{-*/2} & P_{i|i}^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{e}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i} \end{bmatrix}.$$

This implies, as argued before, the existence of a unitary rotation Θ such that

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ P_i H_i^* R_{e,i}^{-*/2} & P_{i|i}^{1/2} \end{bmatrix},$$

which is the array form (12.3.8) for the measurement-update problem.

12.6.3 Time Updates

Assuming $S_i = 0$, the time-update array is rather trivial and can be similarly justified by starting with the time-update error equation rewritten as

$$\begin{bmatrix} F_i & G_i \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{i|i} \\ \mathbf{u}_i \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{i+1|i} \\ \boldsymbol{\beta}_i \end{bmatrix},$$

where we introduced an auxiliary variable $\boldsymbol{\beta}_i$ that is orthogonal to $\tilde{\mathbf{x}}_{i+1|i}$ and such that $\{\tilde{\mathbf{x}}_{i+1|i}, \boldsymbol{\beta}_i\}$ and $\{\tilde{\mathbf{x}}_{i|i}, \mathbf{u}_i\}$ span the same space of random variables. We leave the rest of the analysis to the active reader.

12.7 PAIGE'S FORM OF THE ARRAY ALGORITHM

Simulations and some mathematical analysis show that the performance of all the different variants of the Kalman filter recursions in finite precision implementations depends upon the conditioning of the matrices R_i and to a lesser extent $\{Q_i, F_i\}$ (see, *e.g.*, Verhaegen and Van Dooren (1986)). When the matrices are well conditioned, all the implementations perform satisfactorily and one should choose the most convenient implementation at hand. The best cure for ill-conditioning is a second look at the assumed model and the accuracy to which its parameters are known or the accuracy to which they can reasonably be determined. This accuracy will generally determine the appropriate level of effort to be expended on the choice of algorithm.[4]

With this in mind, we introduce yet another variant of the Kalman filter recursions that avoids the potential loss of accuracy in forming the matrix products $\{F_i P_i^{1/2}, H_i P_i^{1/2}\}$ used in the array algorithms! Paige (1985) described a variant in which such products are not explicitly formed. The algorithm has an interesting form and, in fact, turns out to involve a factorization of the square-root factor $P_i^{1/2}$ itself, as $P_i^{1/2} = A_i^{-1} B_i$, with the $\{A_i, B_i\}$ being separately propagated by the recursions (see the statement of Alg. 12.C.1). Paige's algorithm forms a pre-array of the form

$$\mathcal{P} = \begin{bmatrix} A_i & 0 & B_i & 0 & 0 \\ H_i & 0 & 0 & R_i^{1/2} & 0 \\ -F_i & I & 0 & 0 & G_i Q_i^{1/2} \end{bmatrix}, \qquad A_0 = I, \qquad B_0 = \Pi_0^{1/2},$$

[4] A quotation from p. 2 of a book by N. J. Higham (1996), a valuable reference and sourcebook for numerical analysis, is relevant: "A word of warning: some of the examples from Sec. 1.12 onward are special ones chosen to illustrate particular phenomena. You may never see in practice the extremes of behaviour shown here. Let the examples show you what can happen, but do not let them destroy your confidence in finite precision arithmetic!"

and then transforms it through a sequence of unitary rotations from left *and* right to a post-array of the form

$$\mathcal{Q} = \begin{bmatrix} \times & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & 0 \\ 0 & A_{i+1} & 0 & 0 & B_{i+1} \end{bmatrix}.$$

The algorithm also propagates quantities from which the predicted (and smoothed) estimators can be obtained by a simple matrix-vector multiplication. In addition, Paige proved that his algorithm is "backward" stable, a widely sought feature by numerical analysts.

Remark 4. A numerical algorithm for solving a linear system of equations $Ax = b$ is said to be *backward stable* if the computed solution $\bar{x}$ can be shown to be the *exact* solution of a slightly perturbed system, *viz.*, if $\bar{x}$ satisfies $(A + \delta A)\bar{x} = (b + \delta b)$, with $\|\delta A\| \leq \epsilon_1 \cdot \|A\|$, $\|\delta b\| \leq \epsilon_2 \cdot \|b\|$ and where (ϵ_1, ϵ_2) denote small numbers of the same order of magnitude as the machine precision ϵ — this is essentially the smallest number that can be represented in finite precision arithmetic. [Also, $\|A\|$ denotes a matrix norm, *e.g.*, the maximum singular value of A — see App. A.]

For the filtering problem, Paige (1979b) showed that Alg. 12.C.1 is backward stable in the following sense (in terms of quantities defined in App. 12.C). The computed state estimates $\hat{\bar{x}}_i$ that are obtained for a given observation vector y are the exact solution of an estimation problem with slightly perturbed $\{\mathcal{A}, \mathcal{W}, y\}$ in (12.C.3), *viz.*, with

$$\|\delta\mathcal{A}\| \leq \epsilon_1 \cdot \|\mathcal{A}\|, \qquad \|\delta\mathcal{W}\| \leq \epsilon_2 \cdot \|\mathcal{W}\|, \qquad \|\delta y\| \leq \epsilon_3 \cdot \|y\|.$$

◆

The derivation in Paige (1985) (see also the related works of Paige and Saunders (1977) and Paige (1979a,1979b)) was in the context of deterministic least-squares problems. In App. 12.C, we give a derivation that is carried out directly in the stochastic context. It also allows us to provide the above interpretation for the factors A_i and B_i.

12.8 ARRAY ALGORITHMS FOR THE INFORMATION FORMS

As noted in Ch. 9, the general information filter equations are somewhat complicated, so that they are usually separated into measurement- and time-update formulas. We shall see in a later section that this complication disappears in the square-root formulation of the information filter. But first, let us present the array forms of the corresponding information filters for the time and measurement updates.

We shall assume here that $R_i > 0$, $\Pi_0 > 0$, and that the F_i are invertible. Under these assumptions, we established in Lemma 9.5.1 that the Riccati variable P_i is necessarily positive-definite and, hence, invertible.

12.8.1 Information Array for the Measurement Update

The information form of the measurement-update filter is given by (*cf.* Thm. 9.5.1):

$$P_{i|i}^{-1} = P_i^{-1} + H_i^* R_i^{-1} H_i, \tag{12.8.1}$$

$$P_{i|i}^{-1} \hat{\mathbf{x}}_{i|i} = P_i^{-1} \hat{\mathbf{x}}_i + H_i^* R_i^{-1} \mathbf{y}_i. \tag{12.8.2}$$

Now (12.8.1) has exactly the same form as the time-update equation for going from $P_{i|i}$ to P_{i+1} (see (12.3.1)), so that as in Sec. 12.3.2, we can write down an array algorithm as (recall that $P_i^{-1} = P_i^{-*/2} P_i^{-1/2}$)

$$\begin{bmatrix} P_i^{-*/2} & H_i^* R_i^{-*/2} \end{bmatrix} \Theta_{mu} = \begin{bmatrix} P_{i|i}^{-*/2} & 0 \end{bmatrix}. \tag{12.8.3}$$

Moreover, in this problem we can get a bonus. Noticing that the right-hand side of (12.8.2) can be rewritten as

$$P_i^{-*/2} P_i^{-1/2} \hat{\mathbf{x}}_i \;+\; H_i^* R_i^{-*/2} R_i^{-1/} \mathbf{y}_i,$$

we can identify it as the inner product of the rows

$$\begin{bmatrix} P_i^{-*/2} & H_i^* R_i^{-*/2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \hat{\mathbf{x}}_i^* P_i^{-*/2} & \mathbf{y}_i^* R_i^{-*/2} \end{bmatrix}.$$

Therefore, the array in (12.8.3) can be expanded to

$$\begin{bmatrix} P_i^{-*/2} & H_i^* R_i^{-*/2} \\ \hat{\mathbf{x}}_i^* P_i^{-*/2} & \mathbf{y}_i^* R_i^{-*/2} \end{bmatrix} \Theta_{mu} = \begin{bmatrix} P_{i|i}^{-*/2} & 0 \\ \hat{\mathbf{x}}_{i|i}^* P_{i|i}^{-*/2} & \boldsymbol{\alpha} \end{bmatrix}, \tag{12.8.4}$$

where $\boldsymbol{\alpha}$ is a $1 \times p$ row vector. By equating the norms of the second rows, then some calculations will show that

$$\boldsymbol{\alpha} = \mathbf{e}_i^* R_{e,i}^{-*/2}.$$

The result (12.8.4) was derived by Dyer and McReynolds (1969) in one of the first papers on array methods for state-space estimation; they were inspired by the QR method of Sec. 2.5 for the deterministic least-squares problem and its recursive form described in Prob. 2.17.

12.8.2 Information Array for the Time Update

Recall again from Thm. 9.5.1 that the time-update equations are given by (assuming $S_i = 0$; the case $S_i \neq 0$ is treated in Prob. 12.12)

$$\hat{\mathbf{x}}_{i+1} = F_i \hat{\mathbf{x}}_{i|i}, \qquad P_{i+1} = F_i P_{i|i} F_i^* + G_i Q_i G_i^*.$$

To get the information form for these time-update expressions we simply apply the matrix inversion formula to get

$$P_{i+1}^{-1} = A_i - A_i G_i (Q_i^{-1} + G_i^* A_i G_i)^{-1} G_i^* A_i, \tag{12.8.5}$$

$$P_{i+1}^{-1} \hat{\mathbf{x}}_{i+1} = (I + A_i G_i Q_i G_i^*)^{-1} (F_i^{-*} P_{i|i}^{-1} \hat{\mathbf{x}}_{i|i}), \tag{12.8.6}$$

where we have defined $A_i \triangleq F_i^{-*} P_{i|i}^{-1} F_i^{-1}$. We also define (recall the remark after (12.3.14))

$$Q_i^{-1} + G_i^* A_i G_i \triangleq Q_i^{-r}, \qquad Q_i^r = Q_i - Q_i G_i^* P_{i+1}^{-1} G_i Q_i. \tag{12.8.7}$$

If we now follow the same arguments as in the previous sections, we can verify the validity of the following array algorithm

$$\begin{bmatrix} Q_i^{-*/2} & G_i^* F_i^{-*} P_{i|i}^{-*/2} \\ 0 & F_i^{-*} P_{i|i}^{-*/2} \end{bmatrix} \Theta_{tu} = \begin{bmatrix} Q_i^{-*r/2} & 0 \\ A_i G_i Q_i^{r/2} & P_{i+1}^{-*/2} \end{bmatrix}, \tag{12.8.8}$$

where Θ_{tu} is any unitary rotation that nulls out the (1, 1) block element in the post-array.

Remark 5. Bierman (1977) dubbed the combination of (12.8.4) and (12.8.8) the Square-Root Information Filter (SRIF). ◆

12.8.3 Alternative Derivation via Inversion of Covariance Forms

An alternative useful approach is to note that if A and B are square and invertible matrices, then $A = B\Theta$ also implies that $\Theta^{-1} B^{-1} = A^{-1}$. When Θ is unitary, and hence $\Theta^{-1} = \Theta^*$, we conclude from this that $A^{-*} = B^{-*}\Theta$. That is, the same unitary transformation Θ also maps B^{-*} to A^{-*}. Here are two examples.

First, recall the time-update array (12.5.2), with $S_i = 0$. We may augment the pre-array to a square matrix and then invert it. Thus we can write

$$\begin{bmatrix} F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \\ 0 & Q_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} P_{i+1}^{1/2} & 0 \\ X & Y \end{bmatrix},$$

for some X and Y to be determined. By inverting both sides of the equality we obtain

$$\Theta^* \begin{bmatrix} P_{i|i}^{-1/2} F_i^{-1} & -P_{i|i}^{-1/2} F_i^{-1} G_i \\ 0 & Q_i^{-1/2} \end{bmatrix} = \begin{bmatrix} P_{i+1}^{-1/2} & 0 \\ X_1 & X_2 \end{bmatrix},$$

where, by squaring or otherwise, X_1 and X_2 can readily be determined to be equal to the desired quantities in (12.8.8).

Likewise, consider the measurement-update form (12.5.4), rewritten conveniently with a minus sign in the (1, 2) block entry of the pre-array and with an unidentified entry in the (2, 1) block of the post-array:

$$\begin{bmatrix} R_i^{1/2} & -H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ X & P_{i|i}^{1/2} \end{bmatrix}.$$

Inverting both sides of the equality leads to

$$\Theta^* \begin{bmatrix} R_i^{-1/2} & R_i^{-1/2} H_i \\ 0 & P_i^{-1/2} \end{bmatrix} = \begin{bmatrix} R_{e,i}^{-1/2} & 0 \\ X_1 & P_{i|i}^{-1/2} \end{bmatrix},$$

where X_1 can be identified by squaring. It is not necessary to do this, because we just need to observe that

$$\begin{bmatrix} H_i^* R_i^{-*/2} & P_i^{-*/2} \end{bmatrix} \Theta = \begin{bmatrix} 0 & P_{i|i}^{-*/2} \end{bmatrix},$$

which agrees with the top row of (12.8.4).

12.8.4 Derivation via Dualities When $R_i > 0$ and $Q_i > 0$

As perhaps first noted by Kaminski, Bryson, and Schmidt (1971), the information arrays of the time and measurement updates can also be obtained by appealing to certain duality relations that exist between the quantities in the covariance-based and the information-based descriptions of the algorithms. To show this re-derivation, we shall carry out the arguments in this section for the general case $S_i \neq 0$.

Thus recall the time-update relation (12.1.8) — see also Prob. 12.12,

$$P_{i+1} = F_i^s P_{i|i} F_i^{*s} + G_i Q_i^s G_i^*, \tag{12.8.9}$$

where

$$F_i^s = F_i - G_i S_i R_i^{-1} H_i, \quad Q_i^s = Q_i - S_i R_i^{-1} S_i^*, \tag{12.8.10}$$

and the measurement-update relation (12.8.1) in information form,

$$P_{i|i}^{-1} = P_i^{-1} + H_i^* R_i^{-1} H_i. \tag{12.8.11}$$

Comparing (12.8.9) and (12.8.11) we see that if we make the identifications

$$P_i^{-1} \longleftrightarrow F_i^s P_{i|i} F_i^{*s}, \quad H_i^* \longleftrightarrow G_i, \quad R_i^{-1} \longleftrightarrow Q_i^s, \quad P_{i|i}^{-1} \longleftrightarrow P_{i+1}, \tag{12.8.12}$$

then expression (12.8.9) reduces to expression (12.8.11) and vice versa. Therefore, we can employ the array form (12.5.2) for the time-update algorithm (written here for the general case of $S_i \neq 0$),

$$\begin{bmatrix} F_i^s P_{i|i}^{1/2} & G_i Q_i^{s/2} \end{bmatrix} \Theta = \begin{bmatrix} P_{i+1}^{1/2} & 0 \end{bmatrix},$$

along with the dualities (12.8.12), in order to conclude the following information form for the measurement update (12.8.11):

$$\begin{bmatrix} P_i^{-*/2} & H_i^* R_i^{-*/2} \end{bmatrix} \Theta_{mu} = \begin{bmatrix} P_{i|i}^{-*/2} & 0 \end{bmatrix}, \tag{12.8.13}$$

where Θ_{mu} is a unitary rotation (compare with the first row in (12.8.4)).

Likewise, consider the measurement-update relation (12.1.7) in covariance form,

$$P_{i|i} = P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i, \tag{12.8.14}$$

and the time-update relation (12.8.5) in information form (written here for the general $S_i \neq 0$ case, as described in Prob. 12.12)

$$P_{i+1}^{-1} = A_i - A_i G_i Q_i^r G_i^* A_i, \tag{12.8.15}$$

where

$$A_i = F_i^{-*s} P_{i|i}^{-1} F_i^{-s}, \quad Q_i^{-r} = Q_i^{-s} + G_i^* A_i G_i. \tag{12.8.16}$$

We see that if we invoke the duality relations (12.8.12), along with the additional relation

$$R_{e,i} \longleftrightarrow Q_i^{-r}, \tag{12.8.17}$$

Table 12.1 Dualities between covariance and information variables.

Covariance case	Information case
G_i	H_i^*
$Q_i^s = Q_i - S_i R_i^{-1} S_i^*$	R_i^{-1}
$P_{i\|i}$	P_{i+1}^{-1}
P_i	$F_i^{-*s} P_{i\|i}^{-1} F_i^{-s} = A_i, \quad F_i^s = F_i - G_i S_i R_i^{-1} H_i$
$R_{e,i}$	$Q_i^{-r} = Q_i^{-s} + G_i^* A_i G_i$

then expression (12.8.14) reduces to (12.8.15) and vice versa. Therefore, we can employ the array form (12.5.4) for the measurement-update algorithm,

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ P_i H_i^* R_{e,i}^{-*/2} & P_{i|i}^{1/2} \end{bmatrix},$$

along with the dualities (12.8.12) and (12.8.17), in order to conclude the following information form for the time-update algorithm (12.8.8):

$$\begin{bmatrix} Q_i^{-*/2} & G_i^* F_i^{-*s} P_{i|i}^{-*/2} \\ 0 & F_i^{-*s} P_{i|i}^{-*/2} \end{bmatrix} \Theta_{tu} = \begin{bmatrix} Q_i^{-*r/2} & 0 \\ F_i^{-*s} P_{i|i}^{-1} F_i^{-s} G_i Q_i^{r/2} & P_{i+1}^{-*/2} \end{bmatrix}, \tag{12.8.18}$$

where Θ_{tu} is a unitary rotation (compare with (12.8.8)). Table 12.1 summarizes the duality relations that allow us to translate the covariance forms into information forms and vice versa.

12.8.5 The General Information Filter Form

We now treat the information filter form of the Kalman recursions (12.1.3)–(12.1.4). In particular, we show that although the update expression for P_i^{-1} is somewhat complicated, the array formulation allows us to avoid this difficulty. We shall return to our standard assumption $S_i = 0$.

One convenient way of obtaining the array information filter is to augment the array form of the Kalman filter (12.3.10) in order to construct nonsingular arrays that can be inverted. Thus, with any nonsingular $W \in \mathbb{C}^{m \times m}$, let us form the augmented arrays

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \\ 0 & 0 & W \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \\ Y_1 & Y_2 & Y_3 \end{bmatrix},$$

where forming inner products yields

$$Y_1 = 0, \quad Y_2 = W Q_i^{*/2} G_i^* P_{i+1}^{-*/2}, \quad Y_3 = (WW^* - Y_2 Y_2^*)^{1/2}.$$

Convenient choices for W are $W = I$ or $W = Q_i^{1/2}$. Making the first choice, we can invert and transpose the above equation to obtain the following result, which holds for $R_i > 0$, $\Pi_0 > 0$, F_i invertible, and $S_i = 0$. The "(*)" notation indicates "don't care" entries:

$$\begin{bmatrix} R_i^{-*/2} & 0 & 0 \\ -F_i^{-*}H_i^*R_i^{-*/2} & F_i^{-*}P_i^{-*/2} & 0 \\ Q_i^{*/2}G_i^*F_i^{-*}H_i^*R_i^{-*/2} & -Q_i^{*/2}G_i^*F_i^{-*}P_i^{-*/2} & I \\ -\mathbf{y}_i^*R_i^{-*/2} & \hat{\mathbf{x}}_i^*P_i^{-*/2} & 0 \end{bmatrix} \Theta =$$

$$\begin{bmatrix} R_{e,i}^{-*/2} & -K_{p,i}^*P_{i+1}^{-*/2} & (*) \\ 0 & P_{i+1}^{-*/2} & (*) \\ 0 & 0 & (*) \\ -\mathbf{e}_i^*R_{e,i}^{-*/2} & \hat{\mathbf{x}}_{i+1}^*P_{i+1}^{-*/2} & (*) \end{bmatrix}. \tag{12.8.19}$$

Here we note that Θ can be the same rotation as in the array form of the Kalman filter (12.3.10). Alternatively, we could define Θ directly as any unitary rotation that upper triangularizes the second and third (block) rows of the pre-array.

We further note that the first (block) row in the above equations can be ignored unless we are interested in finding $R_{e,i}$. Moreover, the predicted estimator can be found from the entries of the post-array by solving the triangular system

$$(P_{i+1}^{-1/2})(\hat{\mathbf{x}}_{i+1}) = (P_{i+1}^{-1/2}\hat{\mathbf{x}}_{i+1}).$$

However, this backwards substitution calculation may be avoided by adding an additional (fourth block) row to the arrays, as shown below:

$$\begin{bmatrix} R_i^{-*/2} & 0 & 0 \\ -F_i^{-*}H_i^*R_i^{-*/2} & F_i^{-*}P_i^{-*/2} & 0 \\ Q_i^{*/2}G_i^*F_i^{-*}H_i^*R_i^{-*/2} & -Q_i^{*/2}G_i^*F_i^{-*}P_i^{-*/2} & I \\ 0 & F_iP_i^{1/2} & G_iQ_i^{1/2} \\ -\mathbf{y}_i^*R_i^{-*/2} & \hat{\mathbf{x}}_i^*P_i^{-*/2} & 0 \end{bmatrix} \Theta =$$

$$\begin{bmatrix} R_{e,i}^{-*/2} & -K_{p,i}^*P_{i+1}^{-*/2} & (*) \\ 0 & P_{i+1}^{-*/2} & (*) \\ 0 & 0 & (*) \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \\ -\mathbf{e}_i^*R_{e,i}^{-*/2} & \hat{\mathbf{x}}_{i+1}^*P_{i+1}^{-*/2} & (*) \end{bmatrix},$$

where Θ is any unitary rotation that upper triangularizes the second and third (block) rows of the pre-array. Now that we have available the term $(P_{i+1}^{1/2})$, we can directly form $\hat{\mathbf{x}}_{i+1} = (P_{i+1}^{1/2})(P_{i+1}^{-1/2}\hat{\mathbf{x}}_{i+1})$.

12.8.6 A Geometric Derivation of the Information Filter Form

A geometric motivation for the information arrays (12.8.19) can also be provided. For this purpose, we return to the discussion in Sec. 12.6 and note, using $\hat{\mathbf{u}}_{i|i-1} = 0$ and $\tilde{\mathbf{u}}_{i|i-1} = \mathbf{u}_i$, that we can extend the equations (12.6.4) and (12.6.6) as follows:

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_{i+1} \\ \mathbf{u}_i \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i+1|i-1} \\ \hat{\mathbf{u}}_{i|i-1} \end{bmatrix} + \begin{bmatrix} I & H_i & 0 \\ 0 & F_i & G_i \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \tilde{\mathbf{x}}_{i|i-1} \\ \mathbf{u}_i \end{bmatrix},$$

and

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_{i+1} \\ \mathbf{u}_i \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i+1|i-1} \; cr\hat{\mathbf{u}}_{i|i-1} \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ K_{p,i} & I & 0 \\ (*) & (*) & (*) \end{bmatrix} \begin{bmatrix} \mathbf{e}_i \\ \tilde{\mathbf{x}}_{i+1|i} \\ \boldsymbol{\mu}_i \end{bmatrix},$$

where the (*) denote "don't care" entries. In terms of normalized quantities we then obtain

$$\begin{bmatrix} \bar{\mathbf{v}}_i \\ \bar{\tilde{\mathbf{x}}}_{i|i-1} \\ \bar{\mathbf{u}}_i \end{bmatrix} = \begin{bmatrix} R_i^{1/2} & H_iP_i^{1/2} & 0 \\ 0 & F_iP_i^{1/2} & G_iQ_i^{1/2} \\ 0 & 0 & Q_i^{1/2} \end{bmatrix}^{-1} \left(\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_{i+1} \\ \mathbf{u}_i \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i+1|i-1} \\ \hat{\mathbf{u}}_{i|i-1} \end{bmatrix} \right),$$

and

$$\begin{bmatrix} \bar{\mathbf{e}}_i \\ \bar{\tilde{\mathbf{x}}}_{i+1|i} \\ \bar{\boldsymbol{\mu}}_i \end{bmatrix} = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ K_iR_{e,i}^{-*/2} & P_{i+1}^{1/2} & 0 \\ (*) & (*) & (*) \end{bmatrix}^{-1} \left(\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_{i+1} \\ \mathbf{u}_i \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{y}}_{i|i-1} \\ \hat{\mathbf{x}}_{i+1|i-1} \\ \hat{\mathbf{u}}_{i|i-1} \end{bmatrix} \right).$$

But since $\{\bar{\mathbf{v}}_i\bar{\mathbf{u}}_i, \bar{\tilde{\mathbf{x}}}_{i|i-1}\}$ and $\{\bar{\mathbf{e}}_i, \bar{\tilde{\mathbf{x}}}_{i+1|i}, \bar{\boldsymbol{\mu}}_i\}$ are orthonormal bases for the same space of random variables, we conclude that there exists a unitary rotation Θ such that

$$\Theta^* \begin{bmatrix} R_i^{1/2} & H_iP_i^{1/2} & 0 \\ 0 & F_iP_i^{1/2} & G_iQ_i^{1/2} \\ 0 & 0 & Q_i^{1/2} \end{bmatrix}^{-1} = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ K_iR_{e,i}^{-*/2} & P_{i+1}^{1/2} & 0 \\ (*) & (*) & (*) \end{bmatrix}^{-1}.$$

Explicit inversion of the block matrices in the above equation leads to

$$\Theta^* \begin{bmatrix} R_i^{-1/2} & -R_i^{-1/2}H_iF_i^{-1} & R_i^{-1/2}H_iF_i^{-1}G_i \\ 0 & P_i^{-1/2}F_i^{-1} & -P_i^{-1/2}F_i^{-1}G_i \\ 0 & 0 & Q_i^{-1/2} \end{bmatrix} = \begin{bmatrix} R_{e,i}^{-1/2} & 0 & 0 \\ -P_{i+1}^{-1/2}K_iR_{e,i}^{-1} & P_{i+1}^{-1/2} & 0 \\ (*) & (*) & (*) \end{bmatrix}.$$

By transposing both sides of the equality we obtain an array algorithm that agrees with the top three rows of (12.8.19).

12.9 ARRAY ALGORITHMS FOR SMOOTHING

The array formulation of the earlier sections can be extended to the smoothing algorithms of Ch. 10, as we now elaborate rather briefly for the Bryson-Frazier, Rauch-Tung-Striebel, and Mayne-Fraser formulas (following Park and Kailath (1995b)). An interesting feature of the results is that the apparently most computationally intensive traditional algorithm — the two-filter solution of Mayne and Fraser — has the least complex array form.

12.9.1 Bryson-Frazier Formulas in Array Form

Consider the following array that is obtained from (12.3.13) by incorporating the row $[0 \;\; I \;\; 0]$ into the pre-array:

$$\left[\begin{array}{ccc} R_i^{1/2} & H_iP_i^{1/2} & 0 \\ 0 & F_iP_i^{1/2} & G_iQ_i^{1/2} \\ \hline 0 & I & 0 \\ -\mathbf{y}_i^*R_i^{-*/2} & \hat{\mathbf{x}}_i^*P_i^{-*/2} & 0 \end{array}\right] \Theta_1 = \left[\begin{array}{ccc} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \\ \hline X & Y & Z \\ -\mathbf{e}_i^*R_{e,i}^{-*/2} & \hat{\mathbf{x}}_{i+1}^*P_{i+1}^{-*/2} & (*) \end{array}\right],$$

where, for convenience of notation, the unitary rotation is here denoted by Θ_1.

Cross-products between the first and third rows and between the second and third rows yield the equations

$$X = P_i^{*/2}H_i^*R_{e,i}^{-*/2} = P_i^{-1/2}(P_iH_i^*R_{e,i}^{-*/2}),$$

$$Y = (P_i^{*/2}F_i^* - X\bar{K}_{p,i}^*)P_{i+1}^{-*/2} = P_i^{*/2}(F_i - K_{p,i}H_i)^*P_{i+1}^{-*/2} = P_i^{*/2}F_{p,i}P_{i+1}^{-*/2}.$$

Rewriting equation (10.2.7) using X and Y, we get

$$(P_i^{*/2}\boldsymbol{\lambda}_{i|N}) = (Y)(P_{i+1}^{*/2}\boldsymbol{\lambda}_{i+1|N}) + (X)(R_{e,i}^{-1/2}\mathbf{e}_i),$$

which propagates $P_i^{*/2}\boldsymbol{\lambda}_{i|N}$ instead of $\boldsymbol{\lambda}_{i|N}$. With $P_i^{-1/2}\hat{\mathbf{x}}_i$ also recursively found from the array, we can compute the smoothed estimators $\hat{\mathbf{x}}_{i|N}$ as

$$\hat{\mathbf{x}}_{i|N} = (P_i^{1/2})\left[(P_i^{-1/2}\hat{\mathbf{x}}_i) + (P_i^{*/2}\boldsymbol{\lambda}_{i|N})\right]. \tag{12.9.1}$$

Note that we use parenthesis in these formulas to indicate quantities that can be directly read off from the pre- and post-arrays.

To complete the recursions, we need to compute the error covariances. However, it can soon be realized that it is impossible to find the quantities $\Lambda_{i|N}$ in (10.2.9) by directly using X and Y. Therefore, we separate $(P_i^{*/2}\boldsymbol{\lambda}_{i|N})$ into $(P_i^{*/2}\Lambda_{i|N}^{1/2})$ and $(\Lambda_{i|N}^{-1/2}\boldsymbol{\lambda}_{i|N})$ and form another array using X, Y, and $(P_i^{*/2}\Lambda_{i|N}^{1/2})$:

$$\left[\begin{array}{cc} X & (Y)(P_{i+1}^{*/2}\Lambda_{i+1|N}^{1/2}) \\ \hline (\mathbf{e}_i^*R_{e,i}^{-*/2}) & (\boldsymbol{\lambda}_{i+1|N}^*\Lambda_{i+1|N}^{-*/2}) \end{array}\right] \Theta_2 = \left[\begin{array}{cc} W & 0 \\ \hline \boldsymbol{\mu}^* & \nu^* \end{array}\right],$$

where Θ_2 is any unitary matrix that nulls out the (1,2) entry of the pre-array. Inner or cross-products of the array entries yield

$$WW^* = XX^* + (Y)(P_{i+1}^{*/2}\Lambda_{i+1|N}^{1/2})(\Lambda_{i+1|N}^{*/2}P_{i+1}^{1/2})(Y)^* = P_i^{*/2}\Lambda_{i|N}P_i^{1/2},$$

$$\boldsymbol{\mu} = W^{-1}[(X)(R_{e,i}^{-1/2}\mathbf{e}_i) + (Y)(\Lambda_{i+1|N}^{-1/2}\boldsymbol{\lambda}_{i+1|N})] = \Lambda_{i|N}^{-1/2}\boldsymbol{\lambda}_{i|N}.$$

Hence, the new array recursively provides $(P_i^{*/2}\Lambda_{i|N}^{1/2})$ and $(\Lambda_{i|N}^{-1/2}\boldsymbol{\lambda}_{i|N})$. The resulting equation for $P_{i|N}$ is

$$P_{i|N} = (P_i^{1/2})\{I - (P_i^{*/2}\Lambda_{i|N}^{1/2})(\Lambda_{i|N}^{*/2}P_i^{1/2})\}(P_i^{*/2}).$$

These results give us an array form of the BF algorithm, which we summarize below:

Step 1: With $(P_0^{-1/2}\hat{\mathbf{x}}_0) = 0$ and $(P_0^{1/2}) = \Pi_0^{1/2}$, propagate $(P_i^{-1/2}\hat{\mathbf{x}}_i)$ and $(P_i^{1/2})$ using the forwards recursion:

$$\left[\begin{array}{ccc} R_i^{1/2} & H_iP_i^{1/2} & 0 \\ 0 & F_iP_i^{1/2} & G_iQ_i^{1/2} \\ \hline 0 & I & 0 \\ \mathbf{y}_i^*R_i^{-*/2} & -\hat{\mathbf{x}}_i^*P_i^{-*/2} & 0 \end{array}\right] \Theta_1 = \left[\begin{array}{ccc} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \\ \hline P_i^{*/2}H_i^*R_{e,i}^{-*/2} & P_i^{*/2}F_{p,i}P_{i+1}^{-*/2} & (*) \\ \mathbf{e}_i^*R_{e,i}^{-*/2} & -\hat{\mathbf{x}}_{i+1}^*P_{i+1}^{-*/2} & (*) \end{array}\right],$$

where Θ_1 is any unitary matrix that lower triangularizes the first and second rows of the pre-array. During this forwards recursion, generate and save the following variables for step 2, $\{(P_i^{*/2}H_i^*R_{e,i}^{-*/2}), (P_i^{*/2}F_{p,i}P_{i+1}^{-*/2}), (R_{e,i}^{-1/2}\mathbf{e}_i)\}$, and the following variables for step 3, $\{(P_i^{1/2}), (P_i^{-1/2}\hat{\mathbf{x}}_i)\}$.

Step 2: With $(\Lambda_{N+1|N}^{-1/2}\boldsymbol{\lambda}_{N+1|N}) = 0$ and $(P_{N+1|N}^{*/2}\Lambda_{N+1|N}^{1/2}) = 0$, propagate $(\Lambda_{i|N}^{-1/2}\boldsymbol{\lambda}_{i|N})$ using the backwards recursion:

$$\left[\begin{array}{cc} (P_i^{*/2}F_{p,i}P_{i+1}^{-*/2})(P_{i+1}^{*/2}\Lambda_{i+1|N}^{1/2}) & (P_i^{*/2}H_i^*R_{e,i}^{-*/2}) \\ (\boldsymbol{\lambda}_{i+1|N}^*\Lambda_{i+1|N}^{-*/2}) & (\mathbf{e}_i^*R_{e,i}^{-*/2}) \end{array}\right] \Theta_2 = \left[\begin{array}{cc} (P_i^{*/2}\Lambda_{i|N}^{1/2}) & 0 \\ (\boldsymbol{\lambda}_{i|N}^*\Lambda_{i|N}^{-*/2}) & (*) \end{array}\right],$$

where Θ_2 is any unitary matrix that nulls out the (1,2) entry of the pre-array.

Step 3: Calculate the smoothed estimators and their error covariances via

$$\hat{\mathbf{x}}_{i|N} = (P_i^{1/2})\{(P_i^{-1/2}\hat{\mathbf{x}}_i) + (P_i^{*/2}\Lambda_{i|N}^{1/2})(\Lambda_{i|N}^{-1/2}\boldsymbol{\lambda}_{i|N})\}, \tag{12.9.2}$$

$$P_{i|N} = (P_i^{1/2})\{I - (P_i^{*/2}\Lambda_{i|N}^{1/2})(P_i^{*/2}\Lambda_{i|N}^{1/2})^*\}(P_i^{1/2})^*. \tag{12.9.3}$$

12.9.2 Rauch-Tung-Striebel Formulas in Array Form

The RTS smoothing formulas followed closely from the BF formulas, so it is not surprising that they have similar array algorithms:

Step 1: This is the same as Step 1 in the BF case above.

Step 2: With $(P_{N+1}^{-1/2}\hat{\mathbf{x}}_{N+1}) = 0$ and $(P_{N+1}^{-1/2}P_{N+1}P_{N+1}^{-*/2}) = I$, propagate $(P_i^{-1/2}\hat{\mathbf{x}}_{i|N})$ and $(P_i^{-1/2}P_{i|N}P_i^{-*/2})$, using the following backwards equations:

$$\begin{aligned}(P_i^{-1/2}\hat{\mathbf{x}}_{i|N}) &= (P_i^{-1/2}\hat{\mathbf{x}}_i) + \{(P_i^{*/2}H_i^*R_{e,i}^{-*/2})(R_{e,i}^{-1/2}\mathbf{e}_i)\} \\ &\quad + (P_i^{*/2}F_{p,i}P_{i+1}^{-*/2})\{(P_{i+1}^{-1/2}\hat{\mathbf{x}}_{i+1|N}) - (P_{i+1}^{-1/2}\hat{\mathbf{x}}_{i+1})\}, \\ (P_i^{-1/2}P_{i|N}P_i^{-*/2}) &= I - (P_i^{*/2}H_i^*R_{e,i}^{-*/2})(P_i^{*/2}H_i^*R_{e,i}^{-*/2})^* - \\ &\quad - (P_i^*F_{p,i}P_{i+1}^{-*/2})(P_i^*F_{p,i}P_{i+1}^{-*/2})^* \\ &\quad + (P_i^*F_{p,i}P_{i+1}^{-*/2})(P_{i+1}^{-1/2}P_{i+1|N}P_{i+1}^{-*/2})(P_i^*F_{p,i}P_{i+1}^{-*/2})^*.\end{aligned}$$

Step 3: Calculate the smoothed estimators and their error covariances via

$$\hat{\mathbf{x}}_{i|N} = (P_i^{1/2})(P_i^{-1/2}\hat{\mathbf{x}}_{i|N}), \tag{12.9.4}$$

$$P_{i|N} = (P_i^{1/2})(P_i^{-1/2}P_{i|N}P_i^{-*/2})(P_i^{*/2}). \tag{12.9.5}$$

12.9.3 Two-Filter (or Mayne-Fraser) Array Formulas

When compared with the BF and RTS formulas, the two-filter formula is computationally the most intensive. However, it turns out that its array form is the least-complex.

Step 1 (*Forwards Estimator*): With $(P_0^{1/2}) = \Pi_0^{1/2}$ and $(P_0^{-1/2}\hat{\mathbf{x}}_0) = 0$, employ the array equation

$$\left[\begin{array}{ccc} R_i^{1/2} & H_iP_i^{1/2} & 0 \\ 0 & F_iP_i^{1/2} & G_iQ_i^{1/2} \\ \hline -\mathbf{y}_i^*R_{-*/2} & \hat{\mathbf{x}}_i^*P_i^{-*/2} & 0 \end{array}\right]\Theta^f = \left[\begin{array}{ccc} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \\ \hline -\mathbf{e}_i^*R_{e,i}^{-*/2} & \hat{\mathbf{x}}_{i+1}^*P_{i+1}^{-*/2} & (*) \end{array}\right],$$

where Θ^f is any unitary matrix that lower triangularizes the pre-array. Save the variables $\{P_i^{1/2},\ P_i^{-1/2}\hat{\mathbf{x}}_i\}$.

Step 2 (*Backwards Estimator*): With

$$P_{N+1,\infty|N+1}^{-b*/2} = 0, \qquad P_{N+1,\infty|N+1}^{-b/2}\hat{\mathbf{x}}_{N+1,\infty|N+1}^{b} = 0,$$

employ the array equation

$$\left[\begin{array}{ccc} Q_i^{-*/2} & G_i^*P_{i+1,\infty|i+1}^{-b*/2} & 0 \\ 0 & F_i^*P_{i+1,\infty|i+1}^{-b*/2} & H_i^*R_i^{-*/2} \\ \hline 0 & \hat{\mathbf{x}}_{i+1,\infty|i+1}^{b*}P_{i+1,\infty|i+1}^{-b*/2} & \mathbf{y}_i^*R_i^{-*/2} \end{array}\right]\Theta^b = \left[\begin{array}{ccc} R_{b,i}^{1/2} & 0 & 0 \\ (*) & P_{i,\infty|i}^{-b*/2} & 0 \\ \hline (*) & \hat{\mathbf{x}}_{i,\infty|i}^{b*}P_{i,\infty|i}^{-b*/2} & (*) \end{array}\right],$$

where Θ^b is any unitary matrix that lower triangularizes the pre-array. Save the variables $P_{i,\infty|i}^{-b*/2}$ and $P_{i,\infty|i}^{-b/2}\hat{\mathbf{x}}_{i,\infty|i}^{b}$.

Step 3 (*Smoothed Estimator*): Use the array

$$\left[\begin{array}{cc} P_i^{*/2}P_{i,\infty|i}^{-b*/2} & I \\ 0 & P_i^{1/2} \\ \hline \hat{\mathbf{x}}_{i,\infty|i}^{b*}P_{i,\infty|i}^{-b*/2} & 0 \end{array}\right]\Theta^s = \left[\begin{array}{cc} P_i^{*/2}P_{i|N}^{-*/2} & 0 \\ P_{i|N}^{1/2} & (*) \\ \hline \hat{\mathbf{x}}_{s1,i}^*P_{i|N}^{-*/2} & (*) \\ \hat{\mathbf{x}}_{s2,i}^*P_{i|N}^{-*/2} & (*) \end{array}\right], \tag{12.9.6}$$

where Θ^s is a unitary matrix that nulls out the (1,2) entry of the pre-array and $\hat{\mathbf{x}}_{s1,i}$ and $\hat{\mathbf{x}}_{s2,i}$ denote

$$\hat{\mathbf{x}}_{s1,i} = P_{i|N}P_i^{-1}\hat{\mathbf{x}}_i, \qquad \hat{\mathbf{x}}_{s2,i} = P_{i|N}P_{i,\infty|i}^{-b}\hat{\mathbf{x}}_{i,\infty|i}^{b}.$$

The smoothed estimators and the error covariances can then be computed via

$$\hat{\mathbf{x}}_{s,i} = (P_{i|N}^{1/2})\{(P_{i|N}^{-1/2}\hat{\mathbf{x}}_{s1,i}) + (P_{i|N}^{-1/2}\hat{\mathbf{x}}_{s2,i})\}, \qquad P_{i|N} = (P_{i|N}^{1/2})(P_{i|N}^{1/2})^*. \tag{12.9.7}$$

12.10 COMPLEMENTS

The idea of propagating square-root factors, rather than the matrices themselves, was first introduced by Potter (as cited in Potter and Stern (1963) and Battin (1964, pp. 339–340)) for the measurement-update step of the Kalman filter recursions. However, Potter's equation-based method was not (fully) applicable to the time-update problem, for which one needs an array formulation (see Sec. 12.3.2). Such a time-update algorithm was deduced by Schmidt (1970), who combined it with Potter's measurement-update algorithm (used in the Apollo moon landing system) for application in an airborne navigation system for precision approach and landing. The successful use of square-root algorithms in these and other applications made it possible, as described in the historical review paper of McGee and Schmidt (1985), to "place the square-root Kalman filter above all suspicion when software problems occurred. With the standard formulation, the filter would always have been suspect because of its propensity to diverge or develop negative eigenvalues, which could cause very peculiar transients in the navigation estimate." Other investigators further pursued the square-root implemen-

tations. In particular, Carlson (1973) presented a variation of Potter's algorithm that gave the square-root factor $P_{i|i}^{1/2}$ in a triangular form, thus reducing the number of computations and making the algorithm closer in speed to the standard Kalman filter (see Grewal and Andrews (1993, Sec. 6.3.2)).

Independently of these developments, Golub (1965) and Businger and Golub (1965) elaborated the method of orthogonal triangularization suggested by Householder (1953, Ch. 5) to solve the deterministic least-squares problem; we described this method in Sec. 2.5 and showed in Sec. 4.3 that it also followed naturally from the innovations approach to this problem. The papers of Golub (1965) and Businger and Golub (1965) sparked the development of the information-form array algorithms for state-space filtering and smoothing by Hanson and Lawson (1969) and Dyer and McReynolds (1969) at the Jet Propulsion Laboratories (JPL) in Pasadena, California. Using the duality between covariance and information forms, some extensions and a nice review and several new results (especially the array form of Sec. 12.3.3 for the covariance form measurement updates) were presented by Kaminski, Bryson, and Schmidt (1971). Further developments were made by Agee and Turner (1972) and Carlson (1973), followed by the so-called UD measurement-update algorithm by Bierman (1974), which avoided the use of any arithmetic square roots. Bierman refined and implemented these algorithms in JPL software and was very active in advocating their use in place of the traditional algorithms through his book (Bierman (1977)) and his papers (see, *e.g.*, Bierman and Thornton (1977)). Morf and Kailath (1975) presented a new approach to the study of square-root algorithms, which is the one presented in this chapter. One bonus of their approach was the recognition that the UD algorithm was equivalent to the use of square-root free Givens transformations in place of the conventional Givens transformations. Another bonus is a direct approach to fast square-root algorithms — see Ch. 13.

In recent times, array algorithms have also been derived to great effect, in many different areas, especially adaptive filtering (see, *e.g.*, Sayed and Kailath (1994b)), $\mathcal{H}_\infty$ filtering (see, *e.g.*, Hassibi, Sayed, and Kailath (1999)), and numerical linear algebra (see Kailath and Sayed (1999)).

Sec. 12.4. Numerical Examples. A word of caution is in order for the use of unitary transformations, as described in App. B, when dealing with complex-valued as opposed to real-valued data; complex data is especially important in many communications and signal processing applications. This is because unitary transformations preserve norms and angles and, therefore, the pre- and post-arrays must satisfy certain norm and angle-invariance properties. In the complex case, this imposes a certain phase requirement on the leading entry of the post-array. The details are explained in App. B for different classes of rotation matrices (Householder, Givens, hyperbolic, etc.).

Sec. 12.7. Paige's Form of the Array Algorithm. In this section we provided a newer array formulation that is especially useful when dealing with ill-conditioned data in finite-precision implementations. Rather than propagating a square-root factor of the covariance matrix P_i, this algorithm propagates factors of $P_i^{1/2}$ itself. In this way, it avoids matrix products while forming the pre-arrays and leads to improved (in fact, stable) numerical performance albeit at some increased computational cost. This algorithm was proposed by Paige (1985) as a result of studies by Paige (1979a,1979b) and Paige and Saunders (1977) on the development of a numerically reliable algorithm for the solution of a generalized least-squares problem. Once again we see here a fruitful interaction between the deterministic least-squares problem and the stochastic state-space estimation problem.

PROBLEMS

12.1 **(Uniqueness of square-root factors)** Let P be an $n \times n$ positive-definite matrix.

(a) By following the discussion in App. A on the LDU factorization of strongly regular matrices, show that P has a unique $n \times n$ square-root factor A such that $P = AA^*$ and A is lower triangular and has positive diagonal entries.

Remark. Such an A is called the Cholesky factor of P. ◆

(b) Use the spectral (eigenvector-eigenvalue) decomposition to show that P has a unique $n \times n$ square-root factor A such that A is nonsingular, Hermitian ($A = A^*$), and $P = A^2$.

12.2 **(Nonnegative-definite matrices)** Consider a Hermitian matrix P of the form

$$P = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix}, \quad Q = Q^*, \quad C = C^*.$$

Show that each of the following two conditions is equivalent to $P \geq 0$:

(a) $C \geq 0$, $A - BC^\dagger B^* \geq 0$, and $B(I - CC^\dagger) \geq 0$.

(b) $A \geq 0$, $C - B^* A^\dagger B \geq 0$, and $B^*(I - AA^\dagger) \geq 0$.

Remark. The notation $C^\dagger$ denotes the pseudo-inverse of C, which can be defined in terms of the SVD of C — see App. A. ◆

12.3 **(Rank-deficient square-root factors)** Let P be an $n \times n$ nonnegative-definite matrix, $P \geq 0$. Assume the $(0, 0)$ entry of P is zero, say

$$P = \begin{bmatrix} 0 & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \triangleq \begin{bmatrix} 0 & b \\ b^* & C \end{bmatrix},$$

where b is a row vector.

(a) Show that the first row and column of P have to be identically zero, *i.e.*, P has to be of the form

$$P = \begin{bmatrix} 0 & 0 \\ 0 & C \end{bmatrix}.$$

Consider both cases when C is full rank and when C is rank-deficient.

(b) Assume C is full rank and let $C = \bar{A}\bar{A}^*$ be its unique square-root factorization with a lower triangular $\bar{A}$ that has positive diagonal entries. Verify that

$$P = \begin{bmatrix} 0 & \\ 0 & \bar{A} \end{bmatrix} \begin{bmatrix} 0 & \\ 0 & \bar{A} \end{bmatrix}^*.$$

More generally, follow the discussion in App. A on the LDU factorization of strongly regular matrices to show that any nonnegative matrix P of rank β has a unique $n \times n$ square-root factor A such that $P = AA^*$, A is lower triangular, and A has β positive-diagonal entries and $n - \beta$ zero columns.

12.4 **(Rank-one modifications of the identity matrix)** Let x be a column vector with possibly complex entries, and α a real number. Define the matrix $X = I - \alpha xx^*$, which is simply a rank-one modification of the identity matrix.

(a) Show that $Y = XX^*$ can be expressed in the same form, *viz.*, $XX^* = I - \beta xx^*$, for some real number β. Express β in terms of α and $\|x\|^2$.

(b) Given $Y = I - \beta xx^*$, for some real scalar β, show that it admits a Hermitian square-root factor of the same form, *i.e.*, $Y^{1/2} = I - \alpha xx^*$, for some real scalar α, iff $(1 - \beta\|x\|^2) \geq 0$. In this case, solve for α in terms of β and $\|x\|^2$. In general, will the resulting square-root factor $Y^{1/2}$ be triangular?

12.5 **(Andrews' formula)** Show that the measurement-update equation $P_{i|i} = P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i$, where H_i is a $p \times n$ matrix, can be written as

$$P_{i|i}^{1/2} = P_i^{1/2}\left[I - A_i R_{e,i}^{-*/2}(R_{e,i}^{1/2} + R_i^{1/2})^{-1} A_i^*\right],$$

with $R_{e,i} = R_i + A_i^* A_i$, $A_i = P_i^{*/2} H_i^*$.

Remark. See Andrews (1968). ◆

12.6 **(Potter's formula for information time update)** Refer to the information time-update formula (12.8.5). Assume G_i is a column vector and Q_i is a scalar. Follow the arguments at the beginning of Sec. 12.2 and derive a formula, similar to Potter's formula (12.2.6), relating $P_{i+1}^{-1/2}$ and $P_{i|i}^{-1/2}$.

Remark. See Dyer and McReynolds (1969, Sec. 6.2). ◆

12.7 **(Update of predicted estimators)** Consider the time and measurement arrays (12.3.4) and (12.3.8), respectively. Let Θ_1 denote the rotation matrix in (12.3.8) and let Θ_2 denote the rotation matrix in (12.3.4). We wish to employ these array equations in order to derive the update equation (12.3.10) for the predicted estimators.

(a) Use the time-update array to conclude that

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \begin{bmatrix} \Theta_1 & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix}.$$

(b) Now invoke the measurement-update array to conclude that

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \end{bmatrix},$$

where Θ is given by the product $\text{diag}\{\Theta_1, I\} \cdot \text{diag}\{I, \Theta_2\}$.

12.8 **(Update of filtered estimators)** Consider the setting of Prob. 12.7.

(a) Use the time-update array to conclude that

$$\begin{bmatrix} R_{i+1}^{1/2} & H_{i+1} P_{i+1}^{1/2} & 0 \\ 0 & P_{i+1}^{1/2} & 0 \end{bmatrix} \begin{bmatrix} \Theta_1 & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 & 0 \\ \bar{K}_{f,i+1} & P_{i+1|i+1}^{1/2} & 0 \end{bmatrix}.$$

(b) Use the measurement-update array to conclude that

$$\begin{bmatrix} R_{i+1}^{1/2} & H_{i+1} F_i P_{i|i}^{1/2} & H_{i+1} G_i Q_i^{1/2} \\ 0 & F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \Theta_2 \end{bmatrix} = \begin{bmatrix} R_{i+1}^{1/2} & H_{i+1} P_{i+1}^{1/2} & 0 \\ 0 & P_{i+1}^{1/2} & 0 \end{bmatrix}.$$

(c) Conclude that

$$\begin{bmatrix} R_{i+1}^{1/2} & H_{i+1} F_i P_{i|i}^{1/2} & H_{i+1} G_i Q_i^{1/2} \\ 0 & F_i P_{i|i}^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 & 0 \\ \bar{K}_{f,i+1} & P_{i+1|i+1}^{1/2} & 0 \end{bmatrix},$$

where Θ is given by the product $\text{diag}\{\Theta_1, I\} \cdot \text{diag}\{I, \Theta_2\}$.

12.9 **(A numerical example)** Consider a 3-state 1-input state-space estimation problem with the following values at time i:

$$F_i = \begin{bmatrix} 0.3 & 0.1 & 0.2 \\ -0.2 & 0.4 & 0.3 \\ -0.7 & 0.6 & -0.2 \end{bmatrix}, \quad G_i = \begin{bmatrix} 0.7 \\ 0.5 \\ 0.6 \end{bmatrix}, \quad Q_i = 0.5, \quad P_{i|i}^{1/2} = \begin{bmatrix} 0.4 & 0 & 0 \\ 0.3 & 0.6 & 0 \\ 0.6 & 0.4 & 0.7 \end{bmatrix}.$$

Compute a lower triangular square-root factor $P_{i+1}^{1/2}$ by

(a) Using Givens rotations.

(b) Using Householder transformations.

(c) Using a mixture of Givens and Householder transformations.

(d) Using square-root free rotations.

12.10 **(Hyperbolic update of $P_{i|i}^1/2$)** Consider the measurement-update equation (12.1.7). Argue that $P_{i|i}^{1/2}$ can be evaluated by triangularizing the pre-array in the following equation:

$$\begin{bmatrix} P_i^{1/2} & P_i H_i^* R_{e,i}^{-*/2} \end{bmatrix} \Theta = \begin{bmatrix} X & 0 \end{bmatrix},$$

where the rotation matrix Θ should satisfy $\Theta(I_n \oplus -I_q)\Theta^* = (I_n \oplus -I_q)$. [We say that Θ is $(I_n \oplus -I_q)$−unitary.]

12.11 **(Update of predicted estimators when $S_i \neq 0$)** Consider the equations (12.1.3) and (12.1.4) with $S_i \neq 0$.

(a) Verify that $\{R_{e,i}^{1/2}, K_i R_{e,i}^{-*/2}, P_{i+1}^{1/2}\}$ can be updated by unitarily triangularizing the pre-array in the equation

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ G_i S_i R_i^{-*/2} & F_i P_i^{1/2} & G_i Q_i^{s/2} \end{bmatrix} \Theta = \begin{bmatrix} X & 0 & 0 \\ Y & Z & 0 \end{bmatrix},$$

where $Q_i^s = Q_i - S_i R_i^{-1} S_i^*$. Show this by making the identifications

$$X = R_{e,i}^{1/2}, \quad Y = K_i R_{e,i}^{-*/2} = \bar{K}_{p,i}, \quad Z = P_{i+1}^{1/2}.$$

(b) Obtain the same arrays by using the geometric arguments of Sec. 12.6.

(c) Assume $R_i > 0$ and introduce the quantities

$$F_i^s = F_i - G_i S_i R_i^{-1} H_i, \quad K_i^s = F_i^s P_i H_i^*, \quad K_{p,i}^s = K_i^s R_{e,i}^{-1}, \quad \bar{K}_{p,i}^s = K_{p,i}^s R_{e,i}^{1/2}.$$

Argue that the factors $\{R_{e,i}^{1/2}, \bar{K}_{p,i}^s, P_{i+1}^{1/2}\}$ can be updated by unitarily triangularizing the pre-array in the equation

$$\begin{bmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i^s P_i^{1/2} & G_i Q_i^{s/2} \end{bmatrix} \Theta = \begin{bmatrix} X & 0 & 0 \\ Y & Z & 0 \end{bmatrix}.$$

Identify the $\{X, Y, Z\}$.

(d) Assume R_i is not strictly positive-definite but that $Q_i > 0$, and define $R_i^s = R_i - S_i^* Q_i^{-1} S_i$. Verify the validity of the following array algorithm:

$$\begin{bmatrix} R_i^{s/2} & H_i P_i^{1/2} & S_i^* Q_i^{-*/2} \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \end{bmatrix},$$

where Θ is a unitary rotation that triangularizes the pre-array.

12.12 (Information arrays for the time update when $S_i \neq 0$) When $S_i \neq 0$, the time-update equations are given by

$$\hat{\mathbf{x}}_{i+1} = F_i^s \hat{\mathbf{x}}_{i|i} + G_i S_i R_i^{-1} \mathbf{y}_i, \quad P_{i+1} = F_i^s P_{i|i} F_i^{s*} + G_i Q_i^s G_i^*,$$

where $F_i^s = F_i - G_i S_i R_i^{-1} H_i$ and $Q_i^s = Q_i - S_i R_i^{-1} S_i^*$.

(a) Use the matrix inversion lemma to show that

$$\begin{aligned} P_{i+1}^{-1} &= A_i - A_i G_i (Q_i^{-s} + G_i^* A_i G_i)^{-1} G_i^* A_i, \\ P_{i+1}^{-1} \hat{\mathbf{x}}_{i+1} &= (I + A_i G_i Q_i G_i^*)^{-1} (F_i^{-*s} P_{i|i}^{-1} \hat{\mathbf{x}}_{i|i}) + \\ &\qquad (A_i - G_i (Q_i^{-s} + G_i^* A_i G_i)^{-1} G_i^* A_i) G_i S_i R_i^{-1} \mathbf{y}_i, \end{aligned}$$

where $A_i = F_i^{-*s} P_{i|i}^{-1} F_i^{-s}$.

(b) Define $Q_i^{-s} + G_i^* A_i G_i \triangleq Q_i^{-r}$ and $Q_i^r = Q_i^s - Q_i^s G_i^* P_{i+1}^{-1} G_i Q_i^s$. Verify the validity of the following array equations

$$\begin{bmatrix} Q_i^{-*s/2} & G_i^* F_i^{-*s} P_{i|i}^{-*/2} \\ 0 & F_i^{-*s} P_{i|i}^{-*/2} \end{bmatrix} \Theta_{tu} = \begin{bmatrix} Q_i^{-*r/2} & 0 \\ F_i^{-*s} P_{i|i}^{-1} F_i^{-s} G_i Q_i^{r/2} & P_{i+1}^{-*/2} \end{bmatrix},$$

where Θ_{tu} is any unitary rotation that zeros out the $(1,1)$ block element in the post-array.

(c) Show that for the special case $S_i = 0$, we can add an extra line to the above arrays in order to propagate $P_{i+1}^{-1/2} \hat{\mathbf{x}}_{i+1}$ as well, *i.e.*, verify that the following holds:

$$\begin{bmatrix} Q_i^{-*/2} & G_i^* F_i^{-*} P_{i|i}^{-*/2} \\ 0 & F_i^{-*} P_{i|i}^{-*/2} \\ 0 & \hat{\mathbf{x}}_{i|i}^* P_{i|i}^{-*/2} \end{bmatrix} \Theta_{tu} = \begin{bmatrix} Q_i^{-*r/2} & 0 \\ F_i^{-*} P_{i|i}^{-1} F_i^{-1} G_i Q_i^{r/2} & P_{i+1}^{-*/2} \\ \hat{\mathbf{x}}_{i|i}^* P_{i|i}^{-1} F_i^{-1} G_i Q_i^{r/2} & \hat{\mathbf{x}}_{i+1}^* P_{i+1}^{-*/2} \end{bmatrix}.$$

12.13 (Combined arrays) Verify that the covariance and information form arrays (12.3.10) and (12.8.19), respectively, can be combined as follows (assuming $S_i = 0$, $R_i > 0$, $\Pi_0 > 0$, and F_i invertible):

$$\left[\begin{array}{ccc} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \\ \hline -F_i^{-*} H_i^* R_i^{-*/2} & F_i^{-*} P_i^{-*/2} & 0 \\ Q_i^{*/2} G_i^* F_i^{-*} H_i^* R_i^{-*/2} & -Q_i^{*/2} G_i^* F_i^{-*} P_i^{-*/2} & I \\ \hline -\mathbf{y}_i^* R_i^{-*/2} & \hat{\mathbf{x}}_i^* P_i^{-*/2} & 0 \end{array}\right] \Theta =$$

$$\left[\begin{array}{ccc} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \\ \hline 0 & P_{i+1}^{-*/2} & (*) \\ 0 & 0 & (*) \\ \hline -\mathbf{e}_i^* R_{e,i}^{-*/2} & \hat{\mathbf{x}}_{i+1}^* P_{i+1}^{-*/2} & (*) \end{array}\right],$$

where $(*)$ denotes "don't care" entries, and Θ is any unitary rotation that either lower triangularizes the first two (block) rows or upper triangularizes the third and fourth (block) rows of the pre-array.

Remark. This array algorithm and others (*e.g.*, combined arrays for the time- and measurement-update formulas when $S_i \neq 0$) can be found in Park and Kailath (1995a). The special case of $G_i = 0$ was derived in Sayed and Kailath (1994b) and applied to adaptive RLS filtering. ◆

12.14 (The two-filter smoothing algorithm) Verify the validity of the equations (12.9.6)–(12.9.7) that describe the array form of the two-filter smoothing algorithm.

12.15 (Paige's algorithm) Refer to Alg. 12.C.1.

(a) Apply it to the following data:

$$F_0 = 1, \quad G_0 = 0, \quad H_0 = 1, \quad R_0 = 1, \quad \Pi_0 = 1,$$

and verify that it leads to $A_1 = \sqrt{2}/\sqrt{3}$ and $B_1 = 1/\sqrt{3}$. Conclude that $P_1^{1/2} = 1/\sqrt{2}$.

(b) Apply it now to the following data (which refers to the case studied in Sec. 12.1.3):

$$F_0 = 1, \quad G_0 = 0, \quad H_0 = 1, \quad R_0 = 1,$$

and assume Π_0 is large enough that $\Pi_0 + 1 = \Pi_0$ in finite precision arithmetic.

12.16 **(Uncorrelated variables)** Consider the variable $\tilde{\mathbf{x}}_0^{(1)}$ defined in (12.C.7) and partition the matrices W_0 and Z_0 in (12.C.5) and (12.C.7) into

$$W_0 = \begin{bmatrix} W_1 & W_2 \\ W_3 & W_4 \end{bmatrix}, \quad Z_0 = \begin{bmatrix} Z_1 & Z_2 \\ Z_3 & Z_4 \end{bmatrix}.$$

Recall further that $\mathbf{y}_0 = H_0\Pi_0^{1/2}\tilde{\mathbf{x}}_0 + R^{1/2}\tilde{\mathbf{v}}_0$.

(a) Show that because of the equality (12.C.5), W_4 is nonsingular.

(b) Verify that $< \mathbf{y}_0, \tilde{\mathbf{x}}_0^{(1)} > = H_0\Pi_0^{1/2}Z_1 - R_0^{1/2}Z_3$.

(c) Use the definitions of W_0 and Z_0 to conclude that

$$< \mathbf{y}_0, \tilde{\mathbf{x}}_0^{(1)} > = \begin{bmatrix} H_0 & -I \end{bmatrix} W_0^* \begin{bmatrix} C_5 \\ 0 \end{bmatrix}.$$

(d) Show that $\begin{bmatrix} H_0 & -I \end{bmatrix} W_0^*$ has the form $\begin{bmatrix} H_0 & -I \end{bmatrix} W_0^* = \begin{bmatrix} 0 & \times \end{bmatrix}$. Conclude that $\mathbf{y}_0$ and $\tilde{\mathbf{x}}_0^{(1)}$ are uncorrelated random variables.

Appendices for Chapter 12

12.A THE UD ALGORITHM

An early enthusiast for the use of array algorithms was G. J. Bierman, of the Jet Propulsion Laboratories, Pasadena, California, who strongly advocated their use in place of the usual Kalman filter recursions — see, *e.g.*, Bierman and Thornton (1977). Bierman emphasized that, despite the common belief, array algorithms when properly coded were not more expensive than the usual recursions. He advocated a so-called UD algorithm for the measurement-update problem with scalar measurements. This algorithm avoids the use of arithmetic square roots. As noted in the text, this can be achieved by using the modified Givens transformations described in App. B; a numerical example was given in Sec. 12.4.3. However, since Bierman's algorithm is available in some computer packages, we present it explicitly here, along with an outline of his original derivation.

The measurement-update equation with scalar measurements is

$$P_{i|i} = P_i - P_i h_i^* r_e^{-1}(i) h_i P_i, \tag{12.A.1}$$

with h_i a row vector and $r_e(i) = r(i) + h_i P_i h_i^*$, a scalar. To somewhat reduce the notational burden, we shall omit the subscript i, and write $P_i = P$, $P_{i|i} = P_+$, and consider the factored forms

$$P = LDL^*, \quad P_+ = L_+ D_+ L_+^*, \tag{12.A.2}$$

where L is lower triangular with unit diagonal and D is diagonal. The algorithm will show how to obtain $\{L_+, D_+\}$ given $\{L, D, r_e, h\}$. [For certain reasons, Bierman used the factorizations $P = UDU^*$, $P_+ = U_+ D_+ U_+^*$, with U upper triangular with unit diagonal. For consistency with the material in this chapter, we stay with the LDL^* factorization.]

Next denote

$$hL \triangleq f = \begin{bmatrix} f_1 & f_2 & \dots & f_n \end{bmatrix},$$

$$L = \begin{bmatrix} l_1 & l_2 & \dots & l_n \end{bmatrix}, \quad L_+ = \begin{bmatrix} l_1^+ & l_2^+ & \dots & l_n^+ \end{bmatrix},$$

and

$$D = \text{diag}\{d_i\}, \quad D_+ = \text{diag}\{d_i^+\}.$$

Algorithm 12.A.1 (Bierman's LD Algorithm) *Initialization. Let* $\alpha_{n+1} = r$, $k_{n+1} = 0 \in \mathbb{C}^n$.

- *For $i = n$ down to 1, do*

$$\begin{aligned}\alpha_i &= \alpha_{i+1} + d_i|f_i|^2,\\ d_i^+ &= d_i(\alpha_{i+1}/\alpha_i),\\ \begin{bmatrix}k_i & l_i^+\end{bmatrix} &= \begin{bmatrix}k_{i+1} & l_i\end{bmatrix}\begin{bmatrix}1 & -f_i/d_i\\ d_i f_i^* & 1\end{bmatrix}.\end{aligned}$$

end

Then $\alpha_0 = r_e$ and $k_0 = Ph^*r_e^{-1}$. ∎

Bierman acknowledges in his book (Bierman (1977, p. 78)) that this algorithm is somewhat "awkward in appearence." The derivation is long, and it is not at all evident that it can be applied to the time-update recursion — in fact, Bierman uses a different (so-called weighted Gram-Schmidt) array algorithm for that problem. His derivation proceeds as follows.

From (12.A.1) and (12.A.2) we can write

$$L_+D_+L_+ = L(D - \underbrace{DL^*h^*r_e^{-1}hLD}_{\text{rank one}})L^*.$$

Now Bierman appeals to a formula of Agee and Turner (1972) for rank-one "downdating." He combines this with a numerical improvement suggested by Carlson (1973) — replacing a subtraction by a division — to finally obtain his algorithm, as presented above. Since the Agee-Turner formula only holds for scalar measurements, a sequentialization procedure (see Sec. 9.5.4) has to be used in the vector case.

The approach we used in the main chapter avoids all these difficulties and in fact shows that there can be many useful variants — we just have to rotate the pre-array using any of a number unitary rotations. Of course, our interest is to reduce the number of computations, but this too can be done in several ways, one of which leads to the Bierman equations.

To explore this a little further, consider the following problem. Given two $n \times 1$ columns $\{x_1, x_2\}$ and two scalars $\{a_1, a_2\}$, find two columns $\{x_1^+, x_2^+\}$ and two scalar $\{a_1^+, a_2^+\}$ such that

$$\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}a_1 & \\ & a_2\end{bmatrix}\begin{bmatrix}x_1^*\\ x_2^*\end{bmatrix} = \begin{bmatrix}x_1^+ & x_2^+\end{bmatrix}\begin{bmatrix}a_1^+ & \\ & a_2^+\end{bmatrix}\begin{bmatrix}x_1^{+*}\\ x_2^{+*}\end{bmatrix}, \tag{12.A.3}$$

and the first component of x_2^+ is zero, *i.e.*, $x_2^+(1) = 0$. Without loss of generality, we can assume that $x_1(1) = 1$. Here are some common methods of achieving this.

1. **Three-multiplies form (Gentleman).**

$$\begin{aligned}a_1^+ &= a_1 + a_2|x_2(1)|^2,\\ a_2^+ &= a_1a_2/a_1^+,\\ x_1^+ &= (a_1/a_1^+)x_1 + (a_2x_2^*(1)/a_1^+)x_2,\\ x_2^+ &= -x_2(1)x_1 + x_2.\end{aligned} \tag{12.A.4}$$

   Then we shall obtain $x_1^+(1) = 1$ and $x_2^+(1) = 0$.

2. **Two-multiplies form (Golub, Gentleman).** Eq. (12.A.4) can be re-arranged as follows:

$$\begin{aligned}x_1^+ &= (a_1/a_1^+)x_1 + (a_2x_2^*(1)/a_1^+)(x_2^+ + x_2(1)x_1),\\ &= (1/a_1^+)(a_1 + a_2|x_2(1)|^2)x_1 + (a_2x_2^*(1)/a_1^+)x_2^+,\\ &= x_1 + (a_2x_2^*(1)/a_1^+)x_2^+.\end{aligned} \tag{12.A.5}$$

   This last formula needs one less multiplication than (12.A.4) but it can be numerically unstable if $|a_1^+| \gg |a_1|$.

3. **Agee-Turner's choice (Bierman).** We re-arrange (12.A.4) as follows:

$$\begin{aligned}a_1^+x_1^+ &= a_1x_1 + a_2x_2^*(1)x_2,\\ x_2^+ &= -(x_2(1)a_1x_1/a_1) + x_2.\end{aligned}$$

   This approach still needs only two multiplications, and can be proved to be numerically stable if $\text{sign}(a_1) = \text{sign}(a_2)$; Bierman's variant uses Agee-Turner's form in just this way. However, there are several other stable two-multiplies forms in the literature, with more being discovered frequently.

12.B THE USE OF SCHUR AND CONDENSED FORMS

Though the number of computations in any of the state-space algorithms of this chapter is $O(n^3)$, there can be variations in the coefficient of n^3 depending upon the nature of the matrices $\{F, G, H, Q, R, S\}$. It would seem to be useful, especially in the time-invariant case, to choose these matrices in as simple a form as possible, or to transform them by similarity transformation to such a form, so as to reduce the complexity of subsequent operations such as $\{FP_iF^*, P_iH^*, \text{etc.}\}$.

The problem with similarity transformations, *e.g.*, to companion form matrices for F, is that the operations may introduce numerical errors themselves. This issue has been studied by system theorists and attention has focused first on unitary transformations, which leave Euclidean norms unchanged and therefore do not enhance the "noise" in the computation.

It is known that any $n \times n$ matrix can be transformed to triangular form by using unitary transformations — a result attributed to Schur, so the resulting form is often

called the Schur form (*e.g.*, Golub and Van Loan (1996)). A detailed calculation shows that the leading coefficient of n^3 in the estimation algorithm can be reduced by at least 50% by going to Schur forms.

It has also been proposed that attention should be paid not only to the form of the F matrix, but also to the forms of G and H. Van Dooren and Verhaegen (1985) proposed the use of certain "condensed" forms in which the $\{G, F\}$ and $\{F, H\}$ pairs are reduced to Hessenberg form.

The Schur and Hessenberg forms are shown below.

1. Triangular Schur Form.

$$\left[\begin{array}{c|c} G & F \\ \hline & H \end{array}\right] = \left[\begin{array}{cc|cccccc} \times & \times & \times & \times & \times & \times & \times & \times \\ \times & \times & 0 & \times & \times & \times & \times & \times \\ \times & \times & 0 & 0 & \times & \times & \times & \times \\ \times & \times & 0 & 0 & 0 & \times & \times & \times \\ \times & \times & 0 & 0 & 0 & 0 & \times & \times \\ \times & \times & 0 & 0 & 0 & 0 & 0 & \times \\ \hline & & \times & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times & \times & \times \end{array}\right]$$

2. Controller Hessenberg Form.

$$\left[\begin{array}{c|c} G & F \\ \hline & H \end{array}\right] = \left[\begin{array}{cc|cccccc} \times & \times & \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \times & \times & \times \\ \hline & & \times & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times & \times & \times \end{array}\right]$$

3. Observer Hessenberg Form.

$$\left[\begin{array}{c|c} G & F \\ \hline & H \end{array}\right] = \left[\begin{array}{cc|cccccc} \times & \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times & \times \\ \times & \times & 0 & \times & \times & \times & \times & \times \\ \times & \times & 0 & 0 & \times & \times & \times & \times \\ \times & \times & 0 & 0 & 0 & \times & \times & \times \\ \hline & & 0 & 0 & 0 & 0 & \times & \times \\ & & 0 & 0 & 0 & 0 & 0 & \times \end{array}\right]$$

The "number" of flops needed to obtain these forms is roughly

$$n^2(5kn + m + p), \quad n^2(3n + p + m), \quad \text{and} \quad n^2(3n + p + m),$$

respectively, where k is usually between 1 and 2. These operation counts include the construction of the corresponding state-space transformation yielding the condensed forms. Notice that the three forms have the same number of zeros as each other, namely $n(n-1)/2$.

The flop counts for the filter implementations that are based on these forms are shown in Table 12.2. These counts are for a combined measurement- and time-update operation. We note that for large n, the array forms are in fact somewhat less expensive than the regular Kalman filter recursions; also that the Schur and Hessenberg forms compare quite favorably with using a general full matrix F. For more on these issues, see Van Dooren and Verhaegen (1985) and Laub and Linnemann (1986).

Table 12.2 Comparison of flop counts for different forms.

Form	Covariance Kalman Filter	Array Covariance Filter
Full	$\frac{3}{2}n^3 + \left(\frac{3p}{2} + \frac{m}{2}\right)n^2 + \left(\frac{3p^2}{2} + m^2\right)n + \frac{p^3}{6}$	$\frac{7}{6}n^3 + \left(\frac{5p}{2} + m\right)n^2 + \left(2p^2 + \frac{m^2}{2}\right)n + \frac{2p^3}{3}$
Schur	$\frac{3}{4}n^3 + (2p + m)n^2 + (2p^2 + m^2)n + \frac{p^3}{6}$	$\frac{1}{3}n^3 + \frac{7p}{2}n^2 + \left(2p^2 + \frac{m^2}{2}\right)n$
Hessenberg	$\frac{3}{4}n^3 + (3p + m)n^2 + \left(\frac{3p^2}{2} + m^2\right)n + \frac{2p^3}{3}$	$\frac{1}{3}n^3 + \left(\frac{3p}{2} + m\right)n^2 + \left(p^2 + \frac{m^2}{2}\right)n + \frac{p^3}{2}$

12.C PAIGE'S ARRAY ALGORITHM

We provide in this appendix a stochastic derivation of Paige's algorithm, mentioned in Sec. 12.7. Thus assume that $\Pi_0 > 0$, $Q_i > 0$, and $R_i > 0$, and let $\Pi_0^{1/2}$, $Q_i^{1/2}$, and $R_i^{1/2}$ denote their (lower) triangular square-root factors. We also assume that the spectral norm of each of the matrices G_i, $\|G_i\|$, is not too large so that matrix products of the form $G_i Q_i^{1/2}$ do not contribute to numerical instability. Recall also the dimensions of the matrices involved:

$$F : n \times n, \quad G : n \times q, \quad H : p \times n, \quad Q : q \times q, \quad R : p \times p.$$

We can now rewrite the standard state-space model (12.1.1), with $S_i = 0$, in the equivalent form

$$\begin{cases} \mathbf{x}_{i+1} &= F_i\mathbf{x}_i + G_iQ_i^{1/2}\bar{\mathbf{u}}_i, \\ \mathbf{y}_i &= H_i\mathbf{x}_i - R_i^{1/2}\bar{\mathbf{v}}_i, \quad i \geq 0, \end{cases} \tag{12.C.1}$$

where $\{\bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$ are now white-noise uncorrelated processes with identity covariance matrices, *viz.*, $\bar{\mathbf{u}}_i = Q_i^{-1/2}\mathbf{u}_i$ and $\bar{\mathbf{v}}_i = -R_i^{-1/2}\mathbf{v}_i$ (with a minus sign, for convenience). We also define $\bar{\mathbf{x}}_0 = \Pi_0^{-1/2}\mathbf{x}_0$, which is uncorrelated with $\{\bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$ and has unit covariance as well. That is, we replace the random variables $\{\mathbf{u}_i, \mathbf{v}_i, \mathbf{x}_0\}$ by equivalent random variables $\{\bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i, \bar{\mathbf{x}}_0\}$ with identity covariance matrices. Solving the state-space estimation problem for the above model is obviously equivalent to solving the state-space estimation problem for the earlier model (12.1.1).

If we now expand the above state-space equations from time 0 up to time $i+1$ we obtain

$$\underbrace{\begin{bmatrix} I & & & & \\ H_0 & & & & \\ -F_0 & I & & & \\ & H_1 & & & \\ & -F_1 & I & & \\ & & \ddots & & \\ & & & H_i & \\ & & & -F_i & I \end{bmatrix}}_{\mathcal{A}} \underbrace{\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \vdots \\ \mathbf{x}_i \\ \mathbf{x}_{i+1} \end{bmatrix}}_{\mathbf{x}} = \tag{12.C.2}$$

$$\underbrace{\begin{bmatrix} \Pi_0^{1/2} & & & & & & & \\ & R_0^{1/2} & & & & & & \\ & & G_0Q_0^{1/2} & & & & & \\ & & & R_1^{1/2} & & & & \\ & & & & G_1Q_1^{1/2} & & & \\ & & & & & \ddots & & \\ & & & & & & R_i^{1/2} & \\ & & & & & & & G_iQ_i^{1/2} \end{bmatrix}}_{\mathcal{W}} \underbrace{\begin{bmatrix} \bar{\mathbf{x}}_0 \\ \bar{\mathbf{v}}_0 \\ \bar{\mathbf{u}}_0 \\ \bar{\mathbf{v}}_1 \\ \bar{\mathbf{u}}_1 \\ \vdots \\ \bar{\mathbf{v}}_i \\ \bar{\mathbf{u}}_i \end{bmatrix}}_{\mathbf{z}} + \underbrace{\begin{bmatrix} 0 \\ \mathbf{y}_0 \\ 0 \\ \mathbf{y}_1 \\ 0 \\ \vdots \\ \mathbf{y}_i \\ 0 \end{bmatrix}}_{\mathbf{y}},$$

which we write more compactly as follows:

$$\mathcal{A}\mathbf{x} = \mathcal{W}\mathbf{u} + \mathbf{y}, \tag{12.C.3}$$

with the quantities $\{\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathcal{A}, \mathcal{W}\}$ defined above. Note that all the entries in $\{\mathcal{A}, \mathcal{W}\}$ are defined directly in terms of the parameters $\{F_i, H_i, G_iQ_i^{1/2}, R_i^{1/2}, \Pi_0^{1/2}\}$. This distinguishes the current representation from the one we encountered earlier in Sec. 5.A, where matrix products of the form $\{H_iG_j\}$ or $\{H_i\Phi(i,j)G_j\}$ were used.

To proceed further we first recall a result proven in Prob. 3.4. Assume B is any nonnegative-definite weighting matrix and consider two k-dimensional zero-mean random vectors $\mathbf{x}$ and $\mathbf{y}$. Let $\hat{\mathbf{x}}$ denote the linear least-mean-squares estimator of $\mathbf{x}$ given $\mathbf{y}$. Choose $B = e_ke_k^*$, where e_k is a column vector with $k-1$ leading zeros and last entry equal to unity (*i.e.*, it is the last basis vector). Then Prob. 3.4 tells us that $\hat{\mathbf{x}}$ also minimizes $\|B(\mathbf{x} - \hat{\mathbf{x}})\|^2$. This means that the last entry of $\hat{\mathbf{x}}$ is also the linear least-mean-squares estimator of the last entry of $\mathbf{x}$ given the same vector $\mathbf{y}$.

Returning to the equation (12.C.3), we now pose the problem of linearly estimating the vector $\mathbf{x}$ from the vector $\mathbf{y}$ in the linear least-mean-squares (l.l.m.s.) sense. Let $\hat{\mathbf{x}}$ denote the resulting estimator. Its last entry will then be, in view of what we have recalled from Prob. 3.4, the l.l.m.s. estimator of $\mathbf{x}_{i+1}$ given $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_i\}$. In other words, it will be $\hat{\mathbf{x}}_{i+1}$. If, on the other hand, we delete the last block row from both sides of (12.C.3), then the last entry of $\mathbf{x}$ becomes $\mathbf{x}_i$ and all the $\{\mathbf{y}_j, j = 0, \ldots, i\}$ continue in $\mathbf{y}$. The problem of estimating $\mathbf{x}$ from $\mathbf{y}$ now leads to a new estimator $\hat{\mathbf{x}}$ whose last entry becomes $\hat{\mathbf{x}}_{i|i}$.

Our solution to the problem of estimating $\mathbf{x}$ from $\mathbf{y}$ in (12.C.3) will proceed in this recursive manner. We shall obtain in sequence

$$\hat{\mathbf{x}}_0 \longrightarrow \hat{\mathbf{x}}_{0|0} \longrightarrow \hat{\mathbf{x}}_1 \longrightarrow \hat{\mathbf{x}}_{1|1} \longrightarrow \ldots \longrightarrow \hat{\mathbf{x}}_{i|i} \longrightarrow \hat{\mathbf{x}}_{i+1},$$

which corresponds to the procedure we obtain by implementing the measurement- and time-update forms of the Kalman filter (Sec. 9.2.4).

In order to help fix ideas, let us consider a subproblem of (12.C.3) with only the state vectors $\{\mathbf{x}_0, \mathbf{x}_1\}$ in $\mathbf{x}$, *viz.*, let us consider the following relation:

$$\begin{bmatrix} I & 0 \\ H_0 & 0 \\ -F_0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \end{bmatrix} = \begin{bmatrix} \Pi_0^{1/2} & 0 & 0 \\ 0 & R_0^{1/2} & 0 \\ 0 & 0 & G_0Q_0^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_0 \\ \bar{\mathbf{v}}_0 \\ \bar{\mathbf{u}}_0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{y}_0 \\ 0 \end{bmatrix}. \tag{12.C.4}$$

It allows us to evaluate $\{\hat{\mathbf{x}}_{0|0}, \hat{\mathbf{x}}_1\}$. Since we are interested in proceeding recursively, we need to perform these evaluations in a convenient manner so that the result of this calculation can be used to progress to the next calculation of finding $\{\hat{\mathbf{x}}_{1|1}, \hat{\mathbf{x}}_2\}$, and so on. In order to show how to achieve this, we shall now exhibit a sequence of transformations that will essentially reduce the coefficient matrices multiplying $\text{col}\{\mathbf{x}_0, \mathbf{x}_1\}$ and $\text{col}\{\bar{\mathbf{x}}_0, \bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$ in the above equation into upper triangular forms.

Our first step is to pre-multiply (12.C.4) by $\Theta_0 = (W_0 \oplus I)$, where W_0 is a unitary matrix that performs the QR decomposition

$$W_0 \begin{bmatrix} I \\ H_0 \end{bmatrix} = \begin{bmatrix} C_0 \\ 0 \end{bmatrix}, \quad C_0 : n \times n, \tag{12.C.5}$$

with C_0 upper triangular (and in fact also invertible). Applying Θ to (12.C.4) leads to the equivalent equality

$$\begin{bmatrix} C_0 & 0 \\ 0 & 0 \\ -F_0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \end{bmatrix} = \begin{bmatrix} C_1 & C_2 & 0 \\ C_3 & C_4 & 0 \\ 0 & 0 & G_0 Q_0^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_0 \\ \bar{\mathbf{v}}_0 \\ \bar{\mathbf{u}}_0 \end{bmatrix} + \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ 0 \end{bmatrix}, \tag{12.C.6}$$

where we have defined

$$\begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} = W_0 \begin{bmatrix} \Pi_0^{1/2} & 0 \\ 0 & R_0^{1/2} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \end{bmatrix} = W_0 \begin{bmatrix} 0 \\ \mathbf{y}_0 \end{bmatrix},$$

$C_1 : n \times n, \quad C_2 : n \times p, \quad C_3 : p \times n, \quad C_4 : p \times p, \quad \mathbf{b}_0 : n \times 1, \quad \mathbf{b}_1 : p \times 1.$

The quantities $\{\mathbf{b}_0, \mathbf{b}_1\}$ are random variables that are linearly related to the observation $\mathbf{y}_0$.

The second step is to upper triangularize the submatrix with the $\{C_i\}$ on the right-hand side of (12.C.6). This will allow us to remove the block row with zeros from the left-hand side. More specifically, let Z_0 be a unitary matrix such that

$$\begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} Z_0 = \begin{bmatrix} C_5 & C_6 \\ 0 & C_7 \end{bmatrix}, \quad C_5 : n \times n, \quad C_6 : n \times p, \quad C_7 : p \times p,$$

and define

$$\begin{bmatrix} \bar{\mathbf{x}}_0^{(1)} \\ \bar{\mathbf{v}}_0^{(1)} \end{bmatrix} \triangleq Z_0^* \begin{bmatrix} \bar{\mathbf{x}}_0 \\ \bar{\mathbf{v}}_0 \end{bmatrix}, \quad \bar{\mathbf{x}}_0^{(1)} : n \times 1, \quad \bar{\mathbf{v}}_0^{(1)} : p \times 1. \tag{12.C.7}$$

The random variable $\bar{\mathbf{x}}_0^{(1)}$ can be shown to be uncorrelated with $\mathbf{y}_0$ (see Prob. 12.16)! Moreover, C_7 is also invertible.

Now, expression (12.C.6) becomes

$$\begin{bmatrix} C_0 & 0 \\ 0 & 0 \\ -F_0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \end{bmatrix} = \begin{bmatrix} C_5 & C_6 & 0 \\ 0 & C_7 & 0 \\ 0 & 0 & G_0 Q_0^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_0^{(1)} \\ \bar{\mathbf{v}}_0^{(1)} \\ \bar{\mathbf{u}}_0 \end{bmatrix} + \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ 0 \end{bmatrix}. \tag{12.C.8}$$

The second block row of the above equality shows that $\bar{\mathbf{v}}_0^{(1)}$ can be determined in terms of $\mathbf{b}_1$ and, hence, is also a linear function of the observation $\mathbf{y}_0$. That is, we obtain $C_7 \bar{\mathbf{v}}_0^{(1)} = \mathbf{b}_1$.

We can now reduce (12.C.8) to the form

$$\begin{bmatrix} C_0 & 0 \\ -F_0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \end{bmatrix} = \begin{bmatrix} C_5 & 0 \\ 0 & G_0 Q_0^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_0^{(1)} \\ \bar{\mathbf{u}}_0 \end{bmatrix} + \begin{bmatrix} \mathbf{b}_2 \\ 0 \end{bmatrix}, \tag{12.C.9}$$

where the new $n \times 1$ random variable $\mathbf{b}_2$ is defined by $\mathbf{b}_2 = \mathbf{b}_0 + C_6 \bar{\mathbf{v}}_0^{(1)}$ and is therefore completely determined (linearly) by $\mathbf{y}_0$. At this stage, we can evaluate $\hat{\mathbf{x}}_{0|0}$ by noting that $C_0 \mathbf{x}_0 = C_5 \bar{\mathbf{x}}_0^{(1)} + \mathbf{b}_2$. Since $\bar{\mathbf{x}}_0^{(1)}$ is uncorrelated with $\mathbf{y}_0$, and the l.l.m.s. estimator of $\mathbf{b}_2$ given $\mathbf{y}_0$ is $\mathbf{b}_2$ itself, we obtain $C_0 \hat{\mathbf{x}}_{0|0} = \mathbf{b}_2$. We now proceed to evaluate $\hat{\mathbf{x}}_1$. For this purpose, we pre-multiply (12.C.9) with a unitary matrix T_0 that upper triangularizes the coefficient matrix on the left-hand side, *viz.*,

$$T_0 \begin{bmatrix} C_0 & 0 \\ -F_0 & I \end{bmatrix} = \begin{bmatrix} C_8 & C_9 \\ 0 & A_1 \end{bmatrix}, \quad \{C_8, C_9, A_1\} \text{ are } n \times n.$$

The invertibility of C_0 guarantees the invertibility of A_1. Applying this transformation to both sides of (12.C.9) we obtain

$$\begin{bmatrix} C_8 & C_9 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{13} & C_{14} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_0^{(1)} \\ \bar{\mathbf{u}}_0 \end{bmatrix} + \begin{bmatrix} \mathbf{b}_3 \\ \mathbf{d}_1 \end{bmatrix}, \tag{12.C.10}$$

where

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{13} & C_{14} \end{bmatrix} = T_0 \begin{bmatrix} C_5 & 0 \\ 0 & G_0 Q_0^{1/2} \end{bmatrix}, \quad \{C_{11}, C_{13} : n \times n\}, \quad \{C_{12}, C_{14} : n \times q\},$$

and the new $n \times 1$ variables $\{\mathbf{b}_3, \mathbf{d}_1\}$ are defined by

$$\begin{bmatrix} \mathbf{b}_3 \\ \mathbf{d}_1 \end{bmatrix} \triangleq T_0 \begin{bmatrix} \mathbf{b}_2 \\ 0 \end{bmatrix},$$

and are, again, linear functions of the observation $\mathbf{y}_0$.

Finally, in order to proceed recursively with our derivation, we need to reduce the matrix with the quantities $\{C_{11}, C_{12}, C_{13}, C_{14}\}$ into upper triangular form, as will become clear shortly. So assume we perform this step and use a unitary transformation V_0 such that (note how the dimensions of the partitioned matrices are modified at this point)

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{13} & C_{14} \end{bmatrix} V_0 = \begin{bmatrix} C_{15} & C_{16} \\ 0 & B_1 \end{bmatrix}, \quad C_{15} : n \times q, \quad \{C_{16}, B_1 : n \times n\}.$$

Define also

$$\begin{bmatrix} \bar{\mathbf{x}}_0^{(2)} \\ \bar{\mathbf{x}}_1 \end{bmatrix} \triangleq V_0^* \begin{bmatrix} \bar{\mathbf{x}}_0^{(1)} \\ \bar{\mathbf{u}}_0 \end{bmatrix}, \quad \bar{\mathbf{x}}_0^{(2)} : q \times 1, \quad \bar{\mathbf{x}}_1 : n \times 1. \tag{12.C.11}$$

Then we can write (12.C.10) in the equivalent form:

$$\begin{bmatrix} C_8 & C_9 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \end{bmatrix} = \begin{bmatrix} C_{15} & C_{16} \\ 0 & B_1 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_0^{(2)} \\ \bar{\mathbf{x}}_1 \end{bmatrix} + \begin{bmatrix} \mathbf{b}_3 \\ \mathbf{d}_1 \end{bmatrix}. \tag{12.C.12}$$

Now recall that $\{\mathbf{b}_3, \mathbf{d}_1\}$ are completely determined by the observation $\mathbf{y}_0$. Therefore, the l.l.m.s. estimator of $\text{col}\{\mathbf{b}_3, \mathbf{d}_1\}$ given $\mathbf{y}_0$ is the vector $\text{col}\{\mathbf{b}_3, \mathbf{d}_1\}$ itself. Likewise, $\bar{\mathbf{x}}_1$ is uncorrelated with $\mathbf{y}_0$ since it is a linear combination of two random variables that are themselves uncorrelated with $\mathbf{y}_0$. Hence, the l.l.m.s. estimator of $\bar{\mathbf{x}}_1$ given $\mathbf{y}_0$ is zero. It then follows from (12.C.12) that

$$A_1 \hat{\mathbf{x}}_1 = \mathbf{d}_1, \tag{12.C.13}$$

which completely determines $\{\hat{\mathbf{x}}_{0|0}, \hat{\mathbf{x}}_1\}$. In particular, $\hat{\mathbf{x}}_1 = A_1^{-1} \mathbf{d}_1$.

The matrices A_1 and B_1 derived above have a special significance in terms of the error covariance matrix P_1. Indeed, note from (12.C.12) and (12.C.13) that $A_1(\mathbf{x}_1 - \hat{\mathbf{x}}_1) = B_1\bar{\mathbf{x}}_1$. But since $\{\bar{\mathbf{x}}_0, \bar{\mathbf{u}}_0, \bar{\mathbf{v}}_0\}$ are uncorrelated random variables with unit covariance matrices, and since Z_0 and V_0 are unitary matrices, we conclude from (12.C.7) and (12.C.11) that $\|\bar{\mathbf{x}}_1\|^2 = I$. Consequently, $A_1 P_1 A_1^* = B_1 B_1^*$, which shows that we can take $P_1^{1/2} = A_1^{-1} B_1$. In other words, this procedure propagates a square-root factor of P_1 in factored form!

In order to further appreciate the significance of the above sequence of rotations, let us return to the full problem in (12.C.2) and assume we apply these rotations to the coefficient matrices in this equation. It is easy to see that it reduces to the form:

$$\begin{bmatrix} C_8 & C_9 & & & & \\ 0 & A_1 & & & & \\ & H_1 & & & & \\ & -F_1 & I & & & \\ & & \ddots & & & \\ & & & H_i & & \\ & & & -F_i & I & \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \\ \mathbf{x}_i \\ \mathbf{x}_{i+1} \end{bmatrix} = \tag{12.C.14}$$

$$\begin{bmatrix} C_{15} & C_{16} & & & & & \\ 0 & B_1 & & & & & \\ & & R_1^{1/2} & & & & \\ & & & G_1 Q_1^{1/2} & & & \\ & & & & \ddots & & \\ & & & & & R_i^{1/2} & \\ & & & & & & G_i Q_i^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{v}}_1 \\ \bar{\mathbf{u}}_1 \\ \vdots \\ \bar{\mathbf{v}}_i \\ \bar{\mathbf{u}}_i \end{bmatrix} + \begin{bmatrix} \mathbf{b}_3 \\ \mathbf{d}_1 \\ \mathbf{y}_1 \\ 0 \\ \vdots \\ \mathbf{y}_i \\ 0 \end{bmatrix}.$$

These equations show that we can now consider the subproblem

$$\begin{bmatrix} A_1 & 0 \\ H_1 & 0 \\ -F_1 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} B_1 & 0 & 0 \\ 0 & R_1^{1/2} & 0 \\ 0 & 0 & G_1 Q_1^{1/2} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{v}}_1 \\ \bar{\mathbf{u}}_1 \end{bmatrix} + \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{y}_1 \\ 0 \end{bmatrix}, \tag{12.C.15}$$

which has the same structure as the earlier problem (12.C.4). The new variables $\{A_1, B_1, \mathbf{d}_1\}$ carry the information from the previous step to the current step. We can now repeat the same triangularization procedure as before. This would lead us to (*cf.* (12.C.13))

$$\begin{bmatrix} C_8' & C_9' \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{1|1} \\ \hat{\mathbf{x}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_3' \\ \mathbf{d}_2 \end{bmatrix}, \tag{12.C.16}$$

for some new $\{C_8', C_9', \mathbf{b}_3'\}$ and new $\{A_2, B_2\}$, and the procedure can be continued. Each step provides us with $\hat{\mathbf{x}}_j$ and $\hat{\mathbf{x}}_{j|j}$. In summary, we obtain the following algorithm.

Algorithm 12.C.1 (Backward Stable Algorithm) *Consider the model (12.1.1), with $S_i = 0$, and assume the maximum singular value of G_i is not too large so that products of the form $G_i Q_i^{1/2}$ do not contribute significantly to numerical inaccuracy. The state estimators can be computed recursively in a numerically stable manner as follows. Start with $A_0 = I$, $B_0 = \Pi_0^{1/2}$, $\mathbf{d}_0 = 0$, and repeat for $i \geq 0$:*

1. *Choose a unitary matrix W_i that performs the upper triangularization:*

$$W_i \begin{bmatrix} A_i \\ H_i \end{bmatrix} = \begin{bmatrix} C_0 \\ 0 \end{bmatrix},$$

and compute

$$\begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \end{bmatrix} = W_i \begin{bmatrix} \mathbf{d}_i \\ \mathbf{y}_i \end{bmatrix}, \quad \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} = W_i \begin{bmatrix} B_i & 0 \\ 0 & R_i^{1/2} \end{bmatrix}.$$

2. *Choose a unitary matrix Z_i that performs the upper triangularization:*

$$\begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} Z_i = \begin{bmatrix} C_5 & C_6 \\ 0 & C_7 \end{bmatrix},$$

and determine the variable $\mathbf{v}$ via $C_7\mathbf{v} = \mathbf{b}_1$. Define also $\mathbf{b}_2 = \mathbf{b}_0 + C_6\mathbf{v}$.

3. *Choose a unitary matrix T_i that performs the upper triangularization*

$$T_i \begin{bmatrix} C_0 & 0 \\ -F_i & I \end{bmatrix} = \begin{bmatrix} C_8 & C_9 \\ 0 & A_{i+1} \end{bmatrix},$$

and let

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{13} & C_{14} \end{bmatrix} = T_i \begin{bmatrix} C_5 & 0 \\ 0 & G_0 Q_0^{1/2} \end{bmatrix},$$

and

$$\begin{bmatrix} \mathbf{b}_3 \\ \mathbf{d}_{i+1} \end{bmatrix} = T_i \begin{bmatrix} \mathbf{b}_2 \\ 0 \end{bmatrix}.$$

4. *Choose a unitary matrix V_i that performs the upper triangularization (B_{i+1} is $n \times n$)*

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{13} & C_{14} \end{bmatrix} V_i = \begin{bmatrix} C_{15} & C_{16} \\ 0 & B_{i+1} \end{bmatrix}.$$

5. *Then $\hat{\mathbf{x}}_{i+1} = A_{i+1}^{-1}\mathbf{d}_{i+1}$. Moreover, $P_{i+1}^{1/2} = A_{i+1}^{-1}B_{i+1}$.*

■

CHAPTER 1

Fast Array Algorithms

We now turn to array versions of the CKMS recursions of Ch. 11.Beyond the advantages of array algorithms that were already mentioned in Ch. 12, we shall see some further benefits here. For example, we can (see Sec. 13.2) obtain proofs that avoid the need for the special (generalized Stokes) identities for δP_i that were critical in Ch. 11. Secondly, it becomes even clearer that the critical assumption is the constancy of $\{F, H\}$, and that variations in $\{G_i, Q_i, R_i\}$ can be accommodated. Finally, in App. 13.A we shall show how the fast (generalized Schur algorithm) for factoring matrices with displacement structure results in the fast array algorithms of this chapter.

13.1 A SPECIAL CASE: $P_0 = 0$

It is simplest to begin our study with the special case of constant parameter models with $\Pi_0 = 0$. The reason is that in the general case we shall have to use not only unitary transformations (as done in Ch. 12), but also the generally less familiar J-unitary (or hyperbolic) transformations described in App. B.

Now when $\Pi_0 = 0$, the Riccati recursion

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0 = \Pi_0,$$

with

$$K_{p,i} = (FP_iH^* + GS)R_{e,i}^{-1}, \quad R_{e,i} = R + HP_iH^*,$$

shows that

$$P_1 = G(Q - SR^{-1}S^*)G^* = GQ^sG^* = \bar{L}_0\bar{L}_0^*,$$

where we introduced $Q^s = Q - SR^{-1}S^*$ and

$$\bar{L}_0 = GQ^{s/2}, \quad \text{an } n \times m \text{ matrix.}$$

For simplicity, and without any real loss of generality, we are assuming that the matrices $\{G, Q^s, R\}$ are full rank.

Moreover, when $\Pi_0 = 0$, the sequence $\{P_i\}$ is monotone nondecreasing (*see* Sec. 11.2.1) so that we can write

$$P_{i+1} - P_i = \bar{L}_i\bar{L}_i^*, \tag{13.1.1}$$

for some $n \times m$ matrices $\{\bar{L}_i\}$. It turns out that the $\{\bar{L}_i\}$ can be propagated by unitarily triangularizing the pre-array in the following equation:

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix}\Theta = \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}, \quad \text{say.} \tag{13.1.2}$$

The reason is that we can identify

$$X = R_{e,i+1}^{1/2}, \quad Y = \bar{K}_{p,i+1} = K_{p,i+1}R_{e,i+1}^{-*/2}, \quad Z = \bar{L}_{i+1}. \tag{13.1.3}$$

The claim (13.1.3) can be justified by verifying that

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix}\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix}^* = \begin{bmatrix} XX^* & XY^* \\ YX^* & YY^* + ZZ^* \end{bmatrix}, \tag{13.1.4}$$

with $\{X, Y, Z\}$ as in (13.1.3). The verification is straightforward, but as active readers should check, the fact that $\{F, G, H, Q, R\}$ are constant is important in the algebra.

Of course, the point of all this is that the algorithm in (13.1.2) exploits the constancy so as to allow us to work with smaller arrays, with only $(n + p) \times (n + m)$ elements, as compared to $(n + p) \times (n + p + m)$ elements in the general array algorithm of Ch. 12, see (12.6.13):

$$\begin{bmatrix} R_i^{1/2} & H_iP_i^{1/2} & 0 \\ G_iS_iR_i^{-*/2} & F_iP_i^{1/2} & G_iQ_i^{s/2} \end{bmatrix}\Theta = \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ K_iR_{e,i}^{-*/2} & P_{i+1}^{1/2} & 0 \end{bmatrix}, \tag{13.1.5}$$

which has the same form whether the model parameters $\{F, G, H, Q, R\}$ are constant or not. Of course, the simpler algorithm depended upon knowing the identity (13.1.1), which uses the fact that $\Pi_0 = 0$ and that the model parameters are constant.

13.1.1 Unitary Equivalence and an Alternative Derivation

More insight into the role of the parameter invariance can be gained by studying how the new algorithm (13.1.2)–(13.1.3) can be derived from the array algorithm (12.6.13), which holds for both time-variant and time-invariant models.

For this purpose, it will be convenient to introduce the notation

$$A \approx B \iff AA^* = BB^*, \tag{13.1.6}$$

and read it as saying that the $n \times m$ matrices A and B (with $n \leq m$) are *unitarily equivalent*. In view of Lemma A.5.1, this is equivalent to saying that there exists a unitary matrix Θ relating A to B, say $A = B\Theta$. It is also easy to check that

$$A \approx A, \quad A \approx B \Rightarrow B \approx A, \quad \text{and} \quad A \approx B, \; B \approx C \Rightarrow A \approx C.$$

Returning to the fast recursions, we start with the array algorithm (12.6.13) and observe that it can be written as (note that $\{F, G, H, Q, R\}$ are constant)

$$\begin{bmatrix} R^{1/2} & 0 & HP_i^{1/2} \\ GSR^{-*/2} & GQ^{s/2} & FP_i^{1/2} \end{bmatrix} \approx \begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 \end{bmatrix}. \tag{13.1.7}$$

That is, the pre- and post-arrays are unitarily equivalent matrices. Now by virtue of (13.1.1),

$$\begin{bmatrix} P_{i+1}^{1/2} & 0 \end{bmatrix} \quad \approx \quad \begin{bmatrix} P_i^{1/2} & \bar{L}_i \end{bmatrix}.$$

Using the constancy of F and H, note that we can write

$$\begin{bmatrix} R^{1/2} & 0 & HP_{i+1}^{1/2} & 0 \\ GSR^{-*/2} & GQ^{s/2} & FP_{i+1}^{1/2} & 0 \end{bmatrix} \approx \begin{bmatrix} R^{1/2} & 0 & HP_i^{1/2} & H\bar{L}_i \\ GSR^{-*/2} & GQ^{s/2} & FP_i^{1/2} & F\bar{L}_i \end{bmatrix}. \tag{13.1.8}$$

But since the first three block columns in the right-hand side array of (13.1.8) are the same as the pre-array in (13.1.7), we can write

$$\left[\begin{array}{ccc|c} R^{1/2} & 0 & HP_i^{1/2} & H\bar{L}_i \\ GSR^{-*/2} & GQ^{s/2} & FP_i^{1/2} & F\bar{L}_i \end{array}\right] \approx \left[\begin{array}{ccc|c} R_{e,i}^{1/2} & 0 & 0 & H\bar{L}_i \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 & F\bar{L}_i \end{array}\right]. \tag{13.1.9}$$

On the other hand, by algorithm (12.3.10) again,

$$\begin{bmatrix} R^{1/2} & 0 & HP_{i+1}^{1/2} \\ GSR^{-*/2} & GQ^{s/2} & FP_{i+1}^{1/2} \end{bmatrix} \approx \begin{bmatrix} R_{e,i+1}^{1/2} & 0 & 0 \\ \bar{K}_{p,i+1} & P_{i+2}^{1/2} & 0 \end{bmatrix},$$

and, again by virtue of (13.1.1),

$$\begin{bmatrix} R_{e,i+1}^{1/2} & 0 & 0 & 0 \\ \bar{K}_{p,i+1} & P_{i+2}^{1/2} & 0 & 0 \end{bmatrix} \approx \begin{bmatrix} R_{e,i+1}^{1/2} & 0 & 0 & 0 \\ \bar{K}_{p,i+1} & P_{i+1}^{1/2} & 0 & \bar{L}_{i+1} \end{bmatrix}. \tag{13.1.10}$$

Therefore, the right-hand side arrays in (13.1.9) and (13.1.10) must be unitarily equivalent, *i.e.*,

$$\begin{bmatrix} R_{e,i}^{1/2} & 0 & 0 & H\bar{L}_i \\ \bar{K}_{p,i} & P_{i+1}^{1/2} & 0 & F\bar{L}_i \end{bmatrix} \approx \begin{bmatrix} R_{e,i+1}^{1/2} & 0 & 0 & 0 \\ \bar{K}_{p,i+1} & P_{i+1}^{1/2} & 0 & \bar{L}_{i+1} \end{bmatrix}. \tag{13.1.11}$$

The second and third block columns are the same in both arrays and so they can be ignored, giving

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix} \approx \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} \end{bmatrix}, \tag{13.1.12}$$

which again justifies the algorithm described earlier in (13.1.2)–(13.1.3).

13.2 A GENERAL FAST ARRAY ALGORITHM

The derivation in Sec. 13.1 used the relation (13.1.1), which is a special case of the (generalized Stokes) identity (11.1.6). Here, following Sayed and Kailath (1992,1994a), we shall give a derivation of a general fast algorithm that needs no special identities. However, it does use the idea that we should begin by factoring the increments $\delta P_i = P_{i+1} - P_i$. So we now continue to assume that we are given the constant-parameter state-space model (11.1.1)–(11.1.2) but now Π_0 is not restricted to be zero. At any time instant i, we introduce a (nonunique) factorization of the form[1]

$$P_{i+1} - P_i = \bar{L}_i J_i \bar{L}_i^*, \tag{13.2.1}$$

where $\bar{L}_i$ is an $n \times \alpha_i$ matrix, J_i is an $\alpha_i \times \alpha_i$ signature matrix with as many $\pm 1's$ as $(P_{i+1} - P_i)$ has positive and negative eigenvalues, and $\alpha_i = \text{rank}(P_{i+1} - P_i)$. The time subscript i is used in both J_i and α_i for generality and in order not to assume any prior knowledge about their constancy or not. The fact that we can assume a constant $\{J, \alpha\}$ will instead follow as a consequence of the arguments given below.

The argument proceeds as follows. We form the pre-array

$$\mathcal{A} = \begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix}, \tag{13.2.2}$$

and triangularize it via an $(I \oplus J_i)-$unitary matrix Θ (procedures for achieving such triangularizations are discussed at length in App. B.4), *i.e.*,

$$A\Theta = \begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix} \Theta = \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}, \tag{13.2.3}$$

for some Θ such that

$$\Theta \begin{bmatrix} I & 0 \\ 0 & J_i \end{bmatrix} \Theta^* = \begin{bmatrix} I & 0 \\ 0 & J_i \end{bmatrix}.$$

The first question is whether such a Θ exists. The answer is affirmative and the construction is possible because we know that we can write

$$R_{e,i+1} = R + HP_{i+1}H^* = R_{e,i} + H\bar{L}_i J_i \bar{L}_i^* H^*,$$

or equivalently,

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & J_i \end{bmatrix} \begin{bmatrix} R_{e,i}^{*/2} \\ \bar{L}_i^* H^* \end{bmatrix} = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & J_i \end{bmatrix} \begin{bmatrix} R_{e,i+1}^{*/2} \\ 0 \end{bmatrix}.$$

If we now invoke the result of Lemma A.5.2, we conclude that there always exists an $(I \oplus J_i)-$unitary rotation Θ relating the following arrays:

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \end{bmatrix}. \tag{13.2.4}$$

[1] Compared with the earlier factorization (11.1.10), we see that we are now replacing M_i by a signature matrix and denoting the corresponding factor L_i by $\bar{L}_i$.

Hence, expression (13.2.3) can be regarded as rotating the first block row of the array to the form $\begin{bmatrix} X & 0 \end{bmatrix}$, which we know is possible because of (13.2.4), and then transforming the second block row of the pre-array in (13.2.3) by the *same* transformation, thus resulting in quantities that we denote by Y and Z.

Proceeding with (13.2.3), we can identify the $\{X, Y, Z\}$ terms as follows. Comparing the $(I \oplus J_i)$-"norms" on both sides of the equality (13.2.3) we have

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix} \underbrace{\Theta \begin{bmatrix} I & 0 \\ 0 & J_i \end{bmatrix} \Theta^*}_{I \oplus J_i} \begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix}^* = \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & J_i \end{bmatrix} \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}^*.$$

Therefore (here we use the constancy of H and R),

$$\begin{aligned} XX^* &= R_{e,i} + H\bar{L}_i J_i \bar{L}_i^* H^* &&= R_{e,i} + H[P_{i+1} - P_i]H^*, \\ &= R + HP_iH^* + HP_{i+1}H^* - HP_iH^* &&= R + HP_{i+1}H^* = R_{e,i+1}, \end{aligned}$$

which allows us to identify $X = R_{e,i+1}^{1/2}$, as predicted by (13.2.4). Moreover (we now use the constancy of $\{F, G, S\}$),

$$\begin{aligned} YX^* &= K_i + F\bar{L}_i J_i \bar{L}_i^* H^* &&= K_i + F[P_{i+1} - P_i]H^*, \\ &= [FP_iH^* + GS] + FP_{i+1}H^* - FP_iH^* &&= GS + FP_{i+1}H^* = K_{i+1}. \end{aligned}$$

We can therefore identify $Y = K_{i+1}R_{e,i+1}^{-*/2} = \bar{K}_{p,i+1}$. Finally (we now use the constancy of $\{G, Q\}$),

$$\begin{aligned} YY^* + ZJ_iZ^* &= K_iR_{e,i}^{-1}K_i^* + F\bar{L}_iJ_i\bar{L}_i^*F^*, \\ &= K_iR_{e,i}^{-1}K_i^* + FP_{i+1}F^* - FP_iF^*, \end{aligned}$$

and hence, $ZJ_iZ^* = P_{i+2} - P_{i+1}$. We can then identify $Z = \bar{L}_{i+1}$ and set $J_{i+1} = J_i$ since, by definition, $P_{i+2} - P_{i+1} = \bar{L}_{i+1}J_{i+1}\bar{L}_{i+1}^*$. So in fact, we can choose J_{i+1} to be equal to the first inertia matrix J, defined by the factorization

$$P_1 - P_0 = \left(F\Pi_0F^* + GQG^* - K_0R_{e,0}^{-1}K_0^*\right) - \Pi_0 = \bar{L}_0J\bar{L}_0^*. \tag{13.2.5}$$

Likewise, α_i can be chosen to be equal to the size α of J.

In summary, the above derivation shows that $\{\bar{K}_{p,i}, R_{e,i}^{1/2}\}$ can be recursively updated via the array algorithm

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} \end{bmatrix}, \tag{13.2.6}$$

where Θ is any $(I \oplus J)$-unitary matrix that produces the block zero entry in the post-array. Moreover, the initial conditions are $R_{e,0} = R + H\Pi_0H^*$ and $K_0 = F\Pi_0H^* + GS$, with $(\bar{L}_0, J)$ obtained via the factorization (13.2.5).

We should note that the $(I \oplus J)$-unitary transformation Θ can be implemented in several ways, especially by a sequence of elementary unitary and J-unitary Householder and/or Givens transformations, as described in App. B. As mentioned there, extra care has to be exercised in implementing the hyperbolic transformations.

Remark 1. A review of the derivation of the general array algorithm shows that the crucial fact for obtaining the fast algorithm was the constancy of $\{F, H\}$. It is not really essential that $\{R_i, Q_i, G_i\}$ be time-invariant. Time variations in these matrices can be handled just as for P_i, as noted by Morf and Kailath (1975). ◆

13.3 FROM EXPLICIT EQUATIONS TO ARRAY ALGORITHMS

The derivation in Sec. 13.2 is interesting because it circumvents the special (Stokes) identity for δP_i, the factored form of which led us to the explicit CKMS recursions of Ch. 11. Note also that the array form uses only 3 variables $\{R_{e,i}^{1/2}, \bar{K}_{p,i}, \bar{L}_i\}$ in place of the 4 variables in the explicit equations, $\{R_{e,i}, R_{r,i}, L_i, K_{p,i}\}$. It is interesting to explore the relations between these two algorithms. In fact, we can deduce the explicit CKMS recursions from the fast array recursions by using some characterizations of $(I \oplus J)$-unitary matrices (see Prob. 13.4). However, it will be easier to (at least first) go in the other direction, which we shall do in this section, following the arguments in Kailath, Vieira, and Morf (1978a).

To do so, we first note that the explicit CKMS equations of Thm. 11.1.2 can be organized as follows:

$$\begin{bmatrix} R_{e,i} & HL_i \\ K_i & FL_i \\ L_i^*H^* & R_{r,i} \end{bmatrix} \Sigma = \begin{bmatrix} R_{e,i+1} & 0 \\ K_{i+1} & L_{i+1} \\ 0 & R_{r,i+1} \end{bmatrix}, \tag{13.3.1}$$

where

$$\Sigma = \begin{bmatrix} I & -R_{e,i}^{-1}HL_i \\ -R_{r,i}^{-1}L_i^*H^* & I \end{bmatrix}. \tag{13.3.2}$$

So we can also regard the explicit fast equations as a transformation of a certain pre-array to a certain post-array. However, while we have here an explicit formula (13.3.2) for the transformation matrix Σ, it is conceivable that Σ has enough structure that it need only be implicitly described. For example, if Σ were unitary, then the transformation in (13.3.1) could be achieved without explicitly computing Σ by using, for example, Householder or Givens transformations (as in App. B). Here, it turns out that though Σ is not unitary, by appropriate normalization it can be made unitary with respect to a certain indefinite metric. Once this is achieved, the array version (13.2.6) of the algorithm will become almost self-evident.

To achieve the normalization, we begin by noting the easily checked identity

$$\begin{bmatrix} R_{e,i+1} & 0 \\ 0 & -R_{r,i+1} \end{bmatrix} = \Sigma^* \begin{bmatrix} R_{e,i} & 0 \\ 0 & -R_{r,i} \end{bmatrix} \Sigma. \tag{13.3.3}$$

Now we appeal to the fact (Lemma 11.1.2) that the $\{R_{r,i}\}$ have constant inertia to say that we can write $-R_{r,i}$ in the form

$$-R_{r,i} = R_{r,i}^{1/2} J R_{r,i}^{*/2}, \tag{13.3.4}$$

for some "generalized" square-root factor $R_{r,i}^{1/2}$, and where the $\alpha \times \alpha$ signature matrix J has the form $J = (I_{\alpha_-} \oplus -I_{\alpha_+})$ and $\alpha_\pm$ denotes the number of positive (negative) eigenvalues of $P_1 - P_0$,

$$P_1 - P_0 = F\Pi_0 F^* + GQG^* - K_0 R_{e,0}^{-1} K_0^* - \Pi_0.$$

A factorization of the form (13.3.4) always exists and is studied further in Prob. 13.3.

Now if we define $\bar{J} = (I_p \oplus J) = (I_{p+\alpha_+} \oplus -I_{\alpha_-})$, then we can write

$$\begin{bmatrix} R_{e,i} & 0 \\ 0 & -R_{r,i} \end{bmatrix} = \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ 0 & R_{r,i}^{1/2} \end{bmatrix} \bar{J} \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ 0 & R_{r,i}^{1/2} \end{bmatrix}^*.$$

But then if we define

$$\bar{\Theta} \triangleq \begin{bmatrix} R_{e,i}^{1/2} & 0 \\ 0 & R_{r,i}^{1/2} \end{bmatrix}^* \Sigma \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \\ 0 & R_{r,i+1}^{1/2} \end{bmatrix}^{-*}, \tag{13.3.5}$$

we can see immediately from (13.3.3) that

$$\bar{\Theta}^* \bar{J} \bar{\Theta} = \bar{J}, \quad \bar{J} = (I \oplus J). \tag{13.3.6}$$

The $\{\bar{\Theta}\}$ are nonsingular (their determinants are nonzero) and therefore from (13.3.6) we can write $\bar{\Theta} = \bar{J}\bar{\Theta}^{-*}\bar{J}$, which shows that we also have $\bar{\Theta}\bar{J}\bar{\Theta}^* = \bar{J}$. Therefore, $\bar{\Theta}$ is a $\bar{J}$–unitary matrix. We can now rewrite the array equation (13.3.1) by using (13.3.5) as

$$\begin{bmatrix} R_{e,i} & HL_i \\ K_i & FL_i \\ L_i^* H^* & R_{r,i} \end{bmatrix} \underbrace{\begin{bmatrix} R_{e,i}^{-*/2} & 0 \\ 0 & R_{r,i}^{-*/2} \end{bmatrix} \bar{\Theta} \begin{bmatrix} R_{e,i+1}^{*/2} & 0 \\ 0 & R_{r,i+1}^{*/2} \end{bmatrix}}_{\Sigma} = \begin{bmatrix} R_{e,i+1} & 0 \\ K_{i+1} & L_{i+1} \\ 0 & R_{r,i+1} \end{bmatrix}$$

or, equivalently,

$$\left[\begin{array}{cc} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \\ \hline R_{r,i}^{1/2}\bar{L}_i^* H^* R_{e,i}^{-*/2} & R_{r,i}^{1/2} \end{array}\right] \bar{\Theta} = \left[\begin{array}{cc} R_{e,i+1}^{1/2} & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} \\ \hline 0 & R_{r,i+1}^{1/2} \end{array}\right], \tag{13.3.7}$$

where

$$\bar{K}_{p,i} = K_i R_{e,i}^{-*/2}, \quad \bar{L}_i = L_i R_{r,i}^{-*/2}.$$

We have partitioned the arrays in (13.3.7) because it can be seen that we only need the first two block rows in it to compute $K_{p,i}$ and $R_{e,i}$, knowledge of which completely determines the innovations representation of $\{y_i\}$. So now we again have the fast array algorithm (13.2.6)!

Moreover, though we have an explicit formula (13.3.5) for $\bar{\Theta}$, *viz.* (see also Prob. 13.4),

$$\bar{\Theta} = \begin{bmatrix} R_{e,i}^{*/2} R_{e,i+1}^{-*/2} & -R_{e,i}^{-1/2} H L_i R_{r,i+1}^{-*/2} \\ -R_{r,i}^{-1/2} L_i^* H^* R_{e,i+1}^{-*/2} & R_{r,i}^{*/2} R_{r,i+1}^{-*/2} \end{bmatrix}, \tag{13.3.8}$$

we could in fact use any $\bar{J}$-unitary matrix Θ such that $\Theta\bar{J}\Theta^* = \bar{J}$ and

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix} \Theta \text{ has the form } \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}.$$

It is easy to see that by a now standard procedure, we can make the identifications $X = R_{e,i+1}^{1/2}$, $Y = \bar{K}_{p,i+1}$, and $Z = \bar{L}_{i+1}$. In other words, the $\bar{\Theta}$ that we obtained in (13.3.5) is only one possible choice for the rotation matrix Θ that is needed in (13.2.6).

13.4 STRUCTURED TIME-VARIANT SYSTEMS

We can also of course obtain array forms of the extended CKMS recursions derived in Sec. 11.3 to cover certain structured forms of time-variant models, in particular those having the form (11.3.3)

$$H_i = H_{i+1}\Psi_i, \qquad F_{i+1}\Psi_i = \Psi_{i+1}F_i, \qquad G_{i+1} = \Psi_{i+1}G_i,$$

for some $n \times n$ matrices Ψ_i. [We continue to assume that the covariance matrices R_i, S_i, and Q_i are constant for all i ($R_i = R$, $Q_i = Q$, $S_i = S$).]

As mentioned in Sec. 11.3, the savings in computation are now achieved by considering the generalized difference matrix $\delta_\Psi P_i = P_{i+1} - \Psi_i P_i \Psi_i^*$. Thus assume that we factor (nonuniquely) $\delta_\Psi P_i$ as

$$P_{i+1} - \Psi_i P_i \Psi_i^* = \bar{L}_i J_i \bar{L}_i^*,$$

where $\bar{L}_i$ is $n \times \alpha_i$, and J_i is an $\alpha_i \times \alpha_i$ signature matrix. The time subscript i is again used in both J_i and α_i for generality and in order not to assume any prior knowledge about their constancy or not. This fact will instead follow as a consequence of the arguments given below, which will allow us to drop the time subscript from J_i and α_i and to replace them by a constant signature matrix J and a constant scalar α, respectively.

The array algorithm can be derived in much the same way as we did in Sec. 13.2 (*cf.* Sayed and Kailath (1992,1994a)). We form the pre-array of numbers

$$\begin{bmatrix} R_{e,i}^{1/2} & H_{i+1}\bar{L}_i \\ \Psi_{i+1}\bar{K}_{p,i} & F_{i+1}\bar{L}_i \end{bmatrix}, \tag{13.4.1}$$

and choose any $(I \oplus J_i)$-unitary matrix Θ that triangularizes it, say

$$\begin{bmatrix} R_{e,i}^{1/2} & H_{i+1}\bar{L}_i \\ \Psi_{i+1}\bar{K}_{p,i} & F_{i+1}\bar{L}_i \end{bmatrix} \Theta = \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}.$$

In contrast to the earlier pre-array (13.2.2), the pre-array formed in (13.4.1) now contains time-variant matrices H_{i+1} and F_{i+1}, as well as the matrix Ψ_{i+1}.

By comparing the $(I \oplus J_i)$-"norms" on both sides of the following equality,

$$\begin{bmatrix} R_{e,i}^{1/2} & H_{i+1}\bar{L}_i \\ \Psi_{i+1}\bar{K}_{p,i} & F_{i+1}\bar{L}_i \end{bmatrix} \underbrace{\Theta \begin{bmatrix} I & 0 \\ 0 & J_i \end{bmatrix} \Theta^*}_{I \oplus J_i} \begin{bmatrix} R_{e,i}^{1/2} & H_{i+1}\bar{L}_i \\ \Psi_{i+1}\bar{K}_{p,i} & F_{i+1}\bar{L}_i \end{bmatrix}^* = \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & J_i \end{bmatrix} \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}^*,$$

we get (here we use the condition $H_{i+1}\Psi_i = H_i$ and the fact that R_i is a constant equal to R)

$$\begin{aligned} XX^* &= R_{e,i} + H_{i+1}\bar{L}_i J_i \bar{L}_i^* H_{i+1}^* = R_{e,i} + H_{i+1}[P_{i+1} - \Psi_i P_i \Psi_i^*]H_{i+1}^*, \\ &= R + H_i P_i H_i^* + H_{i+1}P_{i+1}H_{i+1}^* - H_i P_i H_i^* = R + H_{i+1}P_{i+1}H_{i+1}^* = R_{e,i+1}. \end{aligned}$$

We can therefore identify $X = R_{e,i+1}^{1/2}$. Moreover (we now use the conditions on F_i and G_i and the fact that S_i is constant),

$$\begin{aligned} YX^* &= \Psi_{i+1}K_i + F_{i+1}\bar{L}_i J_i \bar{L}_i^* H_{i+1}^*, \\ &= \Psi_{i+1}K_i + F_{i+1}[P_{i+1} - \Psi_i P_i \Psi_i^*]H_{i+1}^*, \\ &= \Psi_{i+1}[F_i P_i H_i^* + G_i S] + F_{i+1}P_{i+1}H_{i+1}^* - \Psi_{i+1}F_i P_i H_i^*, \\ &= \Psi_{i+1}G_i S + F_{i+1}P_{i+1}H_{i+1}^*, \\ &= G_{i+1}S + F_{i+1}P_{i+1}H_{i+1}^* = K_{i+1}. \end{aligned}$$

We can then identify $Y = K_{i+1}R_{e,i+1}^{-*/2} = \bar{K}_{p,i+1}$. Finally (we now use the condition on G_i and the fact that Q_i is constant),

$$\begin{aligned} YY^* + ZJ_iZ^* &= \Psi_{i+1}K_i R_{e,i}^{-1} K_i^* \Psi_{i+1}^* + F_{i+1}\bar{L}_i J_i \bar{L}_i^* F_{i+1}^*, \\ &= \Psi_{i+1}K_i R_{e,i}^{-1} K_i^* \Psi_{i+1}^* + F_{i+1}P_{i+1}F_{i+1}^* - \Psi_{i+1}F_i P_i F_i^* \Psi_{i+1}^*. \end{aligned}$$

Therefore,

$$ZJ_iZ^* = P_{i+2} - \Psi_{i+1}P_{i+1}\Psi_{i+1}^*,$$

which allows us to choose $Z = \bar{L}_{i+1}$ and to set $J_{i+1} = J_i$ since, by definition, $P_{i+2} - \Psi_{i+1}P_{i+1}\Psi_{i+1}^* = \bar{L}_{i+1}J_{i+1}\bar{L}_{i+1}^*$.

This argument therefore shows that we can choose J_i to be equal to the inertia matrix J obtained via the factorization

$$P_1 - \Psi_0 P_0 \Psi_0^* = \left(F_0 \Pi_0 F_0^* + G_0 Q G_0^* - K_0 R_{e,0}^{-1} K_0^*\right) - \Psi_0 \Pi_0 \Psi_0^* = \bar{L}_0 J \bar{L}_0^*.$$

Likewise, α_i can be taken to be equal to the size α of J.

In summary, the resulting extended fast (CKMS) array takes the form

$$\begin{bmatrix} R_{e,i}^{1/2} & H_{i+1}\bar{L}_i \\ \Psi_{i+1}\bar{K}_{p,i} & F_{i+1}\bar{L}_i \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} \end{bmatrix}, \tag{13.4.2}$$

where Θ is any $(I \oplus J)$-unitary matrix that lower triangularizes the pre-array, and $(\bar{L}_0, J)$ are found from the factorization of $P_1 - \Psi_0 P_0 \Psi_0^*$.

13.5 COMPLEMENTS

The array form (13.2.6) of the fast filtering algorithm in Sec. 13.2 was originally derived by Morf and Kailath (1975); see also Silverman (1976) and Kailath, Vieira, and Morf (1978a). Their derivation relied on the explicit expressions of Thm. 11.1.2 and on the constant inertia property of the matrices $\{R_{r,i}\}$, as we explained in Sec. 13.3. The derivation in Sec. 13.2 is due to Sayed and Kailath (1992,1994a); it is self-contained and new in two respects. First, it does not assume prior knowledge of the recursions of Thm. 11.1.2 and therefore does not rely on the explicit formula (11.1.6) for δP_i. Second, the derivation does not assume that the successive differences $P_{i+1} - P_i$ have low rank, or that the inertia of the $\{R_{r,i}\}$ is constant for all i. These properties follow as byproducts of the array-based argument itself. Applications of the extended fast array algorithm (13.4.2) to adaptive RLS filtering can be found in Sayed and Kailath (1994b) and Merched and Sayed (1999).

In Sec. 11.5 we noted that the somewhat mysterious parameter α had a nice interpretation in terms of the displacement rank of the covariance matrix R_y. In App. 13.A we shall show how the displacement structure representation (11.5.6) can be used to derive an array form of the CKMS recursions by means of a generalized Schur algorithm presented in App. F.

PROBLEMS

13.1 **(An identity for Σ)** Establish the identity (13.3.3).

13.2 **(A modified fast algorithm)** Introduce the factorization

$$\begin{bmatrix} H \\ F \end{bmatrix} (P_1 - P_0) \begin{bmatrix} H \\ F \end{bmatrix}^* \triangleq \begin{bmatrix} \bar{L}_{1,0} \\ \bar{L}_{2,0} \end{bmatrix} J \begin{bmatrix} \bar{L}_{1,0} \\ \bar{L}_{2,0} \end{bmatrix}^*,$$

where J is a signature matrix. Refer to the modified fast algorithm of Prob. 11.3 and argue that the corresponding array algorithm has the following form.

Initialization. Compute a matrix $\bar{L}_1$ from the triangularization

$$\begin{bmatrix} R_{e,0}^{1/2} & \bar{L}_{1,0} \\ \bar{K}_{p,0} & \bar{L}_{2,0} \end{bmatrix} \Theta = \begin{bmatrix} A_1 & 0 \\ B_1 & \bar{L}_1 \end{bmatrix},$$

where Θ is any $(I \oplus J)-$unitary matrix that annihilates the (1, 2) entry of the pre-array. Verify that

$$P_2 - P_1 = \bar{L}_1 J \bar{L}_1^*.$$

Verify also that we can take $A_1 = R_{e,1}^{1/2}$ and $B_1 = \bar{K}_{p,1}$.

Recursion. Repeat for $i \geq 1$:

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & F\bar{L}_i \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} \end{bmatrix},$$

where Θ is again any $(I \oplus J)-$unitary matrix that annihilates the $(1,2)$ entry of the pre-array. Verify further that $P_{i+2} - P_{i+1} = \bar{L}_{i+1} J \bar{L}_{i+1}^*$.

13.3 (Generalized square roots) Let A be a full rank $n \times n$ Hermitian matrix with inertia $J = (I_p \oplus -I_q)$. That is, p denotes the number of positive eigenvalues of A and q denotes the number of negative eigenvalues of A (with $p + q = n$).

(a) Introduce the eigendecomposition $A = U\Lambda U^*$. Verify that A admits a generalized Hermitian square-root factor, denoted by $A^{1/2}$, such that

$$A = A^{1/2} J A^{*/2}, \quad A^{1/2} = A^{*/2}.$$

(b) Verfiy that for any $J-$unitary matrix Θ, $X = A^{1/2}\Theta$ is also a generalized (but not necessarily Hermitian) square root of A, *viz.*, $A = XJX^*$.

(c) Is it possible to always guarantee the existence of a lower triangular generalized square-root factor X with unit diagonal entries for any full rank matrix A?

13.4 (An explicit formula for $\bar{\Theta}$) We showed in the text that the matrix (13.3.8), *viz.*,

$$\bar{\Theta} = \begin{bmatrix} R_{e,i}^{*/2} R_{e,i+1}^{-*/2} & -R_{e,i}^{-1/2} H L_i R_{r,i+1}^{-*/2} \\ -R_{r,i}^{-1/2} L_i^* H^* R_{e,i+1}^{-*/2} & R_{r,i}^{*/2} R_{r,i+1}^{-*/2} \end{bmatrix},$$

is $\bar{J} = (I \oplus J)$-unitary, *i.e.*, $\bar{\Theta}\bar{J}\bar{\Theta}^* = \bar{J} = \bar{\Theta}^*\bar{J}\bar{\Theta}$.

(a) Define $\rho_i \triangleq R_{e,i}^{-1/2} H L_i R_{r,i}^{-*/2}$. Show that

$$J + \rho_i^* \rho_i = R_{r,i}^{-1/2} R_{r,i+1}^{1/2} J R_{r,i+1}^{*/2} R_{r,i}^{-*/2},$$

where

$$-R_{r,i} = R_{r,i}^{1/2} J R_{r,i}^{*/2}, \quad -R_{r,i+1} = R_{r,i+1}^{1/2} J R_{r,i+1}^{*/2}.$$

Conclude that $(J + \rho_i^* \rho_i)^{-*/2}$ is a generalized square-root factor of $(J + \rho_i^* \rho_i)^{-1}$.

(b) Show that $\bar{\Theta}$ can be rewritten as form

$$\bar{\Theta} = \begin{bmatrix} I & -\rho_i \\ -\rho_i^* & I \end{bmatrix} \begin{bmatrix} (I + \rho_i J \rho_i^*)^{-*/2} & 0 \\ 0 & (J + \rho_i^* \rho_i)^{-*/2} \end{bmatrix},$$

where ρ_i is also called the *reflection coefficient*.

(c) Show that in the stationary case (*cf.* Sec. 11.2.2) the above expression reduces to the form

$$\bar{\Theta} = \begin{bmatrix} I & -\rho_i \\ -\rho_i^* & I \end{bmatrix} \begin{bmatrix} (I - \rho_i \rho_i^*)^{-*/2} & 0 \\ 0 & -(I - \rho_i^* \rho_i)^{-*/2} \end{bmatrix},$$

and is now $(I_p \oplus -I_p)$-unitary.

Remark. Note that when ρ_i is a scalar, this is exactly the elementary hyperbolic rotation discussed in App. B. ♦

13.5 (Normalized extended recursions) Refer to the extended fast recursions of Thm. 11.3.1.

(a) Verify that they can be expressed in the form

$$\begin{bmatrix} R_{e,i} & H_{i+1} L_i \\ \Psi_{i+1} K_i & F_{i+1} L_i \\ L_i^* H_{i+1}^* & R_{r,i} \end{bmatrix} \Sigma = \begin{bmatrix} R_{e,i+1} & 0 \\ K_{i+1} & L_{i+1} \\ 0 & R_{r,i+1} \end{bmatrix},$$

where Σ is given by

$$\Sigma = \begin{bmatrix} I & -R_{e,i}^{-1} H_{i+1} L_i \\ -R_{r,i}^{-1} L_i^* H_{i+1}^* & I \end{bmatrix},$$

and satisfies the relation

$$\Sigma^* \begin{bmatrix} R_{e,i} & 0 \\ 0 & -R_{r,i} \end{bmatrix} \Sigma = \begin{bmatrix} R_{e,i+1} & 0 \\ 0 & -R_{r,i+1} \end{bmatrix}.$$

(b) By following the normalization procedure used in Sec. 13.3, show how to go from the above equations to the array form (13.4.2).

13.6 (Fast recursions in information form) Refer to the fast information recursions of Prob. 11.5. Follow the arguments of Sec. 13.2 to verify that the array form for these recursions is given by

$$\begin{bmatrix} R_{e,i}^{d/2} & G^* F^{-s*} \bar{L}_i^d \\ \bar{K}_{p,i}^d & F^{-s*} \bar{L}_i^d \end{bmatrix} \Theta = \begin{bmatrix} R_{e,i+1}^{d/2} & 0 \\ \bar{K}_{p,i+1}^d & \bar{L}_{i+1}^d \end{bmatrix},$$

where Θ is $(I \oplus J)$-unitary and J is a signature matrix that is obtained from the factorization:

$$\left(F^{-s*} P_{0|0}^{-1} F^{s*} + H^* R^{-1} H - K_{p,0}^d R_{e,0}^d K_{p,0}^{d*} \right) - P_{0|0}^{-1} = -\bar{L}_0^d J_0 \bar{L}_0^{d*}.$$

13.7 (Another structured model) Consider the state-space model (11.3.1) and assume $G_i = I$ and that there exist matrices $\{\Lambda_i, \Omega_i, \Psi_i\}$ such that the following time-variations hold:

$$\begin{bmatrix} \Lambda_{i+1} & 0 \\ \Omega_{i+1} & \Psi_{i+1} \end{bmatrix} \begin{bmatrix} H_i \\ F_i \end{bmatrix} = \begin{bmatrix} H_{i+1} \\ F_{i+1} \end{bmatrix} \Psi_i,$$

$$\begin{bmatrix} R_{i+1} & S_{i+1} \\ S_{i+1}^* & Q_{i+1} \end{bmatrix} - \begin{bmatrix} \Lambda_{i+1} & 0 \\ \Omega_{i+1} & \Psi_{i+1} \end{bmatrix} \begin{bmatrix} R_i & S_i \\ S_i^* & Q_i \end{bmatrix} \begin{bmatrix} \Lambda_{i+1} & 0 \\ \Omega_{i+1} & \Psi_{i+1} \end{bmatrix}^* = 0.$$

Show that a fast array algorithm for such models is the following:

$$\left[\begin{bmatrix} \Lambda_{i+1} & 0 \\ \Omega_{i+1} & \Psi_{i+1} \end{bmatrix} \begin{bmatrix} R_{e,i}^{1/2} \\ \bar{K}_{p,i} \end{bmatrix} \begin{bmatrix} H_{i+1} \bar{L}_i \\ F_{i+1} \bar{L}_i \end{bmatrix} \right] \Theta = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} \end{bmatrix},$$

where Θ is any $(I \oplus J)$-unitary matrix that produces the block zero entry in the post-array and $\{\bar{L}_0, J\}$ are obtained from the factorization $P_1 - \Psi_0 P_0 \Psi_0^* = \bar{L}_0 J \bar{L}_0^*$. Show further that $P_{i+1} - \Psi_i P_i \Psi_i^* = \bar{L}_i J \bar{L}_i^*$.

13.8 **(Time-variant noise covariances)** Consider again the state-space model (11.3.1) and assume $G_i = I$ and that there exist matrices $\{\Lambda_i, \Omega_i, \Psi_i\}$ such that the following time-variations hold:

$$\begin{bmatrix} \Lambda_{i+1} & 0 \\ \Omega_{i+1} & \Psi_{i+1} \end{bmatrix} \begin{bmatrix} H_i \\ F_i \end{bmatrix} = \begin{bmatrix} H_{i+1} \\ F_{i+1} \end{bmatrix} \Psi_i,$$

and

$$\begin{bmatrix} R_{i+1} & S_{i+1} \\ S_{i+1}^* & Q_{i+1} \end{bmatrix} - \begin{bmatrix} \Lambda_{i+1} & 0 \\ \Omega_{i+1} & \Psi_{i+1} \end{bmatrix} \begin{bmatrix} R_i & S_i \\ S_i^* & Q_i \end{bmatrix} \begin{bmatrix} \Lambda_{i+1} & 0 \\ \Omega_{i+1} & \Psi_{i+1} \end{bmatrix}^* \triangleq \begin{bmatrix} \bar{L}_i^{(1)} & \bar{L}_i^{(2)} \\ \bar{L}_i^{(3)} & \bar{L}_i^{(4)} \end{bmatrix} \begin{bmatrix} J_i^{(1)} & 0 \\ 0 & J_i^{(2)} \end{bmatrix} \begin{bmatrix} \bar{L}_i^{(1)} & \bar{L}_i^{(2)} \\ \bar{L}_i^{(3)} & \bar{L}_i^{(4)} \end{bmatrix}^*.$$

Show that a fast array algorithm for such models is given by

$$\left[\begin{bmatrix} \Lambda_{i+1} & 0 \\ \Omega_{i+1} & \Psi_{i+1} \end{bmatrix} \begin{bmatrix} R_{e,i}^{1/2} \\ \bar{K}_{p,i} \end{bmatrix} \begin{bmatrix} H_{i+1}\bar{L}_i \\ F_{i+1}\bar{L}_i \end{bmatrix} \begin{bmatrix} \bar{L}_i^{(1)} & \bar{L}_i^{(2)} \\ \bar{L}_i^{(3)} & \bar{L}_i^{(4)} \end{bmatrix} \right] \Theta = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} & 0 \end{bmatrix},$$

where Θ is any $(I \oplus J_i \oplus J_i^{(1)} \oplus J_i^{(2)})$-unitary matrix that produces the block zero entries in the post-array and $P_{i+1} - \Psi_i P_i \Psi_i^* = \bar{L}_i J_i \bar{L}_i^*$.

13.9 **(Extended fast recursions)** Consider the structured state-space model (11.3.1)–(11.3.3) with constant $\{R, S, Q\}$ and let R_y denote the covariance matrix of the output process $\{\mathbf{y}_i\}$. In Prob. 11.8 we showed that R_y satisfies a displacement equation of the form

$$\nabla_{Z^p} R_y = \mathcal{G} \begin{bmatrix} I_p & 0 \\ 0 & J \end{bmatrix} \mathcal{G}^*,$$

for some $\mathcal{G}$. Follow the derivation of App. 13.A to show that by applying the generalized Schur algorithm to $\mathcal{G}$, one is led again to the array form (13.4.2) of the extended fast recursions.

Appendix for Chapter 13

13.A COMBINING DISPLACEMENT AND STATE-SPACE STRUCTURES

We showed earlier in App. 9.A that the Kalman filter can be obtained by applying the modified Gram-Schmidt procedure to the Gramian matrix of the observations. We also mentioned in Sec. 11.5 that for matrices with displacement structure (see App. F), there are fast generalized Schur algorithms for factoring such matrices. Here we shall show that by further incorporating state-space structure, a form of the generalized Schur algorithms reduces to the fast CKMS array algorithm.

It will be helpful to review Sec. 11.5 and App. F at this time. In the former we showed that for a process $\{\mathbf{y}_i\}$ arising from a time-invariant state-space model, the displacement $\nabla_{Z^p} R_y$ had the form

$$\nabla_{Z^p} R_y = R_y - Z^p R_y [Z^p]^* = \mathcal{G} \mathcal{J} \mathcal{G}^*, \tag{13.A.1}$$

where

$$\mathcal{J} = \begin{bmatrix} I_p & 0 \\ 0 & J \end{bmatrix} \quad \text{and} \quad \mathcal{G} = \begin{bmatrix} R_{e,0}^{1/2} & 0 \\ H\bar{K}_{p,0} & H\bar{L}_0 \\ HF\bar{K}_{p,0} & HF\bar{L}_0 \\ \vdots & \vdots \end{bmatrix}, \tag{13.A.2}$$

with $\{\bar{K}_{p,0}, \bar{L}_0, J\}$ defined via

$$\bar{K}_{p,0} = K_0 R_{e,0}^{-*/2}, \qquad P_1 - P_0 = \bar{L}_0 J \bar{L}_0^*.$$

Here, J is an $\alpha \times \alpha$ signature matrix and $\bar{L}_0$ is $n \times \alpha$.

For convenience, we repeat here the description of the generalized Schur algorithm from App. F. It is a recursive procedure that starts with the matrix $\mathcal{G}_0 = \mathcal{G}$ and yields successive matrices $\{\mathcal{G}_i, \overline{\mathcal{G}}_i\}$, all with the same number of columns $p + \alpha$. The leading p-columns of each $\overline{\mathcal{G}}_i$ can be used to construct the Cholesky factor of R_y.

We start with $\mathcal{G}_0 = \mathcal{G}$ and repeat for $i \geq 0$:

1. Let g_i denote the top p rows of $\mathcal{G}_i$.
2. Determine a $\mathcal{J}$-unitary matrix Θ_i that reduces g_i to the form $g_i \Theta_i = [X \;\; 0]$, where X is a $p \times p$ lower triangular matrix. That is, a $p \times \alpha$ zero block is introduced in $g_i \Theta_i$. This step can be performed by using any of the elementary transformations described in App. B.
3. Apply the transformation Θ_i to all other rows of $\mathcal{G}_i$. Let $\overline{\mathcal{G}}_i$ denote the resulting intermediate matrix, *i.e.*, $\overline{\mathcal{G}}_i = \mathcal{G}_i \Theta_i$.

ft down the first p columns of $\overline{\mathcal{G}}_i$ by p steps and keep the last α columns ıltered. This results in a new matrix whose top p rows are zero and whose tom rows we denote by $\mathcal{G}_{i+1}$. In matrix language, we can express this procedure ollows:

$$\begin{bmatrix} 0_{p\times(p+\alpha)} \\ \mathcal{G}_{i+1} \end{bmatrix} = \mathcal{Z}^p \mathcal{G}_i \Theta_i \begin{bmatrix} I_p & 0 \\ 0 & 0_\alpha \end{bmatrix} + \mathcal{G}_i \Theta_i \begin{bmatrix} 0_p & 0 \\ 0 & I_\alpha \end{bmatrix}, \quad \mathcal{G}_0 = \mathcal{G}. \tag{13.A.3}$$

e i-th block column of the Cholesky factor of R_y, *viz.*, the block lower triangular trix in the factorization $R_y = \bar{L}\bar{L}^*$, is given by

$$\bar{l}_i = \overline{\mathcal{G}}_i \begin{bmatrix} I_p \\ 0 \end{bmatrix}.$$

ow apply the above procedure to the matrix $\mathcal{G}$ in (13.A.2). The first step involves ing $\mathcal{G}_0$ by a $\mathcal{J}$-unitary matrix Θ_0, which in this case will be the identity matrix first block row of $\mathcal{G}_0$ already has a $p \times \alpha$ zero block. Therefore, $\overline{\mathcal{G}}_0 = \mathcal{G}_0$. Now down the first block column we get

$$\mathcal{G}_1 = \begin{bmatrix} R_{e,0}^{1/2} & H\bar{L}_0 \\ H\bar{K}_{p,0} & HF\bar{L}_0 \\ HF\bar{K}_{p,0} & HF^2\bar{L}_0 \\ \vdots & \vdots \end{bmatrix}.$$

t (block) row of $\mathcal{G}_1$ is $g_1 = \begin{bmatrix} R_{e,0}^{1/2} & H\bar{L}_0 \end{bmatrix}$. The second step of the algorithm uires the determination of a $\mathcal{J}$-unitary matrix Θ_1 that reduces g_1 to the form for some lower triangular X, and the multiplication of all other rows of $\mathcal{G}_1$ by e Θ_1. In order to determine the resulting $\overline{\mathcal{G}}_1$ it will be enough to examine the f applying Θ_1 to the first two (block) rows of $\mathcal{G}_1$ only (and which are denoted ay

$$\mathcal{A}\Theta_1 = \begin{bmatrix} R_{e,0}^{1/2} & H\bar{L}_0 \\ H\bar{K}_{p,0} & HF\bar{L}_0 \end{bmatrix} \Theta_1 \triangleq \begin{bmatrix} X & 0 \\ Y & Z \end{bmatrix}, \quad \text{say.}$$

rmine the unknowns $\{X, Y, Z\}$ in terms of known quantities, we compare entries sides of the equality $A\mathcal{J}A^* = A\Theta_1\mathcal{J}\Theta_1^*A^*$. This leads to the equality

$$XX^* = R_{e,0} + H\bar{L}_0 J \bar{L}_0^* H^* = R_{e,1}.$$

an choose $X = R_{e,1}^{1/2}$. Moreover,

and, hence, we can identify $Y = K_1 R_{e,i}^{-*/2} = \bar{K}_{p,1}$. Finally,

$$YY^* + ZJZ^* = K_0 R_{e,0}^{-1} K_0^* + F\bar{L}_0 J \bar{L}_0^* F^* = P_2 - P_1 = \bar{L}_1 J \bar{L}_1^*,$$

which shows that we can identify Z as $\bar{L}_1$. We thus conclude that the effect of Θ_1 is

$$\begin{bmatrix} R_{e,0}^{1/2} & H\bar{L}_0 \\ \bar{K}_{p,0} & F\bar{L}_0 \end{bmatrix} \Theta_1 = \begin{bmatrix} R_{e,1}^{1/2} & 0 \\ \bar{K}_{p,1} & \bar{L}_1 \end{bmatrix}.$$

Now note that the rows of $\mathcal{G}_1$ have special structure, which is a consequence of the underlying state-space structure for the covariance matrix R_y. More specifically, going from one row of $\mathcal{G}_1$ to another (except for the first row) just changes the power of the F matrix. Hence, we can deduce from the above equality that applying Θ_1 to all other rows of $\mathcal{G}_1$ results in the matrix

$$\overline{\mathcal{G}}_1 = \mathcal{G}_1\Theta_1 = \begin{bmatrix} R_{e,1}^{1/2} & 0 \\ H\bar{K}_{p,1} & H\bar{L}_1 \\ HF\bar{K}_{p,1} & HF\bar{L}_1 \\ \vdots & \vdots \end{bmatrix}.$$

Next we shift down the first p columns to get

$$\mathcal{G}_2 = \begin{bmatrix} R_{e,1}^{1/2} & H\bar{L}_1 \\ H\bar{K}_{p,1} & HF\bar{L}_1 \\ HF\bar{K}_{p,1} & HF^2\bar{L}_1 \\ \vdots & \vdots \end{bmatrix},$$

choose a $\mathcal{J}$-unitary matrix Θ_2, shift down, form Θ_3, and so on.

We see that because of the special state-space structure of the elements of the generator matrix $\mathcal{G}$ of R_y, there is again a significant redundancy in the factorization arrays: the equality of the first two nonzero rows tells enough to fill out all other rows. So the basic recursion is just the following, which coincides with the array form (13.2.6) of the (CKMS) recursions:

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & FL_i \end{bmatrix} \Theta_{i+1} = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} \end{bmatrix}, \tag{13.A.4}$$

where Θ_{i+1} is *any* $(I \oplus J)$-unitary matrix that introduces the block zero entry on the right-hand side, and

Remark. It is satisfying to see how the fast recursions, which were in fact instrumental in the initiation of the displacement structure theory itself (see Kailath (1991a) and Kailath and Sayed (1995)) turn out to be a special case of the resulting theory. However, the displacement structure framework also provides further insights into the nature of the fast recursions and, in fact, the extension to structured time-variant state-space models first arose through this connection (Sayed and Kailath (1992,1994a)): it turns out that the conditions (11.3.3), *viz.*,

$$H_i = H_{i+1}\Psi_i, \qquad F_{i+1}\Psi_i = \Psi_{i+1}F_i, \qquad G_{i+1} = \Psi_{i+1}G_i, \tag{13.A.5}$$

guarantee that R_y will still satisfy a relation of the form (*cf.* Prob.11.8)

$$R_y - \mathcal{Z}^p R_y [\mathcal{Z}^p]^* = \mathcal{G}\mathcal{J}\mathcal{G}^*,$$

but with different $\mathcal{G}$ and $\mathcal{J}$. Consequently, its Cholesky factorization can still be obtained by the generalized Schur algorithm — see Prob. 13.9. ♦

CHAPTER 14

Asymptotic Behavior

In this chapter we focus on the steady-state behavior of the Kalman filter, *i.e.*, its behavior as time progresses to infinity, a topic often regarded as one of the highlights of the theory. It certainly encompasses a rich variety of results, and the analysis is more detailed and challenging than anywhere else in this book. And as in Ch. 8, on the steady-state estimation problem, it calls on a number of fundamental concepts (*e.g.*, detectability, unit-circle controllability) and tools (*e.g.*, the PBH test) from linear system theory. Beginning readers may be content with a reading of the overviews in Sec. 1.5 and Secs. 14.1–14.3.

14.1 INTRODUCTION

Given the large number of results herein, and their sometimes technical nature, it will be useful to begin with a couple of overview sections.

14.1.1 Time-Invariant State-Space Models

Steady-state behavior is of most interest when the underlying state-space model is time-invariant, *i.e.*, $\{F, G, H, Q, S, R\}$ are constant matrices. Moreover, though not always essential, we shall also make the practically important assumption that

$$R > 0.$$

In this case, as noted in Sec. 9.5.1, we could also assume $S = 0$ (provided we replace $\{F, Q\}$ by $\{F^s = F - GSR^{-1}H,\ Q^s = Q - SR^{-1}S^*\}$). However, for notational convenience, we shall initially just assume that

$$S = 0,$$

and discuss the more general case in Sec. 14.7.

Time-invariant state-space models were studied in Ch. 8, where it was shown that as time goes to infinity, the output of any *stable* time-invariant state-space model converges to a stationary process. For such stationary processes, and for estimation problems given a semi-infinite observation interval, and with the additional assumption that the pair $\{F, GQ^{1/2}\}$ is unit-circle controllable or, in less model-dependent language, that the z-spectrum $S_y(z)$ of the output process $\{\mathbf{y}_i\}$ has no unit-circle zeros, we constructed a recursive Wiener filter for determining the innovations process as (*cf.* Sec. 8.4)

$$\begin{aligned} \mathbf{e}_i &= \mathbf{y}_i - H\hat{\mathbf{x}}_i, \\ \hat{\mathbf{x}}_{i+1} &= F\hat{\mathbf{x}}_i + K_p(\mathbf{y}_i - H\hat{\mathbf{x}}_i), \quad i > -\infty, \\ &= F_p\hat{\mathbf{x}}_i + K_p\mathbf{y}_i, \end{aligned} \tag{14.1.1}$$

where $F_p = F - K_pH$, $K_p = FPH^*R_e^{-1}$, $R_e = R + HPH^*$, and P was the unique stabilizing solution to the discrete-time algebraic Riccati equation (DARE)

$$P = FPF^* + GQG^* - K_pR_eK_p^*. \tag{14.1.2}$$

That is, P is such that F_p is a stable matrix. On the other hand, even when specialized to a time-invariant model, the Kalman filter recursion for processes observed for $i \geq 0$ is (*cf.* Thm. 9.2.1)

$$\begin{aligned} \mathbf{e}_i &= \mathbf{y}_i - H\hat{\mathbf{x}}_i, \\ \hat{\mathbf{x}}_{i+1} &= F\hat{\mathbf{x}}_i + K_{p,i}(\mathbf{y}_i - H\hat{\mathbf{x}}_i) \;=\; F_{p,i}\hat{\mathbf{x}}_i + K_p\mathbf{y}_i, \quad i \geq 0, \end{aligned} \tag{14.1.3}$$

where $F_{p,i} = F - K_{p,i}H$, $K_{p,i} = FP_iH^*R_{e,i}^{-1}$, $R_{e,i} = R + HP_iH^*$, and P_i satisfies the Riccati recursion

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0 = \Pi_0. \tag{14.1.4}$$

Note that this filter is *time-variant* and will be so even if Π_0 is chosen so as to make the process $\{\mathbf{y}_i, i \geq 0\}$ stationary over $[0, \infty)$. However, it is natural to expect that as $i \to \infty$, the time-variant Kalman filter will approach the recursive Wiener filter, and that

$$\lim_{i\to\infty} P_i = P, \quad \text{the unique stabilizing solution of the DARE (14.1.2),}$$

no matter what initial condition $\Pi_0 \geq 0$ *we choose.*

These results are in fact true, and moreover the convergence of P_i to P is at an exponential rate. However, the proofs require some effort; in return, though, they allow us to also deduce certain (at least, initially) surprising results, as we shall now describe.

14.1.2 Convergence for Indefinite Initial Conditions

The first surprising fact is that the convergence of P_i to P holds also for certain *indefinite*, and even *negative-semi-definite*, initial conditions Π_0 (provided they are bounded below by a certain negative-semi-definite matrix). This is not just a mathematical curiosity; it has some useful implications.

The first is numerical. Although no physical initial conditions can be indefinite, starting with $\Pi_0 \geq 0$, it is quite possible that numerical effects (especially, round-off errors in the computations) can cause the computed P_i, at some instant $i > 0$, to lose its required nonnegative-definite character (since it is to be the variance matrix of the error vector). The question is whether the algorithm will then break down and give meaningless results thereafter? Or whether if we continue the recursion, P_i will eventually recover its nonnegativeness and, more to the point, ultimately converge to the stabilizing solution, P? Now, since the coefficients in the Riccati recursion under consideration are constant, and since we are interested in $i \to \infty$, we can regard propagating the Riccati recursion with an indefinite P_i (at some time i) as the same as propagating it with an indefinite *initial* condition, $P_0 \triangleq P_i$. Therefore the result that P_i can converge for certain indefinite initial conditions implies that the Riccati recursion can recover from loss of nonnegativeness and still converge to the solution of the DARE. This surprising phenomenon has in fact been observed in practice.

The next issue is to try to understand why the above result is possible. One line of argument is that P_i is an "internal" variable, dependent upon the state-space model we have chosen for the process $\{\mathbf{y}_i\}$. Such internal quantities might be "nonphysical", as long as the statistical properties of "external" variables such as $\{\mathbf{y}_i\}$ and $\{\mathbf{e}_i\}$ are not affected. In the Kalman filter, the "external" quantity influenced by P_i is the innovations variance $R_{e,i} = R + HP_iH^*$, which is uniquely determined by the covariance matrix of the observed process $\{\mathbf{y}_i\}$,

$$R_y = \left[\langle \mathbf{y}_i, \mathbf{y}_j \rangle\right]_{i,j\geq 0}.$$

Now the $\{R_{e,i}\}$ are the diagonal entries in the unique LDL* factorization of R_y, and consequently the positive-definiteness of R_y implies that of the $\{R_{e,i}\}$. But since we assumed $R > 0$, it is clear that we can have $R_{e,i} = R + HP_iH^* > 0$ even if P_i is indefinite! This is in fact often the case, as we shall show in Lemmas 14.5.3 and 14.5.6.

However, an even more interesting fact is that convergence can still occur even if some $R_{e,i}$ fail to be positive-definite, as shown in Lemma 14.5.3. More explicitly, Lemma 14.5.3 describes a set of initial conditions, Π_0, for which we can always guarantee convergence of the Riccati recursion. This set is such that the resulting innovations variances $\{R_{e,i}\}$ are positive-definite for *almost all* $i \geq 0$, except only for a *finite* number of times that is at most equal to the negative inertia of the difference $\Pi_0 - P$! In particular, the result also guarantees that there is a *finite* time instant N_o beyond which the innovations variances $\{R_{e,i}\}$ are positive-definite (*i.e.*, for all $i \geq N_o$). This surprising result highlights an inherent robustness of the filter to possible loss of positive-definiteness in R_y; in loose terms, the result shows that if P_0 is such that the $\{R_{e,i}\}$ recover their positive-definiteness in finite time, then convergence can occur. Numerical examples that demonstrate these possibilities will be given.

In a similar vein, and under a stronger condition, Lemma 14.5.6 describes a smaller set of initial conditions, Π_0, for which P_i converges to P and for which the resulting innovations variances $\{R_{e,i}\}$ are positive-definite for all $i \geq 0$. When F is further assumed to be stable, this set of initial conditions is shown in Lemma 14.6.1 to be

equivalent to the set of all Π_0 for which R_y is strongly positive-definite, *i.e.*, $R_y > \epsilon I$ for some $\epsilon > 0$.[1]

It is interesting to see that in order to better understand the Kalman filter, we often have to go beyond the given (nonunique) state-space models of the process $\{\mathbf{y}_i\}$ to the (of course, unique) covariance function of the process $\{\mathbf{y}_i\}$, the starting point of the Wiener filtering theory. We shall see more examples of this fact as we proceed.

14.1.3 Convergence for Unstable F

There is a further, also initially surprising, property of the Kalman filter. This is that convergence holds even when the matrix F is *unstable*, provided that the model has certain reasonable detectability and controllability properties further discussed below. Thus, even though the process $\{\mathbf{y}_i\}$ is neither stationary, nor even asymptotically stationary, so that the whole premise of the Wiener filtering theory breaks down, it is still possible that the Kalman filter converges to a time-invariant one. The reason is essentially that even though the state variance will tend to infinity when F is unstable, so will the variance of the estimator $\hat{\mathbf{x}}_i$, since it obeys essentially the same recursion as $\mathbf{x}_i$, except for the variance of the driving noise: $\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + K_{p,i}\mathbf{e}_i$. Therefore we may expect that $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ will track each other in such a way that the error variance is finite, leading to a stationary steady-state filter for the *innovations*.

A necessary (and almost sufficient) condition for such behavior to hold is that the pair $\{F, H\}$ be *detectable* (see App. C), *i.e.*, that we can find a matrix K such that $F - KH$ is stable. The reason is that such a K can be used (in place of $K_{p,i}$) to define a suboptimal estimator (see Prob. 14.4), whose error variances are clearly bounded; then the (smaller) optimal error variances, P_i, are a fortiori bounded.

To further show that the $\{P_i\}$ converge to the unique stabilizing solution P, it turns out that we must also assume that $\{F, GQ^{1/2}\}$ is *unit-circle controllable*, *i.e.*, that there exists a K such that $F - GQ^{1/2}K$ has no unit-circle eigenvalues.

14.1.4 Why Study Models with Unstable F?

We begin by showing that the study of models with unstable F is of limited value for pure estimation problems. The point is that the finiteness of the error variance depends upon the state, $\mathbf{x}_i$, and the estimator, $\hat{\mathbf{x}}_i$, tracking each other at the same exponential rate, determined by the matrix F. However, in practice F is rarely known exactly and, when F is unstable, any difference between the true F and the one assumed in building the estimator can be catastrophic.

To see why, consider a state-space model where the true state matrix is given by $F + \delta F$, where F is the nominal state matrix and δF represents the modeling error, *i.e.*,

$$\mathbf{x}_{i+1} = (F + \delta F)\mathbf{x}_i + G\mathbf{u}_i.$$

[1] While $R_y > 0$ always implies that $R_{e,i} > 0$ for all i, the converse statement is not true in general. It will be true when F is stable. This fact marks the distinction between Lemmas 14.5.6 and 14.6.1, and it is discussed in greater detail in Sec. 14.6.

Now the recursion for the predicted state estimator $\hat{\mathbf{x}}_i$, as given by the Kalman filter, will use the nominal value of the state matrix. Thus,

$$\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + K_{p,i}(H\tilde{\mathbf{x}}_i + \mathbf{v}_i),$$

where, as usual, $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$. Subtracting these last two equations shows that the state estimation error satisfies

$$\tilde{\mathbf{x}}_{i+1} = (F - K_{p,i}H)\tilde{\mathbf{x}}_i + \delta F\mathbf{x}_i + G\mathbf{u}_i - K_{p,i}\mathbf{v}_i.$$

Note the existence of the extra term $\delta F\mathbf{x}_i$, as compared to the model error-free state estimation error recursion (9.2.23). Now, when $F + \delta F$ is unstable, $\mathbf{x}_i$ will be exponentially growing, so that even though $F - K_{p,i}H$ will be asymptotically stable (since its limit $F - K_pH$ will be stable), and even though $\mathbf{u}_i$ and $\mathbf{v}_i$ have finite variance, because of the extra driving term $\delta F\mathbf{x}_i$, the variance of $\tilde{\mathbf{x}}_{i+1}$ will be unbounded. So in principle, models with unstable F cannot really be used in the unavoidable presence of modeling errors.

Why then has there been so much study of the Riccati recursion and of the DARE for unstable F? And what led to the study of unstable F in the first place? The answer is that unstable F are encountered in control problems, where uncertainties in F are (fortunately) not so devastating. Let us briefly examine this problem.

The Deterministic Quadratic Regulator Control Problem. When Kalman first posed and solved the estimation problem for state-space models, he noticed (see Kalman (1960a)) that the Riccati recursion for the estimation problem was dual (in a certain sense — see, *e.g.*, Sec. 15.3.6) to a Riccati recursion he had previously obtained (Kalman and Koepcke (1958)) in the solution to the deterministic linear quadratic regulator (LQR) problem of optimal control. Now the interesting *control* problems are those for unstable systems (*i.e.*, unstable F), for which Kalman was able to establish[2] convergence of the recursion under certain controllability and observability assumptions on the state-space model. By duality, the convergence result can be taken to the estimation problem, which is how the case of unstable F was introduced into state-space estimation and heralded as a significant step beyond the Wiener solution.[3] Nevertheless, as noted above, uncertainties in the value of F can be fatal in the estimation problem. Fortunately, however, they need not be so in the control problem.

In the control problem (of Sec. 15.3.6), use of the feedback control law in steady state, say $u_i = -K^c x_i$, modifies the state equation $x_{i+1} = Fx_i + Gu_i$ to the closed-loop equation

$$x_{i+1} = Fx_i - GK^c x_i = (F - GK^c)x_i,$$

where K^c is such that $F - GK^c$ is stable, even if F is unstable. In the presence of model uncertainties, the equation changes to

$$x_{i+1} = (F + \delta F - GK^c)x_i.$$

[2] This was done using Lyapunov techniques, which Kalman again was instrumental in bringing to the attention of control engineers (outside the former Soviet Union) — see Kalman and Bertram (1958).

[3] It has been believed (see Kalman (1963a)) that convergence always held for stable F (the Wiener assumption); however, when $S \neq 0$, we also need the unit-circle controllability condition on $\{F^s, GQ^{s/2}\}$.

Thus, if δF is small enough there is still hope that the closed-loop system matrix can be stable, which is the desired goal in the control problem. [It should be mentioned that there can be cases of parameter perturbations that can reduce the "stability margin" of LQR designs to arbitrarily small values. Still, it appears that in many cases the LQR solution is quite "robust"; for more on such issues, see, *e.g.*, Soroka and Shaked (1984), Dorato, Abdallah, and Cerone (1995, p. 125), and Zhou, Doyle, and Glover (1996).]

The Measurement Feedback Problem and the Separation Principle. The deterministic quadratic regulator problem assumes that the control input u_i is the only input driving the state equation, $x_{i+1} = Fx_i + Gu_i$. Now in modeling, it is useful to assume that the system is also driven by a random exogenous input $\mathbf{w}_i$, so that the state equation becomes stochastic and of the form

$$\mathbf{x}_{i+1} = F\mathbf{x}_i + G_1\mathbf{w}_i + G_2\mathbf{u}_i. \tag{14.1.5}$$

Moreover, one often has access only to some combination of the states, say

$$\mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i, \tag{14.1.6}$$

where $\mathbf{v}_i$ models the random measurement noise. Since in these situations the controller can only use the measurement signal $\mathbf{y}_i$, such problems are referred to as measurement feedback control problems; they are studied in more detail in Sec. 15.5.3.[4]

It turns out that the solution to the measurement feedback problem is still given by a state feedback law, as in the LQR case, where now the (inaccessible) states, $\{\mathbf{x}_i\}$, are replaced by their linear least-mean-squares estimators, $\{\hat{\mathbf{x}}_i\}$, using the measurement signals, $\{\mathbf{y}_i\}$. That is, $\mathbf{u}_i = -K^c\hat{\mathbf{x}}_i$ in steady state. This useful separation of the control and estimation aspects has been called the *separation principle* or the *certainty equivalence* principle (see, *e.g.*, Joseph and Tou (1961), Gunckel and Franklin (1963), Whittle (1963), Wonham (1968a), and also Sec. 15.5).

The point here is that we now have a situation where estimators for models with unstable matrices F are needed. And, moreover, while perturbations in F can be fatal in the pure estimation problem, they need not be so in the combined observer-controller structure. Indeed, the recursion for the steady-state estimator in this case will be given by

$$\hat{\mathbf{x}}_{i+1} = F\hat{\mathbf{x}}_i + G_2\mathbf{u}_i + K_p(\mathbf{y}_i - H\hat{\mathbf{x}}_i). \tag{14.1.7}$$

Combining (14.1.5)–(14.1.7) leads to the following state recursion for the controlled system:

$$\begin{bmatrix}\mathbf{x}_{i+1}\\ \hat{\mathbf{x}}_{i+1}\end{bmatrix} = \begin{bmatrix}F & -G_2K^c\\ K_pH & F - G_2K - K_pH\end{bmatrix}\begin{bmatrix}\mathbf{x}_i\\ \hat{\mathbf{x}}_i\end{bmatrix} + \begin{bmatrix}G_1\\ 0\end{bmatrix}\mathbf{w}_i + \begin{bmatrix}0\\ K_p\end{bmatrix}\mathbf{v}_i.$$

[4] They are often also known as Linear Quadratic Gaussian (LQG) control problems, to reflect the fact that the system and controller are linear, the cost function is quadratic, and that the disturbances are often taken to be Gaussian random variables. We shall, however, not use this terminology since we will not be making the Gaussian assumption.

In the presence of uncertainties in F, the dynamics of the system will be determined by the modes of the matrix

$$\begin{bmatrix}F + \delta F & -G_2K^c\\ K_pH & F + \delta F - G_2K - K_pH\end{bmatrix},$$

which are given by the values λ at which

$$\det\begin{bmatrix}\lambda I - F - \delta F & G_2K^c\\ -K_pH & \lambda I - F - \delta F + G_2K + K_pH\end{bmatrix} = 0.$$

Now recall that the determinant does not change if we replace the second block column by the sum of the two columns, so that the modes also satisfy

$$\det\begin{bmatrix}\lambda I - F - \delta F & \lambda I - F - \delta F + G_2K^c\\ -K_pH & \lambda I - F - \delta F + G_2K^c\end{bmatrix} = 0.$$

Likewise, the determinant does not change if we replace the first block row by the difference of both rows, so that the modes satisfy

$$\det\begin{bmatrix}\lambda I - F - \delta F - K_pH & 0\\ -K_pH & \lambda I - F - \delta F + G_2K^c\end{bmatrix} = 0.$$

Thus, if δF is small enough there is still hope that the closed-loop system matrix will be stable. Still, as in the pure state-feedback case, there are special cases of parameter perturbations that can lead to arbitrarily small stability margins (see, *e.g.*, Doyle (1978) and Zhou, Doyle, and Glover (1996, p. 398)).

Remark 1. Though we shall be studying the asymptotic properties of the Kalman filter by examining the Riccati recursion for P_i, what is ultimately at issue is the convergence of the Kalman gain $K_{p,i}$ and the innovations variance $R_{e,i}$, since these determine the innovations representation. Therefore, our analysis will also extend to the alternative filtering and smoothing algorithms derived in Chs. 11 and 12 (see, *e.g.*, Sec. 14.8, which discusses the exponential convergence of the CKMS recursions). ◆

14.2 SOLUTIONS OF THE DARE

Before providing an overview of the major results of this chapter, we summarize in this section some of the basic background material on the DARE (14.1.2). Detailed proofs, and more results, are provided in App. E.

To begin with, the DARE (14.1.2) is essentially a system of algebraic equations and therefore may have many solutions, possibly even many positive semi-definite solutions (see the examples and further discussion in App. E). Moreover, none of these solutions may be stabilizing, *i.e.*, a solution that results in a stable closed-loop matrix $F_p = F - K_pH$. Therefore, the first question of interest is

Q1. When will the DARE (14.1.2) have a stabilizing solution P? Moreover, is the stabilizing solution unique?

The answer is given by Thm. E.5.1 and is stated below:

A1. The DARE (14.1.2) will have a stabilizing solution if, and only if, $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the unit circle. Moreover, any such stabilizing solution is unique and in fact also positive-semi-definite.

Thm. E.5.1 is proven in App. E. However, here we attempt to give some insight into the reasons for the requirements of detectability and unit-circle controllability. The condition that $\{F, H\}$ be detectable makes sense, since otherwise $F - KH$ will be unstable for *all* K.

The condition of controllability on the unit circle was seen in Sec. 8.3.1 to be the condition required for $S_y(z)$ to have no unit-circle zeros or, equivalently, for F_p to have no unit-circle eigenvalues. To further motivate this condition, suppose that $\{F, GQ^{1/2}\}$ has an uncontrollable mode on the unit circle, *i.e.*, that there exists an eigenvalue λ of F, with corresponding left eigenvector x, such that

$$xF = \lambda x, \quad xGQ^{1/2} = 0, \quad |\lambda| = 1.$$

Now let P be any solution of the DARE (14.1.2). Pre- and post-multiplying the DARE by x and x^* yields,

$$xPx^* = |\lambda|^2 xPx^* + xGQG^*x^* - xK_pR_eK_p^*x^*,$$

from which we infer that $xK_pR_eK_p^*x^* = 0$ and, hence, $xK_p = 0$ since R_e has full rank. (Note that $R_e = R + HPH^*$ has full rank, since our asssumption that the DARE (14.1.2) has a solution implies that $R_e^{-1} = (R + HPH^*)^{-1}$ exists.) This implies that

$$xF_p = x(F - K_pH) = xF - xK_pH = xF = \lambda x,$$

which means that F_p has a unit circle eigenvalue and hence cannot be stable.

The answer to the first question above also states that the stabilizing solution (when it exists) is unique and positive-semi-definite. This may encourage us to think that finding any positive-semi-definite solution of the DARE will be good enough to give us a stable closed-loop matrix, $F_p = F - K_pH$. Unfortunately, the answer is no. Under the detectability and unit-circle controllability assumptions, it turns out that there can exist positive-semi-definite solutions of the DARE that are not stabilizing (see the example below). [When this happens, we cannot assert that the Riccati recursion will converge to the unique stabilizing solution: just use one of the nonstabilizing positive-semi-definite solutions of the DARE as the initial condition for the Riccati recursion.] This discussion leads to:

Q2. When will the stabilizing solution of the DARE (14.1.2) be the only positive-semi-definite solution?

The answer is given by Thm. E.6.1 and is stated below:

A2. Assume that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the unit circle. Then the DARE (14.1.2) will have only one positive-semi-definite solution if, and only if, $\{F, GQ^{1/2}\}$ is *stabilizable*. Moreover, the unique positive-semi-definite solution of the DARE also defines its stabilizing solution.

In other words, to guarantee that there be only one positive-semi-definite solution, we need the additional condition that $\{F, GQ^{1/2}\}$ be stabilizable, *i.e.*, that it be controllable on *and* outside the unit circle (and not just on the unit circle, which was the condition for the existence of a stabilizing solution to the DARE). To illustrate this fact, let us assume that the stabilizability condition is not satisfied and then exhibit a positive-semi-definite solution of the DARE (14.1.2) that is not stabilizing. So consider the zero-initial-condition Riccati recursion,

$$P_{i+1}^0 = FP_i^0F^* + GQG^* - FP_i^0H^*(R + HP_i^0H^*)^{-1}HP_i^0F^*, \quad P_0^0 = 0. \quad (14.2.1)$$

Then the following facts hold:

(a) $\{P_i^0\}$ is a nondecreasing sequence of matrices ($P_{i+1}^0 \geq P_i^0$. [This was shown in Sec. 11.2.1.]

(b) P_i^0 is bounded from above. [This requires that $\{F, H\}$ be detectable. In this case, if we choose any matrix K such that $F - KH$ is stable, we may write $P_i^0 \leq \Pi$, for all $i \geq 0$, where Π is the unique solution of the Lyapunov equation

$$\Pi = F\Pi F^* + GQG^* + KRK^*.$$

Intuitively, the point is that Π can be regarded as the error covariance matrix in estimating the state using an observer with gain matrix K, which cannot outperform the "optimal" Kalman filter (*cf.* Prob. 14.4).]

These two facts imply that the sequence $\{P_i^0\}$ has a limit, $P^0 \geq 0$, say. Now assume that $\{F, GQ^{1/2}\}$ is not stabilizable, *i.e.*, that there exist some $\{\lambda, x\}$ such that,

$$xF = \lambda x, \quad xGQ^{1/2} = 0, \quad |\lambda| \geq 1.$$

Using induction we can show that $xP_i^0 = 0$, for all $i \geq 0$. Indeed for $i = 0$ this is clear since $P_0^0 = 0$. Moreover, if we suppose that $xP_i^0 = 0$ and pre-multiply the zero-initial-condition Riccati recursion (14.2.1) by x we obtain

$$xP_{i+1}^0 = \lambda xP_i^0F^* - \lambda xP_i^0H^*(R_{e,i}^0)^{-1}HP_i^0F^* = 0,$$

which is the desired result (where we have defined $R_{e,i}^0 = R + HP_i^0H^*$). This fact now shows that

$$xF_{p,i}^0 = x(F - FP_i^0H^*(R_{e,i}^0)^{-1}H) = \lambda x - \lambda xP_i^0H^*(R_{e,i}^0)^{-1}H = \lambda x,$$

which means that $F_{p,i}^0 = F - FP_i^0H^*(R_{e,i}^0)^{-1}H$ will have an unstable eigenvalue at λ for all $i \geq 0$. Thus clearly $P^0 \geq 0$, the limit of $\{P_i^0\}$, cannot be stabilizing since F_p^0 will also have an eigenvalue at λ. This then implies that $P^0 \geq 0$ must be different from $P \geq 0$, the desired stabilizing solution of the DARE, so that we have more than one positive-semi-definite solution. This discussion already suggests the following result.

Lemma 14.2.1 (Convergence of P_i^0) *Consider the zero-initial-condition Riccati recursion (14.2.1) and assume $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ unit-circle controllable so that the unique stabilizing solution P exists. Then P_i^0 converges to P if, and only if, $\{F, GQ^{1/2}\}$ is stabilizable.* ■

Proof: The detectability and unit-circle controllability assumptions on $\{F, H\}$ and $\{F, GQ^{1/2}\}$ guarantee that P_i^0 converges to a nonnegative-definite matrix P^0 that satisfies the DARE (14.1.2). Now assume $\{F, GQ^{1/2}\}$ is stabilizable. We then know from Thm. E.6.1 that the DARE (14.1.2) can only have a unique nonnegative-definite solution, which must also coincide with the stabilizing solution P. Therefore, $P^0 = P$. Conversely, assume P_i^0 converges to P. Then $\{F, GQ^{1/2}\}$ must be stabilizable since otherwise, as argued prior to the statement of the lemma, the $\{P_i^0\}$ will converge to a nonnegative-definite solution P^0 that is distinct from the stabilizing solution P. ♦

14.3 SUMMARY OF RESULTS

Since the chapter contains many results and several detailed calculations, it will be useful to summarize here, in more detail than in the introductory section, the major results.

A Sufficient Condition. Our analysis begins with a global identity (Lemma 14.4.2) relating the solution of the constant parameter Riccati recursion (14.1.4) for one initial condition to its solution for another initial condition. Lemma 14.4.2 is then used in Sec. 14.5 to compare the solution of the Riccati recursion for an arbitrary initial condition, P_0, with the stabilizing solution of the corresponding DARE, leading to the equation

$$P_{i+1} - P = F_p^{i+1}\left[I + (P_0 - P)\mathcal{O}_i^p\right]^{-1}(P_0 - P)F_p^{(i+1)*}, \tag{14.3.1}$$

where $\mathcal{O}_i^p$ satisfies the Lyapunov recursion

$$\mathcal{O}_{i+1}^p = F_p^*\mathcal{O}_i^p F_p + H^*R_e^{-1}H, \quad \mathcal{O}_{-1}^p = 0. \tag{14.3.2}$$

Since P is stabilizing, the matrix F_p is stable, and $\lim_{i\to\infty} F_p^i = 0$. Therefore, from relation (14.3.1), a sufficient condition for P_i to converge exponentially to P is that the sequence of matrices

$$T_i \triangleq [I + (P_0 - P)\mathcal{O}_i^p]^{-1}(P_0 - P), \tag{14.3.3}$$

be uniformly bounded, *i.e.*, that

$$\|T_i\|_2 \leq c, \tag{14.3.4}$$

for some finite positive c and for all i, and where $\|\cdot\|_2$ denotes the 2-induced norm (or maximum singular value) of its argument.[5]

The result (14.3.4) is established in Thm. 14.5.1 under the assumptions of a detectable pair $\{F, H\}$ and a unit-circle controllable pair $\{F, GQ^{1/2}\}$. It is further shown in Lemma 14.5.2 that condition (14.3.4) is equivalent to the conditions that the innovations variances $\{R_{e,i}\}$ be nonsingular for all $i \geq 0$ and that $I + (P_0 - P)\mathcal{O}^p$ be nonsingular, where $\mathcal{O}^p$ is given by the unique solution to the Lyapunov equation

$$\mathcal{O}^p = F_p^*\mathcal{O}^p F_p + H^*R_e^{-1}H. \tag{14.3.5}$$

[5] In fact, we show in Prob. 14.1 that the matrices T_i are Hermitian, so that the 2–induced (or spectral) norm, $\|T_i\|_2$, is also equal to the spectral radius, $\rho(T_i)$ (*i.e.*, the largest eigenvalue in magnitude) of T_i. Hence, condition (14.3.4) is equivalent to $\rho(T_i) \leq c$ for all i.

In fact, a stronger conclusion is established in Lemma 14.5.3 where it is shown that (14.3.4) results in variances $\{R_{e,i}\}$ that are not only nonsingular, but are also positive-definite for *almost* all i. Only at a finite number of times, which is at most equal to the negative inertia of the difference $P_0 - P$, will the $R_{e,i}$ lose positive-definiteness! The lemma also guarantees that there is a *finite* time instant N_o beyond which the innovations variances $\{R_{e,i}\}$ will be positive-definite (*i.e.*, for all $i \geq N_o$). This is a surprising result that highlights the inherent robustness of the filter to possible loss of positive-definiteness in the $\{R_{e,i}\}$; it essentially states that if P_0 is such that if the $\{R_{e,i}\}$ can recover their positive-definiteness in finite time, then convergence can occur. We shall illustrate this point with a few numerical examples (following the proof of Lemma 14.5.3).

Condition (14.3.4) does not guarantee convergence for all nonnegative-definite initial conditions P_0. For example, Lemma 14.5.5 shows that the (often assumed case of a) zero initial condition, $P_0 = 0$, does not result in a uniformly bounded sequence $\{T_i\}$ unless the pair $\{F, GQ^{1/2}\}$ is stabilizable (*i.e.*, controllable on and outside the unit circle, rather than just controllable on the unit circle); this is in agreement with Lemma 14.2.1 above. This same stabilizability condition guarantees convergence for all nonnegative-definite initial conditions, P_0, as established in Thm. 14.5.3 and as explained further ahead.

A Simplified Condition. The condition (14.3.4) requires that we check the uniform boundedness of a sequence of matrices $\{T_i\}$. In the general case, this is as difficult to verify as it is to compute the $\{P_i\}$ and verify that they indeed converge to P. However, it turns out that it is sufficient to check the positivity of a *single* matrix, rather than to check the uniform boundedness of an infinite number of matrices.

Sec. 14.5.2 gives the main result in this regard. The result (Thm. 14.5.2) states that if $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the unit circle, then a *sufficient* condition for the convergence of the Riccati recursion is that the initial condition P_0 be such that

$$I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2} > 0, \tag{14.3.6}$$

where $\mathcal{O}^p = \mathcal{O}^{p/2}\mathcal{O}^{p*/2}$ and $\mathcal{O}^{p/2}$ is full rank. The above condition describes one basin of attraction for the stabilizing solution of the DARE (14.1.2), and it is more restrictive than the condition (14.3.4). That is, the set of initial conditions P_0 that satisfy (14.3.6) is a subset of the set of initial conditions that satisfy (14.3.4). However, as compensation, it is shown in Lemma 14.5.6 that condition (14.3.6) results in variances $\{R_{e,i}\}$ that are now positive-definite for *all* $i \geq 0$. This set of initial conditions is further shown in Lemma 14.6.1, and for the case of stable systems F, to be equivalent to the set of all Π_0 for which R_y is strongly positive-definite, *i.e.*, $R_y > \epsilon I$ for some $\epsilon > 0$. As with (14.3.4), Lemma 14.5.8 also shows that the zero initial condition $P_0 = 0$ does not satisfy (14.3.6) unless the pair $\{F, GQ^{1/2}\}$ is stabilizable.

The Dual DARE. Conditions (14.3.4) and (14.3.6) guarantee the convergence of the Riccati variable P_i to P under the assumptions of a detectable $\{F, H\}$ and a unit-circle controllable $\{F, GQ^{1/2}\}$. They do not require the stabilizability of $\{F, GQ^{1/2}\}$; this condition is only needed to guarantee the convergence of the zero-initial-condition Riccati variable, P_i^0.

This issue is studied further in Sec. 14.5.3, where the dual DARE, for $S = 0$,

$$P^a = F^* P^a F + H^* R^{-1} H - F^* P^a G Q^{1/2} (I + Q^{*/2} G^* P^a G Q^{1/2})^{-1} Q^{*/2} G^* P^a F, \tag{14.3.7}$$

is introduced, with $Q^{1/2}$ denoting a square-root factor of Q ($Q = Q^{1/2} Q^{*/2}$). The relevance of the dual DARE to the study of the zero-initial-condition Riccati recursion (14.2.1), as well as to the study of the convergence of the Riccati recursion (14.1.2) for nonnegative-definite initial conditions P_0, arises from the following special case of Thm. E.8.1.

Theorem 14.3.1 (Stabilizing Solution to the Dual DARE) *Assume $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is unit-circle controllable. Then a unique stabilizing solution, P^a, of the dual* DARE *(14.3.7) exists if, and only if, either one of the following conditions holds:*

(i) $\{F, GQ^{1/2}\}$ is stabilizable.

(ii) The matrix $I - P\mathcal{O}^p$ is nonsingular, where $\mathcal{O}^p$ is the unique solution of (14.3.5).

Moreover, when the stabilizing solutions $\{P, P^a\}$ of the DARE *and the dual* DARE *exist (i.e., when $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is stabilizable), they are related via*

$$P^a = \mathcal{O}^p (I - P\mathcal{O}^p)^{-1}. \tag{14.3.8}$$

■

Condition (i) in the theorem therefore shows that the convergence of the zero-initial-condition Riccati recursion is equivalent to the existence of a stabilizing solution to the dual DARE. Condition (ii), on the other hand, can be used to show that stabilizability of $\{F, GQ^{1/2}\}$ also guarantees convergence of P_i to P for all nonnegative-definite initial conditions, P_0. This is because it is established in Thm. 14.5.3 that for stabilizable $\{F, GQ^{1/2}\}$, and in view of (ii) and of the nonnegative definiteness of P^a, condition (14.3.6) is equivalent to requiring

$$I + (P^a)^{*/2} P_0 (P^a)^{1/2} > 0, \tag{14.3.9}$$

where $(P^a)^{1/2}(P^a)^{*/2} = P^a$. The convergence for $P_0 \geq 0$ now follows immediately from (14.3.9). As mentioned earlier, the condition (14.3.9) also allows for some indefinite (and even negative semi-definite) initial conditions P_0. Indeed, under a certain observability assumption, it can be shown that P^a is invertible so that (14.3.9) can be replaced by the more revealing condition that (*cf.* Cor. 14.5.1)

$$P_0 > -(P^a)^{-1}. \tag{14.3.10}$$

We should also emphasize, as mentioned earlier, that conditions (14.3.6) and (14.3.9) have the interesting property of guaranteeing that the $\{R_{e,i}\}$ are positive-definite for all $i \geq 0$.

The main results mentioned above, and the relevant theorems, are collected in Table 14.1; note that as we add more assumptions to the state-space model, the region of convergence expands. We now move steadily to fill out all the details omitted above.

Table 14.1 Convergence of the Riccati recursion when $S = 0$.

Assumptions	Convergence guaranteed for	Relevant thms.
detectable $\{F, H\}$ and unit-circle controllable $\{F, GQ^{1/2}\}$	$\rho[I + (P_0 - P)\mathcal{O}_i^p]^{-1}(P_0 - P)] \leq c$	Thm. 14.5.1
detectable $\{F, H\}$ and unit-circle controllable $\{F, GQ^{1/2}\}$	$I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2} > 0$	Thm. 14.5.2
detectable $\{F, H\}$ and stabilizable $\{F, GQ^{1/2}\}$	$I + (P^a)^{*/2} P_0 (P^a)^{1/2} > 0$	Thm. 14.5.3
detectable $\{F, H\}$ and controllable $\{F, G\}$	$P_0 > -(P^a)^{-1}$	Cor. 14.5.1

14.4 RICCATI SOLUTIONS FOR DIFFERENT INITIAL CONDITIONS

There are many interesting identities involving solutions of Riccati equations. For continuous-time problems, *i.e.*, for Riccati *differential* equations (*cf.* Ch. 16), many of these identities are very old and are fairly easy to derive. The situation is different for Riccati difference equations, for at least two reasons. One is that the difference equations are of more recent origin; the other is that the corresponding identities often have a more complicated form (because additional second-order terms appear that vanish in the continuous-time limit. A good example is the identity (14.4.2) below whose continuous-time analogue is $\dot{P}(t) = \Psi(t, 0)\dot{P}(0)\Psi^*(t, 0)$, where $d\Psi(t, 0)/dt = (F - K(t)H)\Psi(t, 0)$, $\Psi(0, 0) = I$ — these quantities are defined in Ch. 16).

Lemma 14.4.1 (Local Identities) *Suppose $P_i^{(1)}$ and $P_i^{(2)}$ are two solutions to the Riccati recursion (14.1.4) with the same $\{F, G, H\}$ and $\{Q, R\}$ matrices, but with different initial conditions $\Pi_0^{(1)}$ and $\Pi_0^{(2)}$, respectively. Let $\delta P_i = P_i^{(2)} - P_i^{(1)}$. Then, assuming the required inverses exist, we have the following identities:*

$$\delta P_{i+1} = F_{p,i}^{(1)} \delta P_i F_{p,i}^{(2)*}. \tag{14.4.1}$$

$$\delta P_{i+1} = F_{p,i}^{(1)} \left[\delta P_i - \delta P_i H^* (R_{e,i}^{(2)})^{-1} H \delta P_i\right] F_{p,i}^{(1)*}, \tag{14.4.2}$$

where $F_{p,i}^{(m)} = F - K_{p,i}^{(m)} H$, $K_{p,i}^{(m)} = F P_i^{(m)} H^* (R_{e,i}^{(m)})^{-1}$, and $R_{e,i}^{(m)} = R + H P_i^{(m)} H^*$, for $m = 1, 2$. ■

Proof: The proof of the nonsymmetric identitiy (14.4.1) involves straightforward algebraic manipulations; it was perhaps first derived by Nishimura (1966). The proof of (14.4.2) is slightly more involved (see Prob. 14.6). In fact, an equivalent result was proved in Ch. 11 (see Lemma 11.1.3), where it was called a generalized Stokes identity; the formula was first derived in Kailath, Morf, and Sidhu (1973). ◆

Lemma 14.4.1 gives a *local* identity relating solutions of the Riccati recursion for different initial conditions. We now present a *global* identity, perhaps first noted in Lainiotis (1974).

Lemma 14.4.2 (A Global Identity) *Suppose $P_i^{(1)}$ and $P_i^{(2)}$ are two solutions to the Riccati recursion (14.1.4) with the same $\{F, G, H\}$ and $\{Q, R\}$ matrices, but with different initial conditions $\Pi_0^{(1)}$ and $\Pi_0^{(2)}$, respectively. Let $\delta P_{i+1} = P_{i+1}^{(2)} - P_{i+1}^{(1)}$ and, hence, $\delta P_0 = \Pi_0^{(2)} - \Pi_0^{(1)}$. Then, whenever the inverse in (14.4.3) exists, it holds that*

$$\delta P_{i+1} = \Phi_p^{(1)}(i+1,0)\left[I + \delta P_0 \mathcal{O}_i^{(1)}\right]^{-1} \delta P_0 \Phi_p^{(1)*}(i+1,0), \tag{14.4.3}$$

where

$$\Phi_p^{(1)}(i,0) = \begin{cases} F_{p,i-1}^{(1)} F_{p,i-2}^{(1)} \cdots F_{p,0}^{(1)} & i > 0, \\ I & i = 0, \end{cases} \tag{14.4.4}$$

is the state transition matrix of $F_{p,j}^{(1)} = F - K_{p,j}^{(1)} H$, and

$$\mathcal{O}_i^{(1)} = \sum_{j=0}^{i} \Phi_p^{(1)*}(j,0) H^* (R_{e,j}^{(1)})^{-1} H \Phi_p^{(1)}(j,0), \tag{14.4.5}$$

is the observability Gramian of $\{F_{p,j}^{(1)}, (R_{e,j}^{(1)})^{-1/2} H\}$, with $R_{e,j}^{(1)} = (R_{e,j}^{(1)})^{1/2} (R_{e,j}^{(1)})^{*/2}$. ∎

Proof: We prove (14.4.3) by induction. For $i = 0$, using (14.4.2), we have

$$\begin{aligned}
\delta P_1 &= F_{p,0}^{(1)}\left[\delta P_0 - \delta P_0 H^* (R_{e,0}^{(2)})^{-1} H \delta P_0\right] F_{p,0}^{(1)*} \\
&= \Phi_p^{(1)}(1,0)\left[I - \delta P_0 H^* (R_{e,0}^{(2)})^{-1} H\right] \delta P_0 \Phi_p^{(1)*}(1,0) \\
&= \Phi_p^{(1)}(1,0)\left[I - \delta P_0 H^* (R_{e,0}^{(1)} + H \delta P_0 H^*)^{-1} H\right] \delta P_0 \Phi_p^{(1)*}(1,0) \\
&= \Phi_p^{(1)}(1,0)\left[I + \delta P_0 H^* (R_{e,0}^{(1)})^{-1} H\right]^{-1} \delta P_0 \Phi_p^{(1)*}(1,0) \\
&= \Phi_p^{(1)}(1,0)\left[I + \delta P_0 \mathcal{O}_0^{(1)}\right]^{-1} \delta P_0 \Phi_p^{(1)*}(1,0)
\end{aligned}$$

as desired.

Now suppose that (14.4.3) is true for i. We shall show that the identity is true for $i+1$ as well. Indeed, using (14.4.2) and the above arguments we have

$$\begin{aligned}
\delta P_{i+1} &= F_{p,i}^{(1)}\left[\delta P_i - \delta P_i H^* (R_{e,i}^{(2)})^{-1} H \delta P_i\right] F_{p,i}^{(1)*} \\
&= F_{p,i}^{(1)}\left[I + \delta P_i H^* (R_{e,i}^{(1)})^{-1} H\right]^{-1} \delta P_i F_{p,i}^{(1)*} \\
&= F_{p,i}^{(1)}\left[I + \delta P_i H^* (R_{e,i}^{(1)})^{-1} H\right]^{-1} \Phi_p^{(1)}(i,0)\left[I + \delta P_0 \mathcal{O}_{i-1}^{(1)}\right]^{-1} \delta P_0 \Phi_p^{(1)*}(i+1,0).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\left[I + \delta P_i H^* (R_{e,i}^{(1)})^{-1} H\right] \Phi_p^{(1)}(i,0) \\
&= \Phi_p^{(1)}(i,0) + \Phi_p^{(1)}(i,0)\left[I + \delta P_0 \mathcal{O}_{i-1}^{(1)}\right]^{-1} \delta P_0 \underbrace{\Phi_p^{(1)*}(i,0) H^* (R_{e,i}^{(1)})^{-1} H \Phi_p^{(1)}(i,0)}_{\mathcal{O}_i^{(1)} - \mathcal{O}_{i-1}^{(1)}} \\
&= \Phi_p^{(1)}(i,0)\left[I + (I + \delta P_0 \mathcal{O}_{i-1}^{(1)})^{-1} \delta P_0 (\mathcal{O}_i^{(1)} - \mathcal{O}_{i-1}^{(1)})\right],
\end{aligned}$$

so that we may write

$$\begin{aligned}
&\left[I + \delta P_i H^* (R_{e,i}^{(1)})^{-1} H\right]^{-1} \Phi_p^{(1)}(i,0) \\
&\quad = \Phi_p^{(1)}(i,0)\left[I + (I + \delta P_0 \mathcal{O}_{i-1}^{(1)})^{-1} \delta P_0 (\mathcal{O}_i^{(1)} - \mathcal{O}_{i-1}^{(1)})\right]^{-1}.
\end{aligned}$$

Substituting into the expression for δP_{i+1} yields

$$\begin{aligned}
\delta P_{i+1} &= \Phi_p^{(1)}(i+1,0)\left[I + \delta P_0 \mathcal{O}_{i-1}^{(1)} + \delta P_0 (\mathcal{O}_i^{(1)} - \mathcal{O}_{i-1}^{(1)})\right]^{-1} \delta P_0 \Phi_p^{(1)*}(i+1,0) \\
&= \Phi_p^{(1)}(i+1,0)\left[I + \delta P_0 \mathcal{O}_i^{(1)}\right]^{-1} \delta P_0 \Phi_p^{(1)*}(i+1,0),
\end{aligned}$$

which is the desired result. ◆

Remark 2. These algebraic proofs are rather messy, especially in discrete-time (see Ch. 16 for the simpler continuous-time analogs). In fact, the simplest and most insightful proofs of many identities involving the Riccati recursion arise through a physical transmission line scattering model of the estimation problem, as we shall show in Ch. 17. [We may also mention that the global identities of Lemma 14.4.2 apparently first appeared in Lainiotis (1974). Their continuous-time counterparts are far more wellknown, and apparently go back to Sandor (1959); they can also be found in many textbooks, *e.g.*, Brockett (1970) and Reid (1972).] ◆

14.5 CONVERGENCE RESULTS

In this section we derive some general results concerning the convergence of the Riccati recursion (14.1.4). We give requirements on the initial condition P_0 such that the solution of the Riccati recursion converges to the unique stabilizing solution of the DARE (14.1.2). Although the requirements on P_0 may be difficult to verify in the general case, in later sections we shall see that, under some simplifying conditions, they result in more explicit requirements and yield basins of attraction for the initial condition, P_0, that are more general than those currently available in the literature.

14.5.1 A Sufficiency Result

The results presented in this section all use the identity (14.3.1). The first result is given below, where the notation $\rho(X)$ denotes the *spectral radius* of its argument:

$$\rho(X) \triangleq \max\{|\lambda| : \lambda \text{ is an eigenvalue of } X\}.$$

Theorem 14.5.1 (A Sufficiency Result) *Consider the Riccati recursion*

$$P_{i+1} = FP_iF^* + GQG^* - FP_iH^*(R + HP_iH^*)^{-1}HP_iF^*, \tag{14.5.1}$$

with initial condition P_0, and suppose that $\{F, H\}$ is detectable and that $\{F, GQ^{1/2}\}$ is controllable on the unit circle. Then if P_0 is a Hermitian matrix chosen such that the sequence of $n \times n$ Hermitian matrices

$$T_i \triangleq \left[I + (P_0 - P)\mathcal{O}_i^p\right]^{-1}(P_0 - P) \tag{14.5.2}$$

is uniformly bounded (cf. (14.3.4)), where $\mathcal{O}_i^p$ satisfies the recursion (14.3.2), then P_i converges to P, the unique stabilizing solution of the DARE *(14.1.2). Moreover, the convergence of P_i to P is exponential, viz.,*

$$\|P_i - P\| \leq \lambda^{2i}m, \tag{14.5.3}$$

for some matrix norm $\|\cdot\|$, a scalar $0 < \lambda < 1$, and a finite-positive m. ■

Proof: The assumptions of detectability and unit-circle controllability guarantee a stable F_p (*cf.* Thm. E.5.1), so that its spectral radius is smaller than one, say $\rho(F_p) = \gamma < 1$. Now there is a basic result in matrix theory (see Prob. 14.19), which states that for any matrix A with spectral radius $\rho(A)$, there always exists a (submultiplicative) matrix norm $\|A\|$ such that

$$\rho(A) \leq \|A\| \leq \rho(A) + \epsilon, \qquad \text{for any } \epsilon > 0.$$

So choose an $\epsilon > 0$ such that $\lambda \triangleq \gamma + \epsilon < 1$. Then, there exists a matrix norm, defined for any matrix, such that when applied to F_p it results in $\|F_p\| \leq \lambda < 1$. We shall denote this particular norm by $\|\cdot\|_\rho$ so that $\|F_p\|_\rho \leq \lambda < 1$.

Now taking norms of both sides of (14.3.1) we conclude that

$$\|P_{i+1} - P\|_\rho \leq \|F_p\|_\rho^{2(i+1)} \cdot \|T_i\|_\rho \leq m\lambda^{2(i+1)} \to 0 \quad \text{as } i \to \infty,$$

where, by the uniform boundedness assumption (14.3.4) on $\|T_i\|_2$, we have (by equivalence of norms)[6] that $\|T_i\|_\rho$ is also uniformly bounded, say $\|T_i\|_\rho \leq m$ for some m and for all i. ◆

Remark 3 [Only a Sufficient Condition] The convergence of P_i to P can still occur for some unbounded sequences $\{T_i\}$, so that the uniform boundedness condition (14.3.4) is indeed only a sufficient requirement for convergence. One example to this effect is the following. Consider the model below where all quantities are assumed scalar-valued,

$$\mathbf{x}_{i+1} = \mathbf{u}_i, \qquad \mathbf{y}_i = \frac{1}{\sqrt{3}}\mathbf{x}_i + \mathbf{v}_i.$$

[6] Any two matrix norms, say $\{\|A\|_a, \|A\|_b\}$, are equivalent. Indeed, define the (bounded, closed, and convex) set $\mathcal{A} = \{A \text{ such that } \|A\|_a = 1\}$ and let $k = \min_{\mathcal{A}} \|A\|_b$ and $K = \max_{\mathcal{A}} \|A\|_b$. The scalars $\{k, K\}$ are both nonzero and finite. Moreover, for any nonzero matrix B,

$$k \leq \left\| \frac{B}{\|B\|_a} \right\|_b \leq K,$$

so that $k\|B\|_a \leq \|B\|_b \leq K\|B\|_a$.

The system parameters are $f = 0, g = 1, h = \frac{1}{\sqrt{3}}$. Assume further that the noise sequences $\{\mathbf{u}_i, \mathbf{v}_i\}$ are white, uncorrelated, and have variances $r = 1$ and $q = 1$, respectively. The Riccati recursion (14.1.4) in this case collapses to $p_{i+1} = gqg^* = 1$, for all i and regardless of p_0 (the value of p_0 here is the variance of $\mathbf{u}_{-1}$). The stabilizing solution is therefore $p = 1$, which leads to the $r_e = \frac{4}{3}$, $k_p = 0$, and $f_p = 0$. Moreover, $\mathcal{O}_i^p = \frac{1}{4}$ for all $i \geq 0$. It then follows that the sequence $\{T_i\}$ in (14.5.2) is given by

$$T_i = \frac{p_0 - 1}{1 + \frac{1}{4}(p_0 - 1)}.$$

If we choose $p_0 = -3$, then T_i is unbounded for all i. ◆

Remark 4. From (14.3.1) we see that the same conclusion of exponential convergence of P_i to P would still hold if we instead require the uniform boundedness of the following sequence of (now non-Hermitian) $n \times n$ matrices (see Prob. 14.3),

$$X_i \triangleq [I + (P_0 - P)\mathcal{O}_i^p]^{-1}. \tag{14.5.4}$$

We shall continue our discussions with T_i, however, since it is a Hermitian matrix (as established in Prob. 14.1). ◆

The arguments in this chapter rely on the sufficient condition (14.3.4). In the sequel, we shall provide several equivalent conditions to the uniform boundedness of the $\{T_i\}$, which will therefore also guarantee the exponential convergence of P_i to P.

The first result below relies on a generalized square-root factorization of the (possibly indefinite) Hermitian matrix difference $P_0 - P$, *viz.*,

$$P_0 - P = A_0JA_0^*, \tag{14.5.5}$$

where J is an $r \times r$ ($r \leq n$) signature matrix representing the inertia of $P_0 - P$, say

$$J = \begin{bmatrix} I_\alpha & 0 \\ 0 & -I_\beta \end{bmatrix}, \qquad r = \alpha + \beta, \tag{14.5.6}$$

and A_0 is an $n \times r$ full rank matrix.

Lemma 14.5.1 (A First Equivalent Condition) *Consider again the same setting as in Thm. 14.5.1, with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ unit-circle controllable, so that the unique stabilizing solution P exists. The sequence of Hermitian matrices $\{T_i\}$ defined by (14.5.2) is uniformly bounded (cf. (14.3.4)) if, and only if, the $r \times r$ Hermitian matrices*

$$D_i \triangleq J + A_0^*\mathcal{O}_i^pA_0 \tag{14.5.7}$$

are nonsingular for all i, including in the limit as $i \to \infty$. ■

Proof: Using the factorization (14.5.5), and the matrix inversion lemma, we can write each T_i in the form

$$T_i = \left[I + A_0 J A_0^* \mathcal{O}_i^p\right]^{-1} A_0 J A_0^* = A_0 \left[J + A_0^* \mathcal{O}_i^p A_0\right]^{-1} A_0^* = A_0 D_i^{-1} A_0^*. \quad (14.5.8)$$

With this formula, and because A_0 has full rank, we conclude that T_i will be uniformly bounded if, and only if, the sequence of Hermitian matrices $\{D_i^{-1}\}$ is uniformly bounded, say $\|D_i^{-1}\|_2 \leq d$ for all i and for some finite-positive d (see Prob. 14.2 for more details).

Now recall that we mentioned earlier, following (14.3.3), that the spectral norm of a Hermitian matrix is also equal to its spectral radius. Therefore, the above implies that $\rho(D_i^{-1}) \leq d$ for all i. But since the eigenvalues of a matrix and its inverse are the inverses of each other, we conclude from the above that the smallest eigenvalue in magnitude of each D_i is bounded away from zero. Hence, all the D_i are nonsingular, including in the limit as $i \to \infty$. Note that the limit of D_i indeed exists since F_p is stable and, from (14.3.2), $\mathcal{O}_i^p$ converges to the unique solution of the Lyapunov equation (14.3.5) so that

$$D \triangleq \lim_{i \to \infty} D_i = J + A_0^* \mathcal{O}^p A_0. \quad (14.5.9)$$

◆

Remark 5. The nonsingularity of the $\{D_i\}$ for all i, including as $i \to \infty$, is thus equivalent to saying that the sequence $\{D_i\}$ is sufficiently and uniformly nonsingular. That is, there exists a positive number ϵ, such that the smallest eigenvalue in magnitude of each D_i is ϵ-away from zero,

$$\min_{1 \leq j \leq r} |\lambda_j(D_i)| > \epsilon > 0, \quad (14.5.10)$$

for all i. [Here, $\lambda_j(\cdot)$ denotes the j-th eigenvalue of its argument.] ◆

Example 14.5.1 (Convergence with Singular $\{D_i\}$) In accordance with Remark 3, convergence of P_i to P can still occur for a sequence $\{D_i\}$ that has singular entries. Indeed, consider the same model as in the example of Remark 3, and choose again $p_0 = -3$. Then $A_0 = 2$ and $J = -1$ so that $D_i = 0$ for all i. ◆

The convergence condition that the $\{T_i\}$ be uniformly bounded (or the $\{D_i\}$ be nonsingular) for all i is interesting since the matrices P and $\mathcal{O}_i^p$ do not depend upon the initial condition P_0. Nonetheless, these conditions are in terms of quantities $\{D_i, T_i\}$ that do not appear explicitly in the Riccati recursion (14.5.1). The next lemma provides a second equivalent condition for the uniform boundedness of the $\{T_i\}$ in terms of the innovations "variances" $\{R_{e,i}\}$.

First, however, observe from the recursion (14.3.2) for $\mathcal{O}_i^p$ that

$$\mathcal{O}_{i+1}^p = \sum_{j=0}^{i+1} F_p^{j*} H^* R_e^{-1} H F_p^j, \quad (14.5.11)$$

so that

$$\mathcal{O}_{i+1}^p = \mathcal{O}_i^p + F_p^{(i+1)*} H^* R_e^{-1} H F_p^{i+1}. \quad (14.5.12)$$

We then conclude from the definition (14.5.7) that two successive matrices, D_i and D_{i+1}, are related as follows:

$$\begin{aligned} D_{i+1} &= J + A_0^* \mathcal{O}_{i+1}^p A_0, \\ &= J + A_0^* \mathcal{O}_i^p A_0 + A_0^* F_p^{(i+1)*} H^* R_e^{-1} H F_p^{i+1} A_0, \\ &= D_i + A_0^* F_p^{(i+1)*} H^* R_e^{-1} H F_p^{i+1} A_0. \end{aligned} \quad (14.5.13)$$

Moreover, we shall often employ in our arguments a square-root factorization for the $n \times n$ matrix $\mathcal{O}^p$, which is the unique solution of the Lyapunov equation (14.3.5). We shall denote this factorization by

$$\mathcal{O}^p = \mathcal{O}^{p/2} \mathcal{O}^{p*/2}, \quad (14.5.14)$$

where $\mathcal{O}^{p/2}$ is assumed full rank.

Lemma 14.5.2 (A Second Equivalent Condition) *Consider the same setting as in Thm. 14.5.1, with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ unit-circle controllable so that the unique stabilizing solution P exists. Let also P_i denote the Riccati variable given by the recursion (14.5.1). The sequence of Hermitian matrices $\{T_i\}$ defined by (14.5.2) is uniformly bounded (cf. (14.3.4)) if, and only if, the $p \times p$ innovations variances,*

$$R_{e,i} \triangleq R + H P_i H^* \quad (14.5.15)$$

satisfy the following two conditions simultaneously:

(i) The $\{R_{e,i}\}$ are nonsingular for all finite i.

(ii) The matrix $I + (P_0 - P)\mathcal{O}^p$ is nonsingular, where $\mathcal{O}^p$ is the unique solution of the Lyapunov equation (14.3.5).

■

Proof: Proof:We need to prove both directions:
($\Rightarrow$) We first prove that if the $\{T_i\}$ are uniformly bounded, then the $\{R_{e,i}\}$ satisfy (i) and (ii) above. Indeed, by Lemma 14.5.1, the uniform boundedness of the $\{T_i\}$ implies the nonsingularity of the $\{D_i\}$ for all i, including in the limit. Thus, the matrix D defined in (14.5.9) is nonsingular, which means that $J + A_0^* \mathcal{O}^p A_0$ is nonsingular. This conclusion is equivalent to condition (ii) in the lemma. To see this, form the block matrix

$$\begin{bmatrix} J & A_0^* \mathcal{O}^{p/2} \\ \mathcal{O}^{p*/2} A_0 & -I \end{bmatrix},$$

Two lower-upper and upper-lower block triangular factorizations of this matrix show that the following block diagonal matrices

$$\begin{bmatrix} J & 0 \\ 0 & -(I + \mathcal{O}^{p*/2} A_0 J A_0^* \mathcal{O}^{p/2}) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} J + A_0^* \mathcal{O}^p A_0 & 0 \\ 0 & -I \end{bmatrix}, \quad (14.5.16)$$

are congruent, which means that $(I + \mathcal{O}^{p*/2} A_0 J A_0^* \mathcal{O}^{p/2})$ must be nonsingular. Using $\det(I + AB) = \det(I + BA)$, and $A_0 J A_0^* = (P_0 - P)$, we conclude that (ii) must hold.

To establish (i) we proceed as follows. Using (14.3.1) and (14.5.8), we write

$$P_{i+1} - P = F_p^{i+1} T_i F_p^{(i+1)*} = F_p^{i+1} A_0 D_i^{-1} A_0^* F_p^{(i+1)*} \triangleq A_{i+1} D_i^{-1} A_{i+1}^*, \quad (14.5.17)$$

where we have defined $A_{i+1} \triangleq F_p^{i+1} A_0$. Using (14.5.13) we thus have

$$D_{i+1} = D_i + A_{i+1}^* H^* R_e^{-1} H A_{i+1}, \quad (14.5.18)$$

where both $\{D_i, D_{i+1}\}$ are nonsingular by assumption. It then follows from the definition of $R_{e,i}$ in (14.5.15) that

$$R_{e,i+1} = R_e + H(P_{i+1} - P)H^* = R_e + H A_{i+1} D_i^{-1} A_{i+1}^* H^*.$$

Moreover, the detectability and unit-circle controllability assumptions guarantee that $R_e > 0$ so that R_e is invertible. Hence, using the matrix inversion lemma, we conclude that $R_{e,i+1}$ is also nonsingular since its inverse is well defined and given by

$$R_{e,i+1}^{-1} = R_e^{-1} - R_e^{-1} H A_{i+1} D_{i+1}^{-1} A_{i+1}^* H^* R_e^{-1}.$$

Therefore, condition (i) of the lemma is also satisfied.

($\Leftarrow$) Conversely, let us establish that conditions (i) and (ii) of the lemma imply the uniform boundedness of the $\{T_i\}$ or, equivalently, the nonsingularity of the $\{D_i\}$. We use induction. The initial matrix D_{-1} is clearly invertible since it is equal to J. Using (14.5.18) we have

$$D_0 = D_{-1} + A_0^* H^* R_e^{-1} H A_0,$$

so that by the matrix inversion lemma, and by the assumed nonsingularity of $R_{e,0}$,

$$D_0^{-1} = D_{-1}^{-1} - D_{-1}^{-1} A_0^* H^* R_{e,0}^{-1} H A_0 D_{-1}^{-1},$$

showing that D_0 is also invertible. Now assume D_j is invertible for all $j \leq i$ and let us show that D_{i+1} is also invertible because of the nonsingularity of $R_{e,i+1}$. This follows from (14.5.18) since the inverse of D_{i+1} is well defined and given by

$$D_{i+1}^{-1} = D_i^{-1} - D_i^{-1} A_{i+1}^* H^* R_{e,i+1}^{-1} H A_{i+1} D_i^{-1}.$$

This argument shows that D_i is invertible for all finite i. We still need to show that, in the limit, D_i is well defined and also invertible. For this purpose, recall from (14.5.9) that D is given by $D = J + A_0^* \mathcal{O}^p A_0$, and is therefore well defined. The nonsingularity of D now follows from the assumption (ii) and from the congruent matrices (14.5.16). ◆

Remark 6. Both conditions (i) and (ii) in Lemma 14.5.2 are required for the uniform boundedness of the sequence $\{T_i\}$, which is in turn a sufficient condition for the convergence of P_i to P. We illustrate this fact by means of an example that leads to a sequence $\{R_{e,i}\}$ that satisfies only (i) but not (ii) and for which convergence does not occur — so that (i) alone does not necessarily guarantee convergence of P_i to P. Indeed, recall that we argued earlier in Sec. 14.2, in response to Question 2 after recursion (14.2.1), that the zero-initial-condition Riccati recursion converges

to a steady-state value P^0 that is distinct from the stabilizing solution P when $\{F, GQ^{1/2}\}$ is not stabilizable. Now note that the $\{R_{e,i}^0\}$ generated by the zero-initial-condition Riccati recursion are all nonsingular (since $R_{e,i}^0 = R + HP_i^0 H^*$ and $R > 0$, $P_i^0 \geq 0$). Hence, this sequence satisfies condition (i) of Lemma 14.5.2. However, condition (ii) is not satisfied since otherwise convergence would be implied. [An alternative way for establishing that (ii) is not satisfied for this case is to recall the equivalence of conditions (i) and (ii) of Thm. 14.3.1, where the nonstabilizability of $\{F, GQ^{1/2}\}$ implies a singular $I - P\mathcal{O}^p$.] ◆

It turns out that we can be more explicit about the inertia of the innovations "variances" $\{R_{e,i}\}$. The following result establishes that if an initial condition P_0 guarantees a uniformly bounded sequence $\{T_i\}$, then the resulting sequence $\{R_{e,i}\}$ is not only nonsingular, but will be positive-definite for all i except possibly for a finite number of time instants.

In the proof of the following result we write $\text{In}(X) = \{n_+, n_-, n_o\}$ to denote the inertia of a matrix X in terms of the number of its positive eigenvalues (n_+), negative eigenvalues (n_-), and zero eigenvalues (n_o). Now recall from (14.5.6) that β is the negative inertia of the signature matrix J of $(P_0 - P)$.

Lemma 14.5.3 (Inertia of the $\{R_{e,i}\}$) *Consider again the setting of Thm. 14.5.1, with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ unit-circle controllable so that the unique stabilizing solution P exists. Assume the initial condition P_0 is chosen such that the matrices $\{T_i\}$ are uniformly bounded (cf. (14.3.4)). Then the resulting $p \times p$ matrices $\{R_{e,i}\}$ are positive-definite for all time instants i, except for at most β finite time instants where the $R_{e,i}$ are indefinite. In particular, there exists a finite time N_o such that*

$$R_{e,i} > 0 \quad \textit{for all} \quad i \geq N_o.$$

■

Proof: By the assumption of uniform boundedness of $\{T_i\}$ we conclude from Lemma 14.5.1 that the Hermitian matrices $\{D_i\}$ are nonsingular for all i, including $i \to \infty$, so that all the eigenvalues of each D_i are nonzero for all time instants.

Using (14.5.18), and the fact that $R_e > 0$, we obtain that for any row vector a,

$$a D_{i+1} a^* \geq a D_i a^*,$$

which in turn implies that *all* the eigenvalues of the matrices $\{D_i\}$ are nondecreasing with i. With these two conclusions regarding the eigenvalues of the $\{D_i\}$ (*viz.*, nondecreasing and nonzero), we can characterize the inertia of the $\{D_i\}$ for all i. More specifically, we now show that that the inertia of the sequence $\{D_i\}$ is piece-wise constant.

Thus note that the number of positive eigenvalues of D_i can only increase or remain constant with i. Starting from $D_{-1} = J$, which has α positive unity eigenvalues and β negative unity eigenvalues, we conclude that the inertia of D_i can be one of only $(\beta + 1)$ possibilities that are given by

$$\begin{bmatrix} I_\alpha & 0 \\ 0 & -I_\beta \end{bmatrix}, \begin{bmatrix} I_{\alpha+1} & 0 \\ 0 & -I_{\beta-1} \end{bmatrix}, \begin{bmatrix} I_{\alpha+2} & 0 \\ 0 & -I_{\beta-2} \end{bmatrix}, \quad \ldots, \quad I_r. \quad (14.5.19)$$

More explicitly, given that the inertia of D_{-1} is J, two possibilities can happen in view of the nondecreasing nature of the eigenvalues of D_i. One possibility is for the inertia to remain equal to J for all i. The other possibility is for the inertia to remain at J up to some time instant i_1 and then switch to a new inertia, with more positive eigenvalues than J. Let us denote this new inertia by J_1. The new inertia cannot have zero eigenvalues since the $\{D_i\}$ are always nonsingular.

Now starting from i_1, the inertia of the D_i can either remain at J_1 for all i or switch at some future time instant $i_2 > i_1$ to a new inertia J_2, with more positive eigenvalues than J_1, and so on. In the extreme case, only β such transitions in the inertia of D_i can occur. Let us assume that they occur at the time instants $\{i_1, i_2, \ldots, i_M, M \leq \beta\}$ so that the inertia is J for $-1 \leq i < i_1$, J_1 for $i_1 \leq i < i_2$, J_2 for $i_2 \leq i < i_3$, etc. This means that the inertia of the sequence $\{D_i\}$ is piece-wise constant.

With this result in mind, we now form the block matrix W

$$W = \begin{bmatrix} D_i & A_{i+1}^* H^* \\ H A_{i+1} & -R_e \end{bmatrix}. \tag{14.5.20}$$

Two different (block lower-upper and block upper-lower) triangular factorizations of W show that the block diagonal matrices

$$\begin{bmatrix} D_i & 0 \\ 0 & -R_{e,i+1} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} D_{i+1} & 0 \\ 0 & -R_e \end{bmatrix}, \tag{14.5.21}$$

are congruent and therefore have the same inertia. Now if D_i and D_{i+1} belong to the same interval of time over which the inertia remains constant, then the inertia of $\{R_{e,i+1}, R_e\}$ must coincide so that $R_{e,i+1} > 0$ over the same interval. Only when $i+1$ coincides with a transition time i_k, we will have that D_{i+1} has at least one more positive eigenvalue than D_i (say, γ_k more positive eigenvalues), so that we must have

$$\text{In}(R_{e,i_k}) = \{p - \gamma_k, \gamma_k, 0\}, \tag{14.5.22}$$

at i_k. Note further that the last transition time $\{i_M\}$ must be finite since D_i converges to a nonsingular matrix D and thus the $\{D_i\}$ have constant inertia in the limit, as $i \to \infty$. That is, there exists a finite N such that for all $i \geq N$, $\text{In}(D_i) = \text{In}(D)$, and, hence, $i_M < N$. We thus take $N_o = i_M + 1$.

Before concluding this proof, let us recall again that starting with $D_{-1} = J$, the positive inertia of D_{i_1}, at the time when the first transition occurs, can be one more than α, two more or, at most, β more. That is, since we are denoting this additional number of positive eigenvalues by γ_1, we must have

$$1 \leq \gamma_1 \leq \beta.$$

Similarly, the positive inertia of D_{i_2}, at the time when the second transition occurs, can be one more than that of D_{i_1}, two more or, at most, $\beta - \gamma_1$ more. That is, the integer γ_2 must satisfy

$$1 \leq \gamma_2 \leq \beta - \gamma_1.$$

In other words, the integers $\{\gamma_k, 1 \leq k \leq M\}$ that characterize the inertia of the $\{R_{e,i}\}$ during the M transition times $\{i_1, \ldots, i_M\}$ (*cf.* (14.5.22)), must all satisfy the nested relation

$$1 \leq \sum_{j=1}^{k} \gamma_j \leq \beta \quad \text{for all} \quad 1 \leq k \leq M. \tag{14.5.23}$$

◆

The following examples demonstrate the result of the above lemma and the possibility of indefinite innovations "variances" with a convergent P_i.

Example 14.5.2 (Convergence with Indefinite $\{R_{e,i}\}$) Consider the scalar Riccati recursion

$$p_{i+1} = \frac{1}{2} p_i - \frac{1}{2} p_i^2 (1 + p_i)^{-1}, \quad p_0 = \text{ initial condition},$$

with $h = 1, r = 1, g = q = 0$, and $f = 1/\sqrt{2}$. The system parameters are therefore such that f is stable, $\{f, h\}$ is detectable, and $\{f, gq^{1/2}\}$ is unit-circle controllable (since $f - gq^{1/2}k$ has no unit-circle eigenvalue for all k). The inverse of p_i can be seen to satisfy

$$p_{i+1}^{-1} = 2\left(p_i^{-1} + 1\right).$$

Now choose $p_0^{-1} = -3/4$ (which corresponds to $p_0 = -4/3$, a negative value). It follows that

$$p_1^{-1} = \frac{1}{2}, \quad p_2^{-1} = 3, \quad p_3^{-1} = 8, \quad \ldots$$

which shows that $p_i \to 0 \triangleq p$ and $f_{p,i} \to f \triangleq f_p$. The corresponding innovations "variances" are given by

$$r_{e,0} = -\frac{1}{3}, \quad r_{e,1} = 3, \quad r_{e,2} = \frac{4}{3}, \quad r_{e,3} = \frac{9}{8}, \quad \ldots$$

with $r_{e,i} \to 1 = r_e$. Condition (i) of Lemma 14.5.2 is therefore satisfied. Moreover, the corresponding steady-state observability Gramian is given by the solution of

$$\mathcal{O}^p = f_p^2 \mathcal{O}^p + r_e^{-1},$$

which yields $\mathcal{O}^p = 2$. Condition (ii) then evaluates to

$$1 + (p_0 - p)\mathcal{O}^p = 1 - \frac{8}{3} = -\frac{5}{3},$$

which is nonzero. Thus the conditions of Lemma 14.5.2 are satisfied for our choice of p_0. This implies that convergence should occur, as evidenced by the above calculations. Note further that

$$p_0 - p = p_0 = -\frac{4}{3},$$

so that we can take $J = -1$ and $A_0 = \frac{2}{\sqrt{3}}$. It follows, from the recursion

$$\mathcal{O}^p_{i+1} = f_p^2 \mathcal{O}^p_i + r_e^{-1}, \quad \mathcal{O}^p_{-1} = 0,$$

and from the definition $D_i = J + A_0^* \mathcal{O}^p_i A_0$, that

$$D_{-1} = -1, \quad D_0 = \frac{1}{3}, \quad D_1 = 1, \quad D_2 = \frac{4}{3}, \quad \cdots$$

with $D_i \to 5/3$. We thus see that the sequence $\{D_i\}$ changes sign (inertia) at the time instant $i = 0$, which is the same time instant at which $r_{e,0}$ has a negative eigenvalue, as predicted by Lemma 14.5.3. ◆

The next example repeats the above calculation for a state-space model where none of the system parameters $\{F, G, H, R, Q\}$ is zero. While the calculations are not as transparent as in the previous example, the same conclusion will nevertheless hold.

Example 14.5.3 (Convergence with Indefinite $\{R_{e,i}\}$) Consider the scalar Riccati recursion

$$p_{i+1} = \frac{1}{2} p_i - \frac{1}{2} p_i^2 (1 + p_i)^{-1} + 1, \quad p_0 = \text{initial condition},$$

with $h = 1, r = 1, g = q = 1$, and $f = 1/\sqrt{2}$. The system parameters are therefore such that f is stable, $\{f, h\}$ is detectable, and $\{f, gq^{1/2}\}$ is unit-circle controllable. The recursion for p_i can also be written as (by combining the first two terms)

$$p_{i+1} = \frac{1}{2} \left(p_i^{-1} + 1\right)^{-1} + 1.$$

Its stabilizing solution can be found as the nonnegative root of the equation

$$p = \frac{1}{2}(p^{-1} + 1)^{-1} + 1,$$

which leads to the quadratic equation $2p^2 - p - 2 = 0$, so that

$$p = \frac{1 + \sqrt{17}}{4}.$$

Now choose $p_0 = -4/3$, a negative value. It follows that

$$p_1 = 3, \quad p_2 = \frac{11}{8}, \quad p_3 = \frac{49}{38}, \quad \cdots$$

with $p_i \to p$. The corresponding innovations "variances" are given by

$$r_{e,0} = -\frac{1}{3}, \quad r_{e,1} = 4, \quad r_{e,2} = \frac{19}{8}, \quad r_{e,3} = \frac{87}{38}, \quad \cdots$$

with

$$r_e = 1 + p = \frac{5 + \sqrt{17}}{4}.$$

Condition (i) of Lemma 14.5.2 is therefore satisfied. Moreover,

$$f_p = f - f p r_e^{-1} = \frac{1}{\sqrt{2}} \frac{4 + \sqrt{17}}{5 + \sqrt{17}},$$

and the corresponding steady-state observability Gramian is given by the solution of

$$\mathcal{O}^p = f_p^2 \mathcal{O}^p + r_e^{-1},$$

which yields

$$\mathcal{O}^p = \frac{51 + 12\sqrt{17}}{84 + 20\sqrt{17}}.$$

Condition (ii) then evaluates to

$$1 + (p_0 - p)\mathcal{O}^p = -\frac{191 + 47\sqrt{17}}{84 + 20\sqrt{17}},$$

which is nonzero (in fact, negative). Thus the conditions of Lemma 14.5.2 are satisfied for our choice of p_0. This implies that convergence should occur, as evidenced by the above calculations. Note further that

$$p_0 - p = -\frac{1}{12} (19 + 3\sqrt{17}),$$

so that we can take $J = -1$ and

$$A_0 = \sqrt{\frac{1}{12} (19 + 3\sqrt{17})}.$$

It follows, from the recursion

$$\mathcal{O}^p_{i+1} = f_p^2 \mathcal{O}^p_i + r_e^{-1}, \quad \mathcal{O}^p_{-1} = 0,$$

and from the definition $D_i = J + A_0^* \mathcal{O}^p_i A_0$, that

$$D_{-1} = -1, \quad D_0 = \frac{4}{15 + 3\sqrt{17}}, \quad D_1 = \frac{1}{6} \frac{1371 + 331\sqrt{17}}{380 + 92\sqrt{17}}, \quad \cdots$$

with

$$D = \frac{1}{12} \frac{573 + 141\sqrt{17}}{84 + 20\sqrt{17}} > 0.$$

We thus see that the sequence $\{D_i\}$ changes sign (inertia) at the time instant $i = 0$, which is the same time instant at which $r_{e,0}$ has a negative eigenvalue, as predicted by Lemma 14.5.3. ◆

EXAMPLE 14.5.4 (The Case of Unbounded $\{T_i\}$) Recall from Remark 3 that the uniform boundedness of $\{T_i\}$ is only a sufficient condition for convergence of P_i to P. When it holds, the inertia of the resulting $\{R_{e,i}\}$ will be positive-definite except for at most β time instants. When uniform boundedness does not hold, convergence can still occur and the matrices $\{R_{e,i}\}$ can also be indefinite. To see this, consider again the model in Remark 3 where $p_{i+1} = 1$ for all i and regardless of p_0. Choose any $p_0 < -3$ then $r_{e,0} < 0$, while $r_{e,i} = \frac{4}{3}$ for all $i > 1$. ♦

The following result states that the inertia properties of the $\{R_{e,i}\}$ in Lemma 14.5.3 are in fact equivalent to the uniform boundedness of the $\{T_i\}$.

Lemma 14.5.4 (A Third Equivalent Condition) *Consider the same setting as in Thm. 14.5.1, with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ unit-circle controllable so that the unique stabilizing solution P exists. Let J denote the signature of the difference $P_0 - P$ as in (14.5.5)–(14.5.6). The sequence of Hermitian matrices $\{T_i\}$ defined by (14.5.2) is uniformly bounded (cf. (14.3.4)) if, and only if,*

(i) The $\{R_{e,i}\}$ are positive-definite for all finite i, except for at most β time instants (say M of them) where the $R_{e,i}$ have inertia $\{p - \gamma_k, \gamma_k, 0\}$ for some integers γ_k that satisfy (14.5.23).

(ii) The matrix $I + (P_0 - P)\mathcal{O}^P$ is nonsingular. ■

Proof: One direction is immediate. If the sequence $\{T_i\}$ is uniformly bounded, then Lemmas 14.5.2 and 14.5.3 imply that (i) and (ii) above must hold. Conversely, assume the $\{R_{e,i}\}$ have the inertia properties mentioned in (i), then they are nonsingular for all finite i. When combined with (ii), and using Lemma 14.5.2, this result implies that the $\{T_i\}$ must be uniformly bounded. ♦

The next result establishes that the zero initial condition $P_0 = 0$ does not belong to the set of initial conditions P_0 that guarantee a uniform bounded sequence $\{T_i\}$, unless $\{F, GQ^{1/2}\}$ is stabilizable.

Lemma 14.5.5 (Zero Initial Condition) *Let P denote the unique stabilizing solution of the* DARE *(14.1.2) and let $\mathcal{O}^P$ be the unique solution of (14.3.5), where it is assumed that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is unit-circle controllable. Then the following three statements are equivalent:*

(i) $\{F, GQ^{1/2}\}$ is stabilizable.

(ii) The matrix $(I - P\mathcal{O}^P)$ is invertible.

(iii) The zero initial condition $P_0 = 0$ belongs to the set that guarantees a uniform bounded sequence $\{T_i\}$. ■

Proof: We proceed as follows:

(i)⇒(ii): The equivalence of conditions (i) and (ii) in Thm. 14.3.1 shows that the stabilizability of $\{F, GQ^{1/2}\}$ implies the nonsingularity of $(I - P\mathcal{O}^P)$.

(ii)⇒(iii): We now prove that an invertible $(I - P\mathcal{O}^P)$ implies a uniformly bounded sequence $\{T_i\}$ for $P_0 = 0$. Indeed, recall that the sequence $\{R^0_{e,i}\}$ that is generated by the zero-initial-condition Riccati recursion (14.2.1) satisfies condition (i) of Lemma 14.5.2. When coupled with the invertibility of $(I - P\mathcal{O}^P)$, this fact implies that the choice $P_0 = 0$ results in a uniformly bounded sequence $\{T_i\}$.

(iii)⇒(i): Let us now show that if $P_0 = 0$ results in a uniformly bounded sequence $\{T_i\}$, then $\{F, GQ^{1/2}\}$ must be stabilizable. Assume, to the contrary, that $\{F, GQ^{1/2}\}$ is not stabilizable. By the uniform boundedness of $\{T_i\}$ we conclude from Thm. 14.5.1 that the zero-initial-condition Riccati variable P^0_i, generated by (14.2.1), should converge to P. However, by the nonstabilizability assumption, we conclude from the analysis following Question 2 in Sec. 14.2 that P^0_i does not converge to P; a contradiction. ♦

One immediate application of the above result is to reconsider Lemma 14.2.1 regarding the convergence of the zero-initial-condition Riccati recursion (14.2.1) to the unique stabilizing solution P. In our earlier proof of the sufficiency of the stabilizability of $\{F, GQ^{1/2}\}$ we appealed to Thm. E.6.1, which guarantees the existence of a unique nonnegative-definite solution of the DARE (14.1.2). This uniqueness was then used to conclude that P^0 must coincide with P. Alternatively, in light of the equivalence of statements (i) and (iii) in Lemma 14.5.5, we now see that a stabilizable pair $\{F, GQ^{1/2}\}$ implies that the zero initial condition, $P_0 = 0$, leads to a uniformly bounded sequence $\{T_i\}$ so that P^0_i should converge to P.

14.5.2 Simplified Convergence Conditions

In the previous section we studied the convergence of the Riccati recursion (14.5.1) under general conditions and showed that it required the uniform boundedness of the sequence of matrices $\{T_i\}$ or, equivalently, the nonsingularity of the sequences $\{D_i, R_{e,i}\}$. In the general case, these are as difficult to check as it is to compute the $\{P_i\}$ and verify that they indeed converge to P! However, it turns out that it is sufficient to check the positivity of a *single* matrix, rather than to check the nonsingularity or boundedness of an infinite sequence of matrices.

Theorem 14.5.2 (A Sufficient Convergence Condition) *Consider the earlier Riccati recursion (14.5.1) with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ controllable on the unit circle. Moreover, let P denote the unique stabilizing solution of the* DARE *(14.1.2), and let $\mathcal{O}^P \geq 0$ be the unique solution of the Lyapunov equation (14.3.5). Also consider the square-root factorization (14.5.14) of $\mathcal{O}^P$. Then, if the initial condition P_0 is a Hermitian matrix satisfying*

$$I + \mathcal{O}^{P*/2}(P_0 - P)\mathcal{O}^{P/2} > 0, \tag{14.5.24}$$

P_i converges exponentially to P. ■

Proof: Since F_p is stable, the limit matirx D of the sequence $\{D_i\}$ is well defined and given by (14.5.9). Now consider the following block matrix

$$\begin{bmatrix} -I & \mathcal{O}^{P*/2}A_0 \\ A_0^*\mathcal{O}^{P/2} & J \end{bmatrix}. \tag{14.5.25}$$

Two different (lower-upper and upper-lower) block triangular factorizations of the above matrix shows that the matrices

$$\begin{bmatrix} -I & 0 \\ 0 & D \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -I - \mathcal{O}^{p*/2} A_0 J A_0^* \mathcal{O}^{p/2} & 0 \\ 0 & J \end{bmatrix} \tag{14.5.26}$$

are congruent and must therefore have the same inertia. Thus, since $A_0 J A_0^* = P_0 - P$ and using the given condition (14.5.24), the matrices D and J will have the same inertia. Now recall from the proof of Lemma 14.5.3 that the inertia of the $\{D_i\}$ is piece-wise constant, and that it can only move from one inertia to another that has more positive eigenvalues. The above conclusion shows that $D_{-1} = J$ and D have the same inertia, which therefore implies that the inertia of all other D_i is constant and equal to the same J. We then conclude that all the D_i are nonsingular, including in the limit as $i \to \infty$, which guarantees by Lemma 14.5.1 and Thm. 14.5.1 the exponential convergence of P_i to P. ♦

Condition (14.5.24) describes one basin of attraction for the stabilizing solution of the DARE (14.1.2), and it is more restrictive than the condition of Thm. 14.5.1 requiring the uniform boundedness of the $\{T_i\}$. That is, the set of initial conditions P_0 that satisfy (14.5.24) is a subset of the set of initial conditions that result in uniformly bounded $\{T_i\}$. This is because condition (14.5.24) is equivalent to guaranteeing that the inertia of the matrices $\{D_i\}$ is constant and equal to J for all i, which is stronger than requiring the $\{D_i\}$ to be nonsingular for all i.

Moreover, comparing condition (ii) of Lemma 14.5.2 with (14.5.24), we see that the latter is a positivity condition while the former is only a nonsingularity condition on the same matrix. Indeed, since $\det(I+AB) = \det(I+BA)$, we have that $I+(P_0-P)\mathcal{O}^p$ is nonsingular if, and only if, $I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2}$ is nonsingular. But what about condition (i) of Lemma 14.5.2? The result of Thm. 14.5.2 shows that the positivity condition (14.5.24) is enough for convergence. This is because, as we show below, it guarantees the positivity of the innovations variances, $\{R_{e,i}\}$, so that condition (i) of Lemma 14.5.2 is automatically satisfied. The following statement thus strengthens the conclusion of Lemma 14.5.3.

Lemma 14.5.6 (Inertia of $R_{e,i}$) *Consider the setting of Thm. 14.5.2. Then, if the initial condition of the Riccati recursion is chosen such that (14.5.24) holds, the $\{R_{e,i}\}$ will be positive-definite for all i, including in the limit as $i \to \infty$.* ■

Proof: As established in the proof of Thm. 14.5.2, the condition (14.5.24) guarantees that the inertia of the $\{D_i\}$ is constant and equal to J for all i. It then follows from the proof of Lemma 14.5.3, and in particular from the congruence of the block diagonal matrices in (14.5.21), that the $\{R_{e,i}\}$ are positive-definite for all finite i. Moreover, by Thm. 14.5.2, we have that P_i converges to P so that $R_{e,i}$ converges to R_e and will therefore be positive-definite in the limit as well. ♦

A natural question is whether the converse of Lemma 14.5.6 holds. That is, if the $\{R_{e,i}\}$ are positive-definite for all i, will it follow that (14.5.24) holds? The answer is negative, as the following example shows.

Example 14.5.5 ($\{R_{e,i} > 0\}$ Does not Imply Condition (14.5.24)) Consider the scalar Riccati recursion

$$p_{i+1} = \frac{1}{4}p_i - \frac{1}{4}p_i^2(1+p_i)^{-1}, \quad p_0 = \text{ initial condition},$$

with $h = 1$, $r = 1$, $g = q = 0$, and $f = 1/2$. The system parameters are therefore such that f is stable, $\{f, h\}$ is detectable, and $\{f, gq^{1/2}\}$ is unit-circle controllable. The corresponding DARE leads to the equation $4p^2+3p = 0$, which has two solutions at $p = 0$ and $p = -3/4$. The former ($p = 0$) is the stabilizing solution and it leads to $f_p = 1/2$, $r_e = 1$, and $\mathcal{O}^p = 4/3$.

Assume now that we pick $p_0 = -3/4$ (the other solution of the DARE). This choice does not satisfy (14.5.24) since the expression $1 + p_0\mathcal{O}^p$ evaluates to zero. The resulting Riccati variable, however, will be such that $p_i = -3/4$ and $r_{e,i} = 1/4$, for all i. We thus have an example of a situation with $\{r_{e,i} > 0\}$ and p_0 not satisfying (14.5.24). ◆

Another way to see that $\{R_{e,i} > 0\}$ does not imply (14.5.24) is to reconsider the zero-initial-condition Riccati recursion (14.1.2). We argued in Sec. 14.2 that, for nonstabilizable $\{F, GQ^{1/2}\}$, P_i^0 does not converge to P while the resulting $\{R_{e,i}^0\}$ are positive-definite for all i (since $R > 0$ and $P_i^0 \geq 0$). This means that condition (14.5.24) must be violated since otherwise convergence would be implied.

The following statement provides an equivalent characterization of condition (14.5.24). In addition to the positive-definiteness of the $\{R_{e,i}\}$, the lemma shows that the nonsingularity of $I + (P_0 - P)\mathcal{O}^p$ is also needed — compare this statement with that of Lemma 14.5.2.

Lemma 14.5.7 (An Equivalent Condition) *Consider the earlier Riccati recursion (14.5.1) with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ controllable on the unit circle. Moreover, let P denote the unique stabilizing solution of the* DARE *(14.1.2), and let $\mathcal{O}^p \geq 0$ be the unique solution of the Lyapunov equation (14.3.5). Then the initial condition P_0 is a Hermitian matrix satisfying (14.5.24) if, and only if,*

(i) The $\{R_{e,i}\}$ are positive-definite for all finite i.

(ii) The matrix $I + (P_0 - P)\mathcal{O}^p$ is nonsingular.

■

Proof: Assume first that P_0 satisfies (14.5.24) and let us show that conditions (i) and (ii) are satisfied. The first condition is simply the result of Lemma 14.5.6. To establish (ii), we use the equality $\det(I + AB) = \det(I + BA)$ to note that

$$0 < \det[I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2}] = \det[I + (P_0 - P)\mathcal{O}^p],$$

so that $I + (P_0 - P)\mathcal{O}^p$ is invertible.

Now assume that (i) and (ii) above hold and let us show that P_0 must satisfy (14.5.24). Indeed, using the block diagonal matrices (14.5.21) that follow from two block triangular factorizations of the matrix W in (14.5.20), we conclude that $\text{In}(D_i) = \text{In}(D_{i+1})$ for all $i < \infty$. This is because $R_e > 0$ and $R_{e,i} > 0$ by condition (i). We thus conclude that

$$\text{In}(D_i) = J \qquad \text{for all } i < \infty,$$

since $D_{-1} = J$. We shall now prove that because of condition (ii) in the statement of the lemma, it should follow that $\text{In}(D) = J$ as well, where D is the limiting value of the sequence $\{D_i\}$. This limit matrix exists in view of the stability of F_p and it is given by (14.5.9).

Assume, to the contrary, that $\text{In}(D) \neq J$ and let us verify that this leads to a contradiction. To begin with note, as shown at the beginning of the proof of Lemma 14.5.2, that condition (ii) above implies that D is necessarily nonsingular. Moreover, we also showed before, in the proof of Lemma 14.5.3, that the eigenvalues of the matrices $\{D_i\}$ are non-decreasing with i. Thus let λ_i denote one of the *negative* eigenvalues of D_i that becomes positive in the limit, *i.e.,* at $i \to \infty$. Let $\bar{\lambda}$ denote its limit value,

$$\lim_{i \to \infty} \lambda_i = \bar{\lambda} > 0.$$

That is,

$$\lambda_i < 0 \quad \text{for all finite } i, \qquad \lambda_\infty = \bar{\lambda} > 0.$$

By the assumption that $\text{In}(D) \neq J$, and by $\text{In}(D_i) = J$, such an eigenvalue should exist. Therefore, there should exist a finite integer $N(\epsilon)$ such that for any given $\epsilon > 0$, and for all $i > N(\epsilon)$,

$$|\lambda_i - \bar{\lambda}| < \epsilon.$$

Assume we choose $\epsilon = \bar{\lambda}/2$. Then it follows that $\lambda_i \in (\frac{\bar{\lambda}}{2}, \frac{3\bar{\lambda}}{2})$ for all $i > N(\epsilon)$. That is, λ_i assumes positive values for all $i > N(\epsilon)$. This contradicts the fact that $\lambda_i < 0$ for all $i < \infty$ so that the assumption $\text{In}(D) \neq J$ is wrong and we must have $\text{In}(D) = J$. Now recall from the proof of Thm. 14.5.2 that $\text{In}(D) = J$ is equivalent to (14.5.24). ♦

Returning to the case of the zero-initial-condition Riccati recursion (14.1.2), and recalling the result of Lemma 14.5.5, we see that condition (ii) in the lemma we just established can be met by assuming a stabilizable pair $\{F, GQ^{1/2}\}$. In fact, it turns out that stabilizability is a necessary and sufficient condition for the zero-initial-condition, $P_0 = 0$, to satisfy (14.5.24).

Lemma 14.5.8 (Zero Initial Condition) *Consider the setting of Thm. 14.5.2 with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ unit-circle controllable. Then we have that $(I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2}) > 0$ if, and only if, $\{F, GQ^{1/2}\}$ is stabilizable.* ■

Proof: The argument prior to the statement of the lemma for P_i^0 shows that the lack of stabilizability implies that (14.5.24) does not hold. Thus condition (14.5.24) implies stabilizability.

Conversely, assume $\{F, GQ^{1/2}\}$ is stabilizable. Then from the equivalence of conditions (i) and (ii) in Thm. 14.3.1 we conclude that $(I - P\mathcal{O}^p)$ is nonsingular, and from (14.3.8) we conclude that $\mathcal{O}^p(I - P\mathcal{O}^p)^{-1} \geq 0$ (since $P^a \geq 0$).

Now the invertibility of $(I - P\mathcal{O}^p)$ implies the invertibility of $(I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2})$ in view of the relation $\det(I + AB) = \det(I + BA)$. Moreover,

$$\mathcal{O}^p(I - P\mathcal{O}^p)^{-1} = \mathcal{O}^{p/2}(I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2})^{-1} \mathcal{O}^{p*/2}.$$

This equality can be verified by checking that the difference of both expressions is zero. Indeed, let Δ denote the difference. Then

$$\begin{aligned} \Delta &= \mathcal{O}^{p/2}(I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2})^{-1} \big((I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2}) \mathcal{O}^{p*/2} \\ &\qquad - \mathcal{O}^{p*/2}(I - P\mathcal{O}^p) \big) (I - P\mathcal{O}^p)^{-1} \\ &= 0. \end{aligned}$$

It thus follows that

$$\mathcal{O}^{p/2}(I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2})^{-1} \mathcal{O}^{p*/2} \geq 0,$$

which is only possible if

$$I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2} > 0.$$

Indeed, assume $x^*[I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2}]x < 0$ for some nonzero vector x. Then, since $\mathcal{O}^{p*/2}$ has full rank (with generally more columns than rows), we conclude that there exists a nonzero vector y such that $x = \mathcal{O}^{p*/2} y$. It then follows that

$$y^* \mathcal{O}^{p/2}(I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2})^{-1} \mathcal{O}^{p*/2} y < 0,$$

which is a contradiction. ♦

Finally, it should be mentioned that it is also possible to develop alternatives to condition (14.5.24) that would guarantee a different inertia pattern for the $\{D_i\}$ (rather than having a constant inertia that is equal to J for all i, as guaranteed by (14.5.24)). For example, Thm. 14.5.2 would still hold if we replace (14.5.24) by the following condition (in terms of a square-root factor of the initial observability Gramian, $\mathcal{O}_0^p = H^* R_e^{-1} H$, rather than its steady-state value, $\mathcal{O}^p$).

Lemma 14.5.9 (A Second Convergence Condition) *Consider the same setting of Thm. 14.5.2 and let $J = (I_\alpha \oplus -I_\beta)$ denote the signature of the difference $P_0 - P$ as in (14.5.5). Let also H be $p \times n$. If $p \geq \beta$ and the initial condition P_0 is a Hermitian matrix satisfying the inertia condition*

$$\text{In}\left[I + \mathcal{O}_0^{p*/2}(P_0 - P)\mathcal{O}_0^{p/2} \right] = \{p - \beta, \beta, 0\}, \tag{14.5.27}$$

where $\mathcal{O}_0^{p/2} = H^* R_e^{-*/2}$, then P_i converges exponentially to P. Moreover, in this case, all the $\{R_{e,i}\}$ will be positive-definite for $i \geq 1$, while $R_{e,0}$ will have inertia $\{p - \beta, \beta, 0\}$.

■

Proof: Consider the block matrix

$$\begin{bmatrix} J & A_0^* \mathcal{O}_0^{p/2} \\ \mathcal{O}_0^{p*/2} A_0 & -I \end{bmatrix},$$

and perform two lower-upper and upper-lower block triangular factorizations to obtain the congruent matrices

$$\begin{bmatrix} J & 0 \\ 0 & -(I + \mathcal{O}_0^{p*/2}(P_0 - P)\mathcal{O}_0^{p/2}) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} D_0 & 0 \\ 0 & -I \end{bmatrix},$$

where $J = (I_\alpha \oplus -I_\beta)$ is $r \times r$, D_0 is $r \times r$, H is $p \times n$, and I is $p \times p$. Using condition (14.5.27), and the fact that the above block diagonal matrices must have the same inertia, we conclude that D_0 is necessarily positive-definite.

In other words, the initial conditions P_0 that satisfy (14.5.27) will guarantee a positive-definite D_0. Now since, according to the proof of Lemma 14.5.3, the inertia of the successive D_i can only change to one with a larger number of positive eigenvalues, we conclude that all the D_i will be positive-definite for $i \geq 0$. Consequently, all the $\{D_i\}$ will be invertible and, by Lemma 14.5.1, P_i converges to P. The statement regarding the inertia of the $\{R_{e,i}\}$ follows from Lemma 14.5.3 by noting that the inertia of the $\{D_i\}$ switches only at $i = 0$ from J to I. ◆

Remark 7. The reader can check that Exs. 14.5.2 and 14.5.3 that were discussed earlier in Sec. 14.5.1 satisfy (14.5.27) and the conclusions of the lemma. ◆

14.5.3 The Dual DARE and Stabilizablity

Conditions (14.3.4) and (14.3.6) guarantee the convergence of the Riccati variable P_i to P under the assumptions of a detectable $\{F, H\}$ and a unit-circle controllable $\{F, GQ^{1/2}\}$. They do not require the stabilizability of $\{F, GQ^{1/2}\}$; this condition is only needed to guarantee the convergence of the zero-initial-condition Riccati variable, P_i^0, as shown by Lemmas 14.5.5 and 14.5.8. In this section we further clarify the convergence of the zero-initial-condition Riccati recursion, as well as convergence for nonnegative-definite initial conditions, by relating the convergence of the Riccati recursion to the study of the dual DARE (as perhaps first done in Ljung and Kailath (1976c)). For $S = 0$, the *dual* DARE is defined as

$$P^a = F^* P^a F + H^* R^{-1} H - F^* P^a G Q^{1/2} (I + Q^{*/2} G^* P^a G Q^{1/2})^{-1} Q^{*/2} G^* P^a F, \tag{14.5.28}$$

where $Q = Q^{1/2} Q^{*/2}$. When $Q > 0$, the above equation reduces to the equivalent form

$$P^a = F^* P^a F + H^* R^{-1} H - F^* P^a G (Q^{-1} + G^* P^a G)^{-1} G^* P^a F,$$

with Q^{-1}. We shall use (14.5.28) for generality.

Now, equation (14.5.28) will be said to have a stabilizing solution if a P^a satisfying (14.5.28) can be found such that the corresponding closed-loop matrix,

$$F^* - F^* P^a G Q^{1/2} (I + Q^{*/2} G^* P^a G Q^{1/2})^{-1} Q^{*/2} G^*,$$

is stable. Clearly, the existence of a stabilizing P^a is equivalent to the detectability of $\{F^*, Q^{*/2} G^*\}$ (the stabilizability of $\{F, GQ^{1/2}\}$) and the unit-circle controllability of $\{F^*, H^* R^{-*/2}\}$ (the unit-circle observability of $\{F, R^{-1/2} H\}$). These conditions follow immediately from Thm. 14.3.5. In particular, the equivalence of conditions (i) and (ii) in that theorem provide one additional equivalent statement to the three statements that already appear in Lemma 14.5.5, *viz.*, that

(*iv*) A stabilizing solution P^a exists to the dual DARE (14.5.28).

By Lemmas 14.5.5 and 14.5.8, this condition is therefore equivalent to the inclusion of the zero-initial-condition, $P_0 = 0$, in the set that guarantees a uniformly bounded sequence $\{T_i\}$ or the positivity condition $I - \mathcal{O}^{p*/2} P \mathcal{O}^{p/2} > 0$, so that convergence of P_i^0 to P is guaranteed.

Lemma 14.5.10 (Zero Initial Condition) *Assume that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is unit-circle controllable. Then P_i^0 converges to the unique stabilizing solution P of the* DARE *(14.1.2) if, and only if, a unique stabilizing solution, P^a, of the dual* DARE *(14.5.28) exists.* ■

Moreover, when $\{F, GQ^{1/2}\}$ is stabilizable, the various convergence conditions that we exhibited earlier can be re-expressed in terms of the stabilizing solution of the dual DARE. For example, using Lemma 14.5.2, and relation (14.3.8), we can establish the following result.

Lemma 14.5.11 (Convergence and the Dual DARE) *Consider the Riccati recursion (14.5.1) with initial condition P_0, and suppose that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is stabilizable. Then P_i will converge to P, the stabilizing solution of the* DARE *(14.1.2), if the initial condition P_0 is chosen such that*

(i) The matrices $R_{e,i} = R + H P_i H^*$ are nonsingular for all finite i.

(ii) The matrix $I + P_0 P^a$ is nonsingular.

■

Proof: First note that the detectability and stabilizability assumptions guarantee that the stabilizing solutions P and P^a, to the DARE and dual DARE, both exist. Referring back to Lemma 14.5.2, the only fact to prove is condition (ii) of that lemma, *viz.*, that $I + (P_0 - P)\mathcal{O}^p = I - P\mathcal{O}^p + P_0 \mathcal{O}^p$ be nonsingular. But since a stabilizing solution to the dual DARE exists, the matrix $I - P\mathcal{O}^p$ is nonsingular and hence we may write,

$$I - P\mathcal{O}^p + P_0 \mathcal{O}^p = \left[I + P_0 \mathcal{O}^p (I - P\mathcal{O}^p)^{-1}\right](I - P\mathcal{O}^p) = [I + P_0 P^a](I - P\mathcal{O}^p).$$

Thus, $I + (P_0 - P)\mathcal{O}^p$ is nonsingular, if, and only if, $I + P_0 P^a$ is nonsigular. ◆

The next result shows that convergence for nonnegative-definite initial conditions, and even for some indefinite or nonpositive-definite initial conditions, is guaranteed when $\{F, GQ^{1/2}\}$ is stabilizable.

Theorem 14.5.3 (Convergence with Indefinite P_0) *Consider the Riccati recursion (14.5.1) where $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is stabilizable. Suppose, moreover, that the initial condition P_0 is a Hermitian matrix such that*

$$I + (P^a)^{*/2} P_0 (P^a)^{1/2} > 0, \tag{14.5.29}$$

where $P^a = (P^a)^{1/2}(P^a)^{*/2}$ is the unique stabilizing solution to the dual DARE *(14.5.28). Then P_i converges to the unique stabilizing solution, P, of the* DARE *(14.1.2).* ∎

Proof: The stabilizability assumption guarantees, by Lemma 14.5.5, the invertibility of $(I - P\mathcal{O}^p)$. Next using an equality established in the proof of Lemma 14.5.8 we have

$$P^a = \mathcal{O}^p(I - P\mathcal{O}^p)^{-1} = \mathcal{O}^{p/2}(I - \mathcal{O}^{p*/2}P\mathcal{O}^{p/2})^{-1}\mathcal{O}^{p*/2}.$$

Now let $P_0 = M_0 S M_0^*$ denote a factorization of the possibly indefinite matrix P_0, where S is a diagonal signature matrix with as many $\pm 1's$ as P_0 has negative or positive eigenvalues. Moreover, $S^2 = S$.

Then, in view of Thm. 14.5.2, the desired result follows from the following sequence of equivalent positivity statements:

$$\begin{aligned} I + (P^a)^{*/2}P_0(P^a)^{1/2} > 0 &\Leftrightarrow \text{In}[S + M_0^* P^a M_0] = \text{In}(S) \\ &\Leftrightarrow \text{In}(S + M_0^*[\mathcal{O}^{p/2}(I - \mathcal{O}^{p*/2}P\mathcal{O}^{p/2})^{-1}\mathcal{O}^{p*/2}]M_0) = \text{In}(S) \\ &\Leftrightarrow I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2} > 0. \end{aligned}$$

For example, the first equivalent statement can be verified by considering the block matrix

$$\begin{bmatrix} I & P^{a*/2}M_0 \\ M_0^* P^{a/2} & -S \end{bmatrix},$$

and performing two lower-upper and upper-lower block triangular factorizations to conclude that the matrices

$$\begin{bmatrix} I + (P^a)^{*/2}P_0(P^a)^{1/2} & 0 \\ 0 & -S \end{bmatrix}, \quad \begin{bmatrix} I & 0 \\ 0 & -[S + M_0^* P^a M_0] \end{bmatrix}$$

are congruent. Similar constructions establish the other positivity statements. ♦

Note that the above result, although equivalent to (14.5.24) under the stabilizability assumption, it nevertheless shows more clearly that convergence of the Riccati recursion occurs for all positive-semi-definite initial conditions, $P_0 \geq 0$, since $I + (P^a)^{*/2}P_0(P^a)^{1/2}$ will clearly be positive definite. It also shows that convergence can be guaranteed for indefinite, and even negative-definite, initial conditions, as long as (14.5.29) is satisfied.

Under an additional observability condition, (14.5.29) can be made even more explicit. The result is given below. (The proof is simple and is omitted—see Prob. 14.13.)

Corollary 14.5.1. Conditions for Convergence *When $\{F, H\}$ is observable, the condition (14.5.29) is equivalent to $P_0 > -(P^a)^{-1}$. Moreover, in this case we have*

$$-(P^a)^{-1} = P_- \triangleq \textit{the infimum over all solutions to the} \text{ DARE } \textit{(14.1.2)}. \tag{14.5.30}$$

♦

Finally, since conditions (14.5.24) and (14.5.29) are equivalent under stabilizability, it follows that the result of Lemma 14.5.6 also holds for (14.5.29). That is, condition (14.5.29) results in positive-definite innovations variances $\{R_{e,i}\}$.

14.6 THE CASE OF STABLE SYSTEMS

We established earlier in Lemma 14.5.7 that, for $S = 0$, a detectable pair $\{F, H\}$ and a unit-circle controllable pair $\{F, GQ^{1/2}\}$, the condition (14.5.24), *viz.*,

$$I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2} > 0,$$

is equivalent[7] to the two combined conditions of a positive-definite sequence $\{R_{e,i}\}$ *and* a nonsingular matrix $I + (P_0 - P)\mathcal{O}^p$. Now introduce the semi-infinite matrix R_y that is defined by

$$R_y \triangleq LR_eL^*, \tag{14.6.1}$$

where L is given by

$$L \triangleq \begin{bmatrix} I & & & & \\ HK_{p,0} & I & & & \\ HFK_{p,0} & HK_{p,1} & I & & \\ HF^2K_{p,0} & HFK_{p,1} & HK_{p,2} & I & \\ \vdots & \vdots & \vdots & & \ddots \end{bmatrix}, \tag{14.6.2}$$

and $R_e = \text{diag}\{R_{e,0}, R_{e,1}, R_{e,2}, \ldots\}$. The $\{K_{p,i}, R_{e,i}\}$ are the quantities generated by the Kalman filter when initialized with initial condition P_0 — recall the discussion in Sec. 9.4, *viz.*,

$$K_{p,i} = FP_iH^*R_{e,i}^{-1}, \quad R_{e,i} = R + HP_iH^*.$$

We shall show in this section that when F is a *stable* matrix, condition (14.5.24) is also equivalent to the strong positivity of R_y. More specifically, when F is a stable matrix, both conditions of detectability and unit-circle controllability will be automatically satisfied and, in this case, we claim that the following statements will hold:

$$\{P_0 \text{ is such that } I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2} > 0\} \iff \{P_0 \text{ is such that } R_y > \epsilon I\}$$

or, equivalently,

$$\left\{P_0 \text{ is such that } \begin{pmatrix} R_{e,i} > \mu I \text{ and} \\ I + (P_0 - P)\mathcal{O}^p \text{ invertible} \end{pmatrix}\right\} \iff \{P_0 \text{ is such that } R_y > \epsilon I\}$$

for some $\mu > 0$ and $\epsilon > 0$. We shall say that R_y is strongly positive-definite.

Before establishing this result, let us remark that the above equivalence suggests that the (strong) positivity of the sequence $\{R_{e,i}\}$ by itself is not enough to guarantee a (strongly) positive-definite R_y. Both conditions are equivalent only when dealing with finite-horizon problems, *i.e.*,

$$\{R_{e,i} > 0, \quad 0 \leq i \leq N\} \iff R_y > 0,$$

[7] Recall that Lemma 14.5.7 guarantees the positivity of the $R_{e,i}$ for all finite i. But since under condition (14.5.24), P_i converges to $P \geq 0$, and since $R > 0$, we conclude that the limiting value R_e is also positive-definite. Hence, Lemma 14.5.7 actually implies that the resulting sequence $\{R_{e,i}\}$ is positive-definite for all i, including in the limit. This is equivalent to saying that the sequence $\{R_{e,i}\}$ is strongly positive-definite, meaning that there exists a positive number ν, such that $R_{e,i} > \nu I$ for all i.

when R_y is $(N+1)\times(N+1)$ and N is finite. This fact follows from the congruence relation $R_y = LR_eL^*$.

When, on the other hand, $N \to \infty$, the matrices R_e and R_y become semi-infinite. In this case, the (strong) positivity of the sequence $\{R_{e,i}\}$ is not enough to guarantee a (strongly) positive-definite R_y. All that can be concluded is that $R_y \geq 0$. This is because the semi-infinite lower triangular operator L can become singular despite its unit diagonal entries.[8] Consider, for example,

$$L = \begin{bmatrix} 1 & & & & \\ 2 & 1 & & & \\ & 2 & 1 & & \\ & & 2 & 1 & \\ & & & \ddots & \ddots \end{bmatrix}.$$

Its inverse is *unbounded* and given by

$$L^{-1} = \begin{bmatrix} 1 & & & & \\ -2 & 1 & & & \\ 4 & -2 & 1 & & \\ -8 & 4 & -2 & 1 & \\ & & & \ddots & \ddots \end{bmatrix},$$

so that L is singular. This means that there exists a vector b with arbitrarily small Euclidean norm such that $a = L^{-1}b$ is unbounded. This also means that there exists a vector a with arbitrarily large Euclidean norm such that $b = La$ is arbitrarily small. The following example illustrates this possibility in the context of the Riccati recursion.

EXAMPLE 14.6.1 ($\{R_{e,i} > 0\}$ Does Not Imply $R_y > 0$) Consider the model of Ex. 14.5.5, *viz.*,

$$\mathbf{x}_{i+1} = \frac{1}{2}\mathbf{x}_i, \quad \mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i,$$

with $r = 1$ and $q = 0$. Let $f = 1/2$. Choose $p_0 = -3/4$. Then $p_i = -3/4$ and $r_{e,i} = 1/4$ for all i. Thus the sequence $\{r_{e,i}\}$ is (strongly) positive. Let us now evaluate R_y. From the above model

[8] The semi-infinite matrix L can be regarded as a linear operator that takes a vector a into a vector b, say $b = La$. An operator is said to be *invertible* if $La = 0$ implies $a = 0$. By using the fact that L is lower triangular with unit-diagonal entries, we can easily verify that it is indeed invertible. We shall denote its inverse by L^{-1}. The inverse, however, can be a bounded or an unbounded operator. We shall say that L is *singular* if its inverse is unbounded. Thus, for operators, and in contrast to finite-dimensional matrices, we make a distinction between invertibility and singularity. Whenever a matrix is invertible, it is also nonsingular. Operators, on the other hand, can have inverses and be singular at the same time — see the example in the text.

we get

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ f \\ f^2 \\ \vdots \end{bmatrix} \mathbf{x}_0 + \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \end{bmatrix},$$

so that the semi-infinite matrix R_y is given by

$$R_y = p_0 \begin{bmatrix} 1 \\ f \\ f^2 \\ \vdots \end{bmatrix} \begin{bmatrix} 1 \\ f \\ f^2 \\ \vdots \end{bmatrix}^T + I.$$

The vector $\mathrm{col}\{1, f, f^2, \ldots\}$ is bounded since $f = 1/2$. Its squared Euclidean norm is equal to $1/(1-f^2)$, which evaluates to $4/3$. Hence, the nonunity eigenvalue of R_y is equal to $1 + 4p_0/3$, which is zero for our choice of p_0. This shows that R_y is singular for $p_0 = -3/4$. Note further that for this example, the quantity $1 + (p_0 - p)\mathcal{O}^p = 0$ (as shown in Ex. 14.5.5). ◆

Let us now establish the claim that $R_y > 0$ is equivalent to condition (14.5.24) when F is stable. For this purpose, we recall the time-invariant state-space model

$$\begin{cases} \mathbf{x}_{i+1} = F\mathbf{x}_i + G\mathbf{u}_i, & \mathbf{x}_0, \\ \mathbf{y}_i = H\mathbf{x}_i + \mathbf{v}_i, \end{cases}$$

where we continue to assume that $S = 0$. It follows that we can write

$$\underbrace{\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} H \\ HF \\ HF^2 \\ HF^3 \\ \vdots \end{bmatrix}}_{\mathcal{O}} \mathbf{x}_0 + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ HG & 0 & 0 & 0 \\ HFG & HG & 0 & 0 \\ HF^2G & HFG & HG & 0 \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{\Gamma} \underbrace{\begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \\ \vdots \end{bmatrix}}_{\mathbf{u}} + \underbrace{\begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \\ \vdots \end{bmatrix}}_{\mathbf{v}}$$

The operators $\mathcal{O}$ and Γ are bounded in view of the assumption of a stable matrix F. The semi-infinite matrix R_y in (14.6.1) can be verified to be equal to[9]

$$R_y = \mathcal{O}P_0\mathcal{O}^* + \Gamma\mathcal{Q}\Gamma^* + \mathcal{R}, \tag{14.6.3}$$

[9] When $P_0 \geq 0$, this expression for R_y follows by simply computing the covariance matrix of the output vector $\mathbf{y}$ defined above, since in this case $P_0 = \|\mathbf{x}_0\|^2$. The arguments of Sec. 9.4 (and in particular Eq. (9.4.1)) will then confirm that the matrices R_y in (14.6.1) and (14.6.3) are identical and correspond to the output covariance matrix. When P_0 is indefinite, however, P_0 cannot be regarded as the variance of the random variable $\mathbf{x}_0$ in the usual sense and, hence, expression (14.6.3) for R_y cannot be interpreted as the variance of the expression $\mathcal{O}\mathbf{x}_0 + \Gamma\mathbf{u} + \mathbf{v}$. Still, if we define R_y as in (14.6.3), then the algebraic argument of App. 9.A will show that it is this matrix whose triangular factorization (14.6.1) is computed by the Kalman filter recursions, so that the matrices R_y in (14.6.1) and (14.6.3) are again identical.

where we introduced the semi-infinite block-diagonal covariance matrices

$$\mathcal{Q} \triangleq \text{diag}(Q, Q, \ldots), \qquad \mathcal{R} \triangleq \text{diag}(R, R, \ldots).$$

Moreover, R_y is also a bounded operator.

Assume for now that we set $P_0 = P$, the stabilizing solution of the DARE. Then the covariance matrix of the resulting process $\{\mathbf{y}_i\}$ will be

$$\overline{R}_y = \mathcal{O}P\mathcal{O}^* + \Gamma\mathcal{Q}\Gamma^* + \mathcal{R}. \tag{14.6.4}$$

If we run the Kalman filter with this choice of P_0, the convergence of the resulting Riccati recursion is of course trivial since P_i will be equal to P for all i. Moreover, and in view of the discussion in Sec. 9.4 (Eq. (9.4.1)), this will lead to a triangular factorization for $\overline{R}_y$ of the form

$$\overline{R}_y = \overline{L}\,\overline{R}_e\,\overline{L}^*,$$

where $\overline{R}_e = \text{diag}\{R_e, R_e, \ldots\}$ has constant diagonal entries, and $\overline{L}$ is the lower triangular operator

$$\overline{L} \triangleq \begin{bmatrix} I & & & & \\ HK_p & I & & & \\ HFK_p & HK_p & I & & \\ HF^2K_p & HFK_p & HK_p & I & \\ \vdots & \vdots & \vdots & & \ddots \end{bmatrix}. \tag{14.6.5}$$

Here, $R_e = R + HPH^*$ and $K_p = FPH^*R_e^{-1}$. Note that the stability of F also guarantees that $\overline{L}$ is a bounded operator. In addition, the stability of the closed-loop matrix $F_p = F - K_pH$ guarantees that the inverse of $\overline{L}$ is bounded since (as suggested by (9.4.3)),

$$\overline{L}^{-1} = \begin{bmatrix} I & & & & \\ -HK_p & I & & & \\ -HF_pK_p & -HK_p & I & & \\ -HF_p^2K_p & -HF_pK_p & -HK_p & I & \\ \vdots & \vdots & \vdots & & \ddots \end{bmatrix}.$$

We have thus shown that the covariance matrix $\overline{R}_y$ admits a factorization of the form $\overline{L}\,\overline{R}_e\,\overline{L}^*$, with $\overline{L}$ bounded, invertible, and has a bounded inverse.

Returning to the expression (14.6.3) for R_y we can therefore write, by adding and subtracting $\mathcal{O}P\mathcal{O}^*$,

$$R_y = \overline{R}_y + \mathcal{O}(P_0 - P)\mathcal{O}^* = \overline{L}\,\overline{R}_e\,\overline{L}^* + \mathcal{O}(P_0 - P)\mathcal{O}^*.$$

Given this representation for R_y, we can now establish the earlier claim.

Lemma 14.6.1 (Positive Covariance R_y for Stable F) *Consider the earlier Riccati recursion (14.5.1) and assume that F is stable. Moreover, let P denote the unique stabilizing solution of the* DARE *(14.1.2), and let $\mathcal{O}^p \geq 0$ be the unique solution of the Lyapunov equation (14.3.5). Then the initial condition P_0 is a Hermitian matrix satisfying (14.5.24) if, and only if, $R_y > \epsilon I$ for some $\epsilon > 0$ (i.e., if, and only if, R_y is strongly positive-definite).* ■

Proof: First note that all the operators $\{\mathcal{O}, \overline{L}, \overline{L}^{-1}, \overline{R}_e\}$ are bounded. Now we write

$$R_y = \overline{L}\,\overline{R}_e^{1/2}\left[I + \overline{R}_e^{-1/2}\,\overline{L}^{-1}\mathcal{O}(P_0 - P)\mathcal{O}^*\overline{L}^{-*}\,\overline{R}_e^{-*/2}\right]\overline{R}_e^{*/2}\,\overline{L}^*,$$

which allows us to conclude that $R_y > \epsilon I$ if, and only if,

$$I + \overline{R}_e^{-1/2}\,\overline{L}^{-1}\mathcal{O}(P_0 - P)\mathcal{O}^*\overline{L}^{-*}\,\overline{R}_e^{-*/2} > \epsilon' I,$$

for some $\epsilon' > 0$. It is easy to verify by direct calculations, and using the expressions for $\mathcal{O}$ and $\overline{L}^{-1}$, that

$$\overline{L}^{-1}\,\mathcal{O} = \begin{bmatrix} H \\ HF_p \\ HF_p^2 \\ \vdots \end{bmatrix} \triangleq \mathcal{O}_c.$$

That is, $\overline{L}^{-1}\mathcal{O}$ is equal to the observability operator of the closed-loop system $\{F_p, H\}$. Therefore, the above is equivalent to

$$I + \overline{R}_e^{-1/2}\,\mathcal{O}_c(P_0 - P)\mathcal{O}_c^*\,\overline{R}_e^{-*/2} > \epsilon' I.$$

But the nonunity eigenvalues of the above operator are the same as the nonunity eigenvalues of the matrix

$$I + (P_0 - P)\mathcal{O}_c^*\,\overline{R}_e^{-1}\mathcal{O}_c.$$

Moreover, it is immediate to verify that $\mathcal{O}^p = \mathcal{O}_c^*\overline{R}_e^{-1}\mathcal{O}_c$, so that the nonunity eigenvalues of the matrix $I + (P_0 - P)\mathcal{O}^p$ must be positive. These eigenvalues are the same as the nonunity eigenvalues of the matrix

$$I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2},$$

which is the matrix appearing in condition (14.5.24). ◆

The above result suggests that, for stable matrices F, condition (14.5.24) is more a property of the process $\{\mathbf{y}_i\}$ itself, rather than of the state-space matrices $\{F, G, H, Q, R, \Pi_0\}$ used to model it.

A natural question is whether the equivalence result established in the above lemma still holds for unstable F? The answer is negative as shown by the following example.

EXAMPLE 14.6.2 ($R_y > 0$ Does Not Imply (14.5.24) for Unstable F) Consider the unstable state-space model

$$\mathbf{x}_{i+1} = 2\mathbf{x}_i, \quad \mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i,$$

with $r = 1$ and $q = 0$. The system is thus detectable and unit-circle controllable. The Riccati recursion is given by

$$p_{i+1} = 4p_i - \frac{4p_i^2}{1 + p_i}, \quad p_0 = \text{initial condition},$$

with the resulting DARE being equivalent to the equation $p^2 - 3p = 0$. The stabilizing solution in this case can be seen to be $p = 3$, which yields $r_e = 4$, $k_p = 3/2$, and $f_p = 1/2$. Then $O^p = 1/3$ and condition (14.5.24) require that p_0 be such that $1 + 1/3(p_0 - 3) > 0$ or, equivalently, $p_0 > 0$.

Now the covariance matrix R_y can be seen to be unbounded and given by

$$R_y = I + p_0 \mathcal{O}\mathcal{O}^*,$$

where $\mathcal{O} = \text{col}\{1, f, f^2, f^3, \ldots\}$. Note in particular that $p_0 = 0$ still guarantees $R_y > 0$, while this choice for p_0 is ruled out by condition (14.5.24) since it requires $p_0 > 0$. ◆

The following lemma provides the exact equivalence conditions for general matrices F (and, in particular, for unstable matrices F).

Lemma 14.6.2 (Positive Covariance R_y for General F) *Consider the earlier Riccati recursion (14.5.1) and assume that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is unit-circle controllable. Moreover, let P denote the unique stabilizing solution of the* DARE *(14.1.2), and let $\mathcal{O}^p \geq 0$ be the unique solution of the Lyapunov equation (14.3.5). Then the initial condition P_0 is a Hermitian matrix satisfying (14.5.24) if, and only if, the following two conditions hold:*

(i) $R_y > \epsilon I$, for some $\epsilon > 0$, and

(ii) The matrix $I + (P_0 - P)\mathcal{O}^p$ is nonsingular.

■

Proof: Assume first that (i) and (ii) above hold. From (i) we conclude that $R_{e,i} > \nu I$ for all time instants i and for some $\nu > 0$. Therefore, conditions (i) and (ii) of Lemma 14.5.7 hold and P_0 must be such that (14.5.24) is satisfied.

Conversely, assume P_0 is such that condition (14.5.24) holds. Then, by part (i) of Lemma 14.5.7, we conclude that $R_{e,i} > 0$ for all i, and by part (ii) of the same lemma we conclude that $I + (P_0 - P)\mathcal{O}^p$ is nonsingular (which is condition (ii) above). Repeating the argument in the first footnote of this section (just prior to Eq. (14.6.1)) we conclude that it must hold that $R_{e,i} > \nu I$ for all time instants i and for some $\nu > 0$. Let us now show that $R_y > \epsilon I$ for some $\epsilon > 0$.

Indeed, from Thm. 14.5.2, we have that P_i converges to P for such choices of P_0. This means that the Kalman recursions will lead to a factorization for R_y of the form (14.6.1)–(14.6.2).

The matrix L is generally unbounded (since F can be unstable). However, it is invertible and its inverse is given by (as suggested by (9.4.3))

$$L^{-1} = \begin{bmatrix} I & & & \\ -HK_{p,0} & I & & \\ -HF_pK_{p,0} & -HK_{p,1} & I & \\ -HF_p^2K_{p,0} & -HF_pK_{p,1} & -HK_{p,2} & I \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Since F_p is stable, and since $\{K_{p,i}\}$ is a convergent sequence (and, hence, bounded), we conclude that L^{-1} is in fact bounded. This means that L itself is nonsingular. It then follows from $R_y = LR_eL^*$ that $R_y > \epsilon I$ for some $\epsilon > 0$. ◆

Note that the example of an unstable model prior to the statement of the lemma is such that condition (i) is satisfied ($R_y > \epsilon I$) while condition (ii) is violated ($1 + (p_0 - p)\mathcal{O}^p = 0$).

Remark 8. Comparing the statements of Lemmas 14.6.1 and 14.6.2, one might be led to conclude that the assumption of a stable matrix F in Lemma 14.6.1 guarantees that condition (ii) of Lemma 14.6.2 is automatically satisfied so that it is not required in Lemma 14.6.1. This is *not* correct, simply because condition (ii) in the above lemma depends on P_0 and, even for a stable F, one can choose P_0 arbitrarily to violate it. It is the combination of both conditions, a stable F and a strongly positive-definite R_y, that removes the need for condition (ii) above in the stable case. ◆

A Global Approach to Convergence. In any case, the above discussion shows that it is also possible to study the convergence of the Riccati variable, P_i, using a global approach, at least for stable systems F so that several required semi-infinite matrices (or operators) can be guaranteed to be bounded. In this approach, instead of explicitly using the Riccati recursion (14.1.4), as we have done so far in this chapter, we use the definition of P_i as the prediction error Gramian itself, *i.e.*,

$$P_{i+1} = \langle \mathbf{x}_{i+1}, \mathbf{x}_{i+1} \rangle - \langle \mathbf{x}_{i+1}, \mathbf{y}^{(i)} \rangle \langle \mathbf{y}^{(i)}, \mathbf{y}^{(i)} \rangle^{-1} \langle \mathbf{y}^{(i)}, \mathbf{x}_{i+1} \rangle, \tag{14.6.6}$$

where $\mathbf{y}^{(i)} \triangleq \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}$. The state-space model in global form is

$$\begin{cases} \mathbf{x}_{i+1} = \Phi^{(i)}\mathbf{x}_0 + \mathcal{C}^{(i)}\mathbf{u}^{(i)}, \\ \mathbf{y}^{(i)} = \mathcal{O}^{(i)}\mathbf{x}_0 + \Gamma^{(i)}\mathbf{u}^{(i)} + \mathbf{v}^{(i)}, \end{cases}$$

where $\mathbf{u}^{(i)} = \text{col}\{\mathbf{u}_0, \ldots, \mathbf{u}_i\}$ and $\mathbf{v}^{(i)} = \text{col}\{\mathbf{v}_0, \ldots, \mathbf{v}_i\}$, with

$$\left\langle \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}^{(i)} \\ \mathbf{v}^{(i)} \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}^{(i)} \\ \mathbf{v}^{(i)} \end{bmatrix} \right\rangle = \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & \mathcal{Q}^{(i)} & 0 \\ 0 & 0 & \mathcal{R}^{(i)} \end{bmatrix}, \quad \begin{aligned} \mathcal{Q}^{(i)} &= \text{diag}(Q, \ldots, Q), \\ \mathcal{R}^{(i)} &= \text{diag}(R, \ldots, R). \end{aligned}$$

Also $\Phi^{(i)} = F^i$ is the state transition matrix, while $\mathcal{O}^{(i)}$ and $\mathcal{C}^{(i)}$ are the observability and controllability maps,

$$\mathcal{O}^{(i)} = \text{col}\{H, HF, \ldots, HF^{i-1}\} \quad \text{and} \quad \mathcal{C}^{(i)} = \begin{bmatrix} F^{i-1}G & F^{i-2}G & \ldots & G \end{bmatrix},$$

and

$$\Gamma^{(i)} = \begin{bmatrix} 0 & & & \\ HG & 0 & & \\ HFG & HG & 0 & \\ \vdots & \vdots & \vdots & \ddots \\ HF^{i-1}G & HF^{i-2}G & \ldots & HG \quad 0 \end{bmatrix}$$

is the impulse response matrix. In this setting, it is easy to show, for example, that

$$P_{i+1} = \begin{bmatrix} \Phi^{(i)} & \mathcal{C}^{(i)} \end{bmatrix} \left(\begin{bmatrix} \Pi_0^{-1} & 0 \\ 0 & \mathcal{Q}^{-(i)} \end{bmatrix} + \begin{bmatrix} \mathcal{O}^{*(i)} \\ \Gamma^{*(i)} \end{bmatrix} \mathcal{R}^{-(i)} \begin{bmatrix} \mathcal{O}^{(i)} & \Gamma^{(i)} \end{bmatrix} \right)^{-1} \begin{bmatrix} \Phi^{*(i)} \\ \mathcal{C}^{*(i)} \end{bmatrix}.$$

This expression can be used to establish the global identities of Lemma 14.4.2, and thereby to establish the convergence of P_i, without explicitly resorting to the Riccati recursion, though for reasons of space we shall not present the details here.

14.7 THE CASE OF $S \neq 0$

We assumed throughout the chapter that $S = 0$. However, as mentioned several times already (especially in Sec. 9.5.1), when $R > 0$, the results for the case $S \neq 0$ can be deduced from what we have done so far in the chapter by means of the following transformations:

$$F \rightarrow F^s \triangleq F - GSR^{-1}H, \quad \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \rightarrow \begin{bmatrix} Q^s \triangleq Q - SR^{-1}S^* & 0 \\ 0 & R \end{bmatrix}.$$

With these transformations the Riccati recursion with $S \neq 0$,

$$P_{i+1} = FP_iF^* + GQG^* - (FP_iH^* + GS)(R + HP_iH^*)^{-1}(FP_iH^* + GS)^*, \quad (14.7.1)$$

is transformed to one with $S = 0$,

$$P_{i+1} = F^sP_iF^{s*} + GQ^sG^* - F^sP_iH^*(R + HP_iH^*)^{-1}HP_iF^{s*}. \quad (14.7.2)$$

In this case, the DARE and the dual DARE become

$$P = F^sPF^{s*} + GQ^sG^* - F^sPH^*(R + HPH^*)^{-1}HPF^{s*}, \quad (14.7.3)$$

and

$$P^a = F^{s*}P^aF^s + H^*R^{-1}H - F^{*s}P^aGQ^{s/2}(I + Q^{s*/2}G^*P^aGQ^{s/2})^{-1}Q^{s*/2}G^*P^aF^s. \quad (14.7.4)$$

Moreover, F_p becomes

$$F_p = F - K_pH = F^s - K_p^sH, \qquad K_p^s = F^sPH^*R_e^{-1}.$$

This implies that all our convergence results for $S \neq 0$ can be obtained by using the new system matrices $\{F^s, G, H, Q^s, R\}$. Thus the conditions of detectable $\{F, H\}$, unit-circle controllable $\{F, GQ^{1/2}\}$, and stabilizable $\{F, GQ^{1/2}\}$ that were needed in various statements would have to be replaced by the detectability of $\{F^s, H\}$, unit-circle controllability of $\{F^s, GQ^{s/2}\}$, and stabilizability of $\{F^s, GQ^{s/2}\}$, respectively.[10] For example, Thms. 14.5.2 and 14.5.3 would become the following.

Theorem 14.7.1 (A Sufficient Convergence Condition) *Consider the Riccati recursion (14.7.1) with $\{F, H\}$ detectable and $\{F^s, GQ^{s/2}\}$ controllable on the unit circle. Moreover, let P denote the unique stabilizing solution of the* DARE *(14.7.3), and let $\mathcal{O}^p \geq 0$ be the unique solution of the Lyapunov equation (14.3.5). Also consider the square-root factorization (14.5.14) of $\mathcal{O}^p$. Then, if the initial condition P_0 is a Hermitian matrix satisfying*

$$I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2} > 0,$$

P_i converges exponentially to P. ■

Theorem 14.7.2 (Convergence with Indefinite P_0) *Consider the Riccati recursion (14.7.1) where $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is stabilizable. Suppose, moreover, that the initial condition P_0 is a Hermitian matrix such that*

$$I + (P^a)^{*/2}P_0(P^a)^{1/2} > 0,$$

where $P^a = (P^a)^{1/2}(P^a)^{*/2}$ is the unique stabilizing solution to the dual DARE *(14.7.4). Then P_i converges to the unique stabilizing solution, P, of the* DARE *(14.7.3).* ■

Likewise, the equivalence result of Lemma 14.6.1 would still hold for stable systems F except that now the assumption of a unit-circle controllable pair $\{F^s, GQ^{s/2}\}$ is not automatically satisfied and therefore needs to be incorporated into the statement of the lemma.

Lemma 14.7.1 (Positive Covariance R_y for Stable F) *Consider the Riccati recursion (14.7.1) and assume that F is stable. Assume further that $\{F^s, GQ^{s/2}\}$ is unit-circle controllable so that the unique stabilizing solution of the* DARE *(14.7.3) exists. Let $\mathcal{O}^p \geq 0$ be the unique solution of the Lyapunov equation (14.3.5). Then the initial condition P_0 is a Hermitian matrix satisfying (14.5.24) if, and only if, $R_y > \epsilon I$ for some $\epsilon > 0$.* ■

Proof: Expression (14.6.3) for R_y now becomes

$$R_y = \mathcal{O}P_0\mathcal{O}^* + \Gamma\mathcal{Q}\Gamma^* + \Gamma\mathcal{S} + \mathcal{S}^*\Gamma^* + \mathcal{R},$$

where we introduced the semi-infinite block-diagonal matrix $\mathcal{S} \triangleq \text{diag}(S, S, \ldots)$. Following the arguments in the proof of Lemma 14.6.1 we can again establish that the alternative covariance matrix

$$\overline{R}_y = \mathcal{O}P\mathcal{O}^* + \Gamma\mathcal{Q}\Gamma^* + \Gamma\mathcal{S} + \mathcal{S}^*\Gamma^* + \mathcal{R},$$

[10] Observe, however, that the detectability of $\{F, H\}$ is equivalent to the detectability of $\{F^s, H\}$.

corresponding to $P_0 = P$, can be factored as $\overline{R}_y = \overline{L}\,\overline{R}_e\,\overline{L}^*$, with $\overline{L}$ bounded, invertible, and has a bounded inverse. The result now follows as in Lemma 14.6.1. ♦

14.8 EXPONENTIAL CONVERGENCE OF THE FAST RECURSIONS

In the earlier sections, we studied the asymptotic behavior of the Kalman filter and discussed conditions for the convergence of the Riccati variable P_i in the time-invariant case. In order to illustrate that the convergence of P_i is a property of the Kalman filter itself, and not of the method of implementing the filter, we now establish the exponential convergence of the fast recursions of Thm. 11.1.2. Let

$$F^s \triangleq F - GSR^{-1}H, \quad Q^s \triangleq Q - SR^{-1}S^*,$$

and consider the DARE

$$P = F^s P F^{s*} + GQ^s G^* - K_p R_e K_p^*, \tag{14.8.1}$$

with $K_p^s = F^s P H^* R_e^{-1}$ and $R_e = R + HPH^*$. [Recall from (9.5.12) that $K_p = (FPH^* + GS)R_e^{-1}$ is also equal to $K_p = K_p^s + GSR^{-1}$.]

Theorem 14.8.1 (Convergence of the Fast Recursions) *Consider the fast recursions (11.1.14)–(11.1.19) with $R > 0$, $\{F, H\}$ detectable, and $\{F^s, GQ^{s/2}\}$ unit-circle controllable. Assume further that P_0 is chosen according to any of the following conditions:*

- **(a)** *P_0 is such that the sequence of matrices $\{T_i\}$ in (14.5.2) is uniformly bounded (cf. (14.3.4)).*
- **(b)** *P_0 is such that $I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2} > 0$, where $\mathcal{O}^{p/2}$ is a square-root factor of the unique solution to the Lyapunov equation*

$$\mathcal{O}^p = F_p^* \mathcal{O}^p F_p + H^* R_e^{-1} H,$$

with $F_p = F^s - K_p^s H = F - K_p H$.

- **(c)** *$\{F^s, GQ^{s/2}\}$ is stabilizable and P_0 is such that*

$$I + (P^a)^{*/2} P_0 (P^a)^{1/2} > 0, \tag{14.8.2}$$

where P^a is the unique positive-semi-definite solution to the dual Riccati equation (14.7.4).

Then for any such P_0, the matrices K_i and L_i in (11.1.14) and (11.1.15) converge exponentially to $FPH^* + GS$ and zero, respectively. Moreover, the sequences $\{R_{r,i}\}$ and $\{R_{r,i}^{-1}\}$ will be uniformly bounded. ■

Proof: We first note that since the fast recursions (11.1.14)–(11.1.19) are just a reorganization of the usual Riccati recursions, the Riccati variable P_i will still converge for all initial conditions specified by (a)–(c) above (*cf.* Sec. 14.7 and Thms. 14.5.1, 14.5.2, and 14.5.3). Therefore, the difference

$$P_{i+1} - P_i = -L_i R_{r,i}^{-1} L_i^*$$

will converge to zero. However, this does not necessarily mean that L_i and/or $R_{r,i}^{-1}$ converge or are even bounded. In what follows we shall prove that L_i converges exponentially to zero and that $R_{r,i}$ and $R_{r,i}^{-1}$ remain uniformly bounded for all i. This will show that in the recursions (11.1.14)–(11.1.19) all the variables will remain bounded. We also remark that, for the same reason that P_i converges, the gain vector $K_i = FP_iH^* + GS$ converges as well.

Now using the equation for L_i we can write $L_i = \Phi_p(i, 0)L_0$, where $\Phi_p(i, 0)$ is the state-transition matrix associated with $F_{p,i} = (F - K_{p,i}H)$, *i.e.*,

$$\Phi_p(i, 0) = F_{p,i-1} \ldots F_{p,1} F_{p,0}, \qquad \Phi_p(0, 0) = I.$$

The result of part (c) in Prob. 14.8 guarantees that for any P_0 chosen according to (a)–(c) above, $\Phi_p(i, 0) \to 0$ exponentially fast and, consequently, that L_i converges to zero also exponentially.

It remains to show that the sequences $\{R_{r,i}\}$ and $\{R_{r,i}^{-1}\}$ are uniformly bounded. To establish the uniform boundedness of $\{R_{r,i}\}$ we use (11.1.17) to write

$$R_{r,i+1} = R_{r,0} - \sum_{j=0}^{i} L_j^* H^* R_{e,j}^{-1} H L_j,$$

and, hence,

$$R_{r,i+1} = R_{r,0} - \sum_{j=0}^{i} L_0^* X_j^* F_p^{j*} H^* R_{e,j}^{-1} H F_p^j X_j L_0, \tag{14.8.3}$$

where we introduced

$$X_j \triangleq [I + (P_0 - P)\mathcal{O}_j^p]^{-1}.$$

The closed-loop matrix F_p is stable and, hence, its spectral radius is strictly less than one. It then follows from the result of Prob. 14.19 that there exists a matrix norm, denoted by $\|\cdot\|_\rho$, such that $\|F_p\|_\rho = \beta < 1$, for some β (simply choose $\beta = \rho(F_p) + \epsilon$ for small enough ϵ). Computing the norm of $R_{r,i+1}$ in (14.8.3) then leads to

$$\|R_{r,i+1}\|_\rho \leq \|R_{r,0}\|_\rho + \sum_{j=0}^{i} \beta^{2j} \|L_0\|_\rho^2 \, \|X_j\|_\rho^2 \, \|H\|_\rho^2 \, \|R_{e,j}^{-1}\|_\rho. \tag{14.8.4}$$

But we know from Prob. 14.3 that the sequence $\{X_i\}$ is uniformly bounded for any initial condition P_0 that is chosen according to (a)–(c) in the statement of the theorem above. We also know from Lemma 14.5.2 that the $\{R_{e,i}\}$ are nonsingular for all i and, consequently, the sequence $\{R_{e,i}^{-1}\}$ is also uniformly bounded. It thus follows that

$$\|L_0\|_\rho^2 \, \|X_j\|_\rho^2 \, \|H\|_\rho^2 \, \|R_{e,j}^{+1}\|_\rho \leq \gamma < \infty,$$

for some γ and for all j. Using this result in (14.8.4) gives

$$\|R_{r,i+1}\|_\rho \leq \|R_{r,0}\|_\rho + \sum_{j=0}^{i} \beta^{2j}\gamma = \|R_{r,0}\|_\rho + \frac{\gamma}{1 - \beta^2} < \infty,$$

which means that the sequence $\{R_{r,i}\}$ is uniformly bounded.

We now establish the uniform boundedness of the sequence $\{R_{r,i}^{-1}\}$ or, equivalently, the nonsingularity of $R_{r,i}$ for all i, including in the limit as $i \to \infty$.

First recall that if P_0 is chosen according to (a) in the statement of the theorem, then the resulting $\{R_{e,i}\}$ are nonsingular for all i, including in the limit (see Lemma 14.5.2 and the fact that under (a) $R_{e,i}$ converges to $R_e > 0$). Likewise, if P_0 is chosen according to (b) and (c), then the resulting $\{R_{e,i}\}$ are not only nonsingular but also positive-definite for all i, including in the limit (see Lemma 14.5.6 and the remark after Cor. 14.5.1). Now using the decompositions (11.1.22)–(11.1.23) we have that for any finite i,

$$\det R_{r,i+1} = (\det R_{r,i})\,(\det R_{e,i+1}/\det R_{e,i}),$$

and, consequently,

$$\det R_{r,i} = \det R_{r,0}\,\frac{\det R_{e,i}}{\det R_{e,0}}.$$

Now since in all cases (a)–(c), the $R_{e,i}$ are nonsingular, and since also $\det R_{r,0} \neq 0$, we conclude that $\det R_{r,i} \neq 0$ for all finite i. We still need to establish the nonsingularity of $R_{r,i}$ in the limit, as $i \to \infty$. This step requires more effort.

First, in part (b) of Prob. 11.2 we establish the identity

$$R_{r,i+1} = R_{r,0} - L_0^* \mathcal{O}_i L_0,$$

where the observability Gramian $\mathcal{O}_i$ is given by

$$\mathcal{O}_i \triangleq \sum_{j=0}^{i} \Phi_p^*(j,0) H^* R_{e,j}^{-1} H \Phi_p(j,0), \qquad \mathcal{O}_{-1} = 0.$$

Moreover, in Prob. 14.11 we show that $\mathcal{O}_i$ can be expressed in terms of $\mathcal{O}_i^p$ as follows

$$\mathcal{O}_i = \mathcal{O}_i^p \left[I + (P_0 - P)\mathcal{O}_i^p\right]^{-1}. \tag{14.8.5}$$

Here, $\mathcal{O}_i^p$ is the observability Gramian that is obtained via the recursion

$$\mathcal{O}_{i+1}^p = F_p^* \mathcal{O}_i^p F_p + H^* R_e^{-1} H, \qquad \mathcal{O}_{-1}^p = 0.$$

That is,

$$\mathcal{O}_i^p \triangleq \sum_{j=0}^{i} F_p^{j*} H^* R_e^{-1} H F_p^j, \qquad \mathcal{O}_{-1}^p = 0.$$

The matrices $\{I + (P_0 - P)\mathcal{O}_i^p\}$ that appear in (14.8.5) are all invertible for any initial condition that is chosen according to (a)–(c) in the statement of the theorem (*cf.* the results of Ch. 14. In particular, choices (b) and (c) lead, by Lemma 14.5.6 and the remark after Cor. 14.5.1, to a sequence $\{R_{e,i}\}$ and to a matrix $I + (P_0 - P)\mathcal{O}^p$ that satisfy both conditions of Lemma 14.5.2. The same conclusion holds for choice (a)).

Now since F_p is stable, $\mathcal{O}_i^p$ tends to the unique nonnegative-definite solution of the Lyapunov equation $\mathcal{O}^p = F_p^* \mathcal{O}^p F_p + H^* R_e^{-1} H$. Therefore, (14.8.5) implies that $\mathcal{O}_i$ also converges to

$$\lim_{i\to\infty} \mathcal{O}_i = \mathcal{O} \triangleq \mathcal{O}^p \left[I + (P_0 - P)\mathcal{O}^p\right]^{-1}. \tag{14.8.6}$$

Consequently, $R_{r,i}$ converges to $R_r = R_{r,0} - L_0^* \mathcal{O} L_0$. We thus need to establish the nonsingularity of this limit matrix. Recall that $R_{r,0}$ is invertible so that $R_r R_{r,0}^{-1} = I - L_0^* \mathcal{O} L_0 R_{r,0}^{-1}$. Using $\det(I + AB) = \det(I + BA)$, we see that we can equivalently check for the nonsingularity of $I + (P_1 - P_0)\mathcal{O}$. Substituting expression (14.8.6) for $\mathcal{O}$, this last matrix becomes

$$I + (P_1 - P_0)\mathcal{O}^p \left[I + (P_0 - P)\mathcal{O}^p\right]^{-1},$$

which is indeed invertible since, in view of the matrix inversion lemma, its inverse is given by

$$I - (P_1 - P_0)\mathcal{O}^p \left[I + (P_1 - P)\mathcal{O}^p\right]^{-1},$$

and the matrix $I + (P_1 - P)\mathcal{O}^p$ is itself invertible by the result of Prob. 14.14. ◆

14.9 COMPLEMENTS

The study of the asymptotic behavior of the Riccati variable, P_i, is clearly quite challenging, especially when F is unstable. [The special case F stable, $S = 0$, and $\Pi_0 = \bar{\Pi}$ is more straightforward. Now the signal process is stationary, and we may only mention here that several results on the asymptotic behavior of finite-time estimators of stationary processes are available in the mathematical literature on orthogonal polynomials, moment problems, Wiener-Hopf equations; some, of many, references are Szegö (1939), Grenander and Szegö (1958), Geronimus (1961), and Gohberg and Fel'dman (1974).] In the state-space context, the problem was perhaps first addressed by Kalman (1960c) for the Riccati differential equation arising in the quadratic regulator problem, where furthermore the case of F unstable is of major interest (see our discussion in Sec. 14.1.4). These results were then translated by duality to the estimation problem (see Kalman and Bucy (1961) and also Kalman (1963a,1963b)). Not surprisingly, Lyapunov function techniques were used to study stability, but Kalman was the first to note the importance of controllability and observability assumptions (when F is unstable). Kalman's analysis was also for nonnegative-definite initial conditions, $\Pi_0 \geq 0$, and for time-variant systems. These results were followed by several more direct analyses, weakeningm the conditions in various ways. A noncomprehensive list includes Wonham (1968b), Kučera (1972), Bucy and Rodriguez-Canabal (1972), Rodriguez-Canabal (1973), Ljung and Kailath (1976c), several surveys in Bittanti, Laub, and Willems (1991), Byrnes, Lindquist, and McGregor (1991), and most recently Callier, Winkin, and Willems (1994). In fact, while several results in this chapter appear to be new in the discrete-time case (the continuous-time results are simpler — see Sec. 16.7), they are most closely related in form to those in the last reference. Our approach is quite different, and extends to more general Riccati recursions, *e.g.*, those encountered in $\mathcal{H}_\infty$ theory (see Hassibi, Sayed, and Kailath (1999)). It is important to note that our convergence results show that $\{K_{p,i}, R_{e,i}\}$ converge as well, and that $F_p = F - K_p H$ is stable. Therefore our results also extend to the array algorithms and the fast algorithms of Chs. 11–13. [Of course, different implementations may vary in terms of numerical behavior. However, since they all propagate the variable P_i, albeit in different guises, they will all exhibit exponential convergence.]

PROBLEMS

14.1 **(A sequence of matrices)** Consider the sequence of matrices $\{T_i\}$ defined by (14.3.3), where $(P_0 - P)$ is Hermitian but not necessarily invertible or even nonnegative definite.

(a) Show that each T_i is Hermitian.

(b) Now assume $P_0 - P \geq 0$. Show that each T_i is also nonnegative-definite.

14.2 **(Boundedness of the $\{D_i\}$)** Refer to the equality $T_i = A_0 D_i^{-1} A_0^*$ in (14.5.8), where A_0 is $n \times r$ ($r \leq n$) and has full rank. We want to establish that the sequence $\{T_i\}$ is uniformly bounded (*cf.* (14.3.4)) if, and only if, the sequence $\{D_i^{-1}\}$ is also uniformly bounded. One direction is immediate since the boundedness of D_i^{-1} clearly implies he boundedness of T_i. To establish the converse, introduce the SVD of A_0 and show that $D_i^{-1} = A_0^{\dagger} T_i \left(A_0^{\dagger}\right)^*$, where $A_0^{\dagger}$ denotes the pseudo-inverse of A_0 (*cf.* Eq. A.4.3 in App. A). Conclude that the sequence $\{D_i^{-1}\}$ is uniformly bounded.

14.3 **(Boundedness of the $\{X_i\}$)** Consider the same setting of Lemma 14.5.1. Show that the sequence of Hermitian matrices $\{T_i\}$ defined by (14.5.2) is uniformly bounded (*cf.* (14.3.4)) if, and only if, the sequence of matrices $\{X_i\}$ defined by (14.5.4) is uniformly bounded. [*Hint.* Show that the matrices $\{D_i\}$ in (14.5.7) are nonsingular for all i if, and only if, the matrices $\{I + (P_0 - P)\mathcal{O}_i^p\}$ are nonsingular for all i.]

14.4 **(Boundedness of P_i for arbitrary P_0)** Consider the Riccati recursion (14.1.4) with arbitrary positive-semi-definite initial condition,

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0 = \Pi_0 \geq 0,$$

and suppose that $\{F, H\}$ is detectable.

(a) Consider a gain matrix K such that $F - KH$ is stable. Prove that the sequence $\{P_i^e\}$ converges, where

$$P_{i+1}^e = (F - KH)P_i^e(F - KH)^* + \begin{bmatrix} G & -K \end{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} G^* \\ -K^* \end{bmatrix}, \quad P_0^e = \Pi_0.$$

(b) Show that $\{P_i\}$ is bounded from above by the convergent sequence $\{P_i^e\}$, *i.e.*, $P_i \leq P_i^e$ for all i.

(c) While the result of part (b) can be established algebraically, a stochastic interpretation can be obtained as follows. Introduce the (suboptimal) state estimator $\hat{\mathbf{x}}_{i+1}^e = F\hat{\mathbf{x}}_i^e + K[\mathbf{y}_i - H\hat{\mathbf{x}}_i^e]$, and the corresponding state estimation error equation. Verify that P_i^e is the error covariance matrix of the suboptimal estimator, *i.e.*, $P_i^e = \|\tilde{\mathbf{x}}_i^e\|^2$. Conclude, by suboptimality, the result of part (b).

14.5 **(Some increment relations)** For any $n \times n$ matrices P and P_i, define the quantities

$$R_{e,i} = R + HP_iH^*, \quad R_e = R + HPH^*, \quad K_{p,i} = K_iR_{e,i}^{-1}, \quad K_p = KR_e^{-1},$$

where $K_i = FP_iH^* + GS$ and $K = FPH^* + GS$. Define also $F_{p,i} = F - K_{p,i}H$ and $F_p = F - K_pH$, and introduce the difference matrix $\Delta P_i = P_{i+1} - P$.

(a) Show that

$$R_{e,i+1} - R_e = H\Delta P_iH^*,$$

$$K_{i+1} - K = F\Delta P_iH^*,$$

$$K_{p,i+1} - K_p = F_p\Delta P_iH^*R_{e,i+1}^{-1},$$

$$F_{p,i+1} = F_p(I - \Delta P_iH^*R_{e,i+1}^{-1}H) = F_p(I + \Delta P_iH^*R_e^{-1}H)^{-1}.$$

(b) Now assume that P is any solution of the DARE (14.1.2) and that P_i satisfies the Riccati recursion (14.1.4). Show that

$$\begin{aligned} \Delta P_{i+1} &= F_p[\Delta P_i - \Delta P_iH^*R_{e,i+1}^{-1}H\Delta P_i]F_p^*, \\ &= F_p[I + \Delta P_iH^*R_e^{-1}H]^{-1}\Delta P_iF_p^*, \\ &= F_{p,i+1}\Delta P_iF_p^*. \end{aligned}$$

14.6 **(A local identity)** Suppose $P_i^{(1)}$ and $P_i^{(2)}$ are two solutions to the Riccati recursion (14.1.4) with the *same* $\{F, G, H\}$ and $\{Q, R, S\}$, but with *different* initial conditions $\Pi_0^{(1)}$ and $\Pi_0^{(2)}$, respectively. Let $\delta P_i = P_i^{(2)} - P_i^{(1)}$. Introduce further the corresponding quantities

$$\left\{K_i^{(1)}, R_{e,i}^{(1)}, F_{p,i}^{(1)}\right\} \quad \text{and} \quad \left\{K_i^{(2)}, R_{e,i}^{(2)}, F_{p,i}^{(2)}\right\}.$$

Following the arguments of Prob. 14.5, evaluate the increments

$$R_{e,i}^{(2)} - R_{e,i}^{(1)}, \quad K_{i+1}^{(2)} - K_i^{(1)}, \quad K_{p,i}^{(2)} - K_{p,i}^{(1)}, \quad \text{and} \quad F_{p,i}^{(2)} - F_{p,i}^{(1)}.$$

Then show that

$$\delta P_{i+1} = F_{p,i}^{(1)}\left[\delta P_i - \delta P_iH^*(R_{e,i}^{(2)})^{-1}H\delta P_i\right]F_{p,i}^{(1)*}.$$

14.7 **(Inertia of $R_{r,i}$)** Assume $\{F, H\}$ is detectable, $\{F, GQ^{1/2}\}$ is unit-circle controllable, and that P_0 is chosen such that the sequence of matrices $\{T_i\}$ in (14.5.2) is uniformly bounded (*cf.* (14.3.4)). Let β denote the number of negative eigenvalues of $P_0 - P$, where P is the unique stabilizing solution of the DARE (14.8.1).

(a) Use the decompositions (11.1.22)–(11.1.23), and the result of Lemma 14.5.3, to show that the $\{R_{r,i}\}$ have constant inertia equal to that of $R_{r,0}$ for all time instants i, except for at most β finite time instants where the inertia of $R_{r,i}$ can 0be different from that of $R_{r,0}$.

(b) Show further than when P_0 is chosen according to (b) and (c) in the statement of Thm. 14.8.1, then the inertia of $R_{r,i}$ is constant for all i.

14.8 **(Limit of $\Phi_p(i, 0)$)** Let $\Phi_p(i, 0)$ be the state transition matrix that is defined by

$$\Phi_p(i + 1, 0) = [F - K_{p,i}H]\Phi_p(i, 0), \quad \Phi_p(0, 0) = I,$$

where $K_{p,i} = FP_iH^*R_{e,i}^{-1}$, $R_{e,i} = R + HP_iH^*$, and P_i is obtained from the Riccati difference equation

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0.$$

Assume further that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is unit-circle controllable.

(a) Show, by induction, that the following relation always holds:

$$\Phi_p(i,0) = F_p^i[I + (P_0 - P)\mathcal{O}_{i-1}^p]^{-1},$$

where P is the unique stabilizing solution of the DARE (14.1.2) and F_p is the corresponding closed-loop matrix. Also, $\mathcal{O}_{i-1}^p$ is given by (14.5.11), *viz.*,

$$\mathcal{O}_{i-1}^p = \sum_{j=0}^{i-1} F_p^{j*} H^* R_e^{-1} H F_p^j, \qquad \mathcal{O}_{-1}^p = 0.$$

[*Hint.* One proof could be obtained as follows. Assume the result holds for $i-1$ and, hence,

$$\Phi_p(i,0) = F_{p,i-1}\Phi_p(i-1,0) = F_{p,i-1}F_p^{i-1}\left[I + (P_0 - P)\mathcal{O}_{i-2}^p\right].$$

Let us show that the right-hand side expression evaluates to $F_p^i[I+(P_0-P)\mathcal{O}_{i-1}^p]^{-1}$.

(i) Define $X_{i-2} \triangleq \left[I + (P_0 - P)\mathcal{O}_{i-2}^p\right]^{-1}$. Using the increment relations for $F_{p,i-1}$ and $(P_{i-1} - P)$ from Prob. 14.5, show that

$$F_{p,i-1}F_p^{i-1}X_{i-2} = F_p^i\left[X_{i-2} - X_{i-2}(P_0 - P)F_p^{(i-1)*}H^*R_{e,i-1}^{-1}HF_p^{i-1}X_{i-2}\right].$$

(ii) Using recursion (14.5.12) show that

$$X_{i-1} = \left[X_{i-2} - X_{i-2}(P_0 - P)F_p^{(i-1)*}H^*R_{e,i-1}^{-1}HF_p^{i-1}X_{i-2}\right],$$

and, hence, conclude that $\Phi_p(i,0) = F_p^i X_{i-1}$, as desired.]

(b) Assume now that P_0 is chosen such that the sequence of matrices $\{T_i\}$ in (14.5.2) is uniformly bounded (*cf.* (14.3.4)). Use the result of Prob. 14.2 to conclude that $\lim_{i\to\infty}\Phi_p(i,0) = 0$.

14.9 **(Uniform boundedness of $\{\mathcal{O}_i\}$)** Consider the definition of the matrix $\mathcal{O}_i$ from Prob. 11.2 and define the matrix $X_j = [I + (P_0 - P)\mathcal{O}_j^p]^{-1}$. Assume $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is unit-circle controllable, and that the initial condition P_0 is chosen according to any of the conditions (a)–(c) in the statement of Thm. 14.8.1. Use the expression for $\Phi_p(j,0)$ from Prob. 14.8 in terms of X_j and F_p, in addition to the stability of F_p, to show that the sequence $\{\mathcal{O}_i\}$ is uniformly bounded. [*Hint.* Follow the argument in the proof of Thm. 14.8.1 that established the uniform boundedness of $\{R_{r,i}\}$ from (14.8.3).]

14.10 **(Change-in-initial-conditions formula)** Let $\Phi_p(i,i_0)$ be the state transition matrix that is defined by $\Phi_p(i+1,i_0) = [F - K_{p,i}H]\Phi_p(i,i_0)$, $\Phi_p(i_0,i_0) = I$, where $K_{p,i} = FP_iH^*R_{e,i}^{-1}$, $R_{e,i} = R + HP_iH^*$, and P_i is obtained from the Riccati difference equation

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^*, \qquad P_{i_0}.$$

Note that we are assuming constant $\{F, G, H, R, Q\}$. Now assume that we change the initial condition at time $i = i_0$ from P_{i_0} to $P_{i_0}^\sigma$ and denote the resulting variables by $\{P_\cdot^\sigma, K_{p,\cdot}^\sigma, R_{e,\cdot}^\sigma, \Phi_p^\sigma(\cdot,i_0)\}$, *i.e.*,

$$P_{i+1}^\sigma = FP_i^\sigma F^* + GQG^* - K_{p,i}^\sigma R_{e,i}^\sigma K_{p,i}^{\sigma *}, \qquad P_{i_0}^\sigma.$$

Also, $\Phi_p^\sigma(i+1,i_0) = [F - K_{p,i}^\sigma H]\Phi_p^\sigma(i,i_0)$, $\Phi_p^\sigma(i_0,i_0) = I$.

(a) Establish the change-in-initial-conditions formula

$$\Phi_p^\sigma(i,i_0) = \Phi_p(i,i_0)[I + (P_{i_0}^\sigma - P_{i_0})\mathcal{O}_{i-1,i_0}]^{-1},$$

where we define $\mathcal{O}_{i-1,i_0}$ as

$$\mathcal{O}_{i-1,i_0} \triangleq \sum_{k=i_0}^{i-1} \Phi_p^*(k,i_0)H^*R_{e,k}^{-1}H\Phi_p(k,i_0), \qquad \mathcal{O}_{i_0-1,i_0} = 0,$$

and where we are assuming the invertibility of $[I + (P_{i_0}^\sigma - P_{i_0})\mathcal{O}_{i-1,i_0}]$.

(b) Define $\Psi_p(i,i_0) \triangleq \Phi_p^\sigma(i,i_0)\Phi_p(i_0+1,i_0)$. Verify that $\Psi_p(i,i_0)$ satisfies the difference equation

$$\Psi_p(i+1,i_0) = [F - K_{p,i}^\sigma H]\Psi_p(i,i_0), \quad \Psi_p(i_0,i_0) = \Phi_p(i_0+1,i_0).$$

(c) Now assume $P_{i_0}^\sigma = P_{i_0+1}$. By comparing the difference equation for $\Psi_p(\cdot,i_0)$ with that for $\Phi_p(\cdot,i_0)$, show that $\Phi_p^\sigma(i,i_0)\Phi_p(i_0+1,i_0) = \Phi_p(i+1,i_0)$, and deduce the identity

$$\Phi_p(i+1,i_0) = \Phi_p(i,i_0)[I + (P_{i_0+1} - P_{i_0})\mathcal{O}_{i-1,i_0}]^{-1}\Phi_p(i_0+1,i_0).$$

14.11 **(Change-in-initial conditions formula)** Consider again the definition of $\mathcal{O}_i$ in Prob. 11.2 and define the matrix $X_j = [I + (P_0 - P)\mathcal{O}_j^p]^{-1}$. Assume further that the sequence $\{X_j\}$ is uniformly bounded. We know from the discussion in this chapter that this condition can be guaranteed when P_0 is chosen according to any of the conditions (a)–(c) in the statement of Thm. 14.8.1 — see Prob. 14.3.

Now recall that the observability Gramian $\mathcal{O}_i^p$ satisfies the recursion

$$\mathcal{O}_{i+1}^p = F_p^*\mathcal{O}_i^pF_p + H^*R_e^{-1}H, \qquad \mathcal{O}_{-1}^p = 0.$$

We want to establish that $\mathcal{O}_i$ and $\mathcal{O}_i^p$ are related as follows:

$$\mathcal{O}_i = \mathcal{O}_i^p\left[I + (P_0 - P)\mathcal{O}_i^p\right]^{-1},$$

where P is the stabilizing solution of the DARE. We proceed by induction.

(a) The relation clearly holds for $i = -1$ since both $\mathcal{O}_{-1}$ and $\mathcal{O}_{-1}^p$ are zero. Now take $\mathcal{O}_0 = H^*R_{e,0}^{-1}H$ and $\mathcal{O}_0^p = H^*R_e^{-1}H$. Show that

$$\mathcal{O}_0 = \mathcal{O}_0^p\left[I + (P_0 - P)\mathcal{O}_0^p\right]^{-1}.$$

That is, the relation is also satisfied at time 0.

(b) Now assume it holds up to time i and let us show that it holds at time $i+1$ as well. Start with

$$\mathcal{O}_i = \mathcal{O}_i^p\left[I + (P_0 - P)\mathcal{O}_i^p\right]^{-1},$$

and substitute $\mathcal{O}_i$ and $\mathcal{O}_i^p$ by

$$\mathcal{O}_i = \mathcal{O}_{i+1} - \Phi_p^*(i+1,0)H^*R_{e,i+1}^{-1}H\Phi_p(i+1,0),$$

$$\mathcal{O}_i^p = \mathcal{O}_{i+1}^p - F_p^{(i+1)*}H^*R_e^{-1}HF_p^{i+1}.$$

Show that the desired equality $\mathcal{O}_{i+1} = \mathcal{O}_{i+1}^p \left[I + (P_0 - P)\mathcal{O}_{i+1}^p\right]^{-1}$ will be satisfied if, and only if, the quantity defined below

$$\Delta_i \triangleq [I - \mathcal{O}_{i+1}(P_0 - P)]\, F_p^{(i+1)*} H^* R_e^{-1} H F_p^{i+1} - X_i^* F_p^{(i+1)*} H^* R_{e,i+1}^{-1} H F_p^{(i+1)},$$

is zero. Here, $X_i = \left[I + (P_0 - P)\mathcal{O}_i^p\right]^{-1}$. [*Hint.* Use the expression for $\Phi_p(i+1, 0)$ in terms of X_i from Prob. 14.8.]

(c) Use the identity $R_{e,i+1}^{-1} R_e = I - R_{e,i+1}^{-1} H (P_{i+1} - P) H^*$, and expression (14.3.1) for $P_{i+1} - P$ to show that

$$\Delta_i = \Gamma_i \; F_p^{(i+1)*} H^* R_e^{-1} H F_p^{(i+1)},$$

where

$$\Gamma_i \triangleq \left\{\left[I - \mathcal{O}_{i+1}(P_0 - P) - X_i^* + X_i^* F_p^{(i+1)*} H^* R_{e,i+1}^{-1} H F_p^{(i+1)} X_i (P_0 - P)\right]\right\}.$$

(d) Substitute $\mathcal{O}_{i+1}$ by $\mathcal{O}_i + X_i^* F_p^{(i+1)*} H^* R_{e,i+1}^{-1} H F_p^{(i+1)} X_i$ and show that $\Gamma_i = 0$.

Remark. The formula tells us how the observability Gramian changes in response to a change in the initial condition for the Riccati recursion from P_0 to P. A similar formula is derived in Prob. 14.8. We shall encounter these results in a more natural way in the context of scattering theory in Ch. 17 — see Thm. 17.6.1. ◆

14.12 (Inertia properties for the Riccati recursion) Suppose $P_i^{(1)}$ and $P_i^{(2)}$ are two solutions to the Riccati recursion (14.1.4) with the *same* $\{F, G, H\}$ and $\{Q, R, S\}$ matrices and $R > 0$, but with *different* initial conditions $\Pi_0^{(1)}$ and $\Pi_0^{(2)}$, respectively. Define $\delta P_i = P_i^{(2)} - P_i^{(1)}$, and factor it as $\delta P_i = L_i S_i L_i^*$, where L_i and S_i have full rank and S_i is square.

(a) Show that $\delta P_{i+1} = F_{p,i}^{(1)} \left[L_i S_i L_i^* - L_i S_i L_i^* H^* (R_{e,i}^{(2)})^{-1} H L_i S_i L_i^*\right] F_{p,i}^{(1)*}$, and deduce that $L_{i+1} = F_{p,i}^{(1)} L_i$ and $S_{i+1}^{-1} = S_i^{-1} + L_i^* H^* (R_{e,i}^{(1)})^{-1} H L_i$.

(b) By using two different block triangular factorizations of the matrix

$$\begin{bmatrix} -R_{e,i}^{(1)} & HL_i \\ L_i^* H^* & S_i^{-1} \end{bmatrix},$$

show that S_i and S_{i+1} have the same inertia if, and only if, $R_{e,i}^{(1)}$ and $R_{e,i}^{(2)}$ have the same inertia.

(c) Show that if $F_{p,i}^{(1)}$ is invertible, then δP_i and δP_{i+1} have the same inertia if, and only if, $R_{e,i}^{(1)}$ and $R_{e,i}^{(2)}$ have the same inertia.

(d) Show that if $\Pi_0^{(1)} \geq 0$ and $\Pi_0^{(2)} \geq 0$, then S_i and S_{i+1} have the same inertia.

14.13 (The dual DARE) Consider the dual DARE (14.5.28) and assume that $\{F, H\}$ is observable and $\{F, GQ^{1/2}\}$ is stabilizable.

(a) Show that P^a is invertible.

(b) Show that (14.5.24) is equivalent to $P_0 > -(P^a)^{-1}$.

(c) Show that if P^{a1} is an invertible solution to the dual DARE, then $-(P^{a1})^{-1}$ is a solution to the original DARE.

(d) Conclude that $-(P^a)^{-1} = P_-$ ≜ the infimum over all solutions to the DARE (14.1.2).

14.14 (Conditions on P_1) Assume P_0 is an initial condition that results in a sequence $\{D_i\}$ that is nonsingular for all i (*cf.* the statement of Lemma 14.5.1). Let P_1 be the value obtained by the Riccati recursion (14.1.4) at time 1 starting with P_0.

(a) Show that $I + (P_1 - P)\mathcal{O}^p$ is nonsingular. Show also that the sequence of matrices $\left\{\left[I + (P_1 - P)\mathcal{O}_i^p\right]^{-1}(P_1 - P)\right\}$ is uniformly bounded. [*Hint.* Introduce the factorization $P_1 - P = \bar{A}_0 \bar{J} \bar{A}_0^*$, for some $\bar{A}_0$ and for some signature matrix $\bar{J}$ (which in fact should be the signature of D_0^{-1}). Now, as in (14.5.7), define the corresponding sequence $\bar{D}_i = \bar{J} + \bar{A}_0^* \mathcal{O}_i^p \bar{A}_0$, $\bar{D}_{-1} = \bar{J}$, for $i \geq 0$. Show that the nonsingularity of the $\{D_i, i \geq 1\}$ implies the nonsingularity of the $\{\bar{D}_i, i \geq 0\}$.]

(b) In a similar vein, assume P_0 satisfies (14.5.24) and show that P_1 satisfies a similar relation:

$$I + \mathcal{O}^{p*/2}(P_1 - P)\mathcal{O}^{p/2} > 0.$$

[*Hint.* Recall Lemma 14.5.7.]

(c) Likewise, assume P_0 satisfies (14.5.29) and show that P_1 satisfies a similar relation:

$$I + (P^a)^{*/2} P_1 (P^a)^{1/2} > 0.$$

[*Hint.* Recall the proof of Thm. 14.5.3.]

14.15 (A symplectic matrix) Let $F^s = F - GSR^{-1}H$, $Q^s = Q - SR^{-1}S^*$, $F_{p,i} = F - (FP_iH^* + GS)(R + HP_iH^*)^{-1}H$, where the $\{P_i\}$ are generated via the Riccati recursion

$$P_{i+1} = FP_iF^* + GQG^* - (FP_iH^* + GS)(R + HP_iH^*)^{-1}(FP_iH^* + GS)^*, \quad P_0.$$

Assume F^s invertible.

(a) Verify that the invertibility of F^s and R guarantees the invertibility of $F_{p,i}$. In particular, show that $F_{p,i}^{-1} = F^{-s} + P_iH^*R^{-1}HF^{-s}$.

(b) Show that

$$\begin{bmatrix} I & -P_i \\ 0 & I \end{bmatrix} M \begin{bmatrix} I & P_{i+1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} F_{p,i}^{-1} & 0 \\ -H^*R^{-1}HF^{-s} & F_{p,i}^* \end{bmatrix},$$

where

$$M = \begin{bmatrix} F^{-s} & -F^{-s}GQ^sG^* \\ -H^*R^{-1}HF^{-s} & F^{s*} + H^*R^{-1}HF^{-s}GQ^sG^* \end{bmatrix}.$$

(c) Show that

$$M = \begin{bmatrix} I & 0 \\ -H^*R^{-1}H & F^{s*} \end{bmatrix} \begin{bmatrix} F^s & GQ^sG^* \\ 0 & I \end{bmatrix}^{-1}.$$

Verify that M is a symplectic matrix, *i.e.*, one that satisfies

$$J^{-1}M^*J = M^{-1}, \quad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

14.16 (Riccati equation with indefinite coefficients) Assume R is nonsingular, R and Q Hermitian, and that F has no unit-circle eigenvalues. Let $\{\lambda, x\}$ denote a generalized eigenvalue-eigenvector pair of the matrix pencil shown below ,

$$\begin{bmatrix} I & -Q \\ 0 & F^* \end{bmatrix} x = \lambda \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix} x.$$

It was shown in Prob. 8.11 that this pencil does not have any unit circle eigenvalues if, and only if, the following Popov function,

$$S_y(z) = H(zI - F)^{-1}Q(z^{-1}I - F^*)^{-1}H^* + R,$$

is nonsingular on the unit circle. So assume that this is the case.

(a) Follow the discussion that led to Thm. E.7.2 in App. E to conclude that there exist $n \times n$ matrices U and V such that

$$\begin{bmatrix} I & -Q \\ 0 & F^* \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} \Lambda,$$

where Λ is an $n \times n$ matrix with all its eigenvalues inside the unit disc, $|\lambda_i(\Lambda)| < 1$. [That is, U and V form a basis for the stable generalized eigenspace of the matrix pencil and that this space is $n-$dimensional.]

(b) Assume V is invertible.

1. Show that UV^{-1} is Hermitian.
2. Define $K_p = V^{-*}\Lambda^*U^*H^*R^{-1}$. Show that K_p satisfies the equation

$$K_p(R + HUV^{-1}H^*) = FUV^{-1}H^*.$$

3. Show that UV^{-1} satisfies the DARE,

$$P = FPF^* + Q - K_pR_eK_p, \qquad R_e = R + HPH^*.$$

4. Introduce the closed-loop matrix $F_p = F - K_pH$. Show that F_p is stable.
5. Show that the system of equations

$$\begin{aligned} P &= FPF^* + Q - K_pR_eK_p^*, \\ K_pR_e &= FPH^*, \\ R_e &= R + HPH^*, \end{aligned}$$

has a *unique* Hermitian solution P that results in a stable closed-loop matrix F_p.

6. Show further that, in the special case $Q \geq 0$ and $R > 0$, the unique stabilizing solution P is not only Hermitian but also nonnegative-definite.

(c) Show that a Hermitian stabilizing solution P of the above DARE system of equations exists if, and only if, V is invertible.

14.17 (Existence of solutions to DARE) Assume $R > 0$, $Q \geq 0$, $\{F, H\}$ detectable, and $\{F, Q^{1/2}\}$ unit-circle controllable. The last two conditions guarantee that the matrix pencil below does not have unit-circle eigenvalues (*cf.* Lemma E.7.3),

$$\begin{bmatrix} I & -Q \\ 0 & F^* \end{bmatrix} x = \lambda \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix} x.$$

Here, $\{\lambda, x\}$ denote a generalized eigenvalue-eigenvector pair.

It is easy to see that conclusions (b.1)–(b.6) of Prob. 14.16 still hold under the assumption that V is invertible. We now prove that the detectability assumption guarantees an invertible V, so that the system of equations that characterize the DARE in part (b.5) of Prob. 14.16 always has a solution. [The argument in this problem is self-contained and does not rely on the introduction of a Popov function. The argument can also be used to establish the invertibility of V in Thm. E.7.2 of the appendix without relying on the existence of a stabilizing solution P.]

(a) Show that U^*V is Hermitian.

(b) Show that $U^*V \geq 0$.

(c) Assume V is singular, say $Vx = 0$ for some nonzero x. Show that $V\Lambda^k x = 0$ for all $k \geq 1$. Conclude that the pair $\{\Lambda, V\}$ is not observable.

(d) Argue that there exists a nonzero vector y such that $\Lambda y = \lambda y$ and $Vy = 0$, with $|\lambda| < 1$.

(e) Show that Uy is a right eigenvector of F that is orthogonal to H. Conclude that $\{F, H\}$ is not detectable. [Parts (a)–(e) therefore show that a singular V violates the detectability assumption. In other words, detectability of $\{F, H\}$ implies invertibility of V.]

(f) Establish the converse statement, *viz.*, that the invertibility of V implies the detectability of $\{F, H\}$.

(g) Assume now that $R < 0$ and $Q \leq 0$. Assume further that $\{F, H\}$ is detectable and $\{F, Q\}$ is unit-circle controllable. Show again that V is invertible.

14.18 (A matrix pencil) Consider the matrix pencil introduced in Prob. 14.16, and define

$$A \triangleq \begin{bmatrix} I & -Q \\ 0 & F^* \end{bmatrix}, \quad B \triangleq \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix}, \quad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

Verify that $AJA^* = BJB^*$.

14.19 (A useful matrix norm) Let A be an $n \times n$ matrix with eigenvalues $\{\lambda_i\}$. The spectral radius of A, denoted by $\rho(A)$, is defined as

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|.$$

A basic result in matrix theory states that there always exists a similarity transformation T such that $A = TJT^{-1}$, where J is in Jordan canonical form . That is, J is block diagonal, say $J = \text{diag}\{J_1, J_2, \ldots, J_p\}$, where each J_k has the generic form

$$J_k = \begin{bmatrix} \lambda_k & & & \\ 1 & \lambda_k & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_k \end{bmatrix},$$

for some eigenvalue λ_k of A. For any positive scalar ϵ, define the $n \times n$ diagonal matrix $D = \text{diag}\{\epsilon, \epsilon^2, \ldots, \epsilon^n\}$.

(a) Verify that DJD^{-1} has the same form as J except that the unit entries of J are replaced by ϵ. Conclude that the one norm of DJD^{-1} is equal to $\rho(A)+\epsilon$. *Remark.* Recall that the one norm of an $n \times n$ matrix B, denoted by $\|B\|_1$, is defined as the maximum absolute column sum of B,

$$\|B\|_1 \triangleq \max_{1 \le j \le n} \sum_{i=1}^{n} |b_{ij}|.$$

(b) Now for any $n \times n$ matrix B define the function $\|B\|_\rho \triangleq \|DT^{-1}BTD^{-1}\|_1$. Show that the function $\|\cdot\|_\rho$ so defined is a matrix norm, *i.e.*, show that it satisfies the following properties, for any B and C and for any complex scalar α,

1. $\|B\|_\rho \ge 0$ always. Moreover, $\|B\|_\rho = 0$ if, and only if, $B = 0$.
2. $\|\alpha B\|_\rho = |\alpha| \cdot \|B\|_\rho$.
3. Triangle inequality: $\|B + C\|_\rho \le \|B\|_\rho + \|C\|_\rho$.
4. Submultiplicative property: $\|BC\|_\rho \le \|B\|_\rho \cdot \|C\|_\rho$.

(c) Verify that

$$\rho(A) \le \|A\|_\rho \le \rho(A) + \epsilon.$$

Remark. This problem esablishes a result in Horn and Johnson (1985, Lemma 5.6.10)), which states that for any matrix A with spectral radius $\rho(A)$, there always exists a matrix norm $\|A\|$ such that

$$\rho(A) \le \|A\| \le \rho(A) + \epsilon, \quad \text{for any } \epsilon > 0.$$

CHAPTER 15

Duality and Equivalence in Estimation and Control

In various discussions so far, we have shown the value of the geometric point of view, in which we regard random variables as vectors in certain linear spaces. This led naturally to the orthogonality interpretation of optimality, to the concept of sequential orthogonalization (or innovations), to a uniform approach to deterministic and stochastic problems (see, *e.g.*, Sec. 4.1.4 and App. 4.A), and to many simpler proofs as compared to purely algebraic approaches. However, there is an important aspect of linear spaces that we have not yet introduced, *viz.*, the important concepts of duality, dual bases, and orthogonal complements.

We do so in this chapter and provide several applications, *e.g.*, interpreting the information form expressions for the Kalman filter via duality (Sec. 15.2.2), and more importantly, the development of the so-called complementary state-space models of various types (Sec. 15.7). The latter is another example of how the geometric approach provides further insight into the structure of stochastic state-space models. We also present important extensions of the simple equivalence result of Sec. 3.5 to give a complete picture (see Tables 15.1 and 15.2) of equivalence and duality relationships between stochastic and deterministic quadratic cost problems.

Applications of duality to the solution of quadratic control problems are also studied in some detail in this chapter (*e.g.*, Secs. 15.3.5, 15.4.4, 15.5.3, and 15.6.3). These applications show once again that the solutions of estimation problems have value well beyond themselves.

15.1 DUAL BASES

We begin with a linear vector space $\mathcal{V}$ over a given ring of scalars (*e.g.*, the the ring of complex numbers $\mathcal{S}$ or the ring of matrices). In particular, using a ring rather than a field allows us to use matrix-valued inner products (or Gramians), a freedom used

often in previous chapters. Readers may find it useful to quickly review the material in App. 4.A before proceeding.

Now consider a set of linearly independent vectors in $\mathcal{V}$, say

$$\{ \mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_m, \mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_n \},$$

which we shall denote by $\{\mathbf{z}, \mathbf{y}\}$, where

$$\mathbf{z} \triangleq \text{col}\{\mathbf{z}_0, \ldots, \mathbf{z}_m\}, \qquad \mathbf{y} \triangleq \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_n\}.$$

As a matter of fact we do not need to partition the set of independent vectors into a set of $\{\mathbf{z}_i\}$ and a set of $\{\mathbf{y}_i\}$ in order to introduce the concept of dual bases. However, since our major interest is estimation and the set $\{\mathbf{y}_i\}$ will typically designate the observations while the set $\{\mathbf{z}_i\}$ will designate the quantities we want to estimate, we shall find this partitioning to be useful later.

The corresponding Gramian for this set of vectors will be denoted by

$$\left\langle \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix}, \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \right\rangle = \begin{bmatrix} \langle \mathbf{z}, \mathbf{z} \rangle & \langle \mathbf{z}, \mathbf{y} \rangle \\ \langle \mathbf{y}, \mathbf{z} \rangle & \langle \mathbf{y}, \mathbf{y} \rangle \end{bmatrix} = \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}, \tag{15.1.1}$$

which by the linear independence of $\{\mathbf{z}, \mathbf{y}\}$ must be nonsingular

$$\det \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} \neq 0.$$

For the same reason, the matrices R_z and R_y are also nonsingular.

We shall write $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$ to denote the linear space of all vectors generated by

$$a_0\mathbf{z}_0 + \ldots + a_m\mathbf{z}_m + b_0\mathbf{y}_0 + \ldots + b_n\mathbf{y}_n,$$

for any $a_i, b_j \in \mathcal{S}$. The linear independence of the vectors $\{\mathbf{z}_0, \ldots, \mathbf{z}_m, \mathbf{y}_0, \ldots \mathbf{y}_n\}$ then implies that these vectors form a *basis* for $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$. Rather than explicitly write the basis vectors as $\{\mathbf{z}_0, \ldots, \mathbf{z}_m, \mathbf{y}_0, \ldots \mathbf{y}_n\}$, we shall simply say that $\{\mathbf{z}, \mathbf{y}\}$ forms a basis for $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$.

Definition 15.1.1. (Dual Basis) *Given a basis, $\{\mathbf{z}, \mathbf{y}\}$, the dual basis is defined as the pair $\{\mathbf{z}^d, \mathbf{y}^d\}$ with the two properties*

$$\mathcal{L}\{\mathbf{z}^d, \mathbf{y}^d\} = \mathcal{L}\{\mathbf{z}, \mathbf{y}\}, \tag{15.1.2}$$

$$\left\langle \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix}, \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \right\rangle = \begin{bmatrix} \langle \mathbf{z}^d, \mathbf{z} \rangle & \langle \mathbf{z}^d, \mathbf{y} \rangle \\ \langle \mathbf{y}^d, \mathbf{z} \rangle & \langle \mathbf{y}^d, \mathbf{y} \rangle \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \tag{15.1.3}$$

♦

That is, $\mathbf{z}^d$ and $\mathbf{y}^d$ form a basis for the same linear space $\mathcal{L}(\mathbf{z}, \mathbf{y})$, and have the property that $\mathbf{z}^d$ is orthogonal to $\mathbf{y}$ and $\mathbf{y}^d$ is orthogonal to $\mathbf{z}$; moreover, $\mathbf{z}^d$ and $\mathbf{y}^d$ are normalized such that $\langle \mathbf{z}^d, \mathbf{z} \rangle = I$ and $\langle \mathbf{y}^d, \mathbf{y} \rangle = I$.

Note that if $\{\mathbf{z}, \mathbf{y}\}$ were an *orthonormal* basis then the dual basis would simply coincide with the original basis. In general, however, the dual basis will be different and (15.1.3) is referred to as a *bi-orthogonality condition*. We now describe two ways — algebraic and geometric — of finding the dual basis.

15.1.1 Algebraic Specification

Clearly, since $\{\mathbf{z}, \mathbf{y}\}$ and $\{\mathbf{z}^d, \mathbf{y}^d\}$ span the same linear space we must have

$$\begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix},$$

for some nonsingular block matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$. Hence,

$$\left\langle \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix}, \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \right\rangle = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix},$$

where the last equality follows from (15.1.3). Therefore, we see that

$$\begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} = \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix}, \tag{15.1.4}$$

with Gramian

$$\begin{bmatrix} R_{z^d} & R_{z^d y^d} \\ R_{y^d z^d} & R_{y^d} \end{bmatrix} \triangleq \left\langle \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix}, \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} \right\rangle = \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1}. \tag{15.1.5}$$

The above arguments and result are quite specific but may not be very intuitive; this can be remedied by considering the geometric description.

15.1.2 Geometric Specification

Let us first introduce (or recall from Ch. 3) the notations

$$\hat{\mathbf{z}}_{|\mathbf{y}} \triangleq \text{the projection of } \mathbf{z} \text{ onto } \mathcal{L}\{\mathbf{y}\}, \qquad \hat{\mathbf{y}}_{|\mathbf{z}} \triangleq \text{the projection of } \mathbf{y} \text{ onto } \mathcal{L}\{\mathbf{z}\},$$

and the corresponding errors

$$\tilde{\mathbf{z}} \triangleq \tilde{\mathbf{z}}_{|\mathbf{y}} = \mathbf{z} - \hat{\mathbf{z}}_{|\mathbf{y}}, \qquad \tilde{\mathbf{y}} \triangleq \tilde{\mathbf{y}}_{|\mathbf{z}} = \mathbf{y} - \hat{\mathbf{y}}_{|\mathbf{z}}.$$

Now it is straightforward to verify that $\{\tilde{\mathbf{z}}, \mathbf{y}\}$ and $\{\mathbf{z}^d, \mathbf{y}\}$ span the same linear space (*cf.* Prob. 15.1), while the orthogonality principle of least-mean-squares estimation shows that $\langle \tilde{\mathbf{z}}, \mathbf{y} \rangle = 0$. Combining these facts with the property $\langle \mathbf{z}^d, \mathbf{y} \rangle = 0$, we conclude that $\tilde{\mathbf{z}}$ and $\mathbf{z}^d$ must span the same linear space. Thus we must have $\mathbf{z}^d = M\tilde{\mathbf{z}}$ for some nonsingular matrix M. However, note that

$$I = \langle \mathbf{z}^d, \mathbf{z} \rangle = M\langle \tilde{\mathbf{z}}, \mathbf{z} \rangle = M\langle \tilde{\mathbf{z}}, \tilde{\mathbf{z}} + \hat{\mathbf{z}}_{\mathbf{y}} \rangle = M\langle \tilde{\mathbf{z}}, \tilde{\mathbf{z}} \rangle \triangleq MR_{\tilde{z}},$$

so that $M = R_{\tilde{z}}^{-1}$ and, consequently,

$$\mathbf{z}^d = R_{\tilde{z}}^{-1}\, \tilde{\mathbf{z}}_{|\mathbf{y}}, \qquad R_{\tilde{z}} = \|\tilde{\mathbf{z}}_{|\mathbf{y}}\|^2. \tag{15.1.6}$$

Similarly, of course we have

$$\mathbf{y}^d = R_{\tilde{y}}^{-1}\, \tilde{\mathbf{y}}_{|\mathbf{z}}, \qquad R_{\tilde{y}} = \|\tilde{\mathbf{y}}_{|\mathbf{z}}\|^2. \tag{15.1.7}$$

The reader may want to verify that the invertibility of the matrices R_z, R_y, and of the Gramian (15.1.1), guarantees the invertibility of $R_{\tilde{z}}$ and $R_{\tilde{y}}$ (see Prob. 15.2). In simple cases, determining the dual basis from the geometry can be easier than inverting the Gramian matrix.

15.1.3 Some Reasons for Introducing Dual Bases

But why introduce dual bases? There are several reasons. One is the fact that to compute the projection of a vector $\mathbf{x}$ onto $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$, say

$$\hat{\mathbf{x}}_{|\mathbf{y},\mathbf{z}} = A\mathbf{z} + B\mathbf{y}, \tag{15.1.8}$$

for some A and B, requires solving a system of linear equations to determine the coefficient matrices $\{A, B\}$, except, of course, when $\mathbf{y}$ and $\mathbf{z}$ are orthogonal, in which case, the coefficients are obtained by projecting $\mathbf{x}$ separately on $\mathbf{z}$ and $\mathbf{y}$,

$$\hat{\mathbf{x}}_{|\mathbf{y},\mathbf{z}} = \langle \mathbf{x}, \mathbf{z}\rangle \|\mathbf{z}\|^{-2}\mathbf{z} + \langle \mathbf{x}, \mathbf{y}\rangle \|\mathbf{y}\|^{-2}\mathbf{y}, \quad \text{when } \langle \mathbf{y}, \mathbf{z}\rangle = 0. \tag{15.1.9}$$

In the general case, we can obtain a somewhat similar expression by using the dual basis.

Lemma 15.1.1 (Projection via Dual Basis) *The projection of* $\mathbf{x}$ *onto* $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$, $\hat{\mathbf{x}}_{|\mathbf{y},\mathbf{z}}$, *can be written as*

$$\hat{\mathbf{x}}_{|\mathbf{y},\mathbf{z}} = \langle \mathbf{x}, \mathbf{z}^d\rangle \mathbf{z} + \langle \mathbf{x}, \mathbf{y}^d\rangle \mathbf{y}, \tag{15.1.10}$$

where $\{\mathbf{z}^d, \mathbf{y}^d\}$ *is the dual basis.* ∎

Proof: We need to find $\{A, B\}$ in (15.1.8) such that

$$(\mathbf{x} - A\mathbf{z} - B\mathbf{y}) \perp \mathcal{L}\{\mathbf{z}, \mathbf{y}\} = \mathcal{L}\{\mathbf{z}^d, \mathbf{y}^d\}.$$

But

$$0 = \langle \mathbf{x} - A\mathbf{z} - B\mathbf{y}, \mathbf{z}^d\rangle = \langle \mathbf{x}, \mathbf{z}^d\rangle - A\langle \mathbf{z}, \mathbf{z}^d\rangle - B\langle \mathbf{y}, \mathbf{z}^d\rangle = \langle \mathbf{x}, \mathbf{z}^d\rangle - A,$$

so that $A = \langle \mathbf{x}, \mathbf{z}^d\rangle$. Similarly, $B = \langle \mathbf{x}, \mathbf{y}^d\rangle$. ♦

Of course, the apparent simplicity of the formula (15.1.10) is, in general, only conceptual rather than computational. The algebraic equations for $\{A, B\}$ are

$$\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} = \begin{bmatrix} \langle \mathbf{x}, \mathbf{z}\rangle & \langle \mathbf{x}, \mathbf{y}\rangle \end{bmatrix},$$

and inverting the (Gramian) coefficient matrix in the above equation is equivalent to determining the dual basis $\{\mathbf{z}^d, \mathbf{y}^d\}$! Nevertheless, the conceptual simplification makes the introduction of dual bases, and their geometric interpretation, quite useful, as we shall see in the sequel.

Another more practical reason for using dual bases is that they provide alternative ways for solving estimation problems (see Sec. 15.7.6). Thus suppose $\mathbf{x} \in \mathcal{L}\{\mathbf{z}, \mathbf{y}\}$ and that we would like to find $\hat{\mathbf{x}}_{|\mathbf{y}}$. Now since $\{\mathbf{z}, \mathbf{y}\}$ and $\{\mathbf{z}^d, \mathbf{y}\}$ span the same linear space, we also have that $\mathbf{x} \in \mathcal{L}\{\mathbf{z}^d, \mathbf{y}\}$. Using the fact that, by definition, $\mathbf{z}^d$ and $\mathbf{y}$ are orthogonal, we can uniquely decompose $\mathbf{x}$ as $\mathbf{x} = \hat{\mathbf{x}}_{|\mathbf{y}} + \hat{\mathbf{x}}_{|\mathbf{z}^d}$, or

$$\hat{\mathbf{x}}_{|\mathbf{y}} = \mathbf{x} - \hat{\mathbf{x}}_{|\mathbf{z}^d} \triangleq \tilde{\mathbf{x}}_{|\mathbf{z}^d}. \tag{15.1.11}$$

In other words, we can obtain the desired estimator, $\hat{\mathbf{x}}_{|\mathbf{y}}$, by projecting onto the dual basis, $\mathbf{z}^d$, and subtracting the result from $\mathbf{x}$; note also that the *error* in estimating $\mathbf{x}$ from $\mathbf{z}^d$ is the estimator of $\mathbf{x}$ given $\mathbf{y}$.

15.1.4 Estimators via the Dual Basis

We can obtain some more interesting results by combining the algebraic and geometric characterizations of dual bases. The geometric characterization (15.1.6)–(15.1.7) states that

$$\begin{bmatrix} \tilde{\mathbf{z}}_{|\mathbf{y}} \\ \tilde{\mathbf{y}}_{|\mathbf{z}} \end{bmatrix} = \begin{bmatrix} R_{\tilde{z}} & 0 \\ 0 & R_{\tilde{y}} \end{bmatrix} \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix}.$$

Therefore

$$\left\langle \begin{bmatrix} \tilde{\mathbf{z}}_{|\mathbf{y}} \\ \tilde{\mathbf{y}}_{|\mathbf{z}} \end{bmatrix}, \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} \right\rangle = \begin{bmatrix} R_{\tilde{z}} & 0 \\ 0 & R_{\tilde{y}} \end{bmatrix} \left\langle \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix}, \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} \right\rangle = \begin{bmatrix} R_{\tilde{z}} & 0 \\ 0 & R_{\tilde{y}} \end{bmatrix} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1}.$$

But note that

$$\langle \tilde{\mathbf{z}}_{|\mathbf{y}}, \mathbf{z}^d\rangle = \langle \mathbf{z} - R_{zy}R_y^{-1}\mathbf{y}, \mathbf{z}^d\rangle = I - 0 = I,$$

while

$$\langle \tilde{\mathbf{z}}_{|\mathbf{y}}, \mathbf{y}^d\rangle = \langle \mathbf{z} - R_{zy}R_y^{-1}\mathbf{y}, \mathbf{y}^d\rangle = 0 - R_{zy}R_y^{-1}.$$

With similar results for $\langle \tilde{\mathbf{y}}_{|\mathbf{z}}, \mathbf{z}^d\rangle$ and $\langle \tilde{\mathbf{y}}_{|\mathbf{z}}, \mathbf{y}^d\rangle$, we see that

$$\left\langle \begin{bmatrix} \tilde{\mathbf{z}}_{|\mathbf{y}} \\ \tilde{\mathbf{y}}_{|\mathbf{z}} \end{bmatrix}, \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} \right\rangle = \begin{bmatrix} I & -R_{zy}R_y^{-1} \\ -R_{yz}R_z^{-1} & I \end{bmatrix}.$$

But this leads immediately to the matrix identity

$$\begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1} = \begin{bmatrix} R_{\tilde{z}}^{-1} & 0 \\ 0 & R_{\tilde{y}}^{-1} \end{bmatrix} \begin{bmatrix} I & -R_{zy}R_y^{-1} \\ -R_{yz}R_z^{-1} & I \end{bmatrix}, \tag{15.1.12}$$

whose origin is otherwise not so evident. We can also rewrite (15.1.12) as (*cf.* (15.1.5))

$$\begin{bmatrix} R_{z^d} & R_{z^d y^d} \\ R_{y^d z^d} & R_{y^d} \end{bmatrix} = \begin{bmatrix} R_{\tilde{z}}^{-1} & -R_{\tilde{z}}^{-1}R_{zy}R_y^{-1} \\ -R_{\tilde{y}}^{-1}R_{yz}R_z^{-1} & R_{\tilde{y}}^{-1} \end{bmatrix}.$$

In particular, this implies that

$$R_{z^d} = R_{\tilde{z}}^{-1}, \quad R_{y^d} = R_{\tilde{y}}^{-1}$$

and

$$R_{zy}R_y^{-1} = -R_{\tilde{z}}R_{z^d y^d} = -R_{z^d}^{-1}R_{z^d y^d} = -(R_{y^d z^d}R_{z^d}^{-1})^*.$$

But recalling that

$$\hat{\mathbf{z}}_{|\mathbf{y}} = R_{zy}R_y^{-1}\mathbf{y} \quad \text{and} \quad \hat{\mathbf{y}}^d_{|\mathbf{z}^d} = R_{y^d z^d}R_{z^d}^{-1}\mathbf{z}^d,$$

we see that the gain matrix for estimating $\mathbf{z}$ from $\mathbf{y}$ is the negative conjugate transpose of the gain matrix for estimating $\mathbf{y}^d$ from $\mathbf{z}^d$! This is an important result, and we therefore collect the above identities into the following statement.

Lemma 15.1.2 (Dual Computation of Projections) *The projection of $\mathbf{z}$ onto $\mathcal{L}(\mathbf{y})$ can be computed as $\hat{\mathbf{z}}_{|\mathbf{y}} = -R_{z^d}^{-1} R_{z^d y^d} \mathbf{y}$, where $\{R_{z^d}, R_{z^d y^d}\}$ are the Gramians and cross-Gramians of the dual basis vectors $\{\mathbf{z}^d, \mathbf{y}^d\}$. This leads to the identities*

$$R_{zy} R_y^{-1} = -R_{z^d}^{-1} R_{z^d y^d},$$

and

$$\|\tilde{\mathbf{z}}_{|\mathbf{y}}\|^2 = R_{\tilde{z}} = R_{z^d}^{-1}.$$

■

15.2 APPLICATION TO LINEAR MODELS

The previous lemma captures some of the duality between $\{\mathbf{z}, \mathbf{y}\}$ and the dual basis $\{\mathbf{z}^d, \mathbf{y}^d\}$. However, before going further, let us study the application of the above results to the important case where the $\{\mathbf{z}, \mathbf{y}\}$ are related in a linear fashion. We shall see that this problem induces a linear relation between $\mathbf{z}^d$ and $\mathbf{y}^d$ as well.

15.2.1 Linear Models and Dual Bases

Consider again the simple linear model studied earlier in Sec. 3.4,

$$\mathbf{y} = H\mathbf{z} + \mathbf{v} = \begin{bmatrix} H & I \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{v} \end{bmatrix}, \tag{15.2.1}$$

where the Gramian matrix of $\mathbf{z}$ and $\mathbf{v}$ is taken to be block-diagonal,

$$\left\langle \begin{bmatrix} \mathbf{z} \\ \mathbf{v} \end{bmatrix}, \begin{bmatrix} \mathbf{z} \\ \mathbf{v} \end{bmatrix} \right\rangle = \begin{bmatrix} R_z & 0 \\ 0 & R_v \end{bmatrix}, \quad \det R_z \neq 0, \quad \det R_v \neq 0. \tag{15.2.2}$$

It then follows easily from (15.2.1) that

$$\begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} = \begin{bmatrix} I & 0 \\ H & I \end{bmatrix} \begin{bmatrix} R_z & 0 \\ 0 & R_v \end{bmatrix} \begin{bmatrix} I & H^* \\ 0 & I \end{bmatrix},$$

which shows, using (15.1.4) and a little algebra, that

$$\begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} = \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} R_z^{-1}\mathbf{z} - H^* R_v^{-1}\mathbf{v} \\ R_v^{-1}\mathbf{v} \end{bmatrix}.$$

We thus see that

$$\mathbf{y}^d = R_v^{-1}\mathbf{v}, \tag{15.2.3}$$

and that $\mathbf{z}^d$ arises from a *dual* linear model,

$$\mathbf{z}^d = -H^* R_v^{-1}\mathbf{v} + R_z^{-1}\mathbf{z} = -H^*\mathbf{y}^d + R_z^{-1}\mathbf{z} \triangleq -H^*\mathbf{y}^d + \mathbf{v}^d, \tag{15.2.4}$$

where we have introduced $\mathbf{v}^d \triangleq R_z^{-1}\mathbf{z}$ and, hence,

$$\left\langle \begin{bmatrix} \mathbf{y}^d \\ \mathbf{v}^d \end{bmatrix}, \begin{bmatrix} \mathbf{y}^d \\ \mathbf{v}^d \end{bmatrix} \right\rangle = \begin{bmatrix} R_v^{-1} & 0 \\ 0 & R_z^{-1} \end{bmatrix}. \tag{15.2.5}$$

The Gramian matrix of the dual basis is readily seen to be

$$\left\langle \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix}, \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} \right\rangle = \begin{bmatrix} R_{z^d} & R_{z^d y^d} \\ R_{y^d z^d} & R_{y^d} \end{bmatrix} = \begin{bmatrix} R_z^{-1} + H^* R_v^{-1} H & -H^* R_v^{-1} \\ -R_v^{-1} H & R_v^{-1} \end{bmatrix}.$$

We can therefore compute the projection of $\mathbf{y}^d$ onto $\mathbf{z}^d$ as follows

$$\hat{\mathbf{y}}^d_{|\mathbf{z}^d} = R_{y^d z^d d} R_{z^d}^{-1} \mathbf{z}^d = -R_v^{-1} H (R_z^{-1} + H^* R_v^{-1} H)^{-1} \mathbf{z}^d.$$

On the other hand,

$$\hat{\mathbf{z}}_{|\mathbf{y}} = R_{zy} R_y^{-1} \mathbf{y} = R_z H^* (R_v + H R_z H^*)^{-1} \mathbf{y}.$$

But as claimed in Lemma 15.1.2, the coefficient matrices for these two problems must be the negative conjugate transpose of each other, *i.e.*, we must have

$$\begin{aligned} R_z H^* (R_v + H R_z H^*)^{-1} &= -[-R_v^{-1} H (R_z^{-1} + H^* R_v^{-1} H)^{-1}]^*, \\ &= (R_z^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1}, \end{aligned}$$

which is a useful identity that we deduced and verified algebraically in Sec. 3.4 using the matrix inversion lemma of App. A. Referring to that discussion, we observe that the so-called *information* form expressions for the estimators actually correspond to an estimation problem for certain *dual* variables. This fact will become more evident in Sec. 15.2.2 below. First, however, we summarize the above results in the following statement.

Lemma 15.2.1 (Linear Models and Dual Bases) *Suppose $\{\mathbf{z}, \mathbf{y}\}$ satisfy $\mathbf{y} = H\mathbf{z} + \mathbf{v}$, with (15.2.2). Then the dual basis $\{\mathbf{z}^d, \mathbf{y}^d\}$ will also satisfy a linear model, namely,*

$$\mathbf{z}^d = -H^*\mathbf{y}^d + \mathbf{v}^d,$$

where $\mathbf{y}^d = R_v^{-1}\mathbf{v}$, $\mathbf{v}^d = R_z^{-1}\mathbf{z}$, and $\{\mathbf{y}^d, \mathbf{v}^d\}$ satisfy (15.2.5). We also have the identities

$$\begin{aligned} R_{zy} R_y^{-1} &= R_z H^* (R_v + H R_z H^*)^{-1}, \\ &= (R_z^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1} \quad = \quad -R_{z^d}^{-1} R_{z^d y^d}, \end{aligned}$$

and

$$\langle \tilde{\mathbf{z}}_{|\mathbf{y}}, \tilde{\mathbf{z}}_{|\mathbf{y}} \rangle = R_{\tilde{z}} = R_{z^d}^{-1} = (R_z^{-1} + H^* R_v^{-1} H)^{-1}.$$

■

15.2.2 Application to the Measurement Update Problem

The results of the linear model case can be used to provide a deeper insight into the information form versions of the measurement update formula (*cf.* Thm. 9.5.1):

$$\hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_i + K_{f,i}\mathbf{e}_i, \tag{15.2.6}$$

with

$$K_{f,i} \triangleq < \mathbf{x}_i, \mathbf{e}_i > R_{e,i}^{-1} = < \tilde{\mathbf{x}}_i, \mathbf{e}_i > R_{e,i}^{-1} = P_i H_i^* [R_i + H_i P_i H_i^*]^{-1}. \tag{15.2.7}$$

Using the result of Lemma 15.1.2, we can suggest an alternative way for the evaluation of $K_{f,i}$ by considering a dual projection problem. So start with

$$\mathbf{e}_i = H_i \tilde{\mathbf{x}}_i + \mathbf{v}_i.$$

Next using Lemma 15.2.1, a dual basis for $\mathcal{L}\{\tilde{\mathbf{x}}_i, \mathbf{e}_i\}$, $\{\tilde{\mathbf{x}}_i^d, \mathbf{e}_i^d\}$, is given by

$$\begin{cases} \tilde{\mathbf{x}}_i^d = -H_i^* R_i^{-1} \mathbf{v}_i + P_i^{-1} \tilde{\mathbf{x}}_i, \\ \mathbf{e}_i^d = R_i^{-1} \mathbf{v}_i. \end{cases} \tag{15.2.8}$$

Now $\tilde{\mathbf{x}}_{i|\mathbf{e}_i} = \tilde{\mathbf{x}}_{i|i}$, so that, by definition (see (15.1.6)),

$$\tilde{\mathbf{x}}_i^d = \|\tilde{\mathbf{x}}_{i|i}\|^{-2} \tilde{\mathbf{x}}_{i|i} = P_{i|i}^{-1} \tilde{\mathbf{x}}_{i|i}, \qquad \|\tilde{\mathbf{x}}_i^d\|^2 = P_{i|i}^{-1}.$$

The equations for $\mathbf{x}_i^d$ yield

$$P_{i|i}^{-1} = \|\tilde{\mathbf{x}}_i^d\|^2 = H_i^* R_i^{-1} H_i + P_i^{-1}. \tag{15.2.9}$$

We can also write

$$\begin{aligned} P_{i|i}^{-1} \hat{\mathbf{x}}_{i|i} &= P_{i|i}^{-1} (\mathbf{x}_i - \tilde{\mathbf{x}}_{i|i}), \\ &= (P_i^{-1} + H_i^* R_i^{-1} H_i) \mathbf{x}_i - (-H_i^* R_i^{-1} \mathbf{v}_i + P_i^{-1} \tilde{\mathbf{x}}_i), \\ &= P_i^{-1} \hat{\mathbf{x}}_i + H_i^* R_i^{-1} (H_i \mathbf{x}_i + \mathbf{v}_i), \\ &= P_i^{-1} \hat{\mathbf{x}}_i + H_i^* R_i^{-1} \mathbf{y}_i. \end{aligned} \tag{15.2.10}$$

Equations (15.2.9)–(15.2.10) are exactly the information form expressions for the measurement update problem (*cf.* Thm. 9.5.1).

We can also obtain the formula (9.5.29) for $K_{f,i}$ by noting that the problem of projecting $\mathbf{e}_i^d$ onto $\tilde{\mathbf{x}}_i^d$ leads to a coefficient matrix that should be equal to the conjugate transpose of $K_{f,i}$. In other words, it holds that

$$K_{f,i} = -\left(\langle \mathbf{e}_i^d, \tilde{\mathbf{x}}_i^d \rangle \, \|\tilde{\mathbf{x}}_i^d\|^{-2}\right)^* = -\|\tilde{\mathbf{x}}_i^d\|^{-2} \langle \tilde{\mathbf{x}}_i^d, \mathbf{e}_i^d \rangle.$$

But from (15.2.8) we have that $\langle \tilde{\mathbf{x}}_i^d, \mathbf{e}_i^d \rangle = -H_i^* R_i^{-1}$. Combining this result with $\|\tilde{\mathbf{x}}_i^d\|^2 = P_{i|i}^{-1}$, we obtain

$$K_{f,i} = P_{i|i} H_i^* R_i^{-1}, \tag{15.2.11}$$

which is exactly the formula (9.5.29) derived algebraically in Sec. 9.5.

15.2.3 Application to State-Space Models

The result of Lemma 15.2.1 can also be used to characterize the dual of a linear state-space model. Thus assume that the entries $\{\mathbf{z}_i, \mathbf{y}_i\}$ of $\{\mathbf{z}, \mathbf{y}\}$ satisfy a forwards-time state-space model of the general form

$$\begin{cases} \mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{z}_i, \\ \mathbf{y}_i = H_i \mathbf{x}_i + D_i \mathbf{z}_i + \mathbf{v}_i, \quad i \geq 0, \end{cases} \tag{15.2.12}$$

for some matrices $\{F_i, G_i, H_i, D_i\}$, and with *zero* initial condition, $\mathbf{x}_0 = 0$.[1] The $\{\mathbf{z}_i, \mathbf{v}_i\}$ are uncorrelated white noise sequences with variances $\{Q_i, R_i\}$. Now, as we already know from the arguments of App. 5.A, every such state-space description induces a linear relation between the aggregate vectors $\{\mathbf{y}, \mathbf{z}, \mathbf{v}\}$ defined by

$$\mathbf{y} \triangleq \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}, \qquad \mathbf{z} \triangleq \text{col}\{\mathbf{z}_0, \ldots, \mathbf{z}_N\}, \qquad \mathbf{v} \triangleq \text{col}\{\mathbf{v}_0, \ldots, \mathbf{v}_N\}.$$

More specifically, it is easy to verify that (15.2.12) leads to

$$\mathbf{y} = A\mathbf{z} + \mathbf{v}, \tag{15.2.13}$$

where

$$R_z = \|\mathbf{z}\|^2 = \text{diag}\{Q_0, Q_1, \ldots, Q_N\}, \qquad R_v = \|\mathbf{v}\|^2 = \text{diag}\{R_0, R_1, \ldots, R_N\},$$

and where A is the block lower triangular matrix

$$A = \begin{bmatrix} D_0 & & & & \\ H_1 G_0 & D_1 & & & \\ H_2 F_1 G_0 & H_2 G_1 & D_2 & & \\ \vdots & & & \ddots & \\ H_N \Phi(N,1) G_0 & H_N \Phi(N,2) G_1 & H_N \Phi(N,3) G_2 & \ldots & H_N G_{N-1} \quad D_N \end{bmatrix},$$

where, as usual, $\Phi(N,i) = F_{N-1} F_{N-2} \ldots F_i$. It then follows from Lemma 15.2.1 that the (block) vectors $\{\mathbf{z}^d, \mathbf{y}^d\}$ that define the dual basis for $\mathcal{L}(\mathbf{z}, \mathbf{y})$ satisfy the relation

$$\mathbf{z}^d = -A^* \mathbf{y}^d + \mathbf{v}^d, \tag{15.2.14}$$

[1] Nonzero initial conditions are treated in Sec. 15.7.2, where we study backwards-time complementary state-space models. The zero-initial case is more immediate to handle, and the results in this case are sufficient for most of our discussions in the following sections. The model (15.2.12) is a variant, which turns out to be more useful in some later discussions.

for some $\mathbf{v}^d$; the exact (global) definitions of $\{\mathbf{y}^d, \mathbf{v}^d\}$ in terms of the original quantities $\{\mathbf{v}, \mathbf{z}\}$ are given in the statement of the lemma and are not of immediate interest here. What we would like to stress instead is that the individual entries of $\{\mathbf{z}^d, \mathbf{y}^d, \mathbf{v}^d\}$ can be shown to satisfy a state-space model similar to (15.2.12), albeit one that runs backwards in time. Moreover, this model coincides, as one might expect, with what is commonly defined as the dual state-space model in system theory.[2]

To see this, let us denote the individual entries of $\{\mathbf{z}^d, \mathbf{y}^d, \mathbf{v}^d\}$ in the model (15.2.14) by $\{\mathbf{z}_i^d, \mathbf{y}_i^d, \mathbf{v}_i^d\}$. If we now use the block-triangular expression for A, we can see (by direct verification) that they satisfy the backwards model

$$\begin{cases} \mathbf{x}_i^d &= F_i^* \mathbf{x}_{i+1}^d &- H_i^* \mathbf{y}_i^d, \\ \mathbf{z}_i^d &= G_i^* \mathbf{x}_{i+1}^d &- D_i^* \mathbf{y}_i^d + \mathbf{v}_i^d, \quad i \leq N, \end{cases} \tag{15.2.15}$$

with *zero* boundary condition, $\mathbf{x}_{N+1}^d = 0$. In other words, these state-space equations lead to the global linear model (15.2.14).

Comparing (15.2.12) and (15.2.15) we see that the dual model can be obtained from the original model by simply *reversing the direction of time* and by making the substitutions[3]

$$F_i \longleftrightarrow F_i^*, \quad H_i \longleftrightarrow G_i^*, \quad G_i \longleftrightarrow -H_i^*, \quad D_i \longleftrightarrow -D_i^*. \tag{15.2.16}$$

Another important, though obvious, consequence of the above construction is the following: the original model (15.2.12) with $\upsilon_i \equiv 0$ is a state-space realization for the lower triangular mapping A that takes $\mathbf{z}$ to $\mathbf{y}$ in (15.2.13). Likewise, if we set $\mathbf{v}_i^d$ equal to zero in the dual model (15.2.15), then the resulting equations correspond to a state-space realization for the upper triangular mapping $-A^*$ that takes $\mathbf{y}^d$ to $\mathbf{z}^d$ in (15.2.14). In other words, the original model realizes A while its dual realizes $-A^*$.

These are useful facts that will be invoked later in the chapter when we show how duality results can be used to solve certain quadratic control problems by reducing them to related estimation problems. In this process, we shall often need to determine a state-space representation for a mapping $-A^*$ given a representation for A (or vice versa). The above remark indicates that all we need to do is replace one model by its dual.

Similar conclusions of course hold if we start instead with a *backwards-time* model, say

$$\begin{cases} \mathbf{x}_i &= F_i \mathbf{x}_{i+1} + G_i \mathbf{z}_i, \quad \mathbf{x}_{N+1} = 0, \\ \mathbf{y}_i &= H_i \mathbf{x}_{i+1} + D_i \mathbf{z}_i + \mathbf{v}_i, \quad i \leq N. \end{cases} \tag{15.2.17}$$

The corresponding dual model will be *forwards-time*:

$$\begin{cases} \mathbf{x}_{i+1}^d &= F_i^* \mathbf{x}_i^d &- H_i^* \mathbf{y}_i^d, \quad \mathbf{x}_0^d = 0, \\ \mathbf{z}_i^d &= G_i^* \mathbf{x}_i^d &- D_i^* \mathbf{y}_i^d + \mathbf{v}_i^d, \quad i \geq 0. \end{cases} \tag{15.2.18}$$

[2] As mentioned in the previous footnote, we shall study nonzero initial conditions in Sec. 15.7.2, in which case only the trailing entries of $\mathbf{z}^d$ will satisfy a backwards-time state-space model.

[3] Observe the additional minus signs in the substitutions for G_i and D_i.

15.3 DUALITY AND EQUIVALENCE RELATIONSHIPS

In Sec. 15.1.4 we saw that the problem of projecting the dual vector $\mathbf{y}^d$ onto the dual vector $\mathbf{z}^d$ was dual to the problem of projecting $\mathbf{z}$ onto $\mathbf{y}$. In this section, we shall study the consequences of this observation for the solution of deterministic least-squares problems. The main conclusion will be that given a deterministic quadratic minimization problem, one can solve it in two ways: either by constructing an *equivalent* stochastic model where the solution is the same as the original problem (which is what we did in Sec. 3.5, or by constructing a *dual* stochastic problem where the solution is given by the negative conjugate transpose of the solution of the original problem.

15.3.1 Equivalent Stochastic and Deterministic Problems

At this point the reader may want to review Sec. 3.5, where we established the following equivalence result. Consider zero-mean random variables $\{\mathbf{y}, \mathbf{z}\}$ that are assumed to be related linearly as in (15.2.1)–(15.2.2),

$$\mathbf{y} = H\mathbf{z} + \mathbf{v}, \quad \langle \mathbf{z}, \mathbf{z} \rangle = R_z, \quad \langle \mathbf{v}, \mathbf{v} \rangle = R_v, \quad \langle \mathbf{z}, \mathbf{v} \rangle = 0, \tag{15.3.1}$$

with positive-definite Gramians $\{R_z, R_v\}$. The projection of $\mathbf{z}$ onto $\mathcal{L}\{\mathbf{y}\}$ is denoted by $\hat{\mathbf{z}}_{|\mathbf{y}}$ and is given by $\hat{\mathbf{z}}_{|\mathbf{y}} = K_o \mathbf{y}$, where K_o is the solution of

$$\min_K \|\mathbf{z} - K\mathbf{y}\|^2. \tag{15.3.2}$$

We already know that K_o is given by

$$K_o = R_{zy} R_y^{-1} = R_z H^* [R_v + H R_z H^*]^{-1}, \tag{15.3.3}$$

where, from (15.3.1), we used the fact that $R_{zy} = R_z H^*$ and $R_y = R_v + H R_z H^*$. Now in Lemma 15.2.1 we also showed that

$$K_o = (R_z^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1}, \tag{15.3.4}$$

which can also be obtained by direct algebraic manipulation of (15.3.3).

Now recall from the discussion in Sec. 3.5 that this alternative expression for K_o allowed us to conclude that an *equivalent* deterministic least-squares problem can be constructed such that its solution, say $\hat{z}$, will have the same expression as the solution $\hat{\mathbf{z}}_{|\mathbf{y}}$ of the stochastic problem (15.3.2). More specifically, we showed the following. Given a vector (of deterministic observations) y and a matrix H, consider the problem of determining a vector $\hat{z}$ that solves[4]

$$\min_z \left[z^* R_z^{-1} z + \|y - Hz\|^2_{R_v^{-1}} \right]. \tag{15.3.5}$$

Then its solution is given by

$$\hat{z} = (R_z^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1} y \triangleq K_o y. \tag{15.3.6}$$

[4] Recall that the notation $\|a\|_W^2$ stands for the weighted squared norm $a^* W a$, with $W > 0$.

In other words, the gain matrices K_o that are needed in the solutions of both problems (15.3.2) and (15.3.5) are identical. This means that whenever we solve a stochastic problem of the form (15.3.2) we are also solving a deterministic problem of the form (15.3.5), and vice versa. We refer to these problems as *equivalent* problems since their gain matrices K_o are identical. Actually, more can be said about the equivalent cost functions (15.3.2) and (15.3.5), which for reasons of space we defer to Prob. 15.3.

15.3.2 Dual Stochastic and Deterministic Problems

We saw in Lemma 15.2.1 that there is a dual problem to that of estimating $\mathbf{z}$ from $\mathbf{y}$. In particular, the lemma states that the gain matrix K_o in $\hat{\mathbf{z}}_{|\mathbf{y}} = K_o\mathbf{y}$ is also equal to the negative conjugate transpose of the matrix that projects $\mathbf{y}^d$ onto $\mathbf{z}^d$ (where $\{\mathbf{z}^d, \mathbf{y}^d\}$ is the dual basis to $\{\mathbf{z}, \mathbf{y}\}$). This suggests that corresponding to the stochastic and deterministic problems (15.3.2) and (15.3.5), we can also construct *dual* stochastic and deterministic problems whose solutions are related as above. We pursue these ideas in more detail below.

Returning to the assumed linear model (15.3.1), and using the construction of Lemma 15.2.1, we obtain that the dual basis $\{\mathbf{z}^d, \mathbf{y}^d\}$ also satisfies a linear model of the form

$$\mathbf{z}^d = -H^*\mathbf{y}^d + \mathbf{v}^d, \tag{15.3.7}$$

where $\{\mathbf{y}^d, \mathbf{v}^d\}$ are orthogonal with $\|\mathbf{y}^d\|^2 = R_v^{-1}$ and $\|\mathbf{v}^d\|^2 = R_z^{-1}$. The projection of $\mathbf{y}^d$ onto $\mathcal{L}\{\mathbf{z}^d\}$ now requires that we determine the coefficient matrix K_o^d that solves the optimization problem

$$\min_{K^d} \|\mathbf{y}^d - K^d\mathbf{z}^d\|^2. \tag{15.3.8}$$

We already know that K_o^d is given by

$$K_o^d = R_{y^d z^d}R_{z^d}^{-1} = -R_y^d H(R_v^d + H^*R_y^d H)^{-1} = -R_v^{-1}H(R_z^{-1} + H^*R_v^{-1}H)^{-1},$$

where, from (15.3.7), we used the fact that $R_{y^d z^d} = -R_y^d H$ and $R_{z^d} = R_v^d + H^*R_y^d H$. We also know from Lemma 15.1.2, or by comparing with (15.3.4), that

$$K_o^d = -K_o^*. \tag{15.3.9}$$

We thus say that the problems (15.3.2) and (15.3.8) are *dual* since the corresponding gain matrices are the negative conjugate transposes of each other.

Now, we can also introduce the deterministic problem that is *equivalent* to the new stochastic problem (15.3.8), and which would therefore lead to the same gain matrix as K_o^d in (15.3.9). By direct analogy with the equivalent problems (15.3.2) and (15.3.5), we conclude that the required equivalent problem is the following. Given a vector of (deterministic) observations y^d and a matrix $-H^*$, consider the problem of determining a vector $\hat{y}^d$ that solves

$$\min_{y^d} \left[y^{d*}R_v y^d + \|z^d + H^*y^d\|^2_{R_z} \right]. \tag{15.3.10}$$

Then its solution is given by

$$\hat{y}^d = -(R_v + HR_zH^*)^{-1}HR_z z^d \triangleq K_o^d z^d. \tag{15.3.11}$$

In other words, the gain matrices K_o^d that are needed in the solutions of both problems (15.3.8) and (15.3.10) are identical.

We therefore refer to these problems as *equivalent* problems. By the same token, we say that the deterministic problems (15.3.5) and (15.3.10) are the dual of one another, since their gain matrices are the negative conjugate transposes of each other. We also say that the deterministic problem (15.3.10) is dual to the original stochastic problem (15.3.2). [Again, more can be said about the cost functions that are associated with these dual and equivalent problems — see Prob. 15.4.]

15.3.3 Summary of Duality and Equivalence Results

The results of this section are summarized in Table 15.1, which collects the relationships between the two dual stochastic problems and their corresponding deterministic quadratic forms:

1. Vertical transitions [between (i) and (iii) and between (ii) and (iv)] correspond to going from the original bases to the dual bases, so that the solutions are duals (negative conjugate transposes) of each other.
2. Horizontal transitions [between (i) and (ii) and between (iii) and (iv)] correspond to going from matrix-valued to scalar-valued quadratic forms so that these solutions are again dual.
3. Diagonal transitions between (i) and (iv) and between (ii) and (iii) relate problems with the *same* solution which, following Sec. 3.5, we refer to as *equivalent* problems.

Let us explain how these four relations may be used to solve various problems. Suppose we are given an optimization problem of the form (iv), *i.e.*,

$$\min_z \left[z^*R_z^{-1}z + \|y - Hz\|^2_{R_v^{-1}} \right].$$

Then to obtain the solution we may proceed in either one of two ways.[5] We can construct an *equivalent* stochastic model of the form

$$\mathbf{y} = H\mathbf{z} + \mathbf{v},$$

with

$$\left\langle \begin{bmatrix} \mathbf{z} \\ \mathbf{v} \end{bmatrix}, \begin{bmatrix} \mathbf{z} \\ \mathbf{v} \end{bmatrix} \right\rangle = \begin{bmatrix} R_z & 0 \\ 0 & R_v \end{bmatrix}.$$

Then the projection of $\mathbf{z}$ onto $\mathbf{y}$ provides the solution to (iv). Equivalently, we can construct the *dual* linear stochastic model

$$\mathbf{z}^d = -H^*\mathbf{y}^d + \mathbf{v}^d,$$

with

$$\left\langle \begin{bmatrix} \mathbf{y}^d \\ \mathbf{v}^d \end{bmatrix}, \begin{bmatrix} \mathbf{y}^d \\ \mathbf{v}^d \end{bmatrix} \right\rangle = \begin{bmatrix} R_v^{-1} & 0 \\ 0 & R_z^{-1} \end{bmatrix}.$$

[5] A third option is to consider the dual deterministic problem, if it happens that its solution is already known. Our point here is to show how stochastic results can be applied to solve deterministic problems.

Table 15.1 Equivalences and dualities for linear models assuming positive-definite matrices $\{R_z, R_v\}$. The expressions for K_o in entries (i) and (iv) coincide. Likewise, the expressions for K_o^d in entries (ii) and (iii) coincide.

Stochastic problems	Deterministic problems
(i) **Given**: $\mathbf{y} = H\mathbf{z} + \mathbf{v}$ with $\langle \begin{bmatrix} \mathbf{z} \\ \mathbf{v} \end{bmatrix}, \begin{bmatrix} \mathbf{z} \\ \mathbf{v} \end{bmatrix} \rangle = \begin{bmatrix} R_z & 0 \\ 0 & R_v \end{bmatrix}$ **Solve**: $\min_{\hat{\mathbf{z}} \in \mathcal{L}\{\mathbf{y}\}} \|\mathbf{z} - \hat{\mathbf{z}}\|^2$ **Solution**: $\hat{\mathbf{z}} = K_o\mathbf{y}$ with $K_o = R_z H^*(R_v + HR_zH^*)^{-1}$ **Min. cost**: $(R_z^{-1} + H^*R_v^{-1}H)^{-1}$	(ii) $\{z^d, H, R_v, R_z\}$ $\min_{y^d} \left[y^{d*}R_v y^d + \|z^d + H^*y^d\|^2_{R_z} \right]$ $\hat{y}^d = K_o^d z^d$, with $K_o^d = -K_o^*$ $K_o^d = -(R_v + HR_zH^*)^{-1}HR_z$ $z^{d*}(R_z^{-1} + H^*R_v^{-1}H)^{-1}z^d$
(iii) **Given**: $\mathbf{z}^d = -H^*\mathbf{y}^d + \mathbf{v}^d$ with $\langle \begin{bmatrix} \mathbf{y}^d \\ \mathbf{v}^d \end{bmatrix}, \begin{bmatrix} \mathbf{y}^d \\ \mathbf{v}^d \end{bmatrix} \rangle = \begin{bmatrix} R_v^{-1} & 0 \\ 0 & R_z^{-1} \end{bmatrix}$ **Solve**: $\min_{\hat{\mathbf{y}}^d \in \mathcal{L}\{\mathbf{z}^d\}} \|\mathbf{y}^d - \hat{\mathbf{y}}^d\|^2$ **Solution**: $\hat{\mathbf{y}}^d = K_o^d\mathbf{z}^d$ with $K_o^d = -K_o^*$ $K_o^d = -R_v^{-1}H(R_z^{-1} + H^*R_v^{-1}H)^{-1}$ **Min. cost**: $(R_v + HR_zH^*)^{-1}$	(iv) $\{y, H, R_v, R_z\}$ $\min_z \left[z^*R_z^{-1}z + \|y - Hz\|^2_{R_v^{-1}} \right]$ $\hat{z} = K_o y$ $K_o = (R_z^{-1} + H^*R_v^{-1}H)^{-1}H^*R_v^{-1}$ $y^*(R_v + HR_zH^*)^{-1}y$

Now if we find the projection of $\mathbf{y}^d$ onto $\mathbf{z}^d$, we can use the negative conjugate transpose of the matrix that performs this projection to solve problem (iv).[6]

Remark 1. The derivation of the duality and equivalence results of Table 15.1 assumed positive-definite, and hence invertible, matrices $\{R_z, R_v\}$. This assumption can be relaxed without affecting some of the conclusions of this section. Consider, for example, entries (iii) and (iv) of the table, restated in the following way (a similar remark holds for entries (i) and (ii) of the table):

Entry (iii): Given the linear model

$$\mathbf{z}^d = -H^*\mathbf{y}^d + \mathbf{v}^d, \tag{15.3.12}$$

[6] Which approach to use depends upon the application at hand. It was shown in Sayed and Kailath (1994b) that for adaptive filtering problems it is most natural to use the equivalent stochastic model (i), whereas later in the chapter we shall see that for control problems it is more natural to use the dual stochastic model. Note that we say "natural" rather than "essential" — the formulas are simpler with the choices we suggest.

with

$$\langle \begin{bmatrix} \mathbf{y}^d \\ \mathbf{v}^d \end{bmatrix}, \begin{bmatrix} \mathbf{y}^d \\ \mathbf{v}^d \end{bmatrix} \rangle = \begin{bmatrix} R_{yd} & 0 \\ 0 & R_{vd} \end{bmatrix}, \qquad R_{vd} > 0, \quad R_{yd} \geq 0,$$

the projection of $\mathbf{y}^d$ onto $\mathcal{L}\{\mathbf{z}^d\}$ is easily seen to be given by $\hat{\mathbf{y}}^d = K_o^d\mathbf{z}^d$, where

$$K_o^d = -R_{yd}H\left[R_{vd} + H^*R_{yd}H\right]^{-1}. \tag{15.3.13}$$

Observe that we are now allowing for a possibly singular Gramian matrix R_{yd}. The positive-definiteness of R_{vd} is enough to guarantee the invertibility of the coefficient matrix $R_{vd} + H^*R_{yd}H$.

Entry (iv): Given matrices $\{R_{vd} > 0, R_{yd} \geq 0\}$, the solution of the quadratic minimization problem

$$\min_z \left[z^*R_{vd}z + \|y - Hz\|^2_{R_{yd}} \right], \tag{15.3.14}$$

is easily seen to be given by $\hat{z} = K_o y$, where

$$K_o = \left[R_{vd} + H^*R_{yd}H\right]^{-1}H^*R_{yd}. \tag{15.3.15}$$

Here again we are allowing for a possibly singular R_{yd}. However, comparing the expressions (15.3.13) and (15.3.15) for the solutions we see that we still have

$$K_o = -K_o^{*d},$$

so that the above two problems will still be the dual of one another. We shall encounter this case (singular R_{yd}) in studying certain control problems via duality (Secs. 15.3.5 and 15.3.6). [As mentioned above, the same conclusion holds for entries (i) and (ii) of the table when $R_z \geq 0$ and $R_v > 0$.] ◆

Remark 2. The results of Table 15.1 can be extended to the case where the variables $\{\mathbf{z}, \mathbf{y}\}$ are not assumed to be linearly related as in (15.3.1) — see Prob. 15.5. We may recall, however, the conclusion of Prob. 3.21 where we showed that the assumption of linear models is not restrictive for estimation problems involving second-order statistics. ◆

Remark 3. In closing this section we note that most of the results obtained so far (such as projections, dual bases, and duality relations) extend to linear spaces with indefinite inner product. There are, however, some key differences (projections may not always exist and be unique, and if so can only guarantee stationary points) that require further analysis; this perspective is pursued in Hassibi, Sayed, and Kailath (1999) to solve $\mathcal{H}_\infty$ problems in estimation and control. ◆

15.3.4 A Deterministic Optimization Problem via Duality

We shall use duality to develop a result similar to that of Sec. 10.7, where we solved, via equivalence, a certain deterministic problem (*cf.* (10.6.2)) that is usually solved via the Hamiltonian equations. As in Sec. 10.7, the solution to the new deterministic problem will be based on the Bryson-Frazier smoothing formulas of Sec. 10.2, but now for backwards-time state-space models rather than forwards-time models. We shall see that, apart from the fact that one argument uses equivalence while the other uses duality, the derivations here and in Sec. 10.7 share several common features.

We thus start with an optimization problem of the form

$$\min_{\{u_0,\ldots,u_N\}} \left[x_{N+1}^* P_{N+1}^d x_{N+1} + \sum_{i=0}^{N} (y_i - H_i x_i)^* R_i^d (y_i - H_i x_i) + \sum_{i=0}^{N} u_i^* Q_i^d u_i \right], \tag{15.3.16}$$

subject to the state-space constraint

$$x_{i+1} = F_i x_i + G_i u_i, \quad x_0 = 0, \tag{15.3.17}$$

and where $P_{N+1}^d \geq 0$, $R_i^d \geq 0$, and $Q_i^d > 0$ are given weighting matrices. Comparing (15.3.16) with (10.6.2) we see that the main difference is in replacing the term $x_0^* \Pi_0^{-1} x_0$, which is dependent on the initial state vector, by the term $x_{N+1}^* P_{N+1}^d x_{N+1}$, which is dependent on the terminal state vector. Two other differences that are included for convenience are that the weighting matrices $\{P_{N+1}^d, R_i^d, Q_i^d\}$ in (15.3.16) do not appear inverted, which enables us to allow for possibly singular matrices $\{P_{N+1}^d, R_i^d\}$ (as is often the case in formulations of quadratic control problems). Moreover, the state-space constraint (15.3.17) has a zero initial condition; the solution for nonzero x_0 can be obtained from the above, as we shall see in the next section when we study a tracking problem.

Our route will be to first verify that the cost function (15.3.16) can be written in the form (15.3.14), which can then be solved via a dual stochastic problem. For this purpose, we introduce the vectors $u = \text{col}\{u_0, u_1, \ldots, u_N\}$ and $s = \text{col}\{H_0 x_0, H_1 x_1, \ldots, H_N x_N\}$, as well as the block lower triangular matrix

$$B \triangleq \left[\begin{array}{cccc|c} \Phi(N+1,1)G_0 & \Phi(N+1,2)G_1 & \ldots & \Phi(N+1,N)G_{N-1} & G_N \\ \hline 0 & & & & \\ H_1 G_0 & & & & \\ H_2\Phi(2,1)G_0 & H_2 G_1 & & & \\ \vdots & & \ddots & & \\ H_N\Phi(N,1)G_0 & H_N\Phi(N,2)G_1 & \ldots & H_N G_{N-1} & 0 \end{array}\right].$$

Then it is easy to verify, by direct calculation, that

$$\begin{bmatrix} x_{N+1} \\ s \end{bmatrix} = Bu,$$

so that problem (15.3.16) can be equivalently rewritten as

$$\min_u \left[u^* \mathcal{Q}^d u + \left\| \begin{bmatrix} 0 \\ -y \end{bmatrix} + Bu \right\|_{\mathcal{W}^d}^2 \right], \tag{15.3.18}$$

where $y = \text{col}\{y_0, y_1, \ldots, y_N\}$ and

$$\mathcal{Q}^d \triangleq \text{diag}\{Q_0^d, \ldots, Q_N^d\}, \quad \mathcal{W}^d \triangleq \text{diag}\{P_{N+1}^d, R_0^d, \ldots, R_N^d\}. \tag{15.3.19}$$

Problem (15.3.18) has the same form as Prob. (15.3.14). Thus let K_o denote the optimal coefficient matrix that solves (15.3.18), *i.e.,*

$$\hat{u} = K_o \begin{bmatrix} 0 \\ -y \end{bmatrix}. \tag{15.3.20}$$

Then in view of the duality result of Remark 1 above, *the optimal matrix K_o can be determined by projecting $\mathbf{y}^d$ onto $\mathbf{z}^d$ in the dual stochastic model:*

$$\mathbf{z}^d = B^* \mathbf{y}^d + \mathbf{v}^d, \tag{15.3.21}$$

where $\{\mathbf{y}^d, \mathbf{v}^d\}$ are uncorrelated with variances

$$\|\mathbf{v}^d\|^2 = \mathcal{Q}^d, \quad \|\mathbf{y}^d\|^2 = \mathcal{W}^d. \tag{15.3.22}$$

That is, if $\hat{\mathbf{y}}^d = K_o^d \mathbf{z}^d$ then $K_o = -K_o^{d*}$. The reason we introduce the dual problem (15.3.21) is that we can now proceed to determine K_o^d (and hence K_o) rather immediately by employing standard state-space and innovations arguments.

To see this, let us partition the entries of $\mathbf{y}^d$ into uncorrelated entries as

$$\mathbf{y}^d \triangleq \text{col}\{\, \mathbf{x}_{N+1}^d,\ \mathbf{u}_0^d,\ \mathbf{u}_1^d,\ \ldots,\ \mathbf{u}_N^d \,\},$$

where $\|\mathbf{x}_{N+1}^d\|^2 = P_{N+1}^d$ and $\|\mathbf{u}_i^d\|^2 = R_i^d$. Let also $\mathbf{z}^d = \text{col}\{\mathbf{z}_0^d, \mathbf{z}_1^d, \ldots, \mathbf{z}_N^d\}$ denote the corresponding partitioning for $\mathbf{z}^d$, and similarly for $\mathbf{v}^d$. It can then be immediately verified from the definition of the matrix B above, and from the dual model (15.3.21), just as we did for the equivalent model in Sec. 10.7, that the entries of $\{\mathbf{z}^d, \mathbf{y}^d, \mathbf{v}^d\}$ satisfy the following backwards-time state-space model

$$\begin{cases} \mathbf{x}_i^d = F_i^* \mathbf{x}_{i+1}^d + H_i^* \mathbf{u}_i^d, \\ \mathbf{z}_i^d = G_i^* \mathbf{x}_{i+1}^d + \mathbf{v}_i^d, \end{cases} \tag{15.3.23}$$

where the $\{\mathbf{u}_i^d, \mathbf{v}_i^d\}$ are uncorrelated white noise sequences with variances $\{R_i^d, Q_i^d\}$; moreover, both are uncorrelated with $\mathbf{x}_{N+1}^d$.

We are thus reduced to the problem of estimating the $\{\mathbf{x}_{N+1}^d, \mathbf{u}_i^d\}$ from the observations $\{\mathbf{z}_0^d, \ldots, \mathbf{z}_N^d\}$. Denote the estimators by $\{\hat{\mathbf{x}}_{N+1|0}^d, \hat{\mathbf{u}}_{i|0}^d\}$. Then the coefficient matrix K_o^d that we seek is the matrix that performs the mapping

$$\left[\begin{array}{c} \hat{\mathbf{x}}_{N+1|0}^d \\ \hline \hat{\mathbf{u}}_{0|0}^d \\ \hat{\mathbf{u}}_{1|0}^d \\ \vdots \\ \hat{\mathbf{u}}_{N|0}^d \end{array}\right] = K_o^d \begin{bmatrix} \mathbf{z}_0^d \\ \mathbf{z}_1^d \\ \vdots \\ \mathbf{z}_N^d \end{bmatrix} \triangleq \left[\begin{array}{c} k_x^d \\ \hline K_u^d \end{array}\right] \mathbf{z}^d, \tag{15.3.24}$$

where we partitioned K_o^d into a row vector k_x^d and a matrix K_u^d; one corresponds to the map from $\mathbf{z}^d$ to $\hat{\mathbf{x}}_{N+1|0}^d$, while the other corresponds to the map from $\mathbf{z}^d$ to the $\{\hat{\mathbf{u}}_{i|0}^d\}$. The reason for this partitioning is the following.

Once we determine K_o^d then, by duality, we know that the desired solution $\hat{u}$ of (15.3.20) is given by

$$\hat{u} = -K_o^{d*}\begin{bmatrix} 0 \\ -y \end{bmatrix} = K_u^{d*}y, \tag{15.3.25}$$

so that we only need to determine K_u^d; the map from $\mathbf{z}^d$ to the $\{\hat{\mathbf{u}}_{i|0}^d\}$.

This map is immediately obtained since the smoothed estimators $\{\hat{\mathbf{x}}_{N+1|0}^d, \hat{\mathbf{u}}_{i|0}^d\}$ can be determined by simply developing Bryson-Frazier (BF)-like recursions for the backwards-time state-space model (15.3.23). The derivation is identical to what we did in Sec. 10.2, except for the reversed direction of time (see Prob. 10.16). We are thus led to the following recursions (analogous to those of Thm. 10.2.1)).

Theorem 15.3.1 (BF Recursions) *Consider the model (15.3.23) and let, as usual, $\{\hat{\mathbf{x}}_{i|i}^d, \hat{\mathbf{z}}_{i-1}^d, \hat{\mathbf{u}}_{j|i}^d\}$ denote the l.l.m.s. estimators of $\{\mathbf{x}_i^d, \mathbf{z}_{i-1}^d, \mathbf{u}_j^d\}$ given $\{\mathbf{z}_i^d, \ldots, \mathbf{z}_N^d\}$. Let also $\mathbf{e}_i^d = \mathbf{z}_i^d - \hat{\mathbf{z}}_i^d$ denote the backwards-time innovations of $\mathbf{z}_i^d$. Then the estimators $\{\hat{\mathbf{x}}_{i|0}^d, \hat{\mathbf{u}}_{i|0}^d\}$ can be determined as follows:*

$$\begin{cases} \hat{\mathbf{x}}_{i|0}^d &= \hat{\mathbf{x}}_{i|i}^d + P_{i|i}^d\lambda_{i|0}^d, \\ \lambda_{i+1|0}^d &= F_{d,i}^*\lambda_{i|0}^d + G_iR_{e,i}^{-d}\mathbf{e}_i^d, \quad \lambda_{0|0}^d = 0, \\ \hat{\mathbf{u}}_{i|0}^d &= R_i^dH_i\lambda_{i|0}^d, \end{cases} \tag{15.3.26}$$

where $F_{d,i}^* = F_i - G_iR_{e,i}^{-d}K_i^{d*}$, and $\{R_{e,i}^d, P_{i|i}^d\}$ are obtained from the Kalman filter recursions of Prob. 9.14, viz.,

$$\begin{cases} R_{e,i}^d &= G_i^*P_{i+1|i+1}^dG_i + Q_i^d, \qquad K_i^d = F_i^*P_{i+1|i+1}^dG_i, \\ P_{i|i}^d &= F_i^*P_{i+1|i+1}^dF_i + H_i^*R_i^dH_i - K_i^dR_{e,i}^{-d}K_i^{d*}, \quad P_{N+1|N+1}^d = P_{N+1}^d. \end{cases}$$

Moreover, the innovations $\mathbf{e}_i^d$ are computed via

$$\begin{cases} \hat{\mathbf{x}}_{i|i}^d &= F_{d,i}\hat{\mathbf{x}}_{i+1|i+1}^d + K_i^dR_{e,i}^{-d}\mathbf{z}_i^d, \quad \hat{\mathbf{x}}_{N+1|N+1}^d = 0, \\ \mathbf{e}_i^d &= -G_i^*\hat{\mathbf{x}}_{i+1|i+1}^d + \mathbf{z}_i^d. \end{cases} \tag{15.3.27}$$

■

Given the BF solution of Thm. 15.3.1, we now only need to identify the mapping K_u^d (from the $\{\mathbf{z}_i^d\}$ to the $\{\hat{\mathbf{u}}_{i|0}^d\}$) and its negative conjugate transpose, in order to find the desired solution $\hat{u}$ in (15.3.25). This mapping is fully described by the cascade of the two state-space models (15.3.26) and (15.3.27). The second model describes the mapping from the $\{\mathbf{z}_i^d\}$ to the $\{\mathbf{e}_i^d\}$, while the first model describes the mapping from the $\{\mathbf{e}_i^d\}$ to the $\{\hat{\mathbf{u}}_{i|0}^d\}$. We denote this cascade schematically as

$$\{\mathbf{z}_i^d\} \xrightarrow{(15.3.27)} \{\mathbf{e}_i^d\} \xrightarrow{(15.3.26)} \{\hat{\mathbf{u}}_{i|0}^d\}.$$

Observe further that both these state-space models have zero boundary states. Therefore, their dual models can be found immediately just as explained in Sec. 15.2.3. Thus let $\{\hat{u}_i\}$ denote the entries of the solution vector $\hat{u}$ in (15.3.25). Then, according to (15.3.20), we can find $\{\hat{u}_i\}$ by feeding the $\{-y_i\}$ into the following cascade:

$$\{-y_i\} \xrightarrow{\text{dual of (15.3.26)}} \{o_i\} \xrightarrow{\text{dual of (15.3.27)}} -\{\hat{u}_i\},$$

where the $\{o_i\}$ denote intermediate variables. This construction means the following. The $\{-y_i\}$ should be input vectors to the dual model of (15.3.26), which, as explained in Sec. 15.2.3, is readily seen to be of the form

$$\lambda_{i|0}^d = F_{d,i}\lambda_{i+1|0}^d + H_i^*R_i^{d*}y_i, \quad o_i = R_{e,i}^{-d}G_i^*\lambda_{i+1|0}^d, \quad \lambda_{N+1|0}^d = 0.$$

The output $\{o_i\}$ should then excite the dual of (15.3.26) to yield $\{-\hat{u}_i\}$, which is again seen to be

$$\hat{x}_{i+1|i+1}^d = F_{d,i}^*\hat{x}_{i|i}^d + G_io_i, \quad -\hat{u}_i = R_{e,i}^{-d}K_i^{d*}\hat{x}_{i|i}^d - o_i, \quad \hat{x}_{0|0} = 0.$$

Note further that if we substitute $F_{d,i}^* = (F_i - G_iR_{e,i}^{-d}K_i^{d*})$ into the above recursion for $\hat{x}_{i+1|i+1}^d$, we find that

$$\hat{x}_{i+1|i+1}^d = F_i\hat{x}_{i|i}^d + G_i\hat{u}_i, \quad \hat{x}_{0|0} = 0.$$

That is, the variables $\{\hat{x}_{i|i}^d\}$ coincide with the successive state vectors when the optimal signals $\{\hat{u}_i\}$ drive the original state equation (15.3.17). We are thus led to the following statement — compare with Thm. 10.7.1.

Theorem 15.3.2 (Deterministic Optimization Problem) *The solution $\{\hat{u}_i\}$ of the optimization problem (15.3.16) can be recursively computed as follows:*

$$\begin{cases} \lambda_{i|0}^d &= F_{d,i}\lambda_{i+1|0}^d + H_i^*R_i^{d*}y_i, \quad \lambda_{N+1|0}^d = 0, \\ x_{i+1} &= F_ix_i + G_i\hat{u}_i, \quad x_0 = 0, \\ \hat{u}_i &= -R_{e,i}^{-d}K_i^{d*}x_i + R_{e,i}^{-d}G_i^*\lambda_{i+1|0}^d. \end{cases} \tag{15.3.28}$$

■

15.3.5 Application to Linear Quadratic Tracking

In this section we illustrate how the result of Thm. 15.3.2 can be used to solve certain linear quadratic control problems.

Thus consider the time-variant state-space model

$$\begin{cases} x_{i+1} &= F_ix_i + G_iu_i, \quad 0 \le i \le N, \\ s_i &= H_ix_i, \end{cases} \tag{15.3.29}$$

where the $F_i \in \mathcal{C}^{n\times n}$, $G_i \in \mathcal{C}^{n\times m}$, and $H_i \in \mathcal{C}^{q\times n}$ are known matrices. It is also assumed that the initial condition x_0 of (15.3.29) is known. The signal $\{u_i\}$ is referred to as the *control input* and is used to influence the *regulated output* signal, $\{s_i\}$. Roughly speaking, in the tracking problem, the goal is to design the control input in order to keep the regulated signal $\{s_i\}$ as "close" as possible to a given *reference signal*

$$\{y_i\}, \quad 0 \le i \le N, \tag{15.3.30}$$

where, of course, "close" is meant in a certain sense.

It is quite conceivable that if the choice of the $\{u_i\}$ were not constrained in any way, then it should be possible to make the regulated signal $\{s_i\}$ arbitrarily close to $\{y_i\}$. As this may (and typically will) require an arbitrarily large control signal, in order to guarantee the cost-effectiveness of the final control strategy, it is necessary to try to keep the control signal small as well. Therefore we are left with the twofold objective of designing a control law that simultaneously guarantees that the *tracking error* signal, $\{y_i - s_i\}$, and the control signal, $\{u_i\}$, be small.

In the so-called linear quadratic approach, the aforementioned twofold objective is met by choosing a control signal $\{u_i\}$ that solves the following minimization problem:

$$\min_{\{u_0,u_1,\ldots,u_N\}} \left[x_{N+1}^* P_{N+1}^d x_{N+1} + \sum_{i=0}^{N} u_i^* Q_i^d u_i + \sum_{i=0}^{N} (y_i - s_i)^* R_i^d (y_i - s_i) \right], \tag{15.3.31}$$

subject to the state-space constraints (15.3.29), where

$$P_{N+1}^d \ge 0, \quad Q_i^d > 0, \quad R_i^d \ge 0,$$

are given weighting matrices that penalize the final state, the control inputs $\{u_i\}$, and the tracking errors $\{y_i - s_i\}$, respectively. The cost function (15.3.31) is similar to (15.3.16) except for a possible nonzero initial condition x_0 in the state-space constraints (15.3.29).[7] As before, we first rewrite (15.3.31) in a form similar to (15.3.18). For this, we define the observability map

$$\mathcal{O} \triangleq \begin{bmatrix} H_0 \\ H_1\Phi(1,0) \\ \vdots \\ H_N\Phi(N,0) \end{bmatrix}, \tag{15.3.32}$$

[7] The simplest case of such design problems was first encountered and solved in the calculus of variations by Legendre (1810); in the control context, it was perhaps first (re)introduced by Bellman (1957). The solution was shown to have a feedback form, with the state feedback gain vector determined by solving a scalar Riccati differential equation. In 1958, Kalman and Koepcke formulated and solved the general discrete-time quadratic regulator problem (which corresponds to $y_i \equiv 0$) and showed that a feedback solution could be obtained by solving a backwards matrix Riccati recursion. Moreover, by later comparing this recursion to the one Kalman (1960a) had independently obtained for the state-space estimation problem, Kalman noticed a simple duality between the two solutions, a now widely cited result. However, to fully exploit the power of duality, one should be able to use it to avoid the independent solution of each problem. This is the approach we shall take in this section.

where $\Phi(i,j) = F_{i-1}\ldots F_j$ for $i > j$, and $\Phi(i,i) = I$. Then, with the same matrices $\{B, \mathcal{Q}^d, \mathcal{W}^d\}$, and with the same vectors $\{y, u\}$, defined before and after (15.3.18), it can be verified by direct calculation that

$$\begin{bmatrix} x_{N+1} \\ s \end{bmatrix} = \begin{bmatrix} \Phi(N+1,0) \\ \mathcal{O} \end{bmatrix} x_0 + Bu,$$

so that the cost function (15.3.31) can be rewritten as

$$\min_u \left[u^*\mathcal{Q}^d u + \left\| \left(\begin{bmatrix} 0 \\ -y \end{bmatrix} + \begin{bmatrix} \Phi(N+1,0) \\ \mathcal{O} \end{bmatrix} x_0 \right) + Bu \right\|_{\mathcal{W}^d}^2 \right]. \tag{15.3.33}$$

Comparing the above cost with (15.3.18) we see that the only difference is in the definition of the data vector; it was col$\{0, -y\}$ before while now it contains an additional term that is due to the nonzero initial condition x_0. Still the solution to the above problem can be obtained from the duality results of the previous section, as we now explain.

We again consider the backwards-time state-space model (15.3.23) and determine the matrix K_o^d that defines the mapping (15.3.24) from the $\{\mathbf{z}_i^d\}$ to the $\{\hat{\mathbf{x}}_{N+1|0}^d, \hat{\mathbf{u}}_{i|0}^d\}$. Then, by duality, the new solution vector $\hat{u}$ is given by

$$\hat{u} = -K_o^{d*}\left(\begin{bmatrix} 0 \\ -y \end{bmatrix} + \begin{bmatrix} \Phi(N+1,0) \\ \mathcal{O} \end{bmatrix} x_0 \right) = K_u^{d*} y - K_o^{d*} \begin{bmatrix} \Phi(N+1,0) \\ \mathcal{O} \end{bmatrix} x_0.$$

This expression shows that $\hat{u}$ is linear in both the reference vector y and the initial condition x_0. We already know from the result of the previous section that the entries of the first term, $K_u^{d*}y$, can be obtained by means of the BF recursions (15.3.28) when $x_0 = 0$. We now argue that the effect of the second term is simply to change the initial condition of the state equation in (15.3.28) from zero to x_0, so that the solution to the tracking problem will be given by the following statement.

Theorem 15.3.3 (Solution of Tracking Problem) *The solutions $\{\hat{u}_i\}$ of problem (15.3.31) can be obtained as follows:*

$$\begin{cases} \lambda_{i|0}^d = F_{d,i}\lambda_{i+1|0}^d + H_i^* R_i^{d*} y_i, & \lambda_{N+1|0}^d = 0, \\ x_{i+1} = F_i x_i + G_i \hat{u}_i, & x_0, \\ \hat{u}_i = -R_{e,i}^{-d} K_i^{d*} x_i + R_{e,i}^{-d} G_i^* \lambda_{i+1|0}^d, \end{cases} \tag{15.3.34}$$

where $\{K_i^d, R_{e,i}^d, P_{i|i}^d\}$ are as in Thm. 15.3.1. ∎

To justify the change in the initial condition from zero to x_0 in (15.3.34), let us consider the term

$$-K_o^{d*} \begin{bmatrix} \Phi(N+1,0) \\ \mathcal{O} \end{bmatrix} x_0, \tag{15.3.35}$$

which appears in the expression for $\hat{u}$ prior to the statement of the theorem. We simply need to show that the matrix multiplying x_0 is equal to the map from the initial state x_0 to the $\{\hat{u}_i\}$ in the last two equations of (15.3.34). That is, we need to verify that

$$-K_o^{d*}\begin{bmatrix}\Phi(N+1,0)\\ \mathcal{O}\end{bmatrix} = -\operatorname{col}\{R_{e,0}^{-d}K_0^{d*},\ldots,R_{e,N}^{-d}K_N^{d*}\Psi_d^*(0,N)\},$$

where $\Psi_d(i,i)=I$, and $\Psi_d(i,j)=F_{d,i}F_{d,i-1}\ldots F_{d,j-1}$ for $j>i$. To see this, introduce the negative conjugate transpose of the matrix multiplying x_0,

$$K_x^d \triangleq \begin{bmatrix}\Phi^*(N+1,0) & \mathcal{O}^*\end{bmatrix} K_o^d.$$

Now since, by definition, K_o^d is the gain matrix that estimates $\mathbf{y}^d$ from $\mathbf{z}^d$, and since from the state-space model (15.3.23),

$$\mathbf{x}_0^d = \begin{bmatrix}\Phi^*(N+1,0) & \mathcal{O}^*\end{bmatrix}\mathbf{y}^d = \Phi^*(N+1,0)\mathbf{x}_{N+1}^d + \mathcal{O}^*\mathbf{u}^d,$$

we conclude that the matrix K_o^x so defined is simply the gain matrix that estimates the initial state vector $\mathbf{x}_0^d$ from the observation vector $\mathbf{z}^d$. But it follows from the estimator recursion (15.3.27) that

$$\hat{\mathbf{x}}_{0|0}^d = \begin{bmatrix}K_0^dR_{e,0}^{-d} & \Psi_d(0,1)K_1^dR_{e,1}^{-d} & \ldots & \Psi_d(0,N)K_N^dR_{e,N}^{-d}\end{bmatrix}\mathbf{z}^d \triangleq K_x^d\,\mathbf{z}^d,$$

which establishes our claim regarding the value of K_x^d. This means that all we need to do in order to incorporate the effect of the initial condition x_0, is to initialize the recursion for x_i in (15.3.28) with x_0, which is what we did in the statement of Thm. 15.3.3.

Finally, it is worth noting that the solution $\{\hat{u}_i\}$ in Thm. 15.3.3 for the tracking problem actually consists of two components. One component is dependent on the actual state vector, x_i, which is obtained by propagating the *forwards-time* state equation as in (15.3.34). The second component is dependent on the variable $\lambda_{i+1|0}^d$, which is obtained by running a *backwards-time* recursion with boundary condition $\lambda_{N+1|0}^d=0$.

15.3.6 Application to Linear Quadratic Regulation

The linear quadratic regulator (LQR) problem is a special case of the linear quadratic tracking problem where the reference signal is identically zero, $y_i=0$ for all i. In this case, the first equation of the tracking solution (15.3.34) shows that $\lambda_{i|0}^d$ is also identically zero for all i, so that the solution takes the form of a simple state-feedback law:

$$\hat{u}_i = -R_{e,i}^{-d}K_i^{d*}x_i,\quad i=0,\ldots,N. \tag{15.3.36}$$

Moreover, the corresponding minimum cost can be verified to be equal to $x_0^*P_{0|0}^dx_0$ — see Prob. 15.9. Observe that now the optimal control sequence is fully described by a state-feedback law; the variable $\lambda_{i+1|0}^d$ disappears.

Remark 4 [Duality with Estimation]. Note that if we apply the transformations given below to the solution of the above LQR problem,

$$\begin{array}{lll} F_i^* \to F_i & P_{i+1|i+1}^d \to P_i & K_i^dR_{e,i}^{-d} \to K_{p,i}\\ H_i^* \to -G_i & R_i^d \to Q_i & \text{backward time} \to \text{forward time}\\ G_i^* \to H_i & Q_i^d \to R_i & \end{array}$$

then we recover the gain vector $K_{p,i}$ and Riccati recursion for the Kalman filter solution corresponding to the standard forwards-time state-space model (*cf.* the statement of Thm. 9.2.1 with $S_i=0$).

The relation between this standard Kalman filter solution and the recursions in the LQR problem is what led Kalman (1960a) to the observation that the solutions of the state-space estimation problem and the LQR control problem are dual to one another. In the above, we have obtained the same conclusion without independently solving each problem. In fact, we also solved the more general linear quadratic tracking problem by noting its duality to the stochastic smoothing problem.[8] ◆

15.4 DUALITY UNDER CAUSALITY CONSTRAINTS

In the previous sections, when considering the projection of one (standard or dual) variable onto another, and studying their various equivalent and dual problems, we did not assume any casuality constraints. When causality is a restriction, *i.e.*, when we are forced to estimate one set of variables given another in a causal fashion, then certain further issues arise, as we now proceed to explain.

15.4.1 Causal Estimation

Consider again random variables $\{\mathbf{y},\mathbf{z}\}$ satisfying

$$\mathbf{y}=H\mathbf{z}+\mathbf{v},\quad \|\mathbf{z}\|^2=R_z\geq 0,\quad \|\mathbf{v}\|^2=R_v>0,\quad \langle\mathbf{z},\mathbf{v}\rangle=0. \tag{15.4.1}$$

Then, according to entries (i) and (ii) of Table 15.1 (and according to Remark 1), the optimal coefficient matrix K_o that estimates $\mathbf{z}$ from $\mathbf{y}$, say $\hat{\mathbf{z}}=K_o\mathbf{y}$, can be obtained by solving the dual deterministic problem[9]

$$\min_{y^d}\left[\ y^{d*}R_vy^d+\left\|z^d+H^*y^d\right\|_{R_z}^2\ \right]. \tag{15.4.2}$$

That is, if $\hat{y}^d=K_o^dz^d$, then $K_o=-K_o^{d*}$.

[8] In Sayed and Kailath (1994b), a deterministic adaptive filtering problem was solved by going to an *equivalent* stochastic problem rather than a *dual* stochastic problem, as done here for a deterministic control problem. Of course, it is also possible to attack the adaptive filtering problem via a dual stochastic model, and the LQR problem via an equivalent stochastic one. The reason for the particular choices, both here and in Sayed and Kailath (1994b), is that they lead to simpler and better known formulas; for example, in the control case, we would first get recursions involving P_i^{-1}.

[9] We should remind the reader that when we write $\|\mathbf{z}\|^2$, with a boldface letter $\mathbf{z}$, we mean the matrix variance of the random variable $\mathbf{z}$, *i.e.*, $\|\mathbf{z}\|^2=E\mathbf{z}\mathbf{z}^*$. To refer to the scalar $E\mathbf{z}^*\mathbf{z}$ we shall instead write $\operatorname{Tr}\|\mathbf{z}\|^2$, which is the trace of the variance matrix. On the other hand, when we write $\|z\|^2$, with a normal font letter z, we mean the Euclidean norm of the column vector z, *i.e.*, $\|z\|^2=z^*z$; a scalar.

Now let $\{\mathbf{z}_i, \mathbf{y}_i\}$ denote the individual entries of $\{\mathbf{z}, \mathbf{y}\}$. The matrix K_o is the one that minimizes the error variance matrix,

$$\min_K \ \|\mathbf{z} - K\mathbf{y}\|^2 \quad \Longrightarrow \quad K_o, \tag{15.4.3}$$

i.e., $K_o = R_{zy}R_y^{-1}$. However, in causal estimation, *i.e.*, when, for any i, the estimator of $\mathbf{z}_i$ is only allowed to be a linear combination of $\{\mathbf{y}_0, \ldots, \mathbf{y}_i\}$, the above solution K_o cannot be used since the optimal coefficient matrix, say K_f, must now be *lower* triangular, *i.e.*,

$$\begin{bmatrix} \hat{\mathbf{z}}_{0|0} \\ \hat{\mathbf{z}}_{1|1} \\ \hat{\mathbf{z}}_{2|2} \\ \vdots \\ \hat{\mathbf{z}}_{m|m} \end{bmatrix} = \begin{bmatrix} \times & & & & \\ \times & \times & & & \\ \times & \times & \times & & \\ \vdots & & & \ddots & \\ \times & \times & \times & \times & \times \end{bmatrix} \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} \triangleq K_f\ \mathbf{y}. \tag{15.4.4}$$

We studied such causal estimation problems in Secs. 4.1.2–4.1.3, where we solved them by using the Wiener-Hopf technique. In particular, from Lemma 4.1.1 we know that K_f is given by

$$K_f = \left\{R_{zy}L^{-*}R_e^{-1}\right\}_{\text{lower}} L^{-1}, \tag{15.4.5}$$

where LR_eL^* is the lower-diagonal-upper triangular factorization of the Gramian matrix R_y, with L having unit diagonal entries; moreover, the notation $\{\cdot\}_{\text{lower}}$ denotes the lower triangular part of its matrix argument. For the model (15.4.1), we have $R_y = HR_zH^* + R_v$ and $R_{zy} = R_zH^*$.

Recall further from Sec. 4.1.3 that the matrix K_f in (15.4.5) is the optimal solution of the following problem:

$$\min_{\text{lower trian. } K} \ \text{Tr}\ \|\mathbf{z} - K\mathbf{y}\|^2 \quad \Longrightarrow \quad \hat{\mathbf{z}} = K_f\ \mathbf{y}. \tag{15.4.6}$$

Moreover, as shown in Prob. 4.3, the matrix K_f also minimizes the cost function

$$E(\mathbf{z} - K\mathbf{y})^*W(\mathbf{z} - K\mathbf{y}), \tag{15.4.7}$$

over all lower triangular matrices K, and for any nonnegative-definite matrix W. We shall rewrite (15.4.7) more compactly as follows:[10]

$$\text{Tr}\ \|\mathbf{z} - K\mathbf{y}\|_W^2.$$

Therefore, we have that K_f also solves the following problem:

$$\min_{\text{lower trian. } K} \ \text{Tr}\ \|\mathbf{z} - K\mathbf{y}\|_W^2 \quad \Longrightarrow \quad \hat{\mathbf{z}} = K_f\ \mathbf{y}, \tag{15.4.8}$$

for any $W \geq 0$. We shall use these facts now.

[10] Recall that the notation $\|\mathbf{z}\|_R^2$ stands for the variance matrix $E\mathbf{z}R\mathbf{z}^*$ so that its trace is equal to $E\mathbf{z}^*R\mathbf{z}$.

15.4.2 Anticausal Dual Problem

We now describe a dual to the causal estimation problem (15.4.4) (or (15.4.6)). The dual problem will not be deterministic (as is the relation between (15.4.1) and (15.4.2)), but rather stochastic (as is the case between the stochastic problems (i) and (iii) of Table 15.1).

Thus let $\mathbf{a}$ be a zero-mean random variable with a nonnegative-definite covariance matrix $R_a \geq 0$ and with individual entries $\{\mathbf{a}_i\}$. Let also $\mathbf{b}$ be a random variable that is generated from $\mathbf{a}$ *anticausally*, say

$$\mathbf{b} = K^d\ \mathbf{a} = \begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \ddots & \vdots \\ & & & \times \end{bmatrix} \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}.$$

We now pose the problem of determining an optimal K^d by solving

$$\min_{\begin{pmatrix}\text{upper trian. } K^d \\ \text{with } \mathbf{b} = K^d\mathbf{a}\end{pmatrix}} E\left[\ \mathbf{b}^*R_v\mathbf{b} \ + \ (\mathbf{a} + H^*\mathbf{b})^*R_z(\mathbf{a} + H^*\mathbf{b})\ \right].$$

Except for the stochastic nature of the variables $\{\mathbf{a}, \mathbf{b}\}$, this cost function is similar in form to that of the deterministic problem (15.4.2).

In the sequel we shall use a more compact notation to refer to such cost functions. More specifically, we shall rewrite the above cost as

$$\min_{\begin{pmatrix}\text{upper trian. } K^d \\ \text{with } \mathbf{b} = K^d\mathbf{a}\end{pmatrix}} \text{Tr}\left(\ \|\mathbf{b}\|_{R_v}^2 \ + \ \|\mathbf{a} + H^*\mathbf{b}\|_{R_z}^2\ \right). \tag{15.4.9}$$

We shall denote the optimal solution by $\hat{\mathbf{b}} = K_f^d\ \mathbf{a}$.

Let us now verify that the solution K_f^d to the above problem is indeed the dual to K_f in (15.4.6), *i.e.*, $K_f^d = -K_f^*$. This fact can be established rather easily, without even relying on knowledge of the explicit solution K_f in (15.4.5). To see this, let $J(K^d)$ denote the cost function in (15.4.9),

$$J(K^d) \triangleq \text{Tr}\left(\ \|\mathbf{b}\|_{R_v}^2 \ + \ \|\mathbf{a} + H^*\mathbf{b}\|_{R_z}^2\right).$$

Substituting $\mathbf{b}$ for $K^d\mathbf{a}$ and expanding, we obtain that

$$\begin{aligned} J(K^d) &= E\mathbf{a}^*\left(R_z + R_{zy}K^d + K^{d*}R_{yz} + K^{d*}R_yK^d\right)\mathbf{a}, \\ &= \text{Tr}\left[\left(R_z + R_{zy}K^d + K^{d*}R_{yz} + K^{d*}R_yK^d\right)R_a\right], \\ &= \text{Tr}\left(\|\mathbf{z} + K^{d*}\mathbf{y}\|^2\ R_a\right), \\ &= \text{Tr}\left(\|\mathbf{z} + K^{d*}\mathbf{y}\|_{R_a}^2\right). \end{aligned} \tag{15.4.10}$$

Comparing with (15.4.8), and since $R_a \geq 0$, we see that the choice $K_f^{d*} = -K_f$ minimizes the above cost or, equivalently,

$$K_f^d = -K_f^*. \tag{15.4.11}$$

In summary, we have shown that problems (15.4.6) and (15.4.9) admit dual solutions.

15.4.3 Anticausal Estimation and Causal Duality

In the above discussion we started with a causal estimation problem (15.4.6) and exhibited a dual anticausal problem to it (*viz.*, (15.4.9)). We can also start with an anticausal estimation problem and end up with a dual causal problem. Arguments similar to the above will show that the following conclusion holds.

Consider random variables $\{\mathbf{z}^d, \mathbf{y}^d\}$ satisfying

$$\mathbf{z}^d = -H^*\mathbf{y}^d + \mathbf{v}^d, \quad \|\mathbf{y}^d\|^2 = R_{y^d} \geq 0, \quad \|\mathbf{v}^d\|^2 = R_{v^d} > 0, \quad \langle \mathbf{y}^d, \mathbf{v}^d \rangle = 0,$$

and let K_a^d be the *upper* triangular matrix that estimates $\mathbf{y}^d$ anticausally from $\mathbf{z}^d$, *viz.*, $\hat{\mathbf{y}}^d = K_a^d \mathbf{z}^d$. By repeating the argument that led to Lemma 4.1.1 we can again establish that

$$K_a^d = \{R_{y^d z^d} U^{-d*} R_e^{-d}\}_{\text{upper}} U^{-d}, \tag{15.4.12}$$

where $U^d R_e^d U^{d*}$ is the upper-diagonal-lower triangular factorization of the Gramian matrix R_{z^d}, with U^d having unit diagonal entries; moreover, the notation $\{\cdot\}_{\text{upper}}$ denotes the upper triangular part of its matrix argument. [For the above linear model, we have $R_{z^d} = H^* R_{y^d} H + R_{v^d}$ and $R_{y^d z^d} = -R_{y^d} H$.] The matrix K_a^d so determined is the solution of the following optimization problem:

$$\min_{\text{upper trian. } K^d} \quad \text{Tr } \|\mathbf{y}^d - K^d \mathbf{z}^d\|^2 \implies \hat{\mathbf{y}}^d = K_a^d \mathbf{z}^d. \tag{15.4.13}$$

Now let $\mathbf{b}$ be a zero-mean random variable with a covariance matrix $R_b \geq 0$. Let also $\mathbf{a}$ be a random variable that is generated from $\mathbf{b}$ *causally*. That is, $\mathbf{a} = K\,\mathbf{b}$ for some *lower* triangular matrix K. We can then pose the problem of determining K optimally by solving

$$\min_{\begin{pmatrix}\text{lower trian. } K \\ \text{with } \mathbf{a} = K\mathbf{b}\end{pmatrix}} \text{Tr}\left(\|\mathbf{a}\|^2_{R_{v^d}} + \|\mathbf{b} - H\mathbf{a}\|^2_{R_{y^d}} \right) \implies \hat{\mathbf{a}} = K_a \mathbf{b}. \tag{15.4.14}$$

The optimal coefficient matrix is denoted by K_a. An argument similar to the one presented in the previous section will then show that problems (15.4.13) and (15.4.14) are dual to one another, *i.e.*,

$$K_a = -K_a^{d*}. \tag{15.4.15}$$

Remark 5 [More General Dualities]. More generally, it is straightforward to verify from the above arguments that the two problems listed in each item below are dual to one another (observe that we are adding a matrix M in the statement of the problems).

(i) Given the linear model

$$\mathbf{y} = H\mathbf{z} + \mathbf{v}, \quad \|\mathbf{z}\|^2 = R_z \geq 0, \quad \|\mathbf{v}\|^2 = R_v > 0, \quad \langle \mathbf{z}, \mathbf{v} \rangle = 0,$$

then the following problems are dual (*i.e.*, $K_f = -K_f^{d*}$):

$$\begin{cases} \min\limits_{\text{lower trian. } K} \text{Tr } \|M\mathbf{z} - K\mathbf{y}\|^2 \implies K_f \\ \min\limits_{\begin{pmatrix}\text{upper trian. } K^d \\ \text{with } \mathbf{b} = K^d\mathbf{a}\end{pmatrix}} \text{Tr}\left(\|\mathbf{b}\|^2_{R_v} + \|M^*\mathbf{a} + H^*\mathbf{b}\|^2_{R_z} \right) \implies \hat{\mathbf{b}} = K_f^d \mathbf{a} \end{cases}$$

for any matrix M. In the top problem we are estimating $M\mathbf{z}$ causally from $\mathbf{y}$.

(ii) Given the linear model

$$\mathbf{z}^d = -H^*\mathbf{y}^d + \mathbf{v}^d, \quad \|\mathbf{y}^d\|^2 = R_{y^d} \geq 0, \quad \|\mathbf{v}^d\|^2 = R_{v^d} > 0, \quad \langle \mathbf{y}^d, \mathbf{v}^d \rangle = 0,$$

then the following problems are dual (*i.e.*, $K_a = -K_a^{d*}$):

$$\begin{cases} \min\limits_{\text{upper trian. } K^d} \text{Tr } \|M^*\mathbf{y}^d - K^d\mathbf{z}^d\|^2 \implies K_a^d \\ \min\limits_{\begin{pmatrix}\text{lower trian. } K \\ \text{with } \mathbf{a} = K\mathbf{b}\end{pmatrix}} \text{Tr}\left(\|\mathbf{a}\|^2_{R_{v^d}} + \|M\mathbf{b} - H\mathbf{a}\|^2_{R_{y^d}} \right) \implies \hat{\mathbf{a}} = K_a \mathbf{b} \end{cases}$$

for any matrix M. In the top problem we are estimating $M^*\mathbf{y}^d$ anticausally from $\mathbf{z}^d$.

◆

15.4.4 Application to Stochastic Quadratic Control

We now present an application of the duality under causality results to a stochastic quadratic control problem. Thus recall that in the linear quadratic tracking and LQR problems described in Secs. 15.3.5 and 15.3.6, we assumed that the control input u_i is the only input driving the system (or the state equation) in (15.3.29). In many practical problems the system is also driven by an exogenous input, referred to as the driving disturbance.

In such cases we may write the state equation for the system as

$$x_{i+1} = F_i x_i + G_{1,i} w_i + G_{2,i} u_i, \qquad i = 0, \ldots N, \tag{15.4.16}$$

where the $\{u_i\}$ are the control inputs and the $\{w_i\}$ are the exogenous inputs. The $\{w_i\}$ may be interpreted as process noise or driving disturbance. Moreover, suppose we are given a linear combination of the states, $s_i = H_i x_i$, that we intend to regulate to a known reference signal $\{y_i\}$. Here for simplicity we take $y_i = 0$ for all i.

There are various ways of representing the driving disturbance $\{w_i\}$, but following the general approach of this book we shall model all such uncertainty as a random variable (or random process). Therefore we shall henceforth consider the model

$$\begin{cases} \mathbf{x}_{i+1} &= F_i\mathbf{x}_i + G_{1,i}\mathbf{w}_i + G_{2,i}\mathbf{u}_i, \\ \mathbf{s}_i &= H_i\mathbf{x}_i, \end{cases} \qquad i = 0, \ldots N, \tag{15.4.17}$$

where $\mathbf{x}_0$ and the $\{\mathbf{w}_i\}_{i=0}^N$ are now zero-mean random variables with known covariance matrices given by

$$\left\langle \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{w}_i \end{bmatrix}, \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{w}_j \end{bmatrix} \right\rangle = \begin{bmatrix} \Pi_0 & 0 \\ 0 & Q_i^w\delta_{ij} \end{bmatrix}, \quad \Pi_0 \geq 0, \quad Q_i^w \geq 0. \tag{15.4.18}$$

Note further that, since the initial condition and driving disturbance are random, so too will be the state $\{\mathbf{x}_i\}$ and the regulated outputs $\{\mathbf{s}_i\}$. Moreover, once the state and driving disturbance are random variables, it is clear that it is not possible to control the system with a deterministic control input, which is why we have taken $\{\mathbf{u}_i\}$ as random in (15.4.17). The explicit form of $\mathbf{u}_i$ will depend on the information available to the controller, as explained below.

Introduce the column vector $\mathbf{u} = \text{col}\{\mathbf{u}_0, \ldots, \mathbf{u}_N\}$. Similarly for $\mathbf{w}$ and $\mathbf{s}$. In the so-called causal full information stochastic control problem we assume that the control signal $\mathbf{u}_i$ has access to all current and past values of the disturbances, $\{\mathbf{x}_0, \mathbf{w}_0, \ldots, \mathbf{w}_i\}$ (hence the name *full information*). More specifically, we shall require the control vector $\mathbf{u}$ to be generated (causally) via a relation of the form

$$\mathbf{u} = K\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{w} \end{bmatrix}, \tag{15.4.19}$$

for some lower triangular matrix K. The optimal choice for K, say K_a, is determined by solving

$$\min_{\begin{pmatrix}\text{lower trian.}\\ K\end{pmatrix}} E\left[\mathbf{x}_{N+1}^* P_{N+1}^d \mathbf{x}_{N+1} + \sum_{i=0}^N \mathbf{u}_i^* Q_i^d \mathbf{u}_i + \sum_{i=0}^N \mathbf{s}_i^* R_i^d \mathbf{s}_i\right], \tag{15.4.20}$$

where the expectation is over the uncorrelated random variables, $\{\mathbf{x}_0, \mathbf{w}_0, \ldots, \mathbf{w}_N\}$.

Define matrices B_1 and B_2 similar to the matrix B prior to (15.3.18), with the $\{G_i\}$ in B replaced by $\{G_{1,i}\}$ for B_1 and by $\{G_{2,i}\}$ for B_2. Define also the observability map $\mathcal{O}$ as in (15.3.32), and the weighting matrices $\{\mathcal{Q}^d, \mathcal{W}^d\}$ as in (15.3.19). Let further

$$M = \left[\begin{bmatrix} \Phi(N+1, 0) \\ \mathcal{O} \end{bmatrix} \middle| B_1\right]$$

Using the state equations (15.4.17), it is easy to verify that

$$\begin{bmatrix} \mathbf{x}_{N+1} \\ \mathbf{s} \end{bmatrix} = M\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{w} \end{bmatrix} + B_2\mathbf{u},$$

so that the optimization problem (15.4.20) can be rewritten in the equivalent form:

$$\min_{\text{lower trian. } K} \text{Tr}\left(\|\mathbf{u}\|_{\mathcal{Q}^d}^2 + \left\| M\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{w} \end{bmatrix} + B_2\mathbf{u} \right\|_{\mathcal{W}^d}^2\right). \tag{15.4.21}$$

This is of the same form as in item (ii) of Remark 5 above, with $H = -B_2$. Therefore, the optimal coefficient matrix K_a for (15.4.19), *viz.*,

$$\hat{\mathbf{u}} = K_a\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{w} \end{bmatrix}, \tag{15.4.22}$$

can be determined by considering the dual problem of estimating anticausally the variable $M^*\mathbf{y}^d$ from the observation $\mathbf{z}^d$ in the linear model

$$\mathbf{z}^d = B_2^*\mathbf{y}^d + \mathbf{v}^d, \quad \|\mathbf{y}^d\|^2 = \mathcal{W}^d, \quad \|\mathbf{v}^d\|^2 = \mathcal{Q}^d.$$

That is, if $\hat{\mathbf{y}}^d = K_a^d\mathbf{z}^d$, then $K_a = -K_a^{d*}$.

The above linear model in $\{\mathbf{z}^d, \mathbf{y}^d\}$ has exactly the same form as the one we encountered earlier in (15.3.21)–(15.3.22), while studying the linear quadratic tracking problem. We proceed as before and partition $\mathbf{y}^d$ into uncorrelated entries,

$$\mathbf{y}^d = \text{col}\{\ \mathbf{x}_{N+1}^d,\ \mathbf{u}_0^d,\ \mathbf{u}_1^d,\ \ldots,\ \mathbf{u}_N^d\ \},$$

where $\|\mathbf{x}_{N+1}^d\|^2 = P_{N+1}^d$ and $\|\mathbf{u}_i^d\|^2 = R_i^d$. We also partition the entries of $\mathbf{z}$ and $M^*\mathbf{y}^d$ into $\mathbf{z}^d = \text{col}\{\mathbf{z}_0^d, \mathbf{z}_1^d, \ldots, \mathbf{z}_N^d\}$ and $M^*\mathbf{y}^d = \text{col}\{\mathbf{x}_0^d, \mathbf{t}_0^d, \mathbf{t}_1^d, \ldots, \mathbf{t}_N^d\}$. It is then immediately verified from the definition of the matrices $\{B_2, M\}$ above that the entries of $\{\mathbf{z}^d, \mathbf{y}^d, M^*\mathbf{y}^d\}$ satisfy the following backwards-time state-space model:

$$\begin{cases} \mathbf{x}_i^d &= F_i^*\mathbf{x}_{i+1}^d + H_i^*\mathbf{u}_i^d, \\ \mathbf{z}_i^d &= G_{2,i}^*\mathbf{x}_{i+1}^d + \mathbf{v}_i^d, \\ \mathbf{t}_i^d &= G_{1,i}^*\mathbf{x}_{i+1}^d, \end{cases} \tag{15.4.23}$$

where the $\{\mathbf{u}_i^d, \mathbf{v}_i^d\}$ are uncorrelated white noise sequences with variances $\{R_i^d, Q_i^d\}$; moreover, both are uncorrelated with $\mathbf{x}_{N+1}^d$.

The arguments from this point on are almost identical to those presented in Sec. 15.3.4 while minimizing a quadratic cost function via duality. Thus note that we are again reduced to the problem of estimating certain variables, *viz.*, $\{\mathbf{x}_0^d, \mathbf{t}_i^d\}$, from the observations $\{\mathbf{z}_0^d, \ldots, \mathbf{z}_N^d\}$, albeit causally. More specifically, let $\hat{\mathbf{x}}_{0|0}^d$ denote the estimator of $\mathbf{x}_0^d$ given all the observations $\{\mathbf{z}_0^d, \ldots, \mathbf{z}_N^d\}$, and let $\hat{\mathbf{t}}_{i|i}^d$ denote the estimator of $\mathbf{t}_i^d$

given the observations $\{\mathbf{z}_i^d, \ldots, \mathbf{z}_N^d\}$. Then the coefficient matrix K_a^d that we seek is the matrix that performs the mapping

$$\begin{bmatrix} \hat{\mathbf{x}}_{0|0}^d \\ \hline \hat{\mathbf{t}}_{0|0}^d \\ \hat{\mathbf{t}}_{1|1}^d \\ \vdots \\ \hat{\mathbf{t}}_{N|N}^d \end{bmatrix} = K_a^d \begin{bmatrix} \mathbf{z}_0^d \\ \mathbf{z}_1^d \\ \vdots \\ \mathbf{z}_N^d \end{bmatrix} \triangleq \begin{bmatrix} K_x^d \\ \hline \\ K_c^d \end{bmatrix} \mathbf{z}^d, \tag{15.4.24}$$

where we partitioned K_a^d into K_x^d and K_c^d; one corresponds to the map from $\mathbf{z}^d$ to $\hat{\mathbf{x}}_{0|0}^d$, while the other corresponds to the map from $\mathbf{z}^d$ to the $\{\hat{\mathbf{t}}_{i|i}^d\}$.

Once we determine K_a^d then, by duality, we know that the desired solution $\hat{\mathbf{u}}$ of the cost function (15.4.20) is given by

$$\hat{\mathbf{u}} = -K_a^{d*} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{w} \end{bmatrix} = -K_x^{d*}\mathbf{x}_0 - K_c^{d*}\mathbf{w}. \tag{15.4.25}$$

Now the desired estimators $\{\hat{\mathbf{x}}_{0|0}^d, \hat{\mathbf{t}}_{i|i}^d\}$ can be readily determined from the following backwards-time Kalman filtering recursions (*cf.* Prob. 9.14).

Theorem 15.4.1 (Backwards-Time Filtering Recursions) *Consider the model (15.4.23). The estimators $\{\hat{\mathbf{x}}_{0|0}^d, \hat{\mathbf{t}}_{i|i}^d\}$ can be computed recursively as follows:*

$$\begin{cases} \hat{\mathbf{x}}_{i|i}^d = F_{d,i}\hat{\mathbf{x}}_{i+1|i+1}^d + K_i^d R_{e,i}^{-d}\mathbf{z}_i^d, \quad \hat{\mathbf{x}}_{N+1|N+1}^d = 0, \\ \hat{\mathbf{t}}_{i|i} = G_{1,i}^*(I - P_{i+1}^d G_{2,i} R_{e,i}^{-d} G_{2,i}^*)\hat{\mathbf{x}}_{i+1|i+1}^d + G_{1,i}^* P_{i+1}^d G_{2,i} R_{e,i}^{-d}\mathbf{z}_i^d, \end{cases} \tag{15.4.26}$$

where $F_{d,i} = (F_i - G_{2,i} R_{e,i}^{-d} K_i^{d*})^*$,

$$R_{e,i}^d = G_{2,i}^* P_{i+1|i+1}^d G_{2,i} + Q_i^d, \quad K_i^d = F_i^* P_{i+1|i+1}^d G_{2,i},$$

and $P_{i|i}^d$ satisfies the backwards Riccati recursion

$$P_{i|i}^d = F_i^* P_{i+1|i+1}^d F_i + H_i^* R_i^d H_i - K_i^d R_{e,i}^{-d} K_i^{d*}, \quad P_{N+1|N+1}^d = P_{N+1}^d.$$

■

Proof: The only recursion that needs justification is that for $\hat{\mathbf{t}}_{i|i}$. But this follows immediately by noting that

$$\begin{aligned} \hat{\mathbf{t}}_{i|i} &= \hat{\mathbf{t}}_{i|i+1} + \langle \mathbf{t}_i, \mathbf{e}_i^d \rangle R_{e,i}^{-d}\mathbf{e}_i^d \\ &= G_{1,i}^*\hat{\mathbf{x}}_{i+1|i+1}^d + G_{1,i}^* P_{i+1}^d G_{2,i} R_{e,i}^{-d}(\mathbf{z}_i^d - G_{2,i}^*\hat{\mathbf{x}}_{i+1|i+1}^d). \end{aligned}$$

◆

Given the above Kalman filtering solution, we now only need to identify the mapping K_a^d and its negative conjugate transpose, in order to find the desired solution $\hat{\mathbf{u}}$ in (15.4.25). Recall that K_c^d is the mapping from the $\{\mathbf{z}_i^d\}$ to the $\{\hat{\mathbf{t}}_{i|i}^d\}$. This mapping is fully described by the state-space model (15.4.26), which we denote schematically by

$$\{\mathbf{z}_i^d\} \xrightarrow{(15.4.26)} \{\hat{\mathbf{t}}_{i|i}^d\}.$$

Observe further that this model has a zero boundary condition. Therefore, its dual model can be found just as explained in Sec. 15.2.3. Thus let $\{\hat{\mathbf{u}}_i\}$ denote the entries of the solution vector $\hat{\mathbf{u}}$ when $\mathbf{x}_0 = 0$. Then we can find $\{\hat{\mathbf{u}}_i\}$ by feeding the $\{\mathbf{w}_i\}$ into the following cascade:

$$\{\mathbf{w}_i\} \xrightarrow{\text{dual of (15.4.26)}} \{\hat{\mathbf{u}}_i\}.$$

This leads to the following construction:

$$\begin{cases} \boldsymbol{\xi}_{i+1} = F_{d,i}^*\boldsymbol{\xi}_i - (I - G_{2,i}R_{e,i}^{-d}G_{2,i}P_{i+1}^d)G_{1,i}\mathbf{w}_i, \quad \boldsymbol{\xi}_{N+1} = 0, \\ \hat{\mathbf{u}}_i = R_{e,i}^{-d}K_i^{d*}\boldsymbol{\xi}_i - R_{e,i}^{-d}G_{2,i}^* P_{i+1}^d G_{1,i}\mathbf{w}_i^d. \end{cases} \tag{15.4.27}$$

As for the mapping from $\mathbf{z}^d$ to $\hat{\mathbf{x}}_{0|0}^d$, the same argument following the statement of Thm. 15.3.3 will show that

$$\hat{\mathbf{x}}_{0|0}^d = \begin{bmatrix} K_0^d R_{e,0}^{-d} & \Psi_d(0,1)K_1^d R_{e,1}^{-d} & \ldots & \Psi_d(0,N)K_N^d R_{e,N}^{-d} \end{bmatrix} \mathbf{z}^d \triangleq K_x^d \mathbf{z}^d.$$

Therefore, the matrix multiplying $\mathbf{x}_0$ in (15.4.25), which is equal to $-K_x^{d*}$, coincides with the negative of the mapping from $\mathbf{x}_0$ to the $\{\hat{\mathbf{u}}_i\}$ in the equations (15.4.27). This means that all we need to do in order to incorporate the effect of the initial condition $\mathbf{x}_0$, is to initialize the recursion for $\boldsymbol{\xi}_i$ with $-\mathbf{x}_0$. Moreover, it is easy to verify that the state equation for $\boldsymbol{\xi}_i$ can be arranged as

$$\boldsymbol{\xi}_{i+1} = F_i\boldsymbol{\xi}_i - G_{1,i}\mathbf{w}_i - G_{2,i}\hat{\mathbf{u}}_i, \quad \boldsymbol{\xi}_0 = -\mathbf{x}_0,$$

which means that $\boldsymbol{\xi}_i$ coincides with the *negative* of the state vector $\mathbf{x}_i$ when the optimal control signal is used. In summary, we have the following result.

Theorem 15.4.2 (Causal Full Information Controller) *Consider the state space model (15.4.17)–(15.4.18), and assume the control signal $\mathbf{u}_i$ is allowed to be a causal linear function of the initial state and driving disturbances. Then the solution to problem (15.4.20) is given by*

$$\hat{\mathbf{u}}_i = -R_{e,i}^{-d}K_i^{d*}\mathbf{x}_i - R_{e,i}^{-d}G_{2,i}^* P_{i+1}^d G_{1,i}\mathbf{w}_i. \tag{15.4.28}$$

■

Remark 6. The solution of Thm. 15.4.2 has an interesting structure and shows that the resulting optimal control signal $\mathbf{u}_i$ is a function *only* of the current state, $\mathbf{x}_i$, and the current driving disturbance, $\mathbf{w}_i$. This is slightly different from the (now famous) state feedback solution of continuous-time optimal control. The reason for this difference is that we have required a causal controller. If we insist on a strictly causal controller (*i.e.*, $\mathbf{u}_i$ can only depend on $\mathbf{x}_0$ and on previous $\{\mathbf{w}_j\}$), then a true state-feedback controller is obtained. Indeed, in the strictly causal

then the above quadratic expression is equivalent to

$$E\left[\ (K_o\mathbf{b} - \mathbf{a})^* R_z^d\ (K_o\mathbf{b} - \mathbf{a}) + \mathbf{b}^*\Delta\mathbf{b}\ \right].$$

The second term is independent of $\mathbf{a}$ and, hence, we can focus on the minimization of the first term with respect to K. If we use $\mathbf{a} = K\mathbf{c}$ and denote $K_o\mathbf{b}$ by $\mathbf{d}$, then we are reduced to estimating $\mathbf{d}$ causally from $\mathbf{c}$ by solving

$$\min_{\text{lower trian. } K} E\ (\mathbf{d} - K\mathbf{c})^* R_{z^d}(\mathbf{d} - K\mathbf{c}), \qquad \mathbf{d} \triangleq K_o\mathbf{b}. \tag{15.5.10}$$

Now in view of the explanation given for (15.4.7), and noting that here we actually have a positive-definite weighting matrix ($R_{z^d} > 0$), we immediately see that the optimal coefficient K_c is given by (*cf.* (15.4.5))

$$K_c = \left\{R_{dc}L^{-c*}R_e^{-c}\right\}_{\text{lower}} L^{-c} = \left\{K_oR_bL_b^*L^{-c*}R_e^{-c}\right\}_{\text{lower}} L^{-c} \tag{15.5.11}$$

where in the second equality we simply used the fact that

$$R_{dc} = E\mathbf{d}\mathbf{c}^* = K_o(E\mathbf{b}\mathbf{c}^*) = K_oR_bL_b^*.$$

Expression (15.5.11) for K_c is almost identical to the desired expression (15.5.8), except for the matrix K_o appearing in place of K_a. To show that we can replace K_o in the equation (15.5.11) by K_a we only need to argue that we can replace K_o by K_a in problem (15.5.10), *i.e.*, we only need to argue that we can instead solve the problem

$$\min_{\text{lower trian. } K} E\ (\mathbf{f} - K\mathbf{c})^* R_{z^d}(\mathbf{f} - K\mathbf{c}), \qquad \mathbf{f} \triangleq K_a\mathbf{b}, \tag{15.5.12}$$

which amounts to estimating a variable $\mathbf{f}$ causally from $\mathbf{c}$, where $\mathbf{f}$ is defined via $\mathbf{f} = K_a\mathbf{b}$ (*i.e.*, we replace the K_o in the definition for $\mathbf{d}$ in (15.5.11) by K_a and call the resulting variable $\mathbf{f}$ to avoid confusion). Once this is done, then the resulting optimal coefficient is simply the matrix K_c that is given by (15.5.8), as desired.

To establish this fact, we shall rely on the assumed causal dependence of $\mathbf{c}$ on $\mathbf{b}$ and on the assumed uncorrelatedness of the entries of $\mathbf{b}$. These two facts have not yet been used in our derivation and, as we shall see, they are what enable us to replace K_o by K_a, as we now explain.

For this purpose, we first invoke the simple identity (4.1.25), derived earlier in Sec. 4.1.2, that relates the solution of a smoothing problem to the solution of the corresponding causal filtering problem. This identity enables us to relate K_o to K_a, which is a first step towards ultimately replacing K_o by K_a in (15.5.11) and getting (15.5.12). We repeat the short argument for this identity here for emphasis. Using the definition (15.5.9) for K_o, and the factorization $R_{z^d} = U^dR_e^dU^{d*}$, we can write

$$\begin{aligned} K_o &= U^{-d*}R_e^{-d}U^{-d}H^*R_y^dM \\ &= U^{-d*}\left[\left\{R_e^{-d}U^{-d}H^*R_y^dM\right\}_{\text{lower}} + \left\{R_e^{-d}U^{-d}H^*R_y^dM\right\}_{\text{s.upper}}\right] \\ &= K_a + U^{-d*}\left\{R_e^{-d}U^{-d}H^*R_y^dM\right\}_{\text{s.upper}} \\ &\triangleq K_a - U^{-d*}\Lambda, \end{aligned}$$

where $\{\cdot\}_{\text{s.upper}}$ denotes the strictly upper triangular part of its argument, and Λ is a *strictly* upper triangular matrix. In the last two equalities we also used the expression (15.5.4) for K_a.[13]

Substituting $K_o = K_a - U^{-d*}\Lambda$ into the definition for $\mathbf{d}$ in (15.5.10), we see that $\mathbf{d}$ can be expressed as the sum of two components

$$\mathbf{d} = K_a\mathbf{b} - U^{-d*}\Lambda\mathbf{b} \triangleq \mathbf{f} - U^{-d*}\Lambda\mathbf{b},$$

where we are denoting $K_a\mathbf{b}$ by $\mathbf{f}$. Therefore, by linearity of causal estimation, the optimal causal estimator of $\mathbf{d}$ given $\mathbf{c}$ is equal to the sum of the optimal causal estimators of its individual components given $\mathbf{c}$, *viz.*,

$$\hat{\mathbf{d}}_{|c} = \hat{\mathbf{f}}_{|c} - \widehat{(U^{-d*}\Lambda\mathbf{b})}_{|c}. \tag{15.5.13}$$

We shall now argue that the second causal estimator is zero. This will follow as a consequence of two facts: (i) the assumed causal dependence of $\mathbf{c}$ on $\mathbf{b}$, (ii) and the assumed uncorrelatedness of the entries of $\mathbf{b}$.

To see this, simply note that the optimal causal estimator for $U^{-d*}\Lambda\mathbf{b}$ given $\mathbf{c}$ is, by definition, found by solving

$$\min_{\text{lower trian. } K} E\ (U^{-d*}\Lambda\mathbf{b} - K\mathbf{c})^* R_{z^d}\ (U^{-d*}\Lambda\mathbf{b} - K\mathbf{c}),$$

which, by using $R_{z^d} = U^dR_e^dU^{d*}$, is equivalent to solving

$$\min_{\text{lower trian. } K} E\ (\Lambda\mathbf{b} - U^{d*}K\mathbf{c})^* R_e^d\ (\Lambda\mathbf{b} - U^{d*}K\mathbf{c}),$$

which in turn, by introducing the change of variables $\overline{K} = U^{d*}K$, is equivalent to solving

$$\min_{\text{lower trian. } \overline{K}} E\ (\Lambda\mathbf{b} - \overline{K}\mathbf{c})^* R_e^d\ (\Lambda\mathbf{b} - \overline{K}\mathbf{c}). \tag{15.5.14}$$

Observe that, since U^{d*} is invertible and lower triangular, both K and $\overline{K}$ define each other uniquely and are lower triangular. Now the solution to (15.5.14) is easily seen to be $\overline{K}_o = 0$. Indeed, recall from the solution of such causal estimation problems that $\overline{K}_o$ should be given by (*cf.* (15.4.5))

$$\overline{K}_o = \left\{R_{\Lambda b,c}L^{-c*}R_e^{-c}\right\}_{\text{lower}} L^{-c}.$$

Using $R_{\Lambda b,c} = E(\Lambda\mathbf{b})\mathbf{c}^* = \Lambda R_bL_b^*$, we find that

$$\overline{K}_o = \left\{\Lambda R_bL_b^*L^{-c*}R_e^{-c}\right\}_{\text{lower}} L^{-c}.$$

[13] To see why we claim that expression (15.5.12) is the analogue of what we derived earlier in Sec. 4.1.2, it is easier to argue by duality. Recall from the previous footnote that $-K_o^*$, or equivalently K_o^d, is the optimal coefficient matrix for estimating $M^*\mathbf{y}^d$ from $\mathbf{z}^d$ in the linear model $\mathbf{z}^d = -H^*\mathbf{y}^d + \mathbf{v}^d$. Recall also from the argument prior to (15.5.4) that $-K_a^*$, or equivalently K_a^d, is the optimal coefficient matrix for estimating $M^*\mathbf{y}^d$ anticausally from $\mathbf{z}^d$. Hence, relation (15.5.12), which can be rewritten as

$$K_o^d = K_a^d + \Lambda^*U^{-d},$$

is simply the identity (4.1.25): it relates the solution of a smoothing problem to the solution of an anticausal filtering problem. We could therefore have written the above equation directly by invoking the result of Sec. 4.1.2 and then, by duality, obtain (15.5.12) from it.

But the product $\Lambda R_b L_b^* L^{-c*} R_e^{-c}$ is strictly upper triangular, so that its lower part is zero! [Note that for this to hold, we used the fact that R_b is diagonal and that L_b is lower triangular, in addition to the strict upper triangularity of Δ.]

Returning to (15.5.13), we conclude that we only need to estimate $\mathbf{f}$ causally from $\mathbf{c}$, which is precisely the desired problem (15.5.12). In summary, we have shown that a solution K_c to problem (15.5.2) can be found via the two-step procedure explained earlier.

Remark 7. The above argument did not require the lower triangularity of L_w and, hence, the same two-step (separation) procedure will hold for mappings from $\mathbf{w}$ to $\mathbf{c}$ in (15.5.1) that are not necessarily causal. ◆

Remark 8 [A Generalization: c depends on a]. For quadratic control applications, the model for the measurement vector $\mathbf{c}$ in (15.5.1) is usually of the form

$$\mathbf{c} = L_b\,\mathbf{b} + L_a\,\mathbf{a} + L_w\,\mathbf{w}, \tag{15.5.15}$$

for lower triangular matrices $\{L_b, L_w\}$, and for an additional *strictly* lower triangular matrix L_a. That is, $\mathbf{c}$ also depends on $\mathbf{a}$ and, consequently, on the matrix K that we are trying to choose by solving (15.5.2). We shall show below that the solutions to such problems with measurement models of the form (15.5.15) are still given by the same separation principle: we first solve (15.5.2) as if $\mathbf{a}$ were a causal function of $\mathbf{b}$ only, *i.e.*, we obtain $\bar{\mathbf{a}}$, and then we estimate $\bar{\mathbf{a}}$ causally from $\mathbf{c}$.

We shall establish this conclusion by showing that an optimization problem as in (15.5.2) with a model as in (15.5.15), can always be reduced to a similar optimization problem with a model of the form (15.5.1), *viz.*, one in which $\mathbf{c}$ does *not* depend on the unknown K. To show this, we shall employ a simple change of variables, known as a linear fractional transformation (LFT).[14]

Thus consider the optimization problem

$$\min_{\begin{pmatrix}\text{lower trian. } K\\ \text{with } \mathbf{a} = K\mathbf{c}\end{pmatrix}} \mathrm{Tr}\left(\|\mathbf{a}\|^2_{R_{vd}} + \|M\mathbf{b} - H\mathbf{a}\|^2_{R_{yd}} \right) \quad \Longrightarrow \quad \hat{\mathbf{a}} = K_c\,\mathbf{c}, \tag{15.5.16}$$

where $\mathbf{c}$ is now given by the model (15.5.15). Substituting $\mathbf{a} = K\mathbf{c}$ into (15.5.15), we can re-express $\mathbf{c}$ in terms of $\{\mathbf{b}, \mathbf{w}\}$ alone as

$$\mathbf{c} = (I - L_aK)^{-1}L_b\,\mathbf{b} + (I - L_aK)^{-1}L_w\,\mathbf{w}, \tag{15.5.17}$$

where the lower triangular matrix $I - L_aK$, whose inverse is needed in the above expression, is invertible since L_a is strictly lower triangular. Define

$$\mathbf{r} \triangleq L_b\,\mathbf{b} + L_w\,\mathbf{w}. \tag{15.5.18}$$

[14] It is also easy to verify that the same argument we employed above for model (15.5.1) can be applied to this case, except that expressions like (15.5.8) will have the unknown matrix K_c appearing on both sides of the equality (since the variance matrix of $\mathbf{c}$, and consequently $\{L^c, R_e^c\}$, will now depend on K_c). Here we prefer to introduce a useful variable transformation, which is in fact widely used in linear quadratic control under the name YBJK parameterization (after the papers of Youla et al. (1976a,1976b) and Kučera (1974,1975) or Q-parametrization (see, *e.g.*, Green and Limebeer (1995)).

Then $\mathbf{r}$ is a random variable that satisfies a model of the form (15.5.1), *i.e.*, it does not depend on K. Introduce further the change of variables (or LFT):

$$J \triangleq K(I - L_aK)^{-1}. \tag{15.5.19}$$

Then J and K define each other uniquely and are both lower triangular. With these definitions, and using (15.5.17), we can write the condition $\mathbf{a} = K\mathbf{c}$ as

$$\mathbf{a} = K\mathbf{c} = K(I - L_aK)^{-1}L_b\,\mathbf{b} + K(I - L_aK)^{-1}L_w\,\mathbf{w} = J\mathbf{r}, \tag{15.5.20}$$

so that problem (15.5.16) can be equivalently stated as

$$\min_{\begin{pmatrix}\text{lower trian. } J\\ \text{with } \mathbf{a} = J\mathbf{r}\end{pmatrix}} \mathrm{Tr}\left(\|\mathbf{a}\|^2_{R_{vd}} + \|M\mathbf{b} - H\mathbf{a}\|^2_{R_{yd}} \right) \quad \Longrightarrow \quad \hat{\mathbf{a}} = J_c\,\mathbf{r},$$

with $\mathbf{r}$ given by (15.5.18). This formulation is exactly of the same form given by (15.5.1)–(15.5.16), so that the optimal solution, J_c, can be found as follows: we first determine $\bar{\mathbf{a}}$ and then estimate it causally from $\mathbf{r}$. Let K_c be the matrix that corresponds to this J_c via the LFT (15.5.19), which is therefore the desired solution of (15.5.16). It is easy to verify, by direct calculation, that K_c is the optimal matrix that estimates $\bar{\mathbf{a}}$ from $\mathbf{c}$, so that the separation argument is also valid for problem (15.5.16)! ◆

15.5.2 A Separation Principle with Anticausal Dependence on Data

In a similar vein, let now $\mathbf{c}$ be a random vector that is defined anticausally in terms of two other uncorrelated random vectors $\{\mathbf{a}, \mathbf{w}\}$, say

$$\mathbf{c} = U_a\,\mathbf{a} + U_w\,\mathbf{w}, \tag{15.5.21}$$

with $\{U_a, U_w\}$ upper triangular. The Gramian matrices of $\{\mathbf{a}, \mathbf{w}\}$ are denoted by $\{R_a, R_w\}$, with R_a assumed diagonal. We now consider the problem

$$\min_{\begin{pmatrix}\text{upper trian. } K^d\\ \text{with } \mathbf{b} = K^d\mathbf{c}\end{pmatrix}} \mathrm{Tr}\left(\|\mathbf{b}\|^2_{R_v} + \|M^*\mathbf{a} + H^*\mathbf{b}\|^2_{R_z} \right) \quad \Longrightarrow \quad \hat{\mathbf{b}} = K_c^d\,\mathbf{c}, \tag{15.5.22}$$

for any matrix M. That is, we seek a causal estimator for $\mathbf{b}$ that is defined in terms of $\mathbf{c}$ and not $\mathbf{a}$ (compare with the problem in item (i) of Remark 5). The same argument as in the previous section will show that the optimal solution, K_c^d, can again be determined in a two-step procedure:

1. First, we solve the above problem as if $\mathbf{b}$ were required to be a causal function of $\mathbf{a}$. That is, we solve the same problem we studied earlier in item (i) of Remark 5, and determine the corresponding coefficient matrix K_f^d. This yields an intermediate estimate for $\mathbf{b}$, which we shall denote by

$$\bar{\mathbf{b}} = K_f^d\,\mathbf{a}. \tag{15.5.23}$$

As we explained before, in Remark 5, this step is dual to a stochastic causal estimation problem, *viz.*, that of estimating Mz causally from $\mathbf{y}$ in the linear model $\mathbf{y} = H\mathbf{z} + \mathbf{v}$. Moreover, the optimal coefficient K_f^d is given by (*cf.* (15.4.5)):

$$K_f^d = -K_f^* = -L^{-*}\{R_e^{-1}L^{-1}HR_zM^*\}_{\text{upper}}, \tag{15.5.24}$$

where LR_eL^* denotes the upper-diagonal-lower triangular factorization of the Gramian matrix $R_y = R_v + HR_zH^*$.

2. Second, we determine the optimal causal estimator of $\bar{\mathbf{b}}$ given $\mathbf{c}$. We denote this solution by $\hat{\mathbf{b}}$, so that

$$\hat{\mathbf{b}} = K_c^d\,\mathbf{c} = \widehat{\bar{\mathbf{b}}}_{|\mathbf{c}}.$$

Remark 9. As in the case of causal dependence on the data, the matrix U_w need not be upper triangular (*i.e.*, the mapping from $\mathbf{w}$ to $\mathbf{c}$ in (15.5.21) need not be anticausal). Moreover, the same separation construction is valid when (15.5.21) is replaced by a model of the form

$$\mathbf{c} = U_a\,\mathbf{a} + U_b\,\mathbf{b} + U_w\,\mathbf{w}, \tag{15.5.25}$$

for upper triangular matrices $\{U_a, U_w\}$, and for an additional *strictly* upper triangular matrix U_b. ◆

15.5.3 Application to Measurement Feedback Control

We now demonstrate one application of the above results to the stochastic quadratic control problem of Sec. 15.4.4. Thus consider the same setting as in that section, with the same model (15.4.17), except that the control sequence $\{\mathbf{u}_i\}$ is now restricted to being a causal function of noisy observations $\{\mathbf{y}_i\}$ that are generated by

$$\mathbf{y}_i = L_i\mathbf{x}_i + \mathbf{v}_i, \tag{15.5.26}$$

where the $\{L_i\}$ are known matrices, while $\{\mathbf{v}_i\}$ is a white-noise sequence with variance $\{R_i^v\}$ and uncorrelated with $\{\mathbf{x}_0, \mathbf{w}_j\}$. This so-called measurement feedback problem differs from the full-information control problem of Sec. 15.4.4 in that (15.4.19) is now replaced by

$$\mathbf{u} = K\mathbf{y}, \qquad \mathbf{y} \stackrel{\Delta}{=} \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}, \tag{15.5.27}$$

with $\mathbf{y}$ instead of $\text{col}\{\mathbf{x}_0, \mathbf{w}\}$, and for a lower triangular matrix K that should still be determined optimally by solving (15.4.20), *viz.*,

$$\min_{\substack{\text{lower trian. } K \\ \text{with } \mathbf{u} = K\mathbf{y}}} E\left[\mathbf{x}_{N+1}^*P_{N+1}^d\mathbf{x}_{N+1} + \sum_{i=0}^{N}\mathbf{u}_i^*Q_i^d\mathbf{u}_i + \sum_{i=0}^{N}\mathbf{s}_i^*R_i^d\mathbf{s}_i\right]. \tag{15.5.28}$$

Since the cost function did not change, it can therefore be again written in the same form as (15.4.21). Note further that since $\mathbf{y}$ can be expressed causally in terms of the variables $\{\mathbf{x}_0, \mathbf{w}_i, \mathbf{v}_i\}$, we are reduced to an optimization problem of the same form as (15.5.15), *viz.*,[15]

$$\min_{\substack{\text{lower trian. } K \\ \text{with } \mathbf{u} = K\mathbf{y}}} \text{Tr}\left(\|\mathbf{u}\|_{\mathcal{Q}^d}^2 + \left\|M\begin{bmatrix}\mathbf{x}_0\\ \mathbf{w}\end{bmatrix} + B_2\mathbf{u}\right\|_{\mathcal{W}^d}^2\right), \tag{15.5.29}$$

with $H = -B_2$ (where B_2 is defined after (15.4.20)).

The solution is therefore readily seen to be equal to the projection of the full-information control signal (15.4.28) onto $\mathcal{L}\{\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_i\}$. Thus, the minimizing solution is now

$$\hat{\mathbf{u}}_i = -R_{e,i}^{-d}K_i^{d*}\hat{\mathbf{x}}_{i|i} - R_{e,i}^{-d}G_{2,i}^*P_{i+1}^dG_{1,i}\hat{\mathbf{w}}_{i|i}, \tag{15.5.30}$$

where $\hat{\mathbf{x}}_{i|i}$ and $\hat{\mathbf{w}}_{i|i}$ are the l.l.m.s. estimators of $\mathbf{x}_i$ and $\mathbf{w}_i$ given $\{\mathbf{y}_0, \ldots \mathbf{y}_j\}$. Using the state-space model (15.4.17) we conclude that

$$\hat{\mathbf{w}}_{i|i} = 0, \qquad \text{since } \mathbf{w}_i \text{ is uncorrelated with } \{\mathbf{y}_0, \ldots \mathbf{y}_i\}, \tag{15.5.31}$$

so that the desired control signal is

$$\hat{\mathbf{u}}_i = -R_{e,i}^{-d}K_i^{d*}\hat{\mathbf{x}}_{i|i}. \tag{15.5.32}$$

This solution shows that we have the same state-feedback control law as in the strictly causal full information case, except that the (inaccessible) state is replaced by its filtered estimator, which is computed by the Kalman filter recursions

$$\begin{cases}\hat{\mathbf{x}}_{i+1} = F_i\hat{\mathbf{x}}_i + G_{2,i}\hat{\mathbf{u}}_i + K_{p,i}(\mathbf{y}_i - L_i\hat{\mathbf{x}}_i), \quad \hat{\mathbf{x}}_0 = 0,\\ \hat{\mathbf{x}}_{i|i} = \hat{\mathbf{x}}_i + P_iL_i^*R_{e,i}^{-1}(\mathbf{y}_i - L_i\hat{\mathbf{x}}_i),\end{cases} \quad i = 0, 1, \ldots N, \tag{15.5.33}$$

where

$$K_{p,i} = F_iP_iL_i^*R_{e,i}^{-1}, \qquad R_{e,i} = R_i^v + L_iP_iL_i^*, \tag{15.5.34}$$

and P_i satisfies the Riccati recursion

$$P_{i+1} = F_iP_iF_i^* + G_{1,i}Q_i^vG_{1,i}^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0 = \Pi_0. \tag{15.5.35}$$

This fact, *i.e.*, the separation of the state feedback law from the state estimator (and vice versa), is known as the *separation principle* or the *certainty equivalence* principle.

We should mention that it is also possible to study the measurement feedback control problem for the case where the control signal is only allowed to be a *strictly* causal function of the observations. The development is very similar to the one given above and will not be repeated here. The only difference is that the filtered state estimators, $\hat{\mathbf{x}}_{i|i}$, should be replaced by the predicted state estimators, $\hat{\mathbf{x}}_i$, as expected.

[15] Observe from the state-space equations (15.4.17) and (15.5.26) that the measurement vector $\mathbf{y}$ depends causally on the $\{\mathbf{x}_0, \mathbf{w}_j\}$ and the $\{\mathbf{v}_j\}$, while its dependence on the $\{\mathbf{u}_j\}$ is strictly causal.

15.6 DUALITY IN THE FREQUENCY DOMAIN

In this section we extend the concept of duality to the frequency domain, and then apply it to the study of the steady-state LQR problem.

Thus let the notation $H(e^{j\omega})$ denote the DTFT of a sequence $\{h_i\}$, *viz.*,

$$H(e^{j\omega}) \triangleq \sum_{i=-\infty}^{\infty} h_i e^{-j\omega i}, \quad j = \sqrt{-1}, \quad \omega \in [-\pi, \pi],$$

and similarly for $\{Z(e^{j\omega}), Y(e^{j\omega})\}$. We shall assume that the ROC of $H(z)$ includes the unit circle so that $H(z)$ and $H^*(z^{-*})$ can be regarded as the transfer functions of BIBO stable systems.

15.6.1 Duality without Constraints

Now given $\{Z(e^{j\omega}), H(e^{j\omega})\}$, we start by considering the problem of minimizing the following deterministic cost function over all $Z(e^{j\omega})$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{Z^*(e^{j\omega})R_{v^d}Z(e^{j\omega}) + \left|Y(e^{j\omega}) - H(e^{j\omega})Z(e^{j\omega})\right|^2_{R_{y^d}}\right\}d\omega. \tag{15.6.1}$$

where $R_{v^d} > 0$ and $R_{y^d} \geq 0$ are given weighting matrices (compare with the cost (15.3.14) in the time domain). It is immediate to verify by differentiation that the minimizing solution is given by

$$\widehat{Z}(e^{j\omega}) = \left[R_{v^d} + H^*(e^{j\omega})R_{y^d}H(e^{j\omega})\right]^{-1}H^*(e^{j\omega})R_{y^d}Y(e^{j\omega}).$$

So, in the z-transform domain, the optimal filter $K_o(z)$ that yields $\widehat{Z}(z)$ is given by

$$K_o(z) = \left[R_{v^d} + H^*(z^{-*})R_{y^d}H(z)\right]^{-1}H^*(z^{-*})R_{y^d}.$$

The dual stochastic problem can be formulated as follows. Introduce stationary random processes $\{\mathbf{z}_i^d, \mathbf{y}_i^d, \mathbf{v}_i^d\}$ that satisfy the linear model (compare again with (15.3.12)):

$$\mathbf{Z}^d(z) = -H^*(z^{-*})\mathbf{Y}^d(z) + \mathbf{V}^d(z), \tag{15.6.2}$$

where $\{\mathbf{y}_i^d, \mathbf{v}_i^d\}$ are white-noise uncorrelated processes with z-spectra

$$S_{y^d}(z) = R_y^d, \qquad S_v(z) = R_v^d.$$

By (15.6.2) we mean that $\{\mathbf{y}_i^d\}$ drives the filter $-H^*(z^{-*})$ and the noisy output is taken as $\{\mathbf{z}_i^d\}$.

We already know from from the discussion in Sec. 7.3.1 that the optimal smoother that estimates $\{\mathbf{y}_i^d\}$ from all the observations $\{\mathbf{z}_j^d, -\infty < j < \infty\}$, *viz.*,

$$\widehat{\mathbf{Y}}^d(z) = K_o^d(z)\mathbf{Z}^d(z),$$

is given by

$$K_o^d(z) = S_{y^dz^d}(z)S_{z^d}^{-1}(z).$$

Using the linear model (15.6.2) we find that

$$S_{y^dz^d}(z) = -R_y^dH(z), \quad S_{z^d}(z) = R_{v^d} + H^*(z)R_{y^d}H(z),$$

so that the expression for the optimal smoother becomes

$$K_o^d(z) = -K_o^*(z^{-*}). \tag{15.6.3}$$

In other words, we find that the problem of minimizing (15.6.1) is dual to that of estimating $\{\mathbf{y}_i^d\}$ from all the $\{\mathbf{z}_j^d\}$.

Remark 10. In a similar fashion, given $\{Z^d(e^{j\omega}), H(e^{j\omega})\}$, we can start with a cost function of the form

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{Y^{d*}(e^{j\omega})R_vY^d(e^{j\omega}) + \left|Z^d(e^{j\omega}) + H^*(e^{j\omega})Y^d(e^{j\omega})\right|^2_{R_z}\right\}d\omega, \tag{15.6.4}$$

and verify that its minimizing solution, *viz.*,

$$\begin{aligned}\widehat{Y}^d(e^{j\omega}) &= -\left[R_v + H(e^{j\omega})R_zH^*(e^{j\omega})\right]^{-1}H(e^{j\omega})R_zZ^d(e^{j\omega}),\\ &\triangleq K_o^d(e^{j\omega})Z^d(e^{j\omega}),\end{aligned} \tag{15.6.5}$$

is dual to the problem of estimating $\{\mathbf{z}_i\}$ from all the observations $\{\mathbf{y}_j\}$ in the model

$$\mathbf{Y}(z) = H(e^{j\omega})\mathbf{Z}(z) + \mathbf{V}(z), \tag{15.6.6}$$

where $\{\mathbf{z}_i, \mathbf{v}_i\}$ are white-noise uncorrelated processes with z-spectra $S_z(z) = R_z$ and $S_v(z) = R_v$. More specifically, using $\widehat{\mathbf{Z}}(z) = K_o(z)\mathbf{Y}(z) = S_{xy}(z)S_y^{-1}(z)\mathbf{Y}(z)$, and comparing with (15.6.5), we again obtain that

$$K_o(z) = -K_o^{d*}(z^{-*}). \tag{15.6.7}$$

◆

15.6.2 Duality with Causality Constraints

We can also formulate estimation problems in the frequency domain with causality constraints. Thus consider first a cost function of the form

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{A^*(e^{j\omega})R_{y^d}A(e^{j\omega}) + \left|M(e^{j\omega})b - H(e^{j\omega})A(e^{j\omega})\right|^2_{R_{y^d}}\right\}d\omega, \tag{15.6.8}$$

where $\{M(e^{j\omega}), H(e^{j\omega})\}$ are given, b is a known column vector, and the unknown $A(e^{j\omega})$ is now restricted to being the DTFT of a *causal sequence* $\{a_i\}$, *i.e.*, $a_i = 0$ for $i < 0$. The ROCs of $M(z)$ and $H(z)$ are assumed to include the unit circle. It is then required to minimize (15.6.8) over all functions $A(z)$ of the form

$$A(z) = K(z)b,$$

with $K(z)$ a *causal* transfer function (*i.e.*, its impulse response sequence $\{k_i\}$ is zero for $i < 0$). This problem is the frequency domain analogue of the problem in item (i) of Remark 5, except for the expectation symbol (since here we are not requiring b to be a random vector).

To minimize (15.6.8), and also to motivate a dual stochastic problem, let us note first that by using $A(z) = K(z)b$, we can rewrite the cost function in the form

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[K^*(R_{v^d}+H^*R_y^dH)K+M^*R_{y^d}M-M^*R_{y^d}HK-K^*H^*R_{y^d}M\right]\|b\|^2\,d\omega \tag{15.6.9}$$

where we are omiting the argument $e^{j\omega}$ for compactness of notation. We shall denote the optimal solution of this problem by $K_a(z)$.

Now consider stationary random processes $\{\mathbf{z}_i^d, \mathbf{y}_i^d, \mathbf{x}_i^d, \mathbf{v}_i^d\}$ that satisfy the linear model, in the z-transform domain,[16]

$$\begin{cases}\mathbf{Z}^d(z) &= -H^*(z^{-*})\mathbf{Y}^d(z)+\mathbf{V}^d(z),\\ \mathbf{X}^d(z) &= M^*(z^{-*})\mathbf{Y}^d(z),\end{cases} \tag{15.6.10}$$

where $\{\mathbf{y}_i^d, \mathbf{v}_i^d\}$ are white-noise uncorrelated processes with z-spectra $S_{y^d}(z)=R_y^d$ and $S_{v^d}(z)=R_v^d$. Let $K_a^d(z)$ denote the smoother that estimates $\{\mathbf{x}_i^d\}$ *anticausally* from $\{\mathbf{z}_j^d, i\le j<\infty\}$. We know from the discussion in Sec. 7.3.1 that $K_a^d(z)$ is given by

$$K_a^d(z) = -\left\{M^*(z^{-*})R_y^dH(z)L^{-1}(z)\right\}_{ac}R_e^{-1}L^{-*}(z^{-*}), \tag{15.6.11}$$

where the notation $\{\cdot\}_{ac}$ extracts the anticausal part of its argument, and $L(z)$ is the (minimum-phase) canonical spectral factor of $S_{z^d}(z)=H^*(z^{-*})R_{y^d}H(z)+R_{v^d}$, *viz.*,

$$S_{z^d}(z) = L^*(z^{-*})R_eL(z), \qquad R_e>0.$$

In fact, $K_a^d(z)$ is the optimal choice, over all anticausal filters $K^d(z)$, that minimizes the variance of the estimation error

$$\tilde{\mathbf{x}}_i^d \triangleq \left(\mathbf{x}_i^d - \sum_{j=i}^{\infty}k_{i-j}^d\,\mathbf{z}_j^d\right),$$

where we are denoting the impulse response sequence of $K^d(z)$ by $\{k_n^d, n\le 0\}$. That is, $K_a^d(z)$ solves

$$\min_{\begin{pmatrix}\text{anticausal}\\ K^d(z)\end{pmatrix}}\ \frac{1}{2\pi}\int_{-\pi}^{\pi}S_{\tilde{x}^d}(e^{j\omega})d\omega.$$

Using $S_{z^dy^d}(z)=-R_{y^d}H(z)$, $S_{x^d}(z)=M^*(z^{-*})R_{y^d}M(z)$, and the expression for $S_{z^d}(z)$, we can evaluate $S_{\tilde{x}^d}(z)$ and find that $K_z^d(z)$ is the anticausal function that minimizes

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[K^d(R_{v^d}+H^*R_y^dH)K^{d*}+M^*R_{y^d}M+M^*R_{y^d}HK^{d*}+K^dH^*R_{y^d}M\right]\,d\omega. \tag{15.6.12}$$

[16] The BIBO stability of this model, and hence the stationarity of the processes $\{\mathbf{z}_i^d, \mathbf{y}_i^d, \mathbf{x}_i^d\}$, is guaranteed by our assumption that the ROCs of $H(z)$ and $M(z)$ include the unit circle.

Comparing with the cost (15.6.9) we conclude that the optimal solutions are the dual of one another, *i.e.*,

$$K_a(z) = -K_a^{d*}(z^{-*}). \tag{15.6.13}$$

Remark 11. In a similar fashion, consider the cost function

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{B^*(e^{j\omega})R_vB(e^{j\omega})+\left|M^*(e^{j\omega})a+H^*(e^{j\omega})B(e^{j\omega})\right|_{R_z}^2\right\}d\omega, \tag{15.6.14}$$

where $\{M(e^{j\omega}), H(e^{j\omega})\}$ are given, a is a known column vector, and the unknown $B(e^{j\omega})$ is now restricted to being the DTFT of an *anticausal sequence* $\{b_i\}$, *i.e.*, $b_i=0$ for $i>0$. The ROCs of $M(z)$ and $H(z)$ are again assumed to include the unit circle. It is required to minimize (15.6.14) over all functions $B(z)$ of the form $K^d(z)a$, with $K^d(z)$ *anticausal*. Let $K_f^d(z)$ denote the optimal solution.

Now consider stationary random processes $\{\mathbf{z}_i, \mathbf{y}_i, \mathbf{x}_i, \mathbf{v}_i\}$ that satisfy the linear model

$$\begin{cases}\mathbf{Y}(z) &= H(z)\mathbf{Z}(z)+\mathbf{V}(z),\\ \mathbf{X}(z) &= M(z)\mathbf{Z}(z),\end{cases} \tag{15.6.15}$$

where $\{\mathbf{z}_i, \mathbf{v}_i\}$ are white-noise uncorrelated processes with z-spectra $S_z(z)=R_z$ and $S_v(z)=R_v$. Let $K_f(z)$ denote the optimal smoother that estimates $\{\mathbf{x}_i\}$ *causally* from $\{\mathbf{y}_j, j\le i\}$. Then

$$K_f(z) = -K_f^{d*}(z^{-*}). \tag{15.6.16}$$

◆

15.6.3 Application to the Infinite-Horizon LQR Problem

We return to the LQR application of Sec. 15.3.6 and consider its infinite-horizon version. More specifically, we consider the problem of minimizing the cost function

$$J^c = \sum_{i=0}^{\infty}\left[u_i^*Q^du_i + s_i^*R^ds_i\right], \tag{15.6.17}$$

over the variables $\{u_i\}_{i=0}^{\infty}$ and subject to the time-invariant state-space constraints

$$x_{i+1} = Fx_i+Gu_i, \qquad s_i = Hx_i. \tag{15.6.18}$$

We further assume that F is a stable matrix.

To begin with, by invoking the Parseval relation of Prob. 6.10, we readily see that we can interpret the control objective (15.6.17) as one that seeks a causal control sequence $\{u_i\}$ that minimizes the following cost in the frequency domain:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[U^*(e^{j\omega})Q^dU(e^{j\omega}) + S^*(e^{j\omega})R^dS(e^{j\omega})\right]d\omega,$$

where $\{U(e^{j\omega}), S(e^{j\omega})\}$ denote the DTFTs of the causal sequences $\{u_i, s_i\}$, respectively. Now using the state-space model (15.6.18) we can express $S(z)$ in terms of the unknown $U(z)$ as follows:

$$S(z) = M(z)x_0 - P(z)U(z),$$

where we are defining the transfer functions

$$M(z) \triangleq H(zI - F)^{-1}z, \qquad P(z) \triangleq -H(zI - F)^{-1}G.$$

Substituting into the above cost function, we find that it reduces to one of the form

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{U^*Q^dU \;+\; [Mx_0 - PU]^*\,R^d\,[Mx_0 - PU]\right\}d\omega,$$

where we are dropping the argument $e^{j\omega}$ for compactness of notation. This cost function will be of the same form as (15.6.8) with the identifications

$$b \longleftrightarrow x_0, \quad H \longleftrightarrow P, \quad A \longleftrightarrow U, \quad R_{v^d} \longleftrightarrow Q^d, \quad R_{y^d} \longleftrightarrow R^d.$$

This suggests that we should consider the following dual stochastic problem.

Let $\{\mathbf{z}_i^d, \mathbf{y}_i^d, \mathbf{x}_i^d, \mathbf{v}_i^d\}$ be zero-mean stationary random processes that are related in the transform domain as follows:

$$\begin{cases}\mathbf{Z}^d(z) \;=\; -P^*(z^{-*})\mathbf{Y}^d(z) + \mathbf{V}^d(z),\\ \mathbf{X}^d(z) \;=\; M^*(z^{-*})\mathbf{Y}^d(z),\end{cases} \tag{15.6.19}$$

where $\{\mathbf{y}_i^d, \mathbf{v}_i^d\}$ are taken as white-noise uncorrelated stationary processes with z-spectra $S_{y^d}(z) = R^d$ and $S_{v^d}(z) = Q^d$. It is immediate to verify, from the definitions of $\{M(z), P(z)\}$, that the stationary processes $\{\mathbf{z}_i^d, \mathbf{y}_i^d, \mathbf{x}_i^d\}$ satisfy the state-space model

$$\begin{cases}\mathbf{x}_i^d \;=\; F^*\mathbf{x}_{i+1}^d + H^*\mathbf{y}_i^d,\\ \mathbf{z}_i^d \;=\; G^*\mathbf{x}_{i+1}^d + \mathbf{v}_i^d, \quad i < \infty.\end{cases} \tag{15.6.20}$$

Let $K_o^d(z)$ denote the filter that estimates $\mathbf{x}_i^d$ anticausally from $\{\mathbf{z}_j^d, i \le j < \infty\}$. This filter can be readily found from the steady-state Kalman filter (of Thm. 15.3.1)

$$\begin{cases}\hat{\mathbf{x}}_{i|i}^d \;=\; F^*\;\hat{\mathbf{x}}_{i+1|i+1}^d + K^dR_e^{-d}\;\mathbf{e}_i^d,\\ \mathbf{e}_i^d \;=\; \mathbf{z}_i^d - G\hat{\mathbf{x}}_{i+1|i+1}^d,\end{cases}$$

where $R_e^d = G^*P^dG + Q^d$, $K^d = F^*P^dG$, and P^d is the unique nonnegative-definite solution of the DARE

$$P^d \;=\; F^*P^dF + H^*R^dH - K^dR_e^{-d}K^{d*},$$

that stabilizes $F_d = (F - GR_e^{-d}K^{d*})^*$. Taking $z-$transforms we find that

$$\widehat{\mathbf{X}}^d(z) = z^{-1}\left(z^{-1}I - F_d\right)^{-1}K^dR_e^{-d}\;\mathbf{Z}^d(z) \triangleq K_o^d(z)\;\mathbf{Z}^d(z),$$

so that, by duality,

$$\widehat{U}(z) = -K_o^{d*}(z^{-*})x_0 \;=\; -R_e^{-d}K^{d*}\left[I - z^{-1}F_d^*\right]^{-1}x_0,$$

which is the solution to the infinite-horizon LQR problem.

15.7 COMPLEMENTARY STATE-SPACE MODELS

In this section we return to the dual bases $\{\mathbf{z}^d, \mathbf{y}^d\}$ of Sec. 15.1 and study more closely their structure when the obervation vectors $\{\mathbf{y}_i\}$ in $\mathbf{y}$ have an underlying *state-space* structure. Our discussion is motivated by the result of Lemma 15.1.2, which shows that estimation problems can also be solved by projecting onto dual bases. This alternative approach requires that we first determine the structure of the space $\mathcal{L}\{\mathbf{z}^d\}$, which we shall refer to as the *complementary space* since it spans the orthogonal complement space of the observations space in the (larger) space spanned by $\{\mathbf{x}_0, \{\mathbf{u}_i, \mathbf{v}_i\}$. It turns out that this complementary space also admits state-space representations, a special case of which was already discussed in Sec. 15.2.3 for zero initial conditions.

15.7.1 The Standard State-Space Model

We start with the standard state-space model

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i, \quad \mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i, \quad 0 \le i \le N, \tag{15.7.1}$$

where the $\{\mathbf{u}_i, \mathbf{v}_i\}$ are uncorrelated white-noise processes with variances $\{Q_i, R_i\}$ and uncorrelated with $\mathbf{x}_0$, whose variance we denote by Π_0. We shall assume here that $\Pi > 0$, $Q_i > 0$, and $R_i > 0$. We further define (*cf.* Sec. 5.A)

$$\mathcal{O} \triangleq \begin{bmatrix} H_0\Phi(0,0)\\ H_1\Phi(1,0)\\ H_2\Phi(2,0)\\ \vdots\\ H_N\Phi(N,0)\end{bmatrix}, \qquad \Gamma \triangleq \begin{bmatrix} 0 & & & \\ \Gamma_{10} & 0 & & \\ \Gamma_{20} & \Gamma_{21} & 0 & \\ \Gamma_{30} & \Gamma_{31} & \Gamma_{32} & 0\\ \cdot & \cdot & \cdot & \cdot\end{bmatrix},$$

where $\Phi(i,i) = I$, $\Gamma_{ij} = H_i\Phi(i, j+1)G_j$, and

$$\Phi(i,j) \;=\; F_{i-1}F_{i-2}\ldots F_j, \quad i > j.$$

Then with

$$\mathbf{y} = \text{col}\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N\}, \quad \mathbf{u} = \text{col}\{\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_N\}, \quad \mathbf{v} = \text{col}\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_N\},$$

we can write

$$\mathbf{y} = \mathcal{O}\mathbf{x}_0 + \Gamma\mathbf{u} + \mathbf{v} = \begin{bmatrix}\mathcal{O} & \Gamma\end{bmatrix}\begin{bmatrix}\mathbf{x}_0\\ \mathbf{u}\end{bmatrix} + \mathbf{v}, \tag{15.7.2}$$

where the Gramian matrix of the variables $\{\mathbf{x}_0, \mathbf{u}, \mathbf{v}\}$ is given by

$$\left\langle\begin{bmatrix}\mathbf{x}_0\\ \mathbf{u}\\ \mathbf{v}\end{bmatrix}, \begin{bmatrix}\mathbf{x}_0\\ \mathbf{u}\\ \mathbf{v}\end{bmatrix}\right\rangle = \begin{bmatrix}\begin{bmatrix}\Pi_0 & 0\\ 0 & \mathcal{Q}\end{bmatrix} & 0\\ 0 & \mathcal{R}\end{bmatrix}, \tag{15.7.3}$$

with

$$\mathcal{Q} \triangleq (Q_0 \oplus \ldots \oplus Q_N) \quad \text{and} \quad \mathcal{R} \triangleq (R_0 \oplus \ldots \oplus R_N).$$

Expression (15.7.2) shows that the entries of the vector $\mathbf{y}$ lie in the linear space spanned by the variables $\{\mathbf{x}_0, \mathbf{u}_i, \mathbf{v}_i, 0 \le i \le N\}$. Let $\mathbf{z} = \text{col}\{\mathbf{x}_0, \mathbf{u}\}$ and, hence,

$$\mathbf{y} = \begin{bmatrix} \mathcal{O} & \Gamma \end{bmatrix} \mathbf{z} + \mathbf{v}. \tag{15.7.4}$$

In the sequel we shall study the structure of the dual basis to the space $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$.

15.7.2 Backwards Complementary Models

Let $\{\mathbf{z}^d, \mathbf{y}^d\}$ denote the dual basis to the space $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$. Since, by definition, $\mathbf{z}^d$ and $\mathbf{y}$ are orthogonal, and since from Prob. 15.1 we know that $\{\mathbf{z}, \mathbf{y}\}$ and $\{\mathbf{z}^d, \mathbf{y}\}$ span the same linear space, we shall say that $\mathbf{z}^d$ spans the orthogonal complement space to $\mathbf{y}$ in $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$. More specifically, we shall write[17]

$$\mathcal{L}(\mathbf{z}, \mathbf{y}) = \mathcal{L}(\mathbf{y}) \oplus \mathcal{L}(\mathbf{z}^d).$$

The interesting fact is that the entries of $\mathbf{z}^d$ will also obey a state-space model, albeit one that runs backwards in time (see Sec. 15.7.4 for a forwards time equivalent).

To see this, we start by noting that from (15.7.4), and Lemma 15.2.1, the dual basis $\{\mathbf{z}^d, \mathbf{y}^d\}$ is given by

$$\mathbf{z}^d = -\begin{bmatrix} \mathcal{O}^* \\ \Gamma^* \end{bmatrix} \mathbf{y}^d + \mathbf{v}^d, \tag{15.7.5}$$

where

$$\mathbf{y}^d \triangleq \mathcal{R}^{-1}\mathbf{v} \quad \text{and} \quad \mathbf{v}^d \triangleq \begin{bmatrix} \Pi_0^{-1}\mathbf{x}_0 \\ \mathcal{Q}^{-1}\mathbf{u} \end{bmatrix}. \tag{15.7.6}$$

Combining (15.7.5) and (15.7.6) we obtain

$$\mathbf{z}^d = \begin{bmatrix} -\mathcal{O}^*\mathcal{R}^{-1}\mathbf{v} + \Pi_0^{-1}\mathbf{x}_0 \\ \\ -\Gamma^*\mathcal{R}^{-1}\mathbf{v} + \mathcal{Q}^{-1}\mathbf{u} \end{bmatrix}, \tag{15.7.7}$$

which leads to the following statement.

Lemma 15.7.1 (A Backwards Complementary Model) *The entries of the vector $\mathbf{z}^d$ can be decomposed as*

$$\mathbf{z}^d = \begin{bmatrix} \mathbf{z}_0^d \\ \hline \boldsymbol{\eta}^b \end{bmatrix} \triangleq \begin{bmatrix} \xi_0^b + \Pi_0^{-1}\mathbf{x}_0 \\ \hline \boldsymbol{\eta}_0^b \\ \boldsymbol{\eta}_1^b \\ \vdots \\ \boldsymbol{\eta}_N^b \end{bmatrix}, \tag{15.7.8}$$

[17] The notation $\mathcal{L}\{\mathbf{a}, \mathbf{b}\} = \mathcal{L}\{\mathbf{a}\} \oplus \mathcal{L}\{\mathbf{c}\}$ means that $\mathcal{L}\{\mathbf{a}, \mathbf{b}\} = \mathcal{L}\{\mathbf{a}, \mathbf{c}\}$ and $\mathbf{c} \perp \mathcal{L}\{\mathbf{a}\}$.

where the entries $\{\boldsymbol{\eta}_j^b\}$ are generated via the backwards-time state-space model:

$$\begin{cases} \xi_i^b &= F_i^*\xi_{i+1}^b - H_i^*R_i^{-1}\mathbf{v}_i, \quad i = N, N-1, \ldots, 1, 0, \\ \boldsymbol{\eta}_i^b &= G_i^*\xi_{i+1}^b + Q_i^{-1}\mathbf{u}_i, \end{cases} \tag{15.7.9}$$

with $\xi_{N+1}^b = 0$ and where ξ_0^b denotes its final state. ■

Proof: Note from Eq. (15.7.7) that we need to show that

$$\begin{bmatrix} -\mathcal{O}^*\mathcal{R}^{-1}\mathbf{v} + \Pi_0^{-1}\mathbf{x}_0 \\ -\Gamma^*\mathcal{R}^{-1}\mathbf{v} + \mathcal{Q}^{-1}\mathbf{u} \end{bmatrix} = \begin{bmatrix} \xi_0^b + \Pi_0^{-1}\mathbf{x}_0 \\ \boldsymbol{\eta}^b \end{bmatrix}. \tag{15.7.10}$$

It is easy to verify from the state-space model (15.7.9), and from the boundary condition $\xi_{N+1}^b = 0$, that

$$\boldsymbol{\eta}^b = -\Gamma^*\mathcal{R}^{-1}\mathbf{v} + \mathcal{Q}^{-1}\mathbf{u},$$

which is the matrix relation used to define $\boldsymbol{\eta}^b$ in (15.7.10). Likewise, it follows from the state-space recursion (15.7.9), and from $\xi_{N+1}^b = 0$, that $\xi_0^b = -\mathcal{O}^*\mathcal{R}^{-1}\mathbf{v}$, which establishes the equality in the first row of (15.7.10). ◆

The backwards-time state-space model (15.7.9) is well known as the *adjoint* state-space model. In the stochastic framework, it is called (following Weinert and Desai (1981)) a (backwards-time) *complementary* model, since its output spans the orthogonal complement space of the output of the original model (15.7.1).

The duality with the original model is quite explicit. Indeed, we see that if we perform the following transformations

$$F_i \to F_i^*, \quad G_i \to -H_i^*, \quad Q_i \to R_i^{-1}, \quad R_i \to Q_i^{-1}, \tag{15.7.11}$$

and reverse the time direction, we can obtain the dual state-space model from its original.

We close this section with two useful properties of the dual vector $\mathbf{z}^d$ in (15.7.8). Thus let $\hat{\mathbf{z}}_0^d$ denote the projection of the leading entry of $\mathbf{z}^d$ onto the space spanned by its remaining entries, *viz.*, the space $\mathcal{L}\{\boldsymbol{\eta}^b\}$. Let also $\tilde{\mathbf{z}}_0^d$ denote the resulting estimation error. Likewise, let $\hat{\xi}_{0|0}^b$ denote the projection of the initial state vector ξ_0^b of the complementary model (15.7.9) onto the same space, $\mathcal{L}\{\boldsymbol{\eta}^b\}$, with estimation error $\tilde{\xi}_{0|0}^b$.

Lemma 15.7.2 (Two Properties of $\mathbf{z}^d$) *Consider the dual vector $\mathbf{z}^d$ defined in (15.7.8). The following facts hold:*

(i) Each state vector $\mathbf{x}_i$ of the original state-space model (15.7.1) is orthogonal to the current and future outputs of the complementary model (15.7.9),

$$\mathbf{x}_i \perp \boldsymbol{\eta}_j^b \quad \text{for} \quad j = i, i+1, \ldots, N.$$

(ii) The estimation error $\tilde{\mathbf{z}}_0^d$ satisfies

$$\tilde{\mathbf{z}}_0^d \;=\; P_{0|0}^{-b}\,\tilde{\mathbf{x}}_{0|0}^b \;=\; \tilde{\boldsymbol{\xi}}_{0|0}^b + \Pi_0^{-1}\mathbf{x}_0, \tag{15.7.12}$$

where $\tilde{\mathbf{x}}_{0|0}^b \triangleq \mathbf{x}_0 - \hat{\mathbf{x}}_{0|0}^b$ is the backwards error in estimating $\mathbf{x}_0$ using the present and future observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ of (15.7.1) and $P_{0|0}^b$ is the corresponding error Gramian (assumed invertible) (cf. Thm. 9.8.2).

■

Proof: We prove both parts.

(i) The result follows immediately from the orthogonality of the $\{\mathbf{x}_0, \{\mathbf{u}_i\}, \{\mathbf{v}_i\}\}$ and from the facts that

$$\mathbf{x}_i \in \mathcal{L}\{\mathbf{u}_0, \mathbf{v}_0, \ldots \mathbf{u}_{i-1}, \mathbf{v}_{i-1}\} \quad \text{and} \quad \boldsymbol{\eta}_j^b \in \mathcal{L}\{\mathbf{u}_j, \mathbf{v}_j \ldots \mathbf{u}_N, \mathbf{v}_N\}.$$

(ii) This result also follows immediately from the following simple observation. Consider the space $\mathcal{L}\{\mathbf{z}_0^d, \boldsymbol{\eta}^b\}$ and let us compute its dual basis, say $\{\mathbf{a}, \mathbf{b}\}$, in two different ways (one geometric and the other algebraic). By equating the resulting expressions, we shall be able to establish the desired identity.

First, according to the geometric interpretation (15.1.6), the vector $\mathbf{a}$ is given by

$$\mathbf{a} \;=\; M^{-1}\tilde{\mathbf{z}}_0^d, \tag{15.7.13}$$

where M is the Gramian of $\tilde{\mathbf{z}}_0^d$.

Moreover, by applying this same interpretation to the space $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$, where $\mathbf{z} = \text{col}\{\mathbf{x}_0, \mathbf{u}\}$, we know that $\mathbf{z}^d$ itself is given by $\mathbf{z}^d = T^{-1}\,\tilde{\mathbf{z}}_{|\mathbf{y}}$, where T is the Gramian of $\tilde{\mathbf{z}}_{|\mathbf{y}}$. This means that the Gramian of $\mathbf{z}^d$ is equal to T^{-1}. If we now use the algebraic interpretation (15.1.4), we conclude that the desired dual basis $\{\mathbf{a}, \mathbf{b}\}$ is given by

$$\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} \;=\; \left(T^{-1}\right)^{-1}\mathbf{z}^d \;=\; \tilde{\mathbf{z}}_{|\mathbf{y}} \;=\; \begin{bmatrix}\tilde{\mathbf{x}}_{0|\mathbf{y}}\\ \tilde{\mathbf{u}}_{|\mathbf{y}}\end{bmatrix}.$$

But since $\tilde{\mathbf{x}}_{0|\mathbf{y}} = \tilde{\mathbf{x}}_{0|0}^b$, and using (15.7.13), we conclude that

$$\tilde{\mathbf{x}}_{0|0}^b \;=\; M^{-1}\tilde{\mathbf{z}}_0^d.$$

This further shows that $M^{-1} = P_{0|0}^b$, so that we arrive at the desired conclusion that $P_{0|0}^{-b}\tilde{\mathbf{x}}_{0|0}^b = \tilde{\mathbf{z}}_0^d$. Finally, since the $\{\boldsymbol{\eta}_j^d\}$ are only functions of the disturbances $\{\mathbf{u}_k, \mathbf{v}_k\}$, which are uncorrelated with $\mathbf{x}_0$, we obtain that $\hat{\mathbf{x}}_{0|\boldsymbol{\eta}^b} = 0$, so that we also have $\hat{\mathbf{z}}_0^d = \hat{\boldsymbol{\xi}}_{0|0}^b$.

◆

Remark 12. In fact, a stronger conclusion holds. Observe that the choice of the initial time instant 0 has been arbitrary in our discussion so far. We could have selected any other time instant, say j, as our initial time. If we do so, then arguments similar to what we have employed above would allow us to establish the following results as well.

Introduce the column vectors

$$\mathbf{y} = \text{col}\{\mathbf{y}_j, \ldots, \mathbf{y}_N\}, \quad \mathbf{u} = \text{col}\{\mathbf{u}_j, \ldots, \mathbf{u}_N\}, \quad \mathbf{v} = \text{col}\{\mathbf{v}_j, \ldots, \mathbf{v}_N\},$$

and let $\{\mathbf{z}^d, \mathbf{y}^d\}$ denote the dual basis for the space $\mathcal{L}(\mathbf{z}, \mathbf{y})$, where *now* $\mathbf{z} = \text{col}\{\mathbf{x}_j, \mathbf{u}\}$. Then the entries of $\mathbf{z}^d$ can be decomposed as

$$\mathbf{z}^d \triangleq \begin{bmatrix}\mathbf{z}_j^d\\ \hline \boldsymbol{\eta}_j^b\\ \vdots\\ \boldsymbol{\eta}_N^b\end{bmatrix}, \quad \mathbf{z}_j^d \;=\; \boldsymbol{\xi}_j^b + \Pi_j^{-1}\mathbf{x}_j,$$

where $\{\boldsymbol{\eta}_k^b\}$ are still generated by the complementary model (15.7.9), with time running backwards from N down to j.

Moreover, if we let $\hat{\mathbf{z}}_j^d$ ($\hat{\boldsymbol{\xi}}_{j|j}^b$) denote the projector of $\mathbf{z}_j^d$ ($\boldsymbol{\xi}_j^b$) onto the space spanned by the present and future outputs $\{\boldsymbol{\eta}_j^b, \ldots, \boldsymbol{\eta}_N^b\}$, and denote the resulting estimation error by $\tilde{\mathbf{z}}_j^d$ ($\tilde{\boldsymbol{\xi}}_{j|j}^b$), then it also holds that

$$\tilde{\mathbf{z}}_j^d \;=\; P_{j|j}^{-b}\,\tilde{\mathbf{x}}_{j|j}^b \;=\; \tilde{\boldsymbol{\xi}}_{j|j}^b + \Pi_j^{-1}\,\mathbf{x}_j, \tag{15.7.14}$$

where $\tilde{\mathbf{x}}_{j|j}^b \triangleq \mathbf{x}_j - \hat{\mathbf{x}}_{j|j}^b$ is the backwards error in estimating $\mathbf{x}_j$ using the present and future observations $\{\mathbf{y}_j, \ldots, \mathbf{y}_N\}$, $P_{j|j}^b$ is the corresponding error Gramian (assumed invertible), and Π_j is the Gramian of $\mathbf{x}_j$ (*cf.* Thm. 9.8.2). ◆

15.7.3 Direct Derivation of the Hamiltonian Equations

A nice application of the backwards complementary state-space model is to show that the Hamiltonian equations that we encountered in the context of smoothing problems (see expression (10.5.3)) can be derived directly by jointly studying the direct and complementary state-space models. [Recall that in Ch. 10, we derived the Hamiltonian equations as a consequence of various smoothing formulas; the following argument was suggested by G. Verghese — see Wienert and Desai (1981).]

Indeed, combining the state-space models (15.7.1) and (15.7.9), for the original and the complementrary systems, we can write

$$\begin{bmatrix}\mathbf{x}_{i+1}\\ \boldsymbol{\xi}_i^b\end{bmatrix} = \begin{bmatrix}F_i & 0\\ 0 & F_i^*\end{bmatrix}\begin{bmatrix}\mathbf{x}_i\\ \boldsymbol{\xi}_{i+1}^b\end{bmatrix} + \begin{bmatrix}G_i & 0\\ 0 & -H_i^*R_i^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{u}_i\\ \mathbf{v}_i\end{bmatrix},$$

$$\begin{bmatrix}\mathbf{y}_i\\ \boldsymbol{\eta}_i^b\end{bmatrix} = \begin{bmatrix}H_i & 0\\ 0 & G_i^*\end{bmatrix}\begin{bmatrix}\mathbf{x}_i\\ \boldsymbol{\xi}_{i+1}^b\end{bmatrix} + \begin{bmatrix}0 & I\\ Q_i^{-1} & 0\end{bmatrix}\begin{bmatrix}\mathbf{u}_i\\ \mathbf{v}_i\end{bmatrix}.$$

If we now project both sides of the above equations onto the linear space spanned by the observations $\{\mathbf{y}_j,\ j = 0, \ldots, N\}$, and hence determine the smoothed estimators

of the quantities $\{\mathbf{x}_i, \xi_i^b, \eta_i^b, \mathbf{y}_i\}$, we obtain, using the orthogonality of the $\{\eta_i^b\}$ and the $\{\mathbf{y}_j\}$, the following relations:

$$\begin{bmatrix} \hat{\mathbf{x}}_{i+1|N} \\ \hat{\xi}_{i|\mathbf{y}}^b \end{bmatrix} = \begin{bmatrix} F_i & 0 \\ 0 & F_i^* \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{i|N} \\ \hat{\xi}_{i+1|\mathbf{y}}^b \end{bmatrix} + \begin{bmatrix} G_i & 0 \\ 0 & -H_i^* R_i^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_{i|N} \\ \hat{\mathbf{v}}_{i|N} \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{y}_i \\ 0 \end{bmatrix} = \begin{bmatrix} H_i & 0 \\ 0 & G_i^* \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{i|N} \\ \hat{\xi}_{i+1|\mathbf{y}}^b \end{bmatrix} + \begin{bmatrix} 0 & I \\ Q_i^{-1} & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_{i|N} \\ \hat{\mathbf{v}}_{i|N} \end{bmatrix},$$

where we are writing $\hat{\xi}_{i|\mathbf{y}}^b$ to denote the l.l.m.s. estimator of ξ_i^b given $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$. Similarly, we are writing $\hat{\mathbf{x}}_{i|N}$ to denote the l.l.m.s. estimator of $\mathbf{x}_i$ given the same observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ (likewise for $\hat{\mathbf{u}}_{i|N}$ and $\hat{\mathbf{v}}_{i|N}$).

We can now solve for $\{\hat{\mathbf{u}}_{i|N}, \hat{\mathbf{v}}_{i|N}\}$ from the second relation and substitute into the first relation to obtain

$$\hat{\mathbf{x}}_{i+1|N} = F_i \hat{\mathbf{x}}_{i|N} - G_i Q_i G_i^* \hat{\xi}_{i+1|\mathbf{y}}^b, \tag{15.7.15}$$

$$\hat{\xi}_{i|\mathbf{y}}^b = F_i^* \hat{\xi}_{i+1|\mathbf{y}}^b - H_i^* R_i^{-1} [\mathbf{y}_i - H_i \hat{\mathbf{x}}_{i|N}]. \tag{15.7.16}$$

Observe further that since the top entry of $\mathbf{z}^d$ in (15.7.8) is also orthogonal to the $\{\mathbf{y}_j\}$, we must have

$$0 = \hat{\xi}_{0|\mathbf{y}}^b + \Pi_0^{-1} \hat{\mathbf{x}}_{0|N}. \tag{15.7.17}$$

If we introduce the change of variables $\boldsymbol{\lambda}_{i|N} \triangleq -\hat{\xi}_{i|\mathbf{y}}^b$, we conclude that (15.7.15)–(15.7.17) can be rewritten in the equivalent form (*cf.* (10.5.3))

$$\begin{bmatrix} \hat{\mathbf{x}}_{i+1|N} \\ \boldsymbol{\lambda}_{i|N} \end{bmatrix} = \begin{bmatrix} F_i & G_i Q_i G_i^* \\ -H_i^* R_i^{-1} H_i & F_i^* \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{i|N} \\ \boldsymbol{\lambda}_{i+1|N} \end{bmatrix} + \begin{bmatrix} 0 \\ H_i^* R_i^{-1} \end{bmatrix} \mathbf{y}_i, \tag{15.7.18}$$

with boundary conditions specified at the time instants 0 and N,

$$\hat{\mathbf{x}}_{0|N} = \Pi_0 \boldsymbol{\lambda}_{0|N} \quad \text{and} \quad \boldsymbol{\lambda}_{N+1|N} = 0. \tag{15.7.19}$$

15.7.4 Forwards Complementary Models

As is obvious from the defining relation (15.1.3), the dual basis depends on our choice of the original basis. In the state-space context, $\{\mathbf{x}_0, \mathbf{u}, \mathbf{y}\}$ is just one choice of basis for the underlying space $\mathcal{L}\{\mathbf{x}_0, \mathbf{u}, \mathbf{v}\}$, and alternative dual bases are found when one chooses different bases for $\mathcal{L}\{\mathbf{x}_0, \mathbf{u}, \mathbf{v}\}$. In this and the next section, we shall study the consequences of choosing different bases.

For example, when the matrices $\{F_i\}_{i=0}^N$ are nonsingular, $\{\mathbf{x}_{N+1}, \mathbf{u}, \mathbf{y}\}$ is another basis for $\mathcal{L}\{\mathbf{x}_0, \mathbf{u}, \mathbf{v}\}$, since the matrix obtained in the transformation

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{x}_{N+1} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ \mathcal{C} & \Phi_F & 0 \\ \Gamma & \mathcal{O} & I \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{x}_0 \\ \mathbf{v} \end{bmatrix},$$

is nonsingular. Here $\Phi_F \triangleq F_N F_{N-1} \ldots F_0$ and

$$\mathcal{C} \triangleq \begin{bmatrix} \Phi(N+1,1)G_0 & \Phi(N+1,2)G_1 & \ldots & \Phi(N+1,N+1)G_N \end{bmatrix}.$$

Defining $\mathbf{z}' \triangleq \text{col}\{\mathbf{u}, \mathbf{x}_{N+1}\}$, we can consider $\{\mathbf{z}', \mathbf{y}\}$ as a basis for $\mathcal{L}\{\mathbf{x}_0, \mathbf{u}, \mathbf{y}\}$. Our goal now is to find the dual basis $\{\mathbf{z}'^d, \mathbf{y}'^d\}$.

Let us first verify that the vectors $\{\mathbf{z}', \mathbf{y}\}$ satisfy a linear model. Indeed, using (15.7.2) and the relation

$$\mathbf{x}_{N+1} = \Phi_F \mathbf{x}_0 + \mathcal{C}\mathbf{u},$$

we obtain

$$\begin{aligned} \mathbf{y} &= \mathcal{O}\mathbf{x}_0 + \Gamma\mathbf{u} + \mathbf{v}, \\ &= \mathcal{O}(\Phi_F^{-1}\mathbf{x}_{N+1} - \Phi_F^{-1}\mathcal{C}\mathbf{u}) + \Gamma\mathbf{u} + \mathbf{v}, \\ &= \begin{bmatrix} \Gamma - \mathcal{O}\Phi_F^{-1}\mathcal{C} & \mathcal{O}\Phi_F^{-1} \end{bmatrix} \mathbf{z}' + \mathbf{v}, \end{aligned} \tag{15.7.20}$$

which specifies a linear relation between the vectors $\{\mathbf{z}', \mathbf{y}\}$. With this relation in hand, we can now apply the result of Lemma 15.2.1 in order to obtain the dual basis $\{\mathbf{z}'^d, \mathbf{y}^d\}$. For this purpose, we first note that

$$\mathbf{z}' = \begin{bmatrix} \mathbf{u} \\ \mathbf{x}_{N+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \mathcal{C} & \Phi_F \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{x}_0 \end{bmatrix},$$

so that the Gramian of $\mathbf{z}'$ is given by

$$R_{z'} \triangleq \|\mathbf{z}'\|^2 = \begin{bmatrix} I & 0 \\ \mathcal{C} & \Phi_F \end{bmatrix} \begin{bmatrix} \mathcal{Q} & 0 \\ 0 & \Pi_0 \end{bmatrix} \begin{bmatrix} I & 0 \\ \mathcal{C} & \Phi_F \end{bmatrix}^*.$$

It then follows from Lemma 15.2.1, and from the linear relation (15.7.20), that

$$\begin{aligned} \mathbf{z}'^d &= -\begin{bmatrix} \Gamma^* - \mathcal{C}^*\Phi_F^{-*}\mathcal{O}^* \\ \Phi_F^{-*}\mathcal{O}^* \end{bmatrix} \mathcal{R}^{-1}\mathbf{v} + R_{z'}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{x}_{N+1} \end{bmatrix}, \\ &= -\begin{bmatrix} \Gamma^* - \mathcal{C}^*\Phi_F^{-*}\mathcal{O}^* \\ \Phi_F^{-*}\mathcal{O}^* \end{bmatrix} \mathcal{R}^{-1}\mathbf{v} + \begin{bmatrix} I & 0 \\ \mathcal{C} & \Phi_F \end{bmatrix}^{-*} \begin{bmatrix} \mathcal{Q}^{-1} & 0 \\ 0 & \Pi_0^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{x}_0 \end{bmatrix}, \\ &= -\begin{bmatrix} \Gamma^* - \mathcal{C}^*\Phi_F^{-*}\mathcal{O}^* \\ \Phi_F^{-*}\mathcal{O}^* \end{bmatrix} \mathcal{R}^{-1}\mathbf{v} + \begin{bmatrix} \mathcal{Q}^{-1} & -\mathcal{C}^*\Phi_F^{-*}\Pi_0^{-1} \\ 0 & \Phi_F^{-*}\Pi_0^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{x}_0 \end{bmatrix}, \end{aligned}$$

from which we finally conclude that

$$\mathbf{z}'^{d} = \begin{bmatrix} -\mathcal{C}^*\Phi_F^{-*}\Pi_0^{-1}\mathbf{x}_0 - (\Gamma^* - \mathcal{C}^*\Phi_F^{-*}\mathcal{O}^*)\mathcal{R}^{-1}\mathbf{v} + \mathcal{Q}^{-1}\mathbf{u} \\ \\ \Phi_F^{-*}\Pi_0^{-1}\mathbf{x}_0 - \Phi_F^{-*}\mathcal{O}^*\mathcal{R}^{-1}\mathbf{v} \end{bmatrix}. \tag{15.7.21}$$

This expression allows us to obtain a state-space model for the dual basis $\mathbf{z}'^{d}$.

Lemma 15.7.3 (A Forwards Complementary Model) *Assume the $\{F_i\}$ are invertible in the state-space model (15.7.1). The entries of the vector $\mathbf{z}'^{d}$ can be decomposed as*

$$\mathbf{z}'^{d} = \begin{bmatrix} \boldsymbol{\eta}^f \\ \hline \mathbf{z}'^{d}_{N+1} \end{bmatrix} \triangleq \begin{bmatrix} \boldsymbol{\eta}_0^f \\ \vdots \\ \boldsymbol{\eta}_N^f \\ \hline 2\Phi_F^{-*}\Pi_0^{-1}\mathbf{x}_0 - \boldsymbol{\xi}_{N+1}^f \end{bmatrix}, \tag{15.7.22}$$

where the entries $\{\boldsymbol{\eta}_k^f\}$ are generated via the forwards-time state-space model

$$\begin{cases} \boldsymbol{\xi}_{i+1}^f &= F_i^{-*}\boldsymbol{\xi}_i^f + F_i^{-*}H_i^*R_i^{-1}\mathbf{v}_i, \quad i = 0, 1, \ldots N, \\ \boldsymbol{\eta}_i^f &= -G_i^*F_i^{-*}\boldsymbol{\xi}_i^f - G_i^*F_i^{-*}H_i^*R_i^{-1}\mathbf{v}_i + Q_i^{-1}\mathbf{u}_i, \end{cases} \tag{15.7.23}$$

with initial state $\boldsymbol{\xi}_0^f = \Pi_0^{-1}\mathbf{x}_0$, and where $\boldsymbol{\xi}_{N+1}^f$ is its final state. ∎

Proof: It is straightforward to verify from the state-space equations (15.7.23) that

$$\boldsymbol{\eta}^f = -\mathcal{C}^*\Phi_F^{-*}\Pi_0^{-1}\mathbf{x}_0 - (\Gamma^* - \mathcal{C}^*\Phi_F^{-*}\mathcal{O}^*)\mathcal{R}^{-1}\mathbf{v} + \mathcal{Q}^{-1}\mathbf{u},$$

which coincides with the equation defining $\boldsymbol{\eta}^f$ in (15.7.21). Likewise using (15.7.23) we obtain $\boldsymbol{\xi}_{N+1}^f = \Phi_F^{-*}\Pi_0^{-1}\mathbf{x}_0 + \Phi_F^{-*}\mathcal{O}^*\mathcal{R}^{-1}\mathbf{v}$, so that

$$2\Phi_F^{-*}\Pi_0^{-1}\mathbf{x}_0 - \boldsymbol{\xi}_{N+1}^f = \Phi_F^{-*}\Pi_0^{-1}\mathbf{x}_0 - \Phi_F^{-*}\mathcal{O}^*\mathcal{R}^{-1}\mathbf{v},$$

which is the desired result for the second component of $\mathbf{z}'^{d}$ in (15.7.21). ◆

The forwards-time state-space model (15.7.23) is less well known in system theory and is also referred to as a *complementary* state-space model since its output spans the orthogonal complement space of the output of the original model.

We close this section with two useful properties of the dual vector $\mathbf{z}'^{d}$ in (15.7.22). Thus let $\hat{\mathbf{z}}'^{d}_{N+1}$ denote the projection of the trailing entry of $\mathbf{z}'^{d}$ onto the space spanned by its leading entries, *viz.*, the space $\mathcal{L}\{\boldsymbol{\eta}^f\}$. Let also $\tilde{\mathbf{z}}'^{d}_{N+1}$ denote the resulting estimation error.

Lemma 15.7.4 (Two Properties of $\mathbf{z}'^{d}$) *Assume the $\{F_i\}$ are invertible and consider the dual vector $\mathbf{z}'^{d}$ defined in (15.7.22). The following facts hold:*

(i) The state vector $\mathbf{x}_{N+1}$ of the original state-space model (15.7.1) is orthogonal to all outputs of the complementary model (15.7.23),

$$\mathbf{x}_{N+1} \perp \boldsymbol{\eta}_j^f \quad \text{for} \quad j = 0, 1, \ldots, N.$$

(ii) The estimation error $\tilde{\mathbf{z}}'^{d}_{N+1}$ is given by

$$\tilde{\mathbf{z}}'^{d}_{N+1} = P_{N+1}^{-1}\,\tilde{\mathbf{x}}_{N+1},$$

where $\tilde{\mathbf{x}}_{N+1}$ is the error in estimating $\mathbf{x}_{N+1}$ using the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ of (15.7.1), and P_{N+1} is the corresponding error Gramian (assumed invertible) (cf. Thm. 9.2.1).

∎

Proof: We prove both parts.

(i) Recall that $\mathbf{z}' = \text{col}\{\mathbf{u}, \mathbf{x}_{N+1}\}$ and $\mathbf{z}'^{d} = \text{col}\{\boldsymbol{\eta}^f, \mathbf{z}'^{d}_{N+1}\}$. Using the defining property (15.1.3) we must have $\langle \mathbf{z}'^{d}, \mathbf{z}' \rangle = I$, from which we conclude that $\langle \mathbf{x}_{N+1}, \boldsymbol{\eta}^f \rangle = 0$. That is, $\mathbf{x}_{N+1} \perp \boldsymbol{\eta}_j^f$ for $j = 0, 1, \ldots, N$, as desired.

(ii) The argument here is similar to the proof of (15.7.12) in Lemma 15.7.2, *i.e.*, we show that $M^{-1}\tilde{\mathbf{z}}'^{d}_{N+1} = \tilde{\mathbf{x}}_{N+1}$ with M being the Gramian of $\tilde{\mathbf{z}}'^{d}_{N+1}$. This equality then identifies M as being P_{N+1}^{-1}, and the result follows.

◆

Remark 13. Again, a stronger conclusion holds. Observe that the choice of the final time instant $N+1$ has been arbitrary and we could have selected any other time instant, say j, as our final time. If we do so, then arguments similar to what we have employed above would allow us to establish the following results as well.

Introduce the column vectors

$$\mathbf{y} = \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_{j-1}\}, \quad \mathbf{u} = \text{col}\{\mathbf{u}_0, \ldots, \mathbf{u}_{j-1}\}, \quad \mathbf{v} = \text{col}\{\mathbf{v}_0, \ldots, \mathbf{v}_{j-1}\},$$

and let $\{\mathbf{z}'^{d}, \mathbf{y}'^{d}\}$ denote the dual basis for the space $\mathcal{L}(\mathbf{z}, \mathbf{y})$, where now $\mathbf{z} = \text{col}\{\mathbf{u}, \mathbf{x}_j\}$. Then the entries of $\mathbf{z}'^{d}$ can be decomposed as

$$\mathbf{z}^d \triangleq \begin{bmatrix} \boldsymbol{\eta}_0^f \\ \vdots \\ \boldsymbol{\eta}_{j-1}^f \\ \hline \mathbf{z}_j'^{d} \end{bmatrix}, \quad \mathbf{z}_j'^{d} = 2F_{j-1}^{-*}\cdots F_1^{-*}F_0^{-*}\Pi_j^{-1}\mathbf{x}_j - \boldsymbol{\xi}_j^f,$$

where $\{\eta_k^f\}$ are still generated by the complementary model (15.7.23), with time running forwards from 0 down to $j-1$. Moreover, the state vector $\mathbf{x}_i$ of the original state-space model (15.7.1) will also be orthogonal to the past outputs of the complementary model (15.7.23),

$$\mathbf{x}_i \perp \eta_j^f \quad \text{for} \quad j = 0, 1, \ldots, i-1.$$

If we further let $\hat{\mathbf{z}}_j^{\prime d}$ denote the projector of $\mathbf{z}_j^{\prime d}$ onto the space spanned by the past outputs $\{\eta_0^f, \ldots, \eta_{j-1}^f\}$, and denote the resulting estimation error by $\tilde{\mathbf{z}}_j^{\prime d}$, then it also holds that

$$\tilde{\mathbf{z}}_j^{\prime d} = P_j^{-1}\, \tilde{\mathbf{x}}_j, \tag{15.7.24}$$

where $\tilde{\mathbf{x}}_j = \mathbf{x}_j - \hat{\mathbf{x}}_j$ is the prediction error in estimating $\mathbf{x}_j$ using the past observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_{j-1}\}$, and where P_i is the corresponding error Gramian (assumed invertible) (*cf.* Thm. 9.2.1). ◆

15.7.5 The Mixed Complementary Model

If we continute to assume invertible $\{F_i\}$, then we can also take

$$\{\mathbf{u}_0, \ldots, \mathbf{u}_{i-1},\ \mathbf{x}_i,\ \mathbf{u}_i, \ldots, \mathbf{u}_N,\ \mathbf{y}\} \tag{15.7.25}$$

as a basis for $\mathcal{L}\{\mathbf{x}_0, \mathbf{u}, \mathbf{v}\}$. For this so-called mixed basis we can obtain the following result using the arguments presented in Secs. 15.7.2 and 15.7.4.

Lemma 15.7.5 (Mixed Complementary Model) *Consider the standard state-space model (15.7.1) with invertible $\{F_i\}$. Then the dual basis to (15.7.25) is given by*

$$\{\eta_0^f, \ldots, \eta_{i-1}^f,\ \ \xi_i^b + \mathbf{z}_i^{\prime d},\ \ \eta_i^b, \ldots, \eta_N^b,\ \ \mathcal{R}^{-1}\mathbf{v}\}, \tag{15.7.26}$$

where the $\{\eta_j^b\}$ are generated by the backwards-time complementary model (15.7.9), while the $\{\eta_j^f\}$ are generated by the forwards-time complementary model (15.7.23). ■

Proof: This follows immediately from the result of Prob. 15.6, which tells us how to combine the dual bases of two subspaces in order to obtain a dual basis for their union. The simple argument is given in Prob. 15.7. ◆

15.7.6 An Application to Smoothing

We now apply some of the results of the above discussion on complementary models to a smoothing problem. It turns out that by considering projections onto the backwards and forwards complementary models of Lemmas 15.7.1 and 15.7.3 it is possible to derive the standard BF and RTS smoothing formulas that we encountered earlier in Ch. 10. The algebra involved in these derivations is (at times) rather tedious and therefore we shall not present them here. [The interested reader may refer to (Ackner and Kailath (1989a,1989b)) for details.] Instead we shall illustrate the general method by deriving the two-filter smoothing formulas of Sec. 10.4 by projecting onto the mixed complementrary state-space model of Lemma 15.7.5.

Thus consider again the standard state-space model (15.7.1) with invertible $\{F_i\}$, and suppose that we would like to obtain $\hat{\mathbf{x}}_{i|N}$, the smoothed estimator of the state $\mathbf{x}_i$ given all the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$. To do so, consider the mixed basis (15.7.25) and its dual basis (15.7.26). It follows from the general result (15.1.11) that the estimator

$\hat{\mathbf{x}}_{i|N}$, which agrees with $\hat{\mathbf{x}}_{i|\mathbf{y}}$, can also be obtained by evaluating the *error* in projecting $\mathbf{x}_i$ onto the space

$$\mathcal{L}\{\eta_0^f, \ldots, \eta_{i-1}^f,\ \boxed{\xi_i^b + \mathbf{z}_i^{\prime d}},\ \eta_i^b, \ldots, \eta_N^b\}. \tag{15.7.27}$$

Let us therefore determine first the projection of $\mathbf{x}_i$ onto this space, which we shall denote by $\hat{\mathbf{x}}_{i|\mathbf{w}}$ (so that the desired estimator $\hat{\mathbf{x}}_{i|N}$ is equal to $\mathbf{x}_i - \hat{\mathbf{x}}_{i|\mathbf{w}}$).

Now using the interpretations in parts (ii) of Lemmas 15.7.2 and 15.7.4 (which were extended in Remarks 12 and 13 to arbitrary initial and final time instants), we have that the error in estimating ξ_i^b from $\{\eta_i^b, \ldots \eta_N^b\}$ is equal to

$$\tilde{\xi}_{i|i}^b = P_{i|i}^{-b}\, \tilde{\mathbf{x}}_{i|i}^b - \Pi_i^{-1}\mathbf{x}_i.$$

Likewise, the error in estimating $\mathbf{z}_i^{\prime d}$ from $\{\eta_0^f, \ldots, \eta_{i-1}^f\}$ is equal to

$$\tilde{\mathbf{z}}_i^{\prime d} = P_i^{-1}\tilde{\mathbf{x}}_i.$$

Therefore, the space described by (15.7.27) coincides with the following space

$$\mathcal{L}\{\eta_0^f, \ldots, \eta_{i-1}^f,\ \boxed{P_{i|i}^{-b}\tilde{\mathbf{x}}_{i|i}^b - \Pi_i^{-1}\mathbf{x}_i + P_i^{-1}\tilde{\mathbf{x}}_i},\ \eta_i^b, \ldots, \eta_N^b\}. \tag{15.7.28}$$

We shall instead determine the projection of $\mathbf{x}_i$ onto this alternative description of the space (15.7.27). It further follows that $\tilde{\xi}_{i|i}^b \perp \tilde{\mathbf{z}}_i^{\prime d}$. This is because $\tilde{\xi}_{i|i}^b \in \mathcal{L}\{\mathbf{u}_i, \mathbf{v}_i, \ldots, \mathbf{u}_N, \mathbf{v}_N\}$ and $\tilde{\mathbf{z}}_i^{\prime d} \in \mathcal{L}\{\mathbf{x}_0, \mathbf{u}_0, \mathbf{v}_0, \ldots, \mathbf{u}_{i-1}, \mathbf{v}_{i-1}\}$ (see Eqs. (15.7.9) and (15.7.23)).

Now recall from properties (i) in Lemmas 15.7.2 and 15.7.4, which were again extended in Remarks 12 and 13 to arbitrary initial and final time instants, that

$$\mathbf{x}_i \perp \{\eta_0^f, \ldots, \eta_{i-1}^f\} \quad \text{and} \quad \mathbf{x}_i \perp \{\eta_i^b, \ldots \eta_N^b\}.$$

Therefore to find the projection of $\mathbf{x}_i$ onto the space (15.7.28), we need only project $\mathbf{x}_i$ onto the random variable

$$\boldsymbol{\gamma} \triangleq P_{i|i}^{-b}\tilde{\mathbf{x}}_{i|i}^b - \Pi_i^{-1}\mathbf{x}_i + P_i^{-1}\tilde{\mathbf{x}}_i.$$

In other words, the required projection, $\hat{\mathbf{x}}_{i|\mathbf{w}}$, is given by

$$\hat{\mathbf{x}}_{i|\mathbf{w}} = \langle \mathbf{x}_i, \boldsymbol{\gamma}\rangle\, \|\boldsymbol{\gamma}\|^{-2}\, \boldsymbol{\gamma}.$$

It is easy to verify that

$$\langle \mathbf{x}_i, \boldsymbol{\gamma}\rangle = \langle \mathbf{x}_i,\ P_{i|i}^{-b}\tilde{\mathbf{x}}_{i|i}^b - \Pi_i^{-1}\mathbf{x}_i + P_i^{-1}\tilde{\mathbf{x}}_i\rangle = I - I + I = I,$$

and

$$\begin{aligned}
\|\boldsymbol{\gamma}\|^2 &= \|\tilde{\xi}_{i|i}^b + \tilde{\mathbf{z}}_i^{\prime d}\|^2, \\
&= \|\tilde{\xi}_{i|i}^b\|^2 + \|\tilde{\mathbf{z}}_i^{\prime d}\|^2, \quad \text{since } \tilde{\xi}_{i|i}^b \perp \tilde{\mathbf{z}}_i^{\prime d}, \\
&= \|P_{i|i}^{-b}\tilde{\mathbf{x}}_{i|i}^b - \Pi_i^{-1}\mathbf{x}_i\|^2 + \|P_i^{-1}\tilde{\mathbf{x}}_i\|^2, \\
&= (P_{i|i}^{-b} - \Pi_i^{-1}) + P_i^{-1}.
\end{aligned}$$

Therefore, we can finally write

$$\begin{aligned}\hat{\mathbf{x}}_{i|N} &= \mathbf{x}_i - \hat{\mathbf{x}}_{i|\mathbf{w}},\\ &= \mathbf{x}_i - \left[P_{i|i}^{-b} - \Pi_i^{-1} + P_i^{-1}\right]^{-1}\boldsymbol{\gamma},\end{aligned}$$

from which it follows that

$$\left(P_{i|i}^{-b} - \Pi_i^{-1} + P_i^{-1}\right)\hat{\mathbf{x}}_{i|N} = P_{i|i}^{-b}\,\hat{\mathbf{x}}_{i|i}^{b} + P_i^{-1}\,\hat{\mathbf{x}}_i. \tag{15.7.29}$$

Moreover, computing the Gramian of $\tilde{\mathbf{x}}_{i|N}$ from this equality we readily see that

$$P_{i|N} = \left(P_{i|i}^{-b} - \Pi_i^{-1} + P_i^{-1}\right)^{-1},$$

so that (15.7.29) becomes

$$P_{i|N}^{-1}\,\hat{\mathbf{x}}_{i|N} = \left(P_{i|i}^{-b}\,\hat{\mathbf{x}}_{i|i}^{b} + P_i^{-1}\,\hat{\mathbf{x}}_i\right),$$

which is the two-filter smoother (10.4.3) derived earlier in Ch. 10 via a different method — see also Prob. 15.8.

15.8 COMPLEMENTS

Duality and equivalence have played important roles in estimation theory since its inception. Kalman's independent Riccati-based solutions of the state-space linear quadratic control and estimation problems brought attention to it again in system theory.

Of course, duality concepts are also very prominent in optimization theory, especially in (old and new) linear and nonlinear programming theories; see, *e.g.*, Luenberger (1969) and Nesterov and Nemirovskii (1994). We used an approach here that is drawn from elementary linear algebra and developed it in some detail. However, there is certainly much more that can be done to more fully develop, apply, and generalize the results presented here.

Applications to Linear Quadratic Control Problems. There is a vast and old literature on linear quadratic control methods and it is practically impossible to provide here a comprehensive list of all the relevant works. A good sample is Athans and Falb (1966), Bryson and Ho (1969), Åström (1970), Brockett (1970), Kwakernaak and Sivan (1972), Lewis (1986), Whittle (1990,1996), Anderson and Moore (1990), Brogan (1991), Dorato, Abdallah, and Cerone (1995), Green and Limebeer (1995), Zhou, Doyle, and Glover (1996), and Hassibi, Sayed, and Kailath (1999), in addition to the 1971 issue of the *IEEE Transactions on Automatic Control*, edited by M. Athans, on the LQG problem. All these works contain extensive references.

A common method for solving the LQR and tracking problems is via dynamic programming (see, *e.g.*, Anderson and Moore (1990)). Other methods include the calculus of variations (see Gelfand and Fomin (1963), Bryson and Ho (1969), and Kwakernaak and Sivan (1972)) and Pontryagin's minimum principle (see Pontryagin et al. (1962), Athans and Falb (1966), and Brogan (1991)). In the text we took a different route and solved these problems by first constructing dual stochastic estimation problems.

Sec. 15.7. Complementary State-Space Models. Another result we stressed in this chapter is a derivation of several complementary models and a discussion of their relevance to smoothing problems. The concept of complementary models was originally introduced by Weinert and Desai (1981) (see also Desai, Weinert, and Yusypchuk, (1983)), and later generalized by Adams, Willsky, and Levy (1984a,1984b) to noncausal systems. Kailath and Wax (1984) gave an estimation-theoretic interpretation to the derivation of the models of Weinert and Desai (1981). However, all these references described only the special backwards-time complementary model. Ackner and Kailath (1989a,1989b) showed that a closer analysis of the approach of Kailath and Wax (1984) allows the possibility of complementary models that run forwards in time and of others that involve both forwards and backwards time evolution. This study is closely related to the problem of stochastic realization, which has been much studied, especially by Desai and Pal (1984), by Lindquist and Picci (1996), and by Byrnes and Lindquist (1997). This topic again deserves further discussion and study — here we refer only to the books by Faurre, Clerget, and Germain (1979) and Caines (1988).

Sec. 15.7.3. Direct Derivation of the Hamiltonian Equations. In this section we derived the Hamiltonian equations by combining state-space models for the space spanned by the observations and for its complementary space. In so doing, we obtained a direct derivation of the Hamiltonian equations without explicitly resorting to filtering or smoothing equations, as was the case in Ch. 10. Later in Ch. 17 we shall use this fact to develop a self-contained scattering theory approach to state-space estimation. The approach will allow us to unify many of the results we encountered before, especially in smoothing problems, and also to derive several other results that are difficult to obtain by direct methods.

PROBLEMS

15.1 **(Linear spans)** Consider the linear space spanned by $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$. Let $\hat{\mathbf{z}}_{|\mathbf{y}}$ = the projection of $\mathbf{z}$ onto $\mathcal{L}\{\mathbf{y}\}$, and introduce the error $\tilde{\mathbf{z}} = \tilde{\mathbf{z}}_{|\mathbf{y}} = \mathbf{z} - \hat{\mathbf{z}}_{|\mathbf{y}}$. Show that $\{\tilde{\mathbf{z}}, \mathbf{y}\}$ and $\{\mathbf{z}^d, \mathbf{y}\}$ span the same linear space. Show also that $\{\mathbf{z}, \mathbf{y}\}$ and $\{\mathbf{z}^d, \mathbf{y}\}$ span the same linear space. *Hint.* Take any vector in $\mathcal{L}\{\mathbf{z}^d, \mathbf{y}\}$, say $p\mathbf{z}^d + q\mathbf{y}$ for some row vectors $\{p, q\}$, and show that it can be expressed as $a\tilde{\mathbf{z}} + b\mathbf{y}$ for some row vectors $\{a, b\}$. Verify further that the converse statement also holds.]

15.2 **(Invertible Gramians)** Consider the linear space spanned by $\mathcal{L}\{\mathbf{z}, \mathbf{y}\}$, and define $\hat{\mathbf{z}}_{|\mathbf{y}}$ and $\tilde{\mathbf{z}}$ as in Prob. 15.1. Let also $\hat{\mathbf{y}}_{|\mathbf{z}}$ denote the the projection of $\mathbf{y}$ onto $\mathcal{L}\{\mathbf{z}\}$, and define $\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_{|\mathbf{z}} = \mathbf{y} - \hat{\mathbf{y}}_{|\mathbf{z}}$. Let $R_{\tilde{z}}$ and $R_{\tilde{y}}$ denote the Gramian matrices of $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{y}}$, respectively.

(a) Determine $\{R_{\tilde{z}}, R_{\tilde{y}}\}$ in terms of $\{R_z, R_y, R_{zy}\}$.

(b) Assume R_z, R_y, and $\begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}$ are all nonsingular matrices. Show that $R_{\tilde{z}}$ and $R_{\tilde{y}}$ are also nonsingular. [*Hint.* Introduce the block triangular factorizations of the above Gramian matrix of $\{\mathbf{z}, \mathbf{y}\}$.]

15.3 **(Cost functions for equivalent problems)** Consider the cost functions (15.3.2) and (15.3.5) that are associated with two equivalent stochastic and deterministic estimation problems (in that they lead to the same gain matrix K_o).

(a) Verify that the matrix-valued cost function (15.3.2) can be written as

$$\min_K \left(\begin{bmatrix} I & -K \end{bmatrix} \begin{bmatrix} R_z & R_z H^* \\ HR_z & HR_z H^* + R_v \end{bmatrix} \begin{bmatrix} I \\ -K^* \end{bmatrix} \right).$$

(b) Verify also that the scalar-valued cost function (15.3.5) can be written as

$$\min_z \begin{bmatrix} 1 & -z^* \end{bmatrix} \begin{bmatrix} y^* R_v^{-1} y & y^* R_v^{-1} H \\ HR_v^{-1} y & R_z^{-1} + H^* R_v^{-1} H \end{bmatrix} \begin{bmatrix} 1 \\ -z \end{bmatrix},$$

which is also identical to

$$\min_z \begin{bmatrix} z^* & y^* \end{bmatrix} \begin{bmatrix} R_z & R_z H^* \\ HR_z & HR_z H^* + R_v \end{bmatrix}^{-1} \begin{bmatrix} z \\ y \end{bmatrix}.$$

Remark. We thus see that while both Problems (15.3.2) and (15.3.5) lead to the same expression for the optimum gain matrix K_o in $\hat{\mathbf{z}}_{|\mathbf{y}} = K_o \mathbf{y}$ and $\hat{z} = K_o y$, one cost function is matrix-valued and involves the Gramian matrix of $\{\mathbf{z}, \mathbf{y}\}$ while the other cost is scalar-valued and involves the inverse Gramian matrix. ◆

15.4 **(Cost functions for dual problems)** Consider the cost functions (15.3.8) and (15.3.10) that are associated with two equivalent stochastic and deterministic estimation problems (in that they lead to the same gain matrix K_o^d).

(a) Verify that the matrix-valued cost function (15.3.8) can be written as

$$\min_{K^d} \begin{bmatrix} -K^d & I \end{bmatrix} \begin{bmatrix} R_{z^d} & R_{z^d y^d} \\ R_{y^d z^d} & R_{y^d} \end{bmatrix} \begin{bmatrix} -K^{d*} \\ I \end{bmatrix},$$

which is also identical to

$$\min_{K^d} \begin{bmatrix} -K^d & I \end{bmatrix} \begin{bmatrix} R_z & R_z H^* \\ HR_z & HR_z H^* + R_v \end{bmatrix}^{-1} \begin{bmatrix} -K^{d*} \\ I \end{bmatrix}.$$

(b) Verify also that the scalar-valued cost function (15.3.10) can be written as

$$\min_{y^d} \begin{bmatrix} z^{d*} & y^{d*} \end{bmatrix} \begin{bmatrix} R_z & R_z H^* \\ HR_z & HR_z H^* + R_v \end{bmatrix} \begin{bmatrix} z^d \\ y^d \end{bmatrix}.$$

Remark. Comparing the expression of part (a) of this problem with that in part (a) of Prob. 15.3, we see that both are quadratic cost functions in the respective unknowns K and K^d and have similar forms, except that the central matrix appearing in one is the inverse of the central matrix appearing in the other. We therefore see that by replacing the central matrix with its inverse, and by switching the orders of K and I, the solution of a quadratic optimization problem is replaced by its dual. A similar conclusion holds for the expressions of parts (b) in both problems. ◆

15.5 **(General equivalences and dualities)** The discussion in Sec. 15.3, and in Probs. 15.3–15.4, will still hold if we remove the assumption of a linear model (*cf.* (15.3.1)). Indeed, verify the validity of the results shown in Table 15.2.

Table 15.2 General equivalences and dualities, assuming invertible Gramian matrices.

Stochastic problems	Deterministic problems
(i) **Given**: $\{\mathbf{z}, \mathbf{y}\}$ with $\langle \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix}, \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \rangle = \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}$ **Solve**: $\min_{\hat{\mathbf{z}} \in \mathcal{L}\{\mathbf{y}\}} \|\mathbf{z} - \hat{\mathbf{z}}\|^2$ **Solution**: $\hat{\mathbf{z}} = K_o \mathbf{y}$ with $K_o = R_{zy} R_y^{-1}$ **Min. cost**: $R_z - R_{zy} R_y^{-1} R_{yz}$	(ii) $\{z^d, y^d\}$ $\min_{y^d} \begin{bmatrix} z^{d*} & y^{d*} \end{bmatrix} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix} \begin{bmatrix} z^d \\ y^d \end{bmatrix}$ $y_o^d = K_o^d z^d$ with $K_o^d = -R_y^{-1} R_{yz}$ $z^{d*}(R_z - R_{zy} R_y^{-1} R_{yz}) z^d$
(iii) **Given**: $\{\mathbf{z}^d, \mathbf{y}^d\}$ with $\langle \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix}, \begin{bmatrix} \mathbf{z}^d \\ \mathbf{y}^d \end{bmatrix} \rangle = \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1}$ **Solve**: $\min_{\hat{\mathbf{y}}^d \in \mathcal{L}\{\mathbf{z}^d\}} \|\mathbf{y}^d - \hat{\mathbf{y}}^d\|^2$ **Solution**: $\hat{\mathbf{y}}^d = K_o^d \mathbf{z}^d$ with $K_o^d = -R_y^{-1} R_{yz}$ **Min. cost**: R_y^{-1}	(iv) $\{z, y\}$ $\min_z \begin{bmatrix} z^* & y^* \end{bmatrix} \begin{bmatrix} R_z & R_{zy} \\ R_{yz} & R_y \end{bmatrix}^{-1} \begin{bmatrix} z \\ y \end{bmatrix}$ $z_o = K_o y$ with $K_o = R_{zy} R_y^{-1}$ $y^* R_y^{-1} y$

15.6 **(Partitioned bases)** Consider random variables $\mathbf{r}$, $\{\mathbf{y}_a, \mathbf{y}_b\}$, $\{\mathbf{v}_a, \mathbf{v}_b\}$, and $\{\mathbf{z}_a, \mathbf{z}_b\}$ such that

$$\mathbf{y}_a = \begin{bmatrix} H_1 & H_2 \end{bmatrix} \begin{bmatrix} \mathbf{z}_a \\ \mathbf{r} \end{bmatrix} + \mathbf{v}_a, \qquad \mathbf{y}_b = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{z}_b \end{bmatrix} + \mathbf{v}_b,$$

for some matrices $\{H_1, H_2, A_1, A_2\}$. Assume further that all required Gramian matrices are invertible, and that

$$\mathbf{v}_a \perp \{\mathbf{v}_b, \mathbf{z}_a, \mathbf{z}_b, \mathbf{r}\}, \quad \mathbf{v}_b \perp \{\mathbf{z}_a, \mathbf{z}_b, \mathbf{r}\}, \quad \mathbf{z}_b \perp \{\mathbf{z}_a, \mathbf{r}\}.$$

[Note that we are only requiring $\mathbf{r}$ to be orthogonal to $\mathbf{z}_b$ and not to $\mathbf{z}_a$.]

(a) Verify that the dual basis for the linear space $\mathcal{L}\{\mathbf{z}_a, \mathbf{r}, \mathbf{y}_a\}$ is given by $\{\mathbf{z}_a^d, \mathbf{r}_1^d, \mathbf{y}_a^d\}$, where

$$\begin{bmatrix} \mathbf{z}_a^d \\ \mathbf{r}_1^d \end{bmatrix} = -\begin{bmatrix} H_1^* \\ H_2^* \end{bmatrix} R_{v_a}^{-1}\mathbf{v}_a + \begin{bmatrix} R_{z_a} & R_{z_a,r} \\ R_{r,z_a} & R_r \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{z}_a \\ \mathbf{r} \end{bmatrix}, \quad \mathbf{y}_a^d = R_{v_a}^{-1}\mathbf{v}_a.$$

Here we used the obvious notations

$$R_{z_a} = \|\mathbf{z}_a\|^2, \quad R_r = \|\mathbf{r}\|^2, \quad R_{v_a} = \|\mathbf{v}_a\|^2, \quad R_{z_a,r} = \langle \mathbf{z}_a, \mathbf{r} \rangle.$$

(b) Verify also the dual basis for the linear space $\mathcal{L}\{\mathbf{r}, \mathbf{z}_b, \mathbf{y}_b\}$ is given by $\{\mathbf{r}_2^d, \mathbf{z}_b^d, \mathbf{y}_b^d\}$, where

$$\begin{bmatrix} \mathbf{r}_2^d \\ \mathbf{z}_b^d \end{bmatrix} = -\begin{bmatrix} A_1^* \\ A_2^* \end{bmatrix} R_{v_b}^{-1}\mathbf{v}_b + \begin{bmatrix} R_r^{-1} & 0 \\ 0 & R_{z_b}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{z}_b \end{bmatrix}, \quad \mathbf{y}_b^d = R_{v_b}^{-1}\mathbf{v}_b.$$

(c) Now use the combined linear relation

$$\begin{bmatrix} \mathbf{y}_a \\ \mathbf{y}_b \end{bmatrix} = \begin{bmatrix} H_1 & H_2 & 0 \\ 0 & A_1 & A_2 \end{bmatrix} \begin{bmatrix} \mathbf{z}_a \\ \mathbf{r} \\ \mathbf{z}_b \end{bmatrix} + \begin{bmatrix} \mathbf{v}_a \\ \mathbf{v}_b \end{bmatrix},$$

to conclude that the dual basis for the linear space $\mathcal{L}\{\mathbf{z}_a, \mathbf{r}, \mathbf{z}_b, \mathbf{y}_a, \mathbf{y}_b\}$ is given by

$$\left\{\mathbf{z}_a^d, \quad \mathbf{r}_1^d + \mathbf{r}_2^d - R_r^{-1}\mathbf{r}, \quad \mathbf{z}_b^d, \quad R_{v_a}^{-1}\mathbf{v}_a, \quad R_{v_b}^{-1}\mathbf{v}_b\right\}.$$

That is, we keep all the entries of the dual bases of parts (a) and (b) except for one, which is obtained via a proper combination as shown above.

Remark. This problem tells us how to combine the dual bases of two subspaces in order to obtain a dual basis for their union. ◆

15.7 **(A mixed basis)** This problem establishes the statement of Lemma 15.7.5 by using the general result of Prob. 15.6. Thus introduce the column vectors

$$\begin{aligned} \mathbf{y}_a &= \text{col}\{\mathbf{y}_0, \ldots, \mathbf{y}_{i-1}\}, & \mathbf{y}_b &= \text{col}\{\mathbf{y}_i, \ldots, \mathbf{y}_N\}, \\ \mathbf{v}_a &= \text{col}\{\mathbf{v}_0, \ldots, \mathbf{v}_{i-1}\}, & \mathbf{v}_b &= \text{col}\{\mathbf{v}_i, \ldots, \mathbf{v}_N\}, \\ \mathbf{u}_a &= \text{col}\{\mathbf{u}_0, \ldots, \mathbf{u}_{i-1}\}, & \mathbf{u}_b &= \text{col}\{\mathbf{u}_i, \ldots, \mathbf{u}_N\}, \end{aligned}$$

as well as the matrices

$$\mathcal{O}_a \triangleq \text{col}\{H_i\Phi(i,i), \ldots, H_N\Phi(N,i)\}, \quad \mathcal{C}_a \triangleq \begin{bmatrix} \Phi(i,1)G_0 & \ldots & \Phi(i,i)G_{i-1} \end{bmatrix},$$

and the (strictly lower triangular) impulse response matrix

$$\Gamma_a \triangleq \begin{bmatrix} 0 & & & & \\ \Gamma_{a,10} & 0 & & & \\ \Gamma_{a,20} & \Gamma_{a,21} & 0 & & \\ \Gamma_{a,30} & \Gamma_{a,31} & \Gamma_{a,32} & 0 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix},$$

where the entries of Γ_a are given by $\Gamma_{a,jk} \triangleq H_{i+j}\Phi(i+j, i+k+1)G_{i+k}$, with $\Phi(k,j) = F_{k-1}F_{k-2}\ldots F_j$ for $k > j$ and $\Phi(k,k) = I_.$. Define also $\Phi_a = F_{i-1}\ldots F_1F_0$.

(a) Use the state-space model (15.7.1), and the arguments of Secs. 15.7.2–15.7.4, to show that the following linear relations hold:

$$\mathbf{y}_a = \begin{bmatrix} \Gamma_a - \mathcal{O}_a\Phi_a^{-1}\mathcal{C}_a & \mathcal{O}_a\Phi_F^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{u}_a \\ \mathbf{x}_i \end{bmatrix} + \mathbf{v}_a,$$

$$\mathbf{y}_b = \begin{bmatrix} \mathcal{O}_a & \Gamma_a \end{bmatrix} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{u}_b \end{bmatrix} + \mathbf{v}_b.$$

(b) Conclude from the argument that led to Lemma 15.7.3 that the dual basis for the space $\mathcal{L}\{\mathbf{u}^a, \mathbf{x}_i, \mathbf{y}^a\}$ is given by $\{\eta_0^f, \ldots, \eta_{i-1}^f, \mathbf{z}_i'^d, \mathcal{R}_a^{-1}\mathbf{v}_a\}$, where $\mathcal{R}_a = \|\mathbf{v}_a\|^2$.

(c) Conclude from the argument that led to Lemma 15.7.1 that the dual basis for the space $\mathcal{L}\{\mathbf{x}_i, \mathbf{u}^b, \mathbf{y}^b\}$ is given by $\{\mathbf{z}_i^d, \eta_i^b, \ldots, \eta_N^b, \mathcal{R}_b^{-1}\mathbf{v}_b\}$, where $\mathcal{R}_b = \|\mathbf{v}_b\|^2$.

(d) Now use the result of Prob. 15.6 to conclude that (15.7.26) is the dual basis corresponding to (15.7.25).

15.8 **(Two-filter smoothing formula)** Refer to the discussion in Sec. 15.7.5 and consider the basis $\{\mathbf{u}_0, \ldots, \mathbf{u}_i, \mathbf{x}_{i+1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_N, \mathbf{y}\}$.

(a) Show that the dual basis is given by $\{\eta_0^f, \ldots, \eta_i^f, \boxed{\xi_{i+1}^b + \mathbf{z}_{i+1}'^d}, \eta_{i+1}^b, \ldots, \eta_N^b\}$.

(b) Show that $\xi_{i+1}^b + \mathbf{z}_{i+1}'^d = F_i^{-*}\left[\xi_i^b + \mathbf{z}_i'^d\right]$, and conclude that the linear space of part (a) is also spanned by the following basis: $\{\eta_0^f, \ldots, \eta_i^f, \boxed{\xi_i^b + \mathbf{z}_i'^d}, \eta_{i+1}^b, \ldots, \eta_N^b\}$.

(c) Repeat the arguments of Sec. 15.7.6 in order to rederive the two-filter smoothing formula (10.4.1), *viz.*,

$$P_{i|N}^{-1}\,\hat{\mathbf{x}}_{i|N} = P_i^{-b}\,\hat{\mathbf{x}}_i^b + P_{i|i}^{-1}\,\hat{\mathbf{x}}_{i|i}, \quad P_{i|N}^{-1} = \left(P_i^{-b} + P_{i|i}^{-1} - \Pi_i^{-1}\right)^{-1},$$

where $\hat{\mathbf{x}}_i^b$, $\hat{\mathbf{x}}_{i|i}$, P_i^b, and $P_{i|i}$ have the obvious meanings.

15.9 **(Optimal LQR cost)** Refer to the solution of the LQR problem in Sec. 15.3.6. Let $s_i = H_i x_i$.

(a) Show that it holds, for all j,

$$u_j^* Q_j^d u_j + s_j^* R_j^d s_j + x_{j+1}^* P_{j+1|j+1}^d x_{j+1} = x_j^* P_{j|j}^d x_j + (u_j - \hat{u}_j)^* R_{e,j}^d (u_j - \hat{u}_j).$$

(b) Verify that the LQR cost function

$$J_N^d(x_0, u_0, \ldots, u_N) = x_{N+1}^* P_{N+1}^d x_{N+1} + \sum_{i=0}^{N} u_i^* Q_i^d u_i + \sum_{i=0}^{N} s_i^* R_i^d s_i$$

can be expressed as

$$J_N^d(x_0, u_0, \ldots, u_N) = x_0^* P_{0|0}^d x_0 + \sum_{i=0}^{N} (u_i - \hat{u}_i)^* R_{e,i}^d (u_i - \hat{u}_i),$$

and conclude that the minimum cost is equal to $x_0^* P_{0|0}^d x_0$.

Remark. In fact, Legendre (1810) introduced the Riccati differential equation to derive the local identity of part (a) for the simplest scalar and continuous-time version of this problem. ◆

15.10 **(Return difference identity)** Refer to Sec. 15.6.3 on the infinite-horizon LQR problem. Introduce the transfer functions

$$T(z) \triangleq R_e^{-d} K^{d*} (zI - F)^{-1} G, \qquad P(z) \triangleq -H(zI - F)^{-1} G.$$

(a) Establish the identity

$$Q^d + P^*(z^{-*}) R^d P(z) = \{I + T^*(z^{-*})\} R_e^d \{I + T(z)\}.$$

(b) Define $D(z) \triangleq R_e^{d*/2} [I + T(z)] Q^{d-*/2}$, and conclude from part (a) that

$$\sigma_{\min}[D(e^{j\omega})] \geq 1 \quad \text{for all} \quad \omega \in [-\pi, \pi],$$

where $\sigma_{\min}[\cdot]$ denotes the minimum singular value of its argument.

Remark. The matrix $I + T(z)$ is usually called the *return difference matrix*. The inequality in part (b) was first derived in continuous time, and for the single input case, by Kalman (1964) in his studies of the inverse control problem. The identity in part (a) can be regarded as the dual of a canonical spectral factorization of the form (*cf.* Ch. 8):

$$R + H(zI - F)^{-1} G Q G^* (z^{-1} I - F^*)^{-1} H^* = L(z) R_e L^*(z^{-*}),$$

with

$$L(z) = I + H(zI - F)^{-1} K_p, \quad K_p = FPH^* R_e^{-1}, \quad R_e = R + HPH^*.$$

◆

CHAPTER 16

Continuous-Time State-Space Estimation

Most of our discussions so far in this book have been on discrete-time signals and systems. However many physical systems evolve continuously in time, as do many physical signals. In this chapter we cover the fundamentals of state-space filtering and smoothing for continuous-time state-space models. This topic can be, and often is, the subject of a book by itself. Here we shall have to be much briefer. While following the lines of the presentation for the discrete-time case in the earlier chapters, we shall attempt to highlight points that are specific to the continuous-time setting.

There are two basic ways of approaching the study of continuous-time problems — directly or via reduction to an equivalent, or more often an approximate, discrete-time problem. We shall start with the second approach and use it to build up our understanding of the continuous-time formulas. However, to always proceed in this way, *e.g.*, for smoothing problems or for fast algorithms, will be seen to be very cumbersome and a direct treatment is better, as we shall illustrate from Sec. 16.4 onwards.

16.1 CONTINUOUS-TIME MODELS

We first present the standard continuous-time state-space model, and then describe one straightforward way of getting a discrete-time approximation. Then we can apply the results for the discrete-time case and return to the continuous solution in the limit. There are several methods of discretization of continuous-time models, especially when one is interested in numerical computation. Our interest is purely academic — to use a method that allows easy transition between discrete-time and continuous-time formulas.

16.1.1 Standard Continuous-Time Models

The standard continuous-time state-space model is of the form

$$\dot{\mathbf{x}}(t) = F(t)\mathbf{x}(t) + G(t)\mathbf{u}(t), \tag{16.1.1}$$

$$\mathbf{y}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t), \qquad t \geq 0, \tag{16.1.2}$$

where $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$ are white-noise processes such that

$$\left\langle \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{v}(t) \\ \mathbf{x}(0) \end{bmatrix}, \begin{bmatrix} \mathbf{u}(s) \\ \mathbf{v}(s) \\ \mathbf{x}(0) \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q(t)\delta(t-s) & S(t)\delta(t-s) & 0 & 0 \\ S^*(t)\delta(t-s) & R(t)\delta(t-s) & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}, \tag{16.1.3}$$

where $\langle \mathbf{a}(t), \mathbf{b}(s) \rangle = E\mathbf{a}(t)\mathbf{b}^*(s)$, for zero-mean random processes $\{\mathbf{a}(\cdot), \mathbf{b}(\cdot)\}$. These equations are clearly quite analogous to those for the standard discrete-time model (see Sec. 9.1), with the analogy becoming even clearer after we discuss discrete-time approximations in Sec. 16.1.2. The major difference is the presence of continuous-time white-noise processes $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$. Engineers use these obviously nonphysical processes as approximations to wideband noise processes. Here we note only that there are also certain mathematical issues in the treatment of white-noise processes in the standard theory of stochastic processes, which means also that the process $\dot{\mathbf{x}}(t)$ in (16.1.1) cannot be directly handled in the conventional theory.

Special definitions of stochastic integrals have to be introduced and equations such as (16.1.1) have to be regarded as a shorthand for a more formal (integral) version. For *linear* least-mean-squares estimation problems, there is no need to introduce this more formal theory and one can proceed quite satisfactorily with the now-usual methods of working with white-noise processes. The problem is analogous to avoiding the need for delta functions in deterministic system theory by first working with step functions and then taking (formal) derivatives — engineers (and others) have long since learned to work comfortably with delta functions. The same holds for studies of stochastic linear systems. One might also mention that reputed mathematicians such as N. Wiener and J. L. Doob (see, *e.g.*, Doob (1953, pp. 435, 538, 638)) did not hesitate to work formally with white noise when it helped simplify the arguments, *e.g.*, by introducing Fourier transforms of stationary random processes. One should also note that rigorous treatments are available and not particularly difficult, see especially the elegant monograph of Davis (1977). In this book, a less formal treatment will be quite adequate; there are only a couple of points at which special care has to be paid to handling the white-noise processes more carefully, and we shall do this when needed (see, *e.g.*, Prob. 16.19).

16.1.2 Discrete-Time Approximations

There are many methods of obtaining discrete-time approximations to continuous-time problems. Here we shall describe one that has the virtue of being simple and of allowing an easy deduction of continuous-time estimation formulas from their discrete-time counterparts; it is not a particularly good one from the viewpoint of numerical computation.

The procedure is based on approximating a function $m(\cdot)$ as

$$m(t) \approx \sum_{i=-\infty}^{\infty} m_i \; p(t - i\Delta),$$

where we define the rectangular pulse

$$p(t) \triangleq \begin{cases} \frac{1}{\Delta} & 0 \leq t \leq \Delta, \\ 0 & \text{elsewhere,} \end{cases}$$

with $\Delta =$ an arbitrary (usually small) interval, and

$$m_i = \frac{\int_{i\Delta}^{(i+1)\Delta} m(t)p(t - i\Delta)dt}{\int_{i\Delta}^{(i+1)\Delta} p^2(t - i\Delta)dt} = \int_{i\Delta}^{(i+1)\Delta} m(t)dt, \tag{16.1.4}$$

$$\approx m(i\Delta)\;\Delta, \quad \text{if } m(\cdot) \text{ is continuous and } \Delta \text{ is small.} \tag{16.1.5}$$

So the approximation that we shall use for *continuous* functions is

$$m(t) \approx \sum_{i=-\infty}^{\infty} m(i\Delta)\Delta \; p(t - i\Delta).$$

For continuously differentiable functions, we shall have

$$\frac{d}{dt}m(t) = \dot{m}(t) \approx \sum_{i=-\infty}^{\infty} \dot{m}(i\Delta)\Delta \; p(t - i\Delta),$$

$$\approx \sum_{i=-\infty}^{\infty} [m(i\Delta + \Delta) - m(i\Delta)] \; p(t - i\Delta).$$

Consider now a random process $\mathbf{n}(\cdot)$ with $E\mathbf{n}(t) = 0$ and $E\mathbf{n}(t)\mathbf{n}^*(s) = R(t, s)$. Then we have the approximation

$$\mathbf{n}(t) \approx \sum_{i=-\infty}^{\infty} \mathbf{n}_i \; p(t - i\Delta),$$

where

$$\mathbf{n}_i = \int_{i\Delta}^{(i+1)\Delta} \mathbf{n}(t)dt.$$

Note that $E\mathbf{n}_i = 0$, and

$$E\mathbf{n}_i\mathbf{n}_j^* = \int_{j\Delta}^{(j+1)\Delta} \int_{i\Delta}^{(i+1)\Delta} E\mathbf{n}(t)\mathbf{n}^*(s)dtds,$$

$$= \int_{j\Delta}^{(j+1)\Delta} \int_{i\Delta}^{(i+1)\Delta} R(t, s)dtds, \tag{16.1.6}$$

$$\approx \Delta^2 \; R(i\Delta, j\Delta) \qquad \text{if } R(t, s) \text{ is continuous in } t \text{ and } s.$$

If the process $\mathbf{n}(\cdot)$ has continuous sample paths, we can write $\mathbf{n}_i \approx \mathbf{n}(i\Delta)\Delta$ so that, as expected,

$$E\mathbf{n}(i\Delta)\mathbf{n}^*(j\Delta) = \frac{E\mathbf{n}_i\mathbf{n}_j^*}{\Delta^2} \approx \frac{\Delta^2 R(i\Delta, j\Delta)}{\Delta^2} = R(i\Delta, j\Delta).$$

For white-noise $\mathbf{u}(\cdot)$, however, we cannot use the sample-value representation (16.1.5) because the paths of $\mathbf{u}(\cdot)$ are not continuous. We use (16.1.4) instead and write (*cf.* (16.1.6))

$$E\mathbf{u}_i\mathbf{u}_j^* = \int_{j\Delta}^{(j+1)\Delta}\int_{i\Delta}^{(i+1)\Delta} Q(t)\delta(t-s)dtds.$$

This is zero when $i \neq j$, and when $i = j$, it is

$$E\mathbf{u}_i\mathbf{u}_i^* = \int_{i\Delta}^{(i+1)\Delta} Q(t)dt \approx \Delta Q(i\Delta),$$

assuming that $Q(\cdot)$ is a continuous function. In other words, we shall have

$$E\mathbf{u}_i\mathbf{u}_j^* \approx \Delta Q(i\Delta)\delta_{ij}.$$

However to maintain uniformity of notation, we can introduce a "fictitious" sample value $\mathbf{u}(i\Delta)$ defined via

$$\mathbf{u}_i \triangleq \mathbf{u}(i\Delta)\Delta,$$

so that

$$E\mathbf{u}(i\Delta)\mathbf{u}^*(j\Delta) = \frac{E\mathbf{u}_i\mathbf{u}_j^*}{\Delta^2} = \frac{Q(i\Delta)}{\Delta}\delta_{ij}. \tag{16.1.7}$$

So our representation for white noise $\mathbf{u}(\cdot)$ will be

$$\mathbf{u}(t) \approx \sum_{i=-\infty}^{\infty} \mathbf{u}(i\Delta)\Delta\ p(t - i\Delta),$$

where the $\{\mathbf{u}(i\Delta)\}$ are uncorrelated (zero mean) random variables obeying (16.1.7). So also for the white noise process $\mathbf{v}(\cdot)$ with

$$E\mathbf{v}(t) = 0, \qquad E\mathbf{v}(t)\mathbf{v}^*(s) = R(t)\delta(t-s),$$

we shall write

$$\mathbf{v}(t) \approx \sum_{i=-\infty}^{\infty} \mathbf{v}(i\Delta)\Delta\ p(t - i\Delta),$$

where

$$E\mathbf{v}(i\Delta) = 0, \qquad E\mathbf{v}(i\Delta)\mathbf{v}^*(j\Delta) = \frac{R(i\Delta)}{\Delta}\delta_{ij}. \tag{16.1.8}$$

Though the above discussion has tacitly assumed scalar-valued functions, the discussion easily extends to vector-valued functions, as enter in our state-space model (16.1.1)–(16.1.2), to which we now return.

An Approximate State-Space Model. Using the procedure described above, we can approximate the continuous-time state-space model (16.1.1)–(16.1.2) by the discrete-time model

$$\frac{\mathbf{x}(i\Delta + \Delta) - \mathbf{x}(i\Delta)}{\Delta} = F(i\Delta)\mathbf{x}(i\Delta) + G(i\Delta)\mathbf{u}(i\Delta),$$

$$\mathbf{y}(i\Delta) = H(i\Delta)\mathbf{x}(i\Delta) + \mathbf{v}(i\Delta), \qquad i \geq 0,$$

or, equivalently,

$$\mathbf{x}(i\Delta + \Delta) = [I + F(i\Delta)\Delta]\mathbf{x}(i\Delta) + G(i\Delta)\Delta\mathbf{u}(i\Delta), \tag{16.1.9}$$

$$\mathbf{y}(i\Delta) = H(i\Delta)\mathbf{x}(i\Delta) + \mathbf{v}(i\Delta), \qquad i \geq 0, \tag{16.1.10}$$

where $\{\mathbf{u}(\cdot\Delta), \mathbf{v}(\cdot\Delta\}$ are zero-mean random processes such that

$$\left\langle \begin{bmatrix} \mathbf{u}(i\Delta) \\ \mathbf{v}(i\Delta) \\ \mathbf{x}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{u}(j\Delta) \\ \mathbf{v}(j\Delta) \\ \mathbf{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} \frac{Q(i\Delta)}{\Delta}\delta_{ij} & \frac{S(i\Delta)}{\Delta}\delta_{ij} & 0 \\ \frac{S^*(i\Delta)}{\Delta}\delta_{ij} & \frac{R(i\Delta)}{\Delta}\delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \end{bmatrix}. \tag{16.1.11}$$

16.1.3 An Application: State-Variance Recursions

Let $\Pi(i\Delta) = \|\mathbf{x}(i\Delta)\|^2$ denote the covariance matrix of the state of the discrete-time model (16.1.9)–(16.1.10). It then follows from the state-equation (16.1.9) that

$$\Pi(i\Delta + \Delta) = (I + F(i\Delta)\Delta)\Pi(i\Delta)(I + F(i\Delta)\Delta)^* + \Delta G(i\Delta)Q(i\Delta)G^*(i\Delta).$$

Expanding the right-hand side and rearranging terms we obtain

$$\frac{\Pi(i\Delta + \Delta) - \Pi(i\Delta)}{\Delta} = \Pi(i\Delta)F^*(i\Delta) + F(i\Delta)\Pi(i\Delta) + G(i\Delta)Q(i\Delta)G^*(i\Delta) +$$

$$\Delta F(i\Delta)\Pi(i\Delta)F^*(i\Delta).$$

Taking the limit as $\Delta \to 0$ we conclude that the state covariance matrix $\Pi(t) = \|\mathbf{x}(t)\|^2$ for the continuous-time model (16.1.1) satisfies the Lyapunov differential equation

$$\frac{d}{dt}\Pi(t) = \Pi(t)F^*(t) + F(t)\Pi(t) + G(t)Q(t)G^*(t), \qquad \Pi(0) = \Pi_0. \tag{16.1.12}$$

In Prob. 16.19, we show a way of obtaining this result by purely continuous-time arguments.

16.2 THE KALMAN FILTER EQUATIONS GIVEN STATE-SPACE AND COVARIANCE MODELS

We shall start with the continuous-time model (16.1.2)–(16.1.3), replace it with the approximate model (16.1.9)–(16.1.11), write down the corresponding Kalman filter recursions for the innovations of the process $\{\mathbf{y}(i\Delta)\}$, and then determine their continuous-time limit.

The innovations of the process $\{\mathbf{y}(i\Delta)\}$ described by the discrete-time model (16.1.9)–(16.1.10) are given by

$$\mathbf{e}(i\Delta) = \mathbf{y}(i\Delta) - H(i\Delta)\hat{\mathbf{x}}(i\Delta) = H(i\Delta)\tilde{\mathbf{x}}(i\Delta) + \mathbf{v}(i\Delta).$$

As we know, $\{\mathbf{e}(i\Delta)\}$ is a white noise process, so that writing $P(i\Delta) = \|\tilde{\mathbf{x}}(i\Delta)\|^2$, we shall get

$$\langle \mathbf{e}(i\Delta), \mathbf{e}(j\Delta)\rangle = \left[H(i\Delta)P(i\Delta)H^*(i\Delta) + \frac{R(i\Delta)}{\Delta}\right]\delta_{ij} \triangleq \frac{R_e(i\Delta)}{\Delta}\delta_{ij},$$

so that

$$R_e(i\Delta) = R(i\Delta) + \Delta H(i\Delta)P(i\Delta)H^*(i\Delta).$$

The reason for this definition of $R_e(i\Delta)$ is to keep the parallel with our definitions for the white-noise processes $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$ — see (16.1.7) and (16.1.8). The point is that in the limit as

$$\Delta \to 0, \quad i \to \infty, \qquad \text{keeping} \quad i\Delta = t,$$

we obtain

$$\mathbf{e}(i\Delta) \to \mathbf{e}(t), \qquad R_e(i\Delta) \triangleq R(i\Delta) + \Delta H(i\Delta)P(i\Delta)H^*(i\Delta) \to R(t),$$

and

$$\langle \mathbf{e}(i\Delta), \mathbf{e}(j\Delta)\rangle \to \langle \mathbf{e}(t), \mathbf{e}(s)\rangle = R(t)\delta(t-s).$$

So we have the very interesting result that in continuous-time the white-noise innovations process $\{\mathbf{e}(\cdot)\}$ has the same intensity as the measurement white-noise process $\{\mathbf{v}(\cdot)\}$ — a considerable simplification over the discrete-time formula; more on this later.

Now the Kalman filter recursions for the innovations of the model (16.1.9)–(16.1.10) are (see Thm. 9.2.1)

$$\mathbf{e}(i\Delta) = \mathbf{y}(i\Delta) - H(i\Delta)\hat{\mathbf{x}}(i\Delta),$$

$$\hat{\mathbf{x}}(i\Delta + \Delta) = [I + F(i\Delta)\Delta]\hat{\mathbf{x}}(i\Delta) + K_p(i\Delta)\mathbf{e}(i\Delta), \quad \hat{\mathbf{x}}(0) = 0,$$

where

$$K_p(i\Delta) = \{[I + F(i\Delta)\Delta]P(i\Delta)H^*(i\Delta) + G(i\Delta)S(i\Delta)\}\left[\frac{R_e(i\Delta)}{\Delta}\right]^{-1},$$

$$R_e(i\Delta) = R(i\Delta) + \Delta H(i\Delta)P(i\Delta)H^*(i\Delta).$$

To take limits, we first form

$$\frac{\hat{\mathbf{x}}(i\Delta + \Delta) - \hat{\mathbf{x}}(i\Delta)}{\Delta} = F(i\Delta)\hat{\mathbf{x}}(i\Delta) + \frac{K_p(i\Delta)}{\Delta}\mathbf{e}(i\Delta).$$

Now we see that as $\Delta \to 0$,

$$\lim_{\Delta\to 0}\frac{K_p(i\Delta)}{\Delta} = [P(t)H^*(t) + G(t)S(t)]R^{-1}(t) \triangleq K(t), \quad \text{say,}$$

and that the difference equation for $\hat{\mathbf{x}}(i\Delta)$ goes over to the differential equation

$$\dot{\hat{\mathbf{x}}}(t) = F(t)\hat{\mathbf{x}}(t) + K(t)\mathbf{e}(t), \quad \hat{\mathbf{x}}(0) = 0.$$

Finally, for the Riccati recursion

$$P(i\Delta + \Delta) = [I + F(i\Delta)\Delta]P(i\Delta)[I + F(i\Delta)\Delta]^* + \tag{16.2.1}$$
$$G(i\Delta)\Delta\frac{Q(i\Delta)}{\Delta}G^*(i\Delta)\Delta - K_p(i\Delta)\frac{R_e(i\Delta)}{\Delta}K_p^*(i\Delta),$$

we first rewrite it as

$$\frac{P(i\Delta + \Delta) - P(i\Delta)}{\Delta} = F(i\Delta)P(i\Delta) + P(i\Delta)F^*(i\Delta) + G(i\Delta)Q(i\Delta)G^*(i\Delta) -$$
$$- \frac{K_p(i\Delta)}{\Delta}R_e(i\Delta)\frac{K_p^*(i\Delta)}{\Delta} + \Delta F(i\Delta)P(i\Delta)F^*(i\Delta).$$

Then in the limit as $\Delta \to 0$, $i\Delta = t$, we get

$$\dot{P}(t) = F(t)P(t) + P(t)F^*(t) + G(t)Q(t)G^*(t) - K(t)R(t)K^*(t), \quad P(0) = \Pi_0,$$

which is a matrix Riccati differential equation. We have thus obtained the famous continuous-time Kalman (or Kalman-Bucy or Stratonovich-Kalman-Bucy — see the notes) filter equations.

Theorem 16.2.1 (Continuous-Time Filter) *Given a random process $\{\mathbf{y}(t), t \geq 0\}$ with the state-space model (16.1.1)–(16.1.2), its innovations can be determined as*

$$\mathbf{e}(t) = \mathbf{y}(t) - H(t)\hat{\mathbf{x}}(t), \quad \mathbf{e}(0) = \mathbf{y}(0), \tag{16.2.2}$$

$$\dot{\hat{\mathbf{x}}}(t) = F(t)\hat{\mathbf{x}}(t) + K(t)\mathbf{e}(t), \quad \hat{\mathbf{x}}(0) = 0, \quad t \geq 0, \tag{16.2.3}$$

where

$$K(t) = [P(t)H^*(t) + G(t)S(t)]R^{-1}(t), \tag{16.2.4}$$

and

$$\dot{P}(t) = F(t)P(t) + P(t)F^*(t) + G(t)Q(t)G^*(t) - K(t)R(t)K^*(t), \tag{16.2.5}$$

with $P(0) = \Pi_0$. ■

Proof: Given in the discussion preceding the theorem. An alternative direct derivation will be given in Sec. 16.4. ◆

Remark 1. Note that the formulas in Thm. 16.2.1 depend critically upon the fact that $R(\cdot)$, the intensity of the measurement noise process $\mathbf{v}(\cdot)$, is invertible, *i.e.*, is strictly positive-definite. This is in contrast to the discrete-time case, where it was only needed that $E\mathbf{v}_i\mathbf{v}_i^* = R_i \geq 0$. While one can of course study continuous-time problems where $R(\cdot)$ is not strictly positive-definite, the solution (see, *e.g.*, Geesey and Kailath (1973)) will generally involve derivatives of the observed process $\mathbf{y}(\cdot)$ and will therefore tend to be more sensitive to errors of various kinds; we shall not discuss them here (*cf.* Prob. 16.11). ◆

Remark 2. Notice that in our continuous-time model, $\hat{\mathbf{y}}(t)$ is really $\hat{\mathbf{y}}(t|t_-)$, *i.e.*, it is the l.l.m.s.e. of $\mathbf{y}(t)$ given $\{\mathbf{y}(\tau), \tau < t\}$; the reason this makes sense is the presence of the white-noise process $\mathbf{v}(\cdot)$ in our model. The use of white noise can be rigorized by working with an integral model (see, *e.g.*, Davis (1977), Kailath (1971)), but one can go a long way with the traditional direct use of white-noise models for linear estimation problems. ◆

Remark 3. The nonlinear matrix Riccati differential equation (16.2.5) can rarely be solved analytically when $n \neq 1$ (the state dimension); a few of the explicit solutions are described in Sec. 16.3 and in the problems. In general, it will have to be solved numerically, which is in principle straightforward because it is an initial value equation. So for example, using the simplest discretization, we would write for small δ,

$$P(t_0+\delta) = P(t_0) + \delta\,[F(t_0)P(t_0) + P(t_0)F^*(t_0) + G(t_0)Q(t_0)G^*(t_0) - K(t_0)R(t_0)K^*(t_0)] + o(\delta^2), \tag{16.2.6}$$

and successively find $\{P(t_0+\delta), P(t_0+2\delta), \ldots\}$. In practice more sophisticated numerical schemes would be used. ◆

Remark 4. The continuous-time formulas in Thm. 16.2.1 are simpler than those in discrete time (compare with Thm. 9.2.1). The main reason is that $R(\cdot)$ replaces $R_{e,i} = R_i + H_iP_iH_i^*$. Another is that, there is now no distinction between predicted and filtered state estimators, since

$$\hat{\mathbf{x}}(t) \triangleq \hat{\mathbf{x}}(t|t_-) = \text{the l.l.m.s.e. of } \mathbf{x}(t) \text{ given } \{\mathbf{y}(\tau),\ 0 \leq \tau < t\},$$
$$= \lim_{\Delta\to 0, i\Delta = t} \hat{\mathbf{x}}(i\Delta|i\Delta - \Delta) = \lim_{\Delta\to 0, i\Delta = t} \hat{\mathbf{x}}(i\Delta|i\Delta) = \hat{\mathbf{x}}(t|t).$$

The main consequence is that the continuous-time Riccati equation is simpler in form than the discrete-time recursion and in fact the same holds true for discrete approximations to the Riccati differential equation — compare, for example, the recursions (16.2.6) and (16.2.1). This raises the interesting, but not fully studied problem, as to whether one should study continuous-time problems by first discretizing the given state-space model and then using the discrete-time Kalman filter recursions, or whether one should just discretize the continuous-time solution. In general, we feel the latter choice would be preferable. ◆

Our recognition that $\mathbf{e}(\cdot)$ is a white-noise process allows the filter equations to be rearranged to give a causal and causally invertible state-space model for the process $\mathbf{y}(\cdot)$, known as an innovations representation of $\mathbf{y}(\cdot)$.

Theorem 16.2.2 (Innovations Representation of $\mathbf{y}(\cdot)$) *Given the causal, but not causally invertible, state-space model (16.1.1)–(16.1.2), a causal and causally invertible state-space model is given by the equations*

$$\begin{cases} \dot{\hat{\mathbf{x}}}(t) = F(t)\hat{\mathbf{x}}(t) + K(t)\mathbf{e}(t), & \hat{\mathbf{x}}(0) = 0, \\ \mathbf{y}(t) = H(t)\hat{\mathbf{x}}(t) + \mathbf{e}(t), \end{cases} \tag{16.2.7}$$

where $K(\cdot)$ can be found via (16.2.4)–(16.2.1) and $\mathbf{e}(\cdot)$ is a white noise process with $\langle \mathbf{e}(t), \mathbf{e}(s)\rangle = R(t)\delta(t-s)$. Moreover, the mapping from $\mathbf{y}(\cdot)$ to $\mathbf{e}(\cdot)$ (also known as the whitening filter) is seen to be given by

$$\begin{cases} \dot{\hat{\mathbf{x}}}(t) = [F(t) - K(t)H(t)]\hat{\mathbf{x}}(t) + K(t)\mathbf{y}(t), & \hat{\mathbf{x}}(0) = 0, \\ \mathbf{e}(t) = \mathbf{y}(t) - H(t)\hat{\mathbf{x}}(t). \end{cases} \tag{16.2.8}$$

■

Remark 5 [Covariance Factorization]. The above results give us a canonical (*i.e.*, causal and causally invertible) factorization of the covariance function $R_y(t,s)$ of the output process. Thus let $L(t,s)$ denote the impulse response of the system (16.2.7), *i.e.*,

$$L(t,s) = \begin{cases} I\delta(t-s) + H(t)\Phi(t,s)K(s), & t \geq s, \\ 0, & t < s, \end{cases} \tag{16.2.9}$$

where $\Phi(t,s)$ is the state transition matrix that is associated with $F(t)$, *viz.*, it is the unique solution of the linear differential equation (see App. C)

$$\frac{d\Phi(t,s)}{dt} = F(t)\Phi(t,s), \quad \Phi(s,s) = I. \tag{16.2.10}$$

Then it is not hard to see that

$$R_y(t,s) \triangleq \langle \mathbf{y}(t), \mathbf{y}(s)\rangle = \int_0^T L(t,\tau)R(\tau)L^*(s,\tau)d\tau. \tag{16.2.11}$$

◆

Remark 6. Since the covariance factorization (16.2.11) completely determines the impulse response $L(\cdot,\cdot)$, this means that $\{F(\cdot), H(\cdot), K(\cdot), R(\cdot)\}$ must be determined by the covariance function $R_y(t,s)$. In fact, this is true (of course, as already shown in discrete time in Sec. 9.6). We give the appropriate formulas below. We first note that we can express the covariance function as

$$R_y(t,s) = R(t)\delta(t-s) + H(t)\Phi(t,s)N(s)1(t-s) + N^*(t)\Phi^*(s,t)H^*(s)1(s-t), \tag{16.2.12}$$

where $1(t)$ is the Heaviside unit step function, *i.e.*, $1(t) = 1$ for $t \geq 0$ and zero elsewhere, while $N(t) = \Pi(t)H^*(t) + G(t)S(t)$, where $\Pi(t) = \|\mathbf{x}(t)\|^2$ obeys the Lyapunov differential equation (16.1.12). The main step in establishing (16.2.12) is to write (*cf.* App. C), for $t \geq s$,

$$\mathbf{x}(t) = \Phi(t,s)\mathbf{x}(s) + \int_s^t \Phi(t,\tau)G(\tau)\mathbf{u}(\tau)d\tau, \tag{16.2.13}$$

which yields $\langle \mathbf{x}(t), \mathbf{x}(s)\rangle = \Phi(t,s)\Pi(s)$ for $t \geq s$. The rest of the argument is straightforward (and is left to the active reader). ◆

We go on to show that the innovations model (16.2.7) and the whitening model (16.2.8) can be both completely specified having just the covariance parameters $\{R(\cdot), H(\cdot), F(\cdot), N(\cdot)\}$.

Theorem 16.2.3 (Covariance Specifications) *Given a process with state-space model (16.1.1)–(16.1.3), and corresponding covariance function as in (16.2.12), we can again write (16.2.7) with*

$$K(t) = [N(t) - \Sigma(t)H^*(t)]R^{-1}(t), \tag{16.2.14}$$

and $\Sigma(t)$ obeys the Riccati differential equation

$$\frac{d}{dt}\Sigma(t) = F(t)\Sigma(t) + \Sigma(t)F^*(t) + K(t)R(t)K^*(t), \quad \Sigma(0) = 0. \tag{16.2.15}$$

■

Proof: Define $\Sigma(t) = \|\hat{\mathbf{x}}(t)\|^2 = E\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^*(t)$. Then (16.2.15) follows from (16.2.7) by using the fact that $\mathbf{e}(\cdot)$ is white with $\langle \mathbf{e}(t), \mathbf{e}(s)\rangle = R(t)\delta(t-s)$. Now using the orthogonality property that $\tilde{\mathbf{x}}(t) \perp \hat{\mathbf{x}}(t)$, we can write

$$\Pi(t) \triangleq \|\mathbf{x}(t)\|^2 = \|\hat{\mathbf{x}}(t)\|^2 + \|\tilde{\mathbf{x}}(t)\|^2 \triangleq \Sigma(t) + P(t).$$

Therefore,

$$\begin{aligned} K(t) &= [P(t)H^*(t) + G(t)S(t)]R^{-1}(t), \\ &= [\Pi(t)H^*(t) + G(t)S(t) - \Sigma(t)H^*(t)]\,R^{-1}(t), \\ &= [N(t) - \Sigma(t)H^*(t)]R^{-1}(t). \end{aligned}$$

◆

Remark 7 [Semi-separable Kernels and Explicit Solutions]. Using the property that $\Phi(t,s) = \Phi(t,0)\Phi(0,s)$ (*cf.* App. C), we can express the covariance function (16.2.12) in the so-called semi-separable form noted in Sec. 7.9:

$$R_y(t,s) = R(t)\delta(t-s) + A(t)B(s)1(t-s) + B^*(t)A^*(s)1(s-t),$$

where

$$A(t) = H(t)\Phi(t,0), \qquad B(s) = \Phi(0,s)N(s).$$

This result nicely identifies the semi-separable kernels introduced by Shinbrot (1957), and others, as a generalization of the expression for the covariance functions of stationary processes with rational power spectral density functions. We remarked in Sec. 7.9 that a major difficulty in solving finite-time Wiener-Hopf equations of the form

$$R_{xy}(t,s) = \int_0^t h_{xy}(t,\tau)R_y(\tau,s)d\tau, 0 \le s \le t$$

was the fact that explicit solutions were very complicated. This is easy to appreciate now, since even with the compactness afforded by the use of state-space notation, the expression for $h_{xy}(\cdot,\cdot)$ is quite complicated: it is the impulse response of the system described by

$$\dot{\hat{\mathbf{x}}}(t) = [F(t) - K(t)H(t)]\hat{\mathbf{x}}(t) + K(t)\mathbf{y}(t), \quad \hat{\mathbf{x}}(0) = 0.$$

Therefore

$$h_{xy}(t,s) = \Psi(t,s)K(s)1(t-s),$$

where $\Psi(\cdot,\cdot)$ is the solution of

$$\frac{d\Psi(t,s)}{dt} = [F(t) - K(t)H(t)]\Psi(t,s), \quad \Psi(s,s) = I.$$

Given the fact that $K(\cdot)$ is specified via a Riccati differential equation (for $P(\cdot)$ or for $\Sigma(\cdot)$), not easy to solve analytically except in very special cases (see Sec. 16.3), and also that $\Phi(\cdot,\cdot)$ and $\Psi(\cdot,\cdot)$ are not readily expressible in closed form, we can imagine the difficulty that the investigators in the 1950s faced. The logjam was broken by Kalman's introduction of state-space process models and the resulting natural step to seek similar models for the estimators.

We shall return to some further historical remarks in the notes at the end of this chapter. ◆

16.3 SOME EXAMPLES

Explicit solution of the nonlinear Riccati differential equation (16.2.5) is only possible in a few cases. We give the best known examples below. However, even when explicit solution is not possible, useful information can still be gained by certain standard methods, as we shall also illustrate.

Example 16.3.1 (Scalar Parameter Estimation) Consider the system $\dot{\mathbf{x}}(t) = 0$, $\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t)$, so that $\mathbf{x}(t) = \mathbf{x}(0)$ and $\mathbf{y}(t) = \mathbf{x}(0) + \mathbf{v}(t)$, where $\mathbf{x}(0)$ is assumed to be random with mean zero and variance Π_0 and $\mathbf{v}(\cdot)$ is white noise with unit intensity. The corresponding Riccati differential equation (16.2.5) is

$$\dot{P}(t) = -P^2(t), \quad P(0) = \Pi_0,$$

which is readily solved:

$$-\int_0^t \frac{dP(\tau)}{P^2(\tau)} = \int_0^t d\tau,$$

so that $P(t) = \Pi_0(\Pi_0 t + 1)^{-1}$. Therefore, as expected, the random constant $\mathbf{x}(0)$ can be estimated with increasing accuracy as t increases, since $P(t) \to 0$ as $t \to \infty$. ◆

Example 16.3.2 (Exponentially Correlated Processes) Consider a system

$$\dot{\mathbf{x}}(t) = -\alpha\mathbf{x}(t) + \mathbf{u}(t), \quad \alpha > 0,$$

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t), \quad t \ge 0,$$

with $Q(t) = 2\alpha$, $S(t) = 0$, $R(t) = 1$, $\Pi_0 = 1$. The Riccati equation (16.2.5)

$$\dot{P}(t) = -2\alpha P(t) + 2\alpha - P^2(t), \quad P(0) = 1,$$

is of the classical form and it can be checked that the solution is

$$P(t) = \frac{2\alpha}{\gamma + \alpha} \cdot \frac{1 + \left(\frac{\gamma-\alpha}{\gamma+\alpha}\right)e^{-2\gamma t}}{1 - \left(\frac{\gamma-\alpha}{\gamma+\alpha}\right)^2 e^{-2\gamma t}}, \qquad \text{where} \quad \gamma = \sqrt{\alpha(\alpha+2)}.$$

As $t \to \infty$, the process $\mathbf{x}(\cdot)$ tends to a stationary process with power spectral density function

$$S_x(f) = \frac{2\alpha}{\alpha^2 + 4\pi^2 f^2} = \mathcal{F}\{e^{-\alpha|t|}\}.$$

The Riccati variable $P(\cdot)$ also tends to a constant

$$P(t) \to \frac{2\alpha}{\gamma + \alpha} = \frac{2\alpha(\gamma - \alpha)}{\gamma^2 - \alpha^2} = \gamma - \alpha,$$

so that the steady-state filter is

$$\dot{\hat{\mathbf{x}}}(t) = -\alpha\hat{\mathbf{x}}(t) + (\gamma - \alpha)(\mathbf{y}(t) - \hat{\mathbf{x}}(t)) = -\gamma\hat{\mathbf{x}}(t) + (\gamma - \alpha)\mathbf{y}(t),$$

with transfer function from $\mathbf{y}(\cdot)$ to $\hat{\mathbf{x}}(\cdot)$,

$$H(s) = \frac{\gamma - \alpha}{s + \gamma},$$

which turns out to be exactly the Wiener filter for this problem — compare with the result of the example given at the end of App. 7.A. ◆

Example 16.3.3 (No Plant Noise) When $G(t) = 0$, $S(t) = 0$, for $t \geq t_0$ in the state-space model (16.1.1)–(16.1.2), the Riccati equation (16.2.5) becomes homogeneous,

$$\dot{P}(t) = F(t)P(t) + P(t)F^*(t) - P(t)H^*(t)R^{-1}(t)H(t)P(t), \quad P(t_0) = \Pi_0, \tag{16.3.1}$$

In this case, it turns out that the inverse of $P(t)$ satisfies the *linear* differential equation

$$\frac{dP^{-1}(t)}{dt} = -P^{-1}F(t) - F^*(t)P^{-1}(t) + H^*(t)R^{-1}(t)H(t), \quad P^{-1}(t_0) = \Pi_0^{-1},$$

whose solution, when $\Pi_0 > 0$, is readily seen to be (*cf.* Prob. 16.17)

$$P(t) = \Phi(t, t_0)[\Pi_0^{-1} + \mathcal{O}(t, t_0)]^{-1}\Phi^*(t, t_0). \tag{16.3.2}$$

Here $\Phi(t, t_0)$ is the state-transition matrix that was defined earlier in (16.2.10), and $\mathcal{O}(t, t_0)$ is the so-called observability Gramian of the pair $\{F(t), R^{-1/2}(t)H(t)\}$,

$$\mathcal{O}(t, t_0) = \int_{t_0}^{t} \Phi^*(\tau, t_o)H^*(\tau)R^{-1}(\tau)H(\tau)\Phi(\tau, t_0)d\tau,$$

which can also be characterized as the unique solution of the linear matrix differential equation

$$\frac{d\mathcal{O}(t, t_0)}{dt} = F^*(t)\mathcal{O}(t, t_0) + \mathcal{O}(t, t_0)F(t) + H^*(t)R^{-1}(t)H(t), \quad \mathcal{O}(t_0, t_0) = 0.$$

◆

Remark 8. The formula (16.3.2) is not really useful for computational purposes because the functions $\Phi(t, t_0)$ and $\mathcal{O}(t, t_0)$ cannot in general (when $n > 1$) be found in closed form; in fact the Riccati differential equation (16.3.1) is more useful for this purpose since it can be more easily solved numerically, *e.g.*, by discretization. The value of "closed-form" expressions such as (16.3.2) is that they can be used to draw other useful conclusions, as we illustrate in the next example. ◆

Example 16.3.4 (Comparing Riccati Solutions) Consider the Riccati equation (16.2.5), with $S(t) = 0$,

$$\dot{P}(t) = F(t)P(t) + P(t)F^*(t) + G(t)Q(t)G^*(t) - P(t)H^*(t)R^{-1}(t)H(t)P(t), \quad t \geq t_0.$$

Denote by $P_i(\cdot)$ the solutions for the initial conditions $P_i(t_0) = \Pi_{i0}$, $i = 1, 2$. It turns out that

$$\Pi_{10} > \Pi_{20} \quad \text{implies that} \quad P_1(t) > P_2(t), \quad t \geq t_0,$$

where $A > B$ means that the matrix $A - B$ is positive-definite. To see this, note that $\Delta(t) = P_1(t) - P_2(t)$ satisfies the differential equation

$$\dot{\Delta}(t) = F_{cl}(t)\Delta(t) + \Delta(t)F_{cl}^*(t) - \Delta(t)H^*(t)R^{-1}(t)H(t)\Delta(t), \quad \Delta(t_0) = \Pi_{10} - \Pi_{20},$$

where $F_{cl}(t) = F(t) - P_2(t)H^*(t)R^{-1}(t)H(t)$ But this is a homogeneous Riccati differential equation and by appealing to the result of Ex. 16.3.3 and Prob. 16.17 (suitably modified), we can write

$$\Delta(t) = \Psi(t, t_0)[\Delta^{-1}(t_0) + \mathcal{O}(t, t_0)]^{-1}\Psi^*(t, t_0),$$

where $\Psi(t, t_0)$ is the state transition matrix of $F_{cl}(t)$, and

$$\mathcal{O}(t, t_0) = \int_{t_0}^{t} \Psi^*(\tau, t_o)H^*(\tau)R^{-1}(\tau)H(\tau)\Psi(\tau, t_0)d\tau.$$

Now for every t, $\mathcal{O}(t, t_0) \geq 0$ but $\Delta^{-1}(t_0) > 0$, so that $[\Delta^{-1}(t_0) + \mathcal{O}(t, t_0)]$ and its inverse are both positive-definite. Finally, since it is known (see App. C) that $\det \Psi(t, t_0) \neq 0$, it follows that $\Delta(t) > 0$.

By similar arguments one can show that $P_1(t) > P_2(t)$ if $Q_1(t) > Q_2(t)$, or $R_1(t) > R_2(t)$, or both (see Prob. 16.13). The conclusion is that we shall be conservative, *i.e.*, the actual error variance will be smaller than the value $P(\cdot)$ that we compute if we assume values of $\{\Pi(0), Q(\cdot), R(\cdot)\}$ that are greater than the true values.

There are many such interesting Riccati equation identities, many of which involve the three quantities $\{P(\cdot), \Psi(\cdot, t_0), \mathcal{O}(\cdot, t_0)\}$. A very powerful and organized approach to deriving and understanding them is provided by the scattering theory/transmission line analysis in Ch. 17. ◆

16.4 DIRECT SOLUTION USING THE INNOVATIONS PROCESS

In Sec. 16.1.2, we deduced the Kalman filter equations by first discretizing the continuous-time model and then using the results of Thm. 9.2.1. The same approach can be used to obtain continuous-time analogs of the results of Ch. 10 on smoothing and of Ch. 11 on fast algorithms for time-invariant systems. However, this can become somewhat tiresome, especially as we have already seen, the continuous-time formulas can be simpler in form.

In fact, direct continuous-time solutions are quite straightforward when we use the innovations approach. A clue to the simplicity of the continuous-time results can be found in the fact we noted earlier that the continuous-time innovations $\mathbf{e}(\cdot)$ satisfy

$$\langle \mathbf{e}(t), \mathbf{e}(s) \rangle = R(t)\delta(t - s).$$

This important property can also be obtained directly in continuous time, and in fact in a more general form (*i.e.*, not assuming state-space structure).

16.4.1 The Innovations Process

Thus consider a continuous-time process of the form

$$\mathbf{y}(t) = \mathbf{z}(t) + \mathbf{v}(t), \qquad 0 \le t \le T,$$

where $\mathbf{v}(\cdot)$ is white noise with

$$E\mathbf{v}(t) = 0, \quad \langle \mathbf{v}(t), \mathbf{v}(s) \rangle = R(t)\delta(t-s),$$

and $\mathbf{z}(\cdot)$ is a zero-mean finite-variance random process obeying

$$\int_0^T E\|\mathbf{z}(t)\|^2 < \infty, \tag{16.4.1}$$

and

$$\langle \mathbf{v}(t), \mathbf{z}(s) \rangle \triangleq E\mathbf{v}(t)\mathbf{z}^*(s) = 0, \quad \text{for} \quad t \ge s. \tag{16.4.2}$$

That is, the future white noise is always uncorrelated with the past and present of the signal process, a condition that allows the signal to depend on past $\mathbf{y}(\cdot)$, as in feedback communication and control systems. We should note that the process $\mathbf{z}(t) = H(t)\mathbf{x}(t)$ in the state-space model (16.1.1)–(16.1.3) satisfies (16.4.1).

Now define the innovation at time t as

$$\mathbf{e}(t) \triangleq \mathbf{y}(t) - \hat{\mathbf{y}}(t|t_-) = \mathbf{y}(t) - \hat{\mathbf{z}}(t), \tag{16.4.3}$$

where

$$\hat{\mathbf{z}}(t) \triangleq \hat{\mathbf{z}}(t|t_-) = \text{the l.l.m.s. estimator of } \mathbf{z}(t) \text{ given } \{\mathbf{y}(\tau),\ 0 \le \tau < t\}.$$

The reason for the name "innovation" is that the quantity $\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{z}}(t) = \mathbf{y}(t) - \hat{\mathbf{y}}(t|t-)$ can be regarded as the "new information" brought by the current observation $\mathbf{y}(t)$, being given all the past observations $\mathbf{y}(\cdot)$, and the old information deduced therefrom.

Lemma 16.4.1 (Whiteness of the Innovations) *The innovations process* $\mathbf{e}(\cdot)$ *is white with*

$$E\mathbf{e}(t) = 0, \quad \langle \mathbf{e}(t), \mathbf{e}(s) \rangle = R(t)\delta(t-s). \tag{16.4.4}$$

■

Proof: This can be proved in different ways. First note that $\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{z}}(t) = \tilde{\mathbf{z}}(t) + \mathbf{v}(t)$, where $\tilde{\mathbf{z}}(t)$, the error in the estimator of $\mathbf{z}(t)$ given past $\mathbf{y}(\cdot)$, must satisfy the orthogonality condition

$$\langle \tilde{\mathbf{z}}(t), \mathbf{y}(s) \rangle = 0, \quad \langle \tilde{\mathbf{z}}(t), \mathbf{e}(s) \rangle = 0, \quad 0 \le s < t \le T. \tag{16.4.5}$$

Now we can write

$$\langle \mathbf{e}(t), \mathbf{e}(s) \rangle = \langle \mathbf{v}(t), \mathbf{v}(s) \rangle + \langle \mathbf{v}(t), \tilde{\mathbf{z}}(s) \rangle + \langle \tilde{\mathbf{z}}(t), \mathbf{e}(s) \rangle = R(t)\delta(t-s) + A(t,s), \text{ say.}$$

We shall first show that $A(t,s) = 0$ when $t \ne s$. Indeed, note that by using (16.4.1)–(16.4.2) and (16.4.5), we have

$$A(t,s) = \langle \mathbf{v}(t), \tilde{\mathbf{z}}(s) \rangle + \langle \tilde{\mathbf{z}}(t), \mathbf{e}(s) \rangle = 0 + 0, \quad \text{for } t > s.$$

Since we can rewrite $A(t,s)$ as

$$A(t,s) = \langle \tilde{\mathbf{z}}(t), \mathbf{v}(s) \rangle + \langle \mathbf{e}(t), \tilde{\mathbf{z}}(s) \rangle,$$

we see by a similar argument that $A(t,s) = 0$ for $t < s$. When $t = s$, we have

$$\begin{aligned} A(t,t) &= \langle \mathbf{v}(t), \tilde{\mathbf{z}}(t) \rangle + \langle \tilde{\mathbf{z}}(t), \mathbf{e}(t) \rangle = 0 + \langle \tilde{\mathbf{z}}(t), \tilde{\mathbf{z}}(t) \rangle + \langle \tilde{\mathbf{z}}(t), \mathbf{v}(t) \rangle, \\ &= 0 + \langle \tilde{\mathbf{z}}(t), \tilde{\mathbf{z}}(t) \rangle + 0 = \langle \tilde{\mathbf{z}}(t), \tilde{\mathbf{z}}(t) \rangle, \end{aligned}$$

since by assumption $\mathbf{v}(t) \perp \mathbf{z}(s)$, $s \le t$. Now

$$\langle \tilde{\mathbf{z}}(t), \tilde{\mathbf{z}}(t) \rangle = E\int_0^T |\tilde{\mathbf{z}}(t)|^2 dt \le \langle \mathbf{z}(t), \mathbf{z}(t) \rangle < \infty.$$

So $A(\cdot,\cdot)$ is a function that is zero everywhere in the square $[0,T] \times [0,T]$, except possibly along the diagonal where it is always finite. Since lines have zero "measure" in a plane, this suggests that $A(\cdot,\cdot)$ is "equivalent" to the null function in two variables. Or, more precisely, for all continuous-time "test" functions $\phi(\cdot)$, we shall have

$$\int_0^T [R(t)\delta(t-s) + A(t,s)]\phi(s)ds = R(t)\phi(t) + 0 = \int_0^T R(t)\delta(t-s)\phi(s)ds,$$

which shows that $R(t)\delta(t-s) + A(t,s)$ is equivalent to the delta $R(t)\delta(t-s)$. ◆

Remark 9. The above proof was the one given in Kailath (1968). We may remark that more mathematically conventional proofs can be given, and in fact under more general conditions, by working with integrated versions (*i.e.,* white noise replaced by a Wiener process) and using martingale theory — see Kailath (1971) and Meyer (1973). ◆

Remark 10. As a matter of fact, we may note that a proper understanding of the innovations process requires more background than is usual in first engineering courses on random processes. For example, suppose $\{\mathbf{x}_0, \mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$ are jointly Gaussian processes; then so will be the process $\{\mathbf{e}(\cdot)\}$. But then $\mathbf{e}(\cdot)$ and $\mathbf{v}(\cdot)$, which both have the same mean and covariance functions, would be the "same" stochastic process using the usual definition via joint distribution functions — this is clearly not the case. To distinguish between them one needs to introduce the notion of an increasing family of sigma fields (see, *e.g.,* Davis (1977)). As stated before, we shall not pursue these niceties here; they are really only significant for nonlinear problems. However, we might mention that when $\mathbf{y}(\cdot) = \mathbf{z}(\cdot) + \mathbf{v}(\cdot)$, $\mathbf{z}(\cdot)$ non-Gaussian but $\mathbf{v}(\cdot)$ still Gaussian noise, then $\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{z}}(t)$, where $\hat{\mathbf{z}}(t)$ is the (not necessarily linear) optimal least-mean-squares estimator of $\mathbf{z}(t)$ given $\{\mathbf{y}(\tau), \tau < t\}$ is not only white but also Gaussian (see Kailath (1969a,1970) and Frost and Kailath (1971) and the references therein). This striking result depends heavily upon the use of martingale theory. ◆

Another important property of the innovations process (in the linear case) is that it is causally equivalent to the original process. The proof of this result needs some background on integral equations and may be skipped on a first reading; it too was first given in Kailath (1968). For a related proof, see Benes (1976), where a discussion is given of some of the difficulties in extending this equivalence to nonlinear problems. The best result to date on equivalence was given by Allinger and Mitter (1981); this was for the case of completely independent signal and noise processes — the general case is still open.

Lemma 16.4.2 (Causal and Causally Invertible Transformation) *The processes* $\{\mathbf{y}(\cdot), \mathbf{e}(\cdot)\}$ *can be obtained from each other by causal linear operations. Therefore, the processes are "equivalent" (i.e., they contain the same statistical information) as far as linear operations are concerned. [See footnote 1 of Ch. 7 for more discussion of allowable linear operations on* $\mathbf{y}(\cdot)$ *and* $\mathbf{e}(\cdot)$*.]* ∎

Proof: Let $h(t,s)$ denote the optimal causal filter that operates on the observations $\{\mathbf{y}(\tau), 0 \le \tau < t\}$ to give the estimator $\hat{\mathbf{z}}(t)$, *i.e.*,

$$\hat{\mathbf{z}}(t) = \int_0^t h(t,s)\mathbf{y}(s)ds. \tag{16.4.6}$$

To make (16.4.6) well defined, we need to assume that (see, *e.g.*, Doob (1953))

$$\int_0^t |h(t,s)|^2 ds < \infty \quad \text{for every} \quad t.$$

[If $h(t,\cdot)$ has delta functions in it, then $\hat{\mathbf{z}}(t)$ would have infinite variance.] From our assumption that

$$\int_0^T \langle \mathbf{z}(t), \mathbf{z}(t)\rangle dt = \int_0^T E\|\mathbf{z}(t)\|^2 dt < \infty,$$

it can be shown that $h(\cdot,\cdot)$ is square-integrable, *viz.*,

$$\int_0^t \int_0^t |h(t,s)|^2 \, dtds < \infty. \tag{16.4.7}$$

Now let $\mathcal{H}$ denote the integral operator with kernel $h(t,s)$. Using $\mathcal{H}$ we can write more compactly $\hat{\mathbf{z}}(t) = \mathcal{H}\mathbf{y}$, where $\mathbf{y}$ denotes the observations $\{\mathbf{y}(\tau), 0 \le \tau < t\}$. The operator $\mathcal{H}$ is clearly linear and causal.

Let also $\mathcal{I}$ denote the identity operator, which corresponds to the integral operator with kernel $\delta(t-s)$. Then we obtain, symbolically,

$$\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{z}}(t) = (\mathcal{I} - \mathcal{H})\mathbf{y}. \tag{16.4.8}$$

This shows that the transformation from the observations to the innovations is causal.

To prove the converse, we need to show that $(\mathcal{I} - \mathcal{H})$ is invertible and that its inverse is a causal operator. For this purpose, we remark that for square-integrable kernels $h(\cdot,\cdot)$ (*i.e.*, satisfying (16.4.7)), it can be shown that $(\mathcal{I} - \mathcal{H})^{-1}$ exists and is given by the so-called Neumann series (see, *e.g.*, Smithies (1965) and Riesz and Sz.-Nagy (1990, Ch. 4))

$$(\mathcal{I} - \mathcal{H})^{-1} = \mathcal{I} + \mathcal{H} + \mathcal{H}^2 + \mathcal{H}^3 + \ldots, \tag{16.4.9}$$

where the notation $\mathcal{H}^2$ stands for $\mathcal{H}^2\mathbf{y} = \mathcal{H}\mathcal{H}\mathbf{y}$, and so on. It is clear from the causality of $\mathcal{H}$ and from the right-hand side of (16.4.9) that $(\mathcal{I} - \mathcal{H})^{-1}$ is also causal. ◆

Using the innovations process, we can directly address the continuous-time state-space estimation problem.

16.4.2 The Innovations Approach

We start with the standard state-space model (16.1.1)–(16.1.3). Our argument will closely parallel those used in the discrete-time case in Secs. 9.2–9.2.4, which the reader may wish to review at this point. For simplicity, we shall assume that

$$S(t) = 0. \tag{16.4.10}$$

[See Prob. 16.20 for the general case.]

Step 1. [The Innovations] Introduce the innovations

$$\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{y}}(t) = \mathbf{y}(t) - H(t)\hat{\mathbf{x}}(t) = H(t)\tilde{\mathbf{x}}(t) + \mathbf{v}(t), \quad t \ge 0,$$

where

$$\hat{\mathbf{x}}(t) \triangleq \hat{\mathbf{x}}(t|t_-) = \text{the l.l.m.s. estimator of } \mathbf{x}(t) \text{ given } \{\mathbf{y}(\tau),\ 0 \le \tau < t\}.$$

Step 2. [Estimators from the Innovations] Let

$$\hat{\mathbf{x}}(t) = \int_0^t k(t,s)\mathbf{e}(s)ds.$$

Then the orthogonality condition $\tilde{\mathbf{x}}(t) \perp \mathbf{e}(\tau)$ for $0 \le \tau < t$, gives

$$\langle \mathbf{x}(t), \mathbf{e}(\tau)\rangle = \langle \int_0^t k(t,s)\mathbf{e}(s)ds, \mathbf{e}(\tau)\rangle = \int_0^t k(t,s)\langle \mathbf{e}(s), \mathbf{e}(\tau)\rangle \, ds,$$

$$= \int_0^t k(t,s)R(s)\delta(s-\tau)ds = k(t,\tau)R(\tau), \quad 0 \le \tau < t.$$

Therefore, we can write

$$\hat{\mathbf{x}}(t) = \int_0^t \langle \mathbf{x}(t), \mathbf{e}(s)\rangle \, R^{-1}(s)\mathbf{e}(s)ds. \tag{16.4.11}$$

Step 3. **[A Differential Equation]** The formula (16.4.11) appears to be circular, because $\hat{\mathbf{x}}(\cdot)$ appears on both sides of the equality. However, note that what we actually have is an expression for $\hat{\mathbf{x}}(t)$ in terms of earlier values, $\hat{\mathbf{x}}(s)$, $s < t$. This suggests a recursive relationship, which we can seek by differentiating both sides of (16.4.11) to get

$$\dot{\hat{\mathbf{x}}}(t) = \langle \mathbf{x}(t), \mathbf{e}(t) \rangle \; R^{-1}(t)\mathbf{e}(t) + \int_0^t \langle \dot{\mathbf{x}}(t), \mathbf{e}(s) \rangle \; R^{-1}(s)\mathbf{e}(s)ds.$$

But note that the state-space equation (16.1.1) yields

$$\langle \dot{\mathbf{x}}(t), \mathbf{e}(s) \rangle = F(t) \; \langle \mathbf{x}(t), \mathbf{e}(s) \rangle + G(t) \; \langle \mathbf{u}(t), \mathbf{e}(s) \rangle = F(t) \; \langle \mathbf{x}(t), \mathbf{e}(s) \rangle + 0.$$

Then defining

$$K(t) \triangleq \langle \mathbf{x}(t), \mathbf{e}(t) \rangle \; R^{-1}(t), \tag{16.4.12}$$

we have

$$\dot{\hat{\mathbf{x}}}(t) = K(t)\mathbf{e}(t) + F(t) \int_0^t \langle \mathbf{x}(t), \mathbf{e}(s) \rangle \; R^{-1}(s)\mathbf{e}(s)ds \tag{16.4.13}$$

$$+ G(t) \int_0^t \langle \mathbf{u}(t), \mathbf{e}(s) \rangle \; R^{-1}(s)\mathbf{e}(s)ds,$$

$$= K(t)\mathbf{e}(t) + F(t)\hat{\mathbf{x}}(t) + 0. \tag{16.4.14}$$

From (16.4.11) we see that the initial condition is $\hat{\mathbf{x}}(0) = 0$.

This is the continuous-time Kalman filter equation (16.2.3), which we obtained in Sec. 16.1.2 by a limiting argument.

Step 4. **[A Formula for $K(\cdot)$]** The deterministic function $K(\cdot)$ can be computed in different ways. To obtain the usual Kalman filter formula (16.2.4), we introduce

$$P(t) \triangleq \|\tilde{\mathbf{x}}(t)\|^2, \quad \text{the error covariance matrix,}$$

and note that

$$K(t)R(t) = \langle \mathbf{x}(t), \mathbf{e}(t) \rangle = \langle \mathbf{x}(t), H(t)\tilde{\mathbf{x}}(t) + \mathbf{v}(t) \rangle,$$

$$= \langle \mathbf{x}(t), \tilde{\mathbf{x}}(t) \rangle \; H^*(t) + \langle \mathbf{x}(t), \mathbf{v}(t) \rangle.$$

But our assumption that $S(\cdot) = 0$ means that

$$\mathbf{x}(t) \in \mathcal{L}\{\mathbf{x}(0), \mathbf{u}(\tau), \; 0 \le \tau < t\} \perp \mathbf{v}(t).$$

Also,

$$\langle \mathbf{x}(t), \tilde{\mathbf{x}}(t) \rangle = \langle \hat{\mathbf{x}}(t) + \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(t) \rangle = 0 + P(t).$$

Therefore, we have

$$K(t) = P(t)H^*(t)R^{-1}(t),$$

which agrees with expression (16.2.4) when $S(t) = 0$.

Step 5. **[Error Variance Calculations]** It remains to find a method of computing the error-variance matrix $P(\cdot)$. Again, we use an argument similar to one used in the discrete-time case (Sec. 9.2.3). For this we recall the result established in Sec. 16.1.3 (and in Prob. 16.19) that the variance, $\Pi(\cdot) = \|\mathbf{x}(\cdot)\|^2$, of the state $\mathbf{x}(\cdot)$ in the model (16.1.1)–(16.1.3) obeys the linear differential equation

$$\dot{\Pi}(t) = F(t)\Pi(t) + \Pi(t)F^*(t) + G(t)Q(t)G^*(t), \quad \Pi(0) = \Pi_0. \tag{16.4.15}$$

Now, because of the whiteness of the innovations, the Kalman filter estimator equation (16.4.14) is of the same form as (16.1.1)–(16.1.3), *i.e.*,

$$\dot{\hat{\mathbf{x}}}(t) = F(t)\hat{\mathbf{x}}(t) + K(t)\mathbf{e}(t), \quad \hat{\mathbf{x}}(0) = 0,$$

with

$$\langle \mathbf{e}(t), \hat{\mathbf{x}}(0) \rangle = 0, \; \langle \hat{\mathbf{x}}(0), \hat{\mathbf{x}}(0) \rangle = 0, \; \langle \mathbf{e}(t), \mathbf{e}(s) \rangle = R(t)\delta(t - s).$$

Therefore, $\Sigma(t) \triangleq \|\hat{\mathbf{x}}(t)\|^2$ obeys the equation

$$\dot{\Sigma}(t) = F(t)\Sigma(t) + \Sigma(t)F^*(t) + K(t)R(t)K^*(t), \quad \Sigma(0) = 0.$$

Now the orthogonal decomposition $\mathbf{x}(t) = \hat{\mathbf{x}}(t) + \tilde{\mathbf{x}}(t)$, implies that $\Pi(t) = \Sigma(t) + P(t)$, so that

$$\dot{P}(t) = \dot{\Pi}(t) - \dot{\Sigma}(t),$$

$$= F(t)P(t) + P(t)F^*(t) + G(t)Q(t)G^*(t) - K(t)R(t)K^*(t) \tag{16.4.16}$$

with initial condition

$$P(0) = \Pi(0) - \Sigma(0) = \Pi_0.$$

This is the Riccati differential equation, which we obtained in Sec. 16.2 by a limiting argument applied to the discrete-time formulas. The point is that with the innovations approach we can work directly (and more simply) in continuous time. We further illustrate this point in the following sections.

16.5 SMOOTHED ESTIMATORS

The innovations arguments used here follow those of Ch. 10, except that the calculations tend to be simpler.

Let $\hat{\mathbf{x}}(t|T)$ denote the l.l.m.s. estimator of $\mathbf{x}(t)$ given the entire observations $\mathbf{y}(\tau)$ over an interval $[0, T]$, with $t \in [0, T]$. To compute $\hat{\mathbf{x}}(t|T)$, we start by expressing it in terms of the innovations,

$$\hat{\mathbf{x}}(t|T) = \int_0^T k_T(t, s)\mathbf{e}(s)ds, \quad \text{say.}$$

Then the orthogonality condition

$$\mathbf{x}(t) - \hat{\mathbf{x}}(t|T) \perp \mathbf{e}(\tau), \quad 0 \le \tau \le T,$$

and the whiteness of the innovations gives

$$\langle \mathbf{x}(t), \mathbf{e}(\tau)\rangle \;=\; \int_0^T k_T(t,s)\;\langle \mathbf{e}(s), \mathbf{e}(\tau)\rangle\; ds \;=\; k_T(t,\tau)R(\tau), \quad 0 \le \tau \le T.$$

Therefore,

$$\hat{\mathbf{x}}(t|T) \;=\; \int_0^T \langle \mathbf{x}(t), \mathbf{e}(s)\rangle\; R^{-1}(s)\; \mathbf{e}(s)ds, \quad 0 \le t \le T.$$

Comparing this expression with the formula (16.4.11) for $\hat{\mathbf{x}}(t)$ suggests the decomposition

$$\hat{\mathbf{x}}(t|T) \;=\; \hat{\mathbf{x}}(t) + \int_t^T \langle \mathbf{x}(t), \mathbf{e}(s)\rangle\; R^{-1}(s)\; \mathbf{e}(s)ds.$$

Moreover, we can say a little more about the term

$$\langle \mathbf{x}(t), \mathbf{e}(s)\rangle \;=\; \langle \mathbf{x}(t), \tilde{\mathbf{x}}(s)\rangle\; H^*(s) \;+\; \langle \mathbf{x}(t), \mathbf{v}(s)\rangle,$$

when $s > t$. Then $\mathbf{v}(s) \perp \mathbf{x}(t)$, and

$$\langle \mathbf{x}(t), \tilde{\mathbf{x}}(s)\rangle \;=\; \langle \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(s)\rangle \;+\; \langle \hat{\mathbf{x}}(t), \tilde{\mathbf{x}}(s)\rangle \;=\; \langle \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(s)\rangle \;+\; 0,$$

since for $s > t$, the error in the estimator for $\mathbf{x}(s)$ is orthogonal to all the (linear functions of the) data up to time s and in particular to $\hat{\mathbf{x}}(t)$. We shall define $P(t,s) \triangleq \langle \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(s)\rangle$, with of course $P(t) \triangleq \langle \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(t)\rangle = P(t,t)$. Now the following results are almost immediate.

Lemma 16.5.1 (General Smoothing Formulas) *Given $\mathbf{y}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t)$, $0 \le t \le T$, with $\langle \mathbf{v}(t), \mathbf{v}(s)\rangle = R(t)\delta(t-s)$ and $\langle \mathbf{v}(t), \mathbf{x}(s)\rangle = 0$ for $s \le t$, we can write*

$$\hat{\mathbf{x}}(t|T) \;=\; \hat{\mathbf{x}}(t) + \int_t^T P(t,s)H^*(s)R^{-1}(s)\mathbf{e}(s)ds, \tag{16.5.1}$$

where $\mathbf{e}(t) = \mathbf{y}(t) - H(t)\hat{\mathbf{x}}(t)$ and $\langle \mathbf{e}(t), \mathbf{e}(s)\rangle = R(t)\delta(t-s)$. Moreover, if

$$P(t,s) \;=\; \langle \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(s)\rangle, \qquad P(t) \;=\; \langle \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(t)\rangle,$$

then the error covariance matrix is given by

$$P(t|T) \;=\; P(t) - \int_t^T P(t,s)H^*(s)R^{-1}(s)H(s)P^*(t,s)ds. \tag{16.5.2}$$

■

The above results hold whether or not $\mathbf{x}(\cdot)$ has a state-space model; when it does, we can say more.

The Bryson-Frazier (BF) Formulas. When $\mathbf{x}(\cdot)$ has a state-space model, $\dot{\mathbf{x}}(t) = F(t)\mathbf{x}(t) + G(t)\mathbf{u}(t)$, then, as we saw, so does $\hat{\mathbf{x}}(\cdot)$, and so does the filtered error, $\tilde{\mathbf{x}}(t)$,

$$\begin{aligned}\dot{\tilde{\mathbf{x}}}(t) \;&=\; F(t)\tilde{\mathbf{x}}(t) + G(t)\mathbf{u}(t) - K(t)\mathbf{e}(t),\\ &=\; [F(t) - K(t)H(t)]\tilde{\mathbf{x}}(t) + \begin{bmatrix} G(t) & -K(t)\end{bmatrix}\begin{bmatrix}\mathbf{u}(t)\\ \mathbf{v}(t)\end{bmatrix}.\end{aligned} \tag{16.5.3}$$

Lemma 16.5.2 (Formula for $P(t,s)$) *For the model (16.1.1)–(16.1.3) with $S(t) = 0$, it holds that*

$$\langle \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(s)\rangle \;\triangleq\; P(t,s) \;=\; P(t)\Psi^*(s,t), \qquad s \ge t, \tag{16.5.4}$$

where $\Psi(s,t)$ is the state transition matrix of the closed-loop filter, $F(s) - K(s)H(s)$, defined as the unique solution of the linear matrix differential equation

$$\frac{d\Psi(s,t)}{ds} \;=\; [F(s) - K(s)H(s)]\Psi(s,t), \quad \Psi(0,0) = I. \tag{16.5.5}$$

■

Proof: From (16.5.3) we can write, for $s \ge t$,

$$\tilde{\mathbf{x}}(s) \;=\; \Psi(s,t)\tilde{\mathbf{x}}(t) + \int_t^s \Psi(s,\tau)\begin{bmatrix} G(\tau) & -K(\tau)\end{bmatrix}\begin{bmatrix}\mathbf{u}(\tau)\\ \mathbf{v}(\tau)\end{bmatrix} d\tau,$$

where the second term on the right-hand side is uncorrelated with $\tilde{\mathbf{x}}(t)$. Hence,

$$P(t,s) \;=\; \langle \tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(s)\rangle \;=\; \langle \tilde{\mathbf{x}}(t), \Psi(s,t)\tilde{\mathbf{x}}(t)\rangle \;=\; P(t)\Psi^*(s,t).$$

◆

Remark 11 [Notation]. We used the notation $\Phi_p(\cdot,\cdot)$ in the discrete-time case to denote the state-transition matrix of the closed-loop matrix associated with the one-step prediction of the state vector, as opposed to the filtered estimator. In continuous time, there is no distinction between prediction and filtering. So we drop the subscript p and use $\Psi(\cdot,\cdot)$ to avoid confusion with the transition matrix $\Phi(\cdot,\cdot)$ associated with $F(\cdot)$. We also write $F_{cl} = F(t) - K(t)H(t)$ to denote the closed-loop matrix. ◆

Substituting (16.5.4) into (16.5.1) yields

$$\hat{\mathbf{x}}(t|T) \;=\; \hat{\mathbf{x}}(t) + P(t)\boldsymbol{\lambda}(t|T), \quad 0 \le t \le T, \tag{16.5.6}$$

where

$$\boldsymbol{\lambda}(t|T) \;\triangleq\; \int_t^T \Psi^*(s,t)H^*(s)R^{-1}(s)\mathbf{e}(s)ds. \tag{16.5.7}$$

However, as usual in such situations, $\boldsymbol{\lambda}(t|T)$ is best described by the differential equation with respect to t that is obtained from the definition (16.5.7) (and using the property outlined in Prob. 16.14 for transition matrices),

$$\begin{aligned}\dot{\boldsymbol{\lambda}}(t|T) \;&=\; -[F(t) - K(t)H(t)]^*\boldsymbol{\lambda}(t|T) - H^*(t)R^{-1}(t)\mathbf{e}(t), &(16.5.8)\\ &=\; -F^*(t)\boldsymbol{\lambda}(t|T) + H^*(t)R^{-1}(t)H(t)\hat{\mathbf{x}}(t|T) -\\ &\quad - H^*(t)R^{-1}(t)\mathbf{y}(t), &(16.5.9)\end{aligned}$$

with $\boldsymbol{\lambda}(T|T) = 0$. The formula for smoothing obtained by combining (16.5.6) and (16.5.8) is named after Bryson and Frazier (1963), who derived it by variational arguments (which has led to the name "adjoint" variable for $\boldsymbol{\lambda}(t|T)$).

Theorem 16.5.1 (The BF Equations) *Given the model (16.1.1)–(16.1.3) with $S(t) = 0$, we can find the smoothed estimator $\hat{\mathbf{x}}(t|T)$ via (16.5.6) where $\boldsymbol{\lambda}(t|T)$ satisfies (16.5.8) or (16.5.9). Moreover, the smoothed error variance can then be readily computed as*

$$P(t|T) = P(t) - P(t)\left[\int_t^T \Psi^*(s,t)H^*(s)R^{-1}(s)H(s)\Psi(s,t)ds\right]P(t). \quad (16.5.10)$$

■

We gave the discrete-time version of the BF formulas in Sec. 10.2.1, where we also presented some formulas of Rauch, Tung, and Striebel (1965) and the two-filter formulas. Continuous-time versions of the RTS formulas can readily be obtained by our approach.

Theorem 16.5.2 (The RTS Equations) *Given the model (16.1.1)–(16.1.3) with $S(t) = 0$, we can find the smoothed estimator $\hat{\mathbf{x}}(t|T)$ by solving, backwards in time, the equation*

$$\dot{\hat{\mathbf{x}}}(t|T) = F_s(t)\hat{\mathbf{x}}(t|T) - G(t)Q(t)G^*(t)P^{-1}(t)\hat{\mathbf{x}}(t), \quad \hat{\mathbf{x}}(T|T) = \hat{\mathbf{x}}(T), \quad (16.5.11)$$

where $F_s(t) = F(t) + G(t)Q(t)G^*(t)P^{-1}(t)$. The smoothing errors variance obeys the following equation, with boundary condition $P(T|T) = P(T)$,

$$\frac{dP(t|T)}{dt} = F_s(t)P(t|T) + P(t|T)F_s^*(t) - G(t)Q(t)G^*(t). \quad (16.5.12)$$

■

Proof: For the estimator equation, just differentiate (16.5.6) to obtain, after some algebra, the backwards differential equation

$$\dot{\hat{\mathbf{x}}}(t|T) = F(t)\hat{\mathbf{x}}(t|T) + G(t)Q(t)G^*(t)\boldsymbol{\lambda}(t|T). \quad (16.5.13)$$

By further assuming that $P^{-1}(t)$ exists, we obtain the RTS formula (16.5.11). The error variance equation can be derived by differentiating the general formula (16.5.2), using also (16.5.4) and the Riccati equation for $P(\cdot)$. ◆

Remark 12. The RTS formulas would take considerable effort to obtain by a limiting argument applied to the discrete-time approximation. Note also that the formula for $dP(t|T)/dt$ is surprisingly similar, except for the negative sign, to the recursions for a variable obeying a standard state-space equation, *e.g.*, the equation for $\Pi(t) = \|\mathbf{x}(t)\|^2$. This suggests that the smoothing error obeys a backwards-time (hence, the negative sign) state-equation, which it does (see Prob. 16.28), exactly as in discrete time. ◆

The Hamiltonian Equations. We further note that by combining the equations (16.5.9) and (16.5.13) we obtain the so-called continuous-time Hamiltonian equations

$$\begin{bmatrix} \dot{\hat{\mathbf{x}}}(t|T) \\ -\dot{\boldsymbol{\lambda}}(t|T) \end{bmatrix} = \quad (16.5.14)$$

$$\begin{bmatrix} F(t) & G(t)Q(t)G^*(t) \\ -H^*(t)R^{-1}(t)H(t) & F^*(t) \end{bmatrix}\begin{bmatrix} \hat{\mathbf{x}}(t|T) \\ \boldsymbol{\lambda}(t|T) \end{bmatrix} + \begin{bmatrix} 0 \\ H^*(t)R^{-1}(t)\mathbf{y}(t) \end{bmatrix},$$

with boundary conditions

$$\boldsymbol{\lambda}(T|T) = 0, \quad \hat{\mathbf{x}}(0|T) = \hat{\mathbf{x}}(0) + P(0)\boldsymbol{\lambda}(0|T).$$

In Ch. 17 we provide a direct derivation of the Hamiltonian equations by making a deeper study of the structure of state-space models; this alternative derivation does not rely on the filtering and smoothing formulas of the earlier sections, and in fact, in Ch. 17 it will be made the starting point of a "physical" approach to stochastic estimation problems.

The Two-Filter Formulas. Diagonalization of the Hamiltonian equations will yield, after some algebra, the continuous-time analogs of the two-filter formulas (see Kailath and Ljung (1982)). One can give more direct stochastic derivations by using the properties of backwards Markovian state-space models (see App. 16.A), which also lead to other forms of smoothing formulas.

16.6 FAST ALGORITHMS FOR TIME-INVARIANT MODELS

As in discrete time, faster solutions are possible when the model is time-invariant. Note first that the effort required to solve the Riccati differential equation (16.2.5) is the same whether the model is time-variant or not. By one measure, we have to solve $n(n+1)/2$ (since $P(\cdot)$ is Hermitian) coupled nonlinear differential equations for the entries of $P(\cdot)$. On the other hand, if we consider any discretized version, we have to form a matrix product of $n \times n$ matrices $F(\cdot)$ and $P(\cdot)$, which will require $O(n^3)$ flops at each iteration, whether or not $F(\cdot)$ is constant.

To exploit the constancy of the state-space model we have to find a way of computing the gain function $K(\cdot)$ in (16.2.4) that does not require the computation of $P(\cdot)$.

Lemma 16.6.1 (Generalized Stokes Identity) *For a constant-parameter state-space model (16.1.1)–(16.1.3) with $S(t) = 0$, we have the identity*

$$\dot{P}(t) = \Psi(t,0)\dot{P}(0)\Psi^*(t,0), \quad (16.6.1)$$

where $\Psi(t,\cdot)$ is the state-transition matrix of the closed-loop matrix $F - K(t)H$. An immediate consequence of the above congruence relation is that the rank and the inertia of $\dot{P}(t)$ are constant in time and equal to those of $\dot{P}(0)$.[1] ■

Proof: Differentiate the Riccati equation (16.2.5) to obtain, exploiting the constancy, the results $K(t) = P(t)H^*R^{-1}$, $\dot{K}(t) = \dot{P}(t)H^*R^{-1}$, and

$$\begin{aligned}\ddot{P}(t) &= F\dot{P}(t) + \dot{P}(t)F^* + 0 - \dot{P}(t)H^*K(t) - K(t)H\dot{P}(t), \\ &= [F - K(t)H]\dot{P}(t) + \dot{P}(t)[F - K(t)H]^*. \end{aligned} \quad (16.6.2)$$

Assuming temporarily that $K(\cdot)$ is known, we can write the solution of this homogeneous linear differential equation as (see Prob. 16.16)

$$\dot{P}(t) = \Psi(t,0)\dot{P}(0)\Psi^*(t,0), \quad (16.6.3)$$

[1] Recall that a state transition matrix, like $\Psi(t,\cdot)$, is always full rank.

where

$$\frac{d\Psi(t,0)}{dt} = [F - K(t)H]\Psi(t,0), \quad \Psi(0,0) = I.$$

The validity of this result in general can now be checked by differentiating both sides of (16.6.3) to get the equation (16.6.2). ♦

Remark 13. The reason for the name Stokes identity will be explained in Ch. 17 where we discuss a scattering-theoretic (or transmission line) formulation of the estimation problem (see Sec. 17.5). ♦

Since $P(t)$ is Hermitian, so are $\dot{P}(t)$ and $\dot{P}(0)$; therefore we can always write $\dot{P}(0)$ as

$$\dot{P}(0) = L_0 J L_0^*, \tag{16.6.4}$$

where $J = (I_p \oplus -I_q)$ is the signature of $\dot{P}(0)$, *i.e.*, p is the number of positive eigenvalues of $\dot{P}(0)$, while q is the number of negative eigenvalues. Then

$$\begin{aligned} \alpha \triangleq p + q &= \text{the rank of } \dot{P}(0), \\ &= \text{rank of } (F\Pi_0 + \Pi_0 F^* + GQG^* - \Pi_0 H^* R^{-1} H \Pi_0). \end{aligned}$$

Therefore, L_0 will be an $n \times \alpha$ matrix. The factorization (16.6.4) can be obtained in many ways, using $O(n^2\alpha)$ flops. The factorization is not unique, but here let us assume that we have chosen a particular one. Now we can readily derive the following fast algorithm; it is the continuous-time counterpart of what we called earlier the CKMS algorithm in Ch. 11.

Theorem 16.6.1 (Fast Algorithm) *For the constant parameter state-space model (16.1.1)–(16.1.3) with $S(t) = 0$, the gain matrix $K(\cdot)$ of the Kalman filter in Thm. 16.2.1 can be computed by solving the following set of $n(p + \alpha)$ coupled nonlinear equations*

$$\dot{K}(t) = L(t) J L^*(t) H^* R^{-1}, \tag{16.6.5}$$

$$\dot{L}(t) = [F - K(t)H]L(t), \tag{16.6.6}$$

with initial conditions $K(0) = \Pi_0 H^* R^{-1}$ and $L(0) = L_0$. ■

Proof: From $K(t) = P(t)H^*R^{-1}$, we have $\dot{K} = \dot{P}(t)H^*R^{-1}$. Now by the previous lemma,

$$\dot{P}(t) = \Psi(t,0)\dot{P}(0)\Psi^*(t,0) = \Psi(t,0)L_0 J L_0^* \Psi^*(t,0).$$

Define $L(t) = \Psi(t,0)L_0$, from which it follows that

$$\dot{L}(t) = [F - K(t)H]\Psi(t,0)L_0 = [F - K(t)H]L(t),$$

with $L(0) = \Psi(0,0)L_0 = L_0$. Note also that $\dot{P}(t) = L(t)JL^*(t)$. Combining the above formulas gives the desired results. ♦

The significance of this result can be seen by considering two special cases.

Example 16.6.1 ($\Pi_0 = 0$) When $\Pi_0 = 0$, we see from the Riccati equation (16.2.5) that $\dot{P}(0) = GQG^*$, so that

$$\dot{P}(t) = \Psi(t,0)GQG^*\Psi^*(t,0).$$

Assuming for simplicity that G and Q have full rank, m, we can write $\dot{P}(0) = L_0 L_0^*$ with $L_0 = GQ^{1/2}$. Therefore, $\alpha = m$, and the equations are

$$\dot{K}(t) = L(t)L^*(t)H^*R^{-1}, \quad \dot{L}(t) = [F - K(t)H]L(t),$$

with $K(0) = 0$ and $L(0) = GQ^{1/2}$. So we have $n(p + m)$ simultaneous equations, which will generally be much less than the $n(n + 1)2$ equations of the Riccati-based algorithm. ◆

Example 16.6.2 (Stationary Processes) Assume that F is stable (*i.e.*, with eigenvalues in the open left-half plane) and choose $\Pi_0 = \bar{\Pi}$, where $\bar{\Pi}$ is the unique nonnegative-definite solution of the Lyapunov equation

$$0 = F\bar{\Pi} + \bar{\Pi}F^* + GQG^*.$$

Then from (16.4.15), we can see that the state covariance matrix, $\Pi(t) = \|\mathbf{x}(t)\|^2$, is constant, $\Pi(t) = \bar{\Pi}$. Therefore, the processes $\{\mathbf{x}(t), \mathbf{y}(t), t \geq 0\}$ will be stationary. In this case, we can write

$$\dot{P}(0) = F\bar{\Pi} + \bar{\Pi}F^* + GQG^* - \bar{\Pi}H^*R^{-1}H\bar{\Pi} = -L_0 L_0^*,$$

where $L_0 = \bar{\Pi}H^*R^{-*/2}$. This is assuming that H has full rank. Therefore, $\alpha = p$ and the equations (16.6.5)–(16.6.6) become

$$\dot{K}(t) = -L(t)L^*(t)H^*R^{-1}, \quad K(0) = \bar{\Pi}H^*R^{-1},$$

$$\dot{L}(t) = [F - K(t)H]L(t), \quad L(0) = \bar{\Pi}H^*R^{-*/2}.$$

Recalling from (16.2.12) that the covariance function of $\mathbf{y}(\cdot)$ is

$$R_y(t) = R(t)\delta(t) + He^{Ft}\bar{N}1(t) + \bar{N}^*e^{F^*t}1(-t),$$

where $\bar{N} = \bar{\Pi}H^*$, we see that the $\{K(\cdot), L(\cdot)\}$ equations are completely specified by knowledge of $R_y(\cdot)$ in state-space form. With a little more effort, the same can be shown for the general equations of Thm. 16.6.1 (*cf.* the discussion in discrete time in Sec. 11.4). ◆

Remark 14. As in discrete time (*cf.* Ch. 11), the quantity α can be related to the displacement rank, r, of the covariance function $R_y(t,s)$, which is defined as

$$r \triangleq \text{rank}\left[\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)[R_y(t,s) - R\delta(t-s)]\right].$$

For more on this, we refer to Kailath, Ljung, and Morf (1983). ◆

16.7 ASYMPTOTIC BEHAVIOR

We now study conditions under which the Riccati differential equation (16.2.5) of the Kalman filter solution is guaranteed to converge to a steady-state value P (in the time-invariant case). We shall assume, for simplicity, that $S = 0$. The discussion in this section parallels that in Ch. 14 and is therefore very brief. However, we do comment on the differences from the more complicated (or richer, depending upon one's point of view) discrete-time results.

16.7.1 Positive-Semi-Definite Solutions of the CARE

Consider the continuous-time algebraic Riccati equation (CARE)

$$0 = FP + PF^* + GQG^* - PH^*R^{-1}HP. \tag{16.7.1}$$

It may have many positive-semi-definite solutions P, not all of which yield a stable closed-loop matrix F_{cl} (*i.e.*, with eigenvalues in the left-half plane). The next two results clarify under what conditions stabilizing (as well as positive-semi-definite) solutions exist and are unique. The results are established in App. E (Thms. E.9.2 and E.9.3).

Theorem 16.7.1 (Algebraic Riccati Equation) *Consider the* CARE *(16.7.1). Then the following two statements are equivalent:*

(i) $\{F, H\}$ *is detectable and* $\{F, GQ^{1/2}\}$ *is controllable on the imaginary axis.*

(ii) The CARE *has a stabilizing solution* P*, i.e., one for which the matrix* $F - KH$ *is stable, where* $K = PH^*R^{-1}$.

Moreover, any such stabilizing solution is unique and positive-semi-definite. ∎

Theorem 16.7.2 (Unique Positive-Semi-Definite Solution) *Consider the* CARE *(16.7.1) and assume that* $\{F, H\}$ *is detectable and* $\{F, GQ^{1/2}\}$ *is controllable on the imaginary axis. Then the following two statements are equivalent:*

(i) $\{F, GQ^{1/2}\}$ *is stabilizable.*

(ii) The CARE *has a unique positive-semi-definite solution.*

Moreover, the unique positive-semi-definite solution of the CARE *is given by its stabilizing solution.* ∎

Therefore, in order to guarantee a *unique* positive-semi-definite solution, we need the additional condition that $\{F, GQ^{1/2}\}$ be *stabilizable*, *i.e.*, that it be controllable on and to the left of the imaginary axis (and not just on the imaginary axis, which was the condition we required for the existence of a stabilizing solution to the CARE).

16.7.2 Convergence Results

We now invoke the results of Ex. 16.3.4 and Prob. 16.17 (see also the discussion in Sec. 17.4.2). These results allow us to relate the solutions of the Riccati differential equation (16.2.5) for two different initial conditions. So let $\delta P(t) = P^{(2)}(t) - P^{(1)}(t)$ and, hence, $\delta P(0) = \Pi_0^{(2)} - \Pi_0^{(1)}$. Then, when the required inverse exists,

$$\delta P(t) = \Psi^{(1)}(t,0)[I + \delta P(0)\mathcal{O}^{(1)}(t,0)]^{-1}\delta P(0)\left[\Psi^{(1)}(t,0)\right]^*, \tag{16.7.2}$$

where we defined the observability Gramian

$$\mathcal{O}^{(1)}(t,0) \triangleq \int_0^t \left[\Psi^{(1)}(\tau,0)\right]^* H^*R^{-1}H\Psi^{(1)}(\tau,0)d\tau. \tag{16.7.3}$$

Here $\Psi^{(1)}(t,0)$ is the state transition matrix for the closed-loop system that corresponds to the initial condition $\Pi_0^{(1)}$.

The result (16.7.2) is very useful in establishing convergence results for the Riccati differential equation (16.2.5). To see why, suppose P is some solution to the CARE (16.7.1). Then, using (16.7.2), we may write

$$P(t) - P = e^{F_{cl}t}[I + (P(0) - P)\mathcal{O}(t,0)]^{-1}(P(0) - P)\left[e^{F_{cl}t}\right]^*, \tag{16.7.4}$$

where $P(t)$ is the solution to (16.2.5) with initial condition $P(0)$, and $\mathcal{O}(t,0)$ denotes the observability Gramian of the pair $\{F_{cl}, R^{-1/2}H\}$,

$$\mathcal{O}(t,0) = \int_0^t \left[e^{F_{cl}\tau}\right]^* H^*R^{-1}He^{F_{cl}\tau}d\tau. \tag{16.7.5}$$

Now if P is chosen as the stabilizing solution to the CARE (assuming such a solution exists), then the matrix $F_{cl} = F - KH$ is stable and we have $\lim_{t\to\infty} e^{F_{cl}t} = 0$. Then using (16.7.4), we see that if the initial condition $P(0)$ is such that the matrix function

$$T(t) \triangleq [I + (P(0) - P)\mathcal{O}(t,0)]^{-1}(P(0) - P) \tag{16.7.6}$$

is uniformly bounded for all t, then $P(t)$ will converge to P. We shall often find it convenient to define $P_0 - P = A_0JA_0^*$, where A_0 has full column rank and J is a signature matrix, so that we can write $T(t)$ in the following more symmetric form:

$$T(t) = A_0(J + A_0\mathcal{O}(t,0)A_0^*)^{-1}A_0^*. \tag{16.7.7}$$

Now in the discrete-time case, the uniform boundedness of the sequence $\{T_i\}$ defined by (14.3.3) was seen to be a *sufficient* condition for the convergence of the Riccati recursion (*cf.* Thm. 14.5.1). The main reason for this was that the closed-loop state-transition matrix, F_p, could be singular, so that the product of terms on the right-hand side of (14.3.1) could be zero or finite even when T_i is not bounded (see the example following the statement of Thm. 14.5.1). Here, however, it turns out that the uniform boundedness of $T(t)$ is both *necessary and sufficient*. The reason is that the state-transition matrix, $e^{F_{cl}t}$, is nonsingular for all t.

Theorem 16.7.3 (A General Convergence Result) *Consider the Riccati differential equation (16.2.5), with initial condition* $P(0)$*, and suppose that* $\{F, H\}$ *is detectable and* $\{F, GQ^{1/2}\}$ *is controllable on the imaginary axis. Then* $P(t)$ *converges to* P*, the unique stabilizing solution of the* CARE *(16.7.1) if, and only if, the family of matrices*

$$T(t) = [I + (P(0) - P)\mathcal{O}(t,0)]^{-1}(P(0) - P), \tag{16.7.8}$$

is uniformly bounded for all $t \geq 0$*. Moreover, the convergence of* $P(t)$ *to* P *is exponential, viz.,*

$$\|P(t) - P\| \leq me^{2\lambda t}, \tag{16.7.9}$$

for some matrix norm $\|\cdot\|$*, a scalar* $\lambda < 0$*, and a finite positive* m. ∎

Proof: Clearly, if the matrices $T(t)$ are uniformly bounded for all t, we have convergence to P. Therefore we shall focus on the proof of the converse statement: if the $T(t)$ are not uniformly bounded, then we cannot have convergence to P.

Thus, let us first assume that $T(t)$ becomes unbounded for some finite $t = \tau$. Then clearly, since $e^{F_{cl}\tau}$ is invertible, (16.7.4) implies that $P(t) - P$ will become unbounded, and so we cannot have convergence. Thus, assume that $T(t)$ becomes unbounded in the limit, which requires that

$$\lim_{t\to\infty} [J + A_0\mathcal{O}(t,0)A_0^*] = J + A_0\mathcal{O}A_0^*,$$

to be singular, where we have defined

$$\mathcal{O} \triangleq \lim_{t\to\infty} \mathcal{O}(t,0) = \int_0^\infty \left[e^{F_{cl}\tau}\right]^* H^*R^{-1}He^{F_{cl}\tau}d\tau.$$

Our goal is to show that $P(t) - P$ cannot tend to zero. To see this, note first that

$$\begin{aligned}
\mathcal{O}(t,0) &= \mathcal{O} - \int_t^\infty e^{F_{cl}^*\tau}H^*R^{-1}He^{F_{cl}\tau}d\tau \\
&= \mathcal{O} - e^{F_{cl}^*t}\left[\int_t^\infty e^{F_{cl}^*(-t+\tau)}H^*R^{-1}He^{F_{cl}(-t+\tau)}d\tau\right]e^{F_{cl}t} \\
&= \mathcal{O} - e^{F_{cl}^*t}\left[\int_0^\infty e^{F_{cl}^*\tau}H^*R^{-1}He^{F_{cl}\tau}d\tau\right]e^{F_{cl}t} \\
&= \mathcal{O} - e^{F_{cl}^*t}\mathcal{O}e^{F_{cl}t},
\end{aligned} \tag{16.7.10}$$

where, in the third step, we used the change of variables $-t+\tau \to \tau$. Note that (16.7.10) shows the exponential convergence of $\mathcal{O}(t,0)$ to $\mathcal{O}$.

We now invoke the following readily verified fact: for any two matrices A and B, *the matrices AB and BA share the same nonzero eigenvalues.* This implies that, for all finite t, the matrices

$$\mathcal{O}[P(t) - P] = \mathcal{O}e^{F_{cl}t}A_0\left[J + A_0^*\mathcal{O}A_0 - A_0^*e^{F_{cl}^*t}\mathcal{O}e^{F_{cl}t}A_0\right]^{-1}A_0^*e^{F_{cl}^*t}$$

and

$$\begin{aligned}
&\left[J + A_0^*\mathcal{O}A_0 - A_0^*e^{F_{cl}^*t}\mathcal{O}e^{F_{cl}t}A_0\right]^{-1}A_0^*e^{F_{cl}^*t}\mathcal{O}e^{F_{cl}t}A_0 = \\
&\quad -I + \left[J + A_0^*\mathcal{O}A_0 - A_0^*e^{F_{cl}^*t}\mathcal{O}e^{F_{cl}t}A_0\right]^{-1}(J + A_0^*\mathcal{O}A_0)
\end{aligned} \tag{16.7.11}$$

must have the same set of nonzero eigenvalues. But, since $J + A_0^*\mathcal{O}A_0$ is singular, the matrix on the right-hand side of (16.7.11) will have an eigenvalue at -1 for all t, implying that the matrix $\mathcal{O}[P(t) - P]$ will have an eigenvalue at -1 for all finite t. Thus, $P(t) - P$ cannot approach zero, meaning that we cannot have convergence.

This concludes our proof that the uniform boundedness of $T(\cdot)$ is equivalent to the convergence of $P(\cdot)$ to P. The proof of the exponential convergence is identical to the proof of Thm. 14.5.1, and so will not be repeated here. ◆

In discrete time in Thm. 14.5.2 we showed that a sufficient condition for the uniform boundedness of the $\{T_i\}$ (which, in turn, was a sufficient condition for convergence to P) was given by the set of initial conditions for which $I + \mathcal{O}^{p*/2}(P_0 - P)\mathcal{O}^{p/2} > 0$, where

$\mathcal{O}^p$ was given by (14.3.5). The continuous-time counterpart of the latter condition is, in fact, equivalent to the uniform boundedness of $T(t)$. Therefore we obtain the following result, which characterizes the set of intial conditions for which the solution of the Riccati differential equation converges to P.

Theorem 16.7.4 (Basin of Attraction) *Consider the Riccati differential equation (16.2.5), with initial condition $P(0)$, and suppose that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the imaginary axis. Then $P(t)$ converges to P, the unique stabilizing solution of the* CARE *(16.7.1) if, and only if,*

$$\left[I + \mathcal{O}^{*/2}(P(0) - P)\mathcal{O}^{1/2}\right] > 0, \tag{16.7.12}$$

where

$$\mathcal{O} = \mathcal{O}^{1/2}\mathcal{O}^{*/2} = \int_0^\infty \left[e^{F_{cl}\tau}\right]^* H^*R^{-1}He^{F_{cl}\tau}d\tau, \tag{16.7.13}$$

is the unique solution of the Lyapunov equation (cf. App. C)

$$\mathcal{O}F_{cl} + F_{cl}^*\mathcal{O} + H^*R^{-1}H = 0. \tag{16.7.14}$$

■

Proof: We shall show that (16.7.12) is equivalent to the uniform boundedness of $T(t)$. First, using an argument similar to the proof of Lemma 14.5.1, we can verify that $T(t)$ is uniformly bounded if, and only if, the matrices $J + A_0^*\mathcal{O}(t,0)A_0$ are nonsingular for all t, including in the limit as $t \to \infty$.

Now, as in the proof of Lemma 14.5.3, we can establish that the eigenvalues of the matrices $J + A_0^*\mathcal{O}(t,0)A_0$ are nondecreasing functions of t. This implies that the matrices $J + A_0^*\mathcal{O}(t,0)A_0$ will be nonsingular for all t if, and only if, they have constant inertia for all t (since, for the matrices to change inertia, we would require that one of the negative eigenvalues increase and pass through zero, which violates the invertibility assumption).

Therefore, we conclude that the matrices $J + A_0^*\mathcal{O}(t,0)A_0$ are uniformly nonsingular, if, and only if, the matrices J (corresponding to $t = 0$) and $J + A_0^*\mathcal{O}A_0$ (corresponding to $t = \infty$) have the same inertia. Now consider the matrix

$$\begin{bmatrix} -J & A_0^*\mathcal{O}^{1/2} \\ \mathcal{O}^{*/2}A_0 & I \end{bmatrix}.$$

Two different block LDU and UDL factorizations of the above matrix show that the matrices

$$\begin{bmatrix} -J - A_0^*\mathcal{O}A_0 & 0 \\ 0 & I \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -J & 0 \\ 0 & I + \mathcal{O}^{*/2}A_0JA_0^*\mathcal{O}^{1/2} \end{bmatrix}$$

are congruent. Thus, $J + A_0^*\mathcal{O}A_0$ and J have the same inertia if, and only if,

$$I + \mathcal{O}^{*/2}A_0JA_0^*\mathcal{O}^{1/2} = I + \mathcal{O}^{*/2}(P(0) - P)\mathcal{O}^{1/2} > 0.$$

◆

In conclusion, we note that the convergence analysis in continuous time is considerably simpler than in discrete-time. In particular, condition (16.7.12) gives a very simple characterization of the set of *all* initial conditions for which the Riccati differential equation converges to the stabilizing solution of the corresponding CARE.

16.7.3 The Dual CARE

As in discrete time, it is possible to give an alternative condition for the convergence of the Riccati differential equation by introducing the dual CARE,

$$0 = F^*P^a + P^aF + H^*R^{-1}H - P^aGQG^*P^a. \tag{16.7.15}$$

Thm. E.9.4 of App. E gives conditions for the existence of a stabilizing solution to this equation.

Theorem 16.7.5 (Convergence with Indefinite $P(0)$) *Consider the Riccati differential equation (16.2.5) where $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is stabilizable. Then $P(t)$ converges to the unique positive-semi-definite matrix P that satisfies the* CARE *(16.7.15) if, and only if,*

$$I + (P^a)^{*/2}P(0)(P^a)^{1/2} > 0, \tag{16.7.16}$$

where $P^a = (P^a)^{1/2}(P^a)^{*/2}$ is the unique stabilizing solution to the dual CARE *(16.7.15).* ∎

Note that the above result guarantees convergence of the Riccati differential equation for all positive-semi-definite initial conditions, $P(0) \geq 0$, since the matrix $I+(P^a)^{*/2}P(0)(P^a)^{1/2}$ will clearly be positive-definite. In particular, it establishes convergence for the zero-initial condition Riccati differential equation (*i.e.*, with $P(0) = 0$). However, we can also guarantee convergence for indefinite, and even negative definite, initial conditions, as long as (16.7.16) is satisfied.

16.7.4 Exponential Convergence of the Fast Filtering Equations

It follows from the proof of the Chandrasekhar-Kailath equations in Thm. 16.6.1 that $\dot{P}(t) = L(t)JL^*(t)$, where

$$L(t) = \Psi(t, 0)L_0, \tag{16.7.17}$$

and $\Psi(t, 0)$ is the state transition matrix associated with $F - K(t)H$. Now we already know from the results in the previous section that, under suitable conditions on $\{F, G, H, Q, P(0)\}$, $P(t)$ converges to P exponentially fast. So let K denote the steady-state value of $K(\cdot)$, $K = PH^*R^{-1}$. To establish a similar result for $L(\cdot)$, we invoke the following identity (established in Prob. 16.32):

$$\Psi(t, 0) = e^{(F-KH)t}[I + (P(0) - P)\mathcal{O}(t, 0)]^{-1}, \tag{16.7.18}$$

where $\mathcal{O}(t, 0)$ is given by (16.7.5). Substituting into (16.7.17) we obtain

$$L(t) = e^{(F-KH)t}[I + (P(0) - P)\mathcal{O}(t, 0)]^{-1}L_0,$$

which establishes the convergence of $L(t)$ exponentially to zero, since the matrix $[I + (P(0) - P)\mathcal{O}(t, 0)]^{-1}$ is uniformly bounded.

16.8 THE STEADY-STATE FILTER

With the above convergence results at hand, we can now comment more directly on the form of the steady-state Kalman filter. Thus note first that even when specialized to a time-invariant model, the Kalman filter for processes observed for $t \geq 0$ (*cf.* Thm. 16.2.1)

$$\mathbf{e}(t) = \mathbf{y}(t) - H\hat{\mathbf{x}}(t), \quad \mathbf{e}(0) = \mathbf{y}(0), \tag{16.8.1}$$

$$\dot{\hat{\mathbf{x}}}(t) = F\hat{\mathbf{x}}(t) + K(t)\mathbf{e}(t), \quad t \geq 0, \tag{16.8.2}$$

where $K(t) = P(t)H^*R^{-1}$ and $P(t)$ satisfies

$$\dot{P}(t) = FP(t) + P(t)F^* + GQG^* - K(t)RK^*(t), \tag{16.8.3}$$

with $P(0) = \Pi_0$, is *time-variant* and will be so even if Π_0 is chosen so as to make the process $\{\mathbf{y}(\cdot)\}$ stationary over $[0, \infty)$. However, as indicated in Thm. 16.7.5, when $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is stabilizable, and for any $\Pi_0 \geq 0$, the time-variant Kalman filter approaches as $t \to \infty$, the time-invariant implementation

$$\mathbf{e}(t) = \mathbf{y}(t) - H\hat{\mathbf{x}}(t), \tag{16.8.4}$$

$$\dot{\hat{\mathbf{x}}}(t) = F\hat{\mathbf{x}}(t) + K\mathbf{e}(t), \tag{16.8.5}$$

where now $K = PH^*R^{-1}$ and P is the unique stabilizing solution of the CARE

$$0 = FP + PF^* + GQG^* - KRK^*. \tag{16.8.6}$$

In this case, the transfer function from $\mathbf{e}(\cdot)$ to $\mathbf{y}(\cdot)$ will be given by

$$L(s) = H(sI - F)^{-1}K + I. \tag{16.8.7}$$

Now recall that time-invariant state-space models were also studied in App. 8.D, where it is evident that as time goes to infinity, the output of any *stable* time-invariant state-space model converges to a stationary process. For such stationary processes, and for estimation problems given a semi-infinite observation interval, and with the assumption that the pair $\{F, GQ^{1/2}\}$ is controllable on the imaginary axis or, in less model-dependent language, that the s-spectrum $S_y(s)$ of the output process $\{\mathbf{y}(\cdot)\}$ has no zeros on the imaginary axis, we were able to construct the Wiener filter for determining

the associated innovations process as in (8.D.12), which agrees with the above expression for $L(s)$. In other words, we saw in App. 8.D that the above $L(s)$ defines the canonical factor of $S_y(s)$ in (8.D.6) and that, therefore, it can be determined from the unique stabilizing solution of a CARE.[2] While this result was established there only for stable F, the point is that the Wiener and Kalman approaches can both be pursued to yield the same result, after we incorporate Kalman's crucial and fruitful insight as to introducing state-space descriptions. Still, in the latter (Kalman) approach, the stability requirement on F can be replaced by the weaker requirement of a detectable pair $\{F, H\}$.

16.9 COMPLEMENTS

There is of course considerably more that can be said about continuous-time problems, *e.g.,* on square-root algorithms, colored noise problems, displacement structure, extended Kalman filters, duality, complementary models, applications to control, etc. Instead, and apart from a few results in the problems, we content ourselves with a few largely historical notes.

The Paper of Kalman and Bucy (1961). This was the paper that first presented the general continuous-time equations. Moreover, as Kalman (1960a) had already noted in discrete time, a duality to a previously obtained (see Kalman (1960b,1960c)) solution of a quadratic regulator problem was noted and used to obtain certain convergence and stability results for the estimation problem.

Here we comment on the solution technique used in Kalman and Bucy (1961). As mentioned in the notes to Ch. 9, though Kalman (1960a) had used the discrete-time white-noise process $\{\mathbf{e}_i\}$, he did not notice or exploit the obvious equivalence (in discrete time) of $\{\mathbf{e}_i\}$ and $\{\mathbf{y}_i\}$. Moreover, the corresponding results are less obvious in continuous-time.

So following earlier partial results of Carlton and Follin (1956), Hanson (1959), Bucy (1959) that, as traditional at the time, used the Wiener-Hopf equation, Kalman and Bucy (1961) give a proof showing how to solve the equation by exploiting the state-space structure. We give a (streamlined) version of that argument here.

The problem is, for a standard state-space model with $S(\cdot) \equiv 0$, to find

$$\hat{\mathbf{x}}(t) = \int_0^t h_{xy}(t, \tau)\mathbf{y}(\tau)d\tau, \tag{16.9.1}$$

where

$$\int_0^t h_{xy}(t, \tau)R_y(\tau, s)d\tau = R_{xy}(t, s), \quad 0 \le s \le t, \tag{16.9.2}$$

[2] We may mention that there are also alternative methods for determining the optimal gain K, especially for SISO systems, such as the symmetric root-locus method (see, *e.g.,* Letov (1960), Chang (1961), Kwakernaak and Sivan (1972), and Kailath (1980)).

and the covariances $\{R_{xy}(\cdot, \cdot), R_y(\cdot, \cdot)\}$ are given by (with the now familiar notation)

$$R_{xy}(t, s) \triangleq \langle \mathbf{x}(t), \mathbf{y}(s) \rangle = \Phi(t, s)\Pi(s)H^*(s), \quad s \le t, \tag{16.9.3}$$

$$R_y(t, s) \triangleq \langle \mathbf{y}(t), \mathbf{y}(s) \rangle = R(t)\delta(t - s) + K(t, s), \tag{16.9.4}$$

where

$$K(t, s) = \begin{cases} H(t)R_{xy}(t, s), & s \le t, \\ R_{xy}^*(s, t)H^*(s), & s > t. \end{cases} \tag{16.9.5}$$

Substituting the expression for $R_y(\cdot, \cdot)$ into (16.9.2) we find that $h_{xy}(\cdot, \cdot)$ should satisfy

$$h_{xy}(t, s)R(s) + \int_0^t h_{xy}(t, \tau)K(\tau, s)d\tau = R_{xy}(t, s), \quad 0 \le s \le t. \tag{16.9.6}$$

The first step toward solving this equation is to note that

$$\frac{\partial R_{xy}(t, s)}{\partial t} = F(t)R_{xy}(t, s), \quad s < t. \tag{16.9.7}$$

Then differentiating (16.9.6) with respect to t we obtain

$$\frac{\partial h_{xy}(t, s)}{\partial t} R(s) + h_{xy}(t, t)K(t, s) + \int_0^t \frac{\partial h_{xy}(t, \tau)}{\partial t} K(\tau, s)d\tau = F(t)R_{xy}(t, s),$$

or, equivalently, by using expression (16.9.5) for $K(t, s)$ and (16.9.7) for $s \le t$,

$$\frac{\partial h_{xy}(t, s)}{\partial t} R(s) + \int_0^t \frac{\partial h_{xy}(t, \tau)}{\partial t} K(\tau, s)d\tau = [F(t) - h_{xy}(t, t)H(t)]R_{xy}(t, s).$$

Comparing this integral equation for $\partial h_{xy}(\cdot, \cdot)/\partial t$ with that for $h_{xy}(\cdot, \cdot)$ in (16.9.6), we conclude by linearity that

$$\frac{\partial h_{xy}(t, s)}{\partial t} = [F(t) - h_{xy}(t, t)H(t)]h_{xy}(t, s), \quad 0 \le s \le t.$$

Now, for $\tau < t$,

$$K(\tau, t) = R_{xy}^*(t, \tau)H^*(t) = E\mathbf{y}(\tau)\mathbf{x}^*(t)H^*(t),$$

so that from (16.9.6) we can write

$$\begin{aligned} h_{xy}(t, t)R(t) &= R_{xy}(t, t) - E\left[\int_0^t h_{xy}(t, \tau)\mathbf{y}(\tau)d\tau\right]\mathbf{x}^*(t)H^*(t), \\ &= [E\mathbf{x}(t)\mathbf{x}^*(t)]H^*(t) - [E\hat{\mathbf{x}}(t)\mathbf{x}^*(t)]H^*(t), \\ &= [E\tilde{\mathbf{x}}(t)\mathbf{x}(t)]H^*(t) = P(t)H^*(t). \end{aligned}$$

In other words, we find that

$$h_{xy}(t,t) = P(t)H^*(t)R^{-1}(t) \triangleq K(t).$$

It then follows from (16.9.1) that

$$\begin{aligned}\dot{\hat{\mathbf{x}}}(t) &= \frac{\partial}{\partial t}\left[\int_0^t h_{xy}(t,\tau)\mathbf{y}(\tau)d\tau\right],\\ &= h_{xy}(t,t)\mathbf{y}(t) + \int_0^t [F(t) - h_{xy}(t,t)H(t)]h_{xy}(t,\tau)\mathbf{y}(\tau)d\tau,\\ &= F(t)\hat{\mathbf{x}}(t) + P(t)H^*(t)R^{-1}(t)[\mathbf{y}(t) - H(t)\hat{\mathbf{x}}(t)],\end{aligned}$$

which is the differential equation for the state estimator (*cf.* (16.2.3)).

Though this derivation is quite elegant, it is less direct in that it does have to bring in the covariance functions of $\{\mathbf{x}(\cdot), \mathbf{y}(\cdot)\}$. As discussed earlier, see, *e.g.,* Sec. 8.6, one of the valuable insights of the Kalman theory is that given a model, there is no need to go to the covariance functions: all the required statistical information to solve the problem is present in the model — so a direct solution should be possible. In fact, this was the route taken for the discrete-time problem in Kalman (1960a) and in Ch. 9.

Of course the indirect route can be avoided by using the common (in many subjects) method of solving continuous-time problems via discretization; this was the method used in Kalman (1963b), and in Sec. 16.1.2. However, as noted earlier, to keep doing this to obtain smoothing formulas, fast algorithms, convergence and steady-state results, etc., can be laborious and tiresome. The introduction and use of the continuous-time innovations process (Kailath (1968)) allows us to preserve the integrity of the original Kalman insight — if we have a model, we don't need the covariance functions. At least we don't need them to get the estimator algorithms; they may be needed, and in fact are needed (*cf.* Sec. 16.7) to understand the properties of the algorithms.

Moreover, as with Kalman's approach in the discrete-time case, the method of Kalman-Bucy was not easy to extend to the smoothing problem (unlike the innovations approach). However, we can also develop the integral equations approach further. First, however, let us note the important and (generally) much overlooked work of Stratonovich, a pioneer in the use of Markov process models for estimation, communications, and control problems.

The Contributions of R. L. Stratonovich. In the late 1950s, R. L. Stratonovich, working in Moscow, began to urge, in many articles and books, that researchers in estimation theory, detection theory, control, stochastic processes, etc., begin to go beyond the study of Gaussian processes to Markov processes and to observations of Markov processes corrupted by additive white noise (which he called "conditional Markov processes"). This led him to generalize the classical Fokker-Planck equations for the evolution of the transition probability density functions to obtain partial differential equations for the conditional probability density function of the signal process given the observations. Now we recall from App. 3.A that the conditional mean is the least-mean-squares estimator. Now in writing down an equation for the conditional mean from the partial differential equations we find that this equation requires knowledge of conditional variance. But the equation for the conditional variance involves the conditional third moment. And so on. One has an "infinite" chain of equations that is hard to terminate in any reasonable fashion.

Stratonovich recognized that the chain could be terminated for Gaussian Markov processes; in this case, the equation for the conditional covariance is a deterministic Riccati differential equation, whose solution can be used to completely define the evolution equation for the conditional mean, see, *e.g.* Stratonovich (1960a, Eq. (59)), which gives (of course, in a different notation) exactly the continuous-time Kalman-Bucy equations! While Stratonovich was well recognized for his work on the conditional density function,[3] his result for the linear/Gaussian case has generally been overlooked. It was overshadowed by the almost simultaneous appearance of the papers Kalman (1960a) and Kalman and Bucy (1961), which focused directly on the simpler linear problem.

Smoothing via Integral Equations. Let

$$\mathbf{y}(t) = \mathbf{z}(t) + \mathbf{v}(t), \quad 0 \le t \le T,$$

with

$$\langle \mathbf{v}(t), \mathbf{v}(s)\rangle = I\delta(t-s), \quad \langle \mathbf{v}(t), \mathbf{z}(s)\rangle = 0, \quad \langle \mathbf{z}(t), \mathbf{z}(s)\rangle = K(t,s),$$

and assume that $K(t,s)$ is differentiable in t and s. Let

$$\hat{\mathbf{z}}(t|T) \triangleq \int_0^T H_s(t,\tau;T)\mathbf{y}(\tau)d\tau, \tag{16.9.8}$$

and

$$\hat{\mathbf{z}}(t) \triangleq \hat{\mathbf{z}}(t|t) = \int_0^t H_f(t,\tau)\mathbf{y}(\tau)d\tau. \tag{16.9.9}$$

Note that

$$H_f(t,\tau) = H_s(t,\tau;t). \tag{16.9.10}$$

The subscripts f and s of course correspond to the filtering and smoothing problems. Now use of the orthogonality condition leads to the equation

$$H_s(t,s;T) + \int_0^T H_s(t,\tau;T)K(\tau,s)d\tau = K(t,s). \tag{16.9.11}$$

[This is called a Fredholm integral equation of the second kind and $H_s(\cdot,\cdot;T)$ is known as the resolvent of $K(\cdot,\cdot)$.]

Now under the differentiability assumptions, a straightforward calculation leads to the so-called Siegert-Krein-Bellman identity

$$\frac{\partial H_s(t,s;T)}{\partial T} = -H_s(t,T;T)H_s(T,s;T), \tag{16.9.12}$$

$$= -H_f^*(T,t)H_f(T,s), \tag{16.9.13}$$

[3] For example, Bucy writes in the historical remarks in Bucy and Joseph (1968): "For nonlinear filtering, of course, the central ideas and methods are all supplied in Stratonovich (1960a)."

where H_f^* is the operator adjoint to H_f, with kernel

$$H_f^*(t,s) \triangleq H_f(s,t). \tag{16.9.14}$$

The integrated form of this identity is

$$H_s(t,s;T) = H_f(t,s) + H_f^*(t,s) - \int_0^T H_f^*(t,\tau)H_f(T,s)d\tau, \tag{16.9.15}$$

from which it follows immediately that

$$\hat{\mathbf{z}}(t|T) = \hat{\mathbf{z}}(t) + \int_t^T H_f^*(t,\tau)\mathbf{e}(\tau)d\tau, \tag{16.9.16}$$

where $\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{z}}(t)$, the innovations process of $\mathbf{y}(\cdot)$.

These results were obtained (see Kailath (1969b)) before the innovations process results were derived. It will be useful to see how directly the innovations approach can illuminate and extend the above results.

Innovations and Operator Factorization. We shall use the operator notation introduced in Sec. 16.4.1 and our discussion will be brief (see Kailath (1969b)).

(a) We first note that several of the above expressions can be rewritten in operator form as, in an obvious notation,

$$\mathcal{R}_y = \mathcal{I} + \mathcal{K},$$
$$\mathcal{H}_s + \mathcal{H}_s\mathcal{K} = \mathcal{K} \quad \text{or} \quad \mathcal{H}_s(\mathcal{I} + \mathcal{K}) = \mathcal{K}.$$

Therefore, we can write

$$\mathcal{H}_s = \mathcal{K}(\mathcal{I} + \mathcal{K})^{-1} = (\mathcal{I} + \mathcal{K})^{-1}\mathcal{K} = \mathcal{I} - (\mathcal{I} + \mathcal{K})^{-1} = \mathcal{I} - \mathcal{R}_y^{-1},$$

or

$$\mathcal{R}_y^{-1} = (\mathcal{I} + \mathcal{K})^{-1} = \mathcal{I} - \mathcal{H}_s.$$

The operator $\mathcal{H}_s$ is called the resolvent because the solution of the integral equation, with $\mathcal{K}$ and m known,

$$\mathcal{R}_y a = (\mathcal{I} + \mathcal{K})a = m,$$

is

$$a = \mathcal{R}_y^{-1}m = (\mathcal{I} + \mathcal{K})^{-1}m = (\mathcal{I} - \mathcal{H}_s)m.$$

(b) Next observe that the innovations process can be obtained by using $\mathcal{H}_f$, *i.e.*,

$$\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{z}}(t) = (\mathcal{I} - \mathcal{H}_f)\mathbf{y},$$

and the whiteness of the innovations process leads to

$$\langle \mathbf{e}(t), \mathbf{e}(t)\rangle = I = (\mathcal{I} - \mathcal{H}_f)\mathcal{R}_y(\mathcal{I} - \mathcal{H}_f^*).$$

We define

$$(\mathcal{I} - \mathcal{H}_f)^{-1} \triangleq \mathcal{I} + \mathcal{L},$$

where $\mathcal{L}$ is causal because $\mathcal{H}_f$ is causal. Then we can write[4]

$$\mathcal{I} + \mathcal{K} = \mathcal{R}_y = (\mathcal{I} - \mathcal{H}_f)^{-1}(\mathcal{I} - \mathcal{H}_f^*)^{-1} = (\mathcal{I} + \mathcal{L})(\mathcal{I} + \mathcal{L}^*), \tag{16.9.17}$$

the canonical factorization of $\mathcal{R}_y$. We also have the representation

$$\mathcal{K} = \mathcal{L} + \mathcal{L}^* + \mathcal{L}\mathcal{L}^*, \tag{16.9.18}$$

which is often useful.

(c) From (16.9.17) it also follows that

$$(\mathcal{I} - \mathcal{H}_f^*)(\mathcal{I} - \mathcal{H}_f) = \mathcal{R}_y^{-1} = \mathcal{I} - \mathcal{H}_s,$$

which again leads to the identity (16.9.15),

$$\mathcal{H}_s = \mathcal{H}_f + \mathcal{H}_f^* - \mathcal{H}_f^*\mathcal{H}_f. \tag{16.9.19}$$

There is at least one difference, however, in that the present derivation holds under much weaker conditions than the differentiability assumption we made on $K(\cdot,\cdot)$ to get the Siegert-Krein-Bellman identity (16.9.12).

We may note that the factorization identities (16.9.17)–(16.9.19) that we obtained using the innovations process (and ultimately therefore using the theory of martingale stochastic processes) were first obtained by Gohberg and Krein (1964,1970) using some abstract concepts such as factorization with respect to chains of projections (which are analogous to increasing families of sigma fields). The Gohberg-Krein arguments can be applied to generalize the innovations theory in some interesting ways — see Kailath and Duttweiler (1972).

We may note also that by assuming state-space structure for the $K(\cdot,\cdot)$, we can get the standard Riccati, Chandrasekhar-Kailath, and Bryson-Frazier equations. More on this below. However, while on the topic of smoothing, let us note another useful formula that brings out an important simplification available in the stationary case. [In fact, this result was among several others (see Kailath, Vieira, and Morf (1978b) and Kailath (1982)) leading to the concept of displacement structure, which enabled significant extensions of the following result.]

Sobolev's Identity and Time-Invariant Smoothing. Consider again the model

$$\mathbf{y}(t) = \mathbf{z}(t) + \mathbf{v}(t), \quad 0 \le t \le T < \infty,$$

but now with the stationarity assumption

$$\langle \mathbf{z}(t), \mathbf{z}(s)\rangle = K(t-s), \quad \langle \mathbf{z}(t), \mathbf{v}(s)\rangle = 0, \quad \langle \mathbf{v}(t), \mathbf{v}(s)\rangle = I\delta(t-s).$$

[4] This is the appropriate continuous-time analog of the discrete-time formula $R_y = LR_eL^*$. Without the identity operators, the analogous continuous-time factorization is hard to make meaningful.

Although the processes are all stationary, the filters $H_s(\cdot,\cdot;T)$ and $H_f(\cdot,\cdot)$ are time-variant because $T < \infty$. However, at least for the smoothing filter, we can obtain a time-invariant implementation by exploiting the stationarity, which in particular means that

$$K(t,s) = K(T-t, T-s),$$

and therefore that

$$H_s(t,s;T) = H_s(T-t, T-s; T).$$

We can use this fact to show, following Sobolev (1963), that we can write

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) H_s(t,s;T) = A(T,t)A(T,s) - B(T,t)B(T,s), \tag{16.9.20}$$

where

$$A(T,t) = H_s(T,t;T) = H_f(T,t),$$
$$B(T,t) = H_s(t,0;T) = H_f(T,T-t).$$

Interpretation of (16.9.20) then gives (see Levy et al. (1979))

$$H_s(t,s;T) = B(T,t-s) + B(T,s-t) + \int_0^{\min(t,s)} [A(T,t-r)A(T,s-r) - B(T,t-r)B(T,s-r)]\,dr.$$

The point is that the smoothed estimator

$$\hat{\mathbf{z}}(t|T) = \int_0^T H_s(t,s;T)\mathbf{y}(s)ds,$$

can be computed using only time-invariant filters. The paper by Levy et al. (1979) gives more details. Extensions can be made to nonstationary processes with low displacement rank — see App. I of Kailath (1981). We should remark that the discrete-time analog of the Sobolev identity is the famous Gohberg-Semencul identity for the inverse of a Toeplitz matrix (see, *e.g.,* Gohberg and Semencul (1972)). A more detailed discussion of the relationships, via the innovations and orthogonal polynomials, is given in Kailath, Vieira, and Morf (1978b).

Riccati Equations, Wiener-Hopf Equations, and Fast Algorithms. The widespread use of Kalman's solutions of the state-space estimation and control problems led to a great activity in studying the nonlinear Riccati equation. Therefore, not much attention was paid to the underlying Wiener-Hopf equation, despite its use in obtaining the continuous-time equations. One reason was that the Wiener-Hopf equations were formulated in terms of covariance functions, which had to be obtained by additional calculations on the given state-space model. The gap between these two descriptions was partially closed by Thm. 16.2.3, which gives recursive estimators based on covariance data and therefore specifies the solution of the Wiener-Hopf equation (*cf.* Remark 7). This result, first presented in Kailath and Geesey (1971), was timely because in a note

by Casti, Kalaba, and Murthy (1972), attention was drawn to a different method for solving Wiener-Hopf equations of the form

$$h(t,s) + \int_0^t h(t,\tau)A(\tau - s)d\tau = A(t-s), \quad 0 \le s \le t, \tag{16.9.21}$$

where

$$A(t) = \int_0^1 e^{-\alpha|t|} w(\alpha)d\alpha. \tag{16.9.22}$$

The solution was expressed in terms of a pair of auxiliary functions obeying the simultaneous nonlinear partial differential equations

$$\frac{\partial X(t,\alpha)}{\partial t} = -Y(t,\alpha)\int_0^1 Y(t,\beta)w(\beta)d\beta, \tag{16.9.23}$$

$$\frac{\partial Y(t,\alpha)}{\partial t} = -\alpha Y(t,\alpha) - X(t,\alpha)\int_0^1 Y(t,\beta)w(\beta)d\beta, \tag{16.9.24}$$

with initial conditions

$$X(0,\alpha) = 1 = Y(0,\alpha), \quad 0 \le \alpha \le 1. \tag{16.9.25}$$

The stimulus of a 1972 seminar by Casti on these results led Kailath to explore the possibility of similar results for stationary processes with state-space models. This led first to the results described in Ex. 16.6.2, where the $\{K, L\}$ equations can be recognized as being related to the $\{X, Y\}$ equations given above with

$$w(\alpha) = \sum_{i=1}^n \alpha_i \delta(\alpha - \alpha_i), \quad \alpha_i \ge 0,$$

and

$$F = -\text{diag}\{\alpha_1, \alpha_2, \ldots, \alpha_n\}, \quad R = I.$$

Further reading led to the fact that the $\{X, Y\}$ equations first arose in radiative transfer theory. Here, the special form of the kernel $A(\cdot)$ arose from a model of the propagation of light through the earth's atmosphere: $w(\alpha)$ was the intensity of light impinging on the atmosphere from a direction α, and $e^{-\alpha t}$ represents the attenuation at depth t along that direction.

For such problems, astrophysicists had found the Wiener-Hopf technique to be elegant, but computationally difficult. So it was a big step when in 1943 a famous Soviet astronomer, V. A. Ambartzumian, showed that the Wiener-Hopf equation could be solved by reduction to an integro-differential equation of Riccati-type (see, *e.g.,* Chandrasekhar (1950, Sec. 51, Eq. (30))):

$$\frac{\partial}{\partial \tau}P(t,\alpha,\beta) = Q + \int_0^1 P(t,\alpha,\beta')w(\beta')d\beta' + \int_0^1 P(t,\alpha',\beta)w(\alpha')d\alpha' + \int_0^1\int_0^1 P(t,\alpha',\beta)w(\alpha')w(\beta')P(t,\alpha,\beta')d\alpha'd\beta',$$

with given initial conditions. Ambartzumian's derivation used an invariance principle long familiar to electrical engineers: the input impedance of a medium (the atmosphere, in his case) of infinite depth is the same even after a small layer has been removed.

Later S. Chandrasekhar further exploited and extended these invariance principles to solve the Wiener-Hopf-type equation for a finite atmosphere. He introduced the notation $X(\cdot)$ and $Y(\cdot)$ functions and showed that the the solution could be reduced to the solution of the coupled set of nonlinear integro-differential equations given above (see Chandrasekhar (1950, Sec. 56))

The point then was that solving for two functions of two variables was simpler than solving for one function $P(t, \alpha, \beta)$ of three variables. In fact, the simplifications were enough that Chandrasekhar and others were able to obtain important numerical results using just the clumsy hand-computing machines of the 1940s (see, *e.g.,* Chandrasekhar (1948, p. 214) and Sobolev (1963, p. 169)). [As an aside, we may note that Chandrasekhar, who was to win a Nobel Prize many years later, derived these results in Part XXII (1948) of a long sequence of long papers; Part XXII had 480 equations. This style resulted in an affectionate spoof "On the imperturbability of elevator operators: LVII," by S. Candlestickmaker, Institute for Studied Advances, Old Cardigan, Wales. It had the obligatory acknowledgement to the "computers" of the day: "I wish to record my indebtedness to Miss Canna Helpit, who carried out the laborious numerical work."]

The above equations for X and Y attracted considerable attention in transport theory and related fields (see, *e.g.,* Case (1957) and Noble (1964)). Under the name "invariant embedding", it was vigorously pursued by Bellman and his colleagues, see, *e.g.,* Kagiwada and Kalaba (1966), Bellman and Denman (1971), and Bellman and Wing (1975).

But the most complete and elegant development of the Ambartsumian-Chandrasekhar ideas was made by Redheffer (see Redheffer (1962) and the many references therein to his earlier work and that of others), who also made many connections to the work of electrical engineers on transmission lines. We shall describe and exploit Redheffer's work in some detail in Ch. 17.

The approach used by Kailath to obtain the $\{X, Y\}$ equations (based on deriving and factoring a Stokes identity for the Riccati variable) led to extensions beyond the stationary case, as described in Sec. 16.6, and then to the analogous (but more complicated) discrete-time results described in Chs. 11 and 13. These results can provide dramatic simplifications when n is large, as in image processing problems and in distributed parameter systems; we may refer here to Casti and Kirschner (1966) , Sorine (1977), and Burns and Powers (1986).

■ PROBLEMS

16.1 **(A smoothing and filtering problem)** Let $\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t)$ for $t \geq 0$, where $\mathbf{v}(t)$ is a white-noise process with unit variance, $\dot{\mathbf{x}}(t) = \alpha\mathbf{x}(t)$, and $\mathbf{x}_0$ is a Gaussian random variable with zero mean and variance P_0.

(a) Show that

$$P(t|t) = \|\mathbf{x}(t) - \hat{\mathbf{x}}(t|t)\|^2 = \frac{e^{2\alpha t}P_0}{1 + \frac{P_0}{2\alpha}(e^{2\alpha t} - 1)},$$

and examine the limiting behavior as $t \to \infty$ for both $\alpha > 0$ and $\alpha < 0$.

(b) Compute $P(s|t) = \|\mathbf{x}(s) - \hat{\mathbf{x}}(s|t)\|^2$, and comment on the behavior as $t \to \infty$ for both $\alpha > 0$ and $\alpha < 0$.

(c) Compare the results of parts (a) and (b).

16.2 **(A linearizable state-space model)** Consider the system

$$\begin{cases} \dot{\mathcal{E}}(t) = a\mathcal{E}(t), & \mathcal{E}(0) = \mathcal{E}_0 \\ \mathbf{y}(t) = \sum_{k=1}^{N} h_k \mathcal{E}^k(t) + \mathbf{v}(t) \end{cases}$$

where $\mathcal{E}(t)$ is a scalar and $\mathcal{E}^k(t)$ is the k-th power of $\mathcal{E}(t)$.

(a) Show how to choose N states so as to obtain an equivalent N-state linear system

$$\begin{cases} \dot{\mathbf{x}}(t) = F\mathbf{x}(t), & \mathbf{x}(0) = \mathbf{x}_0, \quad \|\mathbf{x}_0\|^2 = \Pi_0, \\ \mathbf{y}(t) = H\mathbf{x}(t) + \mathbf{v}(t), \end{cases}$$

(b) Show that the error covariance matrix is given by

$$P(t) = e^{Ft}\left[\Pi_0^{-1} + A\right]^{-1} e^{F^*t}$$

where A is a matrix with (i, j)-th element

$$\frac{h_i h_j}{a(i + j)}\left[e^{a(i+j)t} - 1\right].$$

16.3 **(An alternative model)** Consider the state equations

$$\dot{\mathbf{x}}(t) = F(t)\mathbf{x}(t) + G(t)\mathbf{u}(t), \qquad \mathbf{y}(t) = \int_0^t \mathbf{x}(\tau)d\tau + \mathbf{v}(t), \qquad t \geq 0,$$

where $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$ are uncorrelated white-noise processes with intensities $\{Q(\cdot), R(\cdot)\}$; both of which are uncorrelated with $\mathbf{x}(0)$, whose variance we denote by Π_0. Show that

$$\dot{\hat{\mathbf{x}}}(t) = F(t)\hat{\mathbf{x}}(t) + K_1(t)[\mathbf{y}(t) - \hat{\mathbf{z}}(t)], \qquad \hat{\mathbf{x}}(0) = 0,$$

where

$$\dot{\hat{\mathbf{z}}}(t) = \hat{\mathbf{x}}(t) + K_2(t)[\mathbf{y}(t) - \hat{\mathbf{z}}(t)], \qquad \hat{\mathbf{z}}(0) = 0,$$

and

$$\begin{aligned} \dot{K}_1(t) &= F(t)K_1(t) + P(t) - K_1(t)K_2(t) && K_1(0) = 0, \\ \dot{K}_2(t) &= K_1(t) + K_1^*(t) - K_2^*(t) && K_2(0) = 0, \\ \dot{P}(t) &= F(t)K_1(t) + K_1(t)F^*(t) - K_1(t)K_1^*(t) + G(t)Q(t)G^*(t) && P(0) = \Pi_0. \end{aligned}$$

16.4 **(Zero estimation error)** Consider the state-space model

$$\dot{\mathbf{x}}(t) = F(t)\mathbf{x}(t) + G(t)\mathbf{v}(t), \qquad \mathbf{y}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t), \quad t \geq 0,$$

where $\mathbf{x}(0) = 0$, $E\mathbf{v}(t) = 0$, $\langle \mathbf{v}(t), \mathbf{v}(\tau) \rangle = R(t)\delta(t-\tau)$.

(a) Show $P(t) = \|\tilde{\mathbf{x}}(t)\|^2 = 0$. Is this obvious? Find the innovations.

(b) How is the result changed if $E\mathbf{x}(0) = 0$, $\langle \mathbf{x}(0), \mathbf{v}(t) \rangle = 0$, $\|\mathbf{x}(0)\|^2 = \Pi_0$?

16.5 **(Estimating a derivative)** Consider the standard state-space model (16.1.2)–(16.1.3). Let $\mathbf{z}(t) = H(t)\mathbf{x}(t)$.

(a) Find an expression for the estimation of the derivative $\dot{\mathbf{z}}(t)$ of $\mathbf{z}(t)$ in terms of the Kalman filter.

(b) Obtain an expression for the m.m.s.e. in estimating $\dot{\mathbf{z}}(t)$. Interpret your result.

16.6 **(A nonminimum phase plant)** Consider the system

$$\dot{\mathbf{x}}_1(t) = \mathbf{u}(t), \quad \dot{\mathbf{x}}_2(t) = -\mathbf{x}_1(t) - 2\mathbf{x}_2(t) + \mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{x}_2(t) + \mathbf{v}(t),$$

where $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$ are uncorrelated zero-mean white-noise processes with intensities q and r, respectively.

(a) Assuming zero initial conditions, show that the transfer function from $\{\mathbf{u}(\cdot)\}$ and $\{\mathbf{v}(\cdot)\}$ to $\{\mathbf{y}(\cdot)\}$ is given by, in terms of Laplace transforms,

$$\mathbf{y}(s) = \frac{s-1}{s(s+2)}\,\mathbf{u}(s) + \mathbf{v}(s),$$

which has a right half-plane (nonminimum phase) zero at $s = 1$.

(b) Using the steady-state Riccati equation, show that $P_{11}(t) \to 2q$, $P_{12}(t) \to -\sqrt{qr}$, $P_{22}(t) \to \sqrt{qr}$ as r/q tends to zero. Thus, even for $r = 0$ (a noise-free measurement), $\mathbf{x}_1$ is not estimated perfectly. [Here, $P_{ij}(t)$ denotes the (i, j)-th entry of the 2×2 matrix $P(t)$.]

Remark. The estimator accuracy and the estimator-error bandwidth are both limited by the location of right half-plane zeros. ◆

16.7 **(Whiteness of the innovations)** Assume knowledge of the Kalman filter equations in Thm. 16.2.1. Use them to show by direct calculation that $\langle \mathbf{e}(t), \mathbf{e}(s) \rangle = R(t)\delta(t-s)$. [*Hint.* See Collins (1968).]

16.8 **(Steady-state filters)** A zero-mean white-noise stationary process $\{\mathbf{u}(\cdot)\}$ with intensity σ^2 is applied to the system $H(s) = 1/s$. We denote the output by $\{\mathbf{x}(\cdot)\}$. Noisy measurements of $\mathbf{x}(\cdot)$ are available, say $\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t)$, where $\{\mathbf{v}(\cdot)\}$ is a zero-mean white-noise stationary process with intensity $N_0/2$ and uncorrelated with $\{\mathbf{u}(\cdot)\}$.

(a) By *formal* use of the Wiener formulas, find the transfer function of the optimum linear filter for estimating $\mathbf{x}(t)$ given $\{\mathbf{y}(\tau), -\infty < \tau < t\}$.

(b) Repeat part (a) for estimating $\dot{\mathbf{x}}(t)$ given the same observations.

(c) Write down the Kalman filter equations for estimating $\mathbf{x}(t)$ from finite time data $\{\mathbf{y}(\tau), 0 < \tau < t\}$. Find also the limiting solution as $t \to \infty$ and compare with the result of part (a).

16.9 **(Estimating an integral)** Consider the model $\mathbf{y}(t) = \mathbf{s}(t) + \mathbf{v}(t)$ for $t \geq 0$, where $\mathbf{v}(\cdot)$ is a zero-mean scalar white-noise process with variance r, and the desired signal is given by

$$\mathbf{s}(t) = \int_0^t \mathbf{u}(\tau)d\tau,$$

where $\mathbf{u}(\cdot)$ is a zero-mean scalar white-noise process, uncorrelated with $\mathbf{v}(\cdot)$, that has variance q.

(a) Find a differential equation (with initial condition) for $\hat{\mathbf{s}}(t)$, the l.l.m.s.e. of $\mathbf{s}(t)$ given $\{\mathbf{y}(\tau), 0 \leq \tau < t\}$. *Hint.* You may want to use the formula

$$\int \frac{dx}{a^2 - x^2} = \frac{1}{2a}\log\frac{a+x}{a-x} + C.$$

(b) What is the corresponding m.m.s.e.?

(c) Show that the steady-state m.m.s.e. is $\sqrt{qr}$.

(d) Show that in steady-state we may write

$$\hat{\mathbf{s}}(t) = \int_{-\infty}^{t} \sqrt{\frac{q}{r}}\, e^{-\sqrt{\frac{q}{r}}(t-\tau)}\, \mathbf{y}(\tau)d\tau.$$

16.10 **(A hybrid system)** Consider the standard state-space model (16.1.2)–(16.1.3). However, the observations process $\mathbf{y}(\cdot)$ is only measured at the discrete times $0 < t_1 < t_2 < t_3 < \dots$ by using some physical sampling device. Let us model the discrete-time measurements as

$$\mathbf{y}(t_k) = \frac{1}{\epsilon}\int_{t_k - \frac{\epsilon}{2}}^{t_k + \frac{\epsilon}{2}} [H(t)\mathbf{x}(t) + \mathbf{v}(t)]dt = H(t_k)\mathbf{x}(t_k) + \mathbf{w}(t_k), \quad k = 1, 2, \dots,$$

where we assume the changes in $\mathbf{x}(\cdot)$ and $H(\cdot)$ over the interval $\left[t_k - \frac{\epsilon}{2}, t_k + \frac{\epsilon}{2}\right]$ are negligible.

(a) Define the discrete-time process $\mathbf{w}_k \triangleq \mathbf{w}(t_k)$. What are the statistics of $\{\mathbf{w}_k\}$?

(b) Find the l.l.m.s. estimator of $\mathbf{x}(t)$ for all t (not just $t_1, t_2, t_3, \dots$) given the $\{\mathbf{y}(t_k), t_k \leq t\}$. [*Hint.* Find the time-update and measurement-update separately.]

16.11 **(Fredholm integral equations)** Consider a zero-mean stationary process $\mathbf{x}(\cdot)$ with autocorrelation function $R_x(\tau) = e^{-\alpha|\tau|}$.

(a) Show that the l.l.m.s. estimator of $\mathbf{x}(t+\lambda)$, $\lambda > 0$, given $\{\mathbf{x}(\tau), 0 \leq \tau \leq t\}$, can be written as

$$\hat{\mathbf{x}}(t+\lambda|t) = \int_0^t h(t, \tau)\mathbf{x}(\tau)d\tau,$$

where

$$\int_0^t h(t, \tau)e^{-\alpha|\tau-\sigma|}d\tau = e^{-\alpha(t+\lambda-\sigma)}, \quad 0 \leq \sigma \leq t.$$

(b) Show that $h(t, \tau) = e^{-\alpha\lambda}\delta(t-\tau)$ solves the above equation and that, as shown earlier in Ex. 3.3.2, $\hat{\mathbf{x}}(t+\lambda|t) = e^{-\alpha\lambda}\mathbf{x}(t)$.

Remark. Integral equations such as the one in part (a) are known as Fredholm integral equations of the *first* kind and are notoriously difficult to solve, especially by numerical methods. The point is that because of the smoothing effect of integration, the space of possible solutions $h(t, \tau)$ is much larger than the space of given functions $e^{-\alpha|t|}$. In our case, the given functions are continuous but the solution is impulsive; therefore discretization schemes for solving the equation may not find the impulse. In other examples, $h(t, \tau)$ can contain derivatives of impulses, in which case numerical and implementation problems can get much worse. On the other hand, if the observed process contains pure white noise, say $\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t)$, with $E\mathbf{v}(t)\mathbf{x}^*(s) = 0$, $E\mathbf{v}(t)\mathbf{v}^*(s) = R(t)\delta(t - s)$, then the relevant integral equation for estimating $\mathbf{x}(t + \lambda)$ from $\{\mathbf{y}(\tau), 0 \le \tau \le t\}$ is of the form

$$h(t, \sigma) + \int_0^t h(t, \tau)R_x(\tau - \sigma)d\tau \;=\; R_x(t + \lambda - \sigma), \quad 0 \le \sigma \le t.$$

In these so-called Fredholm equations of the *second* kind, the unknown function $h(t, \cdot)$ must have the same degree of smoothness as the known functions $R_x(\cdot)$ — making numerical solutions much easier. Also, since the unknown function is smooth (no impulses), implementation is easier, the sensitivity/robustness is better, and so on. Therefore, in continuous-time problems, the assumption of additive pure white noise is both physically and mathematically important. ◆

16.12 **(A special autocorrelation function)** Find the l.l.m.s.e. of $\mathbf{y}(t+\lambda)$, $\lambda > 0$, given $\{\mathbf{y}(s), -\infty < s \le t\}$, where the autocorrelation function of the random process $\{\mathbf{y}(\cdot)\}$ is

$$R_y(t, \tau) = \langle \mathbf{y}(t), \mathbf{y}(\tau) \rangle \;=\; p(t)p(\tau)e^{-\alpha|t-\tau|},$$

where $p(\cdot)$ is a known function such that $0 < p(t) < 1$ for all t.

16.13 **(Changes in noise variances)** In this problem we assume that the process and measurement noise processes are uncorrelated, *i.e.*, $S(t) = 0$.

(a) Let $P(t)$ and $P_Q(t)$ be the solutions of the standard Riccati differential equation with $\{\Pi_0, R(t), Q(t)\}$ and $\{\Pi_0, R(t), Q(t) + \delta Q(t)\}$, respectively. Show that $\delta Q(t) \ge 0$ implies that $Q(t) \ge P(t)$ for $t \ge 0$.

(b) Show that $R(t) > 0$ and $\delta R(t) \ge 0$ implies that $R^{-1}(t) - (R(t) + \delta R(t))^{-1} \ge 0$. *Hint.* Consider the matrix

$$\begin{bmatrix} R(t) + \delta R(t) & I \\ I & R^{-1}(t) \end{bmatrix}.$$

(c) Now suppose $P_R(t)$ is the solution of the standard Riccati differential equation with $\{\Pi_0, R(t) + \delta R(t), Q(t)\}$. Use the result of part (b) to show that $\delta R(t) \ge 0$ implies that $P_R(t) \ge P(t)$ for $t \ge 0$.

16.14 **(Adjoint systems)** Let $\Phi(t, t_0)$ be the state-transition matrix associated with the equation $\dot{x}(t) = F(t)x(t)$.

(a) Show that

$$\frac{d}{dt}\Phi(t_0, t) = -\Phi(t_0, t)F(t),$$

while

$$\frac{d\Phi^*(t_0, t)}{dt} \;=\; -F^*(t)\Phi^*(t_0, t), \quad \Phi^*(t_0, t_0) = I.$$

(b) If $p(t)$ satisfies the so-called adjoint system $\dot{p}(t) = -F^*(t)p(t)$, check that the inner product $p^*(t)x(t)$ is independent of time.

16.15 **(Use of the adjoint system)** Consider the state-space model

$$\dot{\mathbf{x}}(t) = F(t)\mathbf{x}(t) + G(t)\mathbf{u}(t), \quad \mathbf{y}(t) = H(t)\mathbf{x}(t), \quad t \ge 0,$$

with $\{\mathbf{x}(0), \mathbf{u}(t)\}$ uncorrelated zero-mean random variables with variances Π_0 and $Q(t)$, respectively. Moreover, $\{\mathbf{u}(\cdot)\}$ is a white noise process and $\mathbf{y}(t)$ is scalar valued (hence, $H(t)$ is a row vector).

(a) Show that

$$E|\mathbf{y}(t)|^2 = \langle \mathbf{y}(t), \mathbf{y}(t) \rangle = H(t)\Pi(t)H^*(t),$$

where

$$\Pi(t) = \Phi(t, 0)\Pi_0\Phi^*(t, 0) + \int_0^t \Phi(t, \tau)G(\tau)Q(\tau)G^*(\tau)\Phi^*(t, \tau)d\tau,$$

where $\Phi(\cdot, \cdot)$ is the transition matrix associated with $F(\cdot)$.

(b) Show that an alternative formula is

$$E|\mathbf{y}(t)|^2 = \mathbf{p}^*(0)\Pi_0\mathbf{p}(0) + \int_0^t \mathbf{p}^*(\tau)G(\tau)Q(\tau)G^*(\tau)\mathbf{p}(\tau)d\tau,$$

where $\dot{\mathbf{p}}(\tau) = -F^*(\tau)\mathbf{p}(\tau)$ with boundary condition $\mathbf{p}(t) = H^*(t)$ and for $0 \le \tau \le t$.

(c) What is the significance of the second solution?

16.16 **(A linear matrix equation)** If

$$\frac{d}{dt}X(t) = A(t)X(t) + X(t)B(t) + U(t), \quad X(t_0) = X_0,$$

check that we can write

$$X(t) = \Phi_A(t, t_0)X_0\Phi_{B^*}^*(t, t_0) \;+\; \int_{t_0}^t \Phi_A(t, \sigma)U(\sigma)\Phi_{B^*}^*(t, \sigma)d\sigma,$$

where $\Phi_A(\cdot, t_0)$ is the state-transition matrix of $\dot{X}(t) = A(t)X(t)$ and $\Phi_{B^*}(\cdot, t_0)$ is the state-transition matrix of $\dot{X}(t) = B^*(t)X(t)$.

16.17 **(Homogeneous RDE with nonsingular Π_0)** Consider the homogeneous Riccati differential equation

$$\dot{P}(t) = F(t)P(t) + P(t)F^*(t) - P(t)H^*(t)R^{-1}(t)H(t)P(t), \quad P(t_0) = \Pi_0,$$

with nonsingular initial condition, $\Pi_0 > 0$.

(a) Show first that

$$\frac{d}{dt}P^{-1}(t) = -P^{-1}(t)\dot{P}(t)P^{-1}(t),$$

and conclude that $P^{-1}(t)$ satisfies the linear differential equation

$$\frac{dP^{-1}(t)}{dt} = -P^{-1}F(t) - F^*(t)P^{-1}(t) + H^*(t)R^{-1}(t)H(t), \quad P^{-1}(t_0) = \Pi_0^{-1}.$$

(b) Use the results of Probs. 16.14 and 16.16, and the property $\Phi(\tau, t_0)\Phi(t_0, t) = \Phi(\tau, t)$, to show that

$$P^{-1}(t) = \Phi^*(t_0, t)[\Pi_0^{-1} + \mathcal{O}(t, t_0)]\Phi(t_0, t),$$

where $\mathcal{O}(t, t_0)$ is the observability Gramian matrix

$$\mathcal{O}(t, t_0) = \int_{t_0}^{t} \Phi^*(\tau, t_0)H^*(\tau)R^{-1}(\tau)H(\tau)\Phi(\tau, t_0)d\tau,$$

and $\Phi(t, t_0)$ is the state-transition matrix of $F(\cdot)$.

(c) How could one modify the formula for $P^{-1}(t)$ when Π_0 may be singular?

16.18 **(Nonsingular $P(t)$)** Consider the standard state-space model (16.1.2)–(16.1.3). Assume $\Pi_0 > 0$. Show that $P(t) > 0$ for all $t \geq 0$.

16.19 **(Direct derivation of (16.1.12))** Instead of the approach via discretization, we shall try to obtain the differential equation (16.1.12) for $\Pi(t) = \|\mathbf{x}(t)\|^2$ by writing $\dot{\Pi}(t)$ as $\langle \dot{\mathbf{x}}(t), \mathbf{x}(t)\rangle + \langle \mathbf{x}(t), \dot{\mathbf{x}}(t)\rangle$.

(a) Using the representation

$$\mathbf{x}(t) = \Phi(t, 0)\mathbf{x}(0) + \int_0^t \Phi(t, \tau)G(\tau)\mathbf{u}(\tau)d\tau,$$

show that

$$\langle \mathbf{x}(t), \mathbf{u}(t)\rangle = \int_0^t \Phi(t, \tau)G(\tau)Q(t)\delta(\tau - t)d\tau.$$

(b) Note that the impulse function in part (a) occurs at the end of the integration interval. Hence, the argument to evaluate the integral depends on how the impulse function is defined. When a symmetric pulse is used to define $\delta(t)$, half of its area will be to the left of $t = 0$ and the other half to the right of $t = 0$. More generally, assume a nonsymmetric definition such that

$$\int_0^\infty \delta(t)dt = 1 - a, \quad \int_{-\infty}^0 \delta(t)dt = a, \quad \text{for some } 0 \leq a \leq 1.$$

Use these relations to verify that

$$\langle \mathbf{x}(t), \mathbf{u}(t)\rangle = aG(t)Q(t) \quad \text{and} \quad \langle \mathbf{u}(t), \mathbf{x}(t)\rangle = (1 - a)Q(t)G^*(t).$$

[The "nonsymmetric" delta function is not consistent with the inner product property: $\langle \mathbf{x}(t), \mathbf{u}(t)\rangle = \langle \mathbf{u}(t), \mathbf{x}(t)\rangle^*$. So $a = 1/2$ is a desirable (but not essential) choice.]

(c) Conclude that $G(t)\langle \mathbf{u}(t), \mathbf{x}(t)\rangle + \langle \mathbf{x}(t), \mathbf{u}(t)\rangle G^*(t) = G(t)Q(t)G^*(t)$ and, therefore,

$$\dot{\Pi}(t) = F(t)\Pi(t) + \Pi(t)F^*(t) + G(t)Q(t)G^*(t), \quad \Pi(0) = \Pi_0.$$

Remark. Such arguments with delta functions can be avoided if one uses the Ito stochastic calculus. An elementary exposition can be found in Jazwinski (1970). ◆

16.20 **(Kalman filter with $S(t) \neq 0$)** We wish to extend the arguments of Sec. 16.4.2 to the case $S(\cdot) \neq 0$. Thus note from (16.4.12) and (16.4.14) that we need to evaluate $\langle \mathbf{x}(t), \mathbf{e}(t)\rangle$ and $\langle \mathbf{u}(t), \mathbf{e}(s)\rangle$ for $0 \leq t \leq s$.

(a) Verify that

$$\langle \mathbf{x}(t), \mathbf{e}(t)\rangle = P(t)H^*(t) + \int_0^t \Phi(t, \tau)G(\tau)S(\tau)\delta(\tau - t)d\tau,$$

and that

$$\langle \mathbf{u}(t), \mathbf{e}(s)\rangle = \langle \mathbf{u}(t), \tilde{\mathbf{x}}(s)\rangle H^*(s) + S(t)\delta(t - s).$$

(b) Verify further that

$$\langle \mathbf{u}(t), \tilde{\mathbf{x}}(s)\rangle = \langle \mathbf{u}(t), \mathbf{x}(s)\rangle - \langle \mathbf{u}(t), \hat{\mathbf{x}}(s)\rangle,$$

and argue that $\langle \mathbf{u}(t), \mathbf{x}(s)\rangle$ and $\langle \mathbf{u}(t), \hat{\mathbf{x}}(s)\rangle$ are zero everywhere except when $s = t$, where they are finite.

(c) Conclude that

$$\dot{\hat{\mathbf{x}}}(t) = F(t)\hat{\mathbf{x}}(t) + K(t)\mathbf{e}(t), \quad K(t) = P(t)H^*(t) + G(t)S(t).$$

(d) An alternative approach when $S(t) \neq 0$ is to replace $Q(t)$ by $Q^s(t) = Q(t) - S(t)R^{-1}(t)S^*(t)$ and $F(t)$ by $F^s(t) = F(t) - G(t)S(t)R^{-1}(t)H(t)$. Show that in this case, the state-space model (16.1.1)–(16.1.3) is replaced by the equivalent model

$$\dot{\mathbf{x}}(t) = F^s(t)\mathbf{x}(t) + G(t)S(t)R^{-1}(t)\mathbf{y}(t) + G(t)\mathbf{u}^s(t),$$

$$\mathbf{y}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t), \qquad t \geq 0,$$

where $\{\mathbf{u}^s(\cdot), \mathbf{v}(\cdot)\}$ are white-noise processes such that

$$\left\langle \begin{bmatrix} \mathbf{u}^s(t) \\ \mathbf{v}(t) \\ \mathbf{x}(0) \end{bmatrix}, \begin{bmatrix} \mathbf{u}^s(s) \\ \mathbf{v}(s) \\ \mathbf{x}(0) \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q^s(t)\delta(t - s) & 0 & 0 & 0 \\ 0 & R(t)\delta(t - s) & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}.$$

Use this state-space model transformation to rederive the Kalman filter recursions for the general case $S(t) \neq 0$.

16.21 (Square-root algorithm) Consider the Riccati differential equation (16.4.16) and assume $P(t) > 0$ for all t. Let $M(t)$ denote a square-root factor of $P(t)$, *i.e.*, $P(t) = M(t)M^*(t)$.

(a) Show that $M(t)$ satisfies the differential equation (where we are dropping the argument t for convenience of notation)

$$\dot{M}M^* + M\dot{M}^* = \left(F - \frac{1}{2}MM^*H^*R^{-1}H\right)MM^* + \frac{1}{2}GQG^*M^{-*}M^* + $$

$$+MM^*\left(F - \frac{1}{2}MM^*H^*R^{-1}H\right)^* + \frac{1}{2}MM^{-1}GQG^*,$$

with initial condition $M(0) = \Pi_0^{1/2}$.

(b) Verify that a particular solution of the above differential equation for $M(t)$ is the solution $N(t)$ of the following differential equation:

$$\dot{N} = \left(F - \frac{1}{2}NN^*H^*R^{-1}H\right)N + \frac{1}{2}GQG^*N^{-*},$$

with initial condition $N(0) = \Pi_0^{1/2}$.

(c) Let $\Theta(t)$ be the solution of the differential equation

$$\dot{\Theta}(t) = A(t)\Theta(t), \quad \Theta(0) = \Theta_0, \quad \Theta_0\Theta_0^* = I,$$

for any matrix $A(t)$ that is skew-symmetric, *i.e.*, $A(t) + A^*(t) = 0$. Show that $\Theta(t)\Theta^*(t) = I$ for all $t \geq 0$.

Remark. We say that $\Theta(t)$ is a unitary matrix function generated by $A(t)$. In fact, any differentiable unitary matrix function can be generated in this way.] ◆

(d) Square-root factors are also highly nonunique in continuous time. Indeed, if $N(t)$ is one square-root factor, then $N(t)\Theta(t)$ is also a square-root factor for any unitary matrix function $\Theta(t)$. Thus let $M(t) = N(t)\Theta(t)$ denote any such square-root factor for $P(t)$. Verify that $M(t)$ satisfies the differential equation

$$\dot{M} = \left(F - \frac{1}{2}MM^*H^*R^{-1}H\right)M + \left(B + \frac{1}{2}GQG^*\right)M^{-*},$$

where $S(0) = \Pi_0^{1/2}\Theta_0$ and $B(t) = N(t)A(t)N^*(t)$. Verify further that $B(t)$ is skew-symmetric.

Remark. Therefore, for every choice of a skew-symmetric matrix function $A(t)$, and for every initial condition Θ_0, we can determine a square-root factor $M(t)$ for $P(t)$ by solving the above differential equation. In general, the factor $M(t)$ need not be triangular. ◆

16.22 (Triangular square-root factors) Consider the setting of Prob. 16.21, where now we would like to determine a lower-triangular factor $M(t)$. Define

$$Y(t) \triangleq M^{-1}(t)\dot{M}(t) + \dot{M}^*(t)M^{-*}(t).$$

(a) Use the differential equation of part (a) of Prob. 16.21 to show that (where we are again dropping the argument t for convenience of notation)

$$Y = \bar{F} + \bar{F}^* + \bar{G}Q\bar{G}^* - \bar{H}R^{-1}\bar{H}^*,$$

where

$$\bar{F} = M^{-1}FM, \quad \bar{G} = M^{-1}G, \quad \bar{H} = HM.$$

(b) For any matrix C, let the operator $[C]_{+/2}$ denote the lower triangular part of C with its diagonal entries further scaled by $1/2$, *i.e.*,

$$(i,j) - \text{th entry of } [C]_{+/2} = \begin{cases} 0 & \text{if } i < j \\ \frac{1}{2}C_{ii} & \text{if } i = j \\ C_{ij} & \text{if } i > j \end{cases}$$

Now using the fact that $M^{-1}(t)\dot{M}(t)$ is lower triangular, show that $M(t)$ satisfies the differential equation

$$\dot{M} = M[Y]_{+/2}, \quad M(0) = \Pi_0^{1/2}.$$

(c) Show also that $\bar{F}$ obeys a so-called Lax equation

$$\dot{\bar{F}} = \bar{F}[Y]_{+/2} - [Y]_{+/2}\bar{F}.$$

(d) Show that $\dot{\bar{G}} = [Y]_{+/2}\bar{G}$ and $\dot{\bar{H}} = \bar{H}[Y]_{+/2}$.

Remark. More details on such square-root algorithms in continuous time, and extensions to the smoothing problem as well, can be found in Morf, Levy, and Kailath (1978); see also Andrews (1968), Bierman (1973b), and Tapley and Choe (1976). ◆

16.23 (Whiteness of $\mathbf{u}^b(t)$) Consider the process $\mathbf{u}^b(t)$ defined by (16.A.2), *viz.*,

$$\mathbf{u}^b(t) = \mathbf{u}(t) - Q(t)G^*(t)\Pi^{-1}(t)\mathbf{x}(t),$$

where $\mathbf{x}(t)$ satisfies (16.1.1)–(16.1.3) for $t \in [0, T]$ and $\Pi(t)$ denotes its variance, $\Pi(t) = \|\mathbf{x}(t)\|^2$. Use the expression

$$\mathbf{x}(T) = \Phi(T,t)\mathbf{x}(t) + \int_t^T \Phi(T,\tau)G(\tau)\mathbf{u}(\tau)d\tau,$$

and the results of Prob. 16.19, part (c), to establish the following results:

(a) $\langle \mathbf{x}(t), \mathbf{x}(T)\rangle = \Pi(t)\Phi^*(T,t)$.
(b) $\langle \mathbf{u}(t), \mathbf{x}(T)\rangle = Q(t)G^*(t)\Phi^*(T,t)$.
(c) $\langle \mathbf{u}^b(t), \mathbf{x}(t)\rangle = -aQ(t)G^*(t)$.
(d) $\langle \mathbf{u}^b(t), \mathbf{x}(T)\rangle = 0$.
(e) $\langle \mathbf{u}^b(t), \mathbf{u}^b(t)\rangle = Q(t)\delta(0)$.
(f) $\langle \mathbf{u}^b(t), \mathbf{u}^b(s)\rangle = 0$ for $t \neq s$.

Conclude that $\mathbf{u}^b(t)$ is white noise with $\langle \mathbf{u}^b(t), \mathbf{u}^b(s)\rangle = Q(t)\delta(t-s)$.

16.24 **(A backwards-time model)** Given a state-space model

$$\dot{\mathbf{x}}(t) = F\mathbf{x}(t) + G\mathbf{u}(t), \quad \mathbf{y}(t) = H\mathbf{x}(t), \quad 0 \le t \le T,$$

with zero-mean random variables that satisfy $\langle \mathbf{u}(t), \mathbf{x}(0) \rangle = 0$ and $\langle \mathbf{u}(t), \mathbf{u}(s) \rangle = Q(t)\delta(t-s)$, make a change of variables according to

$$\mathbf{x}(t) = \Pi(t)\boldsymbol{\xi}(t), \quad \Pi(t) = \|\mathbf{x}(t)\|^2.$$

Show that we can write

$$\dot{\boldsymbol{\xi}}(t) = -F^*\boldsymbol{\xi}(t) + \Pi^{-1}(t)G(t)\boldsymbol{\mu}(t), \quad \mathbf{y}(t) = H\Pi(t)\boldsymbol{\xi}(t),$$

where $\boldsymbol{\mu}(t) \triangleq \mathbf{u}(t) - G^*(t)Q(t)\Pi^{-1}(t)\mathbf{x}(t)$ is a white noise process, with unit intensity, and uncorrelated with $\boldsymbol{\xi}(T)$.

16.25 **(Invertibility of $\Pi(t)$)** Let $\Pi(t) = \|\mathbf{x}(t)\|^2$ denote the state-covariance matrix for the model (16.1.1). We already know from (16.4.15) and Prob. 16.19 that it satisfies

$$\frac{d\Pi(t)}{dt} = F(t)\Pi(t) + \Pi(t)F^*(t) + G(t)Q(t)G^*(t), \quad \Pi(0) = \Pi_0.$$

Use the result of Prob. 16.16 to verify that each of the following conditions is by itself sufficient to guarantee the invertibility of $\Pi(t)$ for all t:

(a) $\Pi_0 > 0$.

(b) The so-called controllability Gramian

$$\mathcal{C}(t, 0) = \int_0^t \Phi(t, \tau)G(\tau)Q(\tau)G^*(\tau)\Phi^*(t, \tau)d\tau$$

is positive-definite, where $\Phi(\cdot, \cdot)$ is the state-transition matrix of $F(\cdot)$.

16.26 **(Steady-state RTS and Wiener smoothers)** Consider a constant parameter state-space model (16.1.1)–(16.1.3) with a stable matrix F and $S = 0$. Assume the initial covariance matrix Π_0 is the unique solution of

$$F\Pi + \Pi F^* + GQG^* = 0,$$

so that the random processes $\{\mathbf{x}(\cdot), \mathbf{y}(\cdot)\}$ are zero-mean and stationary. Define $\mathbf{z}(t) = H\mathbf{x}(t)$.

(a) Determine the $s-$spectrum $S_y(s)$ and show that its canonical spectral factor is given by

$$L(s) = H(sI - F)^{-1}K + I,$$

where $K = PH^*R^{-1}$ and P is the unique positive-semi-definite steady-state solution of the CARE

$$FP + PF^* - PH^*R^{-1}HP + GQG^* = 0.$$

(b) Let $\hat{\mathbf{z}}(t|T)$ denote the smoothed estimator of $\mathbf{z}(t)$ given observations $\{\mathbf{y}(\tau), 0 \le \tau \le T\}$. Using the steady-state form of the RTS equation (16.5.11), *i.e.*, with $T \to \infty$, show that the transfer function from the observations $\mathbf{y}(\cdot)$ to the signal $\hat{\mathbf{z}}(t|\infty)$ is given by

$$W(s) = -H(sI - F - GQG^*P^{-1})^{-1}GQG^*P^{-1}(sI - F + KH)^{-1}K,$$

where P is assumed to be invertible.

(c) Let P_s denote the steady-state mean-square error of the RTS smoother. Using (16.5.12), it is given by the solution of

$$0 = F_sP_s + P_sF_s^* - GQG^*,$$

where $F_s = F + GQG^*P^{-1}$. Show that

$$HP_sP^{-1}(-sI - F + KH)^{-1}PH^* + HP(sI - F^* + H^*K^*)^{-1}P^{-1}P_sH$$
$$= HP(sI - F^* + H^*K^*)^{-1}P^{-1}GQG^*P^{-1}(-sI - F + KH)^{-1}PH^*.$$

(d) Show that the transfer function of part (b) collapses to $I - RS_y^{-1}(s)$ and is therefore equal to the (noncausal) Wiener solution, *i.e.*, $W(s) = S_z(s)S_y^{-1}(s)$.

(e) Show further, by taking the inverse Laplace transform of the identity in part (c) and by evaluating at time $t = 0$, that

$$HP_sH^* = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{HP}(j\omega\mathrm{I} - \mathrm{F}^* + \mathrm{H}^*\mathrm{K}^*)^{-1}\mathrm{P}^{-1}\mathrm{GQG}^*\mathrm{P}^{-1}(-j\omega\mathrm{I} - \mathrm{F} + \mathrm{KH})^{-1}\mathrm{PH}^*\mathrm{d}\omega,$$

where $j = \sqrt{-1}$. Conclude that the mean-square error of the RTS smoother in computing $\hat{\mathbf{z}}(t|\infty)$ is given by

$$\text{m.s.e.} = HP_sH^* = \frac{1}{2\pi}\int_{-\infty}^{\infty}\big[I - RS_y(s)\big]\,d\omega,$$

which coincides with the expression for the m.s.e. of the (noncausal) Wiener solution.

16.27 **(Steady-state smoothing via the Hamiltonian)** Consider the setting of Prob. 16.26.

(a) Use the Hamiltonian equations (16.5.14) to show that the transfer function from $\mathbf{y}(\cdot)$ to $\hat{\mathbf{z}}(t|\infty)$ is given by

$$\mathrm{W(s)} = \mathrm{H[(sI - F) + GQG^*(-sI - F^*)^{-1}H^*R^{-1}H]^{-1}GQG^*(-sI - F^*)^{-1}H^*R^{-1}}.$$

(b) Show again that $W(s) = I - RS_y^{-1}(s)$.

Remark. Contrast the derivations in Prob. 16.26 and in this problem with the simplicity of the direct solution of the Wiener smoothing problem in Ch. 7. ◆

16.28 **(Markovian property of the smoothing error)** Refer to the discussion that led to Thm. 16.5.2 on the RTS equations for smoothing, and also to the arguments in Sec. 10.3 for the discrete-time RTS equations.

(a) Starting with the state-equation $\dot{\mathbf{x}}(t) = F(t)\mathbf{x}(t) + G(t)\mathbf{u}(t)$, and using the RTS equation (16.5.11), show that the smoothing error satisfies the backwards-time differential equation

$$-\dot{\tilde{\mathbf{x}}}(t|T) = -F_s(t)\tilde{\mathbf{x}}(t|T) - G(t)\mathbf{u}^r(t),$$

where $F_s(t) = F(t) + G(t)Q(t)G^*(t)P^{-1}(t)$ and $\mathbf{u}^r(t)$ is defined by

$$\mathbf{u}^r(t) \triangleq \mathbf{u}(t) - Q(t)G^*(t)P^{-1}(t)\tilde{\mathbf{x}}(t).$$

(b) Show that $\mathbf{u}^r(t)$ is a white-noise process with variance $\langle \mathbf{u}^r(t), \mathbf{u}^r(s) \rangle = Q(t)\delta(t-s)$, and that $\mathbf{u}^r(t)$ is uncorrelated with $\tilde{\mathbf{x}}(T|T)$.

(c) Conclude that (16.5.12) holds.

16.29 **(Fixed-point smoothing)** Consider the state-space model (16.1.1)–(16.1.3) with $S(t) = 0$. Fix a time instant t_o and consider the smoothed estimator $\hat{\mathbf{x}}(t_o|T)$. In this problem we let T increase and therefore derive a solution to the so-called fixed-point smoothing problem. Let also $P(t_o)$ denote the error covariance matrix of the associated Kalman filter at time t_o.

(a) Use (16.5.8) and (16.5.7) to show that

$$\frac{d}{dT}\hat{\mathbf{x}}(t_o|T) = P(t_o)\Psi^*(T,t_o)H^*(T)R^{-1}(T)\mathbf{e}(T) \;=\; P(t_o,T)H^*(T)R^{-1}(T)\mathbf{e}(T),$$

with boundary condition $\hat{\mathbf{x}}(t_o|t_o)$.

(b) Using (16.5.4) we have $P(t_o,T) = P(t_o)\Psi^*(T,t_o)$. Differentiate this expression with respect to T to conclude that

$$\frac{d}{dT}P(t_o,T) = P(t_o)[F(T) - P(T)H^*(T)R^{-1}(T)H(T)], \quad P(t_o,t_o) = P(t_o).$$

This expression can be solved to determine $P(t_o,T)$ for the expression in part (a), without the need to explicitly determine $\Psi(T,t_o)$.

(c) Next use (16.5.2) to conclude that

$$\frac{d}{dT}P(t_o|T) = -P(t_o,T)H^*(T)R^{-1}(T)H(T)P(t_o,T), \quad P(t_o|t_o) = P(t_o).$$

16.30 **(Fixed-lag smoothing)** Consider again the state-space model (16.1.1)–(16.1.3) with $S(t) = 0$. Now choose a positive number T and let t increase. Let $\hat{\mathbf{x}}(t|t+T)$ denote the l.l.m.s. estimator of $\mathbf{x}(t)$ given the observations $\{\mathbf{y}(\tau), 0 \le \tau < t+T\}$.

(a) Show that

$$\hat{\mathbf{x}}(t|t+T) = \hat{\mathbf{x}}(t|t) + P(t)\boldsymbol{\lambda}(t|t+T),$$

where

$$\boldsymbol{\lambda}(t|t+T) = \int_t^{t+T} \Psi^*(\tau,t)H^*(\tau)R^{-1}(\tau)\mathbf{e}(\tau)d\tau.$$

(b) Conclude that

$$\dot{\hat{\mathbf{x}}}(t|t+T) = F(t)\hat{\mathbf{x}}(t|t+T) + G(t)Q(t)G^*(t)P^{-1}(t)[\hat{\mathbf{x}}(t|t+T) - \hat{\mathbf{x}}(t)] + P(t)\Psi^*(t+T,t)H^*(t+T)R^{-1}(t+T)\mathbf{e}(t+T),$$

with a boundary condition $\hat{\mathbf{x}}(t_o|t_o+T)$ that can be obtained from a fixed-point smoothing algorithm. Show also using (16.5.2) that

$$\begin{aligned}\dot{P}(t|t+T) = {} & [F(t) + G(t)Q(t)G^*(t)P^{-1}(t)]P(t|t+T) \\ & + P(t|t+T)[F(t) + G(t)Q(t)G^*(t)P^{-1}(t)]^* - \\ & - P(t)\Psi^*(t+T,t)H^*(t+T)R^{-1}(t+T)H(t+T)\Psi(t+T,t)P(t) \\ & - G(t)Q(t)G^*(t),\end{aligned}$$

with a boundary condition $P(t_o|t_o+T)$.

16.31 **(Stochastic interpretation of $\boldsymbol{\lambda}(t|T)$)** Consider the state-space model (16.1.1)–(16.1.3) with $S(t) = 0$. Show that

$$\hat{\mathbf{u}}(t|T) = \int_t^T \langle \mathbf{u}(t), \mathbf{e}(\tau)\rangle\; R^{-1}(\tau)\mathbf{e}(\tau)d\tau \;=\; Q(t)G^*(t)\boldsymbol{\lambda}(t|T).$$

16.32 **(A change-of-initial condition formula)** Consider the Riccati differential equation

$$\dot{P}(t) = FP(t) + P(t)F^* + GQG^* - P(t)H^*R^{-1}HP(t),$$

and assume we solve it for two different initial conditions at time t_0, say $P(t_0)$ and $P_1(t_0)$. Let $K(t)$ and $K_1(t)$ denote the resulting gain matrices, *i.e.*, $K(t) = P(t)H^*R^{-1}$ and $K_1(t) = P_1(t)H^*R^{-1}$. Let also $\Psi(t,t_0)$ and $\Psi_1(t,t_0)$ be the state transition matrices of $F - K(t)H$ and $F - K_1(t)H$, respectively, *e.g.*,

$$\frac{d}{dt}\Psi(t,t_0) \;=\; [F - K(t)H]\Psi(t,t_0), \quad \Psi(t_0,t_0) = I.$$

Establish the change-in-initial-conditions formula (assuming the required inverse exists):

$$\Psi_1(t,t_0) = \Psi(t,t_0)[I + (P_1(t_0) - P(t_0))\mathcal{O}(t,t_0)]^{-1},$$

where

$$\mathcal{O}(t,t_0) = \int_{t_0}^t \Psi^*(\tau,t_0)H^*R^{-1}H\Psi(\tau,t_0)d\tau.$$

Remark. This result is a special case of Thm. 17.4.1, where several such formulas are derived by employing a scattering theory approach. ◆

16.33 **(Another expression for $\dot{\Psi}(t,t_0)$)** Consider again the matrix $\Psi(t,t_0)$ from Prob. 16.32. Show that it also satisfies the differential equation

$$\frac{d}{dt}\Psi(t,t_0) = \Psi(t,t_0)[F - P(t_0)H^*R^{-1}H - \dot{P}(t_0)\mathcal{O}(t,t_0)].$$

[*Hint.* Evaluate the limit of $[\Psi(t+\Delta,t_0) - \Psi(t,t_0)]/\Delta$, as $\Delta \to 0$.]

16.34 **(Fast algorithm for the adjoint variable)** Fix the time instant t, say $t = t_0$. It then follows from (16.5.7) that

$$\boldsymbol{\lambda}(t_0|T) = \int_{t_0}^T \Psi^*(\tau,t_0)H^*R^{-1}\mathbf{e}(\tau), \quad \boldsymbol{\lambda}(T|T) = 0,$$

where $\Psi(\tau,t_0)$ denotes the state-transition matrix function for the closed-loop system $F_{cl}(\cdot)$, *viz.*,

$$\frac{d}{d\tau}\Psi(\tau,t_0) = F_{cl}(\tau)\Psi(\tau,t_0), \quad \Psi(t_0,t_0) = I.$$

A fast procedure for evaluating $\lambda(t_0|T)$ should therefore involve a fast procedure for evaluating the products $\{\Psi^*(\tau, t_0)H^*\}$. Define the quantities $W(t) = H\Psi(t, t_0)$ and $Z(t) = L^*(t_0)\mathcal{O}(t, t_0)$, with $W(t_0) = H$ and $Z(t_0) = 0$. Use the result of Prob. 16.33 (see also the change-in-initial-conditions formula (17.4.9)) to show that $\{W(\cdot)\}$ can be obtained as follows. Start with $W(t_0) = H$, $Z(t_0) = 0$, and solve

$$\dot{Z}(t) = L^*(t_0)W^*(t)R^{-1}W(t),$$

$$\dot{W}(t) = W(t)[F - P(t_0)H^*R^{-1}H - L(t_0)JZ(t)],$$

where $\{L(t_0), J\}$ are obtained from the factorization $\dot{P}(t_0) = L(t_0)JL^*(t_0)$ and can also be computed via the fast recursions of Thm. 16.6.1.

Remark. For more, see Ljung and Kailath (1977). ◆

16.35 (A nonstandard state-space model) Consider the nonstandard state-space model

$$\begin{cases} \dot{\mathbf{x}}\ (t) = F(t)\mathbf{x}(t) + \int_0^t G(\tau)\mathbf{u}(\tau)d\tau \\ \mathbf{y}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t) \end{cases}$$

where $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$ are uncorrelated white-noise processes with intensities $\{Q(\cdot), R(\cdot)\}$, both of which are uncorrelated with $\mathbf{x}(0)$, whose variance we denote by Π_0. Show that

$$\dot{\hat{\mathbf{x}}}\ (t) = F(t)\hat{\mathbf{x}}(t) + P_1(t)H^*(t)R^{-1}(t)\left[\mathbf{y}(t) - H(t)\hat{\mathbf{x}}(t)\right] + \int_0^t P_2(\tau)H^*(\tau)R^{-1}(\tau)\left[\mathbf{y}(\tau) - H(\tau)\hat{\mathbf{x}}(\tau)\right]d\tau$$

where

$$\begin{aligned} \dot{P}_1 &= FP_1 + P_1F^* + P_2 + P_2^* - P_1H^*R^{-1}HP_1, & P_1(0) &= \Pi_0, \\ \dot{P}_2 &= P_2F^* + P_3 - P_2H^*R^{-1}HP_1, & P_2(0) &= 0, \\ \dot{P}_3 &= -P_2H^*R^{-1}HP_2^* + GQG^*, & P_3(0) &= 0. \end{aligned}$$

16.36 (Fast extended equations) Consider the state-space model

$$\begin{cases} \dot{\mathbf{x}}\ (t) = F(t)\mathbf{x}(t) \\ \mathbf{y}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t) \end{cases}$$

with

$$\left\langle \begin{bmatrix} \mathbf{x}(0) \\ \mathbf{v}(t) \end{bmatrix}, \begin{bmatrix} \mathbf{x}(0) \\ \mathbf{v}(s) \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & R\delta(t-s) & 0 \end{bmatrix}.$$

Note that R is constant but that the system matrices are time-variant. However, assume that the system matrices satisfy the relations

$$\dot{F}\ (t) = \Psi F(t) - F(t)\Psi, \quad \dot{H}\ (t) = -H(t)\Psi,$$

for some constant matrix Ψ.

(a) Define $P_\Psi = \dot{P} - \Psi P - P\Psi^*$ where $P(\cdot)$ is the (usual) error variance matrix. Show that

$$\dot{P}_\Psi\ (t) = (F(t) - K(t)H(t))P_\Psi(t) + P_\Psi(t)(F(t) - K(t)H(t))^*,$$

where $K(t) = P(t)H^*(t)R^{-1}$ is the Kalman filter gain and deduce that $P_\Psi(t)$ has constant inertia, *i.e.*, $P_\Psi(t) = L(t)JL^*(t)$, for all $t \geq 0$, where J is a signature matrix of dimension say, α.

(b) Show that we can find $K(t)$ via the coupled differential equations

$$\begin{cases} \dot{K}\ (t) = L(t)JL^*(t)H^*(t)R^{-1} + \Psi K(t), \\ \dot{L}\ (t) = (F(t) - K(t)H(t))L(t), \end{cases}$$

initialized with $K(0) = P(0)H^*(0)R^{-1}$ and $L(0)$, where $L(0)$ is obtained from the factorization

$$L(0)JL^*(0) = \dot{P}\ (0) - \Psi P(0) - P(0)\Psi^*.$$

Remark. These results are the continuous-time counterparts of the results in discrete time described in Secs. 11.3 and 13.4. ◆

Appendix for Chapter 16

16.A BACKWARDS MARKOVIAN MODELS

The discussions in Sec. 5.4 of the Markov property and of Markovian representations for discrete-time processes go over with obvious changes to the continuous-time case. We therefore focus on only a few key results in the following pages.

16.A.1 Backwards Models via Time Reversal

By reversing time[5] in the model (16.1.1) we get, say over an interval of time $[0, T]$,

$$-\dot{\mathbf{x}}(t) \;=\; -F(t)\mathbf{x}(t) - G(t)\mathbf{u}(t), \qquad T \geq t \geq 0. \tag{16.A.1}$$

This model is not Markovian (and therefore not suitable for estimation purposes) because, for any t, the driving noise $\mathbf{u}(t)$ is not uncorrelated with the "initial" state $\mathbf{x}(T)$. Following Verghese and Kailath (1979), and recalling the discrete-time results (recall Eq. (5.4.14)), we can show that the reversed-time model (16.A.1) can be made Markovian by replacing $\mathbf{u}(t)$ by

$$\mathbf{u}^b(t) \;=\; \mathbf{u}(t) - Q(t)G^*(t)\Pi^{-1}(t)\mathbf{x}(t), \tag{16.A.2}$$

where we are assuming the invertibility of the state covariance matrix $\Pi(t) = \|\mathbf{x}(t)\|^2$, a condition that can be guaranteed by using any of the sufficient conditions stated in Prob. 16.25, such as requiring $\Pi_0 > 0$.

To see why (16.A.2) leads to a Markovian model, we must show that

$$\langle \mathbf{u}^b(t), \mathbf{x}(T) \rangle \;=\; 0, \tag{16.A.3}$$

and that

$$\langle \mathbf{u}^b(t), \mathbf{u}^b(s) \rangle \;=\; Q(t)\delta(t-s),$$

for all $t \in [0, T]$. The proofs are outlined in Probs. 16.19–16.23. Substituting (16.A.2) into the reversed-time state-space equation (16.A.1) yields the backwards state-equation, as noted in the following lemma.

Lemma 16.A.1 (Backwards State-Space Model) *Given the forwards Markovian state-space equation*

$$\dot{\mathbf{x}}(t) \;=\; F(t)\mathbf{x}(t) + G(t)\mathbf{u}(t), \quad 0 \leq t \leq T,$$

with invertible $\Pi_0 > 0$, *and*

$$\left\langle \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{x}(0) \end{bmatrix}, \begin{bmatrix} \mathbf{u}(s) \\ \mathbf{x}(0) \end{bmatrix} \right\rangle = \begin{bmatrix} Q(t)\delta(t-s) & 0 \\ 0 & \Pi_0 \end{bmatrix}.$$

[5] In contrast to the discrete-time case, continuous-time models are always reversible and we do not need to assume the invertibility of $F(\cdot)$.

Then a backwards Markovian state-space equation for $\{\mathbf{x}(\cdot)\}$ *is*

$$-\dot{\mathbf{x}}(t) \;=\; F^b(t)\mathbf{x}(t) + G^b(t)\mathbf{u}^b(t),$$

with

$$F^b(t) \;=\; -[F(t) + G(t)Q(t)G^*(t)\Pi^{-1}(t)], \qquad G^b(t) \;=\; -G(t), \tag{16.A.4}$$

and

$$\left\langle \begin{bmatrix} \mathbf{u}^b(t) \\ \mathbf{x}(T) \end{bmatrix}, \begin{bmatrix} \mathbf{u}^b(s) \\ \mathbf{x}(T) \end{bmatrix} \right\rangle = \begin{bmatrix} Q(t)\delta(t-s) & 0 \\ 0 & \Pi(T) \end{bmatrix}.$$

■

Example 16.A.1 (Wide-Sense Gauss-Markov Processes) A wide-sense Gauss-Markov process $\mathbf{y}(\cdot)$ over $[0, T]$ is one that has a covariance function of the form (see, *e.g.*, Doob (1953)):

$$\langle \mathbf{y}(s), \mathbf{y}(t) \rangle \;=\; \begin{cases} a(s)b(t) & \text{for} \quad s \leq t, \\ a(t)b(s) & \text{for} \quad s \geq t, \end{cases}$$

where $a(\cdot)$ and $b(\cdot)$ are continuous functions, unique up to a constant and $n(\cdot) \triangleq a(\cdot)/b(\cdot)$ is continuous and strictly increasing on $[0, T]$. It can be checked that a forwards Markovian state-space model for $\mathbf{y}(\cdot)$ can be formed as

$$\dot{\mathbf{x}}(t) \;=\; g(t)\mathbf{u}(t), \qquad \mathbf{y}(t) \;=\; b(t)\mathbf{x}(t),$$

where $\mathbf{u}(\cdot)$ is white with unit intensity, $\|\mathbf{x}(0)\|^2 = a(0)/b(0)$, and $g(\cdot)$ is defined by

$$\frac{a(t)}{b(t)} \;=\; \frac{a(0)}{b(0)} \;+\; \int_0^t g^2(s)ds.$$

It can be readily seen that the state variance is $\|\mathbf{x}(t)\|^2 = n(t)$ and, consequently, the backwards state-space model of Lemma 16.A.1 becomes

$$-\dot{\mathbf{x}}(t) \;=\; -\frac{b(t)}{a(t)}g^2(t)\mathbf{x}(t) \;-\; g(t)\mathbf{u}^b(t), \qquad \mathbf{y}(t) \;=\; b(t)\mathbf{x}(t),$$

where $\mathbf{u}^b(\cdot)$ is unit variance and $\|\mathbf{x}(T)\|^2 = n(T)$. The special case $a(t) = t$, $b(t) = 1$, defines the so-called Wiener process. ◆

Example 16.A.2 (Stationary Processes) Assume $\{F, G, H, R, Q\}$ are time-invariant as well as Π, *i.e.*, assume the process $\mathbf{y}(\cdot)$ is stationary. This happens for instance when F is stable (eigenvalues in the left-half plane) and Π_0 is chosen as the unique solution $\bar{\Pi}$ of the Lyapunov equation (*cf.* the discussion in Ch. 8 for the discrete-time case):

$$0 \;=\; F\bar{\Pi} + \bar{\Pi}F^* + GQG^*.$$

Then,

$$F^b(t) \;=\; -F - GQG^*\bar{\Pi}^{-1} \;=\; \bar{\Pi}F^*\bar{\Pi}^{-1} \;=\; F^b, \quad \text{say.}$$

This shows that the backwards model has the same stability properties as the forwards model since F^* and F^b are related via a similarity transformation. ◆

16.A.2 The Backwards-Time Kalman Filters

Just as for the forwards-time model, we can derive Kalman filter equations for processes described by backwards-time models

$$-\dot{\mathbf{x}}(t) = F^b(t)\mathbf{x}(t) + G^b(t)\mathbf{u}^b(t), \tag{16.A.5}$$

$$\mathbf{y}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t), \tag{16.A.6}$$

with the conditions

$$\left\langle \begin{bmatrix} \mathbf{u}^b(t) \\ \mathbf{v}(t) \\ \mathbf{x}(T) \end{bmatrix}, \begin{bmatrix} \mathbf{u}^b(s) \\ \mathbf{v}(s) \\ \mathbf{x}(T) \end{bmatrix} \right\rangle = \begin{bmatrix} Q(t)\delta(t-s) & 0 & 0 \\ 0 & R(t)\delta(t-s) & 0 \\ 0 & 0 & \Pi(T) \end{bmatrix}. \tag{16.A.7}$$

Let

$$\hat{\mathbf{y}}^b(t) = \text{the l.l.m.s. estimator of } \mathbf{y}(t) \text{ given } \{\mathbf{y}(\tau),\ t < \tau \le T\}, \tag{16.A.8}$$

and define the backwards innovations process

$$\mathbf{e}^b(t) = \mathbf{y}(t) - \hat{\mathbf{y}}^b(t) = \mathbf{y}(t) - H(t)\hat{\mathbf{x}}^b(t), \tag{16.A.9}$$

with $\mathbf{e}^b(T) = \mathbf{y}(T)$, $\hat{\mathbf{x}}^b(T) = 0$. Then arguments similar to those in Sec. 16.4.2 yield the following theorem.

Theorem 16.A.1 (The Backwards Kalman Filter) *Consider the backwards-time model standard state-space model (16.A.5)–(16.A.7). Then the innovations $\mathbf{e}^b(t)$ can be evaluated via the following backwards time equations:*

$$\begin{aligned}
\hat{\mathbf{e}}^b(t) &= \mathbf{y}(t) - H(t)\hat{\mathbf{x}}^b(t), \\
-\dot{\hat{\mathbf{x}}}^b(t) &= F^b(t)\hat{\mathbf{x}}^b(t) + K^b(t)\mathbf{e}^b(t), \quad \hat{\mathbf{x}}^b(T) = 0, \\
K^b(t) &= P^b(t)H^*(t)R^{-1}(t), \\
-\dot{P}^b(t) &= F^b(t)P^b(t) + P^b(t)F^{b*}(t) + G^b(t)Q(t)G^{b*}(t) - K^b(t)R(t)K^{b*}(t),
\end{aligned}$$

with boundary condition $P^b(T) = \Pi(T)$. ∎

16.A.3 Application to Smoothing Problems

The direction of time is not relevant in smoothing problems and we should therefore be able to derive smoothing algorithms by employing a backwards Markovian state-space model. We omit the straightforward derivations and just state the following results.

Theorem 16.A.2 (Backwards BF Smoothing Formulas) *Given the state-space model (16.A.5)–(16.A.7), we can find the smoothed estimator $\hat{\mathbf{x}}(t|T)$ via*

$$\hat{\mathbf{x}}(t|T) = \hat{\mathbf{x}}^b(t) + P^b(t)\boldsymbol{\lambda}^b(t|T), \quad 0 \le t \le T, \tag{16.A.10}$$

where $\boldsymbol{\lambda}^b(t|T)$ satisfies

$$\dot{\boldsymbol{\lambda}}^b(t|T) = [F^b(t) - K^b(t)H(t)]^*\boldsymbol{\lambda}^b(t|T) + H^*(t)R^{-1}(t)\mathbf{e}^b(t), \quad \boldsymbol{\lambda}^b(0|0) = 0. \tag{16.A.11}$$

The smoothed error variance can then be readily computed as

$$P(t|T) = P^b(t) - P^b(t)\left[\int_0^t \Psi^{b*}(s,t)H^*(s)R^{-1}(s)H(s)\Psi^b(s,t)ds\right]P^b(t). \tag{16.A.12}$$

The quantities $\{\hat{\mathbf{x}}^b(t), \mathbf{e}^b(t), K^b(t), P^b(t)\}$ are found in a backwards pass by running the Kalman filter equations of Thm. 16.A.1 over the interval $[0, T]$. ∎

Theorem 16.A.3 (The Backwards RTS Equations) *Given (16.A.5)–(16.A.7), we can find the smoothed estimator $\hat{\mathbf{x}}(t|T)$ by solving equation*

$$-\dot{\hat{\mathbf{x}}}(t|T) = F_s^b(t)\hat{\mathbf{x}}(t|T) - G(t)Q(t)G^*(t)P^{-b}(t)\hat{\mathbf{x}}^b(t), \quad \hat{\mathbf{x}}(0|0) = \hat{\mathbf{x}}^b(0), \tag{16.A.13}$$

where $F_s^b(t) = F^b(t) + G^b(t)Q(t)G^{b*}(t)P^{-b}(t)$. The smoothing error variance obeys the following equation, with initial condition $P(0|0) = P^b(0)$,

$$-\frac{dP(t|T)}{dt} = F_s^b(t)P(t|T) + P(t|T)F_s^{b*}(t) - G^b(t)Q(t)G^{b*}(t). \tag{16.A.14}$$

∎

Theorem 16.A.4 (Two-Filter Smoothing Formulas) *Given (16.A.5)–(16.A.7), we can find the smoothed estimator $\hat{\mathbf{x}}(t|T)$ and its error variance by using*

$$[P^{-1}(t) + P^{-b}(t) - \Pi^{-1}(t)]\hat{\mathbf{x}}(t|T) = P^{-1}(t)\hat{\mathbf{x}}(t) + P^{-b}(t)\hat{\mathbf{x}}^b(t), \tag{16.A.15}$$

and

$$P^{-1}(t|T) = P^{-1}(t) + P^{-b}(t) - \Pi^{-1}(t). \tag{16.A.16}$$

∎

Proof: We add the estimator equations (16.5.11) and (16.A.13) of both the forwards and backwards RTS formulas, and use the expressions in Thm. 16.A.1 for $F^b(t)$ and $G^b(t)$ as well as the definitions for $F_s(t)$ and $F_s^b(t)$. This leads to (16.A.15).

To prove (16.A.16), we use (16.A.15) and subtract $[P^{-1}(t)+P^{-b}(t)-\Pi^{-1}(t)]\mathbf{x}(t)$ from both sides to conclude that

$$[P^{-1}(t) + P^{-b}(t) - \Pi^{-1}(t)]\tilde{\mathbf{x}}(t|T) = P^{-1}(t)\tilde{\mathbf{x}}(t) + P^{-b}(t)\tilde{\mathbf{x}}^b(t) + \Pi^{-1}(t)\mathbf{x}(t).$$

Multiplying from the right by $\mathbf{x}^*(t)$ and taking expectations we obtain

$$[P^{-1}(t) + P^{-b}(t) - \Pi^{-1}(t)]P(t|T) = I + I - I = I,$$

which establishes (16.A.16). ◆

A special case of the two-filter formulas is the Mayne-Fraser smoother, which is obtained for the special boundary condition $P^{-b}(T) = 0$ (or $P^b(T) = \Pi(T) = \infty \cdot I$ — the former representation of the boundary condition in terms of the inverse of $P^b(T)$ is more suitable for computational purposes. We denote $P^b(t)$ in this case by $P_\infty^b(t)$).

Theorem 16.A.5 (Mayne-Fraser Formulas) *Given the model (16.A.5)–(16.A.7), we can find the smoothed estimator $\hat{\mathbf{x}}(t|T)$ and its error variance by using*

$$[P^{-1}(t) + P_\infty^{-b}(t)]\hat{\mathbf{x}}(t|T) = P^{-1}(t)\hat{\mathbf{x}}(t) + P_\infty^{-b}(t)\hat{\mathbf{x}}_\infty^b(t), \tag{16.A.17}$$

where $P_\infty^{-b}(\cdot)$ is the solution of the backwards-time differential equation

$$\dot{P}_\infty^{-b}(t) = -F^*(t)P_\infty^{-b}(t) - P_\infty^{-b}(t)F(t) + P_\infty^{-b}(t)G(t)Q(t)G^*(t)P_\infty^{-b}(t) - H^*(t)R^{-1}(t)H(t), \quad P_\infty^{-b}(T) = 0. \tag{16.A.18}$$

Likewise, $\hat{\mathbf{x}}_\infty^b(\cdot)$ is the solution of the backwards Kalman filter of Thm. 16.A.1 with $P^b(\cdot)$ replaced by $P_\infty^b(\cdot)$ and $\Pi^{-1}(t)$ replaced by zero,

$$-\dot{\hat{\mathbf{x}}}_\infty^b(t) = -F(t)\hat{\mathbf{x}}_\infty^b(t) + P_\infty^b(t)H^*(t)R^{-1}(t)[\mathbf{y}(t) - H(t)\hat{\mathbf{x}}_\infty^b(t)], \quad \hat{\mathbf{x}}^b(T) = 0.$$

■

CHAPTER 17

A Scattering Theory Approach

In the preceding chapters we have presented two different methods of solving linear least-squares estimation problems for stochastic processes, starting either with known state-space models or with covariance/spectral data given in state-space form. The latter approach was followed in Ch. 8, while the model-based approach was developed in Chs. 9–15. Here we shall present a third approach, based on an interesting and fruitful analogy with a so-called *scattering/transmission-line theory*, developed in the 1940s and 1950s (especially by Redheffer (1962)) to study the propagation of waves through a medium, as for example sunlight through the earth's atmosphere (radiative transfer theory), or electric and magnetic waves along a transmission line.

As perhaps first noticed by Ambartsumian (1943), Riccati equations arise in a natural and important way in this theory, a fact that inspired Kailath and Ljung (1977) to develop a scattering theory framework for least-squares estimation. Later it was found by Verghese, Friedlander, and Kailath (1980) that a natural starting point was the Hamiltonian equations (16.5.14) for the smoothed estimators of the state vector. This is of particular interest because as we showed in Ch. 15, Sec. 15.7.3, the Hamiltonian equations can be directly obtained by studying the properties of the linear spaces associated with stochastic state-space models. In other words, the results in this chapter give yet another independent route to state-space estimation theory. Moreover, unlike the approach in earlier chapters, it starts with smoothed estimators (as studied in Ch. 10) as opposed to predicted/filtered estimators (Ch. 9).

In Sec. 17.1, we show how to go from the Hamiltonian equations to a scattering or transmission-line model in which the smoothed estimators and the adjoint variable are interpreted as waves traveling in opposite directions, with interactions defined by the parameters $\{F, G, H, Q, R\}$ of the state-space model. Then we quickly demonstrate

that the Riccati variable of the Kalman filter theory can be identified as the so-called left reflection coefficient of the model. Other important objects in the theory, such as the closed-loop state transition matrix and its Gramian also arise naturally. In fact, they arise in a certain combination (the so-called scattering matrix), study of which leads naturally to many interesting identities, including several previously obtained after considerable algebraic manipulations.

For variety, we start with the continuous-time problems and then more briefly present the discrete-time results.

17.1 A GENERALIZED TRANSMISSION–LINE MODEL

We recall the standard continuous-time state-space model:

$$\dot{\mathbf{x}}(s) = F(s)\mathbf{x}(s) + G(s)\mathbf{u}(s), \tag{17.1.1}$$

$$\mathbf{y}(s) = H(s)\mathbf{x}(s) + \mathbf{v}(s), \qquad 0 \le s \le t, \tag{17.1.2}$$

where $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot)\}$ are white-noise uncorrelated zero-mean processes such that for all σ and s in $[0, t]$

$$\left\langle \begin{bmatrix} \mathbf{u}(s) \\ \mathbf{v}(s) \\ \mathbf{x}(0) \end{bmatrix}, \begin{bmatrix} \mathbf{u}(\sigma) \\ \mathbf{v}(\sigma) \\ \mathbf{x}(0) \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q(s)\delta(s-\sigma) & 0 & 0 & 0 \\ 0 & R(s)\delta(s-\sigma) & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}. \tag{17.1.3}$$

We shall focus in the sequel on the subinterval $[\tau, t]$; the left-hand (initial) point is denoted by τ, because we will soon need to allow τ and t to change: t increasing with τ fixed, or τ decreasing with t fixed. Thus let $\mathbf{x}(\tau)$ denote the state vector at time τ and let $\hat{\mathbf{x}}(\tau)$ denote the l.l.m.s. estimator of $\mathbf{x}(\tau)$ given the observations $\{\mathbf{y}(\nu), 0 \le \nu < \tau\}$. Denote the corresponding error variance matrix by $P(\tau)$,

$$P(\tau) = \|\mathbf{x}(\tau) - \hat{\mathbf{x}}(\tau)\|^2.$$

We shall also write, for all s,

$$\hat{\mathbf{x}}(s|t) = \text{the l.l.m.s.e. of } \mathbf{x}(s) \text{ given } \{\mathbf{y}(\sigma),\ 0 \le \sigma \le t\},$$

and denote, as earlier, $\hat{\mathbf{x}}(s|s)$ by $\hat{\mathbf{x}}(s)$.

Now the Hamiltonian equations for this continuous-time model are derived in App. 17.A, where they are shown to be

$$\begin{bmatrix} \dot{\hat{\mathbf{x}}}(s|t) \\ -\dot{\boldsymbol{\lambda}}(s|t) \end{bmatrix} = \begin{bmatrix} 0 \\ H^*(s)R^{-1}(s)\mathbf{y}(s) \end{bmatrix} + \tag{17.1.4}$$

$$\begin{bmatrix} F(s) & G(s)Q(s)G^*(s) \\ -H^*(s)R^{-1}(s)H(s) & F^*(s) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}(s|t) \\ \boldsymbol{\lambda}(s|t) \end{bmatrix},$$

with boundary conditions

$$\boldsymbol{\lambda}(t|t) = 0, \qquad \hat{\mathbf{x}}(\tau|t) = \hat{\mathbf{x}}(\tau) + P(\tau)\boldsymbol{\lambda}(\tau|t). \tag{17.1.5}$$

Next we discretize the Hamiltonian equations, using the Euler approximations:

$$\dot{\hat{\mathbf{x}}}(s|t) \approx \frac{\hat{\mathbf{x}}(s+\Delta|t) - \hat{\mathbf{x}}(s|t)}{\Delta}, \qquad \dot{\boldsymbol{\lambda}}(s|t) \approx \frac{\boldsymbol{\lambda}(s+\Delta|t) - \boldsymbol{\lambda}(s|t)}{\Delta},$$

where Δ is small enough, to obtain

$$\begin{bmatrix} \hat{\mathbf{x}}(s+\Delta|t) \\ \boldsymbol{\lambda}(s|t) \end{bmatrix} = \begin{bmatrix} 0 \\ H^*(s)R^{-1}(s)\Delta \end{bmatrix} \mathbf{y}(s)$$

$$\begin{bmatrix} I + F(s)\Delta & G(s)Q(s)G^*(s)\Delta \\ -H^*(s)R^{-1}(s)H(s)\Delta & I + F^*(s)\Delta \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}(s|t) \\ \boldsymbol{\lambda}(s+\Delta|t) \end{bmatrix}.$$

[It is important to note that the minus sign in the original differential equation leads us to put $\boldsymbol{\lambda}(s|t)$ rather than $\boldsymbol{\lambda}(s+\Delta|t)$ on the left-hand side of the discretized equation.] The difference equation can now be graphically depicted as shown in Fig. 17.1, which suggests that we can regard $\hat{\mathbf{x}}(\cdot|t)$ as a *forward* wave and $\boldsymbol{\lambda}(\cdot|t)$ as a *backward* wave traveling through an incremental section at time t of some generalized transmission line or *scattering medium*, specified by the scattering quantities:

$$\begin{aligned} I + F(s)\Delta &= \text{the incremental forward transmission coefficient,} \\ I + F^*(s)\Delta &= \text{the incremental backward transmission coefficient,} \\ -H^*(s)R^{-1}(s)H(s)\Delta &= \text{the incremental left reflection coefficient,} \\ G(s)Q(s)G^*(s)\Delta &= \text{the incremental right reflection coefficient,} \\ H^*(s)R^{-1}(s)\mathbf{y}(s)\Delta &= \text{the incremental internal backward source excitation.} \end{aligned}$$

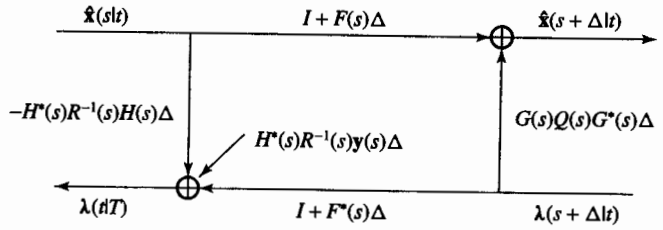


Figure 17.1 An incremental scattering section at time s.

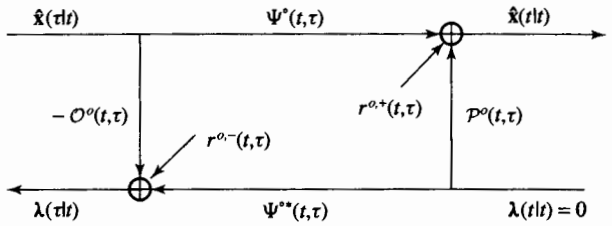


Figure 17.2 A macroscopic scattering section from the initial time τ to the final time t.

Moreover, we can put together several such incremental sections to get a macroscopic section of the scattering medium from time τ up to time t. This is shown in Fig. 17.2 where we can designate

$$\begin{aligned}
\Psi^o(t,\tau) &= \text{the forward transmission operator,}\\
\Psi^{o*}(t,\tau) &= \text{the backward transmission operator,}\\
\mathcal{P}^o(t,\tau) &= \text{the right reflection operator,}\\
-\mathcal{O}^o(t,\tau) &= \text{the left reflection operator,}\\
r^{o,+}(t,\tau) &= \text{the forward internal source,}\\
r^{o,-}(t,\tau) &= \text{the backward internal source.}
\end{aligned}$$

It may not be surprising that the forward and backward transmission operators were taken as conjugates of each other, but this can readily be established (see below and also Prob. 17.1). The choice of notation, *e.g.*, $\mathcal{P}^o(t,\tau)$, may seem strange, but the reasons will soon appear.

Remark 1. The reason for the superscript ^{o} is that when we collapse the section by setting $t=\tau$, there are no reflections, so that

$$\mathcal{P}^o(\tau) \triangleq \mathcal{P}^o(\tau,\tau) = 0, \qquad \mathcal{O}^o(\tau,\tau) = 0,$$

while the transmissions are perfect, *i.e.*,

$$\Psi^o(\tau,\tau) = I = \Psi^{o*}(\tau,\tau).$$

Later we shall show how to incorporate nonzero boundary conditions. ◆

17.1.1 Identifying the Macroscopic Scattering Operators

The scattering operators $\{\Psi^o, \mathcal{P}^o, \mathcal{O}^o\}$ can be identified by considering the effect of adding an incremental section from t to $t+\Delta$ to the right of the composite layer of Fig. 17.2. The result is shown in Fig. 17.3 where, for notational convenience, we have dropped the argument $(t+\Delta)$ from the quantities $\{F, H, G, R, Q, \mathbf{y}\}$. That is, the quantity $(I+F\Delta)$ that appears in the figure should be interpreted as $[I+F(t+\Delta)\Delta]$. Likewise, for the quantities $\{-H^*R^{-1}H\Delta, GQG^*\Delta, H^*R^{-1}\mathbf{y}\Delta\}$.

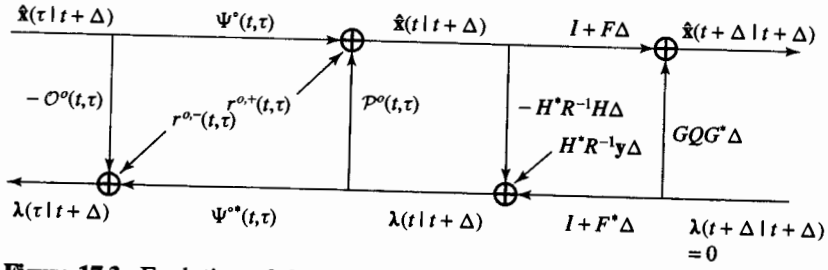


Figure 17.3 Evolution of the scattering quantities. The argument $(t+\Delta)$ has been dropped, for convenience, from the quantities $\{F, H, G, R, Q, \mathbf{y}\}$.

Now by tracing paths through Fig. 17.3, we can write

$$\begin{aligned}
\mathcal{P}^o(t+\Delta,s) = GQG^*\Delta + (I+F\Delta)[&\mathcal{P}^o(t,\tau)\\
&- \mathcal{P}^o(t,\tau)H^*R^{-1}H\mathcal{P}^o(t,\tau)\Delta + o(\Delta)](I+F^*\Delta),
\end{aligned}$$

where $o(\Delta)$ denotes terms that go to zero faster than Δ as $\Delta \to 0$. Therefore, taking the limit as $\Delta \to 0$ and recalling that the parameters $\{F, G, Q, R, H\}$ are all defined at $(t+\Delta)$, we obtain the differential equation

$$\begin{aligned}
\frac{d}{dt}\mathcal{P}^o(t,\tau) = {}& G(t)Q(t)G^*(t) + F(t)\mathcal{P}^o(t,\tau) + \mathcal{P}^o(t,\tau)F^*(t)\\
&- \mathcal{P}^o(t,\tau)H^*(t)R^{-1}(t)H(t)\mathcal{P}^o(t,\tau). \qquad (17.1.6)
\end{aligned}$$

This has exactly the same form as the Riccati differential equation (16.2.1) that arises in the Kalman filter solution, except that its initial condition is zero, $\mathcal{P}^o(\tau,\tau)=0$. Similarly, we can see that

$$\begin{aligned}
\Psi^o(t+\Delta,\tau) &= (I+F\Delta)\Psi^o(t,\tau) - (I+F\Delta)\mathcal{P}^o(t)H^*R^{-1}H\Psi^o(t,\tau) +\\
&\quad + (I+F\Delta)\mathcal{P}^o(t)H^*R^{-1}H\mathcal{P}^o(t)H^*R^{-1}H\Delta^2\Psi^o(t,\tau) - \dots\\
&= (I+F\Delta)[I - \mathcal{P}^o(t)H^*R^{-1}H\Delta + o(\Delta)]\Psi^o(t,\tau).
\end{aligned}$$

It then follows that

$$\frac{\Psi^o(t+\Delta,\tau) - \Psi^o(t,\tau)}{\Delta} = [F - \mathcal{P}^o(t)H^*R^{-1}H]\Psi^o(t,\tau).$$

Taking the limit as $\Delta \to 0$, and recalling again that the parameters $\{F, R, H\}$ are defined at $(t+\Delta)$, we obtain

$$\begin{aligned}
\frac{d}{dt}\Psi^o(t,\tau) &= [F(t) - K^o(t)H(t)]\Psi^o(t,\tau),\\
K^o(t) &\triangleq \mathcal{P}^o(t,\tau)H^*(t)R^{-1}(t), \qquad (17.1.7)
\end{aligned}$$

which identifies $\Psi^o(t,\tau)$ as the state-transition matrix of the closed-loop matrix (*cf.* (16.5.5))

$$F^o_{cl}(t) \triangleq [F(t) - K^o(t)H(t)].$$

Note that $K^o(t)$ is also a function of τ; for convenience, we continue to write $K^o(t)$ rather than $K^o(t,\tau)$.

Finally, similar arguments show that

$$\frac{d}{dt}\mathcal{O}^o(t,\tau) = \Psi^{o*}(t,\tau)H^*(t)R^{-1}(t)H(t)\Psi^o(t,\tau), \qquad \mathcal{O}^o(\tau,\tau) = 0, \tag{17.1.8}$$

which identifies $\mathcal{O}^o(t,\tau)$ as the observability Gramian of $\{R^{-1/2}(\cdot)H(\cdot), F^o_{cl}(\cdot)\}$,

$$\mathcal{O}^o(t,\tau) = \int_\tau^t \Psi^{o*}(s,\tau)H^*(s)R^{-1}(s)H(s)\Psi^o(s,\tau)ds. \tag{17.1.9}$$

We can collect the scattering parameters into a scattering matrix $\mathcal{S}^o(t,\tau)$,

$$\mathcal{S}^o(t,\tau) = \begin{bmatrix} \Psi^o(t,\tau) & \mathcal{P}^o(t,\tau) \\ -\mathcal{O}^o(t,\tau) & \Psi^{o*}(t,\tau) \end{bmatrix}, \tag{17.1.10}$$

which allows the above discussion to be summarized as follows.

Lemma 17.1.1 (Scattering Matrix) *The scattering matrix (17.1.10) satisfies the following differential equation (where we are writing $\{F, G, H, K, R, Q\}$ instead of $\{F(t), G(t), H(t), K^o(t), R(t), Q(t)\}$):*

$$\frac{\partial \mathcal{S}^o(t,\tau)}{\partial t} = \tag{17.1.11}$$

$$\begin{bmatrix} [F - K^oH]\Psi^o(t,\tau) & F\mathcal{P}^o(t,\tau) + \mathcal{P}^o(t,\tau)F^* - K^oRK^{o*} + GQG^* \\ -\Psi^{o*}(t,\tau)H^*R^{-1}H\Psi^o(t,\tau) & \Psi^{o*}(t,\tau)[F - K^oH]^* \end{bmatrix},$$

with initial condition

$$\mathcal{S}^o(\tau,\tau) = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

■

We note also that if we interchange the columns of $\mathcal{S}^o(t,\tau)$ and define

$$\mathcal{X}^o(t,\tau) \triangleq \mathcal{S}^o(t,\tau)\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} = \begin{bmatrix} \mathcal{P}^o(t,\tau) & \Psi^o(t,\tau) \\ \Psi^{o*}(t,\tau) & -\mathcal{O}^o(t,\tau) \end{bmatrix}, \tag{17.1.12}$$

then $\mathcal{X}^o(t,\tau)$ itself satisfies a Riccati differential equation (see Prob. 17.2); this interesting fact will be exploited later in Sec. 17.4.4.

17.1.2 Identifying the Signals

Having identified the macroscopic parameters of the scattering medium, we may proceed to similarly identify the signals $r^{o,+}(t,\tau)$ and $r^{o,-}(t,\tau)$.

First note that with the incident quantities assumed to be zero at the boundaries of Fig. 17.2, *i.e.*, $\hat{\mathbf{x}}(\tau|t) = 0$ and $\boldsymbol{\lambda}(t|t) = 0$, we can identify the emerging waves from the medium as

$$r^{o,+}(t,\tau) = \hat{\mathbf{x}}^o(t|t) \quad \text{and} \quad r^{o,-}(t,\tau) = \boldsymbol{\lambda}^o(\tau|t). \tag{17.1.13}$$

From the boundary conditions (17.1.5) of the Hamiltonian equations, we see that a zero boundary condition for $\hat{\mathbf{x}}(\tau|t)$ occurs when the a priori information is taken to be zero, $\hat{\mathbf{x}}(\tau) = 0$ and $P(\tau) = 0$. Hence, the notation $\hat{\mathbf{x}}^o(t|t)$ and $\boldsymbol{\lambda}^o(\tau|t)$ is used to refer to the quantities $\{\hat{\mathbf{x}}(t|t), \boldsymbol{\lambda}(\tau|t)\}$ that result when the Hamiltonian equations (17.1.4) are solved with zero initial conditions, $\boldsymbol{\lambda}(t|t) = 0$, $\hat{\mathbf{x}}(\tau) = 0$, $P(\tau) = 0$.

To characterize these quantities, we start by adding an incremental section to the right of Fig. 17.2, assuming the above zero boundary conditions. Doing this gives the result shown in Fig. 17.4 (we are again dropping the argument $(t+\Delta)$ from the quantities $\{F, H, G, R, Q, \mathbf{y}\}$, for ease of representation).

Now by tracing the flow through the figure, we can first obtain the relation

$$\boldsymbol{\lambda}^o(\tau|t+\Delta) = \boldsymbol{\lambda}^o(\tau|t) + \Psi^{o*}(t,\tau)H^*R^{-1}[\mathbf{y}(t+\Delta) - H\hat{\mathbf{x}}^o(t|t+\Delta)]\Delta + o(\Delta),$$

which leads to the differential equation

$$\frac{\partial \boldsymbol{\lambda}^o(\tau|t)}{\partial t} = \Psi^{o*}(t,\tau)H^*(t)R^{-1}(t)[\mathbf{y}(t) - H(t)\hat{\mathbf{x}}^o(t|t)]. \tag{17.1.14}$$

This expression does not ring a bell, but we shall return to it presently. Let us first work out the equation for $\hat{\mathbf{x}}^o(t|t)$. From Fig. 17.4, we can write

$$\hat{\mathbf{x}}^o(t+\Delta|t+\Delta) = [I + F\Delta]\hat{\mathbf{x}}^o(t|t+\Delta),$$

$$\hat{\mathbf{x}}^o(t|t+\Delta) = \hat{\mathbf{x}}^o(t|t) + \mathcal{P}^o(t,\tau)H^*R^{-1}[\mathbf{y}(t+\Delta) - H\hat{\mathbf{x}}^o(t|t+\Delta)]\Delta + o(\Delta),$$

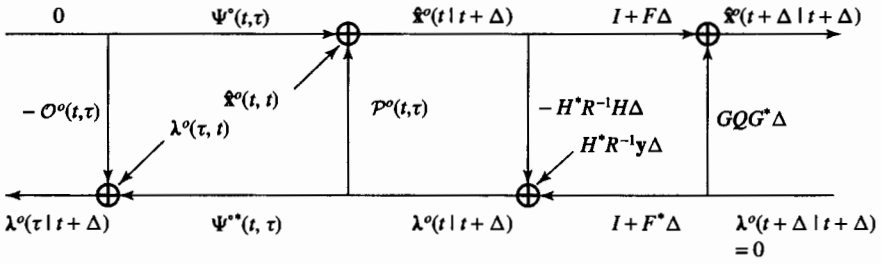


Figure 17.4 Evolution of the scattering quantities for special initial conditions $P(\tau) = 0$ and $\hat{\mathbf{x}}(\tau) = 0$. The argument $(t+\Delta)$ has been dropped for convenience from the quantities $\{F, G, H, R, Q, \mathbf{y}\}$.

from which we obtain a more familiar expression,

$$\frac{d\hat{\mathbf{x}}^o(t|t)}{dt} = F(t)\hat{\mathbf{x}}^o(t|t) + \mathcal{P}^o(t,\tau)H^*(t)R^{-1}(t)[\mathbf{y}(t) - H(t)\hat{\mathbf{x}}^o(t|t)]. \quad (17.1.15)$$

This is in fact the estimator equation of the Kalman filter, except that the variable $\mathcal{P}^o(t,\tau)$ is the solution of the Riccati differential equation (17.1.6) with zero initial condition, $\mathcal{P}^o(\tau,\tau) = 0$.

In other words, we have derived the differential equation for the state estimator as given by the Kalman filter by starting with the scattering formulation of the Hamiltonian equations for smoothed estimators. This is the reverse of the route we took earlier in Chs. 9 and 10.

But the scattering picture has more important benefits, as we shall illustrate by first studying the equations for the *backward* evolution of $\boldsymbol{\lambda}^o(s|t)$ and then all the other quantities as well. The point is that adding an incremental section to the right, *i.e.,* studying the forward evolution of the quantities in the medium, led us to familiar equations for the Kalman filter variable $\hat{\mathbf{x}}^o(t|t)$, but the one for $\boldsymbol{\lambda}^o(\tau|t)$ was unfamiliar—see (17.1.14). Since $\boldsymbol{\lambda}^o(\cdot|t)$ flows in the opposite direction from $\hat{\mathbf{x}}^o(\cdot|\cdot)$, it may be useful to study the backward evolution by adding sections to the left. In fact, this calculation, so natural to consider in the scattering framework, turns out to be very useful.

17.2 BACKWARD EVOLUTION

To study the backward evolution of $\boldsymbol{\lambda}^o(s|t)$ we add an incremental section to the left of Fig. 17.2, and continue to assume zero boundary conditions ($\hat{\mathbf{x}}(\tau) = 0$, $P(\tau) = 0$). The result is shown in Fig. 17.5, where we are denoting the leftmost reflection coefficient of the composite medium by $-\mathcal{O}^o(t,\tau-\Delta)$. We are also dropping the time argument $(\tau-\Delta)$ from all the quantities $\{F, G, H, R, Q, \mathbf{y}\}$, for convenience. From the figure we can write

$$\boldsymbol{\lambda}^o(\tau-\Delta|t) = H^*R\mathbf{y}(\tau-\Delta)\Delta + [I+F^*\Delta](I+\mathcal{O}^o(t,\tau)GQG^*\Delta)\boldsymbol{\lambda}^o(\tau|t) + o(\Delta),$$

which yields, after taking the limit as $\Delta \to 0$,

$$-\frac{\partial\boldsymbol{\lambda}^o(\tau|t)}{\partial\tau} = [F^*(\tau) - \mathcal{O}^o(t,\tau)G(\tau)Q(\tau)G^*(\tau)]\boldsymbol{\lambda}^o(\tau|t) + H^*(\tau)R^{-1}(\tau)\mathbf{y}(\tau), \quad (17.2.1)$$

with boundary condition $\boldsymbol{\lambda}^o(t|t) = 0$. Perhaps not surprisingly by now, this looks like a backwards Kalman filter! But why is $\mathcal{O}^o(t,\tau)$ there? To clarify this, note from the figure that $\mathcal{O}^o(t,\tau)$ satisfies

$$-\mathcal{O}^o(t,\tau-\Delta) = -H^*R^{-1}H\Delta - [I+F^*\Delta]\mathcal{O}^o(t,\tau)[I+F\Delta] + [I+F^*\Delta]\mathcal{O}^o(t,\tau)GQG^*\Delta\mathcal{O}^o(t,\tau)[I+F\Delta] + o(\Delta),$$

so that

$$\frac{\partial\mathcal{O}^o(t,\tau)}{\partial\tau} = -\mathcal{O}^o(t,\tau)F(\tau) - F^*(\tau)\mathcal{O}^o(t,\tau) - H^*(\tau)R^{-1}(\tau)H(\tau) + \mathcal{O}^o(t,\tau)G(\tau)Q(\tau)G^*(\tau)\mathcal{O}^o(t,\tau), \quad (17.2.2)$$

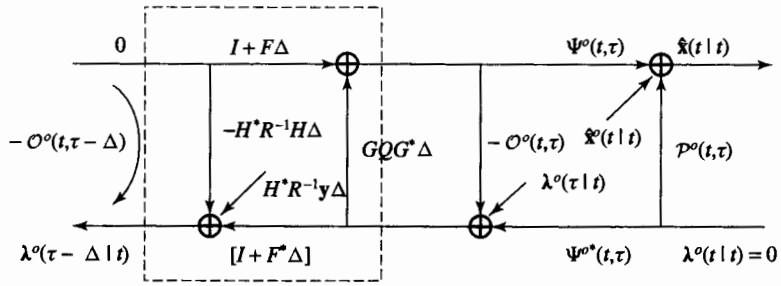


Figure 17.5 Backward evolution with zero initial conditions. The argument $(\tau-\Delta)$ has been dropped for convenience from the quantities $\{F, G, H, R, Q, \mathbf{y}\}$.

with $\mathcal{O}^o(t,t) = 0$. But this equation has exactly the same form as the backwards-time Riccati equation (16.A.18) that we encountered earlier while studying the backwards Kalman filter and the Mayne-Fraser smoothing formula (see Sec. 16.A.3). In fact, we can make the identification $\mathcal{O}^o(t,\tau) = P_\infty^{-b}(\tau)$ where the initial condition $P_\infty^{-b}(t)$ at time t in (16.A.18) is zero.

We may note also that (17.2.2) is "dual" in some sense to the Riccati equation (17.1.6) for the forward evolution of the right reflection operator! This suggests that we explore the backwards equation for the right-reflection coefficient $\mathcal{P}^o(t,\tau)$, which obeyed a forwards Riccati equation (*cf.* (17.1.6)). Again from Fig. 17.5 we can write

$$\mathcal{P}^o(t,\tau-\Delta) = \mathcal{P}^o(t,\tau) + \Psi^o(t,\tau)GQG^*\Psi^{o*}(t,\tau)\Delta + o(\Delta),$$

so that in the limit, we get (see also Prob. 17.6)

$$-\frac{\partial\mathcal{P}^o(t,\tau)}{\partial\tau} = \Psi^o(t,\tau)G(\tau)Q(\tau)G^*(\tau)\Psi^{o*}(t,\tau), \quad \mathcal{P}^o(t,t) = 0, \quad (17.2.3)$$

which should be compared with the forwards equation for $\mathcal{O}^o(t,\tau)$ in (17.1.8).

This is indeed quite striking — through some straightforward signal flow calculations in the generalized transmission line, we not only capture in a simple physical structure the key quantities in the Kalman filter theory, but we get a better insight into their meaning through, *e.g.,* (17.2.1)–(17.2.3).

A Duality between Forward and Backward Evolutions. As done earlier for the forwards evolution equation in Lemma 17.1.1, we can combine the above results to obtain the backwards evolution equation (where we are dropping the argument τ from the matrices $\{F, G, H, Q, R, K^o\}$):

$$-\frac{\partial\mathcal{S}^o(t,\tau)}{\partial\tau} = \quad (17.2.4)$$

$$\begin{bmatrix} \Psi^o(t,\tau)[F-K^oH] & \Psi^o(t,\tau)GQG^*\Psi^{o*}(t,\tau) \\ -\mathcal{O}^o(t,\tau)F - F^*\mathcal{O}^o(t,\tau) - H^*R^{-1}H + \mathcal{O}^o(t,\tau)GQG^*\mathcal{O}^o(t,\tau) & [F-K^oH]^*\Psi^{o*}(t,\tau) \end{bmatrix},$$

with boundary condition

$$\mathcal{S}^o(t,t) = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

Comparing the forward and backward equations for $\mathcal{S}^o(t,\tau)$, (17.1.11) and (17.2.4), yields some interesting identities, which can be made even clearer by introducing the following notation. Let $\Pi^{\pm}$ denote the projection matrices

$$\Pi^+ = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \quad \Pi^- = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}. \tag{17.2.5}$$

Then the forward differential equation for $\mathcal{S}^o(t,\tau)$ in Lemma 17.1.1 can be rewritten more compactly as

$$\frac{\partial \mathcal{S}^o(t,\tau)}{\partial t} = [\Pi^+ + \mathcal{S}^o(t,\tau)\Pi^-]M(t)[\Pi^+\mathcal{S}^o(t,\tau) + \Pi^-], \tag{17.2.6}$$

where $M(\cdot)$ is defined by the parameters of the incremental section,

$$M(t) \triangleq \begin{bmatrix} F(t) & G(t)Q(t)G^*(t) \\ -H^*(t)R^{-1}(t)H(t) & F^*(t) \end{bmatrix}.$$

Now we can check that the backwards evolution for $\mathcal{S}^o(t,\tau)$ can be obtained simply by interchanging the roles of Π^+ and Π^-:

$$-\frac{\partial \mathcal{S}^o(t,\tau)}{\partial \tau} = [\Pi^- + \mathcal{S}^o(t,\tau)\Pi^+]M(\tau)[\Pi^-\mathcal{S}^o(t,\tau) + \Pi^+]. \tag{17.2.7}$$

There is a simple graphical explanation for this fact, which we ask the active reader to pursue.

Remark 2. We may observe that (17.2.6) and (17.2.7) are themselves Riccati equations, which are clearly fundamental in the theory. However, we should note that the usual Riccati variable $\mathcal{P}^o(t,\tau)$ conveys a full picture only for *forward* evolution. In the forward equations (17.1.11), we note that $\mathcal{P}^o(t,\tau)$ defined by the Riccati equation (17.1.6) determines the other quantities $\{\Psi^o(t,\tau), \mathcal{O}^o(t,\tau)\}$. However, this unique role of $\mathcal{P}^o(t,\tau)$ does not carry over to the backward equations, as is evident from (17.2.4). This fact explains why the almost total emphasis in the estimation and control literature on $\mathcal{P}^o(t,\tau)$ (and its natural generalization for nonzero initial conditions), makes it more difficult to study problems where both forward and backward evolution can enter, *e.g.*, smoothing problems. Actually, the exclusive emphasis on $\mathcal{P}^o(t,\tau)$ complicates even problems involving only forward equations, as we shall illustrate in Sec. 17.5.

First however we make a small digression to introduce a notational device, called Redheffer's star product, that will allow us to make further deductions without always going through the detailed path-tracing calculations used for example to obtain (17.1.6) or (17.2.2). ♦

17.3 THE STAR PRODUCT

We here develop, following Redheffer (1962), part of a general calculus of scattering operators. As we have seen in the earlier discussion, the simplest problem in the scattering framework, but one that underlies virtually all others, is that of determining the combined scattering properties of two cascaded scattering sections, as in Fig. 17.6, from knowledge of their individual properties.

The scattering properties of a section, such as Section 1 of Fig. 17.6, may be compactly expressed by defining its associated *scattering matrix* (or operator)

$$\mathcal{S}_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

and its *source vector*

$$s_1 = \begin{bmatrix} r^+ \\ r^- \end{bmatrix}.$$

The quantities $\{a, b\}$ are the forwards and backwards transmission operators, respectively, while $\{c, d\}$ denote the left and right reflection operators.

In terms of these quantities, we can relate the waves emerging from the section to those incident on the section as follows:

$$\begin{bmatrix} k \\ l_1 \end{bmatrix} = \mathcal{S}_1 \begin{bmatrix} k_1 \\ l \end{bmatrix} + s_1.$$

Redheffer denoted the overall scattering matrix $\mathcal{S}$ that corresponds to the cascade of two such sections (see Fig. 17.6) as

$$\mathcal{S} = \mathcal{S}_1 \star \mathcal{S}_2 \triangleq \text{the } \textit{star product} \text{ of } \mathcal{S}_1 \text{ and } \mathcal{S}_2.$$

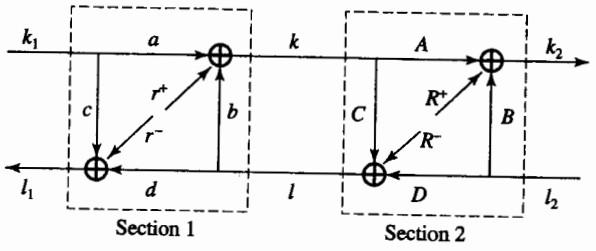


Figure 17.6 Two cascaded scattering layers.

Using elementary flow-graph algebra, or by solving the obvious linear equations, it can readily be seen from Fig. 17.6 that the star product is given by (assuming the required inverses exist)

$$\mathcal{S} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \star \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$
$$= \begin{bmatrix} A(I - bC)^{-1}a & B + Ab(I - Cb)^{-1}D \\ c + dC(I - bC)^{-1}a & d(I - Cb)^{-1}D \end{bmatrix}. \tag{17.3.1}$$

For example, let us compute the $(1,1)$ entry of $\mathcal{S}$. We can do this by assuming zero source vectors and that $l_2 = 0$. Then the output k_2 can be found as

$$k_2 = A[ak_1 + bCak_1 + bCbCak_1 + \ldots] = A(I - bc)^{-1}ak_1.$$

Similarly, we can write for the overall source vector s as $s = s_1 \bullet s_2$, and term it the *dot sum* of s_1 and s_2. Again, reference to Fig. 17.6 easily shows that

$$s = \begin{bmatrix} r^+ \\ r^- \end{bmatrix} \bullet \begin{bmatrix} R^+ \\ R^- \end{bmatrix},$$
$$= \begin{bmatrix} R^+ \\ r^- \end{bmatrix} + \left(\begin{bmatrix} I & b \\ 0 & d \end{bmatrix} \star \begin{bmatrix} A & 0 \\ C & I \end{bmatrix} \right) \begin{bmatrix} r^+ \\ R^- \end{bmatrix},$$
$$= \begin{bmatrix} R^+ \\ r^- \end{bmatrix} + \begin{bmatrix} A(I - bC)^{-1}(r^+ + bR^-) \\ d(I - Cb)^{-1}(R^- + Cr^+) \end{bmatrix}. \tag{17.3.2}$$

A very striking and very useful fact is the following, whose proof we leave to the reader. If $\mathcal{S}$ is invertible, then

$$\mathcal{S}^{-1} \star \mathcal{S} = I = \mathcal{S} \star \mathcal{S}^{-1}. \tag{17.3.3}$$

That is, the ordinary matrix inverse is also the star-product inverse. Therefore if, for example, $\mathcal{S}_1 = \mathcal{T}_1 \star \mathcal{S}$ and $\mathcal{S}_2 = \mathcal{T}_2 \star \mathcal{S}$, then we could compute $\mathcal{S}_2$ from $\mathcal{S}_1$ via $\mathcal{S}_2 = \mathcal{T}_2 \star \mathcal{T}_1^{-1} \star \mathcal{S}_1$. This result has many useful applications, *e.g.*, it allows us to easily derive general change-of-initial condition formulas for Riccati equations — see Sec. 17.4.3.

17.3.1 Evolution Equations

To briefly examine a calculus for such operators, consider a scattering matrix $\mathcal{S}^o$ such that

$$\mathcal{S}^o(t,\tau) = \begin{bmatrix} a^o(t,\tau) & \rho^o(t,\tau) \\ r^o(t,\tau) & \alpha^o(t,\tau) \end{bmatrix}, \quad \text{with} \quad \mathcal{S}^o(\tau,\tau) = I,$$

and define the *infinitesimal generator*

$$M(t) \triangleq \lim_{\Delta \to 0} \frac{\mathcal{S}^o(t+\Delta,t) - \mathcal{S}^o(t,t)}{\Delta} = \begin{bmatrix} f(t) & g(t) \\ h(t) & e(t) \end{bmatrix}, \quad \text{say.} \tag{17.3.4}$$

The quantities $\{f, g, h, e\}$ are known as the *medium parameters* and knowledge of these local parameters suffice to compute the global scattering matrix $\mathcal{S}^o$ via certain differential equations. To see this, we note first that

$$\mathcal{S}^o(t+\Delta,\tau) = \mathcal{S}^o(t,\tau) \star \mathcal{S}^o(t+\Delta,t) = \mathcal{S}^o(t,\tau) \star [I + M(t)\Delta + o(\Delta)]. \tag{17.3.5}$$

Then some simple algebra will lead to the result (we are dropping the arguments (t,τ) from $\{a^o, r^o, \rho^o, \alpha^o\}$)

$$\frac{\partial \mathcal{S}^o(t,\tau)}{\partial t} = \begin{bmatrix} [f(t) + \rho^o h(t)]a^o & g(t) + f(t)\rho^o + \rho^o e(t) + \rho^o h(t)\rho^o \\ \alpha^o h(t) a^o & \alpha^o[e(t) + h(t)\rho^o] \end{bmatrix}.$$

This can be written more compactly as

$$\frac{\partial \mathcal{S}^o(t,\tau)}{\partial t} = [\Pi^+ + \mathcal{S}^o(t,\tau)\Pi^-]M(t)[\Pi^- + \Pi^+\mathcal{S}^o(t,\tau)], \tag{17.3.6}$$

where $\Pi^\pm$ are the projection operators (17.2.5). Note that ρ^o obeys a Riccati differential equation and that so does $\mathcal{S}^o$ itself.

For backward evolution, we begin with

$$\mathcal{S}^o(t,\tau-\Delta) = \mathcal{S}^o(\tau,\tau-\Delta) \star \mathcal{S}^o(t,\tau),$$
$$= [I + M(\tau-\Delta)\Delta + o(\Delta)] \star \mathcal{S}^o(t,\tau),$$
$$= [I + M(\tau)\Delta + o(\Delta)] \star \mathcal{S}^o(t,\tau). \tag{17.3.7}$$

Now some simple algebra leads to the "dual" equation

$$-\frac{\partial \mathcal{S}^o(t,\tau)}{\partial \tau} = \begin{bmatrix} a^o[f(\tau) + g(\tau)r^o] & a^o g(\tau)\alpha^o \\ h(\tau) + e(\tau)r^o + r^o f(\tau) + r^o g(\tau) r^o & [e(\tau) + r^o g(\tau)]\alpha^o \end{bmatrix}, \tag{17.3.8}$$

or, more compactly,

$$-\frac{\partial \mathcal{S}^o(t,\tau)}{\partial \tau} = [\Pi^- + \mathcal{S}^o(t,\tau)\Pi^+]M(\tau)[\Pi^+ + \Pi^-\mathcal{S}^o(t,\tau)]. \tag{17.3.9}$$

Note that now it is r^o and $\mathcal{S}^o$ that obey backward Riccati equations. Note that the backward equation is obtained from the forward equation by replacing $\Pi^\pm$ by $\Pi^\mp$.

17.3.2 General Initial Conditions

With the star-product notation, we can now show how to go beyond our assumption of zero initial conditions. This is almost trivial for the forward equations. Indeed, the scattering matrix $\mathcal{S}^o(t,\tau)$ has initial value $\mathcal{S}^o(\tau,\tau) = I$. Now introduce

$$\Gamma(\tau) = \begin{bmatrix} \Pi_1(\tau) & \Pi_2(\tau) \\ \Pi_3(\tau) & \Pi_4(\tau) \end{bmatrix},$$

and define

$$\mathcal{S}(t,\tau) = \Gamma(\tau) \star \mathcal{S}^o(t,\tau) \triangleq \begin{bmatrix} a(t,\tau) & \rho(t,\tau) \\ r(t,\tau) & \alpha(t,\tau) \end{bmatrix}, \quad \text{say.}$$

Then the new scattering matrix $\mathcal{S}$ has initial value $\Gamma(\tau)$. Moreover, $\mathcal{S}$ obeys the same forward differential equation as $\mathcal{S}^o$. More specifically, note in view of (17.3.5) that

$$\begin{aligned} \mathcal{S}(t+\Delta,\tau) &= \Gamma(\tau) \star \mathcal{S}^o(t+\Delta,\tau) \\ &= \Gamma(\tau) \star \mathcal{S}^o(t,\tau) \star [I + M(t)\Delta + o(\Delta)] \\ &= \mathcal{S}(t,\tau) \star [I + M(t)\Delta + o(\Delta)], \end{aligned}$$

which shows that $\mathcal{S}$ obeys the equation

$$\frac{\partial \mathcal{S}(t,\tau)}{\partial t} = [\Pi^+ + \mathcal{S}(t,\tau)\Pi^-]M(t)[\Pi^- + \Pi^+\mathcal{S}(t,\tau)], \quad \mathcal{S}(\tau,\tau) = \Gamma(\tau). \tag{17.3.10}$$

For the backward evolution, however, things are less obvious[1] because the boundary layer $\Gamma(\tau)$ has to be moved as we move to the left and, consequently, $\mathcal{S}(t,\tau)$ does *not* obey the same backward differential equation as $\mathcal{S}^o(t,\tau)$. The proper equation can nevertheless be calculated as follows. We write, using (17.3.7),

$$\begin{aligned} \mathcal{S}(t,\tau-\Delta) &= \Gamma(\tau-\Delta) \star \mathcal{S}^o(t,\tau-\Delta), \\ &= \Gamma(\tau-\Delta) \star \mathcal{S}^o(\tau,\tau-\Delta) \star \mathcal{S}^o(t,\tau), \\ &= \Gamma(\tau-\Delta) \star [I + M(\tau)\Delta + o(\Delta)] \star \Gamma^{-1}(\tau) \star \Gamma(\tau) \star \mathcal{S}^o(t,\tau). \end{aligned}$$

Now assume the existence of a generator matrix $N(\tau)$ implicitly defined by the equation (we will address the existence issue shortly)

$$\Gamma(\tau-\Delta) \star [I + M(\tau)\Delta + o(\Delta)] = [I + N(\tau)\Delta + o(\Delta)] \star \Gamma(\tau). \tag{17.3.11}$$

Note that when $\Gamma(\cdot) = I$, then we can choose $N(\cdot) = M(\cdot)$. Now we can write

$$\mathcal{S}(t,\tau-\Delta) = [I + N(\tau)\Delta + o(\Delta)] \star \mathcal{S}(t,\tau), \tag{17.3.12}$$

[1] The reader may skip the rest of this section, which involves some detailed algebra; this material will not be used until we come to Sec. 17.5.2 on Stokes identities for homogeneous media with nonzero boundary conditions. The issues here become much simpler in discrete time (see Sec. 17.6.1).

and therefrom obtain the equation

$$-\frac{\partial \mathcal{S}(t,\tau)}{\partial \tau} = [\Pi^- + \mathcal{S}(t,\tau)\Pi^+]N(\tau)[\Pi^+ + \Pi^-\mathcal{S}(t,\tau)]. \tag{17.3.13}$$

So when we have nonzero initial conditions for the (left and right) reflection operators, to go from the forwards to the backwards equations, we not only have to interchange Π^+ and Π^-, but we will also have to use different generator matrices, $M(\cdot)$ and $N(\cdot)$. It remains to discuss the existence and form of $N(\cdot)$. Let us denote (*cf.* (17.3.4)),

$$N(\tau) = \begin{bmatrix} f_N(\tau) & g_N(\tau) \\ h_N(\tau) & e_N(\tau) \end{bmatrix}.$$

The complication is the fact that the exact form of N depends upon our choice for $\Gamma(\tau)$. Let us proceed by writing

$$\Gamma(\tau-\Delta) = \Gamma(\tau) - \frac{d}{d\tau}\Gamma(\tau) \cdot \Delta + o(\Delta).$$

Inserting this into equation (17.3.11), and comparing coefficients of Δ on both sides, will lead to the following set of equations (we are dropping the argument τ from all quantities):

$$\begin{aligned} \Pi_1 g_N \Pi_3 + \Pi_1 f_N &= \Pi_2 h \Pi_1 + f\Pi_1 - \dot{\Pi}_1, \\ \Pi_1 g_N \Pi_4 &= g + f\Pi_2 + \Pi_2 e - \Pi_2 h \Pi_2 - \dot{\Pi}_2, \\ h_N + e_N \Pi_3 + \Pi_2 f_N + \Pi_3 g_N \Pi_3 &= \Pi_4 h \Pi_1 - \dot{\Pi}_3, \\ e_N \Pi_4 + \Pi_3 g_N \Pi_4 &= \Pi_4 h \Pi_2 + \Pi_4 e - \dot{\Pi}_4. \end{aligned}$$

A compact way of writing these equations is

$$\begin{bmatrix} \Pi_1 & 0 \\ \Pi_3 & I \end{bmatrix} N(\tau) \begin{bmatrix} I & 0 \\ \Pi_2 & \Pi_4 \end{bmatrix} = \begin{bmatrix} I & \Pi_2 \\ 0 & \Pi_4 \end{bmatrix} M(\tau) \begin{bmatrix} \Pi_1 & \Pi_2 \\ 0 & I \end{bmatrix} - \frac{d}{d\tau}\Gamma(\tau), \tag{17.3.14}$$

from which we can see that the invertibility of $\{\Pi_1, \Pi_4\}$ will guarantee a unique solution for N, *viz.*,

$$N(\tau) = \tag{17.3.15}$$

$$\begin{bmatrix} \Pi_1^{-1} & 0 \\ -\Pi_3\Pi_1^{-1} & I \end{bmatrix} \left\{ \begin{bmatrix} I & \Pi_2 \\ 0 & \Pi_4 \end{bmatrix} M(\tau) \begin{bmatrix} \Pi_1 & \Pi_2 \\ 0 & I \end{bmatrix} - \frac{d}{d\tau}\Gamma(\tau) \right\} \begin{bmatrix} I & 0 \\ -\Pi_4^{-1}\Pi_2 & \Pi_4^{-1} \end{bmatrix}.$$

If $\Gamma(\tau) = I$, then the backwards equation for $\mathcal{S}$ is just (17.3.7) and we again see that $N = M$. Another important case (see Sec. 17.4.4) is when

$$\Gamma(\tau) = \begin{bmatrix} \Pi(\tau) & \Pi(\tau) \\ \Pi(\tau) & \Pi(\tau) \end{bmatrix},$$

where $\Pi(\tau)$ satisfies the differential equation

$$\frac{d}{d\tau}\Pi(\tau) = f(\tau)\Pi(\tau) + \Pi(\tau)e(\tau) + g(\tau). \tag{17.3.16}$$

In this case, the entries of $N(\tau)$ are readily found to be

$$\begin{aligned} f_N(\tau) &= -e(\tau) - \Pi^{-1}(\tau)g(\tau), \\ e_N(\tau) &= -f(\tau) - g(\tau)\Pi^{-1}(\tau), \\ g_N(\tau) &= h(\tau), \\ h_N(\tau) &= g(\tau). \end{aligned} \tag{17.3.17}$$

A third important case, needed in Sec. 17.5.2, is when

$$\Gamma(\tau) = \begin{bmatrix} I & \Pi(\tau) \\ 0 & I \end{bmatrix}.$$

17.3.3 Chain Scattering or Transmission Matrices

The star-product calculus is certainly an interesting one, but we could widen its scope considerably if by some transformation, say $\hat{S} = f(S)$, we can arrange that

$$f(S_1 \star S_2) = f(S_1)f(S_2),$$

where the product on the right-hand side is the usual matrix product. One of several consequences is that a lot of classical matrix theory (*e.g.*, the use of orthogonal transformations to triangularize and diagonalize matrices) can be carried over to the scattering context.

There are several ways of achieving such transformations, but a particularly simple one is indicated in Fig. 17.7. The reason for the special composition rule for scattering sections is that we have (incident and reflected) waves going forwards and backwards. To go back to the usual matrix composition rule, we can reverse the flow in the lower path of the leftmost figure. It is easy to check that this can be done by inverting the gain and changing the sign of flows into the path — a well-known (Mason's) rule for signal flow graphs. This gives the section shown in the rightmost figure, which is characterized

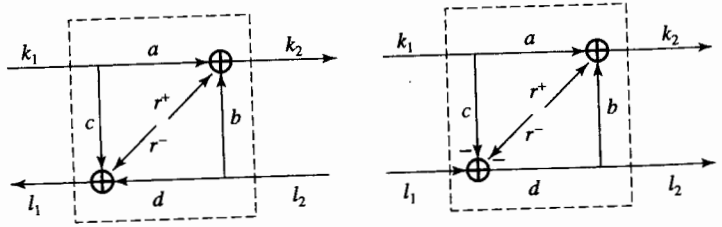


Figure 17.7 Scattering and transmission operators.

by a so-called *chain scattering* or *transmission matrix* (taking input variables on the left to output variables on the right), *viz.*,

$$\begin{bmatrix} k_2 \\ l_2 \end{bmatrix} = \hat{S}\begin{bmatrix} k_1 \\ l_1 \end{bmatrix} + \begin{bmatrix} r^+ - bd^{-1}r^- \\ -d^{-1}r^- \end{bmatrix},$$

where

$$\hat{S} = \begin{bmatrix} a - bd^{-1}c & bd^{-1} \\ -d^{-1}c & d^{-1} \end{bmatrix}.$$

To make this transformation, we must of course assume that d^{-1} exists. [With obvious changes we could adapt to the case where a^{-1} exists.] The mapping from S to $\hat{S}$ is often called an exchange mapping, and we may note that iterating it gives us back the original matrix, *i.e.*,

$$\widehat{(\hat{S})} = S.$$

The property of interest to us is that if $S = S_1 \star S_2$, then $\hat{S} = \hat{S}_1\hat{S}_2$ (in terms of the usual matrix product).

We can readily obtain the analogs of the scattering matrix evolution equations for the transmission matrices, and also several other results. Here we shall only mention some results connected with orthogonal matrices. If Q is unitary, *i.e.*, $QQ^* = I$, then we have

$$\hat{Q} \star \widehat{(Q^*)} = I.$$

Note that $\widehat{(Q^*)} \neq \left(\hat{Q}\right)^*$. A closer look at the exchange mapping shows that

$$QQ^* = I \quad \Longleftrightarrow \quad \hat{Q} \star J \star \left(\hat{Q}\right)^* = J,$$

where $J = (I \oplus -I)$. Similarly, consider the matrix decomposition $AQ = R$, with Q unitary and A upper triangular. The scattering version, $\hat{A} \star \hat{Q} = \hat{R}$, says that given a scattering section $\hat{A}$, there is a J-unitary scattering section section that can remove a reflection from one side. Such results can be used to obtain the square-root estimation algorithms and square-root doubling formulas (see Newkirk (1979)).

17.4 VARIOUS RICCATI FORMULAS

In this section, and the next, we present several results that are more easily obtained by resorting to the scattering framework than through more direct routes.

17.4.1 Incorporating Boundary Conditions

We return to the scattering medium of Fig. 17.2 and recall from the discussion in Secs. 17.1.1 and 17.1.2 that the scattering parameters $\{\mathcal{P}^o, \Psi^o, \mathcal{O}^o\}$, as well as the internal sources $\{\hat{\mathbf{x}}^o(t|t), \boldsymbol{\lambda}^o(\tau|t)\}$, were all defined assuming zero boundary conditions, $\hat{\mathbf{x}}(\tau) = 0$ and $P(\tau) = 0$.

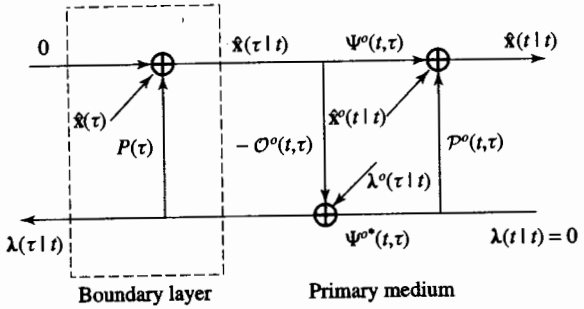


Figure 17.8 Cascade of a boundary layer and the primary medium.

In general, however, these boundary conditions are not zero and are coupled (see (17.1.5)):

$$\hat{\mathbf{x}}(\tau|t) = \hat{\mathbf{x}}(\tau) + P(\tau)\boldsymbol{\lambda}(\tau|t). \tag{17.4.1}$$

This relation can be nicely incorporated into the scattering description by attaching a simple boundary section to the left of the macroscopic section of Fig. 17.2, as shown in Fig. 17.8. We shall denote the parameters and the internal sources of the resulting combined layer by (we now drop the superscript o)

$$\{\Psi(t,\tau), \mathcal{P}(t,\tau), -\mathcal{O}(t,\tau), r^+(t,\tau), r^-(t,\tau)\}.$$

That is,

$$\mathcal{S}(t,\tau) \triangleq \begin{bmatrix} \Psi(t,\tau) & \mathcal{P}(t,\tau) \\ -\mathcal{O}(t,\tau) & \Psi^*(t,\tau) \end{bmatrix} = \begin{bmatrix} I & P(\tau) \\ 0 & I \end{bmatrix} \star \begin{bmatrix} \Psi^o(t,\tau) & \mathcal{P}^o(t,\tau) \\ -\mathcal{O}^o(t,\tau) & \Psi^{o*}(t,\tau) \end{bmatrix},$$

$$\begin{bmatrix} r^+(t,\tau) \\ r^-(t,\tau) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}(\tau) \\ 0 \end{bmatrix} \bullet \begin{bmatrix} \hat{\mathbf{x}}^o(t|t) \\ \boldsymbol{\lambda}^o(\tau|t) \end{bmatrix}.$$

The result of incorporating the boundary layer is depicted in Fig. 17.9. We note at once that the internal sources $\{r^+(t,\tau), r^-(t,\tau)\}$ can be identified as

$$r^+(t,\tau) = \hat{\mathbf{x}}(t|t), \qquad r^-(t,\tau) = \boldsymbol{\lambda}(\tau|t). \tag{17.4.2}$$

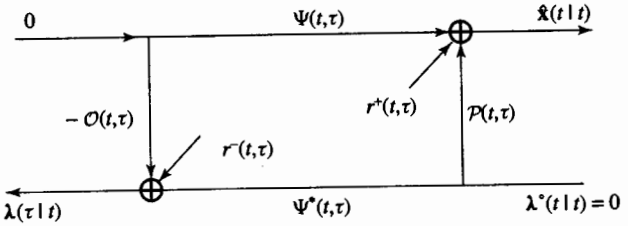


Figure 17.9 The composite layer.

As for $\mathcal{S}(t,\tau)$, note that it has initial value

$$\mathcal{S}(\tau,\tau) = \begin{bmatrix} I & P(\tau) \\ 0 & I \end{bmatrix},$$

while $\mathcal{S}^o(\tau,\tau) = I \oplus I$. Moreover, by repeating a simple argument already used before (at the beginning of Sec. 17.3.2), we can see that $\mathcal{S}(t,\tau)$ obeys the same forward differential equation as $\mathcal{S}^o(t,\tau)$, but with a different initial condition. Therefore, the entries $\{\mathcal{P}(t,\tau), \Psi(t,\tau), \mathcal{O}(t,\tau)\}$ obey the same forward differential equations as $\{\mathcal{P}^o(t,\tau), \Psi^o(t,\tau), \mathcal{O}^o(t,\tau)\}$, but with different initial conditions. That is, (*cf.* (17.1.6)–(17.1.9)),

$$\frac{d}{dt}\mathcal{P}(t,\tau) = G(t)Q(t)G^*(t) + F(t)\mathcal{P}(t,\tau) + \mathcal{P}(t,\tau)F^*(t) - \mathcal{P}(t,\tau)H^*(t)R^{-1}(t)H(t)\mathcal{P}(t,\tau), \tag{17.4.3}$$

$$\frac{d}{dt}\Psi(t,\tau) = [F(t) - K(t)H(t)]\Psi(t,\tau), \tag{17.4.4}$$

$$\frac{d}{dt}\mathcal{O}(t,\tau) = \Psi^*(t,\tau)H^*(t)R^{-1}(t)H(t)\Psi(t,\tau), \qquad \mathcal{O}(\tau,\tau) = 0, \tag{17.4.5}$$

with $\mathcal{P}(\tau,\tau) = P(\tau)$ and $\Psi(\tau,\tau) = I$. Moreover,

$$K(t) \triangleq K(t,\tau) = \mathcal{P}(t,\tau)H^*(t)R^{-1}(t).$$

[Note again that $K(\cdot)$ is also a function of τ.]

17.4.2 Partitioned Formulas

For various reasons, it may be necessary to relate estimators obtained using different initial conditions. Not all such relations are easy to obtain directly. We showed in Sec. 16.7.2 how to relate the solutions of Riccati differential equations for different initial conditions. This is even easier in the scattering formulation. For example, we can use the star product (17.4.2) to relate not only $\mathcal{P}(t,\tau)$ and $\mathcal{P}^o(t,\tau)$, but also $\{\mathcal{O}(t,\tau), \mathcal{O}^o(t,\tau)\}$ and $\{\Psi(t,\tau), \Psi^o(t,\tau)\}$.

Actually, it is instructive to do this directly from the scattering diagrams. Thus, by tracing paths in Fig. 17.8, and assuming the required inverse exists, we obtain

$$\begin{aligned}\mathcal{P}(t,\tau) &= \mathcal{P}^o(t,\tau) + \\ &\quad \Psi^o(t,\tau)P(\tau)\,[I - \mathcal{O}^o(t,\tau)P(\tau) + \mathcal{O}^o(t,\tau)P(\tau)\mathcal{O}^o(t,\tau)P(\tau) + \ldots]\,\Psi^{o*}(t,\tau), \\ &= \mathcal{P}^o(t,\tau) + \Psi^o(t,\tau)P(\tau)[I + \mathcal{O}^o(t,\tau)P(\tau)]^{-1}\Psi^{o*}(t,\tau).\end{aligned}$$

This is exactly the same as (16.7.2).

To relate $\hat{\mathbf{x}}(\tau|t)$ and $\hat{\mathbf{x}}^o(\tau|t)$ is much harder to do algebraically (see, *e.g.*, Lainiotis (1974)), but from Fig. 17.8 we can immediately write that

$$\hat{\mathbf{x}}(t|t) = \hat{\mathbf{x}}^o(t|t) + \Psi^o(t,\tau)\hat{\mathbf{x}}(\tau|t).$$

In fact, we can also write

$$\boldsymbol{\lambda}(\tau|t) = \boldsymbol{\lambda}^o(\tau|t) - \mathcal{O}^o(t,\tau)\hat{\mathbf{x}}(\tau|t).$$

Combining this with the (Hamiltonian) boundary condition (17.4.1) gives the Lainiotis smoothing formula

$$\hat{\mathbf{x}}(\tau|t) = [I + P(\tau)\mathcal{O}^o(t,\tau)]^{-1}[\hat{\mathbf{x}}(\tau) + P(\tau)\boldsymbol{\lambda}^o(\tau|t)]. \tag{17.4.6}$$

We shall show later (Sec. 17.4.4) that this is a forwards-time counterpart of the Mayne-Fraser two-filter smoothing formula. First however, more on changing initial conditions.

17.4.3 General Changes in the Boundary Conditions

We can use the star product to readily obtain more general results. So assume we have obtained the scattering matrix $\mathcal{S}(t,\tau)$ and the source vector $s(t,\tau)$,

$$\mathcal{S}(t,\tau) = \begin{bmatrix} \Psi(t,\tau) & \mathcal{P}(t,\tau) \\ -\mathcal{O}(t,\tau) & \Psi^*(t,\tau) \end{bmatrix}, \qquad s(t,\tau) = \begin{bmatrix} \hat{\mathbf{x}}(t|t) \\ \boldsymbol{\lambda}(\tau|t) \end{bmatrix},$$

for the medium using a boundary condition $P(\tau)$. Now we change the boundary condition on $\mathcal{P}(t,\tau)$ to $P_1(\tau)$ and let $\mathcal{S}_1(t,\tau)$ and $s_1(t,\tau)$ denote the new scattering matrix and the new source vector, say with entries defined by

$$\mathcal{S}_1(t,\tau) = \begin{bmatrix} \Psi_1(t,\tau) & \mathcal{P}_1(t,\tau) \\ -\mathcal{O}_1(t,\tau) & \Psi_1^*(t,\tau) \end{bmatrix}, \qquad s_1(t,\tau) = \begin{bmatrix} \hat{\mathbf{x}}_1(t|t) \\ \boldsymbol{\lambda}_1(\tau|t) \end{bmatrix}.$$

We want to relate the quantities with subscript $_1$ to those without subscript.

Fig. 17.10 shows that placing a layer with scattering matrix (*cf.* the remark below (17.3.3))

$$\mathcal{S}_\delta = \begin{bmatrix} I & P_1(\tau) \\ 0 & I \end{bmatrix} \star \begin{bmatrix} I & P(\tau) \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} I & P_1(\tau) \\ 0 & I \end{bmatrix} \star \begin{bmatrix} I & -P(\tau) \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & \delta P \\ 0 & I \end{bmatrix}$$

where $\delta P = P_1(\tau) - P(\tau)$, and with source vector

$$s_\delta = \begin{bmatrix} \delta\hat{\mathbf{x}} \\ 0 \end{bmatrix}, \qquad \delta\hat{\mathbf{x}} = \mathbf{x}_1(\tau) - \hat{\mathbf{x}}(\tau),$$

to the left of the configuration in Fig. 17.8 effectively changes the initial conditions to $\hat{\mathbf{x}}_1(\tau)$ and $P_1(\tau)$. Fig. 17.11 shows the overall configuration that is obtained by adding the section $\{\mathcal{S}_\delta, s_\delta\}$ to the left of Fig. 17.8.

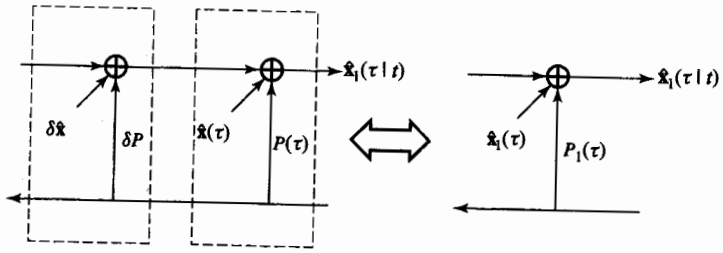


Figure 17.10 Change of initial conditions.

Applying the star-product rules (17.3.1) and (17.3.2) to the relations

$$\mathcal{S}_1(t,\tau) = \mathcal{S}_\delta \star \mathcal{S}(t,\tau), \qquad s_1(t,\tau) = s_\delta \bullet s(t,\tau),$$

we readily obtain that

$$\begin{bmatrix} \Psi_1(t,\tau) & \mathcal{P}_1(t,\tau) \\ -\mathcal{O}_1(t,\tau) & \Psi_1^*(t,\tau) \end{bmatrix} = \tag{17.4.7}$$

$$\begin{bmatrix} \Psi(t,\tau)[I + \delta P\mathcal{O}(t,\tau)]^{-1} & \mathcal{P}(t,\tau) + \Psi(t,\tau)\delta P[I + \mathcal{O}(t,\tau)\delta P]^{-1}\Psi^*(t,\tau) \\ -\mathcal{O}(t,\tau)[I + \delta P\mathcal{O}(t,\tau)]^{-1} & [I + \mathcal{O}(t,\tau)\delta P]^{-1}\Psi^*(t,\tau) \end{bmatrix},$$

and

$$\begin{bmatrix} \hat{\mathbf{x}}_1(t|t) \\ \boldsymbol{\lambda}_1(\tau|t) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}(t|t) + \Psi(t,\tau)[I + \delta P\mathcal{O}(t,\tau)]^{-1}[\delta\hat{\mathbf{x}} + \delta P\boldsymbol{\lambda}(\tau|t)] \\ [I + \mathcal{O}(t,\tau)\delta P]^{-1}[\boldsymbol{\lambda}(\tau|t) - \mathcal{O}(t,\tau)\delta\hat{\mathbf{x}}] \end{bmatrix}. \tag{17.4.8}$$

In summary, we have established the following result.

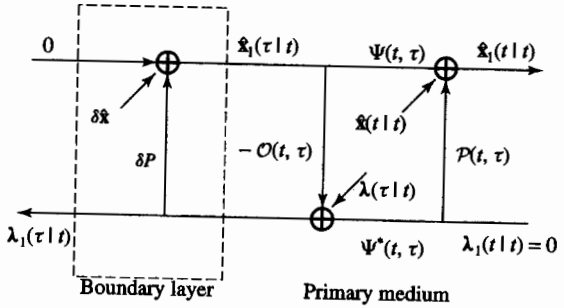


Figure 17.11 Changing initial conditions by adjoining a special boundary layer.

Theorem 17.4.1 (Changes in Initial Conditions) *Consider the state-space model (17.1.1)–(17.1.2) and let*

$$\{\hat{\mathbf{x}}(t|t), \hat{\mathbf{x}}(\tau|t), \boldsymbol{\lambda}(\tau|t), \mathcal{P}(t,\tau), \Psi(t,\tau), \mathcal{O}(t,\tau)\}$$

denote the variables that arise in the solution of the filtering and fixed-interval smoothing problems with initial conditions $\mathcal{P}(\tau,\tau) = P(\tau)$ and $\hat{\mathbf{x}}(\tau)$. Now assume the initial covariance matrix is changed to $P_1(\tau)$ and the initial vector estimator is also changed to $\hat{\mathbf{x}}_1(\tau)$. Let

$$\{\hat{\mathbf{x}}_1(t|t), \hat{\mathbf{x}}_1(\tau|t), \boldsymbol{\lambda}_1(\tau|t), \mathcal{P}_1(t,\tau), \Psi_1(t,\tau), \mathcal{O}_1(t,\tau)\}$$

denote the corresponding variables that arise by solving the same filtering and fixed-interval smoothing problems but with the initial conditions $P_1(\tau)$ and $\hat{\mathbf{x}}_1(\tau)$. Then at any time instant t, the following relations hold (assuming the required inverses exist):

$$\Psi_1(t,\tau) = \Psi(t,\tau)[I + \delta P(\tau)\mathcal{O}(t,\tau)]^{-1}, \tag{17.4.9}$$

$$\mathcal{P}_1(t,\tau) = \mathcal{P}(t,\tau) + \Psi(t,\tau)\delta P(\tau)[I + \mathcal{O}(t,\tau)\delta P(\tau)]^{-1}\Psi^*(t,\tau), \tag{17.4.10}$$

$$\mathcal{O}_1(t,\tau) = \mathcal{O}(t,\tau)[I + \delta P(\tau)\mathcal{O}(t,\tau)]^{-1}, \tag{17.4.11}$$

as well as

$$\hat{\mathbf{x}}_1(t|t) = \hat{\mathbf{x}}(t|t) + \Psi(t,\tau)[I + \delta P(\tau)\mathcal{O}(t,\tau)]^{-1}[\delta\hat{\mathbf{x}}(\tau) + \delta P(\tau)\boldsymbol{\lambda}(\tau|t)],$$

$$\boldsymbol{\lambda}_1(\tau|t) = [I + \mathcal{O}(t,\tau)\delta P(\tau)]^{-1}[\boldsymbol{\lambda}(\tau|t) - \mathcal{O}(t,\tau)\delta\hat{\mathbf{x}}(\tau)],$$

where

$$\delta P(\tau) = P_1(\tau) - P(\tau), \qquad \delta\hat{\mathbf{x}}(\tau) = \hat{\mathbf{x}}_1(\tau) - \hat{\mathbf{x}}(\tau). \tag{17.4.12}$$

■

The reader can make up several variants of this result. Here we return to our promised further examination of the smoothing formula (17.4.6).

17.4.4 Smoothing as an Extended Filtering Problem

The scattering matrix $\mathcal{S}^o(t,\tau)$ admits a stochastic interpretation that will allow us to apply the scattering formulation to obtain a very powerful approach to the smoothing problem that can yield all possible smoothing formulas (by following the discussion in Ljung and Kailath (1976a)). We start with a method originally suggested by Zachrisson (1969) and Willman (1969) to convert a smoothing problem to a filtering problem. Fixing the time τ, we define for any $\tau \le s \le t$, the extended vector

$$\mathbf{z}(s,\tau) \triangleq \begin{bmatrix} \mathbf{x}(s) \\ \mathbf{x}(\tau) \end{bmatrix},$$

where $\mathbf{x}(s)$ satisfies the state equation (17.1.1). Then we can write

$$\frac{d}{ds}\mathbf{z}(s,\tau) = \mathcal{F}(s)\mathbf{z}(s,\tau) + \mathcal{G}(s)\mathbf{u}(s),$$

$$\mathbf{y}(s) = \mathcal{H}(s)\mathbf{z}(s,\tau) + \mathbf{v}(s),$$

where

$$\mathcal{F}(s) = \begin{bmatrix} F(s) & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathcal{G}(s) = \begin{bmatrix} G(s) \\ 0 \end{bmatrix}, \quad \mathcal{H}(s) = \begin{bmatrix} H(s) & 0 \end{bmatrix}.$$

The filtered estimator of $\mathbf{z}(s,\tau)$, given $\{\mathbf{y}(s), 0 \le s \le t\}$, can now be found by the Kalman filter formulas of Thm. 16.2.1 as

$$\frac{d}{ds}\mathbf{z}(s,\tau) = [\mathcal{F}(s) - \mathcal{P}_z(s,t)\mathcal{H}^*(s)R^{-1}(s)\mathcal{H}(s)]\mathbf{z}(s,\tau) + \mathcal{P}_z(t,s)\mathcal{H}^*(s)R^{-1}(s)\mathbf{y}(s),$$

where $\mathcal{P}_z(s,t)$ obeys the Riccati equation

$$\begin{aligned}\mathcal{P}_z(s,t) &= \mathcal{F}(s)\mathcal{P}_z(s,t) + \mathcal{P}_z(s,t)\mathcal{F}^*(s) \\ &\quad + \mathcal{G}(s)\mathcal{Q}(s)\mathcal{G}^*(s) - \mathcal{P}_z(s,t)\mathcal{H}^*(s)R^{-1}\mathcal{H}(s)\mathcal{P}_z(s,t),\end{aligned} \tag{17.4.13}$$

with initial conditions

$$\hat{\mathbf{z}}(\tau,\tau) = \begin{bmatrix} \hat{\mathbf{x}}(\tau) \\ \hat{\mathbf{x}}(\tau) \end{bmatrix}, \quad \mathcal{P}_z(\tau,\tau) = \begin{bmatrix} P(\tau) & P(\tau) \\ P(\tau) & P(\tau) \end{bmatrix} \triangleq \bar{P}. \tag{17.4.14}$$

Note now that

$$\hat{\mathbf{z}}(t,\tau) = \begin{bmatrix} \hat{\mathbf{x}}(t|t) \\ \hat{\mathbf{x}}(\tau|t) \end{bmatrix},$$

so that we here have formulas for both the filtered estimator $\hat{\mathbf{x}}(t|t)$ and the smoothed estimator $\hat{\mathbf{x}}(\tau|t)$. Let $P(t)$ and $P(\tau|t)$ denote the error covariance matrices of the filtered error $\tilde{\mathbf{x}}(t|t)$ and the smoothed error $\tilde{\mathbf{x}}(\tau|t)$, respectively (*cf.* for example, Thms. 16.2.1 and 16.5.1), where the initial time instant is taken to be τ and the initial error variance is $P(\tau)$).

We now denote the entries of $\mathcal{P}_z(t,\tau)$ by

$$\mathcal{P}_z(t,\tau) \triangleq \begin{bmatrix} P_{11}(t,\tau) & P_{12}(t,\tau) \\ P_{21}(t,\tau) & P_{22}(t,\tau) \end{bmatrix}.$$

Then, since

$$\mathcal{P}_z(t,\tau) = \|\mathbf{z}(t,\tau) - \hat{\mathbf{z}}(t,\tau)\|^2,$$

it follows immediately that

$$P_{11}(t,\tau) = P(t), \qquad P_{22}(t,\tau) = P(\tau|t). \tag{17.4.15}$$

That is, we can write more explicitly,

$$\mathcal{P}_z(t,\tau) = \begin{bmatrix} P(t) & P_{12}(t,\tau) \\ P_{21}(t,\tau) & P(\tau|t) \end{bmatrix}, \quad \mathcal{P}_z(\tau,\tau) = \bar{P},$$

where $\bar{P}$ is as in (17.4.14).

Furthermore, define the permutation matrix

$$\mathcal{J} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}.$$

Then it can be checked (see Prob. 17.2, on showing that a permuted version of a scattering matrix satisfies a Riccati differential equation) that $[\bar{P} \star \mathcal{S}^o(t,\tau)]\mathcal{J}$ satisfies the same differential equation as $\mathcal{P}_z(t,\tau)$ and therefore

$$\mathcal{P}_z(t,\tau) = [\bar{P} \star \mathcal{S}^o(t,\tau)]\mathcal{J} \triangleq \bar{\mathcal{S}}(t,\tau)\mathcal{J}. \tag{17.4.16}$$

Here, we introduced the notation $\bar{\mathcal{S}}(t,\tau)$ to denote the scattering matrix that is obtained by adjoining the special initial condition $\bar{P}$ to $\mathcal{S}^o(t,\tau)$. This is in contrast to the notation $\mathcal{S}(t,\tau)$ that we used earlier when the initial condition was taken as $\begin{bmatrix} I & P(\tau) \\ 0 & I \end{bmatrix}$.

Equation (17.4.16) identifies the entries of $\bar{\mathcal{S}}(t,\tau)$ as

$$\bar{\mathcal{S}}(t,\tau) = \begin{bmatrix} P_{12}(t,\tau) & P(t) \\ P(\tau|t) & P_{21}(t,\tau) \end{bmatrix}.$$

In view of the results in Sec. 17.3.1, $\bar{\mathcal{S}}(t,\tau)$ satisfies the same differential equation as $\mathcal{S}^o(t,\tau)$ but with the new initial condition $\bar{P}$. Therefore, we obtain a *stochastic interpretation of the permuted scattering matrix $\bar{\mathcal{S}}(t,\tau)\mathcal{J}$ as the error covariance matrix of the extended vector $\mathbf{z}(t,\tau)$*. Now since $\bar{\mathcal{S}}(t,\tau)$ can be evaluated in several ways, this result suggests several methods for evaluating $\mathcal{P}_z(t,\tau)$ itself and, consequently, for solving the filtering and smoothing problems. We give a few (of many) examples.

First approach. [*The innovations formulas for smoothing.*] Recall that the entries of $\bar{\mathcal{S}}(t,\tau)$ satisfy differential equations of the same form as (17.4.3)–(17.4.5) and, hence, we obtain

$$\dot{P}(t) = F(t)P(t) + P(t)P^*(t) + G(t)Q(t)G^*(t) - P(t)H^*(t)R^{-1}(t)H(t)P(t), \quad P(\tau),$$

$$\frac{d}{dt}P_{12}(t,\tau) = [F(t) - P(t)H^*(t)R^{-1}(t)H(t)]P_{12}(t,\tau), \quad P_{12}(\tau,\tau) = P(\tau),$$

$$\frac{dP(\tau|t)}{dt} = -P_{21}(t,\tau)H^*(t)R^{-1}(t)H(t)P_{12}(t,\tau), \quad P_{21}(\tau,\tau) = P(\tau).$$

We immediately see from the differential equation for $P_{12}(t,\tau)$, and from $P_{12}(t,\tau) = P_{21}^*(t,\tau)$, that we can identify $P_{12}(t,\tau)$ as

$$P_{12}(t,\tau) = \Psi(t,\tau)P(\tau),$$

where $\Psi(t,\tau)$ is the state transition matrix of the closed-loop Kalman filter. Likewise, $P_{21}(t,\tau) = P_{12}^*(t,\tau)$ and therefore

$$P(\tau|t) = P(\tau) - P(\tau)\left[\int_\tau^t \Psi^*(s,\tau)H^*(s)R^{-1}(s)H(s)\Psi(s,\tau)ds\right]P(\tau),$$

which is the basic general formula for the smoothing error variance (16.5.10).

In summary we have shown that the entries of $\bar{\mathcal{S}}(t,\tau)$ and $\mathcal{P}_z(t,\tau)$ are the following:

$$\mathcal{P}_z(t,\tau) = \begin{bmatrix} P(t) & \Psi(t,\tau)P(\tau) \\ P(\tau)\Psi^*(t,\tau) & P(\tau|t) \end{bmatrix}, \quad \bar{\mathcal{S}}(t,\tau) = \begin{bmatrix} \Psi(t,\tau)P(\tau) & P(t) \\ P(\tau|t) & P(t)\Psi^*(t,\tau) \end{bmatrix}.$$

◆

Second approach. [*Finding $\bar{\mathcal{S}}(t,\tau)$ as $\bar{P} \star \mathcal{S}^o(t,\tau)$ gives the partitioned formulas.*] Indeed, we already know that the entries of $\mathcal{S}^o(t,\tau)$ are given by

$$\mathcal{S}^o(t,\tau) = \begin{bmatrix} \Psi^o(t,\tau) & \mathcal{P}^o(t,\tau) \\ -\mathcal{O}^o(t,\tau) & \Psi^{o*}(t,\tau) \end{bmatrix},$$

where $\{\Psi^o, \mathcal{P}^o, \mathcal{O}^o\}$ satisfy the differential equations derived prior to Lemma 17.1.1.

Now we conclude from the above expressions for $\mathcal{P}_z(t,\tau)$ and $\bar{\mathcal{S}}(t,\tau)$, and from $\bar{\mathcal{S}}(t,\tau) = \bar{P} \star \mathcal{S}^o(t,\tau)$, that we must have

$$\begin{bmatrix} \Psi(t,\tau)P(\tau) & P(t) \\ P(\tau|t) & P(t)\Psi^*(t,\tau) \end{bmatrix} = \bar{P} \star \begin{bmatrix} \Psi^o(t,\tau) & \mathcal{P}^o(t,\tau) \\ -\mathcal{O}^o(t,\tau) & \Psi^{o*}(t,\tau) \end{bmatrix},$$

from which we obtain

$$\begin{aligned} P(\tau|t) &= P(\tau) - P(\tau)\mathcal{O}^o(t,\tau)[I + P(\tau)\mathcal{O}^o(t,\tau)]^{-1}P(\tau), \\ &= [\mathcal{O}^o(t,\tau) + P^{-1}(\tau)]^{-1}, \\ P(t) &= \mathcal{P}^o(t,\tau) + \Psi^o(t,\tau)P(\tau)[I + \mathcal{O}^o(t,\tau)P(\tau)]^{-1}\Psi^{o*}(t,\tau), \\ &= \mathcal{P}^o(t,\tau) + \Psi^o(t,\tau)P(\tau|t)\Psi^{o*}(t,\tau), \\ \Psi(t,\tau)P(\tau) &= \Psi^o(t,\tau)[I + P(\tau)\mathcal{O}^o(t,\tau)]^{-1}P(\tau) = \Psi^o(t,\tau)P(\tau|t). \end{aligned} \tag{17.4.17}$$

These are the so-called partitioned formulas obtained by Lainiotis (1974) — see Sec. 17.4.2. ◆

Third approach. [*Finding $\bar{\mathcal{S}}^o(t,\tau)$ by backwards evolution gives the Mayne-Fraser formulas.*] Indeed, define

$$P_\infty^b(\tau) \triangleq \mathcal{O}^{-o}(t,\tau), \quad \hat{\mathbf{x}}_\infty^b(\tau) \triangleq P_\infty^b(\tau)\lambda^o(\tau|t).$$

We noted earlier right after (17.2.2) that $\mathcal{O}^{-o}(t,\tau)$ is indeed equal to the matrix $P_\infty^b(\tau)$ that we encountered in Thm. 16.A.5 while studying the Mayne-Fraser two-filter formulas.

Then (17.4.6) becomes

$$\hat{\mathbf{x}}(\tau|t) = [P^{-1}(\tau) + P_\infty^{-b}(\tau)]^{-1}[P^{-1}(\tau)\hat{\mathbf{x}}(\tau) + P_\infty^{-b}(\tau)\hat{\mathbf{x}}_\infty^b(\tau)].$$

Moreover, using (17.2.1) and (17.2.2), we can check that $\{\hat{\mathbf{x}}_\infty^b(\cdot), P_\infty^b(\cdot)\}$ satisfy

$$-\frac{\mathrm{d}}{\mathrm{d}\tau}\hat{\mathbf{x}}_\infty^b(\tau) = [-F(\tau) - P_\infty^b(\tau)H^*(\tau)R^{-1}(\tau)H(\tau)]\hat{\mathbf{x}}_\infty^b(\tau) + P_\infty^b(\tau)H^*(\tau)R^{-1}(\tau)\mathbf{y}(\tau),$$

$$-\dot{P}_\infty^b(\tau) = -F(\tau)P_\infty^b(\tau) - P_\infty^b(\tau)F^*(\tau) + G(\tau)Q(\tau)G^*(\tau) - P_\infty^b(\tau)H^*(\tau)R^{-1}(\tau)H(\tau)P_\infty^b(\tau),$$

with boundary values

$$\hat{\mathbf{x}}_\infty^b(t|t) \text{ arbitrary} \quad \text{and} \quad P_\infty^b(t) = \infty \cdot I.$$

Note that the differential equation for $\hat{\mathbf{x}}^b_\infty(\tau)$ coincides with the one we encountered earlier in Thm. 16.A.5 (which explains our choice of notation). In fact, the full equivalence is established by noting from (17.4.17) that $[P^{-1}(\tau) + P^{-b}_\infty(\tau)]^{-1}$ is equal to $P(\tau|t)$. ◆

Other Smoothing Formulas. However, the story does not end here, since there are more ways of solving for $\mathcal{P}_z(t, \tau)$. In particular, we see from the scattering framework that the Lainiotis and two-filter approaches are natural counterparts of each other, one forwards in time and the other backwards in time. However, we have not described the backwards counterpart of the first approach. The main ingredient for obtaining this result is to start with a backwards Markovian state-space model that is equivalent to the given forwards state-space model. Now we studied backwards Markovian models in Sec. 16.A and showed in Lemma 16.A.1 how to construct such a model from a given forwards model; the backwards model was further used in Sec. 16.A.2 to derive a backwards Kalman filter. These results on backwards models can also be deduced via the scattering formulation (see, *e.g.,* Prob. 17.7). They can then be used to develop a backwards counterpart of the first approach to smoothing problems that we described in this section. We shall not pursue the details here but rather refer the reader to Ljung and Kailath (1976a). In that reference, various other consequences of the scattering approach to smoothing are presented, including families of smoothing formulas obtained by exploiting the ease in changing initial conditions and also formulas for the estimators; however, not all possibilities are exhausted there either.

Here we proceed to some especially nice consequences of the scattering formulation arising from the fact that time-invariant state-space models correspond to homogeneous scattering media.

17.5 HOMOGENEOUS MEDIA: TIME-INVARIANT MODELS

In homogeneous media, each incremental section has the same transmission and reflection coefficients. This will happen when the original state-space model is time-invariant, *i.e.,* $\{F(t), G(t), H(t), R(t), Q(t)\}$ do not change with t. An important consequence is that the properties of macroscopic sections will depend only upon the "width" of the section and not upon its actual location in the transmission line. That is, a section described by a scattering matrix $\mathcal{S}^o(t, \tau)$ will have the property that

$$\mathcal{S}^o(t, \tau) \;=\; \mathcal{S}^o(t + \sigma, \tau + \sigma),$$

for any σ. This fact has several interesting consequences and applications.

17.5.1 A Doubling Algorithm

An immediate application is to combine the homogeneity assumption with the concept of the star product to obtain a doubling algorithm to rapidly compute the limit of $\mathcal{P}(t, \tau)$ as $t \to \infty$; this was first done, in radiative transfer theory, by Van de Hulst (1963).

Let us fix a (small) interval $[0, t]$ and compute

$$\mathcal{S}^o(t, 0) = \begin{bmatrix} \Psi^o(t, 0) & \mathcal{P}^o(t, 0) \\ -\mathcal{O}^o(t, 0) & \Psi^{o*}(t, 0) \end{bmatrix}.$$

Then by the homogeneity assumption we can compute

$$\mathcal{S}^o(2t, 0) = \mathcal{S}^o(t, 0) \star \mathcal{S}^o(2t, t) = \mathcal{S}^o(t, 0) \star \mathcal{S}^o(t, 0),$$

and continuing in this way,

$$\mathcal{S}^o(2^i t, 0) = \mathcal{S}^o(2^{i-1} t, 0) \star \mathcal{S}^o(2^{i-1} t, 0). \tag{17.5.1}$$

In other words, each time we just take the star product of the result with itself in order to go out to double the time. This simple argument of course immediately yields (by equating the (1,2) elements on both sides of (17.5.1)), the Riccati doubling formula

$$\mathcal{P}^o(2^{i+1}t, 0) = \mathcal{P}^o(2^i t, 0) + \Psi^o(2^i t, 0)\mathcal{P}^o(2^i t, 0)\left[I + \mathcal{O}^o(2^i t, 0)\mathcal{P}^o(2^i t, 0)\right]^{-1}\Psi^{o*}(2^i t, 0).$$

Such a result is not easy to obtain by just studying the Riccati differential equation for $\mathcal{P}^o(t, 0)$, because the quantities $\Psi^o(t, 0)$ and $\mathcal{O}^o(t, 0)$ must also be brought into the picture. This can of course be done (see Lainiotis (1976b) and also Prob. 9.25), but the derivation will not compare in immediacy with the above physical derivation, or even in generality. Thus notice that the scattering formula (17.5.1) also immediately yields the updates

$$\Psi^o(2^{i+1}t, 0) = \Psi^o(2^i t, 0)\left[I + \mathcal{P}^o(2^i t, 0)\mathcal{O}^o(2^i t, 0)\right]^{-1}\Psi^o(2^i t, 0),$$

$$\mathcal{O}^o(2^{i+1}t, 0) = \mathcal{O}^o(2^i t, 0) + \Psi^{o*}(2^i t, 0)\mathcal{O}^o(2^i t, 0)\left[I + \mathcal{P}^o(2^i t, 0)\mathcal{O}^o(2^i t, 0)\right]^{-1}\Psi^o(2^i t, 0).$$

The above doubling formulas for $\{\mathcal{P}^o(2^{i+1}t, 0), \Psi^o(2^{i+1}t, 0), \mathcal{O}^o(2^{i+1}t, 0)\}$ are all algebraically complicated. The conceptually immediate scattering formula (17.5.1) shows that they can be reconstructed whenever explicitly desired just by knowing (or physically re-deriving) the rule for forming a star product.

17.5.2 Generalized Stokes Identities

Another natural consequence of homogeneity is that since

$$\mathcal{S}^o(t, \tau) = \mathcal{S}^o(t - \tau, 0), \tag{17.5.2}$$

we obtain

$$\frac{\partial}{\partial t}\mathcal{S}^o(t, \tau) \;=\; -\frac{\partial}{\partial \tau}\mathcal{S}^o(t, \tau). \tag{17.5.3}$$

This is just the mathematical consequence of the physical fact that in a homogeneous medium, adding a layer to the left of the section described by $\mathcal{S}^o(t, \tau)$ has the same effect as adding the same layer to the right.

Now by using the expressions that were derived earlier for these derivatives (*cf.* (17.1.11) and (17.2.4)), we obtain various special identities. In particular, by equating the (1,2) entries of both sides of (17.5.3), we obtain the formula

$$\frac{\partial}{\partial t}\mathcal{P}^o(t, \tau) \;=\; -\frac{\partial}{\partial \tau}\mathcal{P}^o(t, \tau) \;=\; \Psi^o(t, \tau)GQG^*\Psi^{o*}(t, \tau), \tag{17.5.4}$$

which readers may recognize as a special case, when $P(0) = 0$, of what we called a generalized Stokes identity in Sec. 16.6 — compare with Eq. (16.6.1) when the initial

condition $P(0)$ is zero so that $\dot{P}(0) = GQG^*$. However, the proof there was algebraic, unlike the physical argument used here. Moreover, of course, we also obtain the even more general Stokes identity (17.5.3), which has several other implications.[2] For example, from this general identity, we can obtain a dual set of Chandrasekhar-Kailath equations for fast computation of the observability Gramian $\mathcal{O}^o(t,\tau)$, similar to those described in Sec. 16.6, and below, for fast computation of the error variance $\mathcal{P}^o(t,\tau)$ (see also Prob. 16.34).

Generalized Stokes Identities for Nonzero Initial Conditions. Let us now show how to handle the case of nonzero initial conditions. When $\mathcal{P}(\tau,\tau) \neq 0$, say $\mathcal{P}(\tau,\tau) = \Pi(\tau)$, we must replace $\mathcal{S}^o(t,\tau)$ by

$$\mathcal{S}(t,\tau) = \begin{bmatrix} I & \Pi(\tau) \\ 0 & I \end{bmatrix} \star \mathcal{S}^o(t,\tau). \tag{17.5.5}$$

The forwards evolution equation for $\mathcal{S}(t,\tau)$ then is essentially the same as for $\mathcal{S}^o(t,\tau)$ — we just remove the superscript 'o' from the entries $\{\Psi^o(t,\tau), \mathcal{P}^o(t,\tau), \mathcal{O}^o(t,\tau)\}$ of $\mathcal{S}^o(t,\tau)$ and use the initial conditions

$$\Psi(\tau,\tau) = I, \quad \mathcal{P}(\tau,\tau) = \Pi(\tau), \quad \mathcal{O}(\tau,\tau) = 0.$$

However, the backwards evolution equation gets more complicated when $\Pi(\tau) \neq 0$. Now, as we discussed in Sec. 17.3.2, the appropriate formula depends upon the exact nature of the matrix $\mathcal{S}(t,\tau)$. When it is as above in (17.5.5), then referring to the discussion in Sec. 17.3.2, Eq. (17.3.15), we find that the generator matrix $N(\tau)$ for the backward evolution of $\mathcal{S}(t,\tau)$ is given by

$$N(\tau) = \begin{bmatrix} F - \Pi(\tau)H^*R^{-1}H & F\Pi(\tau) + \Pi(\tau)F^* + GQG^* - \Pi(\tau)H^*R^{-1}H\Pi(\tau) \\ -H^*R^{-1}H & F^* - H^*R^{-1}H\Pi(\tau) \end{bmatrix}$$

$$= \begin{bmatrix} \Psi(t,\tau) & \frac{\partial}{\partial t}\mathcal{P}(t,\tau)\Big|_{t=\tau} \\ -H^*R^{-1}H & \Psi^*(t,\tau) \end{bmatrix}.$$

Now introduce the factorization

$$\frac{\partial}{\partial t}\mathcal{P}(t,\tau)\Big|_{t=\tau} = L(\tau)JL^*(\tau),$$

where $L(\tau)$ is a full rank matrix and J a signature matrix. By using the backward evolution equation (17.3.13) we conclude that

$$-\frac{\partial}{\partial \tau}\mathcal{P}(t,\tau) = \Psi(t,\tau)L(\tau)JL^*(\tau)\Psi^*(t,\tau), \quad \mathcal{P}(\tau,\tau) = \Pi(\tau). \tag{17.5.6}$$

[2] Bellman, Kalaba, and Wing (1960) called (17.5.3) a Stokes identity, since a simple form of it appears in a paper by G. C. Stokes (recall Stokes' theorem for integration over surfaces) on the propagation of light through a pile of plates (Stokes (1862)).

If we further introduce the functions[3]

$$L(t,\tau) \triangleq \Psi(t,\tau)L(\tau), \qquad K(t,\tau) \triangleq \mathcal{P}(t,\tau)H^*R^{-1}, \tag{17.5.7}$$

then by using the forward evolution equation (17.4.4) we conclude that the following equalities should hold:

$$\frac{\partial}{\partial t}L(t,\tau) = [F - K(t,\tau)H]L(t,\tau), \quad L(\tau,\tau) = L(\tau), \tag{17.5.8}$$

$$-\frac{\partial}{\partial \tau}K(t,\tau) = L(t,\tau)JL^*(t,\tau)H^*R^{-1}, \quad K(t,t) = P(t,t)H^*R^{-1}. \tag{17.5.9}$$

This is a set of Chandrasekhar-Kailath equations for $L(\cdot,\cdot)$ and $K(\cdot,\cdot)$. In the general time-variant case, it is however not feasible to use the above equations since the time arguments do not fit. That is, because of the opposite directions of evaluation of (17.5.8) and (17.5.9), at any given intermediate point the values of $L(\cdot,\cdot)$ needed to solve for $K(\cdot,\cdot)$ (and vice versa) will not be available. However, when $\{F, G, Q, H, R\}$ are time-invariant, then (*cf.* (17.5.13)) we have the "Stokes" relation (see below)

$$-\frac{\partial}{\partial \tau}K(t,\tau) = \frac{\partial}{\partial t}K(t,\tau), \qquad K(\tau,\tau) = \Pi(\tau)H^*R^{-1}. \tag{17.5.10}$$

Hence, (17.5.9) can be replaced by

$$\frac{\partial}{\partial t}K(t,\tau) = L(t,\tau)JL^*(t,\tau)H^*R^{-1},$$

$$K(\tau,\tau) = \Pi(\tau)H^*R^{-1}, \tag{17.5.11}$$

and now (17.5.8) and (17.5.11) form the compatible set of Chandrasekhar-Kailath equations.

An algebraic justification for (17.5.10) is the following. Consider the Riccati differential equation for constant state-space models,

$$\frac{d}{dt}\mathcal{P}(t,\tau) = F\mathcal{P}(t,\tau) + \mathcal{P}(t,\tau)F^* + GQG^* - \mathcal{P}(t,\tau)H^*R^{-1}H\mathcal{P}(t,\tau),$$

$$\mathcal{P}(\tau,\tau) = P(\tau),$$

and let $\mathcal{P}(t+\Delta,\tau)$ and $P(\tau+\Delta,\tau)$ denote the values of $\mathcal{P}$ at time instants $t+\Delta$ and $\tau+\Delta$, for some $\Delta > 0$.

[3] An interpretation for $L(t,\tau)$ willfollow from the identity (17.5.12) established further ahead, which shows that, for any t, it holds

$$\frac{\partial}{\partial t}\mathcal{P}(t,\tau) = L(t,\tau)JL^*(t,\tau).$$

Now assume we change the initial condition of the above Riccati differential equation at time τ to the value $P(\tau+\Delta,\tau)$. The homogeneity assumption guarantees that the resulting value of $\mathcal{P}$ at time t will be $\mathcal{P}(t+\Delta,\tau)$. We can now invoke the change in initial conditions formula (17.4.10) to conclude that the following relation must hold:

$$\frac{\mathcal{P}(t+\Delta,\tau)-\mathcal{P}(t,\tau)}{\Delta} = \Psi(t+\Delta,\tau)\,\frac{\mathcal{P}(\tau+\Delta,\tau)-\mathcal{P}(\tau,\tau)}{\Delta}\,[I+\mathcal{O}(t+\Delta,\tau)$$
$$\cdot\,(\mathcal{P}(\tau+\Delta,\tau)-\mathcal{P}(\tau,\tau))]^{-1}\,\Psi^*(t+\Delta,\tau).$$

By taking the limit as $\Delta\to 0$ we conclude that

$$\frac{\partial}{\partial t}\mathcal{P}(t,\tau) = \Psi(t,\tau)\,\frac{\partial}{\partial t}\mathcal{P}(t,\tau)\Big|_{t=\tau}\,\Psi^*(t,\tau), \tag{17.5.12}$$

which is in fact the relation used in Sec. 16.6 to derive the Chandrasekhar-Kailath algorithm. This relation also justifies (17.5.10) since, in view of (17.5.6) and the definition (17.5.7) for $K(t,\tau)$, we have

$$-\frac{\partial}{\partial\tau}K(t,\tau) = \Psi(t,\tau)\,\frac{\partial}{\partial t}\mathcal{P}(t,\tau)\Big|_{t=\tau}\,\Psi^*(t,\tau)H^*R^{-1},$$
$$= \frac{\partial}{\partial t}\mathcal{P}(t,\tau)H^*R^{-1},$$
$$= \frac{\partial}{\partial t}K(t,\tau). \tag{17.5.13}$$

17.6 DISCRETE-TIME SCATTERING FORMULATION

There is a lot more that can be gathered from further study of the continuous-time scattering problem. However, in the interests of brevity, we shall go on to a brief discussion of the discrete-time case.

We start with the standard state-space model

$$\mathbf{x}_{i+1} = F_i\mathbf{x}_i + G_i\mathbf{u}_i, \qquad \mathbf{y}_i = H_i\mathbf{x}_i + \mathbf{v}_i, \tag{17.6.1}$$

where $\{\mathbf{u}_i,\mathbf{v}_i\}$ are uncorrelated white-noise processes with variances $\{Q_i\geq 0, R_i>0\}$, both of which are uncorrelated with $\mathbf{x}_0$, whose variance we denote by Π_0.

Now in Sec. 15.7, we showed that by combining the above state-space model for $\{\mathbf{y}_i\}$ with a complementary state-space model, we were able to directly obtain the following so-called Hamiltonian equations:

$$\begin{bmatrix}\hat{\mathbf{x}}_{i+1|N}\\ \boldsymbol{\lambda}_{i|N}\end{bmatrix} = \begin{bmatrix}F_i & G_iQ_iG_i^*\\ -H_i^*R_i^{-1}H_i & F_i^*\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}_{i|N}\\ \boldsymbol{\lambda}_{i+1|N}\end{bmatrix} + \begin{bmatrix}0\\ H_i^*R_i^{-1}\end{bmatrix}\mathbf{y}_i, \tag{17.6.2}$$

with boundary conditions

$$\hat{\mathbf{x}}_{0|N} = \Pi_0\boldsymbol{\lambda}_{0|N} \quad\text{and}\quad \boldsymbol{\lambda}_{N+1|N} = 0, \tag{17.6.3}$$

and such that for each time instant i

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i + P_i\boldsymbol{\lambda}_{i|N}, \quad 0\leq i\leq N, \tag{17.6.4}$$

where we recall that P_i is the prediction error covariance matrix that satisfies the Riccati recursion (*cf.* Thm. 9.2.1)

$$P_{i+1} = F_iP_iF_i^* + G_iQ_iG_i^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_0=\Pi_0, \tag{17.6.5}$$

with

$$K_{p,i} = F_iP_iH_i^*R_{e,i}^{-1}, \quad R_{e,i} = R_i + H_iP_iH_i^*. \tag{17.6.6}$$

[Relation (17.6.4) can also be derived via a complementary state-space model argument just as we do in App. 17.A by assuming that the initial time instant is i and that we are given an estimator $\hat{\mathbf{x}}_i$ for $\mathbf{x}_i$ with $E\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^*=P_i$.]

Now (unlike continuous time) we can directly go to a transmission line picture in which we can regard $\hat{\mathbf{x}}_{\cdot|N}$ as a *forward* wave and $\boldsymbol{\lambda}_{\cdot|N}$ as a *backward* wave traveling through a section of a scattering medium with reflection and transmission coefficients as shown in Fig. 17.12. For simplicity, we shall start with the assumption that $\Pi_0=0$.

We shall denote the generator matrix and the source vector at time i, respectively, by

$$M_i = \begin{bmatrix}F_i & G_iQ_iG_i^*\\ -H_i^*R_i^{-1}H_i & F_i^*\end{bmatrix} \quad\text{and}\quad m_i = \begin{bmatrix}0\\ H_i^*R_i^{-1}\mathbf{y}_i\end{bmatrix}. \tag{17.6.7}$$

Then we can put together many such sections to get a macroscopic section of the scattering medium, say from time i to time N (see Fig. 17.13), and define the corresponding scattering and signal matrices as

$$\mathcal{S}^o_{N+1,i} = \begin{bmatrix}\Phi^o_{N+1,i} & \mathcal{P}^o_{N+1,i}\\ -\mathcal{O}^o_{N+1,i} & \Phi^{o*}_{N+1,i}\end{bmatrix} \quad\text{and}\quad s^o_{N+1,i} = \begin{bmatrix}r^{o,+}_{N+1,i}\\ r^{o,-}_{N+1,i}\end{bmatrix}, \tag{17.6.8}$$

where the superscript o is used to indicate our assumption that $\Pi_0=0$.

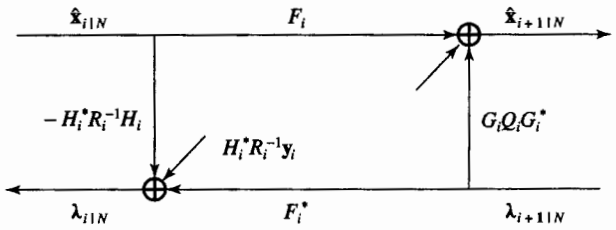


Figure 17.12 A scattering layer for the fixed-interval smoothing problem.

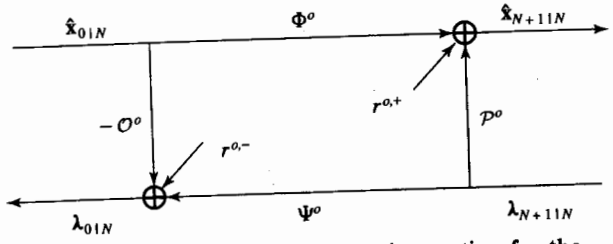


Figure 17.13 A macroscopic scattering section for the fixed-interval smoothing problem with $\Pi_0 = 0$.

17.6.1 Some Features of Discrete-Time Scattering

Another simplification of the discrete-time model is that a macroscopic section from i to N can be described by

$$\mathcal{S}^o_{N+1,i} = M_i \star M_{i+1} \star \ldots \star M_N,$$
$$s^o_{N+1,i} = m_i \bullet m_{i+1} \bullet \ldots \bullet m_N.$$

The superscript o indicates that boundary conditions are not included. No differential equations arise, and in particular, the forward and backward evolution equations are immediate since

$$\mathcal{S}^o_{N+1,i} = \mathcal{S}^o_{N,i} \star M_N, \quad s^o_{N+1,i} = s^o_{N,i} \bullet m_N, \tag{17.6.9}$$
$$\mathcal{S}^o_{N,i-1} = M_{i-1} \star \mathcal{S}^o_{N,i}, \quad s^o_{N,i-1} = m_{i-1} \bullet s^o_{N,i}. \tag{17.6.10}$$

The simplicity of the discrete medium becomes even more apparent when we consider what happens to these equations when we change initial conditions. Let

$$\mathcal{S}_{N+1,i} = \Gamma_i \star \mathcal{S}^o_{N+1,i},$$

denote the scattering matrix that corresponds to the medium $\mathcal{S}^o_{N+1,i}$ with any boundary layer Γ_i attached to the left. Then clearly $\mathcal{S}_{N+1,i}$ satisfies the same forwards equation as $\mathcal{S}^o_{N+1,i}$,

$$\mathcal{S}_{N+1,i} = \mathcal{S}_{N,i} \star M_N, \tag{17.6.11}$$

but with initial condition $\mathcal{S}_{i,i} = \Gamma_i$. For backwards evolution we have to work a little harder,

$$\begin{aligned}\mathcal{S}_{N,i-1} &= \Gamma_{i-1} \star \mathcal{S}^o_{N,i-1} = \Gamma_{i-1} \star M_{i-1} \star \mathcal{S}^o_{N,i},\\ &= \Gamma_{i-1} \star M_{i-1} \star \Gamma_i^{-1} \star \Gamma_i \star \mathcal{S}^o_{N,i},\\ &= N_{i-1} \star \mathcal{S}_{N,i},\end{aligned} \tag{17.6.12}$$

where we defined

$$N_{i-1} \triangleq \Gamma_{i-1} \star M_{i-1} \star \Gamma_i^{-1}. \tag{17.6.13}$$

But this is still much simpler than in continuous time (*cf.* Sec. 17.3.2), where we had to solve a set of four simultaneous equations in order to determine $N(\tau)$ (*cf.* (17.3.14)).

Another distinctive feature of the discrete-time case is that we can factor the matrix M_N in a natural way that leads to the measurement- and time-update forms of the Kalman filter recursions (see Prob. 17.3).

17.6.2 The Scattering Parameters

As in the continuous-time case, we can identify the scattering parameters $\{\mathcal{P}^o, \Phi^o, \mathcal{O}^o\}$ of (17.6.8) by using the forwards equation (17.6.9).

Alternatively, since we already know how to incorporate boundary conditions, we can instead proceed more rapidly and identify the resulting scattering parameters $\{\mathcal{P}_{N+1,i}, \Phi_{N+1,i}, \mathcal{O}_{N+1,i}\}$. The quantities with superscript o will be obtained as special cases by setting the boundary conditions to zero.

More specifically, we combine a boundary section,

$$\mathcal{B} = \begin{bmatrix} I & P_i \\ 0 & I \end{bmatrix}, \tag{17.6.14}$$

with internal source vector

$$s^b = \begin{bmatrix} \hat{\mathbf{x}}_i \\ 0 \end{bmatrix},$$

with the primary medium $(\mathcal{S}^o_{N+1,i}, s^o_{N+1,i})$, as shown in Fig. 17.14. The combined section, say $(\mathcal{S}_{N+1,i}, s_{N+1,i})$, is shown in Fig. 17.15 with scattering parameters $\{\mathcal{P}_{N+1,i}, \Phi_{N+1,i}$, and $\mathcal{O}_{N+1,i}\}$,

$$\mathcal{S}_{N+1,i} = \mathcal{B} \star \mathcal{S}^o_{N+1,i} = \begin{bmatrix} \Phi_{N+1,i} & \mathcal{P}_{N+1,i} \\ -\mathcal{O}_{N+1,i} & \Phi^*_{N+1,i} \end{bmatrix}, \quad s_{N+1,i} = s^b \bullet s^o_{N+1,i} = \begin{bmatrix} r^+_{N+1,i} \\ r^-_{N+1,i} \end{bmatrix},$$

where the star product and the dot sum are exactly as before (Sec. 17.3).

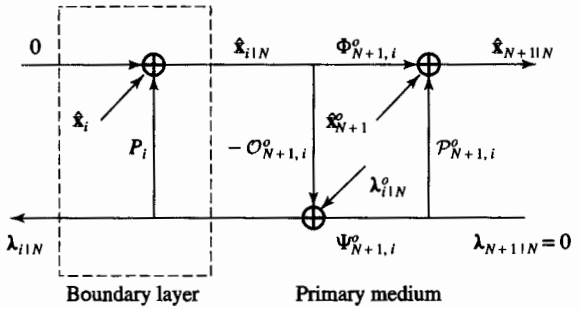


Figure 17.14 A cascade of a boundary layer and the primary medium of sections i through N.

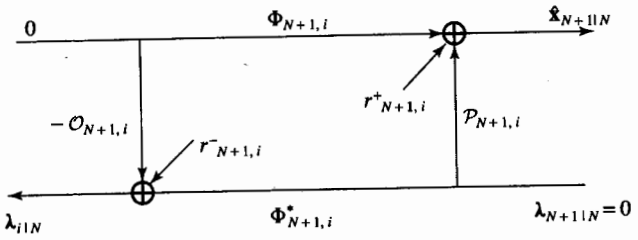


Figure 17.15 The combined layer.

Now from the fact that the incident (or incoming waves) in Fig. 17.15 are zero, we can conclude at once that the internal sources $(r^+_{N+1,i}, r^-_{N+1,i})$ are given by

$$r^+_{N+1,i} = \hat{\mathbf{x}}_{N+1|N} \quad \text{and} \quad r^-_{N+1,i} = \lambda_{i|N}.$$

The scattering parameters of Fig. 17.15 can be identified by resorting to the scattering calculus of Sec. 17.3. So consider the composite layer of Fig. 17.15 and assume we extend it by one more section, which corresponds to considering a smoothing problem of order $N+1$ (see Fig. 17.16 where we are dropping the time subscript $N+1$ from the quantities $\{F, H, G, R, Q, \mathbf{y}\}$).

Then, using (17.6.11), the overall scattering matrix is

$$\begin{bmatrix} \Phi_{N+2,i} & \mathcal{P}_{N+2,i} \\ -\mathcal{O}_{N+2,i} & \Phi^*_{N+2,i} \end{bmatrix} =$$

$$\begin{bmatrix} \Phi_{N+1,i} & \mathcal{P}_{N+1,i} \\ -\mathcal{O}_{N+1,i} & \Phi^*_{N+1,i} \end{bmatrix} \star \begin{bmatrix} F_{N+1} & G_{N+1}Q_{N+1}G^*_{N+1} \\ -H^*_{N+1}R^{-1}_{N+1}H_{N+1} & F^*_{N+1} \end{bmatrix},$$

which yields the recursions (assuming the required inverses exist)

$$\Phi_{N+2,i} = F_{N+1}\left[I + \mathcal{P}_{N+1,i}H^*_{N+1}R^{-1}_{N+1}H_{N+1}\right]^{-1}\Phi_{N+1,i}, \tag{17.6.15}$$

$$\mathcal{P}_{N+2,i} = G_{N+1}Q_{N+1}G^*_{N+1} + F_{N+1}\mathcal{P}_{N+1,i}\left[I + H^*_{N+1}R^{-1}_{N+1}H_{N+1}\mathcal{P}_{N+1,i}\right]^{-1}F^*_{N+1}, \tag{17.6.16}$$

$$\mathcal{O}_{N+2,i} = \mathcal{O}_{N+1,i} + \Phi^*_{N+1,i}H^*_{N+1}R^{-1}_{N+1}H_{N+1}\left[I + \mathcal{P}_{N+1,i}H^*_{N+1}R^{-1}_{N+1}H_{N+1}\right]^{-1}\Phi_{N+1,i}, \tag{17.6.17}$$

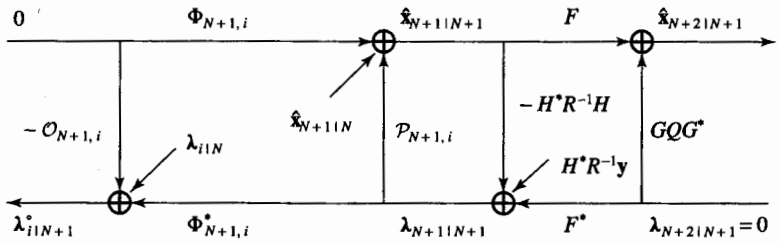


Figure 17.16 Evolution of the scattering quantities.

with the boundary conditions (see Fig. 17.14)

$$\begin{bmatrix} \Phi_{i,i} & \mathcal{P}_{i,i} \\ -\mathcal{O}_{i,i} & \Phi^*_{i,i} \end{bmatrix} = \begin{bmatrix} I & P_i \\ 0 & I \end{bmatrix}.$$

It is now possible, by applying the matrix inversion formula of App. A to the matrices

$$\left[I + H^*_{N+1}R^{-1}_{N+1}H_{N+1}\mathcal{P}_{N+1,i}\right]^{-1} \quad \text{and} \quad \left[I + \mathcal{P}_{N+1,i}H^*_{N+1}R^{-1}_{N+1}H_{N+1}\right]^{-1},$$

to make the following identifications.

Lemma 17.6.1 (Scattering Parameters) *The parameters $\Phi_{N+1,i}$, $\mathcal{P}_{N+1,i}$, and $\mathcal{O}_{N+1,i}$ of Fig. 17.15 are given by*[4]

$$\Phi_{N+1,i} = \Phi_p(N+1, i), \tag{17.6.18}$$

$$\mathcal{P}_{N+1,i} = P_{N+1}, \tag{17.6.19}$$

$$\mathcal{O}_{N+1,i} = \sum_{j=i}^{N} \Phi^*_p(j,i)H^*_j R^{-1}_{e,j}H_j\Phi_p(j,i), \tag{17.6.20}$$

where P_j is the error covariance matrix that is obtained via the Riccati recursion (17.6.5) and $\Phi_p(j,i)$ is the closed-loop transition matrix function

$$\Phi_p(j,i) = F_{p,j-1}F_{p,j-2}\ldots F_{p,i} \quad \textit{for} \quad j > i \quad \textit{and} \quad \Phi_p(i,i) = I, \tag{17.6.21}$$

[4] In earlier chapters we used the notation $\mathcal{O}_N$ to refer to the following observability Gramian (recall, *e.g.*, the statement of Lemma 14.4.2):

$$\mathcal{O}_N = \sum_{j=0}^{N} \Phi^*_p(j,0)H^*_j R^{-1}_{e,j}H_j\Phi_p(j,0),$$

with the subscript N in $\mathcal{O}_N$ coinciding with the upper limit on the summation symbol. In this section, however, we changed the notation slightly and, according to (17.6.20), we would denote the above Gramian by $\mathcal{O}_{N+1,0}$ (with $N+1$ rather than N). This change of notation is used here in order to be consistent with the scattering formulation where we are using the subscript $(N+1)$ to denote the macroscopic scattering parameters that result from cascading sections up to time N. This remark is made here in order to call the reader's attention to this notational change, especially when comparing the change-in-initial conditions formulas of this chapter with formulas we derived in earlier chapters.

with $F_{p,i} = F_i - K_{p,i}H_i$. Expressions for $\{\mathcal{O}^o_{N+1,i}, \mathcal{P}^o_{N+1,i}, \Phi^o_{N+1,i}\}$ follow from the above by employing $P_i = 0$ and using the resulting Riccati variables of (17.6.5) in the above expressions. ∎

Note that $\mathcal{O}_{N+1,i}$ is simply the observability Gramian associated with the sections i through N.

17.6.3 The Kalman Filter and Related Identities

The internal sources of the cascade of Fig. 17.16 can be studied by using the dot sum relation (17.3.2), *viz.*,

$$\begin{bmatrix} \hat{\mathbf{x}}_{N+2|N+1} \\ \boldsymbol{\lambda}_{i|N+1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}_{N+1|N} \\ \boldsymbol{\lambda}_{i|N} \end{bmatrix} \bullet \begin{bmatrix} 0 \\ H^*_{N+1}R^{-1}_{N+1}\mathbf{y}_{N+1} \end{bmatrix} = \tag{17.6.22}$$

$$\begin{bmatrix} F_{N+1}\left[I + P_{N+1}H^*_{N+1}R^{-1}_{N+1}H_{N+1}\right]^{-1}\left[\hat{\mathbf{x}}_{N+1|N} + P_{N+1}H^*_{N+1}R^{-1}_{N+1}\mathbf{y}_{N+1}\right] \\ \boldsymbol{\lambda}_{i|N} + \Phi^*_p(N+1)\left[I + H^*_{N+1}R^{-1}_{N+1}H_{N+1}P_{N+1}\right]^{-1}H^*_{N+1}R^{-1}_{N+1}\mathbf{e}_{N+1} \end{bmatrix},$$

where $\mathbf{e}_{N+1} = \mathbf{y}_{N+1} - H_{N+1}\hat{\mathbf{x}}_{N+1|N}$ is the innovations variable.

The first line in the above equality is nothing but the Kalman filter update for the prediction estimators. Indeed, by applying the matrix inversion formula to $\left[I + P_{N+1}H^*_{N+1}R^{-1}_{N+1}H_{N+1}\right]$ and grouping terms we obtain the more familiar form (see Prob. 17.5)

$$\hat{\mathbf{x}}_{N+2|N+1} = F_{N+1}\hat{\mathbf{x}}_{N+1|N} + K_{p,N+1}[\mathbf{y}_{N+1} - H_{N+1}\hat{\mathbf{x}}_{N+1|N}].$$

The second line in (17.6.22), on the other hand, provides an interesting update relation for the adjoint state as a function of the number of observations N, *viz.*,

$$\boldsymbol{\lambda}_{i|N+1} = \boldsymbol{\lambda}_{i|N} + \Phi^*_p(N+1)\left[I + H^*_{N+1}R^{-1}_{N+1}H_{N+1}P_{N+1}\right]^{-1}H^*_{N+1}R^{-1}_{N+1}\mathbf{e}_{N+1},$$

with initial value $\boldsymbol{\lambda}_{i|i-1} = 0$.

We can derive a similar update relation for $\hat{\mathbf{x}}_{i|N}$ as follows. Starting with (17.6.4) and substituting for $\boldsymbol{\lambda}_{i|N+1}$ from the above expression we obtain

$$\hat{\mathbf{x}}_{i|N+1} = \hat{\mathbf{x}}_{i|N} + P_i\Phi^*_p(N+1)\left[I + H^*_{N+1}R^{-1}_{N+1}H_{N+1}P_{N+1}\right]^{-1}H^*_{N+1}R^{-1}_{N+1}\mathbf{e}_{N+1},$$

with a given initial value $\hat{\mathbf{x}}_{i|i-1}$. This is a so-called fixed-point smoothing formula (see also Prob. 10.1).

We can also, as in continuous time, relate the estimators to those for zero initial conditions. Thus, referring to Fig. 17.14, we denote the values of $\hat{\mathbf{x}}_{N+1|N}$ and $\boldsymbol{\lambda}_{i|N}$ that are obtained under the special initial conditions

$$P_i = 0 \quad \text{and} \quad \hat{\mathbf{x}}_i = 0 \tag{17.6.23}$$

by $\hat{\mathbf{x}}^o_{N+1|N}$ and $\boldsymbol{\lambda}^o_{i|N}$, respectively. We then see from the flow-graph relations implied by the figure that

$$r^{o,+}_{N+1,i} = \hat{\mathbf{x}}^o_{N+1|N} \quad \text{and} \quad r^{o,-}_{N+1,i} = \boldsymbol{\lambda}^o_{i|N}. \tag{17.6.24}$$

With these identifications, we see that

$$\hat{\mathbf{x}}_{i|N} = \hat{\mathbf{x}}_i + P_i[\boldsymbol{\lambda}^o_{i|N} - \mathcal{O}^o_{N+1,i}\hat{\mathbf{x}}_{i|N}],$$

$$\hat{\mathbf{x}}_{N+1|N} = \hat{\mathbf{x}}^o_{N+1|N} + \Phi^o_{N+1,i}\hat{\mathbf{x}}_{i|N}.$$

Then the results in the next lemma follow quite easily.

Lemma 17.6.2 (Partitioned Formulas) *Consider the state-space model (17.6.1) and let $\hat{\mathbf{x}}_{i|N}$ denote the smoothed estimator of $\mathbf{x}_i$ given the observations $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$. Define also $\{\boldsymbol{\lambda}^o_{i|N}, \hat{\mathbf{x}}^o_{N+1|N}\}$ as explained above and let $\{\Phi^o_{N+1,i}, \mathcal{O}^o_{N+1,i}\}$ denote the forward and the left transmission coefficients that result from the scattering cascade with zero boundary conditions (cf. Lemma 17.6.1). Then the following relations hold:*

$$\hat{\mathbf{x}}_{i|N} = [I + P_i\mathcal{O}^o_{N+1,i}]^{-1}\left[\hat{\mathbf{x}}_i + P_i\boldsymbol{\lambda}^o_{i|N}\right], \tag{17.6.25}$$

$$\hat{\mathbf{x}}_{N+1|N} = \hat{\mathbf{x}}^o_{N+1|N} + \Phi^o_{N+1,i}\hat{\mathbf{x}}_{i|N}, \tag{17.6.26}$$

where P_i is the Riccati variable obtained via (17.6.5). Moreover, it holds that

$$\mathcal{P}_{N+1,i} = \mathcal{P}^o_{N+1,i} + \Phi^o_{N+1,i}P_i\left[I + \mathcal{O}^o_{N+1,i}P_i\right]^{-1}\Phi^{o*}_{N+1,i}. \tag{17.6.27}$$

∎

Proof: We still need to establish (17.6.27). For this purpose, note from Fig. 17.14 that $\mathcal{P}_{N+1,i}$ is given by (dropping the subscript $_{N+1,i}$ for simplicity):

$$\begin{aligned}\mathcal{P} &= \mathcal{P}^o + [\Phi^oP_i\Phi^{o*} - \Phi^oP_i\mathcal{O}^oP_i\Phi^{o*} + \Phi^oP_i\mathcal{O}^oP_i\mathcal{O}^oP_i\Phi^{o*} - \ldots], \\ &= \mathcal{P}^o + \Phi^oP_i[I + \mathcal{O}^oP_i]^{-1}\Phi^{o*},\end{aligned}$$

which gives (17.6.27). ◆

17.6.4 General Change of Initial Conditions

As mentioned earlier, one of the features of the scattering framework is the ease with which the effects of changes in initial conditions can be studied. We discussed this problem in Sec. 17.4.3 for continuous-time state-space models, and exactly the same arguments go over to discrete time.

Referring to Fig. 17.14, suppose we have obtained the scattering matrix $\mathcal{S}_{N+1,i}$ and the source vector $s_{N+1,i}$ for the composite layer (*cf.* Lemma 17.6.1),

$$\mathcal{S}_{N+1,i} = \begin{bmatrix} \Phi_p(N+1,i) & P_{N+1} \\ -\mathcal{O}_{N+1,i} & \Phi_p^*(N+1,i) \end{bmatrix} \quad \text{and} \quad s_{N+1,i} = \begin{bmatrix} \hat{\mathbf{x}}_{N+1|N} \\ \boldsymbol{\lambda}_{i|N} \end{bmatrix},$$

with boundary conditions P_i and $\hat{\mathbf{x}}_i$. Now we can change the values of the boundary conditions to $P_i^{(1)}$ and $\hat{\mathbf{x}}_i^{(1)}$ by

$$\Delta P_i = P_i^{(1)} - P_i \qquad \text{and} \qquad \delta\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^{(1)} - \hat{\mathbf{x}}_i,$$

and placing a layer with scattering matrix and source vector,

$$\mathcal{S}_\Delta = \begin{bmatrix} I & \Delta P_i \\ 0 & I \end{bmatrix}, \quad s_\Delta = \begin{bmatrix} \delta\hat{\mathbf{x}}_i \\ 0 \end{bmatrix},$$

to the left of the configuration in Fig. 17.14. That is, the new scattering matrix and source vector can be found via $\mathcal{S}_{N+1,i}^{(1)} = \mathcal{S}_\Delta \star \mathcal{S}_{N+1,i}$ and $s_{N+1,i}^{(1)} = s_\Delta \bullet s_{N+1,i}$. If we denote the scattering parameters of $\mathcal{S}_{N+1,i}^{(1)}$ by

$$\{\Phi_p^{(1)}(N+1,i), \mathcal{P}_{N+1,i}^{(1)}, \Phi_p^{*(1)}(N+1,i), -\mathcal{O}_{N+1,i}^{(1)}\},$$

and the entries of the source vector $s_{N+1,i}^{(1)}$ by $\{\hat{\mathbf{x}}_{N+1|N}^{(1)}, \boldsymbol{\lambda}_{i|N}^{(1)}\}$, we obtain by using (17.3.1)–(17.3.2) the following result.

Theorem 17.6.1 (Changes in Initial Conditions) *Consider the model (17.6.1) and let*

$$\{\hat{\mathbf{x}}_{N+1|N}, \boldsymbol{\lambda}_{i|N}, P_{N+1}, \Phi_p(N+1,i), \mathcal{O}_{N+1,i}\}$$

denote the variables that arise in the solution of the prediction and fixed-interval smoothing problems with initial conditions P_i and $\hat{\mathbf{x}}_i$. Now assume the initial conditions are changed to $P_i^{(1)}$ and $\hat{\mathbf{x}}_i^{(1)}$. Let

$$\{\hat{\mathbf{x}}_{N+1|N}^{(1)}, \boldsymbol{\lambda}_{i|N}^{(1)}, P_{N+1}^{(1)}, \Phi_p^{(1)}(N+1,i), \mathcal{O}_{N+1,i}^{(1)}\}$$

denote the corresponding variables that arise by solving the same prediction and fixed-interval smoothing problems but with the new initial conditions. Then the following relations hold (assuming the required inverses exist):

$$\Phi_p^{(1)}(N+1,i) = \Phi_p(N+1,i)[I + \Delta P_i \mathcal{O}_{N+1,i}]^{-1}, \tag{17.6.28}$$

$$P_{N+1}^{(1)} = P_{N+1} + \Phi_p(N+1,i)\Delta P_i[I + \mathcal{O}_{N+1,i}\Delta P_i]^{-1}\Phi_p^*(N+1,i) \tag{17.6.29}$$

$$\mathcal{O}_{N+1,i}^{(1)} = \mathcal{O}_{N+1,i}[I + \Delta P_i \mathcal{O}_{N+1,i}]^{-1}, \tag{17.6.30}$$

as well as

$$\hat{\mathbf{x}}_{N+1|N}^{(1)} = \hat{\mathbf{x}}_{N+1|N} + \Phi_p(N+1,i)[I + \Delta P_i \mathcal{O}_{N+1,i}]^{-1}[\delta\hat{\mathbf{x}}_i + \Delta P_i \boldsymbol{\lambda}_{i|N}],$$

$$\boldsymbol{\lambda}_{i|N}^{(1)} = [I + \mathcal{O}_{N+1,i}\Delta P_i]^{-1}[\boldsymbol{\lambda}_{i|N} - \mathcal{O}_{N+1,i}\delta\hat{\mathbf{x}}_i],$$

where

$$\Delta P_i = P_i^{(1)} - P_i, \qquad \delta\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^{(1)} - \hat{\mathbf{x}}_i,$$

and the parameters $\{\Phi_p(N+1,i), \mathcal{O}_{N+1,i}\}$ are as defined in Lemma 17.6.1. ■

17.6.5 Backward Evolution

We can also develop backward evolution equations for the scattering parameters by using (17.6.12)–(17.6.13) with $\Gamma_i = \mathcal{B}$ in (17.6.14). For our purposes here it is sufficient to derive these equations with zero initial conditions and, hence, we use (17.6.10) instead.

The situation is depicted in Fig. 17.17 where we placed a section to the left of $\mathcal{S}_{N+1,i}^o$. In the figure, we are dropping the subscript $_{i-1}$ from the $\{F, G, H, Q, R, \mathbf{y}\}$ quantities. We see from Fig. 17.17 that $\boldsymbol{\beta} = \boldsymbol{\lambda}_{i|N}^o - \mathcal{O}_{N+1,i}^o G_{i-1} Q_{i-1} G_{i-1}^* \boldsymbol{\beta}$ and, hence,

$$\boldsymbol{\beta} = [I + \mathcal{O}_{N+1,i}^o G_{i-1} Q_{i-1} G_{i-1}^*]^{-1} \boldsymbol{\lambda}_{i|N}^o.$$

Therefore, referring again to the figure, we can write

$$\begin{aligned} \boldsymbol{\lambda}_{i-1|N}^o &= H_{i-1}^* R_{i-1}^{-1} \mathbf{y}_{i-1} + F_{i-1}^* \boldsymbol{\lambda}_{i|N}^o, \\ &= H_{i-1}^* R_{i-1}^{-1} \mathbf{y}_{i-1} + F_{i-1}^* [I + \mathcal{O}_{N+1,i}^o G_{i-1} Q_{i-1} G_{i-1}^*]^{-1} \boldsymbol{\lambda}_{i|N}^o \end{aligned} \tag{17.6.31}$$

which provides a backward recursion for $\boldsymbol{\lambda}_{i|N}^o$ with boundary condition $\boldsymbol{\lambda}_{N+1|N}^o = 0$.

A backward recursion for the left reflection coefficient can be developed as well. Using (17.6.10) and the star-product rule (17.3.1), we obtain

$$-\mathcal{O}_{N+1,i-1}^o = -H_{i-1}^* R_{i-1}^{-1} H_{i-1} - F_{i-1}^* \mathcal{O}_{N+1,i}^o [I + G_{i-1} Q_{i-1} G_{i-1}^* \mathcal{O}_{N+1,i}^o]^{-1} F_{i-1}, \tag{17.6.32}$$

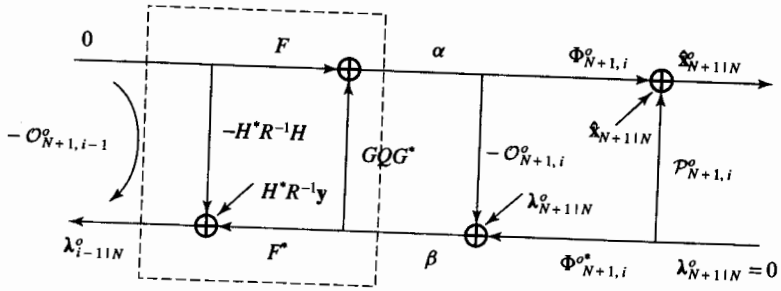


Figure 17.17 Backward evolution.

from which we conclude, by applying the matrix inversion formula to the inverse $\left[I + G_{i-1}Q_{i-1}G_{i-1}^*\mathcal{O}_{N+1,i}^o\right]^{-1}$, that $\mathcal{O}_{N+1,i}^o$ obeys the following backward Riccati recursion:

$$\begin{aligned}\mathcal{O}_{N+1,i-1}^o &= F_{i-1}^*\mathcal{O}_{N+1,i}^o F_{i-1} + H_{i-1}^* R_{i-1}^{-1} H_{i-1} - \\ &\quad - F_{i-1}^*\mathcal{O}_{N+1,i}^o G_{i-1}\left[G_{i-1}^*\mathcal{O}_{N+1,i}^o G_{i-1} + Q_{i-1}^{-1}\right]^{-1} G_{i-1}^*\mathcal{O}_{N+1,i}^o F_{i-1},\end{aligned} \tag{17.6.33}$$

with boundary condition $\mathcal{O}_{N+1,N+1}^o = 0$.

The backward evolution equation (17.6.33) for the left reflection coefficient is exactly dual to the Riccati recursion for the forward evolution of the right reflection coefficient.

17.6.6 Homogeneous Media

Some especially nice consequences of the scattering formulation arise in the special case of homogeneous media, for which the generator matrix M_i in (17.6.7) is independent of time and equal to

$$M = \begin{bmatrix} F & GQG^* \\ -H^*R^{-1}H & F^* \end{bmatrix}.$$

It then follows that quantities like the macroscopic coefficients $\{\mathcal{P}_{N+1,i}^o, \mathcal{O}_{N+1,i}^o\}$ will depend only upon the difference $(N+1-i)$ and not upon the actual values of N and i. A particular consequence is the fact (a generalized Stokes identity) that, for any i,

$$\mathcal{S}_{N,i}^o \star M = M \star \mathcal{S}_{N,i}^o. \tag{17.6.34}$$

This relation can be used to obtain in a very natural way the CKMS algorithm derived after some algebra in Thm. 11.1.2.

The Fast (CKMS) Algorithm. Indeed, it follows from (17.6.34) that

$$\underbrace{\begin{bmatrix} I & P_i \\ 0 & I \end{bmatrix} \star \mathcal{S}_{N,i}^o}_{\mathcal{S}_{N,i}} \star M = \begin{bmatrix} I & P_i \\ 0 & I \end{bmatrix} \star M \star \begin{bmatrix} I & P_i \\ 0 & I \end{bmatrix}^{-1} \star \left(\underbrace{\begin{bmatrix} I & P_i \\ 0 & I \end{bmatrix} \star \mathcal{S}_{N,i}^o}_{\mathcal{S}_{N,i}} \right),$$

which is equivalent to the relation

$$\mathcal{S}_{N,i} \star M = Z_i \star \mathcal{S}_{N,i},$$

where we introduced

$$Z_i \triangleq \begin{bmatrix} I & P_i \\ 0 & I \end{bmatrix} \star M \star \begin{bmatrix} I & -P_i \\ 0 & I \end{bmatrix}.$$

By equating entries on both sides of this equality we can immediately obtain the results of Lemma 11.1.1.

To begin with, it can be verified by direct calculation, using (17.3.1), that

$$\begin{aligned} Z_i &= \begin{bmatrix} F[I + P_iH^*R^{-1}H]^{-1} & GQG^* + FP_i[I + H^*R^{-1}HP_i]^{-1}F^* - P_i \\ -H^*R^{-1}H[I + P_iH^*R^{-1}H]^{-1} & [I + H^*R^{-1}HP_i]^{-1}F^* \end{bmatrix}, \\ &\triangleq \begin{bmatrix} \tilde{F}_i & \tilde{Q}_i \\ -\tilde{H}_i^*\tilde{H}_i & \tilde{F}_i^* \end{bmatrix}, \quad \text{say,} \end{aligned}$$

where we are denoting the individual (block) entries of Z_i by $\{\tilde{F}_i, \tilde{Q}_i, -\tilde{H}_i^*\tilde{H}_i\}$. More specifically, note that the $(2,1)$ entry of Z_i can indeed be written in the form $-\tilde{H}_i^*\tilde{H}_i$, for some $\tilde{H}_i$, since

$$H^*R^{-1}H[I + P_iH^*R^{-1}H]^{-1} = H^*[R + HP_iH^*]^{-1}H,$$

which allows us to take $\tilde{H}_i$ as $\tilde{H}_i = [R + HP_iH^*]^{-1/2}H$. Observe further that by applying the matrix inversion formula to the matrix inverse $[I + H^*R^{-1}HP_i]^{-1}$ that appears in the expression for $\tilde{Q}_i$ we conclude that

$$\begin{aligned} \tilde{Q}_i &= GQG^* + FP_iF^* - FP_iH^*(R + HP_iH^*)^{-1}HP_iF^* - P_i, \\ &= P_{i+1} - P_i \triangleq \delta P_i. \end{aligned}$$

Using the above expression for Z_i we can verify that $Z_i \star \mathcal{S}_{N,i} =$

$$\begin{bmatrix} \Phi_p(N,i)[I + \tilde{Q}_i\mathcal{O}_{N,i}]^{-1}\tilde{F}_i & P_N + \Phi_p(N,i)\tilde{Q}_i[I + \mathcal{O}_{N,i}\tilde{Q}_i]^{-1}\Phi_p^*(N,i) \\ -\tilde{H}_i^*\tilde{H}_i - \tilde{F}_i^*\mathcal{O}_{N,i}[I + \tilde{Q}_i\mathcal{O}_{N,i}]^{-1}\tilde{F}_i & \tilde{F}_i^*[I + \mathcal{O}_{N,i}\tilde{Q}_i]^{-1}\Phi_p^*(N,i) \end{bmatrix}$$

Likewise, we know that

$$\mathcal{S}_{N,i} \star M = \begin{bmatrix} \Phi_p(N+1,i) & P_{N+1} \\ -\mathcal{O}_{N+1,i} & \Phi_p^*(N+1,i) \end{bmatrix},$$

which is analogous to (17.6.15)–(17.6.17). By equating the $(1,2)$ block entries of $Z_i \star \mathcal{S}_{N,i}$ and $\mathcal{S}_{N,i} \star M$, and using $\tilde{Q}_i = \delta P_i$, we obtain the relation (which the reader can verify follows also from the change-in-initial-condition formula (17.6.29))

$$P_{N+1} = P_N + \Phi_p(N,i)\delta P_i[I + \mathcal{O}_{N,i}\delta P_i]^{-1}\Phi_p^*(N,i).$$

In particular, if we take $N = i+1$ then this expression shows that

$$\delta P_{i+1} = F_{p,i}\delta P_i(I + H^*R_{e,i}^{-1}H\delta P_i)^{-1}F_{p,i}^*, \tag{17.6.35}$$

since, by definition,

$$\mathcal{O}_{N+1,i} = \sum_{j=i}^{N} \Phi_p^*(j,i)H^*R_{e,j}^{-1}H\Phi_p^*(j,i),$$

and, consequently, $\mathcal{O}_{i+1,i} = H^* R_{e,j}^{-1} H$. This is exactly relation (11.1.7), which served as the basis for the derivation of the fast recursions in Thm. 11.1.2 (*cf.* Thm. 11.1.1). We can now proceed as in Ch. 11.

Doubling Algorithm. The homogeneity of the medium can also be used to immediately obtain a doubling algorithm, which among other things provides a fast way for computing the steady-state solution of the Riccati recursion (17.6.5).

Let $\mathcal{S}^o_{[i]}$ denote the scattering matrix of a cascade that is composed of i sections, $i \geq 1$, with zero initial conditions. Since the medium is homogeneous, we can write

$$\mathcal{S}^o_{[2^i]} = \mathcal{S}^o_{[2^{i-1}]} \star \mathcal{S}^o_{[2^{i-1}]}. \tag{17.6.36}$$

That is, the scattering matrix of a cascade of 2^i sections can be computed as the star product of the scattering matrix of 2^{i-1} sections with itself! If desired, we can write down explicit formulas. Let $\{\Phi^o_{2^i}, \mathcal{P}^o_{2^i}, \Phi^{o*}_{2^i}, -\mathcal{O}^o_{2^i}\}$ denote the scattering parameters of a cascade of 2^i elementary sections. Then, for $i \geq 0$,

$$\Phi^o_{2^{i+1}} = \Phi^o_{2^i} \left[I + \mathcal{P}^o_{2^i} \mathcal{O}^o_{2^i}\right]^{-1} \Phi^o_{2^i}, \tag{17.6.37}$$

$$\mathcal{P}^o_{2^{i+1}} = \mathcal{P}^o_{2^i} + \Phi^o_{2^i} \mathcal{P}^o_{2^i} \left[I + \mathcal{O}^o_{2^i} \mathcal{P}^o_{2^i}\right]^{-1} \Phi^{o*}_{2^i}, \tag{17.6.38}$$

$$\mathcal{O}^o_{2^{i+1}} = \mathcal{O}^o_{2^i} + \Phi^{o*}_{2^i} \mathcal{O}^o_{2^i} \left[I + \mathcal{P}^o_{2^i} \mathcal{O}^o_{2^i}\right]^{-1} \Phi^o_{2^i}, \tag{17.6.39}$$

with initial conditions given by

$$\Phi^o_1 = F, \quad \mathcal{P}^o_1 = GQG^*, \quad \mathcal{O}^o_1 = H^* R^{-1} H.$$

Algebraic derivations of these formulas are considerably less transparent (see, *e.g.*, Prob. 9.25).

17.7 FURTHER WORK

In addition to the examples and results we have described or cited, there are several other problems whose solutions can be obtained and illuminated by using the above ideas. For example, Levy et al. (1983) have used them to study certain decentralized estimation problems. Other examples arise from limited memory filtering problems (see Bruckstein and Kailath (1985)), whose direct solution is not easy. Also, from control theory (see, *e.g.*, Warrior and Viswanadham (1980)). And so on.

However, the major point is that there is still much more that can be done with this approach, because only a small part of the many existing results in scattering theory, even just in the single paper of Redheffer (1962), has been exploited so far. So there is scope for further work, and a nice way to end our book.

We might add a final thought. Given the rapid pace of technological change, it is conceivable that even the rather complex scattering medium introduced here, with matrix-valued transmission and reflection coefficients, can be implemented in some optical or opto-electronic technology. Then smoothed estimators could be obtained at light speed!

17.8 COMPLEMENTS

The best source is Redheffer (1962), who was a major contributor to the development of a general mathematical formulation of a theory that has origins in, and applications to, a number of fields — radiative transfer, neutron diffusion, microwave circuits, some probability theory calculations, and others. We shall not attempt to reproduce Redheffers's historical review here, except to mention that his earliest reference is Stokes (1862), in which difference equations were obtained for the reflection of light from a pile of identical glass plates. Using the fact that $m+n$ plates can be considered either as m plates followed by n, or vice versa, Stokes obtained certain commutativity relations (*cf.* Sec. 17.5.2). The continuous case was studied by McClelland (1906) who analyzed the effect of adding infinitesimally thin layers to a homogeneous medium. Schmidt (1907) also addressed the same issue by using a method that was later vigorously developed by Bellman, Kalaba, and their collaborators (see, *e.g.*, Bellman and Wing (1975) and also Scott (1974), which lists nearly a 1,000 papers on this topic) under the name "the method of invariant imbedding".

It was Redheffer, who had worked with electrical engineers at the MIT Radiation Laboratories, who recognized the close connections with transmission-line theory (as first studied by Heaviside) and began the development of the abstract theory, part of which we have described here. However, as mentioned earlier, there is much more in Redheffer (1962), as well as in Redheffer and Wang (1970), and in the monograph of Ribaric (1973).

For us, of course, the connection is that the Riccati equation so prominent in the above theory is also fundamental in estimation and control theory. This spurred the investigations described in a series of papers: Ljung, Kailath, and Friedlander (1976), Ljung and Kailath (1976a), Friedlander, Kailath and Ljung (1976), Kailath and Ljung (1977), Kailath (1979), and Verghese, Friedlander, and Kailath (1980). Levy et al. (1983) further applied the scattering approach to decentralized smoothing problems.

The main point is that by this means we effectively "linearize" the analysis of the Riccati equation and find an insightful way of organizing the "calculus" of Riccati equations, especially with regard to the interplay between forwards- and backwards-time evolution.

Finally we should mention that transmission line models are also natural objects in studying problems of signal transmission and reflection in layered-earth media and much work has been done in this area (see, *e.g.*, Claerbout (1976), Robinson and Treitel (1980), and Bruckstein and Kailath (1987a, 1987b)). The generalized Schur algorithm developed in App. F on Displacement Structure is natural for such problems, and for several other apparently very different problems. The difference from the Redheffer theory is that the latter explicitly incorporates state-space structure into the picture.

■ PROBLEMS

17.1 **(Redheffer's composition rule)** Refer to the discussion on Redheffer's star product in Sec. 17.3. Assume $a = d^*$, $A = D^*$ and (b, c, B, C) are all Hermitian. Show that $[A(I - bC)^{-1}a]^* = d(I - Cb)^{-1}D$. Hence, conclude that the forward and backward transmission operators will remain conjugates of each other.

17.2 **(Interchanging the columns of $\mathcal{S}^o(t,\tau)$)** Consider the matrix $\mathcal{X}^o(t,\tau)$ that is obtained from $\mathcal{S}^o(t,\tau)$ by interchanging its columns, as in (17.1.12). Verify that $\mathcal{X}^o(t,\tau)$ satisfies the Riccati differential equation

$$\frac{\partial \mathcal{X}^o(t,\tau)}{\partial t} = \begin{bmatrix} F & 0 \\ 0 & 0 \end{bmatrix} \mathcal{X}^o(t,\tau) + \mathcal{X}^o(t,\tau)\begin{bmatrix} F^* & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} GQG^* & 0 \\ 0 & 0 \end{bmatrix} - \mathcal{X}^o(t,\tau)\begin{bmatrix} H^*R^{-1}H & 0 \\ 0 & 0 \end{bmatrix}\mathcal{X}^o(t,\tau),$$

with initial condition

$$\mathcal{X}^o(\tau,\tau) = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix},$$

and where we are writing $\{F, G, H, R, Q\}$ instead of $\{F(t), G(t), H(t), R(t), Q(t)\}$.

17.3 **(Time and measurement updates via scattering)** Refer to the discussion in Sec. 17.6 and consider the generator matrix M_i in (17.6.7).

(a) Verify that M_i can be factored as

$$M_N = \begin{bmatrix} I & 0 \\ -H_N^*R_N^{-1}H_N & I \end{bmatrix} \star \begin{bmatrix} F_N & G_NQ_NG_N^* \\ 0 & F_N^* \end{bmatrix} \triangleq M_{m,N} \star M_{t,N}, \quad \text{say.}$$

We now implement the composition (17.6.11), *viz.*, $\mathcal{S}_{N,i} \star M_N$, as a sequence of two compositions, $(\mathcal{S}_{N,i} \star M_{m,N}) \star M_{t,N}$. The boundary layer is taken as $\mathcal{B}$ in (17.6.14).

(b) Show that

$$\begin{bmatrix} \Phi_p(N,i) & P_N \\ \times & \times \end{bmatrix} \star M_{m,N} = \begin{bmatrix} [I + P_NH_N^*R_N^{-1}H_N]^{-1}\Phi_p(N,i) & P_{N|N} \\ \times & \times \end{bmatrix},$$

where "$\times$" denote irrelevant entries.

(c) Define $\Phi_f(N,i) = [I + P_NH_N^*R_N^{-1}H_N]^{-1}\Phi_p(N,i)$. Show that

$$\begin{bmatrix} \Phi_f(N,i) & P_{N|N} \\ \times & \times \end{bmatrix} \star M_{t,N} = \begin{bmatrix} \Phi_p(N+1,i) & P_{N+1} \\ \times & \times \end{bmatrix}.$$

(d) Use

$$\begin{bmatrix} \Phi_f(N,i) & P_{N|N} \\ \times & \times \end{bmatrix} \star M_{t,N} \star M_{m,N+1} = \begin{bmatrix} \Phi_f(N+1,i) & P_{N+1|N+1} \\ \times & \times \end{bmatrix}$$

in order to update the filtered quantities $\{\Phi_f(N,i), P_{N|N}\}$ directly.

17.4 **(An expression for the observability Gramian)** Apply the matrix inversion formula of App. A to $\left[I + P_{N+1}H_{N+1}^*R_{N+1}^{-1}H_{N+1}\right]^{-1}$ and establish (17.6.20).

17.5 **(A formula for state prediction)** Consider the update relation

$$\hat{\mathbf{x}}_{N+2|N+1} = F_{N+1}\left[I + P_{N+1}H_{N+1}^*R_{N+1}^{-1}H_{N+1}\right]^{-1}\left[\hat{\mathbf{x}}_{N+1|N} + P_{N+1}H_{N+1}^*R_{N+1}^{-1}\mathbf{y}_{N+1}\right],$$

obtained in (17.6.22). Show that it collapses to the form

$$\hat{\mathbf{x}}_{N+2|N+1} = F_{N+1}\hat{\mathbf{x}}_{N+1|N} + K_{p,N+1}[\mathbf{y}_{N+1} - H_{N+1}\hat{\mathbf{x}}_{N+1|N}].$$

17.6 **(Forward equations from backward equations)** Consider the backwards-time differential equation (17.2.3),

$$-\frac{\partial \mathcal{P}^o(t,\tau)}{\partial \tau} = \Psi^o(t,\tau)G(\tau)Q(\tau)G^*(\tau)\Psi^{o*}(t,\tau), \quad \mathcal{P}^o(t,t) = 0.$$

Integrating both sides with respect to τ, and using the boundary condition, leads to

$$\mathcal{P}^o(t,\tau) = \int_\tau^t \Psi^o(t,\lambda)G(\lambda)Q(\lambda)G^*(\lambda)\Psi^{o*}(t,\lambda)d\lambda.$$

Show from the above expression that $\mathcal{P}^o(t,\tau)$ satisfies the forward Riccati differential equation (17.1.6), *viz.*,

$$\frac{d}{dt}\mathcal{P}^o(t,\tau) =$$

$$G(t)Q(t)G^*(t) + F(t)\mathcal{P}^o(t,\tau) + \mathcal{P}^o(t,\tau)F^*(t) - \mathcal{P}^o(t,\tau)H^*(t)R^{-1}(t)H(t)\mathcal{P}^o(t,\tau),$$

with zero boundary condition, $\mathcal{P}^o(t,t) = 0$.
[*Hint.* Recall from (17.1.7) that $\Psi^o(t,\lambda)$ satisfies

$$\frac{d}{dt}\Psi^o(t,\lambda) = [F(t) - K^o(t)H(t)]\Psi^o(t,\lambda),$$

where $K^o(t) \triangleq \mathcal{P}^o(t,\lambda)H^*(t)R^{-1}(t)$ is itself a function of both t and λ.]

17.7 **(Backwards equations via scattering)** We showed in Sec. 17.3.2 that the scattering matrix $\mathcal{S}(t,\tau)$, with initial conditions included, satisfies a backwards differential equation of the same form as (17.3.8). The entries of the corresponding infinitesimal generator, however, satisfy (17.3.17) when we choose as initial condition

$$\Gamma(\tau) = \begin{bmatrix} \Pi(\tau) & \Pi(\tau) \\ \Pi(\tau) & \Pi(\tau) \end{bmatrix},$$

with $\Pi(\tau)$ satisfying (17.3.16).

In the state-space context we have $f(t) = F(t)$ and

$$g(t) = G(t)Q(t)G^*(t), \quad h(t) = -H^*(t)R^{-1}(t)H(t), \quad e(t) = F^*(t),$$

and we take $\mathcal{S}(t,\tau)$ to be the scattering matrix that results from the composition $\Gamma(\tau) \star \mathcal{S}^o(t,\tau)$. We shall denote this scattering matrix by $\mathcal{S}^b(t,\tau)$ in order to differentiate it from the notation $\mathcal{S}(t,\tau)$ that we used in the text to denote $\begin{bmatrix} I & P(\tau) \\ 0 & I \end{bmatrix} \star \mathcal{S}^o(t,\tau)$.

(a) Use (17.3.17) to identify $\{f_N, g_N, h_N, e_N\}$ in the state-space context as

$$f_N(\tau) = -[F(\tau) + G(\tau)Q(\tau)G^*(\tau)\Pi^{-1}(\tau)]^*,$$
$$g_N(\tau) = -H^*(\tau)R^{-1}(\tau)H(\tau),$$
$$h_N(\tau) = G(\tau)Q(\tau)G^*(\tau),$$
$$e_N(\tau) = -[F(\tau) + G(\tau)Q(\tau)G^*(\tau)\Pi^{-1}(\tau)] \triangleq F^b(\tau).$$

(b) Denote the entries of $\mathcal{S}^b(t,\tau) = \Gamma(\tau) \star \mathcal{S}(t,\tau)$ by

$$\mathcal{S}^b(t,\tau) = \begin{bmatrix} S_{11}(t,\tau) & S_{12}(t,\tau) \\ S_{21}(t,\tau) & S_{22}(t,\tau) \end{bmatrix}.$$

Show that these entries satisfy the backwards differential equations

$$-\frac{d}{d\tau}S_{11}(t,\tau) = S_{11}(t,\tau)[F^{b*}(\tau) - H^*(\tau)R^{-1}(\tau)H(\tau)S_{21}(t,\tau],$$

$$-\frac{d}{d\tau}S_{12}(t,\tau) = -S_{11}(t,\tau)H^*(\tau)R^{-1}(\tau)H(\tau)S_{22}(t,\tau),$$

$$-\frac{d}{d\tau}S_{21}(t,\tau) = G(\tau)Q(\tau)G^*(\tau) + F^b(\tau)S_{21}(t,\tau) + S_{21}(t,\tau)F^{b*}(\tau) - S_{21}(t,\tau)H^*(\tau)R^{-1}(\tau)H(\tau)S_{21}(t,\tau),$$

$$-\frac{d}{d\tau}S_{22}(t,\tau) = [F^b(\tau) - S_{21}(t,\tau)H^*(\tau)R^{-1}(\tau)H(\tau)]S_{22}(t,\tau),$$

with boundary conditions

$$S_{11}(t,t) = \Pi(t), \quad S_{12}(t,t) = \Pi(t), \quad S_{21}(t,t) = \Pi(t), \quad S_{22}(t,\tau) = \Pi(t).$$

(c) Conclude that we can make the identifications:

$$S_{21}(t,\tau) = P^b(\tau), \quad S_{22}(t,\tau) = \Psi^b(t,\tau)\Pi(t), \quad S_{12}(t,\tau) = P(\tau|t),$$

where $P^b(t)$ is described in Thm. 16.A.1 and $\Psi^b(s,t)$ is the state-transition matrix of $F^b(s) - K^b(s)H(s)$, with $K^b(s) = P^b(s)H^*(s)R^{-1}(s)$.

Appendix for Chapter 17

17.A A COMPLEMENTARY STATE-SPACE MODEL

Consider again the standard state-space model (17.1.1)–(17.1.3), *viz.*,

$$\dot{\mathbf{x}}(s) = F(s)\mathbf{x}(s) + G(s)\mathbf{u}(s), \tag{17.A.1}$$

$$\mathbf{y}(s) = H(s)\mathbf{x}(s) + \mathbf{v}(s), \qquad 0 \le s \le t. \tag{17.A.2}$$

For some $\tau \in [0,t]$, let $\hat{\mathbf{x}}(\tau)$ denote the l.l.m.s. estimator of $\mathbf{x}(\tau)$ given the observations $\{\mathbf{y}(\nu), 0 \le \nu < \tau\}$, with the corresponding error variance matrix by $P(\tau) = \|\mathbf{x}(\tau) - \hat{\mathbf{x}}(\tau)\|^2$. Given the above state-space model, it is easy to see that we cannot recover the random variables $\{\mathbf{u}(\cdot), \mathbf{v}(\cdot), \mathbf{x}(\tau)\}$ just from knowledge of the output process $\{\mathbf{y}(\cdot)\}$, unless we have additional information.

It turns out that this additional information can be provided by a model that is said to be complementary to the given model (17.A.1)–(17.A.2) in a certain sense. We studied this issue in Secs. 15.7.2 and 15.7.3 for discrete-time state-space realizations. In fact, the same framework that we developed in Ch. 15 for the description of dual bases can be extended to continuous-time random processes. We shall not pursue the details here, except to note that a backwards complementary model for (17.A.1)–(17.A.2) can be obtained as follows.

We define random variables $\{\boldsymbol{\eta}(s), \boldsymbol{\theta}\}$ by the *backwards-time* state-space model:

$$\dot{\boldsymbol{\xi}}(s) = -F^*(s)\boldsymbol{\xi}(s) - H^*(s)R^{-1}(s)\mathbf{v}(s), \quad \boldsymbol{\xi}(t) = 0, \tag{17.A.3}$$

$$\boldsymbol{\eta}(s) = -G^*(s)\boldsymbol{\xi}(s) + Q^{-1}(s)\mathbf{u}(s), \quad \tau \le s \le t, \tag{17.A.4}$$

with

$$\boldsymbol{\theta} \triangleq -P(\tau)\boldsymbol{\xi}(\tau) + \mathbf{x}(\tau). \tag{17.A.5}$$

Then one can check by direct calculation that $\{\boldsymbol{\theta}, \boldsymbol{\eta}(\cdot)\}$ form a basis for the orthogonal complement space of $\mathcal{Y} = \mathcal{L}\{\mathbf{y}(s),\ \tau \le s \le t\}$ in the larger space $\mathcal{L}\{\mathbf{x}(\tau), \mathbf{u}(s), \mathbf{v}(s),\ \tau \le s \le t\}$. That is,

$$\mathcal{L}\{\boldsymbol{\theta}, \boldsymbol{\eta}(s),\ \tau \le s \le t\} = \mathcal{Y}^{\perp}.$$

Now, as in Sec. 15.7.3, we can combine the equations (17.A.1)–(17.A.2) and (17.A.3)–(17.A.4) and write

$$\begin{bmatrix} \dot{\mathbf{x}}(s) \\ \dot{\boldsymbol{\xi}}(s) \end{bmatrix} = \begin{bmatrix} F(s) & 0 \\ 0 & -F^*(s) \end{bmatrix} \begin{bmatrix} \mathbf{x}(s) \\ \boldsymbol{\xi}(s) \end{bmatrix} + \begin{bmatrix} G(s) & 0 \\ 0 & -H^*(s)R^{-1}(s) \end{bmatrix} \begin{bmatrix} \mathbf{u}(s) \\ \mathbf{v}(s) \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{y}(s) \\ \boldsymbol{\eta}(s) \end{bmatrix} = \begin{bmatrix} H(s) & 0 \\ 0 & -G^*(s) \end{bmatrix} \begin{bmatrix} \mathbf{x}(s) \\ \boldsymbol{\xi}(s) \end{bmatrix} + \begin{bmatrix} 0 & I \\ Q^{-1}(s) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}(s) \\ \mathbf{v}(s) \end{bmatrix}.$$

Then, by elimination, we get

$$\begin{bmatrix}\dot{\mathbf{x}}(s)\\ \dot{\boldsymbol{\xi}}(s)\end{bmatrix} = \begin{bmatrix} G(s)Q(s)\boldsymbol{\eta}(s) \\ -H^*(s)R^{-1}(s)\mathbf{y}(s)\end{bmatrix} + \begin{bmatrix} F(s) & G(s)Q(s)G^*(s) \\ H^*(s)R^{-1}(s)H(s) & -F^*(s)\end{bmatrix}\begin{bmatrix}\mathbf{x}(s)\\ \boldsymbol{\xi}(s)\end{bmatrix} \tag{17.A.6}$$

with the two-point boundary value conditions

$$\boldsymbol{\xi}(t) = 0, \quad \mathbf{x}(\tau) - P(\tau)\boldsymbol{\xi}(\tau) = \boldsymbol{\theta}. \tag{17.A.7}$$

If we now project both sides of (17.A.6)–(17.A.7) onto the space spanned by the observations $\{\mathbf{y}(s),\ 0 \le s \le t\}$, over the entire interval $[0, t]$, we obtain

$$\begin{bmatrix}\dot{\hat{\mathbf{x}}}(s|t)\\ \dot{\hat{\boldsymbol{\xi}}}(s|t)\end{bmatrix} = \begin{bmatrix} 0 \\ -H^*(s)R^{-1}(s)\mathbf{y}(s)\end{bmatrix} + \begin{bmatrix} F(s) & G(s)Q(s)G^*(s) \\ H^*(s)R^{-1}(s)H(s) & -F^*(s)\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}(s|t)\\ \hat{\boldsymbol{\xi}}(s|t)\end{bmatrix},$$

where we used the fact that $\boldsymbol{\eta}(s)$ is not only orthogonal to $\{\mathbf{y}(s),\ \tau \le s \le t\}$ but also to the earlier observations, $\{\mathbf{y}(s),\ 0 \le s < \tau\}$. Moreover, the boundary conditions are given by

$$\hat{\boldsymbol{\xi}}(t|t) = 0, \quad \hat{\mathbf{x}}(\tau|t) - P(\tau)\hat{\boldsymbol{\xi}}(\tau|t) = \hat{\boldsymbol{\theta}},$$

where $\hat{\boldsymbol{\theta}}$ denotes the projection of $\boldsymbol{\theta}$ onto the space spanned by the observations $\{\mathbf{y}(s),\ 0 \le s \le t\}$. Now since $\boldsymbol{\theta}$ is, by construction, orthogonal to the observations $\{\mathbf{y}(s),\ \tau \le s \le t\}$, we see that in order to evaluate $\hat{\boldsymbol{\theta}}$ we only need to consider its projection onto the space spanned by the earlier observations $\{\mathbf{y}(s),\ 0 \le s < \tau\}$. Using (17.A.5), and the easily verified fact that $\boldsymbol{\xi}(\tau)$ is orthogonal to $\{\mathbf{y}(s),\ 0 \le s < \tau\}$, we find that $\hat{\boldsymbol{\theta}} = \hat{\mathbf{x}}(\tau)$ so that the boundary conditions become

$$\hat{\boldsymbol{\xi}}(t|t) = 0, \quad \hat{\mathbf{x}}(\tau|t) - P(\tau)\hat{\boldsymbol{\xi}}(\tau|t) = \hat{\mathbf{x}}(\tau).$$

Finally, if we define $\boldsymbol{\lambda}(s|t) = \hat{\boldsymbol{\xi}}(s|t)$, then the above equations reduce to the Hamiltonian equations (17.1.4).

The point is that the equations are directly obtained by a more detailed study of the state-space model. Then, as shown in this chapter, by further transmission line analysis, we can get many old and new state-space estimation results.

APPENDIX A
Useful Matrix Results

To make our presentation in this book more self-contained, we collect here several facts and formulas from matrix theory. Proofs are generally only given when it is felt that they would enhance understanding and ease of recall. At the end of the appendix, we present a list of some books on matrix theory and linear algebra that we have found interesting and useful (in one way or another).

A.1 SOME MATRIX IDENTITIES

(i) **Block Gaussian Elimination and Schur Complements** Consider a block matrix

$$M = \begin{bmatrix} A & B \\ C & D\end{bmatrix}$$

that we wish to triangularize by a (block) Gaussian elimination procedure. For this, note that

$$\begin{bmatrix} I & 0 \\ X & I\end{bmatrix}\begin{bmatrix} A & B \\ C & D\end{bmatrix} = \begin{bmatrix} A & B \\ XA + C & XB + D\end{bmatrix},$$

so that choosing $X = -CA^{-1}$ gives[1]

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I\end{bmatrix}\begin{bmatrix} A & B \\ C & D\end{bmatrix} = \begin{bmatrix} A & B \\ 0 & \Delta_A\end{bmatrix},$$

where

$$\Delta_A \triangleq D - CA^{-1}B$$

[1] We assume throughout the appendix that inverses exist whenever needed.

is called the *Schur complement of A in M*. Similarly we can find that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & \Delta_A \end{bmatrix}.$$

So also we can obtain

$$\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \Delta_D & 0 \\ C & D \end{bmatrix},$$

and

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \begin{bmatrix} \Delta_D & B \\ 0 & D \end{bmatrix},$$

where $\Delta_D = A - BD^{-1}C$ is the *Schur complement* of D in M.
Schur complements are very useful objects in matrix theory and we shall encounter them often. Good references on the history, properties, and extensions of the Schur complement are Cottle (1974), Ando (1979), and Ouellette (1981).

(ii) **Determinants** Using the product rule for determinants, the results in (i) give

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det A \det (D - CA^{-1}B) = \det A \det \Delta_A, \tag{A.1.1}$$

$$= \det D \det (A - BD^{-1}C) = \det D \det \Delta_D. \tag{A.1.2}$$

These formulas were perhaps first given by I. Schur (1917) and were the reason for the name Schur complement bestowed by E. V. Haynsworth (1968).

(iii) **Block Triangular Factorizations** The results in (i) can be combined to block-diagonaliz M by noting that

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix},$$

and

$$\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \begin{bmatrix} \Delta_D & 0 \\ 0 & D \end{bmatrix}.$$

Then by using the easily verified formula

$$\begin{bmatrix} I & 0 \\ P & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -P & I \end{bmatrix},$$

we can obtain the direct factorization formulas

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}, \tag{A.1.3}$$

$$= \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} \Delta_D & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}. \tag{A.1.4}$$

(iv) **Recursive Triangularization and LDU Decomposition** An alternative way of writing the above formulas is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} A^{-1}\begin{bmatrix} A & B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \Delta_A \end{bmatrix}, \tag{A.1.5}$$

$$= \begin{bmatrix} B \\ D \end{bmatrix} D^{-1}\begin{bmatrix} C & D \end{bmatrix} + \begin{bmatrix} \Delta_D & 0 \\ 0 & 0 \end{bmatrix}, \tag{A.1.6}$$

which also serve to *define* the Schur complements Δ_A and Δ_D. The above formulas can be used recursively to respectively obtain the block "lower-upper" and block "upper-lower" triangular factorizations of the matrix on the left-hand side.
In particular, by choosing A to be scalar and proceeding recursively e can obtain the important LDU decomposition of a *strongly* regular matrix, *i.e.,* one whose leading minors are all nonzero. To demonstrate this so-called Gauss-Schur reduction procedure, let R be an $n \times n$ strongly regular matrix whose individual entries we denote by r_{ij}. Let also l_0 and u_0 denote the first column and the first row of R, respectively. In view of (A.1.5), we see that if we subtract from R the rank-one matrix $l_0 r_{00}^{-1} u_0$, then we obtain a new matrix whose first row and column are zero,

$$R - l_0 r_{00}^{-1} u_0 = \begin{bmatrix} 0 & 0 \\ 0 & R_1 \end{bmatrix}.$$

The matrix R_1 is the Schur complement of R with respect to its $(0,0)$ entry r_{00}. Now, let $\{r_{ij}^{(1)}, l_1, u_1\}$ denote the entries, the first column, and the first row of R_1, respectively, and repeat the above procedure. In general, we can write for the j-th step

$$R_j - l_j \left[r_{00}^{(j)}\right]^{-1} u_j = \begin{bmatrix} 0 & 0 \\ 0 & R_{j+1} \end{bmatrix}.$$

We conclude that we express R in terms of the successive $\{l_i, u_i, r_{00}^{(i)}\}$ as follows:

$$R = l_0 r_{00}^{-1} u_0 + \begin{bmatrix} 0 \\ l_1 \end{bmatrix}\left[r_{00}^{(1)}\right]^{-1}\begin{bmatrix} 0 & u_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ l_2 \end{bmatrix}\left[r_{00}^{(2)}\right]^{-1}\begin{bmatrix} 0 & 0 & u_2 \end{bmatrix} + \dots$$

$$\triangleq LD^{-1}U,$$

where L is lower triangular, D^{-1} is diagonal, and U is upper triangular. The nonzero parts of the columns of L are the $\{l_i\}_{i=0}^{n-1}$, while the nonzero parts of the rows of U are the $\{u_i\}_{i=0}^{n-1}$. Likewise, the entries of D are the $\{r_{00}^{(i)}\}_{i=0}^{n-1}$. We can further normalize the diagonal entries of L and U and define $\bar{L} = LD^{-1}$ and $\bar{U} = D^{-1}U$. In this case, we obtain $R = \bar{L}D\bar{U}$ and the diagonal entries of $\bar{L}$ and $\bar{U}$ are unity.

It is also possible to verify that the $\bar{L}D\bar{U}$ factorization of a strongly regular matrix R is unique. Indeed, assume there exist two decompositions of the form $R = L_1D_1U_1 = L_2D_2U_2$, where $\{L_1, L_2\}$ are lower-triangular with unit diagonal, $\{D_1, D_2\}$ are diagonal, and $\{U_1, U_2\}$ are upper triangular with unit diagonal. Then it must hold that

$$L_2^{-1}L_1 = D_2U_2U_1^{-1}D_1^{-1}.$$

Now the left-hand side matrix in the above equality is lower triangular while the right-hand side matrix is upper triangular. Hence, equality holds only if both matrices are diagonal. But since the diagonal entries of $L_2^{-1}L_1$ are unity, it follows that we must have

$$L_2^{-1}L_1 = D_2U_2U_1^{-1}D_1^{-1} = I, \qquad \text{the identity matrix,}$$

from which we conclude that $L_1 = L_2$, $D_1 = D_2$, and $U_1 = U_2$.

(v) **Inverses of Block Matrices** When the block matrix is invertible, we can use the factorizations in (iii) to write

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix},$$

$$= \begin{bmatrix} A^{-1} + A^{-1}B\Delta_A^{-1}CA^{-1} & -A^{-1}B\Delta_A^{-1} \\ -\Delta_A^{-1}CA^{-1} & \Delta_A^{-1} \end{bmatrix}. \tag{A.1.7}$$

Alternatively, we can write

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix}\begin{bmatrix} \Delta_D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix},$$

$$= \begin{bmatrix} \Delta_D^{-1} & -\Delta_D^{-1}BD^{-1} \\ -D^{-1}C\Delta_D^{-1} & D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1} \end{bmatrix}. \tag{A.1.8}$$

By equating the (1,1) and (2,2) elements in the right-hand sides of (A.1.7) and (A.1.8), we note that

$$\Delta_D^{-1} = A^{-1} + A^{-1}B\Delta_A^{-1}CA^{-1},$$
$$\Delta_A^{-1} = D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1}.$$

(vi) **More Inverse Formulas** Another useful set of formulas can be obtained from the formulas in (v) (and a little algebra):

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & -\Delta_D^{-1}BD^{-1}\Delta_A \\ -D^{-1}C & I \end{bmatrix}\begin{bmatrix} \Delta_D^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix},$$

$$= \begin{bmatrix} I & -A^{-1}B \\ -D^{-1}C & I \end{bmatrix}\begin{bmatrix} \Delta_D^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix},$$

and similarly

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \Delta_D^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix}\begin{bmatrix} I & -BD^{-1} \\ -CA^{-1} & I \end{bmatrix}.$$

We also have formulas analogous to (A.1.5) and (A.1.6):

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix}\Delta_A^{-1}\begin{bmatrix} -CA^{-1} & I \end{bmatrix}, \tag{A.1.9}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -D^{-1}C \end{bmatrix}\Delta_D^{-1}\begin{bmatrix} I & -BD^{-1} \end{bmatrix}. \tag{A.1.10}$$

(vii) **The Matrix Inversion Lemma** For convenience of recall, replacing C by $-D$ and D by C^{-1}, we can rewrite the above formula for Δ_D^{-1} as

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1},$$

which is often called, following Woodbury (1950), the modified matrices formula. The formula was used in estimation problems by Kailath (1960,1961), and then by Ho (1963), who called it the Matrix Inversion lemma, a designation now common in the state-space literature. A nice account of the origins of this formula, and of several variations and extensions, is given by (Henderson and Searle, (1981)). For example, if C is not invertible, we can write

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(I + CDA^{-1}B)^{-1}CDA^{-1}.$$

(viii) **Inertia Properties of Hermitian Matrices** The Hermitian conjugate, A^*, of a matrix A is the complex conjugate of its transpose. Hermitian matrices are (necessarily square) matrices obeying $A^* = A$. For strongly regular Hermitian matrices, the unique LDU decomposition takes the form

$$A = LDL^*, \qquad L = \text{lower triangular with unit diagonal.}$$

The simple proof is instructive. If $A = LDU$, then $A^* = U^*D^*L^* = U^*DL^*$, since D is real-valued. But by uniqueness of triangular factorization, we must have $U = L^*$.

There is another important decomposition for Hermitian matrices, using the fact that such matrices have real eigenvalues, say $\{\lambda_i\}$, and a full set of orthonormal

eigenvectors, say $\{p_i\}$. The spectral (or modal) decomposition of a Hermitian matrix is the representation

$$A = P\Lambda P^* = \sum_{i=1}^{n} \lambda_i p_i p_i^*,$$

where

$$\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_n\}, \quad P = p_1 \ \ldots \ p_n, \quad Ap_i = \lambda_i p_i, \ i = 1, \ldots, n.$$

Now since the eigenvalues of a Hermitian matrix $A = A^*$ are real, we can define the inertia of A as the triple $In\{A\} = \{n_+, n_-, n_0\}$, where n_+ is the number of positive (> 0) eigenvalues of A, n_- is the number of negative (< 0) eigenvalues of A, and n_0 is the number of zero eigenvalues of A. Note that $n_+ + n_- =$ the rank of A, while n_0 is often called the *nullity of A*. The *signature* of A is the pair $\{n_+, n_-\}$. We shall define, S_A, the signature matrix of A, as a a diagonal matrix with n_+ ones $(+1)$ and n_- minus ones (-1) on the diagonal. It is not necessary to compute the eigenvalues of A in order to determine its inertia or its signature matrix. The LDL^* decomposition suffices:

Lemma A.1.1 (Sylvester's Law of Inertia) *For any nonsingular matrix B, it holds that $In\{A\} = In\{BAB^*\}$.* ∎

The matrices A and BAB^* are said to be *congruent* to each other, so Sylvester's law states that congruence preserves inertia. The following useful result follows easily from the above and the factorizations in (iii).

Lemma A.1.2 (Inertia of Block Hermitian Matrices) *Let*

$$M = \begin{bmatrix} A & C^* \\ C & D \end{bmatrix}, \quad A = A^*, \quad D = D^*.$$

(a) *If A is nonsingular, then $In\{M\} = In\{A\} + In\{\Delta_A\}$, where $\Delta_A = D - CA^{-1}C^*$, the Schur complement of A in M.*

(b) *If D is nonsingular, then $In\{M\} = In\{D\} + In\{\Delta_D\}$, where $\Delta_D = A - C^*D^{-1}C$, the Schur complement of D in M.*

∎

(ix) **Positive Definite Matrices** An $n \times n$ Hermitian matrix A is positive-semi-definite (p.s.d. — also often called nonnegative-definite (n.n.d.)), written $A \geq 0$, if it satisfies

$$x^*Ax \geq 0 \quad \text{for all } x \in \mathbb{C}^n.$$

It is positive definite (p.d.), written $A > 0$, if $x^*Ax > 0$ except when $x = 0$. Among the several characterizations of $A \geq 0$, we note the nonnegativity of all its eigenvalues and the fact that *all* minors are nonnegative. For positive-definiteness it is necessary and sufficient that the leading minors be positive. An often more computationally useful characterization is that nonnegative-definite matrices can be factored as $A = LDL^*$ or $A = UDU^*$, where all entries of the diagonal matrix D are nonnegative.

From the results of (viii), we note that a Hermitian block matrix

$$M = \begin{bmatrix} A & C^* \\ C & D \end{bmatrix}$$

is positive-definite if, and only if, either $A > 0$ and $\Delta_A > 0$, or $D > 0$ and $\Delta_D > 0$.

A.2 KRONECKER PRODUCTS

The Kronecker product of two matrices is a useful matrix operation and is defined as follows. Let $A = [a_{ij}]_{i,j=1}^{m}$ and $B = [b_{ij}]_{i,j=1}^{n}$ be two matrices with eigenvalues

$$\lambda(A) = \{\alpha_1, \alpha_2, \ldots, \alpha_m\} \quad \text{and} \quad \lambda(B) = \{\beta_1, \beta_2, \ldots, \beta_n\}.$$

The Kronecker (or tensor) product of A and B, written $A \otimes B$, is an $mn \times mn$ matrix defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \ldots & a_{1m}B \\ a_{21}B & a_{22}B & \ldots & a_{2m}B \\ \vdots & & & \vdots \\ a_{m1}B & a_{m2}B & \ldots & a_{mm}B \end{bmatrix}. \tag{A.2.1}$$

For example, if A is the identity matrix I_m, then $I_m \otimes B$ is simply a block diagonal matrix with B repeated m times along its diagonal.

Kronecker products have many interesting properties. We collect some of them in the following statement, where we use the following notation: $\text{vec}(P)$ denotes a column vector that is obtained by stacking the columns of a matrix P one on top of another, *i.e.*, if $p_1, p_2, \ldots, p_n$ denote the columns of P then $\text{vec}(P) = \text{col}\{p_1, p_2, \ldots, p_n\}$.

Lemma A.2.1 (Properties of Kronecker Products) *Given an $m \times m$ matrix A, an $n \times n$ matrix B, an $m \times k$ matrix C, and an $n \times p$ matrix D, the following facts hold:*

(i) $(A \otimes B)(C \otimes D) = AC \otimes BD$.

(ii) If A and B are invertible, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

(iii) The mn eigenvalues of $(A \otimes B)$ are $\{\alpha_i\beta_j\}$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

(iv) $\det(A \otimes B) = (\det A)^n (\det B)^m$.

(v) For any matrices $\{A, B, C, X\}$ of compatible dimensions, if $C = AXB$, then $vec(C) = (B^T \otimes A)vec(X)$.

∎

Proof: Part (i) can be verified by direct calculation. Part (ii) follows from part (i) by choosing $C = A^{-1}$ and $D = B^{-1}$. Part (iii) also follows from (i) by choosing C as a right eigenvector for A and D as a right eigenvector for B. Part (iv) follows from part (iii). Part (v) follows from the definition of Kronecker products and from the fact that for two column vectors a and b, $\text{vec}(ab^T) = b \otimes a$. ◆

A useful application of Kronecker products is in the study of matrix equations. Consider, for example, the Lyapunov equation $P = FPA + Q$, which is studied in more detail in App. D. Then, using property (v) above, we see that $\text{vec}(P)$ can be obtained by solving $(I - A^T \otimes F)\text{vec}(P) = \text{vec}(Q)$. That is, the use of the Kronecker product notation allows us to express the solution of a Lyapunov equation in the form of the solution of a linear equation. Such reductions are useful and are used, for example, in App. D to derive several useful results concerning Lyapunov equations from well-known properties for linear equations.

A.3 THE REDUCED AND FULL QR DECOMPOSITIONS

An $N \times n$, $n \leq N$, matrix

$$A = [a_0 \ a_1 \ \dots \ a_{n-1}], \qquad a_i \in \mathbb{C}^N,$$

can be written as

$$A = \hat{Q}\hat{R},$$

where $\hat{R} \in \mathbb{C}^{n \times n}$ is upper triangular and

$$\hat{Q} = [q_0 \ q_1 \ \dots \ q_{n-1}], \qquad q_i \in \mathbb{C}^N,$$

has orthonormal columns, *i.e.*, $\langle q_i, q_j \rangle = q_j^* q_i = \delta_{ij}$. When A has rank n, the usual case for us, all diagonal entries of $\hat{R}$ are positive.

The easiest way of understanding this result is to note that the above decomposition is converting (in a "causal" manner) the columns of A, *i.e.*, the set of vectors $\{a_i\}$, to an equivalent set of orthonormal vectors $\{q_i\}$. The natural way of doing this is via the classical Gram-Schmidt procedure: let $q_0 = a_0/\|a_0\|$ and $q_i = r_i/\|r_i\|$ for $i > 0$, where

$$r_i = a_i - \sum_{j=0}^{i-1} \langle a_i, q_j \rangle \, q_j, \qquad \langle a_i, q_j \rangle = q_j^* a_i.$$

This procedure was discussed in Sec. 4.2 for elements in any linear vector space and in particular the space of random variables. When applied to vectors in $\mathbb{C}^N$, and when the $\{a_i\}$ are linearly independent (so that all $\|r_i\| > 0$), the above calculation can be rearranged as

$$a_i = (q_0^* a_i) q_0 + (q_1^* a_i) q_1 + \dots + (q_{i-1}^* a_i) q_{i-1} + \sqrt{r_i^* r_i}\, q_i,$$

or in matrix form,

$$A = \hat{Q}\hat{R}, \tag{A.3.1}$$

where $\hat{Q}$ is as defined above and R is $n \times n$ upper triangular with entries

$$\hat{R} = \begin{bmatrix} \sqrt{r_0^* r_0} & q_0^* a_1 & q_0^* a_2 & \dots & q_0^* a_{n-1} \\ & \sqrt{r_1^* r_1} & q_1^* a_2 & \dots & q_1^* a_{n-1} \\ & & \ddots & & \vdots \\ & & & \ddots & \vdots \\ & & & & \sqrt{r_{n-1}^* r_{n-1}} \end{bmatrix}.$$

The factorization in (A.3.1) is often called the *reduced* QR decomposition of A.

It is sometimes convenient to use a so-called *full* QR decomposition, in which case we append an additional $N - n$ orthonormal columns to $\hat{Q}$ so that it becomes a unitary $N \times N$ matrix. Correspondingly, we append rows of zeros to $\hat{R}$ so that

$$A = QR \triangleq [\hat{Q} \ q_n \ \dots \ q_{N-1}] \begin{bmatrix} \hat{R} \\ 0 \end{bmatrix}.$$

The orthogonalization procedure so described is not reliable numerically due to the accumulation of round-off errors in finite-precision arithmetic. So in practice other methods — modified Gram-Schmidt, Givens and Householder rotations (see App. B) are used. For numerical issues, and for the modifications when A is not full rank, we refer to textbooks on numerical linear algebra.

One application of the QR decomposition of a matrix is a simple proof of the following result (see Lemma A.5.1 below for a more general statement).

Lemma A.3.1 (Basis Rotation) *Given two $n \times m$ ($n \leq m$) full rank matrices A and B. Then $AA^* = BB^*$ if, and only if, there exists an $m \times m$ unitary matrix Θ ($\Theta\Theta^* = I = \Theta^*\Theta$) such that $A = B\Theta$.* ■

Proof: The if implication is immediate. One proof for the converse implication invokes the QR decompositions

$$A^* = Q_A \begin{bmatrix} \hat{R}_A \\ 0 \end{bmatrix}, \qquad B^* = Q_B \begin{bmatrix} \hat{R}_B \\ 0 \end{bmatrix},$$

where Q_A and Q_B are $m \times m$ unitary matrices, and $\hat{R}_A$ and $\hat{R}_B$ are $n \times n$ upper triangular matrices with positive diagonal entries. The equality $AA^* = BB^*$ implies that $AA^* = \hat{R}_A^* \hat{R}_A = \hat{R}_B^* \hat{R}_B$. Hence, $\hat{R}_A^*$ and $\hat{R}_B^*$ are Cholesky factors for the same matrix AA^*. By uniqueness of the triangular factorization (see the discussion in Sec. A.1 (iv)) we must therefore have $\hat{R}_A = \hat{R}_B$. Now define $\Theta = Q_B Q_A^*$. Then $\Theta\Theta^* = I$ and $B\Theta = A$. ◆

THE SINGULAR VALUE DECOMPOSITION AND APPLICATIONS

The singular value decomposition (SVD, for short) is a very powerful analytical and numerical tool with a long history. It states that if A is $n \times m$, then there exist an $n \times n$ unitary matrix U ($UU^* = I$), an $m \times m$ unitary matrix V ($VV^* = I$), and a diagonal matrix Σ with nonnegative entries such that

(i) if $n \leq m$, then Σ is $n \times n$ and

$$A = U[\Sigma \ \ 0]V^*.$$

(ii) if $n \geq m$, then Σ is $m \times m$ and

$$A = U\begin{bmatrix}\Sigma \\ 0\end{bmatrix}V^*.$$

The diagonal entries of Σ are called the singular values of A and are usually ordered in decreasing order, say $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$. If Σ has p nonzero diagonal entries, then A has rank p. The columns of U and V are called the left- and right-singular vectors of A, respectively. The SVD is discussed in detail in many textbooks. Here we mention that it has several important applications, a few of which are described below and in the following sections.

Spectral Norm of a Matrix.
The induced 2-norm of a matrix A, also known as the spectral norm of the matrix, is defined by

$$\|A\|_2 \triangleq \max_{\|x\|\neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|z\|=1} \|Az\|,$$

where $\|x\|$ denotes the Euclidean norm of the vector x. It can be shown that $\|A\|_2$ is equal to the maximum singular value of A, $\|A\|_2 = \sigma_1(A)$.

In fact, there are other ways of defining the norm of a matrix. For example, the so-called Frobenius norm,

$$\|A\|_F \triangleq \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|^2}, \tag{A.4.1}$$

can also be expressed in terms of the singular values of A; it is easy to verify that if A has rank p with nonzero singular values $\{\sigma_1, \ldots, \sigma_p\}$, then

$$\|A\|_F = \sqrt{\sum_{i=1}^{p}\sigma_i^2}.$$

Pseudo-Inverses.
Given an $n \times m$ matrix A, the SVD can be used to define its so-called $m \times n$ (Moore-Penrose) *pseudo-inverse*, $A^\dagger$.[2] Let

$$\Sigma^\dagger \triangleq \text{diagonal } \{\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_p^{-1}, 0, \ldots, 0\}.$$

Then,

(i) when $n \leq m$, we define

$$A^\dagger = V\begin{bmatrix}\Sigma^\dagger \\ 0\end{bmatrix}U^*. \tag{A.4.2}$$

(ii) when $n \geq m$, we define

$$A^\dagger = V[\Sigma^\dagger \ \ 0]U^*. \tag{A.4.3}$$

The pseudo-inverse has the following four (characterizing) properties:

(i) $AA^\dagger A = A$, (ii) $A^\dagger AA^\dagger = A^\dagger$, (iii) $(AA^\dagger)^* = AA^\dagger$, (iv) $(A^\dagger A)^* = A^\dagger A$.

It can be further verified that:

1. When $n \leq m$ and A is full rank, $A^\dagger = A^*(AA^*)^{-1}$.
2. When $n \geq m$ and A is full rank, $A^\dagger = (A^*A)^{-1}A^*$. These two properties can be verified by direct calculation by replacing A by its SVD.
3. $A^\dagger$ is the unique solution to the approximation problem (see, *e.g.*, Golub and Van Loan (1996))

$$\min_{X\in\mathbb{C}^{m\times n}} \|AX - I_n\|_F,$$

where I_n is the $n \times n$ identity matrix and $\|\cdot\|_F$ denotes the so-called Frobenius norm of a matrix, *viz.*, for a matrix $A = [a_{ij}]$,

$$\|A\|_F \triangleq \sqrt{\sum_{i=0}^{M}\sum_{j=0}^{m}|a_{ij}|^2}. \tag{A.4.4}$$

This norm can also be expressed in terms of the singular values of A as follows:

$$\|A\|_F = \sqrt{\sum_{i=1}^{p}\sigma_i^2}.$$

Minimum-Norm Solution of Least-Squares Problems.
Another useful application of the pseudo-inverse is in the determination of the minimum norm solution of a least-squares problem with an infinite number of solutions.

[2] An illuminating (and early) discussion of SVDs and pseudo-inverses is given in a classic book by Lanczos (1956), who used the term *natural inverse*.

Lemma A.4.1 (Minimum-Norm LS Solution) *Given $H \in \mathbb{C}^{N\times n}$ with $N \geq n$ and possibly rank deficient, the problem*

$$\min_{x\in\mathcal{X}} \|x\|, \quad \textit{where} \quad \mathcal{X} = \{x \in \mathbb{C}^n \textit{ such that } \|y - Hx\| \textit{ is minimum}\}$$

has a unique solution that is given by $\hat{x} = H^\dagger y$. ∎

Proof: Assume H has rank $p \leq n$ and introduce its SVD, say $U^*HV = \text{col}\{\Sigma, 0\}$ with $\Sigma = \text{diag}\{\sigma_1, \ldots, \sigma_p, 0, \ldots, 0\}$. Then $\|y - Hx\|^2 = \|U^*y - U^*HVV^*x\|^2$, since V and U are unitary matrices. Define $z = V^*x$ and let $\{z_i\}$ denote its individual entries. Note that x and z define each other uniquely. Let also $\{f_i\}$ denote the individual entries of U^*y. Then $\|y - Hx\|^2 = \sum_{i=1}^{p} |f_i - \sigma_i z_i|^2 + \sum_{i=p+1}^{m} |f_i|^2$. The second term is independent of x. Hence, any solution z has to satisfy $z_i = f_i/\sigma_i$ for $i = 1$ to p and z_i arbitrary for $i = p+1$ to $i = m$. The one with the smallest Euclidean norm requires that these latter values be chosen as zero. In this case, the minimum norm solution becomes $\hat{x} = V\,\text{col}\{f_1/\sigma_1, \ldots, f_p/\sigma_p, 0, \ldots, 0\}$, which is equivalent to $\hat{x} = H^\dagger y$. ◆

Minimum-Norm Solution of Underdetermined Systems.

A similar issue arises in the solution of underdetermined linear systems of equations (see also App. 2.A). Given an $H \in \mathbb{C}^{N\times n}$, a linear system $Hx = y$ is said to be underdetermined whenever $N < n$; it corresponds to a system of equations with more unknowns than equations. Hence, it may either have no solution (when y does not lie in the column span of H) or an infinite number of solutions (which can occur whether H is full rank or not). Among all possible solutions, we can seek the one with minimum Euclidean norm.

Lemma A.4.2 (Minimum Norm Solution) *Consider an underdetermined linear system of equations $Hx = y$, where $H \in \mathbb{C}^{N\times n}$ with $N < n$. Then $x = H^\dagger y$ is the minimum-norm solution when multiple solutions exist.* ∎

Proof: Assume H has rank $p \leq N$ and introduce its SVD, say $H = U\Sigma\ 0V^*$ with $\Sigma = \text{diag}\{\sigma_1, \ldots, \sigma_p, 0, \ldots, 0\}$. Then $Hx = y$ is equivalent to $U^*HVV^*x = U^*y$. Define $z = V^*x$ and let $\{z_i\}_{i=1}^n$ denote its individual entries. Note that x and z define each other uniquely. Let also $\{f_i\}_{i=1}^N$ denote the individual entries of U^*y. We are then reduced to solving

$$\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{bmatrix} \bigcirc \begin{bmatrix} z_1 \\ \vdots \\ z_p \\ z_{p+1} \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} f_1 \\ \vdots \\ f_p \\ f_{p+1} \\ \vdots \\ f_N \end{bmatrix}, \quad N < n.$$

It is clear that if $\{f_j \neq 0\}$ for some $j > p$, then the system of equations is inconsistent and solutions do not exist. Otherwise, $U^*HVz = f$ is satisfied for all $\{z_i\}$ such that $\sigma_i z_i = f_i$ for $i = 1, \ldots, p$, and $\{z_i\}$ arbitrary for $i > p$. Hence, the solution z with the smallest Euclidean norm in this case requires that these latter values be taken as zero. In this case, as in the minimum-norm solution of the LS problem, we again obtain that $x = H^\dagger y$ is the minimum-norm solution to the underdetermined system of equations $Hx = y$ (when multiple solutions exist). ◆

The Condition Number of a Matrix.

The condition number of a matrix A is denoted by $\kappa(A)$ and defined as the ratio of its largest singular value to its smallest singular value,

$$\kappa(A) \triangleq \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}. \tag{A.4.5}$$

Note that $\kappa(A)$ always satisfies $\kappa(A) \geq 1$. A unitary matrix Q achieves the lower bound, $\kappa(Q) = 1$. A close-to-singular matrix, on the other hand, has a small $\sigma_{\min}$ and hence, a very large condition number. A singular matrix has $\sigma_{\min} = 0$ and hence an infinite condition number.

To demonstrate the usefulness of the condition number, consider the solution of a square and invertible linear system of equations $Ax = y$ on a digital computer. If small perturbations in y lead to large variations in the solution, then A is said to be ill-conditioned. On the other hand, if small perturbations in y do not significantly change the solution, then A is said to be well conditioned. We shall see that the condition number provides a measure of the conditioning of A.

Indeed, assume y changes to $y + \delta y$ and denote the new solution by $x + \delta x$. Then

$$x = A^{-1}y \quad \text{and} \quad (x + \delta x) = A^{-1}(y + \delta y),$$

which implies that the perturbation in the solution satisfies $\delta x = A^{-1}\delta y$. Consequently,

$$\|\delta x\| \leq \sigma_{\max}(A^{-1})\ \|\delta y\| = \frac{\|\delta y\|}{\sigma_{\min}(A)}.$$

But since $Ax = y$ and, hence, $\|y\| \leq \sigma_{\max}(A)\|x\|$, we conclude that

$$\|\delta x\| \cdot \|y\| \leq \frac{\|\delta y\|}{\sigma_{\min}(A)} \quad [\sigma_{\max}(A)\|x\|],$$

from which we obtain

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)\ \frac{\|\delta y\|}{\|y\|},$$

so "large" condition numbers are "bad." However, there are several other issues involved, and the pros and cons of the condition number as a measure of numerical behavior are exposed in the literature of numerical linear algebra (and are outside our scope here).

A.5 BASIS ROTATIONS

Here we present several results, some of which appear here for the first time. We begin by noting that the SVD can be used to give an alternative proof (without full rank assumptions) of Lemma A.3.1, which was established via the QR decomposition. It also can be used to prove two less known results on J-unitary rotations.

Lemma A.5.1 (Basis Rotation) *Let A and B be $n \times m$ $(n \leq m)$ matrices. Then $AA^* = BB^*$ if, and only if, there exists an $m \times m$ unitary matrix Θ $(\Theta\Theta^* = I = \Theta^*\Theta)$ such that $A = B\Theta$.* ∎

Proof: One implication is immediate. If there exists a unitary matrix Θ such that $A = B\Theta$, then $AA^* = (B\Theta)(B\Theta)^* = B(\Theta\Theta^*)B^* = BB^*$. One proof for the converse implication follows by invoking the singular value decompositions of A and B,

$$A = U_A \begin{bmatrix} \Sigma_A & 0 \end{bmatrix} V_A^*, \qquad B = U_B \begin{bmatrix} \Sigma_B & 0 \end{bmatrix} V_B^*,$$

where U_A and U_B are $n \times n$ unitary matrices, V_A and V_B are $m \times m$ unitary matrices, and Σ_A and Σ_B are $n \times n$ diagonal matrices with nonnegative entries. The squares of the diagonal entries of Σ_A (Σ_B) are the eigenvalues of AA^* (BB^*). Moreover, U_A (U_B) can be constructed from an orthonormal basis for the right eigenvectors of AA^* (BB^*). Hence, it follows from the identity $AA^* = BB^*$ that we have $\Sigma_A = \Sigma_B$ and $U_A = U_B$. Let $\Theta = V_B V_A^*$. We then get $\Theta\Theta^* = I$ and $B\Theta = A$. ◆

We can establish a similar result when the equality $AA^* = BB^*$ is replaced by $AJA^* = BJB^*$ for some signature matrix J. More specifically, we have the following statement.

Lemma A.5.2 (J-Unitary Transformations) *Let A and B be $n \times m$ matrices (with $n \leq m$), and let $J = (I_p \oplus -I_q)$ be a signature matrix with $p+q = m$. If $AJA^* = BJB^*$ is full rank, then there exists a J-unitary matrix Θ such that $A = B\Theta$.* ∎

Proof: One proof that invokes the so-called hyperbolic SVDs of A and B is given in Ackner (1991). Here we provide an alternative proof that avoids the need for the hyperbolic SVD.[3]

Since AJA^* is Hermitian and invertible, we can factor it as $AJA^* = RSR^*$ where $R \in \mathbb{C}^{n\times n}$ is invertible and $S = (I_\alpha \oplus -I_\beta)$ is a signature matrix (with $\alpha + \beta = n$). We normalize A and B by defining $\bar{A} = R^{-1}A$ and $\bar{B} = R^{-1}B$. Then $\bar{A}J\bar{A}^* = \bar{B}J\bar{B}^* = S$.

Now consider the block triangular factorizations

$$\begin{bmatrix} S & \bar{A} \\ \bar{A}^* & J \end{bmatrix} = \begin{bmatrix} I & \\ \bar{A}^*S & I \end{bmatrix} \begin{bmatrix} S & \\ & J - \bar{A}^*S\bar{A} \end{bmatrix} \begin{bmatrix} I & \\ \bar{A}^*S & I \end{bmatrix}^*,$$

$$= \begin{bmatrix} I & \bar{A}J \\ & I \end{bmatrix} \begin{bmatrix} \underbrace{S - \bar{A}J\bar{A}^*}_{=0} & \\ & J \end{bmatrix} \begin{bmatrix} I & \bar{A}J \\ & I \end{bmatrix}^*.$$

[3] This argument was suggested to the authors by T. Constantinescu.

Using the fact that the central matrices must have the same inertia we conclude that $In\{J - \bar{A}^*S\bar{A}\} = In\{J\} - In\{S\} = \{p-\alpha, q-\beta, n\}$. Similarly, we can show that $In\{J - \bar{B}^*S\bar{B}\} = \{p-\alpha, q-\beta, n\}$.

Define the signature matrix $J_1 \triangleq (I_{p-\alpha} \oplus -I_{q-\beta})$. The above inertia conditions then mean that we can factor $(J - \bar{A}^*S\bar{A})$ and $(J - \bar{B}^*S\bar{B})$ as

$$J - \bar{A}^*S\bar{A} = XJ_1X^*, \qquad J - \bar{B}^*S\bar{B} = YJ_1Y^*, \qquad X, Y \in \mathbb{C}^{m\times m-n}.$$

Finally, introduce the square matrices

$$\Sigma_1 = \begin{bmatrix} \bar{A} \\ X^* \end{bmatrix}, \qquad \Sigma_2 = \begin{bmatrix} \bar{B} \\ Y^* \end{bmatrix}.$$

It is easy to verify that these matrices satisfy $\Sigma_1^*(S \oplus J_1)\Sigma_1 = J$ and $\Sigma_2^*(S \oplus J_1)\Sigma_2 = J$, which further shows that Σ_1 and Σ_2 are invertible. Therefore, we also obtain that $\Sigma_1 J \Sigma_1^* = (S \oplus J_1)$ and $\Sigma_2 J \Sigma_2^* = (S \oplus J_1)$. These relations allow us to relate Σ_1 and Σ_2 as $\Sigma_1 = \Sigma_2[J\Sigma_2^*(S \oplus J_1)\Sigma_1]$. If we set $\Theta = [J\Sigma_2^*(S \oplus J_1)\Sigma_1]$, then it is immediate to check that Θ is J-unitary and, from the equality of the first block row of $\Sigma_1 = \Sigma_2\Theta$, that $\bar{A} = \bar{B}\Omega$. Hence, $A = B\Theta$. ◆

Remark. Consider the example $A = [1\ \ 1]$, $B = [0\ \ 0]$ and $J = \mathrm{diag}\{1, -1\}$. Then $AJA^* = BJB^*$. However, it is easy to see that there *cannot* exist a J-unitary matrix Θ that maps the zero vector B to the nonzero vector A. This case is of course ruled out by the full rank assumption on AJA^* in the statement of Lemma A.5.2. This requirement can be weakened by simply requiring AJA^* to be nonzero, as we show next. ◆

Lemma A.5.3 (A Generalization) *Let A and B be $n \times m$ matrices (with $n \leq m$), and let $J = (I_p \oplus -I_q)$ be a signature matrix with $p+q = m$. If $AJA^* = BJB^*$ is nonzero then there exists a $J-$unitary matrix Θ such that $A = B\Theta$.* ∎

Proof: Assume AJA^* has rank $\lambda \leq n$. Now since AJA^* is Hermitian and nonzero, we can factor it as $AJA^* = RSR^*$ where $R \in \mathbb{C}^{n\times\lambda}$ is full rank and $S = (I_\alpha \oplus -I_\beta)$ is a signature matrix (with $\alpha + \beta = \lambda$). In contrast to the proof of Lemma A.5.3, we now normalize A and B by defining $\bar{A} = R^\dagger A$ and $\bar{B} = R^\dagger B$, where $R^\dagger = (R^*R)^{-1}R^*$ is the pseudo-inverse of R. Then $\bar{A}J\bar{A}^* = \bar{B}J\bar{B}^* = S$.

We can proceed as in the earlier proof and verify that $In\{J - \bar{A}^*S\bar{A}\} = \{p-\alpha, q-\beta, \lambda\} = In\{J - \bar{B}^*S\bar{B}\}$. We also define $J_1 = (I_{p-\alpha} \oplus -I_{q-\beta})$ and introduce the factorizations

$$J - \bar{A}^*S\bar{A} = XJ_1X^*, \qquad J - \bar{B}^*S\bar{B} = YJ_1Y^*, \qquad X, Y \in \mathbb{C}^{m\times m-\lambda}.$$

Finally, we introduce the square matrices

$$\Sigma_1 = \begin{bmatrix} \bar{A} \\ X^* \end{bmatrix}, \qquad \Sigma_2 = \begin{bmatrix} \bar{B} \\ Y^* \end{bmatrix},$$

and show that the desired transformation is $\Theta = [J\Sigma_2^*(S \oplus J_1)\Sigma_1]$. ◆

In the above statements, the arrays A and B are either square or have more columns than rows ($n \leq m$). We can establish a similar result when $n \geq m$ instead. For this purpose, first note that if A is an $n \times m$ and full rank matrix, with $n \geq m$, then its SVD takes the form

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*,$$

where Σ is $n \times n$ and invertible. The *left* inverse of A can then be defined by

$$A^\dagger \triangleq V \begin{bmatrix} \Sigma^{-1} & 0 \end{bmatrix} U^*,$$

and it satisfies $A^\dagger A = I_m$, the identify matrix of size m.

Lemma A.5.4 (J-Unitary Transformations) *Let A and B be $n \times m$ full rank matrices (with $n \geq m$), and let $J = (I_p \oplus -I_q)$ be a signature matrix with $p + q = m$. The relation $AJA^* = BJB^*$ holds if, and only if, there exists a unique $m \times m$ $J-$unitary matrix Θ such that $A = B\Theta$.* ∎

Proof: The "if" statement is immediate. For the converse, note that since A and B are assumed full rank, there exist left inverses $A^\dagger$ and $B^\dagger$ such that $A^\dagger A = I_m$ and $B^\dagger B = I_m$. Now define $\Theta = B^\dagger A$. We claim that Θ is $J-$unitary and maps B to A, as desired.

The proof that $\Theta J \Theta^* = J$ is immediate from the equality $AJA^* = BJB^*$. Just multiply it from the left by $B^\dagger$ and from the right by $(B^\dagger)^*$ and use $B^\dagger B = I_m$.

To prove that $B\Theta = A$, for the above choice of Θ, we start with $AJA^* = BJB^*$ again and insert the term $B^\dagger B$ into the right-hand side to get

$$AJA^* = B(B^\dagger B)JB^* = BB^\dagger(BJB^*) = BB^\dagger AJA^*.$$

Multiplying from the right by $(A^\dagger)^*$ and using $A^\dagger A = I_m$ we obtain $BB^\dagger(AJ) = AJ$. Since J is invertible and its inverse is J, we conclude by multiplying by J from the right that $B(B^\dagger A) = A$, which is the desired result. That is, Θ is J-unitary and maps B to A.

To show that Θ is unique, assume $\bar{\Theta}$ is another J-unitary matrix that maps B to A and write $B\Theta = B\bar{\Theta}$. Now multiply by $B^\dagger$ from the left to conclude that $\Theta = \bar{\Theta}$. ◆

A.6 COMPLEX GRADIENTS AND HESSIANS

Consider a scalar real-valued function $g(z)$ of a scalar complex variable $z = x + jy$, where $j \triangleq \sqrt{-1}$. We can regard $g(z)$ as a real-valued function of two real-valued scalar variables, x and y, say $g(x, y)$. A stationary point (x_o, y_o) of $g(x, y)$ can be determined by setting the partial derivatives of g with respect to both x and y equal to zero, *viz.*, $[\partial g/\partial x]_{x_o} = 0$, $[\partial g/\partial y]_{y_o} = 0$. More compactly, these two conditions are equivalent to the "complex gradient", denoted by $\partial g/\partial z$, being equal to zero, where we define

$$\frac{\partial g}{\partial z} \triangleq \frac{1}{2}\left\{\frac{\partial g}{\partial x} - j\,\frac{\partial g}{\partial y}\right\}.$$

Likewise, we define the complex conjugate gradient of $g(\cdot)$ with respect to z^* as

$$\frac{\partial g}{\partial z^*} = \frac{1}{2}\left\{\frac{\partial g}{\partial x} + j\,\frac{\partial g}{\partial y}\right\}.$$

These definitions are also used to define complex gradients for functions $g(z)$ that are possibly complex-valued (see, *e.g.*, Schwartz (1967) for more details).

The Scalar Case.

We illustrate the above definition by considering several examples of scalar-valued functions $g(z)$ of a *scalar* parameter z.

1. Let $g(z) = z = x + jy$. Then $\partial g/\partial z = (1 - j^2)/2 = 1$ and $\partial g/\partial z^* = (1 + j^2)/2 = 0$.
2. Let $g(z) = z^2 = (x+jy)(x+jy) = (x^2 - y^2) + j2xy$. Then $\partial g/\partial z = 2(x+jy) = 2z$ and $\partial g/\partial z^* = 0$. These results are consistent with what we would expect from the definition of the derivative of a real-valued function of a real variable.
3. Consider next $g(z) = \lambda + \alpha z + \beta z^* + \gamma z z^*$, where $(\lambda, \alpha, \beta, \gamma)$ are complex constants. That is, $g(z) = [\lambda + \alpha x + \beta x + \gamma(x^2 + y^2)] + j\,[\alpha y - \beta y]$. Then $\partial g/\partial z = \alpha + \gamma z^*$ and $\partial g/\partial z^* = \beta + \gamma z$.

The Vector Case.

Now assume that $g(\cdot)$ is a function of a *column vector*

$$z = \mathrm{col}\{z_1, z_2, \ldots, z_n\}, \quad z_i = x_i + jy_i.$$

The gradient of $g(\cdot)$ with respect to z will be defined as the *row* vector

$$\frac{\partial g}{\partial z} \triangleq \begin{bmatrix} \frac{\partial g}{\partial z_1} & \frac{\partial g}{\partial z_2} & \cdots & \frac{\partial g}{\partial z_n} \end{bmatrix}.$$

Likewise, the conjugate gradient of $g(\cdot)$ with respect to z^* is defined as a *column* vector

$$\frac{\partial g}{\partial z^*} \triangleq \mathrm{col}\left\{\frac{\partial g}{\partial z_1^*}, \ldots, \frac{\partial g}{\partial z_n^*}\right\}.$$

Finally, the second derivative of a scalar function $g(\cdot)$ with respect to a column vector variable z is known as the *Hessian* matrix and will be denoted by

$$\frac{\partial^2 g}{\partial z^* \partial z} = \left[\frac{\partial^2 g}{\partial z_i^* \partial z_j}\right]_{i,j=1,\ldots,n}.$$

Consider, for example, $g(z) = \lambda + \alpha z + z^*\beta + z^*\Gamma z$, where λ is a scalar, α is a row vector, β is a column vector, and Γ is a matrix. Then $\partial g/\partial z = \alpha + z^*\Gamma$ and $\partial g/\partial z^* = \beta + \Gamma z$, while the Hessian matrix is $\partial^2 g/\partial z^* \partial z = \Gamma$.

The Real Case.
When z is real, say $z = x$, and say $g(x) = \lambda + \alpha x + x^T\beta + x^T\Gamma x$ (with a column vector x), then the gradient of g with respect to x is defined by

$$\frac{\partial g}{\partial x} \triangleq \left[\frac{\partial g}{\partial x_1} \quad \frac{\partial g}{\partial x_2} \quad \cdots \quad \frac{\partial g}{\partial x_n}\right],$$

so that we obtain $\partial g/\partial x = \alpha + \beta^T + 2x^T\Gamma$. Likewise, the Hessian matrix becomes $\partial^2 g/\partial x^2 = 2\Gamma$. Notice the difference from the complex case.

A.7 FURTHER READING

This appendix is only a brief review of several matrix results useful for this book. Interested readers can find more detailed expositions of matrix theory, linear algebra, and numerical linear algebra in many books. Here are some, among several others, especially in matrix theory and linear algebra.

Matrix Theory. Classic references are Gantmacher (1959, a compendium in two volumes), Bellman (2nd ed.,1970, an elegant presentation with interesting problems and references), Lancaster and Tismenetsky (2nd ed., 1985, a good treatment of stability theory). Ch. 1 of Rao (1973) describes several special results encountered in statistical estimation theory. The two volumes by Horn and Johnson (1985,1991) go into several advanced topics. ◆

Linear Algebra. Strang (1993), Lay (1994), Jänich (1996), Axler (1996), Lax (1997). ◆

Numerical Linear Algebra. Watkins (1991), Patel, Laub, and Van Dooren (1994), Datta (1995), Golub and Van Loan (3rd ed., 1996), Higham (1996), Trefethen and Bau (1997), Demmel (1997), Stewart (1998), Datta (2000). The books of Patel, Laub, and Van Dooren (1994) and of Stewart (1998) are perhaps the closest to the topics in our book. ◆

Deterministic Least-Squares Methods. Gauss (ca. 1820; translation by Stewart (1995)), Lawson and Hanson (1995), Björck (1996). Elementary presentations of deterministic least-squares can be found in almost all books on linear algebra. ◆

Fast Algorithms. Heinig and Rost (1984) and Kailath and Sayed (1999). ◆

APPENDIX B
Unitary and J-Unitary Transformations

In this appendix we review three families of elementary (unitary and hyperbolic) transformations that can be used to annihilate selected entries in a vector and thereby reduce a matrix to triangular form. These are the Householder, Givens, and fast(arithmetic square-root free) Givens transformations. Special care needs to be taken when dealing with complex-valued data as compared to real-valued data, as we shall explain in the sequel.

B.1 HOUSEHOLDER TRANSFORMATIONS

Householder transformations annihilate several entries in a row vector at a time. In the discussion that follows we distinguish between vectors with real-valued entries and vectors with complex-valued entries. We consider the real case first.

Real-Valued Data.
Consider an n-dimensional real-valued row vector x and suppose that we wish to simultaneously annihilate several entries in it by using a symmetric orthogonal matrix Θ (*i.e.*, a transformation Θ that satisfies $\Theta = \Theta^T$ and $\Theta^2 = I$).[1] The desired effect is to transform x to the form

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} \end{bmatrix} \Theta = \alpha\, e_0, \tag{B.1.1}$$

for some real scalar α to be determined, and where e_0 denotes the first basis vector, $e_0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$.

The scalar α cannot be arbitrary and its value can in fact be determined a priori, even before determining the expression for a matrix Θ that achieves (B.1.1). Indeed, it follows from the orthogonality of Θ and from (B.1.1) that

$$x\Theta\Theta^T x^T = xx^T = \|x\|^2 = \alpha^2,$$

[1] Matrices Q that satisfy $Q^2 = I$ are called involutary matrices.

so that we must have $\alpha = \pm\|x\|$. Both values of α are possible (since if Θ achieves $x\Theta = \|x\|e_0$, then $-\Theta$ is orthogonal and achieves $x\Theta = -\|x\|e_0$).

We still need to determine Θ. One way to achieve the transformation is to employ a so-called *Householder reflection*. We shall motivate this fact and also derive an expression for Θ, via a geometric argument.

Thus refer to Fig. B.1. Assume α has been chosen as above. Now given the vector x, we would like to transform it by a matrix Θ such that the resulting vector $x\Theta$ is aligned with the basis vector e_0 (the careful reader will soon realize that the argument applies equally well to alignments along other vector directions). This transformation should keep the norm of x unchanged, and the vector aligned with e_0 is clearly equal to αe_0. The triangle with sides x and αe_0, and base $g = x - \alpha e_0$, is therefore an isosceles triangle.

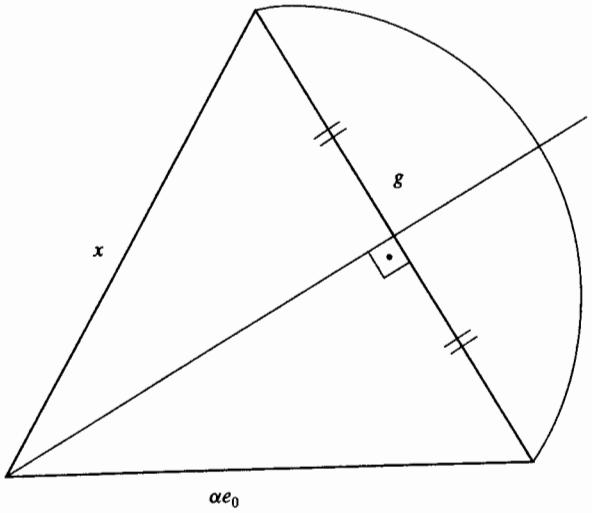


Figure B.1 Geometric interpretation of the Householder transformation.

We thus have $\alpha e_0 = x - g$. But we can also express g in an alternative form as follows. If we drop a perpendicular from the origin of x to the vector g, it will divide the segment g into two equal parts. Moreover, the upper part is nothing but the projection of the vector x onto the vector g and is thus equal to $\langle x, g\rangle \|g\|^{-2} g$. Therefore,

$$\alpha e_0 = x - 2\langle x, g\rangle \|g\|^{-2} g. \tag{B.1.2}$$

This transformation is called a reflection because it reflects the vector x across the perpendicular to g (see also the footnote further ahead).

When x and g are row vectors, we get $\langle x, g\rangle = xg^T$ so that the above equation reduces to

$$\alpha e_0 = x - 2xg^T(gg^T)^{-1}g = x\underbrace{\left[I - 2\frac{g^Tg}{gg^T}\right]}_{\Theta}, \tag{B.1.3}$$

where the matrix denoted by Θ,

$$\Theta \triangleq I - 2\frac{g^Tg}{gg^T}, \tag{B.1.4}$$

is orthogonal and involutary, as desired. It is called an (elementary) Householder transformation.[2]

Next note that to represent the transformation Θ in matrix form we in fact have two choices, depending upon whether we represent vectors as rows (as done above) or as columns. In the first case, we obtain (B.1.4). In the second case, when vectors are represented as columns, we have $\langle x, g\rangle = g^Tx$ and $\|g\|^2 = g^Tg$, so that

$$\alpha e_0^T = x - 2(g^Tx)(g^Tg)^{-1}g = \left(I - 2\frac{gg^T}{g^Tg}\right)x,$$

which leads to

$$\Theta \triangleq I - 2\frac{gg^T}{g^Tg}. \tag{B.1.5}$$

In summary, we have established the following result.

Lemma B.1.1 (Real Householder Transformation) *Given a real-valued row vector x, define $g = x \pm \|x\|e_0$ and Θ as in (B.1.4). Then it holds that $x\Theta = \mp\|x\|e_0$. A similar statement holds for column vectors x with Θ given by (B.1.5), $x\Theta$ replaced by Θx, and e_0 replaced by e_0^T.* ∎

The choice of the sign in $g = x \pm \|x\|e_0$ is usually dictated by the desire to avoid a vector g of small Euclidean norm, since this norm appears in the denominator of the expression defining Θ. This can be guaranteed by choosing the sign in the expression for g to be the same as the sign of the leading entry of the row vector x, *viz.*, the sign of x_1.

Triangularizing a Matrix.

A sequence of Householder transformations of this type can be used to triangularize a given $m \times n$ matrix, say A. For this we first find a transformation Θ_0 to rotate the first

[2] We remark that the multiplication of a vector by an orthogonal matrix Q can correspond to a rotation (if $\det Q = 1$) or a reflection (if $\det Q = -1$). For the Householder matrix Θ in (B.1.4), we see that it is a rank-one modification of the identity matrix and has $(n-1)$ eigenvalues at 1 and a single eigenvalue at -1. Therefore, $\det\Theta = -1$, which confirms again that it is a reflector.

row to lie along e_0, so that we have $A\Theta_0$ of the form (where $\times$ denotes entries whose exact values are not of current interest):

$$A\Theta_0 = \left[\begin{array}{c|ccc} \alpha & 0 & 0 & 0 \\ \hline \times & & & \\ \times & & A_1 & \\ \times & & & \end{array}\right].$$

Now apply a transformation of the type $(1 \oplus \Theta_1)$, where Θ_1 rotates the first row of A_1 so that it lies along e_0 in an $(n-1)$-dimensional space, and so on. A numerical example to this effect is given in Sec. 12.4.2.

This so-called Householder reduction of matrices has been found to be an efficient and stable tool for displaying rank information via matrix triangularization and it is widely used in numerical analysis (see, *e.g.*, Stewart (1973), Golub and Van Loan (1996), Higham (1996)).

Complex-Valued Data.
When the entries of x are complex-valued, we should seek a Hermitian unitary matrix Θ (*i.e.*, one satisfying $\Theta = \Theta^*$ and $\Theta^2 = I$) that leads to the transformation (B.1.1) with a scalar α that is possibly complex-valued. The value of α, however, is still not arbitrary and it will have to meet two constraints in this case: one on its norm and another on its phase.

So let us introduce the polar representation $\alpha = |\alpha|\, e^{j\phi_\alpha}$, where $|\alpha|$ denotes the magnitude of α and ϕ_α its phase, and let us also write $x_1 = |x_1|\, e^{j\phi_{x_1}}$, for the leading entry of the row vector x. It then follows from the unitarity of Θ and from (B.1.1) that

$$x\Theta\Theta^* x^* \;=\; xx^* \;=\; \|x\|^2 \;=\; |\alpha|^2,$$

so that we must have $|\alpha| = \|x\|$. Moreover, it follows from (B.1.1) that $x\Theta x^* = \alpha x_1^*$. But since Θ is Hermitian, we conclude that $x\Theta x^*$ is a real number and, hence, αx_1^* must be real as well. Moreover, since

$$\alpha x_1^* \;=\; |\alpha|\;|x_1|\; e^{j[\phi_\alpha - \phi_{x_1}]},$$

and since $e^{j\beta}$ is a unit-modulus complex number for any β, we conclude that the phases of α and x_1 must be such that

$$e^{j[\phi_\alpha - \phi_{x_1}]} \;=\; \pm 1.$$

This guarantees $\alpha x_1^* = \pm|\alpha|\;|x_1|$; a real quantity. Solving this equation for α we conclude that there are only two possible choices for α when $x_1 \neq 0$ and these are given by

$$\alpha = \pm\|x\| e^{j\phi_{x_1}}.$$

When $x_1 = 0$ we can choose $\alpha = \pm\|x\|$, which is a special case of the above for $\phi_{x_1} = 0$. Once α has been chosen as above, we can repeat the geometric argument of the real-valued case to obtain the following result.

Lemma B.1.2 (Complex Householder Transformation) *Given a complex-valued row vector x with leading entry x_1, define Θ and g as*

$$\Theta \triangleq I - 2\,\frac{g^* g}{gg^*}, \qquad g \triangleq x \pm \|x\| e^{j\phi_{x_1}}\, e_0. \tag{B.1.6}$$

Then it holds that $x\Theta = \mp\|x\| e^{j\phi_{x_1}} e_0$. As in Lemma B.1.1, a similar statement holds for column vectors x. ■

(Algebraic) proof: We provide an alternative algebraic derivation here as well. For this purpose, we write g more compactly as $g = x + \alpha e_0$, where α satisfies $|\alpha|^2 = \|x\|^2$ and αx_1^* is real. Then direct calculation shows that

$$\begin{aligned} gg^* &= 2\|x\|^2 + 2\alpha x_1^*, \\ xg^*g &= x\|x\|^2 + \alpha\|x\|^2 e_0 + \alpha^* x_1 x + \alpha(\alpha x_1^*)e_0, \\ xgg^* &= 2x\|x\|^2 + 2\alpha x_1^* x, \end{aligned}$$

so that

$$x\Theta = \frac{xgg^* - 2xg^*g}{gg^*} = \frac{-(2\|x\|^2 + 2\alpha x_1^*)\alpha e_0}{2\|x\|^2 + 2\alpha x_1^*} = -\alpha e_0.$$

◆

B.2 CIRCULAR OR GIVENS ROTATIONS

If the matrix to be triangularized already has several zeros in it, the Householder method can be wasteful because it may change many of these zero values. Therefore, a "spot" method of introducing zeros can be useful. The Givens rotations provide a nice way of doing this. They do so by pivoting with an entry of choice in order to annihilate another entry. Thus it is enough to explain their operation on 2-dimensional row vectors. We again distinguish between real data and complex data.

Real-Valued Data.
An elementary 2×2 orthogonal rotation Θ (also known as Givens or circular rotation) takes a 1×2 row vector $x = [a \;\; b]$ and rotates it to lie along the basis vector $e_0 = [1 \;\; 0]$. More precisely, it performs the transformation

$$[a \;\; b]\,\Theta = [\alpha \;\; 0], \tag{B.2.1}$$

for some real number α to be determined. An expression for Θ that achieves the transformation (B.2.1) is given by

$$\Theta = \frac{1}{\sqrt{1+\rho^2}} \begin{bmatrix} 1 & -\rho \\ \rho & 1 \end{bmatrix} \quad \text{where} \quad \rho = \frac{b}{a}, \quad a \neq 0. \tag{B.2.2}$$

Indeed, it can be verified by direct calculation that this Θ leads to

$$[a \;\; b\Theta = \pm\sqrt{a^2 + b^2} \;\; 0],$$

where the sign of α depends on whether the value of the square root in the expression (B.2.2) for Θ is chosen to be negative or positive.

Note further that (B.2.2) can be expressed in the alternative form

$$\Theta = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}, \quad c = \frac{1}{\sqrt{1+\rho^2}}, \quad s = \frac{\rho}{\sqrt{1+\rho^2}}, \quad \rho = \frac{b}{a},$$

where c and s can be interpreted as cosine and sine terms. This justifies the name *circular rotation* for Θ, since the effect of Θ is to rotate a vector x along a circle of radius $\|x\|$, by an angle $\theta = \tan^{-1}\rho$, in order to align it with the basis vector $[1 \ 0]$.

When $a = 0$, we simply choose Θ to be the permutation matrix

$$\Theta = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{B.2.3}$$

The net effect in this case is to rotate a by 90 degrees in a clockwise direction (if $b \geq 0$) or in an anticlockwise direction (if $b < 0$).

Readers may note that we could also use elementary reflections defined by matrices of the form

$$\Theta = \frac{1}{\sqrt{1+\rho^2}} \begin{bmatrix} 1 & \rho \\ \rho & -1 \end{bmatrix}, \tag{B.2.4}$$

which have determinant equal to -1.

Triangularizing a Matrix.

Matrix triangularization can also be effected by a product of Givens transformations, each of which introduces a zero in a particular location. For example, suppose that $x = [\times \ \times \ x_i \ \times \ \times \ x_j \ \times \ \times]$, and we wish to null out x_j using x_i. Then let $\rho = x_j/x_i$, and define

$$\Theta = \begin{bmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & \frac{1}{\sqrt{1+\rho^2}} & & & \frac{-\rho}{\sqrt{1+\rho^2}} & & \\ & & & 1 & & & & \\ & & & & 1 & & & \\ & & \frac{\rho}{\sqrt{1+\rho^2}} & & & \frac{1}{\sqrt{1+\rho^2}} & & \\ & & & & & & 1 & \\ & & & & & & & 1 \end{bmatrix}.$$

Except for the ρ terms, all off-diagonal entries are 0. Then we can see that

$$x\Theta = [\times \ \times \ \alpha \ \times \ \times \ 0 \ \times \ \times],$$

where the entries indicated by $\times$ are arbitrary and unchanged by Θ, while the resulting α will be of the general form

$$\alpha = \pm \sqrt{x_i^2 + x_j^2}.$$

To systematically triangularize an $n \times n$ matrix A, apply a sequence of $n-1$ such transformations to zero all entries in the first row of A except for the first element. Proceed to the second row of the thus transformed A matrix and apply a sequence of $n-2$ transformations to zero all entries after the second one. The first row has a zero in every column affected by this sequence of transformations, so it will be undisturbed. Continuing in this fashion for all rows except the last one, we will transform A to lower triangular form. A numerical example is given in Sec. 12.4.1.

In general, the Givens method of triangularization requires more computations than the Householder method. It requires about 30% more multiplications, and it requires one scalar square root per zero produced as opposed to one per column for the Householder method. However, the Givens method is more flexible in preserving zeros already present in the A matrix and can require fewer computations than the Householder method when A is nearly triangular to begin with (see Gentleman (1973)). Moreover, there is a fast version, presented next, that uses 50% fewer multiplications.

Complex-Valued Data.

When the entries of $x = [a \ \ b]$ are possibly complex-valued, the expression for Θ that achieves the transformation (B.2.1) becomes

$$\Theta = \frac{1}{\sqrt{1+|\rho|^2}} \begin{bmatrix} 1 & -\rho \\ \rho^* & 1 \end{bmatrix} \quad \text{where} \quad \rho = \frac{b}{a}, \quad a \neq 0. \tag{B.2.5}$$

In this case, it can be verified that we obtain

$$[a \ \ b]\Theta = [\pm e^{j\phi_a}\sqrt{|a|^2 + |b|^2} \ \ 0],$$

where we are writing $a = |a|e^{j\phi_a}$ to denote the polar representation of a. Again, the sign of α depends on whether the value of the square root in (B.2.5) is chosen to be negative or positive. When $a = 0$, we choose Θ as the permutation matrix (B.2.3).

B.3 FAST GIVENS TRANSFORMATIONS

Gentleman (1973) showed how to use proper prescaling to avoid arithmetic square roots in the Givens rotations; moreover, the resulting algorithm needs only half as many multiplications. We describe the algorithm for real data. The idea is to introduce weighted norms, so that the rotational invariance is expressed as

$$pD_p p^T = qD_q q^T, \tag{B.3.1}$$

where $\{p, q\}$ are given 1×2 row vectors,

$$p = [p_1 \ \ p_2], \qquad q = [q_1 \ \ q_2],$$

and $\{D_p, D_q\}$ are diagonal weighting matrices with positive elements. D_q will be selected so that there are no arithmetic square roots in the actual transformation.

The procedure is as follows. We start with a 1×2 row vector p and a diagonal weighting matrix D_p. We then seek a transformation of the form $D_p^{1/2}\Theta D_q^{-1/2}$, where Θ is a circular rotation of the same form as before, *viz.*,

$$\Theta = \frac{1}{\sqrt{1+\rho^2}}\begin{bmatrix} 1 & -\rho \\ \rho & 1 \end{bmatrix},$$

and D_q is to be chosen so as to avoid the use of square roots in the overall transformation $D_p^{1/2}\Theta D_q^{-1/2}$. Let

$$D_p = (D_{p1} \oplus D_{p2}), \qquad D_q = (D_{q1} \oplus D_{q2}).$$

Let also q denote the resulting transformed vector, $q = pD_p^{1/2}\Theta D_q^{-1/2}$, or, equivalently,

$$pD_p^{1/2}\Theta = qD_q^{1/2}. \tag{B.3.2}$$

It is then clear that this transformation preserves a weighted norm rather than the standard Euclidean norm since we now get, by squaring, $pD_pp^T = qD_qq^T$.

Now assume we wish to get a vector q of the form $q = [1 \;\; 0]$. Then the following equalities must hold, as suggested by (B.3.2),

$$D_{q1} = p_1^2 D_{p1} + p_2^2 D_{p2}, \tag{B.3.3}$$

$$\frac{1}{\sqrt{1+\rho^2}} = p_1 D_{p1}^{1/2} D_{q1}^{-1/2}, \qquad \frac{\rho}{\sqrt{1+\rho^2}} = p_2 D_{p2}^{1/2} D_{q1}^{-1/2},$$

which give the transformation from p to q as

$$D_p^{1/2}\Theta D_q^{-1/2} = \begin{bmatrix} p_1\frac{D_{p1}}{D_{q1}} & -p_2\sqrt{\frac{D_{p1}D_{p2}}{D_{q1}D_{q2}}} \\ p_2\frac{D_{p2}}{D_{q1}} & p_1\sqrt{\frac{D_{p1}D_{p2}}{D_{q1}D_{q2}}} \end{bmatrix}.$$

We still have not specified D_{q2}, and we can now do this so as to eliminate the square roots by choosing

$$D_{q2} = D_{p1}D_{p2}/D_{q1}, \tag{B.3.4}$$

so that the final transformation is

$$D_p^{1/2}\Theta D_q^{-1/2} = \begin{bmatrix} p_1\frac{D_{p1}}{D_{q1}} & -p_2 \\ p_2\frac{D_{p2}}{D_{q1}} & p_1 \end{bmatrix}. \tag{B.3.5}$$

A Numerical Example.
Consider the row vector $x = [0.875 \;\; 0.15 \;\; 1.0]$ and let us reduce it by means of two square-root free transformations to a row vector of the form $\alpha \;\; 0 \;\; 0$, for some α to be determined. The original vector x and the resulting vector αe_0 will ultimately be related by an orthogonal transformation matrix.

We first annihilate the last entry of the row vector x by pivoting with the leading entry. That is, we choose $p = [0.875 \;\; 1.0]$ and start with $D_{p1} = 1$ and $D_{p2} = 1$. (In other words, the weighting factors associated with all columns at the beginning of the procedure are equal to unity.)

Now using (B.3.3) and (B.3.4), we evaluate $\{D_{q1}, D_{q2}\}$,

$$D_{q1} = 0.7656 + 1 = 1.7656, \qquad D_{q2} = 1/1.7656 = 0.5664. \tag{B.3.6}$$

This specifies the transformation matrix (B.3.5) that is necessary for the first step as

$$D_p^{1/2}\Theta D_q^{-1/2} = \begin{bmatrix} 0.4956 & -1 \\ 0.5664 & 0.8750 \end{bmatrix}.$$

Applying this transformation to the row vector x leads to

$$[0.875 \;\; 0.15 \;\; 1.0]\begin{bmatrix} 0.4956 & 0 & -1 \\ 0 & 1 & 0 \\ 0.5664 & 0 & 0.8750 \end{bmatrix} = [1.0000 \;\; 0.15 \;\; 0.0000].$$

Moreover, the new weighting factors that are associated with the columns of the resulting vector on the right-hand side of the above equality are $\{1.7656, 1.0, 0.5664\}$. The first and last of these numbers correspond to the values of the resulting $\{D_{q1}, D_{q2}\}$ in (B.3.6). The middle number is the original weight for the second column since it has not been modified yet.

We now proceed to annihilate the $(1, 2)$ entry of the resulting vector by pivoting with its $(1, 1)$ entry. In this case we have $p_1 = 1.000$, $p_2 = 0.15$ and we thus take $D_{p1} = 1.7656$ and $D_{p2} = 1$. The choice for D_{p1} is, as explained above, the value obtained earlier for D_{q1}, which is the weight associated with the entry that we are now employing as p_1. Hence, the new values for $\{D_{q1}, D_{q2}\}$ become

$$D_{q1} = 1.7656 + (0.15)^2 = 1.7881, \qquad D_{q2} = 1.7656/1.7881 = 0.9874. \tag{B.3.7}$$

Applying the corresponding transformation (B.3.5) to the above postarray leads to

$$[1.0000 \;\; 0.15 \;\; 0.0000]\begin{bmatrix} 0.9874 & -0.15 & 0 \\ 0.0839 & 1.00 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [1.0000 \;\; 0.0000 \;\; 0.0000].$$

The new weighting factors that are associated with the columns of the resulting vector are now $\{1.7881, 0.9874, 0.5664\}$. The first and second of these numbers correspond to the values of the new $\{D_{q1}, D_{q2}\}$ in (B.3.7). The last number is the same as before since the last column was not modified in the second step.

The resulting right-hand side vector $[1 \;\; 0 \;\; 0]$ in the above equation is the desired post-array in normalized form. Its columns are respectively weighted by the numbers $\{(1.7881)^{1/2}, (0.9874)^{1/2}, (0.5664)^{1/2}\}$. This implies that the desired vector $[\alpha \;\; 0 \;\; 0]$

can be obtained by undoing the normalization as follows:

$$[1\ 0\ 0]\begin{bmatrix}(1.7881)^{1/2} & 0 & 0\\ 0 & (0.9874)^{1/2} & 0\\ 0 & 0 & (0.5664)^{1/2}\end{bmatrix} = [1.3772\ 0\ 0].$$

Observe that the resulting $\alpha = 1.3772$ satisfies $\alpha^2 = \|x\|^2$ (apart from numerical errors due to truncation).

Remark 1. Observe from the above example that, except for the initial step, successive steps of square-root free Givens transformations work with vectors p that have $p_1 = 1$ (due to the fact that we enforce $q_1 = 1$ at each step). This normalization is useful because, as noted by Golub and cited in Gentleman (1973), multiplication of an arbitrary pair by the square-root free Givens matrix will require only 2 multiplications (as compared to 4 in the general case). To see this note that, for any row vector $[a\ \ b]$, the resulting entries $\{c, d\}$ of the transformation

$$[a\ \ b]\begin{bmatrix}D_{p1}/D_{q1} & -p_2\\ p_2D_{p2}/D_{q1} & 1\end{bmatrix} = [c\ \ d], \tag{B.3.8}$$

can be evaluated as

$$d = -ap_2 + b, \tag{B.3.9}$$

and

$$\begin{aligned} c &= (aD_{p1} + bp_2D_{p2})/D_{q1} = (aD_{p1} + (d + ap_2)p_2D_{p2})/D_{q1}\\ &= (a(D_{p1} + p_2^2D_{p2}) + dp_2D_{p2})/D_{q1}\\ &= a + d(p_2D_{p2}/D_{q1}). \end{aligned} \tag{B.3.10}$$

We see that d and c can be formed using only 2 (complex) multiplications (of $[a\ \ b]$ by the elements of the transformation matrix). ◆

Remark 2. There are several other variants of such fast rotations, including some that avoid both arithmetic square roots and divisions (see, *e.g.*, Hsieh et al. (1993)). These modified Givens transformations tend to suffer from possible overflow/underflow problems; some self-scaling fast Givens rotations are described in Anda and Park (1994). ◆

B.4 J-UNITARY HOUSEHOLDER TRANSFORMATIONS

In some cases, especially when dealing with fast estimation or factorization algorithms (such as the array form of the CKMS filter of Ch. 13, or the array form of the generalized Schur algorithm of App. 13.A), it is necessary to use J-unitary transformations rather than unitary transformations (see also Prob. 12.10). We explain the necessary modifications here, starting with the case of real data.

Real-Valued Data.
A J-orthogonal transformation Θ is a matrix that satisfies

$$\Theta J\Theta^T = \Theta^T J\Theta = J,$$

where we shall take (other choices are possible) J to be a *signature matrix*, *viz.*, a diagonal matrix with ± 1 entries,

$$J = (I_p \oplus -I_q), \quad p \geq 1, \quad q \geq 1.$$

Like unitary matrices, for which $J = I$, $J-$unitary transformations preserve the squared "J-norm" of a vector. That is, if $y = x\Theta$, then

$$\|y\|_J^2 \triangleq yJy^T = x\Theta J\Theta^T x^T = xJx^T \triangleq \|x\|_J^2.$$

We are using the term "J-norm" loosely because, strictly speaking, the quantity xJx^T does not define a norm; in particular there are vectors x that can have negative squared J-norms, and even nonzero vectors x with zero squared J-norms.

Now consider an $n-$dimensional real-valued row vector x and suppose that we wish to simultaneously annihilate several entries in it by using a J-orthogonal involutary matrix Θ (*i.e.*, a transformation Θ that satisfies $\Theta J\Theta^T = J$ and $\Theta^2 = I$), say

$$[x_1\ x_2\ \ldots\ x_{n-1}]\Theta = \alpha\ e_0, \tag{B.4.1}$$

for some real scalar α. Clearly, the transformation (B.4.1) is only possible if the vector x has a *positive* squared J-norm. Otherwise, when x has a *negative* squared J-norm, we should seek a transformation Θ that reduces x to the form

$$[x_1\ x_2\ \ldots\ x_{n-1}]\Theta = \alpha\ e_{n-1}, \tag{B.4.2}$$

where e_{n-1} is the last basis vector, $e_{n-1} = [0\ \ldots\ 0\ 1]$. This is possible because the squared J-norm of αe_{n-1} is equal to $-|\alpha|^2$, which is negative. We thus consider two cases: $\|x\|_J^2 > 0$ and $\|x\|_J^2 < 0$.

(a) $\|x\|_J^2 > 0$. To find Θ to satisfy (B.4.1), we can use the same geometric argument as in the unitary case (Sec. B.1), except that we replace $\|g\|^2$ by $\|g\|_J^2$ and the inner product $\langle x, g\rangle$ by $\langle x, g\rangle_J$ (which is equal to xJg^T when x and g are row vectors; otherwise it is given by x^TJg when x and g are column vectors).[3] Therefore the first step is to choose $\alpha = \pm\sqrt{\|x\|_J^2}$ and then to write

$$\alpha e_0 = x - 2\langle x, g\rangle_J\ \|g\|_J^{-2}g,$$

which for row vectors gives

$$\alpha e_0 = x - 2xJg^T(gJg^T)^{-1}g = x\underbrace{\left[I - 2J\frac{g^Tg}{gJg^T}\right]}_{\Theta}. \tag{B.4.3}$$

[3] Linear vector spaces with such "inner products" are called indefinite metric spaces — or Krein spaces — and they are studied (in more detail than we shall need here) in Bognar (1974), Azizov and Iohvidov (1989), and also in Hassibi, Sayed, and Kailath (1999).

The matrix denoted by Θ is called a hyperbolic Householder matrix. It is immediate to verify that it is J-unitary and also involutary. In summary, given a real-valued row vector x with $\|x\|_J^2 > 0$, we define $g = x \pm \sqrt{\|x\|_J^2}\, e_0$ and then Θ as in (B.4.3). This leads to $x\Theta = \mp\sqrt{\|x\|_J^2}\, e_0$.

(b) $\|x\|_J^2 < 0$. We now seek Θ to achieve (B.4.2). The first step is to choose $\alpha = \pm\sqrt{-\|x\|_J^2}$, then choose $g = x - \alpha e_{n-1}$, and write

$$\alpha e_{n-1} = x - 2xJg^T(gJg^T)^{-1}g = x\underbrace{\left[I - 2J\frac{g^Tg}{gJg^T}\right]}_{\Theta}. \tag{B.4.4}$$

That is, with $g = x \pm \sqrt{-\|x\|_J^2}e_{n-1}$, the above Θ leads to $x\Theta = \mp\sqrt{-\|x\|_J^2}\, e_{n-1}$.

Complex-Valued Data.
Apart from the fact that we now must use Hermitian transpose, *i.e.*, find Θ such that $\Theta J\Theta^* = J$ and $\Theta^2 = I$, the main issue is the choice of α, which is now complex, say $\alpha = |\alpha|\, e^{j\phi_\alpha}$. We again consider both cases of positive and negative squared J-norms.

(a) $\|x\|_J^2 > 0$. We write the first component of x as $x_1 = |x_1|\, e^{j\phi_{x_1}}$. Then we should choose $\alpha = \pm\sqrt{\|x\|_J^2}\, e^{j\phi_{x_1}}$ where, in the complex case, $\|x\|_J^2 = xJx^*$ for row vectors. Therefore by defining $g = x \pm \sqrt{\|x\|_J^2}\, e^{j\phi_{x_1}}e_0$ and Θ as

$$\Theta \triangleq I - 2\frac{Jg^*g}{gJg^*}, \tag{B.4.5}$$

we obtain $x\Theta = \mp\sqrt{\|x\|_J^2}\, e^{j\phi_{x_1}}e_0$.

(b) $\|x\|_J^2 < 0$. Now we write the last component of x as $x_{n-1} = |x_{n-1}|\, e^{j\phi_{x_{n-1}}}$. Then we choose $\alpha = \pm\sqrt{-\|x\|_J^2}\, e^{j\phi_{x_{n-1}}}$, $g = x \pm \sqrt{-\|x\|_J^2}\, e^{j\phi_{x_{n-1}}}e_{n-1}$, and Θ as in (B.4.5). This leads to $x\Theta = \mp\sqrt{-\|x\|_J^2}\, e^{j\phi_{x_{n-1}}}e_{n-1}$.

B.5 HYPERBOLIC GIVENS ROTATIONS

As mentioned before, elementary transformations allow us to annihilate selected entries in a row vector. In this section we describe the so-called hyperbolic rotations, which are 2×2 rotation matrices that preserve certain J-norms rather than Euclidean norms (as in the case of Givens rotations). We again distinguish between real data and complex data.

Real-Valued Data.
An elementary 2×2 so-called hyperbolic matrix Θ is one that satisfies the relations

$$\Theta J\Theta^T = \Theta^T J\Theta = J,$$

where J is now 2×2 and given by $J = (1 \oplus -1)$.

Given a 1×2 real-valued vector $x = [a \;\; b]$, we desire to determine such a hyperbolic rotation that would perform the transformation

$$[a \;\; b]\Theta = [\alpha \;\; 0], \tag{B.5.1}$$

for some real number α to be determined. Now, contrary to the Givens rotation case of Sec. B.2, it is not always possible to transform a vector x to the above form unless x has positive squared J-norm, *viz.*, unless $|a| > |b|$ since $\|x\|_J^2 = a^2 - b^2$. When x has negative J-norm, *i.e.*, when $|a| < |b|$, we will be able to find a Θ that transforms it to the form

$$[a \;\; b]\Theta = [0 \;\; \alpha]. \tag{B.5.2}$$

We therefore need to distinguish between both cases.

(a) $|a| > |b|$. An expression for a hyperbolic rotation Θ that achieves (B.5.1) is given by

$$\Theta = \frac{1}{\sqrt{1-\rho^2}}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \quad \text{where} \quad \rho = \frac{b}{a}, \quad a \neq 0. \tag{B.5.3}$$

Indeed, it can be verified that applying Θ to x leads to

$$[a \;\; b]\Theta = [\pm\sqrt{a^2 - b^2} \;\; 0],$$

where the sign of α depends on whether the value of the square root in the expression (B.5.3) for Θ is chosen to be negative or positive.

Note further that Θ can be expressed in the alternative form

$$\Theta = \begin{bmatrix} ch & -sh \\ -sh & ch \end{bmatrix}, \quad ch = \frac{1}{\sqrt{1-\rho^2}}, \quad sh = \frac{\rho}{\sqrt{1-\rho^2}},$$

where ch and sh can be interpreted as hyperbolic cosine and sine parameters, respectively. This indeed justifies the name *hyperbolic rotation* for Θ, since the effect of Θ is to rotate a vector x along the *hyperbola* of equation $x^2 - y^2 = a^2 - b^2$, by an angle $\theta = \tanh^{-1}\rho$.

(b) $|b| > |a|$. An expression for a hyperbolic rotation Θ that achieves (B.5.2) is given by

$$\Theta = \frac{1}{\sqrt{1-\rho^2}}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \quad \text{where} \quad \rho = \frac{a}{b}, \quad b \neq 0, \tag{B.5.4}$$

and it can be verified that applying Θ to x leads to

$$[a \;\; b]\Theta = [0 \;\; \pm\sqrt{b^2 - a^2}].$$

Complex-Valued Data.
For complex-valued data, we should seek elementary 2×2 matrices Θ that satisfy

$$\Theta J \Theta^* = \Theta^* J \Theta = J,$$

with $J = (1 \oplus -1)$. Moreover, the scalar α in (B.5.1) or (B.5.2) will in general be complex-valued. We again distinguish between the cases of positive and negative squared J-norms.

(a) $|a| > |b|$. An expression for a hyperbolic rotation Θ that achieves (B.5.1) is given by

$$\Theta = \frac{1}{\sqrt{1-|\rho|^2}} \begin{bmatrix} 1 & -\rho \\ -\rho^* & 1 \end{bmatrix} \quad \text{where} \quad \rho = \frac{b}{a}, \quad a \neq 0. \tag{B.5.5}$$

This choices leads to

$$[a \;\; b]\Theta = [\pm \; e^{j\phi_a}\sqrt{|a|^2 - |b|^2} \;\; 0],$$

where ϕ_a denotes the phase of the possibly complex entry a.

(b) $|b| > |a|$. An expression for a hyperbolic rotation Θ that achieves (B.5.2) is given by

$$\Theta = \frac{1}{\sqrt{1-|\rho|^2}} \begin{bmatrix} 1 & -\rho \\ -\rho^* & 1 \end{bmatrix} \quad \text{where} \quad \rho = \frac{a^*}{b^*}, \quad b \neq 0, \tag{B.5.6}$$

and it can be verified that applying Θ to x now leads to

$$[a \;\; b]\Theta = [0 \pm e^{j\phi_b}\sqrt{|b|^2 - |a|^2}],$$

where ϕ_b denotes the phase of the possibly complex entry b.

B.6 SOME ALTERNATIVE IMPLEMENTATIONS

When implementating the hyperbolic rotations of the previous section in finite precision, numerical errors can accumulate thus leading to unreliable calculations. For example, let $p = q\Theta$, for some row vector q and a hyperbolic rotation Θ. In finite precision, the direct evaluation of the product $q\Theta$ leads to a computed vector $\hat{p}$ whose error relative to the exact value p can be shown to satisfy (assuming floating-point arithmetic — see, *e.g.*, Golub and Van Loan (1996))

$$\|p - \hat{p}\| \leq O(\epsilon) \cdot \|q\| \cdot \|\Theta\|,$$

where $\|\Theta\|$ denotes the spectral norm of Θ (its maximum singular value), ϵ denotes the machine precision , and $O(\epsilon)$ denotes a quantity of the order of the machine precision. But since $\|\Theta\|$ can be large, the above relation shows that the computed quantity $\hat{p}$ is not guaranteed to be evaluated to sufficient accuracy.

There exist alternative implementations of hyperbolic rotations Θ that exhibit better numerical properties. Here we briefly mention two modifications for real-valued data. More details on the value of these alternative methods in the context of fast factorization algorithms can be found in the edited volume by Kailath and Sayed (1999).

Mixed Downdating.
Let Θ be a hyperbolic rotation that is defined by a parameter ρ. Assume that we apply Θ to a generic row vector $[x \;\; y]$ to obtain

$$[x_1 \;\; y_1] = [x \;\; y] \frac{1}{\sqrt{1-\rho^2}} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}. \tag{B.6.1}$$

The equality can be written more explicitly as

$$x_1 = \frac{1}{\sqrt{1-\rho^2}}[x - \rho y], \tag{B.6.2}$$

$$y_1 = \frac{1}{\sqrt{1-\rho^2}}[-\rho x + y]. \tag{B.6.3}$$

Solving for x in terms of x_1 from the first equation and substituting into the second equation we obtain

$$y_1 = -\rho x_1 + \sqrt{1-\rho^2}\, y. \tag{B.6.4}$$

An implementation that is based on (B.6.2) and (B.6.4) is said to be in mixed downdating form. It has better numerical stability properties than a direct implementation of Θ as in (B.6.1) — see Bojanczyk et al. (1987). In the above mixed form, we first evaluate x_1 and then use it to compute y_1. We can obtain a similar procedure that first evaluates y_1 and then uses it to compute x_1. For this purpose, we solve for y in terms of y_1 from (B.6.3) and substitute into (B.6.2) to obtain

$$x_1 = -\rho y_1 + \sqrt{1-\rho^2}\, x. \tag{B.6.5}$$

Eqs. (B.6.3) and (B.6.5) represent the second mixed form.

The OD method.
The OD (Orthogonal-Diagonal) procedure is based on using the SVD of the hyperbolic rotation Θ. Assume $\rho = b/a$ with $|a| > |b|$. Then it is straightforward to verify that any hyperbolic rotation of the form (B.6.1) admits the following eigen-decomposition:

$$\Theta = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{\frac{a+b}{a-b}} & 0 \\ 0 & \sqrt{\frac{a-b}{a+b}} \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}} \triangleq QDQ^T, \tag{B.6.6}$$

where the matrix

$$Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

is orthogonal ($QQ^T = I$).

Due to the special form of the factors (Q, D), a real hyperbolic rotation, with $\rho < 1$, can then be applied to a row vector $[x \;\; y]$ to yield $[x_1 \;\; y_1]$ as follows (note that the first and last steps involve simple additions and subtractions):

$$[x' \;\; y'] \leftarrow [x \;\; y]\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$[x'' \;\; y''] \leftarrow [x' \;\; y']\begin{bmatrix} \frac{1}{2}\sqrt{\frac{a+b}{a-b}} & 0 \\ 0 & \frac{1}{2}\sqrt{\frac{a-b}{a+b}} \end{bmatrix}$$

$$[x_1 \;\; y_1] \leftarrow [x'' \;\; y'']\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

This procedure is numerically stable, as shown in Chandrasekaran and Sayed (1996). An alternative so-called H-procedure is also described in the same reference. It is costlier than the OD method but is more accurate and can be shown to be forward stable, which is a desirable property for finite-precision implementations.

APPENDIX C

Some System Theory Concepts

In this appendix we provide a brief overview of some concepts in linear system theory that are used in the body of the text. More comprehensive treatments can be found in the literature (*e.g.,* Kailath (1980), Callier and Desoer (1991), Antsaklis and Michel (1997), among many others).

C.1 LINEAR STATE-SPACE MODELS

A discrete-time linear state-space model is characterized by an n-dimensional *first-order* linear difference equation of the form

$$\begin{aligned} x_{i+1} &= F_i x_i + G_i u_i, \quad x_{i_0} = \text{ initial condition at time } i_0, \\ y_i &= H_i x_i + K_i u_i, \quad i \geq i_0, \end{aligned}$$

where the $\{F_i, G_i, H_i, K_i\}$ have dimensions $n \times n$, $n \times q$, $p \times n$, and $p \times q$, respectively. Likewise, u_i is $q \times 1$ and y_i is $p \times 1$. The n-dimensional vector x is called the *state* of the system.

In continuous time, the model becomes an n-dimensional first-order linear *differential* equation of the form

$$\begin{aligned} \dot{x}(t) &= F(t)x(t) + G(t)u(t), \quad x(t_0) = \text{ initial condition at time } t_0, \\ y(t) &= H(t)x(t) + K(t)u(t), \quad t \geq t_0, \end{aligned}$$

where $\{F(\cdot), G(\cdot), H(\cdot), K(\cdot)\}$ are now matrix functions of the time variable t and have dimensions $n \times n$, $n \times q$, $p \times n$, and $p \times q$, respectively. Correspondingly, $u(\cdot)$ is $q \times 1$ and $y(\cdot)$ is $p \times 1$.

We shall employ the shorthand notation (F, G, H, K) to refer to a state-space realization in both discrete time or continuous time. Moreover, when the model parameters $\{F, G, H, K\}$ do not vary with time, the realization (F, G, H, K) is said to be time-invariant.

C.2 STATE-TRANSITION MATRICES

For any linear state-space realization (F, G, H, K), an expression for its state vector in terms of the model parameters, the input signal, and the initial condition, can be determined by introducing certain so-called state-transition matrices.

The Discrete-Time Case.

The state-transition matrix in this case is defined by

$$\Phi(i, j) \triangleq F^{i-j} \quad \text{for} \quad i \geq j \quad \text{and} \quad \Phi(i, i) = I.$$

The reason for the name is that with $u(i) = 0$, $\Phi(i, j)$ can be interpreted as the function that transfers the state from time j to time i. Indeed, it is easy to verify by recurrence that for $i \geq j$,

$$x_i = F^{i-j}x_j + \sum_{k=j}^{i-1} F^{i-k-1}Gu_k. \tag{C.2.1}$$

Likewise, for a time-variant realization, $x_{i+1} = F_i x_i + G_i u_i$, it is immediate to verify that for $i \geq j$,

$$x_i = \Phi(i, j)x_j + \sum_{k=j}^{i-1} \Phi(i, k+1)G_k u_k, \tag{C.2.2}$$

where the *state-transition* matrix $\Phi(i, j)$ is now given by the matrix product

$$\Phi(i, j) = F_{i-1} \cdot F_{i-2} \ldots F_j, \qquad \Phi(i, i) = I.$$

In continuous time, we have to work a bit harder. For this reason we shall treat separately the cases of time-invariant and time-variant realizations. In the first case, the state-transition matrix is defined in terms of the exponential matrix function.

The Exponential Matrix Function.

Thus consider the time-invariant continuous-time state equation

$$\dot{x}(t) = Fx(t) + Gu(t), \quad x(t_0). \tag{C.2.3}$$

The solution $x(t)$ is unique and given by the superposition formula

$$x(t) = e^{F(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{F(t-\tau)}Gu(\tau)d\tau, \tag{C.2.4}$$

where the matrix $\Phi(t, t_0) \triangleq e^{F(t-t_0)}$ is called the exponential matrix function and it is defined by the power series

$$e^{Ft} \triangleq \sum_{k=0}^{\infty} \frac{t^k}{k!} F^k. \tag{C.2.5}$$

This series can be shown to converge uniformly and absolutely for any F and, moreover, the resulting function e^{Ft} will have several interesting properties, *e.g.*,

1. $\frac{d}{dt}[e^{Ft}] = Fe^{Ft}$ and $e^{Ft}\big|_{t=0} = I$.
2. e^{Ft} is the unique solution of $\dot{X}(t) = FX(t)$, $X(0) = I$.
3. $e^{F(t+\tau)} = e^{Ft}e^{F\tau}$ and $(e^{Ft})^{-1} = e^{-Ft}$.
4. $e^{Ft} = \mathcal{L}^{-1}(sI - F)^{-1}$ for $\mathrm{Re}(s) > \max_k \mathrm{Re}[\lambda_k(F)]$. Here, $\lambda_k(F)$ denotes the k-th eigenvalue of F. Also $\mathcal{L}(h)$ denotes the unilateral Laplace transform of a right-sided function $h(t)$, *viz.*,

$$\mathcal{L}(h) \triangleq H(s) = \int_{0-}^{\infty} h(t)e^{-st}dt.$$

Likewise, $\mathcal{L}^{-1}(H)$ denotes the inverse Laplace transform of $H(s)$.

Remark. [Fundamental Theorem of Differential Equations] Apart from a superposition justification, we can also verify the validity of (C.2.4) by differentiating it to show that expression (C.2.4) for $x(t)$ satisfies the differential equation (C.2.3), with the same initial condition. Then, by a fundamental theorem of differential equations, which guarantees, under certain conditions, the uniqueness of the solution of a general differential equation (possibly nonlinear) for a given initial condition, we can conclude that $x(t)$ in (C.2.4) is the desired solution. Detailed expositions on the fundamental theorem of differential equations can be found, *e.g.*, in Callier and Desoer (1991), Vidyasagar (1993), or Khalil (1996). ◆

State-Transition Functions for Linear Time-Variant Systems.

As mentioned above, more care is needed in characterizing the state-transition matrix function for linear *time-variant* realizations. Thus consider the model

$$\dot{x}(t) = F(t)x(t) + G(t)u(t), \qquad x(t_0), \tag{C.2.6}$$

with the matrix functions $F(\cdot)$ and $G(\cdot)$ assumed bounded for $t \geq t_0$. Then its unique solution can be represented as

$$x(t) = \Phi(t, t_0)x(t_0) + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)u(\tau)d\tau, \tag{C.2.7}$$

where the state-transition matrix $\Phi(t, \tau)$ is now defined as the unique solution of the differential equation

$$\frac{d}{dt}\Phi(t, \tau) = F(t)\Phi(t, \tau), \qquad \Phi(\tau, \tau) = I. \tag{C.2.8}$$

The validity of the above expression for $x(t)$ can again be verified by differentiating (C.2.7) and using (C.2.8) to conclude that $x(t)$ in (C.2.7) satisfies $\dot{x}(t) = F(t)x(t) + G(t)u(t)$, with the same initial condition $x(t_0)$. Hence, in view of the aforementioned fundamental theorem of differential equations, expression (C.2.7) is the desired solution.

Though in general no explicit expression for $\Phi(\cdot, \cdot)$ is available, its properties make the representation (C.2.7) very useful for analysis purposes (see, *e.g.*, Ch. 16). Among these properties, we list the following:

1. $\Phi(t,t) = I$.
2. $\Phi(t,t_0)$ satisfies the composition rule $\Phi(t,t_0) = \Phi(t,t_1)\Phi(t_1,t_0)$ for all $\{t_1,t_0,t\}$.
3. $\Phi(t,t_0)$ is always invertible and $\Phi^{-1}(t,t_0) = \Phi(t_0,t)$.

C.3 CONTROLLABILITY AND STABILIZABILITY

In the remainder of this appendix we focus on *time-invariant* state-space models and describe several system-theoretic concepts that are often used to characterize a realization (F,G,H,K).

We first remark that by a *stable* matrix, we shall mean one whose eigenvalues are either inside the open unit disc (for discrete-time realizations) or in the open left-half plane (for continuous-time realizations). It will be explained in Sec. D.3 how this notion of stable matrices is required in the characterization of the so-called asymptotic or exponential stability of an unforced system.

We start by introducing the notions of controllability, stabilizability, and unit-circle or imaginary-axis controllability. These depend only on the matrices $\{F,G\}$.

Controllability.
Consider first a special state-equation with a diagonal matrix F, with entries $\{\lambda_i\}$, and a column vector G, with entries $\{\beta_i\}$,

$$F = \text{diag}\{\lambda_1,\ldots,\lambda_n\}, \quad G = \text{col}\{\beta_1,\ldots,\beta_n\}. \tag{C.3.1}$$

Then the pair $\{F,G\}$ is said to be controllable if, and only if, all the $\{\lambda_i\}$ are distinct and all the $\{\beta_i\}$ are nonzero. These conditions guarantee that the input signal (u_i in discrete time or $u(t)$ in continuous time) will be able to influence (or control) the evolution of each individual entry of the state vector (x_i or $x(t)$), and hence the name *controllable* for the pair $\{F,G\}$.

For more general matrices $\{F,G\}$, however, the test for controllability is more involved. In fact, there are several equivalent tests, which we list below. They can all be shown to collapse to the above simple conditions on $\{\lambda_i,\beta_i\}$ when F is diagonal and G is a column vector.

We thus say that, in general, a pair $\{F,G\}$ is controllable if, and only if, any of the following conditions hold:

1. Controllability matrix. The so-called controllability matrix $\mathcal{C}$ defined below has full rank n,
$$\mathcal{C} \triangleq \begin{bmatrix} G & FG & F^2G & \ldots & F^{n-1}G \end{bmatrix}.$$
2. Rank test.
$$\text{rank}\ \begin{bmatrix} \lambda I - F & G \end{bmatrix} = n \quad \text{at all eigenvalues of } F.$$
3. PBH test. There does not exist a left eigenvector of F that is orthogonal to G, *i.e.*, for any $vF = \lambda v$ we obtain $vG \neq 0$.
4. State feedback test. There always exists a matrix L such that $(F - GL)$ is a stable matrix.

5. Controllability Gramian test. This test is restricted to stable matrices F, so that the following so-called controllability Gramian matrices are well defined:
$$W_c \triangleq \int_0^\infty e^{F(t-\tau)}GG^*e^{F^*(t-\tau)}d\tau, \qquad \text{(for continuous-time)},$$
$$W_c \triangleq \sum_{j=0}^{\infty} F^jGG^*F^{*j}, \qquad \text{(for discrete-time)}.$$
Then, for stable F, the pair $\{F,G\}$ is controllable if $W_c > 0$. These conditions are also equivalent to the following:
   (i) The equation $FW_c + W_cF^* = -GG^*$ has a unique positive-definite solution W_c (for continuous-time realizations).
   (ii) The equation $W_c - FW_cF^* = GG^*$ has a unique positive-definite solution W_c (for discrete-time realizations).

Stabilizability.
The eigenvalues of the matrix F are sometimes called the modes of the realization (F,G,H,K). There can be stable or unstable modes. In the discrete-time case, a stable mode is one that satisfies $|\lambda| < 1$. In continuous time, it is one that satisfies $\text{Re}(\lambda) < 0$.

Now consider again the special case of a diagonal matrix F and a column vector G, as in (C.3.1). Then $\{F,G\}$ is said to be stabilizable if, and only if, all the unstable $\{\lambda_i\}$ are distinct and the corresponding $\{\beta_i\}$ are nonzero. These conditions guarantee that the input signal will be able to influence (or control) the evolution of each of the unstable modes of the system, and hence the name *stabilizable* for the pair $\{F,G\}$.

More generally, for any pair $\{F,G\}$, we classify the modes of F as controllable or uncontrollable according to the following criterion. Modes at which

$$\text{rank}\ \left(\begin{bmatrix} \lambda I - F & G \end{bmatrix}\right) < n,$$

are said to be uncontrollable. Otherwise, they are said to be controllable. Now a pair $\{F,G\}$ will be said to be *stabilizable* if all its unstable modes are controllable. That is, if all the modes at which the above rank condition is satisfied are stable modes. The following are equivalent conditions for a stabilizable pair $\{F,G\}$:

1. Rank test.
$$\text{rank}\ \begin{bmatrix} \lambda I - F & G \end{bmatrix} = n \quad \text{at all unstable eigenvalues of } F.$$
2. PBH test. There does not exist a left eigenvector, corresponding to an unstable eigenvalue of F, that is orthogonal to G.
3. State feedback test. A matrix L can always be found such that $(F - GL)$ is stable.

Unit-Circle or Imaginary-Axis Controllability.
Consider once more the special case of a diagonal matrix F and a column vector G, as in (C.3.1). Then $\{F,G\}$ is said to be unit-circle (in discrete-time) or imaginary-axis (in continuous time) controllable if, and only if, all the $\{\lambda_i\}$ that lie on the unit circle

(in discrete time) or on the imaginary axis (in continuous time) are distinct and the corresponding $\{\beta_i\}$ are nonzero. These conditions guarantee that the input signal will be able to influence (or control) the evolution of these modes.

More generally, a pair $\{F, G\}$ is said to be unit-circle controllable (in discrete-time) or imaginary-axis controllable (in continuous time) if any of the following equivalent conditions hold:

1. Rank test.
$$\text{rank } (\lambda I - F \;\; G) = n,$$
at all unit-circle (in discrete time) or imaginary-axis (in continuous time) eigenvalues of F.
2. PBH test. There does not exist a left eigenvector, corresponding to a unit-circle (in discrete time) or imaginary-axis (in continuous time) eigenvalue of F, that is orthogonal to G. More specifically, this means for discrete time that for any left eigenvector v satisfying $vF = \lambda v$, $|\lambda| = 1$, it holds that $vG \neq 0$. For continuous time, the condition means that any left eigenvector v satisfying $vF = \lambda v$, $\text{Re}(\lambda) = 0$, is such that $vG \neq 0$.
3. State feedback test. A matrix L can always be found such that $(F - GL)$ has no unit-circle (in discrete time) or imagniary-axis (in continuous time) eigenvalues.

C.4 OBSERVABILITY AND DETECTABILITY

The concepts of observability, detectability, and unit-circle or imaginary-axis observability are simply dual to the concepts of controllability, stabilizability, and unit-circle or imaginary-axis controllability. Hence, we have the following.

Observability. A pair $\{F, H\}$ is observable if, and only if, $\{F^*, H^*\}$ is controllable.

Detectability. A pair $\{F, H\}$ is detectable if, and only if, (F^*, H^*) is stabilizable.

Unit-Circle or Imaginary-Axis Observability. A pair $\{F, H\}$ is unit-circle observable (in discrete time) or imaginary-axis observable (in continuous time) if, and only if, $\{F^*, H^*\}$ is unit-circle (in discrete time) or imaginary-axis (in continuous time) controllable.

We can again provide a physical interpretation for these definitions in the simple case of a realization with a diagonal matrix F and a row vector H, say

$$\begin{aligned} x_{i+1} &= Fx_i, \qquad y_i = Hx_i, \quad \text{in discrete time,} \\ \dot{x}(t) &= Fx(t), \quad y(t) = Hx(t), \quad \text{in continuous time,} \end{aligned}$$

and

$$F = \text{diag}\{\lambda_1, \ldots, \lambda_n\}, \qquad H^T = \text{col}\{\gamma_1, \ldots, \gamma_n\}.$$

Then the pair $\{F, H\}$ is observable if, and only if, all the $\{\lambda_i\}$ are distinct and all the $\{\gamma_i\}$ are nonzero. These conditions guarantee that ali modes will be observed at the output, and hence the name *observable* for the pair $\{F, G\}$.

Likewise, $\{F, G\}$ will be detectable if, and only if, all the unstable $\{\lambda_i\}$ are distinct and the corresponding $\{\gamma_i\}$ are nonzero. That is, the unstable modes will be observable at the output of the realization. Similarly for unit-circle or imaginary-axis observability.

C.5 MINIMAL REALIZATIONS

A realization (F, G, H, K) is said to be minimal if, and only if, $\{F, G\}$ is controllable and $\{F, H\}$ is observable. An important theorem due to Kalman (1963c) is that if $\{F_k, G_k, H_k, K\}$, $k = 1, 2$, are two minimal realizations of the same system, then there exists an invertible matrix T such that

$$F_1 = T^{-1}F_2T, \quad G_1 = T^{-1}G_2, \quad H_1 = H_2T.$$

APPENDIX D

Lyapunov Equations

Lyapunov equations arise in many contexts in linear system theory. We have encountered them, for example, in Chs. 5 and 8 (see Eqs. (5.3.10) and (8.1.6)) in the form of equations for the steady-state covariance matrix of the state vector. Such Lyapunov equations in fact first arose in the study of the so-called internal (or exponential) stability of linear time-invariant dynamical systems, and later in the characterization of the controllability and observability properties of such systems in terms of their controllability and observability Gramians (*cf.* App. C).

We have studied some of the properties of Lyapunov equations in Prob. 5.14 but we shall pursue them in more detail in this appendix. A good reference for further results on the material here are the books of Lancaster and Tismenetsky (1985) and Lancaster and Rodman (1995).

D.1 DISCRETE-TIME LYAPUNOV EQUATIONS

Given $n \times n$ matrices P, A, F, and Q, we consider first the following linear matrix equation, often called a discrete-time Lyapunov equation (or sometimes a Stein equation),[1]

$$P - FPA = Q. \tag{D.1.1}$$

Let $\lambda_i(A)$ and $\lambda_j(F)$ denote any of the eigenvalues of A and F, with i and j assuming values between 1 and n.

Lemma D.1.1 (Uniqueness of Solutions) *For each Q, the Lyapunov equation (D.1.1) has a unique solution P if, and only if, the eigenvalues of A and F satisfy the conditions*

$$1 - \lambda_i(A)\lambda_j(F) \neq 0 \quad \text{for all} \quad i, j = 1, \ldots, n. \tag{D.1.2}$$

■

[1] The Hermitian equation with $A = F^*$ is encountered in the so-called Lyapunov approach to the stability of linear discrete-time systems — see Sec. D.3. Lyapunov (1892) actually only studied the analogous continuous-time problem discussed in the next section.

Proof: An immediate proof follows by using the Kronecker product notation to reduce the above equation to an equivalent linear equation in the entries of P (as explained after the statement of Lemma A.2.1). Thus using property (v) of this lemma, we can write $(I - A^T \otimes F)\text{vec}(P) = \text{vec}(Q)$. This shows that a unique solution $\text{vec}(P)$ exists as long as the coefficient matrix $(I - A^T \otimes F)$ is nonsingular. In view of property (iii) of Lemma A.2.1, the nonsingularity of $(I - A^T \otimes F)$ is equivalent to (D.1.2). ◆

When $A = F^*$ and $Q = Q^*$ is Hermitian, we have the more traditional Lyapunov equation,

$$P - FPF^* = Q. \tag{D.1.3}$$

Lemma D.1.2 (Properties of the Lyapunov Equation) *Consider the equation (D.1.3), with Q Hermitian. The following facts hold:*

(i) A unique solution P exists if, and only if, $\lambda_i(F)\lambda_j^*(F) \neq 1$, for all i, j. Moreover, the unique solution will be Hermitian.

(ii) If F is a stable matrix (i.e., all its eigenvalues are inside the open unit disc), then P is unique and Hermitian.

(iii) If F is a stable matrix, the unique Hermitian solution P can be expressed in the series form

$$P = \sum_{i=0}^{\infty} F^i Q F^{*i}. \tag{D.1.4}$$

(iv) If F is stable and Q is positive (semi-)definite, then P is unique, Hermitian, and positive-(semi-)definite.

(v) If F is stable, Q is positive-(semi-)definite, and the pair $\{F, Q^{1/2}\}$ is controllable, then P is unique, Hermitian, and positive-definite.

■

Proof: We proceed as follows.

(i) By Lemma D.1.1, the condition $\lambda_i(F)\lambda_j^*(F) \neq 1$ guarantees a unique solution P. The fact that the solution is Hermitian follows by uniqueness and by noting that P^* satisfies the same Lyapunov equation.

(ii) The eigenvalues of a stable matrix F satisfy $\lambda_i(F)\lambda_j^*(F) \neq 1$. Hence, by (i) above, P is unique and Hermitian.

(iii) When F is stable, the series converges absolutely to some matrix $\overline{P}$,

$$\overline{P} \triangleq \sum_{i=0}^{\infty} F^i Q F^{*i}.$$

Now note that

$$\overline{P} - F\overline{P}F^* = \sum_{i=0}^{\infty} F^i Q F^{*i} - \sum_{i=1}^{\infty} F^i Q F^{*i} = Q,$$

so that $\overline{P}$ satisfies the same Lyapunov equation as P. But, by (ii), the Lyapunov equation has a unique Hermitian solution P. Therefore, $P = \overline{P}$.

(iv) The result follows by noting that each term $F^i Q F^{*i}$ in the above series representation is positive-(semi)-definite when Q is positive-(semi)-definite.

(v) The controllability assumption on the pair $\{F, Q^{1/2}\}$ guarantees that the matrix

$$\begin{bmatrix} Q^{1/2} & FQ^{1/2} & F^2Q^{1/2} & \dots & F^{n-1}Q^{1/2} \end{bmatrix} \quad \text{is full rank,}$$

and, consequently, that the controllability Gramian matrix is positive definite, *i.e.*,

$$\sum_{i=0}^{n-1} F^i Q F^{*i} > 0.$$

The positive definiteness of P follows from the fact that

$$P = \sum_{i=0}^{\infty} F^i Q F^{*i} \geq \sum_{i=0}^{n-1} F^i Q F^{*i} > 0.$$

◆

Remark 1. One proof that establishes the absolute convergence of the series in (iii) is the following. There is a basic basic result in matrix theory (see Prob. 14.19), which states that for a given matrix F with spectral radius $\rho(F)$, there always exists a matrix norm $\|\cdot\|$ such that

$$\rho(F) \leq \|F\| \leq \rho(F) + \epsilon, \quad \text{for any } \epsilon > 0.$$

Since F is a stable matrix, we can choose an $\epsilon > 0$ such that $\alpha \triangleq \rho(F) + \epsilon < 1$. Then there will exist a matrix norm with respect to which $\|F\| \leq \alpha < 1$. We shall denote this particular norm by $\|\cdot\|_\rho$. Now using the triangle inequality of norms we obtain

$$\left\| \sum_{i=0}^{\infty} F^i Q F^{*i} \right\|_\rho \leq \sum_{i=0}^{\infty} \|Q\|_\rho \, \alpha^{2i} = \frac{\|Q\|_\rho}{1-\alpha^2} < \infty,$$

which establishes the convergence of the series of part (iii). ◆

D.2 CONTINUOUS-TIME LYAPUNOV EQUATIONS

A similar analysis applies to the following equation, often called a continuous-time Lyapunov equation (or sometimes a Sylvester equation):

$$PA + FP + Q = 0. \tag{D.2.1}$$

Let again $\lambda_i(A)$ and $\lambda_j(F)$ denote any of the eigenvalues of A and F, with i and j assuming values between 1 and n. The following result is also immediate from the properties of Kronecker products.

Lemma D.2.1 (Uniqueness of Solutions) *For any Q, the Lyapunov equation (D.2.1) has a unique solution P if, and only if, the eigenvalues of A and F satisfy the conditions*

$$\lambda_i(A) + \lambda_j(F) \neq 0 \quad \textit{for all} \quad i, j = 1, \dots, n. \tag{D.2.2}$$

■

Proof: The solution P of (D.2.1) is the solution of the linear system of equations

$$(A^T \otimes I + I \otimes F)\text{vec}(P) = -\text{vec}(Q),$$

which shows that $\text{vec}(P)$ exists and is unique as long as the coefficient matrix $(A^T \otimes I + I \otimes F)$ is nonsingular. In view of property (iii) of Lemma A.2.1, the nonsingularity of $(A^T \otimes I + I \otimes F)$ is equivalent to (D.2.2). ◆

When $A = F^*$ and $Q = Q^*$ is Hermitian, we have the more traditional Lyapunov equation,

$$PF^* + FP + Q = 0. \tag{D.2.3}$$

Lemma D.2.2 (Properties of the Lyapunov Equation) *Consider the equation (D.2.3), with Q Hermitian. The following facts hold:*

(i) A unique solution P exists if, and only if, $\lambda_i^*(F) + \lambda_j(F) \neq 0$, for all i, j. Moreover, the unique solution will be Hermitian.

(ii) If F is a stable matrix (i.e., all its eigenvalues are inside the open left-half plane), then P is unique and Hermitian.

(iii) If F is a stable matrix, the unique Hermitian solution P can be expressed in the integral form

$$P = \int_0^\infty e^{F^*t} Q e^{Ft} dt. \tag{D.2.4}$$

(iv) If F is stable and Q is positive-(semi-)definite, then P is unique, Hermitian, and positive-(semi-)definite.

(v) If F is stable, Q is positive-semi-definite, and $(F, Q^{*/2})$ is controllable, then P is unique, Hermitian, and positive-definite.

■

Proof: We proceed in steps.

(i) The uniqueness of the solution follows from Lemma D.2.1. Moreover, since P^* satisfies the same Lyapunov equation, then $P = P^*$.

(ii) The eigenvalues of a stable matrix F satisfy the conditions in (i) and hence, (ii) is a special case of (i).

(iii) Since F is stable, the following matrix is well defined:

$$P \triangleq \int_0^\infty e^{F^*t} Q e^{Ft} dt.$$

We now show that this P satisfies (D.2.3). Indeed, using the property $de^{Ft}/dt = Fe^{Ft}$, we conclude that

$$F^* \left(\int_0^\infty e^{F^*t} Q e^{Ft} dt \right) + \left(\int_0^\infty e^{F^*t} Q e^{Ft} dt \right) F = \int_0^\infty \frac{d}{dt} \left[e^{F^*t} Q e^{Ft} \right] dt = -Q,$$

as desired. The uniqueness of P follows from part (ii).

(iv) The center matrix in the integral representation

$$P \triangleq \int_0^\infty e^{F^*t} Q e^{Ft} dt,$$

is positive-(semi-)definite since Q is positive (semi-)definite. Hence, P is also positive-(semi-)definite.

(v) We know from (iv) that $P \geq 0$. We now want to show that the controllability assumption implies $P > 0$. Assume to the contrary that there exists a nonzero vector x such that $Px = 0$. Then post- and pre-multiplying the Lyapunov equation (D.2.3) by x and x^* leads to $x^*Qx = 0$ so that $Q^{1/2}x = 0$. Now post-multiplying the Lyapunov equation again by x we obtain $PF^*x = 0$. That is, we showed that if $Px = 0$, then $PF^*x = 0$ as well.

We now post- and pre-multiply the Lyapunov equation (D.2.3) by F^*x and x^*F to conclude that $Q^{1/2}F^*x = 0$. Repeating the argument we can establish that $Q^{1/2}F^{*i}x = 0$ for $i \geq 0$. This shows that

$$x^*\left[Q^{*/2} \;\; FQ^{*/2} \;\; \ldots \;\; F^{n-1}Q^{*/2}\right] = 0,$$

which contradicts the controllability of $\{F, Q^{*/2}\}$. ◆

D.3 INTERNAL STABILITY

As mentioned before, Lyapunov equations play an important role in stability analysis. They were introduced as a major tool in this context by Lyapunov in the 1890s in his study of the stability of linear and nonlinear systems. We focus here on the linear case and provide a brief account of the results that are more relevant to our purposes.

We start with the notion of *internal stability* or *asymptotic stability in the sense of Lyapunov*. The unforced equation $x_{i+1} = Fx_i$ is said to be asymptotically stable if $x_i \to 0$ as $i \to \infty$ for all initial conditions (i_0, x_{i_0}). To be more pedantic, we should instead say that the equation $x_{i+1} = Fx_i$ is uniformly (*i.e.*, irrespective of i_0), globally (*i.e.*, irrespective of x_{i_0}), asymptotically stable. But for the linear time-invariant equation $x_{i+1} = Fx_i$, the more specific terminology is equivalent to asymptotic stability, which we therefore adopt in the sequel.

Lemma D.3.1 (Internal Stability) *The unforced system $x_{i+1} = Fx_i$ is asymptotically stable if, and only if, all eigenvalues of F are inside the open unit disc.* ■

Proof: Let λ denote any of the eigenvalues of F. Now since $x_i = F^{i-i_0}x_{i_0}$, by using the canonical Jordan decomposition of the matrix F, we can verify that all entries of the state vector x_i are linear combinations of exponential terms of the form $\{\lambda^i, i\lambda^i, i^2\lambda^i, \ldots\}$, for different eigenvalues of F and depending on the multiplicities of the eigenvalues. It is then clear that x_i tends to zero if, and only if, $|\lambda| < 1$. ◆

The above result can be related to Lyapunov equations via the following statement.

Theorem D.3.1 (Lyapunov Condition) *The unforced system $x_{i+1} = Fx_i$ is asymptotically stable if, and only if, for any positive-definite matrix Q there exists a unique positive-definite matrix P such that*

$$P - F^*PF = Q. \tag{D.3.1}$$

■

Proof: One direction is immediate from part (iv) of Lemma D.1.2. That is, if F is stable, then for any positive-definite Q the solution P of (D.3.1) will be unique and positive-definite.

Conversely, assume there exists a positive-definite P that solves (D.3.1). Let λ denote an eigenvalue of F corresponding to a right eigenvector q, *viz.*, $Fq = \lambda q$. Multiplying (D.3.1) by q^* from the left and by q from the right we obtain

$$q^*Pq[1 - |\lambda|^2] = q^*Qq.$$

The quantities q^*Qq and q^*Pq are both positive since Q and P are positive-definite. Therefore, it must hold that $1 - |\lambda|^2 > 0$. In other words, $|\lambda| < 1$ and F has all its eigenvalues inside the open unit disc. ◆

We can extend the discussion to the continuous-time case as well. In this context, we shall say that the unforced differential equation $\dot{x}(t) = Fx(t)$ is asymptotically stable if $x(t) \to 0$ as $t \to \infty$ for all t_0 and $x(t_0)$.

Lemma D.3.2 (Internal Stability in Continuous-Time) *The unforced system $\dot{x}(t) = Fx(t)$ is asymptotically stable if, and only if, all eigenvalues of F are in the open left-half plane.* ■

Proof: First note that, using the results of App. C, we can express the solution of $\dot{x}(t) = Fx(t)$ in terms of the matrix exponential e^{Ft} as $x(t) = e^{F(t-t_0)}x_0$. Therefore, all the entries of $x(t)$ will involve linear combinations of exponential terms of the form $\{e^{\lambda_i t}, te^{\lambda_i t}, \ldots\}$, for different eigenvalues of F and depending on the multiplicity of each eigenvalue. It is then clear that $x(t) \to 0$ if, and only if, $\mathrm{Re}(\lambda_i) < 0$. ◆

We can also relate this result to a theorem of Lyapunov.

Theorem D.3.2 (Lyapunov Condition in Continuous-Time) *The realization $\dot{x} = Fx$ is asymptotically stable if, and only if, for any positive-definite matrix Q, there exists a unique positive-definite matrix P that solves the Lyapunov equation*

$$F^*P + PF = -Q.$$

■

Proof: Assume F is asymptotically stable. Then by part (iv) of Lemma D.2.2 we have that for any $Q > 0$, a unique $P > 0$ exists.

Conversely, assume there exists a unique positive-definite P that solves $F^*P + PF + Q = 0$. Let $Fv = \lambda v$ be an eigenvalue-eigenvector pair for F. Multiplying the Lyapunov equation by v^* from the left and by v from the right, we obtain

$$(\lambda^* + \lambda)v^*Pv = -v^*Qv,$$

which shows that we must have $\mathrm{Re}(\lambda) < 0$ since $v^*Qv > 0$ and $v^*Pv > 0$. ◆

Remark 2. [Mappings between Discrete Time and Continuous Time] It is useful to recall here that many continuous-time results, including the above stability results, can be obtained from the discrete-time results by use of the well-known bilinear transformations

$$s = \frac{z-1}{z+1}, \qquad z = \frac{1+s}{1-s},$$

which map a point s in the left-half plane into a point z in the unit disc and vice versa. The approrppriate transformations for systems in state-space form are

$$F \longleftarrow (F_d - I)(F_d + I)^{-1}, \quad Q \longleftarrow 2(F_d^* + I)^{-1}Q_d(F_d + I)^{-1}, \quad P \longleftarrow P_d,$$

where the subscript d is used to denote discrete-time quantities (and where it is assumed that none of the eigenvalues of F_d are at -1 so that the inverse matrix $(F_d + I)^{-1}$ exists). Now we can verify that with these substitutions the equation

$$F^*P + PF = -Q,$$

in continuous time goes into

$$P_d - F_d^*P_dF_d = Q_d$$

in discrete time. ◆

APPENDIX E
Algebraic Riccati Equations

Algebraic Riccati equations (ARE) play an important role in many estimation and control problems. Their discrete-time versions (DARE) arise in Chs. 8 and 14 in the context of canonical spectral factorization and steady-state Kalman filtering, the continuous-time versions (CARE) in Ch. 16. These equations have an extensive literature (see the references given at the end of this appendix), as well as several important properties. The most relevant of these for our purposes are studied in this appendix. We start with an overview of the results for the DARE. A quick review of system theoretic concepts from Apps. C and D may be helpful.

E.1 OVERVIEW OF DARE

Consider matrices $\{Q, R, S\}$, of respective dimensions $\{m \times m, p \times p, m \times p\}$, that satisfy

$$Q = Q^*, \quad \begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix} \geq 0, \quad \text{and} \quad R = R^* > 0. \tag{E.1.1}$$

We shall assume (E.1.1) throughout this appendix; note the positivity assumption on R.

Now given matrices $\{F, G, H\}$, of dimensions $\{n \times n, n \times m, p \times n\}$, we shall first be concerned with the discrete-time algebraic Riccati equation (DARE)

$$P = FPF^* + GQG^* - (FPH^* + GS)(R + HPH^*)^{-1}(FPH^* + GS)^*. \tag{E.1.2}$$

In later sections, we study the CARE and the dual DARE and CARE.

As noted in Ch. 8, our main reason for studying the DARE is as a tool for performing canonical spectral factorization. In this case, one is interested in obtaining the so-called *stabilizing* solution,[1] assuming it exists, *i.e.*, a solution P such that the matrix

$$F_p \triangleq F - K_p H, \quad \text{with} \quad K_p = (FPH^* + GS)R_e^{-1}, \; R_e = R + HPH^*$$

is stable (with eigenvalues inside the open unit disc).

The DARE is a highly nonlinear equation, so that solutions P may or may not exist, they may or may not be unique, or indeed they may or may not be stabilizing. The purpose of this appendix is to establish conditions under which stabilizing solutions to the DARE exist. These results are summarized in Table E.1 and then elaborated by several examples. As can be seen, the first result states that the detectability of $\{F, H\}$ is a sufficient condition for the existence of a *semi-stabilizing* solution to the DARE. The *detectability* of $\{F, H\}$ means that there exists a matrix K such that $F - KH$ is stable.

Example E.1.1 Consider a simple scalar DARE with $F = 1$, $G = 0$, $H = 1$, $Q = 1$, $S = 0$, and $R = 1$. Clearly, $\{F, H\}$ is detectable and the DARE becomes

$$P = P + 0 - \frac{P^2}{1 + P},$$

which has multiple solutions at $P = 0$. For these solutions we have $K_p = 0$, so that $F_p = F = 1$, which is semi-stable (also called marginally stable). ◆

The detectability assumption is also clearly necessary for the existence of a *stabilizing* solution (otherwise, $F - K_p H$ cannot be stable since $F - KH$ cannot be stable for *any* K). However, it is not sufficient. The second result in Table E.1 states that the necessary and sufficient condition is that $\{F, H\}$ be detectable and that $\{F^s, GQ^{s/2}\}$ be *controllable on the unit circle*, where

$$F^s \triangleq F - GSR^{-1}H, \qquad Q^s \triangleq Q - SR^{-1}S^*. \tag{E.1.3}$$

This latter condition means that there exists some matrix K such that $F^s - GQ^{s/2}K$ has no unit-circle eigenvalues. Recall also from Lemma 8.3.1 that this ensures that the spectrum $S_y(z)$ contains no unit-circle zeros, a condition required for the existence of a canonical spectral factorization.

Example E.1.2 Note that in Example E.1.1 we did not have a stabilizing solution P since $\{F^s, GQ^{s/2}\} = \{1, 0\}$ was not controllable on the unit circle. However, consider a scalar DARE with $F = 2$, $G = 0$, $H = 1$, $Q = 1$, $S = 0$, and $R = 1$. Now $\{F^s, GQ^{s/2}\} = \{2, 0\}$ is controllable on the unit circle, while the DARE is

$$P = 4P + 0 - \frac{4P^2}{1 + P},$$

which has the two solutions $P_1 = 0$ and $P_2 = 3$. It is easy to see that P_2 is a stabilizing solution, since for this solution we have $K_p = 3/2$ so that $F_p = 1/2$. Moreover, $P_1 = 0$ is not stabilizing since it leads to $F_p = 2$. ◆

[1] It will shortly be shown that, when a stabilizing solution exists, it is unique; hence our reference to "the" stabilizing solution, rather than "a" stabilizing solution.

Table E.1 Solutions to the DARE.

Properties of the solution	Conditions	Relevant thms.	Remarks
Is there a solution of the DARE such that F_p is semi-stable, *i.e.*, $\lvert\lambda(F_p)\rvert \le 1$?	Yes, under detectability (only a sufficient condition).	Lemmas E.3.2 and E.4.1.	At least one such solution is p.s.d.
Is there a solution of the DARE such that F_p is stable, *i.e.*, $\lvert\lambda(F_p)\rvert < 1$?	Yes, iff we have detectability and unit circle controllability of $\{F^s, GQ^{s/2}\}$.	Lemmas E.4.2 and E.4.3 and Thm. E.5.1.	The stabilizing solution is *unique* and p.s.d. However, there can be several p.s.d. solutions.
When is the stabilizing solution of the DARE its *unique* p.s.d. solution?	Iff we have detectability and stabilizability (controllability on and outside the unit circle) of $\{F^s, GQ^{s/2}\}$.	Thm. E.6.1.	
When is the stabilizing solution of the DARE positive definite?	Iff we have detectability and controllability on and inside the unit circle.	Thm. E.6.2.	

It is claimed in Table E.1, that the stabilizing solution of the DARE in the above case is unique and positive-semi-definite. However, this does not necessarily imply that the DARE has a unique positive-semi-definite solution. The necessary and sufficient condition for this to be true is that $\{F, H\}$ be detectable and that $\{F^s, GQ^{s/2}\}$ be *stabilizable*. Here, by stabilizable, we mean that $\{F^s, GQ^{s/2}\}$ be controllable on and outside the unit circle, or, in other words, that there exist some matrix K such that $F^s - GQ^{s/2}K$ be stable.

Example E.1.3 Note that in Example E.1.2 we did not have a unique positive-semi-definite solution (both solutions were nonnegative) because $\{F^s, GQ^{s/2}\} = \{2, 0\}$ was not stabilizable. However, consider a scalar DARE with $F = 1/2$, $G = 0$, $H = 1$, $Q = 1$, $S = 0$, and $R = 1$. Now $\{F^s, GQ^{s/2}\} = \{1/2, 0\}$ is stabilizable and the DARE becomes

$$P = \frac{P}{4} - \frac{P^2}{4(1 + P)},$$

which has the two solutions $P_1 = 0$ and $P_2 = -3/4$. It is easy to see that P_1 is a stabilizing solution, since for this solution we have $K_p = 0$ so that $F_p = 1/2$. On the other hand, P_2 is not a stabilizing solution since it leads to $F_p = 2$. Thus, we now have a unique positive-semi-definite solution. ◆

The final result of Table E.1 deals with the question of the *positive-definiteness* of the unique stabilizing solution of the DARE. The necessary and sufficient condition is that $\{F, H\}$ be detectable and that $\{F^s, GQ^{s/2}\}$ be controllable on and inside the unit circle. This latter condition means that there exists a matrix K such that $F^s - GQ^{s/2}K$ has no eigenvalues on and inside the unit circle. An alternative characterization is that F^s has no stable modes that are uncontrollable.[2]

Example E.1.4 In Ex. E.1.3 the stabilizing solution was not positive-definite ($P_1 = 0$). In that case, we had $\{F^s, GQ^{s/2}\} = \{1/2, 0\}$, which is not controllable on and inside the unit circle. On the other hand, Ex. E.1.2 with $F = 2$, $G = 0$, $H = 1$, $Q = 1$, $S = 0$, and $R = 1$ had a positive-definite stabilizing solution ($P_2 = 3$). In this case, we had $\{F^s, GQ^{s/2}\} = \{2, 0\}$, which is controllable on and inside the unit circle. ◆

In all the above examples we had $S = 0$, so that if we had further assumed that F were stable all the various detectability, stabilizability, and unit-circle controllability conditions would have been automatically met. However, when $S \neq 0$ we cannot claim that F stable is necessary (or sufficient) for the existence of solutions with various properties. The point is that a nonzero S can affect the results by making $\{F^s, GQ^{s/2}\}$ nonstabilizable, or noncontrollable on the unit circle.

Example E.1.5 Consider $F = 0$, $G = 1$, $H = 1$, $Q = 1$, $S = 1$, and $R = 1$, so that the associated DARE becomes

$$P = 0 + 1 - (0 + 1)(1 + P)^{-1}(0 + 1),$$

or, after simplification, $P^2 = 0$. We thus have a repeated solution at $P = 0$, which is clearly not stabilizing since $F_p = -1$. Thus, even though F is stable, the DARE has no stabilizing solution, the reason being that the pair $\{F^s, GQ^s\} = \{-1, 0\}$ is not controllable on the unit circle. ◆

Outline of Arguments.

The outline of our arguments in studying the DARE is as follows. In Sec. E.2 we introduce a linear matrix inequality (LMI) closely related to the Popov function of Ch. 8 and show that, under detectability, it contains a maximal (more precisely, trace-maximizing) element. In Sec. E.3 we show that the maximal element of the LMI is a solution of the DARE. Sec. E.4 shows that the maximal element is a semi-stabilizing solution to the DARE and gives conditions under which the solution becomes stabilizing. Sec. E.5 contains the main result on necessary and sufficient conditions for the existence of a stabilizing solution to the DARE. Sec. E.6 gives some further results on the uniqueness of positive-semi-definite solutions, the existence of positive-definite solutions, and the properties of other solutions, such as the minimal solution. Sec. E.7 gives a description of the method of *invariant subspaces* for computing solutions of the DARE. Sec. E.8 studies the dual DARE and the final Sec. E.9 extends the results to the CARE.

[2] This has the following physical interpretation. If F^s has a stable mode that is uncontrollable, then the input cannot excite this mode, and hence the mode will become zero in steady state. This means that the uncertainty in this mode of the state vector is zero, and that therefore the state estimation error in the direction of this mode will also be zero. Since, P, the stabilizing solution to the DARE has the interpretation of being the steady-state state error covariance matrix, this implies that P must be singular.

Remark. We may mention that the main result of Thm. E.5.1 on the existence and uniqueness of a stabilizing solution to the DARE can be established in several different ways. Although we are starting here with an LMI in order to further highlight the connection between DAREs and Popov functions, the same result can also be obtained as a byproduct of the invariant subspace method used in Sec. E.7 — see the remark after Thm. E.7.2. We shall illustrate this latter approach when dealing with the CARE in Sec. E.9. of the DARE. ◆

E.2 A LINEAR MATRIX INEQUALITY

Recall from Ch. 8 the *Popov* function

$$S_y(z) = \begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}.$$

Given (E.1.1) it is clear that the Popov function is nonnegative-definite on the unit circle, *i.e.*,

$$S_y(e^{j\omega}) \geq 0 \qquad \text{for all} \qquad \omega \in [-\pi, \pi].$$

Moreover, from Lemma 8.2.1, we know that, for any $Z = Z^*$, we can rewrite the Popov function as

$$S_y(z) =$$

$$\begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} -Z + FZF^* + GQG^* & FZH^* + GS \\ HZF^* + S^*G^* & R + HZH^* \end{bmatrix} \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}.$$

Note that even though $S_y(z) \geq 0$, for all $|z| = 1$, we cannot make any assertions on the positivity of the center matrix, denoted by $N(Z)$,

$$N(Z) \triangleq \begin{bmatrix} -Z + FZF^* + GQG^* & FZH^* + GS \\ HZF^* + S^*G^* & R + HZH^* \end{bmatrix}. \tag{E.2.1}$$

However, let us concentrate on those values of Z that lead to a positive-semi-definite $N(Z)$, *i.e.*, let us consider the set

$$\mathcal{L} = \{Z = Z^* \mid N(Z) \geq 0\}. \tag{E.2.2}$$

The condition $N(Z) \geq 0$ is referred to as a *Linear Matrix Inequality* (LMI) and $\mathcal{L}$ is its so-called solution set. It is now straightforward to establish the following properties of the solution set $\mathcal{L}$.

Lemma E.2.1 (Properties of $\mathcal{L}$) *Suppose that $\{F, H\}$ is detectable and let $\mathcal{L}$ be the solution set of the* LMI $N(Z) \geq 0$. *Then $\mathcal{L}$ is (a) nonempty, (b) convex, (c) bounded above, and (d) closed.* ■

Proof: Only the proof of (c) is nonobvious. Thus take any $Z \in \mathcal{L}$. Since $\{F, H\}$ is detectable, consider a matrix K such that $F - KH$ is stable. Now using Sylvester's congruence theorem, $N(Z) \geq 0$ implies

$$\begin{bmatrix} I & -K \\ 0 & I \end{bmatrix} \begin{bmatrix} -Z + FZF^* + GQG^* & FZH^* + GS \\ HZF^* + S^*G^* & R + HZH^* \end{bmatrix} \begin{bmatrix} I & 0 \\ -K^* & I \end{bmatrix} \geq 0.$$

Clearly the $(1,1)$ block entry of the above product of matrices must be nonnegative-definite. Some simple algebra shows that the $(1,1)$ block entry is

$$-Z + (F - KH)Z(F - KH)^* + \underbrace{\begin{bmatrix} I & -K \end{bmatrix} \begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix} \begin{bmatrix} I \\ -K^* \end{bmatrix}}_{\triangleq \tilde{Q} \geq 0} \geq 0,$$

or, equivalently,

$$Z \leq (F - KH)Z(F - KH)^* + \tilde{Q}.$$

Now, since $F - KH$ is stable, the equation $\bar{\Pi} = (F - KH)\bar{\Pi}(F - KH)^* + \tilde{Q}$ has a unique positive-semi-definite solution $\bar{\Pi} \geq 0$. Subtracting the solution of the Lyapunov equation from the above inequality yields

$$Z - \bar{\Pi} \leq (F - KH)(Z - \bar{\Pi})(F - KH)^*.$$

If we reapply this inequality i times we obtain $Z - \bar{\Pi} \leq (F - KH)^i(Z - \bar{\Pi})(F - KH)^{*i}$, and since, by the stability of $F - KH$, $(F - KH)^i \to 0$, we obtain that in the limit, $Z - \bar{\Pi} \leq 0$, meaning that Z is bounded from above. ◆

E.3 EXISTENCE OF SOLUTIONS TO THE DARE

Since $\mathcal{L}$ is a convex, closed, and bounded (from above) set one can define its *trace-maximizing* element,

$$Z_+ \triangleq \arg\left(\max_{Z \in \mathcal{L}} \text{trace } Z\right). \tag{E.3.1}$$

It can further be shown that Z_+, as defined above, not only maximizes the trace, but is in fact the maximizing element of $\mathcal{L}$, *i.e.*, for any $Z \in \mathcal{L}$, we have $Z - Z_+ \leq 0$. Although we shall not need to use this more general property of Z_+, for simplicity we shall henceforth refer to Z_+ as the *maximal* (rather than trace-maximizing) element of $\mathcal{L}$.

Since $0 \in \mathcal{L}$, we have that $Z_+ \geq 0$. Our next claim is that Z_+ satisfies the DARE,

$$P = FPF^* + GQG^* - K_p R_e K_p^*, \tag{E.3.2}$$

where $K_p = (FPH^* + GS)R_e^{-1}$ and $R_e = R + HPH^*$. To verify this fact, we begin by noting that since $R > 0$ and $Z_+ \geq 0$, the matrix $R + HZ_+H^*$ is invertible. Therefore, we can perform the following triangular factorization of $N(Z_+)$:

$$N(Z_+) = \begin{bmatrix} I & K_{p,z} \\ 0 & I \end{bmatrix} \begin{bmatrix} -Z_+ + FZ_+F^* + GQG^* - K_{p,z}R_{e,z}K_{p,z} & 0 \\ 0 & R + HZ_+H^* \end{bmatrix} \begin{bmatrix} I & 0 \\ K_{p,z}^* & I \end{bmatrix},$$

where we have defined $K_{p,z} \triangleq (FZ_+H^* + GS)R_{e,z}^{-1}$ and $R_{e,z} \triangleq R + HZ_+H^*$. Now since $N(Z_+) \geq 0$, the $(1,1)$ block entry in the block diagonal matrix of the above triangular factorization must be positive-semi-definite, *i.e.*,

$$-Z_+ + FZ_+F^* + GQG^* - (FZ_+H^* + GS)(R + HZ_+H^*)^{-1}(FZ_+H^* + GS)^* \geq 0.$$

What we would like to show is that the above expression is indeed zero, so that Z_+ satisfies the DARE (E.3.2). To this end, suppose that the expression is not zero so that we may write

$$-Z_+ + \alpha(Z_+) = Q_2 \geq 0, \tag{E.3.3}$$

for some nonzero $Q_2 \geq 0$, and where we have defined

$$\alpha(Z) = FZF^* + GQG^* - (FZH^* + GS)(R + HZH^*)^{-1}(FZH^* + GS)^*. \tag{E.3.4}$$

Now the matrix function $\alpha(Z)$ has the following monotonicity property.

Lemma E.3.1 (Monotonicity Properties of) *Define*

$$K_{p,z_i} = (FZ_iH^* + GS)R_{e,z_i}^{-1} \text{ and } R_{e,z_i} = R + HZ_iH^*, \quad i = 1, 2.$$

The following facts hold for $\alpha(Z)$:

(i) The condition $Z_1 \geq Z_2$ implies $\alpha(Z_1) \geq \alpha(Z_2)$.

(ii) $\alpha(Z_2) - \alpha(Z_1) = (F - K_{p,z_1}H)(Z_2 - Z_1)(F - K_{p,z_2}H)^*$.

(iii) $\alpha(Z_2) - \alpha(Z_1) = (F - K_{p,z_1}H)(Z_2 - Z_1)(F - K_{p,z_1}H)^* -$

$$(F - K_{p,z_1}H)(Z_2 - Z_1)H^*R_{e,z_2}^{-1}H(Z_2 - Z_1)(F - K_{p,z_1}H)^*.$$

■

Proof: The proof of the equalities (ii) and (iii) is algebraic and is omitted. To prove the first claim, suppose that $Z_1 \geq Z_2$, so that using equality (iii) we may write

$$\alpha(Z_2) - \alpha(Z_1) = (F - K_{p,z_1}H)(Z_2 - Z_1)(F - K_{p,z_1}H)^* - (F - K_{p,z_1}H)(Z_2 - Z_1)H^*R_{e,z_2}^{-1}H(Z_2 - Z_1)(F - K_{p,z_1}H)^*.$$

Both terms on the right-hand side are negative-semi-definite. Therefore, we have $\alpha(Z_2) - \alpha(Z_1) \leq 0$, as desired. ◆

Now we can return to equation (E.3.3) and rewrite it as $-Z_+ - Q_2 + \alpha(Z_+) \geq 0$. But since $Q_2 \geq 0$ we have $\alpha(Z_+ + Q_2) \geq \alpha(Z_+)$, and therefore

$$-Z_+ - Q_2 + \alpha(Z_+ + Q_2) \geq 0. \tag{E.3.5}$$

But this implies that

$$N(Z_+ + Q_2) = \begin{bmatrix} -(Z_+ + Q_2) + F(Z_+ + Q_2)F^* + GQG^* & F(Z_+ + Q_2)H^* + GS \\ H(Z_+ + Q_2)F^* + S^*G^* & R + H(Z_+ + Q_2)H^* \end{bmatrix} \geq 0,$$

since $R + H(Z_+ + Q_2)H^* > 0$, and since the Schur complement of the (2, 2) entry, which is given by (E.3.5), is positive-semi-definite. This in turn implies that $(Z_+ + Q_2) \in \mathcal{L}$, which is a contradiction, since Z_+ is the trace-maximizing element of $\mathcal{L}$ and $\text{trace}(Z_+ + Q_2) > \text{trace}(Z_+)$. Therefore, our assumption that Q_2 is nonzero in (E.3.3) is false. Thus $Q_2 = 0$, meaning that Z_+ satisfies the DARE,

$$-Z_+ + FZ_+F^* + GQG^* - (FZ_+H^* + GS)(R + HZ_+H^*)^{-1}(FZ_+H^* + GS)^* = 0. \tag{E.3.6}$$

We have thus shown the following result.

Lemma E.3.2 (Existence of Solutions to the DARE) *If $\{F, H\}$ is detectable, then there always exists a positive-semi-definite solution to the* DARE

$$P = FPF^* + GQG^* - K_pR_eK_p^*,$$

where $K_p = (FPH^* + GS)R_e^{-1}$ and $R_e = R + HPH^*$. One such solution is given by Z_+, the maximal element of $\mathcal{L}$, defined by (E.3.1). ∎

E.4 PROPERTIES OF THE MAXIMAL SOLUTION

In this section we shall establish some of the properties of Z_+, the maximal solution of the DARE (E.3.2). The first result requires no further assumptions other than those we have made so far.

Lemma E.4.1 (Eigenvalues of $F - K_{p,z}H$) *Assume that $\{F, H\}$ is detectable, so that the maximizing element $Z_+ \geq 0$, which satisfies the* DARE *(E.3.2), exists. Then all the eigenvalues of $(F - K_{p,z}H)$ lie inside the closed unit disc, i.e.,*

$$|\lambda(F - K_{p,z}H)| \leq 1, \tag{E.4.1}$$

where $K_{p,z} = (FZ_+H^* + GS)R_{e,z}^{-1}$ and $R_{e,z} = R + HZ_+H^*$. ∎

Proof: Let x be a right eigenvector of $(F - K_{p,z}H)$ with eigenvalue λ, *i.e.*, $(F - K_{p,z}H)x = \lambda x$. Suppose now that $|\lambda| > 1$. We shall show that this assumption leads to a contradiction so that we must have $|\lambda| \leq 1$. [The proof requires some algebraic manipulations and can be omitted.]

Consider the matrix $(Z_+ + \delta xx^*)$ for some scalar $\delta > 0$. It follows that we can write

$$-(Z_+ + \delta xx^*) + \alpha(Z_+ + \delta xx^*) - [-Z_+ + \alpha(Z_+)] = -\delta xx^* + \alpha(Z_+ + \delta xx^*) - \alpha(Z_+).$$

But using equality (iii) in Lemma E.3.1, we can readily see that

$$\alpha(Z_+ + \delta xx^*) - \alpha(Z_+) = \delta x\left[|\lambda|^2 - \delta|\lambda|^2 x^*H^*(R_{e,z} + \delta Hxx^*H^*)^{-1}Hx\right]x^*.$$

Replacing this last result into the earlier expression yields

$$-(Z_+ + \delta xx^*) + \alpha(Z_+ + \delta xx^*) - [-Z_+ + \alpha(Z_+)] =$$

$$\delta x\left[(|\lambda|^2 - 1) - \delta|\lambda|^2 x^*H^*(R_{e,z} + \delta Hxx^*H^*)^{-1}Hx\right]x^*.$$

Note that when $|\lambda| > 1$, the term $(|\lambda|^2 - 1)$ is positive and therefore we can always choose a small enough $\delta > 0$ so that

$$\left[(|\lambda|^2 - 1) - \delta|\lambda|^2 x^*H^*(R_{e,z} + \delta Hxx^*H^*)^{-1}Hx\right] > 0.$$

This will then imply that

$$-(Z_+ + \delta xx^*) + \alpha(Z_+ + \delta xx^*) \geq -Z_+ + \alpha(Z_+) = 0, \tag{E.4.2}$$

where the equality $-Z_+ + \alpha(Z_+) = 0$ is due to the fact that Z_+ satisfies the DARE (E.3.2). Expression (E.4.2) means that we have a matrix $Z_+ + \delta x^*x \geq Z_+$ for which the Schur complement of the (2, 2) block entry of $N(Z)$ will be nonnegative and, consequently, $N(Z_+ + \delta x^*x) \geq 0$. This contradicts the maximality of Z_+ and, therefore, we must have $|\lambda| \leq 1$, as desired. ◆

We have shown that the existence of Z_+ guarantees that the matrix $F - K_{p,z}H$ has all its eigenvalues on and inside the unit circle. However, we are most often interested in the question of whether $F - K_{p,z}H$ is *stable*, *i.e.*, whether it has all its eigenvalues strictly inside the unit circle. The condition for this to hold is given below, in terms of the matrices $\{F^s, Q^s\}$ defined in (E.1.3).

Lemma E.4.2 (Stable $F - K_{p,z}H$) *Assume that $\{F, H\}$ is detectable so that $Z_+ \geq 0$ exists. Then $F - K_{p,z}H$, where $K_{p,z} = (FZ_+H^* + GS)R_{e,z}^{-1}$ and $R_{e,z} = R + HZ_+H^*$, is stable if, and only if, the pair $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle, i.e., if, and only if, there exists no left eigenvector x of F^s, $xF^s = \lambda x$, such that $|\lambda| = 1$ and such that $xGQ^{s/2} = 0$.* ∎

Proof: When $\{F, H\}$ is detectable we know from Lemma E.4.1 that all eigenvalues of $F - K_{p,z}H$ lie on and inside the unit circle. To prove one direction of the statement of the current lemma, suppose that there exists an eigenvalue on the unit circle, and let x be its corresponding left eigenvector. In other words, $x(F - K_{p,z}H) = \lambda x$ and $|\lambda| = 1$. Since the DARE (E.3.6) can be rewritten as

$$Z_+ = (F - K_{p,z}H)Z_+(F - K_{p,z}H)^* + \begin{bmatrix} G & -K_{p,z} \end{bmatrix}\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\begin{bmatrix} G^* \\ -K_{p,z}^* \end{bmatrix},$$

if we pre- and post-multiply the above expression by x and x^*, respectively, we obtain

$$(1 - |\lambda|^2)xZ_+x^* = x\begin{bmatrix} G & -K_{p,z} \end{bmatrix}\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\begin{bmatrix} G^* \\ -K_{p,z}^* \end{bmatrix}x^*,$$

and since $|\lambda| = 1$,

$$x\begin{bmatrix} G & -K_{p,z} \end{bmatrix}\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\begin{bmatrix} G^* \\ -K_{p,z}^* \end{bmatrix}x^* = 0. \tag{E.4.3}$$

But this implies that

$$x[G(Q - SR^{-1}S^*)G^* + (GSR^{-1} - K_{p,z})R(GSR^{-1} - K_{p,z})^*]x^* = 0,$$

from which we infer that

$$xG(Q - SR^{-1}S^*)G^*x^* = 0 \text{ and } xK_{p,z} = xGSR^{-1}. \tag{E.4.4}$$

We now see that

$$\lambda x = x(F - K_{p,z}H) = xF - xK_{p,z}H = xF - xGSR^{-1}H = x(F - GSR^{-1}H),$$

i.e., x is also a left eigenvector of $F - GSR^{-1}H$ with eigenvalue λ. The first equality in (E.4.4) along with the above result can be rewritten as

$$xG(Q - SR^{-1}S^*)^{1/2} = 0 \text{ and } x(F - GSR^{-1}H) = \lambda x.$$

This of course implies that the pair $\{F^s, GQ^{s/2}\}$ is not controllable on the unit circle.

To prove the other direction assume that the pair $\{F^s, GQ^{s/2}\}$ is not controllable on the unit circle. This implies that there exists a unit modulus λ and a vector x such that

$$xF^s = \lambda F^s, \quad |\lambda| = 1, \quad \text{and} \quad xGQ^{s/2} = 0. \tag{E.4.5}$$

Now with the definitions of F^s and Q^s, it is straightforward to see that the DARE can be written as

$$Z_+ = F^sZ_+F^{s*} + GQ^sG^* - F^sZ_+H^*R_{e,z}^{-1}HZ_+F^{s*}.$$

Moreover, we have $F_{p,z} = F^s - F^sZ_+H^*R_{e,z}^{-1}H$. Premultiplying both sides of the above DARE by x, and using (E.4.5), yields

$$xZ_+ = \lambda xZ_+F^{s*} - \lambda xZ_+H^*R_{e,z}^{-1}HZ_+F^{s*} = \lambda xZ_+\underbrace{(F^{s*} - H^*R_{e,z}^{-1}HZ_+F^{s*})}_{F_{p,z}^*}$$

so that

$$xZ_+F_{p,z}^* = \lambda^{-1}xZ_+. \tag{E.4.6}$$

Now either of the following mutually exclusive cases may happen:

(i) $xZ_+ \neq 0$, *i.e.*, $F_{p,z}$ has a unit-circle eigenvalue at $\lambda^{-*}(= \lambda)$.

(ii) $xZ_+ = 0$. Then $xF_{p,z} = x(F^s - F^sZ_+H^*R_{e,z}^{-1}H) = \lambda x - \lambda xZ_+H^*R_{e,z}^{-1}H = \lambda x$, so that $F_{p,z}$ has a unit-circle eigenvalue at λ.

In either case, $F_{p,z}$ has a unit-circle eigenvalue and the lemma is proved. ◆

Lemma E.4.3 (Unique Stabilizing Solution) *Any stabilizing solution to the* DARE,

$$P = FPF^* + GQG^* - K_pR_eK_p^*,$$

i.e., a solution for which $F - K_pH$ is stable, if one exists, is unique. ■

Proof: Suppose that we have two stabilizing solutions, P_1 and P_2, with corresponding gain matrices $K_{p,1}$ and $K_{p,2}$. Now since $\alpha(P_1) = P_1$ and $\alpha(P_2) = P_2$, using the second result of Lemma E.3.1, we have

$$P_2 - P_1 = (F - K_{p,1}H)(P_2 - P_1)(F - K_{p,2}H)^*.$$

Applying the above equality i times, we obtain

$$P_2 - P_1 = (F - K_{p,1}H)^i(P_2 - P_1)(F - K_{p,2}H)^{i*}.$$

Now since the matrices $F - K_{p,1}H$ and $F - K_{p,2}H$ are both stable, as $i \to \infty$, the above equation becomes $P_2 - P_1 = 0$, showing that the stabilizing solution is unique. ◆

We remark that when $\{F, H\}$ is detectable and when $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle, the unique stabilizing solution is given by $P = Z_+ \geq 0$.

E.5 MAIN RESULT

We are now in a position to state and prove the main result.

Theorem E.5.1 (Algebraic Riccati Equation) *Consider the discrete-time algebraic Riccati equation*

$$P = FPF^* + GQG^* - (FPH^* + GS)(R + HPH^*)^{-1}(FPH^* + GS)^*.$$

Then the following two statements are equivalent.

(i) $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle.

(ii) The DARE *has a stabilizing solution P, i.e., one for which the matrix $F - K_pH$ is stable, where $K_p = (FPH^* + GS)(R + HPH^*)^{-1}$.*

Moreover, any such stabilizing solution is unique and positive-semi-definite. ■

Proof: When $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle, Lemma E.4.2 shows that the solution $P = Z_+$ exists and is stabilizng. This shows that (i) implies (ii). To prove the converse, *i.e.*, that (ii) implies (i), note first that if $\{F, H\}$ is not detectable, then it is clear that a stabilizing solution cannot exist (since $F - KH$ cannot be stable for *any* K). Therefore assume that $\{F, H\}$ is detectable but that $\{F^s, GQ^{s/2}\}$ has an uncontrollable unit-circle eigenvalue, say λ ($|\lambda| = 1$). Then using an argument similar to the one presented in the proof of the second direction of Lemma E.4.2 it can be shown that, for *any* solution P to the DARE, λ must be an eigenvalue of $F - K_pH$. Thus the DARE cannot have a stabilizing solution.

The fact that the stabilizing solution is unique follows from Lemma E.4.3. That it is positive-semi-definite follows from the fact that the unique stabilizing solution is given by $P = Z_+ \geq 0$. ◆

Corollary E.5.1. (F^s Stable) *F^s stable is a sufficient condition for the* DARE *to have a stabilizing solution (which is also unique and positive-semi-definite).* ◆

Proof: When $F^s = F - GSR^{-1}H$ is stable, clearly $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is unit-circle controllable. ◆

E.6 FURTHER REMARKS

So far we have only focused on the maximal (and thereby positive-semi-definite and stabilizing) solution of the DARE. There are, of course, many other solutions to the DARE, but they are of less interest to us since they do not correspond to the canonical factorization of a z-spectrum.

Another question of interest is whether other positive-semi-definite solutions to the DARE exist or not. [This question is especially important in Ch. 14 where we study the convergence of the Riccati recursion to solutions of the DARE.] From Theorem E.5.1 we know that when $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle, then one positive-semi-definite solution is the maximal (and stabilizing) solution Z_+. However, in this case, positive-semi-definite solutions are not necessarily unique. To ensure uniqueness of the positive-semi-definite solution we need a stronger condition, namely that $\{F^s, GQ^{s/2}\}$ be controllable on and outside the unit circle; a condition that is known as *stabilizability*.[3]

Theorem E.6.1 (Positive Semi-Definite Solutions) *Consider the discrete-time algebraic Riccati equation (*DARE*)*

$$P = FPF^* + GQG^* - (FPH^* + GS)(R + HPH^*)^{-1}(FPH^* + GS)^*,$$

and assume that $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle. Then the following two statements are equivalent.

(i) $\{F^s, GQ^{s/2}\}$ is stabilizable.

(ii) The DARE *has a unique positive-semi-definite solution.*

Moreover, the unique positive-semi-definite solution of the DARE *is given by the maximal (and stabilizing) solution Z_+.* ■

Proof: Recall first that, in terms of the definition (E.1.3) of $\{F^s, Q^s\}$, the DARE can be rewritten as

$$P = F^sPF^{s*} + GQ^sG^* - F^sPH^*(R + HPH^*)^{-1}HPF^{s*}.$$

(i)⇒(ii): Suppose that the DARE has a positive-semi-definite solution, $P \geq 0$, that is different from the maximal solution, $P \neq Z_+$. We shall show that $\{F^s, GQ^{s/2}\}$ is not stabilizable. To this end, we write the DARE as

$$P = F_pPF_p^* + GQ^sG^* + K_pRK_p^*,$$

[3] This condition is the dual of detectability: $\{F, G\}$ is stabilizable if, and only if, $\{F^*, G^*\}$ is detectable, *i.e.*, if, and only if, there exists a K such that $F - GK = (F^* - K^*G^*)^*$ is stable.

where $F_p = F^s - K_pH$ and $K_p = F^sPH^*(R+HPH^*)^{-1}$, and note that since $P \neq Z_+$, and since the stabilizing solution is unique (by Lemma E.4.3), the matrix F_p cannot be stable. Let λ (with $|\lambda| \geq 1$) be an unstable mode of F_p with corresponding left eigenvector x. Pre- and post-multiplying the above DARE by x and x^*, yields

$$(1 - |\lambda|^2)xPx^* = xGQ^sG^*x^* + xK_pRK_p^*x^*.$$

Since $(1 - |\lambda|^2) \leq 0$ and $xPx^* \geq 0$ we conclude that

$$xGQ^sG^*x^* = 0 \quad \text{and} \quad xK_pRK_p^*x^* = 0.$$

But since $R > 0$, this implies $xK_p = 0$. This now shows that $\{F^s, GQ^{s/2}\}$ is not stabilizable since $xGQ^{s/2} = 0$ and

$$xF^s = xF^s - xK_pH = x(F^s - K_pH) = xF_p = \lambda x.$$

(ii)⇒(i): Suppose that $\{F^s, GQ^{s/2}\}$ is not stabilizable. We shall now show that the DARE has more than one positive-semi-definite solution. To this end, suppose that λ (with $|\lambda| \geq 1$) is an unstabilizable mode of $\{F^s, GQ^{s/2}\}$ with corresponding left eigenvector x. Consider the maximal solution of the DARE,

$$Z_+ = F^sZ_+F^* + GQ^sG^* - K_{p,z}R_{e,z}K_{p,z}^*,$$

where $K_{p,z} = F^sZ_+H^*R_{e,z}^{-1}$ and $R_{e,z} = R + HZ_+H^*$. Now pre-multiplying the DARE by x yields

$$\begin{aligned} xZ_+ &= \underbrace{xF^s}_{=\lambda x}Z_+F^{s*} + \underbrace{xGQ^s}_{=0}G^* - x\underbrace{K_{p,z}R_{e,z}}_{=F^sZ_+H^*}K_{p,z}^* \\ &= \lambda xZ_+F^{s*} - xF^sZ_+H^*K_{p,z}^* \\ &= \lambda xZ_+F^{s*} - \lambda xZ_+H^*K_{p,z}^* = \lambda xZ_+(F^s - K_{p,z}H)^*. \end{aligned}$$

Defining $F_{p,z} = F^s - K_{p,z}H$, we can thus conclude that

$$F_{p,z}Z_+x^* = \lambda^{-*}Z_+x^*. \tag{E.6.1}$$

We should also note that $Z_+x^* \neq 0$. Indeed if $Z_+x^* = 0$ we obtain

$$xF_{p,z} = xF^s - xF^sZ_+H^*R_{e,z}^{-1}H = xF^s - \lambda xZ_+H^*R_{e,z}^{-1}H = xF^s = \lambda x,$$

which clearly contradicts the fact that $F_{p,z}$ is stable.

We can therefore construct the following positive-semi-definite matrix (unequal to Z_+ since $Z_+x^* \neq 0$)

$$P = Z_+ - \frac{Z_+x^*xZ_+}{xZ_+x^*} \geq 0. \tag{E.6.2}$$

[Note that P can be regarded as the projection operator that projects onto the orthogonal complement space of x in the Z_+ norm. This is one justification of why $P \geq 0$.

The other is by algebraic verification.] We can finally show, after some algebra that uses Lemma E.3.1 and (E.6.1), that

$$\alpha(P) = \alpha(Z_+) - \frac{Z_+ x^* x Z_+}{x Z_+ x^*} = \alpha(Z_+) - Z_+ + P.$$

But since Z_+ is a solution to the DARE we have $\alpha(Z_+) = Z_+$ and therefore $\alpha(P) = P$. This means that P also satisfies the DARE. ◆

One can also give the condition for which the maximal solution of the DARE is positive-definite.

Theorem E.6.2 (Positive Definite Solution) *Consider the discrete-time algebraic Riccati equation*

$$P = FPF^* + GQG^* - (FPH^* + GS)(R + HPH^*)^{-1}(FPH^* + GS)^*,$$

and assume that $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle. Then the following two statements are equivalent.

(i) $\{F^s, GQ^{s/2}\}$ is controllable inside the unit circle.

(ii) The DARE *has a positive-definite solution.*

Moreover, if the above conditions hold, the maximal (and stabilizing) solution Z_+ is one such positive-definite solution. ■

Proof: We prove both directions.

(i)⇒(ii): Suppose that the maximal solution of the DARE, Z_+, is singular. We shall show that $\{F^s, GQ^{s/2}\}$ cannot be controllable inside the unit circle. To this end, write the DARE (E.3.6) as

$$Z_+ = F_{p,z} Z_+ F_{p,z}^* + GQ_s G^* + K_{p,z} R K_{p,z}^*,$$

where $F_{p,z} = F^s - K_{p,z}H$ and $K_{p,z} = F^s Z_+ H^*(R + HZ_+H^*)^{-1}$. Now let x be a vector in the left null space of Z_+, *i.e.*, $xZ_+ = 0$. Pre- and post-multiplying the DARE by x and x^* yields

$$0 = xF_{p,z}Z_+F_{p,z}^*x^* + xGQ^sG^*x^* + xK_{p,z}RK_{p,z}^*x^*,$$

which clearly implies that

$$xF_{p,z}Z_+ = 0, \quad xGQ^{s/2} = 0, \quad xK_{p,z} = 0.$$

The first equality implies that if $x \in \mathcal{N}(Z_+)$, where $\mathcal{N}(Z_+)$ denotes the left null space of Z_+, then $xF_{p,z} \in \mathcal{N}(Z_+)$. In other words, $\mathcal{N}(Z_+)$ is a left invariant subspace of $F_{p,z}$. Any such invariant subspace must contain a left eigenvector of $F_{p,z}$, and we may therefore take x to be such an eigenvector, *i.e.*, $xF_{p,z} = \lambda x$ with $|\lambda| < 1$ since, by Thm. E.5.1, $F_{p,z}$ is stable. Now

$$xF^s = xF^s - xK_{p,z}H = x(F^s - K_{p,z}H) = xF_{p,z} = \lambda x,$$

which along with $xGQ^{s/2} = 0$ shows that $\{F^s, GQ^{s/2}\}$ is not controllable inside the unit circle.

(ii)⇒(i): Suppose that $\{F^s, GQ^{s/2}\}$ is not controllable inside the unit circle. We shall now show that Z_+ is singular. To this end, suppose that λ, with $|\lambda| < 1$, is an uncontrollable mode of $\{F^s, GQ^{s/2}\}$ with corresponding left eigenvector, x, *i.e.*, $xF^s = \lambda x$ and $xGQ^{s/2} = 0$. Then pre-multiplying the DARE

$$Z_+ = F^s Z_+ F^{s*} + GQ^s G^* - F^s Z_+ H^* K_{p,z}^*$$

by x yields

$$xZ_+ = \lambda x Z_+ F^{s*} + 0 - \lambda x Z_+ H^* K_{p,z}^* = \lambda x Z_+ (F^s - K_{p,z}H)^* = \lambda x Z_+ F_{p,z}^*.$$

The above equation shows that if $xZ_+ \neq 0$, then $F_{p,z}$ has an unstable eigenvalue at λ^{-*}. But this cannot be, since $F_{p,z}$ is stable. Therefore $xZ_+ = 0$. ◆

Another interesting solution to the DARE is the minimal solution. In particular, it can be shown that if $\{F, H\}$ is antidetectable (meaning that there exists a constant matrix K such that $F - KH$ is antistable), then there exists a *minimal* solution, $Z_- \leq 0$, to the linear matrix inequality $N(Z) \geq 0$, that also satisfies the DARE. We shall not give the details here, but we can follow the same reasoning that was given for the maximal solution to obtain the following result.

Theorem E.6.3 (Minimal Solution) *Consider the discrete-time algebraic Riccati equation*

$$P = FPF^* + GQG^* - (FPH^* + GS)(R + HPH^*)^{-1}(FPH^* + GS)^*.$$

Then the following two statements are equivalent.

(i) $\{F, H\}$ is antidetectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle.

(ii) The DARE *has an antistabilizing solution P, i.e., one for which the matrix $F - K_p H$ is antistable, where $K_p = (FPH^* + GS)(R + HPH^*)^{-1}$.*

Moreover, any such antistabilizing solution is unique and negative-semi-definite. If, in addition, $\{F^s, GQ^{s/2}\}$ is controllable, then the unique stabilizing solution P is negative-definite. ■

E.7 THE INVARIANT SUBSPACE METHOD

We now review the method of invariant subspaces for solving the discrete-time algebraic Riccati equation (DARE). Apart from leading to further insights and properties of the DARE, the method has also resulted in reliable and efficient numerical solutions for various types of algebraic Riccati equations (see, *e.g.*, Patel, Laub, and Van Dooren (1994)). The literature on numerical solutions to the algebraic Riccati equation is very extensive and still growing.

As mentioned earlier in Sec. 9.5.1 and Ch. 14, when $R > 0$there is no loss of generality in assuming that $S = 0$, so that from now on we deal with the following DARE:

$$P = FPF^* + GQG^* - FPH^*(R + HPH^*)^{-1}HPF^*. \qquad \text{(E.7.1)}$$

The Case of Nonsingular F.
The assumption that F is nonsingular simplifies some of the arguments; the general case is discussed later.

We begin by noting that, after some algebra, we can write the above DARE as a quadratic form in P:

$$[I \ -P]\underbrace{\begin{bmatrix} F^{-1} & -F^{-1}GQG^* \\ -H^*R^{-1}HF^{-1} & F^* + H^*R^{-1}HF^{-1}GQG^* \end{bmatrix}}_{\triangleq M}\begin{bmatrix} P \\ I \end{bmatrix} = 0, \qquad \text{(E.7.2)}$$

It is easy to check that M is a *symplectic* matrix, *i.e.*, it satisfies the relation

$$J^{-1}M^*J = M^{-1}, \quad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}. \qquad \text{(E.7.3)}$$

[To verify this it is convenient to begin with the block LDU decomposition of M

$$M = \begin{bmatrix} I & 0 \\ -H^*R^{-1}H & I \end{bmatrix}\begin{bmatrix} F^{-1} & 0 \\ 0 & F^* \end{bmatrix}\begin{bmatrix} I & -GQG^* \\ 0 & I \end{bmatrix}, \qquad \text{(E.7.4)}$$

which will simplify the algebra.]

Lemma E.7.1 (Eigenvalues of the Symplectic Matrix) *Let λ denote any eigenvalue of the symplectic matrix M. Then*

(i) $\lambda \neq 0$.

(ii) $1/\lambda^*$ *is also an eigenvalue of* M.

(iii) If $\{F, H\}$ *is detectable and* $\{F, GQ^{1/2}\}$ *is unit-circle controllable, then* $|\lambda| \neq 1$.

■

Proof: We proceed as follows:
(i) Using the LDU representation (E.7.4) for M we have

$$\begin{bmatrix} I & -GQG^* \\ 0 & I \end{bmatrix}x = \lambda\begin{bmatrix} F & 0 \\ 0 & F^{-*} \end{bmatrix}\begin{bmatrix} I & 0 \\ H^*R^{-1}H & I \end{bmatrix}x. \qquad \text{(E.7.5)}$$

Now assume $\lambda = 0$. Then

$$\begin{bmatrix} I & -GQG^* \\ 0 & I \end{bmatrix}x = 0,$$

which is only possible for $x = 0$. This contradicts the assumption of a nonzero x.
(ii) This follows from relation (E.7.3), which shows that M^{-1} and M^* are similar matrices.
(iii) Assume $|\lambda| = 1$ and partition x into $x = \text{col}\{x_1, x_2\}$. Then (E.7.5) implies that

$$x_1 - GQG^*x_2 = \lambda F x_1, \qquad F^*x_2 = \lambda H^*R^{-1}Hx_1 + \lambda x_2, \qquad \text{(E.7.6)}$$

which can be combined to yield

$$x_2^*GQG^*x_2 + x_1^*HR^{-1}Hx_1 = 0,$$

so that $x_2^*GQ^{1/2} = 0$ and $R^{-1/2}Hx_1 = 0$. Substituting into (E.7.6) we get $F^*x_2 = \lambda x_2$ and $Fx_1 = \frac{1}{\lambda}x_1$. These equations show that $\{x_2, x_1\}$ are left and right eigenvectors for F that are orthogonal to $GQ^{1/2}$ and H, respectively, thus contradicting the stabilizability and unit-circle controllability assumptions. ◆

It is further convenient to expand the block row and column vectors appearing in (E.7.2) into upper triangular matrices, to obtain

$$\begin{bmatrix} I & -P \\ 0 & I \end{bmatrix}\begin{bmatrix} F^{-1} & -F^{-1}GQG^* \\ -H^*R^{-1}HF^{-1} & F^* + H^*R^{-1}HF^{-1}GQG^* \end{bmatrix}\begin{bmatrix} I & P \\ 0 & I \end{bmatrix}.$$

By (E.7.2), the $(1,2)$ block entry in the above resulting product is zero. Some simple algebra then leads to

$$\begin{bmatrix} I & -P \\ 0 & I \end{bmatrix}M\begin{bmatrix} I & P \\ 0 & I \end{bmatrix} = \begin{bmatrix} F_p^{-1} & 0 \\ -H^*R^{-1}HF^{-1} & F_p^* \end{bmatrix}, \qquad \text{(E.7.7)}$$

where, as usual, we have defined

$$F_p = F - K_pH, \qquad K_p = FPH^*(R + HPH^*)^{-1}. \qquad \text{(E.7.8)}$$

The invertiblity of F and R guarantees the invertibility of F_p. The relation (E.7.7) shows that the eigenvalues of the symplectic matrix M are determined by the eigenvalues of the closed-loop matrix, F_p. More specifically, we have the following easily proved result.

Lemma E.7.2 (Solutions of DARE) *Assume F is invertible. Let P denote any solution of the* DARE *(E.7.1), when it exists. Then P satisfies*

$$M\begin{bmatrix} P \\ I \end{bmatrix} = \begin{bmatrix} P \\ I \end{bmatrix}F_p^*. \qquad \text{(E.7.9)}$$

That is, $\{P, I\}$ forms a basis for an eigenspace of M with eigenvalues given by those of F_p^*. Moreover,

$$M = \begin{bmatrix} I & P \\ 0 & I \end{bmatrix}\begin{bmatrix} F_p^{-1} & 0 \\ -H^*R^{-1}HF^{-1} & F_p^* \end{bmatrix}\begin{bmatrix} I & P \\ 0 & I \end{bmatrix}^{-1}, \qquad \text{(E.7.10)}$$

showing that the eigenvalues of M coincide with those of $\{F_p^*, F_p^{-1}\}$. ■

Now assume $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is unit-circle controllable. Then the DARE (E.7.1) has a unique stabilizing solution P with the corresponding closed-loop matrix F_p stable. Moreover, from part (iii) of Lemma E.7.1, the symplectic matrix M will not have unit-circle eigenvalues; its stable eigenvalues will coincide with the eigenvalues of F_p^* and its unstable eigenvalues will coincide with the eigenvalues of

F_p^{-1}. Thus there exists an invertible similarity transformation T that reduces M to the form

$$M = T \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda^{-*} \end{bmatrix} T^{-1}, \tag{E.7.11}$$

where Λ is composed of Jordan blocks that correspond to the eigenvalues of M that are strictly inside the unit circle. We further partition the entries of T into

$$T = \begin{bmatrix} U & T_{12} \\ V & T_{22} \end{bmatrix},$$

so that using (E.7.11) and Eq. (E.7.7) for M, we get

$$\begin{bmatrix} F_p^{-1} & 0 \\ -H^*R^{-1}HF^{-1} & F_p^* \end{bmatrix} \begin{bmatrix} I & -P \\ 0 & I \end{bmatrix} \begin{bmatrix} U & T_{12} \\ V & T_{22} \end{bmatrix} =$$

$$\begin{bmatrix} I & -P \\ 0 & I \end{bmatrix} \begin{bmatrix} U & T_{12} \\ V & T_{22} \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda^{-*} \end{bmatrix}.$$

This last expression can be rewritten as

$$\begin{bmatrix} F_p^{-1} & 0 \\ -H^*R^{-1}HF^{-1} & F_p^* \end{bmatrix} \begin{bmatrix} U - PV & T_{12} - PT_{22} \\ V & T_{22} \end{bmatrix} =$$

$$\begin{bmatrix} U - PV & T_{12} - PT_{22} \\ V & T_{22} \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda^{-*} \end{bmatrix}.$$

Equating the (1, 1) block entries in the above equation yields

$$F_p^{-1}(U - PV) = (U - PV)\Lambda. \tag{E.7.12}$$

Now since F_p is stable, F_p^{-1} will have all its eigenvalues strictly *outside* the unit circle. On the other hand, by our construction of the eigenvalue decomposition, Λ will have all its eigenvalues strictly *inside* the unit circle.

It can be verified that if two matrices A and B have no common eigenvalues (see the discussion on Lyapunov equations in App. D) then the unique solution to the matrix equation $AX - XB = 0$, is $X = 0$. Applying this result to Eq. (E.7.12) we conclude that we must have

$$U - PV = 0. \tag{E.7.13}$$

We now show that V is invertible. To do so, suppose that V is singular and x is a vector such that $Vx = 0$. Applying x to both sides of (E.7.13) yields $Ux - PVx = Ux = 0$. But this then implies that

$$\begin{bmatrix} U & T_{12} \\ V & T_{22} \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} = 0,$$

which is a contradiction since T is nonsingular. Therefore V is nonsingular, and we may use (E.7.13) to conclude that

$$P = UV^{-1}. \tag{E.7.14}$$

In summary, we have derived the following so-called invariant subspace method for the computation of the stabilizaing solution P.

Theorem E.7.1 (The Invariant Subspace Method) *Consider the* DARE *(E.7.1) with F invertible, $\{F, H\}$ detectable, and $\{F, GQ^{1/2}\}$ unit-circle controllable. Let U and V be any $n \times n$ matrices that form a basis for the stable eigenspace of the symplectic matrix M in (E.7.2), viz.,*

$$M \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} \Lambda, \tag{E.7.15}$$

where Λ is an $n \times n$ matrix with all its eigenvalues inside the unit disc, $|\lambda_i(\Lambda)| < 1$. Then

(i) V is invertible.

(ii) UV^{-1} is the unique stabilizing solution of the DARE *(E.7.1).*

■

Continuing with the assumptions of detectable $\{F, H\}$ and unit-circle controllable $\{F, GQ^{1/2}\}$, let P again denote the unique stabilizing solution of the DARE (E.7.1). With this choice of P, Eq. (E.7.7) implies the similarity relation

$$\begin{bmatrix} I & P \\ 0 & I \end{bmatrix}^{-1} M \begin{bmatrix} I & P \\ 0 & I \end{bmatrix} = \begin{bmatrix} F_p^{-1} & 0 \\ -H^*R^{-1}HF^{-1} & F_p^* \end{bmatrix}. \tag{E.7.16}$$

Moreover, some simple algebra shows that

$$\begin{bmatrix} F_p^{-1} & 0 \\ -H^*R^{-1}HF^{-1} & F_p^* \end{bmatrix} = \begin{bmatrix} I & 0 \\ \mathcal{O}^p & I \end{bmatrix}^{-1} \begin{bmatrix} F_p^{-1} & 0 \\ 0 & F_p^* \end{bmatrix} \begin{bmatrix} I & 0 \\ \mathcal{O}^p & I \end{bmatrix}, \tag{E.7.17}$$

where we have defined $\mathcal{O}^p$ via the solution to the Lyapunov equation,

$$\mathcal{O}^p = F_p^* \mathcal{O}^p F_p + H^* R_e^{-1} H. \tag{E.7.18}$$

[Note that since F_p is stable, the above equation has a unique (positive-semi-definite) solution.] Combining (E.7.16) and (E.7.17), we may write

$$\begin{bmatrix} I - P\mathcal{O}^p & P \\ -\mathcal{O}^p & I \end{bmatrix}^{-1} M \begin{bmatrix} I - P\mathcal{O}^p & P \\ -\mathcal{O}^p & I \end{bmatrix} = \begin{bmatrix} F_p^{-1} & 0 \\ 0 & F_p^* \end{bmatrix}. \tag{E.7.19}$$

In other words, one possible choice for a similarity transformation T that reduces M to a block diagonal matrix of the form $\{\Lambda, \Lambda^{-*}\}$ is

$$T = \begin{bmatrix} I - P\mathcal{O}^p & P \\ -\mathcal{O}^p & I \end{bmatrix}.$$

Using (E.7.19), the antistable invariant subspace of M is spanned by

$$\begin{bmatrix} I - P\mathcal{O}^p \\ -\mathcal{O}^p \end{bmatrix}. \tag{E.7.20}$$

When $\{F, H\}$ is observable (so that it is both detectable and antidetectable), the pair $\{F_p, R_e^{*/2}H\}$ is also observable and therefore the solution to the Lyapunov equation $\mathcal{O}^p = F_p^* \mathcal{O}^p F_p + H^* R_e H$ is positive definite, *i.e.*, $\mathcal{O}^p > 0$. Using an argument similar to the one that led to the proof of Thm. E.7.1, this then allows us to conclude that P_-, the antistabilizing solution to the DARE, is given by

$$P_- = -(I - P\mathcal{O}^p)(\mathcal{O}^p)^{-1} = -(\mathcal{O}^p)^{-1} + P. \tag{E.7.21}$$

Unitary Triangularization (or Schur Form) Method.
The connection between an apparently nonlinear matrix Riccati equation of order n and a linear eigenvalue problem of order $2n$ is classic and dates back at least to Von Escherich (1898). The above "eigenvector" solution method was popularized in the control literature by MacFarlane (1963) and Potter (1966). However, it can encounter severe numerical difficulties when the symplectic matrix, M, has multiple or near multiple eigenvalues.

To ameliorate these difficulties, (unitary) Schur triangularization methods were proposed in (Laub (1979)). The procedure is essentially the same as the above except that instead of determining a nonsingular matrix T of eigenvectors, a unitary matrix U is computed so that M is reduced to triangular form (see, *e.g.*, Golub and Van Loan (1996)), say

$$\begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}^* \begin{bmatrix} F^{-1} & -F^{-1}GQG^* \\ -H^*R^{-1}HF^{-1} & F^* + H^*R^{-1}HF^{-1}GQG^* \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix},$$

where S_{11} is upper triangular with eigenvalues strictly inside the unit disk and S_{22} is upper triangular with eigenvalues strictly outside the unit disk. The key observation is that

$$\begin{bmatrix} U \\ V \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} U_{11} \\ U_{21} \end{bmatrix}$$

span the same invariant subspace so that it is easily seen that P can also be computed via $P = U_{11}U_{21}^{-1}$.

The Case of General F.
The case of singular F can be treated by noting that the symplectic M can be factored and written as

$$M = \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix}^{-1} \begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix},$$

and, therefore, instead of computing eigenvalues and eigenvectors of M, we can compute the generalized eigenvalues and eigenvectors of the pair

$$\left\{ \begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix}, \quad \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix} \right\}. \tag{E.7.22}$$

Lemma E.7.3 (Generalized Eigenvalue Problem) *Let λ denote any generalized eigenvalue of the pair of matrices (E.7.22), i.e.,*

$$\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix} x = \lambda \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix} x, \tag{E.7.23}$$

for some nonzero vector x. Then

(i) $\lambda = 0$ is an eigenvalue if, and only if, F is singular.

(ii) When $\lambda \neq 0$, $1/\lambda^*$ is also an eigenvalue with the same multiplicity. Moreover, assume $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is unit-circle controllable. Then

(iii) $|\lambda| \neq 1$.

(iv) If $\lambda = 0$ is an eigenvalue with multiplicity r, then there are only $2(n-r)$ nonzero eigenvalues (where $n \times n$ is the size of F). [The remaining r eigenvalues can be said to be at infinity.]

■

Proof: We proceed as follows:
(i) Assume F is invertible and that a zero eigenvalue exists. This means that, for some nonzero x,

$$\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix} x = 0,$$

from which it is immediate to conclude that x must be zero; a contradiction. Therefore, F is singular. Conversely, assume F is singular, then

$$\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix}$$

is singular and there exists a nonzero vector x such that

$$\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix} x = 0.$$

This implies that $\lambda = 0$ is a generalized eigenvalue.
(ii) First we note that if $\lambda \neq 0$ is a generalized eigenvalue for a pair $\{A, B\}$ then λ^{-*} is a generalized eigenvalue for the conjugate pair $\{B^*, A^*\}$. Now a nonzero generalized eigenvalue of the pair of matrices (E.7.22) is a value λ such that

$$\det\left(\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix} - \lambda \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix} \right) = 0,$$

or, equivalently,

$$\det \begin{bmatrix} I - \lambda F & -GQG^* \\ -\lambda H^* R^{-1} H & F^* - \lambda I \end{bmatrix} = 0.$$

Now note the similarity transformation

$$\begin{bmatrix} 0 & \frac{1}{\lambda} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} I - \lambda F & -GQG^* \\ -\lambda H^* R^{-1} H & F^* - \lambda I \end{bmatrix} \begin{bmatrix} 0 & 1 \\ \lambda & 0 \end{bmatrix} = \begin{bmatrix} F^* - \lambda I & H^* R^{-1} H \\ -\lambda GQG^* & I - \lambda F \end{bmatrix},$$

so that the right-hand side matrix also drops rank at λ. However, this matrix is equal to

$$\begin{bmatrix} F & 0 \\ H^* R^{-1} H & I \end{bmatrix}^* - \lambda \begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix}^*. \tag{E.7.24}$$

This means that if $\lambda \neq 0$ is a generalized eigenvalue for the pair of matrices (E.7.22), then it is also a generalized eigenvalue for the corresponding conjugate pair. It thus follows that λ^{-*} is also a generalized eigenvalue for the pair of matrices (E.7.22).

(iii) Assume $|\lambda| = 1$ and partition x into $x = \text{col}\{x_1, x_2\}$. Then (E.7.23) implies that

$$x_1 - GQG^* x_2 = \lambda F x_1, \qquad F^* x_2 = \lambda H^* R^{-1} H x_1 + \lambda x_2, \tag{E.7.25}$$

which in turn leads to

$$x_2^* GQG^* x_2 + x_1^* H R^{-1} H x_1 = 0,$$

so that $x_2^* GQ^{1/2} = 0$ and $R^{-1/2} H x_1 = 0$. Substituting into (E.7.25) we get $F^* x_2 = \lambda x_2$ and $F x_1 = \frac{1}{\lambda} x_1$. These equations show that $\{x_2, x_1\}$ are left and right eigenvectors for F that are orthogonal to $GQ^{1/2}$ and $R^{-1/2}H$, respectively, thus contradicting the detectability and unit-circle controllability assumptions.

(iv) The pencil

$$\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix} - \lambda \begin{bmatrix} F & 0 \\ H^* R^{-1} H & I \end{bmatrix} \tag{E.7.26}$$

is regular since its determinant cannot vanish for all λ (in particular, by the result of (iii), it does not vanish for any $|\lambda| = 1$). Hence, there are at most $2n$ generalized eigenvalues. Moreover, it is easy to see that if 0 is a generalized eigenvalue with multiplicity r for the above pencil, then it is also a generalized eigenvalue with multiplicity r for the corresponding conjugate pencil (E.7.24). In this case, the original pencil (E.7.26) is said to have r eigenvalues at ∞. From this we conclude that there are only $2n - 2r$ finite nonzero eigenvalues for the pencil (E.7.26). ◆

We thus see that the generalized eigenvalues of the pair of matrices (E.7.22) can be partitioned into the following form

$$\underbrace{0, \ldots, 0}_{r}, \quad \underbrace{\lambda_1, \ldots, \lambda_{n-r}}_{|\lambda_i|<1}, \quad \underbrace{\lambda_1^{-*}, \ldots, \lambda_{n-r}^{-*}}_{|\lambda_i^{-*}|>1}, \quad \underbrace{\infty, \ldots, \infty}_{r}.$$

Recall that in the nonsingular F case studied earlier, zero eigenvalues did not occur.

The following result establishes that the zero and nonzero generalized eigenvalues of the pair of matrices (E.7.22) is determined by the eigenvalues of the closed-loop matrix F_p.

Lemma E.7.4 (Solutions of DARE) *Let P denote any solution of the* DARE *(E.7.1), when it exists. Then P satisfies*

$$\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix} \begin{bmatrix} P \\ I \end{bmatrix} = \begin{bmatrix} F & 0 \\ H^* R^{-1} H & I \end{bmatrix} \begin{bmatrix} P \\ I \end{bmatrix} F_p^*,$$

where $F_p = F - K_p H$. That is, $\{P, I\}$ forms a basis for a generalized eigenspace of the pair of matrices (E.7.22), with eigenvalues given by those of F_p^*. ■

Proof: Let P be any solution of the DARE (E.7.1) and note that, from the definition of F_p, $F_p = F - K_p H$, we obtain

$$F P F_p^* = P - GQG^*. \tag{E.7.27}$$

It also holds that

$$F - F_p = F_p P H^* R^{-1} H, \tag{E.7.28}$$

as can be checked by verifying that the difference is zero. Combining (E.7.27) and (E.7.28) we obtain the desired result. ◆

The next theorem shows how the unique stabilizing solution can be determined from a basis for the generalized stable eigenspace of the pair of matrices (E.7.22).

Theorem E.7.2 (The Invariant Subspace Method) *Consider again the* DARE *(E.7.1) with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ unit-circle controllable. Let U and V be any $n \times n$ matrices that form a basis for the stable generalized eigenspace of the pair of matrices (E.7.22), viz.,*

$$\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} F & 0 \\ H^* R^{-1} H & I \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} \Lambda, \tag{E.7.29}$$

where Λ is an $n \times n$ matrix with all its eigenvalues inside the unit disc, $|\lambda_i(\Lambda)| < 1$. Then

(i) V is invertible.

(ii) UV^{-1} is the unique stabilizing solution of the DARE *(E.7.1).*

■

Proof: We proceed as follows:

(i) Assume V is singular. The detectability and unit-circle controllability assumptions on $\{F, H\}$ and $\{F, GQ^{1/2}\}$ guarantee that a unique stabilizing solution P to the DARE (E.7.1) exists. Using Lemma E.7.4, we get

$$\begin{bmatrix} I & -GQG^* \\ 0 & F^* \end{bmatrix} \begin{bmatrix} P \\ I \end{bmatrix} = \begin{bmatrix} F & 0 \\ H^*R^{-1}H & I \end{bmatrix} \begin{bmatrix} P \\ I \end{bmatrix} F_p^*,$$

with stable F_p, showing that $\{P, I\}$ also spans the generalized stable eigenspace of the pair of matrices (E.7.22) and, hence, the matrices F_p^* and Λ should have the same canonical Jordan form. Now since

$$\begin{bmatrix} P \\ I \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} U \\ V \end{bmatrix}$$

span the same invariant subspace, they should be related by an invertible transformation, T. Then $I = VT$ and we conclude that $V^{-1} = T^{-1}$.

(ii) From part (i) we have

$$\begin{bmatrix} P \\ I \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} T$$

for some invertible matrix T. Then $P = UT$ and $I = VT$ so that $P = UV^{-1}$. ◆

Remark. An alternative argument that establishes the invertibility of V in part (i) *without* relying on the existence of the stabilizing solution P is given in Prob. 14.17; it is also used further ahead in the proof of part (ii) of Thm. E.9.1 in the case of the CARE. Once this is done, the main results of Thms. E.5.1 and E.6.1 will follow rather immediately. We shall illustrate this alternative route while studing the CARE in Sec. E.9. The arguments in that section will rely exclusively on the results of Thm. E.9.1, which establish several properties of the stable eigenspace of a Hamiltonian matrix; the theorem is the counterpart of Thm. E.7.2 above for the DARE. ◆

Remark. Since

$$\left[\begin{bmatrix} I & 0 \\ 0 & F^* \end{bmatrix} \begin{bmatrix} 0 \\ H^* \end{bmatrix} \right] \left[\begin{bmatrix} F & GQG^* \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right]^{-1} = \begin{bmatrix} M & \begin{bmatrix} 0 \\ H^*R^{-1} \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \end{bmatrix} & 0 \end{bmatrix},$$

we could alternatively study the generalized eigenvalues and eigenvectors of the pair

$$\left\{ \begin{bmatrix} I & 0 & 0 \\ 0 & F^* & H^* \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} F & GQG^* & 0 \\ 0 & I & 0 \\ H & 0 & R \end{bmatrix} \right\} \tag{E.7.30}$$

to compute solutions of the DARE when both F and R are singular, as suggested by Van Dooren (1981) and Ionescu, Oara, and Weiss (1997). ◆

E.8 THE DUAL DARE

We end this appendix with a discussion of the dual DARE, which was shown in Sec. 14.5.3 to be relevant to the study of the convergence of the Riccati recursion for zero initial conditions and also for nonnegative-definite initial conditions.

The *dual* DARE that corresponds to (E.7.1) is defined as (assuming $S = 0$)

$$P^a = F^*P^aF + H^*R^{-1}H - F^*P^aGQ^{1/2}(I + Q^{*/2}G^*P^aGQ^{1/2})^{-1}Q^{*/2}G^*P^aF, \tag{E.8.1}$$

where $Q = Q^{1/2}Q^{*/2}$. When $Q > 0$, the above equation reduces to the equivalent form

$$P^a = F^*P^aF + H^*R^{-1}H - F^*P^aG(Q^{-1} + G^*P^aG)^{-1}G^*P^aF,$$

with Q^{-1}. We shall use (E.8.1) for generality.

Comparing with the original DARE (E.7.1), we see that we have replaced $\{F, G, H, R, Q\}$ in the original DARE by $\{F^*, H^*, GQ^{1/2}, I, R^{-1}\}$ in the dual DARE. Hence, all the statements we derived in the earlier sections for the DARE can be extended to the dual DARE. In particular, equation (E.8.1) will be said to have a stabilizing solution if a P^a satisfying (E.8.1) can be found such that the corresponding closed-loop matrix,

$$F^* - F^*P^aGQ^{1/2}(I + Q^{*/2}G^*P^aGQ^{1/2})^{-1}Q^{*/2}G^*,$$

is stable. Clearly, the existence of a stabilizing P^a is equivalent to the detectability of $\{F^*, Q^{*/2}G^*\}$ (the stabilizability of $\{F, GQ^{1/2}\}$) and the unit-circle controllability of $\{F^*, H^*R^{-*/2}\}$ (the unit-circle observability of $\{F, R^{-1/2}H\}$). More specifically, we have the following result.

Theorem E.8.1 (Stabilizing Solution to the Dual DARE) *The two statements below provide separate conditions for the existence of a stabilizing solution to the dual* DARE *(E.8.1).*

(i) A stabilizing solution P^a of (E.8.1) exists if, and only if, $\{F, GQ^{1/2}\}$ is stabilizable and $\{F, R^{-1/2}H\}$ is observable on the unit circle.

(ii) Assume that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the unit circle so that the DARE *(E.7.1) has a unique stabilizing solution P. Then the dual* DARE *(E.8.1) has a stabilizing solution P^a if, and only if, the matrix $I - P\mathcal{O}^p$ is nonsingular, where $\mathcal{O}^p$ is given by the unique solution of the Lyapunov equation*

$$\mathcal{O}^p = F_p^*\mathcal{O}^pF_p + H^*R_e^{-1}H.$$

Moreover, when the stabilizing solutions $\{P, P^a\}$ of the DARE *and the dual* DARE *exist (i.e., when $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is stabilizable), they are related via*

$$P^a = \mathcal{O}^p(I - P\mathcal{O}^p)^{-1}. \tag{E.8.2}$$

■

Proof: The statement (i) is immediate and follows from Thm. E.5.1. To prove (ii) we need to argue in both directions.

Assume first that $I-P\mathcal{O}^p$ is invertible. We established in the proof of Lemma 14.5.5 that this implies the stabilizability of $\{F, GQ^{1/2}\}$. Combining this condition with the assumed detectability of $\{F, H\}$ we conclude by statement (i) of the current theorem that a stabilizing solution P^a of the dual DARE should exist.

Conversely, assume now that a stabilizing solution P^a exists and let us prove that $(I - P\mathcal{O}^p)$ is invertible. Indeed, by (i), the existence of P^a implies $\{F, GQ^{1/2}\}$ is stabilizable. We now show that this stabilizability condition implies the invertibility of $(I - P\mathcal{O}^p)$.

The argument relies on establishing that the following relation holds

$$\begin{bmatrix} I & -H^*R^{-1}H \\ 0 & F \end{bmatrix}\begin{bmatrix} \mathcal{O}^p \\ I - P\mathcal{O}^p \end{bmatrix} = \begin{bmatrix} F^* & 0 \\ GQG^* & I \end{bmatrix}\begin{bmatrix} \mathcal{O}^p \\ I - P\mathcal{O}^p \end{bmatrix} F_p. \tag{E.8.3}$$

This corresponds to a generalized eigenvalue-eigenvector decomposition of the pair of matrices

$$\left\{\begin{bmatrix} I & -H^*R^{-1}H \\ 0 & F \end{bmatrix}, \quad \begin{bmatrix} F^* & 0 \\ GQG^* & I \end{bmatrix}\right\}, \tag{E.8.4}$$

with the columns of

$$\begin{bmatrix} \mathcal{O}^p \\ I - P\mathcal{O}^p \end{bmatrix}$$

spanning a stable generalized eigenspace (since F_p is a stable matrix by the detectability and unit-circle controllability assumptions). The matrices (E.8.4) play for the dual DARE (E.8.1) the same role as the matrices (E.7.22) play for the original DARE (E.7.1). Thus, by applying the result of part (i) of Thm. E.7.2 for the pair of matrices (E.8.4) in the dual DARE case, we will be able to conclude that $I - P\mathcal{O}^p$ must be invertible. Hence, all we need to do is establish (E.8.3), which follows from some straightforward algebra that we omit here.

Finally, the expression (E.8.2) for the solution of the dual DARE follows from the generalized eigendecomposition (E.8.3) and from part (ii) of Thm. E.7.2 when applied to the dual DARE. ◆

E.9 THE CARE

We now extend our discussions to the continuous-time algebraic Riccati equation (CARE),

$$0 = FP + PF^* + GQG^* - KRK^*, \quad K = PH^*R^{-1}, \quad R > 0 \tag{E.9.1}$$

where the assumption $S = 0$ is made for simplicity (otherwise, just replace F by $F^s = F - GSR^{-1}H$ and Q by $Q^s = Q - SR^{-1}S^*$).

The Hamiltonian Matrix.
With the CARE (E.9.1) we associate the so-called Hamiltonian matrix,

$$M \triangleq \begin{bmatrix} F & -GQG^* \\ -H^*R^{-1}H & -F^* \end{bmatrix}, \tag{E.9.2}$$

which can be easily seen to satisfy the relation

$$J^{-1}MJ = -M^*, \qquad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

This shows that M and $-M^*$ are similar matrices. The following result is now the counterpart of Lemma E.7.3.

Lemma E.9.1 (Eigenvalues of the Hamiltonian Matrix) *Let λ denote any eigenvalue of the Hamiltonian matrix M in (E.9.2), i.e., $Mx = \lambda x$ for some nonzero vector x. Then*

(i) $-\lambda^*$ is also an eigenvalue of M.

(ii) If $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the imaginary axis then $Re(\lambda) \neq 0$.

■

Proof: Part (i) follows immediately from the fact that M and $-M^*$ are similar matrices. We thus focus on establishing part (ii). So assume that $\mathrm{Re}(\lambda) = 0$ and partition x into $x = \mathrm{col}\{x_1, x_2\}$. Then $Mx = \lambda x$ implies that

$$Fx_1 - GQG^*x_2 = \lambda x_1, \qquad -H^*R^{-1}Hx_1 - F^*x_2 = \lambda x_2. \tag{E.9.3}$$

These relations can be combined together to yield

$$x_1^*HR^{-1}H^*x_1 + x_2^*GQG^*x_2 = 0,$$

so that $x_2^*GQ^{1/2} = 0$ and $R^{-1/2}Hx_1 = 0$. Substituting these results into (E.9.3) we get $F^*x_2 = -\lambda x_2$ and $Fx_1 = \lambda x_1$. These equations show that $\{x_1, x_2\}$ are right and left eigenvectors for F that are orthogonal to H and $GQ^{1/2}$, respectively, thus contradicting the assumptions of detectability and controllability on the imaginary axis. ◆

Now let P denote any solution of the CARE (E.9.1) and define the closed-loop matrix

$$F_{cl} \triangleq F - KH, \quad K \triangleq PH^*R^{-1}.$$

Then some straightforward algebra shows that

$$\begin{bmatrix} I & 0 \\ -P & I \end{bmatrix} M^* \begin{bmatrix} I & 0 \\ P & I \end{bmatrix} = \begin{bmatrix} F_{cl}^* & -H^*R^{-1}H \\ 0 & -F_{cl} \end{bmatrix}, \tag{E.9.4}$$

which establishes that the eigenvalues of the Hamiltonian matrix M in (E.9.2) are determined by the eigenvalues of the closed-loop matrix. More specifically, we have the following result.

Lemma E.9.2 (Solutions of CARE) *Let P denote any solution of the* CARE *(E.9.1), when it exists. Then P satisfies*

$$M^* \begin{bmatrix} I \\ P \end{bmatrix} = \begin{bmatrix} I \\ P \end{bmatrix} F_{cl}^*, \tag{E.9.5}$$

where M is given by (E.9.2). That is, $\{I, P\}$ forms a basis for an eigenspace of M^* with eigenvalues given by those of F_{cl}^*. Moreover, the eigenvalues of M^* coincide with those of $\{F_{cl}, -F_{cl}^*\}$. ∎

Now assume $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the imaginary axis. Then by Lemma E.9.1, the Hamiltonian matrix M^* will not have eigenvalues on the imaginary axis. Thus there exists an invertible similarity transformation T that reduces M to the form

$$M^* = T \begin{bmatrix} \Lambda & 0 \\ 0 & -\Lambda^* \end{bmatrix} T^{-1}, \tag{E.9.6}$$

where Λ is composed of Jordan blocks that correspond to the eigenvalues of M^* that are inside the open left-half plane. We further partition the entries of T into

$$T = \begin{bmatrix} V & T_{12} \\ U & T_{22} \end{bmatrix}.$$

The next theorem shows how the unique stabilizing solution P of the CARE (E.9.1) can be determined from a basis for the stable eigenspace of M^*. The result is the counterpart of Thm. E.7.2.

Theorem E.9.1 (The Invariant Subspace Method) *Consider again the* CARE *(E.9.1) with $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ controllable on the imaginary axis. Let U and V be any $n \times n$ matrices that form a basis for the stable eigenspace of the Hamiltonian matrix M^* in (E.9.2), viz.,*

$$\begin{bmatrix} F^* & -H^*R^{-1}H \\ -GQG^* & -F \end{bmatrix} \begin{bmatrix} V \\ U \end{bmatrix} = \begin{bmatrix} V \\ U \end{bmatrix} \Lambda, \tag{E.9.7}$$

where Λ is an $n \times n$ matrix with all its eigenvalues inside the open left-half plane. Then

(i) U^*V is Hermitian.

(ii) V is invertible.

(iii) UV^{-1} is Hermitian and satisfies the CARE *(E.9.1).*

(iv) UV^{-1} is nonnegative-definite.

(v) UV^{-1} is a stabilizing solution of the CARE *(E.9.1).* ∎

Proof: We proceed as follows:

(i) Using $J^{-1}MJ = -M^*$, $J^{-1} = -J$, and $J^* = -J$, we observe that $(JM^*)^* = JM^*$. That is, JM^* is Hermitian. Now pre-multiplying (E.9.7) by $[V^* \;\; U^*]J$ leads to

$$[V^* \;\; U^*] JM^* \begin{bmatrix} V \\ U \end{bmatrix} = [V^* \;\; U^*] J \begin{bmatrix} V \\ U \end{bmatrix} \Lambda.$$

Evaluating the conjugate transpose of both sides of the above equality and using the fact that JM^* is Hermitian, we conclude that we must have

$$\Lambda^* [V^* \;\; U^*] J^* \begin{bmatrix} V \\ U \end{bmatrix} = [V^* \;\; U^*] J \begin{bmatrix} V \\ U \end{bmatrix} \Lambda.$$

This shows that the difference $(U^*V - V^*U)$ satisfies the homogeneous Lyapunov equation

$$\Lambda^*(U^*V - V^*U) + (U^*V - VU^*)\Lambda = 0.$$

Now since Λ is a stable matrix (*i.e.*, all its eigenvalues are in the open left-half plane), we conclude that the unique solution of the above equation is $(U^*V - V^*U) = 0$ (*cf.* Lemma D.2.2). Therefore, $U^*V = V^*U$, as desired.

(ii) Assume V is singular, say $Vx = 0$ for some nonzero x. It then follows from (E.9.7) that

$$F^*V - H^*R^{-1}HU = V\Lambda, \qquad -GQG^*V - FU = U\Lambda. \tag{E.9.8}$$

Pre-multiplying the first equation by x^*U^* and post-multiplying it by x we obtain that

$$-x^*UH^*R^{-1}HUx = x^*U^*V\Lambda x = x^*V^*U\Lambda x = 0,$$

where we used the fact that $U^*V = V^*U$. We thus conclude that $HUx = 0$ since $R > 0$. Now post-multiplying the first equation again by x we conclude that $V\Lambda x = 0$. In other words, we showed that if $Vx = 0$, then $V\Lambda x = 0$. Iterating this argument, we conclude that $V\Lambda^k x = 0$ for $k \geq 1$ so that the pair $\{\Lambda, V\}$ is not observable. By the rank test of App. C, it follows that the matrix

$$\begin{bmatrix} \lambda I - \Lambda \\ V \end{bmatrix}$$

drops rank at some eigenvalue λ of Λ. This is equivalent to saying that there exists a nonzero vector y such that $\Lambda y = \lambda y$ and $Vy = 0$, with $\mathrm{Re}(\lambda) < 0$.

Using the second relation $-GQG^*V - FU = U\Lambda$ we then conclude that Uy is a right eigenvector of F, while the first relation $F^*V - H^*R^{-1}HU = V\Lambda$ allows us to conclude that $HUy = 0$. The two conditions $FUy = \lambda Uy$ and $HUy = 0$ violate the detectability of $\{F, H\}$.

[The converse is in fact also true and it follows from part (v) of this lemma. Once we show that UV^{-1} is a stabilizing solution, this would mean that $F - (UV^{-1})H^*R^{-1}H$ is stable and therefore the pair $\{F, H\}$ must be detectable since there exists a gain matrix, $K = (UV^{-1})H^*R^{-1}$, that results in a stable $F - KH$.]

(iii) The symmetry of $Y \triangleq UV^{-1}$ follows from part (i) since the equality $U^*V = V^*U$, and the invertibility of V, imply that $UV^{-1} = V^{-*}U^*$ so that $Y = Y^*$. Now using (E.9.8) we can easily verify that Y satisfies the equation

$$0 = FY + YF^* + GQG^* - YH^*R^{-1}HY,$$

which coincides with the CARE (E.9.1).

(iv) Using (E.9.8) again we can verify that Y satisfies the equation

$$0 = A^*Y + YA + GQG^* + YH^*R^{-1}HY,$$

where A is the stable matrix $A = V\Lambda V^{-1}$. Using Lemma D.2.2, the solution Y of the above equation can be expressed in the form

$$Y = \int_0^\infty e^{A^*t}[GQG^* + YH^*R^{-1}HY]e^{At}dt,$$

where the center matrix inside the integral is nonnegative definite in view of $Q \geq 0$ and $R > 0$. Therefore, $Y \geq 0$.

(v) Using the first relation in (E.9.8) we easily get that

$$F^* - H^*R^{-1}HUV^{-1} = V\Lambda V^{-1}.$$

This means that $(F - YH^*R^{-1}H)$ is similar to Λ^*, so that the closed-loop matrix is indeed stable. This means that Y is a stabilizing solution of the CARE (E.9.1). ◆

Main Results.

We are now in a position to establish two results that are the counterparts of Thms. E.5.1 and E.6.1.

Theorem E.9.2 (Algebraic Riccati Equation) *Consider the* CARE *(E.9.1). Then the following two statements are equivalent.*

(i) $\{F, H\}$ *is detectable and* $\{F, GQ^{1/2}\}$ *is controllable on the imaginary axis.*

(ii) The CARE *has a stabilizing solution* P*, i.e., one for which the matrix* $F - KH$ *is stable, where* $K = PH^*R^{-1}$.

Moreover, any such stabilizing solution is unique and positive-semi-definite. ■

Proof: We proceed as follows:

(i)⇒(ii). Assume $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the imaginary axis. Then by Thm. E.9.1 there exists a stabilizing solution P, which can be found by determining a basis $\{V, U\}$ for the stable eigenspace of the Hamiltonian matrix M^* and then taking $P = UV^{-1}$. The only fact that remains to be shown is that P is unique. Thus assume that two stabilizing solutions exist, say P_1 and P_2. Let $F_{cl,1}$ and $F_{cl,2}$ denote the corresponding closed-loop matrices. Since both P_1 and P_2 satisfy the CARE (E.9.1), some straightforward algebra shows that the difference $(P_1 - P_2)$ satisfies the Lyapunov equation

$$F_{cl,1}(P_1 - P_2) + (P_1 - P_2)F_{cl,2}^* = 0.$$

The stability of $\{F_{cl,1}, F_{cl,2}\}$ then guarantees (*cf.* Lemma D.2.2) that the only solution of this equation is $P_1 - P_2 = 0$ so that $P_1 = P_2$.

(ii)⇒(i). Now assume that a stabilizing solution P exists. Then $\{F, H\}$ is clearly detectable since $K = PH^*R^{-1}$ leads to a stable matrix $F - KH$. Therefore, assume that $\{F, H\}$ is detectable but that $\{F, GQ^{1/2}\}$ is not controllable on the imaginary axis. This implies that there exists a λ, with $\mathrm{Re}(\lambda) = 0$, and a nonzero vector x such that

$$xF = \lambda F, \quad xGQ^{1/2} = 0. \tag{E.9.9}$$

Premultiplying both sides of the CARE (E.9.1) by x, we conclude that

$$-\lambda xP = xPF_{cl}^*. \tag{E.9.10}$$

Now either of the following mutually exclusive cases may happen:

(i) $xP \neq 0$. In this case (E.9.10) implies that F_{cl} has a unit circle eigenvalue at $-\lambda$.

(ii) $xP = 0$. In this case

$$xF_{cl} = x(F - PH^*R^{-1}H) = xF = \lambda x,$$

so that F_{cl} again has a unit-circle eigenvalue at λ.

In either case, F_{cl} has a unit-circle eigenvalue, which contradicts the fact that P is stabilizing. ◆

The next result shows that in order to guarantee a *unique* positive semi-definite solution, we need the additional condition that $\{F, GQ^{1/2}\}$ be stabilizable, *i.e.*, that it be controllable on and to the left of the imaginary axis (and not just on the imaginary axis, which was the condition we required above for the existence of a stabilizing solution to the CARE).

Theorem E.9.3 (Unique Positive-Semi-Definite Solution) *Consider again the* CARE *(E.9.1) and assume that* $\{F, H\}$ *is detectable and* $\{F, GQ^{1/2}\}$ *is controllable on the imaginary axis. Then the following two statements are equivalent.*

(i) $\{F, GQ^{1/2}\}$ *is stabilizable.*

(ii) The CARE *has a unique positive-semi-definite solution.*

Moreover, the unique positive-semi-definite solution of the CARE *is given by its stabilizing solution.* ■

Proof: We again proceed in steps.

(i)⇒(ii): Suppose that the CARE has a positive-semi-definite solution, $P_1 \geq 0$, that is different from its stabilizing solution P. We shall show that $\{F, GQ^{1/2}\}$ is not stabilizable. To this end, we write the CARE (E.9.1) as

$$0 = F_{cl,1}P_1 + P_1F_{cl,1}^* + GQG^* + K_1RK_1^*, \quad K_1 = P_1H^*R^{-1},$$

where $F_{cl,1} = F - K_1H$. Now since $P_1 \neq P$, and since the stabilizing solution is unique (by Thm. E.9.2), the matrix $F_{cl,1}$ cannot be stable. Let λ (with $\mathrm{Re}(\lambda) \geq 0$) be an unstable mode of $F_{cl,1}$ with corresponding left eigenvector x. Pre- and post-multiplying the above CARE by x and x^*, yields

$$0 = \lambda xP_1x^* + \lambda^*xP_1x^* + xGQG^*x^* + xK_1RK_1^*x^*.$$

Since $\lambda + \lambda^* \geq 0$ and $xPx^* \geq 0$ we conclude that

$$xGQG^*x^* = 0 \quad \text{and} \quad xK_1RK_1^*x^* = 0.$$

But since $R > 0$, this implies $xK_1 = 0$. This now shows that $\{F, GQ^{1/2}\}$ is not stabilizable since $xGQ^{1/2} = 0$ and

$$xF = x(F_{cl,1} + K_1H) = xF_{cl,1} = \lambda x.$$

(ii)⇒(i): Suppose that $\{F, GQ^{1/2}\}$ is not stabilizable. We shall now show that the CARE has more than one positive-semi-definite solution. To this end, suppose that λ (with $\mathrm{Re}(\lambda) \geq 0$) is an unstabilizable mode of $\{F, GQ^{1/2}\}$ with corresponding left eigenvector x,

$$xF = \lambda x, \quad xGQ^{1/2} = 0.$$

Consider the stabilizing solution P of the CARE,

$$0 = FP + PF^* + GQG^* - PH^*R^{-1}HP.$$

Pre-multiplying the CARE by x yields

$$\lambda xP + xPF^* - xPH^*R^{-1}P = 0, \tag{E.9.11}$$

which is equivalent to $\lambda xP = -xPF_{cl}^*$. We should note that $xP \neq 0$ since otherwise $xF_{cl} = xF = \lambda x$, showing that F_{cl} has an unstable eigenvalue, a contradiction.

We can therefore construct the following positive-semi-definite matrix (unequal to P since $xP \neq 0$):

$$P_1 = P - \frac{Px^*xP}{xPx^*} \geq 0. \tag{E.9.12}$$

Some simple algebra, using (E.9.11), then shows that the quantity

$$FP_1 + P_1F^* + GQG^* - P_1H^*R^{-1}HP_1,$$

evaluates to zero. This shows that P_1 satisfies the CARE as well. ◆

The Dual CARE.

The *dual* CARE that corresponds to (E.9.1) is defined as

$$0 = F^*P^a + P^aF + H^*R^{-1}H - P^aGQG^*P^a. \tag{E.9.13}$$

This equation was encountered in Sec. 16.7 while studying the asymptotic behavior of the Riccati differential equation.

Comparing with the original CARE (E.9.1), we see that we have replaced the parameters $\{F, G, H, R^{-1}, Q\}$ in the original CARE by $\{F^*, H^*, G^*, Q, R^{-1}\}$ in the dual CARE. Hence, all the statements we derived in the earlier sections for the CARE can be extended to the dual CARE. In particular, equation (E.9.13) will be said to have a stabilizing solution if a P^a satisfying (E.9.13) can be found such that the corresponding closed-loop matrix,

$$F_{cl}^a \triangleq F^* - P^aGQG^*,$$

is stable. Clearly, the existence of a stabilizing P^a is equivalent to the detectability of $\{F^*, Q^{*/2}G^*\}$ (the stabilizability of $\{F, GQ^{1/2}\}$) and the controllability of $\{F^*, H^*R^{-*/2}\}$ on the imaginary axis (the observability of $\{F, R^{-1/2}H\}$ on the imaginary axis). More specifically, we have the following result, which is the counterpart of Thm. E.8.1.

Theorem E.9.4 (Stabilizing Solution to the Dual CARE) *The two statements below provide separate conditions for the existence of a stabilizing solution to the dual* CARE *(E.9.13).*

(i) A stabilizing solution P^a of (E.9.13) exists if, and only if, $\{F, GQ^{1/2}\}$ is stabilizable and $\{F, R^{-1/2}H\}$ is observable on the imaginary axis.

(ii) Assume that $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is controllable on the imaginary axis so that the CARE *(E.9.1) has a unique stabilizing solution P. Then the dual* CARE *(E.9.13) has a stabilizing solution P^a if, and only if, the matrix $I - P\mathcal{O}$ is nonsingular, where $\mathcal{O}$ is given by the unique solution of the Lyapunov equation*

$$\mathcal{O}F_{cl} + F_{cl}^*\mathcal{O} + H^*R^{-1}H = 0.$$

Moreover, when the stabilizing solutions $\{P, P^a\}$ of the CARE *and the dual* CARE *exist (i.e., when $\{F, H\}$ is detectable and $\{F, GQ^{1/2}\}$ is stabilizable), they are related via*

$$P^a = \mathcal{O}(I - P\mathcal{O})^{-1}. \tag{E.9.14}$$

■

Proof: Consider the Hamiltonian matrix that is associated with the dual CARE,

$$M^a \triangleq \begin{bmatrix} F^* & -H^*R^{-1}H \\ -GQG^* & -F \end{bmatrix} = M^*.$$

Some simple algebra shows that

$$M^{a*}\begin{bmatrix} I - P\mathcal{O} & -P \\ \mathcal{O} & I \end{bmatrix} = \begin{bmatrix} I - P\mathcal{O} & -P \\ \mathcal{O} & I \end{bmatrix}\begin{bmatrix} F_{cl} & 0 \\ 0 & -F_{cl}^* \end{bmatrix}.$$

The matrix

$$\begin{bmatrix} I - P\mathcal{O} & -P \\ \mathcal{O} & I \end{bmatrix}$$

is invertible since its (2, 2) entry is the identity and the Schur complement with respect to it is also the identity matrix. Hence, the above equality is a similarity transformation and it shows that the eigenvalues of M^{a*} are given by those of $\{F_{cl}, -F_{cl}^*\}$. Since F_{cl} is stable, we conclude that $\{I - P\mathcal{O}, \mathcal{O}\}$ is a basis for the stable eigenspace of M^{a*}.

Assume first that $(I - P\mathcal{O})$ is invertible. Then by part (v) of Thm. E.9.1, the matrix $P^a = \mathcal{O}(I - P\mathcal{O})^{-1}$ is a stabilizing solution of the dual CARE. Conversely, assume that a stabilizing solution P^a exists. Then by (E.9.5),

$$M^{a*}\begin{bmatrix} I \\ P^a \end{bmatrix} = \begin{bmatrix} I \\ P^a \end{bmatrix}F_{cl}^{a*}.$$

This shows that $\{I, P^a\}$ is also a basis for the stable eigenspace of M^{a*} (and, hence, that the eigenvalues of F_{cl}^{a*} and F_{cl} should coincide). Therefore, there must exist an invertible transformation T such that

$$\begin{bmatrix} I - P\mathcal{O} \\ \mathcal{O} \end{bmatrix} = \begin{bmatrix} I \\ P^a \end{bmatrix} T,$$

which shows that $(I - P\mathcal{O}) = T$. Thus $(I - P\mathcal{O})$ is nonsingular. ◆

E.10 COMPLEMENTS

The literature on the algebraic Riccati equation is vast (and growing), so we confine ourselves to mentioning the monograph of Lancaster and Rodman (1995), the volume edited by Bittanti, Laub, and Willems (1991), and the reprint volume edited by Patel, Laub, and Van Dooren (1994), where the reader may find additional works and references. The survey papers of Willems (1971) , Singer and Hammarling (1983), Dorato (1983), Shayman (1983), Lancaster and Rodman (1991), Kučera (1991), Laub (1991), as well as the chapter on the Riccati equation in the book of Saberi, Sannuti, and Chen (1995), are also of interest for various reasons. The work by Pappas, Laub, and Sandell (1980) seems to be the original reference on the use of matrix pencils for the study of the DARE, which we elaborated in Sec. E.7. Further discussion on the CARE can also be found in Kučera (1972), Molinari (1977), and Callier, Winkin, and Willems (1994). In this appendix we studied the positive case of the DARE and CARE. Similar results in the indefinite case can be found in some of the problems (*e.g.*, Prob. 14.16) and in Ionescu and Weiss (1993), Kwakernaak and Sebek (1994), and in the monograph Hassibi, Sayed, and Kailath (1999).

APPENDIX F

Displacement Structure

As mentioned at various points in the text, the concept of displacement structure is important in obtaining efficient and insightful solutions of several problems. We briefly introduce the topic here, mainly to show how the $O(Nn^3)$ CKMS recursions of Chs. 11 and 13 can be obtained by combining displacement structure with state-space structure (*cf.* App. 13.A). The results will also justify the claim in Ch. 4 that $N \times N$ Toeplitz matrices can be factored with $O(N^2)$ flops. There is much more to be said about displacement structure, but here we only refer to Kailath and Sayed (1999) and the references therein.

F.1 MOTIVATION

Many problems in engineering and applied mathematics ultimately require the solution of $n \times n$ linear systems of equations. For small-size problems, there is often not much else to do except to use one of the already standard methods of solution such as Gaussian elimination. However, in many applications, n can be very large ($n \sim 1000, n \sim 1{,}000{,}000$) and, moreover, the linear equations may have to be solved over and over again, with different problem/model parameters, until a satisfactory solution to the original physical problem is obtained. In such cases, the $O(n^3)$ burden, *i.e.*, the number of flops required to solve an $n \times n$ linear system of equations, can become prohibitively large. This is one reason why one seeks in various classes of applications to identify special/characteristic matrix structures that may be assumed in order to reduce the computational burden.

The most obvious matrix structures are those that involve explicit patterns among the matrix entries such as Toeplitz, Hankel, Vandermonde, Cauchy, and Pick matrices. Several fast algorithms have been devised over the years to exploit these special structures. However, even more common than these explicit matrix structures, are matrices in which the structure is implicit. For example, in certain least-squares problems one often encounters products of Toeplitz matrices; these products are not generally Toeplitz, but on the other hand, they are not "unstructured." Similarly, in probabilistic calculations the matrix of interest is often not a Toeplitz matrix, but rather its inverse, which is rarely Toeplitz itself, but of course is not unstructured: its inverse is Toeplitz. It

is well known that $O(n^2)$ flops suffice to solve linear systems of equations with an $n \times n$ Toeplitz coefficient matrix; a question is whether one will need $O(n^3)$ flops to invert a non-Toeplitz coefficient matrix whose inverse is known to be Toeplitz. When pressed, one's response clearly must be that it is conceivable that $O(n^2)$ flops will suffice, and this is in fact true.

Such problems suggest the need for a quantitative way of defining and identifying structure in matrices. Over the years, starting with Kailath, Kung, and Morf (1979), it was found that an elegant and useful way to do so is the concept of *displacement structure*. This concept has also been useful for a host of problems apparently far removed from the solution of linear equations, such as the study of constrained and unconstrained rational interpolation, maximum entropy extension, signal detection, digital filter design, nonlinear Riccati differential equations, inverse scattering, certain Fredholm integral equations (see Kailath and Sayed (1995,1999) and the many references therein).

For motivation, consider an $n \times n$ Hermitian Toeplitz matrix,

$$T = [c_{i-j}]_{i,j=0}^{n-1}, \qquad c_i = c_{-i}^*.$$

Since such matrices are completely specified by n entries rather than n^2, one would of course expect a reduction in computational effort for handling problems involving such matrices. However, exploiting the Toeplitz structure is apparently more difficult than it may at first seem. To see this, consider the simple case of a real symmetric 3×3 Toeplitz matrix and apply the first step of the Gaussian elimination procedure to it, namely

$$\begin{bmatrix} c_0 & c_1 & c_2 \\ c_1 & c_0 & c_1 \\ c_2 & c_1 & c_0 \end{bmatrix} - \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} c_0^{-1} \begin{bmatrix} c_0 & c_1 & c_2 \end{bmatrix} = \begin{bmatrix} 0 & \\ & \Delta \end{bmatrix}$$

where the so-called Schur complement matrix Δ is seen to be

$$\Delta = \frac{1}{c_0} \begin{bmatrix} c_0^2 - c_1^2 & c_1 c_0 - c_1 c_2 \\ c_1 c_0 - c_1 c_2 & c_0^2 - c_2^2 \end{bmatrix}.$$

However, Δ is no longer Toeplitz, so the special structure is lost in the very first step of the procedure. The fact is that what is preserved is not the Toeplitz structure, but a deeper notion called "displacement structure."

There are several forms of displacement structure, the earliest of which is the following. Consider an $n \times n$ Hermitian matrix R and the $n \times n$ lower triangular shift matrix Z with ones on the first subdiagonal and zeros elsewhere (*i.e.*, a lower triangular Jordan block with eigenvalue 0). The displacement of R with respect to Z is denoted by $\nabla_Z R$ and defined as the difference

$$\nabla_Z R \triangleq R - ZRZ^*. \tag{F.1.1}$$

The matrix R is said to have displacement structure (or low displacement rank) with respect to Z if the rank of $\nabla_Z R$ is considerably lower than (and independent of) n. For example, a Hermitian Toeplitz matrix has displacement rank 2 with respect to Z.

More generally, let $r \ll n$ denote the rank of $\nabla_Z R$. Then one can write $\nabla_Z R$ as $\nabla_Z R = GJG^*$, where G is an $n \times r$ matrix and J is a signature matrix of the form $J = (I_p \oplus -I_q)$ with $p + q = r$. This representation is highly nonunique since G can be replaced by $G\Theta$ for any J-unitary matrix Θ, *i.e.*, for any Θ such that $\Theta J \Theta^* = J$; this flexibility is actually very useful. The matrix G is said to be a generator matrix for R since, along with $\{Z, J\}$, it completely identifies R. If one labels the columns of G as

$$G = \begin{bmatrix} x_0 & \dots & x_{p-1} & y_0 & \dots & y_{q-1} \end{bmatrix},$$

and lets $L(x)$ denote a lower triangular Toeplitz matrix whose first column is x, then it can be seen that the unique R that solves the displacement equation $\nabla_Z R = GJG^*$ is given by

$$R = \sum_{i=0}^{n-1} Z^i GJG^* Z^{*i} = \sum_{i=0}^{p-1} L(x_i) L^*(x_i) - \sum_{i=0}^{q-1} L(y_i) L^*(y_i). \tag{F.1.2}$$

Such displacement representations of R as a combination of products of lower and upper triangular Toeplitz matrices allow, for example, bilinear forms such as $x^* R y$ to be rapidly evaluated via convolutions (and hence FFTs).

F.2 TWO FUNDAMENTAL PROPERTIES

As mentioned above, a general Toeplitz matrix has displacement rank 2, with in fact $p = 1$ and $q = 1$, as interested readers may wish to check. But there are interesting non-Toeplitz matrices with $p = 1 = q$, *e.g.*, the inverse of a Toeplitz matrix. In fact, this is a special case of the following fundamental result.

Lemma F.2.1 Structure of the Inverse *If R is an invertible matrix that satisfies*

$$R - ZRZ^* = GJG^*, \tag{F.2.1}$$

for some $n \times r$ full rank matrix G, then there must exist an $r \times n$ full rank matrix H such that

$$R^{-1} - Z^* R^{-1} Z = H^* J H. \tag{F.2.2}$$

■

Proof: The block matrix

$$\begin{bmatrix} R & Z \\ Z^* & R^{-1} \end{bmatrix}$$

admits the following block triangular decompositions (*cf.* App. A):

$$\begin{bmatrix} R & Z \\ Z^* & R^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ Z^* R^{-1} & I \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & R^{-1} - Z^* R^{-1} Z \end{bmatrix} \begin{bmatrix} I & 0 \\ Z^* R^{-1} & I \end{bmatrix}^*$$

$$= \begin{bmatrix} I & ZR \\ 0 & I \end{bmatrix} \begin{bmatrix} R - ZRZ^* & 0 \\ 0 & R^{-1} \end{bmatrix} \begin{bmatrix} I & ZR \\ 0 & I \end{bmatrix}^*.$$

Now Sylvester's law of inertia (see also App. A) implies that

$$\text{Inertia}\{R^{-1} - Z^*R^{-1}Z\} = \text{Inertia}\{R - ZRZ^*\}.$$

It follows from (F.2.1) that there must exist a full rank $r \times n$ matrix H such that (F.2.2) is valid. ◆

Observe that the proof of the above result actually holds with Z replaced by any matrix F. Therefore, the conclusion also applies to matrices R that satisfy displacement equations of the form

$$R - FRF^* = GJG^*, \tag{F.2.3}$$

for some $\{F, G, J\}$.

Now a central feature in many of the applications mentioned earlier is the ability to efficiently obtain the so-called triangular LDU factorization of a matrix and of its inverse. Among purely matrix computations, one may mention that this enables fast determination of the so-called QR decomposition and the factorization of composite matrices such as $T_1T_2^{-1}T_3$, T_i being Toeplitz. The LDU factorization of structured matrices is facilitated by the fact that Schur complements inherit displacement structure.

Lemma F.2.2 Structure of Schur Complements *Assume that F is a block lower triangular,*

$$F = \begin{bmatrix} F_1 & 0 \\ F_2 & F_3 \end{bmatrix},$$

and partition R accordingly with F as

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}.$$

Assume that R_{11} is invertible, and introduce the Schur complement $\Delta = R_{22} - R_{21}R_{11}^{-1}R_{12}$. Then it holds that

$$\textit{rank } (\Delta - F_3\Delta F_3^*) \leq \textit{rank } (R - FRF^*). \tag{F.2.4}$$

■

Proof: Note that, by Lemma F.2.1 with F replacing Z, we have

$$\text{rank } (R^{-1} - F^*R^{-1}F) = \text{rank } (R - FRF^*).$$

We now invoke a block matrix formula for R^{-1} (*cf.* App. A),

$$R^{-1} = \begin{bmatrix} R_{11}^{-1} + E\Delta^{-1}P & -E\Delta^{-1} \\ \Delta^{-1}P & \Delta^{-1} \end{bmatrix}, \quad E = R_{11}^{-1}R_{12}, \quad P = R_{21}R_{11}^{-1},$$

and observe that Δ^{-1} is a submatrix of R^{-1}. Hence,

$$\text{rank } (\Delta^{-1} - F_3^*\Delta^{-1}F_3) \leq \text{rank } (R^{-1} - F^*R^{-1}F).$$

But, by Lemma F.2.1 again,

$$\text{rank } (\Delta - F_3\Delta F_3^*) = \text{rank } (\Delta^{-1} - F_3^*\Delta^{-1}F_3),$$

so that rank $(\Delta - F_3\Delta F_3^*) \leq$ rank $(R - FRF^*)$. ◆

By properly defining the submatrices $\{R_{ij}\}$ and $\{F_i\}$, the above result allows us to find the displacement structure of various composite matrices. For example, choosing $R_{11} = -T$, $R_{12} = I = R_{21}$, $R_{22} = 0$, and $F = Z \oplus Z$ gives the previously mentioned result on inverses of structured matrices.

F.3 A GENERALIZED SCHUR ALGORITHM

Computations on structured matrices are made efficient by working not with the n^2 entries of R, but with the nr entries of the generator matrix G. The basic triangular factorization algorithm is Gaussian elimination, which as noted in the first calculation, amounts to finding a sequence of Schur complements. Incorporating structure into the Gaussian elimination procedure was, in retrospect, first done (in the special case $r = 2$) by Schur himself in a remarkable 1917 paper (Schur (1917)) dealing with the apparently very different problem of checking when a power series is bounded in the unit disc.[1]

The algorithm below is one generalization of Schur's original recursion (which we describe in Sec. F.4). It provides an efficient $O(rn^2)$ procedure for the computation of the triangular factors of a Hermitian *positive-definite* matrix R satisfying

$$R - ZRZ^* = GJG^*, \quad G \in C^{n\times r}, \tag{F.3.1}$$

with $J = (I_p \oplus -I_q)$. So let $R = L\,D^{-1}\,L^*$ denote the triangular decomposition of R, where $D = \text{diag}\{d_0, d_1, \ldots, d_{n-1}\}$, and the lower triangular factor L is normalized in such a way that the $\{d_i\}$ appear on its main diagonal. The nonzero part of the consecutive columns of L will be further denoted by l_i. Then it holds that the successive Schur complements of R with respect to its leading $i \times i$ submatrices, denoted by R_i, are also structured and satisfy

$$R_i - Z_iR_iZ_i^* = G_iJG_i^*, \tag{F.3.2}$$

with $(n-i)\times(n-i)$ lower triangular shift matrices Z_i, and where the generator matrices G_i are obtained by the following recursive construction: start with $G_0 = G$ and repeat for $i \geq 0$:

$$\begin{bmatrix} 0 \\ G_{i+1} \end{bmatrix} = \left\{ G_i + Z_iG_i\frac{Jg_i^*g_i}{g_iJg_i^*} \right\} \Theta_i, \tag{F.3.3}$$

where Θ_i is an arbitrary J-unitary matrix, and g_i denotes the top row of G_i. Then

$$l_i = G_iJg_i^*, \qquad d_i = g_iJg_i^*. \tag{F.3.4}$$

The degree of freedom in choosing Θ_i is often very useful. One particular choice leads to the so-called proper form of the generator recursion. Let Θ_i reduce g_i to the form

$$g_i\Theta_i = \begin{bmatrix} \delta_i & 0 & \ldots & 0 \end{bmatrix},$$

[1] We are happy to acknowledge that it was P. Dewilde who introduced us to the work of Schur, which has led to many further collaborations, starting with Dewilde, Vieira, and Kailath (1978).

with a nonzero scalar entry in the leading position. Then

$$\begin{bmatrix} 0 \\ G_{i+1} \end{bmatrix} = Z_i G_i \Theta_i \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + G_i \Theta_i \begin{bmatrix} 0 & 0 \\ 0 & I_{p+q-1} \end{bmatrix}, \tag{F.3.5}$$

with

$$l_i = \delta_i^* G_i \Theta_i \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad d_i = |\delta_i|^2. \tag{F.3.6}$$

In words, this shows that G_{i+1} can be obtained from G_i as follows:

1. Reduce g_i to proper form.
2. Multiply G_i by Θ_i and keep the last columns of $G_i\Theta_i$ unaltered;
3. Shift down the first column of $G_i\Theta_i$ by one position;

Shematically one has the following array picture (for $r = 3$):

$$G_i = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \vdots & \vdots & \vdots \end{bmatrix} \xrightarrow{\Theta_i} \begin{bmatrix} \times' & 0 & 0 \\ \times' & \times' & \times' \\ \times' & \times' & \times' \\ \vdots & \vdots & \end{bmatrix}$$

$$\xrightarrow{\text{shift first column}} \begin{bmatrix} 0 & 0 & 0 \\ \times'' & \times' & \times' \\ \times'' & \times' & \times' \\ \vdots & \vdots & \end{bmatrix} \triangleq \begin{bmatrix} [0\ 0\ 0] \\ G_{i+1} \end{bmatrix}.$$

Algebraic Derivation.
We provide here one derivation for the proper form (F.3.5) of the algorithm. There are of course many other derivations of the above algorithm, and of generalizations of it to other kinds of matrix structures and also to matrices that are not necessarily Hermitian or even positive-definite. The algorithm can also be extended to provide simultaneous factorizations of both a matrix R and its inverse, R^{-1}. For more details on these variations see, *e.g.*, Lev-Ari and Kailath (1986) and Kailath and Sayed (1995,1999),

Now starting with (F.3.1), let g_0 denote the top row of G. The first diagonal entry of R is then given by $d_0 = g_0 J g_0^*$ and is a positive real number. This guarantees the existence of a J-unitary rotation matrix Θ_0 that reduces g_0 to the form (*cf.* Lemma A.5.1)

$$g_0\Theta_0 = [\delta_0\ 0 \dots\ 0],$$

with a single nonzero entry δ_0 in the first position of the post-array (or, more generally, in any of the first p positions of the post-array). The number δ_0 satisfies $|\delta_0|^2 = d_0$ and

the rotation Θ_0 can be implemented in many different ways, as explained in App. B. By applying Θ_0 to the original matrix G, we obtain a post-array $\bar{G}_0$ of the form

$$G\Theta_0 = \bar{G}_0 = \begin{bmatrix} \delta_0 & 0 & \dots & 0 \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \vdots & & \vdots & \\ \times & \times & \times & \times \end{bmatrix},$$

where the top row of $\bar{G}_0$ has a single nonzero entry. We say that $\bar{G}_0$ is in proper form. Note that $\bar{G}_0$ is also a generator for R since

$$GJG^* = G\Theta_0 J \Theta_0^* G^* = \bar{G}_0 J \bar{G}_0^*.$$

Let us denote the columns of $\bar{G}_0$ by

$$\bar{G}_0 = [\bar{x}_0\ \dots\ \bar{x}_{p-1}\ \bar{y}_0\ \dots\ \bar{y}_{q-1}].$$

One consequence of properness is that from the displacement equation

$$R - ZRZ^* = GJG^* = \bar{G}_0 J \bar{G}_0^*$$

we can conclude that l_0, the first column of R, is given by $l_0 = \delta_0^* \bar{x}_0$. Now we are ready to explore the displacement structure of the matrix

$$\widetilde{R}_1 \triangleq R - l_0 d_0^{-1} l_0^* = \begin{bmatrix} 0 & 0 \\ 0 & R_1 \end{bmatrix},$$

which is formed by extending the Schur complement matrix, R_1, by one zero row and by one zero column. It then follows that

$$\begin{aligned} \widetilde{R}_1 - Z\widetilde{R}_1 Z^* &= R - l_0 d_0^{-1} l_0^* - Z(R_0 - l_0 d_0^{-1} l_0^*) Z^* \\ &= \bar{G}_0 J \bar{G}_0^* - l_0 d_0^{-1} l_0^* + Z l_0 d_0^{-1} l_0^* Z^* \\ &= \sum_{i=0}^{p-1} \bar{x}_i \bar{x}_i^* - \sum_{i=0}^{q-1} \bar{y}_i \bar{y}_i^* - \bar{x}_0 \bar{x}_0^* + Z \bar{x}_0 \bar{x}_0^* Z^*, \\ &\triangleq \begin{bmatrix} 0 \\ G_1 \end{bmatrix} J \begin{bmatrix} 0 \\ G_1 \end{bmatrix}^*, \text{ say,} \end{aligned}$$

where we introduced the matrix G_1 defined by

$$\begin{bmatrix} 0 \\ G_1 \end{bmatrix} = [Z\bar{x}_0\ \dots\ \bar{x}_{p-1}\ \bar{y}_0\ \dots\ \bar{y}_{q-1}],$$

and where we used the fact that the top row of the right-hand side matrix is zero. Now from the definition of $\widetilde{R}_1$ above we conclude that

$$R_1 - Z_1 R_1 Z_1^* = G_1 J G_1^*,$$

so that $\{G_1, J\}$ is a generator for the Schur complement R_1. Note further that G_1 is obtained directly from G via the matrix operation

$$\begin{bmatrix} 0 \\ G_1 \end{bmatrix} = ZG\Theta_0 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + G\Theta_0 \begin{bmatrix} 0 & 0 \\ 0 & I_{p+q-1} \end{bmatrix},$$

which is the first step of (F.3.5). We can repeat the argument for G_2, G_3, and so on.

F.4 THE CLASSICAL SCHUR ALGORITHM

Let us now verify that the above algorithm collapses to Schur's original recursion in the special case of matrices with displacement rank 2, Thus let $\mathcal{R}$ denote a semi-infinite structured matrix with a semi-infinite generator matrix $\mathcal{G}$, say

$$\mathcal{R} - \mathcal{Z}\mathcal{R}\mathcal{Z}^* = \mathcal{G}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\mathcal{G}^*, \tag{F.4.1}$$

where $\mathcal{Z}$ denotes the semi-infinite lower triangular shift matrix. We further denote the individual entries of $\mathcal{G}$ by

$$\mathcal{G} = \begin{bmatrix} x_{00} & y_{00} \\ x_{10} & y_{10} \\ x_{20} & y_{20} \\ x_{30} & y_{30} \\ \vdots & \vdots \end{bmatrix},$$

and introduce the power series

$$x_0(z) = x_{00} + x_{10}z + x_{20}z^2 + \cdots$$
$$y_0(z) = y_{00} + y_{10}z + y_{20}z^2 + \cdots$$

[We assume that $\mathcal{R}$ and $\mathcal{G}$ are such that these power series are well defined in some region of the complex plane.] These series are obtained from $\mathcal{G}$ via

$$\begin{bmatrix} x_0(z) & y_0(z) \end{bmatrix} = \begin{bmatrix} 1 & z & z^2 & z^3 & \ldots \end{bmatrix} \mathcal{G}.$$

We now determine an equivalent representation of the generator recursion (F.3.5) in function form. For this purpose, we start by noting that the first step of (F.3.5) takes the form

$$\begin{bmatrix} 0 \\ \mathcal{G}_1 \end{bmatrix} = \mathcal{G}_0\,\Theta_0 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \mathcal{Z}\,\mathcal{G}_0\,\Theta_0 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{F.4.2}$$

where we also denote the individual entries of the resulting $\mathcal{G}_1$ by

$$\begin{bmatrix} 0 \\ \mathcal{G}_1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ x_{11} & y_{11} \\ x_{21} & y_{21} \\ x_{31} & y_{31} \\ \vdots & \vdots \end{bmatrix}.$$

If we further associate with $\mathcal{G}_1$ the two power series,

$$x_1(z) = x_{11} + x_{21}z + x_{31}z^2 + \ldots$$
$$y_1(z) = y_{11} + y_{21}z + y_{31}z^2 + \ldots,$$

then, by multiplying both sides of (F.4.2) by $[1\ z\ z^2\ z^3\ \ldots]$ from the left, and by noting that

$$\begin{bmatrix} 1 & z & z^2 & z^3 & \ldots \end{bmatrix}\mathcal{Z} = z\begin{bmatrix} 1 & z & z^2 & z^3 & \ldots \end{bmatrix},$$

we conclude that the series $\{x_0, (z), y_0(z), x_1(z), y_1(z)\}$ are related as

$$\begin{bmatrix} zx_1(z) & zy_1(z) \end{bmatrix} = \begin{bmatrix} x_0(z) & y_0(z) \end{bmatrix} \Theta_0 \begin{bmatrix} z & 0 \\ 0 & 1 \end{bmatrix}.$$

Now recall that the purpose of the $J-$unitary rotation Θ_0 is to annihilate y_{00}. This can be achieved by considering a hyperbolic rotation of the form (*cf.* the discussion in Sec. B.5)

$$\Theta_0 = \frac{1}{\sqrt{1-|\gamma_0|^2}} \begin{bmatrix} 1 & -\gamma_0 \\ -\gamma_0^* & 1 \end{bmatrix}, \qquad \gamma_0 = \frac{y_{00}}{x_{00}} = \frac{y_0(0)}{x_0(0)}.$$

[The positive-definiteness of $\mathcal{R}$ guarantees $|x_{00}|^2 - |y_{00}|^2 > 0$ and, hence, x_{00} cannot be zero.]

This completes the first step of (F.3.5). The recursive procedure now continues as follows: compute γ_1 as the ratio of y_{11} and x_{11}, multiply the pre-array $\mathcal{G}_1$ by Θ_1 in order to introduce a zero in the first entry of the second column of the post-array, shift down the first column of the post-array, and so on. In function form, for the i-th step, we have

$$z\begin{bmatrix} x_{i+1}(z) & y_{i+1}(z) \end{bmatrix} = \begin{bmatrix} x_i(z) & y_i(z) \end{bmatrix} \Theta_i \begin{bmatrix} z & 0 \\ 0 & 1 \end{bmatrix}, \tag{F.4.3}$$

where Θ_i is an elementary hyperbolic rotation, as above, that is determined by a coefficient $\gamma_i = y_i(0)/x_i(0)$. Now define the function

$$s_i(z) \triangleq \frac{y_i(z)}{x_i(z)}.$$

It then follows easily from (F.4.3) that $s_i(z)$ satisfies the recursion:

$$s_{i+1}(z) = \frac{1}{z}\,\frac{s_i(z) - \gamma_i}{1 - \gamma_i^* s_i(z)}, \qquad \gamma_i = s_i(0). \tag{F.4.4}$$

This is Schur's original recursion (Schur (1917)), which was in fact derived by him while studying power series that are analytic and bounded by unity in the unit disc.

F.5 COMBINING DISPLACEMENT AND STATE-SPACE STRUCTURES

We conclude this appendix by mentioning that displacement structure can also be combined with state-space structure. Recall that we showed in App. 9.A that the Kalman filter can be obtained by applying the modified Gram-Schmidt procedure to the Gramian matrix R_y of the observations. We then showed in Sec. 11.5 that R_y had displacement structure of the form

$$\nabla_{Z^p} R_y = R_y - Z^p R_y [Z^p]^* = \mathcal{G}\mathcal{J}\mathcal{G}^*,$$

where

$$\mathcal{J} = \begin{bmatrix} I_p & 0 \\ 0 & J \end{bmatrix} \quad \text{and} \quad \mathcal{G} = \begin{bmatrix} R_{e,0}^{1/2} & 0 \\ H\bar{K}_{p,0} & H\bar{L}_0 \\ HF\bar{K}_{p,0} & HF\bar{L}_0 \\ \vdots & \vdots \end{bmatrix},$$

with $\{\bar{K}_{p,0}, \bar{L}_0, J\}$ defined via

$$\bar{K}_{p,0} = K_0 R_{e,0}^{-*/2}, \qquad P_1 - P_0 = \bar{L}_0 J \bar{L}_0^*.$$

Here, J is an $\alpha \times \alpha$ signature matrix and $\bar{L}_0$ is $n \times \alpha$. By applying the generalized Schur algorithm to $\{\mathcal{G}, \mathcal{J}\}$, we showed in App. 13.A that it collapses to the array form (13.2.6) of the CKMS recursions, *viz.*,

$$\begin{bmatrix} R_{e,i}^{1/2} & H\bar{L}_i \\ \bar{K}_{p,i} & FL_i \end{bmatrix} \Theta_{i+1} = \begin{bmatrix} R_{e,i+1}^{1/2} & 0 \\ \bar{K}_{p,i+1} & \bar{L}_{i+1} \end{bmatrix},$$

where Θ_{i+1} is any $(I \oplus J)$-unitary matrix that introduces the block zero entry on the right-hand side. In other words, the CKMS algorithm is an efficient procedure that exploits both kinds of structure: state-space and displacement structures.

References

R. Ackner (1991), *Fast Algorithms for Indefinite Matrices and Meromorphic Functions*, Ph.D. thesis, Department of Electrical Engineering, Stanford University, Stanford, CA.

R. Ackner and T. Kailath (1989a), *Complementary models and smoothing*, IEEE Trans. Automat. Contr., AC–34, pp. 963–969.

R. Ackner and T. Kailath (1989b), *Discrete-time complementary models and smoothing*, Int. J. Control, 49, pp. 1665–1682.

M. B. Adams, A. S. Willsky, and B. C. Levy (1984a), *Linear estimation of boundary value stochastic processes — Part 1: The role and construction of complementary models*, IEEE Trans. Automat. Contr., AC–29, pp. 803–811.

M. B. Adams, A. S. Willsky, and B. C. Levy (1984b), *Linear estimation of boundary value stochastic processes — Part 2: 1-D smoothing problems*, IEEE Trans. Automat. Contr., AC–29, pp. 811–821.

W. S. Agee and R. H. Turner (1972), *Triangular decomposition of a positive definite matrix plus a symmetric dyad with application to Kalman filtering*, White Sands Missile Range Technical Report 38.

N. I. Akhiezer (1965), *The Classical Moment Problem and Some Related Questions in Analysis*, Hafner Publishing Company, NY (in Russian, 1961).

D. F. Allinger and S. K. Mitter (1981), *New results on the innovations problem for nonlinear filtering*, Stochastics, 4, pp. 339–348.

D. L. Alspach and H. W. Sorenson (1972), *Nonlinear Bayesian estimation using Gaussian sum approximations*, IEEE Trans. Automat. Contr., AC–17, pp. 439–448.

V. A. Ambartsumian (1943), *Diffuse reflection of light by a foggy medium*, Dokl. Akad. Sci. SSSR, 38, pp. 229–322.

A. A. Anda and H. Park (1994), *Fast plane rotations with dynamic scaling*, SIAM J. Matrix Anal. Appl., 15, pp. 162–174.

B. D. O. Anderson and J. B. Moore (1979), *Optimal Filtering*, Prentice Hall, NJ.

B. D. O. Anderson and J. B. Moore (1990), *Optimal Control: Linear Quadratic Methods*, Prentice Hall, NJ.

T. Ando (1979), *Generalized Schur complements*, Linear Algebra and Its Applications, 27, pp. 173–186.

A. P. Andrews (1968), *A square-root formulation of the Kalman covariance equations*, AIAA Journal, vol. 6, pp. 1165–1166.

A. C. Antoulas, editor, (1991), *Mathematical System Theory: The Influence of R. E. Kalman*, Springer-Verlag, NY.

P. J. Antsaklis and A. N. Michel (1997), *Linear Systems*, McGraw-Hill, NY.

K. J. Åström (1970), *Introduction to Stochastic Control Theory*, Academic Press, NY.

M. Athans, editor, (1971), *Special Issue on the Linear Quadratic Gaussian Problem*, IEEE Trans. Automat. Contr., AC–16.

M. Athans and P. L. Falb (1966), *Optimal Control*, McGraw-Hill, NY.

S. Axler (1996), *Linear Algebra Done Right*, Springer-Verlag, NY.

T. I. Azizov and I. S. Iohvidov (1989), *Linear Operators in Spaces with an Indefinite Metric*, Wiley, NY.

F. A. Badawi, A. Lindquist, and M. Pavon (1979), *A stochastic realization approach to the smoothing problem*, IEEE Trans. Automat. Contr., AC–24, pp. 878–888.

H. Bart, I. Gohberg, and M. A. Kaashoek (1979), *Minimal Factorizations of Matrix and Operator Functions*, Operator Theory: Advances and Applications, vol. 1, Birkhäuser, Basel.

T. R. Bashkow (1957), *The A matrix: A new network description*, IRE Trans. Circuit Theory, vol. 4, pp. 117–120.

R. H. Battin (1962), *A statistical optimizing navigation procedure for space flight*, J. Amer. Rocket Soc., 32, pp. 1681–1692.

R. H. Battin (1964), *Astronautical Guidance*, McGraw-Hill, NY.

F. L. Bauer (1955), *Ein direktes iterations verfahren zur Hurwitz-zerlegung eines polynoms*, Arch. Elektr. Uebertragung, vol. 9, pp. 285–290.

F. L. Bauer (1956), *Beiträge zur entwicklung numerischer verfahren für programmgesteuerte rechenanlagen, II. Direkte faktorisierung eines polynoms*, Sitz. Ber. Bayer. Akad. Wiss., pp. 163–203.

R. T. Behrens and L. L. Scharf (1994), *Signal processing applications of oblique projection operators*, IEEE Trans. Signal Processing, 42, pp. 1413–1424.

J. F. Bellantoni and K. W. Dodge (1967), *A square-root formulation of the Kalman-Schmidt filter*, AIAA Journal, 5, pp. 1309–1314.

R. E. Bellman (1957), *Dynamic Programming*, Princeton University Press, NJ.

R. E. Bellman (1970), *Introduction to Matrix Analysis*, 2nd edition, McGraw-Hill, NY.

R. E. Bellman and E. D. Denman, editors, (1971), *Invariant Embedding*, in Lecture Notes in Operations Research, vol. 52, Springer, NY.

R. E. Bellman and G. M. Wing (1975), *An Introduction to Invariant Imbedding*, Wiley & Sons, NY.

M. G. Bello (1981), *Centralized and Decentralized Map-Updating and Terrain Masking Analysis*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

M. G. Bello, A. S. Willsky, and B. C. Levy (1989), *Construction and applications of discrete-time smoothing error models*, Int. J. Control, 50, pp. 203–223.

M. G. Bello, A. S. Willsky, B. C. Levy, and D. A. Castanon (1986), *Smoothing error dynamics and their use in the solution of smoothing and mapping problems*, IEEE Trans. Inform. Theory, IT–32, pp. 483–495.

A. Ben-Artzi and I. Gohberg (1994), *Orthogonal polynomials over Hilbert modules*, in Operator Theory: Advances and Applications, Birkhauser, Basel, vol. 73, pp. 96–126.

V. E. Benes (1976), *On Kailath's innovations conjecture*, Bell Syst. Tech. J., 55, pp. 981–1001.

G. J. Bierman (1973a), *A comparison of discrete linear filtering algorithms*, IEEE Trans. Aerosp. Electron., 9, pp. 28–37.

G. J. Bierman (1973b), *A square-root data array solution of the continuous-discrete filtering problem*, IEEE Trans. Automat. Contr., AC–18, pp. 675–676.

G. J. Bierman (1974), *Sequential square-root filtering and smoothing for discrete linear systems*, Automatica, 10, pp. 147–158.

G. J. Bierman (1977), *Factorization Methods for Discrete Sequential Estimation*, Academic Press, NY.

G. J. Bierman and C. L. Thornton (1977), *Numerical comparison of Kalman filter algorithms: Orbit determination case study*, Automatica, 13, pp. 23–35.

S. Bittanti, A. J. Laub, and J. C. Willems (1991), *The Riccati Equation*, Springer-Verlag, Berlin.

Å. Björck (1996), *Numerical Methods for Least-Squares Methods*, SIAM, PA.

H. W. Bode and C. E. Shannon (1950), *A simplified derivation of linear least-square smoothing and prediction theory*, in Proc. IRE, vol. 38, pp. 417–425.

J. Bognar (1974), *Indefinite Inner Product Spaces*, Springer-Verlag, NY.

A. W. Bojanczyk, R. P. Brent, P. Van Dooren, and F. R. de Hoog (1987), *A note on downdating the cholesky factorization*, SIAM Journal on Scientific and Statistical Computing, 8, pp. 210–221.

A. van den Bos (1998), *The real-complex normal distribution*, IEEE Trans. Inform. Theory, IT–44, pp. 1670–1672.

S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan (1994), *Linear Matrix Inequalities in System and Control Theory*, Studies in Applied Mathematics, SIAM, PA.

R. W. Brockett (1970), *Finite Dimensional Linear Systems*, Wiley, NY.

W. L. Brogan (1991), *Modern Control Theory*, 3rd edition, Prentice Hall, NJ.

A. Bruckstein and T. Kailath (1987a), *Inverse scattering for discrete transmission-line models*, SIAM Review, 29, pp. 359–389.

A. Bruckstein and T. Kailath (1987b), *An inverse scattering framework for several problems in signal processing*, IEEE ASSP Magazine, 4, pp. 6–20.

O. Brune (1931), *Synthesis of a finite two-terminal network whose driving point impedance is specified as a function of frequency*, J. Math. Phys., 10, pp. 191–236.

A. E. Bryson and M. Frazier (1963), *Smoothing for linear and nonlinear dynamic systems. TDR 63–119*, Tech. Rep., Aero. Sys. Div. Wright-Patterson Air Force Base, Ohio, pp. 353–364.

A. E. Bryson and Y. C. Ho (1969), *Applied Optimal Control*, Blaisdell, Waltham, MA.

R. S. Bucy (1959), *Optimum finite-time filters for a special nonstationary class of inputs*, John Hopkins University Phys. Lab., Internal Memorandum BBD–600.

R. S. Bucy (1970), *Linear and nonlinear filtering*, Proc. IEEE, 58, pp. 854–864.

R. S. Bucy and P. D. Joseph (1968), *Filtering for Stochastic Processes, with Applications to Guidance*, Wiley, NY.

R. S. Bucy and J. M. Rodriguez-Canabal (1972), *A negative definite equilibrium and its induced cone of global existence for the Riccati equation*, SIAM Journal on Mathematical Analysis, 3, pp. 644–646.

R. S. Bucy and K. D. Senne (1971), *Digital synthesis of nonlinear filters*, Automatica, 7, pp. 287–289.

J. Burns and R. K. Powers (1986), *Factorization and reduction methods for optimal control of hereditary systems*, Mat. Aplic. Comp., 5, pp. 203–248.

P. Businger and G. H. Golub (1965), *Linear least-squares solution by Householder transformations*, Numer. Math., 7, pp. 269–276.

C. I. Byrnes and C. I. Lindquist (1997), *On the partial stochastic realization problem*, IEEE Trans. Automat. Contr., vol. AC–42, pp. 1049–1070.

C. I. Byrnes, C. I. Lindquist, and T. McGregor (1991), *Predictability and unpredictability in Kalman filtering*, IEEE Trans. Automat. Contr., vol. AC–36, pp. 563–579.

P. E. Caines (1988), *Linear Stochastic Systems*, Wiley, NY.

F. M. Callier and C. A. Desoer (1991), *Linear System Theory*, Springer-Verlag, NY.

F. M. Callier, J. Winkin, and J. C. Willems (1994), *Convergence of the time-invariant Riccati differential equation and LQ-problem: Mechanisms of attraction*, Int. J. Control, 59, pp. 983–1000.

C. Carathéodory (1907), *Über den variabilitätsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehmen*, Math. Ann., 64, pp. 95–115.

N. A. Carlson (1973), *Fast triangular factorization of the square root filter*, AIAA J., 11, pp. 1259–1265.

A. G. Carlton (1962), *Linear estimation in stochastic processes*, Tech. Rep., Applied Physics Laboratory, John Hopkins University, Silver Springs, MD, Bumblebee Series Rept. No. 311.

A. G. Carlton and J. W. Follin (1956), *Recent developments in fixed and adaptive filtering*, Tech. Rep., AGARDograph 21, NATO Advisory Group for Aerospace Research and Development.

K. M. Case (1957), *On Wiener-Hopf equations*, Ann. Phys., 2, pp. 384–405.

J. L. Casti, R. E. Kalaba, and V. K. Murthy (1972), *A new initial-value method for on-line filtering and estimation*, IEEE Trans. Inform. Theory, IT–18, pp. 515–518.

J. L. Casti and O. Kirschner (1966), *Numerical experiments in linear control theory using generalized S-Y equations*, IEEE Trans. Automat. Contr., AC–21, pp. 792–795.

S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed (1997), *Parameter estimation in the presence of bounded modeling errors*, IEEE Signal Processing Letters, 4, pp. 195–197.

S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed (1998), *Parameter estimation in the presence of bounded data uncertainties*, SIAM J. Matrix Analysis and Applications, 19, pp. 235–252.

S. Chandrasekaran and A. H. Sayed (1996), *Stabilizing the generalized Schur algorithm*, SIAM J. Matrix Analysis and Applications, 17, pp. 950–983.

S. Chandrasekhar (1947a), *On the radiative equilibrium of a stellar atmosphere, Pt XXI*, Astrophys. J., 106, pp. 152–216.

S. Chandrasekhar (1947b), *On the radiative equilibrium of a stellar atmosphere, Pt XXII*, Astrophys. J., 107, pp. 48–72.

S. Chandrasekhar (1948), *On the radiative equilibrium of a stellar atmosphere, Pt XXII (concluded)*, Astrophys. J., 108, pp. 188–215.

S. Chandrasekhar (1950), *Radiative Transfer*, Oxford University Press, London (Dover Publication, NY,1960).

S. S. L. Chang (1961), *Synthesis of Optimum Control Systems*, McGraw-Hill, NY.

C. Chang and T. T. Georgiou (1992), *On a Schur-algorithm based approach to spectral factorization: Connection with the Riccati equation*, Linear Algebra and Its Applications, 171, pp. 233–247.

R. Chen and J. Liu (2000), *Mixture Kalman filters*, to appear in J. Roy. Statistical Soc., Ser. B.

J. Chun and T. Kailath (1991), *Divide and conquer solutions of least squares problems for matrices with displacement structure*, SIAM J. Matrix Anal. Appl., 12, pp. 128–145.

J. F. Claerbout (1976), *Fundamentals of Geophysical Data Processing with Applications to Petroleum Prospecting*, McGraw-Hill, NY.

L. D. Collins (1968), *Realizable whitening filters and state-variable realizations*, Proc. IEEE, vol. 56, pp. 100–101.

T. Constantinescu (1996), *Schur Parameters, Factorization, and Dilation Problems*, Birkhäuser, Berlin, Germany.

R. W. Cottle (1974), *Manifestations of the Schur complement*, Linear Algebra and Its Applications, 8, pp. 189–211.

H. Cramér (1960), *On some classes of nonstationary stochastic processes*, in Proc. 4th Berkeley Symp. Mathematics, Statistics, and Probability, Berkeley, CA, Univ. California Press, pp. 57–78.

S. Darlington (1959), *Nonstationary smoothing and prediction using network theory concepts*, IRE Trans. Inform. Theory (Special Suppl.), IT–5, pp. 1–14.

B. W. Datta (1995), *Numerical Linear Algebra Applications*, Brooks/Cole, CA.

B. W. Datta (2000), *Numerical Methods for Linear Control Systems Design and Analysis*, Academic Press, NY, 1999.

W. D. Davenport and W. L. Root (1958), *Random Signals and Noise*, McGraw-Hill, NY.

M. C. Davis (1963), *Factoring the spectral matrix*, IEEE Trans. Automat. Contr., AC–8, pp. 296–305.

M. H. A. Davis (1977), *Linear Estimation and Stochastic Control*, Halsted Press, NY.

R. J. P. de Figueiredo and Y. G. Jan (1971), *Spline filters*, in Proc. 2nd Symp. on Nonlinear Estimation Theory and its Applications, San Diego, pp. 127–141.

J. W. Demmel (1997), *Applied Numerical Linear Algebra*, SIAM, PA.

U. B. Desai and D. Pal (1984), *A transformation approach to stochastic model reduction*, IEEE Trans. Automat. Contr., AC–29, pp. 1097–1100.

U. B. Desai, H. L. Weinert, and G. Yusypchuk (1983), *Discrete-time complementary models and smoothing algorithms: The correlated case*, IEEE Trans. Automat. Contr., AC–28, pp. 536–539.

A. A. Desalu, L. A. Gould, and F. C. Schweppe (1974), *Dynamic estimation of air pollution*, IEEE Trans. Automat. Contr., AC–19, pp. 904–910.

P. Dewilde, A. Vieira, and T. Kailath (1978), *On a generalized Szego-Levinson realization algorithm for optimal linear predictions on a network synthesis approach*, IEEE Trans. on Circuits and Systems, vol. 25, pp. 663–675.

P. Dewilde and H. Dym (1981a), *Schur recursions, error formulas and convergence of rational estimators for stationary stochastic sequences*, IEEE Trans. Inform. Theory, IT-27, no. 4, pp. 446–461.

P. Dewilde and H. Dym (1981b), *Lossless chain scattering matrices and optimum linear prediction: The vector case*, Circuit Theory and Applications, vol. 9, pp. 135–175.

J. L. Doob (1944), *The elementary Gaussian processes*, Ann. Math. Statist., 15, pp. 229–282.

J. L. Doob (1953), *Stochastic Processes*, John Wiley, NY.

P. Dorato (1983), *Theoretical developments in discrete-time control*, Automatica, 19, pp. 395–400.

P. Dorato, C. Abdallah, and V. Cerone (1995), *Linear Quadratic Control: An Introduction*, Prentice Hall, NJ.

J. C. Doyle (1978), *Guaranteed margins for LQG regulators*, IEEE Trans. Automat. Contr., AC-23, pp. 756–757.

D. B. Duncan and S. D. Horn (1972), *Linear dynamic regression from the viewpoint of regression analysis*, J. Amer. Stat. Assoc., 67, pp. 815–821.

J. Durbin (1960), *The fitting of time series models*, Rev. L'Institut Intl. de Statistique, 28, pp. 233–244.

P. Dyer and S. R. McReynolds (1969), *Extension of square-root filtering to include process noise*, J. Optimiz. Theory Appl., 3, pp. 444–459.

D. J. Edelblute (1996), *Noncircularity*, IEEE Signal Processing Letters, 3, pp. 156–157.

P. Faurre (1970), *Identification par Minimisation d'une Representation Markovienne de Processus Aleatoire*, Springer Lecture Notes in Math., vol. 132.

P. Faurre (1991), *Kalman filtering and the advancement of navigation and guidance*, in A. C. Antoulas, editor, Mathematical System Theory: The Influence of R. E. Kalman, Springer-Verlag, NY.

P. Faurre, M. Clerget, and F. Germain (1979), *Operateurs Rationnels Positifs*, Dunod, Paris, France.

P. Faurre and J. Marmorat (1969), *Un algorithme de realization stochastique*, C. R. Acad. Sci. Paris, Ser. A, 268, pp. 978–981.

R. J. Fitzgerald (1971), *Divergence of the Kalman filter*, IEEE Trans. Automat. Contr., AC-16, pp. 736–747.

D. C. Fraser (1967), *A New Technique for the Optimal Smoothing of Data*, Ph.D. thesis, Dept. of Aero. and Astro., MIT, Cambridge, MA.

D. C. Fraser and J. E. Potter (1969), *The optimum linear smoother as a combination of two optimum linear filters*, IEEE Trans. Automat. Contr., AC-14, pp. 387–390.

M. Fréchet (1950), *Recherches Théoriques Modernes sur le Calcul des Probabilités*, vol. 1, Gauthier-Villars, Paris, 2nd edition.

B. Friedland (1969), *Treatment of bias in recursive filtering*, IEEE Trans. Automat. Contr., AC-14, pp. 359–367.

B. Friedlander (1976), *Scattering Theory and Linear Least-Squares Estimation*, Ph.D. thesis, Department of Electrical Engineering, Stanford University, CA.

B. Friedlander, T. Kailath, and L. Ljung (1976), *Scattering theory and linear least squares estimation — II: Discrete-time problems*, J. Franklin Institute, 301, pp. 71–82.

B. Friedlander, T. Kailath, M. Morf, and L. Ljung (1978), *Extended Levinson and Chandrasekhar equations for general discrete-time linear estimation problems*, IEEE Trans. Automat. Contr., AC-23, pp. 653–659.

P. A. Frost and T. Kailath (1971), *An innovations approach to least-squares estimation, part III: Nonlinear estimation in white Gaussian noise*, IEEE Trans. Automat. Contr., AC-16, pp. 217–226.

F. R. Gantmacher (1959), *The Theory of Matrices*, Chelsea Publishing Company, NY.

C. F. Gauss (1809), *Theory of the Motion of Heavenly Bodies*, Dover, NY. [English translation of *Theoria Motus Corporum Coelestium, 1809.*]

C. F. Gauss (1809), *Theory of the Motion of Heavenly Bodies*, Dover, NY. [English translation of *Theoria Motus Corporum Coelestium, 1809.*]

R. Geesey and T. Kailath (1973), *An innovations approach to least-squares estimation, part V: Innovations representations and recursive estimation in colored noise*, IEEE Trans. Automat. Contr., AC-18, pp. 435–453.

A. Gelb, editor, (1974), *Applied Optimal Estimation*, MIT Press, Cambridge, MA.

A. Gelb, (1986), *Dual contributions of optimal estimation theory in aerospace applications*, IEEE Control Systems Magazine, pp. 3–13.

I. M. Gel'fand and S. V. Fomin (1963), *Calculus of Variations*, Prentice Hall, NJ (translated from Russian edition).

M. Gentleman (1973), *Least squares computations by Givens transformations*, J. Inst. Math. Appl., 12, pp. 329–336.

T. T. Georgiou (1989), *Computational aspects of spectral factorization and the tangential Schur algorithm*, IEEE Trans. Circuits Syst., 36, pp. 103–108.

L. Y. Geronimus (1961), *Orthogonal Polynomials*, Transl. Consultants Bureau, NY [original in Russian, 1958.]

M. R. Gevers and T. Kailath (1973), *Constant, predictable, and degenerate directions of the discrete-time Riccati equation*, Automatica, vol. 9, pp. 699–711.

L. E. Ghaoui and H. Hebret (1997), *Robust solutions to least-squares problems with uncertain data*, SIAM J. Matrix Anal. Appl., 18, pp. 1035–1064.

I. Gohberg and I. Fel'dman (1974), *Convolution Equations and Projection Methods for their Solutions*, vol. 41 of Translations of Mathematical Monographs, Amer. Math. Soc., Providence, RI.

I. Gohberg and M. A. Kaashoek, editors, (1986), *Constructive Methods of Wiener-Hopf Factorization*, Operator Theory: Advances and Applications, vol. 21, Birkhäuser, Verlag.

I. Gohberg and M. G. Krein (1958), *Systems of integral equations on a half line with kernels depending on the difference of the arguments*, Usp. Mat. Nauk., pp. 43–118. [English translation in *Amer. Math. Soc. Transl.* (2), vol. 13, pp. 185–264, 1960.]

I. Gohberg and M. G. Krein (1964), *The factorization of operators in Hilbert space*, Acta Sci. Math. (in Hungarian), 25, pp. 90–123. English translation in *Amer. Math. Soc. Transl.* (2), 51, pp. 155–188, 1966.

I. Gohberg and M. G. Krein (1970), *Theory and Applications of Volterra Operators in Hilbert Space*, Transl. Math. Monog., vol. 24, Amer. Math. Soc. , Providence, RI. [Russian original, Izdat. Nauk., 1967.]

I. Gohberg and A. Semencul (1972), *On the inversion of finite Toeplitz matrices and their continuous analogs*, Mat. Issled., vol. 2, pp. 201–233.

H. H. Goldstine and L. P. Horwitz (1966), *Hilbert space with non-associative scalars, II*, Math. Annal., 164, pp. 291–316.

G. H. Golub (1965), *Numerical methods for solving linear least-squares problems*, Numer. Math., 7, pp. 206–216.

G. H. Golub and C. F. Van Loan (1996), *Matrix Computations*, 3rd edition, The Johns Hopkins University Press, Baltimore.

T. N. T. Goodman, C. A. Micchelli, G. Rodriguez, and S. Seatzu (1997), *Spectral factorization of Laurent polynomials*, Advances in Computational Mathematics, vol. 7, pp. 429–445.

N. Gordon, D. Salmond, and C. Ewing (1995), *Bayesian state estimation for tracking and guidance using the bootstrap filter*, Journal of Guidance, Control, and Dynamics, 18, pp. 1434–1443.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith (1993), *Novel approach to nonlinear non-Gaussian Bayesian state estimation*, IEE Proceedings F (Radar and Signal Processing), 140, pp. 107–113.

M. Green and D. J. N. Limebeer (1995), *Linear Robust Control*, Prentice Hall, NJ.

U. Grenander and M. Rosenblatt (1957), *Statistical Analysis of Stationary Time Series*, Wiley, NY.

U. Grenander and G. Szegö (1958), *Toeplitz Forms and their Applications*, University of California Press.

M. S. Grewal and A. P. Andrews (1993), *Kalman Filtering: Theory and Practice*, Prentice Hall, NJ.

R. E. Griffin and A. P. Sage (1969), *Sensitivity analysis of discrete filtering and smoothing algorithms*, AIAA Journal, 7, pp. 1890–1897.

T. F. Gunckel and G. F. Franklin (1963), *A general solution for linear sampled-data control systems*, J. Basic Eng. Trans. ASME 85D, pp. 197–203.

J. Hajek (1962), *On linear statistical problems in stochastic processes*, Czech. Math. J., 12, pp. 404–444.

T. Hall (1970), *Carl Friedrich Gauss*, Cambridge, MA.

E. J. Hannan (1970), *Multiple Time Series*, Wiley, NY.

E. J. Hannan and M. Deistler (1988), *The Statistical Theory of Linear Systems*, Wiley, NY.

J. E. Hanson (1957), *Some notes on the application of the calculus of variations to smoothing for finite time, etc.*, Tech. Rep., *Appl. Physc. Lab.*, John Hopkins Univ., Baltimore, MD, Internal Memo BBD–346.

R. J. Hanson and C. L. Lawson (1969), *Extensions and applications of the Householder algorithm for solving linear least-squares problems*, Math. Comput., 23, pp. 787–812.

B. Hassibi, A. H. Sayed, and T. Kailath (1996a), *Linear estimation in Krein spaces — Part I: Theory*, IEEE Trans. Automat. Contr., 41, pp. 18–33.

B. Hassibi, A. H. Sayed, and T. Kailath (1996b), *Linear estimation in Krein spaces — Part II: Applications*, IEEE Trans. Automat. Contr., 41, pp. 34–49.

B. Hassibi, A. H. Sayed, and T. Kailath (1996c), *$\mathcal{H}^\infty$ optimality of the LMS algorithm*, IEEE Trans. Signal Processing, 44, pp. 267–280.

B. Hassibi, A. H. Sayed, and T. Kailath (1999), *Indefinite Quadratic Estimation and Control: A Unified Approach to $\mathcal{H}^2$ and $\mathcal{H}^\infty$ Theories*, Studies in Applied Mathematics, vol. 16, SIAM, PA.

S. Haykin (1996), *Adaptive Filter Theory*, Prentice Hall, NJ, 3rd edition.

E. V. Haynsworth (1968), *Determination of the inertia of a partitioned Hermitian matrix*, Linear Algebra and Its Applications, 1, pp. 73–81.

H. Heffes (1966), *The effect of erroneous models on the Kalman filter response*, IEEE Trans. Automat. Contr., AC–11, pp. 541–543.

G. Heinig and K. Rost (1984), *Algebraic Methods for Toeplitz-like Matrices and Operators*, Birkhäuser.

C. W. Helstrom (1965), *Solution of the detection integral equation for stationary filtered white noise*, IEEE Trans. Inform. Theory, IT–11, pp. 335–339.

H. V. Henderson and S. R. Searle (1981), *On deriving the inverse of a sum of matrices*, SIAM Review, 23, pp. 53–60.

N. J. Higham (1996), *Accuracy and Stability of Numerical Algorithms*, SIAM, PA.

Y. C. Ho (1963), *On the stochastic approximation method and optimal filter theory*, J. Math. Anal. Appl., 6, pp. 152–154.

B. L. Ho and R. E. Kalman (1965), *Effective construction of linear state-variable models from input/output functions*, Proc. Third Allerton Conference, pp. 449–459.

R. A. Horn and C. R. Johnson (1985), *Matrix Analysis*, Cambridge University Press.

R. A. Horn and C. R. Johnson (1991), *Topics in Matrix Analysis*, Cambridge University Press.

A. Houacine (1991), *Regularized fast recursive least squares algorithms for adaptive filtering*, IEEE Trans. Signal Processing, 39, pp. 860–870.

A. Houacine and G. Demoment (1986), *Chandrasekhar adaptive regularizer for adaptive filtering*, in Proc. IEEE ICASSP, Tokyo, Japan, pp. 2967–2970.

A. S. Householder (1953), *Principles of Numerical Analysis*, McGraw-Hill, NY.

S. F. Hsieh, K. J. R. Liu, and K. Yao (1993), *A unified square-root-free approach for QRD-based recursive least-squares estimation*, IEEE Trans. Signal Processing, 41, pp. 1405–1409.

S. V. Huffel and J. Vandewalle (1991), *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, PA.

C. E. Hutchinson (1984), *The Kalman filter applied to aerospace and electronics systems*, IEEE Trans. Aerosp. and Electronic Syst., pp. 500–504.

V. Ionescu, C. Oara, and M. Weiss (1997), *General matrix pencil techniques for the solution of algebraic Riccati equations: A unified approach*, IEEE Trans. Automat. Contr., AC–42, pp. 1085–1097.

V. Ionescu and M. Weiss (1993), *Continuous and discrete-time Riccati theory: A Popov function approach*, Linear Algebra and Its Applications, 193, pp. 173–209.

R. Jaffe and E. Rechtin (1955), *Design and performance of phase-lock circuits capable of near-optimum performance over a wide range of input signal and noise levels*, IRE Trans. Inf. Theory, IT–1, pp. 66–76.

K. Jänich (1996), *Linear Algebra*, Springer-Verlag.

A. H. Jazwinski (1969), *Adaptive filtering*, Automatica, 5, pp. 475–485.

A. H. Jazwinski (1970), *Stochastic Processes and Filtering Theory*, Academic Press, NY.

P. D. Joseph and J. T. Tou (1961), *On linear control theory*, Trans. AIEE, vol. 80, pp. 193–196.

H. Kagiwada and R. E. Kalaba (1966), *An initial-value method for Fredholm integral equations of convolution type*, Tech. Rep., RAND Corp. Memo RM–5186-PR.

T. Kailath (1960), *Estimating filters for linear time-invariant channels*, Quarterly Progress Rep. 58, MIT Research Laboratory for Electronics, pp. 185–197.

T. Kailath (1961), *Optimum receivers for randomly varying channels*, Proc. Fourth London Symposium on Information Theory, C. Cherry, editor, Butterworths, London, pp. 189–212.

T. Kailath (1968), *An innovations approach to least-squares estimation, part I: Linear filtering in additive white noise*, IEEE Trans. Automat. Contr., AC–13, pp. 646–655.

T. Kailath (1969a), *A general likelihood ratio formula for random signals in noise*, IEEE Trans. Inform. Theory, IT–15, pp. 350–361.

T. Kailath (1969b), *Application of a resolvent identity to a linear smoothing problem*, SIAM J. Control Optim., 7, pp. 68–74.

T. Kailath (1970), *Likelihood ratios for Gaussian processes*, IEEE Trans. Inform. Theory, IT–16, pp. 276–288.

T. Kailath (1971), *Some extensions of the innovations theorem*, Bell Syst. Tech. J., 50, pp. 1487–1494.

T. Kailath (1972a), *A note on least-squares estimation by the innovations method*, SIAM J. Comput., vol. 10, no. 3, pp. 477–486.

T. Kailath (1972b), *Some Chandrasekhar-type algorithms for quadratic regulator problems*, in Proc. Conference on Decision and Control, New Orleans, pp. 219–223.

T. Kailath (1973), *Some new algorithms for recursive estimation in constant linear systems*, IEEE Trans. Inform. Theory, 19, pp. 750–760.

T. Kailath (1974), *A view of three decades of linear filtering theory*, IEEE Trans. Inform. Theory, 20, pp. 146–181.

T. Kailath (1975), *Supplement to "A survey of data smoothing"*, Automatica, vol. 11, pp. 109–111.

T. Kailath, editor, (1977), *Linear Least-Squares Estimation*, Benchmark Papers in Elec. Eng. and Comp. Science, vol. 17, Academic Press.

T. Kailath (1979), *Redheffer scattering theory and linear state-space estimation problems*, Ricerchedi Automatica, 10, pp. 136–162.

T. Kailath (1980), *Linear Systems*, Prentice Hall, NJ.

T. Kailath (1981), *Lectures on Wiener and Kalman Filtering*, 2nd edition, Springer, NY.

T. Kailath (1982), *Equations of Wiener-Hopf type in filtering theory and related applications*, pp. 63–64 in *Norbert Wiener: Collected Works*, vol. III, Ed. P. Masani, MIT Press.

T. Kailath (1987), *Signal processing applications of some moment problems*, Moments in Mathematics, H. Landau, editor, American Mathematical Society, vol. 37, pp. 71–109.

T. Kailath (1991a), *Remarks on the origin of the dis placement-rank concept*, Applied Mathematics and Computation, vol. 45, pp. 193–206.

T. Kailath (1991b), *From Kalman filtering to innovations, martingales, scattering and other nice things*, in Mathematical System Theory: The Influence of R.E. Kalman, A. C. Antoulas, Ed., Springer-Verlag, Berlin, pp. 55–88.

T. Kailath (1997), *Norbert Wiener and the development of mathematical engineering*, in Proceedings of Symposia in Pure Mathematics, D. Jerison, I. M. Singer, and D. W. Strock, editors, American Mathematical Society, pp. 93–116.

T. Kailath (1999), *Displacement structure and array algorithms*, in Fast Reliable Algorithms for Matrices with Structure, T. Kailath and A. H. Sayed, editors, SIAM, PA.

T. Kailath and D. L. Duttweiler (1972), *An RKHS approach to detection and estimation theory, part III: Generalized innovations representations and a likelihood ratio formula*, IEEE Trans. Inform. Theory., IT–18, pp. 730–745.

T. Kailath and P. A. Frost (1968), *An innovations approach to least-squares estimation, part II: Linear smoothing in additive white noise*, IEEE Trans. Automat. Contr., AC–13, pp. 655–660.

T. Kailath and R. Geesey (1971), *An innovations approach to least-squares estimation, part IV: Recursive estimation given the covariance functions*, IEEE Trans. Automat. Contr., AC–16, pp. 720–727.

T. Kailath, T., S. Y. Kung, and M. Morf, M. (1979), *Displacement ranks of a matrix*, Bulletin of the American Mathematical Society, vol. 1, no. 5, pp. 769–773.

T. Kailath and L. Ljung (1977), *A scattering theory framework for fast least-squares algorithms*, in Multivariate Analysis — IV, North Holland, pp. 387–406.

T. Kailath and L. Ljung (1982), *Two filter smoothing formulae by diagonalization of the Hamiltonian equations*, Int. J. Control, vol. 36, no. 4, pp. 663–673.

T. Kailath, L. Ljung, and M. Morf (1983), *Recursive input-output and state-space solutions for continuous-time linear estimation problems*, IEEE Trans. Automat. Contr., AC–28, pp. 897–906. [See also Proc. IEEE Conf. on Decision and Control, pp. 182–185, Florida, Dec. 1976.]

T. Kailath, M. Morf, and G. S. Sidhu (1973), *Some new algorithms for recursive estimation in constant linear discrete-time systems*, in Proc. Seventh Princeton Conf. Inform. Sci. Systems, Princeton, NJ, pp. 344–352.

T. Kailath and A. H. Sayed (1995), *Displacement structure: Theory and applications*, SIAM Review, 37, pp. 297–386.

T. Kailath and A. H. Sayed, editors, (1999), *Fast Reliable Algorithms for Matrices with Structure*, SIAM, PA.

T. Kailath, A. C. Vieira, and M. Morf (1978a), *Orthogonal transformation (square–root) implementations of the generalized Chandrasekhar and generalized Levinson algorithms*, in Lecture Notes in Control and Information Sciences, A. Bensoussan and J. Lions, Eds., Springer-Verlag, NY, vol. 32, pp. 81–91.

T. Kailath, A. C. Vieira, and M. Morf (1978b), *Inverses of Toeplitz operators, innovations, and orthogonal polynomials*, SIAM Review, vol. 20, no. 1, pp. 106–119.

T. Kailath and M. Wax (1984), *A note on the complementary model of Weinert and Desai*, IEEE Trans. Automat. Contr., AC–29, pp. 551–552.

R. E. Kalman (1960a), *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Eng., 82, pp. 34–45. [Also published as ASME Paper 59-IRD–11.]

R. E. Kalman (1960b), *On the general theory of control systems*, in Proc. of the first IFAC Congress, London, vol. 1, pp. 481–491.

R. E. Kalman (1960c), *Contributions to the theory of optimal control*, Bol. Soc. Mat. Mex., 5, pp. 102–119.

R. E. Kalman (1963a), *Lyapunov functions for the problem of Lur'e in automatic control*, Proc. Nat. Acad. Sci., 49, pp. 201–205.

R. E. Kalman (1963b), *New methods of Wiener filtering theory*, in Proc. 1st Symp. Engineering Applications of Random Function Theory and Probability, J. L. Bogdanoff and F. Kozin, Eds., Wiley, NY, pp. 279–388. [See also the closely related RIAS technical report 61–1, Feb. 1961.]

R. E. Kalman (1963c), *Mathematical description of linear dynamical systems*, SIAM J. Contr., 1, pp. 152–192.

R. E. Kalman (1964), *When is a linear control system optimal?* Trans. ASME, Ser. D (J. Basic Engr.), vol. 86, pp. 51–60.

R. E. Kalman (1965), *Linear stochastic filtering theory – Reappraisal and outlook*, Proc. Symp. on System Theory, J. Fox, editor, Polytechnique Institute of Brooklyn, pp. 197–205.

R. E. Kalman (1968), *Lectures on Controllability and Observability*, CIME, Bologna.

R. E. Kalman (1978), *A retrospective after twenty years: From the pure to the applied*, in Applications of Kalman Filter to Hydrology, Hydraulics and Water Resources, C. Chiu, editor, Detp. Civil Engineering, University of Pittsburgh, pp. 31–54.

R. E. Kalman and J. E. Bertram (1958), *General synthesis procedures for computer control of single and multiloop linear systems*, Trans. AIEE, 77, part 2, pp. 602–609.

R. E. Kalman and R. S. Bucy (1961), *New results in linear filtering and prediction theory*, Trans. ASME, Ser. D, J. Basic Engr., 83, pp. 95–107. [Also published as ASME Paper 60-JAC–12.]

R. E. Kalman, P. L. Falb, and M. A. Arbib (1969), *Topics in Mathematical System Theory*, McGraw-Hill, NY.

R. E. Kalman and R. W. Koepcke (1958), *Optimal synthesis of linear sampling control systems using generalized performance indices*, Trans. ASME, Ser. D., J. Basic Engr., 80, pp. 1820–1826.

P. G. Kaminski (1971), *Square-Root Filtering and Smoothing for Discrete Processes*, Ph.D. thesis, Dept. Aero. and Astro., Stanford University, Stanford, CA.

P. G. Kaminski and A. E. Bryson (1972), *Discrete square-root smoothing*, Proc. AIAA Guidance and Control Conference.

P. G. Kaminski, A. E. Bryson, and S. F. Schmidt (1971), *Discrete square-root filtering: a survey of current techniques*, IEEE Trans. Automat. Contr., AC–16, pp. 727–736.

S. Kayalar and H. L. Weinert (1989), *Oblique projections: Formulas, algorithms, and error bounds*, Math. Contr. Signals Syst., 2, pp. 33–45.

H. K. Khalil (1996), *Nonlinear Systems*, 2nd edition, Prentice Hall, NJ.

L. Kleinrock (1975), *Queuing Systems*, vol. 1: Theory, Wiley, NY.

A. N. Kolmogorov (1939), *Sur l'interpolation et extrapolation des suites stationnaires*, C. R. Acad. Sci., 208, p. 2043.

A. N. Kolmogorov (1941a), *Stationary sequences in Hilbert space (in Russian)*, Bull. Math. Univ. Moscow, 2. [A translation by N. Artin is available in many libraries.]

A. N. Kolmogorov (1941b), *Interpolation and extrapolation of stationary random processes*, Bull. Acad. Sci. USSR, 5. [A translation has been published by the RAND Corp., Santa Monica, CA, as Memo. RM–3090-PR, Apr. 1962.]

M. G. Krein (1958), *Integral equations on a half line with kernels depending upon the difference of the arguments*, Uspehi Mat. Nauk, 13, pp. 3–120. [English translation in *Amer. Math. Soc. Transl.*, vol. 22, pp. 163–288, 1962.]

V. Kučera (1972), *A contribution to matrix quadratic equations*, IEEE Trans. Automat. Contr., 17, pp. 344–347.

V. Kučera (1974), *Closed-loop stability of discrete linear single-variable systems*, Kybernetika, 10, 146–171.

V. Kučera (1975), *Algebraic approach to discrete linear control*, IEEE Trans. Automat. Contr., 20, pp. 116–120.

V. Kučera (1991), *Algebraic Riccati equation: Hermitian and definite solutions*, The Riccati Equation, S. Bittanti, A. J. Laub, and J. C. Willems, editors, Springer-Verlag, Berlin, pp. 53–89.

H. Kwakernaak and M. Sebek (1994), *Polynomial J-spectral factorization*, IEEE Trans. Automat. Contr., AC–39, pp. 315–328.

H. Kwakernaak and R. Sivan (1972), *Linear Optimal Control Systems*, Wiley, NY.

D. G. Lainiotis (1974), *Partitioned estimation algorithms, II: Linear estimation*, Information Sciences, 7, pp. 317–340.

D. G. Lainiotis (1976a), *General backwards Markov models*, IEEE Trans. Automat. Contr., AC–21, pp. 595–599.

D. G. Lainiotis (1976b), *Partitioned Riccati solutions and integration free doubling algorithms*, IEEE Trans. Automat. Contr., AC–21, pp. 677–689.

P. Lancaster and L. Rodman (1991), *Solutions of the continuous and discrete time algebraic Riccati equations: A review*, in The Riccati Equation, S. Bittanti, A. J. Laub, and J. C. Willems, Eds., Springer-Verlag, Berlin, pp. 11–52.

P. Lancaster and L. Rodman (1995), *Algebraic Riccati Equations*, Oxford University Press, NY.

P. Lancaster and M. Tismenetsky (1985), *The Theory of Matrices with Applications*, 2nd edition, Academic Press, NY.

C. Lanczos (1956), *Applied Analysis*, Prentice Hall, NJ. [Reprinted by Dover, NY, 1988.]

J. H. Laning and R. H. Battin (1956), *Random Processes in Automatic Control*, McGraw-Hill, PA.

B. F. La Scala and R. R. Bitmead (1996), *Design of an extended Kalman filter frequency tracker*, IEEE Trans. Signal Processing, 44, pp. 739–742.

A. J. Laub (1979), *A Schur method for solving algebraic Riccati equations*, IEEE Trans. Automat. Contr., AC–24, pp. 913–921.

A. J. Laub (1991), *Invariant subspace methods for the numerical solution of Riccati equations*, The Riccati Equation, S. Bittanti, A. J. Laub, and J. C. Willems, editors, Springer-Verlag, Berlin, pp. 163–197.

A. J. Laub and A. Linnemann (1986), *Hessenberg and Hessenberg/triangular forms in linear system theory*, Int. J. Control, 44, pp. 1523–1547.

C. L. Lawson and R. J. Hanson (1995), *Solving Least-Squares Problems*, SIAM, PA. [A republication of the original 1974 Prentice Hall version.]

P. Lax (1997), *Linear Algebra*, John Wiley, NY.

D. Lay (1994), *Linear Algebra and Its Applications*, Addison-Wesley.

R. C. K. Lee (1960), *Optimal Estimation, Identification, and Control*, MIT Press, Cambridge, MA.

A. M. Legendre (1805), *Nouvelles Méthodes pour la Détermination des Orbites de Comètes*, Courcier, Paris.

A. M. Legendre (1810), *Méthode de moindres quarres, pour trouver le milieu de plus probable entre les résultats des différentes observations*, Mem. Inst. France, pp. 149–154.

E. C. T. Leondes (1970), *Theory and applications of Kalman filtering*, Tech. Rep., AGARDOgraph 139, NATO Advisory Group for Aerospace Research and Development.

A. M. Letov (1960), *Analytic controller design I, II*, Autom. Remote Contr., vol. 21, pp. 303–306.

H. Lev-Ari and T. Kailath (1986), *Triangular factorization of structured Hermitian matrices*, Operator Theory: Advances and Applications, I. Gohberg, editor, Birkhäuser, Boston, vol. 18, pp. 301–324.

H. Lev-Ari, T. Kailath, and J. Cioffi (1984), *Least squares adaptive lattice and transversal filters: A unified geometrical theory*, IEEE Trans. Inform. Theory, IT-30, pp. 222–236.

N. Levinson (1947), *The Wiener r.m.s. (root-mean-square) error criterion in filter design and prediction*, J. Math. Phys., 25, pp. 261–278.

B. C. Levy, D. A. Castanon, G. Verghese, and A. S. Willsky (1983), *A scattering framework for decentralized estimation problems*, Automatica, 19, pp. 373–384.

B. C. Levy, T. Kailath, L. Ljung, and M. Morf (1979), *Fast time-invariant implementations for linear least-squares smoothing filters*, IEEE Trans. Automat. Contr., AC-24, pp. 770–774.

F. L. Lewis (1986), *Optimal Control*, Wiley, NY.

I. S. Lin and S. K. Mitra (1996), *Overlapped block digital filtering*, IEEE Trans. on Circuits and Systems, II: Analog and Digital Signal Processing, vol. 43, no. 8, pp. 586–596.

A. Lindquist (1974), *A new algorithm for optimal filtering of discrete-time stationary processes*, SIAM J. Control, 12, pp. 736–746.

A. Lindquist (1976), *Some reduced-order non-Riccati equations for linear least-squares estimation: The stationary, single output case*, Int. J. Control, 24, pp. 821–842.

A. Lindquist and G. Picci (1996), *Canonical correlation analysis, approximate covariance extension, and identification of stationary time series*, Automatica, vol. 32, pp. 709–733.

L. Ljung (1987), *System Identification: Theory for the User*, Prentice Hall, NJ.

L. Ljung and T. Kailath (1976a), *A unified approach to smoothing formulas*, Automatica, 12, pp. 147–157.

L. Ljung and T. Kailath (1976b), *Backwards Markovian models for second-order stochastic processes*, IEEE Trans. Inform. Theory, IT-22, pp. 488–491.

L. Ljung and T. Kailath (1976c), *The asymptotic behaviour of constant-coefficient Riccati differential equations*, IEEE Trans. Inform. Theory, AC-21, pp. 385–388.

L. Ljung and T. Kailath (1977), *Efficient change of initial conditions, dual Chandrasekhar equations, and some applications*, IEEE Trans. Automat. Contr., AC-22, pp. 443–447.

L. Ljung, T. Kailath, and B. Friedlander (1976), *Scattering theory and linear least squares estimation, Part I: Continuous-time problems*, Proc. IEEE, Special Issue on Recent Trends in System Theory, 64, pp. 131–139.

M. Loève (1963), *Probability Theory*, 3rd edition, Van Nostrand Reinhold, NY.

D. G. Luenberger (1964), *Observing the state of a linear system*, IEEE Trans. Military Electronics, 8, pp. 74–80.

D. G. Luenberger (1969), *Optimization by Vector Space Methods*, Wiley, NY.

D. G. Luenberger (1971), *An introduction to observers*, IEEE Trans. Automat. Contr., AC-16, pp. 596–602.

A. M. Lyapunov (1892), *The general problem of motion stability*. Translated in French, *Ann. Fac. Sci Toulouse*, vol. 9, pp. 203–474, 1907. [Also reprinted in *Ann. Math. Study*, no. 17, Princeton University Press, 1949.]

A. G. J. MacFarlane (1963), *An eigenvector solution of the optimal linear regulator problem*, J. Electron. Contr., 14, pp. 643–654.

A. K. Mahalanabis and K. Xue (1987), *An efficient two-dimensional Chandrasekhar filter for restoration of images degraded by spatial blur and noise*, IEEE Trans. Acoust. Speech Signal Process., ASSP-35, pp. 1603–1610.

P. Masani (1966), *Wiener's contribution to generalized harmonic analysis, prediction theory and filter theory*, Bull. Amer. Math. Soc., 72, pp. 73–125.

P. S. Maybeck (1979), *Stochastic Models, Estimation and Control*, vol. 1, Academic Press, NY.

P. S. Maybeck (1982), *Stochastic Models, Estimation and Control*, vol. 2, Academic Press, NY.

D. Q. Mayne (1966), *A solution to the smoothing problem for linear dynamic systems*, Automatica, 4, pp. 73–92.

A. L. McBride (1973), *On optimum sample-data FM demodulation*, IEEE Trans. Communications, COM-21, pp. 40–50.

J. A. McClelland (1906), *The energy of secondary radiation*, Royal Dublin Soc. Scientific Trans., 9, pp. 9–26.

L. A. McGee and S. F. Schmidt (1985), *Discovery of the Kalman filter as a practical tool for aerospace and industry*, Technical Memorandum 86847, NASA.

J. S. Meditch (1969), *Stochastic optimal linear estimation and control*, McGraw-Hill, NY.

J. S. Meditch (1973), *A survey of data smoothing for linear and nonlinear systems*, Automatica, vol. 9, pp. 151–162.

R. K. Mehra (1970), *On the identification of variances and adaptive Kalman filtering*, IEEE Trans. Autmat. Contr., AC-15, pp. 175–183.

R. Merched and A. H. Sayed (1998), *Block trigonometric transform adaptive filtering*, Proc. Asilomar Conference on Signals, Systems, and Computers, vol. 1, pp. 384–388, Pacific Grove, CA.

R. Merched and A. H. Sayed (1999), *Fast RLS Laguerre adaptive filtering*, Proc. Allerton Conference on Communication, Control, and Computing, Allerton, IL.

P. A. Meyer (1973), *Sur un problème de filtration, Séminaire de probabilitiés, Pt. VII, Lecture Notes in Mathematics*, vol. 321, Springer, NY, pp. 223–247.

S. K. Mitra (1998), *Digital Signal Processing*, McGraw-Hill, NY.

B. P. Molinari (1977), *The time-invariant linear-quadratic optimal control problem*, Automatica, 13, pp. 347–357.

M. Morf (1974), *Fast Algorithms for Multivariable Systems*, Ph.D. thesis, Dept. Electrical Engineering, Stanford University, CA.

M. Morf and T. Kailath (1975), *Square root algorithms for least squares estimation*, IEEE Trans. Automat. Contr., AC-20, pp. 487–497. M. Morf, B. C. Levy, and T. Kailath (1978), *Square root algorithms for the continuous-time linear least-square estimation problem*, IEEE Trans. Automat. Contr., AC-23, pp. 907–911.

M. Morf, G. S. Sidhu, and T. Kailath (1974), *Some new algorithms for recursive estimation in constant, linear, discrete-time systems*, IEEE Trans. Automat. Contr., 19, pp. 315–323.

V. H. Nascimento and A. H. Sayed (1999), *Optimal state regulation for uncertain state-space models*, Proc. American Control Conference, vol. 1, pp. 419–424, San Diego, CA.

F. D. Neeser and J. L. Massey (1993), *Proper complex random processes with applications to information theory*, IEEE Trans. Inform. Theory, IT-39, pp. 1293–1302.

Y. Nesterov and A. Nemirovskii (1994), *Interior-Point Polynomial Algorithms in Convex Programming*, Studies in Applied Mathematics, SIAM, PA.

R. Nevanlinna (1919), *Über beschränkte funktionen die in gegebenen punkten vorgeschreibene funktionswerte bewirkt werden*, Anal. Acad. Sci. Fenn., 13, ser. A., pp. 1–71.

J. A. Newkirk (1979), *Computational Issues in Linear Least-Squares Estimation and Control*, Ph.D. thesis, Dept. Electrical Engineering, Stanford University, CA.

G. C. Newton, L. A. Gould, and J. F. Kaiser (1957), *Analytical Design of Linear Feedback Controls*, Wiley, NY.

R. Nikoukhah, S. L. Campbell, and F. D elebecque (1999), *Kalman filtering for general discrete-time linear systems*, IEEE Trans. Automat. Contr., AC-44, pp. 1829–1839.

T. Nishimura (1966), *On the a priori information in sequential estimation problems*, IEEE Trans. Automat. Contr., AC-11, pp. 197–204.

B. Noble (1964), *The Numerical Solution of Nonlinear Integral Equations and Related Topics in Nonlinear Integral Equations*, Wisconsin Univ. Press, Madison, WI.

A. V. Oppenheim and R. W. Schafer (1998), *Discrete-Time Signal Processing*, second edition, Prentice Hall, NJ.

D. V. Ouellette (1981), *Schur complements and statistics*, Linear Algebra and Its Applications, 36, pp. 187–295.

C. C. Paige (1979a), *Computer solution and perturbation analysis of generalized linear least-squares problems*, Mathematics of Computation, 33, pp. 171–183.

C. C. Paige (1979b), *Fast numerically stable computation for generalized linear least-squares problems*, SIAM J. Numer. Anal., 16, pp. 165–171.

C. C. Paige (1985), *Covariance matrix representation in linear filtering*, in Proc. 1985 Joint Summer Research Conference on Linear Algebra and Its Role in Systems Theory, B. Datta, editor, 1985.

C. C. Paige and M. A. Saunders (1977), *Least-squares estimation of discrete linear dynamic systems using orthogonal transformations*, SIAM J. Numer. Anal., 14, pp. 180–193.

T. Pappas, A. J. Laub, and N. R. Sandell (1980), *On the numerical solution of the discrete-time algebraic Riccati equation*, IEEE Trans. Automat. Contr., 25, pp. 631–641.

P. Park and T. Kailath (1995a), *New square-root algorithms for Kalman filtering*, IEEE Trans. on Automat. Contr., 40, pp. 895–899.

P. Park and T. Kailath (1995b), *Square-root Bryson-Frazier smoothing algorithms*, IEEE Trans. Automat. Contr., 40, pp. 761–766.

R. V. Patel, A. J. Laub, and P. Van Dooren (1994), editors, *Numerical Linear Algebra Techniques for Systems and Control*, IEEE Press reprint volume, IEEE, NY.

E. L. Peterson (1961), *Statistical Analysis and Optimization of Systems*, Wiley, NY.

B. Picinbono (1993), *Random Signals and Systems*, Prentice Hall, NJ.

B. Picinbono (1994), *On circularity*, IEEE Trans. Signal Processing, SP-42, pp. 3473–3482.

G. Pick (1916), *Über die beschränkungen analytischer funktionen, welche durch vorgegebene funktionswerte bewirkt werden*, Math. Ann., 77, pp. 7–23.

S. U. Pillai and T. I. Shim (1993), *Spectrum Estimation and System Identification*, Springer-Verlag, NY.

V. F. Pisarenko and Y. A. Rozanov (1963), *On some problems for stationary processes reducing to equations of the Wiener-Hopf type*, Probl. Pered. Inform., 14, pp. 113–135 (in Russian).

R. L. Plackett (1972), *The discovery of the method of least-squares*, Biometrika, 59, pp. 239–251.

O. R. Polk and S. C. Gupta (1973), *Quasi-optimum digital phase-locked loops*, IEEE Trans. Communications, COM-21, pp. 75–82.

L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko (1962), *The Mathematical Theory of Optimal Processes*, Wiley, NY. [Translated from Russian edition.]

V. M. Popov (1964), *Hyperstability and optimality of automatic systems with several control functions*, Rev. Roum. Sci. Tech. Ser. Electrotech. Energ., 9, pp. 629–690.

V. M. Popov (1973), *Hyperstability of Control Systems*, Springer-Verlag, NY. [Translation of Romanian edition, 1966.]

B. M. Porat (1994), *Digital Processing of Random Signals: Theory and Methods*, Prentice Hall, NJ.

J. E. Potter (1966), *Matrix quadratic solutions*, SIAM J. Appl. Math., 14, pp. 496–501.

J. E. Potter and R. G. Stern (1963), *Statistical filtering of space navigation measurements*, in Proc. AIAA Guidance Contr. Conference.

C. R. Rao (1973), *Linear Statistical Inference and Its Applications*, 2nd edition, Wiley, NY.

H. E. Rauch (1962), *Linear estimation of sampled stochastic processes with random parameters*, Tech. Rep. 2108–1, Stanford Electronics Laboratory, Stanford University, Stanford, CA.

H. E. Rauch, F. Tung, and C. T. Striebel (1965), *Maximum likelihood estimates of linear dynamic systems*, AIAA J., 3, pp. 1445–1450.

R. M. Redheffer (1962), *On the relation of transmission-line theory to scattering and transfer*, J. Math. Phys., 41, pp. 1–41.

R. M. Redheffer and A. P. Wang (1973), *Formal properties of time-dependent scattering processes*, J. Math. Mech., vol. 19, pp. 765–781.

W. T. Reid (1972), *Riccati Differential Equations*, Academic, NY.

M. Ribaric (1973), *Functional-Analytic Concepts and Structures of Neutron Transport Theory*, Ljubljanan and the Institute Jozef Stefan, Ljubljana (edited by the Slovene Academy of Sciences and Arts).

J. F. Riccati (1724), *Animadversationnes in aequationes differentiales secundi gradus*, Eruditorum Quae Lipside Publicantur Supplementa, VIII, pp. 66–73.

F. Riesz and B. Sz.-Nagy (1990), *Functional Analysis*, Dover Publications, NY. [A republication of the original work published by Frederick Ungar Publishing Co., NY, 1955.]

J. Rissanen (1973), *Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with application to factoring positive matrix polynomials*, Mathematics of Computation, vol. 27, pp. 147–154.

J. Rissanen and T. Kailath (1972), *Partial realization of random systems*, Automatica, vol. 8, pp. 389–396.

E. A. Robinson (1963), *Structural properties of stationary stochastic processes with applications*, in Time Series Analysis, M. Rosenblatt, editor, pp. 170–196, Wiley, NY.

E. A. Robinson (1967), *Multichannel Time-Series Analysis with Digital Computer Programs*, Holden-Day, San Francisco, CA.

J. M. Rodriguez-Canabal (1973), *The geometry of the Riccati equation*, Stochastics, 1, pp. 129–149.

H. H. Rosenbrock (1965), *On the connection between discrete linear filters and some formulae of Gauss*, in Acta Congr. Automatique Théorique, Paris, Dunod.

H. L. Royden (1988), *Real Analysis*, 3rd edition, Macmillan, NY.

Y. A. Rozanov (1960), *Spectral properties of multivariate stationary processes and boundary properties of analytic matrices*, Theory Prob. Applics., vol. 5, pp. 362–376. [Reprinted in Kailath (1970).]

Y. A. Rozanov (1967), *Stationary Random Processes*, Holden-Day, San Francisco, CA.

A. Saberi, P. Sannuti, and B. M. Chen (1995), $\mathcal{H}^2$ *Optimal Control*, Prentice Hall, NJ.

A. P. Sage and J. L. Melsa (1971), *Estimation Theory with Applications to Communications and Control*, McGraw-Hill, NY.

D. Saint, X. C. Du, and G. Demoment (1985), *Image restoration using a non-causal state-space model and a fast 2-D Kalman filter,*, Proc. Conf. Mathematics in Signal Processing, Bath, UK.

S. Sandor (1959), *Sur l'equation differentielle matricielle de type Riccati*, Bull. Math. Soc. Sci. Math. Phys. R. P. Roumaine, 3, pp. 229–249.

A. H. Sayed (1992), *Displacement Structure in Signal Processing and Mathematics*, Ph.D. thesis, Department of Electrical Engineering, Stanford University, Stanford, CA.

A. H. Sayed (2000), *A framework for state-space estimation with uncertain models*, submitted for publication.

A. H. Sayed and S. Chandrasekaran (1998), *Estimation in the presence of multiple sources of uncertainties with applications*, Proc. Asilomar Conference on Signals, Systems, and Computers, vol. 2, pp. 1811–1815, Pacific Grove, CA.

A. H. Sayed and S. Chandrasekaran (2000), *Parameter estimation with multiple sources and levels of uncertainties*, to appear in IEEE Trans. Signal Processing.

A. H. Sayed, T. Constantinescu, and T. Kailath (1995), *Square-root algorithms for structured matrices, interpolation, and completion problems*, in vol. 69 of IMA Volumes in Mathematics and Its Applications, A. Bojanczyk and G. Cybenko, editors, Springer-Verlag, pp. 153–184.

A. H. Sayed, B. Hassibi, and T. Kailath (1996), *Inertia conditions for the minimization of quadratic forms in indefinite metric spaces*, in vol. 87 of Operator Theory: Advances and Applications, I. Gohberg, P. Lancaster, and P. N. Shivakumar, editors, Birkhauser, 1996, pp. 309–347.

A. H. Sayed and T. Kailath (1992), *Structured matrices and fast RLS adaptive filtering*, Proc. 2nd IFAC Workshop on Algorithms and Architectures for Real-Time Control, P. J. Fleming and W. H. Kwon, editors, Seoul, Korea, Pergamon Press, pp. 211–216.

A. H. Sayed and T. Kailath (1994a), *Extended Chandrasekhar recursions*, IEEE Trans. Automat. Contr., 39, pp. 619–623.

A. H. Sayed and T. Kailath (1994b), *A state-space approach to adaptive RLS filtering*, IEEE Signal Processing Magazine, 11, pp. 18–60.

A. H. Sayed and T. Kailath (1995), *Oblique state-space estimation algorithms*, Proc. American Control Conference, vol. 3, pp. 1969–1973, Seattle, WA.

A. H. Sayed and T. Kailath (2000), *A survey of spectral factorization algorithms*, to appear in *Applied and Computational Control, Signals and Circuits*, B. Datta, editor, Birkhäuser.

A. H. Sayed, T. Kailath, and H. Lev-Ari (1994), *Generalized Chandrasekhar recursions from the generalized Schur algorithm*, IEEE Trans. Automat. Contr., 39, pp. 2265–2269.

A. H. Sayed, T. Kailath, H. Lev-Ari, and T. Constantinescu (1994), *Recursive solutions of rational interpolation problems via fast matrix factorization*, Integral Equations and Operator Theory, vol. 20, pp. 84–118.

A. H. Sayed, H. Lev-Ari, and T. Kailath (1992), *Lattice filter interpretations of the Chandrasekhar recursions for estimation and spectral factorization*, Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, San Francisco, CA, vol. 4, pp. 549–552.

A. H. Sayed and V. H. Nascimento (1999), *Design criteria for uncertain models with structured and unstructured uncertainties*, in Robustness in Identification and Control, A. Garulli, A. Tesi, and A. Vicino, editors, vol. 245 of Lecture Notes in Control and Information Sciences, pp. 159–173, Springer-Verlag, London.

A. H. Sayed, V. H. Nascimento, and S. Chandrasekaran (1998), *Estimation and control with bounded data uncertainties*, Linear Algebra and Its Applications, vol. 284, pp. 259–306.

A. H. Sayed and M. Rupp (1997), *An l_2-stable feedback structure for nonlinear adaptive filtering and identification*, Automatica, 33, pp. 13–30.

F. H. Schlee, C. J. Standish, and N. F. Toda (1967), *Divergence in the Kalman filter*, AIAA Journal, 5, pp. 1114–1120.

H. W. Schmidt (1907), *Reflexion u. absortion von β strahlen*, Ann. Phys., 23, pp. 671–697.

S. F. Schmidt (1967), *Estimation of state with acceptable accuracy constraints*, Tech. Rep., Analytical Mechanics Associates, Inc., Interim Report 67–4.

S. F. Schmidt (1970), *Computational techniques in Kalman filtering*, in Theory and Applications of Kalman Filtering, AGARDograph 139, Tech. Rep., NATO Advisory Group for Aerospace Research and Development.

G. T. Schmidt (1976), editor, *Practical aspects of Kalman filtering implementation*, Tech. Rep. AGARD-LS–82, NATO Advisory Group for Aerospace Research and Development, London.

I. Schur (1917), *Über potenzreihen die im Inneren des Einheitskreises beschränkt sind*, Journal für die Reine und Angewandte Mathematik, 147, pp. 205–232. [English translation in *Operator Theory: Advances and Applications*, vol. 18, pp. 31–88, I. Gohberg, editor, Birkhäuser, Boston, 1986.]

L. Schwartz (1967), *Cours d'Analyse*, vol. II, Hermann, Paris.

F. C. Schweppe (1973), *Uncertain Dynamic Systems*, Prentice Hall, NJ.

M. Scott (1974), *A bibliography on invariant imbedding and related topics*, Tech. Rep. SLA–74–0284, Sandia Lab., Albuquerque, NM.

M. Shayman (1983), *Geometry of the algebraic Riccati equation, Parts I and II*, SIAM J. Control Optim., 21, pp. 375–394 and 395–409.

O. B. Sheynin (1977), *Laplace's theory of errors*, Archive for History of Exact Sciences, 17, pp. 1–61.

O. B. Sheynin (1977), *C. F. Gauss and the theory of errors*, Archive for History of Exact Sciences, 20, pp. 21–72.

M. Shinbrot (1957), *A generalization of a method for the solution of the integral equation arising in optimization of time-varying linear systems with nonstationary inputs*, IRE Trans. Inform. Theory, IT-3, pp. 220–225.

G. S. Sidhu and U. B. Desai (1976), *New smoothing algorithms based on reversed-time lumped models*, IEEE Trans. Automat. Contr., AC-21, pp. 538–541.

G. S. Sidhu and T. Kailath (1974), *Development of new estimation algorithms by innovations analysis and shift-invariance properties*, IEEE Trans. Inform. Theory, IT-20, pp. 759–762.

A. J. F. Siegert (1957), *A systematic approach to a class of problems in the theory of noise and other random phenomena, pt. II*, IRE Trans. Inform. Theory, IT-13, pp. 38–43.

A. J. F. Siegert (1958), *A systematic approach to a class of problems in the theory of noise and other random phenomena, pt. III*, IRE Trans. Inform. Theory, IT-4, pp. 4–14.

L. Silverman (1976), *Discrete Riccati equations*, in Control and Dynamic Systems, C. T. Leondes, editor, vol. 12, pp. 313–386, Academic Press, NY.

M. A. Singer and S. J. Hammarling (1983), *The algebraic Riccati equation: A summary review of some available results*, Tech. Rep. DITC 23/83, Nat. Phys. Lab..

D. Slepian and T. T. Kadota (1969), *Four integral equations of detection theory*, SIAM J. Appl. Math., 17, pp. 1102–1117.

D. T. M. Slock (1989), *Fast Algorithms for Fixed-Order Recursive Least-Squares Parameter Estimation*, Ph.D. thesis, Department of Electrical Engineering, Stanford University, Stanford, CA.

A. F. M. Smith and A. E. Gelfand (1992), *Bayesian statistics without tears: a sampling-resampling perspective*, Amer. Stat., 46, pp. 84–88.

F. Smithies (1965), *Integral Equations*, Cambridge University Press.

D. L. Snyder (1969), *The State-Variable Approach to Continuous Estimation with Applications to Analog Communications Theory*, MIT Press, Boston, MA.

V. V. Sobolev (1963), *A Treatise on Radiative Transfer*, Van Nostrand, Princeton, NJ. [Translated from the Russian original by S. I. Gaposchkin.]

T. Söderström (1994), *Discrete-Time Stochastic Systems: Estimation and Control*, Prentice Hall, Great Britain.

T. Söderström and P. G. Stoica (1983), *Instrumental Variable Methods for System Identification*, Springer-Verlag, Berlin, Heidelberg.

T. T. Soong (1965), *On a priori statistics in minimum variance estimation problems*, Trans. ASME, J. Basic Eng., series 87D, pp. 109–112.

H. W. Sorenson (1970), *Least-squares estimation: from Gauss to Kalman*, IEEE Spectrum, vol. 7, pp. 63–68.

H. W. Sorenson, editor, (1985), *Kalman Filtering : Theory and Application*, IEEE Press, NY.

H. W. Sorenson and D. L. Alspach (1971), *Recursive Bayesian estimation using Gaussian sums*, Automatica, 7, pp. 465–479.

M. Sorine (1977), *Sur les équations de Chandrasekhar associés au problème de contr^ole d'un système parabolique*, C. R. Aca. Sc., Paris, 285, pp. 863–865.

E. Soroka and U. Shaked (1984), *On the robustness of LQ regulators*, IEEE Trans. Automat. Contr., AC-29, pp. 664–665.

G. W. Stewart (1973), *Introduction to Matrix Computations*, Academic Press, NY.

G. W. Stewart (1995), *Theory of the Combination of Observations Least Subject to Errors*, Classics in Applied Mathematics, SIAM, PA. [Translation of original works by C. F. Gauss under the title *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae.*]

G. W. Stewart (1996), *Afternotes on Numerical Analysis*, SIAM, PA.

G. W. Stewart (1998), *Matrix Algorithms, vol. I: Basic Decompositions*, SIAM, PA.

G. C. Stokes (1862), *On the intensity of the light reflected from or transmitted through a pile of plates*, Proc. Roy. Soc., 11, pp. 545–556.

G. Strang (1988), *Linear Algebra and Its Applications*, 3rd edition, Academic Press, NY.

G. Strang (1993), *Introduction to Linear Algebra*, Wellesley-Cambridge Press, Wellesley, MA.

R. L. Stratonovich (1959), *On the theory of optimal non-linear filtration of random functions*, Th. Prob. and Its Appl., 4, pp. 223–225.

R. L. Stratonovich (1960a), *Application of the theory of Markov processes for optimum filtration of signals*, Radio Eng. Electron. Phys. (USSR), 1, pp. 1–19.

R. L. Stratonovich (1960b), *Conditional Markov processes*, Th. Prob. and Its Appl., 5, pp. 156–178.

R. L. Stratonovich (1966), *Conditional Markov Processes and Their Application to the Theory of Optimal Control*, Elsevier, NY, 1968. [Russin original in 1966.]

R. L. Stratonovich (1970), *Detection and estimation of signals in noise when one or both are non-gaussian*, Proc. IEEE, 58, pp. 670–679.

P. Swerling (1959), *First-order error propagation in a stagewise smoothing procedure for satellite observations*, J. Astronaut. Sci., 6, pp. 46–52.

P. Swerling (1968), *A proposed stagewise differential correction procedure for satellite tracking and prediction*, Tech. Rep., RAND Corp. Rep. P-1292.

P. Swerling (1971), *Modern state estimation methods from the viewpoint of the method of least-squares*, IEEE Trans. Automat. Contr., AC-16, pp. 707–719.

G. Szegö (1939), *Orthogonal Polynomials*, American Mathematical Society Colloquium Publishing, vol. 23, Providence, RI.

B. D. Tapley and C. Y. Choe (1976), *An algorithm for propagating the square-root covariance matrix in triangular form*, IEEE Trans. Automat. Contr., AC-21, pp. 122–123.

L. N. Trefethen and D. Bau (1997), *Numerical Linear Algebra*, SIAM, PA.

H. C. Van de Hulst (1963), *A new look at multiple scattering*, Tech. Rep., Goddard Inst. for Space Studies, NASA, NY.

P. Van Dooren (1981), *A generalized eigenvalue approach for solving Riccati equations*, SIAM J. Sci. St. Comp., 2, pp. 121–135.

P. Van Dooren and M. H. Verhaegen (1985), *On the use of unitary state-space transformations*, Contemporary Math., 47, pp. 447–463.

H. L. Van Trees (1968), *Detection, Estimation, and Modulation Theory, Part I*, Wiley, NY.

G. Verghese, B. Friedlander, and T. Kailath (1980), *Scattering theory and linear least-squares estimation, Part III: The estimates*, IEEE Trans. Automat. Contr., 25, pp. 779–794.

G. Verghese and T. Kailath (1979), *A further note on backwards Markovian models*, IEEE Trans. Inform. Theory, IT-25, pp. 121–124.

M. H. Verhaegen and P. Van Dooren (1986), *Numerical aspects of different Kalman filter implementations*, IEEE Trans. Automat. Contr., AC-31, pp. 907–917.

M. Vidyasagar (1993), *Nonlinear Systems Analysis*, 2nd edition, Prentice Hall, NJ.

G. Von Escherich (1898), *Die zweite variation der einfachen integrale*, Wiener Sitzungsberichte, 8, pp. 1191–1250.

J. E. Wall, A. S. Willsky, and N. R. Sandell (1981), *On the fixed-interval smoothing problem*, Stochastics, 5, pp. 1–42, 1981.

M. C. Wang and G. E. Uhlenbeck (1945), *On the theory of the Brownian motion, II*, Revs. Mod. Phys., 17, pp. 323–342.

J. Warrior and N. Viswanadham (1980), *Scattering theory and the discrete optimal control problem,* Int. J. Systems Science, vol. 11, no. 6, pp. 659–676, England.

D. S. Watkins (1991), *Fundamentals of Matrix Computation*, Wiley, NY.

N. Wax, editor, (1954) *Selected Papers on Noise and Stochastic Processes*, Dover, NY.

H. L. Weinert and U. B. Desai (1981), *On complementary models and fixed-interval smoothing*, IEEE Trans. Automat. Contr., AC–26, pp. 863–867.

P. Whittle (1963), *Prediction and Regulation by Linear Least-Square Methods*, Van Nostrand, Princeton, NJ.

P. Whittle (1983), *Prediction and Regulation by Linear Least-Square Methods*, 2nd edition, University of Minnesota Press, Minneapolis, MN. [1st edition pulished by Van Nostrand, Princeton, NJ, 1963.]

P. Whittle (1990), *Risk Sensitive Optimal Control*, Wiley, NY.

P. Whittle (1991), *Likelihood and cost as path integrals*, J. Roy. Statist. Soc., B53, pp. 505–529.

P. Whittle (1996), *Optimal Control: Basics and Beyond*, Wiley, NY.

B. Widrow and M. E. Hoff (1960), *Adaptive switching circuits*, IRE WESCON Conv. Rec., Pt. 4, pp. 96–104.

B. Widrow and S. D. Stearns (1985), *Adaptive Signal Processing*, Prentice Hall, NJ.

N. Wiener (1949), *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Technology Press and Wiley, NY. [Originally published in 1942 as a classified Nat. Defense Res. Council Report. Also published under the title *Time Series Analysis* by MIT Press, Cambridge, MA.]

N. Wiener and E. Hopf (1931), *On a class of singular integral equations*, Proc. Prussian Acad. Math. — Phys. Ser., p. 696.

N. Wiener and P. Masani (1957), *The prediction theory of multivariate processes, Part I: The regulatity condition*, Acta Math., 98, pp. 111–150.

N. Wiener and P. Masani (1958), *The prediction theory of multivariate processes, Part II: The linear predictor*, Acta Math., pp. 93–139.

R. A. Wiggins and E. A. Robinson (1965), *Recursive solution to the multichannel filtering problem*, J. Geophys. Res., 70, pp. 1885–1891.

J. C. Willems (1971), *Least squares stationary optimal control and the algebraic Rccati equation*, IEEE Trans. Automat. Contr., AC–18, pp. 621–634.

W. W. Willman (1969), *On the linear smoothing problem*, IEEE Trans. Automat. Contr., 14, pp. 116–117.

A. S. Willsky (1974), *Fourier series and estimation on the circle with applications to synchronous communication, part I (analysis), and part II (implementation)*, IEEE Trans. Inform. Theory, IT–20, pp. 577–590.

G. Wilson (1969), *Factorization of the covariance generating function of a pure moving average process*, SIAM J. Numer. Anal., vol. 6, pp. 1–7.

G. Wilson (1972), *The factorization of matricial spectral densities*, SIAM J. Appl. Math., vol. 23, pp. 420–426.

H. Wold (1954), *A Study in the Analysis of Stationary Time Series*, 2nd edition, Almqvist & Wiksell, Stockholm. [1st edition by Almqvist & Wiksell, Upssala, 1938.]

W. M. Wonham (1968a), *On the separation theorem of stochastic control*, SIAM J. Control Optim., 6, pp. 312–326.

W. M. Wonham (1968b), *On a matrix Riccati equation of stochastic control*, SIAM J. Control Optim., 6, pp. 681–697.

M. A. Woodbury (1950), *Inverting modified matrices*, Tech. Rep., Statistical Research Group, Princeton University, NJ, Memorandum Report no. 42.

A. Yaglom (1960), *Effective solutions of linear approximation problems for multivariate stationary processes with a rational spectrum*, Theory Prob. Applics., vol. 5, pp. 239–264. [Reprinted in Kailath (1970).]

A. Yaglom (1962), *Theory of Stationary Random Functions*, Prentice Hall, NJ. [Russian original, 1955.]

V. A. Yakubovic (1962), *The solution of certain matrix inequalities in automatic control theory*, Dokl. Akad. Nauk. SSSR, 143, pp. 1304–1307.

D. C. Youla (1961), *On the factorization of rational matrices*, IRE Trans. Inform. Theory, pp. 172–189.

D. C. Youla, J. J. Bongiorno, and H. A. Jabr (1976a), *Modern Wiener-Hopf design of optimal controllers. Part I: The single input-output case.* IEEE Trans. Automat. Contr., 21, pp. 3–13.

D. C. Youla, J. J. Bongiorno, and H. A. Jabr (1976b), *Modern Wiener-Hopf design of optimal controllers. Part II: The multivariable case.* IEEE Trans. Automat. Contr., 21, pp. 319–338.

U. G. Yule (1927), *On a method for investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers*, Philos. Trans. Roy. Soc. London, Ser. A, 226, pp. 267–298.

L. E. Zachrisson (1969), *On optimal smoothing of continuous-time Kalman processes*, Information Sciences, 1, pp. 143–172.

L. A. Zadeh and J. R. Ragazzini (1950), *An extension of Wiener's theory of prediction*, J. Appl. Phys., 21, pp. 645–655.

M. Zakai (1964), *General error criteria*, IEEE Trans. Inform. Theory, IT–10, pp. 94–95.

K. Zhou, J. C. Doyle, and K. Glover (1996), *Robust and Optimal Control*, Prentice Hall, NJ.

Author Index

D

E

F

G

H

I

J

K

L

M

Subject Index

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Y

Z