A Log Likelihood Gradient Evaluation by Using the Extended Square-Root Information Filter

M.V. Kulikova, I.V. Semushin

School of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3, Wits 2050, Johannesburg, South Africa

E-mail: mkulikova@cam.wits.ac.za

Introduction

The method of maximum likelihood . . .

Introduction

The method of maximum likelihood . . .

- Gradient-search optimization algorithms;

Introduction

The method of maximum likelihood . . .

- Gradient-search optimization algorithms;
- Gradient of the negative Log Likelihood Function;

Introduction

The method of maximum likelihood . . .

- Gradient-search optimization algorithms;
- Gradient of the negative Log Likelihood Function;
- Kalman Filter

Introduction

The method of maximum likelihood . . .

- Gradient-search optimization algorithms;
- Gradient of the negative Log Likelihood Function;
- Kalman Filter $\Longrightarrow$ is known to be unstable;

Introduction

The method of maximum likelihood . . .

- Gradient-search optimization algorithms;
- Gradient of the negative Log Likelihood Function;
- Kalman Filter $\Longrightarrow$ is known to be unstable;

[P. Park&T. Kailath, 1995] The extended Square Root Information Filter (eSRIF):

Introduction

The method of maximum likelihood . . .

- Gradient-search optimization algorithms;
- Gradient of the negative Log Likelihood Function;
- Kalman Filter $\Longrightarrow$ is known to be unstable;

[P. Park&T. Kailath, 1995] The extended Square Root Information Filter (eSRIF):

- avoids numerical instabilities arising from computational errors;

Introduction

The method of maximum likelihood . . .

- Gradient-search optimization algorithms;
- Gradient of the negative Log Likelihood Function;
- Kalman Filter $\Longrightarrow$ is known to be unstable;

[P. Park&T. Kailath, 1995] The extended Square Root Information Filter (eSRIF):

- avoids numerical instabilities arising from computational errors;
- appears to be better suited to parallel implementation and to very large scale integration (VLSI) implementation.

Problem Statement

Consider the discrete-time linear dynamic stochastic system

$$x_{t+1} = F_t x_t + G_t w_t, \qquad t = 0, 1, \ldots, N, \tag{1}$$

$$z_t = H_t x_t + v_t, \qquad t = 1, 2, \ldots, N, \tag{2}$$

Problem Statement

Consider the discrete-time linear dynamic stochastic system

$$x_{t+1} = F_t x_t + G_t w_t, \qquad t = 0, 1, \ldots, N, \tag{1}$$

$$z_t = H_t x_t + v_t, \qquad t = 1, 2, \ldots, N, \tag{2}$$

with the system state $x_t \in \mathbb{R}^n$,

Problem Statement

Consider the discrete-time linear dynamic stochastic system

$$x_{t+1} = F_t x_t + G_t w_t, \qquad t = 0, 1, \dots, N, \tag{1}$$

$$z_t = H_t x_t + v_t, \qquad t = 1, 2, \dots, N, \tag{2}$$

with the system state $x_t \in \mathbb{R}^n$, the state disturbance $w_t \in \mathbb{R}^q$,

Problem Statement

Consider the discrete-time linear dynamic stochastic system

$$x_{t+1} = F_t x_t + G_t w_t, \qquad t = 0, 1, \ldots, N, \tag{1}$$

$$z_t = H_t x_t + v_t, \qquad t = 1, 2, \ldots, N, \tag{2}$$

with the system state $x_t \in \mathbb{R}^n$, the state disturbance $w_t \in \mathbb{R}^q$, the observed vector $z_t \in \mathbb{R}^m$,

Problem Statement

Consider the discrete-time linear dynamic stochastic system

$$x_{t+1} = F_t x_t + G_t w_t, \qquad t = 0, 1, \ldots, N, \tag{1}$$

$$z_t = H_t x_t + v_t, \qquad t = 1, 2, \ldots, N, \tag{2}$$

with the system state $x_t \in \mathbb{R}^n$, the state disturbance $w_t \in \mathbb{R}^q$, the observed vector $z_t \in \mathbb{R}^m$, and the measurement error $v_t \in \mathbb{R}^m$,

Problem Statement

Consider the discrete-time linear dynamic stochastic system

$$x_{t+1} = F_t x_t + G_t w_t, \qquad t = 0, 1, \ldots, N, \tag{1}$$

$$z_t = H_t x_t + v_t, \qquad t = 1, 2, \ldots, N, \tag{2}$$

with the system state $x_t \in \mathbb{R}^n$, the state disturbance $w_t \in \mathbb{R}^q$, the observed vector $z_t \in \mathbb{R}^m$, and the measurement error $v_t \in \mathbb{R}^m$, such that the initial state x_0 and each w_t, v_t of $\{w_t : t = 0, 1, \ldots\}$, $\{v_t : t = 1, 2, \ldots\}$ are taken from mutually independent Gaussian distributions with the following expectations:

Problem Statement

$$\mathbf{E}\left\{\begin{bmatrix} x_0 \\ w_t \\ v_t \end{bmatrix}\right\} = \begin{bmatrix} \bar{x}_0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{E}\left\{\begin{bmatrix} (x_0 - \bar{x}_0) \\ w_t \\ v_t \end{bmatrix}\begin{bmatrix} (x_0 - \bar{x}_0) \\ w_t \\ v_t \end{bmatrix}^T\right\} = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & Q_t & 0 \\ 0 & 0 & R_t \end{bmatrix}$$

and $\mathbf{E}\left\{w_t w_{t'}^T\right\} = 0$, $\mathbf{E}\left\{v_t v_{t'}^T\right\} = 0$ if $t \neq t'$.

Problem Statement

$$\mathbf{E}\left\{\begin{bmatrix} x_0 \\ w_t \\ v_t \end{bmatrix}\right\} = \begin{bmatrix} \bar{x}_0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{E}\left\{\begin{bmatrix} (x_0 - \bar{x}_0) \\ w_t \\ v_t \end{bmatrix}\begin{bmatrix} (x_0 - \bar{x}_0) \\ w_t \\ v_t \end{bmatrix}^T\right\} = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & Q_t & 0 \\ 0 & 0 & R_t \end{bmatrix}$$

and $\mathbf{E}\left\{w_t w_{t'}^T\right\} = 0$, $\mathbf{E}\left\{v_t v_{t'}^T\right\} = 0$ if $t \neq t'$. Assume the system is parameterized by a vector $\theta \in \mathbb{R}^p$ of unknown system parameters. This means that all the above characteristics, namely F_t, G_t, H_t, $P_0 \geq 0$, $Q_t \geq 0$ and $R_t > 0$ can depend upon θ (the corresponding notations $F_t(\theta)$, $G_t(\theta)$ and so on, are suppressed for the sake of simplicity).

Problem Statement

For example, (1), (2) may describe a discrete autoregressive (AR) process observed in the presence of additive noise

$$x_{t+1} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \theta_1 & \theta_2 & \theta_3 & \dots & \theta_p \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \gamma \end{bmatrix} w_t,$$

$$z_t = \begin{bmatrix} 0 & 0 & \dots & 1 \end{bmatrix} x_t + v_t.$$

In this case θ are the unknown AR parameters which need to be estimated.

Problem Statement

The negative Log Likelihood Function (LLF) for system (1), (2) is given by

$$L_\theta\left(Z_1^N\right) = \frac{1}{2}\sum_{t=1}^{N}\left\{\frac{m}{2}\ln(2\pi) + \ln(\det(R_{e,t})) + e_t^T R_{e,t}^{-1} e_t\right\}$$

with $e_t \stackrel{\text{def}}{=} z_t - H_t\hat{x}_t$ being the zero-mean innovations whose covariance is determined as $R_{e,t} \stackrel{\text{def}}{=} E\{e_t e_t^T\} = H_t P_t H_t^T + R_t$ through matrix P_t, the error covariance of the time updated estimate $\hat{x}_t$ generated by the Kalman Filter.

Problem Statement

Let $l_\theta(z_t)$ denote the negative LLF for the t-th measurement z_t in system (1), (2), given measurements $Z_1^{t-1} \stackrel{\text{def}}{=} \{z_1, \ldots, z_{t-1}\}$,

Problem Statement

Let $l_\theta(z_t)$ denote the negative LLF for the t-th measurement z_t in system (1), (2), given measurements $Z_1^{t-1} \stackrel{\text{def}}{=} \{z_1, \ldots, z_{t-1}\}$, then

$$l_\theta(z_t) = \frac{1}{2}\left\{\frac{m}{2}\ln(2\pi) + \ln(\det(R_{e,t})) + e_t^T R_{e,t}^{-1} e_t\right\}. \qquad (3)$$

Problem Statement

Let $l_\theta(z_t)$ denote the negative LLF for the t-th measurement z_t in system (1), (2), given measurements $Z_1^{t-1} \stackrel{\text{def}}{=} \{z_1, \dots, z_{t-1}\}$, then

$$l_\theta(z_t) = \frac{1}{2}\left\{\frac{m}{2}\ln(2\pi) + \ln(\det(R_{e,t})) + e_t^T R_{e,t}^{-1} e_t\right\}. \quad (3)$$

By differentiating (3) we obtain

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[\ln(\det(R_{e,t}))\right] + \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[e_t^T R_{e,t}^{-1} e_t\right] \; i = 1, \dots p. \quad (4)$$

Problem Statement

Let $l_\theta(z_t)$ denote the negative LLF for the t-th measurement z_t in system (1), (2), given measurements $Z_1^{t-1} \stackrel{\text{def}}{=} \{z_1, \ldots, z_{t-1}\}$, then

$$l_\theta(z_t) = \frac{1}{2}\left\{\frac{m}{2}\ln(2\pi) + \ln(\det(R_{e,t})) + e_t^T R_{e,t}^{-1} e_t\right\}. \qquad (3)$$

By differentiating (3) we obtain

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[\ln(\det(R_{e,t}))\right] + \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[e_t^T R_{e,t}^{-1} e_t\right] i = 1, \ldots p. \quad (4)$$

As can be seen the computation of (3) and (4) leads to implementation of a Kalman Filter (and its derivative with respect to each parameter) which is known to be unstable.

The eSRIF

[P. Park&T. Kailath, 1995]: assume that $\Pi_0 > 0$, $R_t > 0$ and F_t are invertible. Given $P_0^{-T/2} = \Pi_0^{-T/2}$ and $P_0^{-T/2}\hat{x}_0 = \Pi_0^{-T/2}\bar{x}_0$, then

$$O_t \left[\begin{array}{ccc|c} R_t^{-T/2} & -R_t^{-T/2}H_tF_t^{-1} & R_t^{-T/2}H_tF_t^{-1}G_tQ_t^{T/2} & -R_t^{-T/2}z_t \\ 0 & P_t^{-T/2}F_t^{-1} & -P_t^{-T/2}F_t^{-1}G_tQ_t^{T/2} & P_t^{-T/2}\hat{x}_t \\ 0 & 0 & I_q & 0 \end{array}\right] = \left[\begin{array}{ccc|c} R_{e,t}^{-T/2} & 0 & 0 & -\bar{e}_t \\ -P_{t+1}^{-T/2}K_{p,t} & P_{t+1}^{-T/2} & 0 & P_{t+1}^{-T/2}\hat{x}_{t+1} \\ * & * & * & * \end{array}\right]$$

where O_t is any orthogonal transformation such that the matrix on the right-hand side of the formula is block lower triangular.

The eSRIF

The eSRIF is a modification of the conventional one:

$$O_t \begin{bmatrix} R_t^{-T/2} & -R_t^{-T/2} H_t F_t^{-1} & R_t^{-T/2} H_t F_t^{-1} G_t Q_t^{T/2} \\ 0 & P_t^{-T/2} F_t^{-1} & -P_t^{-T/2} F_t^{-1} G_t Q_t^{T/2} \\ 0 & 0 & I_q \end{bmatrix} = \begin{bmatrix} R_{e,t}^{-T/2} & 0 & 0 \\ -P_{t+1}^{-T/2} K_{p,t} & P_{t+1}^{-T/2} & 0 \\ * & * & * \end{bmatrix}$$

where O_t is any orthogonal transformation such that the matrix on the right-hand side of the formula is lower triangular.

The eSRIF

The eSRIF is a modification of the conventional one:

$$O_t \left[\begin{array}{ccc|c} R_t^{-T/2} & -R_t^{-T/2} H_t F_t^{-1} & R_t^{-T/2} H_t F_t^{-1} G_t Q_t^{T/2} & -R_t^{-T/2} z_t \\ 0 & P_t^{-T/2} F_t^{-1} & -P_t^{-T/2} F_t^{-1} G_t Q_t^{T/2} & P_t^{-T/2} \hat{x}_t \\ 0 & 0 & I_q & 0 \end{array} \right]$$

$$= \left[\begin{array}{ccc|c} R_{e,t}^{-T/2} & 0 & 0 & -\bar{e}_t \\ -P_{t+1}^{-T/2} K_{p,t} & P_{t+1}^{-T/2} & 0 & P_{t+1}^{-T/2} \hat{x}_{t+1} \\ * & * & * & * \end{array} \right]$$

where O_t is any orthogonal transformation such that the matrix on the right-hand side of the formula is

The eSRIF

Remark. The predicted estimate now can be found from the entries of the post-array by solving the triangular system

$$\left(\tilde{P}_{t+1}^{-1/2}\right)\left(\hat{x}_{t+1}^{-}\right) = \left(\tilde{P}_{t+1}^{-1/2}\hat{x}_{t+1}^{-}\right). \tag{5}$$

The LLG in terms of the eSRIF

The Log Likelihood Gradient (LLG):

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[\ln(\det(R_{e,t}))\right] + \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[e_t^T R_{e,t}^{-1} e_t\right], \quad i = 1, 2, \dots p,$$

The LLG in terms of the eSRIF

The Log Likelihood Gradient (LLG):

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[\ln(\det(R_{e,t}))\right] + \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[e_t^T R_{e,t}^{-1} e_t\right], \quad i = 1, 2, \ldots p,$$

in terms of the eSRIF is given by

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}\left[\ln(\det(R_{e,t}^{1/2}))\right] + \bar{e}_t{}^T \frac{\partial \bar{e}_t}{\partial \theta_i}, \quad i = 1, 2, \ldots p,$$

The LLG in terms of the eSRIF

The Log Likelihood Gradient (LLG):

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[\ln(\det(R_{e,t}))\right] + \frac{1}{2}\frac{\partial}{\partial \theta_i}\left[e_t^T R_{e,t}^{-1} e_t\right], \quad i = 1, 2, \ldots p,$$

in terms of the eSRIF is given by

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}\left[\ln(\det(R_{e,t}^{1/2}))\right] + \bar{e}_t{}^T \frac{\partial \bar{e}_t}{\partial \theta_i}, \quad i = 1, 2, \ldots p,$$

where $R_{e,t}^{1/2}$ is a square-root factor of the matrix $R_{e,t}$, i. e. $R_{e,t} = R_{e,t}^{T/2} R_{e,t}^{1/2}$, and $\bar{e}_t$ are the normalized innovations, i. e. $\bar{e}_t = R_{e,t}^{-T/2} e_t$.

The LLG in terms of the eSRIF

Taking into account that matrix $R_{e,t}^{1/2}$ is upper triangular, we can show

The LLG in terms of the eSRIF

Taking into account that matrix $R_{e,t}^{1/2}$ is upper triangular, we can show

$$\frac{\partial}{\partial \theta_i}\left[\ln\left(\det(R_{e,t}^{1/2})\right)\right] = \mathbf{tr}\left[R_{e,t}^{-1/2}\,\frac{\partial\left(R_{e,t}^{1/2}\right)}{\partial \theta_i}\right], \quad i = 1, 2, \ldots, p$$

where $\mathbf{tr}\left[\,\cdot\,\right]$ is a trace of matrix.

The LLG in terms of the eSRIF

Taking into account that matrix $R_{e,t}^{1/2}$ is upper triangular, we can show

$$\frac{\partial}{\partial\theta_i}\left[\ln\left(\det(R_{e,t}^{1/2})\right)\right] = \mathbf{tr}\left[R_{e,t}^{-1/2}\,\frac{\partial\left(R_{e,t}^{1/2}\right)}{\partial\theta_i}\right], \quad i = 1, 2, \ldots, p$$

where $\mathbf{tr}\,[\,\cdot\,]$ is a trace of matrix.

Finally, we obtain the expression for the LLG in terms of the eSRIF:

$$\frac{\partial l_\theta(z_t)}{\partial\theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2}\,\frac{\partial\left(R_{e,t}^{-1/2}\right)}{\partial\theta_i}\right] + \bar{e}_t^T\frac{\partial\bar{e}_t}{\partial\theta_i},\; i = 1, 2, \ldots, p. \quad (6)$$

LLG Evaluation

$$\frac{\partial l(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[\ R_{e,t}^{1/2}\ \ \frac{\partial\left(R_{e,t}^{-1/2}\right)}{\partial \theta_i}\right] + \ \bar{e}_t^T\ \ \frac{\partial \bar{e}_t}{\partial \theta_i},\ i = 1, \ldots, p. \qquad (6),$$

LLG Evaluation

$$\frac{\partial l(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2} \frac{\partial \left(R_{e,t}^{-1/2} \right)}{\partial \theta_i} \right] + \bar{e}_t^T \frac{\partial \bar{e}_t}{\partial \theta_i}, \; i = 1, \ldots, p. \qquad (6),$$

LLG Evaluation

$$\frac{\partial l(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2} \frac{\partial \left(R_{e,t}^{-1/2} \right)}{\partial \theta_i} \right] + \bar{e}_t^T \frac{\partial \bar{e}_t}{\partial \theta_i}, \quad i = 1, \ldots, p. \qquad (6),$$

The eSRIF:

$$O_t \left[\begin{array}{ccc|c} R_t^{-T/2} & -R_t^{-T/2} H_t F_t^{-1} & R_t^{-T/2} H_t F_t^{-1} G_t Q_t^{T/2} & -R_t^{-T/2} z_t \\ 0 & P_t^{-T/2} F_t^{-1} & -P_t^{-T/2} F_t^{-1} G_t Q_t^{T/2} & P_t^{-T/2} \hat{x}_t \\ 0 & 0 & I_q & 0 \end{array} \right]$$

$$= \left[\begin{array}{ccc|c} R_{e,t}^{-T/2} & 0 & 0 & -\bar{e}_t \\ -P_{t+1}^{-T/2} K_{p,t} & P_{t+1}^{-T/2} & 0 & P_{t+1}^{-T/2} \hat{x}_{t+1} \\ * & * & * & * \end{array} \right]$$

LLG Evaluation

$$\frac{\partial l(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2} \frac{\partial \left(R_{e,t}^{-1/2} \right)}{\partial \theta_i} \right] + \bar{e}_t^T \frac{\partial \bar{e}_t}{\partial \theta_i}, \quad i = 1, \ldots, p. \qquad (6),$$

The eSRIF:

$$O_t \left[\begin{array}{ccc|c} R_t^{-T/2} & -R_t^{-T/2} H_t F_t^{-1} & R_t^{-T/2} H_t F_t^{-1} G_t Q_t^{T/2} & -R_t^{-T/2} z_t \\ 0 & P_t^{-T/2} F_t^{-1} & -P_t^{-T/2} F_t^{-1} G_t Q_t^{T/2} & P_t^{-T/2} \hat{x}_t \\ 0 & 0 & I_q & 0 \end{array} \right]$$

$$= \left[\begin{array}{ccc|c} R_{e,t}^{-T/2} & 0 & 0 & -\bar{e}_t \\ -P_{t+1}^{-T/2} K_{p,t} & P_{t+1}^{-T/2} & 0 & P_{t+1}^{-T/2} \hat{x}_{t+1} \\ * & * & * & * \end{array} \right]$$

LLG Evaluation

$$\frac{\partial l(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2} \frac{\partial\left(R_{e,t}^{-1/2}\right)}{\partial \theta_i} \right] + \bar{e}_t^T \frac{\partial \bar{e}_t}{\partial \theta_i}, \; i = 1, \dots, p. \qquad (6),$$

The eSRIF:

$$O_t \left[\begin{array}{ccc|c} R_t^{-T/2} & -R_t^{-T/2} H_t F_t^{-1} & R_t^{-T/2} H_t F_t^{-1} G_t Q_t^{T/2} & -R_t^{-T/2} z_t \\ 0 & P_t^{-T/2} F_t^{-1} & -P_t^{-T/2} F_t^{-1} G_t Q_t^{T/2} & P_t^{-T/2} \hat{x}_t \\ 0 & 0 & I_q & 0 \end{array} \right]$$

$$= \left[\begin{array}{ccc|c} R_{e,t}^{-T/2} & 0 & 0 & -\bar{e}_t \\ -P_{t+1}^{-T/2} K_{p,t} & P_{t+1}^{-T/2} & 0 & P_{t+1}^{-T/2} \hat{x}_{t+1} \\ * & * & * & * \end{array} \right]$$

LLG Evaluation

Lemma 1. Let
$$QA = L \tag{7}$$
where Q is any orthogonal transformation such that the matrix on the right-hand side of formula (7) is lower triangular and A is a nonsingular matrix. If the elements of A are differentiable functions of a parameter θ then the upper triangular matrix U in

$$Q'_\theta Q^T = \bar{U}^T - \bar{U} \tag{8}$$

is, in fact, the strictly upper triangular part of the matrix $QA'_\theta L^{-1}$:

$$QA'_\theta L^{-1} = \bar{L} + D + \bar{U} \tag{9}$$

where $\bar{L}$, D and $\bar{U}$ are, respectively, strictly lower triangular, diagonal and strictly upper triangular.

LLG Evaluation

Lemma 1 easily leads to a method for computing L'_θ.

LLG Evaluation

Lemma 1 easily leads to a method for computing L'_θ.

By differenting $QA = L$ we obtain

$$L'_\theta = QA'_\theta + Q'_\theta A \tag{10}$$

LLG Evaluation

Lemma 1 easily leads to a method for computing L'_θ.

By differenting $QA = L$ we obtain

$$L'_\theta = QA'_\theta + Q'_\theta A \tag{10}$$

Multiply by L^{-1}:

$$L'_\theta L^{-1} = QA'_\theta L^{-1} + Q'_\theta A L^{-1}$$

LLG Evaluation

Lemma 1 easily leads to a method for computing L'_θ.

By differenting $QA = L$ we obtain

$$L'_\theta = QA'_\theta + Q'_\theta A \qquad (10)$$

Multiply by L^{-1}:

$$L'_\theta L^{-1} = QA'_\theta L^{-1} + Q'_\theta A L^{-1} \Rightarrow L'_\theta L^{-1} = QA'_\theta L^{-1} + Q'_\theta Q^T \qquad (11)$$

LLG Evaluation

Lemma 1 easily leads to a method for computing L'_θ.
By differenting $QA = L$ we obtain

$$L'_\theta = QA'_\theta + Q'_\theta A \tag{10}$$

Multiply by L^{-1}:

$$L'_\theta L^{-1} = QA'_\theta L^{-1} + Q'_\theta A L^{-1} \Rightarrow L'_\theta L^{-1} = QA'_\theta L^{-1} + Q'_\theta Q^T \tag{11}$$

By substituting (8) and (9) into (11), we find that

$$L'_\theta L^{-1} = \underbrace{(L + D + U)}_{QA'_\theta L^{-1}} + \underbrace{(U^T - U)}_{Q'_\theta Q^T}$$

LLG Evaluation

Lemma 1 easily leads to a method for computing L'_θ.

By differenting $QA = L$ we obtain

$$L'_\theta = QA'_\theta + Q'_\theta A \tag{10}$$

Multiply by L^{-1}:

$$L'_\theta L^{-1} = QA'_\theta L^{-1} + Q'_\theta A L^{-1} \Rightarrow L'_\theta L^{-1} = QA'_\theta L^{-1} + Q'_\theta Q^T \tag{11}$$

By substituting (8) and (9) into (11), we find that

$$L'_\theta L^{-1} = \underbrace{(L + D + U)}_{QA'_\theta L^{-1}} + \underbrace{(U^T - U)}_{Q'_\theta Q^T} \quad \Longrightarrow \quad L'_\theta = (L + D + U^T)L^{-1}$$

Algorithm LLG-eSRIF

I. For each θ_i, $i = 1, 2, \ldots, p$, apply the eSRIF

$$O_t \left[\begin{array}{ccc|c} R_t^{-T/2} & -R_t^{-T/2} H_t F_t^{-1} & R_t^{-T/2} H_t F_t^{-1} G_t Q_t^{T/2} & -R_t^{-T/2} z_t \\ 0 & P_t^{-T/2} F_t^{-1} & -P_t^{-T/2} F_t^{-1} G_t Q_t^{T/2} & P_t^{-T/2} \hat{x}_t \\ 0 & 0 & I_q & 0 \end{array} \right]$$

$$= \left[\begin{array}{ccc|c} R_{e,t}^{-T/2} & 0 & 0 & -\bar{e}_t \\ -P_{t+1}^{-T/2} K_{p,t} & P_{t+1}^{-T/2} & 0 & P_{t+1}^{-T/2} \hat{x}_{t+1} \\ * & * & * & * \end{array} \right]$$

where O_t is any orthogonal transformation such that the matrix on the right-hand side of the formula is block lower triangular.

Algorithm LLG-eSRIF

II. For each θ_i, $i = 1, 2, \ldots, p$, calculate

$$O_t \left[\begin{array}{ccc|c} \frac{\partial}{\partial\theta_i}\left(R_t^{-T/2}\right) & \frac{\partial}{\partial\theta_i}\left(S_t^{(1)}\right) & \frac{\partial}{\partial\theta_i}\left(S_t^{(2)}\right) & \frac{\partial}{\partial\theta_i}\left(S_t^{(3)}\right) \\ 0 & \frac{\partial}{\partial\theta_i}\left(S_t^{(4)}\right) & \frac{\partial}{\partial\theta_i}\left(S_t^{(5)}\right) & \frac{\partial}{\partial\theta_i}\left(S_t^{(6)}\right) \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$= \left[\begin{array}{ccc|c} X_i & Y_i & M_i & L_i \\ N_i & V_i & W_i & K_i \\ * & * & * & * \end{array}\right],$$

O_t is the same orthogonal transformation as in the eSRIF and

$$S_t^{(1)} = -R_t^{-T/2} H_t F_t^{-1}, \quad S_t^{(2)} = R_t^{-T/2} H_t F_t^{-1} G_t Q_t^{T/2}, \quad S_t^{(3)} = -R_t^{-T/2}$$

$$S_t^{(4)} = P_t^{-T/2} F_t^{-1}, \quad S_t^{(5)} = -P_t^{-T/2} F_t^{-1} G_t Q_t^{T/2}, \quad S_t^{(6)} = P_t^{-T/2} \hat{x}_t$$

Algorithm LLG-eSRIF

III. For each θ_i, $i = 1, 2, \ldots, p$, compute

$$J_i = \begin{bmatrix} X_i & Y_i & M_i \\ N_i & V_i & W_i \end{bmatrix} \begin{bmatrix} R_{e,t}^{-T/2} & 0 & 0 \\ -P_{t+1}^{-T/2} K_{p,t} & P_{t+1}^{-T/2} & 0 \\ * & * & * \end{bmatrix}^{-1}.$$

Algorithm LLG-eSRIF

IV. For each θ_i, $i = 1, 2, \ldots, p$, split the matrices

$$J_i = \overbrace{\left[\underbrace{L_i + D_i + U_i}_{m+n} \Bigg| \quad * * * \quad \right]}^{m+n+q} \Big\} {\scriptstyle m+n}$$

where L_i, D_i and U_i are the strictly lower triangular, diagonal and strictly upper triangular parts of J_i, respectively.

Algorithm LLG-eSRIF

V. For each θ_i, $i = 1, 2, \ldots, p$, compute the following quantities:

$$\begin{bmatrix} \dfrac{\partial R_{e,t}^{-T/2}}{\partial \theta_i} & 0 \\ -\dfrac{\partial\left(\tilde{P}_{t+1}^{-T/2} K_{p,t}\right)}{\partial \theta_i} & \dfrac{\partial \tilde{P}_{t+1}^{-T/2}}{\partial \theta_i} \end{bmatrix} = \left[L_i + D_i + U_i^T\right] \tag{12}$$

$$\times \begin{bmatrix} R_{e,t}^{-T/2} & 0 \\ -\tilde{P}_{t+1}^{-T/2} K_{p,t} & \tilde{P}_{t+1}^{-T/2} \end{bmatrix},$$

Algorithm LLG-eSRIF

$$\frac{\partial \bar{e}_t}{\partial \theta_i} = \left[\frac{\partial R_{e,t}^{-T/2}}{\partial \theta_i} - X_i \right] R_{e,t}^{T/2} \bar{e}_t + Y_i F_t \hat{x}_t - L_i, \qquad (13)$$

Algorithm LLG-eSRIF

$$\frac{\partial \bar{e}_t}{\partial \theta_i} = \left[\frac{\partial R_{e,t}^{-T/2}}{\partial \theta_i} - X_i \right] R_{e,t}^{T/2} \bar{e}_t + Y_i F_t \hat{x}_t - L_i, \qquad (13)$$

$$\frac{\partial S_{t+1}^{(6)}}{\partial \theta_i} = \left[\frac{\partial \left(\tilde{P}_{t+1}^{-T/2} K_{p,t} \right)}{\partial \theta_i} + N_i \right] R_{e,t}^{T/2} \bar{e}_t + \left[\frac{\partial \tilde{P}_{t+1}^{-T/2}}{\partial \theta_i} - V_i \right] F_t \hat{x}_t + K_i. \qquad (14)$$

Algorithm LLG-eSRIF

VI. Finally, we have all values to compute the LLG according to (6):

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2} \frac{\partial \left(R_{e,t}^{-1/2} \right)}{\partial \theta_i} \right] + \bar{e}_t^T \frac{\partial \bar{e}_t}{\partial \theta_i}, \; i = 1, \ldots, p.$$

Algorithm LLG-eSRIF

VI. Finally, we have all values to compute the LLG according to (6):

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2} \quad \frac{\partial \left(R_{e,t}^{-1/2} \right)}{\partial \theta_i} \right] + \bar{e}_t^T \quad \frac{\partial \bar{e}_t}{\partial \theta_i} \quad , \; i = 1, \ldots, p.$$

The LLG-eSRIF: Stage I

The LLG-eSRIF: Stages II — IV

Algorithm LLG-eSRIF

VI. Finally, we have all values to compute the LLG according to (6):

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2} \; \frac{\partial \left(R_{e,t}^{-1/2} \right)}{\partial \theta_i} \right] + \bar{e}_t^T \; \frac{\partial \bar{e}_t}{\partial \theta_i} \quad , \; i = 1, \ldots, p.$$

The LLG-eSRIF: Stage I

The LLG-eSRIF: Stages II — IV

Algorithm LLG-eSRIF

VI. Finally, we have all values to compute the LLG according to (6):

$$\frac{\partial l_\theta(z_t)}{\partial \theta_i} = -\,\mathbf{tr}\left[R_{e,t}^{1/2} \, \frac{\partial \left(R_{e,t}^{-1/2} \right)}{\partial \theta_i} \right] + \bar{e}_t^T \, \frac{\partial \bar{e}_t}{\partial \theta_i}, \quad i = 1, \ldots, p.$$

The LLG-eSRIF: Stage I

The LLG-eSRIF: Stages II — IV

Algorithm LLG-eSRIF

The LLG-eSRIF consists of two parts:

Algorithm LLG-eSRIF

The LLG-eSRIF consists of two parts:

- The source filtering algorithm, i.e. eSRIF.

Algorithm LLG-eSRIF

The LLG-eSRIF consists of two parts:

- The source filtering algorithm, i.e. eSRIF.
- The "differentiated" part: Stages II — IV.

Algorithm LLG-eSRIF

The LLG-eSRIF consists of two parts:

- The source filtering algorithm, i.e. eSRIF.
- The "differentiated" part: Stages II — IV.

Thus the Algorithm LLG-eSRIF is ideal for simultaneous

Algorithm LLG-eSRIF

The LLG-eSRIF consists of two parts:

- The source filtering algorithm, i.e. eSRIF.
- The "differentiated" part: Stages II — IV.

Thus the Algorithm LLG-eSRIF is ideal for simultaneous

- state estimation

Algorithm LLG-eSRIF

The LLG-eSRIF consists of two parts:

- The source filtering algorithm, i.e. eSRIF.
- The "differentiated" part: Stages II — IV.

Thus the Algorithm LLG-eSRIF is ideal for simultaneous

- state estimation
- parameter identification.

Algorithm LLG-eSRIF

Remark. Since, the matrices in LLG (6) are triangular, only the diagonal elements of $R_{e,t}^{1/2}$ and $\frac{\partial\left(R_{e,t}^{-1/2}\right)}{\partial\theta_i}$ need to be computed. Hence, the Algorithm LLG-eSRIF allows the $m \times m$-matrix inversion of $R_{e,t}$ to be avoided in the evaluation of LLG.

Ill-Conditioned Example Problems

Problem 1 Given:

$$P_0 = \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix}, H = \begin{bmatrix} 1, & 0 \end{bmatrix}, R = e^2\theta,\ F = I_2,\ Q = 0,$$

$$G = \begin{bmatrix} 0, & 0 \end{bmatrix}^T, 0< e \ll 1$$

where θ is an unknown parameter, I_2 is an identity 2×2 matrix; to simulate roundoff we assume

Ill-Conditioned Example Problems

Problem 1 Given:

$$P_0 = \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix}, H = \begin{bmatrix} 1, & 0 \end{bmatrix}, R = e^2\theta,\ F = I_2,\ Q = 0,$$

$$G = \begin{bmatrix} 0, & 0 \end{bmatrix}^T, 0 < e \ll 1$$

where θ is an unknown parameter, I_2 is an identity 2×2 matrix; to simulate roundoff we assume $e + 1 \neq 1$ but $e^2 + 1 \stackrel{r}{=} 1$.

Ill-Conditioned Example Problems

Problem 1 Given:

$$P_0 = \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix}, H = \begin{bmatrix} 1, & 0 \end{bmatrix}, R = e^2\theta,\ F = I_2,\ Q = 0,$$

$$G = \begin{bmatrix} 0, & 0 \end{bmatrix}^T, \text{0< e « 1}$$

where θ is an unknown parameter, I_2 is an identity 2×2 matrix; to simulate roundoff we assume $e + 1 \neq 1$ but $e^2 + 1 \stackrel{r}{=} 1$.

Calculate: $(P_1)'_\theta$ at the point $\theta = 1$.

Ill-Conditioned Example Problems

Problem 1 Given:

$$P_0 = \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix}, H = \begin{bmatrix} 1, & 0 \end{bmatrix}, R = e^2\theta, \; F = I_2, \; Q = 0,$$

$$G = \begin{bmatrix} 0, & 0 \end{bmatrix}^T, 0 < e \ll 1$$

where θ is an unknown parameter, I_2 is an identity 2×2 matrix; to simulate roundoff we assume $e + 1 \neq 1$ but $e^2 + 1 \stackrel{r}{=} 1$.

Calculate: $(P_1)'_\theta$ at the point $\theta = 1$.

For the $\theta = 1$, this example illustrates the initialization problems [Kaminski P.G., Bryson A.E., Schmidt S.F., 1971] that result when $H_1\Pi_0H_1^T + R_1$ is rounded to $H_1\Pi_0H_1^T$.

Comparison (Problem 1)

Filter	Exact Answer	Rounded Answe

Comparison (Problem 1)

Filter	Exact Answer	Rounded Answe
'diff' KF	$(P_1)'_\theta\big\rvert_{\theta=1} = \begin{bmatrix} \frac{e^2}{1+e^2} & 0 \\ 0 & 1 \end{bmatrix}$	$\stackrel{r}{=} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

Comparison (Problem 1)

Filter	Exact Answer	Rounded Answe
'diff' KF	$(P_1)'_\theta\Big\|_{\theta=1} = \begin{bmatrix} \frac{e^2}{1+e^2} & 0 \\ 0 & 1 \end{bmatrix}$	$\stackrel{r}{=} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$
'diff' IF	$(P_1^{-1})'_\theta\Big\|_{\theta=1} = -\begin{bmatrix} \frac{1+e^2}{e^2} & 0 \\ 0 & 1 \end{bmatrix}$	$\stackrel{r}{=} -\begin{bmatrix} \frac{1}{e^2} & 0 \\ 0 & 1 \end{bmatrix}$

Comparison (Problem 1)

Filter	Exact Answer	Rounded Answe
'diff' KF	$(P_1)'_\theta\Big\vert_{\theta=1} = \begin{bmatrix} \frac{e^2}{1+e^2} & 0 \\ 0 & 1 \end{bmatrix}$	$\overset{r}{=} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$
'diff' IF	$(P_1^{-1})'_\theta\Big\vert_{\theta=1} = -\begin{bmatrix} \frac{1+e^2}{e^2} & 0 \\ 0 & 1 \end{bmatrix}$	$\overset{r}{=} -\begin{bmatrix} \frac{1}{e^2} & 0 \\ 0 & 1 \end{bmatrix}$
LLG-eSRIF	$\left(P_1^{-1/2}\right)'_\theta\Big\vert_{\theta=1} = -\frac{1}{2}\begin{bmatrix} \frac{\sqrt{1+e^2}}{e} & 0 \\ 0 & 1 \end{bmatrix}$	$\overset{r}{=} -\frac{1}{2}\begin{bmatrix} \frac{1}{e} & 0 \\ 0 & 1 \end{bmatrix}$

Ill-Conditioned Example Problems

Problem 2 Given:

$$P_0 = \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix}, H = \begin{bmatrix} 1, & 1 \end{bmatrix}, R = e^2\theta,\ F = I_2,\ Q = 0,$$

$$G = \begin{bmatrix} 0, & 0 \end{bmatrix}^T, 0 < e \ll 1$$

where θ is an unknown parameter, I_2 is an identity 2×2 matrix; to simulate roundoff we assume

Ill-Conditioned Example Problems

Problem 2 Given:

$$P_0 = \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix}, H = \begin{bmatrix} 1, & 1 \end{bmatrix}, R = e^2\theta,\ F = I_2,\ Q = 0,$$

$$G = \begin{bmatrix} 0, & 0 \end{bmatrix}^T, 0 < e \ll 1$$

where θ is an unknown parameter, I_2 is an identity 2×2 matrix; to simulate roundoff we assume $e + 1 \neq 1$ but $e^2 + 1 \stackrel{r}{=} 1$.

Ill-Conditioned Example Problems

Problem 2 Given:

$$P_0 = \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix}, H = \begin{bmatrix} 1, & 1 \end{bmatrix}, R = e^2\theta,\ F = I_2,\ Q = 0,$$

$$G = \begin{bmatrix} 0, & 0 \end{bmatrix}^T, 0 < e \ll 1$$

where θ is an unknown parameter, I_2 is an identity 2×2 matrix; to simulate roundoff we assume $e + 1 \neq 1$ but $e^2 + 1 \overset{r}{=} 1$.

Calculate: $(P_1)'_\theta$ at the point $\theta = 1$.

Comparison (Problem 2)

Filter	Exact Answer	Rounded Answer

Comparison (Problem 2)

Filter	Exact Answer	Rounded Answer
'diff' KF	$\frac{1}{2+e^2}\begin{bmatrix} 1+e^2 & -1 \\ -1 & 1+e^2 \end{bmatrix}$	$\overset{r}{=} \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

Comparison (Problem 2)

Filter	Exact Answer	Rounded Answer
'diff' KF	$\frac{1}{2+e^2}\begin{bmatrix} 1+e^2 & -1 \\ -1 & 1+e^2 \end{bmatrix}$	$\stackrel{r}{=} \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$
'diff' IF	$-\frac{1}{e^2}\begin{bmatrix} 1+e^2 & 1 \\ 1 & 1+e^2 \end{bmatrix}$	$\stackrel{r}{=} -\frac{1}{e^2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

Comparison (Problem 2)

Filter	Exact Answer	Rounded Answer
'diff' KF	$\frac{1}{2+e^2}\begin{bmatrix} 1+e^2 & -1 \\ -1 & 1+e^2 \end{bmatrix}$	$\stackrel{r}{=} \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$
'diff' IF	$-\frac{1}{e^2}\begin{bmatrix} 1+e^2 & 1 \\ 1 & 1+e^2 \end{bmatrix}$	$\stackrel{r}{=} -\frac{1}{e^2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
LLG-eSRIF	$-\frac{1}{2}\begin{bmatrix} \sqrt{\frac{2+e^2}{1+e^2}} & \frac{1}{e\sqrt{1+e^2}} \\ 0 & \frac{\sqrt{1+e^2}}{e} \end{bmatrix}$	$\stackrel{r}{=} -\frac{1}{2}\begin{bmatrix} \sqrt{2} & \frac{1}{e} \\ 0 & \frac{1}{e} \end{bmatrix}$

Numerical Results

Example. Let the test system (1), (2) be defined as follows:

$$x_{t+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & e^{-\Delta t/\tau} \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_t,$$

$$z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t + v_t$$

where τ is a unknown parameter which needs to be estimated.

Numerical Results

Example. Let the test system (1), (2) be defined as follows:

$$x_{t+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & e^{-\Delta t/\tau} \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_t,$$

$$z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t + v_t$$

where τ is a unknown parameter which needs to be estimated. For the test problem, τ^*=10 was chosen as the true value of parameter τ.

Conclusion

In this paper, the new algorithm for evaluating the Log Likelihood Gradient (score) of linear discrete-time dynamic systems has been developed. The necessary theory has been given and substantiated by the computational experiments. Two ill-conditioned example problems have been constructed to show the superior perfomance of the Algorithm LLG-eSRIF over the conventional approach.