6.854J / 18.415J Advanced Algorithms
Fall 2008

18.415/6.854 Advanced Algorithms	October 8, 2008

Lecture 10

Lecturer: Michel X. Goemans

Last lecture we introduced the basic formulation of a linear programming problem, namely the problem with the objective of minimizing the expression $c^T x$ (where $c \in \mathbb{R}^n, x \in \mathbb{R}^n$) subject to the constraints $Ax = b$ where $A \in \mathbb{R}^{mxn}, b \in \mathbb{R}^m$) and $x \geq 0$. We then introduced the dual linear program, with the objective of maximizing $b^T y$, subject to the constraints that $A^T y \leq c$. Eventually, we were able to relate the two forms via the Theorem of Strong Duality, which states that if either the primal or the dual has a feasible solution then their values are equal:

$$w := \min\{c^T x : Ax = b, x \geq 0\} = \max\{b^T y : A^T y \leq c\} =: z.$$

Today, we further explore duality by justifying the Theorem of Strong Duality via a physical argument, introducing rules for constructing dual problems for non-standard linear programming formulations, and further discussing the notion of complementary slackness mentioned in the last lecture. We then shift gears and discuss the geometry of linear programming, which leads us to the Simplex Method of solving linear programs.

1 The Dual

1.1 Physical Justification of the Dual

Consider the standard dual form of a linear program. The set of feasible solutions y that satisfy the constraints $A^T y \leq c$ form a polyhedron in $\mathbb{R}^n$; this is the intersection of m halfspaces. Consider a tiny ball within this polyhedron at position y. To maximize $b^T y$, we move the ball as far as possible in the direction of b within the confines of our polyhedron. This is analogous to having a force, say gravity, acting on the ball in the b direction.

We now switch over entirely to the physical analogy. At equilibrium, the ball ends up at a point y maximizing $b^T y$ over $A^T y \leq c$, and the gravity force b is in equilibrium with the forces exerted against the ball by the 'walls' of our polyhedron. These wall forces are normal to the hyperplanes defining them, so for the hyperplane defined by $a_j^T y \leq c$ (where a_j is the jth column of A), the force exerted on the ball can be expressed as $-x_j a_j$ for some magnitude multiplier $x_j \geq 0$. As stated previously, our ball is at equilibrium (there is no net force on it), and so we find

$$b - \sum_j x_j a_j = 0.$$

We also note that for any wall which our ball is not touching, there is no force exerted by that wall on the ball. This is equivalent to saying

$$x_j = 0 \text{ if } a_j^T y < c_j.$$

We now argue that these multipliers x_j form an optimum solution to the primal linear program. We first note that

$$b - \sum_j x_j a_j = 0$$

is equivalent to $Ax = b$, and that the multipliers x_j are either zero or positive, and thus $x \geq 0$. This shows that our x_j's yield a feasible solution to the primal, now we need to prove that the x_j's

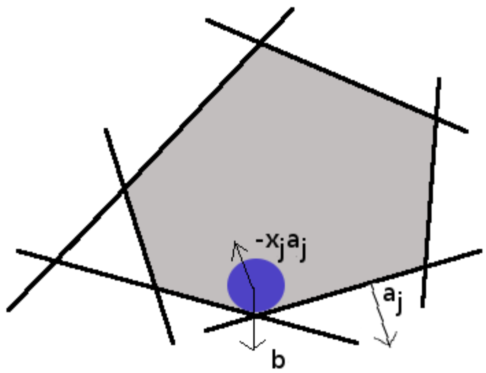

Figure 1: Physical visualization of the dual with $n = 2$ (two dimensions), $m = 6$ (six hyperplanes), and b as gravity. The dual is maximized when our $b^T y$ ball is at the lowest point of the polyhedron.

minimize the primal. For this, we will show that the value $c^T x$ equals $b^T y$, and therefore by weak duality, this will mean that x is a minimizer for the primal. The value $c^T x$ is:

$$c^T x = \sum_j c_j x_j = \sum_j (a_j^T y) x_j,$$

since x_j is non-zero only where $a_j^T y = c_j$ (a non-zero force is only exerted by a wall on our ball if the ball is touching that wall), and thus

$$c^T x = \sum_j (a_j^T y) x_j = y^T (\sum_j a_j x_j) = y^T b = b^T y.$$

1.2 Rules for Writing a Dual

So far, we have dealt only with the dual of the standard primal linear programming problem, minimizing $c^T x$ such that $Ax = b$ and $x \geq 0$. What if we are confronted with a non-standard linear program, such as a program that involves inequalities on the $a_{ij} x_j$, or non-positivity constraints on the x_j? We have two options. The first is to massage the linear program into the standard primal form, immediately convert to the standard dual, and then potentially massage the dual problem into a form more suitable to our original problem. This can be a long, frustrating process, however, and so instead we present a set of standard rules for converting any linear program into its dual form.

Consider a linear problem with the objective of minimizing $\sum_j c_j x_j$ subject to the following constraints:

$$\sum_j a_{ij} x_j \begin{cases} = b_i & i \in I_= \\ \geq b_i & i \in I_\geq \\ \leq b_i & i \in I_\leq \end{cases} \tag{1}$$

$$x_j \begin{cases} \geq 0 & j \in J_+ \\ \leq 0 & j \in J_- \\ \in \mathbb{R} & j \in J_0. \end{cases} \tag{2}$$

Earlier, the way we obtained the dual was to get a lower bound (or an upper bound if it was a maximization problem) on the objective function of the primal, and to maximize this upper bound. We claim that the same process leads to the dual of maximizing $\sum_i b_i y_i$ subject to the constraints:

$$\sum_i a_{ij} y_i \begin{cases} \leq c_j & j \in J_+ \\ \geq c_j & j \in J_- \\ = c_j & j \in J_0 \end{cases} \tag{3}$$

$$y_i \begin{cases} \geq 0 & i \in I_\geq \\ \leq 0 & i \in I_\leq \\ \in \mathbb{R} & i \in I_= \end{cases} \tag{4}$$

Weak duality is pretty straightforward. Constraints (4) on y_i guarantee that, when multiplying constraint (1) by y_i and summing them over i, we get

$$\sum_i y_i \sum_j a_{ij} x_j \geq \sum_i y_i b_i. \tag{5}$$

Similarly, constraints (3) together with constraints (2) imply that

$$\sum_j c_j x_j \geq \sum_j x_j \sum_i a_{ij} y_i. \tag{6}$$

The left-hand-side of (5) being equal to the right-hand-side of (6) (after rearranging the summation), we get weak duality that

$$c^T x \geq b^T y.$$

And strong duality also holds provided that either the primal or the dual has a feasible solution.

1.3 Complementary Slackness

Complementary slackness allows to easily check when a feasible primal and dual solutions are simultaneously optimal. Consider the primal

$$\min\{c^T x : Ax = b, x \geq 0\}.$$

Consider an alternative definition of the dual LP obtained by adding slack variables:

$$\max\{b^T y : A^T y + Is = c, s \geq 0\},$$

where $s \in \mathbb{R}^n$. Given a feasible primal solution x and a feasible dual solution (y, s), we see that the difference in their value is

$$c^T x - b^T y = s^T x + y^T Ax - y^T b = s^T x,$$

and this quantity better be 0 if x is optimum for the primal and (y, s) is optimal for the dual. Notice that $x \geq 0$ and $s \geq 0$, and therefore $x^T s = 0$ if and only if $x_j s_j = 0$ for all j. Thus, for the 2 solutions to be simultaneously optimum in the primal and in the dual, we need that, for all j, $x_j = 0$ whenever $s_j > 0$ (or equivalently that $s_j = 0$ whenever $x_j > 0$).

Summarizing, we have:

Theorem 1 *Let x^* be feasible in the primal, and (y^*, s^*) be feasible in the dual. Then the following are equivalent.*

1. *x^* is optimal in the primal, and (y^*, s^*) is optimal in the dual,*
2. *For all j: $x_j^* > 0 \implies s_j^* = 0$,*

3. For all j: $x_j^* s_j^* = 0$,

4. $\sum_j x_j^* s_j^* = 0$.

For a general pair of primal-dual linear programs as given in (1)-(2) and (3)-(4), complementary slackness says that, for x to be optimal in the primal and for y to be optimal in the dual, we must have that

1. $y_i = 0$ whenever $\sum_j a_{ij} x_j \neq b_i$ and,
2. $x_j = 0$ whenever $\sum_i a_{ij} y_i \neq c_j$.

2 The Geometry of Linear Programming

We now switch gears and discuss the geometry of linear programming. First, we define a polyhedral set $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ as the finite intersection of halfspaces. We then define a *vertex* of polyhedral set P to be any $x \in P$ such that $x + y \in P \wedge x - y \in P \implies y = 0$. Intuitively, a vertex is a "corner" of a polyhedral set. We can state this geometric definition also algebraically. Given an index set $J \subseteq \{1, 2, \cdots, n\}$, A_J denotes the $m \times |J|$ submatrix of A consisting of all columns of A indexed by J.

Lemma 2 *For $P = \{x : Ax = b, x \geq 0\}$ and $x \in P$, x is a vertex of P if and only if A_J has linearly independent colums for $J = \{j : x_j > 0\}$.*

Proof: For both directions, we prove the contrapositive.

$\Leftarrow$: Assuming x is not a vertex implies that $\exists y \neq 0 : x + y, x - y \in P$. Therefore $A(x + y) = b, A(x - y) = b$, which implies that $Ay = 0$. However, because membership in P requires points to be non-negative, we have that if $x_j = 0$ then $y_j = 0$. Thus, if we let $w = y_J$ (i.e. w corresponds to the components of y in J), we see that $w \neq 0$ and $A_J w = 0$, which implies that A_J has linearly dependent columns.

$\Rightarrow$: If A_J has linearly dependent columns, then $\exists w \neq 0 : A_J w = 0$. This implies you can construct a y via zero padding such that $y \neq 0$ and $Ay = 0, y_j = 0$ for $j \notin J$. Thus, $A(x + \epsilon y) = A(x - \epsilon y) = b$ for any $\epsilon \in \mathbb{R}$. We also note that $x_j \pm \epsilon y_j \geq 0$ if $\epsilon \leq \frac{x_j}{|y_j|}$, which is strictly greater than 0. Therefore, if we choose $\epsilon = \min_{j : y_j \neq 0} \frac{x_j}{|y_j|}$, we have that $x \pm \epsilon y \in P$, and thus x is a not a vertex of P. $\square$

We can take the notions in this lemma a step further by introducing the notions of a basis, a basic solution, and a basic feasible solution. For what follows, we assume that $rank(A) = m$ (if that's not the case, then either there is no solution to $Ax = b$ and our problem is infeasible, or there exists a redundant constraint (possibly more than one) in $Ax = b$ which can be removed).

Definition 1 *For a polyhedral set $P = \{x : Ax = b, x \geq 0\}$, a basis B is a subset of $\{1...n\}$ such that $|B| = m$ and A_B is invertible (i.e. $rank(A_B) = m$).*

Definition 2 *x is a basic solution of P if $\exists$ basis $B : x_B = A_B^{-1} b, x_N = 0$ for $N = \{1...n\} \setminus B$.*

Note that by this definition, $A_B x_B + A_N x_N = b$ must be true, but x could be negative and therefore infeasible.

Definition 3 *x is a basic feasible solution (bfs) if it is a basic solution such that $x \geq 0$.*

We are now ready to prove the following theorem relating vertices to basic feasible solutions.

Theorem 3 *Given a polyhedral set $P = \{x : Ax = b, x \geq 0\}$ such that $rank(A) = m$, and a point $x \in P$, x is a vertex of P if and only if it is a basic feasible solution of P.*

Proof: Will be provided in Lecture 11. □

There are several notable remarks to make pertaining to this theorem:

- The vertex to basic feasible solution relationship is one-to-many, or in other words, there may be multiple basic feasible solutions that correspond to a single vertex.
- The number of vertices of P is less than or equal to the number of bases of P. This follows from the first remark, and the fact that some bases may be infeasible. Therefore, the number of vertices of P is upper bounded by $\binom{n}{m}$. However, a stricter upper bound has been shown using a more detailed analysis, namely the number of vertices of P is upper bounded approximately by $\binom{n-\frac{m}{2}}{\frac{m}{2}}$.

We now know that finding basic feasible solutions of P is equivalent to finding vertices of P. Why is this important? Because there must an optimum solution to our linear programming problem that is a vertex of the polyhedral set defined by the linear constraints. More formally,

Theorem 4 *Given a polyhedral set $P = \{x : Ax = b, x \geq 0\}$, if $\min\{c^T x : x \in P\}$ is finite (the program is feasible and bounded), and $x \in P$, then $\exists$ vertex x' of $P : c^T x' \leq c^T x$.*

Proof: Will be provided in Lecture 11. □

This theorem directly leads us to the insight behind the Simplex Method for solving linear programs by finding the best vertex.

3 Sketch of the Simplex Method

Here is a very basic sketch of how the simplex method works.

1. Choose a basic feasible solution x corresponding to the basis B.
2. While x is not an optimal solution, choose j and k such that the new basis $B' = B \setminus \{j\} \cup \{k\}$ forms a bfs x' with $c^T x \leq c^T x$.

There are several important remarks to make about this method:

- It is not clear that j and k will always exist. But they do, and this can be shown.
- As defined, x and x' will either be equal or will be 'adjacent' vertices on P.
- The reason it is called a 'method' and not an algorithm is because we haven't specified yet how to choose j and k if several choices exist. The choice of j and k is referred to as a pivoting rule; many pivoting rules have been proposed.
- As such, there is no guarantee that $c^T x' < c^T x$, namely we could have $c^T x' = c^T x$; in fact we could even have $x' = x$ since we could switch from one basis to another representing the same vertex. There is therefore the risk that we repeat the same basis and the algorithm never terminates. And this can happen for some of the pivoting rules. There exist however *anticycling* pivoting rules which guarantee that the same basis is never repeated. With such a rule, the simplex method will terminate since there are finitely many bases.
- The running time of the simplex method depends on the number of bases considered before finding an optimal one.

- For all currently known pivoting rules, there is at least one instance that will cause the simplex method to run in exponential time. (This is in contrast with the simplex method in practice for which the number of iterations is usually good. A partial explanation of this sharp contrast between the worst-case behavior and a typical behavior is highlighted in the work of Spielman and Teng on smoothed analysis.)

We will cover other algorithms that will guarantee a polynomial running time in the worst-case; they will however not proceed from vertex to vertex of the polyhedral set.

There is a lower bound on the number of iterations of the Simplex Method, which is the number of edges in the path from the starting vertex of P to the optimum vertex of P. For a given P, this lower bound will be the *diameter* of P, the maximum over all pairs of vertices of the length of the shortest path between them. In 1957, Hirsch conjectured that the diameter of a polyhedral set is upper bounded by $n - d$, where d is the dimension of the space, and n is the number of hyperplanes defining P. While this has not been proven true in the general case, the following results have been found:

- The conjecture is not true in the unbounded case, namely there exist unbounded polyhedra with diameter $n - d + \lfloor \frac{d}{5} \rfloor$.
- No polynomial bound on the diameter is known for the general case (even for just bounded polyhedra).
- Kalai and Kleitman derived a subexponential bound $n^{O(\log d)}$ on the diameter.
- If the Hirsch Conjecture can be proven for $n = 2d$, then the conjecture holds for all n.
- The Hirsch Conjecture is true for polytopes with all their vertces in $\{0, 1\}^d$.