ORIE 6300 Mathematical Programming I November 25, 2008

Lecture 24

Lecturer: David P. Williamson *Scribe: Rolf Waeber*

Interior-Point Methods

The ellipsoid method is not practically efficient for large scale problems, though very important theoretically, especially in combinatorial optimization. It can be viewed as an existence proof for an efficient algorithm. It inspired the search for a practically efficient and theoretically polynomial time algorithm.

Interior-point methods were initially devised by Karmakar (1984) (although they are closely related to barrier methods used for linear and nonlinear programming since the 1950s). Since then, great development has led to more sophisticated interior-point methods that are competitive with (and are sometimes faster than) the simplex method.

Interior-point methods represent a significant development in the theory and practice of linear programming. They combine the advantages of the simplex method and of the ellipsoid algorithm. From a theoretical point of view, they lead to efficient (polynomial time) algorithms and use interesting geometric ideas; from a practical point of view, they allow the solution to large scale problems that arise in many applications.

Consider the standard form LP and its dual:

$$\begin{array}{rl} \min & c^T x \\ s.t. & Ax = b \\ & x \geq 0 \end{array} \qquad \begin{array}{rl} \max & b^T y \\ s.t. & A^T y \leq c. \\ & \end{array}$$

We assume A is an $m \times n$ matrix with rank m.

Note that $A^T y \leq c$ is equivalent to $A^T y + s = c, s \geq 0$. Therefore consider:

$$\begin{array}{rl} \min & c^T x \\ s.t. & Ax = b \\ & y \geq 0 \end{array} \qquad \begin{array}{rl} \max & b^T y \\ s.t. & A^T y + s = c \\ & s \geq 0. \end{array}$$

We define feasible regions:

$$\begin{aligned} \mathcal{F}(P) &= \{x \in R^n : Ax = b, \, x \geq 0\} \\ \mathcal{F}^\circ(P) &= \{x \in R^n : Ax = b, \, x > 0\} \\ \mathcal{F}(D) &= \{(y, s) \in R^m \times R^n : A^T y + s = c, \, s \geq 0\} \\ \mathcal{F}^\circ(D) &= \{(y, s) \in R^m \times R^n : A^T y + s = c, \, s > 0\} \end{aligned}$$

Interior-point methods generate a sequence of points in $\mathcal{F}^\circ(P)$ or in $\mathcal{F}^\circ(P) \times \mathcal{F}^\circ(D)$ converging to an optimal solution. In practice, we get with in 10^{-8} of optimal after 10-50 iterations. These iterations are more expensive either than a simplex pivot or an ellipsoid iteration. However, in $O(n \ln \frac{1}{\varepsilon})$ iterations, they come within a $(1 + \varepsilon)$ factor of the optimal value. Some interior-point methods only need $O(\sqrt{n} \ln \frac{1}{\varepsilon})$ iterations, but usually these algorithms work worse in practice.

One idea to generate this sequence: given a feasible point $\bar{x} \in \mathcal{F}^{\circ}(P)$, we want to "improve" it, using the "steepest descent" approach to computing the next iteration. The idea is, to find an improving direction $\bar{d}$, such that we keep the constraints $Ax = b$. We want $x = \bar{x} + \alpha\bar{d}$, such that

$$\begin{aligned} A(\bar{x} + \alpha\bar{d}) &= b \\ A\bar{x} + \alpha A\bar{d} &= b, \end{aligned}$$

which means we require $A\bar{d} = 0$. To make sure that α indeed controls the step length, we condition that $\|d\| \leq 1$ (so then $\|x - \bar{x}\| \leq \alpha$). We want $\bar{d}$ to be the "steepest descent" direction, so $\bar{d}$ should be the solution of the following, where we have $u = c$:

$$\begin{array}{ll} \min & u^T d \\ s.t. & Ad = 0 \\ & \|d\| \leq 1. \end{array}$$

Lemma 1 *Suppose u^T is not a linear combination of the rows of A ($\nexists y$ such that $A^T y = u$), then the solution to the optimal descent problem is:*

$$\bar{d} = -\frac{P_A u}{\| P_A u \|}$$

where

$$P_A = I - A^T(AA^T)^{-1}A.$$

Note that if there exists a y such that $A^T y = u = c$, then $(y, 0)$ is feasible for the dual. Since $\bar{x}$ is feasible for the primal, we have that $\bar{x}$ obeys complementary slackness with respect to y, which implies that $\bar{x}$ is optimal (and indeed any feasible x is optimal).

Proof: First, we check if everything in the lemma is well-defined. Since we assumed that A has full rank, AA^T is positive definite and therefore $(AA^T)^{-1}$ exists.

We start by showing that $P_A u \neq 0$. Suppose, for a contradiction, that

$$P_A u = 0$$

$$\Rightarrow (I - A^T(AA^T)^{-1}A)u = 0$$

$$\Rightarrow u = A^T(AA^T)^{-1}Au$$

Since u is now $A^T z$ for vector $z = (AA^T)^{-1}Au$, we see that u^T is a linear combination of the rows of A, which contradicts our assumption, and proves the claim.

Consider any d such that $Ad = 0$. Then since $P_A^T = I - A^T(AA^T)^{-1}A = P_A$, we have

$$(P_A u)^T d = u^T {P_A}^T d = u^T(I - A^T(AA^T)^{-1}A)d = u^T d$$

since $Ad = 0$.

So, rather than considering our original linear objective function, it is equivalent to solve the optimization problem

$$\begin{array}{ll} \min & (P_A u)^T d \\ s.t. & Ad = 0 \\ & \| d \| \leq 1. \end{array}$$

Suppose that we ignore the constraint $Ad = 0$ for the moment. Now, we are simply optimizing over the unit ball. By the Cauchy-Schwarz theorem, we know that

$$-\|x\|\|y\| \leq x^T y \leq \|x\|\|y\|.$$

Cauchy-Schwarz gives us a lower bound on the objective function:

$$\begin{aligned}(P_A u)^T d &\geq -\|P_A u\| \, \|d\| \\ &\geq -\|P_A u\| \, .\end{aligned}$$

So $-\|P_A u\|$ is the smallest objective function we can hope to get.

If we set

$$\bar{d} = -\frac{P_A u}{\|P_A u\|}$$

then

$$(P_A u)^T \bar{d} = -\frac{\|P_A u\|^2}{\|P_A u\|} = -\|P_A u\|,$$

so that this gives us the best possible objective function.

Also,

$$\begin{aligned}A\bar{d} &= -\frac{AP_A u}{\|P_A u\|} \\ &= -\frac{A(I - A^T(AA^T)^{-1}A)u}{\|P_A u\|} \\ &= -\frac{(A - AA^T(AA^T)^{-1}A)u}{\|P_A u\|} \\ &= -\frac{(A - A)u}{\|P_A u\|} \\ &= 0\end{aligned}$$

It follows that $\bar{d}$ optimizes the descent direction optimization problem. Furthermore, since $u^T \bar{d} = (P_A u)^T \bar{d} = -\|P_A u\| < 0$, it follows that $u^T(x + d) < u^T x$, so that taking a step in the direction of $\bar{d}$ improves the objective function value. $\square$

Note, that our steepest descent

$$\bar{d} = -\frac{P_A u}{\|P_A u\|}$$

does not depend on x. This is not a problem, as long as we are not close to the boundary (where here the boundary corresponds to the $x \geq 0$ constraints, since $Ax = b$ is always satisfied). For example, for $\bar{x} = e$ (the all 1's vector):

$$e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

(if e is feasible). Our next idea is to rescale the problem, such that we look like we are at e.

Let our current iterate at this step be x. Transform this to $\hat{x} = e$ by re-scaling as defined below. By the steepest descent step, get $\bar{d}$ in this transformed space. We then perform the inverse of our transform to map the new point back to our original space (as described below).

Given $\bar{x} \in \mathcal{F}^o(P)$, let

$$\bar{X} = Diag(\bar{x}) = \begin{bmatrix} \bar{x}_1 & 0 & \cdots & 0 \\ 0 & \bar{x}_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \bar{x}_n \end{bmatrix}$$

and consider the linear transformation,

$$x \to \hat{x} = \bar{X}^{-1}x.$$

This transforms $\bar{x}$ to e and our original optimization problem becomes

$$\begin{array}{rl} \min & c^T(\bar{X}\hat{x}) = (\bar{X}c)^T\hat{x} \\ s.t. & A(\bar{X}\hat{x}) = b \Rightarrow (A\bar{X})\hat{x} = b \\ & \hat{x} \geq 0. \end{array}$$

Therefore, in our transformed space, we compute the descent direction

$$\hat{d} = -\frac{P_{A\bar{X}}\bar{X}c}{\|P_{A\bar{X}}\bar{X}c\|}.$$

Note, $\hat{d}$ solves

$$\begin{array}{rl} \min & (\bar{X}c)^T d \\ s.t. & (A\bar{X})d = 0 \\ & \| d \| \leq 1. \end{array}$$

If we now map $\hat{d}$ from the transformed space back to our original space, we have derived our new descent direction:

$$\bar{d} = \bar{X}\hat{d} = -\frac{\bar{X}P_{A\bar{X}}\bar{X}c}{\|P_{A\bar{X}}\bar{X}c\|}.$$

Note, $\bar{d}$ solves

$$\begin{array}{rl} \min & c^T d \\ s.t. & Ad = 0 \\ & \|\bar{X}^{-1}d\| \leq 1. \end{array}$$

The direction $\bar{d}$ is called *affine-scaling direction* and was introduced by Dikin (1967). This direction gives a good algorithm, but it is not known if the algorithm terminates in polynomial time.

Note, that the direction of $\bar{d}$, i.e. $-\bar{x}P_{A\bar{X}}(\bar{X}c)$, is also the solution to the

$$\begin{array}{rl} \min & c^T d + \frac{1}{2}d^T\bar{X}^{-2}d = c^T d + \frac{1}{2}\|\bar{X}^{-1}d\|^2 \\ s.t. & Ad = 0. \end{array}$$

This shows the tradeoff between improvement in the objective function and the step length.

We can ask the question if there exists a function F such that $\nabla^2 F = \bar{X}^{-2}$? Note, $\bar{X}^{-2}$ has the simple form

$$\bar{X}^{-2} = \begin{bmatrix} \frac{1}{\bar{x}_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{x}_2^2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \frac{1}{\bar{x}_n^2} \end{bmatrix}.$$

Such a function exists. It is called the *logarithmic barrier function* F:

$$F(x) = -\sum_{j=1}^{n} \ln(x_j).$$

Next time we will look at the *central path*, points which minimize the function $c^T x + \mu F(x)$.