ORIE 6300 Mathematical Programming I December 2, 2008

Lecture 25

Lecturer: David P. Williamson *Scribe: Md. Saifur Rahman*

1 Interior Point Methods

Last time we introduced the *logarithmic barrier function*

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

$F(x)$ measures how central x is. The point x that minimizes $F(x)$ s.t. $Ax = b$ is called the *analytic center* of $P = \{x : Ax = b\}$.

Idea: Consider

$$B_\mu(x) := c^T x + \mu F(x)$$

for $\mu > 0$ defined on $\mathcal{F}^\circ(P) = \{x \in \mathbb{R}^n : Ax = b, x > 0\}$. The minimizer of $B_\mu(x)$ is near optimal for the original problem if μ is small, and near the analytic center of the feasible region if μ is very large. We are interested in minimizing $B_\mu(x)$ over $\mathcal{F}^\circ(P)$ and the following theorem gives necessary and sufficient conditions for the existence of such minimizers.

Theorem 1 *(a) For B_μ to have a minimizer on $\mathcal{F}^\circ(P)$, it is necessary and sufficient for $\mathcal{F}^\circ(P)$ and $\mathcal{F}^\circ(D) = \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^T y + s = c, s > 0\}$ to be nonempty.*

(b) If $\mathcal{F}^\circ(P)$ and $\mathcal{F}^\circ(D)$ are nonempty, a necessary and sufficient condition for $x \in \mathcal{F}^\circ(P)$ to be a unique minimizer of B_μ is that $\exists\ (y, s) \in \mathcal{F}^\circ(D)$ such that

$$\begin{aligned} A^T y + s &= c \\ Ax &= b \\ XSe &= \mu e \end{aligned} \tag{1}$$

where $X = diag(x)$, $S = diag(s)$, and $e = [1, 1, 1, ..., 1]^T \in \mathbb{R}^n$.

Insight: If we have $x, y, s, x \geq 0, s \geq 0$ for (1) and $\mu = 0$, then $XSe = \mu e \Rightarrow x_i s_i = \mu = 0, \forall i$, and hence x, y, s are optimal by the complementary slackness condition, since $x_j > 0 \Rightarrow s_j = 0 \Rightarrow$ the constraint i is met with equality.

Proof: (*sufficiency of (a):*) Let $\hat{x} \in \mathcal{F}^\circ(P)$, and $(\hat{y}, \hat{z}) \in \mathcal{F}^\circ(D)$. Then,

$$\begin{aligned} B_\mu(x) &= c^T x + \mu F(x) \\ &= (A^T \hat{y} + \hat{s})^T x + \mu F(x) \quad (\ \because\ \hat{y} = (interior)\ dual\ feasible \Rightarrow c^T = (A^T \hat{y} + \hat{s})^T) \\ &= \hat{y}^T A x + \hat{s}^T x + \mu F(x) \\ &= \hat{y}^T b + \hat{s}^T x + \mu F(x) \quad (\ \because x \in \mathcal{F}^\circ(P) \Rightarrow Ax = b) \\ &= b^T \hat{y} + \sum_j (\hat{s}_j x_j - \mu \ln(x_j)). \end{aligned}$$

Note that $\hat{s}_j x_j - \mu \ln(x_j) \to \infty$ as $x_j \to 0$ or $x_j \to \infty$. Thus, for all x_j, we can compute a lower bound, $\underline{x}_j$, and a upper bound, $\bar{x}_j$ that intersect with $\mathcal{F}^\circ(P)$ as shown in the Figure 1 below, i.e. $\exists\ \underline{x}_j$ and $\bar{x}_j$ such that $0 < \underline{x}_j < x_j < \bar{x}_j$ for all $x \in \mathcal{F}^\circ(P)$ such that $B_\mu(x) \leq B_\mu(\hat{x})$. Now, since B_μ is continuous function on a compact set $C = \{x \in \mathcal{F}^\circ(P) : \underline{x} \leq x \leq \bar{x}\}$, by Weierstrass's theorem, there exists a minimizer of B_μ on C, and by construction, this is also a minimizer over $\mathcal{F}^\circ(P)$.

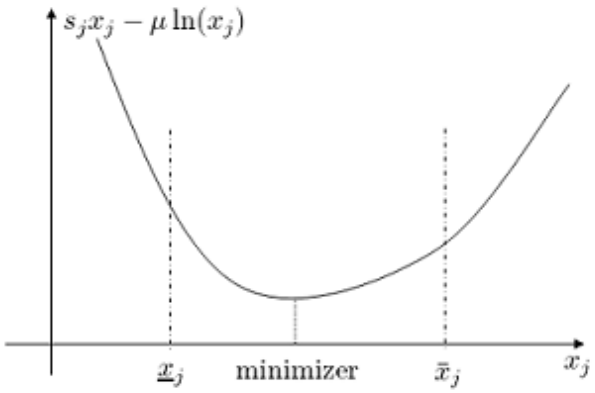

Figure 1: Plot of $\hat{s}_j x_j - \mu \ln(x_j)$ vs x_j

(*(b) and necessity of (a):*) Suppose x is the minimizer of B_μ over $\mathcal{F}^\circ(P)$. Then from the discussion on affine-scaling direction, $\exists\ y$ s.t. $A^T y = c + \mu \nabla F(x) = \nabla B_\mu(x)$, since otherwise there exists a direction to decrease $\nabla B_\mu(x)$. Hence, $\exists\ y$ s.t.

$$\begin{aligned} A^T y &= c + \mu \nabla F(x) \\ &= c + \mu(-X^{-1}e) \\ &= c - \mu \begin{bmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_n \end{bmatrix}. \end{aligned}$$

Now set

$$s = \mu \begin{bmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_n \end{bmatrix} > 0$$

Then since $A^T y + s = c$, this implies $(y, s) \in \mathcal{F}^\circ(D)$. Moreover, we have $x_i s_i = \mu$ for all i, so that $XSe = \mu e$. Thus, (1) holds.

Finally, we shall show that if (1) holds for $x \in \mathcal{F}^\circ(P)$, and $(y, s) \in \mathcal{F}^\circ(D)$, then x is the minimizer of B_μ. Consider a function

$$G(x) = (c - A^T y)^T x + \mu F(x).$$

The gradient of $G(x)$ is

$$\begin{aligned}
\nabla G(x) &= c - A^T y + \mu \nabla F(x) \\
&= c - A^T y - \mu \begin{bmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_n \end{bmatrix} \\
&= c - A^T y - s \\
&= 0,
\end{aligned}$$

by the fact that $(y, s) \in \mathcal{F}^\circ(D)$. Hence, $G(x)$ has a zero gradient at x. Since both $F(x)$ and $G(x)$ are convex functions, and thus have a unique minimizer, x becomes a unique minimizer of $G(x)$ over $\mathcal{F}^\circ(P)$. Now, since $Ax = b$ over $\mathcal{F}^\circ(P)$, we have

$$\begin{aligned}
G(x) &= c^T x - y^T A x + \mu F(x) \\
&= B_\mu(x) - y^T b.
\end{aligned}$$

Thus, over $\mathcal{F}^\circ(P)$, $B_\mu(x)$ and $G(x)$ differ only by a fixed constant, so minimizing $G(x)$ is equivalent to minimizing $B_\mu(x)$. Therefore, x is a minimizer of B_μ over $\mathcal{F}^\circ(P)$.

□

Given some μ, we define the solution (x, y, s) to the system (1) as points on the *central path*. Let $(x(\mu), y(\mu), s(\mu))$ denote solution for a given μ, then $\{x(\mu) : \mu > 0\}$ is called the *primal central path*, and $\{x(\mu), y(\mu), s(\mu) : \mu > 0\}$ is called the *primal-dual central path*. We now proceed to find the solution to the system (1). Note that the first two sets of equations define a system that are linear but the last is quadratic.

2 Newton's Method

Our method for solving this system (1) is based on Newton's method for finding a root of a function. In the one-dimensional, unconstrained case, we have a function $f(x)$, and we begin with an initial point x_0. We then repeatedly update the point. In iteration k, the tangent line to f at the current point x_k is described by $y = f'(x_k)(x - x_k) + f(x_k) = f'(x_k)\Delta x + f(x_k)$, where $\Delta x = x - x_k$. We find a value of Δx such that $y = 0$, and set $x_{k+1} \leftarrow x_k + \Delta x$. Figure 2 below shows an example of an update in Newton's method. We repeat this process until the value $f(x_k)$ of the function at the current point is sufficiently close to zero.

Interior-point methods apply the same approach to the system (1). We define a function $F(x, y, s)$ of the primal and dual solutions (x, y, s) to the linear programs as

$$F(x, y, s) = \begin{pmatrix} A^T y + s - c \\ Ax - b \\ XSe - \mu e. \end{pmatrix}$$

Our goal is to find solutions (x, y, s) such that $F(x, y, s) = 0$ and $(x, s) \geq 0$. To do this, we make use of the Jacobian J, which is a matrix of partial derivatives of F.

$$J(x, y, s) = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix}$$

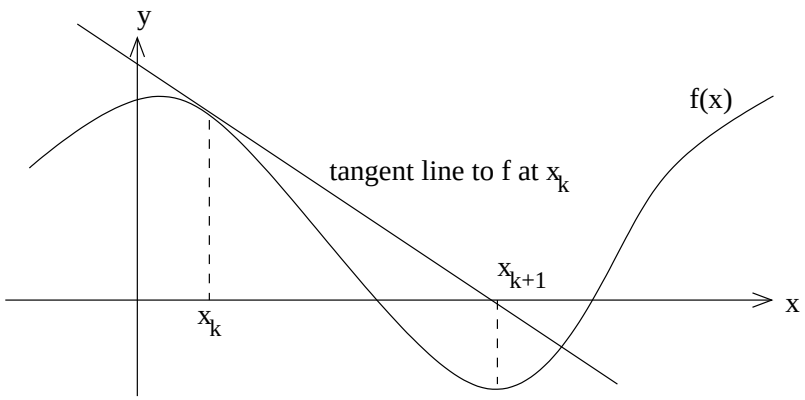

Figure 2: An update by Newton's method in the one-dimensional, unconstrained case.

Given these definitons, a *Newton direction* $(\Delta x, \Delta y, \Delta s)$ is a solution to the following equation:

$$J(x, y, s) \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} + F(x, y, s) = 0.$$

Over the course of updating the solutions (x, y, s) to the linear programs, we will reduce μ, causing our solutions to converge to optimal solutions to the linear programs. Suppose that in some iteration, the current solutions are (x, y, s). We consider the choice $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i s_i = \frac{1}{n} x^T s$; that is, given our current solution, solving the system for this value of μ makes the $x_i s_i$ equal for all i. To balance the movement towards the central path against the movement toward optimal solutions, we maintain a *centering parameter* $\sigma \in [0, 1]$. If $\sigma = 1$, then our update will move towards the center of the feasible region. On the other hand, if $\sigma = 0$, then our update step is in the direction of optimal solutions to the linear programs. A step with $\sigma = 1$ is referred to as a centering step, and a step with $\sigma = 0$ is referred to as an affine-scaling step. The choice of the centering parameter σ provides us with a trade-off between moving towards the central path and moving toward optimal solutions to the linear programs.

We assume we have (x, y, s) so that $x \in \mathcal{F}^\circ(P)$ and $(y, s) \in \mathcal{F}^\circ(D)$. We update our current solutions in the direction $(\Delta x, \Delta y, \Delta s)$ that solves the following linear system:

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -XSe + \sigma\mu e \end{pmatrix}$$

Some interior-point methods examine only the feasible region of the primal linear program, and as such are known as pure primal interior-point methods. The interior-point algorithm that we presented here examines the feasible regions of both the primal and the dual linear programs, and so it is referred to a *primal-dual interior-point algorithm.*

Primal-Dual Interior-Point Algorithm

$(x^0, y^0, s^0) \leftarrow$ initial feasible point $(x^0, s^0 > 0)$
$\mu^0 \leftarrow \frac{1}{n}(x^0)^T s^0$
$k \leftarrow 0$
While $\mu^k > \epsilon$

$$\text{Solve } \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k S^k e + \sigma^k \mu^k e \end{pmatrix}$$

$(x^{k+1}, y^{k+1}, s^{k+1}) \leftarrow (x^k, y^k, s^k) + \alpha^k(\Delta x^k, \Delta y^k, \Delta s^k)$
where α^k is some scaling parameter such that $x^{k+1}, s^{k+1} > 0$
$\mu^{k+1} \leftarrow \frac{1}{n}(x^{k+1})^T s^{k+1}$
$k \leftarrow k + 1$

This general framework omits several details that must be addressed in any implementation of an interior-point algorithm for linear programming. In particular, we have not specified how the centering parameter σ^k is chosen, as different interior-point algorithms use different methods to select σ^k. Furthermore, other important issues to be addressed are: how to set the threshold ϵ, when do we terminate, how do we know that we are making progress, how to find an initial solution $(x^\circ, y^\circ, s^\circ)$, and how to choose the scaling parameter α^k. We will discuss some of this in the next lecture.