Lecture 22

Lecturer: David P. Williamson — *Scribe: Pu Yang*

1 Interior-Point Methods

1.1 Barrier Function Minimization & Introduction of Central Path

Recall the primal and dual LP we considered in last class:

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

We defined the interior of primal feasible region P and dual feasible region D as follows:

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

We also introduced the *logarithmic barrier function*:

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

which is defined on $\mathcal{F}^\circ(P)$. $F(x)$ measures how central x is. The point x that minimizes $F(x)$ over $\mathcal{F}^\circ(P)$ is called the *analytical center* of P.

Consider the function $B_\mu(x) = c^T x + \mu F(x)$ for $\mu > 0$, defined on $\mathcal{F}^\circ(P)$. The $x \in \mathcal{F}^\circ(P)$ minimizing $B_\mu(x)$ is close to the anlytical center when μ is large, and close to the optimal solution to $c^T x$ when μ is small. So as $\mu \to 0$, $x \in P$ minimizing $B_\mu(x)$ converges to the optimal solution to $c^T x$. We can use this idea to find the optimal solution to $c^T x$.

We need first to check whether a minimizer of $B_\mu(x)$ exists on $\mathcal{F}^\circ(P)$. The following theorem gives necessary and sufficient conditions for the existence of such a minimizer.

Theorem 1

(i) *For B_μ to have a minimizer on $\mathcal{F}^\circ(P)$, it is necessary and sufficient for $\mathcal{F}^\circ(P)$ and $\mathcal{F}^\circ(D)$ to be non-empty.*

(ii) *If $\mathcal{F}^\circ(P)$ and $\mathcal{F}^\circ(D)$ are non-empty, a necessray and sufficient condition for $x \in \mathcal{F}^\circ(P)$ to be the 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* $e = (1, \dots, 1)^T \in \mathbb{R}^n$*.*

The last condition in (ii) is equivalent as $x_j s_j = \mu, \forall j$. When $\mu = 0$, this condition is equivalent that x, y, s satisfy the complementery slackness condition, which ensures (x, y, s) to be optimal.

Proof: We first prove sufficiency of (i). Assume $\exists \hat{x} \in \mathcal{F}^\circ(P)$, $(\hat{y}, \hat{s}) \in \mathcal{F}^\circ(D)$, then:

$$\begin{aligned} B_\mu(x) &= c^T x + \mu F(x) \\ &= (A^T \hat{y} + \hat{s})^T x + \mu F(x) \\ &= \hat{y}^T A x + \hat{s}^T x + \mu F(x) \\ &= \hat{y}^T b + \hat{s}^T x + \mu F(x) \\ &= \hat{y}^T b + \sum_{j=1}^{n} (\hat{s}_j x_j - \mu \ln x_j) \end{aligned}$$

Note that $\forall j$, $\hat{s_j} x_j - \mu \ln x_j \to \infty$ as $x_j \to 0$ or $x_j \to \infty$ (Figure 1). Therefore, for each j, we can find $\underline{x}_j > 0$ and $\overline{x}_j > 0$ such that for all $x \in \mathcal{F}^\circ(P)$ s.t. $B_\mu(x) \le B_\mu(\hat{x})$, x satisfies $0 < \underline{x}_j \le x_j \le \overline{x}_j$, $\forall j$. Since B_μ is a continuous function over a non-empty, closed and bounded set $C = \{x \in \mathcal{F}^\circ(P) : \underline{x} \le x \le \overline{x}\}$, by Weierstrass's theorem, there exists a minimizer of B_μ on C, and by construction, this is also a minimizer over $\mathcal{F}^\circ(P)$.

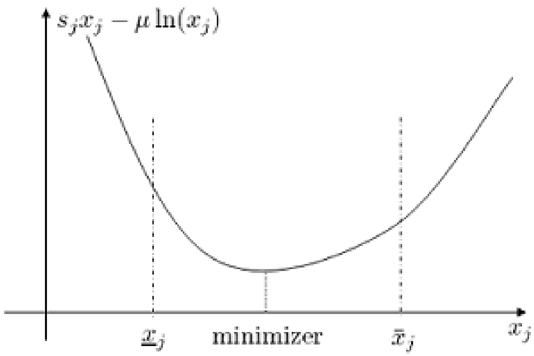

Figure 1: Plot of $\hat{s}_j x_j - \mu \ln x_j$ vs x_j

Next, we prove necessity of (i) and (ii).

Suppose x is a minimizer of B_μ on $\mathcal{F}^\circ(P)$, then there exists y s.t. $A^T y = c + \mu \nabla F(x) = \nabla B_\mu(x)$, since otherwise by Lemma 1 of last lecture, there would exist a direction to decrease $B_\mu(x)$ along $\nabla B_\mu(x)$.

Thus, $\exists y$ s.t.

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

Now set $s = \mu \begin{pmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_n \end{pmatrix} > 0.$

Since $A^T y + s = c$, this implies $(y, s) \in \mathcal{F}^\circ(D)$. Thus both $\mathcal{F}^\circ(P)$ and $\mathcal{F}^\circ(D)$ are non-empty. Moreover, we have $x_j s_j = \mu$ for all j, so that $XSe = \mu e$, which shows (1) holds.

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

Consider the function:

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

The gradient of $G(x)$ is:

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

implied by the fact $(y, s) \in \mathcal{F}^\circ(D)$.

Since G is a convex function over $\mathcal{F}^\circ(P)$, then x is the unique minimizer of G over that region. Also, by $x \in \mathcal{F}^\circ(P)$ we have $Ax = b$ and:

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

Thus, $B_\mu(x)$ and $G(x)$ differ only by a constant $y^T b$ over $\mathcal{F}^\circ(P)$. Hence minimizing B_μ is equivalent as minimizing G and x is the unique minimizer of B_μ over $\mathcal{F}^\circ(P)$. $\square$

Let $x(\mu)$, $y(\mu)$, $s(\mu)$ be solutions to (1) for some fixed 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 have as $\mu \to 0$, the central path will converge to an optimal solution of the original LP. In next lecture, we're going to talk about some "path-following" methods in which we follow a central path to find an optimal solution to the LP.

1.2 Potential Function Reduction Method

Now we turn our steer towards another important class of interior-point method: the methods of potential function reduction. We're going to pick some potential function $G(x, s)$ with the property that when $G(x, s)$ is sufficiently small, (x, y, s) must be close to an optimal solution. The idea of the potentail reduction method is to start with some feasible (x, y, s) and try to decrease the potential function G in each iteration until G is small enough so that we are close to an optimal solution.

One choice of G is:

$$\begin{aligned} G_q(x,s) &= q\ln(x^Ts) + F(x) + F(s) \\ &= q\ln(x^Ts) - \sum_{j=1}^n \ln x_j - \sum_{j=1}^n \ln s_j \\ &= q\ln(x^Ts) - \sum_{j=1}^n \ln x_j s_j \end{aligned}$$

for some q.

What q shall we choose? The following lemma shows we can not choose q too small.

Lemma 2 *If* $q = n$, *then* $G_n(x,s) \geq n\ln n$.

Proof: Recall that for a list of values $t_1, \ldots, t_n$ where $t_i > 0\ \forall i$, the arithmatic mean is always no less than the geometric mean, i.e.

$$\begin{aligned} & \left(\prod_{j=1}^n t_j\right)^{\frac{1}{n}} \leq \frac{1}{n}\left(\sum_{j=1}^n t_j\right) \\ \Rightarrow\ & \frac{1}{n}\sum_{j=1}^n \ln t_j \leq \ln\left(\sum_{j=1}^n t_j\right) - \ln n \\ \Rightarrow\ & n\ln\left(\sum_{j=1}^n t_j\right) - \sum_{j=1}^n \ln t_j \geq \ln n \end{aligned}$$

Let $t_j = s_j x_j$, and we have the desired result. □

Keep in mind we want that when $G(x,s)$ is sufficiently small, (x,y,s) is close to an optimal solution. When does (x,y,s) approach an optimal solution? Note that for $x \in \mathcal{F}^\circ(P)$ and $(y,s) \in \mathcal{F}^\circ(D)$:

$$x^Ts = x^T(c - A^Ty) = x^Tc - (Ax)^Ty = c^Tx - b^Ty$$

If $x^Ts = 0$, then $c^Tx = b^Ty$ and x and (y,s) are optimal. Also note if $x^Ts \leq \epsilon$, then $c^Tx - b^Ty \leq \epsilon$. Since $b^Ty \leq c^Tx$, this implies the primal and dual are within ϵ of their optimal values.

Therefore, we want that $G_q(x,s)$ sufficiently small implies x^Ts close to 0. Hence we may ask if there exists some q s.t. $G_q(x,s) \to -\infty$ as $x^Ts \to 0$. If such q exists, then $G_q(x,s)$ is a desirable potential function. By Lemma 2, we have:

$$\begin{aligned} G_q(x,s) &= G_n(x,s) - (n-q)\ln(x^Ts) \\ &\geq -(n-q)\ln(x^Ts) + n\ln n \end{aligned}$$

Therefore, for $q < n$, $G_q(x,s) \to +\infty$ as $x^Ts \to 0$. Thus, we want to choose some $q > n$, and the following lemma shows such q exists:

Lemma 3 *If* $G_q(x,s) \leq -\sqrt{n}\ln\frac{1}{\epsilon}$ *for* $q = n + \sqrt{n}$, *then* $x^Ts \leq \epsilon$.

Proof:

$$-\sqrt{n}\ln\frac{1}{\epsilon} = G_q(x,s) = (n+\sqrt{n})\ln(x^Ts) - \sum_{j=1}^{n}\ln x_js_j \geq \sqrt{n}\ln(x^Ts) + n\ln n$$

$$\Rightarrow \quad \ln(x^Ts) \leq -\ln\frac{1}{\epsilon} - \sqrt{n}\ln n \leq -\ln\frac{1}{\epsilon}$$

$$\Rightarrow \quad x^Ts \leq e^{-\ln\frac{1}{\epsilon}} = \epsilon$$

□

Therefore, if we can find a good starting point (x, y, s) and can reduce $G_q(x, s)$ by a constant in each iteration, then in $O(\sqrt{n}\ln\frac{1}{\epsilon})$ iterations we can find a solution whose value is within ϵ of the value of the optimal solution.