Mathematical Programming	**Lecture 26**
OR 630 Fall 2005	**November 30, 2005**
Notes by Serkan Kirac	

.
The log barrier function is defined as $F(x) := -\ln(x) := -\sum \ln(x_j)$.

a) It is a (strictly) convex function on $\mathcal{F}^\circ(\mathcal{P})$ or on $R^n_{++} := \{x \in R^n : x > 0\}$;

b) $F(x) \to +\infty$ if $x \to \bar{x} \in R^n_+ \backslash R^n_{++}$;

c) $F(\bar{X}^{-1}x) = \ln(\bar{x}) + F(x)$.

We can use this to compare points in $\mathcal{F}^\circ(\mathcal{P})$:

i) $c^T x$ measures its objective function value;

ii) $F(x)$ measures its "centrality" in the feasible region.

This motivates the penalized function $\theta_\mu(x) := c^T x + \mu F(x)$ for $\mu > 0$ defined on $\mathcal{F}^\circ(\mathcal{P})$.
μ is small - the minimizer is near optimal;
μ is large - the minimizer is "central" in the feasible region.

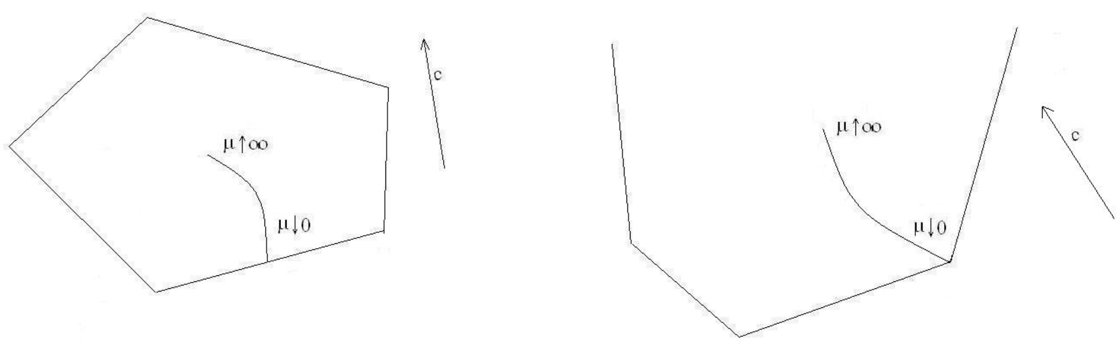

Figure 1: *Example for bounded and unbounded polyhedra.*

Theorem 1 ***a)*** *A necessary and sufficient condition for θ_μ to have a minimizer on $\mathcal{F}^\circ(\mathcal{P})$ is that $\mathcal{F}^\circ(\mathcal{P})$ and $\mathcal{F}^\circ(\mathcal{D})$ be nonempty.*
b) *If these conditions hold, a necessary and sufficient condition to $x \in \mathcal{F}^\circ(\mathcal{P})$ to be a minimizer*

(in fact "the" minimizer) is that there is $(y, s) \in \mathcal{F}^\circ(\mathcal{D})$ such that

$$\begin{array}{rrll} & A^T y + s & = c, & s > 0 \\ Ax & & = b, & x > 0 \quad (*) \\ & XSe & = \mu e, & \end{array}$$

where $X :=$ Diag(x), $S :=$ Diag(s), $e = (1, 1, 1..., 1)^T \in R^n$

(Note: the minimizer is unique because $F(x) = -\ln(x)$ is strictly convex)
Proof: (Sufficiency of (a)). Assume $\hat{x} \in \mathcal{F}^\circ(\mathcal{P})$, $(\hat{y}, \hat{s}) \in \mathcal{F}^\circ(\mathcal{D})$. Then $\theta_\mu(x) = c^T x + \mu F(x) = (A^T \hat{y} + \hat{s})^T x + \mu F(x) = b^T \hat{y} + \hat{s}^T x + \mu F(x) = b^T \hat{y} + \sum(\hat{s}_j x_j - \mu \ln(x_j))$.

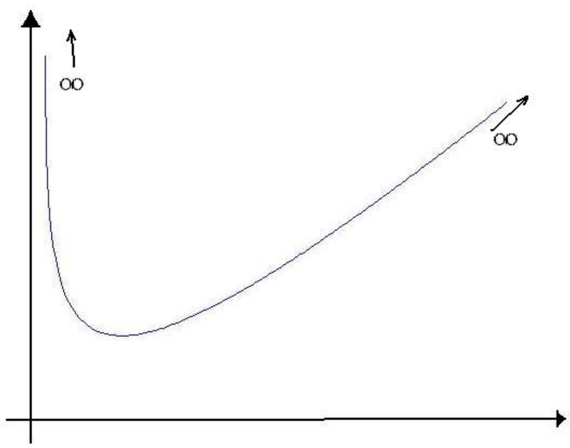

Figure 2: *Graph of $\hat{s}_j x_j - \mu \ln(x_j)$*

For every $x \in \mathcal{F}^\circ(\mathcal{P})$, if $\theta_\mu(x) \le \theta_\mu(\hat{x})$ then $0 < \underline{x}_j \le x_j \le \bar{x}_j < \infty$ for all j where $\underline{x}_j$ is a lower bound and $\bar{x}_j$ an upper bound.
We will apply Weierstrass's Theorem: A continuous function attains its minimum on a compact set.
Apply this to minimize θ_μ (it is continuous) over $\{x \in \mathcal{F}^\circ(\mathcal{P}) : \underline{x} \le x \le \bar{x}\}$. So there exists a minimizer.

((b) and necessity of (a)). Suppose that x is a minimizer of θ_μ over $\mathcal{F}^\circ(\mathcal{P})$. Then $\nabla\theta_\mu(x) = c + \mu\nabla F(x)$ must be in the range of A^T. If not, by the proposition of lecture of 11/29, $-P_A \nabla\theta_\mu(x)$ is a direction where θ_μ decreases to first order so x is not a minimizer. I.e., $c + \mu\nabla F(x) = c + \mu(-X^{-1}e) =: c - \mu x^{-1} = A^T y$ for some y ($x^{-1} := (\frac{1}{x_1}; \frac{1}{x_2}; ....)$). Now set $s = \mu x^{-1} > 0$ and obtain (*). Part (a) is done, since $x \in \mathcal{F}^\circ(\mathcal{P})$ and $(y, s) \in \mathcal{F}^\circ(\mathcal{D})$.
Finally suppose conditions (*) hold. Then $x \in \mathcal{F}^\circ(\mathcal{P})$ and $\nabla\theta_\mu(x) = c - \mu x^{-1} = A^T y$. So, $(c - A^T y)^T x + \mu F(x)$ has zero gradient at x and (since $F(x)$ is a convex function) x is a global minimizer of $(c - A^T y)^T x + \mu F(x)$ over R^n_{++} and hence over $\mathcal{F}^\circ(\mathcal{P})$. But this function is $\theta_\mu(x) - b^T y$, so x is a minimizer of θ_μ over $\mathcal{F}^\circ(\mathcal{P})$. $\square$

Note: Conditions (*) are symmetric between (P) and (D). No surprise that (*) gives necessary and sufficient conditions for (y, s) to solve

$$\begin{aligned} \max \quad & b^T y - \mu F(s) \\ & A^T + s = c \\ & s > 0. \end{aligned}$$

Corollary 1: (assuming $\mathcal{F}^\circ(\mathcal{P}), \mathcal{F}^\circ(\mathcal{D})$ are nonempty) The solutions $(x(\mu), y(\mu), s(\mu))$ to (*) satisfy

$$\begin{array}{rcccc} x(\mu) & \in & \mathcal{F}^\circ(\mathcal{P}), & & \\ (y(\mu), s(\mu)) & \in & \mathcal{F}^\circ(\mathcal{D}), & & \\ c^T x(\mu) - b^T y(\mu) & = & x(\mu)^T s(\mu) & = & n\mu, \end{array}$$

for any $\mu > 0$.

So, as $\mu \downarrow 0$,
$c^T x(\mu) \to v(P)$, optimal value of (P), and
$b^T x(\mu) \to v(D)$, optimal value of (D).

Corollary 2: If $\mathcal{F}^\circ(\mathcal{P}), \mathcal{F}^\circ(\mathcal{D})$ are nonempty, then at least one of them is unbounded ($x_j(\mu) s_j(\mu) \to \infty$ as $\mu \to \infty$).
$\{x(\mu) : \mu > 0\}$ is called the primal central path;
$\{(y(\mu), s(\mu)) : \mu > 0\}$ is called the dual central path; and
$\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ is called the primal-dual central path.

Introduction to Interior Point Methods

(For more see Bertsimas-Tsitsiklis and S. Wright, Primal-Dual Interior-Point Methods, SIAM, 1997.)

a) Affine-Scaling Methods: Originally due to Dikin('67) and rediscovered several times after Karmarkar. Take a step along the affine-scaling direction of a certain length at each iteration. (Convergence proved for different variants, but not thought to be polynomial time.)

b) Potential-Reduction Methods: Based on reducing a potential function at every iteration.

Primal potential function:
$\phi_q(x : \zeta) := q \ln(c^T x - \zeta) + F(x)$ (cf. $\theta_\mu(x) = c^T x + \mu F(x)$) $q \geq n$, $\zeta \leq \zeta_* = v(D)$, $x \in \mathcal{F}^\circ(\mathcal{P})$ (Karmarkar '84). We want to decrease $\phi_q(x : \zeta)$ by a constant at every iteration. If we can, we can get "$O(n \ln \frac{1}{\varepsilon})$" iterations to get ε-optimal solutions (the quotes indicate that there are other terms in the bound depending on initialization, etc.).

Primal-dual potential function:

$\phi_q(x,y,s) := q\ln(x^T s) + F(x) + F(s)$ for $q \geq n$, defined on $\mathcal{F}^\circ(\mathcal{P}) \times \mathcal{F}^\circ(\mathcal{D})$ (Tanabe '87, Todd-Ye '90).

Proposition 1: $\phi_n(x,y,s) \geq n\ln(n)$, with equality if and only if all $x_j s_j$'s are equal.
Proof: $\phi_n(x,y,s) = n\ln(n) + n\ln\frac{x^T s}{n} - \sum\ln(x_j s_j) = n\ln(n) + n\ln(A) - n\ln(\prod x_j s_j)^{\frac{1}{n}} = n\ln(n) + n\ln(A) - n\ln(G) \geq n\ln(n)$, where A is the arithmetic mean and G the geometric mean of the $x_j s_j$'s. $\square$

Proposition 2: $\phi_q(x,y,s)$ is unbounded below on $(x,y,s) \in \mathcal{F}^\circ(\mathcal{P}) \times \mathcal{F}^\circ(\mathcal{D})$ if $q > n$.
Proof: Look at $(x(\mu), y(\mu), s(\mu))$ with $\phi_q(x,y,s) = (q-n)\ln(n\mu) + n\ln(n) \to -\infty$ as $\mu \downarrow 0$. $\square$

Theorem 2 *If $(x,y,s) \in \mathcal{F}^\circ(\mathcal{P}) \times \mathcal{F}^\circ(\mathcal{D})$, then $c^T x - b^T y = x^T s \leq \exp(\frac{\phi_q(x,y,s)}{q-n})$. (This leads to "$O((q-n)\ln\frac{1}{\varepsilon})$" steps to get $x^T s \leq \varepsilon$ if we can decrease ϕ_q by a constant every iteration. This can be achieved for $q \geq n + \sqrt{n}$, leading to "$O(\sqrt{n}\ln\frac{1}{\varepsilon})$" iteration algorithms (Ye '91, Freund '91, Kojima-Mizuno-Yoshise '92).)*

Proof: $\phi_q(x,y,s) := q\ln(x^T s) + F(x) + F(s) = (q-n)\ln(x^T s) + \phi_n(x,y,s) \geq (q-n)\ln(x^T s) + n\ln(n) \geq (q-n)\ln(x^T s)$. $\square$