Semidefinite Programming	**Lecture 15**
OR 6327 Spring 2012	**March 13, 2012**
Scribe: Shanshan Zhang	

Let's consider the barrier function $F(X) := -\text{lndet}(X)$. Note that the set

$$\{(x_0; \bar{x}) \in \mathbb{R}^{1+n} : x_0 \geq ||\bar{x}||_2\}$$

is equal to the set

$$\{(x_0; \bar{x}) \in \mathbb{R}^{1+n} : \begin{pmatrix} x_0 & \bar{x}^T \\ \bar{x} & x_0 I \end{pmatrix} \succeq 0\}.$$

Then, for points in the interior of this cone, we have

$$\begin{aligned} -\ln\det\begin{pmatrix} x_0 & \bar{x}^T \\ \bar{x} & x_0 I \end{pmatrix} &= -\ln\det(x_0 I) - \ln(x_0 - \frac{1}{x_0}\bar{x}^T\bar{x}) \\ &= -n\ln x_0 - \ln(x_0 - \frac{1}{x_0}\bar{x}^T\bar{x}) \\ &= -(n-1)\ln x_0 - \ln(x_0^2 - ||\bar{x}||^2). \end{aligned}$$

But, $-\ln(x_0^2 - ||\bar{x}||^2)$ is sufficient. It is a 2-logarithmically homogeneous self-concordant barrier.

More generally, if $P \in \mathbb{R}^{m\times n}$, $m \geq n$, then for $t > \|P\|$,

$$\begin{aligned} -\ln\det\begin{pmatrix} tI_m & P \\ P^T & tI_n \end{pmatrix} &= -\ln\det(tI_m) - \ln\det(tI_n - P^T(tI_m)^{-1}P) \\ &= -m\ln t - \ln\det(tI_n - \frac{1}{t}P^TP) \\ &= -(m-1)\ln t - \ln\det(tI_n - \frac{1}{t}P^TP) - \ln t. \end{aligned}$$

But the first term $-(m-1)\ln t$ is not necessary; the remaining terms give a $(n+1)$-logarithmically homogeneous self-concordant barrier function. This shows that for certain subsets of the semidefinite cone, there are better barrier functions than F; but in general F above is "optimal."

For the second-order cone, the Carathéodory number (the minimum number of extreme rays to represent any point as a nonnegative linear combination in the cone) is 2, and the optimal barrier (above) has parameter 2. For SDP the numbers are n and n. This is true for all "symmetric" cones. Notice that the number is much smaller than the dimension of the cone ($1+n$ and $\frac{n(n+1)}{2}$ respectively).

Consider the barrier problems:

$$(BP_\mu) : \min C \bullet X + \mu F(X) : \mathcal{A}X = b,$$

$$(BD_\mu) : \max b^T y - \mu F(S) : \mathcal{A}^* y + S = C,$$

for $\mu > 0$. Suppose X is an optimal solution to (BP_μ), with finite value, so $X \succ 0$. Then by Lagrange's theorem, there is some $y \in \mathbb{R}^m$ with $C + \mu F(X) = \mathcal{A}^* y$. And setting $S = -\mu F'(X) = \mu X^{-1}$, we have a solution to

$$\begin{aligned} \mathcal{A}^* y + S &= C, \qquad S \succ 0 \\ \mathcal{A}X &= b, \qquad X \succ 0 \\ XS &= \mu I, \end{aligned}$$

which we denote by (1). Similarly, if (y, S) is an optimal solution to (BD_μ), then we get $b - \mathcal{A}X = 0$ and $\mu F'(S) - X = 0$ and again we have a solution to (1).

Theorem 1 *If $\mathcal{F}^\circ(P)$ and $\mathcal{F}^\circ(D)$ are nonempty and the A_i's are linearly independent, then for every $\mu > 0$, there is a unique solution $(X(\mu), y(\mu), S(\mu))$ to (1). Moreover, $X(\mu)$ solves (BP_μ), $(y(\mu), S(\mu))$ solves (BD_μ), and the associated duality gap $X(\mu) \bullet S(\mu) = n\mu$. If one (or both) of $\mathcal{F}^\circ(P)$ and $\mathcal{F}^\circ(D)$ is empty, then there is no solution to (1) and no solution to (BP_μ) and (BD_μ) for any $\mu > 0$.*

Proof: The conclusion if $\mathcal{F}^\circ(P)$ or $\mathcal{F}^\circ(D)$ is empty is easy, so assume $\hat{X} \in \mathcal{F}^\circ(P)$ and $(\hat{y}, \hat{S}) \in \mathcal{F}^\circ(D)$. Then (BP_μ) can be written as

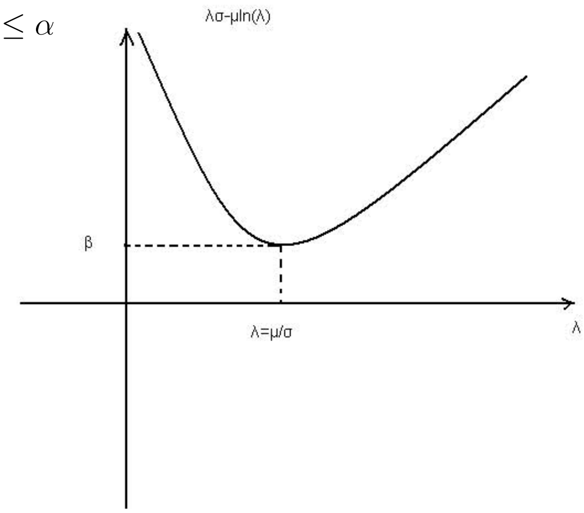
$$\begin{aligned} \min \qquad & C \bullet X + \mu F(X) \\ & \mathcal{A}X = b \\ & C \bullet X + \mu F(X) \leq C \bullet \hat{X} + \mu F(\hat{X}) \end{aligned}$$

Since $C \bullet X - C \bullet \hat{X} = (C \bullet X - b^T \hat{y}) - (C \bullet \hat{X} - b^T \bullet \hat{y}) = \hat{S} \bullet X - \hat{S} \bullet \hat{X}$, then the above system can be written as

$$\begin{aligned} \min \qquad & C \bullet X + \mu F(X) \\ & \mathcal{A}X = b \\ & \hat{S} \bullet X + \mu F(X) \leq \hat{S} \bullet \hat{X} + \mu F(\hat{X}) := \alpha. \end{aligned}$$

Choose σ with $\hat{S} \succeq \sigma I$, so any feasible X satisfies

$$\sigma I \bullet X + \mu F(X) \leq \hat{S} \bullet X + \mu F(X) \leq \alpha$$
$$\Longrightarrow \sum_j (\sigma \lambda_j(X) - \mu \ln \lambda_j(X)) \leq \alpha$$



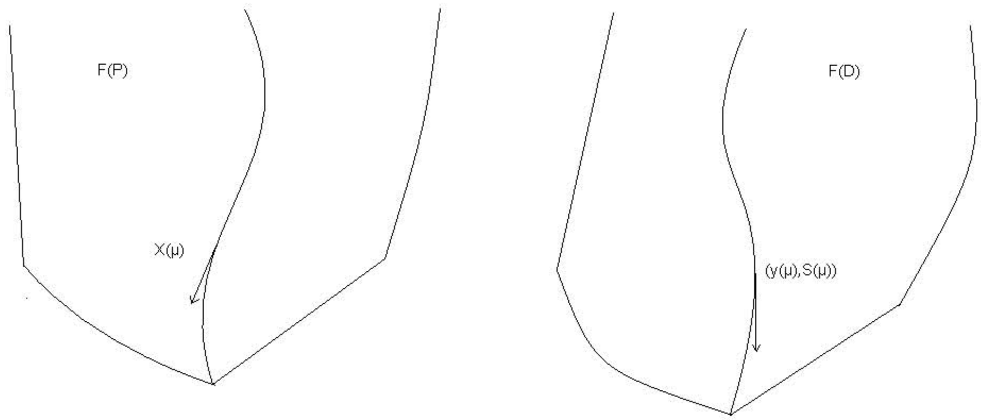
It is easy to see each term is minimized by $\lambda_j = \lambda = \frac{\mu}{\sigma}$. Let β be the minimum of $\lambda_j - \mu \ln \lambda_j$. Then for each j,

$$\sigma \lambda_j - \mu \ln \lambda_j \leq \alpha - (n-1)\beta.$$

This shows the existence of $\underline{\lambda}$ and $\bar{\lambda}$ such that $0 < \underline{\lambda} \leq \lambda_j \leq \bar{\lambda} \leq \infty$ for all j. Then our problem is to minimize a continuous function ($\underline{\lambda} > 0$) on a compact set ($\bar{\lambda} < \infty$). So there exists an optimal X, say $X(\mu)$. Then by the optimality conditions, we have a solution to (1). Conversely, any solution to (1) gives an optimal solution to (BP_μ) because it is convex. Since (BP_μ) has a strictly convex objective function, $X(\mu)$ is unique. So $S(\mu)$ is unique ($= \mu X^{-1}$). Then, since the A_i's are linearly independent, $y(\mu)$ is unique. Finally, $(y(\mu), S(\mu))$ solves (BD_μ) because this has a concave objective to be maximized with linear constraints, and it satisfies optimality conditions with $X = X(\mu)$ as multipliers. And moreover, $X(\mu)S(\mu) = \mu I$, so $X(\mu) \bullet S(\mu) = n\mu$. □

It can be shown (with more work) that all $X(\mu)$ and all $(y(\mu), S(\mu))$ lie in fixed compact sets, say for all $\mu \leq 1$, so there are limit points as $\mu \downarrow 0$. By taking limits, this shows strong duality again.

We would like to think of $\{X(\mu)\}$ and $\{(y(\mu), S(\mu))\}$ as paths. A path $\{X(\mu)\}$ in $\mathcal{F}(P)$ corresponds to a path $\{(y(\mu), S(\mu))\}$ in $\mathcal{F}(D)$.



To do this, we will use the implicit function theorem. $(X(\mu), y(\mu), S(\mu))$ solves

$$G(X, y, S; \mu) = \begin{pmatrix} \mathcal{A}^* y + S - C \\ \mathcal{A}X - b \\ XS - \mu I \end{pmatrix} = 0.$$

Since the dimensions don't match, we need to symmetrize somehow the last equation. We will see lots of ways to do it, but for now, write it as

$$S + \mu F'(X) = S - \mu X^{-1} = 0,$$

Then we get

$$D_{X,y,S} G(X, y, S; \mu) = \begin{pmatrix} 0 & \mathcal{A}^* & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \mathcal{E} & 0 & \mathcal{F} \end{pmatrix},$$

where $\mathcal{E} : \mathbb{M}^n \to \mathbb{M}^n$ is defined by $\mathcal{E}(H) = \mu X^{-1} H X^{-1}$ and $\mathcal{F} : \mathbb{M}^n \to \mathbb{M}^n$ is the identity. To apply the implicit function theorem, we will need DG invertible. We will show next time: if the A_i's are linearly independent, $\mathcal{E}$ is invertible and $\mathcal{E}^{-1}\mathcal{F}$ is positive definite (not necessarily self-adjoint), then DG is invertible.