Semidefinite Programming	**Lecture 14**
OR 6327 Spring 2012	**March 8, 2012**
Scribe: Marcus Lim	

Recall:

$$F(X) = \begin{cases} -\ln\det X & \text{if} X \succ 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Differential properties of F $(X \succ 0, H \in \mathbb{M}^n\ )$:

$$\begin{aligned} F(X+H) &= -\ln\det(X+\epsilon H) \\ &= -\ln\left[\det(X)\det(I+\epsilon X^{-1}H)\right] \\ &= F(X) - \ln\det(I+\epsilon X^{-1}H) \\ &= F(X) - \ln\left(1+\epsilon\text{trace}\,(X^{-1}H)+O(\epsilon^2)\right) \\ &= F(X) - \epsilon\text{trace}\,(X^{-1}H)+O(\epsilon^2) \\ &= F(X) + \epsilon\left(-X^{-1}\right)\bullet H + O(\epsilon^2). \end{aligned}$$

So the directional derivative:

$$\begin{aligned} DF(X)[H] &= -X^{-1}\bullet H \text{ and} \\ F'(X) &= -X^{-1} \end{aligned}$$

where the latter is defined by $DF(X)[H] =: F'(X)\bullet H$. This corresponds to:

$$\begin{aligned} f(\xi) &= -\ln(\xi), \\ f'(\xi) &= -1/\xi, \\ f''(\xi) &= 1/\xi^2. \end{aligned}$$

Next consider

$$\begin{aligned} (X+H)^{-1} &= \left(X\left(I+\epsilon X^{-1}H\right)\right)^{-1} \\ &= \left(I+\epsilon X^{-1}H\right)^{-1}X^{-1} \\ &= \left(I-\epsilon X^{-1}H+O\left(\epsilon^2\right)\right)X^{-1} \\ &= X^{-1}-\epsilon X^{-1}HX^{-1}+O(\epsilon^2). \end{aligned}$$

The second equality is derived as follows: We know that $(1-\epsilon)^{-1} = 1+\epsilon+\epsilon^2+\ldots$ for $|\epsilon|<1$. Correspondingly, we have Neumann's Lemma: If $\|E\|_2 < 1$

$$(I-E)^{-1} = I+E+E^2+\ldots$$

Now we can obtain the second derivative of $F(X)$:

$$\begin{aligned} D^2F(X)[H,H] &= X^{-1}HX^{-1}\bullet H, \text{ and more generally,} \\ D^2F(X)[H,J] &= X^{-1}HX^{-1}\bullet J. \end{aligned}$$

We can then define $F''(X)$ by

$$D^2F(X)[H,J] =: (F''(X)H) \bullet J$$

and then

$$F''(X)H = X^{-1}HX^{-1}.$$

Finally, we can similarly obtain the third derivative:

$$\begin{aligned} D^3F(X)[H,J,K] &= -X^{-1}KX^{-1}HX^{-1} \bullet J - X^{-1}HX^{-1}KX^{-1} \bullet J \\ &= -2X^{-1}HX^{-1}JX^{-1} \bullet K. \end{aligned}$$

Recall: F is convex iff $\phi : R \to R$ defined by $\phi(\alpha) := F(X + \alpha H)$ is convex for any $X \succ 0$ and $H \in \mathbb{M}^n$, and F is strictly convex if it is twice differentiable and $\phi''(0) > 0$ for all such X and $H \neq 0$.

Definition 1 *F is a barrier function for $\mathbb{M}^n_+$ if $F(X) \to +\infty$ as $X \to \bar{X} \in \partial\mathbb{M}^n_+ = \mathbb{M}^n_+ \setminus \mathbb{M}^n_{++}$.*

Definition 2 *F is a self-concordant function if it is convex, thrice differentiable, and for any $X \succ 0, H \in \mathbb{M}^n$, and ϕ as above*

$$|\phi'''(0)| \leq 2(\phi''(0))^{3/2}.$$

[This bound on the third derivative is to assure that Newton's method, which converges in one step for quadratics, converges fast for such problems. (Nesterov & Nemirovski)]

Definition 3 *F defined on a cone K is θ-logarithmically homogeneous if for all $X \in K, \tau > 0$*

$$F(\tau X) = F(X) - \theta \ln \tau.$$

Theorem 1 *The log determinant function F above is a strictly convex, self-concordant, and n-logarithmically homogeneous barrier function.*

Proof: Recall that $F(X) = -\sum_j \ln \lambda_j(X)$. If $X \to \bar{X} \in \partial\mathbb{M}^n_+$, then $\lambda(X) \to \lambda(\bar{X})$, but one of the components of $\lambda(\bar{X})$ is 0. So, F is a barrier function. Next,

$$\begin{aligned} F(\tau X) &= -\ln\det(\tau X) \\ &= -\ln(\tau^n \det X) \\ &= F(X) - n\ln\tau \end{aligned}$$

for $X \succ 0, \tau > 0$, so F is n-logarithmically homogeneous.

For strict convexity, choose any $X \succ 0$ and nonzero $H \in \mathbb{M}^n$, and define $\phi(\alpha) := F(X+\alpha H)$. Then

$$\begin{aligned} \phi''(0) &= D^2F(X)[H,H] \\ &= X^{-1}HX^{-1} \bullet H. \end{aligned}$$

Let $V = X^{-1/2}$, so

$$\begin{aligned}\phi''(0) &= \text{trace}\,(V^2 H V^2 H) \\ &= \text{trace}\,((VHV)\,(VHV)) \\ &= \|VHV\|_F^2 > 0 \qquad\qquad \text{since } H \neq 0.\end{aligned}$$

Finally, to prove self-concordance, we let X, H, and ϕ be as above. Then,

$$\begin{aligned}\phi'''(0) &= -2\ \text{trace}\,(X^{-1}HX^{-1}HX^{-1}H) \\ &= -2\ \text{trace}\,((VHV)\,(VHV)\,(VHV))\,.\end{aligned}$$

Let $\lambda = \lambda(VHV)$. Then the eigenvalues of $(VHV)^3$ are $\lambda_j^3, j = 1, , n$. So,

$$\begin{aligned}|\phi'''(0)| &= 2\left|\sum_j {\lambda_j}^3\right| \\ &< 2\sum_j |\lambda_j|^3 \\ &= 2\,\|\lambda\|_3^3\,.\end{aligned}$$

Similarly,

$$\begin{aligned}\phi''(0) &= \text{trace}\,(X^{-1}HX^{-1}H) \\ &= \text{trace}\,((VHV)\,(VHV)) \\ &= \sum_j \lambda_j^2 = \|\lambda\|_2^2\,.\end{aligned}$$

The result follows since $\|\lambda\|_3 \leq \|\lambda\|_2$. $\square$

Using $F''(X)$, we can define two norms on $\mathbb{M}^n$:

$$\|V\|_X := \sqrt{F''(X)V \bullet V} = \left\|X^{-1/2} V X^{-1/2}\right\|_F,$$

the X-norm of V, and the dual X-norm of U:

$$\|U\|_X^* := \sqrt{U \bullet F''(X)^{-1} U} = \left\|X^{1/2} U X^{1/2}\right\|_F.$$

Note:

$$\begin{aligned}U \bullet V &= \text{trace}\,(UV) \\ &= \text{trace}\,(X^{1/2}UX^{1/2}X^{-1/2}VX^{-1/2}) \\ &\leq \|U\|_X^* \cdot \|V\|_X. \qquad\qquad (\text{Cauchy} - \text{Schwarz}).\end{aligned}$$

The following derives from the n-logarithmic homogeneity of F:

Proposition 1 *The following properties hold:*

(a) $F'(\tau X) = \frac{1}{\tau} F'(X), \qquad F''(\tau X) = \frac{1}{\tau^2} F''(X), \ldots$

(b) $F'(X) \bullet X = -n, \qquad F''(X)X = -F'(X).$

(c) $\|X\|_X = \sqrt{n}, \qquad \|F'(X)\|_X^* = \sqrt{n}.$

Proof: We have $F(\tau X) \equiv F(X) - n \ln \tau$. Taking the derivative w.r.t. X once and twice gives (a). Taking the derivative w.r.t. τ and then setting $\tau = 1$ gives the first part of (b). Taking the derivative of the first part of (b) with respect to X gives the second part. [Note that this says that the Newton step for minimizing F at X is X itself.] For (c),

$$\begin{aligned} \|X\|_X &= \sqrt{F''(X)X \bullet X} \\ &= \sqrt{-F'(X) \bullet X} \\ &= \sqrt{n}, \\ \|F'(X)\|_X^* &= \sqrt{F'(X) \bullet F''(X)^{-1} F'(X)} \\ &= \sqrt{F'(X) \bullet (-X)} \\ &= \sqrt{n}. \qquad \square \end{aligned}$$

Note that (b) and (c) show that the X-norm of the Newton step at X is (bounded by) $\sqrt{n}$, so F is an n-self-concordant barrier.

Given a convex function F, its (modified) convex conjugate F_* is defined by

$$\begin{aligned} F_*(S) &:= \sup_X \{-S \bullet X - F(X)\} \\ &= \sup_X \{-S \bullet X + \ln \det X : X \succ 0\}. \end{aligned}$$

The derivative of the (concave) term to be maximized is:

$$-S + X^{-1}, \qquad \text{which is 0 if} X = S^{-1}.$$

So

$$\begin{aligned} F_*(S) &= -S \bullet S^{-1} + \ln \det(S^{-1}) \\ &= -n - \ln \det S, \qquad\qquad \text{if } S \succ 0 \end{aligned}$$

(and $+\infty$ if $S \not\succ 0$). So $F_*(S) = -n + F(S)$ differs by a constant from the original function F.