Semidefinite Programming	**Lecture 2**
OR 6327 Spring 2012	**January 26, 2012**
Scribe: Tia Sondjaja	

Today, we will cover a few important facts about symmetric matrices and look at the problem of minimizing the maximum eigenvalue of a matrix as an SDP problem in dual form.

First, we recall the primal and dual forms of SDP:

$$\begin{array}{lcll} \min & C \bullet X & & \\ s.t. & \mathcal{A}X = b & \text{(i.e., } A_i \bullet X = b_i, \forall i = 1, \ldots, m) & (P) \\ & X \succeq 0, & & \end{array}$$

$$\begin{array}{lcll} \max & b^T y & & \\ s.t. & \mathcal{A}^* y + S = C & \text{(where } \mathcal{A}^* y = \sum_i y_i A_i) & (D) \\ & S \succeq 0. & & \end{array}$$

Everything you ever want to know about symmetric matrices

Fact 1 *If $P, Q \in \mathbb{R}^{m \times n}$, then*

$$\begin{aligned} P \bullet Q &:= \operatorname{trace}(P^T Q) = \operatorname{trace}(Q P^T) \\ &= \operatorname{trace}(Q^T P) = \operatorname{trace}(P Q^T) \\ &= \sum_i \sum_j p_{ij} q_{ij}, \end{aligned}$$

even though $P^T Q$ and QP^T have different sizes ($n \times n$ and $m \times m$ respectively).

Fact 2 *$\mathcal{A}$ and $\mathcal{A}^*$ are adjoint mappings:*

$$(\mathcal{A}X)^T y = (\mathcal{A}^* y) \bullet X.$$

Fact 3 *If P is a nonsingular $n \times n$ real matrix, then U is psd (respectively, pd) if and only if PUP^T is psd (resp., pd).*

Fact 4 *Suppose $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix; then*

$$\left(QUQ^T\right) \bullet \left(QVQ^T\right) = U \bullet V.$$

More generally, if $P \in \mathbb{R}^{n \times n}$ is nonsingular, then

$$\left(P^{-T} U P^{-1}\right) \bullet \left(PVP^T\right) = U \bullet V.$$

(Generalization of $\left(Q^T u\right)^T \left(Q^T v\right) = \left(P^{-1} u\right)^T (Pv) = u^T v$.)

Proof:

$$\begin{aligned}
\left(P^{-T}UP^{-1}\right)\bullet\left(PVP^{T}\right) &= \operatorname{trace}\left(P^{-T}UP^{-1}PVP^{T}\right), \quad \text{by definition}\\
&= \operatorname{trace}\left(P^{-T}UVP^{T}\right)\\
&= \operatorname{trace}\left(UVP^{T}P^{-T}\right), \quad \text{by Fact 1}\\
&= \operatorname{trace}\left(UV\right) = U\bullet V.
\end{aligned}$$

□

Note. Facts 3 and 4 show that (P) is equivalent to:

$$\begin{aligned}
\min \quad & \left(P^{-T}CP^{-1}\right)\bullet\hat{X}\\
s.t. \quad & \left(P^{-T}A_iP^{-1}\right)\bullet\hat{X} = b_i,\ \forall i=1,\ldots,m\\
& \hat{X}\succeq 0.
\end{aligned}$$

This problem arises from the change of variables $\hat{X}=PXP^T$. Thus the primal variable X transforms in a different way from the data C and the A_i's and the dual slack matrix S transforms in the same way as the data.

Fact 5 *If $Y\in\mathbb{M}^n$, then there are an orthogonal $Q\in\mathbf{R}^{n\times n}$ and a diagonal $\Lambda\in\mathbf{R}^{n\times n}$ such that $U=Q\Lambda Q^T$.*

Notation. For a diagonal matrix $\Lambda\in\mathbf{R}^{n\times n}$ whose diagonal entries are $\lambda_1,\ldots,\lambda_n$, we write:

$$\Lambda=\operatorname{Diag}(\lambda),$$

where λ is the vector $(\lambda_1;\ldots;\lambda_n)$. We also use the notation

$$\operatorname{diag}(U):=(u_{11};\ldots;u_{nn}),$$

for any matrix $U\in\mathbf{R}^{n\times n}$. This is the adjoint mapping of Diag.

If $Q=[q_1,\ldots,q_n]$ (that is, q_i is its ith column vector), then

$$UQ=Q\Lambda=[\lambda_1q_1,\ldots,\lambda_nq_n].$$

Hence, looking at the ith column of UQ and $Q\Lambda$,

$$Uq_i=\lambda_iq_i.$$

So, the q_i's and λ_i's are the eigenvectors and eigenvalues of U. We call $Q\Lambda Q^T$ the *eigenvalue decomposition* of U.

We will usually assume that $\lambda_1\geq\ldots\geq\lambda_n$, and then

$$\lambda=:\lambda(U),\ \Lambda=:\Lambda(U).$$

Fact 6 *The following are norms on $\mathbb{M}^n$:*

- *The 2-norm/operator norm,*

$$\begin{aligned} ||U||_2 &:= \max\{||Uz||_2 \mid ||z|| = 1\} \\ &= ||\Lambda(U)||_2 \\ &= \max\{|\lambda_i(U)|\} \\ &= ||\lambda(U)||_\infty; \end{aligned}$$

- *the Frobenius norm,*

$$\begin{aligned} ||U||_F &:= (U \bullet U)^{1/2}) \\ &= (\Lambda(U) \bullet \Lambda(U))^{1/2}) \\ &= ||\lambda(U)||_2 \\ &= ||\text{vec}(U)||_2 = ||\text{svec}(U)||_2, \end{aligned}$$

- *the nuclear norm or trace norm,*

$$||U||_* \;:=\; ||\lambda(U)||_1.$$

Note. The following observation motivates why $||\cdot||_*$ is called the trace norm:

$$\begin{aligned} \text{trace}\,(U) &= \sum_i u_{ii} = I \bullet U = I \bullet \Lambda(U) \\ &= \sum_i \lambda_i(U). \end{aligned}$$

Fact 7 (Theorem 1) *For $U \in \mathbb{M}^n$, the following are equivalent:*

(a) U is psd (resp., pd);

(b) $z^T Y z \geq 0$ for all $z \in \mathbf{R}^n$ (resp., $z^T Y z > 0$ for all nonzero $z \in \mathbf{R}^n$);

(c) $\lambda(u) \geq 0$ (resp., $\lambda(u) > 0$); and

(d) $U = P^T P$ for some $P \in \mathbf{R}^{n\times n}$ (resp., for some nonsingular $P \in \mathbf{R}^{n\times n}$)

Proof: (a) $\Leftrightarrow$ (b) by definition. For (b) $\Leftrightarrow$ (c), note that if $U = Q\Lambda Q^T$, then $z^T U z = z^T Q \Lambda Q^T z = \sum_i \lambda_i \tilde{z}_i$, where $\tilde{z} = Q^T z$. Next, (d) $\Rightarrow$ (b) because $z^T U z = z^T P^T P z = ||Pz||_2^2 \geq 0$ for all z. For the reverse, let $U = Q\Lambda Q^T$ and let $P = Q\Lambda^{1/2} Q^T$, where $\Lambda^{1/2} := \text{Diag}\,(\sqrt{\lambda_1}; \ldots; \sqrt{\lambda_n})$. $\square$

Example 1 (Eigenvalue optimization) *Suppose that $U(y)$ depends linearly (affinely) on $y \in \mathbf{R}^m$ and we want to choose y to minimize the maximum eigenvalue of $U(y)$. Introduce $\eta \in \mathbf{R}$ and note that*

$$\begin{aligned} & \lambda_{\max}(U) \leq \eta \\ \Leftrightarrow\quad & \lambda_{\max}(U - \eta I) \leq 0 \\ \Leftrightarrow\quad & \lambda_{\min}(\eta I - U) \geq 0 \\ \Leftrightarrow\quad & \eta I - U \succeq 0. \end{aligned}$$

So, the problem of finding $\min(\lambda_{\max}(U(y)))$ *can be formulated as the following SDP in dual form:*

$$\begin{array}{rll} -\max & & -\eta \\ & s.t. & -\eta I + U(y) \preceq 0. \end{array}$$

Corollary 1 *Every psd matrix* U *has a (unique) psd square root* $U^{1/2}$*, with* $U^{1/2}U^{1/2} = U$*. Every pd matrix* U *is nonsingular, and its inverse is pd.*

Proof: If $U = Q\Lambda Q^T$, then set:

$$U^{1/2} := Q\Lambda^{1/2}Q^T,$$

where $\Lambda^{1/2}$ is as in the previous proof. If U is pd, then Λ has positive diagonal entries, so

$$\Lambda^{-1} = \text{Diag}\,(\lambda_1^{-1}; \ldots ; \lambda_n^{-1})$$

exists, and $Q\Lambda^{-1}Q^T$ is the inverse of U, and is pd. (We won't prove uniqueness.) $\square$

Corollary 2 *If* $0 \neq u \in \mathbf{R}^n$*, then* uu^T *is psd. (And all psd rank-one matrices are of this form—see HW1.)*

Corollary 3 $\mathbb{M}^n_+$ *and* $\mathbb{M}^n_{++}$ *are convex cones.* $\mathbb{M}^n_+$ *is closed and* <u>*pointed*</u> *(that is, contains no one-dimensional subspaces) and its interior is* $\mathbb{M}^n_{++}$*.*

Proof: $\mathbb{M}^n_+$ is defined by the homogeneous linear (in U) inequalities

$$z^T U z \geq 0, \ \forall z \in \mathbf{R}^n.$$

Hence, we see that $\mathbb{M}^n_+$ is a closed convex cone.

$\mathbb{M}^n_{++}$ is defined by the strict homogeneous linear inequalities

$$z^T U z > 0, \ \forall 0 \neq z \in \mathbf{R}^n,$$

so $\mathbb{M}^n_{++}$ is a convex cone.

If $U \in (\mathbb{M}^n_+) \cap (-\mathbb{M}^n_+)$, then $\lambda(U) \geq 0$ and $\lambda(U) \leq 0$. So, U must be 0. This shows that $\mathbb{M}^n_+$ is pointed.

If $U \in \mathbb{M}^n_{++}$, then $\hat{\lambda} := \lambda_{\min}(U) > 0$. Let $V \in \mathbb{M}^n$ have $||V||_2 \leq \hat{\lambda}$. Then,

$$\begin{aligned} z^T(U+V)z &= z^T((U - \hat{\lambda}I) + \hat{\lambda}I + V)z \\ &\geq \hat{\lambda} z^T z + z^T V z \\ &\geq \hat{\lambda} - \hat{\lambda} = 0, \end{aligned}$$

for $\|z\|_2 = 1$. So, $\mathbb{M}^n_{++} \subseteq \text{int}(\mathbb{M}^n_+)$. Now, suppose that $U \notin \mathbb{M}^n_{++}$. Then, there exists a nonzero vector z such that $z^T U z \leq 0$. But then

$$z^T(U - \epsilon z z^T)z \leq 0 - \epsilon (z^T z)^2 < 0$$

for all $\epsilon > 0$, which shows that $U \notin \text{int}(\mathbb{M}^n_+)$. Hence, $\text{int}(\mathbb{M}^n_+) = \mathbb{M}^n_{++}$. $\square$

Fact 8 *If* $U \succeq 0$ *(resp.,* $U \succ 0$*), then* $u_{jj} \geq 0$ *(resp.,* $u_{jj} > 0$*) for all* $j = 1, \ldots, n$*, and if* $u_{jj} = 0$*, then* $u_{jk} = 0$ *for all* k*.*

Fact 9 *If* $U \succeq 0$*, then* PUP^T *is psd for all* $P \in \mathbf{R}^{m \times n}$*. If* $U \succ 0$ *and* P *has full row rank, then* $PUP^T \succ 0$*. Hence, if* P *is a permutation matrix, we see that every principal rearrangement of* U *is psd (resp., pd) if* U *is psd (resp., if* U *is pd).*

If

$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \succeq 0, \ (\textit{resp.}, \ \succ 0),$$

then $U_{11} \succeq 0$ *(resp.,* $\succ 0$*), and similarly for every principal submatrix of* U*.*