Semidefinite Programming	**Lecture 3**
OR 6327 Spring 2012	**January 31, 2012**
Scribe: Xiaoting Zhao	

Fact 10: U is psd (respectively, pd) if and only if all its principal minors (2^n of them) are nonnegative (positive); U is pd if and only if all its leading principal minors are positive; and U is pd if and only if it has a Cholesky factorization, i.e., $U = LL^T$, where $L \in \mathbb{R}^{n\times n}$ is lower triangular with positive diagonal entries.

Fact 11: Suppose $U = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \in \mathbb{M}^n_+$, and suppose $A \succ 0$; then U is psd (pd) if and only if $C - B^TA^{-1}B$ is psd (pd). ($C - B^TA^{-1}B$ is called the *Schur complement* of A in U.)

Proof: If $A \succ 0$, then

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ B^TA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^TA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

since $\begin{bmatrix} I & 0 \\ B^TA^{-1} & I \end{bmatrix}$ is a nonsingular matrix. □

This can be used to show the existence of a Cholesky factorization of a pd matrix, since its top left entry must be positive and can be chosen as A, and then the equation above shows how the result can be established by induction.

Fact 12 (Representing quadratics): For $U \in \mathbb{M}^n$, $x^TUx = U \bullet xx^T$ is a quadratic function of x but a linear function of $X = xx^T$.

Fact 13: If U and V are psd, then $U \bullet V \geq 0$ (used in the Weak Duality Lemma). Indeed, we have the following theorem.

Theorem 1 *The cone $K = \mathbb{M}^n_+$ is self-dual, i.e., equal to*

$$K^* := \{V \in \mathbb{M}^n : U \bullet V \geq 0 \quad \forall U \in K\}.$$

Proof:
First show $K \subseteq K^*$ with the following two methods:

- By facts 3-5, we can assume wlog that U is *diagonal*, and then $U \bullet V = \sum_j u_{jj}\ v_{jj} \geq 0$ (we don't get a double sum because U is diagonal).
- $U \bullet V$ = trace (UV) = trace $(U^{1/2}VU^{1/2}) \geq 0$ since U has a psd square root and $U^{1/2}VU^{1/2}$ is psd, so has nonnegative trace. (The trace is the sum of the eigenvalues, which are nonnegative for a psd matrix.)

Now show $K^* \subseteq K$:

Suppose $V \notin K$, so there is $z \in \mathbb{R}^n$ with $z^T V z < 0$. But $zz^T \in K$ and $zz^T \bullet V = z^T V z < 0$, which shows $V \notin K^*$. $\square$

Note: If $U \succ 0$, $V \succeq 0$, $V \neq 0$, then $U \bullet V > 0$.

Fact 14: If $U \succ 0$, and $\beta \geq 0$, then $\{V \in \mathbb{M}^n_+ : \ U \bullet V \leq \beta\}$ is compact.

Proof: $U \succ 0$ implies $\hat{\lambda} := \lambda_{\min}(U) > 0$. Then, if V is feasible,

$$\begin{aligned} \beta \geq U \bullet V &= (U - \hat{\lambda} I + \hat{\lambda} I) \bullet V \\ &\geq \hat{\lambda} I \bullet V = \hat{\lambda} ||V||_*, \end{aligned}$$

since $U - \hat{\lambda} I$ is psd and V is psd as well. So $||V||_* \leq \beta / \hat{\lambda} < \infty$, showing the feasible region is bounded, and clearly it is closed. $\square$

Fact 15: If U, V are psd with $U \bullet V = 0$, then $UV = 0$ (analogous to complementary slackness in linear programming).

(*Note:* the converse is trivial.)

Proof: We have

$$0 = \text{trace}\,(UV) = \text{trace}\,(V^{1/2} U V^{1/2}) = (U^{1/2} V^{1/2}) \bullet (U^{1/2} V^{1/2}) = ||(U^{1/2} V^{1/2})||_F^2.$$

So $U^{1/2} V^{1/2} = 0$, so $UV = 0$. $\square$

Fact 16: If $U, V \in \mathbb{M}^n$, they commute *iff* UV is symmetric, *iff* they can be simultaneously diagonalized, i.e., we can write $U = Q \Lambda Q^T$, $V = Q M Q^T$ (see HW1).

Applications: Matrix Optimization

Recall:

$$(P) \qquad \begin{array}{ll} \min_y \quad \lambda_{\max}(U(y)) \equiv - \max -\eta & \\ \qquad\qquad\qquad\qquad\qquad -\eta\, I + U(y) & \preceq 0. \end{array}$$

What about minimizing $||U(y)||_?$?

- If "F" (for Frobenius), this is a least-squares problem.

- If "2", then this can be modeled as

$$(P) \qquad \begin{array}{rrl} -\max & & -\eta \\ & -\eta\ I + U(y) & \preceq 0 \\ & -\eta\ I - U(y) & \preceq 0 \end{array}$$

whose constraints can be written as

$$\eta \begin{bmatrix} -I & 0 \\ 0 & -I \end{bmatrix} + \begin{bmatrix} U(y) & 0 \\ 0 & -U(y) \end{bmatrix} \preceq 0.$$

Now move away from symmetry, and consider $\min ||R(y)||_?$, where $R(y) \in \mathbb{R}^{m \times n}$ depends affinely on y.

- $||R||_2 := \max \{||Rz||_2 : ||z||_2 = 1\}$.
- $||R||_F := \left(\sum_i \sum_j r_{ij}^2\right)^{1/2} = (R \bullet R)^{1/2}$.

Theorem 2 *For any $R \in \mathbb{R}^{m \times n}$, there is an orthogonal $P \in \mathbb{R}^{mxm}$, an orthogonal $Q \in \mathbb{R}^{m \times n}$, and a "diagonal" $\Sigma =$"Diag(σ)" in $\mathbb{R}^{m \times n}$, where $\sigma = (\sigma_1; ...; \sigma_l) \geq 0$, $l = \min\{m, n\}$, with $R = P\Sigma Q^T$.*

Proof:
WLOG, assume $m \geq n$ (o.w. consider R^T); then $R^T R \in \mathbb{M}_+^n$, so $R^T R = Q \Lambda Q^T$ with $\Lambda \succeq 0$, say

$$\Lambda = \begin{bmatrix} \hat{\Lambda} & 0 \\ 0 & 0 \end{bmatrix}, \ \hat{\Lambda} \in \mathbb{M}_{++}^r.$$

Define $\bar{\Sigma} := \Lambda^{1/2} = \begin{bmatrix} \hat{\Lambda}^{1/2} & 0 \\ 0 & 0 \end{bmatrix} =: \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{M}^n$ and $\bar{P} := RQ \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \\ 0 & I \end{bmatrix} \in \mathbb{R}^{m \times n}$.
Then

$$\begin{aligned} \bar{P}^T \bar{P} &= \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \\ 0 & I \end{bmatrix} Q^T R^T R Q \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{\Sigma}^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

So $\bar{P}$ has r columns of orthonormal vectors, and $(n - r)$ columns of zero vectors.
We know

$$\begin{aligned} R &= \bar{P} \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & I \end{bmatrix} Q^T = \bar{P} \bar{\Sigma} Q^T \quad \text{since the last } n - r \text{ columns of } \bar{P} \text{ are zeros} \\ &= P \begin{bmatrix} \bar{\Sigma} \\ 0 \end{bmatrix} Q^T, \end{aligned}$$

where $P \in \mathbb{R}^{n\times n}$ is orthogonal, with its first $(n-r)$ columns those of $\bar{P}$. Then we have the desired decomposition. $\square$

The factorization in the theorem is called the singular value decomposition. If $\sigma_1 \geq ... \geq \sigma_l$, then $\sigma =: \sigma(R)$, whose elements are called the singular values of R.

Note:

- $||R||_2 = \sigma_1 = ||\sigma(R)||_\infty$,
- $||R||_F = ||\sigma(R)||_2$,
- $||R||_* := ||\sigma(R)||_1$.

Also, $||\sigma||_0$ = number of positive σ_i's = rank(R) (not a norm).

Note: if $U \in \mathbb{M}^n$, $\sigma(U) = |\lambda(U)|$. Indeed, we have

$$U = Q\Lambda Q^T = \bar{Q}|\Lambda|Q^T,$$

where $\bar{Q}$ only differs from Q by the sign of some of its columns, and $|\Lambda|$ is the diagonal matrix with the absolute values of the entries of Λ. This is its singular value decomposition.