Semidefinite Programming	Lecture 4
OR 6327 Spring 2012	February 2, 2012
Scribe: Ilker Birbil	

Consider the matrix $R \in \mathbb{R}^{m\times n}$ from the last lecture and its singular value decomposition given by $R = P\Sigma Q^T$, where $P \in \mathbb{R}^{m\times m}$ and $Q \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and $\Sigma =$ "Diag (σ)" $\in \mathbb{R}^{m\times n}$. We assume for $\sigma = (\sigma_1; \cdots; \sigma_l)$ that $\sigma_1 \geq \cdots \geq \sigma_l \geq 0$ with $l = \min\{m, n\}$. We have

$$\|R\|_2 = \sigma_1,\ \|R\|_F = \|\sigma\|_2 \text{ and } \|R\|_* = \|\sigma\|_1.$$

Proposition 1 *The eigenvalues of*

$$\begin{bmatrix} 0 & R \\ R^T & 0 \end{bmatrix} \in \mathbb{M}^{m+n}$$

are $\pm\sigma_1, \cdots, \pm\sigma_n, 0, \cdots, 0.$

Proof: $R = P\Sigma Q^T$ implies that

$$\begin{bmatrix} P^T & 0 \\ 0 & Q^T \end{bmatrix}\begin{bmatrix} 0 & R \\ R^T & 0 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} 0 & \Sigma \\ \Sigma^T & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \bar{\Sigma} \\ 0 & 0 & 0 \\ \bar{\Sigma} & 0 & 0 \end{bmatrix},$$

where we assume that $m \geq n$ and $\Sigma = \begin{bmatrix} \bar{\Sigma} \\ 0 \end{bmatrix}$. Also,

$$\begin{bmatrix} \bar{I} & 0 & \bar{I} \\ 0 & I & 0 \\ \bar{I} & 0 & -\bar{I} \end{bmatrix}\begin{bmatrix} 0 & 0 & \bar{\Sigma} \\ 0 & 0 & 0 \\ \bar{\Sigma} & 0 & 0 \end{bmatrix}\begin{bmatrix} \bar{I} & 0 & \bar{I} \\ 0 & I & 0 \\ \bar{I} & 0 & -\bar{I} \end{bmatrix} = \begin{bmatrix} \bar{\Sigma} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\bar{\Sigma} \end{bmatrix},$$

where $\bar{I} := \frac{1}{\sqrt{2}} I_n$ and $I := I_{m-n}$. $\square$

Hence, minimizing $\|R(y)\|_2$ is equivalent to

$$\begin{array}{rll} -\max & -\eta & \\ & \begin{bmatrix} -\eta I_m & R(y) \\ R(y)^T & -\eta I_n \end{bmatrix} & \preceq \ 0. \end{array}$$

Proposition 2 *Suppose* $R \in \mathbb{R}^{m\times n}$ *with* $m \geq n$*; then*

$$\begin{array}{rll} 2\|R\|_* = & \min & I \bullet U + I \bullet V \\ & & \begin{bmatrix} U & R \\ R^T & V \end{bmatrix} \succeq 0, \end{array}$$

where U *and* V *are symmetric matrices.*

Proof: Again assume that $R = P\Sigma Q^T$ and $\Sigma = \begin{bmatrix} \bar{\Sigma} \\ 0 \end{bmatrix}$. Note that

$$\begin{bmatrix} U & R \\ R^T & V \end{bmatrix} \succeq 0 \iff \begin{bmatrix} \hat{U} & \Sigma \\ \Sigma^T & \hat{V} \end{bmatrix} \succeq 0,$$

where $\hat{U} = P^TUP$ and $\hat{V} = Q^TVQ$. So, minimizing $\text{trace}\,(U) + \text{trace}\,(V)$ is equivalent to minimizing $\text{trace}\,(\hat{U}) + \text{trace}\,(\hat{V})$. That is, we want to solve

$$\begin{array}{ll} \min & I \bullet \hat{U} + I \bullet \hat{V} \\ & \begin{bmatrix} \hat{U} & \Sigma \\ \Sigma^T & \hat{V} \end{bmatrix} \quad \succeq 0. \end{array}$$

If

$$\hat{U} = \begin{bmatrix} \bar{U} & \check{U} \\ \check{U}^T & \tilde{U} \end{bmatrix},$$

then we want to check whether

$$\begin{bmatrix} \bar{U} & \bar{\Sigma} & \check{U} \\ \bar{\Sigma} & \hat{V} & 0 \\ \check{U}^T & 0 & \tilde{U} \end{bmatrix} \succeq 0.$$

First the necessary conditions: $\tilde{u}_{jj} \geq 0$ for all j; $\bar{u}_{ii} \geq 0$, $\hat{v}_{ii} \geq 0$, and $\bar{u}_{ii}\hat{v}_{ii} \geq \sigma_i^2$ for all i. These conditions are <u>sufficient</u> if we set $\check{U} = 0$ and the off-diagonal entries of $\bar{U}$ and $\hat{V}$ to zero. By the arithmetic mean-geometric mean inequality, the trace is minimized by setting

$$\bar{u}_{ii} = \hat{v}_{ii} = \sigma_i \text{ for } i = 1, \cdots, n, \text{ and } \tilde{U} = 0.$$

This completes the proof. □

Thus, min $\|R(y)\|_*$ is equivalent to

$$\begin{array}{ll} \min_{U,V,y} & I \bullet U + I \bullet V \\ & \begin{bmatrix} U & R(y) \\ R(y)^T & V \end{bmatrix} \succeq 0. \end{array}$$

Maybe we are interested in

$$\begin{array}{ll} \min & \text{rank}(R) \\ & \mathcal{A}R = b. \end{array}$$

An example of this form is the minimum rank completion problem:

$$\begin{array}{ll} \min & \text{rank}(R) \\ & r_{ij} = l_{ij}, \quad ij \in K. \end{array}$$

Such problems arise in collaborative filtering, e.g., the Netflix problem, where we are trying to interpret the ranking matrix R as the result of a small number of factors, i.e., write it as PQ where P has a small number of columns.

Note that $\|R\|_* \leq \text{rank}(R)$ for all R with $\|R\|_2 \leq 1$. In fact $\|R\|_*$ is the convex envelope of $\text{rank}(R)$ on this set.

Another motivation for replacing the rank objective by the nuclear norm comes from examples. Consider first the following LP problem

$$\begin{array}{ll} \min & e^T x \\ & u^T x = \beta, \\ & x \geq 0, \end{array}$$

where $e = (1; 1; \cdots; 1)$, $u > 0$ and $\beta > 0$. The optimal solution of this problem is sparse, with just one nonzero component. In general, min $\|x\|_1$ is a proxy for getting the sparsest solution. Analogously, consider

$$\begin{array}{ll} \min & I \bullet X \\ & U \bullet X = \beta, \\ & X \succeq 0, \end{array}$$

with $U \succ 0$ and $\beta > 0$. If $U = Q \Lambda Q^T$ with $\Lambda = \Lambda(U)$, then the optimal X is given by $\left(\frac{\beta}{\lambda_1}\right) q_1 q_1^T$, with rank one. So minimizing $\|R\|_*$ is a proxy for minimizing the rank of a matrix, and we can approximate the minimum-rank problems above by instead minimizing the nuclear norm.

LP and some NLPs

Consider first an LP in dual form:

$$\begin{array}{ll} \max & b^T y \\ & A^T y \leq c. \end{array}$$

This is equivalent to

$$\begin{array}{ll} \max & b^T y \\ & \text{Diag}\,(c - A^T y) \succeq 0, \text{ or} \\ & C - \mathcal{A}^* y \succeq 0, \end{array}$$

where $C = \text{Diag}\,(c)$ and $A_i = \text{Diag}\,(a_{i1}; \cdots; a_{in})$ for all i. This is an SDP problem in dual form.

Suppose we now have

$$\begin{array}{ll} \min & c^T x \\ & Ax = b, \\ & x \geq 0. \end{array}$$

By considering the diagonal matrix $X = \text{Diag}\,(x)$, we can write

$$\begin{array}{ll} \min_{X \in \mathbb{M}^n} & C \bullet X \\ & A_i \bullet X = b_i, \quad i = 1, \cdots, m \\ & X \succeq 0, \end{array}$$

with C and the A_is as above. However, at the optimal solution X is not necessarily a diagonal matrix. This problem has both block-diagonal and sparsity structures. Without loss of generality, we can assume that X has the same block diagonal structure as C and the A_is (see HW1).

However, for general sparsity structure, we cannot assume that X has the same structure. For example, if

$$X = \begin{bmatrix} 1 & 1 & ? \\ 1 & 1 & 1 \\ ? & 1 & 1 \end{bmatrix},$$

then we would need nonzeros in the missing parts marked by '?' to make X psd. However, the dual slack S always inherits the sparsity of C and the A_is.

More examples using block-diagonal structure: suppose we want to solve

$$\begin{array}{ll} \min & \frac{(b^T y + \beta)^2}{d^T y + \delta} \\ & A^T y \le c, \end{array}$$

where we assume that $A^T y \le c$ implies $d^T y + \delta > 0$. Then,

$$\eta \ge \frac{(b^T y + \beta)^2}{d^T y + \delta} \iff \begin{bmatrix} d^T y + \delta & b^T y + \beta \\ b^T y + \beta & \eta \end{bmatrix} \succeq 0,$$

using the Schur complement. Thus, we obtain

$$\begin{array}{ll} \min & \eta \\ & \mathrm{Diag}\left(\mathrm{Diag}\,(c - A^T y), \begin{bmatrix} d^T y + \delta & b^T y + \beta \\ b^T y + \beta & \eta \end{bmatrix} \right) \succeq 0. \end{array}$$

Exercise: Extend this derivation to min $\frac{\|B^T y + b\|_2^2}{d^T y + \delta}$.

Consider an SDP problem in <u>inequality</u> form:

$$\begin{array}{ll} \min & C \bullet X \\ & A_i \bullet X \le b_i, \quad i = 1, \cdots, m \\ & X \succeq 0. \end{array}$$

Add slack variables $\xi = (\xi_i)_{i=1}^m$ and write the problem as

$$\begin{array}{ll} \min & \hat{C} \bullet \hat{X} \\ & \hat{A}_i \bullet \hat{X} = b_i, \quad i = 1, \cdots, m \\ & \hat{X} \succeq 0, \end{array}$$

where

$$\hat{C} = \begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{M}^{n+m}$$

and

$$\hat{A}_i = \begin{bmatrix} A_i & 0 \\ 0 & e_i e_i^T \end{bmatrix}, \; i = 1, \cdots, m$$

$$\left(\text{and without loss of generality } \hat{X} = \begin{bmatrix} X & 0 \\ 0 & \mathrm{Diag}\,(\xi) \end{bmatrix} \right).$$