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 \( \Sigma = \text{"Diag (} \sigma \text{"} ) \in \mathbb{R}^{m \times n} \). 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

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

**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

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

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

\[
-\max \begin{bmatrix}
-\eta I_m & R(y) \\
R(y)^T & -\eta I_n
\end{bmatrix} \succeq 0.
\]

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

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

where \( U \) and \( V \) are symmetric matrices.
Proof: Again assume that \( R = P\Sigma Q^T \) and \( \Sigma = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \). 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 \( \hat{U} = P^T UP \) and \( \hat{V} = Q^T V Q \). So, minimizing trace \( (U) + \text{trace} (V) \) is equivalent to minimizing trace \( (\hat{U}) + \text{trace} (\hat{V}) \). That is, we want to solve

\[
\min I \cdot \hat{U} + I \cdot \hat{V} \begin{bmatrix} \hat{U} & \Sigma \\ \Sigma^T & \hat{V} \end{bmatrix} \succeq 0.
\]

If

\[
\hat{U} = \begin{bmatrix} \bar{U} & \hat{U} \\ \hat{U}^T & \bar{U} \end{bmatrix},
\]

then we want to check whether

\[
\begin{bmatrix} \bar{U} & \Sigma & \bar{U} \\ \Sigma & \hat{V} & 0 \\ \bar{U}^T & 0 & \bar{U} \end{bmatrix} \succeq 0.
\]

First the necessary conditions: \( \bar{u}_{ij} \geq 0 \) 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 sufficient if we set \( \bar{U} = 0 \) 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

\[
\bar{u}_{ii} = \hat{v}_{ii} = \sigma_i \text{ for } i = 1, \ldots, n, \text{ and } \bar{U} = 0.
\]

This completes the proof. \( \square \)

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

\[
\min_{U, V, y} I \cdot U + I \cdot V \begin{bmatrix} U & R(y) \\ R(y)^T & V \end{bmatrix} \succeq 0.
\]

Maybe we are interested in

\[
\min \text{rank}(R) \quad AR = b.
\]

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

\[
\min \text{rank}(R) \quad r_{ij} = l_{ij}, \quad ij \in K.
\]

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

\[
\begin{align*}
\min & \quad e^T x \\
\text{subject to} & \quad u^T x = \beta, \\
& \quad x \geq 0,
\end{align*}
\]

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, \( \|x\|_1 \) is a proxy for getting the sparest solution. Analogously, consider

\[
\begin{align*}
\min & \quad I \cdot X \\
\text{subject to} & \quad U \cdot X = \beta, \\
& \quad X \succeq 0,
\end{align*}
\]

with \( U > 0 \) and \( \beta > 0 \). If \( U = QAQ^T \) 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:

\[
\begin{align*}
\max & \quad b^T y \\
\text{subject to} & \quad A^T y \leq c.
\end{align*}
\]

This is equivalent to

\[
\begin{align*}
\max & \quad b^T y \\
\text{subject to} & \quad \text{Diag}\left( c - A^T y \right) \succeq 0, \\
& \quad C - A^* y \succeq 0,
\end{align*}
\]

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

Suppose we now have

\[
\begin{align*}
\min & \quad c^T x \\
\text{subject to} & \quad Ax = b, \\
& \quad x \geq 0.
\end{align*}
\]

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

\[
\begin{align*}
\min_{X \in \mathbb{M}^n} & \quad C \cdot X \\
& \quad A_i \cdot X = b_i, \quad i = 1, \cdots, m \\
& \quad X \succeq 0,
\end{align*}
\]

with \( C \) and the \( A_i \)s 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 \( A_i \)s (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 $A_i$s.

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

$$\min \frac{(b^T y + \beta)^2}{d^T y + \delta}$$

$$A^T y \leq c,$$

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

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

using the Schur complement. Thus, we obtain

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

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

Consider an SDP problem in inequality form:

$$\min \ C \cdot X$$

$$A_i \cdot X \leq b_i, \quad i = 1, \ldots, m$$

$$X \succeq 0.$$

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

$$\min \ \hat{C} \cdot \hat{X}$$

$$\hat{A}_i \cdot \hat{X} = b_i, \quad i = 1, \ldots, m$$

$$\hat{X} \succeq 0,$$

where

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

and

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

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