EEOR 4650: Homework 4

Ruofan Meng, Columbia UID: rm3726

November 6, 2020

Estimated time spent on homework: 300

Question 5.6

(a)

Firstly, because $x_{ls} = \text{armin}\|Ax - b\|_2$, we have $\|Ax_{ls} - b\|_2 \leq \|Ax_{ch} - b\|_2$.
Secondly, for the fact $\frac{1}{\sqrt{m}}\|z\|_2 \leq \|z\|_\infty \leq \|z\|_2$, by taking $z_{ls} = Ax_{ls} - b$ and $z_{ch} = Ax_{ch} - b$ respectively, we could get $\|Ax_{ls} - b\|_\infty \leq \|Ax_{ls} - b\|_2$ and $\|Ax_{ch} - b\|_2 \leq \sqrt{m}\|Ax_{ch} - b\|_\infty$.
Combine the three inequalities, we have $\|Ax_{ls} - b\|_\infty \leq \sqrt{m}\|Ax_{ch} - b\|_\infty$.

(b)

For $\hat{v}$, when $r_{ls} \neq 0$, we have $\|r_{ls}\|_1 > 0$, then $\|-r_{ls}\|_1 \leq \|r_{ls}\|_1 \Leftrightarrow \frac{\|-r_{ls}\|_1}{\|r_{ls}\|_1} = \|-\frac{r_{ls}}{\|r_{ls}\|_1}\|_1 \leq 1 \Leftrightarrow \|\hat{v}\|_1 \leq 1$. Also,

$$A^T\hat{v} = \frac{A^T(-r_{ls})}{\|r_{ls}\|_1} = \frac{A^T(Ax_{ls} - b)}{\|r_{ls}\|_1} = \frac{A^TA(A^TA)^{-1}A^Tb - A^Tb}{\|r_{ls}\|_1} = 0$$

so $\hat{v}$ is feasible.
For $\tilde{v}$, $\|\tilde{v}\|_1 = \|-\hat{v}\|_1 = \|\hat{v}\|_1 \leq 1$. Also, $A^T\tilde{v} = -A^T\hat{v} = 0$. So $\hat{v}$ is feasible.

To compare these bounds, we will see that $b^T\hat{v} \leq \frac{1}{\sqrt{m}}\|Ax_{ls}-b\|_\infty \leq b^T\tilde{v}$.

$$
\begin{aligned}
b^T\tilde{v} &= \frac{1}{\|r_{ls}\|_1} b^T r_{ls} \\
&= \frac{1}{\|r_{ls}\|_1}(r_{ls}+Ax_{ls})^T r_{ls} \\
&= \frac{1}{\|r_{ls}\|_1}(\|r_{ls}\|_2^2 + r_{ls}^T A x_{ls}) \\
&= \frac{1}{\|r_{ls}\|_1}(\|r_{ls}\|_2^2 + (b-Ax_{ls})^T A x_{ls}) \\
&= \frac{1}{\|r_{ls}\|_1}(\|r_{ls}\|_2^2 + b^T A x_{ls} - x_{ls}^T A^T A x_{ls}) \\
&= \frac{1}{\|r_{ls}\|_1}(\|r_{ls}\|_2^2 + b^T A(A^TA)^{-1}A^Tb - b^TA(A^TA)^{-1}A^TA(A^TA)^{-1}A^Tb) \\
&= \frac{1}{\|r_{ls}\|_1}(\|r_{ls}\|_2^2 + b^T A(A^TA)^{-1}A^Tb - b^TA(A^TA)^{-1}A^Tb) \\
&= \frac{1}{\|r_{ls}\|_1}\|r_{ls}\|_2^2
\end{aligned}
$$

We also have $\frac{\|r_{ls}\|_1}{m} \leq \frac{\|r_{ls}\|_2}{\sqrt{m}} \Leftrightarrow \|r_{ls}\|_1 \leq \sqrt{m}\|r_{ls}\|_2$ because of the Root-Mean Square-Arithmetic Mean inequality, so $b^T\tilde{v} \geq \frac{1}{\sqrt{m}}\|r_{ls}\|_2 \geq \frac{1}{\sqrt{m}}\|r_{ls}\|_\infty = \frac{1}{\sqrt{m}}\|Ax_{ls}-b\|_\infty$. With $b^T\hat{v} = -b^T\tilde{v} \leq -\frac{1}{\sqrt{m}}\|Ax_{ls}-b\|_\infty \leq \frac{1}{\sqrt{m}}\|Ax_{ls}-b\|_\infty$, we can compare the three bounds as $b^T\hat{v} \leq \frac{1}{\sqrt{m}}\|Ax_{ls}-b\|_\infty \leq b^T\tilde{v}$, which shows $b^T\tilde{v}$ is the best lower bound of the three.

Question 5.10 (a)
With $X = \sum_{i=1} px_iv_iv_i^T$, the original problem can be converted to

$$
\begin{aligned}
\min_{X,x} \quad & \log\det X^{-1} \\
\text{s.t.} \quad & X = \sum_{i=1}^{p} x_i v_i v_i^T, \\
& \mathbf{1}^T x = 1, \\
& x \succeq 0
\end{aligned}
$$

The Lagrangian function is $L(X,x,Y,\lambda,\mu) = \log\det X^{-1} + \mathbf{Tr}Y(X - \sum_{i=1}^p x_iv_iv_i^T) + \lambda(\mathbf{1}^Tx - 1) - \mu^T x$ with $\mu \succeq 0$.

The dual function is

$$
\begin{aligned}
g(Y,\lambda,\mu) &= \inf_{X,x} L(X,x,Y,\lambda,\mu) \\
&= \inf_{X,x} \log\det X^{-1} + \mathbf{Tr} Y(X - \sum_{i=1}^{p} x_i v_i v_i^T) + \lambda(\mathbf{1}^T x - 1) - \mu^T x \\
&= \inf_{X,x} \log\det X^{-1} + \mathbf{Tr} YX - \sum_{i=1}^{p} x_i \mathbf{Tr} Y v_i v_i^T + \lambda(\mathbf{1}^T x - 1) - \mu^T x \\
&= \inf_{X,x} \log\det X^{-1} + \mathbf{Tr} YX - \sum_{i=1}^{p} x_i v_i^T Y v_i + \lambda(\mathbf{1}^T x - 1) - \mu^T x \\
&= \log\det X^{-1} + \mathbf{Tr} YX + (\lambda\mathbf{1} - v - \mu)^T x - \lambda \\
&= \log\frac{1}{\det X} + \mathbf{Tr} YX + (\lambda\mathbf{1} - v - \mu)^T x - \lambda \\
&= -\log\det X + \mathbf{Tr} YX + (\lambda\mathbf{1} - v - \mu)^T x - \lambda
\end{aligned}
$$

where v is the vector whose i's entry is $v_i^T Y v_i$. Obviously, $\inf_{X,x} L(X,x,Y,\lambda,\mu) = -\infty$ if $\lambda\mathbf{1} - v - \mu \neq 0$. When $\lambda\mathbf{1} - v - \mu = 0$, $L(X,x,Y,\lambda,\mu) = -\log\det X + \mathbf{Tr} YX - \lambda$.
With the fact that $\log\det X$ is a concave function of X, $-\log\det X$ is convex function of X, $\mathbf{Tr} YX$ is affine function of X, hence $L(X,x,Y,\lambda,\mu)$ is convex in (X,x) given any (Y,λ,μ) when $\lambda\mathbf{1} - v - \mu = 0$.
For $X \in \mathbf{S}^n_{++}$, note $\mathbf{Tr} YX = \mathbf{Tr}(YX)^T = \mathbf{Tr} X^T Y^T = \mathbf{Tr} XY^T = \mathbf{Tr} Y^T X$. Let $\tilde{Y} = \frac{1}{2}(Y + Y^T)$, we have $\mathbf{Tr}\tilde{Y}X = \frac{1}{2}(\mathbf{Tr} Y^T X + \mathbf{Tr} YX) = \mathbf{Tr} YX$ and $L(X,x,Y,\lambda,\mu) = -\log\det X + \mathbf{Tr}\tilde{Y}X - \lambda$.
To find a local minimum, we take the derivative with respect to X and get $\frac{\partial L}{\partial X} = -X^{-1} + \tilde{Y}$.
We will see that if $\tilde{Y} \notin \mathbf{S}^n_{++}$, $\inf_{X,x} L(X,x,Y,\lambda,\mu) = -\infty$. Note that $\tilde{Y}$ is a symmetric matrix, suppose $\tilde{Y} \notin \mathbf{S}^n_{++}$ and its spectral decomposition is $\tilde{Y} = PDP^T$, then it must have an eigenvalue satisfying $m_i \leq 0$. Take $X = PA(t)P^T$, where $A(t)$ is a diagonal matrix with $A_{ii} = t$ and $A_{jj} = 1$ for $j \neq i$. When $\lambda\mathbf{1} - v - \mu = 0$ and $t > 0$, $X \in \mathbf{S}^n_{++}$,

$$
\begin{aligned}
L(X,x,Y,\lambda,\mu) &= -\log\det PA(t)P^T + +\mathbf{Tr}\tilde{Y}X - \lambda \\
&= -\sum_{j\neq i} \log 1 - \log t + \mathbf{Tr} PDP^T PA(t)P^T - \lambda \\
&= -\log t + \mathbf{Tr} DA(t) - \lambda \\
&= -\log t + \sum_{j\neq i} m_j + m_i t
\end{aligned}
$$

With $m_i \leq 0$, $L(X,x,Y,\lambda,\mu) \to -\infty$ as $t \to +\infty$.
When $\tilde{Y} \in \mathbf{S}^n_{++}$, at $X = \tilde{Y}^{-1}$, the derivative with respect to X is 0 and because L is convex in X, it must attain its minimum when $X = \tilde{Y}^{-1}$.
Therefore, $g(Y,\lambda,\mu) = \log\det\tilde{Y} + n - \lambda$ when $\lambda\mathbf{1} - v - \mu = 0$ and $\tilde{Y} \in \mathbf{S}^n_{++}$, otherwise

$g(Y, \lambda, \mu) = -\infty$. Therefore, the dual problem is

$$\begin{array}{ll} \max\limits_{Y, \lambda, \mu} & \log\det \tilde{Y} + n - \lambda \\ \text{s.t.} & \lambda \mathbf{1} - v - \mu = 0, \\ & Y + Y^T \succ 0, \\ & \mu \succeq 0 \end{array}$$

which is equivalent to

$$\begin{array}{ll} \max\limits_{Z, \lambda} & \log\det Z + n - \lambda \\ \text{s.t.} & v_i^T Z v_i \leq \lambda, \\ & Z \succ 0 \end{array}$$

because $v_i^T Y v_i = v_i^T Y^T v_i$ and $v_i^T Y v_i = \frac{1}{2} v_i^T (Y + Y^T) v_i$.

Question 6.1

The Lagrangian function should be

$$L(x, y, \lambda, \mu, \nu) = -c^T x + \sum_{i=1}^{m} y_i \log y_i - \lambda^T x + \mu^T (Px - y) + \nu(\mathbf{1}^T x - 1)$$

for $\lambda \succeq 0$. The dual function should be

$$\begin{aligned} g(\lambda, \mu, \nu) &= \inf_{x,y} L(x, y, \lambda, \mu, \nu) \\ &= \inf_{x,y} -c^T x + \sum_{i=1}^{m} y_i \log y_i - \lambda^T x + \mu^T (Px - y) + \nu(\mathbf{1}^T x - 1) \\ &= \inf_{x,y} (P^T \mu - c - \lambda + \nu \mathbf{1})^T x + \sum_{i=1}^{m} y_i (\log y_i - \mu_i) - \nu \end{aligned}$$

Obviously, when $P^T\mu - c - \lambda + \nu\mathbf{1} \neq 0$, $g(\lambda, \mu, \nu)$ is unbounded below. When $P^T\mu - c - \lambda + \nu\mathbf{1} = 0$, $g(\lambda, \mu, \nu) = \sum_{i=1}^{m} y_i(\log y_i - \mu_i) - \nu$. Also, $\frac{\partial g}{\partial y_i} = \log y_i + 1 - \mu_i$ and $\frac{\partial^2 g}{\partial y_i^2} = \frac{1}{y_i} > 0$. So g is convex in each y_i and attains its minimum when $y_i = e^{\mu_i - 1}$ for all i.

Thus we have

$$g(\lambda, \mu, \nu) = -\sum_{i=1}^{m} e^{\mu_i - 1} - \nu$$

when $P^T\mu - c - \lambda + \nu\mathbf{1} = 0$ and $g(\lambda, \mu, \nu) = -\infty$ otherwise.

The dual problem should be

$$\begin{array}{ll} \max\limits_{\lambda, \mu, \nu} & -\sum\limits_{i=1}^{m} e^{\mu_i - 1} - \nu \\ \text{s.t.} & P^T \mu - c - \lambda + \nu \mathbf{1} = 0, \\ & \lambda \succeq 0 \end{array}$$

Regard λ as a slack variable, we transfer constraints $P^T\mu - c - \lambda + \nu\mathbf{1} = 0$ and $\lambda \succeq 0$ into $P^T\mu + \nu\mathbf{1} \succeq c$ and then discard λ. With $\omega = \mu + \nu\mathbf{1}$, the dual problem is equivalent to

$$\begin{aligned} \max_{\omega,\,\nu} \quad & -\sum_{i=1}^m e^{\omega_i - \nu - 1} - \nu \\ \text{s.t.} \quad & P^T\omega \succeq c \end{aligned}$$

where $P^T\omega = P^T(\mu + \nu\mathbf{1}) = P^T\mu + \nu\mathbf{1}$.
Because

$$\max_{\omega,\nu} - \sum_{i=1}^m e^{\omega_i - \nu - 1} - \nu = \max_{\omega} \max_{\nu} -e^{-\nu-1}\sum_{i=1}^m e^{\omega_i} - \nu$$

and for $h(\nu) = -e^{-\nu-1}\sum_{i=1}^m e^{\omega_i} - \nu$, we have $h' = e^{-\nu-1}\sum_{i=1}^m e^{\omega_i} - 1$ and $h'' = -e^{-\nu-1}\sum_{i=1}^m e^{\omega_i} < 0$. Thus $h(\nu)$ is a concave function and attains its maximum when $h' = 0 \Leftrightarrow v = \log(\sum_{i=1}^m e^{\omega_i}) - 1$.

$$\begin{aligned} \max_{\nu} h(\nu) &= h(\log(\sum_{i=1}^m e^{\omega_i}) - 1) \\ &= 1 - \log(\sum_{i=1}^m e^{\omega_i}) - \sum_{i=1}^m e^{w_i - 1 + 1 - \log(\sum_{i=1}^m e^{\omega_i})} \\ &= 1 - \log(\sum_{i=1}^m e^{\omega_i}) - e^{-\log(\sum_{i=1}^m e^{\omega_i})}\sum_{i=1}^m e^{w_i} \\ &= 1 - \log(\sum_{i=1}^m e^{\omega_i}) - (\sum_{i=1}^m e^{\omega_i})^{-1}\sum_{i=1}^m e^{w_i} \\ &= -\log(\sum_{i=1}^m e^{\omega_i}) \end{aligned}$$

Therefore, the original dual problem is equivalent to

$$\begin{aligned} \max_{\omega} \quad & -\log(\sum_{i=1}^m e^{\omega_i}) - 1 \\ \text{s.t.} \quad & P^T\omega \succeq c \end{aligned}$$

Question 6.2
(i)
Let $y_i = A_i x - b_i$ and $t_i = c_i^T x + d_i$, the original problem is equivalent to

$$\begin{aligned} \min_{x,\,y,\,t} \quad & f^T x \\ \text{s.t.} \quad & \|y_i\|_2 \le t_i, \\ & y_i = A_i x - b_i, \\ & t_i = c_i^T x + d_i \end{aligned}$$

The Lagrangian function is

$$
\begin{aligned}
& L(x, y, t, \lambda, \mu, \nu) \\
= & f^T x + \sum_{i=1}^m \lambda_i(\|y_i\|_2 - t_i) + \sum_{i=1}^m \mu_i^t(A_i x - b_i - y_i) + \sum_{i=1}^m \nu_i(c_i^T x + d_i - t_i) \\
= & (f + \sum_{i=1}^m A_i^T \mu_i + \nu_i c_i)^T x + \sum_{i=1}^m (\lambda_i \|y_i\|_2 - \mu_i^T y_i) - \sum_{i=1}^m (\lambda_i + \nu_i) t_i - \sum_{i=1}^m \mu_i b_i + \sum_{i=1}^m \nu_i d_i \\
= & (f + \sum_{i=1}^m A_i^T \mu_i + \nu_i c_i)^T x + \sum_{i=1}^m (\lambda_i \|y_i\|_2 - \mu_i^T y_i) - (\lambda + \nu)^T t + \nu^T d - \sum_{i=1}^m \mu_i b_i
\end{aligned}
$$

for $\lambda_i \succeq 0$.
The dual function should be

$$
\begin{aligned}
& g(\lambda, \mu, \nu) \\
= & \inf_{x,y,t} L(x, y, t, \lambda, \mu, \nu) \\
= & \inf_{x,y,t} (f + \sum_{i=1}^m A_i^T \mu_i + \nu_i c_i)^T x + \sum_{i=1}^m (\lambda_i \|y_i\|_2 - \mu_i^T y_i) - (\lambda + \nu)^T t + \nu^T d - \sum_{i=1}^m \mu_i b_i
\end{aligned}
$$

Obviously, L is unbounded below unless $f + \sum_{i=1}^m A_i^T \mu_i + \nu_i c_i = 0$ and $\lambda + \nu = 0$. When these two conditions are satisfied, $L(x, y, t, \lambda, \mu, \nu) = \sum_{i=1}^m (\lambda_i \|y_i\|_2 - \mu_i^T y_i) + \nu^T d - \sum_{i=1}^m \mu_i b_i$. We will see that only if $\lambda_i \geq \|\mu_i\|_2$ for all i, L is bounded below.
Suppose for some i, $0 \leq \lambda_i < \|\mu_i\|_2$, we could take $y_i = \theta \mu_i$ and $y_j = 0$ for $j \neq i$. Then $L = \lambda_i \|\mu_i\|_2 - \theta \|\mu_i\|_2^2 + \nu^T d - \mu^T b = \|\mu_i\|_2 (\lambda_i - \theta \|\mu_i\|_2) + \nu^T d - \mu^T b \to -\infty$ as $\theta \to +\infty$. And when $\lambda_i \geq \|\mu_i\|_2$ for all i, $L = \sum_{i=1}^m (\lambda_i - \cos \pi_i \|\mu_i\|_2) \|y_i\|_2 + \nu^T d - \mu^T b$ where π_i are the angle between μ_i and y_i. With $\lambda_i - \cos \pi_i \|\mu_i\|_2 \geq 0$ the minimum of L is 0 in each direction y. And thus across all directions, the minimum of L is 0, which means $g(\lambda, \mu, \nu) = \nu^T d - \sum_{i=1}^m \mu_i b_i$.
Therefore, the dual problem is

$$
\begin{aligned}
\max_{\lambda, \mu, \nu} \quad & \nu^T d - \sum_{i=1}^m \mu_i b_i \\
\text{s.t.} \quad & f + \sum_{i=1}^m A_i^T \mu_i + \nu_i c_i = 0, \\
& \lambda + \nu = 0, \\
& \|\mu_i\|_2 \leq \lambda_i
\end{aligned}
$$

which is equivalent to

$$\begin{aligned} \max_{\lambda,\,\mu} \quad & -\lambda^T d - \sum_{i=1}^m \mu_i b_i \\ \text{s.t.} \quad & f + \sum_{i=1}^m A_i^T \mu_i = \lambda_i c_i, \\ & \|\mu_i\|_2 \le \lambda_i \end{aligned}$$

(ii)
Let $K_i = \{(y,t) \in \mathbf{R}^{m_i+1} \mid \|x\|_2 \le t\}$, a proper cone, the original SOCP is equivalent to its conic form

$$\begin{aligned} \min_{x} \quad & f^T x \\ \text{s.t.} \quad & \begin{bmatrix} A_i x - b_i \\ c_i^T x + d_i \end{bmatrix} \succeq_{K_i} 0 \end{aligned}$$

because $\|A_i x - b_i\|_2 \le c_i^T x + d_i \Leftrightarrow ((A_i x - b_i)^T, c_i^T x + d_i) \in K_i$.
The Lagrangian function should be

$$L(x,\mu,\lambda) = f^T x - \sum_{i=1}^m \mu_i^T (A_i x - b_i) - \sum_{i=1}^m \lambda_i (c_i^T x + d_i)$$

for $\begin{bmatrix} \mu_i \\ \lambda_i \end{bmatrix} \succeq_{K_i^*} 0 \Leftrightarrow \begin{bmatrix} \mu_i \\ \lambda_i \end{bmatrix} \succeq_{K_i} 0 \Leftrightarrow \|\mu_i\|_2 \le \lambda_i$. The dual function should be

$$\begin{aligned} g(\mu,\lambda) &= \inf_x L(x,\mu,\lambda) \\ &= \inf_x f^T x - \sum_{i=1}^m \mu_i^T (A_i x - b_i) - \sum_{i=1}^m \lambda_i (c_i^T x + d_i) \\ &= \inf_x (f - \sum_{i=1}^m A_i^T \mu_i - \lambda c_i)^T x + \sum_{i=1}^m \mu_i^T b_i - \sum_{i=1}^m \lambda_i d_i \\ &= \inf_x (f - \sum_{i=1}^m A_i^T \mu_i - \lambda c_i)^T x + \sum_{i=1}^m \mu_i^T b_i - \lambda^T d \end{aligned}$$

Obviously, $g(\mu,\lambda) = -\infty$ unless $f - \sum_{i=1}^m A_i^T \mu_i - \lambda c_i = 0$. The dual problem should be

$$\begin{aligned} \max_{\mu,\,\lambda} \quad & \sum_{i=1}^m \mu_i^T b_i - \lambda^T d \\ \text{s.t.} \quad & f - \sum_{i=1}^m A_i^T \mu_i - \lambda c_i = 0, \\ & \|\mu_i\|_2 \le \lambda_i \end{aligned}$$

which is equivalent to (1) with change of variables.

Question 6.3
(i)

$$
\begin{aligned}
\|X\|_{2\star} &= \{\sup \mathbf{Tr} X^T Y \mid \|Y\|_2 \le 1\} \\
&= \left\{\sup \mathbf{Tr} X^T Y \mid \sup_{x \in \{x \mid \|x\|_2 = 1\}} x^T Y^T Y x \le 1\right\} \\
&= \left\{\sup \mathbf{Tr} X^T Y \mid Y^T Y \preceq I_n\right\} \qquad (1)
\end{aligned}
$$

(1) holds because when $Y^TY \preceq I_n$, $x^T(I_n - Y^TY)x = \|x\|_2^2 - x^TY^TYx \ge 0$ for all x satisfying $\|x\|_2 = 1$, and then we have $x^TY^TYx \le 1$ and $\sup_{x\in\{x \mid \|x\|_2=1\}} x^TY^TYx \le 1$. When $\sup_{x\in\{x \mid \|x\|_2=1\}} x^TY^TYx \le 1$, it must be that $x^TY^TYx \le 1$ for all x such that $\|x\|_2 = 1$, otherwise the supremium will be greater than 1. For any $v \ne 0$, $\|\frac{v}{\|v\|_2}\|_2 = 1$, thus we have $\frac{v^T}{\|v\|_2} Y^TY \frac{v}{\|v\|_2} \le 1 \Leftrightarrow v^TY^TYv \le v^Tv \Leftrightarrow v^T(I_n - Y^TY)v \ge 0 \Leftrightarrow I_n \succeq Y^TY$.
Based on the property of Schur complement, $I_n \succeq Y^TY \Leftrightarrow \begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix} \succeq 0$.
So the original problem is equivalent to SDP1:

$$
\begin{aligned}
\max_Y \quad & \mathbf{Tr} X^T Y \\
\text{s.t.} \quad & \begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix} \succeq 0
\end{aligned}
$$

Let Y_{ij} be the row i and column j entry of Y and E_{ij} be the matrix with the row i and column j entry being 0 and all other entries being 0, the above problem is equivalent to SDP2:

$$
\begin{aligned}
\max_Y \quad & \sum_{i=1}^m \sum_{j=1}^n X_{ij} Y_{ij} \\
\text{s.t.} \quad & \begin{bmatrix} I_m & 0 \\ 0 & I_n \end{bmatrix} + \sum_{i=1}^m \sum_{j=1}^n Y_{ij} \begin{bmatrix} 0 & E_{ij} \\ E_{ij}^T & 0 \end{bmatrix} \succeq 0
\end{aligned}
$$

which is an SDP.
To derive the dual, we use form SDP1. Note that $\max_Y \mathbf{Tr} X^T Y = -\min_Y \mathbf{Tr} X^T(-Y)$. Let SDP3 be

$$
\begin{aligned}
\min_Y \quad & \mathbf{Tr} X^T (-Y) \\
\text{s.t.} \quad & \begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix} \succeq 0
\end{aligned}
$$

It is equivalent to SDP4

$$\begin{array}{ll} \min\limits_{Y} & \mathbf{Tr} X^T Y \\ \text{s.t.} & \begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix} \succeq 0 \end{array}$$

because $I_n - Y^T Y = I_n - (-Y)^T(-Y)$. And the Lagrangian of SDP4 should be

$$L(Y,\lambda) = \mathbf{Tr} X^T Y - \mathbf{Tr} \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}^T \begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix}$$

for $\lambda \succeq 0$.
The dual function of SDP4 then should be

$$\begin{aligned} g(\lambda) &= \inf_Y \mathbf{Tr} X^T Y - \mathbf{Tr} \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}^T \begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix} \\ &= \inf_Y \mathbf{Tr} X^T Y - \mathbf{Tr} \lambda_{11}^T I_m - \mathbf{Tr} \lambda_{12}^T Y - \mathbf{Tr} \lambda_{21}^T Y^T - \mathbf{Tr} \lambda_{22}^T I_n \\ &= \inf_Y \mathbf{Tr} (X - 2\lambda_{12})^T Y - \mathbf{Tr} \lambda_{11} - \mathbf{Tr} \lambda_{22} \end{aligned}$$

Obviously, $g(\lambda) = -\infty$ when $X - 2\lambda_{12} \neq 0$ and $g(\lambda) = -\mathbf{Tr}\lambda_{11} - \mathbf{Tr}\lambda_{22}$ when $X - 2\lambda_{12} = 0$.
The dual problem of SDP4 should be

$$\begin{array}{ll} \max\limits_{\lambda} & -\mathbf{Tr} \lambda_{11} - \mathbf{Tr} \lambda_{22} \\ \text{s.t.} & \begin{bmatrix} \lambda_{11} & \frac{1}{2} X \\ \frac{1}{2} X^T & \lambda_{22} \end{bmatrix} \succeq 0 \end{array}$$

The dual problem of the original problem SDP1 should be

$$\begin{array}{ll} \min\limits_{\lambda} & \mathbf{Tr} \lambda_{11} + \mathbf{Tr} \lambda_{22} \\ \text{s.t.} & \begin{bmatrix} \lambda_{11} & \frac{1}{2} X \\ \frac{1}{2} X^T & \lambda_{22} \end{bmatrix} \succeq 0 \end{array}$$

(ii)

In [2]:
```
m=10
n=15
np.random.seed(123)
X=np.random.randn(m,n)
```

In [3]:
```
lmbda = cp.Variable((m+n,m+n), symmetric=True)
constraints = [lmbda >> 0]
constraints += [lmbda[:m,m:]==X/2]
prob = cp.Problem(cp.Minimize(cp.trace(lmbda[:m,:m])+cp.trace(lmbda[m:,m:])), constraints)
prob.solve()
```

Out[3]: 38.518160114179366

In [4]:
```
a=np.linalg.svd(X)
np.sum(a[1])
```

Out[4]: 38.518160113977956

The analytic solution gives 38.518160113977956 while the upper bound gives 38.518160114179366, slightly greater than the solution.
(iii)

In [22]:
```
X = cp.Variable((m,n), symmetric=False)
lmbda = cp.Variable((m+n,m+n), symmetric=True)
M0=M_observed.reshape(m,n)
constraints = [lmbda >> 0]
constraints += [lmbda[:m,m:]==X/2]
constraints += [X[~np.isnan(M0)]==M0[~np.isnan(M0)]]
prob = cp.Problem(cp.Minimize(cp.trace(lmbda[:m,:m])+cp.trace(lmbda[m:,m:])), constraints)
prob.solve()
```

Out[22]: 27.085822325441626

In [28]:
```
np.linalg.norm(X.value-M,ord='fro')
```

Out[28]: 0.05278834819665631

The optimization problem is

$$\begin{aligned}\min_{X,\lambda} \quad & \mathbf{Tr}\lambda_{11} + \mathbf{Tr}\lambda_{22} \\ \text{s.t.} \quad & \begin{bmatrix} \lambda_{11} & \frac{1}{2}X \\ \frac{1}{2}X^T & \lambda_{22} \end{bmatrix} \succeq 0, \\ & X_{ij} = M_{ij}\end{aligned}$$

Because for any fixed X, the optimization will give a tight upper bound of $\|X\|_{2\star}$, when minimizing over X and λ, it will roughly give a solution which minimizes $\|X\|_{2\star}$.
Let the test measure be the Frobenius norm $\|X - M\|_F = \sqrt{\mathbf{Tr}(X-M)^T(X-M)}$, the final result is $\|X - M\|_F = 0.05278834819665631$, which means the root mean squre of the deviation is about 0.05. With $\frac{\sum_{i,j}|M_{ij}|}{mn} = 0.28$, the result is acceptable considering that the deviation is just about 20% of the entries of the original matrix.