Homework 7

Submit answers for problems 1–5.

1. Exercise T5.26.
2. Exercise A5.22.
3. Exercise A5.4.
4. Exercise T5.30.
5. Exercise A15.12.
6. Exercise A5.15.

Solution. This is a convex problem with three equality constraints

$$\begin{array}{ll}\text{minimize} & f_0(X) \\ \text{subject to} & h_1(X) = 0, \\ & h_2(X) = 0, \\ & h_3(X) = 0,\end{array}$$

where $f_0(X) = -\log\det X$ with domain $\mathbf{S}^n_{++}$, and

$$\begin{array}{rcl} h_1(X) & = & \mathbf{tr}\left(\left[\begin{array}{cc} I & 0 \\ 0 & 0 \end{array}\right] X\right) - \alpha, \\ h_2(X) & = & \dfrac{1}{2}\,\mathbf{tr}\left(\left[\begin{array}{cc} 0 & I \\ I & 0 \end{array}\right] X\right) - \beta, \\ h_3(X) & = & \mathbf{tr}\left(\left[\begin{array}{cc} 0 & 0 \\ 0 & I \end{array}\right] X\right) - \gamma. \end{array}$$

The optimality conditions are:

$$X \in \mathbf{S}^n_{++}, \qquad h_1(X) = h_2(X) = h_3(X) = 0, \qquad \nabla f_0(X) + \sum_{i=1}^{3} \nu_i \nabla h_i(X) = 0.$$

The third condition is

$$\begin{array}{rcl} 0 & = & -X^{-1} + \nu_1 \left[\begin{array}{cc} I & 0 \\ 0 & 0 \end{array}\right] + \dfrac{\nu_2}{2}\left[\begin{array}{cc} 0 & I \\ I & 0 \end{array}\right] + \nu_3 \left[\begin{array}{cc} 0 & 0 \\ 0 & I \end{array}\right] \\ & = & -X^{-1} + \left[\begin{array}{cc} \nu_1 I & (\nu_2/2) I \\ (\nu_2/2) I & \nu_3 I \end{array}\right], \end{array}$$

where ν_1, ν_2, ν_3 are multipliers for the three equality constraints. Solving for X gives

$$X = \begin{bmatrix} \nu_1 I & (\nu_2/2) I \\ (\nu_2/2) I & \nu_3 I \end{bmatrix}^{-1}.$$

The inverse is given by

$$\begin{bmatrix} \nu_1 I & (\nu_2/2) I \\ (\nu_2/2) I & \nu_3 I \end{bmatrix}^{-1} = \begin{bmatrix} \lambda_1 I & \lambda_2 I \\ \lambda_2 I & \lambda_3 I \end{bmatrix}$$

where

$$\begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_2 & \lambda_3 \end{bmatrix} = \begin{bmatrix} \nu_1 & (\nu_2/2) \\ (\nu_2/2) & \nu_3 \end{bmatrix}^{-1}.$$

Hence the optimal X must be of the form

$$X = \begin{bmatrix} \lambda_1 I & \lambda_2 I \\ \lambda_2 I & \lambda_3 I \end{bmatrix}.$$

The three coefficients $\lambda_1, \lambda_2, \lambda_3$ (equivalently, the multipliers ν_1, ν_2, ν_3) are easily determined from the feasibility conditions $\mathbf{tr}\, X_1 = \alpha$, $\mathbf{tr}\, X_2 = \beta$, $\mathbf{tr}\, X_3 = \gamma$:

$$\lambda_1 = \alpha/n, \qquad \lambda_2 = \beta/n, \qquad \lambda_3 = \gamma/n.$$

We conclude that the optimal solution is

$$X = \frac{1}{n} \begin{bmatrix} \alpha I & \beta I \\ \beta I & \gamma I \end{bmatrix}.$$

7. Exercise A5.28.

Solution. Consider the ith constraint

$$\sup_{a_i \in P_i} \max \{a_i^T x - b_i, \ -a_i^T x + b_i\} \leq t_i. \tag{1}$$

On the left-hand side we note that

$$\sup_{a_i \in P_i} \max \{a_i^T x - b_i, \ -a_i^T x + b_i\} = \max \{ \sup_{a_i \in P_i} (a_i^T x - b_i), \ \sup_{a_i \in P_i} (-a_i^T x + b_i)\}.$$

Therefore constraint (1) is satisfied if and only if

$$\sup_{a_i \in P_i} (a_i^T x - b_i) \leq t_i \qquad \text{and} \qquad \sup_{a_i \in P_i} (-a_i^T x + b_i) \leq t_i. \tag{2}$$

From linear programming duality,

$$\sup_{C_i a_i \preceq d_i} a_i^T x = \inf_{\substack{C_i^T z_i = x \\ z_i \succeq 0}} d_i^T z_i, \qquad \sup_{C_i a_i \preceq d_i} -a_i^T x = \inf_{\substack{C_i^T w_i = -x \\ w_i \succeq 0}} d_i^T w_i,$$

so (2) holds if and only if there exist z_i, w_i with

$$d_i^T z_i - b_i \le t_i, \quad C_i^T z_i = x, \quad z_i \succeq 0, \qquad d_i^T w_i + b_i \le t_i, \quad C_i^T w_i = -x, \quad w_i \succeq 0.$$

Substituting these conditions for (1) results in the QP

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m t_i^2 \\ \text{subject to} & d_i^T z_i - b_i \le t_i, \quad d_i^T w_i + b_i \le t_i, \quad i = 1, \ldots, m \\ & x = C_i^T z_i = -C_i^T w_i, \quad i = 1, \ldots, m \\ & z_i \succeq 0, \quad w_i \succeq 0, \quad i = 1, \ldots, m. \end{array}$$

The variables are $x \in \mathbf{R}^n$, $t_1, \ldots, t_m$, $z_1, \ldots, z_m$, $w_1, \ldots, w_m$.

8. *Minimizing sum of largest constraint violations.* Let $f_1, \ldots, f_m$ be convex functions. We use the notation $f_{[1]}(x), \ldots, f_{[m]}(x)$ for the function values $f_1(x), \ldots, f_m(x)$ sorted in decreasing order:

$$f_{[1]}(x) \ge f_{[2]}(x) \ge \cdots \ge f_{[m]}(x).$$

Note that the ordering depends on x. Define

$$\phi(x) = \sum_{i=1}^r \max\{f_{[i]}(x), 0\}$$

where r is an integer between 1 and m. To compute $\phi(x)$, we evaluate $f_1(x), \ldots, f_m(x)$, sort them in descending order, replace the negative values with 0, and add the first r values.

(a) Show that $\phi(x)$ is the optimal value of the linear program

$$\begin{array}{ll} \text{maximize} & y_1 f_1(x) + \cdots + y_m f_m(x) \\ \text{subject to} & 0 \preceq y \preceq \mathbf{1} \\ & y_1 + \cdots + y_m \le r, \end{array}$$

with variable $y \in \mathbf{R}^m$.

(b) Is ϕ a convex function?

(c) Show that

$$\phi(x) = \inf_{t \ge 0} \left(rt + \sum_{i=1}^m \max\{f_i(x) - t, 0\} \right).$$

(d) Suppose $t^\star$ is optimal for the one-dimensional minimization problem in this expression for $\phi(x)$ in part (c). Show that $\#\{i \mid f_i(x) > t^\star\} \le r$. In other words, no more than r of the function values $f_i(x)$ exceed $t^\star$.

Remark. The function ϕ has been proposed as a penalty or loss function for classification problems, portfolio optimization, ellipsoidal fitting problems, and other applications. The result in part (c) shows that we can minimize $f_0(x)+\phi(x)$ over x by solving the optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) + rt + \sum_{i=1}^m \max\{f_i(x), 0\} \\ \text{subject to} & t \geq 0 \end{array}$$

with variables x and t.

Solution.

(a) Suppose $f_1(x) \geq f_2(x) \geq \cdots \geq f_m(x)$. If $f_r(x) > 0$, the choice $y_1 = \cdots = y_r = 1$ and $y_{r+1} = \cdots = y_m = 0$ is optimal. Otherwise, let j be the largest j with $f_j(x) > 0$, and choose $y_1 = \cdots = y_j = 1$, $y_{j+1} = \cdots = y_m = 0$. If $f_i(x) \leq 0$ for all i, then $y = 0$ is optimal.

(b) For fixed $y \succeq 0$, the function $y_1 f_1(x) + \cdots + y_m f_m(x)$ is convex. The result in part (a) shows that

$$\phi(x) = \sup_{y \in C} (y_1 f_1(x) + \cdots + y_m f_m(x))$$

where $y = \{y \mid 0 \preceq y \preceq \mathbf{1},\ \mathbf{1}^T y \leq r\}$. Therefore ϕ is convex.

(c) The minimization problem in this expression is equivalent to the LP

$$\begin{array}{ll} \text{minimize} & rt + \mathbf{1}^T u \\ \text{subject to} & f_i(x) - t \leq u_i, \quad i = 1, \ldots, m \\ & u \succeq 0 \\ & t \geq 0 \end{array}$$

with variables t and u. The problem in part (a) is the dual of this. Define the Lagrangian

$$\begin{aligned} L(t,u,y,z,w) &= rt + \mathbf{1}^T u + \sum_{i=1}^m y_i (f_i(x) - t - u_i) - z^T u - wt \\ &= \sum_{i=1}^m y_i f_i(x) + (r - \mathbf{1}^T y - w)t + (\mathbf{1} - y - z)^T u. \end{aligned}$$

The infimum is $-\infty$ unless $\mathbf{1}^T y + w = r$ and $y + z = \mathbf{1}$, so the the dual is

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^m y_i f_i(x) \\ \text{subject to} & y + z = \mathbf{1} \\ & \mathbf{1}^T y + w = r \\ & y \succeq 0, \quad z \succeq 0, \quad w \geq 0. \end{array}$$

Eliminating the slack variables z and w gives the LP in part (a).

(d) The statement about $t^\star$ can be shown by complementary slackness. If $f_i(x) > t^\star$, we have $u_i > 0$ in the primal LP. Complementary slackness requires $z_i = 0$, hence $y_i = 1$. Therefore

$$\#\{i \mid f_i(x) > t^\star\} = \sum_{i:f_i(x)>t^\star} y_i \le \sum_{i=1}^m y_i \le r.$$

We can also see this directly from the one-dimensional minimization problem in t. We are minimizing a piecewise-linear function of t. The slopes of the segments increase from $r - m$ (if $t < f_{[m]}(x)$) to r (if $t > f_{[1]}(x)$). If $r < m$, the slope is zero on the interval $\left[f_{[r+1]}(x),\ f_{[r]}(x)\right]$. With the $t \ge 0$ constraint, the optimal set is

$$I = \left[\max\{f_{[r+1]}(x), 0\},\ \max\{f_{[r]}(x), 0\}\right].$$

If $r = m$, the optimal set is $I = \left[0, \max\{f_{[m]}(x), 0\}\right]$. Since $t^\star \in I$, not more than r values of $f_i(x)$ are greater than $t^\star$.

9. Exercise A5.8.

Solution. The unconstrained problem can be written as an SDP

$$\begin{array}{ll} \text{minimize} & c^T x + t \\ \text{subject to} & F(x) \preceq tI \\ & t \ge 0. \end{array} \tag{3}$$

The dual of this problem is

$$\begin{array}{ll} \text{maximize} & \mathbf{tr}(F_0 Z) \\ \text{subject to} & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, m \\ & \mathbf{tr}\, Z + s = 1 \\ & Z \succeq 0, \quad s \ge 0. \end{array} \tag{4}$$

The difference with the original dual problem is the addition of an upper bound $\mathbf{tr}\, Z \le 1$ (written as $\mathbf{tr}\, Z + s = 1$ for $s \ge 0$). We see that $Z^\star$ satisfies this constraint, with $s > 0$. Therefore it is optimal for (4). By complementary slackness we have $t = 0$ at the optimum of the primal problem (3). The optimal x for (3) is therefore optimal for the original SDP.