Homework #4

Problem 1: Power Cone

Besides the second order cone and positive semidefinite cone, power cone is also very popular and can be used to model many convex programs. For a given \( \alpha = (\alpha_1, \ldots, \alpha_m) > 0 \) with \( \sum_{i=1}^{m} \alpha_i = 1 \), the power cone is defined as

\[
K_\alpha = \{ (x, y) \in \mathbb{R}_+^m \times \mathbb{R} : |y| \leq x_1^{\alpha_1} \cdots x_m^{\alpha_m} \}.
\]

Exercise 1.1 (Cone) Show that \( K_\alpha \) is a cone.

Solution We show that

- \( K_\alpha \) is nonempty, because \( |0| \leq 0^{\alpha_1} \cdots 0^{\alpha_m} = 0 \Rightarrow 0 \in K_\alpha \).

- \( K_\alpha \) is closed w.r.t. addition, because for any given \( (x_1, y_1), (x_2, y_2) \in K_\alpha \), we have

\[
|y_1 + y_2| \leq |y_1| + |y_2| \quad \text{(by triangle inequality)}
\leq (x_1^1)^{\alpha_1} \cdots (x_m^1)^{\alpha_m} + (x_1^2)^{\alpha_1} \cdots (x_m^2)^{\alpha_m}
\leq (x_1^1 + x_1^2)^{\alpha_1} \cdots (x_m^1 + x_m^2)^{\alpha_m}.
\]

- \( K_\alpha \) is closed w.r.t. multiplication, because for any given \( (x, y) \in K_\alpha \), and any \( \lambda \in \mathbb{R}_+ \), we have

\[
|\lambda y| = \lambda |y| \leq \lambda (x_1^{\alpha_1} \cdots x_m^{\alpha_m}) = (\lambda x_1)^{\alpha_1} \cdots (\lambda x_m)^{\alpha_m}
\Rightarrow (\lambda x, \lambda y) \in K_\alpha.
\]

Therefore, \( K_\alpha \) is a convex cone.

Exercise 1.2 (Dual Cone) Follow the steps below to show that the dual cone of power cone is

\[
K_\alpha^* = P_\alpha := \{ (u, v) \in \mathbb{R}_+^m \times \mathbb{R} : |v| \leq \left( \frac{u_1}{\alpha_1} \right)^{\alpha_1} \cdots \left( \frac{u_m}{\alpha_m} \right)^{\alpha_m} \}.
\]

(a) Use Jensen’s inequality to prove the arithmetic-geometric mean inequality:

\[
\sum_{i=1}^{m} \alpha_i y_i \geq \prod_{i=1}^{m} y_i^{\alpha_i}, \quad \forall y \in \mathbb{R}_+^m.
\]

(b) Use the above inequality to show that \( P_\alpha \subseteq K_\alpha^* \).

(c) Show that \( P_\alpha \supseteq K_\alpha^* \).

Hint: consider the particular point \( (\bar{x}, \bar{y}) \) with \( \bar{x}_i = \frac{\alpha_i}{u_i} \) and \( \bar{y} = -\text{sgn}(v) \prod_{i=1}^{m} \left( \frac{\alpha_i}{u_i} \right)^{\alpha_i} \).
Solution

(a) Let \( f(y) = \log y, \forall y > 0 \), which is a concave function. By Jensen’s inequality, \( \forall y_i > 0, i = 1, \ldots, m \), we have

\[
\log \left( \sum_{i=1}^{m} \alpha_i y_i \right) \geq \sum_{i=1}^{m} \alpha_i \log y_i
\]

\[
\Rightarrow e^{\log \left( \sum_{i=1}^{m} \alpha_i y_i \right)} \geq e^{\sum_{i=1}^{m} \alpha_i \log y_i}
\]

\[
\Rightarrow \sum_{i=1}^{m} \alpha_i y_i \geq \prod_{i=1}^{m} y_i^{\alpha_i}.
\]

In addition, when at least one \( y_i = 0 \), the arithmetic-geometric mean inequality holds automatically. Therefore, \( \forall y \in \mathbb{R}^n_+ \), we have

\[
\sum_{i=1}^{m} \alpha_i y_i \geq \prod_{i=1}^{m} y_i^{\alpha_i}.
\]

(b) For any \( (u, v) \in P_\alpha \), we have \( |v| \leq \left( \frac{u_1}{\alpha_1} \right)^{\alpha_1} \cdots \left( \frac{u_m}{\alpha_m} \right)^{\alpha_m} \). For any \( (x, y) \in K_\alpha \), we have \( |y| \leq x_1^{\alpha_1} \cdots x_m^{\alpha_m} \). Therefore,

\[
u^T x + vy = \sum_{i=1}^{m} u_i x_i + vy \geq \sum_{i=1}^{m} u_i x_i - \left( \frac{u_1 x_1}{\alpha_1} \right)^{\alpha_1} \cdots \left( \frac{u_m x_m}{\alpha_m} \right)^{\alpha_m}
\]

\[
\geq \sum_{i=1}^{m} u_i x_i - \sum_{i=1}^{m} u_i x_i (\text{by the arithmetic-geometric mean inequality}) = 0.
\]

So for any given \( (u, v) \in P_\alpha \), we have \( u^T x + vy \geq 0, \forall (x, y) \in K_\alpha \), i.e., \( P_\alpha \subseteq K_\alpha^* \).

(c) For any \( (u, v) \in K_\alpha^* \), we can define the point \((\bar{x}, \bar{y})\) by \( \bar{x} = \frac{\alpha_i}{u_i} \) and \( \bar{y} = -\text{sgn}(v) \prod_{i=1}^{m} \left( \frac{u_i}{\alpha_i} \right)^{\alpha_i} \). Note that

\[
|\bar{y}| = \prod_{i=1}^{m} \left( \frac{\alpha_i}{u_i} \right)^{\alpha_i} = \prod_{i=1}^{m} \bar{x}_i^{\alpha_i} \Rightarrow (\bar{x}, \bar{y}) \in K_\alpha
\]

Since \( (u, v) \in K_\alpha^* \), we have

\[
u^T \bar{x} + v\bar{y} = \sum_{i=1}^{m} u_i \bar{x}_i + v\bar{y} = \sum_{i=1}^{m} \alpha_i + v \left[ -\text{sgn}(v) \prod_{i=1}^{m} \left( \frac{\alpha_i}{u_i} \right)^{\alpha_i} \right] \geq 0
\]

\[
\Rightarrow |v| \leq \left( \frac{u_1}{\alpha_1} \right)^{\alpha_1} \cdots \left( \frac{u_m}{\alpha_m} \right)^{\alpha_m}.
\]

So \( \forall (u, v) \in K_\alpha^* \), we have \( (u, v) \in P_\alpha \), i.e., \( K_\alpha^* \subseteq P_\alpha \).

Problem 2: SOCP Reformulations

Exercise 2.1 (Log-Chebyshev Problem) In HW 1, we have shown that the following optimization problem

\[
\min_x f(x) := \max_{k=1, \ldots, n} |\log(a_k^T x) - \log(b_k)|
\]

s.t. \( 0 \leq x_i \leq 1, i = 1, \ldots, m \)
where $a_k \in \mathbb{R}^m, b_k \in \mathbb{R}, k = 1, \ldots, n$ are given, is equivalent to the following convex optimization problem

$$\min_x \max_{k=1,\ldots,n} h(a_k^T x / b_k)$$

$$\text{s.t. } 0 \leq x_i \leq 1, \ i = 1, \ldots, m$$

where $h(u) = \max(u, 1/u)$ for $u > 0$. Now reformulate the above problem into an SOCP.

Use the fact that any hyperbolic constraints

$$z^2 \leq xy, x \geq 0, y \geq 0$$

can be rewritten as an second order conic constraint

$$\left\| \begin{bmatrix} 2z \\ x - y \end{bmatrix} \right\|_2 \leq x + y.$$

**Solution** We first use epigraph formulation to reformulation the original convex program, and then transfer it into an SOCP.

$$\min_x \max_{k=1,\ldots,n} h(a_k^T x / b_k)$$

$$\text{s.t. } 0 \leq x_i \leq 1, \ i = 1, \ldots, m$$

$$\Leftrightarrow \min t$$

$$\text{s.t. } \frac{1}{t} \leq \frac{a_k^T x}{b_k} \leq t, \ k = 1, \ldots, n$$

$$0 \leq x_i \leq 1, \ i = 1, \ldots, m$$

$$\Leftrightarrow \min t$$

$$\text{s.t. } \left\| \begin{bmatrix} 2z \\ t - \frac{a_k^T x}{b_k} \end{bmatrix} \right\|_2 \leq t + \frac{a_k^T x}{b_k}, \ k = 1, \ldots, n$$

$$\frac{a_k^T x}{b_k} \leq t, \ k = 1, \ldots, n$$

$$0 \leq x_i \leq 1, \ i = 1, \ldots, m.$$  

**Exercise 2.2 (Sparse Group Lasso)** Reformulate the sparse group lasso model as an SOCP:

$$\min_w \left\{ \|Xw - y\|^2_2 + \lambda \sum_{i=1}^{p} \|w_i\|_2 \right\}$$

where $X \in \mathbb{R}^{m \times n}, y \in \mathbb{R}^m$ are the input data, $w = [w_1; \ldots; w_p]$ is the decision variable, with block $w_i \in \mathbb{R}^{n_i}$ and $n_1 + \ldots + n_p = n$. 

3
Solution

\[
\min_w \left\{ \|Xw - y\|_2^2 + \lambda \sum_{i=1}^{p} \|w_i\|_2 \right\}
\]

\[\iff\]

\[
\min_{w, t_0, t_1, \ldots, t_p} t_0 + \lambda \sum_{i=1}^{p} t_i
\]

s.t. \[t_0 \geq \|Xw - y\|_2^2 \] (epigraph formulation)
\[t_i \geq \|w_i\|_2, \forall i = 1, \ldots, p\]

\[\iff\]

\[
\min_{w, t_0, t_1, \ldots, t_p} t_0 + \lambda \sum_{i=1}^{p} t_i
\]

s.t. \[
\begin{bmatrix}
2(Xw - y) \\
t_0 - 1 \\
t_0 + 1
\end{bmatrix} \succeq L_n \geq 0 \] (SOCP)
\[\begin{bmatrix}
w_i \\
t_i
\end{bmatrix} \succeq L_{n+1} \geq 0, \forall i = 1, \ldots, p.
\]

Problem 3: SDP Reformulations

Exercise 3.1 (Spectral Norm Minimization) The spectral norm of a general matrix \(X \in \mathbb{R}^{m \times n}\) is defined as the largest singular value of \(X\), i.e. the square root of the largest eigenvalue of the positive-semidefinite matrix \(X^T X\):

\[
\|X\| := \sigma_{\max}(X) = \lambda_{\max}(X^T X)
\]

Reformulate the spectral norm minimization problem below as an SDP,

\[
\min_{x \in \mathbb{R}^p} \sigma_{\max} \left( \sum_{i=1}^{p} x_i A_i - B \right)
\]

where \(A_1, \ldots, A_p, B \in \mathbb{S}_n\). Assume the domain of the objective \(f\) is \(\text{dom}(f) = \{x \in \mathbb{R}^p : A(x) \succ 0\}\).

Hint: Use Schur complement lemma.

Solution Note that

\[
t \geq \sigma_{\max} \left( \sum_{i=1}^{p} x_i A_i - B \right)
\]

\[\iff\]

\[
t^2 \geq \lambda_{\max} \left[ \left( \sum_{i=1}^{p} x_i A_i - B \right)^T \left( \sum_{i=1}^{p} x_i A_i - B \right) \right]
\]

\[\iff\]

\[
t^2 I - \left( \sum_{i=1}^{p} x_i A_i - B \right)^T \left( \sum_{i=1}^{p} x_i A_i - B \right) \succeq 0
\]

By Schur complement lemma, the original problem can be reformulated as an SDP:

\[
\min_{x \in \mathbb{R}^p, t \in \mathbb{R}} t
\]

s.t. \[
\sum_{i=1}^{p} t_{i_n} \left( \sum_{i=1}^{p} x_i A_i - B \right)^T t_{i_m} \succeq 0
\]

where \(I_n, I_m\) are identity matrices of size \(n \times n\) and \(m \times m\).

Exercise 3.2 (Inverse Matrix Minimization) Reformulate the following minimization problem as an SDP:

\[
\min_{x} f(x) := \max_{1 \leq k \leq K} c_k^T A(x)^{-1} c_k
\]

where \(A(x) = \sum_{i=1}^{p} x_i A_i - B\), and \(A_1, \ldots, A_p, B \in \mathbb{S}^n\). Assume the domain of the objective \(f\) is \(\text{dom}(f) = \{x \in \mathbb{R}^p : A(x) \succ 0\}\).
Solution

\[
\min_x f(x) := \max_{1 \leq k \leq K} c_k^T A(x)^{-1}c_k
\]

\[\iff\]

\[
\min_{t, x} \quad t
\]

\[\text{s.t.} \quad t \geq c_k^T A(x)^{-1}c_k, \quad \forall k = 1, \ldots, K
\]

\[\iff\]

\[
\min_{t, x} \quad t
\]

\[\text{s.t.} \quad \begin{bmatrix} t & c_k^T \\ c_k & A(x) \end{bmatrix} \succeq 0, \quad \forall k = 1, \ldots, K,
\]

(SDP)

where the last equivalence comes from Schur complement lemma, and \(A(x) \succ 0\).

Problem 4: SDP Duality

We denote by \(S_r(A)\) the sum of the largest \(r\) eigenvalues of a symmetric matrix \(A \in S^n (1 \leq r \leq n)\), i.e.,

\[
S_r(A) = \sum_{i=1}^{r} \lambda_i(A)
\]

where \(\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)\) are the eigenvalues of \(A\). In Lecture 15, we have shown that

\[
S_1(A) = \max_X \{ \text{Tr}(AX) : \text{Tr}(X) = 1, \; X \succeq 0 \}.
\]

Now use similar argument to show that

\[
S_k(A) = \max_X \{ \text{Tr}(AX) : \text{Tr}(X) = r, \; 0 \preceq X \preceq I \}.
\]

Solution

First of all, we have

\[
S_k(A) = \sum_{i=1}^{k} \lambda_i(A) = \max_{v_1, \ldots, v_k} \quad \sum_{i=1}^{k} v_i^T A v_i
\]

s.t.

\[
v_i^T v_j = \begin{cases} 1, & i = j \\ 0, & j \neq j \end{cases}
\]

\[= \max_V \quad \text{Tr}(V A V^T)
\]

s.t.

\[
VV^T = I_k, \quad \text{where } V = \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix}.
\]

(1)

Because \(VV^T = I_k \Rightarrow 0 \preceq VV^T \preceq I\) and \(\text{Tr}(VV^T) = k\).

Consider the following SDP which is a convex relaxation of the program above

\[
\max_X \quad \text{Tr}(AX)
\]

s.t.

\[
0 \preceq X \preceq I,
\]

\[
\text{Tr}(X) = k.
\]

(2)
The optimal solution of the convex relaxation (2) provides an upper bound for our original problem (1). By SDP duality, the dual of (2) is

\[
\min_{Z, s} \quad \text{Tr}(Z) + ks \\
\text{s.t.} \quad Z + sI \succeq A, \\
\quad Z \succeq 0.
\]  

(3)

Suppose \( A = U \text{diag}(\lambda) U^T \), where \( U \) is an orthonormal matrix of eigenvectors and \( \lambda \) is the vector of eigenvalues sorted in decreasing order, i.e., \( \lambda_i \geq \lambda_{i+1} \), then let

\[
X^* = U \text{diag}(1_k, 0_{n-k}) U^T \\
Z^* = U \text{diag}(\lambda - \lambda_k 1_k)_+ U^T \\
s^* = \lambda_k,
\]

and it is easy to see that \( X^*, Z^* \) and \( s^* \) are feasible for both (2) and (3).

Moreover, we have

\[
\text{Tr}(AX^*) = \text{Tr} (U \text{diag}(\lambda) U^T U \text{diag}(1_k, 0_{n-k}) U^T) \\
= \text{Tr} (\text{diag}(\lambda) \text{diag}(1_k, 0_{n-k})) \\
= \sum_{i=1}^{k} \lambda_i \\
= S_k(A),
\]

and also

\[
\text{Tr}(Z^*) + ks^* = \sum_{i=1}^{n} (\lambda_i - \lambda_k)_+ + k\lambda_k \\
= \sum_{i=1}^{k} ((\lambda_i - \lambda_k)_+ + \lambda_k) \\
= S_k(A).
\]

By strong duality, we know that

\[
S_k(A) = \max_X \{ \text{Tr}(AX) : \text{Tr}(X) = k, \ 0 \preceq X \preceq I \}.
\]