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

$$K_\alpha = \{(x, y) \in \mathbf{R}^m_+ \times \mathbf{R} : \ |y| \leq x_1^{\alpha_1} \cdots x_m^{\alpha_m}\}.$$

Exercise 1.1 (Cone) Show that K_α is a cone.

Solution We show that

- K_α is nonempty, because
$$|0| \leq 0^{\alpha_1} \cdots 0^{\alpha_m} = 0 \ \Rightarrow \ \mathbf{0} \in K_\alpha.$$
- K_α is closed w.r.t. addition, because for any given $(x^1, y^1), (x^2, y^2) \in K_\alpha$, we have
$$\begin{aligned} |y^1 + y^2| &\leq |y^1| + |y^2| \quad \text{(by triangle inequality)} \\ &\leq (x_1^1)^{\alpha_1} \cdots (x_m^1)^{\alpha_m} + (x_1^2)^{\alpha_1} \cdots (x_m^2)^{\alpha_m} \\ &\leq (x_1^1 + x_1^2)^{\alpha_1} \cdots (x_m^1 + x_m^2)^{\alpha_m}. \end{aligned}$$
- K_α is closed w.r.t. multiplication, because for any given $(x, y) \in K_\alpha$, and any $\lambda \in \mathbf{R}_+$, we have
$$\begin{aligned} & |\lambda y| = \lambda |y| \leq \lambda (x_1^{\alpha_1} \cdots x_m^{\alpha_m}) = \lambda^{\alpha_1 + \cdots + \alpha_m} (x_1^{\alpha_1} \cdots x_m^{\alpha_m}) = (\lambda x_1)^{\alpha_1} \cdots (\lambda x_m)^{\alpha_m} \\ \Rightarrow \ & (\lambda x, \lambda y) \in K_\alpha. \end{aligned}$$

Therefore, K_α is a convex cone.

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

$$K_\alpha^* = P_\alpha := \left\{ (u, v) \in \mathbf{R}^m_+ \times \mathbf{R} : \ |v| \leq \left(\frac{u_1}{\alpha_1}\right)^{\alpha_1} \cdots \left(\frac{u_m}{\alpha_m}\right)^{\alpha_m} \right\}.$$

(a) Use Jensen's inequality to prove the arithmetic-geometric mean inequality:
$$\sum_{i=1}^m \alpha_i y_i \geq \prod_{i=1}^m y_i^{\alpha_i}, \qquad \forall y \in \mathbf{R}^m_+.$$

(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

$$\begin{aligned} &\log\left(\sum_{i=1}^m \alpha_i y_i\right) \geq \sum_{i=1}^m \alpha_i \log y_i \\ \Rightarrow\quad & e^{\log\left(\sum_{i=1}^m \alpha_i y_i\right)} \geq e^{\sum_{i=1}^m \alpha_i \log y_i} \\ \Rightarrow\quad & \sum_{i=1}^m \alpha_i y_i \geq \prod_{i=1}^m y_i^{\alpha_i}. \end{aligned}$$

In addition, when at least one $y_i = 0$, the arithmetic-geometric mean inequality holds automatically. Therefore, $\forall y \in \mathbf{R}_+^n$, 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 $|v| \leq \left(\frac{u_1}{\alpha_1}\right)^{\alpha_1} \ldots \left(\frac{u_m}{\alpha_m}\right)^{\alpha_m}$. For any $(x, y) \in K_\alpha$, we have $|y| \leq x_1^{\alpha_1} \cdots x_m^{\alpha_m}$. Therefore,

$$\begin{aligned} u^T x + vy = \sum_{i=1}^m u_i x_i + vy \quad &\geq \quad \sum_{i=1}^m u_i x_i - \left(\frac{u_1 x_1}{\alpha_1}\right)^{\alpha_1} \ldots \left(\frac{u_m x_m}{\alpha_m}\right)^{\alpha_m} \\ &\geq \quad \sum_{i=1}^m u_i x_i - \sum_{i=1}^m u_i x_i \quad \text{(by the arithmetic-geometric mean inequality)} \\ &= \quad 0. \end{aligned}$$

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

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

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

$$\begin{aligned} & u^T \bar{x} + v\bar{y} = \sum_{i=1}^m u_i \bar{x}_i + v\bar{y} = \sum_{i=1}^m \alpha_i + v\left[-sgn(v) \prod_{i=1}^m \left(\frac{\alpha_i}{u_i}\right)^{\alpha_i}\right] \geq 0 \\ \Rightarrow\quad & |v| \leq \left(\frac{u_1}{\alpha_1}\right)^{\alpha_1} \ldots \left(\frac{u_m}{\alpha_m}\right)^{\alpha_m}. \end{aligned}$$

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

$$\begin{aligned} \min_x \quad & f(x) := \max_{k=1,\ldots,n} |\log(a_k^T x) - \log(b_k)| \\ \text{s.t.} \quad & 0 \leq x_i \leq 1, i = 1, \ldots, m \end{aligned}$$

where $a_k \in \mathbf{R}^m, b_k \in \mathbf{R}, k = 1, \ldots, n$ are given, is equivalent to the following convex optimization problem

$$\begin{aligned} \min_x \quad & \max_{k=1,\ldots,n} h(a_k^T x / b_k) \\ \text{s.t.} \quad & 0 \le x_i \le 1, i = 1, \ldots, m \end{aligned}$$

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

$$z^2 \le xy, x \ge 0, y \ge 0$$

can be rewritten as an second order conic constraint

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

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

$$\begin{aligned} & \min_x \quad \max_{k=1,\ldots,n} h(a_k^T x / b_k) \\ & \text{s.t.} \quad 0 \le x_i \le 1, \; i = 1, \ldots, m \\ \Leftrightarrow \quad & \min \quad t \\ & \text{s.t.} \quad \frac{1}{t} \le \frac{a_k^T x}{b_k} \le t, \; k = 1, \ldots, n \qquad \text{(epigraph formulation)} \\ & \qquad 0 \le x_i \le 1, \; i = 1, \ldots, m \\ \Leftrightarrow \quad & \min \quad t \\ & \text{s.t.} \quad \left\| \begin{bmatrix} 2 \\ t - \frac{a_k^T x}{b_k} \end{bmatrix} \right\|_2 \le t + \frac{a_k^T x}{b_k}, \; k = 1, \ldots, n \\ & \qquad \frac{a_k^T x}{b_k} \le t, \; k = 1, \ldots, n \qquad \text{(SOCP)} \\ & \qquad 0 \le x_i \le 1, \; i = 1, \ldots, m. \end{aligned}$$

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

Solution

$$
\begin{aligned}
& \min_{w} && \left\{ \|Xw - y\|_2^2 + \lambda \sum_{i=1}^{p} \|w_i\|_2 \right\} \\
\Leftrightarrow\ & \min_{w, t_0, t_1, \dots, t_p} && t_0 + \lambda \sum_{i=1}^{p} t_i \\
& \text{s.t.} && t_0 \geq \|Xw - y\|_2^2 && \text{(epigraph formulation)} \\
& && t_i \geq \|w_i\|_2, \ \forall i = 1, \dots, p \\
\Leftrightarrow\ & \min_{w, t_0, t_1, \dots, t_p} && t_0 + \lambda \sum_{i=1}^{p} t_i \\
& \text{s.t.} && \begin{bmatrix} 2(Xw - y) \\ t_0 - 1 \\ t_0 + 1 \end{bmatrix} \geq_{L^{n+2}} 0 && \text{(SOCP)} \\
& && \begin{bmatrix} w_i \\ t_i \end{bmatrix} \geq_{L^{n_i+1}} 0, \quad \forall i = 1, \dots, p.
\end{aligned}
$$

Problem 3: SDP Reformulations

Exercise 3.1 (Spectral Norm Minimization) The spectral norm of a general matrix $X \in \mathbf{R}^{m \times n}$ 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,

$$\min_{x \in \mathbf{R}^p} \ \sigma_{\max}\Big(\sum_{i=1}^{p} x_i A_i - B \Big)$$

where $A_1, \dots, A_p, B \in \mathbf{R}^{m \times n}$.
Hint: Use Schur complement lemma.

Solution Note that

$$
\begin{aligned}
& t \geq \sigma_{\max} \left(\textstyle\sum_{i=1}^{p} x_i A_i - B \right) \\
\Longleftrightarrow\ & t^2 \geq \lambda_{max} \left[\left(\textstyle\sum_{i=1}^{p} x_i A_i - B \right)^T \left(\textstyle\sum_{i=1}^{p} x_i A_i - B \right) \right] \\
\Longleftrightarrow\ & t^2 I - \left(\textstyle\sum_{i=1}^{p} x_i A_i - B \right)^T \left(\textstyle\sum_{i=1}^{p} x_i A_i - B \right) \succeq 0
\end{aligned}
$$

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

$$
\begin{aligned}
& \min_{x \in \mathbf{R}^p, t \in \mathbf{R}} && t \\
& \text{s.t.} && \begin{bmatrix} t I_n & \left(\sum_{i=1}^{p} x_i A_i - B \right)^T \\ \sum_{i=1}^{p} x_i A_i - B & t I_m \end{bmatrix} \succeq 0
\end{aligned}
$$

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:

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

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

Solution

$$
\begin{aligned}
& \min_{x} \quad f(x) := \max_{1\le k\le K} c_k^T \mathcal{A}(x)^{-1} c_k \\
\Leftrightarrow \quad & \min_{t,\ x} \quad t \\
& \text{s.t.} \quad t \ge c_k^T \mathcal{A}(x)^{-1} c_k, \ \forall k = 1, \ldots, K \\
\Leftrightarrow \quad & \min_{t,\ x} \quad t \\
& \text{s.t.} \quad \begin{bmatrix} t & c_k^T \\ c_k & \mathcal{A}(x) \end{bmatrix} \succeq 0, \ \forall k = 1, \ldots, K, \qquad \text{(SDP)}
\end{aligned}
$$

where the last equivalence comes from Schur complement lamma, and $\mathcal{A}(x) \succ 0$.

Problem 4: SDP Duality

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

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

where $\lambda_1(A) \ge \lambda_2(A) \ge \cdots \ge \lambda_n(A)$ 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

$$
\begin{aligned}
S_k(A) = \sum_{i=1}^{k} \lambda_i(A) \quad &= \max_{v_1,\ldots,v_k} \quad \sum_{i=1}^{k} v_i^T A v_i \\
& \qquad \text{s.t.} \qquad v_i^T v_j = \begin{cases} 1, & i = j \\ 0, & j \ne j \end{cases} \\
&= \max_{V} \qquad \text{Tr}(VAV^T) \\
& \qquad \text{s.t.} \qquad VV^T = I_k, \quad \text{where } V = \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix}. \qquad (1)
\end{aligned}
$$

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

$$
\begin{aligned}
\max_{X} \quad & \text{Tr}(AX) \\
\text{s.t.} \quad & 0 \preceq X \preceq I, \qquad (2) \\
& \text{Tr}(X) = k.
\end{aligned}
$$

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

$$\begin{array}{ll} \min_{Z,\ s} & \mathrm{Tr}(Z) + ks \\ \text{s.t.} & Z + sI \succeq A, \\ & Z \succeq 0. \end{array} \tag{3}$$

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

$$\begin{aligned} X^* &= Udiag(\mathbf{1}_k, \mathbf{0}_{n-k})U^T \\ Z^* &= Udiag((\lambda - \lambda_k \mathbf{1})_+)U^T \\ s^* &= \lambda_k, \end{aligned}$$

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

Moreover, we have

$$\begin{aligned} \mathrm{Tr}(AX^*) &= \mathrm{Tr}\left(Udiag(\lambda)U^T Udiag(\mathbf{1}_k, \mathbf{0}_{n-k})U^T\right) \\ &= \mathrm{Tr}\left(diag(\lambda)diag(\mathbf{1}_k, \mathbf{0}_{n-k})\right) \\ &= \sum_{i=1}^{k} \lambda_i \\ &= S_k(A), \end{aligned}$$

and also

$$\begin{aligned} \mathrm{Tr}(Z^*) + ks^* &= \sum_{i=1}^{n} (\lambda_i - \lambda_k)_+ + k\lambda_k \\ &= \sum_{i=1}^{k} ((\lambda_i - \lambda_k)_+ + \lambda_k) \\ &= S_k(A). \end{aligned}$$

By strong duality, we know that

$$S_k(A) = \max_X \left\{ \mathrm{Tr}(AX) : \ \mathrm{Tr}(X) = k, \ 0 \preceq X \preceq I \right\}.$$