SI251 Convex Optimization
Homework 1

Instructor: Professor Ye Shi

Due on 23 Oct 23:59 UTC+8

Note:

- Please provide enough calculation process to get full marks.
- Please submit your homework to Gradescope with entry code: **J7DK2D**.
- Please check carefully whether the question number on the gradescope corresponds to each question.

Exercise 1. Convex sets (40 pts)

1. (20 pts) Please prove that the following sets are convex:

   1) $\mathbf{S} = \{\mathbf{x} \in \mathbb{R}^m \mid \mid p(t) \mid \leq 1 \text{ for } |t| \leq \pi/3\}$, where $p(t) = x_1 \cos t + x_2 \cos 2t + \cdots + x_m \cos mt$. (5 pts)
   2) (Ellipsoids) $\left\{\mathbf{x} | \sqrt{(\mathbf{x}-\mathbf{x}_c)^T \mathbf{P}(\mathbf{x}-\mathbf{x}_c)} \leq r\right\}$ $(\mathbf{x}_c \in \mathbb{R}^n, r \in \mathbb{R}, \mathbf{P} \succeq \mathbf{0})$. (5 pts)
   3) (Symmetric positive semidefinite matrices) $\mathbb{S}_+^{n\times n} = \left\{\mathbf{P} \in \mathbb{S}^{n\times n} | \mathbf{P} \succeq \mathbf{0}\right\}$. (5 pts)
   4) The set of points closer to a given point than a given set, i.e.,

   $$\left\{\mathbf{x} \mid \|\mathbf{x}-\mathbf{x}_0\|_2 \leq \|\mathbf{x}-\mathbf{y}\|_2 \text{ for all } \mathbf{y} \in \mathbf{S}\right\},$$

   where $\mathbf{S} \subseteq \mathbb{R}^n$. (5 pts)

Solution:

(1.1)

The set S can be expressed as the intersection of an infinite number of slabs: $S = \cap_{|t|\leq\frac{\pi}{3}} S_t$, where

$$S_t = \{x | -1 \leq (\cos t, \cdots, \cos mt)^T x \leq 1\}$$

A slab is an intersection of two halfspaces, hence it is a convex set. So S is convex.

(1.2)

Proof: First, note that

$$\|\alpha x\|_P = \sqrt{(\alpha x)^\top P(\alpha x)} = |\alpha| \cdot \sqrt{x^\top P x} = |\alpha| \cdot \|x\|_P \qquad (1)$$

Next, we observe that $\|x\|_P \geq 0$ and $\|x\|_P = 0$ iff $x = 0$ by the definition of $P \succ 0$. The third component of proving $\|\cdot\|_P$ is a norm is to show the triangle inequality holds. By the definition of the Mahalanobis norm, we have

$$\|x+y\|_P^2 = (x+y)^\top P(x+y) = x^\top P x + y^\top P y + 2x^\top P y. \qquad (2)$$

Since $P \succ 0, P$ has the eigendecompostion $P = U\Lambda U^\top$, where U is an orthogonal matrix, Λ is a diagonal matrix with all diagonal entries being positive. Hence, $\Lambda^{1/2}$ is well defined, so is $P^{1/2}$ (defined as $U\Lambda^{1/2}U^\top$). From (2) and the definition of $\|\cdot\|_P$, it then follows that

$$\begin{aligned} \|x+y\|_P^2 &= \|x\|_P^2 + \|y\|_P^2 + 2x^\top P^{\frac{1}{2}} P^{\frac{1}{2}} y \\ &\leq \|x\|_P^2 + \|y\|_P^2 + 2\left\|P^{\frac{1}{2}}x\right\|_2 \cdot \left\|P^{\frac{1}{2}}y\right\|_2 \\ &= \|x\|_P^2 + \|y\|_P^2 + 2\|x\|_P \cdot \|y\|_P \end{aligned} \qquad (3)$$

where the inequality follows from the Cauchy-Schwarz inequality, and the last equality holds since $\left\|P^{1/2}x\right\|_2 = \sqrt{x^\top P x} = \|x\|_P$. Note that (3) can be rewritten as

$$\|x+y\|_P^2 \leq (\|x\|_P + \|y\|_P)^2$$

which is equivalent to the triangle inequality. Therefore, $\|\cdot\|_P$ is a norm.

Given that the Mahalanobis norm is indeed a norm, we can now show that an ellipsoid centered at x is a convex set.

Proof: Since $(y-x)^\top P(y-x) = \|y-x\|_P^2$, we can redefine ellipsoid as

$$\mathcal{E}(\mathrm{x}) = \left\{ y \in \mathbb{R}^d : \|y-x\|_P^2 \le r, P \succ 0, x \in \mathbb{R}^d \right\},$$

or, equivalently,

$$\mathcal{E}(\mathrm{x}) = \left\{ y \in \mathbb{R}^d : \|y-x\|_P \le r, P \succ 0, \mathrm{x} \in \mathbb{R}^d \right\}.$$

To show $\mathcal{E}(x)$ is convex, we need to show that for any $y_1, y_2 \in \mathcal{E}(x)$ and any $\alpha \in [0,1]$, $\alpha y_1 + (1-\alpha)y_2 \in \mathcal{E}(x)$, i.e. $\|\alpha y_1 + (1-\alpha)y_2 - x\|_P \le r$ holds. This is equivalent to showing

$$\|\alpha y_1 - \alpha x + (1-\alpha)y_2 - (1-\alpha)x\|_P \le r. \qquad (4)$$

Applying the triangle inequality gives

$$\begin{aligned}\|\alpha y_1 - \alpha x + (1-\alpha)y_2 - (1-\alpha)x\|_P &\le \|\alpha y_1 - \alpha x\|_P + \|(1-\alpha)y_2 - (1-\alpha)x\|_P \\ &= \alpha \cdot \|y_1 - x\|_P + (1-\alpha) \cdot \|y_2 - x\|_P \\ &\le r,\end{aligned}$$

where the last inequality follows from the assumption that $y_1, y_2 \in \mathcal{E}(x)$. Hence, inequality (4) holds and $\mathcal{E}(\mathrm{x})$ is convex.

(1.3)

Let $A \succeq 0, B \succeq 0$ and $\lambda \in [0,1]$. For any $y \in \mathbb{R}^n$, we have

$$y^T(\lambda A + (1-\lambda)B)y = \lambda y^T A y + (1-\lambda) y^T B y \ge 0$$

So $S_+^{n\times n}$ is convex.

(1.4)

This set is convex because it can be expressed as

$$\bigcap_{y \in S} \{x \mid \|x - x_0\|_2 \le \|x - y\|_2\},$$

i.e., an intersection of halfspaces. (For fixed y, the set $\{x \mid \|x - x_0\|_2 \le \|x - y\|_2\}$ is a halfspace).

2. (10 pts) Example of convex set

- Show that if $a, b \geq 0$ and $0 \leq \theta \leq 1$, then $a^\theta b^{1-\theta} \leq \theta a + (1-\theta)b$. Hint: Use concavity of log functions.
- Show the $\mathbf{S}_n = \{x \in \mathbb{R}^n_+ | \prod_{i=1}^n x_i \geq 1\}$ is convex.

Solution: This inequality still holds when $a = 0$ or $b = 0$. So we consider the case when $a > 0$ and $b > 0$. According to what we have learned in class, the logarithmic function $\log(x)$ is a concave function. According to the definition of a concave function, for any $x, y > 0$ and $0 \leq \lambda \leq 1$, we have:

$$\log(\lambda x + (1-\lambda)y) \geq \lambda \log(x) + (1-\lambda)\log(y)$$

Let $x = a, y = b$, so we have

$$\log(a^\theta b^{1-\theta}) = \log(a^\theta) + \log(b^{1-\theta}) = \theta \log(a) + (1-\theta)\log(b)$$

According to the concavity of the logarithmic function, we have:

$$\log(\theta a + (1-\theta)b) \geq \theta \log(a) + (1-\theta)\log(b)$$

$$e^{\log(\theta a + (1-\theta)b)} \geq e^{\theta \log(a) + (1-\theta)\log(b)}$$

$$\theta a + (1-\theta)b \geq a^\theta b^{1-\theta}$$

Assume that $\prod_i x_i \geq 1$ and $\prod_i y_i \geq 1$. Using the inequality we proved above, we have

$$\prod_i (\theta x_i + (1-\theta)y_i) \geq \prod x_i^\theta y_i^{1-\theta} = (\prod_i x_i)^\theta (\prod_i y_i)^{1-\theta} \geq 1, \theta \in [0,1]$$

Thus we prove that S_n is a convex set.

3. (10 pts) Consider a convex set C defined as $C = \{x \in \mathbb{R}^n : ||x||_2 \leq \sqrt{n}r\}$. Proof that $\sum_{i=1}^{n} x_i \leq nr$ for all $x \in C$ using the Supporting Hyperplane Theorem (**Otherwise you will get almost zero point**).

Solution:

First, observe that the point $x_0 = [r, r, r, ..., r]^T$ is a boundary point of C. This is because the Euclidean norm of x_0 is:

$$||x_0||_2 = \sqrt{r^2 + r^2 + r^2 + ... + r^2} = \sqrt{n}r. \tag{1}$$

Consider the vector $a = [1, 1, 1, ..., 1]^T \in R^n$ which is parallel to $\nabla ||x||_2$ in the point x_0, The hyperplane defined by the equation$\{x | a^T x = a^T x_0\}$ is the supporting hyperplane to C where $a^T x_0$ can be computed as follows:

$$a^T x_0 = [1, 1, 1, ..., 1][r, r, r, ..., r]^T = nr. \tag{2}$$

Therefore, for all $x \in C$, we have

$$\begin{aligned} a^T x &\leq a^T x_0 \\ [1, 1, 1, ..., 1]x &\leq [1, 1, 1, ..., 1][r, r, r, ..., r]^T \\ \sum_{i=1}^{n} x_i &\leq nr. \end{aligned} \tag{3}$$

which completes the proof that $\sum_{i=1}^{n} x_i \leq nr$ for all $x \in C$.

Exercise 2. Convex functions (30 pts)

1. (5 pts) Monotone Mappings. A function $\psi : \mathbb{R}^n \to \mathbb{R}^n$ is called monotone if, for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom}\, \psi$,

$$(\psi(\mathbf{x}) - \psi(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}) \geq 0.$$

Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is a differentiable convex function. Prove that its gradient ∇f is monotone.

Solution:

Due to that f is a differentiable convex function, we have

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \\ f(\mathbf{x}) &\geq f(\mathbf{y}) + \nabla f(\mathbf{y})^T(\mathbf{x} - \mathbf{y}). \end{aligned}$$

Add them together, we have

$$\begin{aligned} f(\mathbf{y}) + f(\mathbf{x}) &\geq f(\mathbf{x}) + f(\mathbf{y}) + (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}))^T(\mathbf{x} - \mathbf{y}) \\ (\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}) &\geq 0. \end{aligned}$$

So gradient ∇f is monotone.

2. (10 pts) For the following function, find the range of values of β that makes the function convex:

$$f(x, y, z) = 5x^2 + 5y^2 + 4z^2 - 6xz + 2\beta xy - 4yz$$

Solution: First we need to transform the function $f(x, y, z)$ into quadratic form. For real symmetric matrices, semi-positive definiteness can also be determined by having all of their leading principal minors be non-negative. So to guarantee the convexity of $f(x, y, z)$, all the principal minors should be non-negative.

$f(x, y, z)$ can be written as:

$$f(x, y, z) = [x \quad y \quad z] \begin{bmatrix} 5 & \beta & -3 \\ \beta & 5 & -2 \\ -3 & -2 & 4 \end{bmatrix} [x \quad y \quad z]^T = [x \quad y \quad z]\mathbf{Q}[x \quad y \quad z]^T$$

First-order principal minors: They are the diagonal elements of the matrix $\mathbf{Q}$, which are all non-negative.

Second-order principal minors: including

1)

$$\begin{vmatrix} 5 & \beta \\ \beta & 5 \end{vmatrix} = 25 - \beta^2 \geq 0 \Longrightarrow \beta \in [-5, 5]$$

2)

$$\begin{vmatrix} 5 & -3 \\ -3 & 4 \end{vmatrix} = 20 - 9 = 11 > 0$$

3)

$$\begin{vmatrix} 5 & -2 \\ -2 & 4 \end{vmatrix} = 20 - 4 = 16 > 0$$

Third-order principal minors:

$$\begin{vmatrix} 5 & \beta & -3 \\ \beta & 5 & -2 \\ -3 & -2 & 4 \end{vmatrix} = -4\beta^2 + 12\beta + 35 \geq 0 \Longrightarrow \beta \in [\frac{3}{2} - \sqrt{11}, \frac{3}{2} + \sqrt{11}]$$

Combining the above conditions on β, the range of values of β that guarantees convexity of f is:

$$f \in [\frac{3}{2} - \sqrt{11}, \frac{3}{2} + \sqrt{11}]$$

3. (15 pts) Let $f = (f_1, \cdots, f_m)$, where each $f_i : \mathbf{C} \to \mathbb{R}$ for $i = 1, \cdots, m$ is a convex function and $\mathbf{C} = dom(f)$. Also, consider a function $g : \mathbb{R}^m \to \mathbb{R}$ that is both convex and monotonically nondecreasing over the set $\{f(x) \,|\, x \in \mathbf{C}\}$. This means for any l_1, l_2 in this set with $l_1 \leq l_2$, it holds that $g(l_1) \leq g(l_2)$. Proof that the function h, specified by $h(x) = g(f(x))$ is convex using the definition of convex functions $(f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y))$.

Solution:

Let $x, y \in \mathbb{R}^n$ and let $\alpha \in [0, 1]$. By the definitions of h and f, we have

$$
\begin{aligned}
h(\alpha x + (1-\alpha)y) &= g(f(\alpha x + (1-\alpha)y)) \\
&= g(f_1(\alpha x + (1-\alpha)y), \ldots, f_m(\alpha x + (1-\alpha)y)) \\
&\leq g(\alpha f_1(x) + (1-\alpha) f_1(y), \ldots, \alpha f_m(x) + (1-\alpha) f_m(y)) \\
&= g(\alpha f_1(x), \ldots, \alpha f_m(x) + (1-\alpha) f_1(y), \ldots, (1-\alpha) f_m(y)) \\
&\leq \alpha g(f_1(x), \ldots, f_m(x)) + (1-\alpha) g(f_1(y), \ldots, f_m(y)) \\
&= \alpha g(f(x)) + (1-\alpha) g(f(y)) \\
&= \alpha h(x) + (1-\alpha) h(y).
\end{aligned} \tag{4}
$$

where the first inequality follows by convexity of each f_i and monotonicity of g, while the second inequality follows by convexity of g.

Exercise 3. Convex Optimization Problems (30 pts)

1. (30 pts) Consider the following quadratic programming

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \preceq \mathbf{b}. \end{aligned}$$

Here, we only introduce error to the matrix $\mathbf{P}$ while maintaining other parameters known. This leads to robust variation of quadratic programming. The robust quadratic programming is defined as

$$\begin{aligned} \min \quad & \sup_{\mathbf{P}\in\mathcal{E}} \left(\frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x}\right) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \end{aligned}$$

where $\mathcal{E}$ is the set of possible matrices $\mathbf{P}$.

For each of the following sets $\mathcal{E}$, express the robust QP as a convex problem in a standard form (e.g., QP, QCQP, SOCP, SDP).

(a) A finite set of matrices: $\mathcal{E} = \{\mathbf{P}_1, \dots, \mathbf{P}_K\}$, where $\mathbf{P}_i \in \mathbb{S}^n_+$, $i = 1, \dots, K$. (5 pts)

(b) A set specified by a nominal value $\mathbf{P}_0 \in \mathbb{S}^n_+$ plus a bound on the eigenvalues of the deviation $\mathbf{P} - \mathbf{P}_0$:

$$\mathcal{E} = \{\mathbf{P} \in \mathbb{S}^n | -\gamma\mathbf{I} \preceq \mathbf{P} - \mathbf{P}_0 \preceq \gamma\mathbf{I}\},$$

where $\gamma \in \mathbb{R}$ and $\mathbf{P}_0 \in \mathbb{S}^n_+$. (10 pts)

(c) An ellipsoid of matrices:

$$\mathcal{E} = \left\{\mathbf{P}_0 + \sum_{i=1}^{K}\mathbf{P}_i u_i \mid ||\mathbf{u}||_2 \le 1\right\}.$$

You can assume $\mathbf{P}_i \in \mathbb{S}^n_+$, $i = 0, \dots, K$. (15 pts)

Solution:

(a) We can set the upper bound of $\frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x}$ as t. Thus we have

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{P}_i\mathbf{x} + \mathbf{q}^T\mathbf{x} \le t, \qquad i = 1, \dots, K \\ & \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \end{aligned}$$

which is a QCQP in the variables $\mathbf{x}$ and t.

(b) For given $\mathbf{x}$, let $\Delta\mathbf{P} = \mathbf{P} - \mathbf{P}_0$, the supremum of $\mathbf{x}^T\Delta\mathbf{P}\mathbf{x}$ over $-\gamma\mathbf{I} \le \Delta\mathbf{P} \le \gamma\mathbf{I}$ is given by

$$\sup_{-\gamma\mathbf{I}\le\Delta\mathbf{P}\le\gamma\mathbf{I}} \mathbf{x}^T\Delta\mathbf{P}\mathbf{x} = \gamma\mathbf{x}^T\mathbf{x}.$$

Therefore we can express the robust QP as

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^T(\mathbf{P}_0 + \gamma\mathbf{I})\mathbf{x} + \mathbf{q}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \end{aligned}$$

which is a QP.

(c) Step 1: For given $\mathbf{x}$, we express the optimization variables as $\mathbf{P} = \mathbf{P}_0 + \sum_{i=1}^{K} \mathbf{P}_i u_i$. We have the objective function as

$$\frac{1}{2}\left(\mathbf{x}^T\mathbf{P}_0\mathbf{x} + \sup_{||\mathbf{u}||_2 \leq 1} \sum_{i=1}^{K} u_i(\mathbf{x}^T\mathbf{P}_i\mathbf{x})\right) + \mathbf{q}^T\mathbf{x}.$$

Due to that

$$\sup_{||\mathbf{u}||_2 \leq 1} \sum_{i=1}^{K} u_i\left(\mathbf{x}^T\mathbf{P}_i\mathbf{x}\right) = \left(\sum_{i=1}^{K}\left(\mathbf{x}^T\mathbf{P}_i\mathbf{x}\right)^2\right)^{\frac{1}{2}}.$$

So we can express the objective function as

$$\frac{1}{2}\left(\mathbf{x}^T\mathbf{P}_0\mathbf{x} + \left(\sum_{i=1}^{K}\left(\mathbf{x}^T\mathbf{P}_i\mathbf{x}\right)^2\right)^{\frac{1}{2}}\right) + \mathbf{q}^T\mathbf{x}.$$

Step 2: We further define the upper bound of $\frac{1}{2}\mathbf{x}^T\mathbf{P}_i\mathbf{x}$ as y_i, then we have

$$\begin{array}{ll} \min & \frac{1}{2}\mathbf{x}^T\mathbf{P}_0\mathbf{x} + ||\mathbf{y}||_2 + \mathbf{q}^T\mathbf{x} \\ \text{s.t.} & \frac{1}{2}\mathbf{x}^T\mathbf{P}_i\mathbf{x} \leq y_i, \quad i = 1, \ldots, K \\ & \mathbf{A}\mathbf{x} \preceq \mathbf{b}. \end{array}$$

Step 3: We further reduce the problem to an SOCP problem. Here, first provide the upper bound of $||\mathbf{y}||_2$

$$\begin{array}{ll} \min & \frac{1}{2}\mathbf{x}^T\mathbf{P}_0\mathbf{x} + \mathbf{q}^T\mathbf{x} + t \\ \text{s.t.} & \frac{1}{2}\mathbf{x}^T\mathbf{P}_i\mathbf{x} \leq y_i, \quad i = 1, \ldots, K \\ & ||\mathbf{y}||_2 \leq t \\ & \mathbf{A}\mathbf{x} \preceq \mathbf{b}. \end{array}$$

Then provide the upper bound of $\frac{1}{2}\mathbf{x}^T\mathbf{P}_0\mathbf{x} + \mathbf{q}^T\mathbf{x}$

$$\begin{array}{ll} \min & u + t \\ \text{s.t.} & \frac{1}{2}\mathbf{x}^T\mathbf{P}_i\mathbf{x} \leq y_i, \quad i = 1, \ldots, K \\ & \frac{1}{2}\mathbf{x}^T\mathbf{P}_0\mathbf{x} + \mathbf{q}^T\mathbf{x} \leq u \\ & ||\mathbf{y}||_2 \leq t \\ & \mathbf{A}\mathbf{x} \preceq \mathbf{b}. \end{array}$$

Finally, reduce the quadratic constraint to an SOC constraint.

$$\begin{array}{ll} \min & u + t \\ \text{s.t.} & \left\|\begin{bmatrix} \mathbf{P}_i^{\frac{1}{2}}\mathbf{x} \\ \frac{1}{2} - y_i \end{bmatrix}\right\|_2 \leq \frac{1}{2} + y_i, \quad i = 1, \ldots, K \\ & \left\|\begin{bmatrix} \mathbf{P}_0^{\frac{1}{2}}\mathbf{x} \\ \frac{1}{2} + \mathbf{q}^T\mathbf{x} - u \end{bmatrix}\right\|_2 \leq \frac{1}{2} - \mathbf{q}^T\mathbf{x} + u \\ & ||\mathbf{y}||_2 \leq t \\ & \mathbf{A}\mathbf{x} \preceq \mathbf{b}. \end{array}$$