SI251 - Convex Optimization homework 1

Deadline: 2024-03-27 23:59:59

1. You can use Word, Latex or handwriting to complete this assignment. If you want to submit a handwritten version, scan it clearly.
2. The **report** has to be submitted as a PDF file to Gradescope, other formats are not accepted.
3. The submitted file name is **student_id+your_student_name.pdf**.
4. Late policy: You have 4 free late days for the quarter and may use up to 2 late days per assignment with no penalty. Once you have exhausted your free late days, we will deduct a late penalty of 25% per additional late day. Note: The timeout period is recorded in days, even if you delay for 1 minute, it will still be counted as a 1 late day.
5. You are required to follow ShanghaiTech's academic honesty policies. You are not allowed to copy materials from other students or from online or published resources. Violating academic honesty can result in serious sanctions.

Any plagiarism will get Zero point.

1 Convex sets

1. Please prove that the following sets are convex:

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

   $$\left\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\right\},$$

   where $S \in R^n$. (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

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

or, equivalently,

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

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. $\|\alpha y_1 + (1-\alpha)y_2 - x\|_P \leq r$ holds. This is equivalent to showing

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

Applying the triangle inequality gives

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

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

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

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

(1.4)

This set is convex because it can be expressed as

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

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

2. (15 pts) For a given norm $\|\cdot\|$ on $\mathbf{R}^n$, the dual norm, denoted $\|\cdot\|_*$, is defined as

$$\|y\|_* = \sup_{x\in\mathbf{R}^n} \{y^T x \mid \|x\| \leq 1\}.$$

Show that the dual of Euclidean norm is the Euclidean norm, i.e., $\sup_{x\in\mathbf{R}^n}\{z^T x \mid \|x\|_2 \leq 1\} = ||z||_2$.

Solution: According to the definition of dual norm, we have the following optimization problem about the dual of Euclidean norm:

$$\begin{aligned}
\max \quad & z^T x \\
\text{s.t.} \quad & ||x||_2 \leq 1
\end{aligned}$$

Since $z^T x \leq ||z||_2 ||x||_2$ (Cauchy-Schwarz inequality) and $||x||_2 \leq 1$, we have $\max z^T x = ||z||_2 \cdot 1 = ||z||_2$

3. (15 pts) Define a norm cone as

$$\mathcal{C} \equiv \left\{(x, t) : x \in \mathbb{R}^d, t \geq 0, \|x\| \leq t\right\} \subseteq \mathbb{R}^{d+1}$$

Show that the norm cone is convex by using the definition of convex sets.

Solution: To show $\mathcal{C}$ is convex, we need to show that for any $(x_1, t_1), (x_2, t_2) \in \mathcal{C}$ and any $\lambda \in [0, 1], \lambda(x_1, t_1) + (1-\lambda)(x_2, t_2) \in \mathcal{C}$. Since $\|x_1\| \leq t_1, \|x_2\| \leq t_2, 0 \leq \lambda \leq 1$, we have $\|\lambda x_1 + (1-\lambda)x_2\| \leq \|\lambda x_1\| + \|(1-\lambda)x_2\| \leq \lambda t_1 + (1-\lambda)t_2$. Hence $\lambda(x_1, t_1) + (1-\lambda)(x_2, t_2) \in \mathcal{C}$ and $\mathcal{C}$ is convex.

2 Convex functions

4. (18 pts) Let $C \subset \mathbb{R}^n$ be convex and $f : C \to R^\star$. Show that the following statements are equivalent:

   (a) $\mathrm{epi}(f)$ is convex.

   (b) For all points $x_i \in C$ and $\{\lambda_i | \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1, i = 1, 2, \cdots, n\}$, we have

$$f\Big(\sum_{i=1}^n \lambda_i x_i\Big) \leq \sum_{i=1}^n \lambda_i f(x_i).$$

   (c) For $\forall x, y \in C$ and $\lambda \in [0, 1]$,

$$f\Big((1-\lambda)x + \lambda y\Big) \leq (1-\lambda)f(x) + \lambda f(y).$$

Solution:

To see that (a) implies (b) we note that, for all $i = 1, 2, \cdots, n$, $(f(x_i), x_i) \in \mathrm{epi}(f)$. Since this latter set is convex, we have that

$$\sum_{i=1}^n \lambda_i \left(f\left(x_i\right), x_i\right) = \left(\sum_{i=1}^n \lambda_i f\left(x_i\right), \sum_{i=1}^n \lambda_i x_i\right) \in \mathrm{epi}(f),$$

which, in turn, implies that

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f\left(x_i\right).$$

This establishes (b). It is obvious that (b) implies (c). So it remains only to show that (c) implies (a) in order to establish the equivalence. To this end, suppose that $\left(z_1, x_1\right), \left(z_2, x_2\right) \in \mathrm{epi}(f)$ and take $0 \leq \lambda \leq 1$. Then

$$(1-\lambda)\left(z_1, x_1\right) + \lambda\left(z_2, x_2\right) = \left((1-\lambda)z_1 + \lambda z_2, (1-\lambda)x_1 + \lambda x_2\right),$$

and since $f\left(x_1\right) \leq z_1$, $f\left(x_2\right) \leq z_2$, $(1-\lambda) > 0$, and $\lambda > 0$, so that

$$(1-\lambda)f\left(x_1\right) + \lambda f\left(x_2\right) \leq (1-\lambda)z_1 + \lambda z_2.$$

Hence, by the assumption (c), $f\left((1-\lambda)x_1 + \lambda x_2\right) \leq (1-\lambda)z_1 + \lambda z_2$, which shows this latter point is in $\mathrm{epi}(f)$.

5. (14 pts) Monotone Mappings. A function $\psi : \mathbf{R}^n \to \mathbf{R}$ is called monotone if for all $x, y \in \mathbf{dom}\psi$,

$$(\psi(x) - \psi(y))^T(x - y) >= 0. \tag{1}$$

Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is a differentiable convex function. Show that its gradient ∇f is monotone. Is the converse true, i.e., is every monotone mapping the gradient of a convex function?

Solution:

f is a differentiable convex function, which implies:

$$f(x) \geq f(y) + \nabla f(y)^T(x-y), \quad f(y) \geq f(x) + \nabla f(x)^T(y-x)$$

for arbitrary $x, y \in \mathbf{dom}f$. Combining the two inequalities gives

$$f(x) + f(y) \geq f(y) + \nabla f(y)^T(x-y) + f(x) + \nabla f(x)^T(y-x)$$

$$0 \geq \nabla f(y)^T(x-y) + \nabla f(x)^T(y-x)$$

$$\nabla f(x)^T(x-y) \geq \nabla f(y)^T(x-y)$$

$$(\nabla f(x) - \nabla f(y))^T(x-y) \geq 0$$

which shows that the gradient ∇f is monotone.

The converse not true in general. As a counterexample, consider

$$\phi(x) = \begin{bmatrix} x_1 \\ x_1/3 + x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

ϕ is monotone because

$$(x-y)^T \begin{bmatrix} 1 & 0 \\ 1/3 & 1 \end{bmatrix} (x-y) = (x-y)^T \begin{bmatrix} 1 & 1/6 \\ 1/6 & 1 \end{bmatrix} (x-y) \geq 0$$

for all x, y. However, there does not exist a function $f : \mathbb{R}^2 \leftarrow \mathbb{R}$ such that $\psi(x) = \nabla f(x)$, because such a function would have to satisfy

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial \psi_1}{\partial \psi_2} = 0, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial \psi_2}{\partial \psi_1} = \frac{1}{3}$$

6. (18 pts) Please determine whether the following functions are convex, concave or none of those, and give a detailed explanation for your choice.

   1)
   $$f_1(x_1, x_2, \cdots, x_n) = \begin{cases} -(x_1 x_2 \cdots x_n)^{\frac{1}{n}}, & \text{if } x_1, \cdots, x_n > 0 \\ \infty & \text{otherwise;} \end{cases}$$

   2) $f_2(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$, where $0 \leq \alpha \leq 1$, on $\mathbb{R}^2_{++}$;

   3) $f_3(x, u, v) = -\log(uv - x^T x)$ on $\mathbf{dom}f = \{(x, u, v) | uv > x^T x, \ \ u, v > 0\}$.

Solution:

(1) Denote $X = \mathrm{dom}\,(f_1)$. It can be seen that f_1 is twice continuously differentiable over X and its Hessian matrix is given by

$$\nabla^2 f_1(x) = \frac{f_1(x)}{n^2} \begin{bmatrix} \frac{1-n}{x_1^2} & \frac{1}{x_1 x_2} & \cdots & \frac{1}{x_1 x_n} \\ \frac{1}{x_2 x_1} & \frac{1-n}{x_2^2} & \cdots & \frac{1}{x_2 x_n} \\ & & \vdots & \\ \frac{1}{x_n x_1} & \frac{1}{x_1 x_2} & \cdots & \frac{1-n}{x_n^2} \end{bmatrix}$$

for all $x = (x_1, \ldots, x_n) \in X$. From this, direct computation shows that for all $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$ and $x = (x_1, \ldots, x_n) \in X$, we have

$$z' \nabla^2 f_1(x) z = \frac{f_1(x)}{n^2} \left(\left(\sum_{i=1}^{n} \frac{z_i}{x_i} \right)^2 - n \sum_{i=1}^{n} \left(\frac{z_i}{x_i} \right)^2 \right)$$

Note that this quadratic form is nonnegative for all $z \in \mathbb{R}^n$ and $x \in X$, since $f_1(x) < 0$, and for any real numbers $\alpha_1, \ldots, \alpha_n$, we have

$$(\alpha_1 + \cdots + \alpha_n)^2 \leq n \left(\alpha_1^2 + \cdots + \alpha_n^2 \right)$$

in view of the fact that $2\alpha_j \alpha_k \leq \alpha_j^2 + \alpha_k^2$. Hence, $\nabla^2 f_1(x)$ is positive semidefinite for all $x \in X$, and it follows from the second-order conditions that f_1 is convex.

(2) The Hessian is

$$\begin{aligned} \nabla^2 f_2(x) &= \begin{bmatrix} \alpha(\alpha-1) x_1^{\alpha-2} x_2^{1-\alpha} & \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} \\ \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} & (1-\alpha)(-\alpha) x_1^{\alpha} x_2^{-\alpha-1} \end{bmatrix} \\ &= \alpha(1-\alpha) x_1^{\alpha} x_2^{1-\alpha} \begin{bmatrix} -1/x_1^2 & 1/x_1 x_2 \\ 1/x_1 x_2 & -1/x_2^2 \end{bmatrix} \\ &= -\alpha(1-\alpha) x_1^{\alpha} x_2^{1-\alpha} \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix} \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix}^T \\ &\preceq 0. \end{aligned}$$

Hence, f_2 is concave.

(3) We can express f_3 as

$$f_3(x, u, v) = -\log u - \log \left(v - x^T x / u \right).$$

The first term is convex. The function $v - x^T x/u$ is concave because v is linear and $x^T x/u$ is convex on $\{(x, u) \mid u > 0\}$. Therefore the second term in f_3 is convex: it is the composition of a convex decreasing function $-\log t$ and a concave function. Therefore f_3 is convex.