SI151A
Convex Optimization and its Applications in Information Science, Fall 2024
Homework 1

Due on Nov. 4, 2024, before class

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points ($\leq 20\%$) of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Do your homework by yourself. Any form of plagiarism will lead to 0 point of this homework. If more than one plagiarisms during the semester are identified, we will prosecute all violations to the fullest extent of the university regulations, including but not limited to failing this course, academic probation, or expulsion from the university.
- If you have any doubts regarding the grading, you need to contact the instructor or the TAs within two days since the grade is announced.

1. Which of the following sets are convex?

1. A *slab*, i.e., a set of the form $\{x \in \mathbb{R}^n \mid \alpha \leq a^T x \leq \beta\}$. (4 points)
2. A *rectangle*, i.e., a set of the form $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, \dots, n\}$. A rectangle is sometimes called a *hyperrectangle* when $n > 2$. (4 points)
3. A *wedge*, i.e., $\{x \in \mathbb{R}^n \mid a_1^T x \leq b_1, a_2^T x \leq b_2\}$. (4 points)
4. The set of points closer to a given point than a given set, i.e.,
$$\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\}$$
where $S \subseteq \mathbb{R}^n$. (4 points)
5. The set of points closer to one set than another, i.e.,
$$\{x \mid \mathrm{dist}(x, S) \leq \mathrm{dist}(x, T)\},$$
where $S, T \subseteq \mathbb{R}^n$, and
$$\mathrm{dist}(x, S) = \inf\{\|x - z\|_2 \mid z \in S\}.$$
(4 points)

Solution:

1. A slab is an intersection of two halfspaces, hence it is a convex set (and a polyhedron).
2. As in part 1, a rectangle is a convex set and a polyhedron because it is a finite intersection of halfspaces.
3. A wedge is an intersection of two halfspaces, so it is convex set. It is also a polyhedron. It is a cone if $b_1 = 0$ and $b_2 = 0$.
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).
5. In general this set is not convex, as the following example in $\mathbb{R}$ shows. With $S = \{-1, 1\}$ and $T = \{0\}$, we have
$$\{x \mid \mathrm{dist}(x, S) \leq \mathrm{dist}(x, T)\} = \{x \in \mathbb{R} \mid x \leq -1/2 \text{ or } x \geq 1/2\}$$
which clearly is not convex.

2. Convex functions.

For each of the following functions determine whether it is convex, concave, quasiconvex, or quasiconcave.

1. $f(x) = e^x - 1$ on $\mathbb{R}$. (4 points)
2. $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$. (4 points)
3. $f(x_1, x_2) = \frac{1}{x_1 x_2}$ on $\mathbb{R}^2_{++}$. (4 points)

Show that the following function $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

4. $f(x) = \|Ax - b\|$, where $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, and $\|\cdot\|$ is a norm on $\mathbb{R}^n$. (4 points)
5. $f(x) = -\log\left(-\log\left(\sum_{i=1}^n e^{a_i^T x + b_i}\right)\right)$ on $\mathbf{dom}\ f = \{x | \sum_{i=1}^n e^{a_i^T x + b_i} < 1\}$. *(hint: You can use the fact that* $\log\left(\sum_{i=1}^n e^{y_i}\right)$ *is convex.)* (4 points)

Solution:

1. $f(x) = e^x - 1$ on $\mathbb{R}$.
   Solution. Strictly convex, and therefore quasiconvex. Also quasiconcave but not concave.
2. $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$.
   Solution. The Hessian of f is
   $$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$
   which is neither positive semidefinite nor negative semidefinite. Therefore, f is neither convex nor concave. It is quasiconcave, since its superlevel sets
   $$\{(x_1, x_2) \in \mathbb{R}^2_{++} \mid x_1 x_2 \geq \alpha\}$$
   are convex. It is not quasiconvex.
3. $f(x_1, x_2) = \frac{1}{x_1 x_2}$ on $\mathbb{R}^2_{++}$.
   Solution. The Hessian of f is
   $$\nabla^2 f(x) = \frac{1}{x_1^2 x_2^2} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \succeq 0.$$
   Therefore, f is convex and quasiconvex. It is not quasiconcave or concave.
4. $f(x) = \|Ax - b\|$, where $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, and $\|\cdot\|$ is a norm on $\mathbb{R}^n$.
   Solution. f is the composition of a norm, which is convex, and an affine function.
5. $f(x) = -\log\left(-\log\left(\sum_{i=1}^n e^{a_i^T x + b_i}\right)\right)$ on $\mathbf{dom}\ f = \{x \mid \sum_{i=1}^n e^{a_i^T x + b_i} < 1\}$. *(hint: You can use the fact that* $\log\left(\sum_{i=1}^n e^{y_i}\right)$ *is convex.)*
   Solution. $g(x) = \log \sum_{i=1}^m e^{a_i^T x + b_i}$ is convex (composition of the log-sum-exp function and an affine mapping), so $-g$ is concave. The function $h(y) = -\log y$ is convex and decreasing. Therefore $f(x) = h(-g(x))$ is convex

3. Let $S \subseteq \mathbb{R}^2$ be the set defined by $S = \{(x, y) \in \mathbb{R}^2_+ \mid y \leq \sqrt{x}\}$. Prove that S is a convex set. (20 points)

Proof: Let $S = \{(x, y) \in \mathbb{R}^2_+ \mid y \leq \sqrt{x}\}$. We want to show that S is a convex set.
To establish this, we need to show that for any two points (x_1, y_1) and (x_2, y_2) in S, the line segment joining them is also contained in S.
Suppose $(x_1, y_1), (x_2, y_2) \in S$. Then:
$$y_1 \leq \sqrt{x_1} \quad \text{and} \quad y_2 \leq \sqrt{x_2}.$$
Consider a point (x, y) on the line segment joining (x_1, y_1) and (x_2, y_2) defined as follows:
$$(x, y) = (1 - \lambda)(x_1, y_1) + \lambda(x_2, y_2),$$
where $\lambda \in [0, 1]$.
This gives:
$$x = (1 - \lambda)x_1 + \lambda x_2,$$
$$y = (1 - \lambda)y_1 + \lambda y_2.$$
We need to show that $y \leq \sqrt{x}$.
Since the function $f(t) = \sqrt{t}$ is concave on $\mathbb{R}_+$, we have:
$$\sqrt{(1 - \lambda)x_1 + \lambda x_2} \geq (1 - \lambda)\sqrt{x_1} + \lambda\sqrt{x_2}.$$
Therefore:
$$y = (1 - \lambda)y_1 + \lambda y_2 \leq (1 - \lambda)\sqrt{x_1} + \lambda\sqrt{x_2}.$$
Utilizing the concavity of the square root function, we have:
$$y \leq (1 - \lambda)\sqrt{x_1} + \lambda\sqrt{x_2} \leq \sqrt{(1 - \lambda)x_1 + \lambda x_2} = \sqrt{x}.$$
Thus, $(x, y) \in S$. Hence, the line segment joining (x_1, y_1) and (x_2, y_2) is contained in S.
Therefore, S is a convex set. $\square$

4. Let $f(X) = ||X||_2$ be the spectral norm of a matrix $X \in \mathbb{R}^{m\times n}$, defined as the largest singular value of X.

   1. Prove that $f(X)$ is convex. (10 points)
   2. Prove that the nuclear norm $f(X) = \sum_{i=1}^{r} \sigma_i(X)$, where $\sigma_i(X)$ are the singular values of X, is convex. (10 points)

Proof: Assume that triangle inequality and homogeneity hold for all norms, i.e., we have

$$\|X + Y\| \leq \|X\| + \|Y\|,$$

and

$$\|\lambda X\| = \lambda\|X\|,$$

for all norms. Thus, for any $\lambda \in [0, 1]$, we have

$$\|\lambda X + (1 - \lambda)Y\| \leq \lambda\|X\| + (1 - \lambda)\|Y\|.$$

Thus, with triangle inequality and homogeneity, the convexity of both norms can be easily proved. Now we need to prove those two properties for both norms. Since the homogeneity is straightforward, we only prove the triangle inequality here.

The triangle inequality for the spectral norm states that for any two matrices A and B of the same dimensions, the following inequality holds:

$$\|A + B\|_2 \leq \|A\|_2 + \|B\|_2.$$

The spectral norm of a matrix M is defined as:

$$\|M\|_2 = \sup_{\|x\|_{l2}=1} \|Mx\|_{l2},$$

where the supremum is taken over all vectors x with the Euclidean norm $\|x\|_{l2} = 1$.

To prove the triangle inequality, consider an arbitrary vector x with $\|x\|_{l2} = 1$. We have:

$$\|(A + B)x\|_{l2} = \|Ax + Bx\|_{l2}.$$

By the triangle inequality for vectors, we know that:

$$\|Ax + Bx\|_{l2} \leq \|Ax\|_{l2} + \|Bx\|_{l2}.$$

Using the definition of the spectral norm, we have:

$$\|Ax\|_{l2} \leq \|A\|_2 \quad \text{and} \quad \|Bx\|_{l2} \leq \|B\|_2$$

Therefore, we can write:

$$\|Ax + Bx\|_{l2} \leq \|A\|_2 + \|B\|_2.$$

Taking the supremum over all vectors x with $\|x\|_2 = 1$, we obtain:

$$\|A + B\|_2 = \sup_{\|x\|_{l2}=1} \|(A + B)x\|_{l2} \leq \sup_{\|x\|_{l2}=1} (\|Ax\|_{l2} + \|Bx\|_{l2}) \leq \|A\|_2 + \|B\|_2.$$

This completes the proof of the triangle inequality for the spectral norm.

The triangle inequality of the nuclear norm can be easily proved with the inequality of singular values for matrix sums.

5. Consider the ridge regression problem:

$$\min_{x\in\mathbb{R}^d} \frac{1}{2n}||Ax - b||^2 + \frac{\lambda}{2}||x||^2$$

where $A \in \mathbb{R}^{n\times d}$, $b \in \mathbb{R}^n$, $\|\cdot\|$ is the L_2 norm, and $\lambda > 0$. Show that the objective function is strongly convex (5 points) and find the strong convexity constant in terms of λ and the smallest eigenvalue of A^TA, i.e., assume that the strong convexity constant is m, express m in terms of σ and λ where σ is the smallest eigenvalue of A^TA.(15 points) *(hint: you can use the second-order differentiability of the strongly convex function.)*

Proof: We can rewrite the objective function as:

$$f(x) = \frac{1}{2n}(Ax-b)^T(Ax-b) + \frac{\lambda}{2}x^Tx$$

Thus, the objective function becomes:

$$f(x) = \frac{1}{2n}x^TA^TAx - \frac{1}{n}b^TAx + \frac{1}{2n}b^Tb + \frac{\lambda}{2}x^Tx$$

The constant term $\frac{1}{2n}b^Tb$ can be ignored for convexity purposes, as it does not affect the strong convexity.

To analyze strong convexity, we compute the Hessian of the objective function $f(x)$.

The Hessian of the full objective function is:

$$\nabla^2 f(x) = \frac{1}{n}A^TA + \lambda I$$

To show that the function is strongly convex, we need to show that the Hessian $\nabla^2 f(x)$ is positive definite and find the lower bound on its smallest eigenvalue.

* The matrix A^TA is symmetric and positive semi-definite. Let the smallest eigenvalue of A^TA be σ^2_{min}, where $\sigma_{\text{min}} \geq 0$ is the smallest singular value of A.

Thus, the eigenvalues of $\frac{1}{n}A^TA$ are $\frac{\sigma_i^2}{n}$, where σ_i^2 are the eigenvalues of A^TA, and $\frac{\sigma^2_{\text{min}}}{n}$ is the smallest eigenvalue.

* The identity matrix I has all eigenvalues equal to 1, so λI has eigenvalues equal to λ.

The eigenvalues of $\nabla^2 f(x) = \frac{1}{n}A^TA + \lambda I$ are the sums of the eigenvalues of $\frac{1}{n}A^TA$ and λI.

Thus, the smallest eigenvalue of $\nabla^2 f(x)$ is:

$$\lambda_{\text{min}}(\nabla^2 f(x)) = \frac{\sigma^2_{\text{min}}}{n} + \lambda$$

Since the Hessian is positive definite, the objective function is strongly convex. The strong convexity constant m is given by the smallest eigenvalue of the Hessian:

$$m = \frac{\sigma^2_{\text{min}}}{n} + \lambda$$

Thus, the strong convexity constant of the ridge regression problem is:

$$m = \frac{\sigma^2_{\text{min}}}{n} + \lambda$$

where σ_{min} is the smallest singular value of A, n is the number of data points, and $\lambda > 0$ is the regularization parameter.