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

Name: **Zhou Shouchen**
Student ID: 2021533042

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.

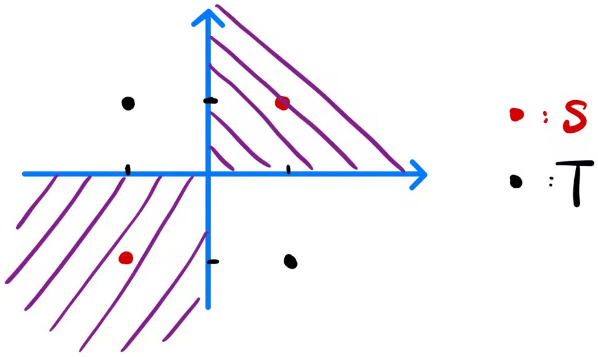
- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points ($\leq 20\%$) of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Do your homework by yourself. Any form of plagiarism will 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 $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, \ldots, n\}$. 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. Let $C_1 = \{x \in \mathbb{R}^n \mid (-a)^T x \leq -\alpha\}$, and $C_2 = \{x \in \mathbb{R}^n \mid a^T x \leq \beta\}$.
   C_1 and C_2 are two halp-spaces, which are convex sets.
   And since the intersection of two convex sets is also a convex set.
   So $C = C_1 \cap C_2 = \{x \in \mathbb{R}^n \mid \alpha \leq a^T x \leq \beta\}$ is a convex set.
   So the slab is a convex set.
2. Let $C = \{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, \ldots, n\}$.
   $\forall x, y \in C, \theta \in [0, 1]$, we have $\forall i = 1, \ldots, n, \alpha_i \leq x_i, y_i \leq \beta_i$.
   Let $z = \theta x + (1 - \theta) y$, then $\forall i = 1, \ldots, n$, we have:
   $$\begin{aligned} z_i &= \theta x_i + (1 - \theta) y_i \geq \theta \alpha_i + (1 - \theta) \alpha_i = \alpha_i \\ z_i &= \theta x_i + (1 - \theta) y_i \leq \theta \beta_i + (1 - \theta) \beta_i = \beta_i \end{aligned}$$
   i.e. $\alpha_i \leq z_i \leq \beta_i, \forall i = 1, \ldots, n$, so $z \in C$.
   So C is a convex set.
3. Let $C_1 = \{x \in \mathbb{R}^n \mid a_1^T x \leq b_1\}$, and $C_2 = \{x \in \mathbb{R}^n \mid a_2^T x \leq b_2\}$. So C_1, C_2 are two half-spaces, which are convex sets.
   And since the intersection of two convex sets is also a convex set.
   So $C = C_1 \cap C_2 = \{x \in \mathbb{R}^n \mid a_1^T x \leq b_1, a_2^T x \leq b_2\}$ is a convex set.
   So the wedge is a convex set.
4. Let $C = \left\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\right\}$.
   $\forall x \in C$, and for a fixed y, we have
   $$\begin{aligned} \|x - x_0\|_2 &\leq \|x - y\|_2 \\ \|x - x_0\|_2^2 &\leq \|x - y\|_2^2 \\ (x - x_0)^T (x - x_0) &\leq (x - y)^T (x - y) \\ (y - x_0)^T x &\leq \frac{1}{2} \left(\|y\|_2^2 - \|x_0\|_2^2\right) \end{aligned}$$
   From the definition, we know that for a fixed y, $(y - x_0)^T x \leq \frac{1}{2} \left(\|y\|_2^2 - \|x_0\|_2^2\right)$ is a half-space S_y.
   So $\forall y \in S$, we could see that $C = \bigcap_{y \in S} S_y$.
   And since each S_y is a half-space, which is a convex set. And from the theorem we have known, that the intersection of convex sets is also a convex set, so C is a convex set.



5. $C = \{x \mid \text{dist}(x, S) \leq \text{dist}(x, T)\}$ is not convex.
   We can construct a counter-example:
   Let $n = 2$, and $S = \{(2, 2), (-2, -2)\} \subseteq \mathbb{R}^2$, $T = \{(2, -2), (-2, 2)\} \subseteq \mathbb{R}^2$, as it's shown above.
   And it is clear that $C = \{x \mid \text{dist}(x, S) \leq \text{dist}(x, T)\} = \{(x, y) \mid (x \geq 0 \wedge y \geq 0) \vee (x \leq 0 \wedge y \leq 0)\}$, which is not a convex set.
   For example, we have $(x_1, y_1) = (2, 0) \in C, (x_2, y_2) = (0, -2) \in C, \theta = \dfrac{1}{2} \in [0, 1]$,
   but $\theta(x_1, y_1) + (1 - \theta)(x_2, y_2) = (1, -1) \notin C$.
   So we have constructed a counter-example to show that C is not a convex set.

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

We firstly introduce two Lemmas:

Lemma1: If $f(x)$ is convex, then $f(x)$ is quasiconvex.
Proof:
Let $S_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \le \alpha\}, \forall \alpha \in \mathbb{R}$
Then $\forall x, y \in S_\alpha$, we have $f(x), f(y) \le \alpha$, and $\forall \theta \in [0, 1]$, we have

$$\begin{aligned} f(\theta x + (1-\theta)y) &\le \theta f(x) + (1-\theta) f(y) \quad \text{(by convexity of f)} \\ &\le \theta\alpha + (1-\theta)\alpha \\ &= \alpha \end{aligned}$$

Which means that $\theta x + (1-\theta)y \in S_\alpha$, so S_α is convex $\forall \alpha \in \mathbb{R}$, so $f(x)$ is quasiconvex.

Lemma2: If $a, b \ge 0$ and $\theta \in [0, 1]$, then $\theta a + (1-\theta)b \ge a^\theta b^{1-\theta}$.
Proof:
1. If $a = 0$ or $b = 0$, it obviously takes the equality.
2. If $a, b > 0$, then let $f(x) = \log x$. We can get that $f'(x) = \dfrac{1}{x} > 0, f''(x) = -\dfrac{1}{x^2} < 0$, so $f(x)$ is concave and non-decreasing.
Then by Jensen's inequality and the monotonicity of $f(x)$, we have

$$\begin{aligned} \log(\theta a + (1-\theta)b) &\ge \theta \log a + (1-\theta)\log b \\ &= \log a^\theta + \log b^{(1-\theta)} \\ &= \log a^\theta b^{(1-\theta)} \\ \Rightarrow \theta a + (1-\theta)b &\ge a^\theta b^{1-\theta} \end{aligned}$$

So we have proved that $\theta a + (1-\theta)b \ge a^\theta b^{1-\theta}, \forall a, b \ge 0, \theta \in [0, 1]$.

1. $f(x) = e^x - 1, f'(x) = e^x, f''(x) = e^x > 0$, so $f(x)$ is convex.
   And from the Lemma1 we have proved, $f(x)$ is also quasiconvex.
   Since $f(x)$ is not a linear function, and it is convex, so it is not concave.
   Let $S_\alpha = \{x \mid f(x) = e^x \ge \alpha\}$, $\forall \alpha \in \mathbb{R}$. Then $\forall x, y \in S_\alpha$, we have $e^x \ge \alpha, e^y \ge \alpha$, and $\forall \theta \in [0, 1]$, we have
   <1>. If $\alpha \le 0$, then $e^{\theta x + (1-\theta)y} > 0 \ge \alpha$.
   <2>. If $\alpha > 0$, then

$$\begin{aligned} e^{\theta x + (1-\theta)y} &= e^{\theta x} \cdot e^{(1-\theta)y} \\ &= (e^x)^\theta \cdot (e^y)^{1-\theta} \\ &\ge \alpha^\theta \cdot \alpha^{1-\theta} \\ &= \alpha \end{aligned}$$

   Which means that $\theta x + (1-\theta)y \in S_\alpha, \forall \alpha \in R$, so S_α is convex $\forall \alpha \in R$.
   So $f(x)$ is quasiconcave.
   So above all, we have proved that $f(x) = e^x - 1$ is convex,not concave, quasiconvex, and quasiconcave.

2. $f(x_1,x_2)=x_1x_2$, $\nabla f(x_1,x_2)=\begin{bmatrix}x_2\\x_1\end{bmatrix}$, $\nabla^2 f(x_1,x_2)=\begin{bmatrix}0&1\\1&0\end{bmatrix}$, $\left|\nabla^2 f(x_1,x_2)\right|=-1<0$, so $f(x_1,x_2)$ is not convex.

   Similarly, let $g(x_1,x_2)=-f(x_1,x_2)=-x_1x_2$, $\nabla g(x_1,x_2)=\begin{bmatrix}-x_2\\-x_1\end{bmatrix}$, $\nabla^2 g(x_1,x_2)=\begin{bmatrix}0&-1\\-1&0\end{bmatrix}$, $\left|\nabla^2 g(x_1,x_2)\right|=-1<0$, so $f(x_1,x_2)$ is not concave.

   Let $S_\alpha=\{(x_1,x_2)\mid x_1x_2\le\alpha\}$, then $\forall\alpha>0,\alpha\ne1$, we have $(1,\alpha)\in S_\alpha,(\alpha,1)\in S_\alpha$, and let $\theta=\frac{1}{2}\in[0,1]$, we have

$$\left(\frac{1+\alpha}{2}\right)^2-\alpha=\frac{(\alpha-1)^2}{4}>0$$

   Which means that $\theta(1,\alpha)+(1-\theta)(\alpha,1)=\left(\frac{1+\alpha}{2},\frac{1+\alpha}{2}\right)\notin S_\alpha$, so S_α is not convex for some α, so $f(x_1,x_2)$ is not quasiconvex.

   Let $S'_\alpha=\{(x_1,x_2)\mid x_1x_2\ge\alpha\}$, then $\forall\alpha\in\mathbb{R}$: $\forall(x_1,x_2),(y_1,y_2)\in S'_\alpha$, we have $x_1x_2\ge\alpha, y_1y_2\ge\alpha$, and $\forall\theta\in[0,1]$, from Lemma2, we have:

$$\theta x_1+(1-\theta)y_1\ge x_1^\theta y_1^{1-\theta}$$
$$\theta x_2+(1-\theta)y_2\ge x_2^\theta y_2^{1-\theta}$$

   And we times these two inequality, we can get that

$$[\theta x_1+(1-\theta)y_1]\,[\theta x_2+(1-\theta)y_2]\ge\left(x_1^\theta y_1^{1-\theta}\right)\left(x_2^\theta y_2^{1-\theta}\right)\tag{1}$$

   Let $(z_1,z_2)=\theta\,(x_1,x_2)+(1-\theta)\,(y_1,y_2)=[\theta x_1+(1-\theta)y_1,\theta x_2+(1-\theta)y_2]$, then we have

$$\begin{aligned}z_1z_2&=[\theta x_1+(1-\theta)y_1]\cdot[\theta x_2+(1-\theta)y_2]\\&\ge\left(x_1^\theta y_1^{1-\theta}\right)\left(x_2^\theta y_2^{1-\theta}\right)\qquad\text{(from inequality we proved above (1))}\\&=(x_1x_2)^\theta\,(y_1y_2)^{1-\theta}\\&\ge\alpha^\theta\alpha^{1-\theta}\\&=\alpha\end{aligned}$$

   So above all, we have proved that $\forall\alpha\in\mathbb{R},S'_\alpha$ is convex, so $f(x_1,x_2)$ is quasiconcave.

   So above all, we have proved that $f(x_1,x_2)=x_1x_2$ is not convex, not concave, not quasiconvex, and quasiconcave.

3. $f(x_1,x_2)=\frac{1}{x_1x_2}$, $\nabla f(x_1,x_2)=\begin{bmatrix}-\frac{1}{x_1^2x_2}\\-\frac{1}{x_1x_2^2}\end{bmatrix}$, $\nabla^2 f(x_1,x_2)=\begin{bmatrix}\frac{2}{x_1^3x_2}&\frac{1}{x_1^2x_2^2}\\\frac{1}{x_1^2x_2^2}&\frac{2}{x_1x_2^3}\end{bmatrix}$.

   Since $\frac{2}{x_1^3x_2}>0$, and $\left|\nabla^2 f(x_1,x_2)\right|=\frac{3}{x_1^4x_2^4}>0$, so $f(x_1,x_2)$ is convex.

   From Lemma1 we have proved, $f(x_1,x_2)$ is also quasiconvex.

   Since $f(x_1,x_2)$ is not a linear function, and it is convex, so it is not concave.

   Let $S_\alpha=\{(x_1,x_2)\mid\frac{1}{x_1x_2}\ge\alpha\}$, then $\forall\alpha>0,\alpha\ne1$, we have $\left(1,\frac{1}{\alpha}\right),\left(\frac{1}{\alpha},1\right)\in S_\alpha$, and let $\theta=\frac{1}{2}\in[0,1]$, we have

$$\left(\frac{1}{\frac{1}{2}\cdot1+\frac{1}{2}\cdot\frac{1}{\alpha}}\right)^2-\alpha=\alpha\left(\frac{4}{\alpha+\frac{1}{\alpha}+2}-1\right)<0$$

   So we have given counterexample that S_α is not convex for some α, so $f(x_1,x_2)$ is not quasiconvex.

   So above all, $f(x_1,x_2)$ is convex, not concave, quasiconvex, and not quasiconcave.

4. Since $\|\cdot\|$ is a norm on $\mathbb{R}^n$, so we have $\forall\theta\in[0,1],x\in\mathbb{R}^n,\|\theta x\|=\theta\|x\|$, and $\forall x,y\in\mathbb{R}^n$, we have $\|x+y\|\le\|x\|+\|y\|$.

   So $\forall x,y\in\mathbb{R}^n,\theta\in[0,1]$, we have

$$\begin{aligned}f(\theta x+(1-\theta)y)&=\|A(\theta x+(1-\theta)y)-b\|\\&=\|\theta(Ax-b)+(1-\theta)(Ay-b)\|\\&\le\|\theta(Ax-b)\|+\|(1-\theta)(Ay-b)\|\qquad\text{(by triangle inequality)}\\&=\theta\|Ax-b\|+(1-\theta)\|Ay-b\|\\&=\theta f(x)+(1-\theta)f(y)\end{aligned}$$

   So we have proved that $f(x)=\|Ax-b\|$ is convex.

5. Let $\phi(y) = \log\left(\sum\limits_{i=1}^{n} e^{y_i}\right)$, and we have known that $\phi(y)$ is convex.

   From the property of composition with affain functions, we can get that $\phi\left(a_i^T x + b\right)$ is convex.

   Let $g : \mathbb{R}^n \to \mathbb{R}$, $g(x) = -\phi(a_i^T x + b) = -\log\left(\sum\limits_{i=1}^{n} e^{a_i^T x + b_i}\right)$, which is a concave function.

   And let $h : \mathbb{R}_{++} \to \mathbb{R}$, $h(x) = -\log x$, and $h'(x) = -\dfrac{1}{x} < 0, h''(x) = \dfrac{1}{x^2} > 0$, so $h(x)$ is a convex function, and it is non-increasing.

   From the composition with scalar functions, since $f(x) = h\left(g(x)\right)$, where $h(x)$ is convex and non-increasing, $g(x)$ is concave, so $f(x)$ is convex.

   So above all, we have proved that $f(x) = -\log\left(-\log\left(\sum\limits_{i=1}^{n} e^{a_i^T x + b_i}\right)\right)$ 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)

Solution

Let $f(x) = -\sqrt{x}$, then we have

$$
\begin{aligned}
f'(x) &= -\frac{1}{2\sqrt{x}} \\
f''(x) &= \frac{1}{4} x^{-\frac{3}{2}}
\end{aligned}
$$

And since $x \in \mathbb{R}_+$, so we have $f''(x) \geq 0$, which means that $f(x)$ is a convex function.
So $g(x) = \sqrt{x}$ is a concave function.
Then $\forall (x_1, y_1), (x_2, y_2) \in S, \theta \in [0, 1]$, we have

$$
\begin{aligned}
y_1 &\leq \sqrt{x_1} \\
y_2 &\leq \sqrt{x_2}
\end{aligned}
$$

So we have

$$
\begin{aligned}
\theta y_1 + (1-\theta) y_2 &\leq \theta \sqrt{x_1} + (1-\theta)\sqrt{x_2} \\
&= \theta g(x_1) + (1-\theta) g(x_2) \\
&\leq g(\theta x_1 + (1-\theta) x_2) \qquad \text{(concavity of } g(x)\text{)} \\
&= \sqrt{\theta x_1 + (1-\theta) x_2}
\end{aligned}
$$

Which means that $\theta(x_1, y_1) + (1-\theta)(x_2, y_2) = (\theta x_1 + (1-\theta) x_2, \theta y_1 + (1-\theta) y_2) \in S$.
So above all, we have proved that S is a convex set.

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)

   Solution

   1. $f(X) = \sqrt{\lambda_{\max}(X^TX)} = \sqrt{\sup_{y\in\mathbb{R}^n} \dfrac{y^T(X^TX)y}{\|y\|_2^2}} = \sup_{y\in\mathbb{R}^n} \sqrt{\dfrac{(Xy)^T(Xy)}{\|y\|_2^2}} = \sup_{y\in\mathbb{R}^n} \dfrac{\|Xy\|_2}{\|y\|_2}$

      Let $g(X,y) = \dfrac{\|Xy\|_2}{\|y\|_2}$, then for a fixed y, and $\forall X_1, X_2 \in \mathbb{R}^{m\times n}$, $\forall\theta \in [0,1]$, we have

$$
\begin{aligned}
g(\theta X_1 + (1-\theta)X_2, y) &= \frac{\|\left[\theta X_1 + (1-\theta)X_2\right] y\|_2}{\|y\|_2} \\
&\leq \frac{\|\theta X_1 y\|_2 + \|(1-\theta)X_2 y\|_2}{\|y\|_2} \\
&= \theta\frac{\|X_1 y\|_2}{\|y\|_2} + (1-\theta)\frac{\|X_2 y\|_2}{\|y\|_2} \\
&= \theta f(X_1, y) + (1-\theta) f(X_2, y)
\end{aligned}
$$

      So when y is fixed, $g(X,y)$ is convex in X.

      From the property of pointwise supremum, we have for each y, $g(X,y) = \dfrac{\|Xy\|_2}{\|y\|_2}$ is convex in X.

      Then $f(X) = \sup_{y\in\mathbb{R}^n} g(X,y)$ is convex.

   2. Similarly with the relation between L_1 norm and L_∞ norm, we could guess that the nuclear norm is the dual norm of the spectral norm. And we could prove our guess.

      Define $\|X\|_* = \sum_{i=1}^{r} \sigma_i(X)$ to be the nuclear norm, and $\|X\|_2 = \sigma_{\max}(X)$ to be the spectral norm.

      i.e. we want to prove that

$$
\|X\|_* = \sup_{\|Z\|_2\leq 1} < Z, X >
$$

      We firstly apply SVD to X, and we have $X = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with singular values on the diagonal. And suppose that X has r singular values.

      Define $P = UV^T = UI_nV^T$, then we could find that all the singular values of P are 1, i.e. $\|P\|_2 = 1$. Then we can get that

$$
\begin{aligned}
\sup_{\|Z\|_2\leq 1} < Z, X > &= \sup_{\|Z\|_2\leq 1} \mathbf{Tr}(Z^TX) \\
&\geq \mathbf{Tr}(P^TX) \qquad (\|P\|_2 = 1 \leq 1) \text{ satisfies the constraint} \\
&= \mathbf{Tr}((UV^T)^TU\Sigma V^T) \\
&= \mathbf{Tr}(V^TVU^TU\Sigma) \qquad (\mathbf{Tr}(AB) = \mathbf{Tr}(BA)) \\
&= \mathbf{Tr}(\Sigma) \\
&= \sum_{i=1}^{r} \sigma_i(X) \\
&= \|X\|_*
\end{aligned}
$$

      i.e. we have proved that

$$
\|X\|_* \leq \sup_{\|Z\|_2\leq 1} < Z, X > \tag{2}
$$

      Then we can prove the other direction of the inequality.

$$
\begin{aligned}
\sup_{\|Z\|_2\leq 1} < Z, X > &= \sup_{\|Z\|_2\leq 1} \mathbf{Tr}\left(Z^T\left(U\Sigma V^T\right)\right) \\
&= \sup_{\|Z\|_2\leq 1} \mathbf{Tr}\left(\left(V^TZ^TU\right)\Sigma\right) \qquad (\mathbf{Tr}(AB) = \mathbf{Tr}(BA))
\end{aligned}
$$

$$\begin{aligned}
\sup_{\|Z\|_2\leq 1} < Z, X > &= \sup_{\|Z\|_2\leq 1} \sum_{i=1}^{r} \sigma_i(X) \cdot \left(V^T Z^T U\right)_{ii} \\
&= \sup_{\|Z\|_2\leq 1} \sum_{i=1}^{r} \sigma_i(X) \left(u_i^T Z v_i\right) && (u_i, v_i \text{ are the } i\text{-th column of } U, V) \\
&= \sup_{\|Z\|_2\leq 1} \sum_{i=1}^{r} \sigma_i(X) \left\|u_i^T Z v_i\right\|_2 \\
&\leq \sup_{\|Z\|_2\leq 1} \sum_{i=1}^{r} \sigma_i(X) \|u_i\|_2 \|Z\|_2 \|v_i\|_2 && \text{(Cauchy-Schwarz Inequality)} \\
&\leq \sum_{i=1}^{r} \sigma_i(X) && (\|Z\|_2 \leq 1, \|u_i\|_2 = \|v_i\|_2 = 1) \\
&= \|X\|_*
\end{aligned}$$

i.e. we have proved that

$$\|X\|_* \geq \sup_{\|Z\|_2\leq 1} < Z, X > \tag{3}$$

So combine the inequalities (2) and (3) we can get that the nuclear norm is the dual norm of the spectral norm. i.e.

$$\|X\|_* = \sup_{\|Z\|_2\leq 1} < Z, X > \tag{4}$$

And then we prove the triangle inequality of the nuclear norm: $\forall X, Y \in \mathbb{R}^{m\times n}$,

$$\begin{aligned}
\|X+Y\|_* &= \sup_{\|Z\|_2\leq 1} < Z, X+Y > && \text{(By the equation (4) we have proved)} \\
&= \sup_{\|Z\|_2\leq 1} (< Z, X > + < Z, Y >) \\
&\leq \sup_{\|Z\|_2\leq 1} < Z, X > + \sup_{\|Z\|_2\leq 1} < Z, Y > && \text{(By the property of supremum)} \\
&= \|X\|_* + \|Y\|_*
\end{aligned}$$

So we have proved that $\forall X, Y \in \mathbb{R}^{m\times n}$,

$$\|X+Y\|_* \leq \|X\|_* + \|Y\|_* \tag{5}$$

Suppose that the i-th eigenvalue of $X^T X$ is $\lambda_i(X^T X)$, and the i-th singular value of X is $\sigma_i(X)$. Which means that $\sigma_i(X) = \sqrt{\lambda_i(X^T X)}$.
Then we have the i-th singular value of θX, where $\theta \in [0, 1]$ is

$$\sigma_i(\theta X) = \sqrt{\lambda_i((\theta X)^T(\theta X))} = \sqrt{\lambda_i(\theta^2 X^T X)} = \theta\sqrt{\lambda_i(X^T X)} = \theta\sigma_i(X)$$

So we have

$$\|\theta X\|_* = \sum_{i=1}^{r} \sigma_i(\theta X) = \sum_{i=1}^{r} \theta\sigma_i(X) = \theta\sum_{i=1}^{r} \sigma_i(X) = \theta\|X\|_*$$

So above all, we have proved that $\forall\theta \in [0, 1]$,

$$\|\theta X\|_* = \theta\|X\|_* \tag{6}$$

With the above conclusions, we could prove that $\forall X_1, X_2 \in \mathbb{R}^{m\times n}, \theta \in [0, 1]$

$$\begin{aligned}
f(\theta X_1 + (1-\theta)X_2) &= \|\theta X_1 + (1-\theta)X_2\|_* \\
&\leq \|\theta X_1\|_* + \|(1-\theta)X_2\|_* && \text{(By the inequality (5) we have proved)} \\
&= \theta\|X_1\|_* + (1-\theta)\|X_2\|_* && \text{(By the equality (6) we have proved)} \\
&= \theta f(X_1) + (1-\theta)f(X_2)
\end{aligned}$$

So above all, we have proved that the nuclear norm is convex.

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 $A^T A$, 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 $A^T A$.(15 points) *(hint: you can use the second-order differentiability of the strongly convex function.)*

Solution

(1) Let $f(x) = \frac{1}{2n}\|Ax - b\|^2 + \frac{\lambda}{2}\|x\|^2, g(x) = \frac{1}{2n}\|Ax - b\|^2 + \frac{\lambda}{2}\|x\|^2 - \frac{\mu}{2}\|x\|^2$, we want to show that $f(x)$ is strongly convex, i.e. $g(x)$ is convex for some $\mu > 0$.

$$\nabla g(x) = \frac{1}{n}A^T Ax + (\lambda - \mu)x$$
$$\nabla^2 g(x) = \frac{1}{n}A^T A + (\lambda - \mu)I$$

To make sure $g(x)$ is convex, we should have

$$\nabla^2 g(x) \succeq 0$$

Let $\lambda_{\min}(\cdot)$ denote the smallest eigenvalue of a matrix, then we have

$$\lambda_{\min}\left(\frac{1}{n}A^T A + (\lambda - \mu) I\right) = \lambda_{\min}\left(\frac{1}{n}A^T A\right) + \lambda - \mu \geq 0$$
$$\Rightarrow 0 < \mu \leq \lambda + \lambda_{\min}\left(\frac{1}{n}A^T A\right)$$

Since $\forall x \in \mathbb{R}^n, x^T(A^T A)x = \|Ax\|^2 \geq 0$, so $\lambda_{\min}\left(\frac{1}{n}A^T A\right) \geq 0$.

And since $\lambda > 0$, so $\lambda_{\min}\left(\frac{1}{n}A^T A\right) + \lambda > 0$, so there must exist some $\mu > 0$ such that $g(x)$ is convex, which means $f(x)$ is strongly convex.

(2) And among all the μ that makes $g(x)$ convex, we want to find the biggest one, which should be the convexity constant of $f(x)$, i.e.

$$\mu = \lambda + \frac{\sigma}{n}$$

where σ is the smallest eigenvalue of $A^T A$.

So above all, we have shown that $f(x)$ is $\left(\lambda + \frac{\sigma}{n}\right)$-strongly convex.