Introduction to Optimization: Homework #2

Due on October 29, 2020

Professor Andre Milzarek

Peng Deng

Assignment A2.1

Consider the unconstrained optimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} f(\boldsymbol{x}) = \frac{1}{3}x_1^3 - x_1\left(\frac{3}{2}+x_2^2\right)+x_2^4 \tag{1}$$

a) Is the function f coercive?

b) Compute the gradient and Hessian of f and calculate all stationary points.

c) For each stationary point $\boldsymbol{x}^*$ found in part b) investigate whether $\boldsymbol{x}^*$ is a local maximizer, local minimizer, or saddle point and explain your answer.

d) Does the mapping f possess any strict local or global minimizer?

Solution

Subproblem a)

The function f is not coervive.

Proof: We set $x_2 = 0$, so

$$f(\boldsymbol{x}) = f(x_1, 0) = \frac{1}{3}x_1^3 - \frac{3}{2}x_1 \qquad (\forall x_1 \in \mathbb{R}) \tag{2}$$

when $x_1 \to -\infty$, the norm

$$\|\boldsymbol{x}\| = \sqrt{x_1^2+x_2^2} = \sqrt{x_1^2+0^2} \to \infty \tag{3}$$

and because $\frac{1}{3}x_1^3$ is dominant comparing with $\frac{3}{2}x_1$, so $f(\boldsymbol{x}) = f(x_1, 0) \to -\infty$ when $x_1 \to -\infty$. In all we can sum up that in the along the direction $(x_1, 0), x_1 \to -\infty$, we have

$$\lim_{\|\boldsymbol{x}\|\to\infty} f(\boldsymbol{x}) = \frac{1}{3}x_1^3 - \frac{3}{2}x_1 = -\infty \neq +\infty \tag{4}$$

So the function f is not coercive.

Subproblem b)

○The gradient of f can be derived as follow:

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} x_1^2 - x_2^2 - \frac{3}{2} \\ -2x_1x_2 + 4x_2^3 \end{pmatrix} \tag{5}$$

○ The Hessian of f can be derived as follow:

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 & -2x_2 \\ -2x_2 & -2x_1 + 12x_2^2 \end{pmatrix} \tag{6}$$

○ In order to calculate the stationary points, we set the gradient of f equals $\mathbf{0}$:

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} x_1^2 - x_2^2 - \frac{3}{2} \\ -2x_1x_2 + 4x_2^3 \end{pmatrix} = \mathbf{0} \quad \Longrightarrow \quad \begin{cases} x_1^2 - x_2^2 - \frac{3}{2} = 0 \\ -2x_1x_2 + 4x_2^3 = 0 \end{cases} \tag{7}$$

By solving the equation, we can derive all the stationary points:

$$P_1\left(-\frac{\sqrt{6}}{2}, 0\right) \quad P_2\left(\frac{\sqrt{6}}{2}, 0\right) \quad P_3\left(\frac{3}{2}, -\frac{\sqrt{3}}{2}\right) \quad P_4\left(\frac{3}{2}, \frac{\sqrt{3}}{2}\right)$$

Assignment A2.1 continued on next page. . .

Subproblem c)
In order to investigate whether the stationary point $\boldsymbol{x}^*$ is a local maximizer, local minimizer, or saddle point, we have to calculate the Hessian matrix at these stationary points.

$$\begin{aligned}
&\text{For } P_1 : \nabla^2 f(\boldsymbol{x}^*) = \begin{pmatrix} 2x_1 & -2x_2 \\ -2x_2 & -2x_1 + 12x_2^2 \end{pmatrix} = \begin{pmatrix} -\sqrt{6} & 0 \\ 0 & \sqrt{6} \end{pmatrix} \\
&\text{For } P_2 : \nabla^2 f(\boldsymbol{x}^*) = \begin{pmatrix} 2x_1 & -2x_2 \\ -2x_2 & -2x_1 + 12x_2^2 \end{pmatrix} = \begin{pmatrix} \sqrt{6} & 0 \\ 0 & -\sqrt{6} \end{pmatrix} \\
&\text{For } P_3 : \nabla^2 f(\boldsymbol{x}^*) = \begin{pmatrix} 2x_1 & -2x_2 \\ -2x_2 & -2x_1 + 12x_2^2 \end{pmatrix} = \begin{pmatrix} 3 & \sqrt{3} \\ \sqrt{3} & 6 \end{pmatrix} \\
&\text{For } P_4 : \nabla^2 f(\boldsymbol{x}^*) = \begin{pmatrix} 2x_1 & -2x_2 \\ -2x_2 & -2x_1 + 12x_2^2 \end{pmatrix} = \begin{pmatrix} 3 & -\sqrt{3} \\ -\sqrt{3} & 6 \end{pmatrix}
\end{aligned} \tag{8}$$

◦ For P_1 and P_2, the eigenvalue of the Hessian matrix is $-\sqrt{6}$ and $\sqrt{6}$, so one eigenvalue is positive and one eigenvalue is negative, so the Hessian matrix is indefinite. So P_1 and P_2 are saddle points.
◦ For P_3 and P_4, we set the eigenvalue of the Hessian matrix is λ_1 and λ_2, so we have:

$$\begin{cases} \lambda_1 + \lambda_2 = 3 \times 6 = 18 \\ \lambda_1 \times \lambda_2 = \det\left(\nabla^2 f(\boldsymbol{x}^*)\right) = 15 \end{cases} \tag{9}$$

From equation 9, we can derive that both λ_1 and λ_2 are positive, so the Hessian matrix at P_3 and P_4 are positive definite, so P_3 and P_4 are local minimizers.

Subproblem d)
According to c), we know that the Hessian matrix at P_3 and P_4 are positive definite, so f possesses two strict local minimizer:

$$P_3\left(\frac{3}{2}, -\frac{\sqrt{3}}{2}\right) \quad P_4\left(\frac{3}{2}, \frac{\sqrt{3}}{2}\right)$$

According to a), we know that $f \to -\infty$ in certain direction, so f doesn't have any global minimizer.

Assignment A2.2

Consider the unconstrained optimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^3} f(\boldsymbol{x}) = x_1^4 - 2x_1^2 + x_2^2 + 2x_2x_3 + 2x_3^2 \tag{10}$$

a) Is the mapping f coercive?

b) Verify whether the function f is convex on $\mathbb{R}^3$.

c) Find all stationary points of f and classify them according to whether they are saddle points, strict / non-strict, local / global, minimum / maximum points.

Subproblem a)
The mapping f is coercive.

Proof:

$$
\begin{aligned}
f(\boldsymbol{x}) &= x_1^4 - 2x_1^2 + x_2^2 + 2x_2x_3 + 2x_3^2 \\
&= \frac{1}{2}x_1^4 + \frac{1}{2}x_1^4 - 2x_1^2 + \frac{3}{4}x_2^2 + \frac{1}{4}x_2^2 + 2x_2x_3 + \frac{4}{3}x_3^2 + \frac{2}{3}x_3^2 \\
&= \frac{1}{2}x_1^4 + \frac{1}{2}x_1^4 - 2x_1^2 + \left(\frac{\sqrt{3}}{2}x_2 + \frac{2}{\sqrt{3}}x_3\right)^2 + \frac{1}{4}x_2^2 + \frac{2}{3}x_3^2 \\
&\geq \frac{1}{2}x_1^4 + \frac{1}{4}x_2^2 + \frac{2}{3}x_3^2 + \left[\frac{1}{2}x_1^4 - 2x_1^2\right] \\
&\geq \frac{1}{2}x_1^4 + \frac{1}{4}x_2^2 + \frac{2}{3}x_3^2 + \left[\min_{x\in\mathbb{R}}\left(\frac{1}{2}x^4 - 2x^2\right)\right]
\end{aligned}
\tag{11}
$$

Now, we consider the function $g(x) = \frac{1}{2}x^4 - 2x^2$, $x \in \mathbb{R}$. By plotting the graph of $g(x)$ as Figure 1, we can find that $g(x)$ has minimum. By calculate the derivative of $g(x)$ and set $g'(x) = 0$, we can find the stationary point $A(-\sqrt{2}, -2)$, $B(\sqrt{2}, -2)$, $C(0, 0)$.

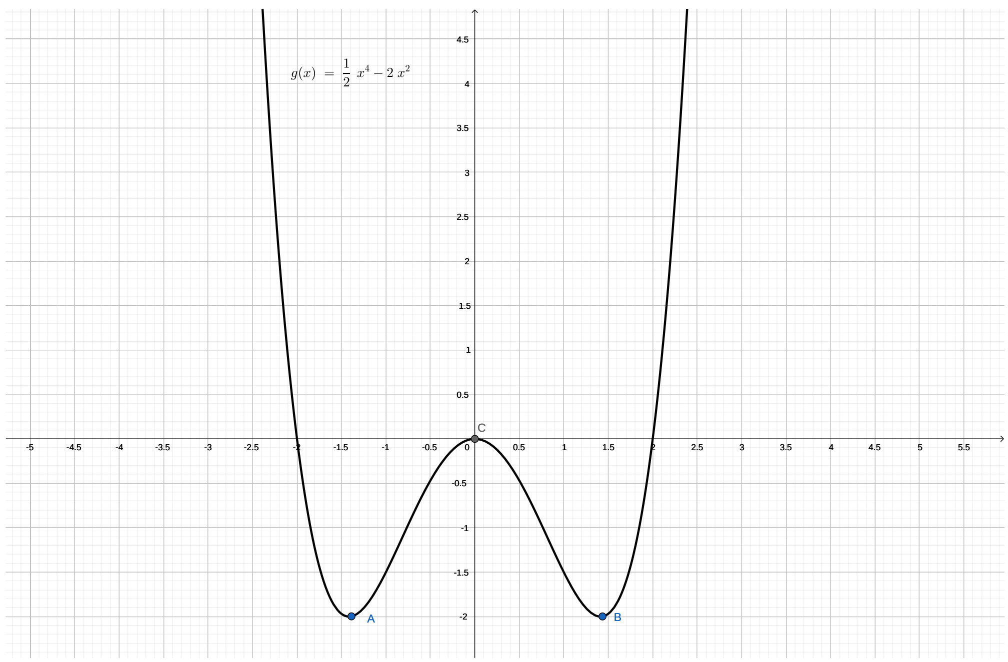

Figure 1 Graph of function $g(x) = \frac{1}{2}x^4 - 2x^2$

Combining the graph, we can derive the global minimum of $g(x)$ is $-\sqrt{2}$. So eqution 11 can be sequentially written as follow:

$$
\begin{aligned}
f(\boldsymbol{x}) &\geq \frac{1}{2}x_1^4 + \frac{1}{4}x_2^2 + \frac{2}{3}x_3^2 + \left[\min_{x\in\mathbb{R}}\left(\frac{1}{2}x^4 - 2x^2\right)\right] \\
&\geq \frac{1}{2}x_1^4 + \frac{1}{4}x_2^2 + \frac{2}{3}x_3^2 + -\sqrt{2}
\end{aligned}
\tag{12}
$$

So we have proved that the mapping f is coercive.

Subproblem b)

In order to verify whether the function f is convex, we would like to calculate the gradient and Hessian matrix of $f(x)$ as follow.

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{pmatrix} = \begin{pmatrix} 4x_1^3 - 4x_1 \\ 2x_2 + 2x_3 \\ 2x_2 + 4x_3 \end{pmatrix}$$
$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3 \partial x_3} \end{pmatrix} = \begin{pmatrix} 12x_1^2 - 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix} \tag{13}$$

If f is convex, then the Hessian matrix has to be positive semidefinite, which means the determinants of all principal minors of the Hessian matrix should ≥ 0. However, the determinant of the follow pincipal minor is not always ≥ 0.

$$\begin{aligned} \left|12x_1^2 - 4\right| \geq 0 \quad & \left(x_1 \in \left[-\infty, \frac{\sqrt{3}}{3}\right] \cup \left[\frac{\sqrt{3}}{3}, +\infty,\right]\right) \\ \left|12x_1^2 - 4\right| < 0 \quad & \left(x_1 \in \left[-\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}\right]\right) \end{aligned} \tag{14}$$

Thus, $f(x)$ is not convex.

Subproblem c)

In order to find the stationary points, we set the gradient of f equals 0.

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 4x_1^3 - 4x_1 \\ 2x_2 + 2x_3 \\ 2x_2 + 4x_3 \end{pmatrix} = 0$$
$$\Longrightarrow \begin{cases} 4x_1^3 - 4x_1 = 0 \\ 2x_2 + 2x_3 = 0 \\ 2x_2 + 4x_3 = 0 \end{cases} \tag{15}$$

By solving the equation, we can derive all the stationary points:

$$P_1\,(0,0,0) \quad P_2\,(1,0,0) \quad P_3\,(-1,0,0)$$

◦ For point P_1: because the Hessian matrix at point P_1, say $\nabla^2 f(P_1)$ is a real symmetric matrix, we can use the eigenvalues of $\nabla^2 f(P_1)$ to judge the definiteness of matrix $\nabla^2 f(P_1)$.

$$\nabla f^2(P_1) = \begin{pmatrix} 12x_1^2 - 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix} = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix} \tag{16}$$

The characteristic polynomial of $\nabla^2 f(P_1)$ can be derived as follow

$$\left|\nabla^2 f(P_1) - \lambda I\right| = \begin{vmatrix} -4-\lambda & 0 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 2 & 4-\lambda \end{vmatrix} = (-4-\lambda)(\lambda^2 - 6\lambda + 4) \tag{17}$$

Thus, the eigenvalues of $\nabla^2 f(P_1)$ can be derived as $\lambda_1 = -4, \lambda_2 = 3+\sqrt{5}, \lambda_3 = 3-\sqrt{5}$. So some eigenvalue is positive, some eigenvalue is negative, thus the Hessian matrix $\nabla^2 f(P_1)$ is **indefinite**. So P_1 is a saddle point.

◦ For point P_2 : because the Hessian matrix at point P_2, say $\nabla^2 f(P_2)$ is a real symmetric matrix, we can use the eigenvalues of $\nabla^2 f(P_2)$ to judge the definiteness of matrix $\nabla^2 f(P_2)$.

$$\nabla f^2(P_2) = \begin{pmatrix} 12x_1^2 - 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix} \tag{18}$$

The characteristic polynomial of $\nabla^2 f(P_2)$ can be derived as follow

$$\left|\nabla^2 f(P_2) - \lambda I\right| = \begin{vmatrix} 8-\lambda & 0 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 2 & 4-\lambda \end{vmatrix} = (8-\lambda)(\lambda^2 - 6\lambda + 4) \tag{19}$$

Thus, the eigenvalues of $\nabla^2 f(P_2)$ can be derived as $\lambda_1 = 8, \lambda_2 = 3+\sqrt{5}, \lambda_3 = 3-\sqrt{5}$. So $\forall \lambda_i > 0$, thus matrix $\nabla^2 f(P_2)$ is **positive definite**. So P_2 is a strict local minimizer.

◦ For point P_3 : The characteristic polynomial of $\nabla^2 f(P_3)$ is the same as $\nabla^2 f(P_2)$, so the eigenvalues of $\nabla^2 f(P_3)$ is the same as $\nabla^2 f(P_2)$. So $\forall \lambda_i > 0$, thus matrix $\nabla^2 f(P_3)$ is **positive definite**. So P_3 is also a strict local minimizer.

◦ For global: because the function is coercive, so there is at least one global minimizer. The global minimizer must be local minimizer, so the global minimier is among P_2 and P_3. The value of f at P_2 and P_3 is:

$$f(P_2) = f(P_3) = -1 \tag{20}$$

Thus, both P_2 and P_3 are global minimizer, and non-strict.

To sum up, P_1 is a saddle point. P_2 is a strict local minimizer and a non-strict global minimizer. P_3 is a strict local minimizer and a non-strict global minimizer

Assignment A2.3

In this exercise, we study convexity of various sets.

a) Verify whether the following sets are convex or not and explain your answer!

$$\begin{aligned} X_1 &= \left\{x \in \mathbb{R}^n : \alpha \le \left(a^\top x\right)^2 \le \beta\right\}, \quad \alpha, \beta \in \mathbb{R}, \alpha \le \beta, a \in \mathbb{R}^n \\ X_2 &= \{x \in \mathbb{R}^n : \|x-a\|_2 \le \|x-b\|_2\}, \quad a, b \in \mathbb{R}^n, a \ne b \end{aligned} \tag{21}$$

b) Let the set of all positive semidefinite and symmetric $n \times n$ matrices be denoted by $\mathbb{S}^n_+$. Show that the set $X := \left\{A \in \mathbb{R}^{n\times n} : A \in \mathbb{S}^n_+ \text{ and } \operatorname{tr}(A) = 1\right\}$ is a convex subset of $\mathbb{R}^{n\times n}$.

c) Decide whether the following statements are true or false. Explain your answer and either present a proof / verification or a counter-example.

- The union of two convex sets $X_1, X_2 \subset \mathbb{R}^n, X_1 \ne X_2$ is never a convex set.
- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a concave. Then, the set $X := \{x \in \mathbb{R}^n : f(x) \ge 0\}$ is convex.

Assignment A2.3 continued on next page...

Subproblem a)

∘ X_1 is not convex. We can present a counter example as $\alpha = 1, \beta = 5$ and $n = 2$. And we set $a^\top = (1, 1)$. So we can rewrite X_1 as follow:

$$X_1 = \left\{x \in \mathbb{R}^2 : 1 \leq (x_1 + x_2)^2 \leq 5\right\} \tag{22}$$

We can draw the graph of X_1 as 2, so we can see that if we choose one point as $A \in X_1$ and another point $B \in X_1$, some points on the connecting line of A and B(the red line) are not in set X_1. Thus, set X_1 is not convex.

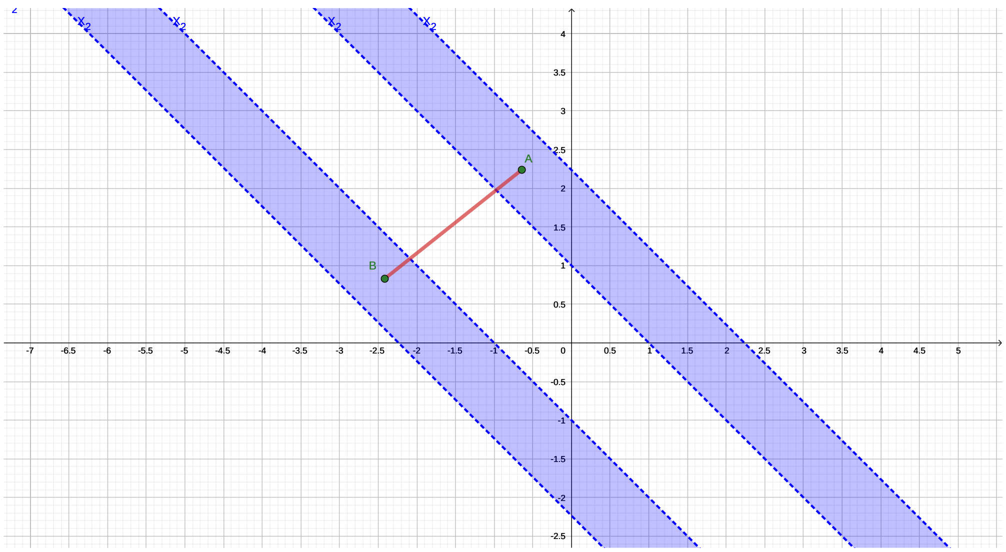

Figure 2 Graph of set X_2

∘ X_2 is convex.

Proof: We can equally write set X_2 as follow:

$$\begin{aligned} X_2 &= \left\{x \in \mathbb{R}^n : \|x - a\|^2 \leq \|x - b\|^2\right\} \\ &= \left\{x \in \mathbb{R}^n : -2a^\top x + \|x\|^2 + \|a\|^2 \leq -2b^\top x + \|x\|^2 + \|b\|^2\right\} \\ &= \left\{x \in \mathbb{R}^n : -2a^\top x + \|a\|^2 \leq -2b^\top x + \|b\|^2\right\} \\ &= \left\{x \in \mathbb{R}^n : (2b^\top - 2a^\top)x \leq \|b\|^2 - \|a\|^2\right\}, \quad a, b \in \mathbb{R}^n, a \neq b \end{aligned} \tag{23}$$

We can see from equation 23 that set X_2 is a half space, so set X_2 is convex.

Subproblem b)

Proof: Let $A, B \in X$, so we have:

$$\begin{aligned} &x^\top A x \geq 0, \quad \forall x \in \mathbb{R}^n \\ &x^\top B x \geq 0, \quad \forall x \in \mathbb{R}^n \\ &\operatorname{tr}(A) = \operatorname{tr}(B) = 1 \end{aligned} \tag{24}$$

Then we choose $\lambda \in [0, 1]$ arbitrary. Let $C = \lambda A + (1 - \lambda)B$. Then we can get:

$$\begin{aligned} x^\top C x &= x^\top \left[\lambda A + (1 - \lambda)B\right] x \\ &= \lambda x^\top A x + (1 - \lambda) x^\top B x \\ &\geq 0 \quad (\forall x \in \mathbb{R}^n) \end{aligned} \tag{25}$$

Besides, we can also get:

$$\begin{aligned}\operatorname{tr}(C) &= \operatorname{tr}(\lambda A + (1-\lambda)B) \\ &= \lambda \operatorname{tr}(A) + (1-\lambda)\operatorname{tr}(B) \\ &= \lambda + (1-\lambda) \\ &= 1\end{aligned} \tag{26}$$

Thus, we can find that C is positive semidefinite and $\operatorname{tr}(C) = 1$, so $C \in X$, so set X is convex.

Subproblem c)

○ For statement 1, the statement is **false**, we can present a counter-example as follow:

$$\begin{aligned}X_1 &= \left\{x \in \mathbb{R}^n : a^\top x \le b\right\} \quad (a \in \mathbb{R}^n, b \in \mathbb{R}) \\ X_2 &= \left\{x \in \mathbb{R}^n : a^\top x \ge b\right\} \\ &= \left\{x \in \mathbb{R}^n : (-a)^\top x \le -b\right\} \quad (a \in \mathbb{R}^n, -b \in \mathbb{R})\end{aligned} \tag{27}$$

Thus, X_1 and X_2 are two half space, so they are both convex and $\subset \mathbb{R}^n$. And we can see the union of X_1 and X_2 is $\mathbb{R}^n$, so the union is convex, which is contradicted with the statement.

○ For statement 2, the statement is **true**.

Proof: Let $x, y \in X$, and choose $\lambda \in [0,1]$. Because f is concave, then we have:

$$\begin{aligned}f[\lambda x + (1-\lambda)y] &\ge \lambda f(x) + (1-\lambda) f(y) \\ &\ge 0\end{aligned} \tag{28}$$

Thus, $\lambda x + (1-\lambda)y \in X$, so X is convex.

Assignment A2.4

In this exercise, convexity properties of different functions are investigated.

a) Verify that the following functions are convex over the specified domain:

- $f : \mathbb{R}_{++} \to \mathbb{R}, f(x) := \sqrt{1 + x^{-2}}$, where $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$
- $f : \mathbb{R}^n \to \mathbb{R}, f(x) := \frac{1}{2}\|Ax - b\|^2 + \mu\|Lx\|^2$, where $A \in \mathbb{R}^{m\times n}, L \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^m$, and $\mu > 0$ are given.
- $f : \mathbb{R}^{n+1} \to \mathbb{R}, f(x, y) := \frac{\lambda}{2}\|x\|^2 + \sum_{i=1}^{m} \max\left\{0, 1 - b_i\left(a_i^\top x + y\right)\right\}$, where $a_i \in \mathbb{R}^n$ and $b_i \in \{-1, 1\}$ are given data points for all $i = 1, \ldots, m$ and $\lambda > 0$ is a parameter.

b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex mapping and set $g(x) := (f(x))^2$. Is the function g convex? Explain your answer and either present a brief verification or a counter-example.

Is the mapping $x \mapsto \frac{1}{2}\left(\|x\|^2 - 1\right)^2$ convex?

Subproblem a)

○ For function 1.

Proof: Because f is twice cont.diff. on the set $\mathbb{R}_{++}$. So we can calculate $f''(x)$ as follow:

$$\begin{aligned}f''(x) &= (f'(x))' \\ &= \left(-\frac{1}{\sqrt{1+x^{-2}}}x^{-3}\right)' \\ &= \frac{3x^{-4} + 2x^{-6}}{(1+x^{-2})^{\frac{3}{2}}} \\ &> 0\end{aligned} \tag{29}$$

Thus, the function f is convex.

◦ For function 2.

Proof: Firstly, we would like to prove that $g(x) = \|x\|^2, x \in \mathbb{R}^n$ is convex. We know that $\|x\|, x \in \mathbb{R}^n$ is convex and $h(x) = x^2, \{x \in \mathbb{R} : x \geq 0\}$ is convex and non-decreasing. So the composition of $h(x)$ and $\|x\|$, say $g(x) = \|x\|^2, x \in \mathbb{R}^n$ is convex.

Then, since we know that $g(x) = \|x\|^2, x \in \mathbb{R}^n$ is convex, and $Ax - b$ and Lx are linear functions, so $\|Ax - b\|^2$ and $\|Lx\|^2$ are both convex.

Finally, $f(x) := \frac{1}{2}\|Ax - b\|^2 + \mu\|Lx\|^2$ is a linear combination of convex set and the coefficients are ≥ 0 $\left(\frac{1}{2} \geq 0, \mu \geq 0\right)$. Thus, according to the lemma **Sum Rule**, the function f is convex.

◦ For function 3.

Proof: From the proof above, we have derived that $\|x\|^2$ is convex. Thus, if we can prove that $\sum_{i=1}^m \max\left\{0, 1 - b_i\left(a_i^\top x + y\right)\right\}$ is convex, then the linear combination $f(x, y) := \frac{\lambda}{2}\|x\|^2 + \sum_{i=1}^m \max\left\{0, 1 - b_i\left(a_i^\top x + y\right)\right\}$would be convex, because the coefficients $\frac{\lambda}{2}$ and 1 are ≥ 0. So we prove $\sum_{i=1}^m \max\left\{0, 1 - b_i\left(a_i^\top x + y\right)\right\}$ is convex as follow.

For $\sum_{i=1}^m \max\left\{0, 1 - b_i\left(a_i^\top x + y\right)\right\}$, we can know that if $\left[1 - b_i\left(a_i^\top x + y\right)\right]$ is convex, the max function would be convex due the the lemma **Taking Maximum** and the Constant equation **0** is convex. Then the sum function $\sum_{i=1}^m \max\left\{0, 1 - b_i\left(a_i^\top x + y\right)\right\}$ would be convex due to the linear combination with non-negative coefficient 1.

Due the the analysis above, all we have to do is to prove that $g(x) = 1 - b_i\left(a_i^\top x + y\right)$ is convex. We prove this as follow:

Let (x_1, y_1) and $(x_2, y_2) \in \mathbb{R}^{n+1}$, and $\lambda \in [0, 1]$ be arbitrary. Then we can get

$$\begin{aligned} g\left(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2\right) &= 1 - b_i\left[a_i^\top(\lambda x_1 + (1-\lambda)x_2) + \lambda y_1 + (1-\lambda)y_2\right] \\ &= \lambda\left[1 - b_i\left(a_i^\top x_1 + y_1\right)\right] + (1-\lambda)\left[1 - b_i\left(a_i^\top x_2 + y_2\right)\right] \\ &= \lambda g(x_1, y_1) + (1-\lambda)\, g(x_2, y_2) \end{aligned} \tag{30}$$

Thus, we have proved that $g(x) = 1 - b_i\left(a_i^\top x + y\right)$ is convex. To sum up, we have proved that the function $f(x, y) := \frac{\lambda}{2}\|x\|^2 + \sum_{i=1}^m \max\left\{0, 1 - b_i\left(a_i^\top x + y\right)\right\}$ is convex.

Subproblem b)

◦ The function g is not convex, we can present a counter-example as follow:

Firstly, we would like to choose $f(x) = x^2 - 1\,(x \in \mathbb{R})$, we can easily derive that $f(x)$ is convex because $f''(x) = 2$ is always greater than 0. So we can get

$$\begin{aligned} g(x) &= (f(x))^2 \\ &= \left(x^2 - 1\right)^2 \end{aligned} \tag{31}$$

By taking the derivative of $g(x)$, we can get:

$$\begin{aligned} g'(x) &= 4x^3 - 4x \\ g''(x) &= 12x^2 - 4 \end{aligned} \tag{32}$$

So we can see, in some cases, $g''(x)$ can be negative, $\left(\text{e.g. } x = \frac{1}{2}, g''(x) = -1\right)$. So The function $g(x)$ is not convex.

◦ The mapping is not convex, we present a counter-example as follow:

We denote the mapping by notation $f(x) = \frac{1}{2}\left(\|x\|^2 - 1\right)^2$ $(x \in \mathbb{R}^n)$. We choose $n = 1$, and we can get:

$$f(x) = \frac{1}{2}\left(x^2 - 1\right)^2 \quad (x \in \mathbb{R}) \tag{33}$$

By taking the derivative of $f(x)$, we can get:

$$\begin{aligned} f'(x) &= 2x^3 - 2x \\ f''(x) &= 6x^2 - 2 \end{aligned} \tag{34}$$

So we can see, in some cases, $f''(x)$ can be negative, $\left(\text{e.g. } x = \frac{1}{2}, f''(x) = -\frac{1}{2}\right)$. So The mapping $f(x)$ is not convex.

Assignment A2.5

We consider the parametrized optimization problem

$$\min_x f_\beta(x) := \frac{1}{2}\|x-b\|^2 + \frac{\beta}{2}\left(\sum_{i=1}^n x_i\right)^2, \quad x \in \mathbb{R}^n \tag{35}$$

where $b \in \mathbb{R}^n$ is given and $\beta \geq 0$ is a parameter.

a) Calculate the gradient and Hessian of f_β.

b) Show that the mapping f_β is strongly convex for all $\beta \geq 0$.

c) Show that f_β has a unique stationary point x^*_β and compute it explicitly. Determine whether x^*_β is a local minimizer, a local maximizer, or a saddle point of problem (35).

d) For $\beta \to \infty$, the solutions x^*_β converge to a point x^*. Calculate the limit $x^* = \lim_{\beta\to\infty} x^*_\beta$ explicitly and show that x^* satisfies the constraint $\mathbb{1}^\top x^* = \sum_{i=1}^n x^*_i = 0$.

Subproblem a)
We can rewrite f_β as follow:

$$f_\beta(x) := \frac{1}{2}\sum_{i=1}^n (x_i - b_i)^2 + \frac{\beta}{2}\left(\sum_{i=1}^n x_i\right)^2, \quad x \in \mathbb{R}^n \tag{36}$$

Thus, we can derive the gradient of f_β:

$$\nabla f_\beta(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \begin{pmatrix} x_1 - b_1 + \beta \sum_{i=1}^n x_i \\ x_2 - b_2 + \beta \sum_{i=1}^n x_i \\ \dots \\ x_n - b_n + \beta \sum_{i=1}^n x_i \end{pmatrix} \tag{37}$$

We can derive the Hessian of f_β:

$$\nabla^2 f_\beta(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} = \begin{pmatrix} 1+\beta & \beta & \dots & \beta \\ \beta & 1+\beta & \dots & \beta \\ \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \dots & 1+\beta \end{pmatrix} \tag{38}$$

Subproblem b)

Firstly, we would prove that the Hessian $\nabla^2 f_\beta(x)$ is positive definite as follow. We can calculate the determinant of the i^{th} leading principal minors of Hessian $\nabla^2 f_\beta(x)$:

$$\begin{aligned}
\begin{vmatrix} 1+\beta & \beta & \dots & \beta \\ \beta & 1+\beta & \dots & \beta \\ \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \dots & 1+\beta \end{vmatrix}_i &= \begin{vmatrix} 1 & \beta & \beta & \dots & \beta \\ 0 & 1+\beta & \beta & \dots & \beta \\ 0 & \beta & 1+\beta & \dots & \beta \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \beta & \beta & \dots & 1+\beta \end{vmatrix}_{i+1} = \begin{vmatrix} 1 & \beta & \beta & \dots & \beta \\ -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1 \end{vmatrix}_{i+1} \\
&= \begin{vmatrix} 1+n\beta & \beta & \beta & \dots & \beta \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix}_{i+1} = (1+i\beta) \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{vmatrix}_i = 1+i\beta > 0
\end{aligned} \tag{39}$$

From euation 39, we can find that all the determinants of the leading principal minors of Hessian $\nabla^2 f_\beta(x)$ are positive, so we can derive that f_β is positive definite and the minimal eigenvalue of $\nabla^2 f_\beta(x)$, say λ_{min} is positive. Thus, according to the lemma, we have:

$$x^\top \nabla^2 f_\beta(x) x \geq \lambda_{min} \|x\|^2 \tag{40}$$

Thus, we can say f_β is λ_{min}-strongly convex.

Subproblem c)

In order to derive the stationary point, we set the gradient $\nabla f_\beta(x) = 0$.

$$\begin{aligned}
\nabla f_\beta(x) &= \begin{pmatrix} x_1 - b_1 + \beta \sum_{i=1}^n x_i \\ x_2 - b_2 + \beta \sum_{i=1}^n x_i \\ \dots \\ x_n - b_n + \beta \sum_{i=1}^n x_i \end{pmatrix} = 0 \\
&\implies \begin{cases} x_1 - b_1 + \beta \sum_{i=1}^n x_i = 0 \\ x_2 - b_2 + \beta \sum_{i=1}^n x_i = 0 \\ \dots \\ x_n - b_n + \beta \sum_{i=1}^n x_i = 0 \end{cases}
\end{aligned} \tag{41}$$

Sum up all the equations, we have:

$$\begin{aligned}
&\sum_{i=1}^n x_i - \sum_{i=1}^n b_i + n\beta \sum_{i=1}^n x_i = 0 \\
&\implies (n\beta + 1) \sum_{i=1}^n x_i = \sum_{i=1}^n b_i \\
&\implies \sum_{i=1}^n x_i = \frac{1}{n\beta + 1} \sum_{i=1}^n b_i
\end{aligned} \tag{42}$$

Put the equation 42 back into equation 41, we can get:

$$
\begin{aligned}
&\begin{cases}
x_1 - b_1 + \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i = 0 \\
x_2 - b_2 + \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i = 0 \\
\qquad \cdots \\
x_n - b_n + \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i = 0
\end{cases} \\
\Longrightarrow &\begin{cases}
x_1 = b_1 - \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i \\
x_2 = b_2 - \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i \\
\qquad \cdots \\
x_n = b_n - \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i
\end{cases}
\end{aligned} \tag{43}
$$

Thus, we have gotten the stationary point x_β^*. Because we know that the Hessian $\nabla^2 f_\beta(x)$ is always positive definite, so the point x_β^* is local minimizer.

Subproblem d)
By taking the limit of x_β^*, we can get x^* as follow:

$$
\begin{cases}
x_1^* = \lim_{\beta\to\infty} x_1 = \lim_{\beta\to\infty} \left[b_1 - \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i\right] = b_1 - \frac{1}{n}\sum_{i=1}^{n} b_i \\
x_2^* = \lim_{\beta\to\infty} x_2 = \lim_{\beta\to\infty} \left[b_2 - \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i\right] = b_2 - \frac{1}{n}\sum_{i=1}^{n} b_i \\
\qquad \cdots \\
x_n^* = \lim_{\beta\to\infty} x_n = \lim_{\beta\to\infty} \left[b_n - \frac{\beta}{n\beta+1}\sum_{i=1}^{n} b_i\right] = b_n - \frac{1}{n}\sum_{i=1}^{n} b_i
\end{cases} \tag{44}
$$

Thus, we can get:

$$
\begin{aligned}
\mathbb{1}^\top x^* &= \sum_{i=1}^{n} x_i^* \\
&= \sum_{i=1}^{n} b_i - n\frac{1}{n}\sum_{i=1}^{n} b_i \\
&= 0
\end{aligned} \tag{45}
$$