1 [Q1]

Last class, we mainly talked about unconstrained optimization, especailly the convex set and convex function. In convex set part, we talked the definition of convex set and gave some concrete examples. Like convex cone, A convex cone is a cone that is convex and a proper cone is a convex cone that does not contain entire lines and has nonempty interior. In convex function part, we talked about the definition and more examples of convex function. Besides, operations that preserve convexity were introduced too. I learned how to distinguish a strictly convex from convex functions. There will be no linear region in srtictly convex functions.

2 [Q2]

done

3 [Q3]

Assume we have a concave function $f(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}$ which is concave if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, then, it must satisfy,

$$f\left(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) \geq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}), \quad \text{for all } 0 \leq \theta \leq 1$$

Now consider $g(\boldsymbol{x}) = -f(\boldsymbol{x})$, according to the definition, we have,

$$-f\left(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) \leq \theta\left[-f(\boldsymbol{x})\right] + (1-\theta)\left[-f(\boldsymbol{y})\right], \quad \text{for all } 0 \leq \theta \leq 1$$

Apparently, it satisifes the definition of convex function, and $-f(\boldsymbol{x})$ is convex and vice versa. Hence, we can conclude that,

$$f(\boldsymbol{x}) \text{ is convex} \Longleftrightarrow -f(\boldsymbol{x}) \text{ is convex}$$

4 [Q4]

4.1 (a)

$$\begin{aligned}
f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) &= \|\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\| && (1)\\
&\leq \|\theta \boldsymbol{x}\| + \|(1-\theta)\boldsymbol{y}\| && (2)\\
&= |\theta|\, \|\boldsymbol{x}\| + |(1-\theta)|\, |\boldsymbol{y}\| && (3)\\
&= \theta |\boldsymbol{x}\| + (1-\theta)\|\boldsymbol{y}\| && (4)\\
&= \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) && (5)
\end{aligned}$$

Note that (1) to (2) obeys the triangle inequality, and since we have $0 \leq \theta \leq 1$, (3) to (4) holds. Hence, $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ is convex function.

4.2 (b)

No norm can be strictly convex. The equality holds as long as $\boldsymbol{x} = k\boldsymbol{y}$, where k is a positive constant.

$$\begin{aligned} \|\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\| &= \|\theta\boldsymbol{x} + (1-\theta)k\boldsymbol{x}\| && (6) \\ &= \theta\|\boldsymbol{x}\| + (1-\theta)k\|\boldsymbol{x}\| && (7) \\ &= \theta f(\boldsymbol{x}) + (1-\theta)f(\boldsymbol{y}) && (8) \end{aligned}$$

5 [Q5]

5.1 (a)

Since $f(\boldsymbol{x})$ is convex, then we have,

$$f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta)f(\boldsymbol{y}), \quad \text{for all } 0 \leq \theta \leq 1$$

Set $g(\boldsymbol{x}) = f(\boldsymbol{x}) - \alpha$, then we can have,

$$\begin{aligned} g(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) &= f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) - \alpha && (9) \\ &\leq \theta f(\boldsymbol{x}) + (1-\theta)f(\boldsymbol{y}) - \alpha && (10) \\ &= \theta\left[f(\boldsymbol{x}) - \alpha\right] + (1-\theta)\left[f(\boldsymbol{y}) - \alpha\right] && (11) \\ &= \theta g(\boldsymbol{x}) + (1-\theta)g(\boldsymbol{y}) && (12) \end{aligned}$$

Note that (9) to (10) is convexity of $f(\boldsymbol{x})$.
If we find 2 points $\boldsymbol{x}, \boldsymbol{y}$ which satisifes $g(\boldsymbol{x}) \leq 0, g(\boldsymbol{y}) \leq 0$, then

$$g(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta g(\boldsymbol{x}) + (1-\theta)g(\boldsymbol{y}) \leq 0$$

which means $\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y} \in S_\alpha$
Hence for all $\alpha \in \mathbb{R}$, set $S_\alpha = \{\boldsymbol{x} : f(\boldsymbol{x}) \leq \alpha\}$ is convex.

5.2 (b)

Set the optimizer as $\boldsymbol{x}^\star$ which satisifes $\forall\boldsymbol{x} \in dom f, f(\boldsymbol{x}^\star) \leq f(x)$ Define the optimizer set as $S = \{\boldsymbol{x}^\star | \forall x, f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})\}$.
Let's define 2 points $a, b \in$ S such that,

$$\forall\boldsymbol{x}, f(\boldsymbol{a}) \leq f(\boldsymbol{x}) \text{ and } f(\boldsymbol{b}) \leq f(\boldsymbol{x})$$

Since $f(\boldsymbol{x})$ is convex, we have the third point $\theta\boldsymbol{a} + (1-\theta)\boldsymbol{b}$ such that,

$$f(\theta\boldsymbol{a} + (1-\theta)\boldsymbol{b}) \leq \theta f(\boldsymbol{a}) + (1-\theta)f(\boldsymbol{b}), \text{ for all } 0 \leq \theta \leq 1$$

Combine the 3 inequalities, we can have,

$$\begin{aligned} f(\theta\boldsymbol{a} + (1-\theta)\boldsymbol{b}) &\leq \theta f(\boldsymbol{a}) + (1-\theta)f(\boldsymbol{b}) && (13) \\ &\leq \theta f(\boldsymbol{x}) + (1-\theta)f(\boldsymbol{x}) && (14) \\ &= f(\boldsymbol{x}) && (15) \end{aligned}$$

Hence, we can conclude that the third point $\theta\boldsymbol{a} + (1-\theta)\boldsymbol{b} \in$ set S. and this optimizer set is convex.

5.3 (c)

In [Q4], we have proved that $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ is convex. Thus, using the conclusion from [Q5](a), it's obvious that set $\mathcal{B} = \{\boldsymbol{x} : |\boldsymbol{x}| \leq 1\}$ must be convex.

5.4 (d)

It's not convex. There's the counterexample, let S_α

$$S_\alpha = \{x : -e^x \leq \alpha\}$$

There's no doubt S_α is convex set, assume $x, y \in S_\alpha$, and we have

$$-e^x \leq \alpha, \quad -e^y \leq \alpha$$

And there's the third point $\theta x + (1-\theta)y$ which satisfies,

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

However, $f(x) = -e^x$ is not convex. Hence, the statement if the question doesn't hold.

6 [Q6]

6.1 (a)

$$\begin{aligned} f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) &= \max\{f_1(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}), f_2(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y})\} && (16) \\ &\leq \max\{\theta f_1(\boldsymbol{x}) + (1-\theta)f_1(\boldsymbol{y}), \theta f_2\boldsymbol{x} + (1-\theta)f_2(\boldsymbol{y})\} && (17) \\ &\leq \theta\max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\} + (1-\theta)\max\{f_1(\boldsymbol{y}), f_2(\boldsymbol{y})\} && (18) \end{aligned}$$

where, (16) to (17) is due to $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ is convex.
Hence, $f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ is convex.

6.2 (b)

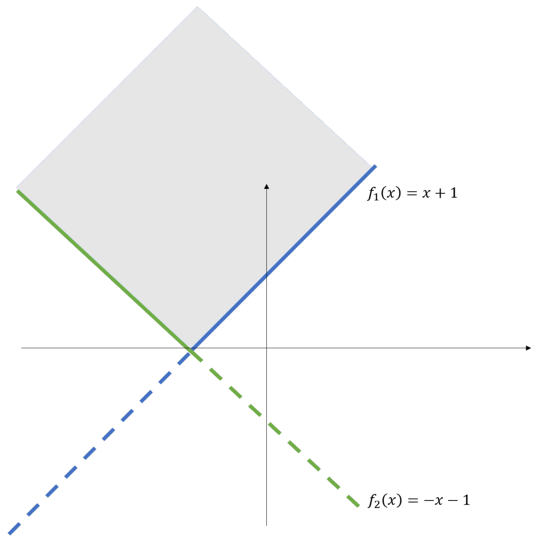

Figure 1: Illustration of $f(x) = \max\{f_1(x), f_2(x)\}$

In Figure 1, I set $f_1(x) = x + 1$ and $f_2(x) = -x - 1$, and the shade region is $f(x) = \max\{f_1(x), f_2(x)\}$. Clearly, the shaded region, namely $f(x)$ is convex.

6.3 (c)

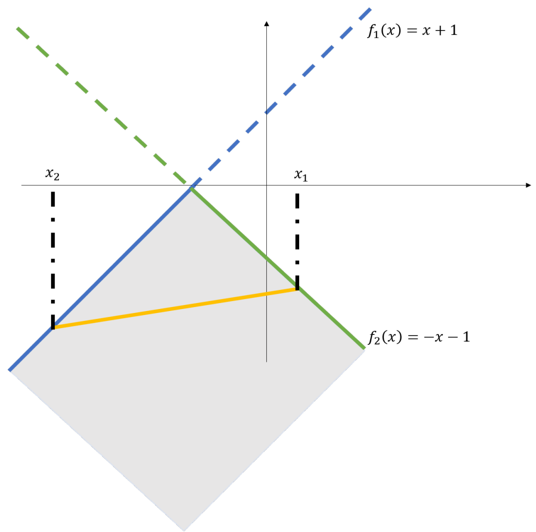

Figure 2: Illustration of $f(x) = \min\{f_1(x), f_2(x)\}$

Here the yellow line intersected the shaded region with 2 points x_1, x_2, where

$$f(\theta x_1 + (1-\theta)x_2) \geq \theta f(x_1) + (1-\theta) f(x_2)$$

Apparently, this is a concave function. It's not a convex function.

7 [Q7]

7.1 (a)

According to the Linear Algebra, we know,

$$\lambda_{min}(\boldsymbol{X}) = \min_{\|y\|=1} y^T \boldsymbol{X} y$$

Since $0 \leq \theta \leq 1, y^T \boldsymbol{X_1} y \geq 0$ and $y^T \boldsymbol{X_2} y \geq 0$, we can have,

$$\begin{aligned}
\lambda_{min}\left[\theta \boldsymbol{X_1} + (1-\theta)\boldsymbol{X_2}\right] &= \min_{\|y\|=1} y^T \left[\theta \boldsymbol{X_1} + (1-\theta)\boldsymbol{X_2}\right] y && (19)\\
&= \min_{\|y\|=1} \left[\theta y^T \boldsymbol{X_1} y + (1-\theta) y^T \boldsymbol{X_2} y\right] && (20)\\
&\geq \theta \min_{\|y\|=1} y^T \boldsymbol{X_1} y + (1-\theta) \min_{\|y\|=1} y^T \boldsymbol{X_2} y && (21)\\
&= \theta \lambda_{min}(\boldsymbol{X_1}) + (1-\theta)\lambda_{min}(\boldsymbol{X_2}) && (22)
\end{aligned}$$

where, (20) to (21) is due to **min** is a concave function.

7.2 (b)

Let's denote λ_{max} is the largest eigenvalue of of symmetric matrix X, then we can have,

$$y^T \boldsymbol{X} y \leq \lambda_{max} \|y\|$$

And the set can be described as $\mathbb{S}_+^N = \left\{\boldsymbol{X} | y^T \boldsymbol{X} y \leq \lambda_{max} \|y\|\right\}$ Since $f(\boldsymbol{X}) = y^T \boldsymbol{X} y$ is an affine function which is convex, we can deduce from the conclusion from [Q5](a) that $\mathbb{S}_+^N$ is a convex set.

The following part proves that $f(\boldsymbol{X}) = y^T \boldsymbol{X} y$ is an affine function.

$$\begin{aligned}
f(\boldsymbol{X}) &= y^T \boldsymbol{X} y\\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} y_i x_{ij} y_j
\end{aligned}$$

where, y_i indicates the ith element of vector y and x_{ij} indicates the ith row and jth coloumn of matrix $\boldsymbol{X}$.

Apparently, it's a linear combination of x_{ij} with its cofficient $\|y\|$. Hence it's an affine function and also a convex function.

7.3 (c)

I suppose that as long as these convex functions are affine functions like,

$$f_m(\boldsymbol{X}) = y_m^T \boldsymbol{X} y_m$$

then it satisfies the requirements. Hence, the set of convex functions must be,

$$S_f = \left\{f_m(\boldsymbol{X}) | f_m(\boldsymbol{X}) = y_m^T \boldsymbol{X} y_m)\right\}$$

And b_m set is

$$S_b = \{b_m | b_m = \lambda_{\max} \|y_m\|\}$$

8 [Q8]

8.1 (a)

$$f'(x) = 2ax + b$$
$$f''(x) = 2a$$

8.2 (b)

$$f'(x) = \sum_{m=1}^{M} -\frac{a_m}{1 + e^{a_m x}}$$
$$f''(x) = \sum_{m=1}^{M} \frac{a_m^2 e^{a_m x}}{(1 + e^{amx})^2}$$

9 [Q9]

9.1 (a)

$$\nabla f(x) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$$
$$\nabla^2 f(x) = 2\boldsymbol{A}$$

9.2 (b)

$$\nabla f(x) = \begin{bmatrix} \sum_{m=1}^{M} \frac{-a_{m1}}{1+e^{a_m^T \boldsymbol{x}}} \\ \sum_{m=1}^{M} \frac{-a_{m2}}{1+e^{a_m^T \boldsymbol{x}}} \\ \vdots \\ \sum_{m=1}^{M} \frac{-a_{mN}}{1+e^{a_m^T \boldsymbol{x}}} \end{bmatrix}$$

where, a_{mi} means the ith element of column vector a_m

$$\nabla^2 f(x) = \begin{bmatrix} \sum_{m=1}^{M} \frac{a_{m1}^2 e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} & \sum_{m=1}^{M} \frac{a_{m1} a_{m2} e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} & \cdots & \sum_{m=1}^{M} \frac{a_{m1} a_{mN} e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} \\ \sum_{m=1}^{M} \frac{a_{m2} a_{m1} e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} & \sum_{m=1}^{M} \frac{a_{m2}^2 e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} & \cdots & \sum_{m=1}^{M} \frac{a_{m2} a_{mN} e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{m=1}^{M} \frac{a_{mN} a_{m1} e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} & \sum_{m=1}^{M} \frac{a_{mN} a_{m2} e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} & \cdots & \sum_{m=1}^{M} \frac{a_{mN}^2 e^{a_m^T \boldsymbol{x}}}{(1+e^{a_m^T \boldsymbol{x}})^2} \end{bmatrix}$$

where, a_{mi} means the ith element of column vector a_m

10 [Q10]

10.1 (a)

convex

10.2 (b)

convex

10.3 (c)

it's not convex and concave

10.4 (d)

convex

10.5 (e)

it's not convex and concave

10.6 (f)

convex