Reference sol for hw4

1. (i) it is not a convex set.

counterexample: given two points $\vec{x} = (\cos\theta, \sin\theta) \in C_1$, $\vec{y} = (1, 0) \in C_1$ when $n=2$,

then $\left\| \frac{\vec{x}+\vec{y}}{2} \right\| = \left\| \left(\frac{\cos\theta + 1}{2}, \frac{\sin\theta}{2} \right) \right\|$

$$= \sqrt{\frac{\cos^2\theta + \sin^2\theta + 1}{4}}$$

$$= \sqrt{\tfrac{1}{2}} < 1$$

that is $\frac{\vec{x}+\vec{y}}{2} \notin C_1$.

(ii) it is a convex set.

proof: let $f(\vec{x}) = \max_{i=1,2,\cdots n} x_i$, $\vec{x} \in \mathbb{R}^n$. then

$\forall \vec{x}, \vec{y} \in \mathbb{R}^n$, $\lambda \in [0,1]$. we have

$$f(\lambda\vec{x} + (1-\lambda)\vec{y}) = \max_{i=1,2,\cdots n} \lambda x_i + (1-\lambda) y_i$$

$$= \lambda \max_{i=1,2,\cdots n} x_i + (1-\lambda) \max_{i=1,2,\cdots,n} y_i$$

$$= \lambda f(\vec{x}) + (1-\lambda) f(\vec{y})$$

Thus $f(\vec{x})$ is a convex function.

On the other hand $C_2 = \mathrm{lev}(f, 1)$, thus C_2 is convex

(iii) it is not a convex set.

counterexample: give two points $\vec{x} = (1, 2) \in C_3$, $\vec{y} = (2, 1) \in C_3$ when $n=2$.

then $\frac{\vec{x}+\vec{y}}{2} = \left(\frac{3}{2}, \frac{3}{2}\right) \notin C_3$

(iv) it is a convex set.

proof: $\vec{x} \in C_4 \Leftrightarrow \log\left(\prod_{i=1}^{n} x_i\right) \geq \log(1) \Leftrightarrow \sum_{i=1}^{n} \log(x_i) \geq 0$

$\Leftrightarrow -\sum_{i=1}^{n} \log(x_i) \leq 0$. Since $-\log(x_i)$ is convex $\forall x_i$, thus.

$f(x) = -\sum_{i=1}^{n} \log(x_i)$ is a convex funition. thus. $C_4 = \mathrm{lev}(f(x), 0)$ is convex (v). it is not a convex set.

counterexample: Given $C_5 = [0 \quad 1]$ and $C_6 = [3, 4]$. then.

$x = 1 \in C_5 \cup C_6$, $y = 3 \in C_5 \cup C_6$. but.

$\frac{x+y}{2} = 2 \notin C_5 \cup C_6$.

2. First, we have

$$A = \begin{bmatrix} 0 & -4 & 1 & 0 \\ 2 & -2 & 0 & 1 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$$

it is obvious that column one and column two are linearly independent; (case 1)

column one and column three are linearly independent; (case 2)

column two and column three are linearly independent; (case 3)

column two and column four are linearly independent; (case 4)

column three and column four are linearly independent; (case 5)

let us consider each case and find the corresponding solution to $A\vec{x} = \vec{b}$.

Case 1:

$$\left[\begin{array}{cccc|c} 0 & -4 & 1 & 0 & 6 \\ 2 & -2 & 0 & 1 & 1 \end{array}\right] \xrightarrow{\text{row}1 \rightleftarrows \text{row}2} \left[\begin{array}{cccc|c} 2 & -2 & 0 & 1 & 1 \\ 0 & -4 & 1 & 0 & 6 \end{array}\right]$$

$$\xrightarrow{\text{row}1 - \frac{1}{2}\cdot\text{row}2} \left[\begin{array}{cccc|c} 2 & 0 & -\frac{1}{2} & 1 & -2 \\ 0 & -4 & 1 & 0 & 6 \end{array}\right] \xrightarrow[-\frac{1}{4}\cdot\text{row}2]{\frac{1}{2}\cdot\text{row}1} \left[\begin{array}{cccc|c} 1 & 0 & -\frac{1}{4} & \frac{1}{2} & -1 \\ 0 & 1 & -\frac{1}{4} & 0 & -\frac{3}{2} \end{array}\right]$$

Thus we have the basic solution of $A\vec{x} = \vec{b}$ for case 1: $\vec{x} = \begin{bmatrix} -1 \\ -\frac{3}{2} \\ 0 \\ 0 \end{bmatrix}$

Case 2:

$$\left[\begin{array}{cccc|c} 0 & -4 & 1 & 0 & 6 \\ 2 & -2 & 0 & 1 & 1 \end{array}\right] \xrightarrow{\frac{1}{2}\cdot\text{row}2} \left[\begin{array}{cccc|c} 0 & -4 & 1 & 0 & 6 \\ 1 & -1 & 0 & \frac{1}{2} & \frac{1}{2} \end{array}\right]$$

Thus we have the basic solution of $A\vec{x} = \vec{b}$ for case 2: $\vec{x} = [\frac{1}{2} \quad 0 \quad 6 \quad 0]^T$

Case 3:

$$\left[\begin{array}{cccc|c} 0 & -4 & 1 & 0 & 6 \\ 2 & -2 & 0 & 1 & 1 \end{array}\right] \xrightarrow{\text{row}1 - 2\cdot\text{row}2} \left[\begin{array}{cccc|c} 0 & 0 & 1 & -2 & 4 \\ 2 & -2 & 0 & 1 & 1 \end{array}\right]$$

$$\xrightarrow{-\frac{1}{2}\text{row}2} \left[\begin{array}{cccc|c} 0 & 0 & 1 & -2 & 4 \\ -1 & 1 & 0 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right]$$

Thus we have the basic solution of $A\vec{x} = \vec{b}$ for case 3: $\vec{x} = [0, -\frac{1}{2}, 4, 0]^T$

Case 4: $\left[\begin{array}{cccc|c} 0 & -4 & 1 & 0 & 6 \\ 2 & -2 & 0 & 1 & 1 \end{array}\right] \xrightarrow{row2 - \frac{1}{2} row1} \left[\begin{array}{cccc|c} 0 & -4 & 1 & 0 & 6 \\ 0 & 0 & \frac{1}{2} & 1 & -2 \end{array}\right]$

$\xrightarrow{-\frac{1}{4} row1} \left[\begin{array}{cccc|c} 0 & 1 & -\frac{1}{4} & 0 & -\frac{3}{2} \\ 0 & 0 & \frac{1}{2} & 1 & -2 \end{array}\right]$

Thus we have the basic solution of $A\vec{x} = \vec{b}$ for case 4: $\vec{x} = [0, -\frac{3}{2}, 0, -2]^T$.

Case 5: $\left[\begin{array}{cccc|c} 0 & -4 & 1 & 0 & 6 \\ 2 & -2 & 0 & 1 & 1 \end{array}\right]$

Thus we have the basic solution of $A\vec{x} = \vec{b}$ for case 5: $\vec{x} = [0, 0, 6, 1]^T$

Therefore, the basic feasible solutions to the system are

$$\vec{x} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ 6 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 0 \\ 0 \\ 6 \\ 1 \end{bmatrix}$$

3. proof: let $g(\vec{x}) = f(\vec{x}) - f(\vec{0})$, if we can show that $g(\vec{x})$ is linear, then we can conclude that $f(\vec{x}) = g(\vec{x}) + f(\vec{0})$ is affine.

Let us first claim that $g(\vec{x})$ is linear provided that $f(\vec{x})$ is both convex and concave.

To show $g(\vec{x})$ is linear, it is equivalent to show that

(a) $g(\alpha\vec{x}) = \alpha g(\vec{x})$, $\forall \alpha \in \mathbb{R}$

(b) $g(\vec{x}+\vec{y}) = g(\vec{x}) + g(\vec{y})$, $\forall \vec{x}, \vec{y} \in \mathbb{R}^n$

First, note that $g(\vec{x})$ is both convex and concave provided that $f(\vec{x})$ is both convex and concave

(i) $\alpha \in [0,1]$, then $g(\alpha\vec{x}) = g(\alpha\vec{x} + (1-\alpha)\vec{0}) = \alpha g(\vec{x}) + (1-\alpha) g(\vec{0})$

$= \alpha g(\vec{x}) + 0 = \alpha g(\vec{x})$

(ii) $\alpha > 1$, then $g(\vec{x}) = g(\frac{1}{\alpha}\alpha\vec{x}) = g(\frac{1}{\alpha}(\alpha\vec{x}) + (1-\frac{1}{\alpha})\vec{0}) = \frac{1}{\alpha} g(\alpha\vec{x})$

i.e. $g(\alpha\vec{x}) = \alpha g(\vec{x})$

Note that combine (i), and (ii), and we conclude that

$$g(\vec{x}+\vec{y}) = g(2\frac{\vec{x}+\vec{y}}{2}) = 2g(\frac{1}{2}\vec{x}+\frac{1}{2}\vec{y}) = 2g(\frac{1}{2}\vec{x}) + 2g(\frac{1}{2}\vec{y}) = g(\vec{x}) + g(\vec{y})$$

Thus we further have

$$0 = g(\vec{0}) = g(\vec{x} - \vec{x}) = g(\vec{x}) + g(-\vec{x}), \text{ that is, } g(-\vec{x}) = -g(\vec{x}).$$

(iii) $\alpha < 0$. we can get

$$g(\alpha\vec{x}) = -g(-\alpha\vec{x}) = \alpha g(\vec{x})$$

Therefore, we have $g(\vec{x})$ is linear. i.e. $g(\vec{x}) = \vec{a}^T\vec{x}$, $\vec{a} \in \mathbb{R}^n$, then let $\vec{b} = f(\vec{0})$, we now get

$$f(\vec{x}) = \vec{a}^T\vec{x} + \vec{b} \quad \text{is an affine function.}$$

4. proof: argue by contradiction. Suppose that there exists $\vec{z} \in \{\vec{x} \in \mathbb{R}^2 : \|\vec{x}\|_2 = 1\}$, but $\vec{z} \notin \text{ext}(S)$.

Then we can find $\vec{x}, \vec{y} \in S$, $\vec{x} \neq \vec{y}$ s.t. $\vec{z} = \lambda\vec{x} + (1-\lambda)\vec{y}$, $\lambda \in (0,1)$

Without loss of generality, we can assume $\vec{x} = (r_1, 0)$, $0 \le r_1 \le 1$, $\vec{y} = (r_2\cos\theta, r_2\sin\theta)$ $0 \le r_2 \le 1$, $\theta \in [0, 2\pi]$ (Note: if $\vec{x}$ is not on the x-axis, we can always shift and rotate the coordinate such that $\vec{x}$ locates at the x-axis)

Since

$$\begin{aligned}\|\vec{z}\|_2 &= \|\lambda\vec{x} + (1-\lambda)\vec{y}\|_2 \\ &= \|(\lambda r_1 + (1-\lambda)r_2\cos\theta,\ (1-\lambda)r_2\sin\theta\|_2 \\ &= \sqrt{(\lambda r_1 + (1-\lambda)r_2\cos\theta)^2 + ((1-\lambda)r_2\sin\theta)^2} \\ &= \sqrt{\lambda^2 r_1^2 + (1-\lambda)^2 r_2^2 + 2\lambda(1-\lambda)r_1 r_2\cos\theta}\end{aligned}$$

$\cos\theta = 1$, $\|\vec{z}\|_2 = \lambda r_1 + (1-\lambda)r_2 < 1$ (since $\vec{x} \neq \vec{y}$)

$\cos\theta < 1$, $\|\vec{z}\|_2 < \lambda r_1 + (1-\lambda)r_2 \le 1$

thus $\vec{z} \in \text{int}(S)$ which contradicts with $\vec{z} \in \text{bdry}(S)$.

Therefore, we get that $\text{ext}(S) = \{\vec{x} \in \mathbb{R}^2 : \|\vec{x}\|_2 = 1\}$

5. (a) proof one: first note that $f(\vec{x})$ is twice continuous differentiable, then we can use the second-order characterization to show the convexity of $f(\vec{x})$.

Since $\nabla f(\vec{x}) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right]^T$ and

$$\frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}}$$

then we have

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{cases} -\dfrac{e^{x_i + x_j}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2}, & j \neq i \\[2ex] \dfrac{e^{x_i}\left(\sum_{k=1}^{n} e^{x_k}\right) - e^{2x_i}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2}, & j = i \end{cases}$$

Thus we have

$$\nabla^2 f(\vec{x}) = \mathrm{diag}\left(\left[\frac{e^{x_1}}{\sum_{k=1}^{n} e^{x_k}}, \frac{e^{x_2}}{\sum_{i=1}^{n} e^{x_k}}, \cdots, \frac{e^{x_n}}{\sum_{i=1}^{n} e^{x_k}}\right]\right) - \begin{bmatrix} \frac{e^{x_1}}{\sum_{k=1}^{n} e^{x_k}} \\ \vdots \\ \frac{e^{x_n}}{\sum_{k=1}^{n} e^{x_k}} \end{bmatrix} \underbrace{\left[\frac{e^{x_1}}{\sum_{i=1}^{n} e^{x_k}} \cdots\cdots \frac{e^{x_n}}{\sum_{i=1}^{n} e^{x_k}}\right]}_{\vec{z}^T}$$

i.e. $\nabla^2 f(\vec{x}) = \mathrm{diag}(\vec{z}^T) - \vec{z}\vec{z}^T$

then $\forall \vec{v} \in \mathbb{R}^n$, we have

$$\begin{aligned} \vec{v}^T \nabla^2 f(\vec{x}) \vec{v} &= \vec{v}^T \left[\mathrm{diag}(\vec{z}^T) - \vec{z}\vec{z}^T\right] \vec{v} \\ &= \frac{\sum_{i=1}^{n} v_i^2 e^{x_i}}{\sum_{k=1}^{n} e^{x_k}} - \frac{\left(\sum_{i=1}^{n} e^{x_i} v_i\right)^2}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2} \\ &= \frac{\sum_{i=1}^{n} e^{x_i} \sum_{i=1}^{n} v_i^2 e^{x_i} - \left(\sum_{i=1}^{n} v_i e^{x_i}\right)^2}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2} \end{aligned}$$

let $\vec{u} = [\, e^{x_1/2} \quad e^{x_2/2} \quad \cdots \quad e^{x_n/2} \,]^T$

$\vec{v} = [\, v_1 e^{x_1/2} \quad v_2 e^{x_2/2} \quad \cdots \quad e^{x_n/2} \,]^T$

Then from cauchy schwaz, we have

$$\vec{u}^T \vec{v} \leq \|\vec{u}\| \, \|\vec{v}\|$$

that is.

$$\sum_{i=1}^{n} v_i e^{x_i} \leq \sqrt{\sum_{i=1}^{n} e^{x_i}} \sqrt{\sum_{i=1}^{n} v_i^2 e^{x_i}}$$

thus

$$\left(\sum_{i=1}^{n} v_i e^{x_i}\right)^2 \leq \left(\sum_{i=1}^{n} e^{x_i}\right)\left(\sum_{i=1}^{n} v_i^2 e^{x_i}\right)$$

Therefore. $\vec{v}^T \nabla^2 f(\vec{x}) \vec{v} \geq 0$, i.e. $\nabla^2 f(\vec{x})$ is positive semi-definite.

Thus $f(\vec{x})$ is indeed a convex function.

proof two: hint: you can also use the definition to verify that

$$f(\lambda \vec{x} + (1-\lambda)\vec{y}) \leq \lambda f(\vec{x}) + (1-\lambda) f(\vec{y}), \quad \forall \lambda \in [0,1]$$

Try to construct $f(\vec{x})$ and $f(\vec{y})$ from $f(\lambda\vec{x} + (1-\lambda)\vec{y})$

a useful tool is Hölder inequality. (we didn't provide this in our lecture, but the problem will be simple if you know this inequality).

Anyway, the Cauchy-schwarz inequality is enough for this problem.

(b) proof: In our lecture, we show that $g(f(\vec{x}))$ is convex if $g(t)$ is a non-descreasing convex function and $f(\vec{x})$ is a convex function.

Similary, we can show that $g(f(\vec{x}))$ is concave if $g(t)$ is a non-descreasing concave function and $f(\vec{x})$ is a concave function.

See the proof here: let $h(\vec{x}) = g(f(\vec{x}))$, then

$\forall \lambda \in [0,1]$, $h(\lambda\vec{x} + (1-\lambda)\vec{y}) = g(f(\lambda\vec{x} + (1-\lambda)\vec{y}))$

$$\geq g(\lambda f(\vec{x}) + (1-\lambda) f(\vec{y})) \geq \lambda h(\vec{x}) + (1-\lambda) h(\vec{y})$$

Thus. $h(\vec{x})$ is concave.

Now, back to our problem, we have

$g_i(\vec{x})$ is concave., $\ln(t)$ is a nondescreasing concave function.

Thus $\ln g_i(\vec{x})$ is a concave function, then $-\mu \ln g_i(\vec{x})$ is a convex function. Combine with the fact that $f(\vec{x})$ is a convex function and summation preserves the convexity. Then

$\beta(\vec{x}) = f(\vec{x}) - \mu \sum_{i=1}^{m} \ln g_i(\vec{x})$ is convex.

obviously, $S = \{ \vec{x} : g_i(\vec{x}) > 0,\ i=1,\cdots,m \}$ is convex since.

$\forall\ \vec{x}, \vec{y} \in S,\ \lambda \in [0,1]$, we have

$$g_i(\lambda\vec{x} + (1-\lambda)\vec{y}) \geq \lambda g_i(\vec{x}) + (1-\lambda) g_i(\vec{y})$$
$$> 0$$

i.e. $\lambda\vec{x} + (1-\lambda)\vec{y} \in S$.

6. (a) proof by mathematic induction.

① $k=1$, the result is trivial since $f(\lambda_1 \vec{x}_1) = f(\vec{x}_1) = \lambda_1 f(\vec{x}_1)$ $[\lambda_1 = 1]$

② suppose $k=n$, we have

$$f\left(\sum_{i=1}^{k} \lambda_i \vec{x}_i\right) \leq \sum_{i=1}^{k} \lambda_i f(\vec{x}_i).$$

then.

$$\begin{aligned}
f\left(\sum_{i=1}^{n+1} \lambda_i \vec{x}_i\right) &= f\left(\sum_{i=1}^{n} \lambda_i \vec{x}_i + \lambda_{n+1} \vec{x}_{n+1}\right) \\
&= f\left((1-\lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} \vec{x}_i + \lambda_{n+1} \vec{x}_{n+1}\right), \quad \lambda_{n+1} \neq 1 \\
&\leq (1-\lambda_{n+1}) f\left(\sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} \vec{x}_i\right) + \lambda_{n+1} f(\vec{x}_{n+1}) \\
&\leq (1-\lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} f(\vec{x}_i) + \lambda_{n+1} f(\vec{x}_{n+1}) \\
&= \sum_{i=1}^{n+1} f(\vec{x}_i)
\end{aligned}$$

if $\lambda_{n+1} = 1$, then the result is trivial.

(b) let $f(\vec{x}) = \ln(\vec{x})$, $\vec{x} > 0$.

Since $\ln(\vec{x})$ is a non-decreasing function over $\vec{x} > 0$, then

To show $\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n}$ for $x_i > 0$. is equivalent to

show $\ln\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \geq \ln\left[\left(\prod_{i=1}^{n} x_i\right)^{1/n}\right] = \frac{1}{n}\ln\left(\prod_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n}\ln x_i$

it it equivalent to show

$$-\ln\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \leq -\frac{1}{n}\sum_{i=1}^{n}\ln x_i \quad (*)$$

since $-\ln(\vec{x})$ is a convex function over $\vec{x} > 0$, then from (a), we

we indeed have $(*)$ valid with $\lambda_i = \frac{1}{n}$.

If there exist some $x_i = 0$, then the inequality is trivial since.

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n} = 0$$