1 Question 1

(6 points) Given functions $g : \mathbb{R}^n \to \mathbb{R}^k$, $h : \mathbb{R}^k \to \mathbb{R}$, construct function f as following

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), .., g_k(x)).$$

prove the following:

1. Assume g_i is a convex function for every $1 \le i \le k$, h is a convex function and h is monotonically non-decreasing for each component. Then f is a convex function.

2. Assume g_i is a concave function for every $1 \le i \le k$, h is a convex function and h is monotonically non-increasing for each component. Then f is a convex function.

Proof:

For 1:

Since $g_1, \ldots, g_k$ are all convex functions, $dom\ g_1, \ldots, dom\ g_k$ are all convex sets. Note that since $dom\ f = dom\ g = dom\ g_1 \cap \ldots \cap dom\ g_k$, where $dom\ g_1, \ldots, dom\ g_k$ are all convex sets, we know $dom\ f$ is also convex by Proposition 3.8.1.

Also, for any $x, y \in dom\ f, 0 \le t \le 1$, $tx + (1-t)y \in dom\ f$, since $dom\ f$ is convex. Then, we have

$$f(tx + (1-t)y) = h\big(g(tx + (1-t)y)\big) = h\left(\begin{bmatrix} g_1(tx + (1-t)y) \\ \vdots \\ g_k(tx + (1-t)y) \end{bmatrix}\right)$$

Since g_i are convex for any $1 \le i \le k$, we have that $g_i(tx + (1-t)y) \le tg_i(x) + (1-t)g_i(y), \forall 1 \le i \le k$. Also, since h is monotonically non-decreasing for each component, we have

$$f(tx + (1-t)y) = h\left(\begin{bmatrix} g_1(tx + (1-t)y) \\ \vdots \\ g_k(tx + (1-t)y) \end{bmatrix}\right) \le h\left(\begin{bmatrix} tg_1(x) + (1-t)g_1(y) \\ \vdots \\ tg_k(x) + (1-t)g_k(y) \end{bmatrix}\right)$$

$$= h\left(t\begin{bmatrix} g_1(x) \\ \vdots \\ g_k(x) \end{bmatrix} + (1-t)\begin{bmatrix} g_1(y) \\ \vdots \\ g_k(y) \end{bmatrix}\right) = h\big(tg(x) + (1-t)g(y)\big)$$

Since h is convex, we have

$$f(tx + (1-t)y) = h\big(tg(x) + (1-t)g(y)\big) \le th\big(g(x)\big) + (1-t)h\big(g(y)\big)$$
$$= tf(x) + (1-t)f(y)$$

Therefore, f is a convex function.

For 2:

Let $p_i = -g_i, \forall 1 \le i \le k$. Then, we know any p_i is convex since any f_i is concave. Then, $dom\ p_1, \ldots, dom\ p_k$ are convex sets, so $dom\ g_1, \ldots, dom\ g_k$ are convex sets as well since $dom\ p_i = dom\ g_i, \forall 1 \le i \le k$.

Note that since $dom\ f = dom\ g = dom\ g_1 \cap \ldots \cap dom\ g_k$, where $dom\ g_1, \ldots, dom\ g_k$ are all convex sets, we know $dom\ f$ is also convex by Proposition 3.8.1.

Also, for any $x, y \in dom\ f, 0 \le t \le 1$, $tx + (1-t)y \in dom\ f = dom\ g$, since

$dom\ f$ is convex. Then, we have

$$
\begin{aligned}
tg_i(x) + (1-t)g_i(y) &= tg_i(x) + g_i(y) - tg_i(y) \\
&= -t[-g_i(x)] - [-g_i(y)] + t[-g_i(y)] = -tp_i(x) - p_i(y) + tp_i(y) \\
&= -[tp_i(x) + (1-t)p_i(y)], \forall 1 \le i \le k
\end{aligned}
$$

Note that since any p_i is convex,

$$tp_i(x) + (1-t)p_i(y) \ge p_i(tx + (1-t)y), \forall 1 \le i \le k$$

$$-[tp_i(x) + (1-t)p_i(y)] \le -p_i(tx + (1-t)y) = g_i(tx + (1-t)y), \forall 1 \le i \le k$$

Therefore,

$$tg_i(x) + (1-t)g_i(y) = -[tp_i(x) + (1-t)p_i(y)] \le g_i(tx + (1-t)y), \forall 1 \le i \le k$$

Also, since h is monotonically non-increasing for each component,

$$h\big(g(tx + (1-t)y)\big) = h\left(\begin{bmatrix} g_1(tx + (1-t)y) \\ \vdots \\ g_k(tx + (1-t)y) \end{bmatrix}\right) \le h\big(tg(x) + (1-t)g(y)\big)$$

Thus,

$$f(tx + (1-t)y) = h\big(g(tx + (1-t)y)\big) \le h\big(tg(x) + (1-t)g(y)\big)$$

Since h is convex,

$$
\begin{aligned}
f(tx + (1-t)y) &\le h\big(tg(x) + (1-t)g(y)\big) \le th\big(g(x)\big) + (1-t)h\big(g(y)\big) \\
&= tf(x) + (1-t)f(y)
\end{aligned}
$$

Therefore, f is a convex function.

2 Question 2

(4 points) Let $x \in \mathbb{R}^n$. Show

1. $f(x) = \ln(\sum_{i=1}^{n} exp(x_i))$ is a convex function

(Hint: show that $\nabla^2 f(x) = \text{diag}(\nabla f(x)) - \nabla f(x)^T \nabla f(x)$ where diag is an operator that reformulate a vector as a diagonal matrix.)

Proof:

For $f(x)$, we need $\sum_{i=1}^{n} e^{x_i} > 0$. Note that for any $x_i \in R$, $e^{x_i} > 0, \forall 1 \leq i \leq n$, we have $\sum_{i=1}^{n} e^{x_i} > 0$ for any $x_i \in R$. Therefore, $dom\ f = R^n$.

Let $u, v \in R^n$, we know for any $0 \leq t \leq 1$, $tu + (1-t)v \in R^n$, so $dom\ f = R^n$ is convex trivially.

Then, we have

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{e^{x_1}}{\sum_{i=1}^{n} e^{x_i}} \\ \vdots \\ \frac{e^{x_n}}{\sum_{i=1}^{n} e^{x_i}} \end{bmatrix}$$

$$\nabla f(x) \nabla f(x)^T = \begin{bmatrix} \frac{e^{x_1}}{\sum_{i=1}^{n} e^{x_i}} \\ \vdots \\ \frac{e^{x_n}}{\sum_{i=1}^{n} e^{x_i}} \end{bmatrix} \begin{bmatrix} \frac{e^{x_1}}{\sum_{i=1}^{n} e^{x_i}} & \cdots & \frac{e^{x_n}}{\sum_{i=1}^{n} e^{x_i}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{e^{x_1} e^{x_1}}{(\sum_{i=1}^{n} e^{x_i})^2} & \cdots & \frac{e^{x_1} e^{x_n}}{(\sum_{i=1}^{n} e^{x_i})^2} \\ \vdots & \ddots & \vdots \\ \frac{e^{x_n} e^{x_1}}{(\sum_{i=1}^{n} e^{x_i})^2} & \cdots & \frac{e^{x_n} e^{x_n}}{(\sum_{i=1}^{n} e^{x_i})^2} \end{bmatrix}$$

$$diag(\nabla f(x)) = \begin{bmatrix} \frac{e^{x_1}}{\sum_{i=1}^{n} e^{x_i}} & & \\ & \ddots & \\ & & \frac{e^{x_n}}{\sum_{i=1}^{n} e^{x_i}} \end{bmatrix}$$

Then, the hessian matrix for any $x \in dom\, f$ satisfies:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{e^{x_1}(\sum_{i=1}^n e^{x_i}) - e^{x_1}e^{x_1}}{(\sum_{i=1}^n e^{x_i})^2} & \cdots & -\frac{e^{x_1}e^{x_n}}{(\sum_{i=1}^n e^{x_i})^2} \\ \vdots & \ddots & \vdots \\ -\frac{e^{x_n}e^{x_1}}{(\sum_{i=1}^n e^{x_i})^2} & \cdots & \frac{e^{x_n}(\sum_{i=1}^n e^{x_i}) - e^{x_n}e^{x_n}}{(\sum_{i=1}^n e^{x_i})^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{e^{x_1}}{\sum_{i=1}^n e^{x_i}} - \frac{e^{x_1}e^{x_1}}{(\sum_{i=1}^n e^{x_i})^2} & \cdots & -\frac{e^{x_1}e^{x_n}}{(\sum_{i=1}^n e^{x_i})^2} \\ \vdots & \ddots & \vdots \\ -\frac{e^{x_n}e^{x_1}}{(\sum_{i=1}^n e^{x_i})^2} & \cdots & \frac{e^{x_n}}{\sum_{i=1}^n e^{x_i}} - \frac{e^{x_n}e^{x_n}}{(\sum_{i=1}^n e^{x_i})^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{e^{x_1}}{\sum_{i=1}^n e^{x_i}} & & \\ & \ddots & \\ & & \frac{e^{x_n}}{\sum_{i=1}^n e^{x_i}} \end{bmatrix} - \begin{bmatrix} \frac{e^{x_1}e^{x_1}}{(\sum_{i=1}^n e^{x_i})^2} & \cdots & \frac{e^{x_1}e^{x_n}}{(\sum_{i=1}^n e^{x_i})^2} \\ \vdots & \ddots & \vdots \\ \frac{e^{x_n}e^{x_1}}{(\sum_{i=1}^n e^{x_i})^2} & \cdots & \frac{e^{x_n}e^{x_n}}{(\sum_{i=1}^n e^{x_i})^2} \end{bmatrix}$$

$$= diag\big(\nabla f(x)\big) - \nabla f(x) \nabla f(x)^T$$

Then, for any $y \neq 0 \in R^n$, we all have'

$$y^T \nabla^2 f(x) y = y^T \big[diag\big(\nabla f(x)\big) - \nabla f(x) \nabla f(x)^T\big] y$$

$$= y^T diag\big(\nabla f(x)\big) y - y^T \nabla f(x) \nabla f(x)^T y$$

For convenience, we let

$$\nabla f(x) = \begin{bmatrix} \frac{e^{x_1}}{\sum_{i=1}^n e^{x_i}} \\ \vdots \\ \frac{e^{x_n}}{\sum_{i=1}^n e^{x_i}} \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = d$$

So

$$diag\big(\nabla f(x)\big) = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} = D$$

Since $0 < e^{x_j} < \sum_{i=1}^n e^{x_i}, \forall 1 \leq j \leq n$, we know $0 < d_j = \frac{e^{x_j}}{\sum_{i=1}^n e^{x_i}} < 1$.

Then,

$$y^T\nabla^2 f(x)y = y^T diag\big(\nabla f(x)\big)y - y^T\nabla f(x)\nabla f(x)^T y = y^T D y - y^T \mathrm{dd}^\mathrm{T} y$$

$$= [y_1 \quad \cdots \quad y_n]\begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} - (y^T d)^2$$

$$= [y_1 d_1 \quad \cdots \quad y_n d_n]\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} - \left([y_1 \quad \cdots \quad y_n]\begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}\right)^2$$

$$= \left(\sum_{i=1}^{n} y_i^2 d_i\right) - \left(\sum_{i=1}^{n} y_i d_i\right)^2$$

$$= \left(\sum_{i=1}^{n} y_i^2 \frac{e^{x_i}}{\sum_{m=1}^{n} e^{x_m}}\right) - \left(\sum_{i=1}^{n} y_i \frac{e^{x_i}}{\sum_{m=1}^{n} e^{x_m}}\right)^2$$

$$= \frac{1}{\sum_{m=1}^{n} e^{x_m}}\left(\sum_{i=1}^{n} y_i^2 e^{x_i}\right) - \left(\frac{1}{\sum_{m=1}^{n} e^{x_m}}\right)^2\left(\sum_{i=1}^{n} y_i e^{x_i}\right)^2$$

$$= \frac{(\sum_{m=1}^{n} e^{x_m})(\sum_{i=1}^{n} y_i^2 e^{x_i}) - (\sum_{i=1}^{n} y_i e^{x_i})^2}{(\sum_{m=1}^{n} e^{x_m})^2}$$

$$= \frac{\left\|\begin{bmatrix} \sqrt{e^{x_1}} \\ \vdots \\ \sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 \left\|\begin{bmatrix} y_1\sqrt{e^{x_1}} \\ \vdots \\ y_n\sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 - \left([y_1\sqrt{e^{x_1}} \quad \cdots \quad y_n\sqrt{e^{x_n}}]\begin{bmatrix} \sqrt{e^{x_1}} \\ \vdots \\ \sqrt{e^{x_n}} \end{bmatrix}\right)^2}{(\sum_{m=1}^{n} e^{x_m})^2}$$

$$= \frac{\left\|\begin{bmatrix} \sqrt{e^{x_1}} \\ \vdots \\ \sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 \left\|\begin{bmatrix} y_1\sqrt{e^{x_1}} \\ \vdots \\ y_n\sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 - \left\|\begin{bmatrix} \sqrt{e^{x_1}} \\ \vdots \\ \sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 \left\|\begin{bmatrix} y_1\sqrt{e^{x_1}} \\ \vdots \\ y_n\sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 \cos^2(\theta)}{(\sum_{m=1}^{n} e^{x_m})^2}$$

$$= \frac{\left\|\begin{bmatrix} \sqrt{e^{x_1}} \\ \vdots \\ \sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 \left\|\begin{bmatrix} y_1\sqrt{e^{x_1}} \\ \vdots \\ y_n\sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 (1-\cos^2(\theta))}{(\sum_{m=1}^{n} e^{x_m})^2}$$

Note that since $0 \le \cos^2(\theta) \le \cos(\theta) \le 1$, $1-\cos^2(\theta) \ge 0$, we have

$$y^T\nabla^2 f(x)y = \frac{\left\|\begin{bmatrix} \sqrt{e^{x_1}} \\ \vdots \\ \sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 \left\|\begin{bmatrix} y_1\sqrt{e^{x_1}} \\ \vdots \\ y_n\sqrt{e^{x_n}} \end{bmatrix}\right\|_2^2 (1-\cos^2(\theta))}{(\sum_{m=1}^{n} e^{x_m})^2} \ge 0$$

Therefore, $\nabla^2 f(x)$ is PSD and $dom\ f$ is a convex set, which means f is a convex function.

3 Question 3

(6 points) We say $\bar{x}$ is a saddle point if $\nabla^2 f(\bar{x})$ has both negative and positive eigenvalues. Now given function

$$f(x) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4$$

Compute all its stationary point and justify if they are local optimum (minimizer or maximizer), saddle points, or global optimum(minimizer or maximizer)?

Solution:

First, we check the convexity of $f(x)$:

For $z = \begin{bmatrix} -\frac{1}{2} \\ 0 \end{bmatrix} \in R^2 = dom\, f$, $\nabla^2 f(z) = \begin{bmatrix} 4 + 12 * \left(-\frac{1}{2}\right) + 12\left(-\frac{1}{2}\right)^2 & -2 \\ -2 & 2 \end{bmatrix} =$

$\begin{bmatrix} 4 - 6 + 3 & -2 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ -2 & 2 \end{bmatrix}$

Consider $l = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \in R^2$, then $l^T \nabla^2 f(z) l = [-1 \quad -1] \begin{bmatrix} 1 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix} =$

$[1 \quad 0] \begin{bmatrix} -1 \\ -1 \end{bmatrix} = -1 < 0$, which means $\nabla^2 f(z)$ is not PSD by definition, so f is not a convex function by Proposition 4.9.1.

Then,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 \\ 2x_2 - 2x_1 \end{bmatrix}$$

Solve

$$\nabla f(x) = \begin{bmatrix} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 \\ 2x_2 - 2x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 = 0 \ (1) \\ 2x_2 - 2x_1 = 0 \ (2) \end{cases}$$

From (2),

$$2x_2 - 2x_1 = 0$$
$$x_1 = x_2$$

Let $x_1 = x_2 = t$, and substitute it into (1):

$$4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 = 4t - 2t + 6t^2 + 4t^3 = 4t^3 + 6t^2 + 2t = 0$$

When t = 0, t satisfies the equation, so $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is a solution.

When $t \neq 0$,

$$4t^3 + 6t^2 + 2t = 0$$
$$2t^2 + 3t + 1 = 0$$
$$t = \frac{-3 \pm \sqrt{3^2 - 4 * 2 * 1}}{2 * 2} = \frac{-3 \pm \sqrt{9 - 8}}{4} = \frac{-3 \pm 1}{4}$$
$$t = -1 \ or \ t = -\frac{1}{2}$$

Therefore, $\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}-1\\-1\end{bmatrix}$, and $\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\end{bmatrix}$ are all stationary points of $f(x)$.

Then, to decide if they are local optimum, saddle points, or global minimizer, we compute the Hessian matrix of $f(x)$:

$$\nabla^2 f(x) = \begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2}\\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2}\end{bmatrix} = \begin{bmatrix}4+12x_1+12x_1^2 & -2\\ -2 & 2\end{bmatrix}$$

For $\boldsymbol{x} = [0,0]^T$,

$$\nabla^2 f\left(\begin{bmatrix}0\\0\end{bmatrix}\right) = \begin{bmatrix}4 & -2\\ -2 & 2\end{bmatrix}$$

$$\nabla^2 f\left(\begin{bmatrix}0\\0\end{bmatrix}\right) v = \lambda v$$

$$\left(\nabla^2 f\left(\begin{bmatrix}0\\0\end{bmatrix}\right) - \lambda I\right) v = 0$$

$$\begin{bmatrix}4-\lambda & -2\\ -2 & 2-\lambda\end{bmatrix} v = 0$$

$$\det\left(\begin{bmatrix}4-\lambda & -2\\ -2 & 2-\lambda\end{bmatrix}\right) = (4-\lambda)(2-\lambda) - 4 = 0$$

$$8 - 6\lambda + \lambda^2 - 4 = 0$$

$$\lambda^2 - 6\lambda + 4 = 0$$

$$\lambda = \frac{6 \pm \sqrt{36-16}}{2} = \frac{6 \pm 2\sqrt{5}}{2} = 3 \pm \sqrt{5} > 0$$

Since all eigenvalues of $\nabla^2 f\left(\begin{bmatrix}0\\0\end{bmatrix}\right)$ are positive, $\nabla^2 f\left(\begin{bmatrix}0\\0\end{bmatrix}\right)$ is positive definite by Proposition 2.19, which means $\begin{bmatrix}0\\0\end{bmatrix}$ is a strict local minimizer by Theorem 5.4.

For $\boldsymbol{x} = [-1,-1]^T$,

$$\nabla^2 f\left(\begin{bmatrix}-1\\-1\end{bmatrix}\right) = \begin{bmatrix}4 & -2\\ -2 & 2\end{bmatrix}$$

Thus, by the calculation above, $\nabla^2 f\left(\begin{bmatrix}-1\\-1\end{bmatrix}\right)$ also has $\lambda = 3 \pm \sqrt{5} > 0$. Hence, $\nabla^2 f\left(\begin{bmatrix}-1\\-1\end{bmatrix}\right)$ is positive definite by Proposition 2.19, which means $\begin{bmatrix}0\\0\end{bmatrix}$ is a strict local minimizer by Theorem 5.4.

For $\boldsymbol{x} = \left[-\frac{1}{2}, -\frac{1}{2}\right]^T$,

$$\nabla^2 f\left(\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\end{bmatrix}\right) = \begin{bmatrix}1 & -2\\ -2 & 2\end{bmatrix}$$

Then, we calculate the eigenvalues of it:

$$\nabla^2 f\left(\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}\right) v = \lambda v$$

$$\left(\nabla^2 f\left(\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}\right) - \lambda I\right) v = 0$$

$$\begin{bmatrix} 1-\lambda & -2 \\ -2 & 2-\lambda \end{bmatrix} v = 0$$

$$(1-\lambda)(2-\lambda) - 4 = 0$$

$$\lambda^2 - 3\lambda - 2 = 0$$

$$\lambda = \frac{3 \pm \sqrt{9+8}}{2} = \frac{3 \pm \sqrt{17}}{2} = \frac{\sqrt{9} \pm \sqrt{17}}{2}$$

$$\lambda_1 = \frac{\sqrt{9} - \sqrt{17}}{2} < 0, \lambda_2 = \frac{\sqrt{9} + \sqrt{17}}{2} > 0$$

Since $\nabla^2 f\left(\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}\right)$ has both positive and negative eigenvalues, $\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$ is a saddle point.

Then, we compare 2 local minimizers to decide global minimizers:

$$f\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = 0 = f\left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}\right) = 2 + 1 - 2 - 2 + 1 = 0$$

Then, we note that

$$2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4 = (x_1 - x_2)^2 + x_1^2(x_1 + 1)^2 \geq 0$$

Therefore, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$ are global minimizers (and also local minimizers trivially), and $\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$ is not global or local minimizer, since $f\left(\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}\right) = 2 * \frac{1}{4} + \frac{1}{4} - 2 * \frac{1}{4} - 2 * \frac{1}{8} + \frac{1}{16} = \frac{1}{16} > 0.$

Finally, since for $x = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$,

$$\lim_{x_1 \to \infty} f(x) = \lim_{x_1 \to \infty} 2x_1^2 + 2x_1^3 + x_1^4 = +\infty$$

The function has no global maximizer.

Therefore, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$ are global minimizers (and also local minimizers trivially), and $\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$ is a saddle point.

4 Question 4

(4 points) Let $A \in \mathbb{R}^{n\times n}$ be a symmstric matrix. Consider the following optimization problem

$$\begin{aligned} p_{\min} = &\quad \min y^T A y \\ \text{s. t.} &\quad \|y\|_2 = 1 \end{aligned}$$

and

$$\begin{aligned} p_{\max} = &\quad \max y^T A y \\ \text{s. t.} &\quad \|y\|_2 = 1 \end{aligned}$$

Prove: $p_{\min} = \lambda_{\min}$ and $p_{\max} = \lambda_{\max}$ where $\lambda_{\min}$ is the minimum eigenvalue of A and $\lambda_{\max}$ is the maximum eigenvalue of A.

(Hint: show that $\lambda_{\min}\|y\|_2^2 \le y^T A y \le \lambda_{\max}\|y\|_2^2$)

Proof:

Let $(\lambda_i, v_i), \forall 1 \le i \le n$ be mutually orthogonal and linearly independent normalized eigenpairs. Such eigenpairs exist because A is symmetric according to Proposition 2.15.

Then, for any $y \in R^n$ such that $\|y\|_2 = 1$, y can be represented as the linear combination of eigenvectors of A:

$$y = \sum_{i=1}^{n} \alpha_i v_i \,, \alpha_i \in R, \forall 1 \le i \le n$$

Then, we have

$$y^T A y = y^T A \sum_{i=1}^{n} \alpha_i v_i = y^T \sum_{i=1}^{n} \alpha_i A v_i = y^T \sum_{i=1}^{n} \alpha_i \lambda_i v_i \le y^T \sum_{i=1}^{n} \alpha_i \lambda_{max} v_i$$

$$= y^T \lambda_{max} \sum_{i=1}^{n} \alpha_i v_i = y^T \lambda_{max} y = \lambda_{max}\|y\|_2^2 = \lambda_{max}$$

Similarly,

$$y^T A y = y^T \sum_{i=1}^{n} \alpha_i \lambda_i v_i \ge y^T \sum_{i=1}^{n} \alpha_i \lambda_{min} v_i = y^T \lambda_{min} \sum_{i=1}^{n} \alpha_i v_i = y^T \lambda_{min} y$$

$$= \lambda_{min}\|y\|_2^2 = \lambda_{min}$$

That said,

$$\lambda_{min} \le y^T A y \le \lambda_{max}$$

Then, let (λ_{min}, v_{min}) be the eigenpair with the smallest eigenvalue λ_{min} and $\|v_{min}\|_2 = 1$, and we have

$$v_{min}^T A v_{min} = v_{min}^T \lambda_{min} v_{min} = \lambda_{min}\|v_{min}\|_2^2 = \lambda_{min}$$

Similarly, let (λ_{max}, v_{max}) be the eigenpair with the largest eigenvalue λ_{max} and $\|v_{max}\|_2 = 1$, and we have

$$v_{max}^T A v_{max} = v_{max}^T \lambda_{max} v_{max} = \lambda_{max}\|v_{max}\|_2^2 = \lambda_{max}$$

Therefore, $p_{min} = \lambda_{min}$ and $p_{max} = \lambda_{max}$.

5 Question 5

(4 points) Let function $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2-class function, which also has a Lipschitz gradient, that is,

$$||\nabla f(x) - \nabla f(y)||_2 \leq L||x - y||_2$$

where L is a positive number.

Prove $||\nabla^2 f(x)||_2 \leq L$.

(Hint: use Taylor theorem, and show the maximum absolute value of the eigenvalues of $\nabla^2 f(x)$ is bounded.)

Proof:

Method 1: If we can use Taylor Theorem for the second derivative:

Let $x \in R^n$, v_{max} be the normalized eigenvector of $\nabla^2 f(x)$ with the corresponding eigenvalue λ_{max}, where $|\lambda_{max}| = \rho\left(\nabla^2 f(x)\right)$ is the largest among all eigenvalues.

Then, let $y = x - cv_{max} \in R^n, c \in R$.

By Taylor Theorem for the second derivative,

$$\nabla f(x) = \nabla f(y) + (x-y)^T \nabla^2 f\big(y + \theta(x-y)\big), \exists 0 < \theta < 1$$

$$\nabla f(x) - \nabla f(y) = (x-y)^T \nabla^2 f\big(y + \theta(x-y)\big)$$

$$\left\|(x-y)^T \nabla^2 f\big(y + \theta(x-y)\big)\right\|_2 = \|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2$$

When $c \to 0$, we get $x \to y$, which gives

$$\|(x-y)^T \nabla^2 f(x)\|_2 = \|\nabla^2 f(x)^T (x-y)\|_2 = \|\nabla^2 f(x)(x-y)\|_2 \leq L\|x-y\|_2$$

$$\left\|\nabla^2 f(x) \frac{(x-y)}{\|x-y\|_2}\right\|_2 = \left\|\nabla^2 f(x) \frac{cv_{max}}{\|cv_{max}\|_2}\right\|_2 = \left\|\frac{\lambda_{max} cv_{max}}{\|cv_{max}\|_2}\right\|_2$$

$$= \frac{\left|\lambda_{max}\right|}{\|cv_{max}\|_2}\|cv_{max}\|_2 = \left|\lambda_{max}\right| = \rho\left(\nabla^2 f(x)\right) = \|\nabla^2 f(x)\|_2 \leq L$$

Method 2: Just use Taylor Theorem in the lecture note:
Let $x \in R^n$, v_{max} be the normalized eigenvector of $\nabla^2 f(x)$ with the corresponding eigenvalue λ_{max}, where $|\lambda_{max}| = \rho\big(\nabla^2 f(x)\big)$ is the largest among all eigenvalues.
Then, let $y = x - cv_{max} \in R^n, c \in R$.
By Taylor Theorem, there exists some $0 < \theta, \alpha < 1$, such that

$$f(x) = f(y) + (x-y)^T \nabla f(y) + \frac{1}{2}(x-y)^T \nabla^2 f\big(y + \theta(x-y)\big)(x-y) \ (*)$$

$$f(y) = f(x) + (y-x)^T \nabla f(x) + \frac{1}{2}(y-x)^T \nabla^2 f\big(x + \alpha(y-x)\big)(y-x) \ (**)$$

Let $(*) + (**)$:

$$\begin{aligned}&(x-y)^T \nabla f(x) - (x-y)^T \nabla f(y)\\ &\qquad = \frac{1}{2}(x-y)^T \nabla^2 f\big(y + \theta(x-y)\big)(x-y)\\ &\qquad + \frac{1}{2}(y-x)^T \nabla^2 f\big(x + \alpha(y-x)\big)(y-x)\end{aligned}$$

$$\begin{aligned}&\left|(x-y)^T\big(\nabla f(x) - \nabla f(y)\big)\right|\\ &\qquad = \left|\frac{1}{2}(x-y)^T \nabla^2 f\big(y + \theta(x-y)\big)(x-y)\right.\\ &\qquad \left. + \frac{1}{2}(x-y)^T \nabla^2 f\big(x + \alpha(y-x)\big)(x-y)\right|\end{aligned}$$

When $c \to 0$, we get $x \to y$, which gives

$$\begin{aligned}&\left|(x-y)^T\big(\nabla f(x) - \nabla f(y)\big)\right|\\ &\qquad = \left|\frac{1}{2}(x-y)^T \nabla^2 f(x)(x-y) + \frac{1}{2}(x-y)^T \nabla^2 f(x)(x-y)\right|\\ &\qquad = |(x-y)^T \nabla^2 f(x)(x-y)|\end{aligned}$$

$$|(x-y)^T \nabla^2 f(x)(x-y)| = |cv_{max}^T \nabla^2 f(x) cv_{max}| = \left|cv_{max}^T \lambda_{\max} cv_{max}\right|$$

$$= \lambda_{\max} \|cv_{max}\|_2^2 = \left|(x-y)^T\big(\nabla f(x) - \nabla f(y)\big)\right|$$

By Schwartz Inequality,

$$\lambda_{\max} \|cv_{max}\|_2^2 = \left|(x-y)^T\big(\nabla f(x) - \nabla f(y)\big)\right| \le \|x-y\|_2 \|\nabla f(x) - \nabla f(y)\|_2$$

$$\le \|x-y\|_2 L \|x-y\|_2 = L\|x-y\|_2^2 = L\|cv_{max}\|_2^2$$

Therefore, since $\|cv_{max}\|_2^2 \ge 0$,

$$\|\nabla^2 f(x)\|_2 = \rho\big(\nabla^2 f(x)\big) = \lambda_{\max} \le L$$