Problem Set #4

Fiona Fan, in collaboration with Zongze, Winston and Shirley

Exercise 6.6

We first find the critical points. Observe that $f(x,y) = 3x^2y + 4xy^2 + xy$

let $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$

FOC: $Df(\mathbf{x}) = [fx, f_y] = [6xy + 4y^2 + y, 3x^2 + 8xy + x] = [0, 0]$

$\Rightarrow \begin{cases} 6xy + 4y^2 + y = 0 \\ 3x^2 + 8xy + x = 0 \end{cases}$

The critical points are:

$\mathbf{x_1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{x_2} = \begin{bmatrix} 0 \\ -\frac{1}{4} \end{bmatrix}, \mathbf{x_3} = \begin{bmatrix} -\frac{1}{3} \\ 0 \end{bmatrix}, \mathbf{x_4} = \begin{bmatrix} -\frac{1}{9} \\ -\frac{1}{12} \end{bmatrix}$

$D^2 f = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 6y & 6x + 8y + 1 \\ 6x + 8y + 1 & 8x \end{bmatrix}$

Then,

$D^2 f(\mathbf{x_1}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \Rightarrow$ saddle point

$D^2 f(\mathbf{x_2}) = \begin{bmatrix} -\frac{3}{2} & -1 \\ -1 & 0 \end{bmatrix} \quad \Rightarrow$ saddle point

$D^2 f(\mathbf{x_3}) = \begin{bmatrix} 0 & -1 \\ -1 & -\frac{8}{3} \end{bmatrix} \quad \Rightarrow$ saddle point

$D^2 f(\mathbf{x_4}) = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{8}{9} \end{bmatrix} < 0 \quad \Rightarrow$ local maximum

Exercise 6.7

(1) Since $Q = A^T + A$, and A is a square matrix, $Q^T = (A^T)^T + A^T = A + A^T = Q$. So Q is symmetriic. Observe that $\mathbf{x}^T A\mathbf{x} = \langle \mathbf{x}, A\mathbf{x} \rangle$, and $\mathbf{x}^T A^T \mathbf{x} = (A\mathbf{x})^T \mathbf{x} = \langle A\mathbf{x}, \mathbf{x} \rangle$. Since here we restrict the field to be $\mathbb{R}$, we have $\langle \mathbf{x}, A\mathbf{x} \rangle = \langle A\mathbf{x}, \mathbf{x} \rangle$. So it follows that

$$\mathbf{x}^T Q\mathbf{x} = \mathbf{x}^T(A^T + A)\mathbf{x} = \mathbf{x}^T A^T \mathbf{x} + \mathbf{x}^T A\mathbf{x} = 2\mathbf{x}^T A\mathbf{x}.$$

Thus we have

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T\mathbf{x} + c.$$

(2) The first order necessary condition implies that if $\mathbf{x}^*$ is a minimizer, then it must be $Df(\mathbf{x}^*) = Q^T\mathbf{x}^* - \mathbf{b} = 0$. Hence $Q^T\mathbf{x}^* = \mathbf{b}$.

(3) Observe that $D^2 f(\mathbf{x}^*) = Q$, and since Q is positive definite, it follows from the second order sufficient condition that $\mathbf{x}^*$ is a minimizer, which is also the solution to the liner system $Q^T\mathbf{x}^* = \mathbf{b}$.

Exercise 6.11

Proof. Observe that $f''(x) = 2a > 0$, so the x^* that satisfies $f'(x^*) = 2ax^* + b = 0$ is the minimizer. $\forall x_0$, by Newton's method,

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = x_0 - \frac{2ax_0 + b}{2a} = -\frac{b}{2a}.$$

Since the quadratic function can also be expressed as $f(x) = a(x + \frac{b}{2a})^2 + \frac{4ac-b^2}{4a}$, it follows that $x_1 = -\frac{b}{2a}$ is the unique minimizer. $\square$

Exercise 7.1

Proof. Since $conv(S)$ is the set of all convex combinations of vectors in S, it follows immediately that this is a convex set. $\square$

Exercise 7.2

Proof. A hyperplane is a set of the form $P = \{\mathbf{x} \in V | \langle \mathbf{a}, \mathbf{x} \rangle = b\}$, for some $b \in \mathbb{R}$ and $\mathbf{a} \neq \mathbf{0}$. Take $\mathbf{x}, \mathbf{y} \in P$. Let $\lambda \in [0, 1]$. Observe that

$$\langle \mathbf{a}, \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \rangle = \lambda\langle \mathbf{a}, \mathbf{x} \rangle + (1-\lambda)\langle \mathbf{a}, \mathbf{y} \rangle = \lambda b + (1-\lambda)b = b.$$

So $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in P, \forall \mathbf{x}, \mathbf{y}$. Thus a hyperplane is a convex set. $\square$

Proof. A half space is a set of the form $H = \{\mathbf{x} \in \mathbb{R}^n | \langle \mathbf{a}, \mathbf{x} \rangle \leq b\}$, for some $b \in \mathbb{R}$ and $\mathbf{a} \neq \mathbf{0}$. Take $\mathbf{x}, \mathbf{y} \in H$. Let $\lambda \in [0, 1]$. Observe that

$$\langle \mathbf{a}, \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \rangle = \lambda\langle \mathbf{a}, \mathbf{x} \rangle + (1-\lambda)\langle \mathbf{a}, \mathbf{y} \rangle \leq \lambda b + (1-\lambda)b = b.$$

So $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in H, \forall \mathbf{x}, \mathbf{y}$. Thus a half space is a convex set. $\square$

Exercise 7.2
(1) Since we restrict the field to be $\mathbb{R}$, we have $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle, \forall \mathbf{x}, \mathbf{y}$.

$$\begin{aligned}
RHS &= \langle \mathbf{x}-\mathbf{p}, \mathbf{x}-\mathbf{p} \rangle + \langle \mathbf{p}-\mathbf{y}, \mathbf{p}-\mathbf{y} \rangle + 2\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y} \rangle \\
&= (\langle \mathbf{x}-\mathbf{p}, \mathbf{x}-\mathbf{p} \rangle + \langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y} \rangle) + (\langle \mathbf{p}-\mathbf{y}, \mathbf{p}-\mathbf{y} \rangle + \langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y} \rangle) \\
&= \langle \mathbf{x}-\mathbf{p}, \mathbf{x}-\mathbf{p}+\mathbf{p}-\mathbf{y} \rangle + \langle \mathbf{p}-\mathbf{y}+\mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y} \rangle \\
&= \langle \mathbf{x}-\mathbf{p}, \mathbf{x}-\mathbf{y} \rangle + \langle \mathbf{x}-\mathbf{y}, \mathbf{p}-\mathbf{y} \rangle \\
&= \langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{p} \rangle + \langle \mathbf{x}-\mathbf{y}, \mathbf{p}-\mathbf{y} \rangle \\
&= \langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{p}+\mathbf{p}-\mathbf{y} \rangle \\
&= \langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y} \rangle = \|\mathbf{x}-\mathbf{y}\|^2 = LHS
\end{aligned}$$

(2) Since $2\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y} \rangle > 0$ and $\|\mathbf{p}-\mathbf{y}\|^2 \geq 0, we have$

$$\begin{aligned}
\|\mathbf{x}-\mathbf{y}\|^2 &= \|\mathbf{x}-\mathbf{p}\|^2 + \|\mathbf{p}-\mathbf{y}\|^2 + 2\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y} \rangle \\
&> \|\mathbf{x}-\mathbf{p}\|^2 + 0 + 0 \\
&= \|\mathbf{x}-\mathbf{p}\|^2, \forall \mathbf{y} \neq \mathbf{p}
\end{aligned}$$

(3) Let $\mathbf{z} = \lambda\mathbf{y} + (1-\lambda)\mathbf{p}$. Then

$$\begin{aligned}
LHS &= \langle \mathbf{x}-\lambda\mathbf{y}-(1-\lambda)\mathbf{p}, \mathbf{x}-\lambda\mathbf{y}-(1-\lambda)\mathbf{p} \rangle \\
&= \langle \mathbf{x}, \mathbf{x} \rangle - 2\lambda\langle \mathbf{x}, \mathbf{y} \rangle - 2(1-\lambda)\langle \mathbf{x}, \mathbf{p} \rangle + 2\lambda(1-\lambda)\langle \mathbf{p}, \mathbf{y} \rangle + \lambda^2\langle \mathbf{y}, \mathbf{y} \rangle + (1-\lambda)^2\langle \mathbf{p}, \mathbf{p} \rangle
\end{aligned}$$

$$
\begin{aligned}
RHS &= \langle \mathbf{x}-\mathbf{p}, \mathbf{x}-\mathbf{p}\rangle + 2\lambda\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + \lambda^2\langle \mathbf{y}-\mathbf{p}, \mathbf{y}-\mathbf{p}\rangle \\
&= \langle \mathbf{x}, \mathbf{x}\rangle - 2\langle \mathbf{x}, \mathbf{p}\rangle + \langle \mathbf{p}, \mathbf{p}\rangle + 2\lambda\langle \mathbf{x}, \mathbf{p}\rangle - 2\lambda\langle \mathbf{p}, \mathbf{p}\rangle + 2\lambda\langle \mathbf{p}, \mathbf{y}\rangle \\
&- 2\lambda\langle \mathbf{x}, \mathbf{y}\rangle + \lambda^2\langle \mathbf{y}, \mathbf{y}\rangle + \lambda^2\langle \mathbf{p}, \mathbf{p}\rangle - 2\lambda^2\langle \mathbf{p}, \mathbf{y}\rangle \\
&= \langle \mathbf{x}, \mathbf{x}\rangle - 2\lambda\langle \mathbf{x}, \mathbf{y}\rangle - 2(1-\lambda)\langle \mathbf{x}, \mathbf{p}\rangle + 2\lambda(1-\lambda)\langle \mathbf{p}, \mathbf{y}\rangle + \lambda^2\langle \mathbf{y}, \mathbf{y}\rangle + (1-2\lambda+\lambda^2)\langle \mathbf{p}, \mathbf{p}\rangle \\
&= \langle \mathbf{x}, \mathbf{x}\rangle - 2\lambda\langle \mathbf{x}, \mathbf{y}\rangle - 2(1-\lambda)\langle \mathbf{x}, \mathbf{p}\rangle + 2\lambda(1-\lambda)\langle \mathbf{p}, \mathbf{y}\rangle + \lambda^2\langle \mathbf{y}, \mathbf{y}\rangle + (1-\lambda)^2\langle \mathbf{p}, \mathbf{p}\rangle \\
&= LHS
\end{aligned}
$$

(4) Suppose $\mathbf{p}$ is the projection of $\mathbf{x}$ onto convex set C. Pick $\forall \mathbf{y} \in C$. Since C is convex , it follows that $z = \lambda\mathbf{y} + (1-\lambda)\mathbf{p} \in C, \lambda \in [0,1]$. Since $\mathbf{p}$ is the projection of $\mathbf{x}$, by definition $\|\mathbf{x}-\mathbf{z}\| \geq \|\mathbf{x}-\mathbf{p}\|$, and hence $\|\mathbf{x}-z\|^2 \geq \|\mathbf{x}-\mathbf{p}\|^2$.
By (3), when $\lambda \neq 0$,

$$
2\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle = \lambda \|\mathbf{y}-\mathbf{p}\|^2 = \frac{\|\mathbf{x}-\mathbf{z}\|^2 - \|\mathbf{x}-\mathbf{p}\|^2}{\lambda} \leq 0, \forall \mathbf{y}.
$$

Since $\|\mathbf{y}-\mathbf{p}\|^2 \geq 0$, and λ can be arbitrarily small, we have $2\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle \geq 0$, and hence $\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle \geq 0$.
When $\lambda = 0$, we have $z = \mathbf{y}$ and $\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}-\mathbf{p}\|^2$. Since the projection is unique, we have $\mathbf{y} = \mathbf{p}$. In this case $\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle = 0$.
Hence, $\forall \mathbf{y} \in C, \langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle \geq 0$.

To show the converse direction, we can see that by (2), if $\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle \geq 0$, then $\|\mathbf{x}-\mathbf{y}\| > \|\mathbf{x}-\mathbf{p}\|, \forall \mathbf{y} \in C, \mathbf{y} \neq \mathbf{p}$. It then follows that $\mathbf{p}$ is the projection of $\mathbf{x}$ onto C.

Exercise 7.8

Proof. Let $\mathbf{x_1}, \mathbf{x_2} \in \mathbb{R}^n$ be two aribitrary vectors. Let $\lambda \in [0,1]$. Observe that

$$
\begin{aligned}
g(\lambda\mathbf{x_1} + (1-\lambda)\mathbf{x_2}) &= f(A[\lambda\mathbf{x_1} + (1-\lambda)\mathbf{x_2}] + b) \\
&= f(\lambda A\mathbf{x_1} + (1-\lambda)A\mathbf{x_2} + b) \\
&= f(\lambda[A\mathbf{x_1} + b] + (1-\lambda)[A\mathbf{x_2} + b]) \\
&\leq \lambda f(A\mathbf{x_1} + b) + (1-\lambda)f(A\mathbf{x_2} + b) \\
&= \lambda g(\mathbf{x_1}) + (1-\lambda)g(\mathbf{x_2})
\end{aligned}
$$

Hence g is a convex function. □

Exercise 7.8
(1)

Proof. Suppose A, B are positive definite matrices. Let $\lambda \in [0,1]$. Observe that $\forall \mathbf{x}, \mathbf{x}^T(\lambda A)\mathbf{x} + \mathbf{x}^T(1-\lambda)B\mathbf{x} = \lambda\mathbf{x}^T A\mathbf{x} + (1-\lambda)\mathbf{x}^T B\mathbf{x} > 0$. So the set of positive definite matrices is a convex set. □

(2)

Proof. (a) This follows immediately from Lemma 7.2.7.
(b) Observe that

$$\begin{aligned}
g(t) &= -\log\{\det[tA + (1-t)B]\} = -\log\{\det[tS^H S + (1-t)S^H (S^H)^{-1} B S^{-1} S]\} \\
&= -\log\{\det[S^H (tI + (1-t)(S^H)^{-1} B S^{-1}) S]\} \\
&= -\log\{\det[S^H] \det[tI + (1-t)(S^H)^{-1} B S^{-1}] \det[S]\} \\
&= -\log\{\det(S^H S) \det[tI + (1-t)(S^H)^{-1} B S^{-1}]\} \\
&= -\log\{\det(A) \det[tI + (1-t)(S^H)^{-1} B S^{-1}]\} \\
&= -\log(\det(A)) - \log(\det[tI + (1-t)(S^H)^{-1} B S^{-1}])
\end{aligned}$$

(c) We need the following facts:

1. If λ_i is an eigenvalue of M, then $t + (1-t)\lambda_i$ is an eigenvalue of $tI + (1-t)M$.
2. $\det(M) = \prod_i \lambda_i$, where each λ_i is an eigenvalue of M.

Hence, $\det[tI + (1-t)(S^H)^{-1} B S^{-1}] = \prod_{i=1}^n t + (1-t)\lambda_i$. Therefore we have

$$g(t) = -\log(\det(A)) - \log(\prod_{i=1}^n t + (1-t)\lambda_i) = -\log(\det(A)) - \sum_{i=1}^n \log(t + (1-t)\lambda_i).$$

(d)

$$g'(t) = \sum_{i=1}^n -\frac{1-\lambda_i}{t + (1-t)\lambda_i}.$$

$$g''(t) = \sum_{i=1}^n -\frac{(1-\lambda_i)(\lambda_i - 1)}{(t + (1-t)\lambda_i)^2} = \frac{(\lambda_i - 1)^2}{(t + (1-t)\lambda_i)^2} \geq 0.$$

Since $g''(t) \geq 0, \forall t \in [0,1]$, $g(t)$ is convex. So $f(X)$ is convex. □

Exercise 7.13

Proof. By contradiction, assume f is not constant. Then there exist $a \neq b$ such that $f(a) \neq f(b)$. Without loss of generality we assume $a < b$ and $f(a) < f(b)$. Now pick any point c such that $c > b$. Let $\lambda = \frac{c-b}{c-a}, 1 - \lambda = \frac{b-a}{c-a}$. Observe that $\lambda a + (1-\lambda)c = b$. Since f is a convex function, it follows that $\lambda f(a) + (1-\lambda) f(c) \geq f(\lambda a + (1-\lambda)c) = f(b)$. So we have

$$\begin{aligned}
f(c) &\geq \frac{f(b) - \lambda f(a)}{1-\lambda} = \frac{f(b) - \frac{c-b}{c-a} f(a)}{\frac{b-a}{c-a}} = \frac{(c-a)f(b) - (c-b)f(a)}{b-a} \\
&= \frac{(c-a)(f(b) - f(a)) + (b-a)f(a)}{b-a} \\
&= f(a) + (c-a)\frac{f(b) - f(a)}{b-a}
\end{aligned}$$

Let $c \to \infty$, we see that $f(c) \to \infty$. This is contradicted to the fact that f is bounded above. Hence f must be a constant function. □

Exercise 7.20

Proof. Since f is convex and $-f$ is convex, we have $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \forall \lambda \in [0,1]$,

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \tag{1}$$

$$-f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq -\lambda f(\mathbf{x}) - (1-\lambda) f(\mathbf{y}) \tag{2}$$

Multiply the second equation by (-1),

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).$$

This implies

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

Hence f is affine. □

Exercise 7.21

Proof. We first show that the second claim implies the first claim.
Let Ω denote the feasible set. Suppose $\mathbf{x}^* \in \Omega$ is a local minimizer of $f(\mathbf{x})$, then in this neighborhood, $\forall \mathbf{x}, f(\mathbf{x}) \geq f(\mathbf{x}^*)$. Since ϕ is a strictly increasing function, it follows that $\phi(f(\mathbf{x})) \geq \phi(f(\mathbf{x}^*))$. So $\mathbf{x}^*$ is a local minimizer of $\phi(f(\mathbf{x}))$.
Then we show that the first claim implies the second claim.
Suppose $\mathbf{x}^*$ is a local minimizer of $\phi(f(\mathbf{x}))$. Then $\forall \mathbf{x}, f(\mathbf{x}) \geq f(\mathbf{x}^*)$ in its neighborhood. By definition this means that $\mathbf{x}^*$ is a local minimizer of $f(\mathbf{x})$. □