Problem Set #4

Reiko Laski

Exercise 6.6

$f(x,y) = 3x^2y + 4xy^2 + xy$

$$Df(x,y) = \begin{bmatrix} 6xy + 4y^2 + y \\ 3x^2 + 8xy + x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{aligned} \implies x &= -\frac{1}{3}, & y &= 0 \\ x &= -\frac{1}{9}, & y &= -\frac{1}{12} \\ x &= 0, & y &= -\frac{1}{4} \\ x &= 0, & y &= 0 \end{aligned}$$

$$D^2f(x,y) = \begin{bmatrix} 6y & 6x + 8y + 1 \\ 6x + 8y + 1 & 8x \end{bmatrix}$$

$$D^2f\Big(-\frac{1}{3},\ 0\Big) = \begin{bmatrix} 0 & -1 \\ -1 & -\frac{8}{3} \end{bmatrix} \implies \lambda = -\frac{1}{3},\ 3$$

$$D^2f\Big(-\frac{1}{9},\ -\frac{1}{12}\Big) = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{8}{9} \end{bmatrix} \implies \lambda = -1.08,\ -0.31$$

$$D^2f\Big(0,\ -\frac{1}{4}\Big) = \begin{bmatrix} -\frac{3}{2} & -1 \\ -1 & 0 \end{bmatrix} \implies \lambda = \frac{1}{2},\ -2$$

$$D^2f(0,0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \implies \lambda = -1,\ 1$$

$(-\frac{1}{3},\ 0) \implies$ saddle point
$(-\frac{1}{9},\ -\frac{1}{12}) \implies$ maximizer
$(0,\ -\frac{1}{4}) \implies$ saddle point
$(0,\ 0) \implies$ saddle point

Exercise 6.7

(i) For any square matrix A the matrix $Q = A^T + A$ is symmetric, and $\mathbf{x}^TQ\mathbf{x} = \mathbf{x}^TA^T\mathbf{x} + \mathbf{x}^TA\mathbf{x} = 2\mathbf{x}^TA\mathbf{x}$. Then

$$\mathbf{x}^TA\mathbf{x} - \mathbf{b}^T\mathbf{x} + c = \frac{1}{2}\mathbf{x}^TQ\mathbf{x} - \mathbf{b}^T\mathbf{x} + c.$$

Proof:

Let A be a square matrix. Then $Q = A + A^T \implies Q^T = (A + A^T)^T = A + A^T = Q$, so $Q = A + A^T$ is symmetric. Then

$$\begin{aligned} \mathbf{x}^TQ\mathbf{x} &= \mathbf{x}^T(A + A^T)\mathbf{x} \\ &= \mathbf{x}^TA\mathbf{x} + \mathbf{x}^TA^T\mathbf{x} \\ &= \mathbf{x}^TA\mathbf{x}. \end{aligned}$$

By substitution, we can show that the objective function of the problem is

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

(ii) Any minimizer $\mathbf{x}^*$ of f is a solution of the equation

$$Q^T\mathbf{x}^* = \mathbf{b}.$$

Proof:
This follows from the FONC: $f'(\mathbf{x}) = Q^T\mathbf{x} - b$.

(iii) The quadratic minimization problem will have a unique solution if and only if Q is positive definite, and in that cause, the mnimizer is the solution of the linear system in part (i).

Exercise 6.11
Consider the quadratic function $f(x) = ax^2 + bx + c$, where $a > 0$, and $b, c \in \mathbb{R}$. For any initial guess $x_0 \in \mathbb{R}$, one iteration of Newton's method lands at the unique minimizer of f.
Proof:
As shown in the textbook, for an intial guess x_0 we have that

$$\begin{aligned} q(x_0) &= f(x_0) = ax_0^2 + bx_0 + c \\ q'(x_0) &= f'(x_0) = 2ax_0 + b \\ q''(x_0) &= f''(x_0) = 2a \end{aligned}$$

Solving for x_1, we find that

$$x_1 = x_0 - \frac{2ax_0 + b}{2a} = x_0 - x_0 - \frac{b}{2a} = -\frac{b}{2a}.$$

Recall that this is the formula for the minimizer of a parabola, and thus Newton's method solved for the unique minimizer in one iteration.

Exercise 7.1
If S is a nonempty subset of V, then $conv(S)$ is convex.
Proof:
Let $\mathbf{x}, \mathbf{y} \in S$ s.t.

$$\mathbf{x} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \in S \tag{1}$$
$$\mathbf{y} = \beta_1\mathbf{x}_1 + \cdots + \beta_k\mathbf{x}_k \in S \tag{2}$$

where $x_1, ..., x_k \in S, \sum_{i=1}^k \alpha_i = 1, \sum_{i=1}^k \beta_i = 1$. Multiplying equation (1) by λ and equation (2) by $1 - \lambda$, we get

$$\begin{aligned} \lambda\mathbf{x} &= \lambda\alpha_1\mathbf{x}_1 + \cdots + \lambda\alpha_k\mathbf{x}_k \\ (1-\lambda)\mathbf{y} &= (1-\lambda)\beta_1\mathbf{x}_1 + \cdots + (1-\lambda)\beta_k\mathbf{x}_k \end{aligned}$$

where $\sum_{i=1}^{k} \lambda\alpha_i = \lambda \sum_{i=1}^{k} \alpha_i = \lambda$ and $\sum_{i=1}^{k}(1-\lambda)\beta_i = (1-\lambda)\sum_{i=1}^{k} \beta_i = 1-\lambda$. Then we know that

$$\begin{aligned}&\lambda\alpha_1\mathbf{x}_1 + \cdots + \lambda\alpha_k\mathbf{x}_k + (1-\lambda)\beta_1\mathbf{x}_1 + \cdots + (1-\lambda)\beta_k\mathbf{x}_k \\ &= \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S\end{aligned}$$

This shows that $conv(S)$ is convex.

Exercise 7.2
(i) A hyperplane is convex.
Proof:
Let P be a hyperplane and let $\mathbf{x}, \mathbf{y} \in P$, i.e. $\langle \mathbf{a}, \mathbf{x} \rangle = b$ and $\langle \mathbf{a}, \mathbf{y} \rangle = b$. Then for some $0 \le \lambda \le 1$,

$$\begin{aligned}\langle \mathbf{a}, \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \rangle &= \langle \mathbf{a}, \lambda\mathbf{x} \rangle + \langle \mathbf{a}, (1-\lambda)\mathbf{y} \rangle \\ &= \lambda\langle \mathbf{a}, \mathbf{x} \rangle + (1-\lambda)\langle \mathbf{a}, \mathbf{y} \rangle \\ &= \lambda b + (1-\lambda)b \\ &= b\end{aligned}$$

which means that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in P$.

(ii) A half space is convex.
Proof:
Let H be a half space and let $\mathbf{x}, \mathbf{y} \in H$, i.e. $\langle \mathbf{a}, \mathbf{x} \rangle = c$ and $\langle \mathbf{a}, \mathbf{y} \rangle = d$ where $c, d \le b$. Then for some $0 \le \lambda \le 1$,

$$\begin{aligned}\langle \mathbf{a}, \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \rangle &= \langle \mathbf{a}, \lambda\mathbf{x} \rangle + \langle \mathbf{a}, (1-\lambda)\mathbf{y} \rangle \\ &= \lambda\langle \mathbf{a}, \mathbf{x} \rangle + (1-\lambda)\langle \mathbf{a}, \mathbf{y} \rangle \\ &= \lambda c + (1-\lambda)d \\ &\le \lambda b + (1-\lambda)b \\ &= b\end{aligned}$$

which means that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in H$.

Exercise 7.4
Let $C \subset \mathbb{R}^n$ be nonempty, closed, and convex. A point $\mathbf{p} \in C$ is the projection of $\mathbf{x}$ onto C if and only if

$$\langle \mathbf{x} - \mathbf{p}, \mathbf{p} - \mathbf{y} \rangle \ge 0, \ \forall \mathbf{y} \in C. \tag{3}$$

Proof:
First, prove the following statements:
(i) $\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x} - \mathbf{p}\|^2 + \|\mathbf{p} - \mathbf{y}\|^2 + 2\langle \mathbf{x} - \mathbf{p}, \mathbf{p} - \mathbf{y} \rangle$.

Subproof:

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

(ii) If the equality in (i) holds, then $\|\mathbf{x}-\mathbf{y}\| > \|\mathbf{x}-\mathbf{p}\|$ for all $\mathbf{y}\in C, \mathbf{y}\neq\mathbf{p}$.
Subproof:
If the equality in part (i) holds, then $\langle\mathbf{x}-\mathbf{p},\mathbf{p}-\mathbf{y}\rangle\geq 0$. Then

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

(iii) If $\mathbf{z}=\lambda\mathbf{y}+(1-\lambda)\mathbf{p}$, where $0\leq\lambda\leq 1$, then

$$\|\mathbf{x}-\mathbf{z}\|^2=\|\mathbf{x}-\mathbf{p}\|^2+2\lambda\langle\mathbf{x}-\mathbf{p},\mathbf{p}-\mathbf{y}\rangle+\lambda^2\|\mathbf{y}-\mathbf{p}\|^2.$$

Subproof:

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

(iv) If $\mathbf{p}$ is a projection of $\mathbf{x}$ onto the convex set C, then $\langle\mathbf{x}-\mathbf{p},\mathbf{p}-\mathbf{y}\rangle\geq 0$ for all $\mathbf{y}\in C$. *Subproof:*
From part (iii), we know that

$$\begin{aligned}&\|\mathbf{x}-\mathbf{z}\|^2=\|\mathbf{x}-\mathbf{p}\|^2+2\lambda\langle\mathbf{x}-\mathbf{p},\mathbf{p}-\mathbf{y}\rangle+\lambda^2\|\mathbf{y}-\mathbf{p}\|^2\\ &\implies \|\mathbf{x}-\mathbf{z}\|^2-\|\mathbf{x}-\mathbf{p}\|^2=2\lambda\langle\mathbf{x}-\mathbf{p},\mathbf{p}-\mathbf{y}\rangle+\lambda^2\|\mathbf{y}-\mathbf{p}\|^2\end{aligned}$$

Since $\mathbf{p}=proj_C\mathbf{x}$ and $\mathbf{z}\in C$,

$$\begin{aligned}&\|\mathbf{x}-\mathbf{p}\|\leq\|\mathbf{x}-\mathbf{z}\|\\ &\implies \|\mathbf{x}-\mathbf{p}\|^2\leq\|\mathbf{x}-\mathbf{z}\|^2\\ &\implies 0\leq\|\mathbf{x}-\mathbf{z}\|^2-\|\mathbf{x}-\mathbf{p}\|^2\end{aligned}$$

Therefore,

$$0\leq 2\langle\mathbf{x}-\mathbf{p},\mathbf{p}-\mathbf{y}\rangle+\lambda\|\mathbf{y}-\mathbf{p}\|^2,\quad\forall\mathbf{y}\in C,\lambda\in[0,1].$$

Exercise 7.8
If $f:\mathbb{R}^m\to\mathbb{R}$ is convex, if $A\in M_{m\times n}(\mathbb{R})$, and if $\mathbf{b}\in\mathbb{R}^m$, then the function

$g : \mathbb{R}^m \to \mathbb{R}$ given by $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$ is convex.
Proof:
Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}$. Then we know that

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$$

We want to know if g is convex:

$$\begin{aligned} g(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) &= f(A(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + \mathbf{b}) \\ &= f(A\lambda \mathbf{x}_1 + A(1-\lambda)\mathbf{x}_2 + \mathbf{b}) \\ &= f(\lambda(A\mathbf{x}_1 + \mathbf{b}) + (1-\lambda)(A\mathbf{x}_2 + \mathbf{b})) \\ &\leq \lambda f(A\mathbf{x}_1 + \mathbf{b}) + (1-\lambda) f(A\mathbf{x}_2 + \mathbf{b}) \\ &= \lambda g(\mathbf{x}_1) + (1-\lambda) g(\mathbf{x}_2) \end{aligned}$$

Exercise 7.12
(i) The set $PD_n(\mathbb{R})$ of positive-definite matrices in $M_n(\mathbb{R})$ is convex.
Proof:
Let $A, B \in PD_n(\mathbb{R})$. Then for all $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^T A \mathbf{x} > 0 \text{ and } \mathbf{x}^T B \mathbf{x} > 0$$

We want to know if $\lambda A + (1-\lambda)B \in PD_n(\mathbb{R})$:

$$\mathbf{x}^T(\lambda A + (1-\lambda)B)\mathbf{x} = \lambda \mathbf{x}^T A \mathbf{x} + (1-\lambda)\mathbf{x}^T B \mathbf{x} > 0$$

Therefore the set $PD_n(\mathbb{R})$ is convex.

(ii) The function $f(X) = -\log(\det(X))$ is convex on $PD_n(\mathbb{R})$.
Proof: To do this, we must prove the following:
(a) The function f is convex if for every $A, B \in PD_n(\mathbb{R})$ the function $g(t) : [0,1] \to \mathbb{R}$ given by $g(t) = f(tA + (1-t)B)$ is convex.
Subproof: Let g be convex. Then we know that

$$\begin{aligned} & g(\lambda t_1 + (1-\lambda)t_2) \leq \lambda g(t_1) + (1-\lambda) g(t_2) \\ &\implies f[(\lambda t_1 + (1-\lambda)t_2)A + (1 - (\lambda t_1 + (1-\lambda)t_2))B] \\ &\qquad\qquad \leq \lambda f(t_1 A + (1-t_1)B) + (1-\lambda) f(t_2 A + (1-t_2)B) \\ &\implies f[(\lambda t_1 + (1-\lambda)t_2)A + (1 - (\lambda t_1 + (1-\lambda)t_2))B] \\ &\qquad\qquad \leq \lambda f(t_1 A + (1-t_1)B) + (1-\lambda) f(t_2 A + (1-t_2)B) \end{aligned}$$

Expanding LHS of the equation,

$$\begin{aligned} & f[(\lambda t_1 + (1-\lambda)t_2)A + (1 - (\lambda t_1 + (1-\lambda)t_2))B] \\ &= f[(\lambda t_1 + (1-\lambda)t_2)A + (\lambda + (1-\lambda) - (\lambda t_1 + (1-\lambda)t_2))B] \\ &= f[\lambda t_1 A + (1-\lambda)t_2 A + \lambda B + (1-\lambda)B - \lambda t_1 B + (1-\lambda)t_2 B] \\ &= f[\lambda t_1 A + (1-\lambda)t_2 A + \lambda(1-t_1)B + (1-\lambda)(1-t_2)B] \\ &= f[\lambda(t_1 A + (1-t_1)B) + (1-\lambda)(t_2 A + (1-t_2)B)] \end{aligned}$$

The inequlity above becomes

$$f[\lambda(t_1A+(1-t_1)B)+(1-\lambda)(t_2A+(1-t_2)B)] \\ \leq \lambda f(t_1A+(1-t_1)B)+(1-\lambda)f(t_2A+(1-t_2)B)$$

Letting $X = t_1A + (1-t_1)B$ and $Y = t_2A + (1-t_2)B$, we get

$$f[\lambda X + (1-\lambda)Y] \leq \lambda f(X) + (1-\lambda) f(Y)$$

and we can see that f is a convex function.

(b) There is an S such that $S^HS = A$ and

$$\begin{aligned} g(t) &= -\log(\det(S^H(tI+(1-t)(S^H)^{-1}BS^{-1})S)) \\ &= -\log(\det(A)) - \log(\det(tI+(1-t)(S^H)^{-1}BS^{-1})S)) \end{aligned}$$

Subproof:
Since A is positive definite, we know that there exists an S such that $A = S^HS$.

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

Then

$$\begin{aligned} g(t) &= -\log(\det(S^HS)\det(tI+(1-t)(S^H)^{-1}BS^{-1})) \\ &= -\log(\det(S^HS)) - \log(\det(tI+(1-t)(S^H)^{-1}BS^{-1})) \\ &= -\log(\det(A)) - \log(\det(tI+(1-t)(S^H)^{-1}BS^{-1})) \end{aligned}$$

(c)

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

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of $(S^H)^{-1}BS^{-1}$.
Subproof: Let $\lambda_1, ..., \lambda_n$ be the eigenvalues of $(S^H)^{-1}BS^{-1}$. Then we know that

$$\det(tI+(1-t)(S^H)^{-1}BS^{-1}) = \prod_{i=1}^{n}(t+(1-t)\lambda_i)$$

and the last equality in part (b) becomes

$$\begin{aligned} g(t) &= -\log(\det(tI+(1-t)(S^H)^{-1}BS^{-1})) - \log(\det(A)) \\ &= -\log(\prod_{i=1}^{n}(t+(1-t)\lambda_i)) - \log(\det(A)) \\ &= -\sum_{i=1}^{n}\log(t+(1-t)\lambda_i) - \log(\det(A)) \end{aligned}$$

(d) $g''(t) \geq 0$ for all $t \in [0,1]$.
Subproof:

$$\begin{aligned}
g(t) &= -\sum_{i=1}^{n} \log(t + (1-t)\lambda_i) - \log(\det(A)) \\
\implies g'(t) &= -\sum_{i=1}^{n} \frac{1-\lambda_i}{t+(1-t)\lambda_i} \\
\implies g''(t) &= -\sum_{i=1}^{n} \frac{-(1-\lambda)^2}{(t+(1-t)\lambda_i)^2} = \sum_{i=1}^{n} \frac{(1-\lambda)^2}{(t+(1-t)\lambda_i)^2} \geq 0
\end{aligned}$$

Exercise 7.13
If $f : \mathbb{R}_n \to \mathbb{R}$ is convex and bounded above, then f is constant.
Proof:
On the contrary, suppose f is not constant. WLOG, assume that $f(\mathbf{x}_1) < f(\mathbf{x}_3)$ for some $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. Since f is convex, we know that for there exists an $\mathbf{x}_2$ s.t.

$$\begin{aligned}
f(\mathbf{x}_2) = f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_3) &\leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_3) \\
&< f(\mathbf{x}_3)
\end{aligned}$$

Since f is bounded above by hypothesis, we know that there exists some $M \in \mathbb{R}$ s.t. $f(\mathbf{x}) \leq M$ for all $\mathbf{x} \in \mathbb{R}^n$. Letting $f(\mathbf{x}_2) = M$, we know that $f(\mathbf{x}_3) > M$, which is a contradiction. Therefore, f must be constant.

Exercise 7.20
If $f : \mathbb{R}_n \to \mathbb{R}$ is convex and $-f$ is also convex, then f is affine.
Proof:
Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. Since f is convex,

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$$

Since $-f$ is also convex,

$$\begin{aligned}
&- f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq -\lambda f(\mathbf{x}_1) - (1-\lambda) f(\mathbf{x}_2) \\
&\implies f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \geq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)
\end{aligned}$$

Therefore,

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) = \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$$

We can define another function $g(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{0})$, which has the property $g(\mathbf{0}) = 0$.

$$\begin{aligned}
g(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) &= f((\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)) - f(\mathbf{0}) \\
&= \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) - \lambda f(\mathbf{0}) - (1-\lambda) f(\mathbf{0}) \\
&= \lambda (f(\mathbf{x}_1) - f(\mathbf{0})) + (1-\lambda)(f(\mathbf{x}_2) - f(\mathbf{0})) \\
&= \lambda g(\mathbf{x}_1) + (1-\lambda) g(\mathbf{x}_2)
\end{aligned}$$

To show that f is affine, it suffices to show that g is linear. It is obvious that $g(\lambda\mathbf{x}) = \lambda g(\mathbf{x})$ for $\lambda \in [0,1]$. For $\lambda > 1$, note that $\frac{1}{\lambda} \in (0,1)$.

$$\begin{aligned}
g(\mathbf{x}) &= g\Big(\lambda\frac{1}{\lambda}\mathbf{x} + \Big(1-\frac{1}{\lambda}\Big)(\mathbf{0})\Big) \\
&= \frac{1}{\lambda}g(\lambda\mathbf{x}) + (1-\frac{1}{\lambda})g(\mathbf{0}) \\
&= \frac{1}{\lambda}g(\lambda\mathbf{x})
\end{aligned}$$

$$\implies \lambda g(\mathbf{x}) = g(\lambda\mathbf{x})$$

We use the above equality to show that $g(\mathbf{x}_1 + \mathbf{x}_2) = g(\mathbf{x}_1) + g(\mathbf{x}_2)$.

$$\begin{aligned}
g(\mathbf{x}_1 + \mathbf{x}_2) &= g\Big(\frac{1}{2}(2\mathbf{x}_1 + \frac{1}{2}(2\mathbf{x}_2)\Big) \\
&= \frac{1}{2}g(2\mathbf{x}_1) + \frac{1}{2}g(2\mathbf{x}_2) \\
&= g(\mathbf{x}_1) + g(\mathbf{x}_2)
\end{aligned}$$

This shows that g is linear $\implies$ f is affine.

Exercise 7.21

If $D \subset \mathbb{R}$ with $f : \mathbb{R}_n \to D$, and if $\phi : D \to \mathbb{R}$ is a strictly increasing function, then $\mathbf{x}^*$ is a local minimizer for the problem

$$\begin{array}{ll}
\text{minimize} & \phi \circ f(\mathbf{x}) \\
\text{subject to} & G(\mathbf{x}) \preceq \mathbf{0} \\
 & H(\mathbf{x}) = \mathbf{0}
\end{array}$$

if and only if $\mathbf{x}^*$ is a local minimizer for the problem

$$\begin{array}{ll}
\text{minimize} & f(\mathbf{x}) \\
\text{subject to} & G(\mathbf{x}) \preceq \mathbf{0} \\
 & H(\mathbf{x}) = \mathbf{0}
\end{array}$$

Proof:

Let $\mathbf{x}^*$ be a minimizer of f. Then $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x}$ in an open neighborhood U around $\mathbf{x}^*$. Since ϕ is strictly increasing, we know that $\phi(f(\mathbf{x}^*)) \leq \phi(f(\mathbf{x}))$ for all $\mathbf{x} \in U$. Therefore $\mathbf{x}^*$ is a minimizer of $\phi \circ f$.

Now let $\mathbf{x}^*$ be a minimizer of $\phi \circ f$. Then $\phi(f(\mathbf{x}^*)) \leq \phi(f(\mathbf{x}))$ for all $\mathbf{x}$ in an open neighborhood V. Since ϕ is strictly increasing, we know that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in V$. Therefore $\mathbf{x}^*$ is a minimizer of f.