2.6

(a) $f(X) = a^T X b \quad X \in \mathbb{R}^{m\times n} \quad a \in \mathbb{R}^m,\ b \in \mathbb{R}^n$

任取 $V \in \mathbb{R}^{m\times n},\ t \in \mathbb{R} \quad X + tV \in \mathbb{R}^{m\times n}$

$$\lim_{t\to 0} \frac{f(X+tV) - f(X)}{t} = \lim_{t\to 0} \frac{a^T(X+tV)b - a^T X b}{t}$$

$$= a^T V b = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j V_{ij} = \langle S, V\rangle,$$ 其中 $S \in \mathbb{R}^{m\times n}$，且 $(S)_{ij} = a_i b_j$

(由矩阵内积定义)

而易知 $ab^T = \begin{pmatrix} a_1b_1 & \cdots & a_1b_n \\ \vdots & \ddots & \vdots \\ a_mb_1 & \cdots & a_mb_n \end{pmatrix} = S$ 故有 $\nabla f(X) = ab^T$.

(b) $f(X) = tr(X^T A X),\quad X \in \mathbb{R}^{m\times n} \quad A \in \mathbb{R}^{m\times m}$

任取 $V \in \mathbb{R}^{m\times n},\ t \in \mathbb{R}$ 则 $X + tV \in \mathbb{R}^{m\times n}$

$$\lim_{t\to 0} \frac{f(X+tV) - f(X)}{t} = \lim_{t\to 0} \frac{tr((X+tV)^T A(X+tV)) - tr(X^T A X)}{t}$$

$$= \lim_{t\to 0} \frac{tr[(X^T + tV^T)(AX + tAV)] - tr(X^T A X)}{t}$$

$$= \lim_{t\to 0} \frac{tr[X^T A X + t(V^T A X + X^T A V) + t^2 V^T A V] - tr(X^T A X)}{t}$$

$$= \lim_{t\to 0} \frac{tr[t(V^T A X + X^T A V) + t^2 V^T A V]}{t}$$ (tr(·) 线性)

(tr(·) 线性)

$$= tr(V^T A X + X^T A V)$$

$$= tr(X^T(A^T + A)V) = tr(V[X^T(A^T + A)])$$

$$= \langle (A + A^T)X, V\rangle$$ 故 $\nabla f(X) = (A + A^T)X$.

(c) $f(X) = \ln(\det(X))$, $X \in \mathbb{R}^{n\times n}$ 且 $\{X \mid \det(X) > 0\}$

直接计算 Frechet 导数. 对 $1\le i\le n$, $1\le j\le n$

由于 $\det X = \sum_{j=1}^{n} (-1)^{i+j} x_{ij} M_{ij}$, ($M_{ij}$ 为 x_{ij} 余子式) 为关于 x_{ij} 一次函数

有 $\frac{\partial f}{\partial x_{ij}} = \frac{(-1)^{i+j} M_{ij}}{\det X}$　　记 $B = \left(\frac{\partial f}{\partial x_{ij}}\right)_{n\times n} = \nabla f(x)$

$$B^T = \frac{1}{\det X}\left((-1)^{i+j} M_{ji}\right)_{n\times n}$$

$$(XB^T)_{ij} = \frac{1}{\det X}\sum_{k=1}^{n} (-1)^{k+j} x_{ik} M_{jk} = \begin{cases} 1 & i=j \\ 0 & i\ne j \end{cases} \quad \text{(线性代数)}$$

$$(B^T X)_{ij} = \frac{1}{\det X}\sum_{k=1}^{n} (-1)^{i+k} M_{ki} x_{kj} = \begin{cases} 1 & i=j \\ 0 & i\ne j \end{cases} \quad \text{(线性代数)}$$

$XB^T = B^T X = I_n$ 故 $B^T = X^{-1}$ 　$B = (X^{-1})^T$

$$\nabla f(x) = B = (X^{-1})^T.$$

2.9

(a) $f(x) = \ln \sum_{k=1}^{n} e^{x_k}$ 　　$f \in C^2(\mathbb{R}^n)$

$$\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{n\times n}$$

$i\ne j$ 时, $\frac{\partial^2 f}{\partial x_i \partial x_j} = \left(\frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}}\right)_{x_j} = -\frac{e^{x_i} e^{x_j}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2}$

$i=j$ 时, $\frac{\partial^2 f}{\partial x_i^2} = \left(\frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}}\right)_{x_i} = \frac{e^{x_i}\left(\sum_{k=1}^{n} e^{x_k} - e^{x_i}\right)}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2}$

$$\nabla^2 f(x) = \frac{1}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2}\begin{pmatrix} e^{x_1}(e^{x_2}+\cdots+e^{x_n}) & -e^{x_1}e^{x_2} & \cdots & -e^{x_1}e^{x_n} \\ \vdots & \ddots & & \vdots \\ -e^{x_1}e^{x_n} & \cdots & \cdots & e^{x_n}(e^{x_1}+\cdots+e^{x_{n-1}}) \end{pmatrix}$$

故 $\nabla^2 f(x)$ 是对角占优对称阵 故 $\nabla^2 f(x) \succeq 0$ 故 f 是凸函数

(b) $f(x) = \sqrt[n]{x_1 \cdots x_n}$, $x > 0$ 有 $f \in C^2((\mathbb{R}_+)^n)$

有 $\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{n \times n} = \frac{\sqrt[n]{x_1 \cdots x_n}}{n^2} \begin{pmatrix} \frac{(1-n)}{x_1^2} & \frac{1}{x_1 x_2} & \cdots & \frac{1}{x_1 x_n} \\ \vdots & \ddots & & \vdots \\ \frac{1}{x_n x_1} & \frac{1}{x_n x_2} & \cdots & \frac{(1-n)}{x_n^2} \end{pmatrix} = \frac{f(x)}{n^2} B$

设 $y = (y_1 \cdots y_n)^T \in \mathbb{R}^n$

则 $y^T B y = \sum_{i=1}^n \sum_{j=1}^n \frac{y_i y_j}{x_i x_j} - n \sum_{i=1}^n \frac{y_i^2}{x_i^2} = \left(\sum_{i=1}^n \frac{y_i}{x_i}\right)^2 - n \sum_{i=1}^n \frac{y_i^2}{x_i^2} \leq 0$

(Cauchy – Schwarz)

有 $-\nabla^2 f(x) \succeq 0$ $f(x)$ 为凹函数

(c) $f(x) = \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$, $p \in (0,1)$, $x > 0$

有 $f(x) \in C^2(\mathbb{R}_+^n)$ 记 $s = \sum_{i=1}^n x_i^p$

计算 $\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{n \times n} = (1-p)\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}-2} \begin{pmatrix} x_1^{2p-2} - x_1^{p-2} s & x_1^{p-1} x_2^{p-1} & \cdots & x_1^{p-1} x_n^{p-1} \\ \vdots & \ddots & & \vdots \\ x_n^{p-1} x_1^{p-1} & \cdots & & x_n^{2p-2} - x_n^{p-2} s \end{pmatrix}$

$= (1-p)\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}-2} B$

记 $y = (y_1 \cdots y_n)^T \in \mathbb{R}^n$

则 $y^T B y = \sum_{i=1}^n \sum_{j=1}^n x_i^{p-1} x_j^{p-1} y_i y_j - \sum_{i=1}^n x_i^{p-2} y_i^2 s$

$= \left(\sum_{i=1}^n x_i^{p-1} y_i\right)^2 - \sum_{i=1}^n x_i^{p-2} y_i^2 \sum_{i=1}^n x_i^p \leq 0$ (Cauchy – Schwartz)

故 $B \preceq 0$ 有 $-\nabla^2 f(x) = (1-p)\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}-2} B \succeq 0$

$f(x)$ 为凹函数