

1. Gradients and Hessians

(a) 首先计算 $f(x) = \frac{1}{2}x^T Ax + b^T x$

$$\begin{aligned} f(x) &= \frac{1}{2}x^T Ax + b^T x \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i A_{ij} x_j + \sum_{i=1}^n b_i x_i \end{aligned}$$

接着计算 $\frac{\partial f(x)}{\partial x_k}$, 注意 A 为对称矩阵, 记 A 的第 k 行为 A_k

$$\begin{aligned} \frac{\partial f(x)}{\partial x_k} &= \frac{1}{2} \sum_{i=1}^n x_i A_{ik} + \frac{1}{2} \sum_{j=1}^n A_{kj} x_j + b_k \\ &= \sum_{i=1}^n x_i A_{ik} + b_k \\ &= A_k x + b_k \end{aligned}$$

所以

$$\nabla f(x) = \begin{bmatrix} A_1 x + b_1 \\ \dots \\ A_n x + b_n \end{bmatrix} = Ax + b$$

(b) 计算 $\frac{\partial f(x)}{\partial x_k}$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial g(h(x))}{\partial x_k} = \frac{\partial g(h(x))}{\partial h(x)} \frac{\partial h(x)}{\partial x_k} = g'(h(x)) \frac{\partial h(x)}{\partial x_k}$$

所以

$$\nabla f(x) = \begin{bmatrix} g'(h(x)) \frac{\partial h(x)}{\partial x_1} \\ \dots \\ g'(h(x)) \frac{\partial h(x)}{\partial x_n} \end{bmatrix} = g'(h(x)) \nabla h(x)$$

(c) 接着(a)计算 $\nabla^2 f(x)$, 我们计算 $\frac{\partial^2 f(x)}{\partial x_l \partial x_k}$

$$\frac{\partial^2 f(x)}{\partial x_l \partial x_k} = \frac{\partial (\sum_{i=1}^n x_i A_{ik} + b_k)}{\partial x_l} = A_{lk}$$

所以

$$\nabla^2 f(x) = A$$

(d) 记 $h(x) = a^T x$, 所以 $f(x) = g(h(x))$, 所以利用(b)计算 $\nabla f(x)$,

$$\begin{aligned}\frac{\partial h(x)}{\partial x_k} &= a_k \\ \nabla h(x) &= a \\ \frac{\partial f(x)}{\partial x_k} &= g'(a^T x) a_k \\ \nabla f(x) &= g'(a^T x) a\end{aligned}$$

接着计算 $\frac{\partial^2 f(x)}{\partial x_l \partial x_k}$

$$\begin{aligned}\frac{\partial^2 f(x)}{\partial x_l \partial x_k} &= \frac{\partial(g'(a^T x) a_k)}{\partial x_l} \\ &= a_k \frac{\partial(g'(a^T x))}{\partial(a^T x)} \frac{\partial(a^T x)}{\partial x_l} \\ &= g''(a^T x) a_l a_k\end{aligned}$$

所以

$$\nabla^2 f(x) = g''(a^T x) a a^T$$

2. Positive definite matrices

(a)任取 $x \in \mathbb{R}^n$, 那么

$$x^T A x = x^T z z^T x = (z^T x)^T (z^T x) \geq 0$$

(b)考虑 A 的零空间, 任取 $x \in N(A)$, 那么

$$\begin{aligned}Ax &= z z^T x = 0 \\ \text{两边左乘 } x^T \text{ 可得} \\ x^T z z^T x &= 0 \\ (z^T x)^T (z^T x) &= 0 \\ z^T x &= 0\end{aligned}$$

这说明 $x \in N(z^T)$ 。反之, 任取 $x \in N(z^T)$, 那么

$$\begin{aligned}z^T x &= 0 \\ z z^T x &= 0\end{aligned}$$

从而 $x \in N(A)$, 因此

$$N(A) = N(z^T)$$

因为 $z \in \mathbb{R}^n$, 所以 $\text{rank}(z) \leq 1$, 因为 z 非零, 所以 $\text{rank}(z) \geq 1$, 从而 $\text{rank}(z) = 1$, 利用这个结论以及 $N(A) = N(z^T)$ 来计算 $\text{rank}(A)$

$$\begin{aligned}\operatorname{rank}(N(A)) &= \operatorname{rank}(N(z^T)) \\ n - \operatorname{rank}(A) &= n - \operatorname{rank}(z^T) \\ \operatorname{rank}(A) &= \operatorname{rank}(z^T) = \operatorname{rank}(z) = 1\end{aligned}$$

(c)任取 $x \in \mathbb{R}^m$, 那么

$$x^T BAB^T x = (B^T x)^T A (B^T x)$$

记 $z = B^T x$, 结合 A 的半正定性可得

$$x^T BAB^T x = z^T A z \geq 0$$

所以 BAB^T 半正定

3. Eigenvectors, eigenvalues, and the spectral theorem

(a)对 $A = T \Lambda T^{-1}$ 两边右乘 T 可得

$$AT = T\Lambda$$

考虑两边的第 i 列得到

$$At^{(i)} = \lambda_i t^{(i)}$$

所以 A 的特征值即其对应的向量为 $(\lambda_i, t^{(i)})$

(b)注意 U 为正交矩阵, 对 $A = U \Lambda U^T$ 两边右乘 U 可得

$$AU = \Lambda U$$

考虑两边的第 i 列得到

$$Au^{(i)} = \lambda_i u^{(i)}$$

(c)取 x_i , 使得

$$x_i = U \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{i-1 \uparrow 0}^T$$

计算 $x_i^T A x_i$ 可得

$$\begin{aligned}x_i^T A x_i &= \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{i-1 \uparrow 0} U^T U \Lambda U^T U \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{i-1 \uparrow 0}^T \\ &= \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{i-1 \uparrow 0} \Lambda \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{i-1 \uparrow 0}^T \\ &= \lambda_i \\ &\geq 0\end{aligned}$$

所以结论得证。