1.Gradients and Hessians

(a)首先计算$f(x)=\frac{1}{2}x^TAx+b^Tx$

$$
\begin{aligned}
f(x) &= \frac{1}{2}x^TAx+b^Tx \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}x_iA_{ij}x_j+\sum_{i=1}^{n}b_ix_i
\end{aligned}
$$

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$A_k$

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

所以

$$
\nabla f(x)=\begin{bmatrix} A_1x+b_1 \\ \dots \\ A_nx+b_n \end{bmatrix}=Ax+b
$$

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

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

所以

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

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

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

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

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

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

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

(b)考虑A的零空间，任取$x \in N(A)$，那么

$$Ax = zz^T x = 0$$
$$\text{两边左乘}x^T\text{可得}$$
$$x^T z z^T x = 0$$
$$(z^T x)^T (z^T x) = 0$$
$$z^T x = 0$$

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

$$z^T x = 0$$
$$zz^T x = 0$$

从而$x \in N(A)$，因此

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

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

$$\text{rank}(N(A)) = \text{rank}(N(z^T))$$
$$n - \text{rank}(A) = n - \text{rank}(z^T)$$
$$\text{rank}(A) = \text{rank}(z^T) = \text{rank}(z) = 1$$

(c)任取$x \in \mathbb{R}^m$，那么

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

记$z=B^Tx$，结合A的半正定性可得

$$x^TBAB^Tx=z^TAz\ge 0$$

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

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

(b)取x_i，使得

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

计算$x_i^TAx_i$可得

$$\begin{aligned}x_i^TAx_i&=[\underbrace{0,\ldots 0}_{i-1个0},1,0,\ldots,0]U^TU\Lambda U^TU[\underbrace{0,\ldots 0}_{i-1个0},1,0,\ldots,0]^T\\&=[\underbrace{0,\ldots 0}_{i-1个0},1,0,\ldots,0]\Lambda[\underbrace{0,\ldots 0}_{i-1个0},1,0,\ldots,0]^T\\&=\lambda_i\\&\ge 0\end{aligned}$$

所以结论得证。