Problem Set #0

1 [0 points] Gradients and Hessians

(a) $\nabla f(x) = \nabla f_1(x) + \nabla f_2(x)$, where

$$f_1(x) = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} A_{ij} x_i = \sum_{i=1}^{n} \sum_{j=1}^{n} x_j A_{ij} x_i$$

$$f_2(x) = \sum_{i=1}^{n} b_i x_i$$

thus,

$$\begin{aligned}\frac{\partial f_1}{\partial x_k} &= \frac{1}{2}\left(\sum_{1\le j\le n, j\ne k} x_j A_{kj} + \sum_{1\le i\le n, i\ne k} x_i A_{ik} + 2A_{kk}\right) \\ &= \frac{1}{2}\left(\sum_{1\le i\le n} x_i A_{ki} + \sum_{1\le i\le n} x_i A_{ik}\right)\end{aligned}$$

since A is symmetric, we have:

$$\frac{\partial f_1}{\partial x_k} = \sum_{1\le i\le n} x_i A_{ki} = A_{k,:} x$$

finally,

$$\nabla f_1(x) = Ax$$

On the other hand, $\nabla f_2(x) = b$, thus $\nabla f(x) = Ax + b$.

(b) According to the chain rule,

$$\frac{\partial f}{\partial x_k} = \frac{\mathrm{d}f}{\mathrm{d}g} \cdot \frac{\mathrm{d}g}{\mathrm{d}h} \cdot \frac{\partial h}{\partial x_k} = \frac{\mathrm{d}g}{\mathrm{d}h} \cdot \frac{\partial h}{\partial x_k}$$

thus,

$$\nabla f(x) = g'(h(x)) \nabla h(x)$$

(c) According to (a),

$$\frac{\partial f}{\partial x_i} = A_{i,:} x + b_i$$

thus,

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = A_{ij}$$

therefore, $\nabla^2 f(x) = A$.

(d) According to (b) and the chain rule,

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

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = a_i a_j \cdot g''(a^T x)$$

thus, $\nabla f(x) = g''(a^T x)aa^T$ (11 symbols in total).

2 [0 points] Positive definite matrices

(a) Notice that $A^T = (zz^T)^T = zz^T = A$, and $\forall x \in \mathbb{R}^n, x^T Ax = x^T zz^T x = (x^T z)(x^T z)^T = (x^T z)^2 \geq 0$. □

(b) Let $z = [z_1, z_2, \ldots, z_n]^T$, then $A_{ij} = z_i z_j$. Thus,

$$\begin{aligned} Ax = 0 &\Longleftrightarrow \sum_{j=1}^{n} z_i z_j x_j = 0 \\ &\Longleftrightarrow z_i \sum_{j=1}^{n} z_j x_j = 0 \end{aligned}$$

in which i ranges from 1 to n.

Since z is a non-zero vector, $\exists k, z_k \neq 0$. Therefore,

$$Ax = 0 \Longleftrightarrow \sum_{j=1}^{n} z_j x_j = 0 \Longleftrightarrow z^T x = 0$$

$N(A) = \{x | x \in \mathbb{R}^n, Ax = 0\} = \{x | x \in \mathbb{R}^n, z^T x = 0\}$.

$r(A) = r(zz^T) = 1$.

(c) BAB^T is PSD. Proof is as follows.

Firstly, BAB^T is symmetric since $(BAB^T)^T = BA^T B^T = BAB^T$.

Secondly, $\forall x \in \mathbb{R}^m, x^T BAB^T x = (x^T B)A(x^T B)^T \geq 0$. □

3 [0 points] Eigenvectors, eigenvalues, and the spectral theorem

(a) $A = T\Lambda T^{-1} \Longrightarrow AT = T\Lambda$, where

$$AT = [At^{(1)}, At^{(2)}, \ldots, At^{(n)}]$$

$$T\Lambda = [\lambda_1 t^{(1)}, \lambda_2 t^{(2)}, \ldots, \lambda_n t^{(n)}]$$

□

(b) Proof. Notice that $U^T = U^{-1}$. □

(c) Proof. A is PSD $\Longrightarrow$ A is symmetric $\Longrightarrow$ $A = U\Lambda U^T$ and $Au^{(i)} = \lambda_i u^{(i)}$.

According to the properties of PSD and orthogonal matrix,

$$u^{(i)T} A u^{(i)} = \lambda_i ||u^{(i)}||_2^2 = \lambda_i \geq 0$$

$\square$