Problem Set #0: Linear Algebra & Multivariable Calculus

Trung H. Nguyen

1. **Gradients and Hessians**

(a) $f(x) = \dfrac{1}{2}x^T Ax + b^T x$, where A symmetric matrix $\in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$.
We have:

$$x^T Ax = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} A_{ij}x_j = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j$$

$$b^T x = \sum_{i=1}^{n} b_i x_i$$

Thus, $f(x) = \dfrac{1}{2}x^T Ax + b^T x$ has $\nabla f(x) = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{bmatrix}$ where,

$$\begin{aligned}
\frac{\partial f(x)}{\partial x_k} &= \frac{\partial}{\partial x_k}\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j + \sum_{i=1}^{n} b_i x_i\right] \\
&= \frac{\partial}{\partial x_k}\left[\frac{1}{2}\left(\sum_{i\neq k}^{n} A_{ik}x_i x_k + \sum_{j\neq k}^{n} A_{kj}x_k x_j + A_{kk}x_k^2 + \sum_{i\neq k}^{n}\sum_{j\neq k}^{n} A_{ij}x_i x_j\right) + \sum_{i=1}^{n} b_i x_i\right] \\
&= \frac{1}{2}\left(\sum_{i\neq k}^{n} A_{ik}x_i + \sum_{j\neq k}^{n} A_{kj}x_j + 2A_{kk}x_k\right) + b_k \\
&= \sum_{i=1}^{n} A_{ki}x_i + b_k
\end{aligned}$$

or $\nabla f(x) = Ax + b$

(b) $f(x) = g(h(x))$ where $g\colon \mathbb{R} \to \mathbb{R}$ is differentiable and $h\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable. By Chain rule:

$$\nabla f(x) = \nabla g(h(x)) = \begin{bmatrix} \dfrac{\partial g(h(x))}{\partial x_1} \\ \vdots \\ \dfrac{\partial g(h(x))}{\partial x_n} \end{bmatrix}$$

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

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

(c) $f(x) = \dfrac{1}{2} x^T A x + b^T x$, where A symmetric matrix $\in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$. By part (a), we have

$$\frac{\partial f(x)}{\partial x_k} = \sum_{i=1}^{n} A_{ki} x_i + b_k \Rightarrow \frac{\partial^2 f(x)}{\partial x_k \partial x_j} = A_{kj}$$

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

(d) $f(x) = g(a^T x)$, where $g \colon \mathbb{R} \mapsto \mathbb{R}$ continuously differentiable and $b \in \mathbb{R}^n$. By part (b), we have:

$$\nabla f(x) = g'(a^T x) \nabla (a^T x) = g'(a^T x) a \Rightarrow \frac{\partial f(x)}{\partial x_i} = g'(a^T x) a_i$$

Taking second derivative, we have:

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

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

2. **Positive definite matrices**

Let $A \in \mathbb{R}^{n \times n}$, we have:

$$A \succeq 0 \Leftrightarrow \begin{cases} A = A^T \\ x^T A x \geq 0, \forall x \in \mathbb{R}^n \end{cases}$$

$$A \succ 0 \Leftrightarrow \begin{cases} A = A^T \\ x^T A x > 0, \forall x \in \mathbb{R}^n \end{cases}$$

(a) $z \in \mathbb{R}^n, A = zz^T$. We have for $\forall x \in \mathbb{R}^n$:

$$x^T A x = x^T z z^T x = \left(x^T z\right)^2 \geq 0 \Leftrightarrow A \succeq 0$$

(b) $z \in \mathbb{R}^n, z \neq \vec{0}, A = zz^T$

Since $A = zz^T, C(A)$ is subspace spanned by z and then has dimension of 1, or $rank(A) = 1$. Therefore, $N(A)$ has dimension $n-1$ and contains all vectors that perpendicular to z.

(c) $A \in \mathbb{R}^{n\times n}, A \succeq 0, B \in \mathbb{R}^{m\times n}$. We have $\forall x \in \mathbb{R}^n$:

$$x^T BAB^T x = \left(B^T x\right)^T A \left(B^T x\right) \succeq 0 \,(\text{since } A \succeq 0)$$

3. **Eigenvectors, eigenvalues, the spectral theorem**

$A \in \mathbb{R}^{n\times n}$ is diagonalizable $A = T\Lambda T^{-1}$, where $\begin{cases} T \text{ invertible} \\ \Lambda = diag\left(\lambda_1, \cdots, \lambda_n\right) \end{cases}$

(a) $A \in \mathbb{R}^{n\times n}$ diagonalizable, $t^{(i)}$: the ith column of T. Since A is diagonalizable, then we have:

$$\begin{aligned} A &= T\Lambda T^{-1} \\ AT &= T\Lambda \\ \left[At^{(1)} \cdots At^{(n)}\right] &= \left[t^{(1)} \cdots t^{(n)}\right] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \\ \left[At^{(1)} \cdots At^{(n)}\right] &= \left[\lambda_1 t^{(1)} \cdots \lambda_n t^{(n)}\right] \end{aligned}$$

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

(b) A symmetric, $U = \left[u^{(1)} \cdots u^{(n)}\right]$ orthogonal, $A = U\Lambda U^{-1}$.

$$\begin{aligned} \text{Since } U \text{ is orthogonal} &\Rightarrow U^T U = I \\ &\Rightarrow U^T = U^{-1} \\ &\Rightarrow Au^{(i)} = \lambda_i u^{(i)} \,(\text{by part (a)}) \end{aligned}$$

(c) $A \succeq 0$. Since $A \succeq 0$ then for all (eigenvalue, eigenvector) pairs (x_i, λ_i) of A, we have:

$$\begin{aligned} & x_i^T A x_i \geq 0 \\ \Rightarrow\ & x^T \lambda_i x_i \geq 0 \end{aligned}$$

Therefore, $\lambda_i \geq 0 \left(\text{since } x_i^T x_i \geq 0\right)$