CS229 Fall 2018 Homework

Tien-Dat Do

April 21, 2025

Problem Set #0: Linear Algebra and Multivariable Calculus

Problem 1: Gradients and Hessians

(a) We have

$$\frac{1}{2}x^T Ax = \frac{1}{2}\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j A_{ij} \quad (1)$$

And

$$b^T x = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} b_i x_i \quad (2)$$

From (1) and (2), we have

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

Since A is a symmetric matrix, we have

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \dots + A_{1n}x_n + b_1 \\ A_{21}x_1 + A_{22}x_2 + \dots + A_{2n}x_n + b_2 \\ \vdots \\ A_{n1}x_1 + A_{n2}x_2 + \dots + A_{nn}x_n + b_n \end{bmatrix} = Ax + b$$

(b) We have

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

$$\nabla f = \nabla g(h(x)) = \begin{bmatrix} \frac{\partial g(h(x)}{\partial x_1} \\ \frac{\partial g(h(x)}{\partial x_2} \\ \vdots \\ \frac{\partial g(h(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} g'(h(x)) \frac{\partial h(x)}{x_1} \\ g'(h(x)) \frac{\partial h(x)}{x_2} \\ \vdots \\ g'(h(x)) \frac{\partial h(x)}{x_n} \end{bmatrix} = g'(h(x)) \begin{bmatrix} \frac{\partial h(x)}{x_1} \\ \frac{\partial h(x)}{x_2} \\ \vdots \\ \frac{\partial h(x)}{x_n} \end{bmatrix} = g'(h(x)) \nabla h$$

(c) We have

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) = \frac{\partial}{\partial x_j} \left(\sum_{k=1}^{n} A_{ik} x_k + b_k \right) = A_{ij}$$

So that,

$$\nabla^2 f = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} = A$$

(d) We have

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

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

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

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

Problem 2: Positve definite matrices

(a) We have

$$\begin{aligned}
&x^T A x = x^T z z^T x = (z^T x)^T (z^T x) = \|z^T x\|_2^2 \geq 0, \quad \forall x \in \mathbb{R}^n \\
&A^T = (zz^T)^T = zz^T = A
\end{aligned}$$

So, $A \succeq 0$

(b) Null-space:

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\} = \{x \in \mathbb{R}^n \mid z^T x = 0\}$$

Rank:

$$\mathcal{R}(A) = \mathcal{R}(zz^T) \leq min(\mathcal{R}(z), \mathcal{R}(z^T)) = 1$$

Since z is a non-zero vector, so $\mathcal{R}(A) = 1$

(c) We have

$$\begin{aligned}
&(BAB^T)^T = (AB^T)^T B^T = BAB^T \\
&x^T BAB^T x = (B^T x)^T A (B^T x) \geq 0
\end{aligned}$$

since A is PSD, so BAB^T is PSD.

Problem 3: Eigenvectors, Eigenvalues and the spectral theorem

(a) We have

$$\begin{aligned}
A = T\Lambda T^{-1} &\Leftrightarrow AT = \Lambda T \\
&\Leftrightarrow A \begin{bmatrix} t^{(1)} & t^{(2)} & \dots & t^{(n)} \end{bmatrix} = \begin{bmatrix} t^{(1)} & t^{(2)} & \dots & t^{(n)} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \\
&\Leftrightarrow \begin{bmatrix} At^{(1)} & At^{(2)} & \dots & At^{(n)} \end{bmatrix} = \begin{bmatrix} \lambda_1 t^{(1)} & \lambda_2 t^{(2)} & \dots & \lambda_n t^{(n)} \end{bmatrix}
\end{aligned}$$

$\Rightarrow \; At^{(i)} = \lambda_i t^{(i)} \quad \forall i = \overline{1, n}$

(b) We have

$$
\begin{aligned}
A = U\Lambda U^T &\Leftrightarrow AU = U\Lambda \\
&\Leftrightarrow A\begin{bmatrix} u^{(1)} & u^{(2)} & \dots & u^{(n)} \end{bmatrix} = \begin{bmatrix} u^{(1)} & u^{(2)} & \dots & u^{(n)} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \\
&\Leftrightarrow \begin{bmatrix} Au^{(1)} & Au^{(2)} & \dots & Au^{(n)} \end{bmatrix} = \begin{bmatrix} \lambda_1 u^{(1)} & \lambda_2 u^{(2)} & \dots & \lambda_n u^{(n)} \end{bmatrix}
\end{aligned}
$$

$$\Rightarrow \; Au^{(i)} = \lambda_i u^{(i)} \quad \forall i = \overline{1, n}$$

(c) Since A is PSD, so we have

$$\left(t^{(i)}\right)^T At^{(i)} \geq 0 \Leftrightarrow \left(t^{(i)}\right)^T \lambda_i t^{(i)} \geq 0 \Leftrightarrow \|t(i)\|_2^2 \; \lambda_i \geq 0 \Rightarrow \lambda_i \geq 0 \; \forall i = \overline{1, n}$$