CS229 Problem Set #0 Solution

Jaeyun Kim

December 31, 2025

1. Gradients and Hessians

(a)

$$\nabla f(x) = \nabla\left(\frac{1}{2}x^\top Ax + b^\top x\right) = \nabla\left(\frac{1}{2}x^\top Ax\right) + \nabla(b^\top x)$$

Let's verify the gradient of the quadratic form.

$$\nabla\left(\frac{1}{2}x^\top Ax\right) = \frac{1}{2}\nabla\left(\sum_{i=1}^{n} x_i \sum_{j=1}^{n} A_{ij}x_j\right) = \frac{1}{2}\nabla\left(\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j\right)$$

Now, let's look at the k-th component of the gradient by taking the partial derivative with respect to x_k

$$\frac{\partial}{\partial x_k}\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\frac{\partial}{\partial x_k}(x_i x_j)$$

We consider three cases for the indices in the summation:

1. Case 1: $i = k$ and $j \neq k$
2. Case 2: $i \neq k$ and $j = k$
3. Case 3: $i = k$ and $j = k$

Then, the derivative is:

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

$$\text{Therefore, } \nabla(\frac{1}{2}x^\top Ax) = \frac{1}{2}\left(Ax + A^\top x\right) = Ax \quad \text{(since } A \text{ is a symmetric matrix)}$$

Next, for the linear term:

$$\nabla(b^\top x) = \nabla \sum_{i=1}^{n} b_i x_i = b$$

Therefore, combining both results:

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

(b)

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

(c)

$$\begin{aligned}
\nabla^2 f(x) &= \begin{bmatrix} | & | & & | \\ \frac{\partial \nabla f(x)}{\partial x_1} & \frac{\partial \nabla f(x)}{\partial x_2} & \cdots & \frac{\partial \nabla f(x)}{\partial x_n} \\ | & | & & | \end{bmatrix} \\
&= \begin{bmatrix} | & | & & | \\ \frac{\partial (Ax+b)}{\partial x_1} & \frac{\partial (Ax+b)}{\partial x_2} & \cdots & \frac{\partial (Ax+b)}{\partial x_n} \\ | & | & & | \end{bmatrix} \\
&= \begin{bmatrix} | & | & & | \\ A^{(1)} & A^{(2)} & \cdots & A^{(n)} \\ | & | & & | \end{bmatrix} \\
&= A
\end{aligned}$$

(d)

Let $h(x) = a^\top x$

- **Gradient**: $\nabla f(x) = \nabla g(h(x)) = g'(h(x)) \nabla h(x) = g'(a^\top x) a$

- **Hessian**:

$$
\begin{aligned}
\text{Let } h(x) &= a^\top x \\
\frac{\partial^2 g(h(x))}{\partial x_i \partial x_j} &= \frac{\partial}{\partial x_j} g'(h(x)) \frac{\partial h(x)}{\partial x_i} \\
&= g''(h(x)) \frac{\partial h(x)}{\partial x_j} \frac{\partial h(x)}{\partial x_i} \\
&= g''(a^\top x) a_j a_i
\end{aligned}
$$

Therefore, the Hessian matrix of $f(x)$ is

$$
\nabla^2 f(x) = g''(a^\top x) \begin{bmatrix} a_1 a_1 & \dots & a_1 a_n \\ \vdots & \ddots & \vdots \\ a_n a_1 & \dots & a_n a_n \end{bmatrix} = g''(a^\top x) a a^\top
$$

2. Positive definite matrices

(a)

$$
x^\top A x = x^\top z z^\top x = (z^\top x)^\top (z^\top x) = \|z^\top x\|^2 \geq 0 \quad \therefore A \succeq 0
$$

(b)

$$
\begin{aligned}
\text{Null}(A) &= \{x \mid Ax = 0\} = \{x \mid z^\top x = 0\} \quad (\because z \neq \vec{0}) \\
\text{Nullity}(A) &= \dim\Big(\{x \mid z^\top x = 0\}\Big) = n - 1 \\
\text{Rank}(A) &= 1 \quad \because \text{(rank-nullity theorem)}
\end{aligned}
$$

(c)

$$
\begin{gathered}
x^\top B A B^\top x = (B^\top x)^\top A (B^\top x) \geq 0 \quad (\because A \succeq 0) \\
\therefore B A B^\top \succeq 0
\end{gathered}
$$

3. Eigenvectors, eigenvalues, and the spectral theorem

(a)

$$
\begin{gathered}
A = T \Lambda T^{-1} \\
AT = T\Lambda \\
\begin{bmatrix} | & | & & | \\ At^{(1)} & At^{(2)} & \dots & At^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 t^{(1)} & \lambda_2 t^{(2)} & \dots & \lambda_n t^{(n)} \\ | & | & & | \end{bmatrix} \\
\therefore At^{(i)} = \lambda_i t^{(i)}
\end{gathered}
$$

(b)

$$\text{Since } U \text{ is an orthogonal matrix, } UU^\top = I$$

$$A = U\Lambda U^\top$$
$$AU = U\Lambda$$
$$\begin{bmatrix} | & | & & | \\ Au^{(1)} & Au^{(2)} & \cdots & Au^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 u^{(1)} & \lambda_2 u^{(2)} & \cdots & \lambda_n u^{(n)} \\ | & | & & | \end{bmatrix}$$

$$\therefore Au^{(i)} = \lambda_i u^{(i)} \implies u^{(i)} \text{is an eigenvector of } A$$

(c)

$$\begin{aligned} x^\top Ax &= x^\top U\Lambda U^\top x \\ &= (U^\top x)^\top \Lambda (U^\top x) \geq 0 \quad (\forall x \in \mathbb{R}^n, \text{ since } A \succeq 0) \end{aligned}$$

Let $c = U^\top x$. Since $U^\top$ is an orthogonal matrix, columns of $U^\top$ span $\mathbb{R}^n$. Therefore, c can be any vector in $\mathbb{R}^n$.

$$c^\top \Lambda c = \sum_{i=1}^{n} c_i^2 \lambda_i \geq 0 \quad (\forall c \in \mathbb{R}^n).$$

For any $j \in \{1, \ldots, n\}$, choose c such that $c_j = 1$ and $c_k = 0$ for $j \neq k$. Then,

$$\begin{aligned} & c^\top \Lambda c = \sum_{i=1}^{n} c_i^2 \lambda_i = \lambda_j \geq 0. \\ \therefore \quad & \lambda_i(A) \geq 0 \quad \text{for each } i \end{aligned}$$