PROBLEM SET 0

(1) **Gradients and Hessians**

Recall that a matrix $A \in \mathbb{R}^{n\times n}$ is symmetric if $A^T = A$, that is, $A_{ij} = A_{ji}$ for all i, j. Also recall the gradient $\nabla f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$, which is the n-vector of partial derivatives

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \cdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \text{ where } x = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}.$$

The hessian $\nabla^2 f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ symmetric matrix of twice partial derivatives,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f(x) \end{bmatrix}.$$

(a) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. What is $\nabla f(x)$? What is $/nabla^2 f(x)$?

Solution.

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

so

$$\frac{1}{2} x^T Ax = \frac{1}{2} \sum_{i=1}^{n} x_i^2 a_{ii} + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j a_{ij}.$$

Hence

$$\frac{\partial}{\partial x_k} f(x) = x_k a_{kk} + \sum_{j=k+1}^{n} x_j a_{kj} + \sum_{i=1}^{k-1} x_i a_{ik} + b_k = \sum_{i=1}^{n} x_i a_{ki} + b_k,$$

where $k = 1, 2, \cdots, n$.

Therefore

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

Since

$$\frac{\partial^2}{\partial x_k \partial x_l} f(x) = a_{kl},$$

where $l = 1, 2, \cdots, n$, we have

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

$\square$

(b) Let $f(x) = g(h(x))$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable. What is $\nabla f(x)$?

Solution.

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

$\square$

(c) Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a vector. What are $\nabla f(x)$ and $\nabla^2 f(x)$?

Solution.

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

Since

$$\frac{\partial^2}{\partial x_k \partial x_l} g(\sum_{i=1}^{n} a_i x_i) = a_k a_l g''(a^T x),$$

then

$$\nabla^2 g(a^T x) = g''(a^T x)aa^T.$$

$\square$

(2) **Positive definite matrices**

A matrix $A \in \mathbb{R}^{n\times n}$ is **positive semi-definite (PSD)**, denoted $A \succeq 0$, if $A = A^T$ and $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$. A matrix A is **positive definite**, denoted $A \succ 0$, if $A = A^T$ and $x^T Ax > 0$ for all $x \neq 0$.

(a) Let $z \in \mathbb{R}^n$ be an n-vector. Show that $A = zz^T$ is positive semi-definite.

Proof.

$$A = zz^T = \begin{bmatrix} z_1z_1 & z_1z_2 & \cdots & z_1z_n \\ z_2z_1 & z_2z_2 & \cdots & z_2z_n \\ \vdots & \vdots & \ddots & \vdots \\ z_nz_1 & z_nz_2 & \cdots & z_nz_n \end{bmatrix},$$

so $A = A^T$. For any $x \in \mathbb{R}^n$,

$$x^T Ax = x^T zz^T x = (x^T z)(x^T z)^T = (x^T z)^2 \geq 0$$

cause $x^T z \in \mathbb{R}$. $\square$

(b) Let $z \in \mathbb{R}^n$ be an non-zero n-vector. Let $A = zz^T$. What is the null-space of A? What is the rank of A?

Solution. Suppose $x \in null(A)$, then $Ax = 0$ which means $x^T Ax = (x^T z)^2 = 0$. Then we have $x^T z = 0$. So

$$Null(A) = \{x \in \mathbb{R}^n \mid x^T z = 0\}.$$

For any $x \in \mathbb{R}$, $Ax = z(z^T x)$, so

$$rank(A) = 1.$$

$\square$

(c) Let $A \in \mathbb{R}^{n\times n}$ be positive semi-definite and $B \in \mathbb{R}^{m\times n}$ be arbitrary, where $m, n \in \mathbb{N}$. Is BAB^T PSD? If so, prove it. If not, give a counterexample.

Proof.

$$(BAB^T)^T = BA^T B^T = BAB^T.$$

For any $x \in \mathbb{R}^m$,

$$x^T BAB^T x = (B^T x)^T A(B^T x) \geq 0.$$

So BAB^T is PSD. $\square$

(3) **Eigenvectors, eigenvalues, and the spectral theorem**

The eigenvalues of an $n \times n$ matrix $A \in \mathbb{R}^{n\times n}$ are the roots of the characteristic polynomial $p_A(\lambda) = \det(\lambda I - A)$, which may be complex. They are also defined as the values $\lambda \in \mathbb{C}$ for which there exists a vector $x \in \mathbb{C}^n$ such that $Ax = \lambda x$. We call such a pair (x, λ) an **eigenvector, eigenvalue** pair. In this question, we use the notation $\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ to denote the diagonal matrix with diagonal entries $\lambda_1, \cdots, \lambda_n$.

(a) Suppose that the matrix A is diagonalizable, that is, $A = T\Lambda T^{-1}$ for an invertible matrix $T \in \mathbb{R}^{n\times n}$, where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ Use the notation $t^{(i)}$ for the columns of T, so that $T = [t^{(1)} \cdots t^{(n)}]$, where $t^{(i)} \in \mathbb{R}^n$. Show that $At^{(i)} = \lambda_i t^{(i)}$, so that the eigenvalues/eigenvectors pairs of A are $(t^{(i)}, \lambda_i)$.

Proof. Use the notation e_i to denote the vector in $\mathbb{R}^n$ which all entries are 0 except the i-th entry. We have

$$t^{(i)} = Te_i,$$

then

$$At^{(i)} = T\Lambda T^{-1}Te_i = T\lambda_i e_i = \lambda_i t^{(i)}.$$

$\square$

A matrix $U \in \mathbb{R}^{n\times n}$ is orthogonal if $U^T U = I$. The spectral theorem, perhaps one of the most important theorems in linear algebra, states that if $A \in \mathbb{R}^{n\times n}$ is symmetric, then A is **diagonalizable by a real orthogonal matrix**. That is, there are a diagonal matrix $\Lambda \in \mathbb{R}^{n\times n}$ and orthogonal matrix $U \in \mathbb{R}^{n\times n}$ such that

$A = U\Lambda U^{-1} = U\Lambda U^T$, which means $U^T A U = \Lambda$. Let $\lambda_i = \lambda_i(A)$ denote the i-th eigenvalue of A.

(b) Let A be symmetric. Show that if $U = [u^{(1)}\ u^{(2)}\ \cdots\ u^{(n)}]$ is orthogonal, and $A = U\Lambda U^T$, then $u^{(i)}$ is an eigenvector of A and $Au^{(i)} = \lambda_i u^{(i)}$, where $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$.

Proof.

$$Au^{(i)} = U\Lambda U^T U e_i = U\lambda_i e_i = \lambda_i u^{(i)}.$$

□

(c) Show that if A is PSD, then $\lambda_i(A) \geq 0$ for each i.

Proof. A is PSD, then for each $(u^{(i)}, \lambda_i)$,

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

Hence $\lambda_i \geq 0$ for each i. □