1. [0 points] **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

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

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,

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

(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 $\nabla f(x)$?

(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 $\nabla f(x)$?

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

(d) 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 $\nabla^2 f(x)$? (*Hint:* your expression for $\nabla^2 f(x)$ may have as few as 11 symbols, including $'$ and parentheses.)

(a) $\cancel{2} Ax + b$

(b) $h'(x) \cdot g'(h(x))$

(c) A

(d) $\cancel{\nabla f(x)}$ $\nabla f(x) = a \cdot g'(a^T x)$

$\nabla^2 f(x) = \cancel{a^2 \cdot g''(a^T x)}$ $a^T a g''(a^T x)$

2. **[0 points] 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 $x \neq 0$, that is, all non-zero vectors x. The simplest example of a positive definite matrix is the identity I (the diagonal matrix with 1s on the diagonal and 0s elsewhere), which satisfies $x^T Ix = \|x\|_2^2 = \sum_{i=1}^n x_i^2$.

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

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

   (c) Let $A \in \mathbb{R}^{n\times n}$ be positive semidefinite 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 with explicit A, B.

(a) rank$(A)=1$ since every row/column is proportional to z.
The rank of A is equal to the number of non-zero eigenvalues of A.
$Az = z(z^Tz) = (z^Tz)z = \lambda z$
(z^Tz is a scalar and a scalar is commutative)
$\lambda = z^Tz = \|z\|^2 \geq 0$
A is a symmetric matrix and the definiteness of a symmetric matrix depends entirely on the sign of its eigenvalues.
Only non-zero eigenvalue $\lambda > 0$
A is positive semidefinite

(b) $N(A) = \mathbb{R}^n - R(A)$
$R(A) = \{v \in \mathbb{R}^n : v = xz, x \in \mathbb{R}\}$
rank$(A) = 1$

(c) Let $x \in \mathbb{R}^m$
$x^T BAB^T x = y^T A y \, \cancel{(B^Tx)} \, (B^T x = y) \geq 0$
BAB^T is PSD.

3. **[0 points] 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 (in general) 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 $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ to denote the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, that is,

$$\mathrm{diag}(\lambda_1, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

(a) Suppose that the matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable, that is, $A = T\Lambda T^{-1}$ for an invertible matrix $T \in \mathbb{R}^{n\times n}$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. 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/eigenvector pairs of A are $(t^{(i)}, \lambda_i)$.

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 symetric, that is, $A = A^T$, 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 $U^T A U = \Lambda$, or, equivalently,

$$A = U\Lambda U^T.$$

Let $\lambda_i = \lambda_i(A)$ denote the ith eigenvalue of A.

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

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

(a) $A = T\Lambda T^{-1} \Rightarrow AT = T\Lambda$

$AT = [At^{(1)} \cdots At^{(n)}]$ $T\Lambda = [T\lambda^{(1)} \cdots T\lambda^{(n)}] = [t^{(1)}\lambda_1 \cdots t^{(n)}\lambda_n]$

$[At^{(1)} \cdots At^{(n)}] = [t^{(1)}\lambda_1 \cdots t^{(n)}\lambda_n]$

(b) $A = U\Lambda U^T \Rightarrow AU = U\Lambda U^T U = U\Lambda$

$AU = [Au^{(1)} \cdots Au^{(n)}]$, $U\Lambda = [U\lambda^{(1)} \cdots U\lambda^{(n)}] = [u^{(1)}\lambda_1 \cdots u^{(n)}\lambda_n]$

$[Au^{(1)} \cdots Au^{(n)}] = [u^{(1)}\lambda_1 \cdots u^{(n)}\lambda_n]$

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

$$X^T A X = X^T U \Lambda U^T X = \hat{X}^T \Lambda \hat{X} = \sum_{i=1}^{n} \lambda_i \hat{X}_i^2 \geq 0$$

To $\sum_{i=1}^{n} \lambda_i \hat{X}_i^2 \geq 0$ be true for all $\hat{X}$,

$\lambda_i(A) \geq 0$ must be true.

4. **[0 points] Probability and multivariate Gaussians**

   Suppose $X = (X_1, ..X_n)$ is sampled from a multivariate Gaussian distribution with mean μ in $\mathbb{R}^n$ and covariance Σ in S_+^n (i.e. Σ is positive semidefinite). This is commonly also written as $X \sim \mathcal{N}(\mu, \Sigma)$.

   (a) Describe the random variable $Y = X_1 + X_2 + \ldots + X_n$. What is the mean and variance? Is this a well known distribution, and if so, which?

   (b) Now, further suppose that Σ is invertible. Find $\mathbb{E}[X^T\Sigma^{-1}X]$. (Hint: use the property of trace that $x^TAx = \mathrm{tr}(x^TAx)$).

(a) Y is the linear combination of the elements of a multivariate normal random variable.

mean : μ

variance : Σ

This is a multivariate Gaussian distribution.

(b) $\mathbb{E}[X^T\Sigma^{-1}X]$

$= \mathbb{E}[tr(X^T\Sigma^{-1}X)]$

$= \mathbb{E}[tr(XX^T\Sigma^{-1})]$

$= tr(\mathbb{E}[XX^T\Sigma^{-1}])$

$= tr(\Sigma^{-1}\mathbb{E}[XX^T])$

$= tr(\Sigma^{-1}(\Sigma + \mu\mu^T))$

$= tr(I + \Sigma^{-1}\mu\mu^T)$

$= tr(I) + tr(\Sigma^{-1}\mu\mu^T)$

$= n + tr(\mu^T\Sigma^{-1}\mu)$

$= n + \mu^T\Sigma^{-1}\mu$