CS 229, Summer 2022
Problem Set #0: Linear Algebra, Multivariable Calculus, and Probability

---

Ungraded.

Notes:
(1) These questions require thought, but do not require long answers. Please be as concise as possible.
(2) If you have a question about this homework, we encourage you to post your question on our Ed at `https://edstem.org/us/courses/23539/discussion/`.
(3) If you missed the first lecture or are unfamiliar with the collaboration or honor code policy, please read the policy before you start.
(4) This specific homework is ***not graded***, but we encourage you to solve each of the problems to brush up on your linear algebra and probability. Some of them may even be useful for subsequent problem sets. It also serves as your introduction to using Gradescope for submissions. We strongly suggest you use LaTex to submit your psets (not ony is it helpful for this class, but it is a good skill to learn). However, if you are scanning your document by cellphone, please use a scanning app such as CamScanner. There will not be any late days allowed for this particular assignment.

Honor code: We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down the solutions independently, and without referring to written notes from the joint session. That being said, if students are submitting in a pair, they act as one unit - they may share resources (such as notes) with each other and write the solutions together. Note that both of the two students should fully understand all the answers in their submission, even though only one of them needs to write up a solution to a question. In other words, each student must understand the solution well enough in order to reconstruct it by him/herself. In addition, each student should write on the problem set the set of people with whom s/he collaborated. Further, because we occasionally reuse problem set questions from previous years, we expect students not to copy, refer to, or look at the solutions in preparing their answers. It is an honor code violation to intentionally refer to a previous year's solutions.

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.)

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 $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$. A matrix A is *positive definite*, denoted $A \succ 0$, if $A = A^T$ and $x^T A x > 0$ 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 $x^T I x = \|x\|_2^2 = \sum_{i=1}^n x_i^2$.

   (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.

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.

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 $x^T A x = \mathrm{tr}(x^T A x)$).