CS 229, Autumn 2016
Problem Set #0 Solutions: Linear Algebra and Multivariable Calculus

---

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 Piazza forum, at `https://piazza.com/stanford/autumn2016/cs229`. (3) If you missed the first lecture or are unfamiliar with the collaboration or honor code policy, please read the policy on Handout #1 (available from the course website) before starting work. (4) This specific homework is *not graded*, but we encourage you to solve each of the problems to brush up on your linear algebra. Some of them may even be useful for subsequent problem sets. It also serves as your introduction to using Gradescope for submissions.

If you are scanning your document by cellphone, please check the Piazza forum for recommended cellphone scanning apps and best practices.

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

   Answer: In short, we know that $\nabla(\frac{1}{2} x^T A x) = Ax$ for a symmetric matrix A, while $\nabla(b^T x) = b$. Then $\nabla f(x) = Ax + b$ when A is symmetric. In more detail, we have

$$\frac{1}{2} x^T A x = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j,$$

so for each $k = 1, \ldots, n$, we have

$$\begin{aligned}
\frac{\partial}{\partial x_k}\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n A_{ij}x_ix_j &\overset{(i)}{=} \frac{\partial}{\partial x_k}\frac{1}{2}\sum_{i=1, i\neq k}^n A_{ik}x_ix_k + \frac{\partial}{\partial x_k}\frac{1}{2}\sum_{j=1, j\neq k}^n A_{kj}x_kx_j + \frac{\partial}{\partial x_k}\frac{1}{2}A_{kk}x_k^2 \\
&\overset{(ii)}{=} \frac{1}{2}\sum_{i=1, i\neq k}^n A_{ik}x_i + \frac{1}{2}\sum_{j=1, j\neq k}^n A_{kj}x_j + A_{kk}x_k \\
&= \sum_{i=1}^n A_{ki}x_i
\end{aligned}$$

where step (i) follows because $\frac{\partial}{\partial x_k}A_{ij}x_ix_j = 0$ if $i \neq k$ and $j \neq k$, step (ii) by the definition of a partial derivative, and the final equality because $A_{ij} = A_{ji}$ for all pairs i, j. Thus $\nabla(\frac{1}{2}x^TAx) = Ax$. To see that $\nabla b^Tx = b$, note that

$$\frac{\partial}{\partial x_k}b^Tx = \frac{\partial}{\partial x_k}\sum_{i=1}^n b_ix_i = \frac{\partial}{\partial x_k}b_kx_k = b_k.$$

(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 $\nabla f(x)$?

Answer: In short, if g' is the derivative of g, then the chain rule gives

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

Expanding this by components, we have for each $i = 1, \ldots, n$ that

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

by the chain rule. Stacking each of these in a column vector, we obtain

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

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

Answer: We have $\nabla^2 f(x) = A$. To see this more formally, note that $\nabla^2(b^Tx) = 0$, because the second derivatives of b_ix_i are all zero. Let $A = [a^{(1)} \ \cdots \ a^{(n)}]$, where $a_i \in \mathbb{R}^n$ is an n-vector (because A is symmetric, we also have $A = [a^{(1)} \ a^{(2)} \ \cdots \ a^{(n)}]^T$). Then we use part (1a) to obtain

$$\frac{\partial}{\partial x_k}(\frac{1}{2}x^TAx) = {a^{(k)}}^T x = \sum_{i=1}^n A_{ik}x_i,$$

and thus

$$\frac{\partial^2}{\partial x_k x_i}(\frac{1}{2}x^TAx) = \frac{\partial}{\partial x_i}{a^{(k)}}^T x = A_{ik}.$$

(d) Let $f(x) = g(a^Tx)$, 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.)

Answer: We use the chain rule (part (1b)) to see that $\nabla f(x) = g'(a^T x)a$, because $\nabla(a^T x) = a$. Taking second derivatives, we have

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

Expanding this in matrix form, we have

$$\nabla^2 f(x) = g''(a^T x)\begin{bmatrix} a_1^2 & a_1 a_2 & \cdots & a_1 a_n \\ a_2 a_1 & a_2^2 & \cdots & a_2 a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n a_1 & a_n a_2 & \cdots & a_n^2 \end{bmatrix} = g''(a^T x)aa^T.$$

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.

Answer: Take any $x \in \mathbb{R}^n$. Then $x^T Ax = x^T zz^T x = (x^T z)^2 \geq 0$.

(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?

Answer: If $n = 1$, the null space of A is empty. The rank of A is always 1, as the null-space of A is the set of vectors orthogonal to z. That is, if $z^T x = 0$, then $x \in \mathsf{Null}(A)$, because $Ax = zz^T x = 0$. Thus, the null-space of A has dimension $n - 1$ and the rank of A is 1.

(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.

Answer: Yes, BAB^T is positive semidefinite. For any $x \in \mathbb{R}^m$, we may define $v = B^T x \in \mathbb{R}^n$. Then

$$x^T BAB^T x = (B^T x)^T A(B^T x) = v^T Av \geq 0,$$

where the inequality follows because $v^T Av \geq 0$ for any vector v.

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 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 $(t^{(i)}, \lambda_i)$.

Answer: The matrix T is invertible, so if we let $t^{(i)}$ be the ith column of T, we have

$$I_{n\times n} = T^{-1}T = T^{-1}\left[t^{(1)} \ t^{(2)} \ \cdots \ t^{(n)}\right] = \left[T^{-1}t^{(1)} \quad T^{-1}t^{(2)} \quad \cdots \quad T^{-1}t^{(n)}\right]$$

so that

$$T^{-1}t^{(i)} = [\ \underbrace{0 \ \cdots \ 0}_{i-1 \text{ times}} \ \ 1 \ \ \underbrace{0 \ \cdots \ 0}_{n-i \text{ times}}\ ]^T \in \{0,1\}^n,$$

the ith standard basis vector, which we denote by $e^{(i)}$ (that is, the vector of all-zeros except for a 1 in its ith position. Thus

$$\Lambda T^{-1}t^{(i)} = \Lambda e^{(i)} = \lambda_i e^{(i)}, \ \text{ and } \ T\Lambda T^{-1}t^{(i)} = \lambda_i T e^{(i)} = \lambda_i t^{(i)}.$$

A matrix $U \in \mathbb{R}^{n\times n}$ is orthogonal if $U^TU = I$. 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 $U^TAU = \Lambda$, 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 $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

Answer: Once we see that $U^{-1} = U^T$ because $U^TU = I$, this is simply a repeated application of part (3a).

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

Answer: Let $x \in \mathbb{R}^n$ be any vector. We know that $A = A^T$, so that $A = U\Lambda U^T$ for an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ by the spectral theorem. Take the ith eigenvector $u^{(i)}$. Then we have

$$U^Tu^{(i)} = e^{(i)},$$

the ith standard basis vector. Using this, we have

$$0 \leq {u^{(i)}}^T Au^{(i)} = (U^Tu^{(i)})^T\Lambda U^Tu^{(i)} = {e^{(i)}}^T\Lambda e^{(i)} = \lambda_i(A).$$