CS229 Fall 2018

Problem Set #0: Linear Algebra and Multivariable Calculus

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)

$$\nabla f(x) = \begin{bmatrix} (Ax)_1 + b_1 \\ \vdots \\ (Ax)_n + b_n \end{bmatrix} = \begin{bmatrix} (Ax+b)_1 \\ \vdots \\ (Ax+b)_n \end{bmatrix} = Ax + b$$

(b)

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

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

(c)

$$\nabla^2 f(x) = \nabla(\nabla f(x)) = \nabla(Ax+b)\nabla^2 f(x) = \nabla(\nabla f(x)) = \nabla(Ax+b)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}(Ax+b) \\ \vdots \\ \frac{\partial}{\partial x_n}(Ax+b) \end{bmatrix}$$

$$= \begin{bmatrix} A_1^\top \\ \vdots \\ A_n^\top \end{bmatrix} = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} = A$$

(d)

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1} g(a^\top x) \\ \vdots \\ \frac{\partial}{\partial x_n} g(a^\top x) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial g(a^\top x)}{\partial a^\top x} \cdot \frac{\partial a^\top x}{\partial x_1} \\ \vdots \\ \frac{\partial g(a^\top x)}{\partial a^\top x} \cdot \frac{\partial a^\top x}{\partial x_n} \end{bmatrix} = \frac{\partial g(a^\top x)}{\partial a^\top x} \begin{bmatrix} \frac{\partial a^\top x}{\partial x_1} \\ \vdots \\ \frac{\partial a^\top x}{\partial x_n} \end{bmatrix}$$

$$= \frac{\partial g(a^\top x)}{\partial a^\top x} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \frac{\partial g(a^\top x)}{\partial a^\top x} \cdot a$$

$$(\nabla^2 f(x))_{ij} = \frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} g(a^\top x)$$

$$= \frac{\partial}{\partial x_i} \frac{\partial g(a^\top x)}{\partial a^\top x} \frac{\partial a^\top x}{\partial x_j} = \frac{\partial}{\partial a^\top x} \frac{\partial a^\top x}{\partial x_i} \frac{\partial g(a^\top x)}{\partial a^\top x} \frac{\partial a^\top x}{\partial x_j}$$

$$= \frac{\partial}{\partial a^\top x} \frac{\partial g(a^\top x)}{\partial a^\top x} \frac{\partial a^\top x}{\partial x_i} \frac{\partial a^\top x}{\partial x_j} = \frac{\partial^2 g(a^\top x)}{(\partial a^\top x)^2} \frac{\partial a^\top x}{\partial x_i} \frac{\partial a^\top x}{\partial x_j}$$

$$= \frac{\partial^2 g(a^\top x)}{(\partial a^\top x)^2} \cdot a_i a_j$$

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

$$= \frac{\partial^2 g(a^\top x)}{(\partial a^\top x)^2} \cdot a a^\top$$

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)

1. $(zz^\top)^\top = zz^\top$
2. $x^\top zz^\top x = (z^\top x)^\top z^\top x = (z^\top x)^2$ (by $z^\top x \in \mathbb{R}$) ≥ 0

(b)

1. Null-space of ***A***

$zz^\top x = 0$

$$\rightarrow \begin{bmatrix} z_1(z_1x_1 + \cdots + z_nx_n) \\ \vdots \\ z_n(z_1x_1 + \cdots + z_nx_n) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\rightarrow z^\top x = 0$

$\therefore \left\{x \in \mathbb{R}^n : \ z^\top x = 0\right\}$

2. Rank of ***A***

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

$$\Rightarrow \begin{bmatrix} z_1 & z_2 & \cdots & z_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$ (by row operation)

$\therefore rank(A) = 1$

(c)

1. $(BAB^\top)^\top = BA^\top B^\top = BAB^\top$ (by $A = A^\top$)
2. $x^\top BAB^\top x = (B^\top x)^\top AB^\top x = y^\top Ay \ (y \in \mathbb{R}^n) \geq 0$

(by $x^\top Ax \geq 0$, for all x)

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

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 $U^T AU = \Lambda$, 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)

$$A = T\Lambda T^{-1}$$
$$AT = T\Lambda$$
$$A\begin{bmatrix} t^{(1)} & \cdots & t^{(n)} \end{bmatrix} = \begin{bmatrix} t^{(1)} & \cdots & t^{(n)} \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$
$$\begin{bmatrix} At^{(1)} & \cdots & At^{(n)} \end{bmatrix} = \begin{bmatrix} \lambda_1 t^{(1)} & \cdots & \lambda_n t^{(n)} \end{bmatrix}$$
$$\Rightarrow At^{(i)} = \lambda_i t^{(i)}$$

(b)

$$A = A^\top, U^\top U = I$$
$$\to A = U\Lambda U^\top$$
$$AU = U\Lambda$$
$\Rightarrow Au^{(i)} = \lambda_i u^{(i)}$ (by the previous (a))

(c)

A is PSD

$\rightarrow A = A^\top,\ x^\top Ax \geq 0$, for all $x \in \mathbb{R}^n$

$Au^{(i)} = \lambda_i u^{(i)}$ (by the previous (b))

$u^{(i)\top}Au^{(i)} = \lambda_i u^{(i)\top}u^{(i)}$

$u^{(i)\top}Au^{(i)} \geq 0 \rightarrow \lambda_i u^{(i)\top}u^{(i)} = \lambda_i \geq 0$