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

Godfray Qiu

January 6, 2017

---

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

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

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,

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

(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 $\nabla f(x)$?

We have $\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} \frac{1}{2} x^T Ax \\ \vdots \\ \frac{\partial}{\partial x_n} \frac{1}{2} x^T Ax \end{bmatrix} + \begin{bmatrix} \frac{\partial}{\partial x_1} b^T x \\ \vdots \\ \frac{\partial}{\partial x_n} b^T x \end{bmatrix}$, considering

$$\frac{\partial x^T Ax}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_{j=1}^n \sum_{k=1}^n x_j a_{jk} x_k = \frac{\partial}{\partial x_i} \left(\sum_{j\neq i}^n \sum_{k\neq i}^n x_j a_{jk} x_k + \sum_{k\neq i}^n x_i a_{ik} x_k + \sum_{j\neq i}^n x_j a_{ji} a_i + x_i a_{ii} x_i \right)$$

$$= 0 + \sum_{k\neq i}^n a_{ik} x_k + \sum_{j\neq i}^n x_j a_{ji} + 2a_{ii} x_i = \sum_{k=1}^n a_{ik} x_k + \sum_{j=1}^n x_j a_{ji} = A_{i*}^T x + A_{*i}^T x = \big((A + A^T)x\big)_i$$

that is, the i-th element of $\frac{\partial x^T Ax}{\partial x_i}$ is exact inner product of i-th row vector of A and x plus i-th column vector of A and x (also i-th row of A^T and x), and we get the vectorized representation. For a symmetric matrix A, we have $A^T = A$, and $\frac{\partial}{\partial x_i}\frac{1}{2}x^TAx = \frac{1}{2}(A^T + A)x = Ax$. Likewise and more easier,

$$\frac{\partial b^T x}{\partial x_i} = \frac{\partial}{\partial x_i}\sum_{j=1}^{n} b_j x_j = b_i.$$

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

(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 $\nabla f(x)$?

By rule of chain of derivatives,

$$\nabla f(x) = \frac{\partial}{\partial x_i} g(h(x)) = \frac{\mathrm{d}g(h(x))}{\mathrm{d}h(x)} \times \frac{\partial h(x)}{\partial x_i} = g'(h(x)) \times \nabla h(x).$$

(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 $\nabla^2 f(x)$?

From 1(a), $\frac{\partial}{\partial x_i}\left(\frac{1}{2}x^TAx + b^Tx\right) = Ax + b$, and

$$\frac{\partial^2}{\partial x_i \partial x_j}\left(\frac{1}{2}x^TAx + b^Tx\right) = \frac{\partial}{\partial x_j}(Ax + b) = \frac{\partial Ax}{\partial x_j} = \frac{\partial}{\partial x_j}\sum_{k=1}^{n} a_{ik}x_k = a_{ij}.$$

Still, A is symmetric, and $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = A_{ij}$, $\nabla^2 f(x) = A$.

(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.*)

By rule of chain of derivatives and 1(a),

$$\frac{\partial}{\partial x_i} g(a^Tx) = \frac{\mathrm{d}g(a^Tx)}{\mathrm{d}(a^Tx)} \times \frac{\partial a^Tx}{\partial x_i} = g'(a^Tx)a_i$$

and $\nabla f(x) = g'(a^Tx)a$;

$$\frac{\partial^2}{\partial x_i \partial x_j} g(a^Tx) = \frac{\partial}{\partial x_j} g'(a^Tx)a_i = \frac{\mathrm{d}g'(a^Tx)}{\mathrm{d}a^Tx} \times \frac{\partial a^Tx}{\partial x_j} = g''(a^Tx)a_i a_j$$

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

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 $x^TAx \geq 0$ for all $x \in \mathbb{R}^n$. A matrix A is *positive definite*, denoted $A \succ 0$, if $A = A^T$ and $x^TAx > 0$ for all $x \neq 0$, that is, all none-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 $x^TIx = \|x\|_2^2 = \sum_{i=1}^{n} x_i^2$.

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

A is symmetric, and for any vector $x \in \mathbb{R}^n$, $x^TAx = x^T(zz^T)x = (x^Tz)(z^Tx) = (z^Tx)(z^Tx) = (z^Tx)^2 \geq 0$. Therefore, $A = zz^T$ is positive semi-definite.

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

Take an n-vector x. Let $Ax = zz^T x = 0$. Thus, $x^T zz^T x = 0 \Rightarrow z^T x = 0, \mathcal{N}(A) = \{x | z^T x = 0\}$.

The rank of A is always 1. If we do elementary operations on A, we get a matrix with the first row of z^T, and all other elements equivalent to 0. From another point of view, the dimension of null-space of A is $n-1$, by adding which the rank of A always equals n.

(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 with explicit A, B.

BAB^T is PSD. Take vector $x \in \mathbb{R}^m$ and we have $x^T BAB^T x = (B^T x)^T A (B^T x) = y^T Ay$, ($y = B^T x \in \mathbb{R}^n$). A is postive semi-definite, and $y^T Ay \geq 0$, meaning $x^T (BAB^T) x \geq 0$.

---

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 (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 diag($\lambda_1, \cdots, \lambda_n$) to denote the diagonal matrix with diagonal entries $\lambda_1, \cdots, \lambda_n$, that is,

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

(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 $A = \text{diag}(\lambda_1, \cdots, \lambda_n)$ 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 $(t^{(i)}, \lambda_i)$.

From $A = T\Lambda T^{-1}$, $AT = T\Lambda(T^{-1}T) = T\Lambda$, that is,

$$[At^{(1)}, \cdots, At^{(n)}] = [t^{(1)} \cdots t^{(n)}] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1 t^{(1)}, \cdots, \lambda_n t^{(n)}]$$

Therefore, $At^{(i)} = \lambda_i t^{(i)}$ and $(t^{(i)}, \lambda_i)$ is a pair of eigenvalues/eigenvector of A.

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, 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 $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$.

For orthogonal matrix U, $U^T U = UU^T = I$, that is U is invertible and $U^{-1} = U^T$. The same with 3(a).

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

A is PSD, for any vector $u \in \mathbb{R}^n$, $u^T Au \geq 0$. Take a pair of eigenvalues/eigenvector of A denoted as $(u^{(i)}, \lambda_i)$, that is $Au^{(i)} = \lambda_i u^{(i)}$, and λ_i is a (real) number,

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

Thus, $\lambda_i(A) \geq 0$.