Problem Set #0: Linear Algebra and Multivariable Calculus

Problem_set_0.pdf

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,

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

(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 $h : \mathbb{R}^{n\times n} \to \mathbb{R}$ is differentiable. What is $\nabla f(x)$?

(c) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a 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.*

1.a) Assuming that $g(x) = x^T Ax$ and $h(x) = b^T x$ such that $f(x) = \frac{1}{2}g(x) + h(x)$.

Given the properties below: 1. $\nabla_x(g(x) + h(x)) = \nabla_x f(x) + \nabla_x g(x)$ 2. For $\lambda \in \mathbb{R}, \nabla_x(\lambda f(x)) = \lambda \nabla_x f(x)$

So, $\nabla_x f(x) = \frac{1}{2}\nabla_x g(x) + \nabla_x h(x)$

i)

$$h(x) = b^T x = \sum_{i=1}^{n} b_i x_i \nabla_x h(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} b_1x_1 + b_2x_2 + \cdots + b_nx_n \\ \frac{\partial}{\partial x_2} b_1x_1 + b_2x_2 + \cdots + b_nx_n \\ \vdots \\ \frac{\partial}{\partial x_n} b_1x_1 + b_2x_2 + \cdots + b_nx_n \end{bmatrix} = \begin{bmatrix} b1 \\ b2 \\ \vdots \\ bn \end{bmatrix} = b$$

ii)

$$g(x) = x^T Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_ix_j$$

$$\begin{aligned}
\nabla_x g(x) &= \frac{\partial}{\partial x_k}\left[\sum_{i=1}^n \sum_{j=1}^n A_{ij}x_i x_j\right] \\
&= \frac{\partial}{\partial x_k}\left[\sum_{i\neq k}^n \sum_{j\neq k}^n A_{ij}x_i x_j + \sum_{i\neq k}^n A_{ik}x_i x_k + \sum_{j\neq k}^n A_{kj}x_k x_j + A_{kk}x_k^2\right] \\
&= \sum_{i\neq k}^n A_{ik}x_i + \sum_{j\neq k}^n A_{kj}x_j + 2A_{kk}x_k \\
&= \sum_{i=1}^n A_{ik}x_i + \sum_{j=1}^n A_{kj}x_j = 2\sum_{i=1}^n A_{ki}x_i \\
&= 2Ax \text{ (if } A \text{ is symmetric)}
\end{aligned}$$

iii) Hence,

$$\begin{aligned}
\nabla_x f(x) &= \frac{1}{2}\nabla_x g(x) + \nabla_x h(x) \\
&= Ax + b
\end{aligned}$$

1.b)

$$\nabla_x f(x) = \nabla_x g(h(x)) = \begin{bmatrix} \frac{d}{dh}g(h)\frac{\partial}{\partial x_1}h(x) \\ \vdots \\ \frac{d}{dh}g(h)\frac{\partial}{\partial x_n}h(x) \end{bmatrix}$$

1.c) Continued from **1.a)**,

$$\nabla_x^2 f(x) = \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{\partial}{\partial x_k}\left[\frac{\partial f(x)}{\partial x_l}\right] = \frac{\partial}{\partial x_k}\left[2\sum_{i=1}^n A_{li}x_i\right] = 2A_{lk} = 2A_{kl} = 2A$$

1.d)

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

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

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 $x^T A x \geqslant 0$ 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.

2.a) $i)$ Proving that $A = A^T$

Considering $z \in \mathbb{R}^n$, so $A = zz^T$ is:

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \rightarrow A = \begin{bmatrix} z_1z_1 & z_1z_2 & \cdots & z_1z_n \\ z_2z_1 & z_2z_2 & \cdots & z_2z_n \\ \vdots & \vdots & \ddots & \vdots \\ z_nz_1 & z_nz_2 & \cdots & z_nz_n \end{bmatrix} \rightarrow A^T = \begin{bmatrix} z_1z_1 & z_2z_1 & \cdots & z_nz_1 \\ z_1z_2 & z_2z_2 & \cdots & z_nz_2 \\ \vdots & \vdots & \ddots & \vdots \\ z_1z_n & z_2z_n & \cdots & z_nz_n \end{bmatrix}$$

Therefore,

$$A = A^T$$

$ii)$ Proving that $x^TAx \geqslant 0$

$$x^TAx = x^Tzz^Txx^Tz = \sum_{i=1}^{n} x_iz_i = \sum_{i=1}^{n} z_ix_i = z^Tx$$

Replacing $\lambda = x^Tz = z^T = \lambda \in \mathbb{R}$. It give us,

$$x^TAx = x^Tzz^Tx = \lambda^2 \geqslant 0$$

Therefore, A is positive semidefinite because

$$A = A^Tx^TAx \geqslant 0$$

2.b) $i)$ Be

$$A = zz^T = \begin{bmatrix} z_1z_1 & z_1z_2 & \cdots & z_1z_n \\ z_2z_1 & z_2z_2 & \cdots & z_2z_n \\ \vdots & \vdots & \ddots & \vdots \\ z_nz_1 & z_nz_2 & \cdots & z_nz_n \end{bmatrix} \rightarrow row_n := row_n - row_1 * \frac{z_n}{z_1}$$

$$\rightarrow \begin{bmatrix} z_1z_1 & z_1z_2 & \cdots & z_1z_n \\ z_2z_1 - z_1z_1 * z_2/z_1 & z_2z_2 - z_1z_2 * z_2/z_1 & \cdots & z_2z_n * z_2/z_1 - z_1z_n * z_2/z_1 \\ \vdots & \vdots & \ddots & \vdots \\ z_nz_1 - z_1z_1 * z_n/z_1 & z_nz_2 - z_1z_2 * z_n/z_1 & \cdots & z_nz_n - z_1z_n * z_n/z_1 \end{bmatrix}$$

$$= \begin{bmatrix} z_1z_1 & z_1z_2 & \cdots & z_1z_n \\ z_2z_1 - z_1z_2 & z_2z_2 - z_2z_2 & \cdots & z_2z_n - z_nz_2 \\ \vdots & \vdots & \ddots & \vdots \\ z_nz_1 - z_1z_n & z_nz_2 - z_2z_n & \cdots & z_nz_n - z_nz_n \end{bmatrix} = \begin{bmatrix} z_1z_1 & z_1z_2 & \cdots & z_1z_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Therefore, $Rank(A) = 1$

$ii)$ If $n = 1$, the ***nullspace*** of A is empty. The rank of A is always 1, as the *nullspace* of A is the set of vectors orthogonal to z. That is, if $z^Tx = 0$, then $x \in \mathcal{N}(A)$, because $Ax = zz^Tx = 0$. Thus, the *nullspace* of A has dimension $n - 1$.

From $\rightarrow$

2.c) Given

1. $A \in \mathbb{R}^{n\times n}$ is PSD, so $A = A^T$, and $x^T Ax \geq 0$, and
2. $B \in \mathbb{R}^{m\times n}$

Let's denote $C = BAB^T$. We need to prove two parts in order to show that C is PSD, too.
i)

$$\begin{aligned} C^T &= (BAB^T)^T \\ &= (B^T)^T A^T B^T \\ &= BA^T B^T \\ &= BAB^T \\ &= C \end{aligned}$$

ii)

$$\begin{aligned} x^T Cx &= x^T (BAB^T)x \\ &= (x^T B)A(B^T x) \\ &= (B^T x)^T A(B^T x) \\ &= y^T Ay \\ &\geq 0 \end{aligned}$$

Here we transformed $x^T Cx$ into one that leverages the property of A, $x^T Ax \geq 0$, by setting $y = B^T x$. Therefore, BAB^T is PSD.

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,

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

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

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,

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

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

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

3.a) Given $A = T\Lambda T^{-1}$, so

$$\begin{aligned} &AT = T\Lambda \\ \rightarrow &A[t^{(1)} \cdots t^{(n)}] = [t^{(1)} \cdots t^{(n)}]\Lambda \\ \rightarrow &At^{(i)} = t^{(i)}\lambda_i \end{aligned}$$

The last step is because Λ is a diagonal matrix with all off-diagonal values being zeros. Therefore,

$$At^{(i)} = \lambda_i t^{(i)}$$

$(t^{(i)}, \lambda_i)$ is an eigenvalue/eigenvector pair.

3.b) Given $A = A^T$, and U is orthogonal, and $A = U\Lambda U^T$, then

$$AU = U\Lambda U^TU = U\Lambda$$

Following the same logic as in **3(a)**, we get

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

so $u^{(i)}$ is an eigenvector.

P.S.: $U^TU = I = U^{-1}U$, therefore $U^T = U^{-1}$

3.c) 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 egenvector $u^{(i)}$. Then we have

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

the ith standard basis vector. Using this, we have

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

From $\rightarrow$