Mathematics for AI
Homework 4

Read these first:

i You may write your solutions on paper, under a word processing software (MS-word, Libre Office, etc.), or under LaTeX.

ii If writing on paper, you must use a scanner device or a Camera Scanner (CamScanner) software to scan the document and submit a *single* PDF file. Also, write your answers neatly, in an organized and legible manner on paper.

iii Up to 15% extra score will be given to solutions written under LaTeX, provided that you follow either of the following conventions:

(a) Represent scalars with normal (italic) letters (a, A), vectors with bold lower-case letters ($\mathbf{a}$, using `\mathbf{a}`), and matrices with bold upper-case letters ($\mathbf{A}$, using `\mathbf{A}`), or

(b) represent scalars with normal (italic) letters (a, A), vectors with bold letters ($\mathbf{a}$, $\mathbf{A}$), and matrices with typewriter upper-case letters ($\mathtt{A}$, using `\mathtt{A}`).

(c) You latex document must contain a *title*, a *date*, and your name as the author.

(d) In all cases, you must submit a *single* PDF file.

(e) If writing under LaTeX, you must submit the *.tex* source (and other necessary source files if there are any) in addition to the PDF file.

Here is a short tutorial on LaTeX: `https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes`

Questions

For each questions, you may use the results of the previous questions (but not the following ones). Throughout this document, the operations diag() and Diag(·) are defined as follows:

- $\text{diag}(\mathtt{A})$ creates a vector $\in \mathbb{R}^n$ from the diagonal elements of the matrix $\mathtt{A} \in \mathbb{R}^{m\times n}$, and
- $\text{Diag}(\mathbf{x})$ creates an $n\times n$ diagonal matrix whose diagonal elements are the entries of $\mathbf{x} \in \mathbb{R}^n$.

Notice that both these operations are linear.

Adjoint

Consider two inner product spaces $\mathcal{U}$ and $\mathcal{V}$. A mapping $f^*\colon \mathcal{V} \to \mathcal{U}$ is called the *adjoint* of the linear map $f\colon \mathcal{U} \to \mathcal{V}$ if

$$\langle \mathbf{y}, f(\mathbf{x})\rangle = \langle f^*(\mathbf{y}), \mathbf{x}\rangle \,,$$

for all $\mathbf{x} \in \mathcal{U}$ and $\mathbf{y} \in \mathcal{V}$.

1. Show that for the linear map $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(\mathbf{x}) = \mathtt{A}\mathbf{x}$ with $\mathtt{A} \in \mathbb{R}^{m\times n}$ the adjoint is defined by $f^*(\mathbf{y}) = \mathtt{A}^T\mathbf{y}$.
2. Show that the diag(·) and Diag(·) operations defined above are adjoints of each other (with respect to the ordinary dot product defined in previous assignments).

Jacobian

3. Derive the Jacobian matrix for the following with respect to $\mathbf{x} \in \mathbb{R}^n$ using the directional derivative method.
   (a) $\text{Diag}(\mathbf{x})\,\mathbf{x}$,
   (b) $\text{Diag}(\mathbf{x})\,\mathbf{a}\,\mathbf{a}^T\,\mathbf{x}$ where $\mathbf{a} \in \mathbb{R}^n$,
   (c) $\mathtt{A}\,\text{Diag}(\mathbf{x})\,\mathbf{x}$ where $\mathtt{A} \in \mathbb{R}^{n\times n}$,
   (d) $(\mathbf{x}^T\mathtt{A}\mathbf{x})^2\,\mathtt{A}\,\mathbf{x}$ where $\mathtt{A} \in \mathbb{R}^{n\times n}$,

Quadratic Forms

4. Consider the quadratic form $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined as $f(\mathbf{x}) = \mathbf{x}^T\tilde{\mathtt{A}}\mathbf{x}$, where

$$\tilde{\mathtt{A}} = \begin{bmatrix} 1 & 2 \\ 4 & -1 \end{bmatrix}.$$

   Find a *symmetric* matrix $\mathtt{A}$ such that $f(\mathbf{x}) = \mathbf{x}^T\mathtt{A}\mathbf{x}$.
5. Show that for every function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined as $f(\mathbf{x}) = \mathbf{x}^T\tilde{\mathtt{A}}\mathbf{x}$, there exist a *symmetric* matrix $\mathtt{A}$ such that $f(\mathbf{x}) = \mathbf{x}^T\mathtt{A}\mathbf{x}$.

Hessian

6. Find the Gradient and the Hessian for the following with respect to $\mathbf{x} \in \mathbb{R}^n$

   (a) $\operatorname{diag}(\mathbf{x}\,\mathtt{A}\,\mathbf{x}^T)$,

   (b) $\mathbf{x}^T\,\operatorname{Diag}(\mathbf{x})\,\mathbf{x}$,

   (c) $(\mathbf{x}^T\mathtt{A}\mathbf{x})^2$ where $\mathtt{A} \in \mathbb{R}^{n\times n}$ is a *symmetric* matrix,

   (d) $(\mathbf{x}^T\mathtt{A}\mathbf{x})/(\mathbf{x}^T\mathtt{B}\mathbf{x})$ where $\mathtt{A}, \mathtt{B} \in \mathbb{R}^{n\times n}$ are *symmetric*,

   (e) $\sum_{i=1}^{n} \sqrt{\mathbf{x}^T\mathtt{A}_i\mathbf{x}}$ where $\mathtt{A}_1, \mathtt{A}_2, \ldots, \mathtt{A}_n \in \mathbb{R}^{n\times n}$ are (symmetric) positive definite.