Mathematics for AI
Homework 4

Read these first:

i To achieve the full score, you need to write your solutions using LaTeX. If you choose to write your solutions on paper or in a word processing software (e.g., MS Word, LibreOffice), you can receive up to 90% of the score.

ii If writing on paper, you must use a scanner or a camera scanning app (e.g., CamScanner) to scan the document and submit it as a *single* PDF file. Ensure your answers are written neatly, organized, and legible on paper.

iii When using LaTeX, follow one of these two conventions:

  (a) Represent scalars with italic letters (a, A), vectors with bold lowercase letters ($\mathbf{a}$, using `\mathbf{a}`), and matrices with bold uppercase letters ($\mathbf{A}$, using `\mathbf{A}`), or

  (b) Represent scalars with italic letters (a, A), vectors with bold letters ($\mathbf{a}$, $\mathbf{A}$), and matrices with typewriter uppercase letters ($\mathtt{A}$, using `\mathtt{A}`).

iv Your LaTeXdocument must include a *title*, a *date*, and your name as the *author*.

v If writing on paper, submit a *single* PDF file; do not send multiple image files.

vi If using LaTeX, submit the *.tex* source file (along with any other required source files) 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($\cdot$) are defined as follows:

- 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
- 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.

Multivariate Calculus

1. Show that for a symmetric matrix $\mathtt{B}$ the gradient of $1/(\mathbf{x}^T\mathtt{B}\mathbf{x})$ with respect to $\mathbf{x}$ is $-2\mathtt{B}\mathbf{x}/(\mathbf{x}^T\mathtt{B}\mathbf{x})^2$ wherever the gradient exists at $\mathbf{x}$.
2. Show that for symmetric matrices $\mathtt{A}$ and $\mathtt{B}$ the gradient of $f(\mathbf{x}) = (\mathbf{x}^T\mathtt{A}\mathbf{x})/(\mathbf{x}^T\mathtt{B}\mathbf{x})$ with respect to x is equal to
$$2(\mathtt{A}\mathbf{x}(\mathbf{x}^T\mathtt{B}\mathbf{x}) - \mathtt{B}\mathbf{x}(\mathbf{x}^T\mathtt{A}\mathbf{x}))/(\mathbf{x}^T\mathtt{B}\mathbf{x})2 = 2(\mathtt{A}\mathbf{x} - f(\mathbf{x})\,\mathtt{B}\mathbf{x})/(\mathbf{x}^T\mathtt{B}\mathbf{x}),$$
if the gradient exists at $\mathbf{x}$.
3. Let $\mathtt{A}$ be symmetric. Calculate the gradient of $\exp(-\mathbf{x}^T\mathtt{A}\mathbf{x})$ with respect to $\mathbf{x}$.
4. Let $\mathtt{A}$ be (symmetric) positive definite. Compute the gradient of $\log(1 + \mathbf{x}^T\mathtt{A}\mathbf{x})$ with respect to $\mathbf{x}$.
5. Consider the function $f(\mathbf{x}) = \mathbf{x}^T\mathtt{A}\mathbf{x}/\|\mathbf{x}\|^2 = \mathbf{x}^T\mathtt{A}\mathbf{x}/(\mathbf{x}^T\mathbf{x})$ defined for a symmetric matrix $\mathtt{A}$. Show that the critical points of f include $\mathbf{x} = \mathbf{0}$ plus the eigenvectors of $\mathtt{A}$. The critical points of a function f are those at which the gradient is zero or nonexistant.
6. Consider the function $f(\mathbf{x}) = \mathbf{x}^T\mathtt{A}\mathbf{x}/(\mathbf{x}^T\mathtt{B}\mathbf{x})$ defined for symmetric matrices $\mathtt{A}$ and $\mathtt{B}$. Show that if $\mathtt{B}$ is invertible then the critical points of f are either the points for which $\mathbf{x}^T\mathtt{B}\mathbf{x} = 0$ or the eigenvectors of $\mathtt{B}^{-1}\mathtt{A}$.

Jacobian

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

8. 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 $f(\mathbf{x}) = \mathbf{x}^T \mathtt{A}\mathbf{x}$.

9. Show that for every function $f\colon \mathbb{R}^n \to \mathbb{R}$ defined as $f(\mathbf{x}) = \mathbf{x}^T \tilde{\mathtt{A}}\mathbf{x}$ with $\tilde{\mathtt{A}} \in \mathbb{R}^{n\times n}$, there exists a *symmetric* matrix $\mathtt{A}$ such that $f(\mathbf{x}) = \mathbf{x}^T \mathtt{A}\mathbf{x}$.

Hessian

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

   (a) $\mathrm{trace}(\mathbf{x}\, \mathtt{A}\, \mathbf{x}^T)$,

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

   (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.