Mathematics for AI
Homework 3

**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 questions).

## Positive Definite Matrices

For all question in this section, by *positive definite* we mean *symmetric positive definite*.

1. Prove that a symmetric matrix is positive definite if and only if all its eigenvalues are positive. (Remember from the class that the eigen-decomposition of a symmetric matrix is in the form of $\mathtt{A} = \mathtt{V\Lambda V}^{-1}\mathtt{V\Lambda V}^{T}$.)

2. Show that the diagonal elements of a positive definite matrix are all positive definite.

3. Remember from the class that an operation $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ defined on a vector space $\mathcal{V}$ is an *inner product* if

   (a) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ for all $\mathbf{u} \in \mathcal{V}$,

   (b) $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$,

   (c) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v} \in \mathcal{V}$,

   (d) $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$.

   Let $\mathtt{A} \in \mathbb{R}^{n\times n}$ be any *positive definite* matrix. Show that the operation $\langle \cdot, \cdot \rangle_{\mathtt{A}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

   $$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathtt{A}} = \mathbf{u}^T \mathtt{A} \mathbf{v}$$

   is indeed an inner product.

## Singular Value Decomposition

4. Let $\mathtt{A}$ be a nonsingular square matrix and $A = \mathtt{U\Sigma V}^T$ be its (full) SVD. Prove that $\det(\mathtt{U})\det(\mathtt{V}) = \text{sign}(\det(\mathtt{A}))$, that is $\det(\mathtt{U})\det(\mathtt{V}) = 1$ if $det(\mathtt{A}) > 0$ and $\det(\mathtt{U})\det(\mathtt{V}) = 1$ if $det(\mathtt{A}) < 0$.

5. Show that for a symmetric positive definite matrix the eigenvalue decomposition $\mathtt{A} = \mathtt{V\Lambda V}^{-1} = \mathtt{V\Lambda V}^T$ is the same as the singular value decomposition.

6. Find a way to obtain the SVD of a symmetric matrix from its eigenvalue decomposition $\mathtt{A} = \mathtt{V\Lambda V}^T$. Notice that the diagonal elements of $\mathtt{\Lambda}$ might be negative.

7. Consider a matrix $\mathtt{A} \in \mathbb{R}^{m\times n}$ and two orthogonal matrices $\mathtt{P} \in \mathbb{R}^{m\times m}$ and $\mathtt{Q} \in \mathbb{R}^{n\times n}$. Show that the singular values of $\mathtt{PAQ}$ is the same as the singular values of $\mathtt{A}$.

## Matrix inner product

8. Perhaps the simplest way to define an inner product between a pair of matrices $\mathtt{A}, \mathtt{B} \in \mathbb{R}^{m\times n}$ is $\langle \mathtt{A}, \mathtt{B}\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ij}$. This is the same as vectorizing the matrices and taking their dot product, and is sometimes called the *Frobenius Inner Product*.

   (a) Prove that real matrices $\langle \mathtt{A}, \mathtt{B}\rangle = \text{trace}(\mathtt{A}^T\mathtt{B}) = \text{trace}(\mathtt{B}^T\mathtt{A}) = \text{trace}(\mathtt{A}\mathtt{B}^T)$, where $\text{trace}(\mathtt{S}) = \sum_i S_{ii}$ gives the sum of the diagonal elements of a square matrix $\mathtt{S}$.

   (b) Prove that $\langle \mathtt{AB}, \mathtt{C}\rangle = \left\langle \mathtt{B}, \mathtt{A}^T\mathtt{C}\right\rangle = \left\langle \mathtt{A}, \mathtt{C}\mathtt{B}^T\right\rangle$ Hint: $(\mathtt{AB})^T = \mathtt{B}^T\mathtt{A}^T$.

   Note: Same results hold for complex matrices by replacing the transpose operation with conjugate transpose: $\langle \mathtt{AB}, \mathtt{C}\rangle = \langle \mathtt{B}, \mathtt{A}^*\mathtt{C}\rangle = \langle \mathtt{A}, \mathtt{C}\mathtt{B}^*\rangle$.

## Matrix Norms

9. Show that the squred Frobenius norm is the same as the Frobenius inner product of a matrix by itself, that is $\|\mathtt{A}\|_F^2 = \langle \mathtt{A}, \mathtt{A}\rangle$.

10. A matrix norm is called *Unitarily Invariant* if $\|\mathtt{A}\| = \|\mathtt{UAV}\|$ for any orthogonal matrices $\mathtt{U}$ and $\mathtt{V}$ of compatible size. Using the above and the properties of matrix inner product prove that the Frobenius norm is unitarily invariant. Notice that for orthogonal matrices we have $\mathtt{U}^T\mathtt{U} = \mathtt{U}\mathtt{U}^T = \mathtt{I}$. (A more general definition that also works for complex matrices is when $\mathtt{U}$ and $\mathtt{V}$ are unitary, that is $\mathtt{U}^*\mathtt{U} = \mathtt{U}\mathtt{U}^* = \mathtt{I}$).

11. Use Question 7 to prove that the *spectral norm* and *nuclear norm* are also unitarily invariant.

## Multivariate Calculus

12. 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 $-2\mathtt{B}\mathbf{x}/(\mathbf{x}^T\mathtt{B}\mathbf{x})^2$ (if the gradient exists at $\mathbf{x}$).

13. 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 $\mathbf{x}$.

14. Let $\mathtt{A}$ be symmetric. Calculate the gradient of $\exp(-\mathbf{x}^T\mathtt{A}\mathbf{x})$ with respect to $\mathbf{x}$.

15. Let $\mathtt{A}$ be (symmetric) positive definite. Compute the gradient of $\log(1 + \mathbf{x}^T\mathtt{A}\mathbf{x})$ with respect to $\mathbf{x}$.

16. 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 $f$ are exactly the eigenvectors of $\mathtt{A}$. The critical points of a function $f$ are points $\mathbf{x}$ at which the gradient is zero or nonexistant.

17. 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 $\mathtt{B}^{-1}\mathtt{A}$.