Problem 1 - PSD matrix properties

Recall the definition that a symmetric matrix $M \in \mathbb{R}^{n\times n}$ is a positive semidefinite (PSD) matrix if $x^\top Mx \geq 0$ for all $x \in \mathbb{R}^n$. We use the notation $M \succeq 0$ or $M \in \mathbb{S}^n_+$ to denote that M is PSD.

For this problem, assume $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix.

(a) Show that if $A \succeq 0$ then all diagonal entries of A are non-negative, denoted as $A_{ii} \geq 0$ for $i = 1, ..., n$. *Hint: use a vector e_i, defined as the vector containing 1 at the ith position and 0 elsewhere.*

> Solution: Given that A is PSD, the quadratic form $x^\top Ax \geq 0$ applies for any arbitrary vector $x \in \mathbb{R}^n$. Therefore, we can design a vector that will pull out A_{ii}, the ith unit vector, and also fit this form.
>
> Say our e_i is a column vector of dimension $n \times 1$. We have $Ae_i = a_i$, where a_i is the ith column of A (also a column vector of dimension $n \times 1$). Since e_i is a column vector, e_i^T is a row vector. Multiplying $e_i^\top a_i$ then pulls out the ith element of the ith column, which is A_{ii}. So $e_i^\top Ae_i = A_{ii} \geq 0$. ■

(b) Show that if $A \in \mathbb{S}^n_+$, then there exists $P \in \mathbb{S}^n_+$ such that $A = P^2$. The matrix P is called a PSD square root of A.

> Solution: Since A is symmetric and PSD, it has non-negative eigenvalues. Furthermore, its eigenvectors are orthogonal due to the spectral theorem covered in Lecture 7. Hence, A can be diagonalized as $A = U\Lambda U^\top$, where the diagonal matrix of eigenvalues Λ has all non-negative diagonal entries, and each column of U is the corresponding eigenvector with length 1. By construction, U is an orthogonal matrix, and therefore $U^\top = U^{-1}$.
>
> Define a matrix $A^{1/2} = U\Lambda^{1/2}U^\top$, where $\Lambda^{1/2}$ is a diagonal matrix with the square roots of the eigenvalues of A. Observe that $A^{1/2}$ is PSD since its eigenvalues are still non-negative. Substituting $A^{1/2}$ for P, we have
>
> $$\begin{aligned}(A^{1/2})^\top A^{1/2} &= (U\Lambda^{1/2}U^\top)^\top(U\Lambda^{1/2}U^\top)\\ &= U\Lambda^{1/2}U^\top U\Lambda^{1/2}U^\top\\ &= U\Lambda^{1/2}(U^{-1}U)\Lambda^{1/2}U^\top\\ &= U\Lambda U^\top\\ &= A.\end{aligned}$$
>
> ■

(c) Show that for any matrix $Q \in \mathbb{R}^{m \times n}$, if $A = Q^\top Q$ then $A \in \mathbb{S}^n_+$.

(Note: Consider the special case when Q is symmetric PSD and $A = Q^2$. Then (c) and (b) together show that a matrix is PSD if and only if it has a PSD square root.)

SOLUTION: Simply use the quadratic form and the fact that norms are non-negative:

$$x^\top A x = x^\top Q^\top Q x = (Qx)^\top (Qx) = ||Qx||_2^2 \geq 0.$$

■

(d) Let $B \in \mathbb{R}^{m \times n}$ be an arbitrary matrix, meaning it may not be symmetric, PSD, or square. From the previous parts, we know that matrices $BB^\top \in \mathbb{R}^{m \times m}$ and $B^\top B \in \mathbb{R}^{n \times n}$ are PSD, and therefore have real non-negative eigenvalues. Prove that the non-zero eigenvalues of $BB^\top$ are the same as the non-zero eigenvalues of $B^\top B$.

(Note: These forms of PSD matrices are useful when constructing singular value decompositions, which we'll cover more later in the course. The point of this problem is to show that we can choose either construction, since they yield equivalent non-zero eigenvalues.)

SOLUTION: We can show the non-zero eigenvalues are the same by showing that non-zero eigenvalues of $B^\top B$ are also non-zero eigenvalues of $BB^\top$ and vice versa (both directions to give equivalence).

Let $\lambda \neq 0, v \neq 0$ be an eigenpair of $B^\top B$. By definition, we have

$$(B^\top B)v = \lambda v$$

where both sides are non-zero. Multiplying by B and regrouping yields

$$(BB^\top)Bv = \lambda Bv$$

where we know Bv is non-zero from the first equation. Therefore, $\lambda \neq 0, Bv \neq 0$ is an eigenpair of $BB^\top$, so both $B^\top B$ and $BB^\top$ have non-zero eigenvalue λ.

Similarly, if we instead let $\lambda \neq 0, v \neq 0$ be an eigenpair of $BB^\top$, we have

$$(BB^\top)v = \lambda v$$

and

$$(B^\top B)B^\top v = \lambda B^\top v$$

so $\lambda \neq 0, B^\top v \neq 0$ is an eigenpair of $B^\top B$, so both $B^\top B$ and $BB^\top$ have non-zero eigenvalue λ. ■

Problem 2 - Gradient and Hessian

Consider the function

$$f(x) = e^{x_1} + e^{x_2} + \cdots + e^{x_n}$$

for a vector $x \in \mathbb{R}^n$.

(a) Calculate the gradient of $f(x)$.

SOLUTION: The function is

$$f(x) = e^{x_1} + e^{x_2} + \cdots + e^{x_n} = \sum_{i=1}^{n} e^{x_i}.$$

Taking the partial derivative of $f(x)$ with respect to any x_i gives

$$\frac{\delta f(x)}{\delta x_i} = \frac{\delta}{\delta x_i}(e^{x_1} + e^{x_2} + \cdots + e^{x_n}) = e^{x_i}.$$

The gradient is then

$$\nabla f(x) = \begin{bmatrix} e^{x_1} \\ e^{x_2} \\ \vdots \\ e^{x_n} \end{bmatrix}.$$

■

(b) Calculate the Hessian of $f(x)$.

SOLUTION: Each element of the Hessian can be represented as

$$H_{ij} = \frac{\delta^2 f(x)}{\delta x_i \delta x_j}$$

with two cases: diagonal elements where $i = j$, and off-diagonal elements where $i \neq j$. For the diagonal elements, we have

$$H_{ii} = \frac{\delta}{\delta x_i}\left(\frac{\delta f(x)}{\delta x_i}\right) = \frac{\delta}{\delta x_i}\left(e^{x_i}\right) = e^{x_i}.$$

For the off-diagonal elements, we have

$$H_{ij} = \frac{\delta}{\delta x_j}\left(\frac{\delta f(x)}{\delta x_i}\right) = \frac{\delta}{\delta x_j}\left(e^{x_i}\right) = 0.$$

The Hessian is then

$$H_{f(x)} = \begin{bmatrix} e^{x_1} & 0 & \dots & 0 \\ 0 & e^{x_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{x_n} \end{bmatrix} = \text{diag}(e^{x_1}, e^{x_2}, ..., e^{x_n}).$$

■

(c) Determine whether the Hessian of $f(x)$ is PSD.

SOLUTION: We show the Hessian is PSD using the quadratic form with some $v \in \mathbb{R}^n$

$$v^\top H_{f(x)} v = \sum_{i=1}^{n} e^{x_i} v^\top v = \sum_{i=1}^{n} e^{x_i} ||v||_2^2 \geq 0$$

where the final inequality holds because $e^{x_i} > 0$ and norms are non-negative for real numbers. ■

Problem 3 - Eigenvectors of a symmetric 2x2 matrix

Let $p, q \in \mathbb{R}^n$ be two linearly independent vectors, with unit norm ($||p||_2 = ||q||_2 = 1$). Define the matrix $A := pq^\top + qp^\top$. In your derivations, it may be useful to use the notation $c := p^\top q$.

(a) Show that A is symmetric.

Solution: Show symmetry by showing $A^\top = A$. We have

$$A^\top = (pq^\top + qp^\top)^\top = qp^\top + pq^\top = A.$$

■

(b) Show that $p+q$ and $p-q$ are eigenvectors of A, and determine the corresponding eigenvalues.

Solution: If $p+q$ and $p-q$ are eigenvectors of A, there exist some eigenvalues λ_1, λ_2 for which $A(p+q) = \lambda_1(p+q)$ and $A(p-q) = \lambda_2(p-q)$. We can find them by expanding

$$\begin{aligned}
A(p+q) &= (pq^\top + qp^\top)p + (pq^\top + qp^\top)q \\
&= pq^\top p + qp^\top p + pq^\top q + qp^\top q \\
&= p(q^\top p + q^\top q) + q(p^\top p + p^\top q) \\
&= p(q^\top p + 1) + q(1 + p^\top q) \\
&= (p^\top q + 1)p + q(1 + p^\top q) \\
&= (c+1)(p+q)
\end{aligned}$$

So $c+1$ is the corresponding eigenvalue for $p+q$. Similarly, we have

$$\begin{aligned}
A(p-q) &= (pq^\top + qp^\top)p - (pq^\top + qp^\top)q \\
&= p(q^\top p - q^\top q) - q(p^\top p - p^\top q) \\
&= (c-1)(p-q)
\end{aligned}$$

So $c-1$ is the corresponding eigenvalue for $p-q$.

A quicker and more elegant solution would be to observe that

$$Ap = cp + q, \quad Aq = p + cq$$

from which we can immediately factor

$$A(p+q) = Ap + Aq = cp + q + p + cq = (c+1)(p+q)$$

and

$$A(p-q) = Ap - Aq = cp + q - p - cq = (c-1)(p-q).$$

■

(c) Determine the nullspace and rank of A.

SOLUTION: For vector $x \in \mathbb{R}^n$ in the nullspace of A, we must have $Ax = 0$. Expanding, we have

$$Ax = p(q^\top x) + q(p^\top x) = 0.$$

Noting that $(q^\top x)$ and $(p^\top x)$ are scalars, and that p and q are linearly independent, this implies

$$q^\top x = p^\top x = 0.$$

So the nullspace of A is the set of vectors orthogonal to p and q,

$$\text{Null}(A) = \text{span}(p, q)^\perp.$$

Since A is a symmetric matrix, we have the following relationships between the row space and nullspace of A:

$$\mathcal{R}(A) = \mathcal{R}(A^\top) = \text{Null}(A)^\perp = (\text{span}(p, q)^\perp)^\perp = \text{span}(p, q).$$

Since p and q are orthogonal, the rank of A is 2. ∎

(d) Use the previous two parts to find an eigenvalue decomposition of A, in terms of p, q.

SOLUTION: We already know that $p + q$ and $p - q$ are eigenvectors with eigenvalues $1 + c$ and $1 - c$, respectively (where $c = p^\top q$). A has rank 2, so we need to verify that these two eigenvalues are nonzero.

We have $p - q \neq 0$, meaning $||p - q||_2^2 \geq 0$ and $||p||_2^2 + ||q||_2^2 - 2p^\top q \geq 0$. Therefore, $c < 1$. A similar proof for $p + q$ gives $-c < 1$. So we have $|c| < 1$, and the linearly independent vectors $p + q$ and $p - q$ do not belong to the nullspace of A.

We then write the eigenvalue decomposition as

$$A = (c + 1)v_1 v_1^\top + (c - 1)v_2 v_2^\top$$

where we have normalized vectors

$$v_1 = \frac{p + q}{||p + q||_2}, \quad v_2 = \frac{p - q}{||p - q||_2}.$$

As an extension, observe that we can further simplify by expanding the squared denominators of the normalized vectors

$$||p + q||_2^2 = p^\top p + 2p^\top q + q^\top q = 2(1 + c)$$
$$||p - q||_2^2 = p^\top p - 2p^\top q + q^\top q = 2(1 - c)$$

to yield normalized vectors of the form

$$v_1 = \frac{1}{\sqrt{2(1+c)}}(p+q)$$
$$v_2 = \frac{1}{\sqrt{2(1-c)}}(p-q)$$

The eigenvalue decomposition can be written just in terms of p and q as

$$A = \frac{1}{2}\left((p+q)(p+q)^\top - (p-q)(p-q)^\top\right).$$

$\blacksquare$