Homework 2

Self grades are due at 11 PM on September 20th, 2023.

1. Norms

(a) Show that the following inequalities hold for any vector $\vec{x} \in \mathbb{R}^n$:

$$\frac{1}{\sqrt{n}} \|\vec{x}\|_2 \leq \|\vec{x}\|_\infty \leq \|\vec{x}\|_2 \leq \|\vec{x}\|_1 \leq \sqrt{n} \|\vec{x}\|_2 \leq n \|\vec{x}\|_\infty . \tag{1}$$

As an aside: note that we can interpret different norms as different ways of computing distance between two points $\vec{x}, \vec{y} \in \mathbb{R}^2$. The ℓ_2 norm is the distance as the crow flies (i.e. point-to-point distance), the ℓ_1 norm, also known as the Manhattan distance is the distance you would have to cover if you were to navigate from $\vec{x}$ to $\vec{y}$ via a rectangular street grid, and the ℓ_∞ norm is the maximum distance that you have to travel in either the north-south or the east-west direction.

Solution: We have

$$\|\vec{x}\|_2^2 = \sum_{i=1}^{n} x_i^2 \leq n \cdot \max_i \, x_i^2 = n \cdot \|\vec{x}\|_\infty^2 . \tag{2}$$

Also, $\|\vec{x}\|_\infty \leq \sqrt{x_1^2 + \ldots + x_n^2} = \|\vec{x}\|_2$. The inequality $\|\vec{x}\|_2 \leq \|\vec{x}\|_1$ is obtained after squaring both sides, and checking that

$$\sum_{i=1}^{n} x_i^2 \leq \sum_{i=1}^{n} x_i^2 + \sum_{i \neq j} |x_i x_j| = \left(\sum_{i=1}^{n} |x_i| \right)^2 = \|\vec{x}\|_1^2 . \tag{3}$$

The condition $\|\vec{x}\|_1 \leq \sqrt{n} \|\vec{x}\|_2$ is due to the Cauchy-Schwarz inequality

$$\left|\vec{z}^\top \vec{y}\right| \leq \|\vec{y}\|_2 \cdot \|\vec{z}\|_2 , \tag{4}$$

applied to the two vectors $y = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^\top$ and $\vec{z} = |\vec{x}| = \begin{bmatrix} |x_1| & \cdots & |x_n| \end{bmatrix}$.

Finally, $\sqrt{n} \|\vec{x}\|_2 \leq n \|\vec{x}\|_\infty$, is achieved by an algebraic manipulation of the first derived bound using the fact that $\sqrt{n} = \frac{n}{\sqrt{n}}$.

(b) We define the *cardinality* of the vector $\vec{x}$ as the number of non-zero elements in $\vec{x}$. This is also commonly known as the ℓ_0 norm of the vector $\vec{x}$, denoted by $\|\vec{x}\|_0$. Show that for any non-zero vector x,

$$\|\vec{x}\|_0 \geq \frac{\|\vec{x}\|_1^2}{\|\vec{x}\|_2^2} . \tag{5}$$

Find all vectors $\vec{x}$ for which the lower bound is attained.

Solution: Let us apply the Cauchy-Schwarz inequality with $\vec{z} = |\vec{x}|$ again, and with $\vec{y}$ a vector with $y_i = 1$ if $x_i \neq 0$, and $y_i = 0$ otherwise. We have $\|\vec{y}\|_2 = \sqrt{k}$, with $k = \|\vec{x}\|_0$. Hence

$$\left|\vec{z}^\top \vec{y}\right| = \|\vec{x}\|_1 \leq \|\vec{y}\|_2 \cdot \|\vec{z}\|_2 = \sqrt{k} \cdot \|\vec{x}\|_2 , \tag{6}$$

which proves the result. The bound is attained for vectors with k non-zero elements, all with the same magnitude.

2. Distinct Eigenvalues, Orthogonal Eigenspaces

Let $A \in \mathbb{S}^n$ (i.e. the set of $n \times n$ symmetric matrices) and $(\lambda_1, \vec{u}_1), (\lambda_2, \vec{u}_2), \lambda_1 \neq \lambda_2$ be distinct eigen-pairs of A. Show that $\langle \vec{u}_1,\ \vec{u}_2 \rangle = 0$, i.e eigenspaces corresponding to distinct eigenvalues are mutually orthogonal.

Note: This exercise is part of the proof of the spectral theorem.

Solution: It is useful to note the following equality:

$$\langle A\vec{x},\ \vec{y} \rangle = \vec{x}^\top A^\top \vec{y} = \left\langle \vec{x},\ A^\top \vec{y} \right\rangle . \tag{7}$$

Now comes the massaging of equations:

$$\begin{aligned}
\lambda_1 \langle \vec{u}_1,\ \vec{u}_2 \rangle &= \langle \lambda_1 \vec{u}_1,\ \vec{u}_2 \rangle && \text{Linearity of inner product} && (8)\\
&= \langle A\vec{u}_1,\ \vec{u}_2 \rangle && A\vec{u}_1 = \lambda_1 \vec{u}_1 && (9)\\
&= \left\langle \vec{u}_1,\ A^\top \vec{u}_2 \right\rangle && \text{Equation 7} && (10)\\
&= \langle \vec{u}_1,\ A\vec{u}_2 \rangle && A \in \mathbb{S}^n && (11)\\
&= \langle \vec{u}_1,\ \lambda_2 \vec{u}_2 \rangle && A\vec{u}_2 = \lambda_2 \vec{u}_2 && (12)\\
&= \lambda_2 \langle \vec{u}_1,\ \vec{u}_2 \rangle . && \text{Linearity of inner product} && (13)\\
\implies \lambda_1 \langle \vec{u}_1,\ \vec{u}_2 \rangle &= \lambda_2 \langle \vec{u}_1,\ \vec{u}_2 \rangle && && (14)\\
\underbrace{(\lambda_1 - \lambda_2)}_{\neq 0} \langle \vec{u}_1,\ \vec{u}_2 \rangle &= 0. && && (15)\\
& && && (16)
\end{aligned}$$

Thus, $\langle \vec{u}_1,\ \vec{u}_2 \rangle = 0$ for any $\vec{u}_1, \vec{u}_2$ corresponding to different eigenvalues. Stated differently, unique eigenvalues correspond to orthogonal eigenvectors.

This, in combination with the fact that the geometric multiplicity and algebraic multiplicity of a symmetric matrix are equal, allows us to construct an orthonormal set of eigenvectors. First, find all the distinct eigenvalues and their respective eigenvectors. Then, for all eigenvalues with algebraic multiplicity > 1, we know that the respective eigenspace is spanned by k linearly independent eigenvectors. Utilizing Gram-Schmidt, we can construct an orthonormal set of eigenvectors from this basis for this eigenspace. Putting the eigenvectors from these two cases together, we have constructed the U matrix of the decomposition.

3. Eigenvectors of a Symmetric Matrix

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

(a) Show that A is symmetric.

Solution: We have

$$A^\top = (\vec{p}\vec{q}^\top + \vec{q}\vec{p}^\top)^\top = \vec{q}\vec{p}^\top + \vec{p}\vec{q}^\top = A. \tag{17}$$

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

Solution: We have

$$A\vec{p} = c\vec{p} + \vec{q}, \quad A\vec{q} = \vec{p} + c\vec{q}, \tag{18}$$

from which we obtain

$$A(\vec{p} - \vec{q}) = (c - 1)(\vec{p} - \vec{q}), \quad A(\vec{p} + \vec{q}) = (c + 1)(\vec{p} + \vec{q}). \tag{19}$$

Thus $\vec{u}_\pm := \vec{p} \pm \vec{q}$ is an (un-normalized) eigenvector of A, with eigenvalue $c \pm 1$.

(c) Determine the nullspace and rank of A.

Solution: If $\vec{x} \in \mathbb{R}^n$ is in the nullspace of A we must have: $A\vec{x} = 0$.

$$0 = A\vec{x} = \vec{p}(\vec{q}^\top \vec{x}) + \vec{q}(\vec{p}^\top \vec{x}). \tag{20}$$

Since $(\vec{q}^\top \vec{x})$ and $(\vec{p}^\top \vec{x})$ are scalars we can rewrite this as:

$$0 = A\vec{x} = (\vec{q}^\top \vec{x})\vec{p} + (\vec{p}^\top \vec{x})\vec{q} = 0. \tag{21}$$

However, since $\vec{p}, \vec{q}$ are linearly independent, the fact that a linear combination of $\vec{p}, \vec{q}$ is zero implies that $\vec{p}^\top \vec{x} = \vec{q}^\top \vec{x} = 0$. Hence, the nullspace of A is contained in the set of vectors orthogonal to $\vec{p}$ and $\vec{q}$, i.e., $\mathcal{N}(A) \subset \text{span}(\vec{p}, \vec{q})^\perp$. Next, observe that $\text{rank}(A) \leq \text{rank}(\vec{p}\vec{q}^\top) + \text{rank}(\vec{q}\vec{p}^\top) \leq 1 + 1 = 2$. Thus, by the rank-nullity theorem, $\dim(\mathcal{N}(A)) = n - \text{rank}(A) \geq n - 2$. Meanwhile, $\text{span}(\vec{p}, \vec{q})^\perp$ is an $(n-2)$-dimensional space containing $\mathcal{N}(A)$, so the dimension of $\mathcal{N}(A)$ must be $n - 2$ as well. Thus, $\mathcal{N}(A) = \text{span}(\vec{p}, \vec{q})^\perp$, as desired. To compute the rank of A, we start with the fundamental theorem of linear algebra and the fact that A is symmetric:

$$\mathcal{R}(A) = \mathcal{R}\left(A^\top\right) = \mathcal{N}(A)^\perp = (\text{span}(\vec{p}, \vec{q})^\perp)^\perp = \text{span}(\vec{p}, \vec{q}). \tag{22}$$

And since p and q are linearly independent, $\text{rank}(A) = 2$.

(d) Find an eigenvalue decomposition of A, in terms of $\vec{p}, \vec{q}$. *HINT: Use the previous two parts.*

Solution: Since the rank is 2, we need to find a total of two non-zero eigenvalues. First, we check that $\lambda = c \pm 1$ is not 0. We have $\vec{p} - \vec{q} \neq 0$ which implies $\|\vec{p} - \vec{q}\|_2^2 > 0$ which means $\|\vec{p}\|_2^2 + \|\vec{q}\|_2^2 - 2\vec{p}^\top \vec{q} > 0$. Therefore, we have $c < 1$ and through a similar proof with $\vec{p} + \vec{q}$, we have $-c < 1$. From these two facts, we get $|c| < 1$. Thus, we have found two linearly independent eigenvectors $\vec{u}_\pm = \vec{p} \pm \vec{q}$ that do not belong to the nullspace. Then, the eigenvalue decomposition is

$$A = (c - 1)\vec{v}_- \vec{v}_-^\top + (c + 1)\vec{v}_+ \vec{v}_+^\top, \tag{23}$$

where $\vec{v}_{\pm}$ are the normalized vectors $\vec{v}_{\pm} = \vec{u}_{\pm} / \|\vec{u}_{\pm}\|_2$. Since

$$\|\vec{p} \pm \vec{q}\|_2^2 = \vec{p}^\top \vec{p} \pm 2\vec{p}^\top \vec{q} + \vec{q}^\top \vec{q} = 2(1 \pm c), \tag{24}$$

we have

$$\vec{v}_{\pm} = \frac{1}{\sqrt{2(1 \pm c)}} (\vec{p} \pm \vec{q}), \tag{25}$$

so that the eigenvalue decomposition becomes

$$A = \frac{1}{2} \left((\vec{p} + \vec{q})(\vec{p} + \vec{q})^\top - (\vec{p} - \vec{q})(\vec{p} - \vec{q})^\top \right). \tag{26}$$

(e) **(OPTIONAL)** Now consider general vectors $\vec{p}_{\text{new}}, \vec{q}_{\text{new}}$ that are scaled versions of $\vec{p}, \vec{q}$. Note that $\vec{p}_{\text{new}}, \vec{q}_{\text{new}}$ are not necessarily norm 1. Define the matrix $A_{\text{new}} := \vec{p}_{\text{new}} \vec{q}_{\text{new}}^\top + \vec{q}_{\text{new}} \vec{p}_{\text{new}}^\top$.

Write A_{new} as a function of $\vec{p}, \vec{q}$ and the norms of $\vec{p}_{\text{new}}, \vec{q}_{\text{new}}$, and the eigenvalues of matrix A_{new} as a function of $\vec{p}, \vec{q}$ and the norms of $\vec{p}_{\text{new}}, \vec{q}_{\text{new}}$.

Solution: We can write the vectors $\vec{p}_{\text{new}}, \vec{q}_{\text{new}}$ as scaled versions of the norm 1 vectors:

$$\vec{p}_{\text{new}} = \|\vec{p}_{\text{new}}\|_2 \, \vec{p}, \; \vec{q}_{\text{new}} = \|\vec{q}_{\text{new}}\|_2 \, \vec{q}. \tag{27}$$

Then the matrix A_{new} can be written as

$$A_{\text{new}} = \vec{p}_{\text{new}} \vec{q}_{\text{new}}^\top + \vec{q}_{\text{new}} \vec{p}_{\text{new}}^\top = \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 \left(\vec{p}\vec{q}^\top + \vec{q}\vec{p}^\top \right) = \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 A. \tag{28}$$

Since A_{new} is just a scaled version of A, whose eigenvalues we have already determined in previous parts, the eigenvalues are scaled accordingly. The eigenvalues of A_{new} are given by,

$$\lambda_{\pm} = \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 (c \pm 1) = \vec{p}_{\text{new}}^\top \vec{q}_{\text{new}} \pm \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2. \tag{29}$$

The unit norm eigenvectors of A_{new} are same as that of A which gives,

$$A_{\text{new}} = \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 A = \frac{\|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2}{2} \left((\vec{p} + \vec{q})(\vec{p} + \vec{q})^\top - (\vec{p} - \vec{q})(\vec{p} - \vec{q})^\top \right). \tag{30}$$

4. PSD Matrices

In this problem, we will analyze properties of positive semidefinite (PSD) matrices. A matrix M is a PSD matrix if all its eigenvalues are non-negative, and we denote that as $M \succeq 0$.
Assume $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix.

(a) Show that $\forall \vec{x} \in \mathbb{R}^n, \vec{x}^\top A\vec{x} \geq 0 \iff$ all eigenvalues of A are non-negative.

Solution: $\implies$:

i. Solution 1: We can plug in the Spectral Decomposition here:

$$\vec{x}^\top A\vec{x} = \vec{x}^\top U\Sigma U^\top \vec{x} = \vec{v}^\top \Sigma \vec{v} \geq 0, \tag{31}$$

where $\vec{v} := U^\top \vec{x}$ is a rotated version of $\vec{x}$ since U is orthonormal. Now, we just need to convert that final quadratic into any eigenvalue of A, and we can do that by choosing a $\vec{v}$ that pulls out whichever eigenvalue we want (e.g. if we want the first eigenvalue, we can choose the first unit vector). To be thorough, we can then realize that the set of $\vec{x}$'s such that $U^\top \vec{x} = \vec{e}_i$ for any unit vector, will pull out the ith eigenvalue, thus satisfying definition 2.

ii. Solution 2: We can just use the definition of an eigenvalue:

$$\vec{x}^\top A\vec{x} = \vec{x}\lambda\vec{x} = \lambda\vec{x}^\top \vec{x} = \lambda \|\vec{x}\|_2^2 \tag{32}$$

Since norms/anything squared is always non-negative, in order for $\lambda \|\vec{x}\|_2^2 \geq 0$, λ must be non-negative.

$\impliedby$: Using the Spectral Decomposition again, we arrive at the equation $\vec{v}^\top \Sigma \vec{v}$, which we can expand further:

$$\vec{v}^\top \Sigma \vec{v} = \sum_i \lambda_i v_i^2 \geq 0, \tag{33}$$

where the last inequality came from the fact that anything squared is non-negative and all eigenvalues are non-negative by assumption of the problem.

Now we will show that a symmetric matrix A is positive semidefinite if and only if there exists a symmetric matrix $P \in \mathbb{R}^{n \times n}$ such that $A = P^\top P$.

(b) First, show that A having non-negative eigenvalues allows us to decompose $A = P^\top P$ where $P \succeq 0$.

Solution: With all non-negative eigenvalues, we are able to define a matrix $A^{\frac{1}{2}} = U\Sigma^{\frac{1}{2}}U^\top$, where $\Sigma^{\frac{1}{2}}$ is a diagonal matrix with the square roots of A's eigenvalues. Note that $A^{\frac{1}{2}}$ is PSD since its eigenvalues are still non-negative. Thus, with $P = A^{\frac{1}{2}}$, we can show the following:

$$P^\top P = (A^{\frac{1}{2}})^\top A^{\frac{1}{2}} = (U\Sigma^{\frac{1}{2}}U^\top)^\top U\Sigma^{\frac{1}{2}}U^\top = U\Sigma^{\frac{1}{2}}U^\top U\Sigma^{\frac{1}{2}}U^\top = U\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}U^\top = U\Sigma U^\top = A. \tag{34}$$

(c) Now, show that any matrix of the form $A = P^\top P$ is positive semidefinite, i.e. $A \succeq 0$.

Solution: We can plug in $A = P^\top P$ into the quadratic form as follows:

$$\vec{x}^\top A\vec{x} = \vec{x}^\top P^\top P\vec{x} = \langle P\vec{x},\ P\vec{x}\rangle = \|P\vec{x}\|_2^2 \geq 0. \tag{35}$$

(d) Show that if $A \succeq 0$ then all diagonal entries of A are non-negative, $A_{ii} \geq 0$.

Solution: The quadratic form $\vec{x}^\top A\vec{x} \geq 0$ applies for all vectors $\vec{x}$. Therefore, let's choose a vector that will pull out A_{ii}: the ith unit vector. $A\vec{e}_i$ pulls out the ith column $\vec{a}_i$, followed by $\vec{e}_i^\top \vec{a}_i$, which will pull out the ith element of the ith column. Therefore, $\vec{e}_i^\top A\vec{e}_i = A_{ii} \geq 0$.

5. SVD Transformation

In this problem we will interpret the linear map corresponding to a matrix $A \in \mathbb{R}^{n\times n}$ by looking at its singular value decomposition, $A = UDV^\top$. Recall that here $U, D, V \in \mathbb{R}^{n\times n}$ and U, V are orthonormal matrices while D is a diagonal matrix. We will first look at how $V^\top, D$ and U each separately transform the unit circle $C = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ and then look at their effect as a whole. This problem has an associated jupyter notebook, `svd_transformation.ipynb` that contains several parts (b, c, d, e) of the problem. These sub-parts can be answered in the notebook itself in the space provided and can be submitted as an attachment to this PDF using the "Download as PDF" feature that Jupyter notebook supports.

(a) Show that $V^\top\vec{x}$ represents $\vec{x}$ in the basis defined by the columns of V. Recall: $V^\top V = I$.

Solution: Since the columns of V form a basis for $\mathbb{R}^n$, we can represent $\vec{x} \in \mathbb{R}^n$ as $V\vec{z}$ for some $\vec{z}$ in $\mathbb{R}^n$. Then,

$$V^\top\vec{x} = V^\top V\vec{z} \tag{36}$$

$$= \vec{z}. \tag{37}$$

The last equality follows since V is an orthonormal matrix.

For the rest of the problem we restrict ourselves to the case where $A \in \mathbb{R}^{2\times 2}$ and move to the Jupyter notebook.

Solution: The solutions for the rest of the parts can be found in the Jupyter notebook solution.

6. Homework Process

With whom did you work on this homework? List the names and SIDs of your group members.

NOTE: If you didn't work with anyone, you can put "none" as your answer.