Homework 2

Self grades are due at 11 PM on February 9th, 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. **Gram Schmidt**

Any set of n linearly independent vectors in $\mathbb{R}^n$ could be used as a basis for $\mathbb{R}^n$. However, certain bases could be more suitable for certain operations than others. For example, an orthonormal basis could facilitate solving linear equations.

(a) Given a matrix $A \in \mathbb{R}^{n\times n}$, it could be represented as a multiplication of two matrices

$$A = QR,$$

where Q is an orthonormal in $\mathbb{R}^n$ and R is an upper-triangular matrix. For the matrix A, describe how Gram-Schmidt process could be used to find the Q and R matrices, and apply this to

$$A = \begin{bmatrix} 3 & -3 & 1 \\ 4 & -4 & -7 \\ 0 & 3 & 3 \end{bmatrix}$$

to find an orthogonal matrix Q and an upper-triangular matrix R.

Solution: Let a_i and q_i denote the columns of A and Q, respectively. Using Gram-Schmidt, we obtain an orthogonal basis q_i for the column space of A.

$$p_1 = a_1, q_1 = \frac{p_1}{\|p_1\|_2} \tag{17}$$

$$p_2 = a_2 - (a_2^\top q_1)q_1, \quad q_2 = \frac{p_2}{\|p_2\|_2} \tag{18}$$

$$p_3 = a_3 - (a_3^\top q_1)q_1 - (a_3^\top q_2)q_2, q_3 = \frac{p_3}{\|p_3\|_2} \tag{19}$$

$$\vdots \tag{20}$$

Rearranging terms, we have

$$a_1 = r_{11}q_1 \tag{21a}$$

$$a_i = r_{i1}q_1 + \cdots + r_{ii}q_i, \quad i = 2, ..., n, \tag{21b}$$

where each q_i has unit norm, and $r_{ij}q_j$ denotes the projection of a_i onto the vector q_j for $j \neq i$.

Stacking a_i horizontally into A and rewriting (21a-b) in matrix notation, we obtain $A = QR$. For the given matrix, we have

$$A = \begin{bmatrix} 0.6 & 0 & 0.8 \\ 0.8 & 0 & -0.6 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 5 & -5 & -5 \\ 0 & 3 & 3 \\ 0 & 0 & 5 \end{bmatrix}.$$

Note that an equivalent factorization is $A = (-Q)(-R)$.

(b) Given an invertible matrix $A \in \mathbb{R}^{n\times n}$ and an observation vector $b \in \mathbb{R}^n$, the solution to the equality

$$Ax = b$$

is given as $x = A^{-1}b$. For the matrix $A = QR$ from part (a), assume that we want to solve

$$Ax = \begin{bmatrix} 8 \\ -6 \\ 3 \end{bmatrix}.$$

By using the fact that Q is an orthonormal matrix, find v such that

$$Rx = v.$$

Then, given the upper-triangular matrix R in part (a) and v, find the elements of x sequentially.

Solution: We note that $Q^{-1} = Q^T$.

$$\begin{aligned} Ax &= b \\ QRx &= b \\ Q^\top QRx = Rx &= Q^\top b. \end{aligned}$$

Thus

$$v = Q^\top b = \begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix}.$$

Given R and v, we can find x by back-substitution:

$$\begin{bmatrix} 5 & -5 & -5 \\ 0 & 3 & 3 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix} \implies x_3 = 2 \implies x_2 = -1 \implies x_1 = 1 \implies x = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}.$$

(c) Given an invertible matrix $B \in \mathbb{R}^{n\times n}$ and an observation vector $c \in \mathbb{R}^n$, find the computational cost of finding the solution z to the equation $Bz = c$ by using the QR decomposition of B. Assume that Q and R matrices are available, and adding, multiplying, and dividing scalars take one unit of "computation".

As an example, computing the inner product $a^\top b$ is said to be $\mathcal{O}(n)$, since we have n scalar multiplication for each $a_i b_i$. Similarly, matrix vector multiplication is $\mathcal{O}(n^2)$, since matrix vector multiplication can be viewed as computing n inner products. The computational cost for inverting a matrix in $\mathbb{R}^n$ is $\mathcal{O}(n^3)$, and consequently, the cost grows rapidly as the set of equations grows in size. This is why the expression $A^{-1}b$ is usually not computed by directly inverting the matrix A. Instead, the QR decomposition of A is exploited to decrease the computational cost.

Solution: We count the number of operations in back substitution. Solving the initial equation

$$r_{nn}x_n = \bar{b}_n$$

takes 1 multiplication. Solving each subsequent equation takes one more multiplication and one more addition than the previous. In total, we have $1 + 3 + 5 + \cdots$ of operations, which is on the order of $\mathcal{O}(n^2)$. Thus, matrix multiplication and back substitution are both $O(n^2)$. Given the QR decomposition of A, we can solve $Ax = b$ in $O(n^2)$ times.

4. 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 $\vec{p} + \vec{q}$ and $\vec{p} - \vec{q}$ are eigenvectors of A, and determine the corresponding eigenvalues.

Solution: We have

$$A\vec{p} = c\vec{p} + \vec{q}, \quad A\vec{q} = \vec{p} + c\vec{q}, \tag{22}$$

from which we obtain

$$A(\vec{p} - \vec{q}) = (c-1)(\vec{p} - \vec{q}), \quad A(\vec{p} + \vec{q}) = (c+1)(\vec{p} + \vec{q}). \tag{23}$$

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

(b) Determine the nullspace and rank of A.

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

$$0 = A\vec{x} = \vec{p}(\vec{q}^\top \vec{x}) + \vec{q}(\vec{p}^\top \vec{x}). \tag{24}$$

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

$$0 = A\vec{x} = (\vec{q}^\top \vec{x})\vec{p} + (\vec{p}^\top \vec{x})\vec{q} = 0. \tag{25}$$

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 the set of vectors orthogonal to $\vec{p}$ and $\vec{q}$, i.e., $\mathcal{N}(A) = \mathrm{span}(\vec{p}, \vec{q})^\perp$. We have from the fundamental theorem of linear algebra and the fact that A is symmetric,

$$\mathcal{R}(A) = \mathcal{R}\left(A^\top\right) = \mathcal{N}(A)^\perp = (\mathrm{span}(\vec{p}, \vec{q})^\perp)^\perp = \mathrm{span}(\vec{p}, \vec{q}). \tag{26}$$

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

(c) 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

$$A = (c-1)\vec{v}_-\vec{v}_-^\top + (c+1)\vec{v}_+\vec{v}_+^\top, \tag{27}$$

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

$$\|\vec{p} \pm \vec{q}\|_2^2 = \vec{p}^\top \vec{p} \pm 2\vec{p}^\top \vec{q} + \vec{q}^\top \vec{q} = 2(1 \pm c), \tag{28}$$

we have

$$\vec{v}_\pm = \frac{1}{\sqrt{2(1 \pm c)}}(\vec{p} \pm \vec{q}), \tag{29}$$

so that the eigenvalue decomposition becomes

$$A = \frac{1}{2}\left((\vec{p} + \vec{q})(\vec{p} + \vec{q})^\top - (\vec{p} - \vec{q})(\vec{p} - \vec{q})^\top\right). \tag{30}$$

(d) **(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:

$$\vec{p}_{\text{new}} = \|\vec{p}_{\text{new}}\|_2\, \vec{p},\ \vec{q}_{\text{new}} = \|\vec{q}_{\text{new}}\|_2\, \vec{q}. \tag{31}$$

Then the matrix A_{new} can be written as

$$A_{\text{new}} = \vec{p}_{\text{new}}\vec{q}_{\text{new}}^{\top} + \vec{q}_{\text{new}}\vec{p}_{\text{new}}^{\top} = \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 \left(\vec{p}\vec{q}^{\top} + \vec{q}\vec{p}^{\top}\right) = \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 A. \tag{32}$$

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,

$$\lambda_{\pm} = \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 (c \pm 1) = \vec{p}_{\text{new}}^{\top}\vec{q}_{\text{new}} \pm \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 . \tag{33}$$

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

$$A_{\text{new}} = \|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2 A = \frac{\|\vec{p}_{\text{new}}\|_2 \|\vec{q}_{\text{new}}\|_2}{2} \left((\vec{p}+\vec{q})(\vec{p}+\vec{q})^{\top} - (\vec{p}-\vec{q})(\vec{p}-\vec{q})^{\top}\right). \tag{34}$$

5. 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:

$$\vec{x}^\top A \vec{x} = \vec{x}^\top U \Sigma U^\top \vec{x} = \vec{v}^\top \Sigma \vec{v} \geq 0, \tag{35}$$

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:

$$\vec{x}^\top A \vec{x} = \vec{x} \lambda \vec{x} = \lambda \vec{x}^\top \vec{x} = \lambda \left\| \vec{x} \right\|_2^2 \tag{36}$$

Since norms/anything squared is always non-negative, in order for $\lambda \left\| \vec{x} \right\|_2^2 \geq 0$, 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:

$$\vec{v}^\top \Sigma \vec{v} = \sum_i \lambda_i v_i^2 \geq 0, \tag{37}$$

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:

$$P^\top P = (A^{\frac{1}{2}})^\top A^{\frac{1}{2}} = (U \Sigma^{\frac{1}{2}} U^\top)^\top U \Sigma^{\frac{1}{2}} U^\top = U \Sigma^{\frac{1}{2}} U^\top U \Sigma^{\frac{1}{2}} U^\top = U \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}} U^\top = U \Sigma U^\top = A. \tag{38}$$

(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:

$$\vec{x}^\top A \vec{x} = \vec{x}^\top P^\top P \vec{x} = \langle P\vec{x},\ P\vec{x} \rangle = \left\| P\vec{x} \right\|_2^2 \geq 0. \tag{39}$$

(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, $\vec{e}_i^\top A \vec{e}_i = A_{ii} \geq 0$.

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.

