STAT 30900: Homework 1

Caleb Derrickson

October 20, 2023

Collaborators: The TA's of the class, as well as Kevin Hefner, Alexander Cram, and Steven Lee.

Contents

Problem 1

Problem 1, part a

Let $A \in \mathbb{C}^{m \times n}$ and $1 \leq p \leq \infty$. Find the closed form expressions for

$$\|A\|_{1,p} := \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_1} \quad \text{and} \quad \|A\|_{p,\infty} := \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_p}$$

Your expressions should agree with the matrix 1-norm and matrix ∞-norm when $p = 1$ and ∞ respectively.

Solution:

Let A_k denote the k-th column of A. Then,

$$
\begin{aligned}
\|A\mathbf{x}\|_p &= \left\| \sum_{k=1}^n x_k A_k \right\|_p & \text{(Definition of vector norm.)} \\[2mm]
&\leq \sum_{k=1}^n \|x_k A_k\|_p & \text{(Triangle Inequality.)} \\[2mm]
&= \sum_{k=1}^n |x_k| \|A_k\|_p & \text{(Absolute Homogeneity of norm.)} \\[2mm]
&\leq \max\{\|A_k\|_p : 1 \leq k \leq n\} \left(\sum_{k=1}^n |x_k| \right) & \text{(All columns of A are } \leq \text{ max column.)} \\[2mm]
&= \max\{\|A_k\|_p : 1 \leq k \leq n\} \|\mathbf{x}\|_1 & \text{(Definition of 1-norm.)} \\[2mm]
\implies \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_1} &\leq \max\{\|A_k\|_p : 1 \leq k \leq n\} & \text{(Rearranging.)} \quad (1)
\end{aligned}
$$

From 1, this means the argument inside the max is less than or equal to the max column of A. The question now is when this is equal to $\|A\|_{1,p}$. We are free to choose any nonzero $\mathbf{x}$, so a natural one to choose would be the one to pick the max column when multiplied to A. Meaning if $\mathbf{x} = e_M$, where M is the index corresponding to the max column of A, then equality it obtained. Note this is given as the definition of $\|A\|_1$ when $p = 1$.

We next wish to find a closed form expression for $\|A\|_{p,\infty}$. It follows the same processes done in the previous section, with some slight modifications.

$$\|A\mathbf{x}\|_{\infty} = \max\{|(A\mathbf{x})_j| : 1 \leq j \leq n\} \qquad \text{(Given.)}$$

$$= \max\left\{\left|\sum_{k=1}^{n} x_k A_{ik}\right| : 1 \leq i \leq m\right\} \qquad \text{(Matrix-vector multiplication.)}$$

$$\leq \max\left\{\sum_{k=1}^{n} |x_k A_{ik}| : 1 \leq i \leq m\right\} \qquad \text{(Triangle inequality.)}$$

$$= \max\left\{\sum_{k=1}^{n} |x_k||A_{ik}| : 1 \leq i \leq m\right\} \qquad \text{(Absolute Homogeneity.)}$$

$$= \sum_{k=1}^{n} |x_k||A_{Mk}| \qquad \text{(\textit{M} is max row of \textit{A}.)}$$

$$\leq |x_M| \sum_{k=1}^{n} |A_{Mk}| \qquad \text{(\textit{M} is max element of \textbf{x}.)}$$

$$= \|\mathbf{x}\|_{\infty} \sum_{k=1}^{n} |A_{Mk}| \qquad \text{(Definition of } \infty\text{-norm.)}$$

$$\leq \|\mathbf{x}\|_p \sum_{k=1}^{n} |A_{Mk}| \qquad \text{(}\infty\text{-norm is included in all } p\text{-norms.)}$$

$$\implies \|A\mathbf{x}\|_{\infty} \leq \|\mathbf{x}\|_p \|A\|_{\infty} \qquad \text{(Definition of } \infty\text{-norm for matrices.)}$$

Thus, we can see $\|A\|_{p,\infty}$ is the max of $\|A\|_{\infty}$, which is just $\|A\|_{\infty}$. We note that $\|A\|_{\infty}$ is the max row sum of the matrix A. Thus, both expressions have been found.

Problem 1, part b

Let $\mathbb{S}^n := \{A \in \mathbb{R}^{n \times n} : A^T = A\}$. Recall the Gram matrix from Homework **0**, Problem **4**:
For $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{R}^n$, we write

$$G(\mathbf{x}_1, ..., \mathbf{x}_n) := \begin{bmatrix} \mathbf{x}_1^T \mathbf{x}_1 & \mathbf{x}_1^T \mathbf{x}_2 & ... & \mathbf{x}_1^T \mathbf{x}_n \\ \mathbf{x}_2^T \mathbf{x}_1 & \mathbf{x}_2^T \mathbf{x}_2 & ... & \mathbf{x}_2^T \mathbf{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n^T \mathbf{x}_1 & \mathbf{x}_n^T \mathbf{x}_2 & ... & \mathbf{x}_n^T \mathbf{x}_n \end{bmatrix} \in \mathbb{S}^n.$$

Consider the set $\mathbb{G}^n := \{G(\mathbf{x}_1, ..., \mathbf{x}_n) \in \mathbb{S}^n : \|\mathbf{x}_1\|_2 \leq 1, ..., \|\mathbf{x}_n\|_2 \leq 1\}$. Prove that for any $A \in \mathbb{S}^n$,

$$\|A\|_G := \max\{|\text{tr}(AG)| : G \in \mathbb{G}^n\}$$

defines a norm on S^n. Show that if $A = \text{diag}(a_{11}, ..., a_{nn}) \in \mathbb{S}^n$, then

$$\|A\|_G = \max\left(\sum_{i=1}^{n} \max(a_{ii}, 0), - \sum_{i=1}^{n} \min(a_{ii}, 0) \right).$$

Solution:

As A quick reminder, a norm $\|\cdot\|$ should satisfy the following properties for scalar $\alpha \in \mathbb{R}$ and matrices $A, B \in \mathbb{R}^{n \times n}$:

---

(1) $\|A\| \geq 0$.

(2) $\|A\| = 0 \iff A = 0$.

(3) $\|\alpha A\| = |\alpha| \|A\|$.

(4) $\|A + B\| \leq \|A\| + \|B\|$.

Note that I'm taking $A, B \in \mathbb{R}^{n \times n}$ since the set $\mathbb{S}^n \subset \mathbb{R}^{n \times n}$. We will prove each property below.

(1) $\|A\| \geq 0$.

By the definition of $\|A\|_G$, we are taking the maximum over a set of strictly positive elements - the absolute value guarantees this. Thus this property is immediately obeyed.

(2) $\|A\| = 0 \iff A = 0$.

The backward implication is straightforward, as if $A = 0$, then $AG = 0$, thus the trace is taken over the zero matrix, meaning we are taking the maximum over the set containing only zeros, thus $\|A\|_G = 0$.

5

The forward implication requires jumping through some extra hoops. We first note if the max of a set is zero, and every element is positive, then all elements are equal to zero. Thus, $|\text{tr}(AG)| = 0$ for all G. Then the trace of AG, without the absolute value, is zero for all G. If we let $\mathbf{x}_1 = [1, 0, ..., 0]$, and $\mathbf{x}_2 = ... = \mathbf{x}_n = 0$ (this can be done, since all vectors have norms ≤ 1), then the only nonzero entry in G is $g_{11} = 1$. This, when multiplied with A, will give us only the first column of A, with the rest zeros. In this case, $tr(AG) = a_{11} = 0$. This means $a_{11} = 0$ for all $G \in \mathbb{G}^n$, and repeating this for any choice of $\mathbf{x}_i$ will result in $a_{ii} = 0$. Thus the diagonal elements are shown to be zero.

Next, take any two vectors $\mathbf{x}_i = \mathbf{x}_j = [1, 0, ..., 0]$. Note if $i = j$, then it reduces to the previous case, so without loss of generality, choose $i \neq j$. This implies the only nonzero entries in G are g_{ji} and g_{ij}, which are both equal 1. Thus, $G = E_{ij} + E_{ji}$, where the E matrices are from the first homework, problem 5. Note that the E_{ij}, when multiplied by A on the left (AE_{ij}) will retrieve the i-th column of A and place it in the j-th column, with zero everywhere else. Thus, the following holds:

$$\text{tr}(AG) = \text{tr}(A(E_{ij} + E_{ji})) \qquad \text{(Explained above.)}$$

$$= \text{tr}(AE_{ij}) + \text{tr}(AE_{ji}) \qquad \text{(Trace operator is linear.)}$$

$$= a_{ij} + a_{ji} \qquad \text{(diagonal of } i\text{-th row is } a_{ii}, \text{ when moved, diagonal is } a_{ji}.)$$

$$\implies 2a_{ij} = 2a_{ji} = 0 \qquad \text{(A is symmetric.)}$$

$$\implies a_{ij} = a_{ji} = 0 \qquad \text{(Simplifying.)}$$

Thus, the off diagonal entries of A are zero for all $G \in \mathbb{G}^n$. This, when paired with the diagonal terms being zero, means $A = 0$.

(3) $\|\alpha A\| = |\alpha| \|A\|$.

We will go straight into calculations.

$$\|\alpha A\|_G = \max\{|\mathrm{tr}(\alpha AG)| : G \in \mathbb{G}^n\} \qquad \text{(Given.)}$$

$$= \max\{|\alpha \mathrm{tr}(AG)| : G \in \mathbb{G}^n\} \qquad \text{(Trace operator is linear.)}$$

$$= \max\{|\alpha||\mathrm{tr}(AG)| : G \in \mathbb{G}^n\} \qquad \text{(Property of absolute value.)}$$

$$= |\alpha| \max\{|\mathrm{tr}(AG)| : G \in \mathbb{G}^n\} \qquad (|\alpha| \text{ is independent of the max.)}$$

$$= |\alpha| \|A\|_G \qquad \text{(Definition.)}$$

---

(4) $\|A + B\| \le \|A\| + \|B\|$.

We will go straight into calculations.

$$\|A + B\|_G = \max\{|\mathrm{tr}((A + B)G)| : G \in \mathbb{G}^n\} \qquad \text{(Given.)}$$

$$= \max\{|\mathrm{tr}(AG + BG)|\} \qquad \text{(Distribution of matrices.)}$$

$$= \max\{|\mathrm{tr}(AG) + \mathrm{tr}(BG)|\} \qquad \text{(Trace operator is linear.)}$$

$$\le \max\{|\mathrm{tr}(AG)| + |\mathrm{tr}(BG)|\} \qquad \text{(Triangle inequality.)}$$

$$= \max\{|\mathrm{tr}(AG)|\} + \max\{|\mathrm{tr}(BG)|\} \qquad \text{(Arguments of max are real and positive.)}$$

$$\implies \|A + B\| \le \|A\| + \|B\| \qquad \text{(Definition.)}$$

For the second part of part b, we need to show if A is diagonal, then,

$$\|A\|_G = \max \left(\sum_{i=1}^{n} \max(a_{ii}, 0), - \sum_{i=1}^{n} \min(a_{ii}, 0) \right).$$

We can first check the form of the absolute trace product, $|\text{tr}(AG)|$.

$$|\text{tr}(AG)| = \left| \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} g_{ki} \right| \qquad \text{(Given.)}$$

$$= \left| \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} \mathbf{x}_i^t \mathbf{x}_j \right| \qquad \text{(Definition of Gram Matrix.)}$$

$$= \left| \sum_{i=1}^{n} a_{ii} \mathbf{x}_i^t \mathbf{x}_i \right| \qquad \text{(Only term that survives is } k = i.\text{)}$$

$$\leq \sum_{i=1}^{n} |a_{ii}| |\mathbf{x}_i^t \mathbf{x}_i| \qquad \text{(Triangle Inequality.)}$$

$$\leq \sum_{i=1}^{n} |a_{ii}| \|\mathbf{x}_i\|_2 \qquad \text{Cauchy-Schwartz.)}$$

$$\leq \sum_{i=1}^{n} |a_{ii}| \qquad (\|\mathbf{x}_i\|_2 \leq 1.) \qquad (2)$$

Thus, $\|A\|_G$ will have the form 2. By the definition of the absolute value function,

$$|a_{ii}| = \begin{cases} a_{ii}, & \text{if } a_{ii} \geq 0 \\ -a_{ii}, & \text{if } a_{ii} < 0 \end{cases}$$

Therefore, when taking the max of a_{ii} and zero, it is equivalent to the absolute value function. We can then write $a_{ii} \leq \max(a_{ii}, 0)$ for all a_{ii}. When plugging this into 2, we can immediately see the following:

$$\|A\|_G = \sum_{i=1}^{n} |a_{ii}|$$

$$\|A\|_G \geq \sum_{i=1}^{n} |\max(a_{ii}, 0)|$$

$$\|A\|_G \geq \sum_{i=1}^{n} \max(a_{ii}, 0)$$

Also, noting that $\min(a_{ii}, 0) \leq 0$, we can write

$$\|A\|_G = \sum_{i=1}^{n} |a_{ii}|$$

$$\|A\|_G \geq \sum_{i=1}^{n} |\min(a_{ii}, 0)|$$

$$\|A\|_G \geq -\sum_{i=1}^{n} \min(a_{ii}, 0)$$

Therefore, $\|A\|_G$ has been shown to be greater than or equal to both sums, so $\|A\|_G$ is greater than or equal to the max of them, which is what we wanted to show.

Problem 2

Let $A \in \mathbb{C}^{n \times n}$ and $\|\cdot\|_p : \mathbb{C}^{n \times n} \to [0, \infty)$ be the matrix p-norm for some $p \in [1, \infty]$.

Problem 2, part a

Show that if $\|A\|_p < 1$, then $\mathbb{I} - A$ is invertible and furthermore,

$$\frac{1}{1 + \|A\|_p} \leq \|(\mathbb{I} - A)^{-1}\|_p \leq \frac{1}{1 - \|A\|_p}. \tag{3}$$

Solution:

Let $\mathbf{v} \in \mathbb{C}^{n \times n} \setminus \{0\}$ be an eigenvector of A with eigenvalue λ. Then $A\mathbf{v} = \lambda \mathbf{v} \iff \|A\mathbf{v}\|_p = \|\lambda \mathbf{v}\|_p$. Note the matrix p-norm is taken as a max over all $\mathbf{x} \neq 0$. Along with $\|A\| < 1$, we get,

$$1 > \max_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A\mathbf{v}\|}{\|\mathbf{v}\|}$$

then $\|A\mathbf{v}\| < \|\mathbf{v}\|$. So 1 is then not an eigenvector of A. This implies that $(A - \mathbb{I})\mathbf{v} = (\lambda - 1)\mathbf{v}$. Note that $\mathbf{v}$ was taken as a nonzero vector, so $(1 - \lambda)\mathbf{v} \neq 0$. This means that 0 is not an eigenvalue of $A - \mathbb{I}$, making the determinant nonzero, thus $A - \mathbb{I}$ is invertible.

We then wish to show 3,

$$\frac{1}{1 + \|A\|_p} \leq \|(\mathbb{I} - A)^{-1}\|_p \leq \frac{1}{1 - \|A\|_p}.$$

To prove the first inequality, we note that if $\mathbf{v}$ is an eigenvector of A with eigenvalue λ, (and A invertible) then $\mathbf{v}$ is an eigenvector of A with eigenvalue λ^{-1}. This is shown by the following:

$$A\mathbf{v} = \lambda \mathbf{v} \iff A^{-1}A\mathbf{v} = A^{-1}\lambda\mathbf{v} \iff A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}.$$

Thus, take $\mathbf{v}$ as an eigenvector of A with eigenvalue $(1 - \lambda)$. The above claim implies that $\mathbf{v}$ is also an eigenvector of $(\mathbb{I} - A)^{-1}$ with eigenvalue $(1 - \lambda)^{-1}$. Furthermore, since $\|\cdot\|_p$ is consistent, we have that the spectral radius $\rho(A) = \max\{|\lambda(A)|\}$ is less than or equal to $\|A\|$, ie $\rho(A) \leq \|A\|$. The following steps can then be justified

$$(\mathbb{I} - A)^{-1}\mathbf{v} = \frac{1}{1 - |\lambda|}\mathbf{v} \qquad \text{(Given.)}$$

$$\left\|(\mathbb{I} - A)^{-1}\mathbf{v}\right\| = \frac{1}{1 - |\lambda|}\|\mathbf{v}\| \qquad \text{(Taking norm and eigenvalue is positive.)}$$

$$\left\|(\mathbb{I} - A)^{-1}\mathbf{v}\right\| \leq \left\|(\mathbb{I} - A)^{-1}\right\|\|\mathbf{v}\| \qquad \text{(Norm Consistency of } \|\cdot\|_p.) \qquad (4)$$

$$\implies \left\|(\mathbb{I} - A)^{-1}\right\| > \frac{1}{1 - |\lambda|} \qquad \text{(Applying 4 and cancelling } \|v\|)$$

$$\left\|(\mathbb{I} - A)^{-1}\right\| > \frac{1}{1 + |\lambda|} \qquad (1 - |\lambda| < 1 + |\lambda|.)$$

$$\left\|(\mathbb{I} - A)^{-1}\right\| \geq \frac{1}{1 + \|A\|_p} \qquad (\rho(A) \leq \|A\|.)$$

This solves the first inequality of 3. We then can show the second inequality via the Taylor expansion of $\left\|(\mathbb{I} - A)^{-1}\right\|_p$. Noting further that $\|A\|_p < 1$, the following steps can be taken:

$$\left\|(\mathbb{I} - A)^{-1}\right\|_p \qquad \text{(Given.)}$$

$$= \left\|\sum_{k=0}^{\infty} A^k\right\|_p \qquad \text{(Taylor Expansion.)}$$

$$\leq \sum_{k=0}^{\infty} \left\|A^k\right\|_p \qquad \text{(Triangle inequality.)}$$

$$\leq \sum_{k=0}^{\infty} \|A\|_p^k \qquad \text{(Submultiplication of } \|\cdot\|_p.)$$

$$\implies \left\|(\mathbb{I} - A)^{-1}\right\|_p \leq \frac{1}{1 - \|A\|_p} \qquad \text{(Taylor Expansion Equivalent.)}$$

This proves both inequalities, thus 3 has been shown. □

Problem 2, part b

Suppose A is invertible. show that and $X \in \mathbb{C}^{n \times n}$ with

$$\|X - A\|_p < \frac{1}{\|A^{-1}\|_p}$$

must also be invertible.

Solution:

We first should note that if a matrix product AB is invertible, then both A and B are invertible. This can be reasoned by a rank argument: if AB is invertible, then AB has full rank. If B mapped anything other than $\mathbf{0}$ to $\mathbf{0}$, then AB would map a nonzero vector to zero, which is a contradiction. Thus, B has full rank, meaning that any vector in its target space can be represented by an element in the domain of B scaled by matrix B (i.e. $B\mathbf{x} = \mathbf{y}$). If we supposed that A mapped $\mathbf{y}$ to the zero vector for some arbitrary $\mathbf{y}$, then AB would map $\mathbf{y}$ to zero, which can only be the zero vector. Thus $\mathbf{y} = \mathbf{0}$, meaning A and B have full rank, thus both are invertible. This will be used in the following steps:

$$\|X - A\|_p < \frac{1}{\|A^{-1}\|_p} \qquad \text{(Given.)}$$

$$\|X - A\|_p \|A^{-1}\|_p < 1 \qquad \text{(Rearranging.)}$$

$$\left\|(X - A)A^{-1}\right\|_p \leq \|X - A\|_p \|A^{-1}\|_p \qquad \text{(Norm submultiplicative.)} \qquad (5)$$

$$\implies \left\|XA^{-1} - AA^{-1}\right\|_p < 1 \qquad \text{(Applying }\textcolor{red}{5}\text{.)}$$

$$\left\|XA^{-1} - \mathbb{I}\right\|_p < 1 \qquad \text{(Simplifying.)}$$

$$\implies \mathbb{I} - XA^{-1} + \mathbb{I} \text{ is invertible.} \qquad \text{(By part a.)}$$

$$\implies XA^{-1} \text{ is invertible.} \qquad \text{(Simplifying.)}$$

$$\implies X \text{ is invertible.} \qquad \text{(By proof above.)}$$

The steps above thus prove X is invertible. $\square$

Problem 2, part c

Let $\|\cdot\| : \mathbb{C}^{n \times n} \to [0, \infty)$ be an arbitrary norm that may not be submultiplicative. Suppose $\|A\| < 1$, can we conclude that $\mathbb{I} - A$ is invertible?

Solution:

As a counterexample, take A to be

$$A = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

under the Hölder ∞-norm. Thus $\|A\| < 1$, but

$$\mathbb{I} - A = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

which is clearly not invertible, since $\mathbb{I} - A$ does not have full rank.

Problem 3

Recall that in the lectures, we mentioned that

   (i) there are matrix norms that are not submutiplicative and an example is the Hölder ∞-norm;

  (ii) we may always construct a norm that approximates the spectral radius of a given matrix A as closely as we want.

Problem 3, part a

Let $\|\cdot\| : \mathbb{C}^{m \times n} \to \mathbb{R}$ be a norm, defined for all $m, n \in \mathbb{N}$. show that there always exist a $c > 0$ such that the constant multiple $\|\cdot\|_c := c\|\cdot\|$ defines a submultiplicative norm, i.e.,

$$\|AB\|_c \leq \|A\|_c \|B\|_c$$

for any $A \in \mathbb{C}^{m \times n}$ and $B \in \mathbb{C}^{n \times p}$ (even if $\|\cdot\|$ does not have this property). Note that the constant c will depend on $m, n, p \in \mathbb{N}$ in general. Find the constant c for the Hölder ∞-norm.

Solution:

    By the Equivalence of Norms over a Finite Dimensional Vector Space, (in this case, $\mathbb{C}^{m \times n}$) there will exist some $0 < a_1 < a_2$ such that

$$a_1 \|C\|_* \leq \|C\| \leq a_2 \|C\|_* \tag{6}$$

    for any norm $\|\cdot\|_*$ over our space and matrix C. Consider the Frobenius norm, $\|\cdot\|_F$. We know that the Frobenius norm obeys submultiplicivity, so the following steps are justified:

$$
\begin{aligned}
\|AB\| &\leq a_2 \|AB\|_F && \text{(Equivalence.)} \\[2mm]
&\leq a_2 \|A\|_F \|B\|_F && \text{(Submultiplicivity of F.)} \\[2mm]
&\leq \frac{a_2}{a_1}\|A\|\frac{1}{a_1}\|B\| && \text{(Equivalence.)} \\[2mm]
&= \frac{a_2}{a_1^2}\|A\|\|B\| && \text{(Rewriting.)}
\end{aligned}
$$

Therefore, let $c = \frac{a_2}{a_1^2} > 0$. Thus $\|AB\|_c \leq \|A\|_c\|B\|_c$.

We next wish to find the constant c for the Hölder ∞-norm. We will be looking at its equivalence to the Frobenius norm.

$$\|AB\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{p}\left|\sum_{k=1}^{n}A_{ik}B_{kj}\right|^2 \qquad \text{(Finding left inequality.)}$$

$$\leq \sum_{i=1}^{m} p \max_j \left\{\left|\sum_{k=1}^{n}A_{ik}B_{kj}\right|^2\right\} \qquad \text{(Taking max over B's columns.)}$$

$$\leq mp \max_{ij}\left\{\left|\sum_{k=1}^{n}A_{ik}B_{kj}\right|^2\right\} \qquad \text{(Taking max over A's rows.)}$$

$$\implies \|AB\|_F \leq \sqrt{mp}\max_{ij}\left\{\left|\sum_{k=1}^{n}A_{ik}B_{kj}\right|\right\} \qquad \text{(Square root function is monotonic.)}$$

$$\iff \|AB\|_F \leq \sqrt{mp}\|AB\|_{H,\infty} \qquad (7)$$

$$\|AB\|_{H,\infty}^2 = \max_{ij}\left\{\left|\sum_{k=1}^{n}A_{ik}B_{kj}\right|\right\} \qquad \text{(Finding right inequality.)}$$

$$\leq \max_i \left\{\sum_{j=1}^{p}\left|\sum_{k=1}^{n}A_{ik}B_{kj}\right|\right\} \qquad \text{(Max over j is one of the sums.)}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{p}\left|\sum_{k=1}^{n}A_{ik}B_{kj}\right| \qquad \text{(Same argument.)}$$

$$\implies \|AB\|_{H,\infty}^2 \leq \|AB\|_2^2 \qquad \text{(Definition of Frobenius norm.)}$$

$$\iff \|AB\|_{H,\infty} \leq \|AB\|_2 \qquad (8)$$

In our calculations above, we can find values for a_1, a_2. Taking the factors in front of each norm, we get $a_1 = (\sqrt{mn})^{-1}$ and $a_2 = 1$. Note that these are not unique, since another valid value for a_2 is 2. It is entirely possible there could be smaller bounds which we could argue. These are just values I've found to work. Thus c will take the value $(mn)^{-1}$.

Problem 3, part b

Let $J \in \mathbb{C}^{n \times n}$ be in Jordan form. Let $\varepsilon > 0$. Verify that $D_\varepsilon^{-1} J D_\varepsilon$ replaces the ones along the superdiagonal by $1, \varepsilon, \varepsilon^2, ..., \varepsilon^{n-1}$.

Note for any Jordan matrix J can be rewritten as a diagonal matrix A plus a nilpotent matrix N of some degree. I.e., $J = A + N$, where N has value 1 only along the superdiagonal (hence N had nilpotency degree n). Therefore, we can rewrite $D_\varepsilon^{-1} J D_\varepsilon$ as the following:

$$D_\varepsilon^{-1} J D_\varepsilon = D_\varepsilon^{-1}(A + N)D_\varepsilon = D_\varepsilon^{-1} A D_\varepsilon + D_\varepsilon^{-1} N D_\varepsilon$$

Note that $D_\varepsilon^{-1} = \operatorname{diag}(1, \varepsilon^{-1}, \varepsilon^{-2}, ...\varepsilon^{1-n})$. Since A is diagonal, each column will be multiplied then divided by the same degree of ε, i.e. column 1 of A will be multiplied by ε^0 then divided by ε^0, column 2 of A will be multiplied by ε^1 then divided by ε^1, and so on, doing nothing in the process. So $D_\varepsilon^{-1} A D_\varepsilon = A$. We just need to show that the nilpotent matrix N will be multiplied by ε.

Denote γ_i as the i-th row of D_ε, and η_j as the j-th column of N. Note $N, D_\varepsilon \in \mathbb{C}^{n \times n}$. So $i, j \leq n$; $\eta_j = e_j$ for $j \neq 1$ and $\eta_1 = 0$; and $\gamma_i = \varepsilon^{i-1} e_{i+1}$, where e_i is the i-th canonical basis vector. The following steps are then justified.

$$N D_\varepsilon = \sum_{i=1}^{n} \eta_i \gamma_i^t \qquad \text{(Matrix product interpretation.)}$$

$$= \sum_{i=1}^{n-1} e_i \varepsilon^{i-1} e_{i+1}^t \qquad \text{(Plugging in).}$$

$$= \operatorname{superdiag}(\varepsilon^2, \varepsilon^3, ..., \varepsilon^{n-1}) \qquad \text{(superdiag is shorthand.)}$$

I couldn't express the outer product $e_i e_{i+1}^t$ elegantly, so please mind my shorthand notation of superdiag(-). The important thing to note here is that by our construction of η_i, the column of D_ε containing the linear ε is terminated, leaving us with the leading term being quadratic, ε^2. We will then multiply on the left by D_ε^{-1} in a similar manner.

$$D_\varepsilon^{-1}ND_\varepsilon = \sum_{i=1}^{n-1} \varepsilon^{1-i}(ND_\varepsilon)_i^t \qquad \text{(Matrix product interpretation.)}$$

$$= \sum_{i=1}^{n-1} \varepsilon^{1-i} e_{i+1}\varepsilon^i e_i^t \qquad \text{(Plugging in.)}$$

$$= \sum_{i=1}^{n-1} \varepsilon^1 (e_i^t)^* (e_{i+1})^* \qquad \text{(Antisymmetry of inner product.)}$$

$$= \sum_{i=1}^{n-1} \varepsilon e_i e_{i+1}^t \qquad \text{(Basis vectors are real.)}$$

$$D_\varepsilon^{-1}ND_\varepsilon = \text{superdiag}(\varepsilon, \varepsilon, ..., \varepsilon) \qquad \text{(Shorthand.)}$$

Therefore we have proven the claim.

Problem 3, part c

Show that

$$\left\| D_\varepsilon^{-1}JD_\varepsilon \right\|_\infty \leq \rho(J) + \varepsilon$$

Solution:

It was shown in part b that the left and right multiplications of D_ε and its inverse did nothing to the diagonal of J. Since J is an upper triangular matrix, its eigenvalues are located along its diagonal. Thus $\rho(J) = |\lambda_M|$ for some $1 \leq M \leq k$. Note it could be the case that all eigenvalues are negative, thus when taking the ∞-norm, a different eigenvalue could be taken. Suppose m is chosen such that $\left\| D_\varepsilon^{-1}JD_\varepsilon \right\|_\infty = |\lambda_m + \varepsilon|$. By the triangle inequality, $|\lambda_m + \varepsilon| \leq |\lambda_m| + \varepsilon$, since ε is positive. But $|\lambda_m| \leq \max_{i \leq k}\{|\lambda_i|\} = \rho(J)$, thus $\left\| D_\varepsilon^{-1}JD_\varepsilon \right\|_\infty \leq \rho(J) + \varepsilon$.

Problem 3, part d

Hence, or otherwise, show that for any given $A \in \mathbb{C}^{n \times n}$ and $\varepsilon > 0$, there exists a norm $\|\cdot\|_*$ on $\mathbb{C}^n$ such that

$$\|A\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_*}{\|\mathbf{x}\|_*}$$

has the property that

$$\rho(A) \leq \|A\| \leq \rho(A) + \varepsilon$$

Solution:

For the sake of summary, I will explain my thought process on this problem. In essence, we need to find the norm $\|\cdot\|_*$ which, depending on the matrix A, will have the property given. If we define $\|\mathbf{v}\|_* = \left\|D_\varepsilon^{-1} X^{-1} \mathbf{v}\right\|_\infty$, then by the process below, we will see that by the previous part, the property will hold. Note that I will be looking at an equivalent form of the norm, that is, $\|A\| = \max\{\|A\mathbf{x}\|_* : \|\mathbf{x}\|_* = 1\}$.

$$\|A\mathbf{x}\|_* = \left\|D_\varepsilon^{-1} X^{-1} A\mathbf{x}\right\|_\infty \qquad \text{(Given.)}$$

$$= \left\|D_\varepsilon^{-1} X^{-1} A X D_\varepsilon \mathbf{y}\right\|_\infty \qquad \text{(By change of variable } y = D_\varepsilon^{-1} X^{-1} \mathbf{x}.\text{)}$$

$$= \left\|D_\varepsilon^{-1} X^{-1} (X J X^{-1}) X D_\varepsilon \mathbf{y}\right\|_\infty \qquad \text{(Jordan expansion of } A.\text{)}$$

$$= \left\|D_\varepsilon^{-1} J D_\varepsilon \mathbf{y}\right\|_\infty \qquad \text{(Simplifying.)}$$

$$\leq \left\|D_\varepsilon^{-1} J D_\varepsilon\right\|_\infty \|\mathbf{y}\|_\infty \qquad \text{(Consistency of norm.)}$$

$$= \left\|D_\varepsilon^{-1} J D_\varepsilon\right\|_\infty \left\|D_\varepsilon^{-1} X^{-1} \mathbf{x}\right\|_\infty \qquad \text{(Definition of } \mathbf{y}.\text{)}$$

$$= \left\|D_\varepsilon^{-1} J D_\varepsilon\right\|_\infty \|\mathbf{x}\|_* \qquad \text{(Definition of } \|\cdot\|_*.\text{)}$$

Thus, $\frac{\|A\mathbf{x}\|_*}{\|\mathbf{x}\|_*}$ is of the form $\left\|D_\varepsilon^{-1} J D_\varepsilon\right\|_\infty$. This is independent of $\mathbf{x}$, thus $\|A\| = \left\|D_\varepsilon^{-1} J D_\varepsilon\right\|_\infty$. From the previous part, we can see that $\|A\|$ is bounded above by $\rho(A) + \varepsilon$. Furthermore, since the ∞-matrix norm is consistent, $\|A\|$ is bounded below by $\rho(A)$ (As shown in class). Thus the property holds. $\square$

18

Problem 4

Note that for simplicity I will provide the results before the code block(s).

Let $A = [a_{ij}]$ be an $n \times n$ matrix with entries

$$a_{ij} = \begin{cases} n + 1 - \max(i, j) & i \leq j + 1, \\ 0 & i > j + 1 \end{cases}$$

This is an example of an *upper Hessenberg* matrix: it is upper triangular except the entries on the subdiagonal $a_{j+1,j}$ may also be non-zero. For $n = 12$ and $n = 25$, do the following:

Problem 4, part a

Compute $\|A\|_\infty$ and $\|A\|_1$.

Solution:

The methods to compute the 1-norm and ∞-norm are provided below, as well as the method to compute the Hessenberg matrix. The code which uses these functions are provided in the second block. . When running the code below, I get the following:

- Size 12 Hessenberg matrix: 1-norm: 55.0, inf-norm: 90.0

- Size 25 Hessenberg matrix: 1-norm: 195.0, inf-norm: 350.0

```python
#Methods block
def compute_hessenberg_matrix(n):
    """

    Returns the Hessenberg matrix of given size
    """

    A = np.empty((n, n))
    for i in range(n):
        for j in range(n):
            if i <= j + 1:
                A[i][j] = n + 1 - max(i,j)
            else:
                A[i][j] = 0
    return A


def compute_inf_norm(A):
    """
```

```python
    Returns the infinity norm of the given matrix
    """

    row_sums = []
    for i in range(A.shape[0]):
        row_sum = 0
        for j in range(A.shape[0]):
            row_sum += abs(A[i][j])
        row_sums.append(row_sum)

    return max(row_sums)


def compute_1_norm(A):
    """
    Returns the 1-norm of the given matrix
    """

    column_sums = []
    for i in range(A.shape[0]):
        column_sum = 0
        for j in range(A.shape[0]):
            column_sum += abs(A[j][i])
        column_sums.append(column_sum)

    return max(column_sums)
```

```python
#"Main"
hessenberg = [compute_hessenberg_matrix(x) for x in [12,25]]

#Computing the matrix inf and 1 norms for matrices
for matrix in hessenberg:
    print("Size {} Hessenberg matrix: 1-norm: {}, inf-norm: {}".format(matrix.shape[0],
        compute_1_norm(matrix), compute_inf_norm(matrix)))
```

Problem 4, part b

Compute $\rho(A)$ and $\|A\|_2$. You may use any built-in functions of your program.

Solution:

The code I used to calculate this is given below. The result when running this block is given as

- Size 12

  Spectral Radius: 35.61

  Matrix 2-norm 54.92

- Size 25

  Spectral Radius: 81.61

  Matrix 2-norm 194.43

```python
#Computing the spectral radius and 2-norm
for matrix in hessenberg:
    eigenvals = [x for x in abs(np.linalg.eigvals(matrix))]
    print("Size {} ".format(matrix.shape[0]))
    print("Spectral Radius: {:.2f}".format(max(eigenvals)))
    print("Matrix 2-norm {:.2f}".format(np.linalg.norm(matrix, ord=2)))
```

Problem 4, part c

Using Gerschgorin's theorem, describe the domain that contains all of the eigenvalues.

Solution:

The order of code shown will be given as: methods used, block to implement methods, then plotting. The resulting plots are given below. It seems that the radius of the balls increase with eigenvalue, resulting in the largest eigenvalue enclosing all smaller eigenvalues. Note that although the number of discs corresponds to the number of eigenalues, each disc does not have to contain at least one eigenvalue. I couldn't get the color bar to as I wanted, so just trust me on this.
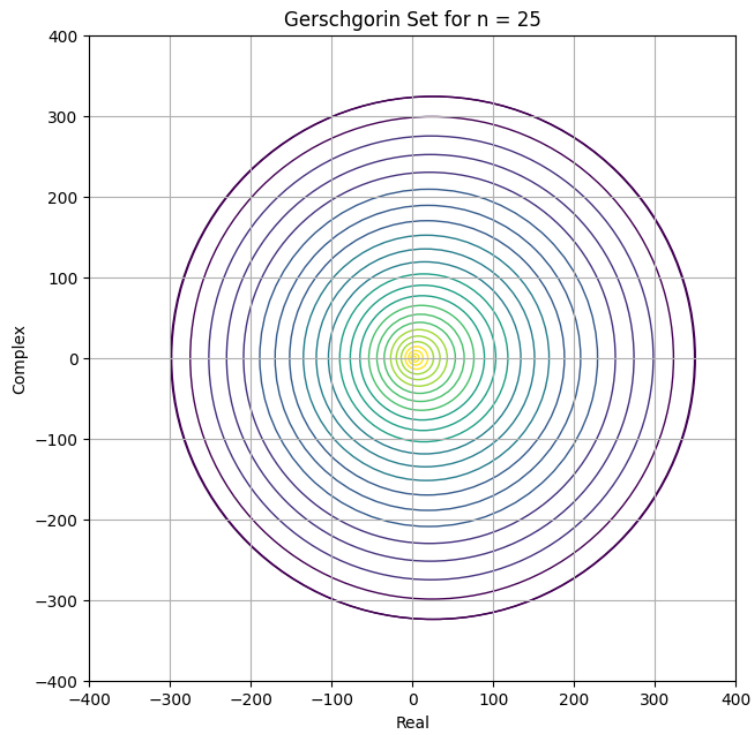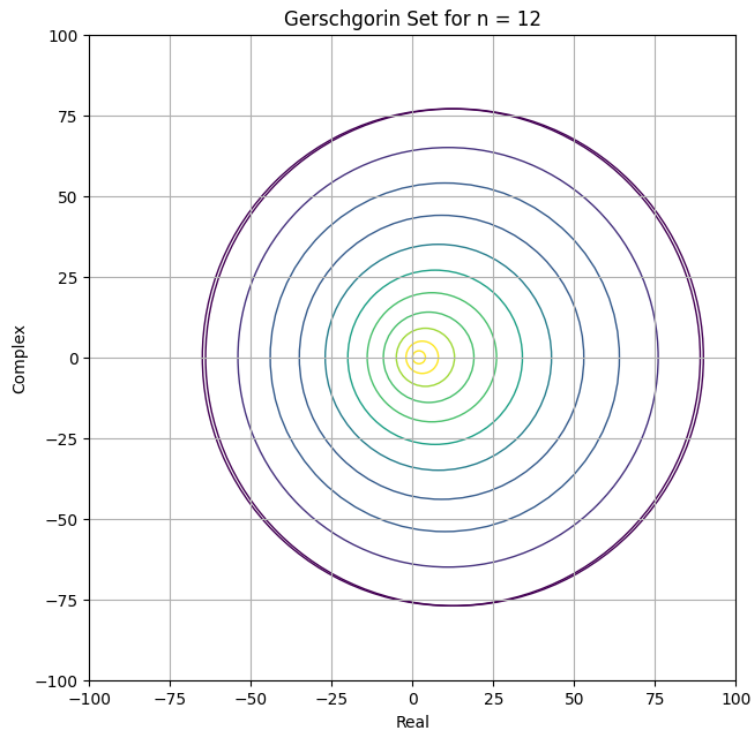
Figure 1: Gerschgorin sets for Hessenberg matrices of n = 12, 25. Larger values → darker colors.

```python
#Methods Block
def compute_gerschgorin_radius(row, diag_entry):
"""
Returns the summed absolute values of entries in the given row,
minus the specified entry index
"""
row_sum = 0
for i in range(len(row)):
    if i != diag_entry:
        row_sum += abs(row[i])

return row_sum
```

```python
#Computing tuples containing the diagonal entry and its radius for Gerschgorin's Theorem

gerschgorin_balls = []
for matrix in hessenberg:
matrix_balls = []
for i in range(matrix.shape[0]):
    matrix_balls.append((matrix[i][i], compute_gerschgorin_radius(matrix[i], i)))
    gerschgorin_balls.append(matrix_balls)
```

```python
#Plotting block
#Setting Color scale
#Larger values -> darker colors
viridis = mpl.colormaps['viridis'].resampled(8)
color12 = viridis(np.linspace(0, 1, 12))
color25 = viridis(np.linspace(0, 1, 25))

fig, ax = plt.subplots(2, figsize=(24,16))

#Plotting the gerschgorin plot for n = 12
i = 0
for tup in gerschgorin_balls[0]:
    center = (tup[0], 0)
    circle = Circle(center, tup[1], fill=False, color=color12[i])
    ax[0].add_patch(circle)
    i +=1

ax[0].set_title("Gerschgorin Set for n = 12")
ax[0].set_aspect('equal', adjustable='box')
```

23

```python
ax[0].set_xlim(-100, 100)
ax[0].set_xlabel("Real")
ax[0].set_ylim(-100, 100)
ax[0].set_ylabel("Complex")
ax[0].grid(True)


#Plotting the gerschgorin plot for n = 25
i = 0
for tup in gerschgorin_balls[1]:
    center = (tup[0], 0)
    circle = Circle(center, tup[1], fill=False, color=color25[i])
    ax[1].add_patch(circle)
    i+=1


ax[1].set_title("Gerschgorin Set for n = 25")
ax[1].set_aspect('equal', adjustable='box')
ax[1].set_xlim(-400, 400)
ax[1].set_xlabel("Real")
ax[1].set_ylim(-400, 400)
ax[1].set_ylabel("Complex")
ax[1].grid(True)
plt.show()
```

Problem 4, part d

Compute all of the eigenvalues and singular values of *A*. How many of the eigenvalues are real and how many are complex? You may use any built-in functions of your program.

The code I used to calculate this is shown below. The results from this code block is given as:

- Size 12

  Real Eigenvalues: 12, Complex Eigenvalues: 0

  Real Singular: 12, Complex Singular: 0

- Size 25

  Real Eigenvalues: 13, Complex Eigenvalues: 12

  Real Singular: 25, Complex Singular: 0

```python
for matrix in hessenberg:
    eigenvals = np.linalg.eigvals(matrix)
    real_eigenvals = 0
    complex_eigenvals = 0
    for val in eigenvals:
        if isinstance(val, complex):
            if val.imag != 0:
                complex_eigenvals+=1
            else:
                real_eigenvals+=1
        else:
            real_eigenvals+=1;

    real_sing = 0
    complex_sing = 0
    for sing in svdvals(matrix):
        if isinstance(sing, complex):
            if sing.imag != 0:
                complex_sing+=1
            else:
                real_sing+=1
        else:
            real_sing+=1;

    print("Size {}".format(matrix.shape[0]))
    print("Real Eigenvalues: {}, Complex Eigenvalues: {}".format(real_eigenvals,
        complex_eigenvals))
    print("Real Singular: {}, Complex Singular: {}".format(real_sing, complex_sing))
```

Problem 5

You are not allowed to use the svd for this problem, i.e., no arguments should depend on the svd of A or A^*.
Let W be a subspace of $\mathbb{C}^n$. The subspace $W^\perp$ below is called the orthogonal complement of W.

$$W^\perp = \{\mathbf{v} \in \mathbb{C}^n : \mathbf{v}^*\mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\}$$

For any subspace $W \subseteq \mathbb{C}^n$, we write $P_W \in \mathbb{C}^{n \times n}$ for an orthogonal projection onto W.

Problem 5, part a

Show that $\mathbb{C}^n = W \oplus W^\perp$ and that $W = (W^\perp)^\perp$.

Solution:

To show $\mathbb{C}^n = W \oplus W^\perp$, we first need to show $W \cap W^\perp = \{0\}$. Let $\mathbf{v} \in W$ and $\mathbf{v} \in W^\perp$, then $\mathbf{v}^*\mathbf{v} = 0 \implies \|\mathbf{v}\|^2 = 0 \iff \mathbf{v} = 0$. $\therefore W \cap W^\perp \subseteq \{0\}$. It is then easy to see that $\mathbf{0} \in W, W^\perp$, since they are vector spaces. So $W \cap W^\perp = \{0\}$.

Next we need to show $\mathbb{C}^n = W + W^\perp$. Let $\mathbf{v} \in \mathbb{C}^n$, $w_1 \in W$ st $\mathbf{v} = P_W\mathbf{w}_1$, and $\mathbf{x} \in W$. We want to show $\mathbf{w}_2 = \mathbf{v} - P_W\mathbf{v} \in W^\perp$.

$$\mathbf{w}_2^*\mathbf{x} = \mathbf{v}^*\mathbf{x} - \mathbf{v}^*P_W^*\mathbf{x} \qquad \text{(Definition of inner product.)}$$

$$= \mathbf{v}^*\mathbf{x} - \mathbf{v}^*P_W\mathbf{x} \qquad (P_W^* = P_W.)$$

$$= \mathbf{v}^*\mathbf{x} - \mathbf{v}^*\mathbf{x} \qquad (\mathbf{x} \in W.)$$

$$\implies \mathbf{w}_2^*\mathbf{x} = 0$$

Thus $\mathbf{w}_2 \in W^\perp$. Next take $\mathbf{w}_1 \in W$, $\mathbf{w}_2 \in W^\perp$.. Then $\mathbf{w}_1 + \mathbf{w}_2 \in \mathbb{C}^n$ as W and $W^\perp$ are vector spaces. Thus $\mathbb{C}^n = W \oplus W^\perp$.

We then wish to show $(W \perp)^\perp = W$. Note that the above proof was done for any subspace $W \in \mathbb{C}^n$. Thus is we swap W with $W \perp$, we get $\mathbb{C}^n = W^{perp} \oplus (W^\perp)^\perp$. Since the direct sum operator is commutative, this implies $W = (W^\perp)^\perp$.

Problem 5, part b

Let $A \in \mathbb{C}^{m \times n}$. Show that
$$\ker(A^*) = \operatorname{im}(A)^\perp \quad \text{and} \quad \operatorname{im}(A^*) = \ker(A)^\perp.$$

Solution:

Let's take the first inequality. For the forward implication, take $\mathbf{w} \in \ker(A^*)$, $\mathbf{y} \in \operatorname{im}(A)$. Thus $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{C}^n$. Then,
$$\mathbf{v}^*\mathbf{y} = \mathbf{v}^*A\mathbf{x} = \mathbf{x}^*A^*\mathbf{v} = \mathbf{x}^*(A^*\mathbf{v}) = 0$$

So $\mathbf{w} \in \operatorname{im}(A)^\perp$.

For the second inequality, let $\mathbf{v} \in \operatorname{im}(A)$. Thus there $A\mathbf{u} = \mathbf{v}$ for some $\mathbf{u} \in \mathbb{C}^n$. So $\mathbf{v}^* = \mathbf{u}^*A^*$. Thus the following steps can be justified:

$$\mathbf{u}^*\mathbf{v} = \mathbf{v}^*\mathbf{w}$$

$$\implies \mathbf{u}^*A^*\mathbf{w} = 0$$

$$\implies \mathbf{u}^*(A^*\mathbf{w}) = 0$$

$$\implies A^*\mathbf{w} = 0$$

Therefore $\mathbf{w} \in \ker(A^*)$, completing the equality. Note to get the next equality, we simply replace A with A^* and taking the orthogonal complement to both sides. I.e., $\ker(A) = \operatorname{im}(A^*)^\perp$ by the first equality, then $\ker(A)^\perp = (\operatorname{im}(A^*)^\perp)^\perp = \operatorname{im}(A^*)$. Thus the second equality is obtained.

Problem 5, part c

Deduce the Fredholm alternative

$$\mathbb{C}^m = \ker(A^*) \oplus \text{im}(A) \quad \text{and} \quad \mathbb{C}^n = \text{im}(A^*) \oplus \ker(A).$$

In other words, any $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$ can be written uniquely as

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_1, \quad \mathbf{x}_0 \in \ker(A), \mathbf{x}_1 \in \text{im}(A^*), \mathbf{x}_0^* \mathbf{x}_1 = 0$$

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{y}_1, \quad \mathbf{y}_0 \in \ker(A^*), \mathbf{y}_1 \in \text{im}(A), \mathbf{y}_0^* \mathbf{x}_1 = 0$$

Solution:

Since $\ker(A^*)$ is equal to $\text{im}(A)^\perp$ from the previous part, then $\mathbb{C}^m = \text{im}(A) \oplus \text{im}(A)^\perp = \text{im}(A) \oplus \ker(A^*)$. We can similarly swap $\ker(A)^\perp$ with $\text{im}(A^*)$ to get $\mathbb{C}^n = \text{im}(A^*) \oplus \ker(A)$.

Problem 5, part d

Show that

$$\mathbf{x}_0 = P_{\text{ker}(A)}\mathbf{x}, \quad \mathbf{x}_1 = P_{\text{im}(A^*)}\mathbf{x}, \quad \mathbf{y}_0 = P_{\text{ker}(A^*)}\mathbf{y}, \quad \mathbf{y}_1 = P_{\text{im}(A)}\mathbf{y}$$

Solution:

(i)

$$P_{\text{ker}(A)}(\mathbf{x}) = P_{\text{ker}(A)}\mathbf{x}_0 + P_{\text{ker}(A)}\mathbf{x}_1$$

$$= \mathbf{x}_0 + P_{\text{ker}(A)}\mathbf{x}_1$$

$$= \mathbf{x}_0$$

Note that since $\mathbf{x}_0$ is in the kernel of A, the projection does nothing to it. Also, since $\mathbf{x}_1$ is in the image of A^*, it is in the orthogonal complement of the kernel of A, thus not in the kernel of A if $\mathbf{x}_1$ is $\mathbf{0}$. If it is zero, then its projection is still zero, so it is just adding nothing in either case. Thus equality is shown.

(ii)

$$P_{\text{im}(A^*)}(\mathbf{x}) = P_{\text{im}(A^*)}\mathbf{x}_0 + P_{\text{im}(A^*)}\mathbf{x}_1$$

$$= P_{\text{im}(A^*)}\mathbf{x}_0 + \mathbf{x}_1$$

$$= \mathbf{x}_1$$

Note this holds by the same argument given in (i), using the image of A^* is equal to the orthogonal complement of the image of A.

(iii)

$$P_{\text{ker}(A^*)}(\mathbf{y}) = P_{\text{ker}(A^*)}\mathbf{y}_0 + P_{\text{ker}(A^*)}\mathbf{y}_1$$

$$= \mathbf{y}_0 + P_{\text{ker}(A^*)}\mathbf{y}_1$$

$$= \mathbf{y}_0$$

Note this is the same as (i), except using the second equality in part c.

(iv)

$$P_{\text{im}(A)}(\mathbf{y}) = P_{\text{im}(A)}\mathbf{y}_0 + P_{\text{im}(A)}\mathbf{y}_1$$

$$= P_{\text{im}(A)}\mathbf{y}_0 + \mathbf{y}_1$$

$$= \mathbf{y}_1$$

Note this is the same as (ii), except using the second equality in part c.

Problem 5, part e

Consider the least squares problem for some $\mathbf{b} \in \mathbb{C}^m$,

$$\min_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{b} - A\mathbf{x}\|_2. \tag{9}$$

show that for any $\mathbf{x} \in \mathbb{C}^n$,

$$\|\mathbf{b} - A\mathbf{x}\|_2 \geq \|\mathbf{b}_0\|_2$$

where $\mathbf{b}_0 = P_{\ker(A^*)}\mathbf{b}$. Deduce that $\mathbf{x} \in \mathbb{C}^n$ is a solution to the least squares problem if and only if

$$A\mathbf{x} = \mathbf{b}_1 \quad \text{or, equivalently} \quad \mathbf{b} - A\mathbf{x} = \mathbf{b}_0. \tag{10}$$

Why is $A\mathbf{x} = \mathbf{b}_1$ consistent?

Solution:

Since $\mathbf{b} \in \mathbb{C}^m$, we can represent $\mathbf{b}$ as $\mathbf{b} = \mathbf{b}_0 + \mathbf{b}_1$, where $\mathbf{b}_0 \in \ker(A^*)$ and $\mathbf{b}_1 \in \text{im}(A)$. Thus expanding 9 yields the following:

$$\|\mathbf{b} - A\mathbf{x}\|_2^2 \qquad\qquad \text{(Given.)}$$

$$= \|\mathbf{b}_0 + \mathbf{b}_1 - A\mathbf{x}\|_2^2 \qquad\qquad \text{(From part c.)}$$

$$= (b_0 + b_1 - Ax)^*(b_0 + b_1 - Ax) \qquad\qquad \text{(Expanding.)}$$

$$= b_0^* b_0 + b_0^* b_1 - b_0^* Ax + b_1^* b_0 + b_1^* b_1 - b_1^* Ax - (Ax)^* b_0 - (Ax)^* b_1 + (Ax)^*(Ax) \qquad\qquad \text{(Expanding.)}$$

$$= b_0^* b_0 - b_0^* Ax + b_1^* b_1 - b_1^* Ax - (Ax)^* b_0 - (Ax)^* b_1 + (Ax)^*(Ax) \qquad\qquad \text{(Orthogonality.)}$$

$$= b_0^* b_0 - A^* b_0 x + b_1^* b_1 - b_1^* Ax - x^* A^* b_0 - (Ax)^* b_1 + (Ax)^*(Ax) \qquad\qquad \text{(Rearranging.)}$$

$$= b_0^* b_0 + b_1^* b_1 - b_1^* Ax - (Ax)^* b_1 + (Ax)^*(Ax) \qquad\qquad (A^* b_0 = 0.)$$

$$= \|\mathbf{b}_0\|_2^2 + \|\mathbf{b}_1 - A\mathbf{x}\|_2^2 \qquad\qquad \text{(Grouping).}$$

$$\geq \|\mathbf{b}_0\|_2^2 + \min_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{b}_1 - A\mathbf{x}\|_2^2 \qquad\qquad \text{(Property of min.)}$$

$$= \|\mathbf{b}_0\|_2^2 + \min_{\mathbf{x} \in \mathbb{C}^n} \|A(\mathbf{c}_1 - \mathbf{x})\|_2^2 \qquad\qquad (\mathbf{b}_1 \in \text{im}(A).)$$

$$= \|\mathbf{b}_0\|_2^2 \qquad\qquad (\text{min is when } \mathbf{x} = \mathbf{c}_1.)$$

Thus $\|\mathbf{b} - A\mathbf{x}\|_2 \geq \|\mathbf{b}_0\|_2$ by the above calculations. Next we wish to show that $\mathbf{x}$ is a solution to 9 if and only if $A\mathbf{x} = \mathbf{b}_1$. If $\mathbf{x}$ is a solution to 9, then $\mathbf{x}$ is the minimum of $\|A(\mathbf{c}_1 - \mathbf{x})\|$ for some $A\mathbf{c}_1 = \mathbf{b}_1$. But the minimum only happens when $\mathbf{c}_1 - \mathbf{x} = 0$, thus $\mathbf{c}_1 = \mathbf{x}$, or $A\mathbf{x} = \mathbf{b}_1$. Similarly, if $\mathbf{b} - A\mathbf{x} = \mathbf{b}_0$, then $\mathbf{b}_0 + \mathbf{b}_1 - A\mathbf{x} = \mathbf{b}_0$. Simplifying yields $\mathbf{b}_1 - A\mathbf{x} = 0$, and since $\mathbf{b}_1 \in \mathrm{im}(A)$, $\mathbf{b}_1 = A\mathbf{c}_1$ for some $\mathbf{c}_1 \in \mathbb{C}^n$. then $A(\mathbf{c}_1 - \mathbf{x}) = 0$, which means $\mathbf{c}_1 = \mathbf{x}$. Thus $\mathbf{x}$ is a solution to 9. Note $A\mathbf{x} = \mathbf{b}_1$ is consistent since $\mathbf{b}_1 \in \mathrm{im}(A)$.

Problem 5, part f

Show that 10 is equivalent to the normal equation

$$A^*A\mathbf{x} = A^*\mathbf{b} \tag{11}$$

Solution:

For the forward direction, we assume $A\mathbf{x} = \mathbf{b}_1$ and will derive 11. Noting that $\mathbf{b}_0 \in \ker(A^*)$, we can "add a zero" to the right hand side when multiplying through by A^* to get $A^*A\mathbf{x} = A^*\mathbf{b}_1 + A^*\mathbf{b}_0$. The right hand side thus equals $A(\mathbf{b}_0 + \mathbf{b}_1)$, which simplifies to $A^*\mathbf{b}$. This implies 11 holds.

Next we assume 11 and derive 10. We expand the right hand side of 11 to get $A^*A\mathbf{x} = A^*(\mathbf{b}_0 + \mathbf{b}_1)$. Noting that $\mathbf{b}_0 \in \ker(A^*)$, we can simplify and rearrange to get $A^*(A\mathbf{x} - \mathbf{b}_1) = 0$. This means $A\mathbf{x} - \mathbf{b}_1 = 0$, or $A\mathbf{x} = \mathbf{b}_1$, which is 10. Thus 10 is equivalent to 11.

Problem 5, part g

Show that the pseudoinverse solution

$$\min \left\{ \|\mathbf{x}\|_2 : \mathbf{x} \in \operatorname*{argmin}_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{b} - A\mathbf{x}\|_2 \right\}$$

is given by

$$\mathbf{x}_1 = P_{\operatorname{im}(A^*)}\mathbf{x}$$

where $\mathbf{x} \in \mathbb{C}^n$ satisfies 10.

Solution:

Since we are given $\mathbf{x}$ satisfies 10, then we know that $\mathbf{x} \in \operatorname{argmin}\{\|\mathbf{b} - A\mathbf{x}\|_2\}$. So, the minimum over this set will have the form justified below.

$$\|\mathbf{x}\|_2^2 = \|\mathbf{x}_0 + \mathbf{x}_1\|_2^2 \qquad\qquad\qquad \text{(By part c.)}$$

$$= (\mathbf{x}_0 + \mathbf{x}_1)^*(\mathbf{x}_0 + \mathbf{x}_1) \qquad\qquad \text{(Definition of 2-norm for vectors.)}$$

$$= \mathbf{x}_0^*\mathbf{x}_0 + \mathbf{x}_0^*\mathbf{x}_1 + \mathbf{x}_1^*\mathbf{x}_0 + \mathbf{x}_1^*\mathbf{x}_1 \qquad\qquad \text{(Expanding.)}$$

$$= \mathbf{x}_0^*\mathbf{x}_0 + \mathbf{x}_1^*\mathbf{x}_1 \qquad\qquad \text{(Orthogonality of spaces.)}$$

Thus, $\min\{\|\mathbf{x}\|_2\} = \min\{\mathbf{x}_0^*\mathbf{x}_0 + \mathbf{x}_1^*\mathbf{x}_1\}$. Since $\mathbf{x}_0$ is the projection of $\mathbf{x}$ onto the kernel of A, this part will equal zero when A is applied (equivalently, $\mathbf{x}$ satisfies 10). Then, the minimum of this set will have zero projection onto the kernel of A, so the solution to the pseudoinverse problem is given by $\mathbf{x}_1 = P_{\operatorname{im}(A^*)}\mathbf{x}$.

Problem 5, part h

Let $A \in \mathbb{C}^{n \times n}$ be normal, i.e., $A^*A = AA^*$. Show that

$$\ker(A^*) = \ker(A) \quad \text{and} \quad \text{im}(A^*) = \text{im}(A)$$

and deduce that for a normal matrix,

$$\mathbb{C}^n = \ker(A) \oplus \text{im}(A).$$

Solution: Note that by applying the Fredholm alternative of $\mathbb{C}^n$ immediately implies $\mathbb{C}^n = \ker(A) \oplus \text{im}(A)$ if the two conditions above hold. I will thus show them to be true for $A\mathbb{C}^{n \times n}$.

$A^*\mathbf{x} = 0$	($\mathbf{x} \in \ker(A^*)$.)
$\implies \|A^*\mathbf{x}\|_2 = 0$	(Definition of 2-norm).
$(A^*\mathbf{x})^*A^*\mathbf{x} = 0$	(Definition of 2-norm).
$\mathbf{x}^*AA^*\mathbf{x} = 0$	(Applying conjugate.)
$\mathbf{x}^*A^*A\mathbf{x} = 0$	(A is normal.)
$(A\mathbf{x})^*A\mathbf{x} = 0$	(Applying conjugate.)
$\implies \|A\mathbf{x}\|_2 = 0$	(Definition of 2-norm).
$A\mathbf{x} = 0$	($\mathbf{x} \in \ker(A)$.)

Note I wrote this proof to specifically be symmetric, i.e. to get the forward implication you go down and for the backward implication you go up. They are pretty much the exact same proof, so I thought I could condense it down into one proof.

We next need to show the same for the images of A and A^*. If $\mathbf{y} \in \text{im}(A)$, this means there exists some $\mathbf{x} \in \mathbb{C}^n$ with $\mathbf{y} = A\mathbf{x}$. Applying the hermitian operator to both sides yields $\mathbf{y}^* = \mathbf{x}^*A^*$. We can note that $\mathbf{x}^*A^*$ is some linear combination of the rows of A^*, which means that $\mathbf{y}^*$ is in the row space of A, which means $\mathbf{y}$ is in the column space of A^*. This proof can be reversed in the exact same order to show that if $\mathbf{y} \in \text{im}(A^*)$, then $\mathbf{y} \in \text{im}(A)$, thus $\text{im}(A) = \text{im}(A^*)$.

36

Problem 6

Let $\mathbf{x} \in \mathbb{C}^m$, $\mathbf{y} \in \mathbb{C}^n$, and $A = \mathbf{x}\mathbf{y}^* \in \mathbb{C}^{m \times n}$.

Problem 6, part a

Show that

$$\|A\|_F = \|A\|_2 = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

and that

$$\|A\|_\infty = \|\mathbf{x}\|_\infty \|\mathbf{y}\|_1.$$

What can you say about $\|A\|_1$?

Solution:

Note that the Frobenius norm and martix 2-norm are given as

$$\|A\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2$$

$$\|A\|_2 = \max_{\mathbf{z} \neq 0} \frac{\|A\mathbf{z}\|_2}{\|\mathbf{z}\|_2}$$

If $A = \mathbf{x}\mathbf{y}^*$, then we can imagine each column of A as the vector $\mathbf{x}$ multiplied by the corresponding row element of $\mathbf{y}$. In other words,

$$A = a_{ij} = \mathbf{x}_i \mathbf{y}_j^*$$

This is somewhat obvious, but I want to be thorough. We will next show both norms are equivalent to the product of the two norms on $\mathbf{x}$ and $\mathbf{y}$. Note we will be looking at the squared norms for each, since they are equivalent.

$$\|A\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2 \qquad \text{(Given.)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} |x_i \bar{y}_j|^2 \qquad \text{(By observation.)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} (\overline{x_i \bar{y}_j})(x_i \bar{y}_j) \qquad \text{(Absolute value over } \mathbb{C}.)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \bar{x}_i y_j x_i \bar{y}_j \qquad \text{(Conjugation.)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \bar{x}_i x_i y_j \bar{y}_j \qquad \text{(Commutivity over } \mathbb{C}.)$$

$$= \sum_{i=1}^{n} |x_i|^2 \sum_{j=1}^{m} |y_j|^2 \qquad \text{(Rearranging and norm over } \mathbb{C}.)$$

$$\implies \|A\|_F^2 = \|\mathbf{x}\|_2^2 \|\mathbf{y}\|_2^2 \qquad \text{(Definition of 2-norm.)}$$

Thus we have shown the Frobenius norm case. Next we will show the 2-norm of A is also equivalent.

$$\|A\|_2^2 = \max_{\mathbf{z} \neq 0} \frac{\|A\mathbf{z}\|_2}{\|\mathbf{z}\|_2^2} \qquad \text{(Given.)}$$

$$= \frac{\|\mathbf{x}\mathbf{y}^*\mathbf{z}\|_2}{\|\mathbf{z}\|_2^2} \qquad \text{(Inspection of argument and given.)}$$

$$= \frac{(\mathbf{x}\mathbf{y}^*\mathbf{z})^*(\mathbf{x}\mathbf{y}^*\mathbf{z})}{\mathbf{z}^*\mathbf{z}} \qquad \text{(Definition of 2-norm.)}$$

$$= \frac{\mathbf{z}^*\mathbf{y}\mathbf{x}^*\mathbf{x}\mathbf{y}^*\mathbf{z}}{\mathbf{z}^*\mathbf{z}} \qquad \text{(Hermitian definition.)}$$

$$= \frac{\|\mathbf{x}\|_2^2 \mathbf{z}^*\mathbf{y}\mathbf{y}^*\mathbf{z}}{\mathbf{z}^*\mathbf{z}} \qquad \text{(Definition of 2-norm and commutivity.)}$$

$$= \frac{\|\mathbf{x}\|_2^2 \|\mathbf{y}\|_2^2 \|\mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2} \qquad \text{(Definition of 2-norm.)}$$

$$\implies \|A\|_2^2 = \|\mathbf{x}\|_2^2 \|\mathbf{y}\|_2^2 \qquad \text{(Simplifying.)}$$

Thus we have shown both norms are equivalent to the product of the 2-norms of the relevant vectors $\mathbf{x}$ and $\mathbf{y}$. Note that by transitivity property of equality, we can say $\|A\|_F = \|A\|_2$ when $A = \mathbf{x}\mathbf{y}^*$. Next we want to show $\|A\|_\infty = \|\mathbf{x}\|_\infty \|\mathbf{y}\|_1$.

$$\|A\|_\infty = \max_{1 \le i \le m} \left\{ \sum_{j=1}^n |a_{ij}| \right\} \qquad \text{(Given.)}$$

$$= \max_{1 \le i \le m} \left\{ \sum_{j=1}^n |x_i \bar{y}_j| \right\} \qquad \text{(By observation.)}$$

$$= \max_{1 \le i \le m} \left\{ \sum_{j=1}^n \sqrt{\bar{x}_i y_j x_i \bar{y}_j} \right\} \qquad \text{(Absolute value over } \mathbb{C}\text{.)}$$

$$= \max_{1 \le i \le m} \left\{ \sum_{j=1}^n \sqrt{\bar{x}_i x_i y_j \bar{y}_j} \right\} \qquad \text{(Commutivity over } \mathbb{C}\text{.)}$$

$$= \max_{1 \le i \le m} \left\{ \sqrt{\bar{x}_i x_i} \sum_{j=1}^n \sqrt{y_j \bar{y}_j} \right\} \qquad \text{Property of square root and rearranging.)}$$

$$= \max_{1 \le i \le m} \left\{ \sqrt{\bar{x}_i x_i} \right\} \left(\sum_{j=1}^n \sqrt{y_j \bar{y}_j} \right) \qquad \text{(Sum has no dependence on } i\text{.)}$$

$$= \max_{1 \le i \le m} \left\{ |x_i| \right\} \left(\sum_{j=1}^n |y_j| \right) \qquad \text{(Absolute value definition.)}$$

$$\implies \|A\|_\infty = \|\mathbf{x}\|_\infty \|\mathbf{y}\|_1 \qquad \text{(Definition of } 1, \infty\text{- norms.)}$$

Thus we have shown the claim. I would imagine the $\|A\|_1$ is swapping norms around, meaning $\|A\|_1 = \|\mathbf{x}\|_1 \|\mathbf{y}\|_\infty$. The proofs are very similar, except instead of pulling $|x_i|$ out of the sum, you pull $|y_j|$ out, then the result follows. For the sake of brevity I will omit this proof, but just note the way to show it is how I explained it.

Problem 6, part b

Let $\mathbf{x}_1, ..., \mathbf{x}_r \in \mathbb{C}^m$ be linearly independent and $\mathbf{y}_1, ..., \mathbf{y}_r$ be linearly independent. Let

$$A = \mathbf{x}_1\mathbf{y}_1^* + ... + \mathbf{x}_r\mathbf{y}_r^*$$

Show that $\text{rank}(A) = r$. Show that this is not necessarily true if we drop either of the linear independence conditions.

Solution:

There might be more elegant ways to show this, but I will prove this by induction on r. First, denote A_k as the first k summations of the rank-1 matrices.

Case 1: r = 1.

Then $A_1 = \mathbf{x}_1\mathbf{y}_1^*$. This is obviously rank 1, since its columns are linear combinations of the $\mathbf{x}_1$ vector. Thus the base case holds.

Next Suppose this holds for some $k \le r$. We will show the $k + 1$ case follows.

Case $k + 1$: r = k + 1

Suppose false, that is, $\text{rank}(A_{k+1}) = k$. This means that A_{k+1} does not have full rank, which is equivalent to saying there exists some nonzero $\mathbf{v} \in \mathbb{C}^n$ such that $A_{k+1}\mathbf{v} = 0$ (I.e, something maps to the kernel that is not the zero vector). Upon expansion of A_{k+1}, we see that

$$(A_k + \mathbf{x}_{k+1}\mathbf{y}_{k+1}^*)\mathbf{v} = 0$$

$$\iff \mathbf{x}_1\mathbf{y}_1^*\mathbf{v} + ... + \mathbf{x}_{k+1}\mathbf{y}_{k+1}^*\mathbf{v} = 0$$

By associativity, $\mathbf{x}_i\mathbf{y}_i^*\mathbf{v} = \mathbf{x}_i(\mathbf{y}_i^*\mathbf{v})$ for all $i \le k + 1$. Observe that $\mathbf{y}_i^*\mathbf{v}$ is some value in $\mathbb{C}$. These can be swapped around with the $\mathbf{x}_i$'s to get

$$(\mathbf{y}_1^*\mathbf{v})\mathbf{x}_1 + ... + (\mathbf{y}_{k+1}^*\mathbf{v})\mathbf{x}_{k+1} = 0$$

Since $\mathbf{y}_i^*\mathbf{v}$ are just scalar constants over $\mathbb{C}$, and $\mathbf{x}_1, ..., \mathbf{x}_{k+1}$ are linearly independent (by assumption), we get that $\mathbf{y}_i^*\mathbf{v} = 0$ for all $i \le k + 1$. Thus the following steps hold:

$$\sum_{i=1}^{k+1} \mathbf{y}_i^* \mathbf{v} = 0 \qquad\qquad \text{(Sum over zeros is zero.)}$$

$$\sum_{i=1}^{k+1} \mathbf{v}^* \mathbf{y}_i = 0 \qquad\qquad \text{(Conjugation.)}$$

$$\mathbf{v}^* \sum_{i=1}^{k+1} \mathbf{y}_i = 0 \qquad\qquad (\mathbf{v}^* \text{ is independent of the sum.})$$

We can note the sum can be thought of as a linear combination of the $\mathbf{y}_i$'s with coefficient 1. Since the $\mathbf{y}_i$'s are linearly independent, and with nonzero coefficients, their linear combination is nonzero. This means that $\mathbf{v}^* = 0$, which is a contradiction, since $\mathbf{v}$ was taken to be nonzero. Therefore we have proven the induction step. $\square$

Note the proof relies on the fact that both $\mathbf{x}_i$'s and $\mathbf{y}_i$'s are linearly independent. If at least one set of vectors were linearly *dependent*, then $\mathbf{v}$ is not necessarily zero.

Problem 6, part c

Given an $0 \neq A \in \mathbb{C}^{m \times n}$, show that

$$\text{rank}(A) = \min \left\{ r \in \mathbb{N} : A = \sum_{i=1}^{r} \mathbf{x}_i \mathbf{y}_i^* \right\}$$

In other words, the rank of a matrix is the smallest r so that it may be expressed as a sum of r rank-1 matrices.

Solution:

Suppose we have two sets of vectors $\mathbf{x}_1, ..., \mathbf{x}_k \in \mathbb{C}^m$ and $\mathbf{y}_1, ..., \mathbf{y}_k \in \mathbb{C}^n$ such that

$$A = \sum_{i=1}^{k} \mathbf{x}_i \mathbf{y}_i^*.$$

Note that neither set of vectors are not necessarily linearly independent. Suppose further that from these two sets we can pick $r \leq k$ linearly independent vectors (if the two sets have a maximally different number of linearly independent, pick the minimum of the two). Then we can say that

$$\text{rank}\left(\sum_{i=1}^{r} \mathbf{x}_i \mathbf{y}_i^* \right) = r$$

From the last part, we showed that $\text{rank}(A) = r$, rather, A can be expressed as the sum of the outer product of two linearly independent vectors. We can recover the $\geq r$ outer products by scaling the first r vectors, since they are each $\geq r$ case is a linear combination of the previous vectors. Thus the following steps hold:

$$\text{rank}(A) = \text{rank}(\sum_{i=1}^{k} \mathbf{x}_i \mathbf{y}_i^*) \qquad\qquad \text{(Given,)}$$

$$= \text{rank}(\sum_{i=1}^{r} \mathbf{x}_i \mathbf{y}_i^* + \sum_{i=r+1}^{k} \mathbf{x}_i \mathbf{y}_i^*) \qquad\qquad \text{(Breaking up summation.)}$$

$$\leq \text{rank}(\sum_{i=1}^{r} \mathbf{x}_i \mathbf{y}_i^*) + \text{rank}(\sum_{i=r+1}^{k} \mathbf{x}_i \mathbf{y}_i^*) \qquad\qquad \text{(Homework 0, problem 1b.)}$$

$$= r + \text{rank}(\sum_{i=r+1}^{k} \mathbf{x}_i \mathbf{y}_i^*) \qquad\qquad \text{(By construction.)}$$

$$\implies \text{rank}(A) \leq \text{rank}(A) + \text{rank}(\sum_{i=r+1}^{k} \mathbf{x}_i \mathbf{y}_i^*) \qquad\qquad \text{(By last part.)}$$

Then equality is attained when $\text{rank}(\sum_{i=r+1}^{k} \mathbf{x}_i \mathbf{y}_i^*) = 0$, which is when none of the vectors $\mathbf{x}_{r+1}, ..., \mathbf{x}_k$ and $\mathbf{y}_{r+1}, ..., \mathbf{y}_k$ are linearly independent. Therefore,

$$\text{rank}(A) = \min \left\{ r \in \mathbb{N} : A = \sum_{i=1}^{r} \mathbf{x}_i \mathbf{y}_i^* \right\}$$

Thus the claim has been proven.

Problem 6, part d

Show the following generalization of part a,

$$\|A\|_F \leq \sqrt{\mathrm{rank}(A)}\|A\|_2$$

Solution:

Before I show the inequality, I would like to write an equivalent form for $\|A\|_2$. Looking at the inside of the max and for some $\mathbf{z}$ nonzero, we can write to following:

$$\frac{\|A\mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2} \qquad\qquad \text{(Given.)}$$

$$= \frac{\left\|\sum_{i=1}^r \mathbf{x}_i \mathbf{y}_i^* \mathbf{z}\right\|_2^2}{\|\mathbf{z}\|_2^2} \qquad\qquad \text{(By previous parts.)}$$

$$\leq \frac{\sum_{i=1}^r \left\|\mathbf{x}_i \mathbf{y}_i^* \mathbf{z}\right\|_2^2}{\|\mathbf{z}\|_2^2} \qquad\qquad \text{(Triangle inequality.)}$$

$$\leq \frac{\sum_{i=1}^r \|\mathbf{x}_i\|_2^2 \|\mathbf{y}_i\|_2^2 \|\mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2} \qquad\qquad \text{(Cauchy - Schwartz inequality.)}$$

$$= \sum_{i=1}^r \|\mathbf{x}_i\|_2^2 \|\mathbf{y}_i\|_2^2 \qquad\qquad \text{(Simplifying.)}$$

$$\implies \|A\|_2^2 = \sum_{i=1}^r \|\mathbf{x}_i\|_2^2 \|\mathbf{y}_i\|_2^2 \qquad\qquad (\|A\|_2^2 \text{ will have this form.}) \qquad (12)$$

Thus, the following steps are justified:

$$\|A\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \qquad \text{(Given.)}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \left| \sum_{k=1}^{r} x_k^i \bar{y}_k^j \right|^2 \qquad \text{(By inspection of elements of A.)}$$

$$\leq \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{r} |x_k^i \bar{y}_k^j|^2 \qquad \text{(Triangle inequality.)}$$

$$= \sum_{k=1}^{r} \sum_{i=1}^{m} \sum_{j=1}^{n} |x_k^i \bar{y}_k^j|^2 \qquad \text{(Rearranging sums.)}$$

$$= \sum_{k=1}^{r} \sum_{i=1}^{m} \sum_{j=1}^{n} |x_k^i|^2 |y_k^j|^2 \qquad \text{(Same as in part a.)}$$

$$= \sum_{k=1}^{r} \left[\left(\sum_{i=1}^{m} |x_k^i|^2 \right) \left(\sum_{j=1}^{n} |y_k^j|^2 \right) \right] \qquad \text{(Separating sums.)}$$

$$= \sum_{k=1}^{r} \|\mathbf{x}_k\|_2^2 \|\mathbf{y}_k\|_2^2 \qquad \text{(Definition of 2-norm.)}$$

$$\leq \sum_{k=1}^{r} \max \left\{ \|\mathbf{x}_l\|_2^2 \|\mathbf{y}_l\|_2^2 : 1 \leq l \leq r \right\} \qquad \text{(Sum of max is more than sum over all.)}$$

$$= r \cdot \max \left\{ \|\mathbf{x}_l\|_2^2 \|\mathbf{y}_l\|_2^2 : 1 \leq l \leq r \right\} \qquad \text{(Max is constant over sum.)}$$

$$= \text{rank}(A) \|\mathbf{x}_l\|_2^2 \|\mathbf{y}_l\|_2^2 \qquad \text{(Definition of rank}(A) \text{ and } l \text{ index is max.)}$$

$$\leq \text{rank}(A) \sum_{i=1}^{r} \|\mathbf{x}_i\|_2^2 \|\mathbf{y}_i\|_2^2 \qquad \text{(Max is one of the indices in the sum.)}$$

$$\implies \|A\|_F^2 = \text{rank}(A) \|A\|_2^2 \qquad \text{(From 12.)}$$

$$\iff \|A\|_F = \sqrt{\text{rank}(A)} \|A\|_2 \qquad \text{(Square root is monotonic.)}$$

Thus the statement has been shown. $\square$

Problem 6, part e

Show that with the nuclear norm we get instead

$$\|A\|_* \le \text{rank}(A)\|A\|_2$$

Solution:

For the sake of completion, I will include the nuclear norm of a matrix below, as well as an equivalent expression for the matrix 2-norm.

$$\|A\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(A), \tag{13}$$

$$\|A\|_2 = \max\{\sigma_1, ..., \sigma_n\} = \|\boldsymbol{\sigma}\|_\infty, \tag{14}$$

---

where $\sigma_i(A)$ is the i-th singular value of A when SVD is applied and $\boldsymbol{\sigma}$ is the vector of singular values of A. Let $r = \text{rank}(A)$, then 13 can be summed up to the r, since the singular values beyond r will be zero. Thus, the following steps are justified:

$$\|A\|_* = \sum_{i=1}^{r} \sigma_i \qquad \text{(Given.)}$$

$$\le \sum_{i=1}^{r} \max\{\sigma_1, ..., \sigma_n\} \qquad \text{(sum of max} \ge \text{sum over all.)}$$

$$= r \cdot \max\{\sigma_1, ..., \sigma_n\} \qquad \text{(Max is constant over sum.)}$$

$$\implies \|A\|_* \le \text{rank}(A) \cdot \|A\|_2 \qquad \text{(By 14.)} \tag{15}$$

Thus, the statement has been shown to hold.