

STAT 30900: Homework 0

Caleb Derrickson

October 8, 2023

Collaborators: I had the help of Kevin Hefner and Alexander Cram, as well as TA Andrew Dennehy on problems 5 and 6.

1. (a) For $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, show that $\text{im}(AB) \subseteq \text{im}(A)$ and $\ker(B) \subseteq \ker(AB)$.

i. Let $\mathbf{z} \in \text{im}(AB)$. We wish to show that $\mathbf{z} \in \text{im}(A)$.

Since $\mathbf{z} \in \text{im}(AB)$, $\exists \mathbf{y} \in \mathbb{R}^p$ such that $AB\mathbf{y} = \mathbf{z}$. We can then say $B\mathbf{y} \in \text{im}(A)$, as \exists a vector (namely, \mathbf{z}) where $A(B\mathbf{y}) = \mathbf{z}$. $B\mathbf{y} \in \text{im}(A)$, thus $\exists \mathbf{x} \in \mathbb{R}^m$ where $\mathbf{x} = B\mathbf{y}$. Therefore, $A\mathbf{x} = \mathbf{z}$, meaning $\text{im}(AB) \subseteq \text{im}(A)$. \square

Note that $\text{im}(A)$ and $\text{im}(AB)$ are equal when the kernel of B is empty, i.e. $\dim(B) = \text{rank}(B)$.

ii. Let $\mathbf{z} \in \ker(B)$. We wish to show $\mathbf{z} \in \ker(AB)$.

Since $\mathbf{z} \in \ker(B)$, $B\mathbf{z} = 0$. Applying the matrix A to both yields $AB\mathbf{z} = A0 = 0$, so $(AB)\mathbf{z} = 0$. Thus $\mathbf{z} \in \ker(AB)$, meaning $\ker(B) \subseteq \ker(AB)$. \square

Note that equality occurs when the kernel of A is empty, meaning no other vectors are mapped to the zero vector aside from the ones which were previously mapped by B .

(b) For $A, B \in \mathbb{R}^{n \times n}$, show that

i. $\text{rank}(AB) \leq \min \{\text{rank}(A), \text{rank}(B)\}$,

Note it is sufficient to show $\text{rank}(AB) \leq \text{rank}(A)$ and $\text{rank}(AB) \leq \text{rank}(B)$. The former follows directly from the first proof of the previous part, noting that $\text{im}(AB) \subseteq \text{im}(A)$ implies $\dim(\text{im}(AB)) \leq \dim(\text{im}(A))$. The latter follows from the second proof of the first part. Using the rank-nullity theorem, we can rewrite $\text{rank}(AB)$ as

$$\text{rank}(AB) = n - \text{null}(AB)$$

which from showing $\ker(B) \subseteq \ker(AB)$, implies that $\text{null}(AB) \geq \text{null}(B)$. Therefore,

$$\text{rank}(AB) \leq n - \text{null}(B)$$

which the left side can be rewritten as $\text{rank}(B)$. Thus $\text{rank}(AB) \leq \text{rank}(B)$, meaning since $\text{rank}(AB)$ is less than or equal to both ranks, it is less than the minimum of the two. \square

ii. $\text{null}(AB) \leq \text{null}(A) + \text{null}(B),$

To prove this we will work backwards,

$$\begin{aligned}
 & \text{null}(A) + \text{null}(B) && \text{(Given.)} \\
 &= n - \text{rank}(A) + n - \text{rank}(A) && \text{(Rank-nullity.)} \\
 &= n - \text{rank}(A) - \text{rank}(B) + \text{rank}(AB) + \text{null}(AB) && \text{(Rank-nullity.)} \\
 &\geq n - \text{rank}(A) - \text{rank}(AB) + \text{rank}(AB) + \text{null}(AB) && \text{(null}(AB) \geq \text{null}(B).) \\
 &= n - \text{rank}(A) + \text{null}(AB) && \text{(Simplifying.)} \\
 &= \text{null}(A) + \text{null}(AB) && \text{(Rank-nullity.)} \\
 &\geq \text{null}(AB) && \text{(0} \leq \text{null}(A).)
 \end{aligned}$$

This proves $\text{null}(AB) \leq \text{null}(A) + \text{null}(B).$ \square

iii. $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B).$

This is a direct implication of the Dimension theorem, i.e. for vector space V and sub-spaces U, W of V ,

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

Substituting $U = \text{im}(A)$, $W = \text{im}(B)$, and $V = \mathbb{R}^{n \times n}$,

$$\dim(\text{im}(A)) + \dim(\text{im}(B)) - \dim(\text{im}(A \cap B)) = \dim(\text{im}(A+B)).$$

Since the dimension of a subspace of a vector space is positive, we can throw away the intersecting term to yield,

$$\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B).$$

\square

(c) For $A, B \in \mathbb{R}^{n \times n}$, show that if $AB = 0$, then $\text{rank}(A) + \text{rank}(B) \leq n$.

Since $AB = 0$, for any vector $\mathbf{x} \in \mathbb{R}^n$, $AB\mathbf{x} = 0$. Thus $\text{null}(AB) = n$, and $\text{rank}(AB) = 0$. From the previous part, we can use the second inequality to prove this one.

$$\begin{aligned}
\text{null}(AB) &\leq \text{null}(A) + \text{null}(B) && \text{(Given.)} \\
n &\leq \text{null}(A) + \text{null}(B) && (\text{null}(AB) = n.) \\
n &\leq n - \text{rank}(A) + n - \text{rank}(B) && \text{(Rank-nullity.)} \\
-n &\leq -\text{rank}(A) - \text{rank}(B) && \text{(Simplifying.)} \\
n &\geq \text{rank}(A) + \text{rank}(B) && \text{(Multiplying by -1.)}
\end{aligned}$$

Thus the inequality is proven. \square

2. (a) Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$. Show that,

$$\text{rank}\left(\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\right) = \text{rank}(A) + \text{rank}(B)$$

Let X be the provided block matrix, for simplicity. Among the first m rows of X , suppose we can find up to l linearly independent vectors. Repeating this for the last p rows, suppose we can find up to r linearly independent vectors. Collect these row vectors into a subspace of X , noting that this subspace of X can represent all independent rows of A and B . Since this is a subspace of X , we can say $\text{rank}(X) \geq \text{rank}(A) + \text{rank}(B)$.

On the other hand, we can note that there X can have up to a max of $\text{rank}(A) + \text{rank}(B)$ linearly independent vectors, since if there were an additional linearly independent vector, it would have to note be in either A nor B , which is impossible. Thus $\text{rank}(X) \leq \text{rank}(A) + \text{rank}(B)$, which means they are equal. \square

(b) For $\mathbf{x} = [x_1, x_2, \dots, x_m]^T \in \mathbb{R}^m$ and $\mathbf{y} = [y_1, y_2, \dots, y_n]^T \in \mathbb{R}^n$, observe that $\mathbf{xy}^T \in \mathbb{R}^{m \times n}$. Let $A \in \mathbb{R}^{m \times n}$. Show that $\text{rank}(A) = 1$ iff $A = \mathbf{xy}^T$ for some nonzero $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$.

For the forward implication, we need to construct \mathbf{x} and \mathbf{y} from A , given $\text{rank}(A) = 1$. By this constriction, and by definition, we can only find 1 linearly independent column of A . Let \mathbf{y} be this column, and without loss of generality, choose the first column as the linearly independent column. Denote \mathbf{x} as the collection of multiples of \mathbf{y} needed to represent each column of A (i.e., $A_{*,j} = x_j \mathbf{y}$). We can note that $\mathbf{x} \in \mathbb{R}^m$, thus we can write $A = \mathbf{xy}^T$.

For the backward implication, suppose $\exists \mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$, and $A \in \mathbb{R}^{m \times n}$ with $A = \mathbf{xy}^T$. Expanding \mathbf{xy}^T ,

$$\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{pmatrix} \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix} = \begin{pmatrix} x_1 \mathbf{y} & x_2 \mathbf{y} & \dots & x_m \mathbf{y} \end{pmatrix}$$

This leads to A containing columns of multiples of \mathbf{y} , thus one can find only one linearly independent column of A . Therefore $\text{rank}(A) = 1$. \square

3. Let $A \in \mathbb{R}^{m \times n}$.

(a) Show that

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

Give an example to show this is not true over a finite field (e.g. a field of two elements $\mathbb{F}_2 = \{0, 1\}$ with binary arithmetic).

i. $\ker(A^T A) = \ker(A)$:

- Let $\mathbf{x} \in \ker(A)$. Thus $A\mathbf{x} = 0$, implying $A^T A\mathbf{x} = 0$. Therefore $\ker(A) \subseteq \ker(A^T A)$.
- Next let $\mathbf{y} \in \ker(A^T A)$. so $A^T A\mathbf{y} = 0$. By Associativity, $A^T(A\mathbf{y}) = 0$, thus $A\mathbf{y} \in \ker(A^T)$. Note that the kernel of the transpose of A is equal to its orthogonal image, i.e. $\ker(A^T) = \text{im}(A)^\perp$. Furthermore, since our vector is of the form $A\mathbf{z}$ for some $\mathbf{z} \in \mathbb{R}^n$, $A\mathbf{y} \in \text{im}(A)$. Thus $A\mathbf{y} \in \text{im}(A)^\perp \cap \text{im}(A) = \{0\}$. Then $A\mathbf{y} = 0$, so $\mathbf{y} \in \ker(A)$, completing the equality.

ii. $\text{im}(A^T A) = \text{im}(A^T)$:

- Let $\mathbf{y} \in \text{im}(A^T A)$. Then $\exists \mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{y} = A^T A\mathbf{x}$. By Associativity, $\mathbf{y} = A^T(A\mathbf{x})$. Since \mathbf{y} is of the form $A^T \mathbf{z}$ for some $\mathbf{z} \in \mathbb{R}^m$, $\mathbf{y} \in \text{im}(A^T)$.
- Next let $\mathbf{z} \in \text{im}(A^T)$. $\exists \mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{z} = A^T \mathbf{y}$. Consider $\mathbf{x} \in \text{im}(A)$. $\exists \mathbf{w} \in \mathbb{R}^n$ with $\mathbf{x} = A\mathbf{w}$. Choose \mathbf{y} such that $\mathbf{y} = A\mathbf{w} \in \text{im}(A)$. Then $\mathbf{z} = A^T A\mathbf{w}$, implying $\mathbf{z} \in \text{im}(A^T A)$ and completing the equality.

Giving an example where $\ker(A^T A)$ is not equal to $\ker(A)$, take A to be

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; \quad A^T = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Then,

$$A^T A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Calculating the kernels of each,

$$\ker(A) : \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix} \implies \ker(A) = \{(x, 0) : x \in \mathbb{R}\}$$

$$\ker(A^T A) : \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix} \implies \ker(A^T A) = \{(0, y) : y \in \mathbb{R}\}$$

Therefore, $\ker(A) \neq \ker(A^T A)$.

(b) Show that

$$A^T A \mathbf{x} = A^T \mathbf{b}$$

always has a solution (even if $A \mathbf{x} = \mathbf{b}$ has no solution). Give an example to show that this is not true over a finite field.

Since $\text{im}(A^T A) = \text{im}(A^T)$, any vector in the image of $\text{im}(A^T A)$ is in the image of $\text{im}(A^T)$. so if $A^T \mathbf{b} \in \text{im}(A^T)$, then $A^T \mathbf{b} \in \text{im}(A^T A)$, thus $\exists \mathbf{x} \in \mathbb{R}^n$ such that $A^T A \mathbf{x} = A^T \mathbf{b}$. Thus even if $A \mathbf{x} = \mathbf{b}$ has no solution, $A^T A \mathbf{x} = A^T \mathbf{b}$ will have a solution.

Over a finite field, we can choose

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Note this system has no solutions. Then, $A^T A = A^T = A$, keeping our original system with no solutions.

4. Let $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$. Let $r \leq n$ and $G_r = [g_{ij}] \in \mathbb{R}^{r \times r}$ be the matrix with

$$g_{ij} = \mathbf{v}_i^T \mathbf{v}_j$$

for $i, j = 1, \dots, r$. This is called a Gram matrix.

(a) Show that $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent iff $\text{null}(G_r) = 0$.

As this is an if and only if statement, we need to prove both the forward and backward directions.

For the forward direction, define matrix M such that its rows contain the corresponding transposed vector, i.e.,

$$M_{i,*} = \mathbf{v}_i^T.$$

Where $i = 1, \dots, r$. Note that $M \mathbf{v}_1$ represents the first column of G_r . The statement $\text{null}(G_r) = 0$ is equivalent to saying $\text{rank}(G_r) = r$, or that the columns of G_r are linearly independent. Thus, All we need to do is show that $M \mathbf{v}_1, \dots, M \mathbf{v}_r$ are linearly independent. This statement is fortunately equivalent to showing the $\text{rank}(M) = r$, which since $\mathbf{v}_1, \dots, \mathbf{v}_r$ are given as linearly independent, is guaranteed. Thus the forward direction has been proven.

The backward direction is quite simple to prove. Since $\text{null}(G_r) = 0$, this means that $\text{rank}(G_r) = r$, $r \leq n$. This is equivalent to saying its columns are linearly independent, or $M \mathbf{v}_1, \dots, M \mathbf{v}_n$ are linearly independent. We just now need to show that $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent. Take $c_1, \dots, c_n \in \mathbb{R}$. We just need to show that $c_1 \mathbf{v}_1 + \dots + c_n \mathbf{v}_n = 0$ iff $c_1 = \dots = c_n = 0$. Applying M to both sides and invoking the linearity of M gives

$$c_1 M \mathbf{v}_1 + \dots + c_n M \mathbf{v}_n = M 0 = 0$$

Since $M \mathbf{v}_1, \dots, M \mathbf{v}_n$ are linearly independent, this equation is zero iff $c_1 = \dots = c_n = 0$, thus proving $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent. This means the backward direction has been proven, and equality holds. \square

- (b) Show that $G_r = \mathbb{I}_r$ iff $\mathbf{v}_1, \dots, \mathbf{v}_r$ are pairwise orthogonal unit vectors, i.e., $\|\mathbf{v}_i\|_2 = 1$ for all $i = 1, \dots, r$ and $\mathbf{v}_i^T \mathbf{v}_j = 0$ for all $i \neq j$. If this holds, show that

$$\sum_{i=1}^r (\mathbf{v}^T \mathbf{v}_i)^2 \leq \|\mathbf{v}\|_2^2 \quad (4.1)$$

for all $\mathbf{v} \in \mathbb{R}^n$. What can you say about $\mathbf{v}_1, \dots, \mathbf{v}_r$ if equality always holds in (4.1) for all $\mathbf{v} \in \mathbb{R}^n$?

For the forward implication we have that $G_r = \mathbb{I}_r$. Note that we can express the identity matrix as δ_{ij} , where δ_{ij} is the Kronecker-delta symbol, for $i, j = 1, \dots, r$. Thus $g_{ij} = \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$ for $i, j = 1, \dots, r$. Note additionally that $\mathbf{v}_i^T \mathbf{v}_j$ is simply the inner product of the two vectors $\mathbf{v}_i, \mathbf{v}_j$. So for $i = j$, $g_{ii} = \delta_{ii} = 1 = \langle \mathbf{v}_i | \mathbf{v}_i \rangle = \|\mathbf{v}_i\|_2^2$. So each vector has length 1. for $i \neq j$, $g_{ij} = \delta_{ij} = 0 = \langle \mathbf{v}_i | \mathbf{v}_j \rangle$. Thus $\mathbf{v}_1, \dots, \mathbf{v}_r$ are orthogonal with unit length, implying $G_r = \mathbb{I}_r$.

For the backward implication we have $\mathbf{v}_1, \dots, \mathbf{v}_r$ are pairwise orthogonal unit vectors. We then want to show $G_r = \mathbb{I}_r$. Note that $g_{ij} = \mathbf{v}_i^T \mathbf{v}_j = \langle \mathbf{v}_i | \mathbf{v}_j \rangle$. Since $\mathbf{v}_1, \dots, \mathbf{v}_r$ are pairwise orthogonal, $\langle \mathbf{v}_i | \mathbf{v}_j \rangle = \|\mathbf{v}_i\|_2^2 \delta_{ij}$, and since $\mathbf{v}_1, \dots, \mathbf{v}_r$ are unit vectors, $\|\mathbf{v}_i\|_2^2 = 1$ for all $i = 1, \dots, r$. Therefore $g_{ij} = \langle \mathbf{v}_i | \mathbf{v}_j \rangle = \delta_{ij} = \mathbb{I}_r$ for $i, j = 1, \dots, r$. This proves the backward implication, proving the statement. \square

To prove (4.1), we can expand out $\|\mathbf{v}\|_2^2$ and compare summations.

$$\|\mathbf{v}\|_2^2 = \sum_{i=1}^n \mathbf{v}_i^2 = \sum_{i=1}^n (e_i^T \mathbf{v})^2$$

Here we're just taking the inner product of the vector \mathbf{v} with the standard basis vector e_i , $i = 1, \dots, n$. Since $\mathbf{v}_1, \dots, \mathbf{v}_r$ are assumed to be pairwise orthogonal unit vectors, these \mathbf{v}_i 's can be swapped with r standard unit vectors (or there exists some mapping from the \mathbf{v}_i 's to the standard unit vectors).

$$\implies \sum_{i=1}^r (\mathbf{v}^T \mathbf{v}_i)^2 = \sum_{i=1}^r (\mathbf{v}^T e_i)$$

Since $r \leq n$, the summation running to r will always be less than or equal to the summation running to n , with equality holding only when $r = n$.

5. Let $A, B \in \mathbb{R}^{m \times n}$. For any $i, j \in \{1, \dots, n\}$, let $E_{ij} \in \mathbb{R}^{p \times q}$ denote the matrix with one in the (i, j) -th entry and zeros everywhere else.

- (a) Describe the matrix $E_{ij} A E_{kl} \in \mathbb{R}^{m \times n}$ where $E_{ij} \in \mathbb{R}^{p \times m}$ and $E_{kl} \in \mathbb{R}^{n \times q}$.

Although having the same end matrix, which E-matrix you choose to apply to A first will result in a different intermediate matrix. Via associativity, I will break this down into two cases.

i. Case 1: $(E_{ij} A) E_{kl}$

This results in an intermediate matrix of size p by n , where the j -th row of A is extracted and emplaced as the i -th row of the resulting matrix. I will provide an example. Say, A is 2 by 3, and $i =$

1, $j = 2$. The resulting intermediate product is then,

$$E_{12}A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 4 & 5 & 6 \\ 0 & 0 & 0 \end{pmatrix}$$

ii. Case 2: $E_{ij}(AE_{kl})$

This results in an intermediate matrix of size m by q , where the k -th column of A is extracted and emplaced as the l -th column of the resulting matrix. I'll keep with the example given in the previous case, where $k = 1$ and $l = 3$.

$$AE_{13} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 4 \end{pmatrix}$$

Where the final matrix is

$$E_{12}AE_{13} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 0 \end{pmatrix}$$

As you can see, Applying E_{12} to our intermediate matrix places the second row into the first row of the final matrix. Note the final matrix is of size p by q .

Therefore, E_{ij} extracts the j -th row of A and places it into the i -th row, and E_{kl} places the k -th column into the l -th column. As shown above, this results in different intermediate matrices, but the final matrix is guaranteed to be the same due to Associativity.

- (b) Find expressions $\text{tr}(A^T B)$, $\text{tr}(AB^T)$, $\text{tr}(BA^T)$, $\text{tr}(B^T A)$ in terms of entries of A and B . What if $A = B$?
What if $B = E_{ij} \in \mathbb{R}^{m \times n}$?

Note that for any matrix $C \in \mathbb{R}^{m \times n}$, $\text{tr}(C) = \text{tr}(C^T)$. This is due to the diagonal of C not being disrupted when transposing. Due to this, it suffices to find expressions $\text{tr}(A^T B)$ and $\text{tr}(AB^T)$.

Taking the standard formula for matrix multiplication,

$$(AB)_{ij} = \sum_{k=1}^m a_{ik} b_{kj}$$

setting $j = i$ and summing over i will give us a formula for the trace of the product of two matrices.

$$\implies \text{tr}(AB) = \sum_{i=1}^m \sum_{k=1}^m a_{ik} b_{ki}$$

so all we need to do is swap indices to suit our input matrices.

$$\text{tr}(A^T B) = \sum_{i=1}^m \sum_{k=1}^m a_{ki} b_{ki}$$

$$\text{tr}(AB^T) = \sum_{i=1}^m \sum_{k=1}^m a_{ik} b_{ik}$$

So if $A = B$, we can write

$$\text{tr}(A^T A) = \sum_{i=1}^m \sum_{k=1}^m a_{ki}^2$$

$$\text{tr}(AA^T) = \sum_{i=1}^m \sum_{k=1}^m a_{ik}^2$$

Note that by this calculation, $\text{tr}(AB^T) = \text{tr}(A^T B)$, up to a relabelling of indices. If $B = E_{ij} \in \mathbb{R}^{m \times n}$, it breaks the formula I've given, since B did not depend on indices. Instead, it will have to be evaluated separately. Reminder that E on the right of A swaps columns, E on the left swaps rows.

For the case of $\text{tr}(A^T B)$, the resulting matrix will have one non-zero column, where the i -th column of A is represented as the j -th column. Since A is being transposed, the i -th row is being represented as the j -th column. This results in the element in the i -th row, j -th column being the trace. So $\text{tr}(AE_{ij}) = a_{ij}$.

For the case of $\text{tr}(AB^T)$, the resulting matrix will have one non-zero column, where the j -th column of A is represented as the i -th column. These are swapped since E is transposed ($E_{ij} \rightarrow E_{ji}$). This means the element on the diagonal is a_{ii} , therefore $\text{tr}(AE_{ji}) = a_{ii}$.

(c) Suppose $m = n$ and $\text{tr}(A) = \text{tr}(A^2) = \dots = \text{tr}(A^n) = 0$. Show that $A^n = 0$.

Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A , with associated eigenvector \mathbf{v} . Note that $A^2 \mathbf{v} = A(A\mathbf{v}) = A\lambda \mathbf{v} = \lambda^2 \mathbf{v}$, thus \mathbf{v} is an eigenvector of A^k with eigenvalue λ^k . We are given $\text{tr}(A) = \text{tr}(A^2) = \dots = \text{tr}(A^n) = 0$. Since the trace of a matrix is the sum of its eigenvalues, we can write the following system of equations:

$$\lambda_1 + \dots + \lambda_n = 0$$

$$\lambda_1^2 + \dots + \lambda_n^2 = 0$$

...

$$\lambda_1^n + \dots + \lambda_n^n = 0$$

This system of equations is called the symmetric powers of A . Since these are all equal 0, this will help tremendously in the next step.

Writing out the characteristic polynomial of A , we get,

$$\chi_A(x) = (x - \lambda_1)(x - \lambda_2)\dots(x - \lambda_n)$$

which, when expanded, is of the form,

$$x^n - \left(\sum_i \lambda_i \right) x^{n-1} + \left(\sum_{i < j \leq n} \lambda_i \lambda_j \right) x^{n-2} + \dots + (-1)^n e_n(\lambda_1, \dots, \lambda_n),$$

Where e_i is an Elementary symmetric polynomial, defined as the sum of all distinct products of n distinct variables. Note that a property of elementary symmetric polynomials is they can be expressed as sums and products of the symmetric powers stated above. Since the symmetric powers are all equivalent to zero, this means each Elementary symmetric polynomial is a sum and product of zero, thus all coefficients of our characteristic polynomial equal zero. This leaves us with,

$$\chi_A(x) = x^n.$$

By The Cayley-Hamilton theorem, the matrix A satisfies its characteristic polynomial, that is $\chi_A(A) = 0$. Thus $A^n = 0$, which proves the statement. \square

(d) Suppose $m = n$. Show that

$$\text{i. } \text{tr}[(AB)^k] = \text{tr}[(BA)^k].$$

This can be done via induction on k .

Base Case: $k = 1$

Then,

$$(AB)_{ii} = \sum_{k=1}^n a_{ik} b_{ki}, \quad (BA)_{ii} = \sum_{k=1}^n b_{ik} a_{ki}$$

These summations are equivalent, up to relabelling of indices. Thus,

$$\text{tr}[(AB)] = \sum_{i=1}^n (AB)_{ii} = \sum_{k=1}^n a_{ki} b_{ik}, \quad \text{tr}[(BA)] = \sum_{i=1}^n (BA)_{ii} = \sum_{k=1}^n b_{ik} a_{ki}$$

$$\implies \text{tr}[(AB)] = \text{tr}[(BA)]$$

Thus the base case has been proven. Next, assume the claim holds for some $l < k$, we will show the $l + 1$ case holds. I.e., we wish to show

$$\text{tr}[(AB)^{l+1}] = \text{tr}[(BA)^{l+1}]$$

$$\begin{aligned}
\text{tr}[(AB)^{l+1}] &= \text{tr}[(AB)^l(AB)] && \text{(Matrix Exponent.)} \\
&= \text{tr}[(BA)^l(AB)] && \text{(Inductive Hypothesis.)} \\
&= \text{tr}[(BA)^l(AB)^T] && \text{(From Problem 5, part b.)} \\
&= \text{tr}[(BA)^l(B^T A^T)] && \text{(Matrix Transpose Property.)} \\
&= \text{tr}[(BA)^l(BA)] && \text{(Problem 5, part b applied to A and B.)} \\
&= \text{tr}[(BA)^{l+1}] && \text{(Matrix Exponent.)}
\end{aligned}$$

Therefore, the statement holds by induction. \square

ii. and if $AB = 0$, Then $\text{tr}[(A + B)^k] = \text{tr}(A^k) + \text{tr}(B^k)$, for all $k \in \mathbb{N}$

Since in the previous part we showed that matrices commute under the trace operator, we are free to use binomial expansion on $(A + B)^k$.

$$\implies \text{tr}[(A + B)^k] = \text{tr}\left[\sum_{j=0}^k \binom{k}{j} A^{k-j} B^j\right].$$

From this expansion, we can take out the two cases where $j = 0$ and $j = k$ to get our leading terms.

$$\text{tr}\left[\sum_{j=0}^k \binom{k}{j} A^{k-j} B^j\right] = \text{tr}\left[A^k + B^k + \sum_{j=1}^{k-1} \binom{k}{j} A^{k-j} B^j\right].$$

We can note that under the summation, each term will have an AB product in it, thus making the entire term 0. We can discard the summation in that case.

$$\text{tr}[(A + B)^k] = \text{tr}[A^k + B^k].$$

And finally, since the trace function is a linear operator, we can break this apart into two pieces.

$$\implies \text{tr}[(A + B)^k] = \text{tr}(A^k) + \text{tr}(B^k),$$

thus proving the statement. \square

6. Let $A \in \mathbb{C}^{n \times n}$. Recall that A is diagonalizable iff there exists an invertible $X \in \mathbb{C}^{n \times n}$ such that $X^{-1}AX = \Lambda$, a diagonal matrix.

(a) Show that A is diagonalizable if and only if its minimal polynomial is of the form

$$m_A(x) = (x - \lambda_1) \cdots (x - \lambda_d)$$

where $\lambda_1, \dots, \lambda_d \in \mathbb{C}$ are all distinct. Hence deduce for a diagonalizable matrix, the degree of its polynomial equals the number of distinct eigenvalues.

Since any square matrix can be written in Jordan Canonical Form, i.e.

$$A = XJX^{-1}.$$

Where J is composed of block matrices along its diagonal of the form $J_k = \lambda_k I + N$, where N is a nilpotent matrix of nilpotency p and λ_k is the k -th eigenvalue of A . There are a few things to note about the Jordan Canonical form of a matrix:

- i. The number of total blocks inside J is equal to the number of distinct eigenvectors, thus each Jordan block has one eigenvector.
- ii. For the minimum polynomial μ_A of A , we need only investigate J itself. That is, $\mu_A(A) = X\mu_A(J)X^{-1}$.
- iii. The size of a Jordan block J_k depends on how many eigenvectors are associated with the eigenvalue λ_k . That is, the eigenspace of λ_k has size equal to the size of J_k .

For the backward implication, we suppose the minimal polynomial is of the form

$$\mu_A(x) = (x - \lambda_1)(\dots)(x - \lambda_d)$$

where $\lambda_1, \dots, \lambda_d$ are the only distinct eigenvalues of A . By the Cayley-Hamilton Theorem, $\mu_A(A) = 0$, i.e.

$$\mu_A(A) = (A - \lambda_1 I)(A - \lambda_2 I)(\dots)(A - \lambda_d I) = 0$$

since $\mu_A(A) = 0$, then A must zero out all non-diagonal terms of J , leaving J with entries only along the diagonal. Thus A is diagonalizable.

For the forward implication, we assume A is diagonalizable. It's Jordan Canonical form is thus a diagonal matrix. Its characteristic polynomial is thus

$$\chi_A(x) = (x - \lambda_1)^{\alpha_1}(\dots)(x - \lambda_d)^{\alpha_d}$$

where $\alpha_1, \dots, \alpha_d > 1$ are the algebraic multiplicities of the eigenvalue λ_i .

Claim: $\mu_A(x) = (x - \lambda_1)(\dots)(x - \lambda_d)$, i.e. $\alpha_1 = \dots = \alpha_i = 1$.

Suppose false, meaning $\exists g < d$ such that $\mu_A(x) = (x - \lambda_1)(\dots)(x - \lambda_g)$. Consider the Jordan block associated with the i -th eigenvalue, J_i with $g < i \leq d$. Since $\lambda_1, \dots, \lambda_d$ are distinct eigenvalues, λ_k cannot zero out the diagonal values of the Jordan block J_i for $k \leq g$, thus making this impossible. Therefore,

$\mu_A(x) = (x - \lambda_1)(\dots)(x - \lambda_d)$. Now we just need to show $\alpha_1 = \dots = \alpha_i = 1$. This is simple, since

$$\chi_A(J) = (J_1 - \lambda_1 I)^{\alpha_1}(\dots)(J_d - \lambda_d I)^{\alpha_d} = 0.$$

Since $J_k = \lambda_k + N$, $J_k - \lambda_k I = N$. And since A is diagonalizable, $N = 0$, meaning A has no off-diagonal terms in its Jordan form. Thus $\alpha_1 = \dots = \alpha_i = 1$, finishing the proof. \square

(b) Let A be diagonalizable. Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{C}^n$ be n linearly independent right eigenvectors of A , and $\mathbf{y}_1, \dots, \mathbf{y}_n \in \mathbb{C}^n$ be n linearly independent left eigenvectors of A . Show that we may choose $\mathbf{x}_1, \dots, \mathbf{x}_n$ and $\mathbf{y}_1, \dots, \mathbf{y}_n$ so that any vector $\mathbf{v} \in \mathbb{C}^n$ can be expressed as

$$\mathbf{v} = \sum_{i=1}^n (\mathbf{y}_i^T \mathbf{v}) \mathbf{x}_i.$$

If we write $X = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{C}^{n \times n}$ and $Y = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{C}^{n \times n}$, what is the relation between X and Y ?

By some manipulation of the summation,

$$\begin{aligned} \mathbf{v} &= \sum_{i=1}^n (\mathbf{y}_i^T \mathbf{v}) \mathbf{x}_i && \text{(Given.)} \\ &= \sum_{i=1}^n \mathbf{x}_i (\mathbf{y}_i^T \mathbf{v}) && \text{(Vector-scalar commutation.)} \\ &= \sum_{i=1}^n \mathbf{x}_i \mathbf{y}_i^T \mathbf{v} && \text{(Associativity.)} \\ &= \sum_{i=1}^n (\mathbf{x}_i \mathbf{y}_i^T) \mathbf{v} && \text{(Associativity.)} \\ &= \left[\sum_{i=1}^n (\mathbf{x}_i \mathbf{y}_i^T) \right] \mathbf{v} && \text{(Distributive Law.)} \end{aligned}$$

Thus we can note that

$$\sum_{i=1}^n (\mathbf{x}_i \mathbf{y}_i^T) = I.$$

This should hold for any choice of linearly independent eigenvectors, provided we are in the basis of the eigenvectors of A . If we were to write $X = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{C}^{n \times n}$ and $Y = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{C}^{n \times n}$, then we can interpret the above summation as multiplying the i -th column of X with the i -th row of Y and summing over all results. This is one of the ways to interpret Matrix Multiplication, therefore: $XY = I$. Since X and Y are written to have linearly independent columns, both are invertible, thus $X = Y^{-1}$ and vice versa.