Exam 1

• This exam covers Topics 1 - 3, Topic 4 will not be covered here.

• I will write $(a_1, a_2, \ldots, a_n)$ in place of $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ to save space on occasion. The book writes $\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T$ for the same purpose.

• I use $\mathrm{NS}(A)$ for the null space of A, $\mathrm{RS}(A)$ for the row space of A, $\mathrm{CS}(A)$ for the column space of A. Note that $\mathrm{CS}(A) = \mathrm{rng}(A)$ is the range of A.

• When I say you can use some fact from another part of the exam, this means that you can use the fact whether or not you have completed that part of the exam correctly.

• If you have questions ask via Remind or email.

• Use only arguments that you fully understand. I am aware that you can find solutions to some of these online. I am also aware that some of these solutions use concepts and theorems far past what we have covered in Ch 1 - 3. If you use such ideas, then I will ask you to verbally explain your solution so that I can verify that you understand what you have submitted as your own work. In short, **this is an exam** and the usual expectations of academic honesty apply.

Part I: True/False (5 points each; 25 points)

For each of the following mark as true or false.

a) _____ $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for an $n \times n$ matrices A and B, where

$$\operatorname{tr}(C) = \sum_{i=1}^{n} C_{ii} = \text{the sum of the diagonal elements of } C.$$

This is true. This is just a computation. $(AB)_{ii} = \sum_{k=1}^{n} A_{ik}B_{ki}$, so

$$\operatorname{tr}(AB) = \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik}B_{ki}$$

and

$$\operatorname{tr}(BA) = \sum_{i=1}^{n} \sum_{k=1}^{n} B_{ik}A_{ki} = \sum_{k=1}^{n} A_{ki}B_{ik} = \operatorname{tr}(AB).$$

b) _____ $\operatorname{tr}(AB) = \operatorname{tr}(A)\operatorname{tr}(B)$ for an $n \times n$ matrices A, B, and C.

Interestingly, this is false, as an example can show. In fact, generating any three random 2×2 matrices with entries from $\{-1, 0, 1\}$ are likely to work. Try this using MATLAB: round(rand(2)). The first three matrices I got this way were:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \quad AB = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

So, $\operatorname{tr}(AB) = -1$ while $\operatorname{tr}(A)\operatorname{tr}(B) = (-1)(0) = 0$.

c) _____ If W is a subspace of a vector space V and $\mathcal{B}$ is a basis for V, then B can be restricted to a basis for W.

This is false. Let $W = \operatorname{span}\{(1, 1)\} \subseteq \mathbb{R}^2 = V$. The standard basis for $\mathbb{R}^2$ can not be restricted to a basis for W.

d) _____ If W is a subspace of a vector space V, then there is a subspace U so that $V = W \oplus U$.

This notation is a little hard to find in your text: $V = U + W$ means that for all $\boldsymbol{v} \in V$, there is $\boldsymbol{u} \in U$ and $\boldsymbol{w} \in W$ so that $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}$. $V = U \oplus V$ means $V = U + V$ **and** $U \cap W = \{\mathbf{0}\}$, equivalently, for every $\boldsymbol{v} \in V$, there is a **unique** $\boldsymbol{u} \in U$ and $\boldsymbol{w} \in W$ so that $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}$.

This is true. Let $\mathcal{B}_W$ be a basis for W and extend $\mathcal{B}_W$ to $\mathcal{B}_V$ a basis for V. Then let $U = \operatorname{span}(\mathcal{B}_V - \cap B_W)$. It is clear that $V = W \oplus U$.

e) _____ For any $m \times n$ matrices A and B,

$$B = EA \text{ for some invertible } E \iff \operatorname{NS}(A) = \operatorname{NS}(B).$$

This is true. ($\Rightarrow$) is trivial, since if $B = EA$, then $B\boldsymbol{x} = \mathbf{0} \iff EA\boldsymbol{x} = \mathbf{0} \iff A\boldsymbol{x} = E^{-1}\mathbf{0} = 0$. ($\Leftarrow$) is discussed below in the "Proofs" section.

Part II: Definitions and Theorems (5 points each; 25 points)

a) Define what it means for a set of vectors $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ from a real vector space V to span V.

$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ **spans** V iff for all $\boldsymbol{v} \in V$, $\boldsymbol{v}$ is a linear combination of the vectors in $\mathcal{B}$, that is $\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$ for some coefficients $\alpha_i \in \mathbb{R}$.

b) Define what it means for a set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ from a real vector space V to be linearly independent.

A set of vectors $\mathcal{B}$ is **linearly independent** iff $\sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i = \mathbf{0}$, then $\alpha_i = 0$ for all i. Equivalently, any linear combination of the vectors that gives $\mathbf{0}$ must be trivial.

There are equivalent definitions, but essentially, this is what you need. You can say something like, for all i, $\boldsymbol{v}_i \notin \operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_{i+1}, \ldots, \boldsymbol{v}_n\}$

c) Define what it means for a set of vectors $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ to be a basis for a vector space V.

$\mathcal{B}$ has must be a linearly independent and span V.

d) State the Rank-Nullity Theorem.

If A is an $m \times n$ matrix, then $n = \dim(\mathrm{RS}(A)) + \dim(\mathrm{NS}(A)) = \operatorname{rank}(A) + \operatorname{nullity}(A)$.

e) What conditions must be checked to verify that $W \subseteq V$ is a subspace of a vector space. V

Closure under addition and scalar multiplication must be checked.

Part III: Computational (15 points each; 45 point)

a) Given that A is a 3×4 matrix and

$$\mathrm{NS}(A) = \mathrm{span}\left(\left\{\begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}\right\}\right)$$

compute $\mathrm{rref}(A)$. Make sure to explain how you arrive at your result. You may use (a) from the "Proofs" part below.

Notation: $\mathrm{rref}(A)$ means the reduced row echelon form of A. This is unique, a general echelon form is not unique. From $\mathrm{rref}(A)$ there is a simple way to read off a basis for $\mathrm{NS}(A)$, this exercise asks you to reverse that process.

We know $A\boldsymbol{x} = \mathbf{0} \iff \mathrm{rref}(A)\boldsymbol{x} = \mathbf{0} \iff \boldsymbol{x} \in \mathrm{NS}(A)$. From what we are given we see $\boldsymbol{x} \in \mathrm{NS}(A)$ iff

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = r \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} r + 2s \\ -2r + 3s \\ r \\ s \end{bmatrix}$$

Working backwards from what we usually do we see $x_4 = s, x_3 = r$, and so $x_2 = -2x_3 + 3x_4$ and $x_1 = x_3 + 2x_4$. This gives the system

$$\begin{aligned} x_1 - x_3 - 2x_4 &= 0 \\ x_2 + 2x_3 - 3x_4 &= 0 \end{aligned}$$

This corresponds to $B\boldsymbol{x} = \mathbf{0}$ for

$$B = \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

B is rref and so by Proofs (a) $B = \mathrm{rref}(A)$.

b) For the same (unknown) A used in (a) for each of $\mathrm{RS}(A)$ and $\mathrm{CS}(A)$ find a basis if possible and explain how you know that you have found a basis; if it is not possible to find a basis, then explain why it is not.

$\mathrm{RS}(A)$: Here we know $\mathrm{RS}(A) = \mathrm{RS}(\mathrm{rref}(A))$ so the non-zero rows of $\mathrm{rref}(A)$ form a basis for $\mathrm{RS}(A)$.

$\mathrm{CS}(A)$: You know that the first two columns of $\mathrm{rref}(A)$ are where the pivots are and so the first two columns of A would be a basis for $\mathrm{CS}(A)$, but you have no way of finding these and you know nothing about $\mathrm{CS}(A)$ other than $\dim(\mathrm{CS}(A)) = 2$. For example, $\mathrm{rref}(A_1) = \mathrm{rref}(A_2) = B$ for

$$A_1 = \begin{bmatrix} 1 & 1 & 1 & -5 \\ 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & -3 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 0 & 1 & -1 & -2 \\ 1 & 0 & -1 & -2 \\ 1 & 2 & 3 & -8 \end{bmatrix},$$

but

$$\mathrm{CS}(A_1) = \mathrm{span}\left(\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}\right\}\right) \text{ and } \mathrm{CS}(A_2) = \mathrm{span}\left(\left\{\begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\2\end{bmatrix}\right\}\right),$$

and $\begin{bmatrix}1\\0\\2\end{bmatrix} \notin \mathrm{CS}(A_1)$ so $\mathrm{CS}(A_1) \neq \mathrm{CS}(A_2)$.

c) Show that the upper-triangular $n \times n$ matrices form a subspace of all $n \times n$ matrices and find a basis for this subspace.

Let U be the collection of upper-triangular $n \times n$ matrices, that is $A \in U \iff A(i,j) = 0$ for $n \geq i > j \geq 1$.

It is clear that $A + B \in U$ and $\alpha A \in U$ for any $A, B \in U$. Let $A, B \in U$:

$$(A + B)(i,j) = A(i,j) + B(i,j) = 0 + 0 = 0 \text{ for } n \geq i > j \geq 1$$

$$(\alpha A)(i,j) = \alpha A(i,j) = 0 \text{ for } n \geq i > j \geq 1$$

A basis for U is $\{E_{i,j} \mid 1 \leq i \leq j \leq n\}$ where $E_{i,j}(l,m) = \delta_{(i,j),(l,m)}$. It is clear that if $A \in U$, then $A = \sum_{1 \leq i \leq j \leq n} A(i,j) E_{i,j}$.

Note: $\delta_{a,b} = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases}$.

Part IV: Proofs (15 points each; 60 points)

Provide complete arguments/proofs for the following.

a) Show that if A and B are 3×4 rref matrices, then

$$\boxed{A = B \iff \mathrm{NS}(A) = \mathrm{NS}(B).}$$

Note: the 3×4 is a red-herring, this holds for arbitrary $m \times n$ matrices. If this helps you , then just prove this more general result. Also notice that this actually gives

$$\mathrm{rref}(A) = \mathrm{rref}(B) \iff \mathrm{NS}(A) = \mathrm{NS}(B)$$

since $\mathrm{NS}(\mathrm{rref}(A)) = \mathrm{NS}(A)$, trivially.

($\Rightarrow$) is trivial.

For ($\Leftarrow$) there are at least two proof that I would accept.

Proof 1: Using the method of Part III (a) given a basis $\mathcal{B}$ for $\mathrm{NS}(A)$, row operations can be used to get a basis in one of the following forms, depending on the dimension of $\mathrm{NS}(A)$:

$$\left\{\begin{bmatrix} a \\ b \\ c \\ 1 \end{bmatrix}\right\} \qquad \left\{\begin{bmatrix} a_1 \\ b_1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} a_2 \\ b_2 \\ 0 \\ 1 \end{bmatrix}\right\} \qquad \left\{\begin{bmatrix} a_1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} a_2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} a_3 \\ 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

Clearly this basis is unique. For example, suppose

$$\mathrm{span}\left\{\begin{bmatrix} a_1 \\ b_1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} a_2 \\ b_2 \\ 0 \\ 1 \end{bmatrix}\right\} = \mathrm{span}\left\{\begin{bmatrix} a_1' \\ b_1' \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} a_2' \\ b_2' \\ 0 \\ 1 \end{bmatrix}\right\}$$

then $(a_1', b_1', 1, 0) \in \mathrm{span}\{(a_1, b_1, 1, 0), (a_1', b_1', 0, 1)\}$. This means

$$(a_1', b_1', 1, 0) = \alpha(a_1, b_1, 1, 0) + \beta(a_1', b_1', 0, 1)$$

for some scalars α and β, but then clearly $\alpha = 1$ and $\beta = 0$ so that $(a_1', b_1', 1, 0) = (a_1, b_1, 1, 0)$. Similarly for the other vector.

This is exactly the basis (by the uniqueness just proved) that you get from $\mathrm{rref}(A)$ and we know how to reconstruct $\mathrm{rref}(A)$ from this basis. So if $\mathrm{NS}(A) = \mathrm{NS}(B)$, then we would reconstruct the same $\mathrm{rref}(A)$ and $\mathrm{rref}(B)$, hence $\mathrm{rref}(A) = \mathrm{rref}(B)$.

Proof 2: We prove the contrapositive, namely, suppose $A \neq B$, then $\mathrm{NS}(A) \neq \mathrm{NS}(B)$. To start recall $\mathrm{RS}(A) = \mathrm{NS}(A)^{\perp} = \mathrm{NS}(B)^{\perp} = \mathrm{RS}(B)$. We also know that if A is in rref form, then the non-zero rows of A are a basis for $\mathrm{RS}(A)$ and hence for $\mathrm{RS}(B)$.

Suppose $A \neq B$ and let i be the first row on which they differ. Let $r = \text{rank}(A)$, this is $\dim(A)$ and as A is rref we know this is the number of non-zero rows of A.

Claim 1: The pivots (leading 1's) occur at the same places in A and B.

If this fails, let i be the first row where $R_i(A) \neq R_i(B)$, where $R_i(A) = i^{\text{th}}$ row of A. We may assume the first 1 of $R_i(A)$ occurs before the first 1 of $R_i(B)$. (Else just swap the roles of A and B.) Let k be the position of the pivot (leading 1) in $R_i(A)$. We know $R_i(A) = \sum_{i=1}^{r} c_j R_j(B)$, then clearly $c_j = 0$ for $j = 1, \ldots, i-1$. This is because all entries up to $k-1$ are 0's in $R_i(A)$ while all of $R_j(B)$ for $j < k$ have a leading 1 before k.

So $R_i(A) = \sum_{j=k}^{r} c_j R_j(B)$, but for $j \geq k$ we know the first k entries of $R_j(B)$ are 0's so it is impossible to get a 1 in the k^{th} position.

Claim 2: $R_i(A) = R_i(B)$ for all $i \leq r$.

This claim is actually trivial given the first. We know $R_i(A) = \sum_{j=1}^{r} c_j R_j(B)$, but $R_i(A)$ has 0 at all pivot places except for that of $R_i(B)$, so we must have $R_i(A) = c_i R_i(B)$. Since the leading non-zero element is 1 we have $C_i = 1$, so $R_i(A) = R_i(B)$.

b) **Prove:** For $m \times n$ matrices A and B define $A \underset{\text{row}}{\sim} B$ to mean that you can get from A to B by a series of elementary row operations. Use the $m \times n$ version of (a), namely: $\text{rref}(A) = \text{rref}(B) \iff \text{NS}(A) = \text{NS}(B)$ to show that

$$\boxed{A \underset{\text{row}}{\sim} B \iff \text{rref}(A) = \text{rref}(B)}$$

Remark: Using elementary matrices one can show

$$A \underset{\text{row}}{\sim} B \iff A = EB \text{ for some invertible matrix } E$$

This is done in the text and in my notes. So you get here

$$A \underset{\text{row}}{\sim} B \iff \text{rref}(A) = \text{rref}(B) \iff A = EB \text{ for some invertible matrix } E$$

($\Rightarrow$) (**Proof 1**) Clearly if $A \underset{\text{row}}{\sim} B$, then $\text{NS}(A) = \text{NS}(B)$ since elementary row ops generate equivalent systems of equations. (This was the whole point of row operations in the first place!) Similarly, $\text{NS}(A) = \text{NS}(\text{rref}(A))$ and $\text{NS}(B) = \text{NS}(\text{rref}(B))$ so using (a) above

$$\text{NS}(\text{rref}(A)) = \text{NS}(\text{rref}(B)) \implies \text{rref}(A) = \text{rref}(B)$$

as desired.

($\Rightarrow$) (**Proof 2**) You may use that $\text{rref}(A)$ is unique. This is something you have been given. Thus pass from A to B to $\text{rref}(B)$ through a series of row ops, then you know $\text{rref}(B)$ is an RREF form of A and by uniqueness $\text{rref}(A) = \text{rref}(B)$.

Note: If you use proof 1, then you actually prove the uniqueness of rref(A) here. This is one of the standard proofs.

($\Leftarrow$). This is really trivial. Do a series of operations to get from A to rref(A) and from B to rref(B), then just reverse the series from B to rref(B) = rref(A) to get back to B. Combining these you get a series of row ops that goes from A to rref(A) and then from rref(A) = rref(B) back to B.

c) **Prove:** Let A be an $m \times n$ matrix, $\mathbb{R}^n = \mathrm{NS}(A) \oplus \mathrm{RS}(A)$.

Recall: $V = U \oplus W$ means $V = U + W = \{\boldsymbol{u} + \boldsymbol{w} \mid \boldsymbol{u} \in U \text{ and } \boldsymbol{w} \in W\}$ and $U \cap W = \{\mathbf{0}\}$.

By the rank-nullity theorem if $\mathcal{B}$ a basis for RS(A) and $\mathcal{C}$ a basis for NS(A), then $\mathcal{B} \cup \mathcal{C}$ has size n.

If we can show that $\mathcal{B} \cup \mathcal{C}$ is linearly independent, then we have that $\mathcal{B} \cup \mathcal{C}$ is a basis for $\mathbb{R}^n$ and so $\mathrm{RS}(A) \oplus \mathrm{NS}(A) = \mathbb{R}^n$.

We just need to see that $\mathrm{RS}(A) \cap \mathrm{NS}(A) = \{\mathbf{0}\}$.

Proof 1: Suppose $\boldsymbol{v} \in \mathrm{NS}(A) \cap \mathrm{RS}(A)$ and $\boldsymbol{v} \neq \mathbf{0}$, then $A\boldsymbol{v} = \mathbf{0}$. Say the i^{th} row of A is $\boldsymbol{u}_i^T$, that is

$$A = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix}$$

Since $\boldsymbol{v} \in \mathrm{RS}(A)$, $\boldsymbol{v} = c_1\boldsymbol{u}_1 + \cdots + c_m\boldsymbol{u}_m$ and since $\boldsymbol{v} \in \mathrm{NS}(A)$

$$A\boldsymbol{v} = c_1 A\boldsymbol{u}_1 + \cdots + c_m A\boldsymbol{u}_m = \mathbf{0}$$

Consider

$$A\boldsymbol{u}_i = \begin{bmatrix} \boldsymbol{u}_1^T\boldsymbol{u}_i \\ \boldsymbol{u}_2^T\boldsymbol{u}_i \\ \vdots \\ \boldsymbol{u}_i^T\boldsymbol{u}_i \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

But $u_i^T u_i = \sum_{j=1}^n u_{ij}^2 = 0$ means that $\boldsymbol{u}_i = \mathbf{0}$, thus we have $\boldsymbol{u}_i^T = \mathbf{0}$ for all i and hence $A = \mathbf{0}$, but then $\boldsymbol{v} = \mathbf{0}$ which is a contradiction.

Proof 2 (if you already know about inner product and orthogonality): The simplest thing here is to note that since $A\boldsymbol{x} = \mathbf{0}$ for an $\boldsymbol{x} \in \mathrm{NS}(A)$, then $\boldsymbol{r}_i \perp \boldsymbol{x}$ where $\boldsymbol{r}_i^T$ is the i^{th} row of A. So $\mathrm{RS}(A) \perp \mathrm{NS}(A)$ and thus $\mathrm{RS}(A) \cap \mathrm{NS}(A) = \{\mathbf{0}\}$.

d) **Prove:** If A is an $n \times n$ matrix and $A^k = \mathbf{0}$ for any k, then $A^n = \mathbf{0}$.

Proof 1: To do this show

i) Show $\mathrm{NS}(A^{m+1}) \supseteq \mathrm{NS}(A^m)$ for all m.

ii) Show that if $\mathrm{NS}(A^{m+1}) = \mathrm{NS}(A^m)$, then $\mathrm{NS}(A^n) = \mathrm{NS}(A^m)$ for all $n \geq m$.

It is clear that $\mathrm{NS}(A^{m+1}) \supseteq \mathrm{NS}(A^m)$, since $A^m \boldsymbol{x} = \mathbf{0} \implies A(A^m \boldsymbol{x}) = \mathbf{0} \implies A^{m+1}\boldsymbol{x} = \mathbf{0}$. So (i) is shown,

For (ii) suppose $\mathrm{NS}(A^m) = \mathrm{NS}(A^{m+1})$, then $A^{m+2}\boldsymbol{x} = \mathbf{0} \implies A^{m+1}(A\boldsymbol{x}) = \mathbf{0} \implies A^m(A\boldsymbol{x}) = \mathbf{0} \implies A^{m+1}\boldsymbol{x} = \mathbf{0}$. So $\mathrm{NS}(A^{m+2}) \subseteq \mathrm{NS}(A^{m+1})$, but then $\mathrm{NS}(A^{m+2}) = \mathrm{NS}(A^{m+1}) = \mathrm{NS}(A^m)$. Now just keep going to get $\mathrm{NS}(A^k) = \mathrm{NS}(A^m)$ for all $k \geq m$.

This means we have

$$\mathrm{NS}(A^0) \subsetneq \mathrm{NS}(A^1) \subsetneq \mathrm{NS}(A^2) \subsetneq \cdots \subsetneq \mathrm{NS}(A^{m-1}) \subsetneq \mathrm{NS}(A^m) = \mathrm{NS}(A^{m+1}) = \cdots$$

The m at which $\mathrm{NS}(A^k) = \mathrm{NS}(A^m)$ for all $m \geq k$ must itself be $\leq n$.

If $A^k = \mathbf{0}$ for any k, then $\mathrm{NS}(A^k) = \mathbb{R}^n$ is maximal and thus $m \leq k$ and $\mathrm{NS}(A^m) = \mathbb{R}^n$. Since $m \leq n$, $\mathrm{NS}(A^n) = \mathbb{R}^n$ and so $A^n = \mathbf{0}$.

Proof 2: This is just like proof 1, but uses a descending tower of images rather than an ascending tower of null spaces.

Note: The *image*, *range*, and *column space* of A are all different names for the same space, in notation $\mathrm{Img}(A) = \mathrm{rng}(A) = \mathrm{CS}(A)$.

It is clear that $\mathrm{Img}(A^{k+1}) \subseteq \mathrm{Img}(A^k)$, since

$$y \in \mathrm{Img}(A) \iff (\exists \boldsymbol{x})(y = A^{k+1}\boldsymbol{x} = A^k(A\boldsymbol{x})) \implies (\exists z)(y = A^k z) \implies y \in \mathrm{Img}(A^k)$$

In fact if $\mathrm{Img}(A^{k+1}) = \mathrm{Img}(A^k)$, then $\mathrm{Img}(A^l) = \mathrm{Img}(A^k)$ for $l \geq k$. (This is as in Proof 1.) So we have a strictly decreasing sequence of images until the image stabilizes. If $A^k = 0$ for some k, then the image stabilizes at $\{\mathbf{0}\}$. That is

$$\mathbb{R}^n \supsetneq \mathrm{Img}(A) \supsetneq \mathrm{Img}(A^2) \cdots \supsetneq \mathrm{Img}(A^k) = \{0\}$$

and thus

$$n > \dim(\mathrm{Img}(A)) > \dim(\mathrm{Img}(A^2)) \cdots > \dim(\mathrm{Img}(A^k)) = 0$$

Clearly then the least such k is $\leq n$ and so $A^k = \mathbf{0} \implies A^n = \mathbf{0}$.

Proof 3: You can use induction. To do this we need to prove something that sounds slightly stronger:

P_n : For any $n \times n$ matrix A, if $A^m = \mathbf{0}$ for any $m > n$, then $A^n = 0$.

base case: (n = 1) If $A^m = [a]^m = [a^m] = [0]$, for m>1, then $a = 0$, so $A^1 = [a] = [0]$ as needed.

inductive step: Suppose P_{n-1}: For any $m > n-1$, $A^m = \mathbf{0} \implies A^{n-1} = \mathbf{0}$ for all $(n-1) \times (n-1)$ matrices. We want to prove P_n.

Assume A is an $n \times n$ matrix and $A^m = 0$ for some $m > n$. Notice that $\ker(A) \neq \{\mathbf{0}\}$, since if $\ker(A) = \{0\}$, then $A : \mathbb{R}^n \to \mathbb{R}^n$ is injective and thus A^m is also injective, so $\ker(A^m) = \{\mathbf{0}\}$. This obviously contradicts $A^m = \mathbf{0}$.

Let $\boldsymbol{v}_1 \in \ker(A)$ and let $\mathcal{B} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a basis for $\mathbb{R}^n$. So letting $B = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ we have

$$A = BA'B^{-1}$$

where

$$A' = \begin{bmatrix} 0 & a'_{12} & a'_{13} & \cdots & a'_{1n} \\ 0 & & & & \\ \vdots & & \hat{A} & & \\ 0 & & & & \end{bmatrix}$$

where $\hat{A}$ is the indicated $(n-1) \times (n-1)$ submatrix of A'.

A' is the matrix of $L(\boldsymbol{x}) = A\boldsymbol{x}$ with respect to the basis $\mathcal{B}$. Notice that $A^m = \mathbf{0}$ means $L^m(\boldsymbol{x}) = \mathbf{0}$ for all $\boldsymbol{x}$ and hence $A'\boldsymbol{x} = \mathbf{0}$ for all $\boldsymbol{x}$, a finally this means $A'^m = \mathbf{0}$.

Notice that A' has the block form

$$\begin{bmatrix} 0 & \boldsymbol{b}^T \\ \mathbf{0} & \hat{A} \end{bmatrix}$$

and

$$(A')^2 = \begin{bmatrix} 0 & \boldsymbol{b}^T \\ \mathbf{0} & \hat{A} \end{bmatrix} \begin{bmatrix} 0 & \boldsymbol{b}^T \\ \mathbf{0} & \hat{A} \end{bmatrix} = \begin{bmatrix} 0 & \boldsymbol{b}^T\hat{A} \\ \mathbf{0} & \hat{A}^2 \end{bmatrix}$$

By a second, obvious induction,

$$(A')^k = \begin{bmatrix} 0 & \boldsymbol{b}^T\hat{A}^{k-1} \\ \mathbf{0} & \hat{A}^k \end{bmatrix}$$

We assume $\hat{A'}^m = \mathbf{0}$ so $\hat{A}^m = \mathbf{0}$ and by induction $\hat{A}^{n-1} = \mathbf{0}$ and thus

$$(A')^n = \begin{bmatrix} 0 & \boldsymbol{b}^T\hat{A}^{n-1} \\ \mathbf{0} & \hat{A}^n \end{bmatrix} = \mathbf{0}$$

Proof 4: This is not a proof that I would expect to see except that if you do a Google search on the problem this comes up early in the results. This proof uses material we have not covered and will not cover in this course.

Some students find this proof and then can't quite carry it out, so I will also indicate the main error.

It is true that is λ is an eigenvalue for A and $\boldsymbol{x}$ and eigenvector for λ, then $A^k\boldsymbol{x} = \lambda^n\boldsymbol{x} = 0\boldsymbol{x} = 0$, so $\lambda^n = 0$ and hence $\lambda = 0$. But this does not mean that $A = 0$. In fact, here is a matrix where $A^n = 0$,

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

You can think about what A "does"

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \underset{A}{\rightarrow} \begin{bmatrix} 0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \underset{A}{\rightarrow} \cdots \underset{A}{\rightarrow} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_1 \end{bmatrix} \underset{A}{\rightarrow} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

So $A^n = 0$, but $A^m \neq 0$ for any $m < n$ and the only eigenvalue is $\lambda = 0$.

What you can argue is that if $p(t) = a_n x^n + \cdots a_1 x + a_0$ is the characteristic function, then $p(A) = 0$, Cayley-Hamilton theorem. Since 0 is the only eigenvalue $p(x) = a_n x^n$, so we know $a_n A^n = 0$, thus $A^n = 0$.

Proof 5: From the Schur decomposition (also later in the class, but one we will discuss), you know $A = PUP^{-1}$ where P is invertible and U is upper-triangular. (Actually, P will be better, *unitary*.)

Now we get $A^k = (PUP^{-1})(PUP^{-1}) \cdots (PUP^{-1}) = PU(P^{-1}P)U(P^{-1}P)U \cdots (P^{-1}P)UP^{-1} = PUIUI \cdots UIP^{-1} = PU^kP^{-1}$.

OK, do now we analyze U^k. The diagonal elements are just U^k_{ii}, since $U^2_{ii} = \sum_{k=1}^{n} U_{ik}U_{ki} = \sum_{k<i} 0 \cdot U_{ki} + U^2_{ii} + \sum_{k>i} U_{ik} \cdot 0 = U^2_{ii}$. Continue basically just like this.

Now $A^k = PU^kP^{-1} = 0$ so $U^k = 0$. But this means that $U^k_{ii} = 0$ and thus $U_{ii} = 0$ for all i. So U is all 0 on the diagonal.

An $n \times n$ actually has $2n - 1$ diagonals, the diagonals:

$$\text{diag}_{-(n-1)}, \text{diag}_{-(n-2)}, \ldots, \text{diag}_0, \ldots \text{diag}_{(n-2)}, \text{diag}_{n-1}.$$

The entries of diag_k are those a_{ij}, so that $j - i = k$. diag_0 has all a_{ii} so it is the usual diagonal. Now a matrix is upper-triangular if $\text{diag}_k = \mathbf{0}$ for $k < 0$. We see that our U satisfied $\text{diag}_k = 0$ for $k \leq 0$.

We will show that U^m satisfies $\text{diag}_k = \mathbf{0}$ for $k \leq m-1$. Thus U^n satisfies $\text{diag}_k = 0$ for $k \leq n-1$, so $U^n = \mathbf{0}$. Thus we have $A^n = PU^nP^{-1} = P\mathbf{0}P^{-1} = \mathbf{0}$. This is what we wanted to prove.

We must show that U^m satisfies $\text{diag}_k = \mathbf{0}$ for $k \leq m-1$. This is basically induction, it is true for $m = 1$, as indicated above. Suppose it is true for m, then we want to show it true for $m+1$. So take $A = U^m$. We know, $a_{ij} = 0$ for all $j - i \leq m - 1$.

Compute $(AU)_{ij}$ where $j-i=m$, this is the m^{th} diagonal of $AU=U^{m+1}$. $j-i=m$ is the same as $j=i+m$:

$$(AU)_{ij}=\sum_{k=1}^{n}a_{ik}u_{kj}=\sum_{k=1}^{m}a_{ik}u_{kj}+\sum_{k=i+m+1}^{n}a_{ik}u_{kj}=\sum_{k=1}^{i+m-1}0\cdot u_{kj}+\sum_{k=i+m}^{n}a_{ik}\cdot 0=0$$