- New concepts:
  - Vector space sums
  - FTLA
  - FTLA step-by-step (question by question) proof
  - Rank-nullity theorem
  - Characterizing solutions to linear systems in terms of rank and nullity

Definition. *Given two vector subspaces* $(\mathcal{U},\mathcal{S},+)$, $(\mathcal{V},\mathcal{S},+)$ *of the space* $(\mathcal{W},\mathcal{S},+)$, *the sum is the set* $\mathcal{U}+\mathcal{V}=\{\boldsymbol{u}+\boldsymbol{v}|\ \boldsymbol{u}\in\mathcal{U},\boldsymbol{v}\in\mathcal{V}\}$.

Definition. *Given two vector subspaces* $(\mathcal{U},\mathcal{S},+)$, $(\mathcal{V},\mathcal{S},+)$ *of the space* $(\mathcal{W},\mathcal{S},+)$, *the direct sum is the set* $\mathcal{U}\oplus\mathcal{V}=\{\boldsymbol{u}+\boldsymbol{v}|\ \exists!\boldsymbol{u}\in\mathcal{U},\exists!\boldsymbol{v}\in\mathcal{V}\}$. *(unique decomposition)*

Definition. *Given two vector subspaces* $(\mathcal{U},\mathcal{S},+)$, $(\mathcal{V},\mathcal{S},+)$ *of the space* $(\mathcal{W},\mathcal{S},+)$, *the intersection is the set*

$$\mathcal{U}\cap\mathcal{V}=\{\boldsymbol{x}|\boldsymbol{x}\in\mathcal{U},\boldsymbol{x}\in\mathcal{V}\}.$$

Definition. *Two vector subspaces* $(\mathcal{U},\mathcal{S},+)$, $(\mathcal{V},\mathcal{S},+)$ *of the space* $(\mathcal{W},\mathcal{S},+)$ *are orthogonal subspaces, denoted* $\mathcal{U}\perp\mathcal{V}$ *if* $\boldsymbol{u}^T\boldsymbol{v}=0$ *for any* $\boldsymbol{u}\in\mathcal{U},\boldsymbol{v}\in\mathcal{V}$.

Definition. *Two vector subspaces* $(\mathcal{U},\mathcal{S},+)$, $(\mathcal{V},\mathcal{S},+)$ *of the space* $(\mathcal{W},\mathcal{S},+)$ *are orthogonal complements, denoted* $\mathcal{U}=\mathcal{V}^{\perp}$, $\mathcal{V}=\mathcal{U}^{\perp}$ *if they are orthogonal subspaces and* $\mathcal{U}\cap\mathcal{V}=\{\mathbf{0}\}$, *i.e., the null vector is the only common element of both subspaces.*

- A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$, $\boldsymbol{A}: \mathbb{R}^n \to \mathbb{R}^m$
- The transpose $\boldsymbol{A}^T \in \mathbb{R}^{n \times m}$ is a linear mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$, $\boldsymbol{A}^T: \mathbb{R}^m \to \mathbb{R}^n$
- To each matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ associate four fundamental subspaces:
  1. *Column space*, $C(\boldsymbol{A}) = \{\boldsymbol{b} \in \mathbb{R}^m | \exists \boldsymbol{x} \in \mathbb{R}^n \text{ such that } \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}\} \subseteq \mathbb{R}^m$, the part of $\mathbb{R}^m$ *reachable* by linear combination of columns of $\boldsymbol{A}$
  2. *Left null space*, $N(\boldsymbol{A}^T) = \{\boldsymbol{y} \in \mathbb{R}^m | \boldsymbol{A}^T\boldsymbol{y} = 0\} \subseteq \mathbb{R}^m$, the part of $\mathbb{R}^m$ *not reachable* by linear combination of columns of $\boldsymbol{A}$
  3. *Row space*, $R(\boldsymbol{A}) = C(\boldsymbol{A}^T) = \{\boldsymbol{c} \in \mathbb{R}^n | \exists \boldsymbol{y} \in \mathbb{R}^m \text{ such that } \boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{y}\} \subseteq \mathbb{R}^n$, the part of $\mathbb{R}^n$ *reachable* by linear combination of rows of $\boldsymbol{A}$
  4. *Null space*, $N(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} = 0\} \subseteq \mathbb{R}^n$, the part of $\mathbb{R}^n$ *not reachable* by linear combination of rows of $\boldsymbol{A}$

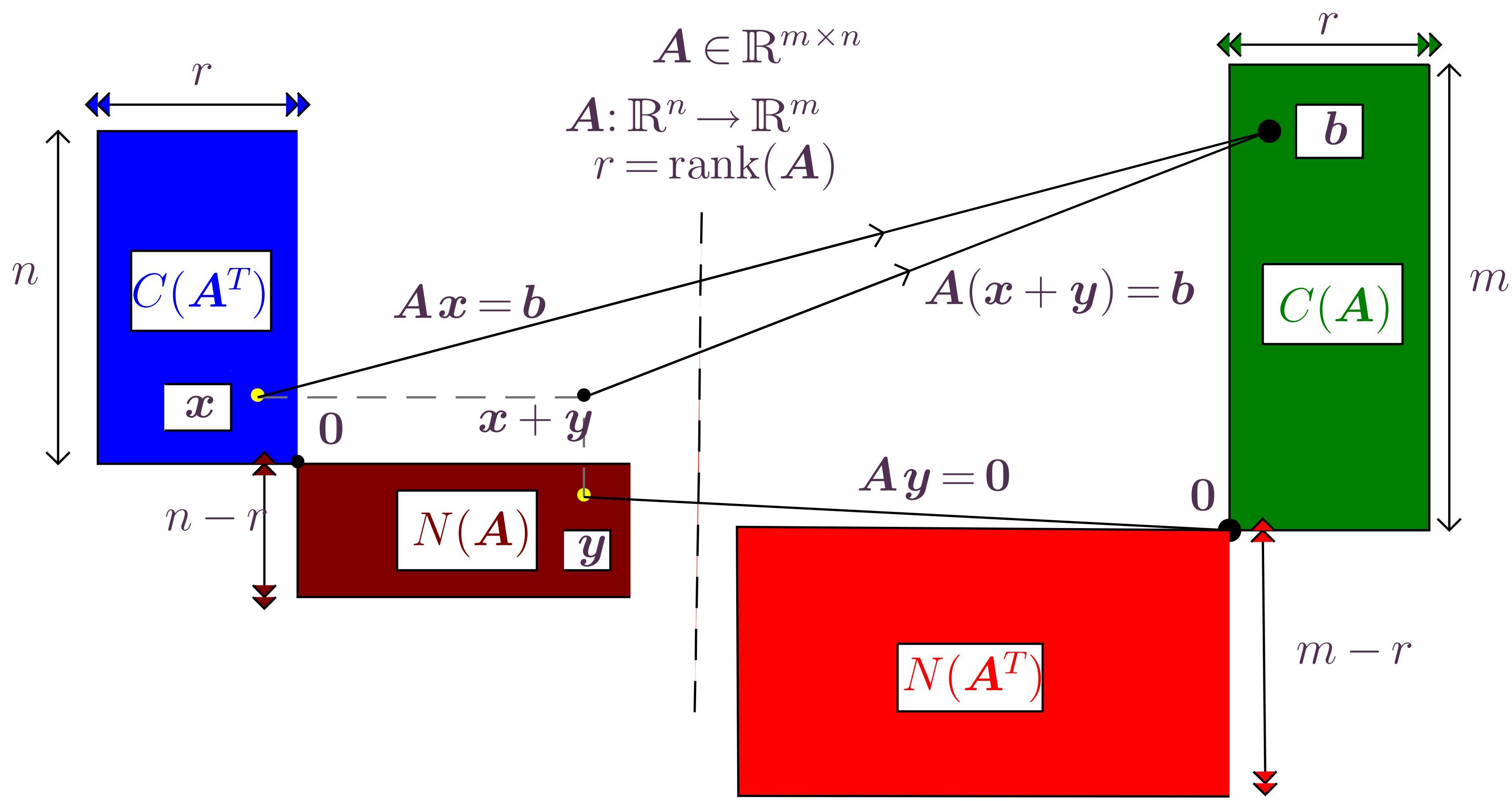
Theorem. *Given the linear mapping associated with matrix* $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ *we have:*

1. $C(\boldsymbol{A}) \oplus N(\boldsymbol{A}^T) = \mathbb{R}^m$, *the direct sum of the column space and left null space is the codomain of the mapping*
2. $C(\boldsymbol{A}^T) \oplus N(\boldsymbol{A}) = \mathbb{R}^n$, *the direct sum of the row space and null space is the domain of the mapping*
3. $C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T)$ *and* $C(\boldsymbol{A}) \cap N(\boldsymbol{A}^T) = \{\mathbf{0}\}$, *the column space is orthogonal to the left null space, and they are orthogonal complements of one another,*

$$C(\boldsymbol{A}) = N(\boldsymbol{A}^T)^\perp, \;\; N(\boldsymbol{A}^T) = C(\boldsymbol{A})^\perp \; .$$

4. $C(\boldsymbol{A}^T) \perp N(\boldsymbol{A})$ *and* $C(\boldsymbol{A}^T) \cap N(\boldsymbol{A}) = \{\mathbf{0}\}$, *the row space is orthogonal to the null space, and they are orthogonal complements of one another,*

$$C(\boldsymbol{A}^T) = N(\boldsymbol{A})^\perp, \;\; N(\boldsymbol{A}) = C(\boldsymbol{A}^T)^\perp \; .$$



$\mathbb{R}^n = C(A^T) \oplus N(A)$ usually: $m \geqslant n$ $\mathbb{R}^m = N(A^T) \oplus C(A)$

$C(A^T) \perp N(A)$ $N(A^T) \perp C(A)$

- Understanding the FTLA is essential to applications
- Proofs of the FTLA help in:
  - building the ability to recognize rigorous mathematical arguments as opposed to intuition
  - gaining an appreciation of the interplay between construction of mathematical concepts (formalized in a definition) and interaction between these concepts (propositions and theorems).
- Recall: linear algebra seeks construction of complex objects, vectors $\boldsymbol{b} \in \mathbb{R}^m$ through linear combination of n column vectors organized into a matrix

  $\boldsymbol{A} = [\ \boldsymbol{a}_1 \ \ \ldots \ \ \boldsymbol{a}_n\ ]$, $\boldsymbol{b} = T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{x} \in \mathbb{R}^n$ scaling coefficients.

  $T: \mathbb{R}^n \to \mathbb{R}^m$ a linear mapping from domain $\mathbb{R}^n$ to codomain $\mathbb{R}^m$.

  A proof of the FTLA is now presented as answers (first informal, and then rigorous) to a series of natural questions arising from the initial goal:

  1. What vectors can be obtained by the linear combination $\boldsymbol{A}\boldsymbol{x}$?
  2. Is there only one way to obtain a vector $\boldsymbol{b}$ by linear combination?
  3. Is there a preferred way to describe vectors $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$?
  4. Is there a preferred way to describe vectors $\boldsymbol{y}$ that satisfy $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$?
  5. Is there anything different about organizing vectors into rows?

- Set of reachable vectors: defined by *column space* of $\boldsymbol{A}$ (range of mapping T)

$$C(\boldsymbol{A}) = \{\boldsymbol{b} | \exists \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}\} \subseteq \mathbb{R}^m$$

- $C(\boldsymbol{A})$ has structure, it is a vector subspace of $\mathbb{R}^m$, $C(\boldsymbol{A}) \leq \mathbb{R}^m$
  Proof. $\forall \boldsymbol{u}, \boldsymbol{v} \in C(\boldsymbol{A}) \Rightarrow \exists \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ such that $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{v} = \boldsymbol{A}\boldsymbol{y}$ by definition of $C(\boldsymbol{A})$. Then $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{y})$, hence $\boldsymbol{u} + \boldsymbol{v} \in C(\boldsymbol{A})$.
- "Size" of a vector space has been characterized by the concept of dimension. Give a distinct name to the "size" of the set of reachable vectors.

Definition. *The rank of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the dimension of the column space*

$$r = \dim C(\boldsymbol{A})$$

- Suppose $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$, could $\boldsymbol{b}$ also be obtained differently, as $\boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y})$?
- Subtracting the two linear combinations leads to $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$, and the *null space*

$$N(\boldsymbol{A}) = \{\boldsymbol{y} | \boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}\} \subseteq \mathbb{R}^n.$$

- $N(\boldsymbol{A})$ has structure, it is a vector subspace of $\mathbb{R}^n$, $N(\boldsymbol{A}) \leq \mathbb{R}^n$
  Proof. $\forall \boldsymbol{u}, \boldsymbol{v} \in N(\boldsymbol{A}) \Rightarrow \boldsymbol{A}\boldsymbol{u} = \boldsymbol{0}, \boldsymbol{A}\boldsymbol{v} = \boldsymbol{0} \Rightarrow \boldsymbol{A}(\boldsymbol{u}+\boldsymbol{v}) = \boldsymbol{0} \Rightarrow \boldsymbol{u}+\boldsymbol{v} \in N(\boldsymbol{A})$.
- Obviously, $\boldsymbol{0} \in N(\boldsymbol{A})$, but the null space might also contain non-zero vectors
- Even when $\boldsymbol{A} \neq \boldsymbol{0}$ (i.e., $\boldsymbol{A}$ is not the zero matrix) and $\boldsymbol{y} \neq \boldsymbol{0}$ (i.e., $\boldsymbol{y}$ is not the zero vector), there still might be choices of $\boldsymbol{A}$ and $\boldsymbol{y}$ such that $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$. This is different from $a, x \in \mathbb{R}$, where $ax = 0 \Rightarrow a = 0 \text{ or } x = 0$.
- Give a distinct name to the "size" of the set of such vectors.

Definition. *The nullity of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the dimension of the null space*

$$z = \dim N(\boldsymbol{A})$$

- Since $r = \dim C(\boldsymbol{A})$ and $C(\boldsymbol{A})$ is a vector subspace of $\mathbb{R}^m$, $r \leqslant m$.
- The above implies that only r of the n columns of $\boldsymbol{A}$ are linearly independent
- Gather the linearly independent columns as the first r columns

$$\boldsymbol{A} = [\ \boldsymbol{A}_r \quad \boldsymbol{A}_{n-r}\ ], \boldsymbol{A}_r \in \mathbb{R}^{m \times r}, \boldsymbol{A}_{n-r} \in \mathbb{R}^{m \times (n-r)},$$

  a block decomposition of $\boldsymbol{A}$, with the index denoting number of columns.
- Since columns $\boldsymbol{A}_{n-r}$ are linearly dependent on those of $\boldsymbol{A}_r$, $\boldsymbol{A}_{n-r} = \boldsymbol{A}_r \boldsymbol{B}_{n-r}$

$$\boldsymbol{A} = \boldsymbol{A}_r [\ \boldsymbol{I}_r \quad \boldsymbol{B}_{n-r}\ ] = \boldsymbol{A}_r \boldsymbol{X}, \boldsymbol{X} \in \mathbb{R}^{r \times n}$$

  The above states: "all the column vectors of $\boldsymbol{A}$ can be expressed as linear combinations of r linearly independent columns, $\boldsymbol{A}_r$.

- Consider now $N(\boldsymbol{A})$: $\boldsymbol{A}\boldsymbol{u}=\boldsymbol{A}_r\boldsymbol{X}\boldsymbol{u}=\boldsymbol{0}$
- Let $\boldsymbol{v}=\boldsymbol{X}\boldsymbol{u}$. $\boldsymbol{A}_r$ with linearly independent columns, $\boldsymbol{A}_r\boldsymbol{v}=\boldsymbol{0}\Rightarrow\boldsymbol{v}=0\Rightarrow\boldsymbol{X}\boldsymbol{u}=\boldsymbol{0}$. Write this out in blocks

$$[\ \boldsymbol{I}_r\ \ \boldsymbol{B}_{n-r}\ ]\left[\begin{array}{c}\boldsymbol{u}_r\\ \boldsymbol{u}_{n-r}\end{array}\right]=\boldsymbol{0}\Rightarrow\boldsymbol{u}=\left[\begin{array}{c}-\boldsymbol{B}_{n-r}\\ \boldsymbol{I}_{n-r}\end{array}\right]\boldsymbol{u}_{n-r}=\boldsymbol{Y}_{n-r}\boldsymbol{u}_{n-r}.$$

This states that columns of $\boldsymbol{Y}_{n-r}$ are a spanning set for $N(\boldsymbol{A})$, $\boldsymbol{u}=\boldsymbol{Y}_{n-r}\boldsymbol{u}_{n-r}$.

- Is it a minimal spanning set, i.e., a basis? Consider

$$\boldsymbol{Y}_{n-r}\boldsymbol{w}=\boldsymbol{0}\Rightarrow\left[\begin{array}{c}-\boldsymbol{B}_{n-r}\\ \boldsymbol{I}_{n-r}\end{array}\right]\boldsymbol{w}=\left[\begin{array}{c}-\boldsymbol{B}_{n-r}\boldsymbol{w}\\ \boldsymbol{w}\end{array}\right]=\left[\begin{array}{c}\boldsymbol{0}\\ \boldsymbol{0}\end{array}\right]\Rightarrow\boldsymbol{w}=\boldsymbol{0}.$$

Indeed columns of $\boldsymbol{Y}_{n-r}$ are linearly independent, establishing that

$$z=\dim N(\boldsymbol{A})=n-r$$

Theorem. *(Rank-nullity theorem) For* $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $r+z=n$

- Matrix vector multiplication $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$, expresses a linear combination of columns. Multiple linear combination of columns: $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{X}$.
- Organizing data into column vectors is an arbitrary choice, hence linear combinations of rows should also be possible, and indeed are expressed as

$$\boldsymbol{c}^T = \boldsymbol{y}^T\boldsymbol{A} \Rightarrow \boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{y}.$$

  $\boldsymbol{c}^T$: the row vector obtained by linear combination of rows of $\boldsymbol{A}$ with scaling coefficients gathered in the row vector $\boldsymbol{y}^T$. Multiple linear combinations of rows

$$\boldsymbol{C} = \boldsymbol{Y}\boldsymbol{A}$$

  Rows of $\boldsymbol{C}$ are linear combination of rows of $\boldsymbol{A}$ with scalings as rows of $\boldsymbol{Y}$.
- The set of vectors reachable by linear combination of rows of $\boldsymbol{A}$ is $C(\boldsymbol{A}^T)$

$$C(\boldsymbol{A}^T) = \{\boldsymbol{c} | \exists \boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{y}\} \leq \mathbb{R}^n,$$

  the *row space* or column space of the transpose, and is a subspace of $\mathbb{R}^n$.
- Let $p = \dim C(\boldsymbol{A}^T)$, the dimension of the row space.

Proposition. *The dimension of the column space equals that of the row space*

$$r = \dim C(\boldsymbol{A}) = \dim C(\boldsymbol{A}^T) = p.$$

Proof. Interpret $\boldsymbol{A} = \boldsymbol{A}_r \boldsymbol{X}$ as stating that the rows of $\boldsymbol{A}$ can be obtained linear combinations of the rows of $\boldsymbol{X}$ with scalings contained in $\boldsymbol{A}_r$. Since $\boldsymbol{X} \in \mathbb{R}^{r \times n}$,

$$p = \dim C(\boldsymbol{A}^T) \leqslant r = \dim C(\boldsymbol{A}).$$

The above is true for any matrix $\boldsymbol{M}$, $\dim C(\boldsymbol{M}^T) \leqslant \dim C(\boldsymbol{M})$. Choose $\boldsymbol{M} = \boldsymbol{A}^T$

$$\dim C((\boldsymbol{A}^T)^T) \leqslant \dim C(\boldsymbol{A}^T) \Rightarrow r = \dim C(\boldsymbol{A}) \leqslant \dim C(\boldsymbol{A}^T) = p.$$

Since $p \leqslant r$ and $r \leqslant p$, it results that $r = p$.

Results up to now can be used to prove:

- $C(\boldsymbol{A})$ is orthogonal to $N(\boldsymbol{A}^T)$.
  Proof. $\boldsymbol{u} \in C(\boldsymbol{A}) \Rightarrow \boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{v} \in N(\boldsymbol{A}^T) \Rightarrow \boldsymbol{A}^T\boldsymbol{v} = \boldsymbol{0}$. Compute $\boldsymbol{u}^T\boldsymbol{v} = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{v} = \boldsymbol{x}^T\boldsymbol{0} = \boldsymbol{0}$.
- $\boldsymbol{0}$ is the only vector both in $C(\boldsymbol{A})$ and $N(\boldsymbol{A}^T)$.
  Proof. Assume there might be $\boldsymbol{b} \in C(\boldsymbol{A})$ and $b \in N(\boldsymbol{A}^T)$ and $\boldsymbol{b} \neq \boldsymbol{0}$. Since $\boldsymbol{b} \in C(\boldsymbol{A})$, $\exists \boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$. Since $\boldsymbol{b} \in N(\boldsymbol{A}^T)$, $\boldsymbol{A}^T\boldsymbol{b} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0}$. Note that $\boldsymbol{x} \neq 0$ since $\boldsymbol{x} = 0 \Rightarrow \boldsymbol{b} = 0$, contradicting assumptions. Multiply equality $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ on left by $\boldsymbol{x}^T$,

$$\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{b}^T\boldsymbol{b} = \|\boldsymbol{b}\|^2 = 0 \Rightarrow \boldsymbol{b} = 0.$$

- $C(\boldsymbol{A}) \oplus N(\boldsymbol{A}^T) = \mathbb{R}^m$. Rank-nullity theorem states $\dim C(\boldsymbol{A}) + \dim N(\boldsymbol{A}^T) = m$, thereby covering the entire codomain, $\mathbb{R}^m$.