OPTIONAL 02

Linear Algebra 02

Orthogonality, Rank, Invertibility

Data 100/Data 200, Spring 2023 @ UC Berkeley

Narges Norouzi and Lisa Yan

Content credit: The wealth of instructional knowledge from Math 54, Math 91, and EECS16A

Today's Roadmap

Optional 02, Data 100 Spring 2023

Assumptions and Disclaimers

- This Linear Algebra optional lecture is primarily geared towards co-enrolled Math 54 and Data 100 students in Spring 2023.
  - It may also be useful as a refresher to other students.
  - At the time of this review session, I am assuming you have covered all of "Week 1 curriculum" on [Prof. Lott's page](#) (linear combinations, spans, Gaussian elimination).
- This is part one in a multipart series.
  - Last time: Vector properties and Matrix multiplication
  - This time: Matrix properties
- These optional lectures are **not** a substitute for a linear algebra course. We only cover basics.
- The goal of these optional lectures is to prepare you for the linear algebra requirements for the Ordinary Least Squares unit.
  - Linear algebra will be needed in proofs on Homework 06: Regression.
  - On the Spring 2023 midterm, we will try to de-emphasize linear algebra—as in, we will expect you to be familiar with equations (and how to use them), but we will not ask you to prove anything.

Credits for this lecture, and other Linear Algebra Resources

- Material credit:
  - Math 54 and its textbook (Lay, Lay, McDonald)
  - The new Math 91 course (bCourses)
  - EECS 16A (website)

- https://ds100.org/sp23/resources/#calculus-and-linear-algebra
  - Includes non-Berkeley sources like Khan Academy, 3Brown1Blue, etc.

Main Takeaways of Today

If $\mathbb{X} \in \mathbb{R}^{n \times p}$, then we can decompose $\vec{y} \in \mathbb{R}^n$ uniquely as $\vec{y} = \hat{y} + (\vec{y} - \hat{y})$, where

- $\hat{y} \in \text{span}\,\{\mathbb{X}\}$ is the **orthogonal projection** of y onto the **column space** of $\mathbb{X}$.
- $(\vec{y} - \hat{y}) \perp \text{span}\,\{\mathbb{X}\}$.
- $\hat{y}$ is the **best approximation** to $\vec{y}$ by elements of $\text{span}\,\{\mathbb{X}\}$.

For a square matrix in $\mathbb{R}^{n \times n}$,
full rank = full column rank = full row rank = n.

For a **square** matrix $A \in \mathbb{R}^{n \times n}$:

- Matrix A is **invertible**.
- A has linearly independent columns.
- A is **full rank**.
- The columns of A span $\mathbb{R}^n$.
- For every $b \in \mathbb{R}^n$, $A\vec{x} = \vec{b}$ has a **unique** solution $x = A^{-1}b$.

Transpose: Review and Properties

Optional 02, Data 100 Spring 2023

Transpose Properties

THEOREM 3

Let A and B denote matrices whose sizes are appropriate for the following sums and products.

a. $(A^T)^T = A$

b. $(A+B)^T = A^T + B^T$

c. For any scalar r, $(rA)^T = rA^T$

d. $(AB)^T = B^T A^T$

PRACTICE PROBLEMS

1. Since vectors in $\mathbb{R}^n$ may be regarded as $n \times 1$ matrices, the properties of transposes in Theorem 3 apply to vectors, too. Let

$$A = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Compute $(A\mathbf{x})^T$, $\mathbf{x}^T A^T$, $\mathbf{x}\mathbf{x}^T$, and $\mathbf{x}^T\mathbf{x}$. Is $A^T\mathbf{x}^T$ defined?

A B C D E

$$Ax = \underset{2\times 2}{\begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix}}\underset{2\times 1}{\begin{bmatrix} 5 \\ 3 \end{bmatrix}} = \underset{2\times 1}{\begin{bmatrix} -4 \\ 2 \end{bmatrix}}$$

A $(Ax)^T = \begin{bmatrix} -4 & 2 \end{bmatrix}$

B $x^T A^T \rightarrow \underset{1\times 2}{\overset{x^T}{\begin{bmatrix} 5 & 3 \end{bmatrix}}}\underset{2\times 2}{\overset{A^T}{\begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix}}} = \underset{1\times 2}{\begin{bmatrix} -4 & 2 \end{bmatrix}}$

$(AB)^T = B^T A^T$

$(Ax)^T = x^T A^T$

C $\underset{2\times 1}{x}\,\underset{1\times 2}{x^T} \rightarrow = \begin{bmatrix} 5 \\ 3 \end{bmatrix}\begin{bmatrix} 5 & 3 \end{bmatrix} = \underset{2\times 2}{\begin{bmatrix} 25 & 15 \\ 15 & 9 \end{bmatrix}}$

D $x^T x = \underset{1\times 2}{\begin{bmatrix} 5 & 3 \end{bmatrix}}\underset{2\times 1}{\begin{bmatrix} 5 \\ 3 \end{bmatrix}} = \underset{1\times 1}{\begin{bmatrix} 34 \end{bmatrix}} = \underset{\mathbb{R}}{34}$

1×1

E $\underset{2\times 2}{A^T}\,\underset{1\times 2}{x^T}$ nope

Orthogonality

Optional 02, Data 100 Spring 2023

Orthogonality links Geometry to Algebra

To "compare" two vectors:

- Draw them in a geometric space.
- Do some algebraic manipulation with them.

perpendicular.

The dot product links algebra and geometry together via a concept called **orthogonality**.

$$[\qquad\qquad] \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} \to [\ ] \in \mathbb{R}$$

Orthogonality links Geometry to Algebra

To "compare" two vectors:

- Draw them in a geometric space.
- Do some algebraic manipulation with them.

The dot product links algebra and geometry together via a concept called **orthogonality**.

$(u^T v)^T = v^T u^{TT} = v^T u$

Two vectors **u** and **v** in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$. $= u^T v = v^T u$

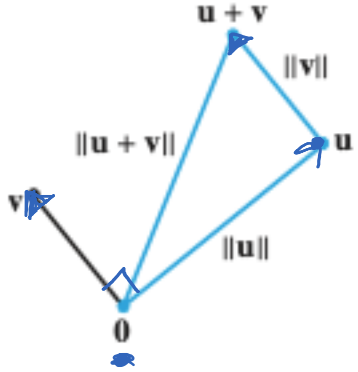
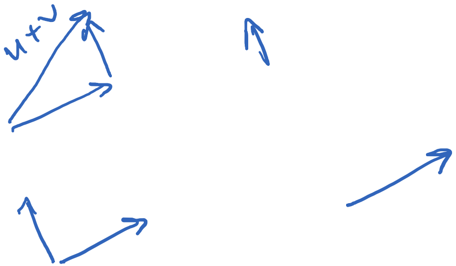


FIGURE 6

The Pythagorean Theorem

Two vectors **u** and **v** are orthogonal if and only if $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.



$\begin{bmatrix} 2^c \\ 3^c \\ 0^c \end{bmatrix}$

We will write:
$u \perp v$

Orthogonality Exercises

Two vectors **u** and **v** in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

We will write:
$u \perp v$

Are the following vectors orthogonal to each other?

1. $\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{v} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

2. $\vec{u} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}, \vec{v} = \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$

3. $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \vec{v} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$

Orthogonality Exercises

Two vectors **u** and **v** in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

We will write:
$u \perp v$

Are the following vectors orthogonal to each other?

1. $\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{v} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

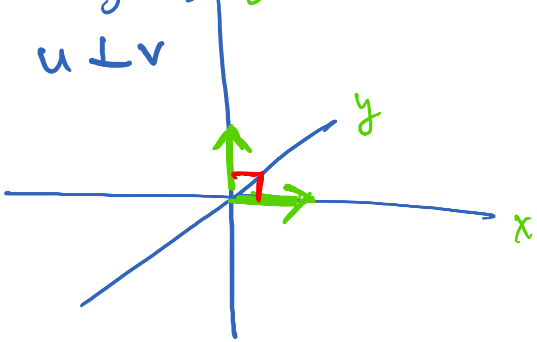
2. $\vec{u} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}, \vec{v} = \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$

3. $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \vec{v} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$

$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0$

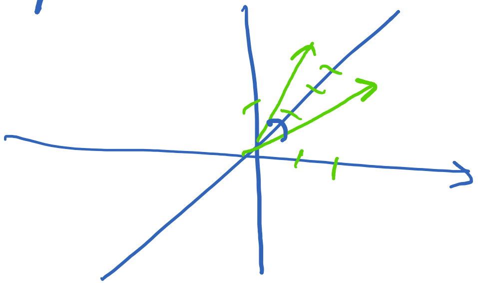
yes, 3

$u \perp v$



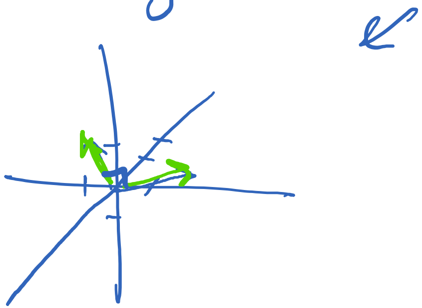
$\begin{bmatrix} 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} = 9$

$u \not\perp v$ no



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

$u \perp v$ yes



Orthogonal Projections

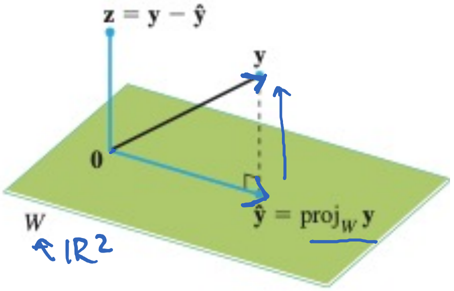


FIGURE 2
Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

An Orthogonal Projection

Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$. We wish to write

$$\mathbf{y} = \boxed{\hat{\mathbf{y}}} + \boxed{\mathbf{z}} \tag{1}$$

where $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$. See Figure 2.

The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of y onto u**, and the vector $\mathbf{z}$ is called the **component of y orthogonal to u**.

$$\boxed{\hat{\mathbf{y}} = \text{proj}_L\, \mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}} \tag{2}$$

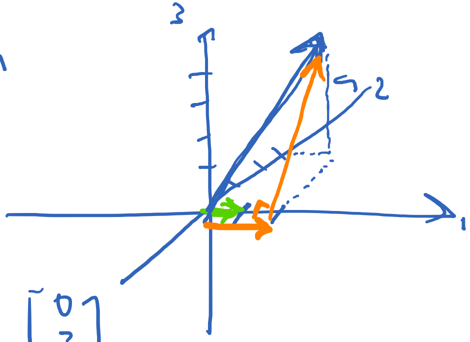
$\mathbb{R}$ $\mathbb{R}^n$

Let $\vec{u} = \begin{bmatrix}1\\0\\0\end{bmatrix}$. Let $\vec{y} = \begin{bmatrix}2\\3\\4\end{bmatrix}$. $\rightarrow$

$$\frac{\vec{y}^T\vec{u}}{\vec{u}^T\vec{u}} = \frac{2}{1} = 2$$

$$2\vec{u} = \begin{bmatrix}2\\0\\0\end{bmatrix}$$

$$\rightarrow \quad y = \begin{bmatrix}2\\3\\4\end{bmatrix} = \overset{\hat{y}}{\begin{bmatrix}2\\0\\0\end{bmatrix}} + \overset{\tilde{}}{\begin{bmatrix}0\\3\\4\end{bmatrix}}$$



(Lay 6.2)

Orthogonal decomposition: (orthogonal projection) + (orthogonal vector)

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \boxed{\hat{\mathbf{y}}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^\perp$. In fact, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p \tag{2}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

$\hat{y} \in \text{span}\{u_1, u_2\}$

$z \perp \text{span}\{u_1, u_2\}$

There's a lot of terminology here, but here's the general idea:

Let $\overrightarrow{u_1} = \begin{bmatrix}1\\0\\0\end{bmatrix}, \overrightarrow{u_2} = \begin{bmatrix}0\\1\\0\end{bmatrix}$. Let $\vec{y} = \begin{bmatrix}1/2\\1/3\\1/4\end{bmatrix}$.

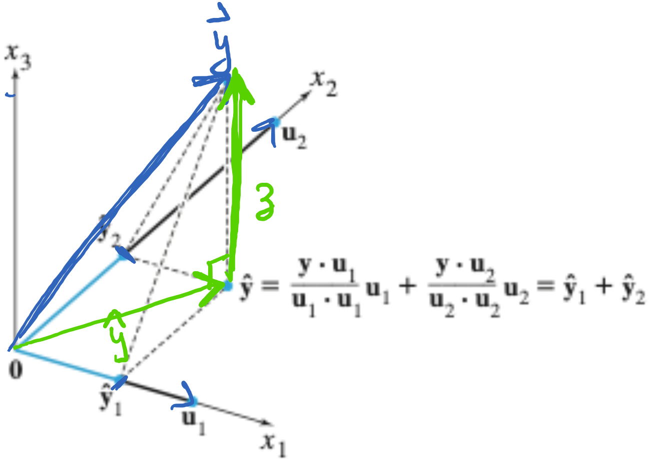


FIGURE 3 The orthogonal projection of $\mathbf{y}$ is the sum of its projections onto one-dimensional subspaces that are mutually orthogonal.

$\hat{y}$: **orthogonal projection** of $\boldsymbol{y}$ onto $\text{span}\{u_1, u_2\}$

$z = y - \hat{y}$: **orthogonal** to $\text{span}\{u_1, u_2\}$.

(Lay 6.3)

Orthogonal Projection Practice

PRACTICE PROBLEMS

1. Let $\mathbf{u}_1 = \begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Use the fact that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal to compute $\text{proj}_W\, \mathbf{y}$.

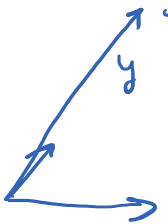
SOLUTION TO PRACTICE PROBLEMS

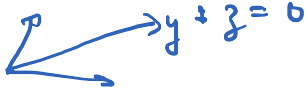
1. Compute

$$\text{proj}_W\, \mathbf{y} = \boxed{\frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \frac{88}{66}\mathbf{u}_1 + \frac{-2}{6}\mathbf{u}_2$$

$$= \boxed{\frac{4}{3}}\begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix} = \underbrace{\begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix} = \mathbf{y}}$$

$\frac{4}{3} u_1 - \frac{1}{3} u_2$





$y = \text{proj}_W\, y$

(Lay 6.3)

Vector Decomposition

Optional 02, Data 100 Spring 2023

Vector Decompositions

Any vector $\vec{y} \in \mathbb{R}^n$ can be written as a sum of two vectors (or infinite vectors, for that matter).

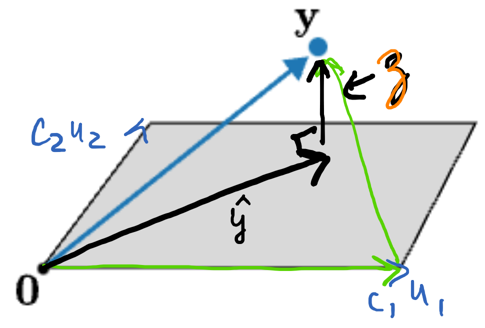
$$\vec{y} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} -3 \\ -4 \\ 0 \end{bmatrix} + \begin{bmatrix} 5 \\ 7 \\ 4 \end{bmatrix}$$



$$\vec{y} = \boxed{\hat{y}} + z$$

$$= \hat{y} + (\vec{y} - \hat{y})$$

$$= \cancel{y} + \vec{y} - \cancel{y}$$

Column Spaces

Suppose $\mathbb{X} \in \mathbb{R}^{n \times p}$ is a matrix with columns $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_p$.

$\text{span}\{\mathbb{X}\}$ is the **span of the columns** of $\mathbb{X}$, AKA the **column space** of $\mathbb{X}$.

Intuitively, $\text{span}\{\mathbb{X}\} = \text{span}\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_p\}$ is all the vectors you can "reach" using the columns of $\mathbb{X}$ via linear combination:

$$\boxed{\text{span}\{\mathbb{X}\}} = \left\{ \sum_{j=1}^{p} \theta_j \vec{x}_j \,\middle|\, \theta_1, \theta_2, \ldots, \theta_p \in \mathbb{R} \right\}$$

$\mathbb{X}$ has a row space, too, but we will only use $\text{span}\{\mathbb{X}\}$ to refer to the column space.

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

$$\underbrace{\theta_1}_{\mathbb{R}} \vec{x}_1 + \theta_2 \vec{x}_2 + \ldots + \theta_p \vec{x}_p$$

$$\begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_p \end{bmatrix}$$

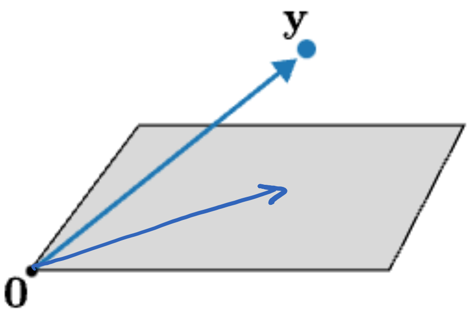
Notation/Reasoning check

$\perp$ (orthogonality) is a relationship between:
- (vector, vector)
- (vector, subspace)

Let $\boldsymbol{y} \in \mathbb{R}^n$ and $\mathbb{X} \in \mathbb{R}^{n \times p}$.

Which of the following notation/reasoning makes sense?

1. Define $\hat{y}$ as the **orthogonal projection** of $\vec{y}$ onto the column space of $\mathbb{X}$.
   A. $\hat{y} \in \mathbb{R}^n$, B. $\hat{y} \in \mathbb{R}^p$ C. Neither
   A. $\hat{y} \perp \mathbb{R}^n$, B. $\hat{y} \perp \mathbb{R}^p$ C. Neither
   A. $\hat{y} \perp \boldsymbol{y}$ B. $\hat{y} \perp \mathbb{X}$ C. Neither
   A. $\hat{y} \in \text{span}\,\{\mathbb{X}\}$, B. $\hat{y} \perp \text{span}\,\{\mathbb{X}\}$ C. Neither
2. Define $\vec{z} = \boldsymbol{y} - \hat{y}$:
   A. $\vec{z} \in \mathbb{R}^n$, B. $\vec{z} \in \mathbb{R}^p$ C. Neither
   A. $\vec{z} \perp \mathbb{R}^n$, B. $\vec{z} \perp \mathbb{R}^p$ C. Neither
   A. $\vec{z} \perp \boldsymbol{y}$ B. $\vec{z} \perp \hat{y}$ C. Neither
   A. $\vec{z} \perp \boldsymbol{y}$ B. $\vec{z} \perp \mathbb{X}$ C. Neither
   A. $\vec{z} \in \text{span}\,\{\mathbb{X}\}$, B. $\vec{z} \perp \text{span}\,\{\mathbb{X}\}$ C. Neither



Notation/Reasoning check

> ⊥ (orthogonality) is a relationship between:
> - (vector, vector)
> - (vector, subspace)

Let $\boldsymbol{y} \in \mathbb{R}^n$ and $\mathbb{X} \in \mathbb{R}^{n \times p}$.

$\underset{\in \mathbb{R}^n}{y} = \underset{\in \mathbb{R}^n}{\hat{y}} + \underset{\in \mathbb{R}^n}{z}$ $\qquad \mathbb{X} = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_p \end{bmatrix}$

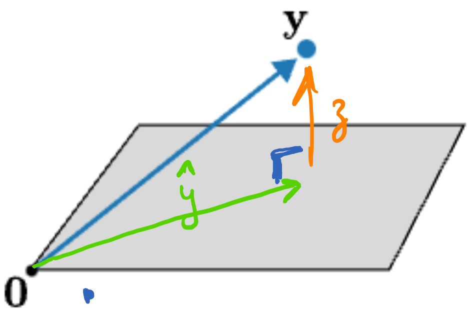
Which of the following notation/reasoning makes sense?

1. Define $\hat{y}$ as the **orthogonal projection** of $\vec{y}$ onto the column space of $\mathbb{X}$.

A. $\hat{y} \in \mathbb{R}^n$,	B. $\hat{y} \in \mathbb{R}^p$	C. Neither
A. $\hat{y} \perp \mathbb{R}^n$,	B. $\hat{y} \perp \mathbb{R}^p$	C. Neither
~~A.~~ $\hat{y} \perp \boldsymbol{y}$	~~B.~~ $\hat{y} \perp \mathbb{X}$ makes no sense	C. Neither
A. $\hat{y} \in \operatorname{span}\{\mathbb{X}\}$,	~~B.~~ $\hat{y} \perp \operatorname{span}\{\mathbb{X}\}$	C. Neither

2. Define $\vec{z} = \boldsymbol{y} - \hat{y}$:

A. $\vec{z} \in \mathbb{R}^n$,	B. $\vec{z} \in \mathbb{R}^p$	C. Neither
A. $\vec{z} \perp \mathbb{R}^n$,	B. $\vec{z} \perp \mathbb{R}^p$	C. Neither
A. $\vec{z} \perp \boldsymbol{y}$	B. $\vec{z} \perp \hat{y}$	C. Neither
~~A.~~ $\vec{z} \perp \boldsymbol{y}$	~~B.~~ $\vec{z} \perp \mathbb{X}$	C. Neither
~~A.~~ $\vec{z} \in \operatorname{span}\{\mathbb{X}\}$,	B. $\vec{z} \perp \operatorname{span}\{\mathbb{X}\}$	C. Neither

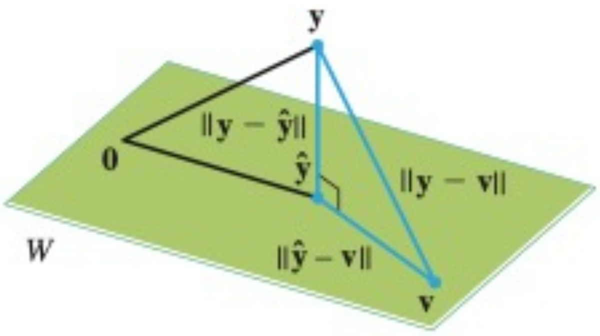


Column Spaces and Orthogonal Decompositions

span $\{\mathbb{X}\}$ is the **column space** of $\mathbb{X}$.

If $\mathbb{X} \in \mathbb{R}^{n \times p}$, then we can decompose $\vec{y} \in \mathbb{R}^{n}$ uniquely as $\boxed{\vec{y} = \hat{y} + (\vec{y} - \hat{y})}$ where

- $\hat{y} \in \text{span}\,\{\mathbb{X}\}$ is the orthogonal projection of y onto the column space of $\mathbb{X}$.
- $(\vec{y} - \hat{y}) \perp \text{span}\,\{\mathbb{X}\}$.



(Lay 6.3)

Column Spaces and Orthogonal Decompositions

span $\{\mathbb{X}\}$ is the **column space** of $\mathbb{X}$.

If $\mathbb{X} \in \mathbb{R}^{n \times p}$, then we can decompose $\vec{y} \in \mathbb{R}^n$ uniquely as $\boxed{\vec{y} = \hat{y} + (\vec{y} - \hat{y})}$ where

- $\hat{y} \in \text{span}\{\mathbb{X}\}$ is the orthogonal projection of y onto the column space of $\mathbb{X}$.
- $(\vec{y} - \hat{y}) \perp \text{span}\{\mathbb{X}\}$.
- $\hat{y}$ is the **best approximation** to $\vec{y}$ by elements of $\text{span}\{\mathbb{X}\}$.

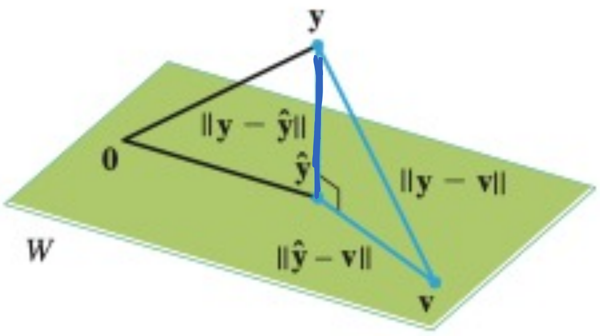


FIGURE 4 The orthogonal projection of **y** onto W is the closest point in W to **y**.

(Lay 6.3)

Proof by Numerical Example

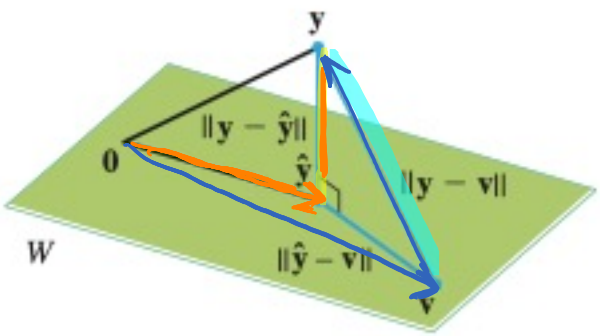


FIGURE 4 The orthogonal projection of **y** onto W is the closest point in W to **y**.

$$\boldsymbol{y} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, W = \text{span}\,\{A\}, \qquad \boldsymbol{v} = \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}$$

$$\hat{y} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$$ is orthogonal projection of $\boldsymbol{y}$ onto $\text{span}\,\{A\}$.

$$\|y - \hat{y}\| = \left\| \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} \right\| = \sqrt{4^2} = \sqrt{16} = 4$$

$$\|y - v\| = \left\| \begin{bmatrix} -2 \\ 2 \\ 4 \end{bmatrix} \right\| = \sqrt{2^2 + 2^2 + 4^2} = \sqrt{24}$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$$

for all **v** in W distinct from $\hat{\mathbf{y}}$.

General Proof

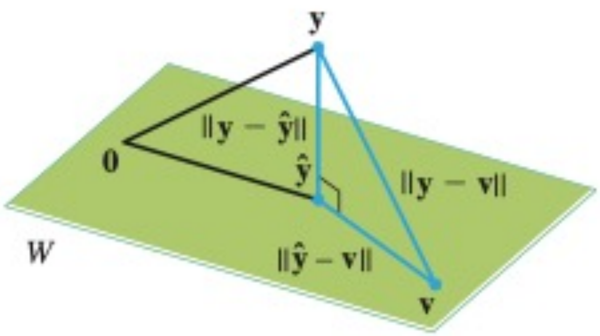


FIGURE 4 The orthogonal projection of $\mathbf{y}$ onto W is the closest point in W to $\mathbf{y}$.

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$$

for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

PROOF Take $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$. See Figure 4. Then $\hat{\mathbf{y}} - \mathbf{v}$ is in W. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W. In particular, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$ (which is in W). Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2$$

(See the colored right triangle in Figure 4. The length of each side is labeled.) Now $\|\hat{\mathbf{y}} - \mathbf{v}\|^2 > 0$ because $\hat{\mathbf{y}} - \mathbf{v} \neq \mathbf{0}$, and so inequality (3) follows immediately. ■

(Lay 6.3)

Lay 2.1

EECS 16A Lec3A

Matrix Inverses

Optional 02, Data 100 Spring 2023

Main Takeaways of Today

If $\mathbb{X} \in \mathbb{R}^{n \times p}$, then we can decompose $\vec{y} \in \mathbb{R}^n$ uniquely as $\vec{y} = \hat{y} + (\vec{y} - \hat{y})$, where

- $\hat{y} \in \text{span}\{\mathbb{X}\}$ is the **orthogonal projection** of y onto the **column space** of $\mathbb{X}$.
- $(\vec{y} - \hat{y}) \perp \text{span}\{\mathbb{X}\}$.
- $\hat{y}$ is the **best approximation** to $\vec{y}$ by elements of $\text{span}\{\mathbb{X}\}$.

For a square matrix in $\mathbb{R}^{n \times n}$,
full rank = full column rank = full row rank = n.

For a **square** matrix $A \in \mathbb{R}^{n \times n}$:

- Matrix A is **invertible**.
- A has linearly independent columns.
- A is **full rank**.
- The columns of A span $\mathbb{R}^n$.
- For every $b \in \mathbb{R}^n$, $A\vec{x} = \vec{b}$ has a **unique** solution $x = A^{-1}b$.

Invertible Operations

Invertible means we can recover input from output

Is $f(x)=0$ invertible? No!

Is eating a sandwich invertible? (not really...)

Is iPad scribbling invertible? Yes! (undo)

Example: inverse of a stapler

$$P = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \qquad P^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

P^{-1} is the **inverse** of P.

One way to see this is that for any vector $b \in \mathbb{R}^2$,

$b \longrightarrow_P Pb \longrightarrow_{P^{-1}} b$:

$$Pb = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.5b_1 \\ 0.5b_2 \end{bmatrix}$$

$$P^{-1}(Pb) = b$$

$$P^{-1}(Pb) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 0.5b_1 \\ 0.5b_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = b$$

Therefore, P is also the inverse of P^{-1}:

$$b \longrightarrow_{P^{-1}} \underbrace{P^{-1}b}_{\begin{bmatrix} 2b_1 \\ 2b_2 \end{bmatrix}} \longrightarrow_P b:$$

$$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} 2b_1 \\ 2b_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Defining the Matrix Inverse

- Definition: Let $P, Q \in \mathbb{R}^{N\times N}$ be square matrices.
  - P is the inverse of Q if $\boxed{PQ = QP = I}$

$I = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ & & 1 \\ 0 & & & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

Matrix multiply is generally not commutative (order matters), but inverses are!

We say that $P = Q^{-1}$ and $Q = P^{-1}$

Does commutative imply inverses? No!

$P = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ $\quad P^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

$P P^{-1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$P^{-1} P = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$b \xrightarrow[P]{} Pb \xrightarrow[P^{-1}]{} \underbrace{P^{-1}P}\, b = Ib = b$

Inverses are:

- Unique—either *the (one and only)* inverse exists, or it doesn't.
- Commutative—even though matrix multiplication typically isn't.

$P^{-1}P = PP^{-1}$

Finding the Matrix Inverse

$3 \cdot \frac{1}{3} = 1$

$$\mathbf{A}\vec{x} = \vec{b} \longrightarrow \vec{x} = \mathbf{A}^{-1}\vec{b}$$

$Ax = b$

$\underbrace{A^{-1}A}\, x = A^{-1} b$

$I x = A^{-1} b$

$x = A^{-1} b$

- We can use Gaussian Elimination (**Gauss-Jordan method**) to find the inverse of a square matrix
- Once we have the inverse, we can use it to solve system of equations

So matrix inverse is like division? Sort of, but matrix division doesn't technically exist

What if Ax=b has infinite solutions? No way to predict x from b, so A is not invertible

We will use NumPy to solve for matrix inverses (see lab).

Inverse Properties

Equivalent statements:

(full version in Lay 2.3)

For a **square** matrix $A \in \mathbb{R}^{n\times n}$:

- Matrix A is **invertible**.
- A has linearly independent columns.
- A is **full rank**.
- The columns of A span $\mathbb{R}^n$.
- For every $b \in \mathbb{R}^n$, $Ax = b$ has a **unique** solution $x = A^{-1}b$.

$$A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$$

$$Ax = b$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\begin{tikzcd} x \arrow[r, bend right, "A"'] & b \arrow[l, bend right, "A^{-1}"'] \end{tikzcd}$$

$$b = x_1 \vec{a}_1 + x_2 \vec{a}_2 + x_3 \vec{a}_3 + \cdots + x_n \vec{a}_n$$

Proof: Invertibility of Linear Transformations

Theorem: A is invertible, if and only if (iff) the columns of A are linearly independent. (unique sol'n)

1. If columns of A are lin. dep. then A^{-1} does not exist
2. If A^{-1} exists, then the cols. of A are linearly independent

What we know:

cols of A are lin. dep.

↳ $\exists \vec{\alpha} \neq \vec{0}$, s.t. $A\vec{\alpha} = \vec{0}$

there exists

some non-trivial combo of cols(A) → $\vec{0}$

To Show:

A^{-1} does not exist

Proof by contradiction: Assume A^{-1} exists

$\underbrace{A^{-1}A}\vec{\alpha} = A^{-1}\vec{0}$

$I\vec{\alpha} = \vec{0} \longrightarrow$ But $\vec{\alpha} \neq \vec{0}$! Hence A^{-1} does not exist!

Lay 2.9, 4.6

EECS 16A Lec3B

Rank

Optional 02, Data 100 Spring 2023

Main Takeaways of Today

If $\mathbb{X} \in \mathbb{R}^{n \times p}$, then we can decompose $\vec{y} \in \mathbb{R}^n$ uniquely as $\vec{y} = \hat{y} + (\vec{y} - \hat{y})$, where

- $\hat{y} \in \text{span}\{\mathbb{X}\}$ is the **orthogonal projection** of y onto the **column space** of $\mathbb{X}$.
- $(\vec{y} - \hat{y}) \perp \text{span}\{\mathbb{X}\}$.
- $\hat{y}$ is the **best approximation** to $\vec{y}$ by elements of $\text{span}\{\mathbb{X}\}$.

✅

For a square matrix in $\mathbb{R}^{n \times n}$,
full rank = full column rank = full row rank = n.

For a **square** matrix $A \in \mathbb{R}^{n \times n}$:

- Matrix A is **invertible**.
- A has linearly independent columns.
- A is **full rank**.
- The columns of A span $\mathbb{R}^n$.
- For every $b \in \mathbb{R}^n$, $Ax = b$ has a **unique** solution $x = A^{-1}b$.

✅ (almost)

Rank (1/2)

Column

- $A \in \mathbb{R}^{N\times M}$,$\text{Rank}\{A\} = \dim\{\text{Span}\{\text{cols}(A)\}\}$

$$A = \begin{bmatrix} a & 0 \\ 0 & b \\ 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} a & 0 & 0 \\ b & 0 & 0 \\ c & 0 & 0 \end{bmatrix} \quad Ax$$

2   2   1

For each of the above matrices, what is the rank?

Rank (1/2)

- $A \in \mathbb{R}^{N \times M}$,$\text{Rank}\{A\} = \dim\{\text{Span}\{\text{cols}(A)\}\}$

$$A = \begin{bmatrix} a & 0 \\ 0 & b \\ 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad A = \begin{array}{cc} N & \\ \begin{bmatrix} a & 0 & 0 \\ b & 0 & 0 \\ c & 0 & 0 \end{bmatrix} & M \end{array}$$

2 2 1

- $\boxed{\text{Rank}\{A\} = \dim\{\text{Span}\{\text{cols}(A)\}\} \leq min(M, N)}$

Can rank be larger than input dim(A)? No!

Rank (2/2)

- $A \in \mathbb{R}^{N \times M}$,$\text{Rank}\{A\} = \dim\{\text{Span}\{cols\{A\}\}\}$

$n \downarrow \overset{n}{\longrightarrow} \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$ $\vec{a}_i \in \mathbb{R}^n$

- $\text{Rank}\{A\} = \dim\{\text{Span}\{A\}\} \leq min(M, N)$

- $\text{Rank} = L$, mean the matrix $A \in \mathbb{R}^{N \times M}$ has L independent rows & columns

- $\boxed{\text{Rank}\{A\} + \dim\{\text{Null}\{A\}\} = M}$

Rank-Nullity Theorem

(Lay 2.9, 4.6 The Rank Theorem)

"Full rank" means rank is max possible (min(M,N))

For a square matrix in $\mathbb{R}^{n \times n}$,
full rank = full column rank = full row rank = n.

Main Takeaways of Today

If $\mathbb{X} \in \mathbb{R}^{n \times p}$, then we can decompose $\vec{y} \in \mathbb{R}^n$ uniquely as $\vec{y} = \hat{y} + (\vec{y} - \hat{y})$, where

- $\hat{y} \in \text{span}\{\mathbb{X}\}$ is the **orthogonal projection** of y onto the **column space** of $\mathbb{X}$.
- $(\vec{y} - \hat{y}) \perp \text{span}\{\mathbb{X}\}$.
- $\hat{y}$ is the **best approximation** to $\vec{y}$ by elements of $\text{span}\{\mathbb{X}\}$.

✅

For a square matrix in $\mathbb{R}^{n \times n}$,
full rank = full column rank = full row rank = n.

✅

For a **square** matrix $A \in \mathbb{R}^{n \times n}$:

- Matrix A is **invertible**.
- A has linearly independent columns.
- A is **full rank**.
- The columns of A span $\mathbb{R}^n$.
- For every $b \in \mathbb{R}^n$, $Ax = b$ has a **unique** solution $x = A^{-1}b$.

Restated: If $\text{rank}(A) = n$, then A is invertible.

Main Takeaways of Today

If $\mathbb{X} \in \mathbb{R}^{n\times p}$, then we can decompose $\vec{y} \in \mathbb{R}^n$ uniquely as $\vec{y} = \hat{y} + (\vec{y} - \hat{y})$, where

- $\hat{y} \in \mathrm{span}\,\{\mathbb{X}\}$ is the **orthogonal projection** of y onto the **column space** of $\mathbb{X}$.
- $(\vec{y} - \hat{y}) \perp \mathrm{span}\,\{\mathbb{X}\}$.
- $\hat{y}$ is the **best approximation** to $\vec{y}$ by elements of $\mathrm{span}\,\{\mathbb{X}\}$.

For a square matri ... n,
full rank = full co ... = full row ra

For a **square** m

- Matrix A i ... e.
- A has lin ... ndependent colum ... .
- A is **full rank**.
- The columns of A span $\mathbb{R}^n$.
- For every $b \in \mathbb{R}^n$, $Ax = b$ has a **unique** solution $x = A^{-1}b$.

Lay 2.1, 2.3

Proof Practice

Optional 02, Data 100 Spring 2023

Practice

In Exercises 9 and 10, mark each statement True or False. Justify each answer.

(Lay 2.2)

9. a. In order for a matrix B to be the inverse of A, both equations $AB = I$ and $BA = I$ must be true. T

b. If A and B are $n \times n$ and invertible, then $A^{-1}B^{-1}$ is the inverse of AB. F

c. If A is invertible, then the inverse of A^{-1} is A itself. T

For a **square** matrix $A \in \mathbb{R}^{n\times n}$:

- Matrix A is **invertible**.
- A has linearly independent columns.
- A is **full rank**.
- The columns of A span $\mathbb{R}^n$.
- For every $b \in \mathbb{R}^n$, $A\vec{x} = \vec{b}$ has a **unique** solution $x = A^{-1}b$.

(a) Let $B = A^{-1}$

$AB = AA^{-1} = A^{-1}A = I$

$BA = A^{-1}A = I$

(b) If $A^{-1}B^{-1}$ is inverse of AB, then $(A^{-1}B^{-1})AB \overset{?}{=} I$

$A^{-1}\underbrace{B^{-1}A}B$

seems fishy

$B^{-1}A^{-1}$ is inverse of AB

$(B^{-1}A^{-1})AB \overset{?}{=} I$

$B^{-1}\underbrace{A^{-1}A}B$

$B^{-1}IB$

$B^{-1}B = I$

(c) $A^{-1}A = I$

$AA^{-1} = I$ ✓

15. Can a square matrix with two identical columns be invertible? Why or why not? (Lay 2.3) No

16. Is it possible for a 5×5 matrix to be invertible when its columns do not span $\mathbb{R}^5$? Why or why not?

(15) 2 identical cols

- lin indep cols does not hold
- A is not invertible

$A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_p & \vec{a}_p \end{bmatrix}$

$\mathbb{R}^{n\times(p+1)}$

(16) Cols of A span $\mathbb{R}^n$ does not hold

- A is not invertible

Practice

17. If A is invertible, then the columns of A^{-1} are linearly independent. Explain why. T

(Lay 2.3)

19. If the columns of a 7×7 matrix D are linearly independent, what can you say about solutions of $D\mathbf{x} = \mathbf{b}$? Why?

20. If $n \times n$ matrices E and F have the property that $EF = I$, then E and F commute. Explain why.

For a **square** matrix $A \in \mathbb{R}^{n \times n}$:

- Matrix A is **invertible**.
- A has linearly independent columns.
- A is **full rank**.
- The columns of A span $\mathbb{R}^n$.
- For every $b \in \mathbb{R}^n$, $A\vec{x} = \vec{b}$ has a **unique** solution $x = A^{-1}b$.

(17)
- A is invertible $\in \mathbb{R}^{n \times n}$
  ↳ A^{-1} must exist → $AA^{-1} = I$, $A^{-1}A = I$
- ↳ A^{-1} invertible $\in \mathbb{R}^{n \times n}$
- A^{-1} has lin independent columns

(19)
- D lin independent cols
- D is invertible
- $Dx = b$ has a unique solution $x = D^{-1}b$

(20)
- $EF = I$ (1)
- E is invertible
- Inverse of E is F

- F is invertible
- Inverse of F is E → $FE = I$ (2)

$EF = I = FE$

$EF = FE$ ∎

OPTIONAL 02

Linear Algebra 02

Content credit: The wealth of instructional knowledge from Math 54, Math 91, and EECS16A