EECS 245, Spring 2026

LEC 8

Linear Transformations, Inverses, and Projections

→ Read: Ch. 6

Agenda

More videos, please watch!

- Big idea 1: deeply understand the conditions under which a matrix is invertible
- What is an inverse?
- Linear transformations, determinants
- Big idea 2: projecting onto the column space

how we will get back to regression!

Announcements

- HW 5 due tonight
- HW 6 due Sunday
- Lab 6 solutions up
- Check out MT 1 scores + Grade Report on Gradescope

Suppose A is an invertible matrix.
What does that mean?

- → A must be square, $n \times n$
- → There exists a unique matrix, A^{-1}, such that

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

- → $\text{rank}(A) = n$
- → A's cols are linearly independent

Example: 2 x 2 matrix

Inverse of a 2x2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is

$$A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

determinant of A, det(A)

e.g. $A = \begin{bmatrix} 1 & 5 \\ 2 & 12 \end{bmatrix}$

$$A^{-1} = \frac{1}{2} \begin{bmatrix} 12 & -5 \\ -2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & -5/2 \\ -1 & 1/2 \end{bmatrix}$$

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

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

$$\begin{bmatrix} 1 & 5 \\ 2 & 12 \end{bmatrix}$$

if this was a 10 instead of a 12, inverse doesn't exist!

observation:
$\det(A) = 0 \rightarrow A$ not invertible

$$A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\frac{1}{10-10}$$

determinant of A, det(A)

What if we consider a 3x3 matrix?

$$\underbrace{\begin{bmatrix} 5 & 7 & 3 \\ 7 & 1 & 4 \\ 0 & 2 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}}_{A^{-1}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{I}$$

$$5a + 7d + 3g = 1$$

$$7a + f + 4i = 0$$

$$\vdots$$

could solve for A^{-1} this way

Linear transformation : a type of function

Suppose T is a function.
T is a linear transformation if it satisfies these 2 properties:

① $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$ for all $\vec{x}, \vec{y} \in \text{domain}(T)$

② $T(c\vec{x}) = c\,T(\vec{x})$ for all $\vec{x} \in \text{domain}(T)$, $c \in \mathbb{R}$

① + ② $T(a\vec{x}+b\vec{y}) = aT(\vec{x}) + bT(\vec{y})$

Example: $f(x) = 2x+5$

Is this a linear transformation?

No!!!!!

this is an "affine" transformation

For it to be an LT, it must be the case that

$$cf(x) = f(cx) \text{ for all } c, x$$

but, e.g. if $c=0$,

$$0f(x) = 0$$
$$f(0x) = f(0) = 5$$

but $0 \neq 5$

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ 2x_1 \\ 3x_1 - 4x_2 \end{bmatrix}_{3\times 1}$$

$$= \underbrace{\begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & -4 \end{bmatrix}_{3\times 2}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{2\times 1} = A\vec{x}$$

$$T(\vec{x}) = A\vec{x}$$

$$T(\vec{x}+\vec{y})$$

$$= A(\vec{x}+\vec{y})$$

$$= A\vec{x} + A\vec{y}$$

Big idea : All linear transformations are matrix-vector multiplications!

$$T(\vec{x}) = A_{n \times d}\, \vec{x}_{d \times 1}$$

$$T : \underbrace{\mathbb{R}^d}_{\text{domain}} \longrightarrow \underbrace{\mathbb{R}^n}_{\text{codomain}}$$

If $d=n$, meaning A square,

$T: \mathbb{R}^n \to \mathbb{R}^n$

$\Rightarrow$ focus on square matrices

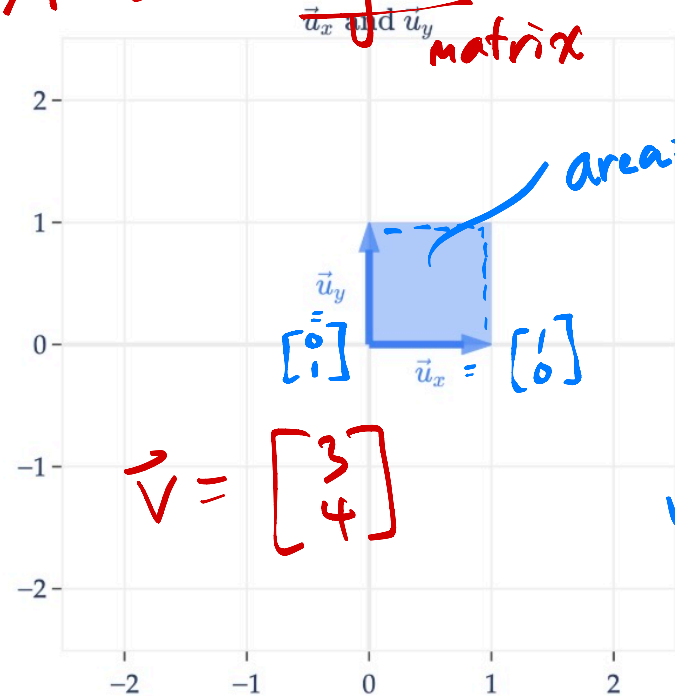
$f(\vec{v}) = A\vec{v}$

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

$A\vec{u}_x = A\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$

$A\vec{u}_y = \begin{bmatrix}0\\-1/3\end{bmatrix}$

A is a diagonal matrix

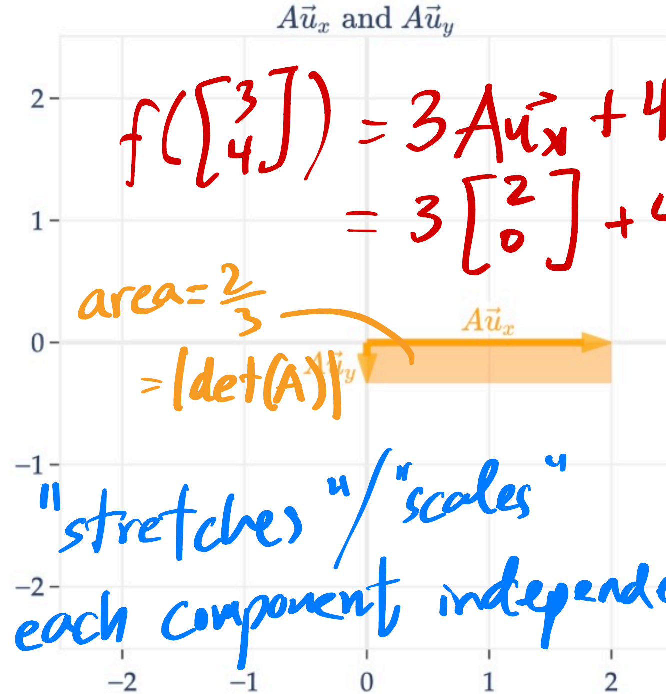


area = 1

$\vec{u}_y = \begin{bmatrix}0\\1\end{bmatrix}$ $\quad \vec{u}_x = \begin{bmatrix}1\\0\end{bmatrix}$

$\vec{v} = \begin{bmatrix}3\\4\end{bmatrix}$

A



$$f\left(\begin{bmatrix}3\\4\end{bmatrix}\right) = 3A\vec{u}_x + 4A\vec{u}_y = 3\begin{bmatrix}2\\0\end{bmatrix} + 4\begin{bmatrix}0\\-1/3\end{bmatrix} = \begin{bmatrix}6\\-4/3\end{bmatrix}$$

area $= \frac{2}{3} = |\det(A)|$

"stretches"/"scales" each component independently!

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -1/3 \end{bmatrix} \qquad A' = \begin{bmatrix} 0 & 2 \\ -1/3 & 0 \end{bmatrix}$$

$$\det(A) = -\frac{2}{3} \qquad \det(A') = \frac{2}{3}$$

Suppose $\vec{v} \in \mathbb{R}^2$, e.g. $\vec{v} = \begin{bmatrix} c \\ d \end{bmatrix}$

Note that $\vec{v} = c\vec{u}_x + d\vec{u}_y$

$\vec{u}_x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$\vec{u}_y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

what is $f(\vec{v})$?

$$\begin{aligned} f(\vec{v}) &= A\vec{v} \\ &= A\left(c\vec{u}_x + d\vec{u}_y\right) \\ &= c\left(A\vec{u}_x\right) + d\left(A\vec{u}_y\right) \end{aligned}$$

Key idea: $A\vec{v}$ is a linear combination of $A\vec{u}_x$ and $A\vec{u}_y$!

B "rotates"

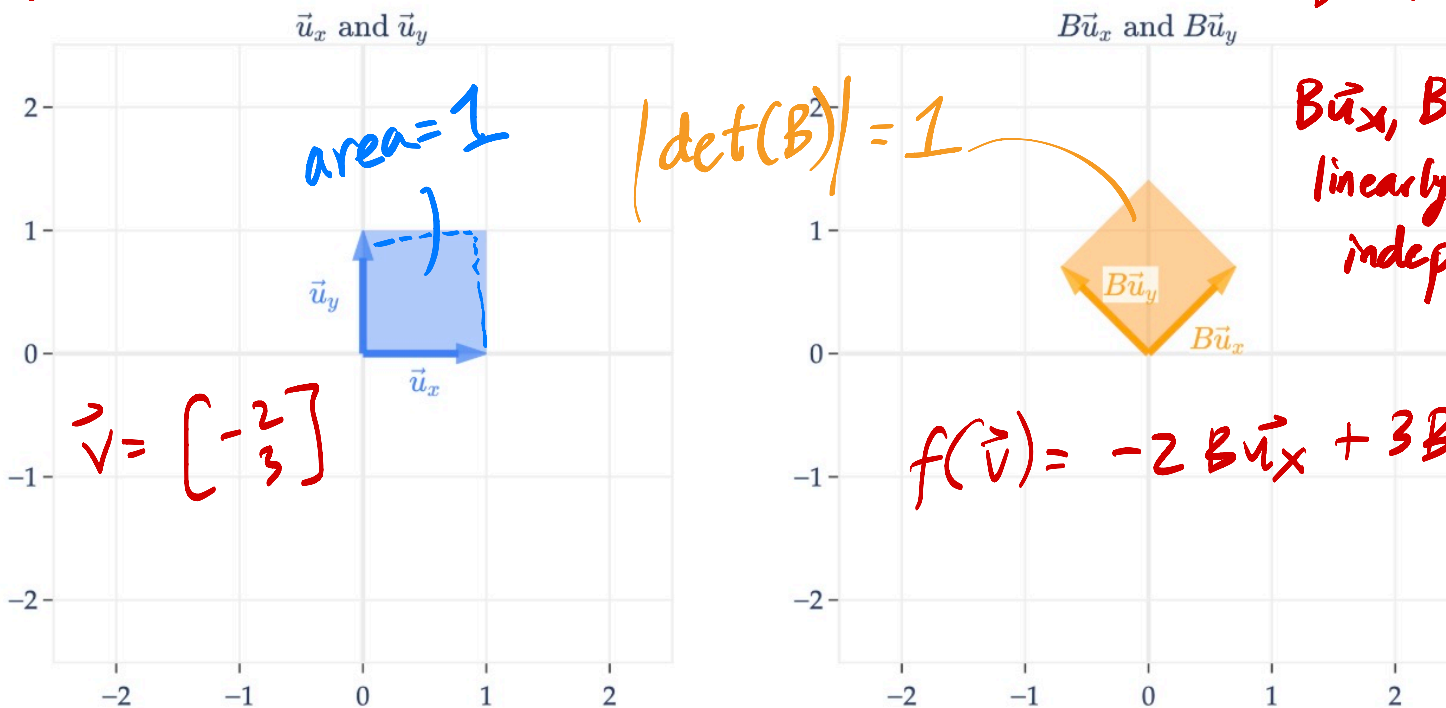
$\|B\vec{v}\| = \|\vec{v}\|$

$$B = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

$B^TB = BB^T = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

"orthogonal" matrix

$B^{-1} = B^T$



area = 1

$|\det(B)| = 1$

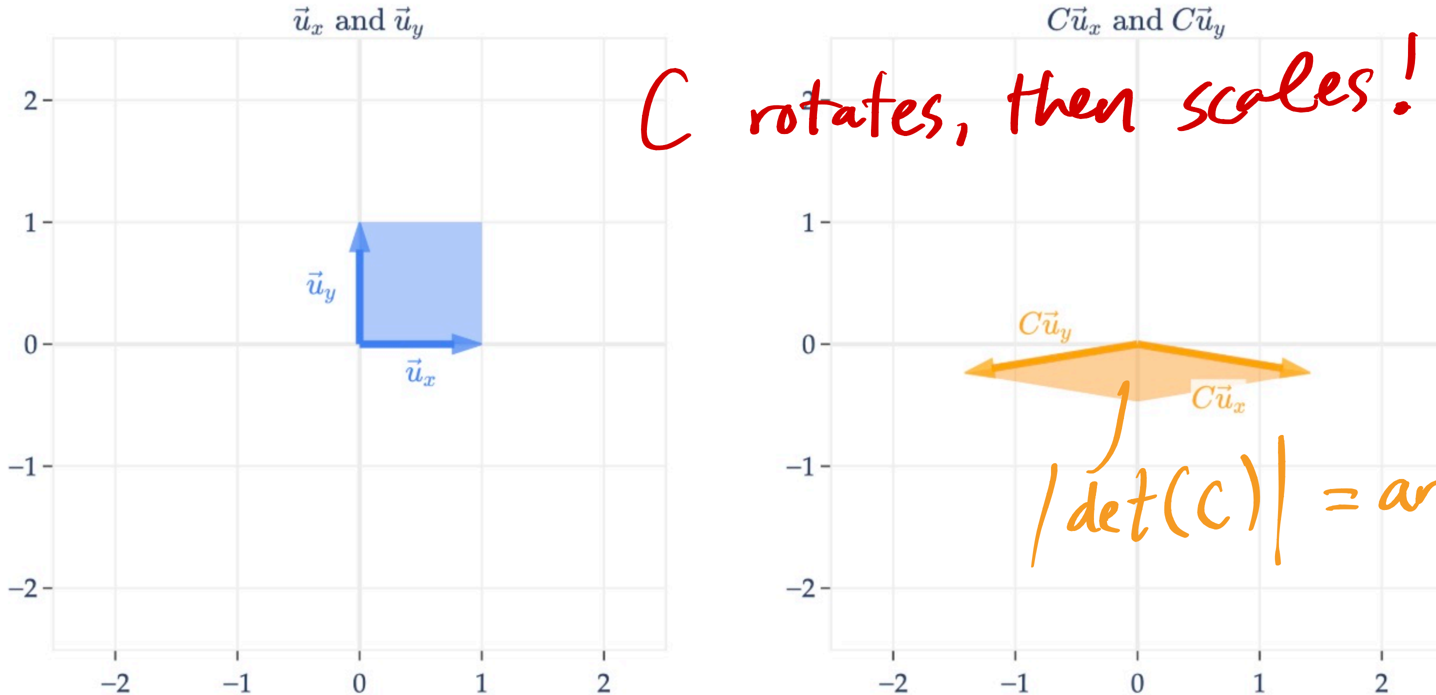
$B\vec{u}_x, B\vec{u}_y$ linearly independent!

$\vec{v} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$

$f(\vec{v}) = -2B\vec{u}_x + 3B\vec{u}_y$

$$C = AB = \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & -1/3 \end{bmatrix}}_{\text{scale}} \underbrace{\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}}_{\text{rotate}} \vec{v}$$

C rotates, then scales!



$|\det(C)| = \text{area}$

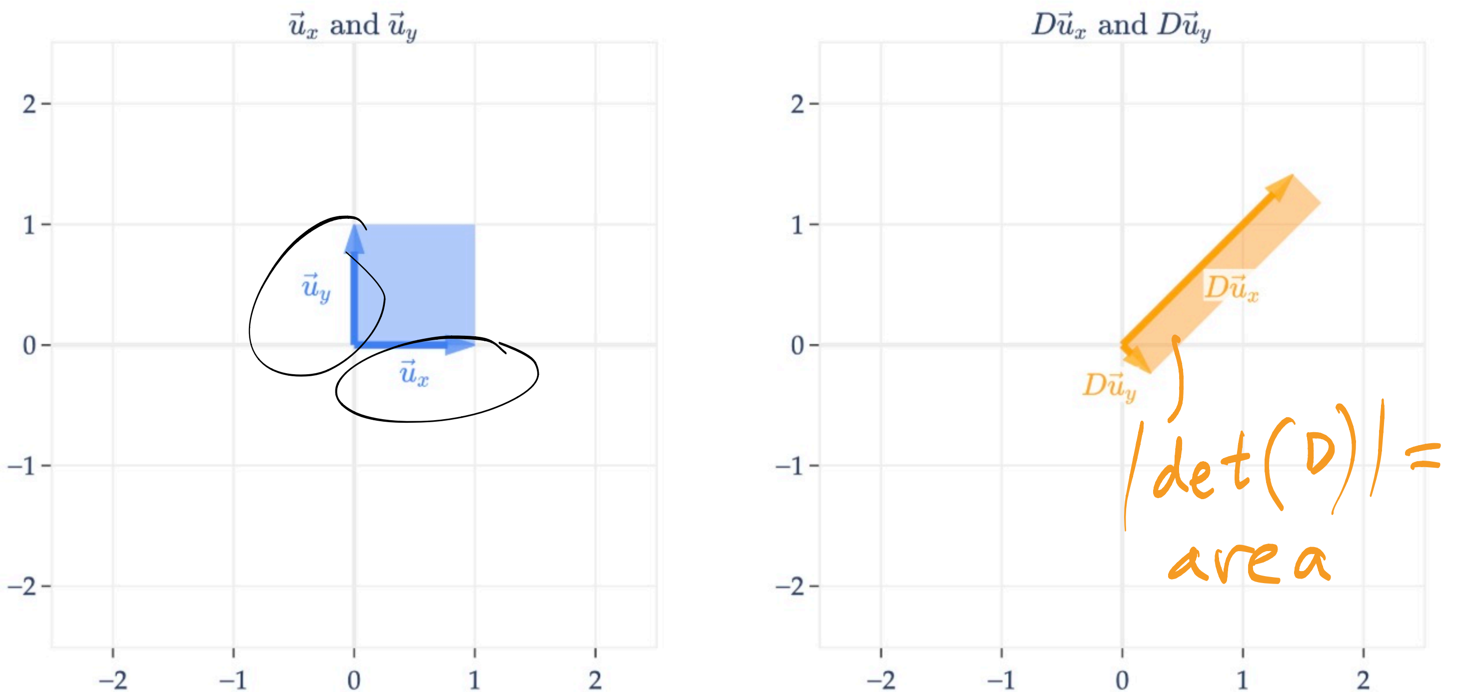
$C = AB$ is different from

$D\vec{v} = \underline{BA\vec{v}}$

A scales

B rotates

$$D = \begin{bmatrix} \sqrt{2} & \sqrt{2}/6 \\ \sqrt{2} & -\sqrt{2}/6 \end{bmatrix} = BA$$

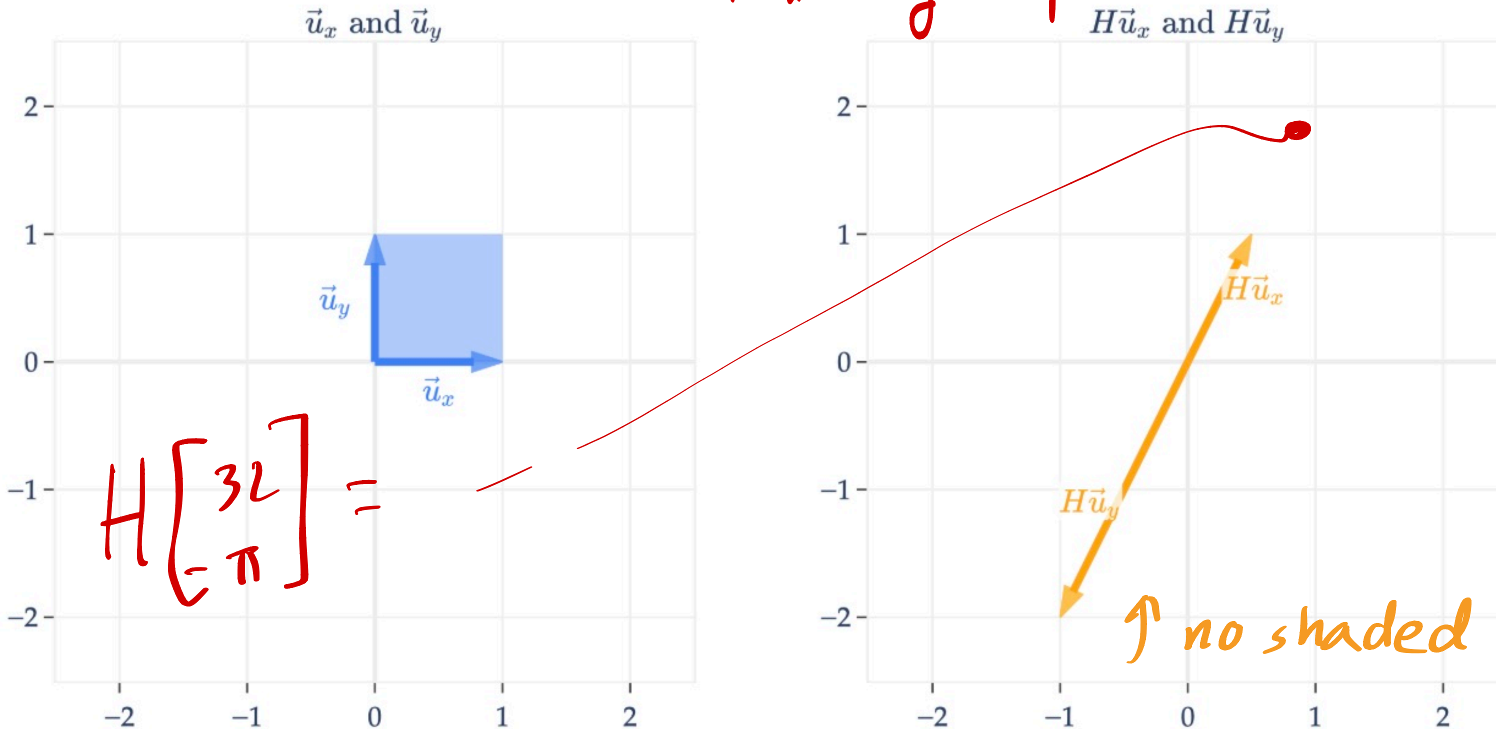


$|\det(D)| =$ area

$H\vec{v}$

$$\underbrace{H = \begin{bmatrix} 1/2 & -1 \\ 1 & -2 \end{bmatrix}}_{\text{not linearly independent!}}$$

colsp(H) = span of H's cols = a line!

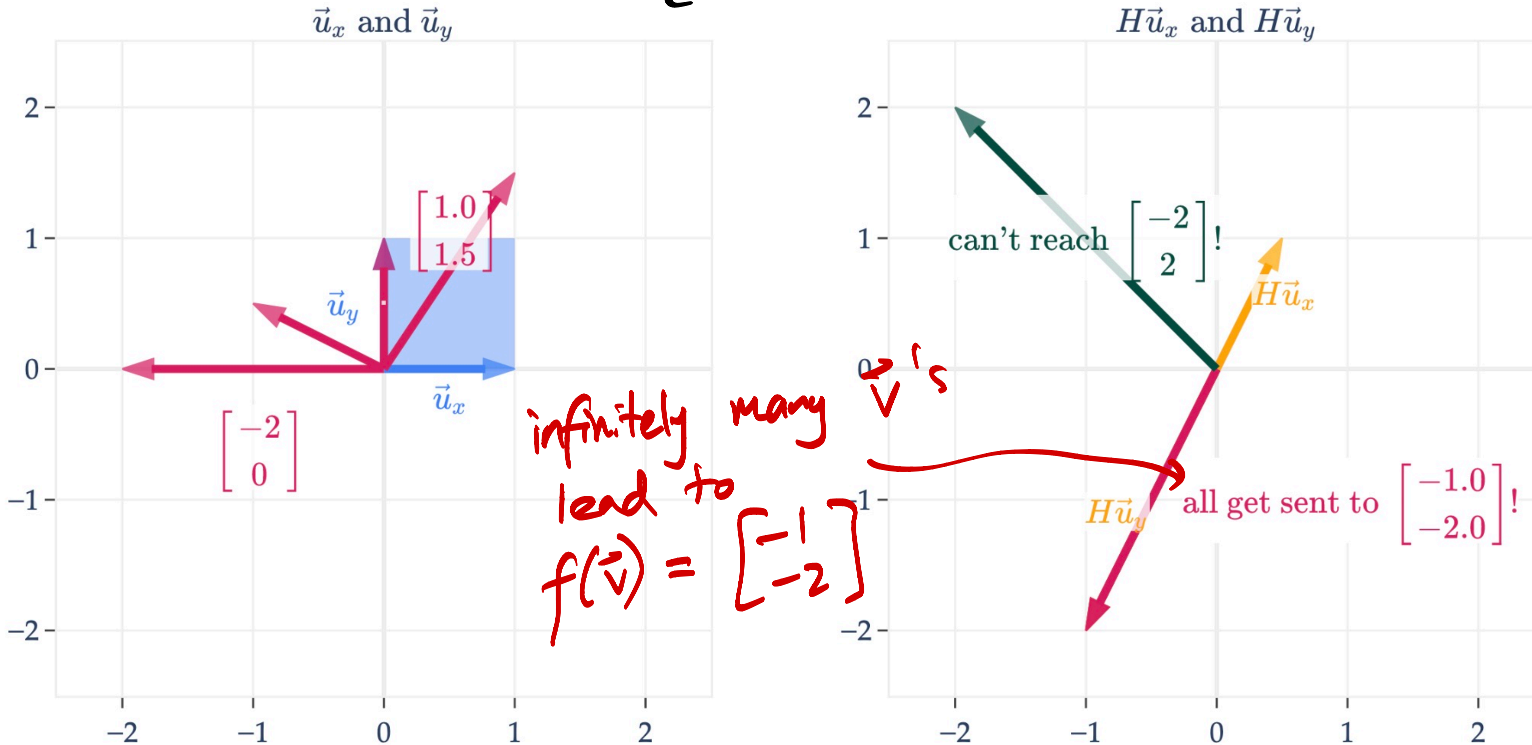


$H\begin{bmatrix} 32 \\ -\pi \end{bmatrix} =$

no shaded area!

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

$$f(\vec{v}) = H\vec{v}$$



infinitely many $\vec{v}$'s lead to $f(\vec{v}) = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$

Big idea: $n \times n$ matrix A is invertible

if and only if

the transformation

$$T(\vec{x}) = A\vec{x}$$

is invertible

→ in other words, for every $\vec{y} \in \mathbb{R}^n$, there is exactly one $\vec{x}$ such that

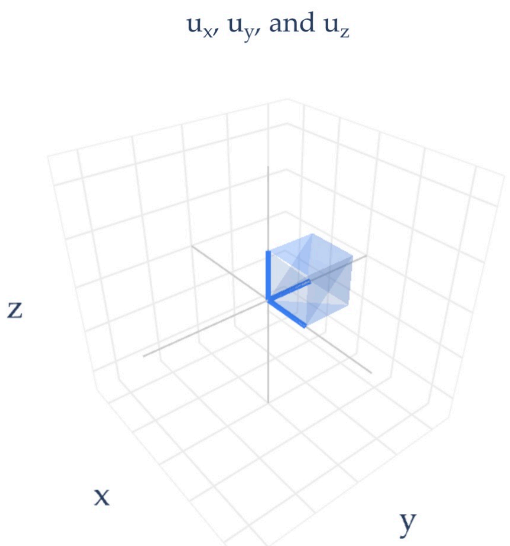
$$T(\vec{x}) = A\vec{x} = \vec{y}$$

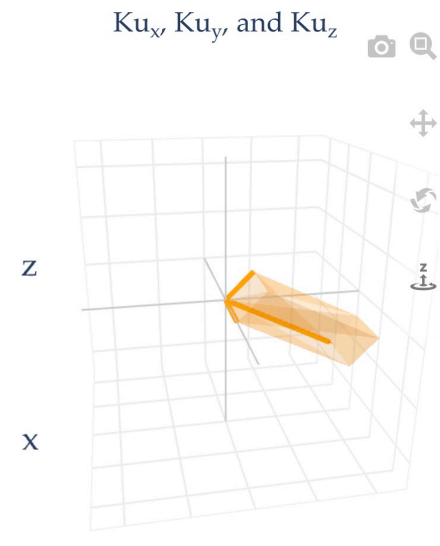
Determinants:

- swap 2 rows, det multiplies by -1
- swap 2 cols, det multiplies by -1
- if A is 2×2, $|\det(A)|$ is area enclosed by the parallelogram formed by A's columns ($A\vec{u}_x$ and $A\vec{u}_y$)
- 3×3: $|\det(A)|$ measures volume of parallelepiped

the unit cube to a parallelepiped (the generalization of a parallelogram to three dimensions).

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





for an arbitrary $n \times n$, $|\det(A)|$ measures n-dimensional "volume"

Recap: when is an $n \times n$ matrix invertible?

A is invertible if and only if:

all 7 are equivalent!

① there exists A^{-1} such that $AA^{-1} = A^{-1}A = I$

② $\det(A) \neq 0$

③ $\operatorname{rank}(A) = n$

④ A has n linearly independent columns

⑤ A has n linearly independent rows

⑥ $\operatorname{nullsp}(A) = \{\vec{0}\}$ "trivial null space"

⑦ For every $\vec{b} \in \mathbb{R}^n$, there is exactly one $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{b}$

$A\vec{x}$ linear comb. of A's cols!

$$A_{n\times n}\vec{x}_{n\times 1} = \vec{b}_{n\times 1}$$

System of equations : n unknowns $(x_1, x_2, \ldots, x_n)$

n equations (one per row of A)

Q : What is the solution to the system of eq'ns?

A : If A is invertible, then

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

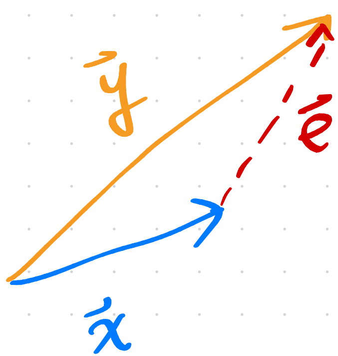
$$\vec{x} = A^{-1}\vec{b}$$

big idea: if A invertible, there is exactly one linear combination of A's columns that makes $\vec{b}$.

Practical consideration : actually finding A^{-1} is expensive computationally, so we prefer to solve systems directly

→ Ch. 6.2

Big idea 2



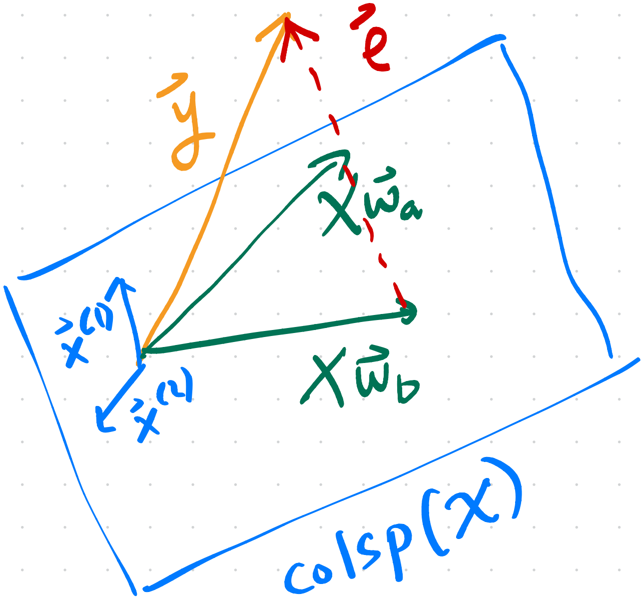
Projection of $\vec{y}$ onto $\vec{x}$

$$\vec{p} = \underbrace{\left(\frac{\vec{x}\cdot\vec{y}}{\vec{x}\cdot\vec{x}}\right)}_{k^*} \vec{x}$$

Why? This vector minimizes

$$\|\vec{e}\|^2 = \|\vec{y} - k\vec{x}\|^2$$

Recall: the shortest possible $\vec{e}$ is orthogonal to $\vec{x}$



$$X_{n \times d} = \begin{bmatrix} | & | & & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} & \cdots & \vec{x}^{(d)} \\ | & | & & | \end{bmatrix}$$

$\vec{y} \in \mathbb{R}^n$

Remember, $X\vec{w}$ is a linear combination of X's columns!

$$X\vec{w} = w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)} + \dots + w_d \vec{x}^{(d)}$$

= set of all linear combinations of X's columns

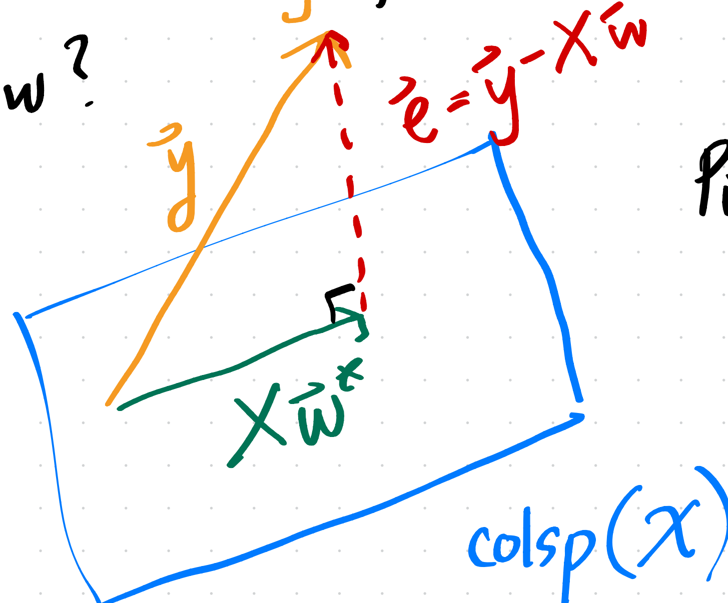
= set of possible values of $\underline{X\vec{w}}$, $\vec{w} \in \mathbb{R}^d$

Goal: Minimize $\|\vec{e}\|^2 = \|\vec{y} - X\vec{w}\|^2$

function of $\vec{w}$ only!

Assume $X, \vec{y}$ fixed; we can only control $\vec{w}$

How?



Pick $\vec{w}$ such that $\vec{y} - X\vec{w}$ is orthogonal to <u>every vector in colsp(X)</u>!

How do we find the $\vec{w}$ such that $\vec{y} - X\vec{w}$ orthogonal to colsp(X)?

$$X_{n \times d} = \begin{bmatrix} | & | & & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} & \cdots & \vec{x}^{(d)} \\ | & | & & | \end{bmatrix}_{n \times d}$$

$$\vec{x}^{(i)} \in \mathbb{R}^n$$

How? Find $\vec{w}$ so that $\vec{y} - X\vec{w}$ orthogonal to each col of X

$$\vec{x}^{(1)} \cdot (\vec{y} - X\vec{w}) = \vec{0}$$

$$\vec{x}^{(2)} \cdot (\vec{y} - X\vec{w}) = \vec{0}$$

$$\vdots$$

d equations

Aside:

$$A\vec{x}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 7 & 12 & 4 \\ 1 & 1 & 0 \\ 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} 7 \\ 0 \\ 4 \end{bmatrix}$$

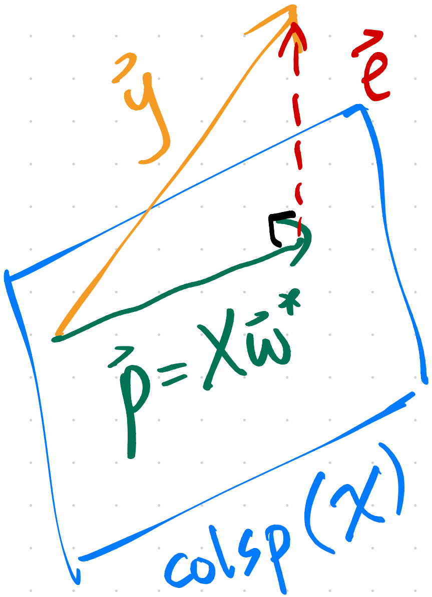
Two interpretations:

1. dot product of the rows of A with $\vec{x}$
2. linear combination of A's columns, with coefficients from $\vec{x}$

Idea: what if we compute X^T?

$$X^T_{d\times n} = \begin{bmatrix} — \vec{x}^{(1)T} — \\ — \vec{x}^{(2)T} — \\ \vdots \\ — \vec{x}^{(d)T} — \end{bmatrix}$$

$$X^T_{d\times n}(\vec{y} - X\vec{w})_{n\times 1} = \begin{bmatrix} \vec{x}^{(1)T}(\vec{y} - X\vec{w}) \\ \vec{x}^{(2)T}(\vec{y} - X\vec{w}) \\ \vdots \\ \vec{x}^{(d)T}(\vec{y} - X\vec{w}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{0}_d$$



All we need to do now is find $\vec{w}$ so that:

$$X^T(\underbrace{\vec{y} - X\vec{w}}_{\text{error vector}}) = \vec{0}$$

This will give us the error vector with the shortest possible length!

$$X^T(\vec{y} - X\vec{w}) = \vec{0}$$

Remember, $\vec{w} \in \mathbb{R}^d$

$$X^T\vec{y} - X^TX\vec{w} = \vec{0}$$

$$X^TX\vec{w} = X^T\vec{y}$$

"normal equation(s)"

Case 1: X^TX is invertible (happens when X's cols are linearly independent)

$$\underbrace{(X^TX)^{-1}X^TX}_{I}\vec{w} = (X^TX)^{-1}X^T\vec{y}$$

unique solution

$$\vec{w}^* = (X^TX)^{-1}X^T\vec{y}$$

most important equation of the semester

When is X^TX invertible?

$X : n \times d$ $\qquad X^TX : d \times d$

Hugely important fact: for any X, $\text{rank}(X) = \text{rank}(X^TX)$!

proof: see video

For X^TX to be invertible,

$\Rightarrow$ $\text{rank}(X^TX)$ must be d

$\Rightarrow$ so $\text{rank}(X)$ must be d too

$\Rightarrow$ so X's columns must be linearly independent!

Example

$$X = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 1 \\ 0 & -1 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 0 \\ 4 \\ 5 \end{bmatrix}$$

Q: which linear combination of X's cols is closest to $\vec{y}$?

A: $X\vec{w}^*$, where $\vec{w}^* = (X^TX)^{-1}X^T\vec{y}$

$$\underbrace{X^TX = \begin{bmatrix} 6 & 3 \\ 3 & 3 \end{bmatrix}}_{\text{contains dot products of cols}} \qquad \ldots\ldots \vec{w}^* = \begin{bmatrix} 2 \\ -8/3 \end{bmatrix}$$

of all linear combinations of X's columns,

the vector closest to $\vec{y}$ is

$$X\vec{w}^* = \underbrace{(2)}\underbrace{\begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}}_{\vec{x}^{(1)}} + \left(-8/3\right)\underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix}}_{\vec{x}^{(2)}} = \vec{p}$$

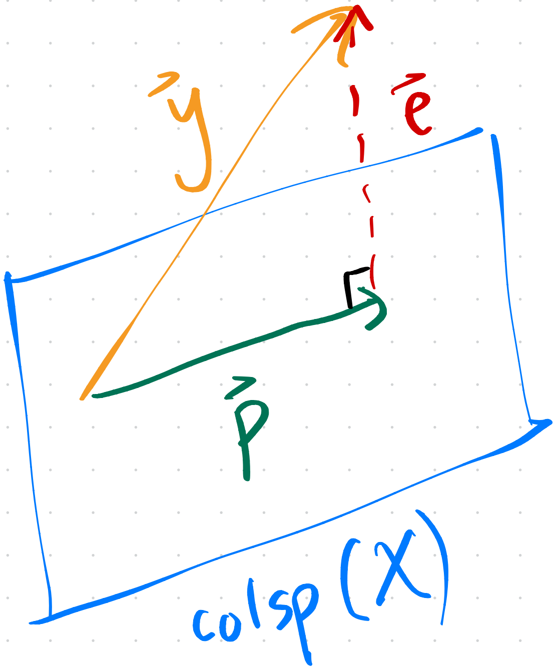
orthogonality: $\underbrace{\vec{p}}_{\text{in colsp}(X)}$ orthogonal to $\vec{y} - \vec{p}$

$$X^T X \vec{w} = X^T \vec{y}$$ "normal equations"

→ In Case 1, when X's cols were linearly independent, $X^T X$ was invertible, and there was a unique $\vec{w}^*$

→ Case 2: X's cols are linearly dependent, so $X^T X$ not invertible, then ∞ many $\vec{w}^*$'s that satisfy $X^T X \vec{w} = X^T \vec{y}$!

infinitely many solutions to the normal equations



... but even if X's cols are linearly dependent, there is still only one best $\vec{p}$

- there are just infinitely many ways to describe $\vec{p}$ as a linear combination of X's columns