

Spans and projections

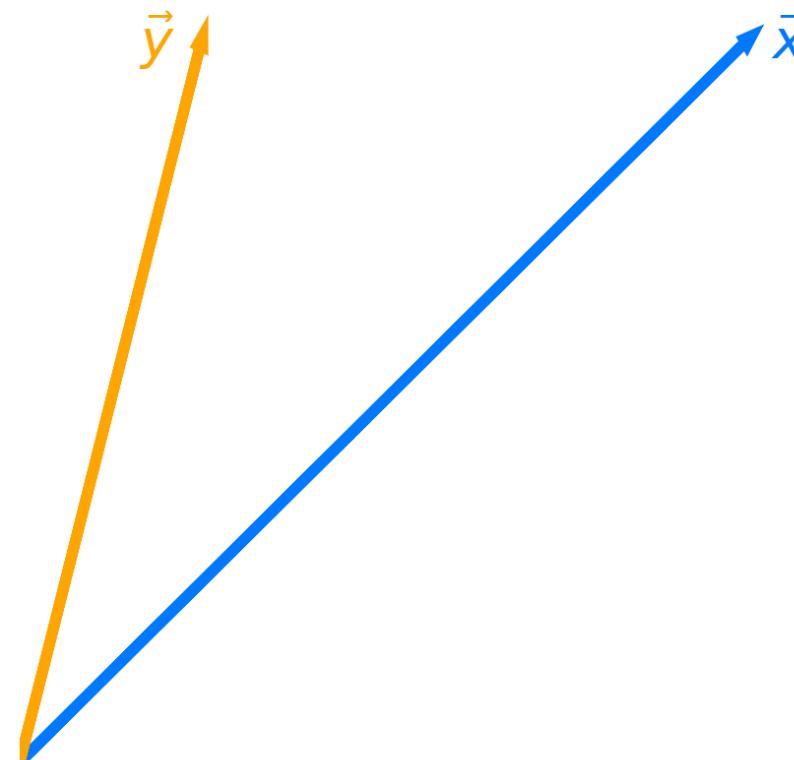
Projecting onto a single vector

- Let \vec{x} and \vec{y} be two vectors in \mathbb{R}^n .
- The span of \vec{x} is the set of all vectors of the form:

$$w\vec{x}$$

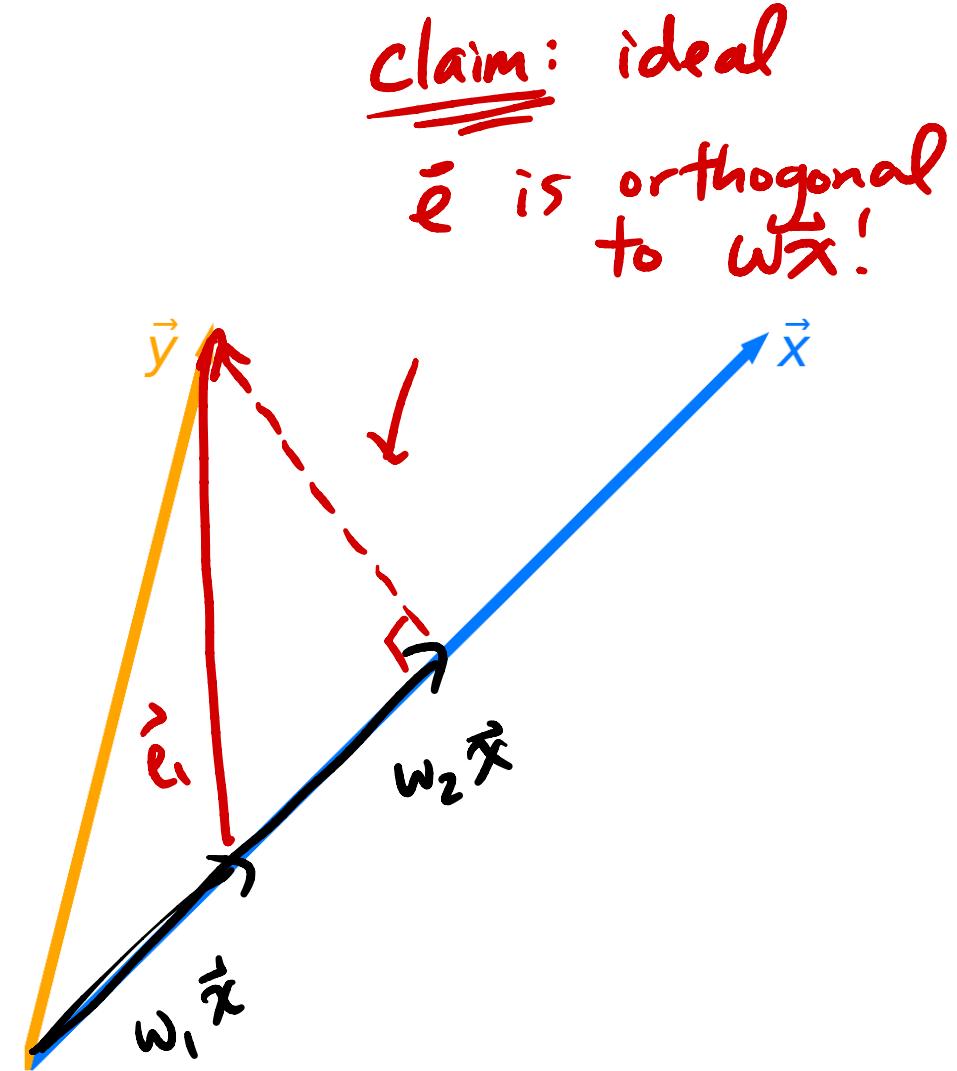
where $w \in \mathbb{R}$ is a scalar.

- **Question:** What vector in $\text{span}(\vec{x})$ is closest to \vec{y} ?
- The vector in $\text{span}(\vec{x})$ that is closest to \vec{y} is the _____
projection of \vec{y} onto $\text{span}(\vec{x})$.



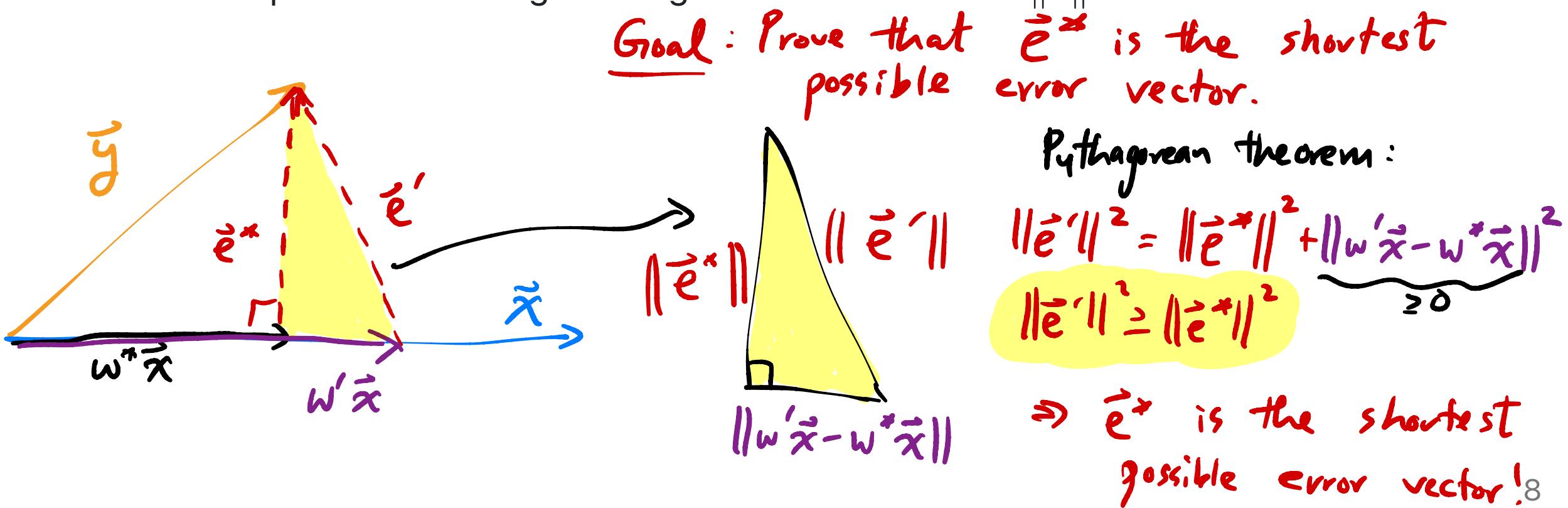
Projection error

- Let $\vec{e} = \vec{y} - w\vec{x}$ be the **projection error**: that is, the vector that connects \vec{y} to $\text{span}(\vec{x})$.
- Goal:** Find the w that makes \vec{e} as short as possible.
 - That is, minimize:
$$\|\vec{e}\|$$
 - Equivalently, minimize:
$$\|\vec{y} - w\vec{x}\|$$
- Idea:** To make \vec{e} has short as possible, it should be **orthogonal** to $w\vec{x}$.



Minimizing projection error

- Goal: Find the w that makes $\vec{e} = \vec{y} - w\vec{x}$ as short as possible.
- Idea: To make \vec{e} as short as possible, it should be orthogonal to $w\vec{x}$.
- Can we prove that making \vec{e} orthogonal to $w\vec{x}$ minimizes $\|\vec{e}\|$?



Minimizing projection error

- Goal: Find the w that makes $\vec{e} = \vec{y} - w\vec{x}$ as short as possible.
- Now we know that to minimize $\|\vec{e}\|$, \vec{e} must be orthogonal to $w\vec{x}$.
- Given this fact, how can we solve for w ?

$$\vec{e} \text{ orthogonal to } w\vec{x} \Rightarrow$$

$$w\vec{x} \cdot \vec{e} = 0$$

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

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

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

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

$$\vec{x} \cdot \vec{y} = w(\vec{x} \cdot \vec{x})$$

$$\Rightarrow w = \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}}$$

The w that makes the error vector as short as possible!!!

Orthogonal projection

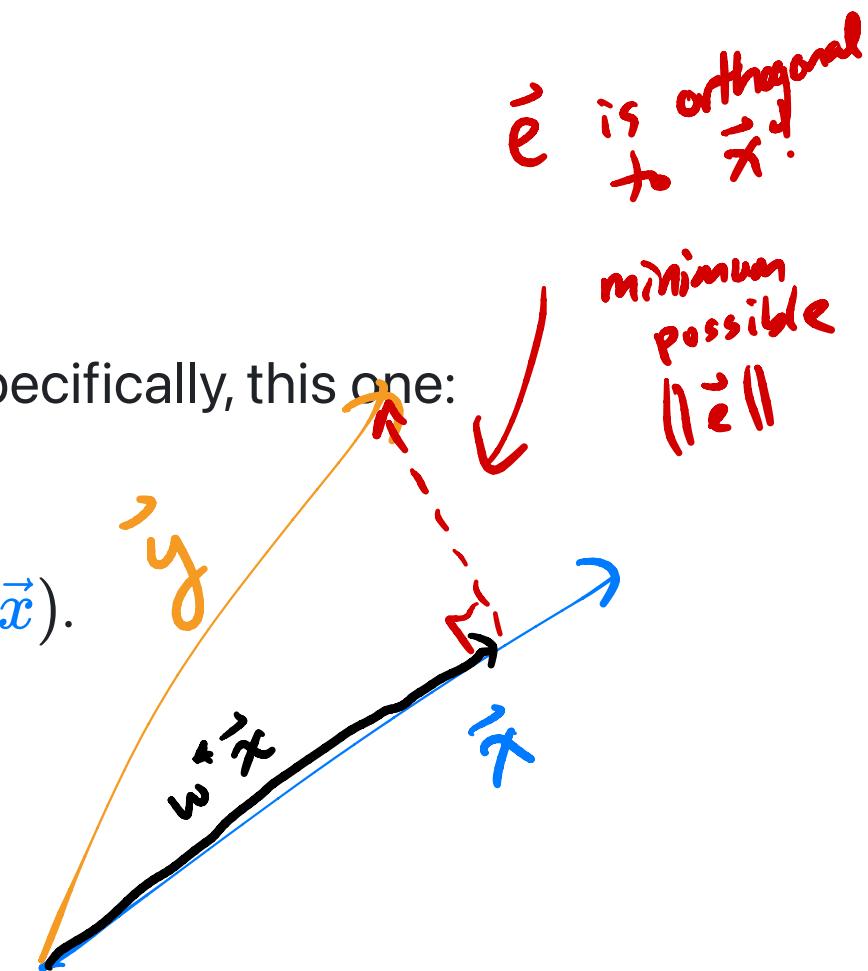
- Question: What vector in $\text{span}(\vec{x})$ is closest to \vec{y} ?
- Answer: It is the vector $w^* \vec{x}$, where:

$$w^* = \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}}$$

- Note that w^* is the solution to a minimization problem, specifically, this one:

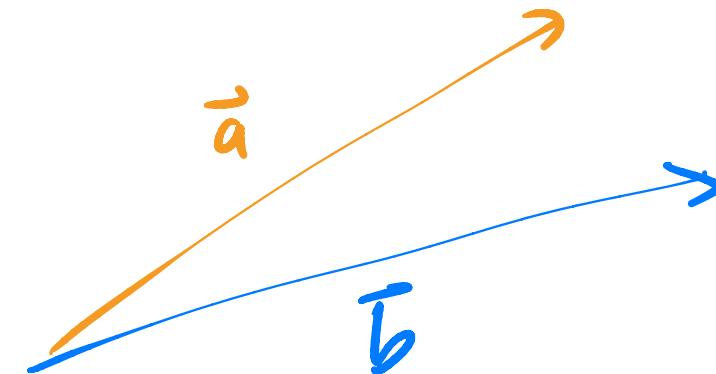
$$\text{error}(w) = \|\vec{e}\| = \|\vec{y} - w\vec{x}\|$$

- We call $w^* \vec{x}$ the **orthogonal projection** of \vec{y} onto $\text{span}(\vec{x})$.
 - Think of $w^* \vec{x}$ as the "shadow" of \vec{y} .



Exercise

Let $\vec{a} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} -1 \\ 9 \end{bmatrix}$.



What is the orthogonal projection of \vec{a} onto $\text{span}(\vec{b})$?

Your answer should be of the form $w^* \vec{b}$, where w^* is a scalar.

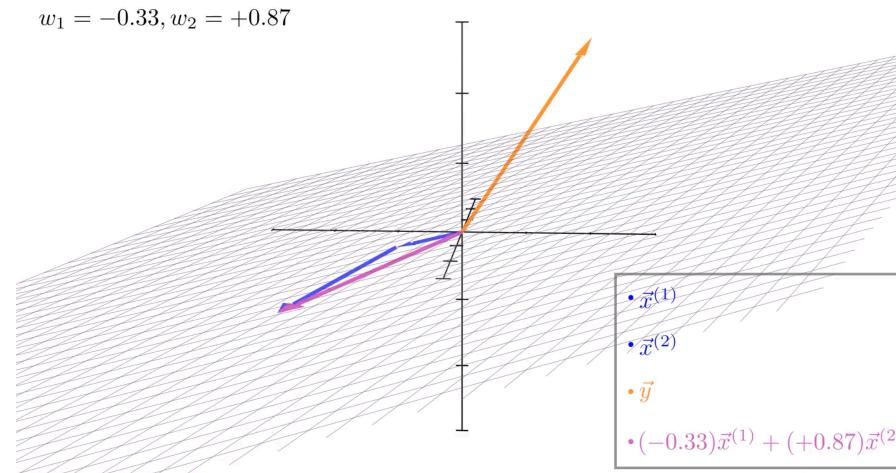
$$w^* = \frac{\vec{b} \cdot \vec{a}}{\vec{b} \cdot \vec{b}} = \frac{(-1)(5) + (9)(2)}{(-1)^2 + (9)^2} = \boxed{\frac{13}{82}}$$

Orthogonal projection of \vec{a} onto $\text{span}(\vec{b})$ is

$$\boxed{\frac{13}{82} \vec{b}}$$

Moving to multiple dimensions

- Let's now consider three vectors, \vec{y} , $\vec{x}^{(1)}$, and $\vec{x}^{(2)}$, all in \mathbb{R}^n .
- **Question:** What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to \vec{y} ?
 - Vectors in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ are of the form $w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}$, where $w_1, w_2 \in \mathbb{R}$ are scalars.
- Before trying to answer, let's watch  [this animation that Jack, one of our tutors, made.](#)



Minimizing projection error in multiple dimensions

- **Question:** What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to \vec{y} ?

- That is, what vector minimizes $\|\vec{e}\|$, where:

$$\vec{e} = \vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}$$

- **Answer:** It's the vector such that $w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}$ is **orthogonal** to \vec{e} .
- **Issue:** Solving for w_1 and w_2 in the following equation is difficult:

$$\underbrace{\left(w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)} \right)}_{\substack{\text{any vector in} \\ \text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})}} \cdot \underbrace{\left(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)} \right)}_{\vec{e}} = 0$$

any vector in
 $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$
can be written in
this form!

Minimizing projection error in multiple dimensions

- It's hard for us to solve for w_1 and w_2 in:

$$\left(w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)} \right) \cdot \underbrace{\left(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)} \right)}_{\vec{e}} = 0$$

- Observation:** All we really need is for $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ to individually be orthogonal to \vec{e} .
 - That is, it's sufficient for \vec{e} to be orthogonal to the spanning vectors themselves.
- If $\vec{x}^{(1)} \cdot \vec{e} = 0$ and $\vec{x}^{(2)} \cdot \vec{e} = 0$, then:

$$\begin{aligned} (\omega_1 \vec{x}^{(1)} + \omega_2 \vec{x}^{(2)}) \cdot \vec{e} &= \omega_1 \vec{x}^{(1)} \cdot \vec{e} + \omega_2 \vec{x}^{(2)} \cdot \vec{e} \\ &= \omega_1 (\vec{x}^{(1)} \cdot \vec{e}) + \omega_2 (\vec{x}^{(2)} \cdot \vec{e}) \\ &= \omega_1 (0) + \omega_2 (0) \\ &= \boxed{0} \end{aligned}$$

Minimizing projection error in multiple dimensions

- **Question:** What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to \vec{y} ?
- **Answer:** It's the vector such that $w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}$ is **orthogonal** to $\vec{e} = \vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}$.
- **Equivalently,** it's the vector such that $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ are both orthogonal to \vec{e} :

$$\begin{aligned}\vec{x}^{(1)} \cdot (\underbrace{\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}}_{\vec{e}}) &= 0 \\ \vec{x}^{(2)} \cdot (\underbrace{\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}}_{\vec{e}}) &= 0\end{aligned}$$

need to find
 w_1^* , w_2^* !

- This is a system of two equations, two unknowns (w_1 and w_2), but it still looks difficult to solve.

Now what?

- We're looking for the scalars w_1 and w_2 that satisfy the following equations:

$$\vec{x}^{(1)} \cdot \left(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)} \right) = 0$$
$$\vec{x}^{(2)} \cdot \underbrace{\left(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)} \right)}_{\vec{e}} = 0$$

- In this example, we just have two spanning vectors, $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$.
- If we had any more, this system of equations would get extremely messy, extremely quickly.
- **Idea:** Rewrite the above system of equations **as a single equation, involving matrix-vector products.**

Matrices

Matrices

- An $n \times d$ **matrix** is a table of numbers with n rows and d columns.
- We use upper-case letters to denote matrices.

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix}$$

the set of
matrices with
2 rows and
3 columns

- Since A has two rows and three columns, we say $A \in \mathbb{R}^{2 \times 3}$.
- Key idea: Think of a matrix as **several column vectors**, stacked next to each other.

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad \begin{bmatrix} 5 \\ 5 \end{bmatrix} \quad \begin{bmatrix} 8 \\ -3 \end{bmatrix}$$

Matrix addition and scalar multiplication

- We can add two matrices only if they have the same dimensions.
- Addition occurs elementwise:

$$\begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 7 & 11 \\ -1 & 6 & -1 \end{bmatrix}$$

- Scalar multiplication occurs elementwise, too:

$$2 \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} = \begin{bmatrix} 4 & 10 & 16 \\ -2 & 10 & -6 \end{bmatrix}$$

Matrix-matrix multiplication

- Key idea: We can multiply matrices A and B if and only if:

$$\# \text{ columns in } A = \# \text{ rows in } B$$

- If A is $n \times d$ and B is $d \times p$, then AB is $n \times p$.
- Example: If A is as defined below, what is $A^T A$?

$$A^T = \begin{bmatrix} 2 & -1 \\ 5 & 5 \\ 8 & -3 \end{bmatrix} \quad 3 \times 2$$

$\xrightarrow{\hspace{10em}}$

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} \quad 2 \times 3$$
$$A^T A = \begin{bmatrix} 5 & 5 & 19 \\ 5 & 25 & \end{bmatrix} \quad 3 \times 3$$

Question 🤔

Answer at q.dsc40a.com

Assume A , B , and C are all matrices. Select the **incorrect** statement below.

- A. $A(B + C) = AB + AC$.
- B. $A(BC) = (AB)C$.
- C. $AB = BA$.
- D. $(A + B)^T = A^T + B^T$.
- E. $(AB)^T = B^T A^T$.

$$\begin{array}{ccc} A_{5 \times 7} & B_{7 \times 5} & \rightarrow 5 \times 5 \\ & & \downarrow \text{different dimensions!} \\ B_{7 \times 5} & A_{5 \times 7} & \rightarrow 7 \times 7 \end{array}$$

Matrix-vector multiplication

- A vector $\vec{v} \in \mathbb{R}^n$ is a matrix with n rows and 1 column.

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

- Suppose $A \in \mathbb{R}^{n \times d}$.
 - What must the dimensions of \vec{v} be in order for the product $A\vec{v}$ to be valid?
 - What must the dimensions of \vec{v} be in order for the product $\vec{v}^T A$ to be valid?

$$A_{n \times d} \quad \vec{v}_{d \times 1} \quad \Rightarrow \quad \vec{v} \in \mathbb{R}^d \quad d \text{ components}$$

$$\vec{v}^T_{1 \times n} \quad A_{n \times d} \quad \Rightarrow \quad \vec{v} \in \mathbb{R}^n \quad n \text{ components}$$

One view of matrix-vector multiplication

- One way of thinking about the product $A\vec{v}$ is that it is the dot product of \vec{v} with every row of A .
- Example: What is $A\vec{v}$?

$$\begin{aligned} & 2(2) + (-1)(5) + (-5)(8) \\ & = 4 - 5 - 40 = -41 \end{aligned}$$

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} \quad 2 \times 3$$

$$\vec{v} = \begin{bmatrix} 2 \\ -1 \\ -5 \end{bmatrix} \quad 3 \times 1$$

$$A\vec{v} = \begin{bmatrix} -41 \\ 8 \end{bmatrix}$$

$$\begin{aligned} & 2(-1) + (-1)(5) + (-5)(-3) \\ & = -2 - 5 + 15 = 8 \end{aligned}$$

$$\vec{v} \in \mathbb{R}^3$$

$$A\vec{v} \in \mathbb{R}^2$$

Another view of matrix-vector multiplication

- Another way of thinking about the product $A\vec{v}$ is that it is a **linear combination of the columns of A , using the weights in \vec{v}** .
- Example: What is $A\vec{v}$?

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix}$$
$$\vec{v} = \begin{bmatrix} 2 \\ -1 \\ -5 \end{bmatrix}$$
$$A\vec{v} = 2 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + (-1) \begin{bmatrix} 5 \\ 5 \end{bmatrix} + (-5) \begin{bmatrix} 8 \\ -3 \end{bmatrix} = \begin{bmatrix} -41 \\ 8 \end{bmatrix}$$

a linear combination
of the columns of A !

Matrix-vector products create linear combinations of columns!

- **Key idea:** It'll be very useful to think of the matrix-vector product $A\vec{v}$ as a linear combination of the columns of A , using the weights in \vec{v} .

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1d} \\ a_{21} & a_{22} & \dots & a_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nd} \end{bmatrix}_{n \times d} \quad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}_{d \times 1}$$

↓

$$A\vec{v} = v_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} + v_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} + \dots + v_d \begin{bmatrix} a_{1d} \\ a_{2d} \\ \vdots \\ a_{nd} \end{bmatrix}$$

⇒ result is
a vector in
 \mathbb{R}^n !

Spans and projections, revisited

$$\omega_1 \vec{x}^{(1)} + \omega_2 \vec{x}^{(2)}$$

Moving to multiple dimensions

- Let's now consider three vectors, \vec{y} , $\vec{x}^{(1)}$, and $\vec{x}^{(2)}$, all in \mathbb{R}^n .
- Question:** What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to \vec{y} ?
 - That is, what values of w_1 and w_2 minimize $\|\vec{e}\| = \|\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}\|$?

Answer : ω_1 and ω_2 such that :

$$\vec{x}^{(1)} \cdot \vec{e} = 0$$

$$\vec{x}^{(2)} \cdot \vec{e} = 0$$

Matrix-vector products create linear combinations of columns! *the same!*

$$\vec{x}^{(1)} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} \quad \vec{x}^{(2)} = \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

$$\omega_1 \vec{x}^{(1)} + \omega_2 \vec{x}^{(2)}$$

$$\vec{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}$$

- Combining $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ into a single matrix gives:

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & \frac{-1}{0} \\ \frac{5}{0} & \frac{0}{4} \\ \frac{3}{4} & \frac{9}{4} \end{bmatrix}$$

$$X\vec{\omega} = \omega_1 \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} + \omega_2 \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix}$$

the same!

- Then, if $\vec{\omega} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, linear combinations of $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ can be written as $X\vec{\omega}$.
- The **span of the columns of X** , or $\text{span}(X)$, consists of all vectors that can be written in the form $X\vec{\omega}$.

Minimizing projection error in multiple dimensions

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

- **Goal:** Find the vector $\vec{w} = [w_1 \quad w_2]^T$ such that $\|\vec{e}\| = \|\vec{y} - X\vec{w}\|$ is minimized.
- As we've seen, \vec{w} must be such that:

$$\vec{x}^{(1)} \cdot (\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}) = 0$$

$$\vec{x}^{(2)} \cdot \underbrace{(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)})}_{\vec{e}} = 0$$

$$w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}$$

- How can we use our knowledge of matrices to rewrite this system of equations as a single equation?

Simplifying the system of equations, using matrices

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

$$\vec{x}^{(1)} \cdot (\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}) = 0$$

$$\vec{x}^{(2)} \cdot (\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}) = 0$$

$$\omega_1 \vec{x}^{(1)} + \omega_2 \vec{x}^{(2)} = \vec{x}\vec{\omega}$$

$$\Rightarrow \vec{e} = \vec{y} - \omega_1 \vec{x}^{(1)} - \omega_2 \vec{x}^{(2)} = \vec{y} - \vec{x}\vec{\omega}$$

\vec{e}

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

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

Simplifying the system of equations, using matrices

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

1. $w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}$ can be written as $X\vec{w}$, so $\vec{e} = \vec{y} - X\vec{w}$.
2. The condition that \vec{e} must be orthogonal to each column of X is equivalent to condition that $X^T \vec{e} = 0$.

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

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

↓ combine into
a single
equation

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

$$\vec{x}^T \vec{e} = \begin{bmatrix} -\vec{x}^{(1)T} \\ -\vec{x}^{(2)T} \end{bmatrix} \vec{e} = \begin{bmatrix} \vec{x}^{(1)T} \vec{e} \\ \vec{x}^{(2)T} \vec{e} \end{bmatrix} = \vec{0}$$

$$X = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix}_{3 \times 2} = \begin{bmatrix} 1 & 1 \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ 1 & 1 \end{bmatrix}$$

$$X^T = \begin{bmatrix} -\vec{x}^{(1)T} \\ -\vec{x}^{(2)T} \end{bmatrix}_{2 \times 3}$$

$$= \begin{bmatrix} 2 & 5 & 3 \\ -1 & 0 & 4 \end{bmatrix}_{2 \times 3}$$

example

$$\begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix}$$

rows of X^T are the
columns of X !!!

The normal equations

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

- **Goal:** Find the vector $\vec{w} = [w_1 \quad w_2]^T$ such that $\|\vec{e}\| = \|\vec{y} - X\vec{w}\|$ is minimized.
- We now know that it is the vector \vec{w}^* such that:

$$\begin{aligned} X^T \vec{e} &= 0 \\ X^T(\vec{y} - X\vec{w}^*) &= 0 \\ X^T \vec{y} - X^T X \vec{w}^* &= 0 \\ \implies X^T X \vec{w}^* &= X^T \vec{y} \end{aligned}$$

← previous slide

- The last statement is referred to as the **normal equations**.

The general solution to the normal equation

$$\textcolor{blue}{X} \in \mathbb{R}^{n \times d} \quad \vec{\textcolor{brown}{y}} \in \mathbb{R}^n$$

- **Goal, in general:** Find the vector $\vec{w} \in \mathbb{R}^d$ such that $\|\vec{e}\| = \|\vec{\textcolor{brown}{y}} - \textcolor{blue}{X}\vec{w}\|$ is minimized.
- We now know that it is the vector \vec{w}^* such that:

$$\begin{aligned} \textcolor{blue}{X}^T \vec{e} &= 0 \\ \implies \textcolor{blue}{X}^T \textcolor{blue}{X} \vec{w}^* &= \textcolor{blue}{X}^T \vec{\textcolor{brown}{y}} \end{aligned}$$

- Assuming $\textcolor{blue}{X}^T \textcolor{blue}{X}$ is invertible, this is the vector:

$$\boxed{\vec{w}^* = (\textcolor{blue}{X}^T \textcolor{blue}{X})^{-1} \textcolor{blue}{X}^T \vec{\textcolor{brown}{y}}}$$

- This is a big assumption, because it requires $\textcolor{blue}{X}^T \textcolor{blue}{X}$ to be **full rank**.
- If $\textcolor{blue}{X}^T \textcolor{blue}{X}$ is not full rank, then there are infinitely many solutions to the normal equations, $\textcolor{blue}{X}^T \textcolor{blue}{X} \vec{w}^* = \textcolor{blue}{X}^T \vec{\textcolor{brown}{y}}$.

What does it mean?

- **Original question:** What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to \vec{y} ?
- **Final answer:** It is the vector $\mathbf{X}\vec{w}^*$, where:

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

- Revisiting our example:

$$\mathbf{X} = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

- Using a computer gives us $\vec{w}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y} \approx \begin{bmatrix} 0.7289 \\ 1.6300 \end{bmatrix}$.
- So, the vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ closest to \vec{y} is $0.7289\vec{x}^{(1)} + 1.6300\vec{x}^{(2)}$.

An optimization problem, solved

- We just used linear algebra to solve an **optimization problem**.
- Specifically, the function we minimized is:

$$\text{error}(\vec{w}) = \|\vec{y} - \mathbf{X}\vec{w}\|$$

- This is a function whose input is a vector, \vec{w} , and whose output is a scalar!
- The input, \vec{w}^* , to $\text{error}(\vec{w})$ that minimizes it is:

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

- We're going to use this frequently!