

# Section 5.3 : Diagonalization

Chapter 5 : Eigenvalues and Eigenvectors

Math 1554 Linear Algebra

**Motivation:** it can be useful to take large powers of matrices, for example  $A^k$ , for large  $k$ .

**But:** multiplying two  $n \times n$  matrices requires roughly  $n^3$  computations. Is there a more efficient way to compute  $A^k$ ?

# Topics and Objectives

## Topics

1. Diagonal, similar, and diagonalizable matrices
2. Diagonalizing matrices

## Learning Objectives

For the topics covered in this section, students are expected to be able to do the following.

1. Determine whether a matrix can be diagonalized, and if possible diagonalize a square matrix.
2. Apply diagonalization to compute matrix powers.

# Similar Matrices

## Definition

Two  $n \times n$  matrices  $A$  and  $B$  are **similar** if there is a <sup>invertible</sup> matrix  $P$  so that  $A = PBP^{-1}$ .

## Theorem

If  $A$  and  $B$  similar, then they have the same characteristic polynomial.

If time permits, we will explain or prove this theorem in lecture. Note:

- Our textbook introduces similar matrices in Section 5.2, but doesn't have exercises on this concept until 5.3.
- Two matrices,  $A$  and  $B$ , do not need to be similar to have the same eigenvalues. For example,

$$A = \underline{P} \cdot \underline{B} \cdot \underline{P}^{-1} = \underline{O}$$

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = B \quad \text{Not similar.}$$

$$\phi_A = \lambda^2$$

$$\phi_B = \lambda^2$$

proof

$$\phi_A(\lambda) = \det(A - \lambda I), \quad A = P \cdot B \cdot P^{-1}$$

$$= \det(PBP^{-1} - \lambda \cdot P \cdot I \cdot P^{-1})$$

$$= \det(P \cdot (B - \lambda I) \cdot P^{-1})$$

$$= \underline{\det(P)} \cdot \det(B - \lambda I) \cdot \underline{\det(P^{-1})}$$

$$I = P \cdot I \cdot P^{-1}$$

$$= \underline{\det(P) \cdot \det(P^{-1})} \phi_B(\lambda)$$

$$= \det(P \cdot P^{-1}) \stackrel{=1}{=} \phi_B(\lambda)$$

$$= \phi_B(\lambda)$$



## Additional Examples (if time permits)

1. True or false.
  - a) If  $A$  is similar to the identity matrix, then  $A$  is equal to the identity matrix.
  - b) A row replacement operation on a matrix does not change its eigenvalues.
2. For what values of  $k$  does the matrix have one real eigenvalue with algebraic multiplicity 2?

$$\begin{pmatrix} -3 & k \\ 2 & -6 \end{pmatrix}$$

# Diagonal Matrices

*square*

A <sup>*square*</sup> matrix is **diagonal** if the **only non-zero elements**, if any, are on the main **diagonal**.

The following are all diagonal matrices.

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad [2], \quad I_n, \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

We'll only be working with diagonal square matrices in this course.

## Powers of Diagonal Matrices

If  $A$  is diagonal, then  $A^k$  is easy to compute. For example,

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 0.5 \end{pmatrix}$$

$$A^2 = \begin{pmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 3^2 & 0 \\ 0 & (\frac{1}{2})^2 \end{pmatrix}$$

$$A^k = \begin{pmatrix} 3^k & 0 \\ 0 & (\frac{1}{2})^k \end{pmatrix}$$

Note  $A, B$  : diagonal  
But what if  $A$  is not diagonal?

Is  $AB$  diagonal? Yes.

$$\begin{bmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{bmatrix} \begin{bmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_n \end{bmatrix} = \begin{bmatrix} a_1 b_1 & & 0 \\ & \ddots & \\ 0 & & a_n b_n \end{bmatrix}$$

①  $A$  is similar to diagonal.

$$A = P \cdot \underset{\substack{\uparrow \\ \text{diagonal}}}{D} \cdot P^{-1}$$

Section 5.3 Slide 26

$$A^2 = (P D P^{-1}) (P D P^{-1}) = P D \cdot \overset{I}{(P \cdot P^{-1})} \cdot D P^{-1}$$

$$= P \cdot \underbrace{D \cdot D}_{D^2} \cdot P^{-1} = P \cdot D^2 \cdot P^{-1}$$

$$A^k = P \cdot D^k \cdot P^{-1}$$

②: When is this possible?

$$\textcircled{2} \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad A^2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

$$A^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a+b \\ 0 & 1 \end{pmatrix}$$

Note If  $A, B$  upper triangular,  
 So is  $A+B$  (Exercise)

Goal :  $A = P \cdot D \cdot P^{-1}$ ,  $D$ : diagonal.

Use eigenvalues & eigenvectors.  $A \in \mathbb{R}^{n \times n}$

Suppose  $\lambda_1, \lambda_2, \dots, \lambda_n$  eigenvalues  
 $\begin{matrix} | & | & & | \\ v_1 & v_2 & & v_n \end{matrix}$  eigenvectors

$$\begin{cases} Av_1 = \lambda_1 v_1 \\ Av_2 = \lambda_2 v_2 \\ \vdots \\ Av_n = \lambda_n v_n \end{cases} \quad n \text{ equations} \Rightarrow 1 \text{ equation.}$$

$$A \cdot \underbrace{\begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{bmatrix}}_{=P} = \begin{bmatrix} | & | & \dots & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \dots & \lambda_n v_n \\ | & | & & | \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{bmatrix}}_{=P} \underbrace{\begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}}_{=D}$$

$AP = PD$ . If  $P$  is invertible.

$$A = P D P^{-1}$$



# Diagonalization

Suppose  $A \in \mathbb{R}^{n \times n}$ . We say that  $A$  is **diagonalizable** if it is similar to a diagonal matrix,  $D$ . That is, we can write

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

# Diagonalization

$$A = P \cdot D \cdot P^{-1} \Leftrightarrow \overset{[v_1 \dots v_n]}{P} \text{ is invertible}$$

Theorem

If  $A$  is diagonalizable  $\Leftrightarrow A$  has  $n$  linearly independent eigenvectors.

Note: the symbol  $\Leftrightarrow$  means “if and only if”.

Also note that  $A = PDP^{-1}$  if and only if

$$A = [\vec{v}_1 \ \vec{v}_2 \ \dots \ \vec{v}_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} [\vec{v}_1 \ \vec{v}_2 \ \dots \ \vec{v}_n]^{-1}$$

where  $\vec{v}_1, \dots, \vec{v}_n$  are linearly independent eigenvectors, and  $\lambda_1, \dots, \lambda_n$  are the corresponding eigenvalues (**in order**).

Q: When we have  $n$  lin. indep. eigenvectors.

## Example 1

Diagonalize if possible.

$$\begin{pmatrix} 2 & 6 \\ 0 & -1 \end{pmatrix}$$

① Eigenvalues :  $\phi_A(\lambda) = \det(A - \lambda I) = \lambda^2 - (\overset{\text{trace}}{2+(-1)}\lambda + \overset{\text{det}}{(-2)})$   
 $= \lambda^2 - \lambda - 2 = 0 \quad \therefore \lambda = 2, -1.$

②  $\lambda = 2$  :  $E_2 = \text{Nul}(A - 2I)$

$$A - 2I = \begin{pmatrix} 0 & 6 \\ 0 & -3 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \begin{array}{l} 0 \cdot x + 1 \cdot y = 0 \\ \therefore y = 0. \end{array}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix} = x \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

lin. indep

③  $\lambda = -1$  :  $A + I = \begin{pmatrix} 3 & 6 \\ 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} \quad \begin{array}{l} x + 2y = 0 \\ x = -2y \end{array}$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2y \\ y \end{bmatrix} = y \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

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

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^2$$

## Example 2

Diagonalize if possible.

$$\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

Special case

## Distinct Eigenvalues

all algebraic multiplicities are 1

### Theorem

If  $A$  is  $n \times n$  and has  $n$  distinct eigenvalues, then  $A$  is diagonalizable.

Why does this theorem hold?

Thm

$\lambda_1, \lambda_2, \dots, \lambda_n$  : distinct eigenvalues

$\begin{matrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \end{matrix}$

$\Rightarrow \{v_1, \dots, v_n\}$  linearly independent.

Is it necessary for an  $n \times n$  matrix to have  $n$  distinct eigenvalues for it to be diagonalizable?

Sketchy of Proof

Assume  $(k-1)$  distinct eigenvalues  $\Rightarrow$  lin. indep.

Goal :

$k$  distinct  $\Rightarrow$  lin. indep.

$v_1, v_2, \dots, v_k$

$\begin{matrix} | & | & & | \\ \lambda_1 & \lambda_2 & \dots & \lambda_k \end{matrix}$  : distinct.

$$\lambda_1 (a_1 v_1 + a_2 v_2 + \dots + a_k v_k) = 0 \quad : \text{WANT : } a_1 = a_2 = \dots = a_k = 0$$

$$A (a_1 v_1 + \dots + a_k v_k) = 0$$

$$a_1 \cancel{\lambda_1} v_1 + a_2 (\lambda_2) v_2 + \dots + a_k \lambda_k v_k = 0$$

$$- (a_1 \cancel{\lambda_1} v_1 + a_2 (\lambda_1) v_2 + \dots + a_k \lambda_1 v_k = 0$$

$$a_2(\lambda_2 - \lambda_1) \cdot v_2 + a_3(\lambda_3 - \lambda_1)v_3 + \dots + a_k(\lambda_k - \lambda_1)v_k = 0$$

$$\Rightarrow a_2 \underbrace{(\lambda_2 - \lambda_1)}_{\neq 0} = \dots = a_k \underbrace{(\lambda_k - \lambda_1)}_{\neq 0} = 0$$

$$\Rightarrow a_2 = \dots = a_k = 0$$

$$\Rightarrow a_1 v_1^{\neq 0} = 0 \quad \therefore a_1 = 0$$

## Non-Distinct Eigenvalues

Theorem. Suppose

- $A$  is  $n \times n$
- $A$  has distinct eigenvalues  $\lambda_1, \dots, \lambda_k$ ,  $k \leq n$
- $a_i =$  algebraic multiplicity of  $\lambda_i$
- $d_i =$  dimension of  $\lambda_i$  eigenspace (“geometric multiplicity”)

Then

1.  $d_i \leq a_i$  for all  $i$
2.  $A$  is diagonalizable  $\Leftrightarrow \sum d_i = n \Leftrightarrow d_i = a_i$  for all  $i$
3.  $A$  is diagonalizable  $\Leftrightarrow$  the eigenvectors, for all eigenvalues, together form a basis for  $\mathbb{R}^n$ .

## Example 3

The eigenvalues of  $A$  are  $\lambda = 3, 1$ . If possible, construct  $P$  and  $D$  such that  $AP = PD$ .

$$A = \begin{pmatrix} 7 & 4 & 16 \\ 2 & 5 & 8 \\ -2 & -2 & -5 \end{pmatrix}$$



## Additional Example (if time permits)

Note that

$$\vec{x}_k = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \vec{x}_{k-1}, \quad \vec{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad k = 1, 2, 3, \dots$$

generates a well-known sequence of numbers.

Use a diagonalization to find a matrix equation that gives the  $n^{\text{th}}$  number in this sequence.

# Chapter 5 : Eigenvalues and Eigenvectors

## 5.5 : Complex Eigenvalues

# Topics and Objectives

## Topics

1. Complex numbers: addition, multiplication, complex conjugate
2. Complex eigenvalues and eigenvectors.
3. Eigenvalue theorems

## Learning Objectives

1. Use eigenvalues to determine identify the rotation and dilation of a linear transform.
2. Rotation dilation matrices.
3. Find complex eigenvalues and eigenvectors of a real matrix.
4. Apply theorems to characterize matrices with complex eigenvalues.

## Motivating Question

What are the eigenvalues of a rotation matrix?

# Imaginary Numbers

**Recall:** When calculating roots of polynomials, we can encounter square roots of negative numbers. For example:

$$x^2 + 1 = 0$$

The roots of this equation are:

We usually write  $\sqrt{-1}$  as  $i$  (for “imaginary”).

## Addition and Multiplication

The imaginary (or complex) numbers are denoted by  $\mathbb{C}$ , where

$$\mathbb{C} = \{a + bi \mid a, b \text{ in } \mathbb{R}\}$$

We can identify  $\mathbb{C}$  with  $\mathbb{R}^2$ :  $a + bi \leftrightarrow (a, b)$

We can add and multiply complex numbers as follows:

$$(2 - 3i) + (-1 + i) =$$

$$(2 - 3i)(-1 + i) =$$

## Complex Conjugate, Absolute Value, Polar Form

We can **conjugate** complex numbers:  $\overline{a + bi} = \underline{\hspace{2cm}}$

The **absolute value** of a complex number:  $|a + bi| = \underline{\hspace{2cm}}$

We can write complex numbers in **polar form**:  $a + ib = r(\cos \phi + i \sin \phi)$

## Complex Conjugate Properties

If  $x$  and  $y$  are complex numbers,  $\vec{v} \in \mathbb{C}^n$ , it can be shown that:

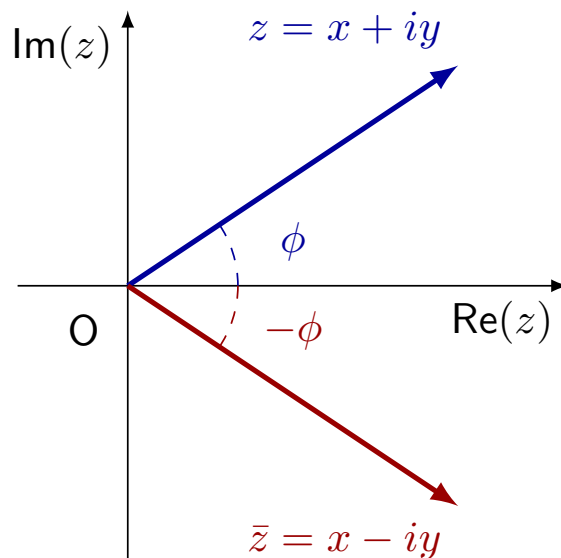
- $\overline{(x + y)} = \bar{x} + \bar{y}$
- $\overline{A\vec{v}} = A\vec{v}$
- $\text{Im}(x\bar{x}) = 0$ .

**Example** True or false: if  $x$  and  $y$  are complex numbers, then

$$\overline{(xy)} = \bar{x} \bar{y}$$

## Polar Form and the Complex Conjugate

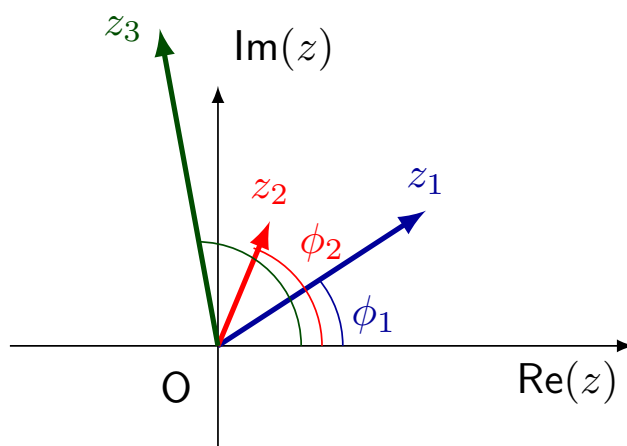
Conjugation reflects points across the real axis.





## Euler's Formula

Suppose  $z_1$  has angle  $\phi_1$ , and  $z_2$  has angle  $\phi_2$ .



The product  $z_1 z_2$  has angle  $\phi_1 + \phi_2$  and modulus  $|z| |w|$ . Easy to remember using Euler's formula.

$$z = |z| e^{i\phi}$$

The product  $z_1 z_2$  is:

$$z_3 = z_1 z_2 = (|z_1| e^{i\phi_1})(|z_2| e^{i\phi_2}) = |z_1| |z_2| e^{i(\phi_1 + \phi_2)}$$

# Complex Numbers and Polynomials

## Theorem: Fundamental Theorem of Algebra

Every polynomial of degree  $n$  has exactly  $n$  complex roots, counting multiplicity.

## Theorem

1. If  $\lambda \in \mathbb{C}$  is a root of a real polynomial  $p(x)$ , then the conjugate  $\bar{\lambda}$  is also a root of  $p(x)$ .
2. If  $\lambda$  is an eigenvalue of real matrix  $A$  with eigenvector  $\vec{v}$ , then  $\bar{\lambda}$  is an eigenvalue of  $A$  with eigenvector  $\vec{\bar{v}}$ .

## Example

Four of the eigenvalues of a  $7 \times 7$  matrix are  $-2$ ,  $4 + i$ ,  $-4 - i$ , and  $i$ .  
What are the other eigenvalues?

## Example

The matrix that rotates vectors by  $\phi = \pi/4$  radians about the origin, and then scales (or dilates) vectors by  $r = \sqrt{2}$ , is

$$A = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

What are the eigenvalues of  $A$ ? Find an eigenvector for each eigenvalue.

## Example

The matrix in the previous example is a special case of this matrix:

$$C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

Calculate the eigenvalues of  $C$  and express them in polar form.

## Example

Find the complex eigenvalues and an associated complex eigenvector for each eigenvalue for the matrix.

$$A = \begin{pmatrix} 1 & -2 \\ 1 & 3 \end{pmatrix}$$

# Section 6.1 : Inner Product, Length, and Orthogonality

Chapter 6: Orthogonality and Least Squares

Math 1554 Linear Algebra

# Topics and Objectives

## Topics

1. Dot product of vectors
2. Magnitude of vectors, and distances in  $\mathbb{R}^n$
3. Orthogonal vectors and complements
4. Angles between vectors

## Learning Objectives

1. Compute (a) dot product of two vectors, (b) length (or magnitude) of a vector, (c) distance between two points in  $\mathbb{R}^n$ , and (d) angles between vectors.
2. Apply theorems related to orthogonal complements, and their relationships to Row and Null space, to characterize vectors and linear systems.

## Motivating Question

For a matrix  $A$ , which vectors are orthogonal to all the rows of  $A$ ? To the columns of  $A$ ?



# The Dot Product

The dot product between two vectors,  $\vec{u}$  and  $\vec{v}$  in  $\mathbb{R}^n$ , is defined as

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

**Example 1:** For what values of  $k$  is  $\vec{u} \cdot \vec{v} = 0$ ?

$$\vec{u} = \begin{pmatrix} -1 \\ 3 \\ k \\ 2 \end{pmatrix}, \quad \vec{v} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ -3 \end{pmatrix}$$

## Properties of the Dot Product

The dot product is a special form of matrix multiplication, so it inherits linear properties.

### Theorem (Basic Identities of Dot Product)

Let  $\vec{u}, \vec{v}, \vec{w}$  be three vectors in  $\mathbb{R}^n$ , and  $c \in \mathbb{R}$ .

1. (Symmetry)  $\vec{u} \cdot \vec{w} = \underline{\hspace{2cm}}$
2. (Linear in each vector)  $(\vec{v} + \vec{w}) \cdot \vec{u} = \underline{\hspace{2cm}}$
3. (Scalars)  $(c\vec{u}) \cdot \vec{w} = \underline{\hspace{2cm}}$
4. (Positivity)  $\vec{u} \cdot \vec{u} \geq 0$ , and the dot product equals  $\underline{\hspace{2cm}}$

# The Length of a Vector

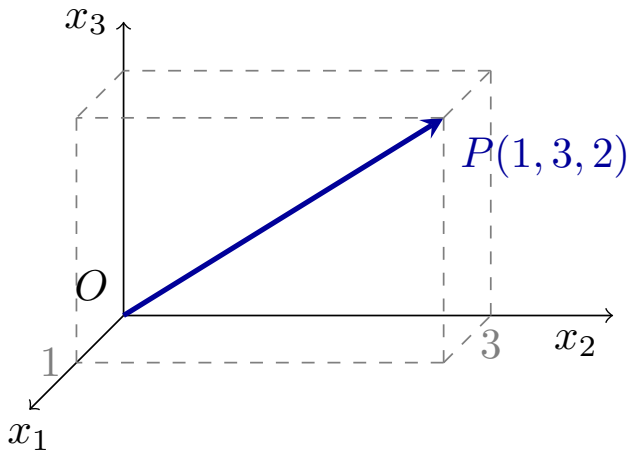
## Definition

The **length** of a vector  $\vec{u} \in \mathbb{R}^n$  is

$$\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

**Example:** the length of the vector  $\overrightarrow{OP}$  is

$$\sqrt{1^2 + 3^2 + 2^2} = \sqrt{14}$$



## Example

Let  $\vec{u}, \vec{v}$  be two vectors in  $\mathbb{R}^n$  with  $\|\vec{u}\| = 5$ ,  $\|\vec{v}\| = \sqrt{3}$ , and  $\vec{u} \cdot \vec{v} = -1$ . Compute the value of  $\|\vec{u} + \vec{v}\|$ .

## Length of Vectors and Unit Vectors

**Note:** for any vector  $\vec{v}$  and scalar  $c$ , the length of  $c\vec{v}$  is

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

### Definition

If  $\vec{v} \in \mathbb{R}^n$  has length one, we say that it is a **unit vector**.

For example, each of the following vectors are unit vectors.

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \vec{y} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \vec{v} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

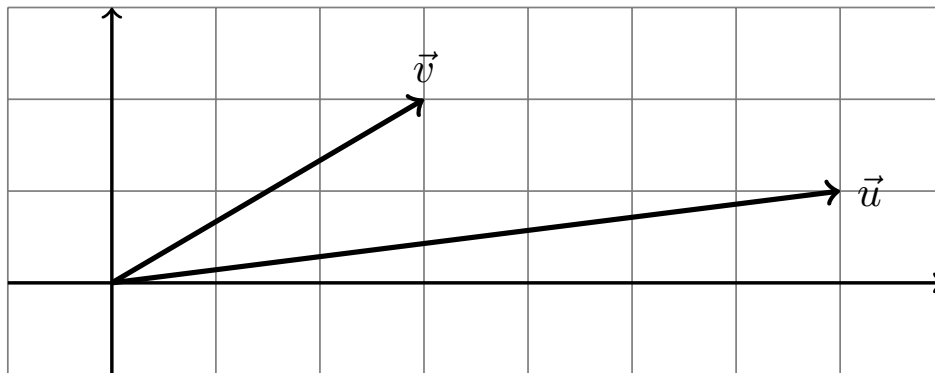
## Distance in $\mathbb{R}^n$

### Definition

For  $\vec{u}, \vec{v} \in \mathbb{R}^n$ , the **distance** between  $\vec{u}$  and  $\vec{v}$  is given by the formula

$$\|\vec{u} - \vec{v}\|$$

**Example:** Compute the distance from  $\vec{u} = \begin{pmatrix} 7 \\ 1 \end{pmatrix}$  and  $\vec{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ .



# The Cauchy-Schwarz Inequality

## Theorem: Cauchy-Bunyakovsky–Schwarz Inequality

For all  $\vec{u}$  and  $\vec{v}$  in  $\mathbb{R}^n$ ,

$$|\vec{u} \cdot \vec{v}| \leq \|\vec{u}\| \|\vec{v}\|.$$

Equality holds *if and only if*  $\vec{v} = \alpha \vec{u}$  for  $\alpha = \frac{\vec{u} \cdot \vec{v}}{\vec{u} \cdot \vec{u}}$ .

**Proof:** Assume  $\vec{u} \neq 0$ , otherwise there is nothing to prove.

Set  $\alpha = \frac{\vec{u} \cdot \vec{v}}{\vec{u} \cdot \vec{u}}$ . Observe that  $\vec{u} \cdot (\alpha \vec{u} - \vec{v}) = 0$ . So

$$\begin{aligned} 0 &\leq \|\alpha \vec{u} - \vec{v}\|^2 = (\alpha \vec{u} - \vec{v}) \cdot (\alpha \vec{u} - \vec{v}) \\ &= \alpha \vec{u} \cdot (\alpha \vec{u} - \vec{v}) - \vec{v} \cdot (\alpha \vec{u} - \vec{v}) \\ &= -\vec{v} \cdot (\alpha \vec{u} - \vec{v}) \\ &= \frac{\|\vec{u}\|^2 \|\vec{v}\|^2 - |\vec{u} \cdot \vec{v}|^2}{\|\vec{u}\|^2} \end{aligned}$$

# The Triangle Inequality

## Theorem: Triangle Inequality

For all  $\vec{u}$  and  $\vec{v}$  in  $\mathbb{R}^n$ ,

$$\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|.$$

**Proof:**

$$\begin{aligned}\|\vec{u} + \vec{v}\|^2 &= (\vec{u} + \vec{v}) \cdot (\vec{u} + \vec{v}) \\ &= \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\vec{u} \cdot \vec{v} \\ &\leq \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\|\vec{u}\|\|\vec{v}\| \\ &= (\|\vec{u}\| + \|\vec{v}\|)^2\end{aligned}$$



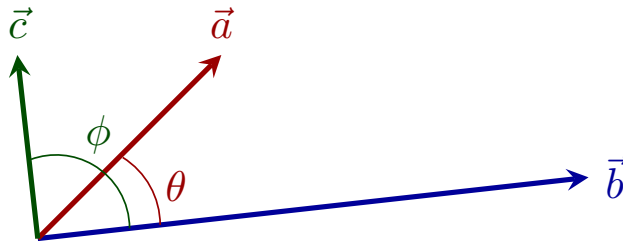
# Angles

## Theorem

$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos \theta$ . Thus, if  $\vec{a} \cdot \vec{b} = 0$ , then:

- $\vec{a}$  and/or  $\vec{b}$  are \_\_\_\_\_ vectors, or
- $\vec{a}$  and  $\vec{b}$  are \_\_\_\_\_.

For example, consider the vectors below.



# Orthogonality

## Definition (Orthogonal Vectors)

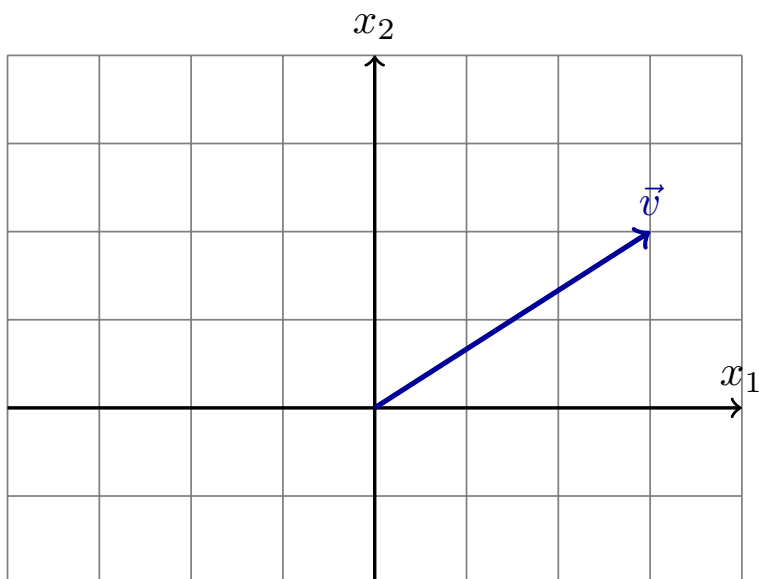
Two vectors  $\vec{u}$  and  $\vec{w}$  are **orthogonal** if  $\vec{u} \cdot \vec{w} = 0$ . This is equivalent to:

$$\|\vec{u} + \vec{w}\|^2 =$$

Note: The zero vector in  $\mathbb{R}^n$  is orthogonal to every vector in  $\mathbb{R}^n$ . But we usually only mean non-zero vectors.

## Example

Sketch the subspace spanned by the set of all vectors  $\vec{u}$  that are orthogonal to  $\vec{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ .



# Orthogonal Compliments

## Definitions

Let  $W$  be a subspace of  $\mathbb{R}^n$ . Vector  $\vec{z} \in \mathbb{R}^n$  is **orthogonal** to  $W$  if  $\vec{z}$  is orthogonal to every vector in  $W$ .

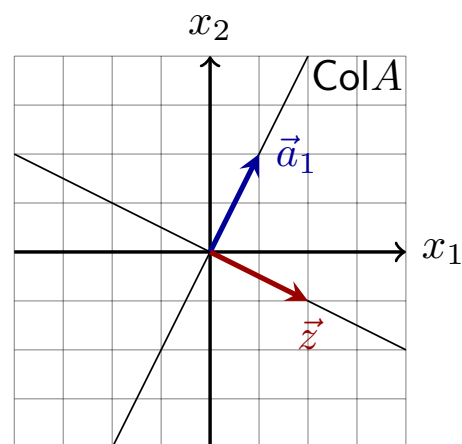
The set of all vectors orthogonal to  $W$  is a subspace, the **orthogonal complement** of  $W$ , or  $W^\perp$  or ' $W$  perp.'

$$W^\perp = \{\vec{z} \in \mathbb{R}^n : \vec{z} \cdot \vec{w} = 0 \text{ for all } \vec{w} \in W\}$$

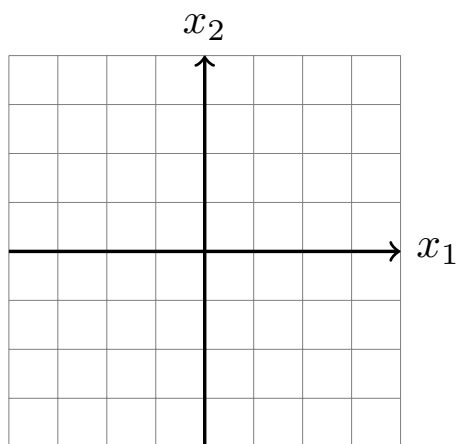
## Example

Example: suppose  $A = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$ .

- $\text{Col}A$  is the span of  $\vec{a}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$
- $\text{Col}A^\perp$  is the span of  $\vec{z} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$

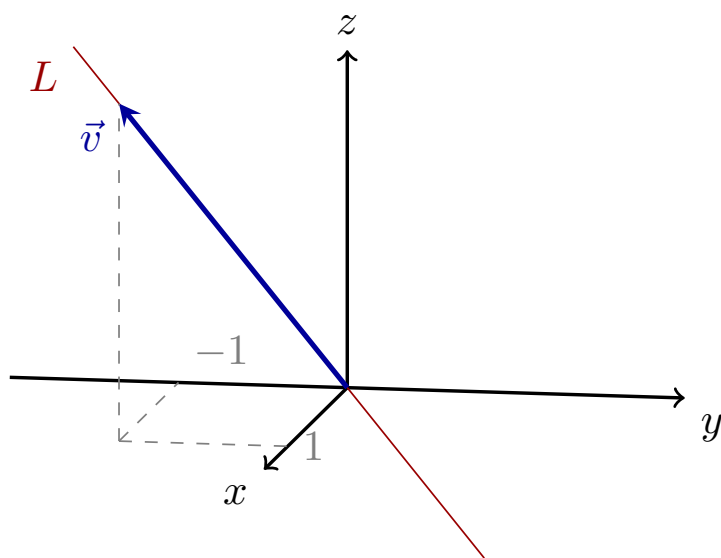


Sketch  $\text{Null}A$  and  $\text{Null}A^\perp$  on the grid below.



## Example

Line  $L$  is a subspace of  $\mathbb{R}^3$  spanned by  $\vec{v} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ . Then the space  $L^\perp$  is a plane. Construct an equation of the plane  $L^\perp$ .



Can also visualise line and plane with CalcPlot3D: [web.monroecc.edu/calcNSF](http://web.monroecc.edu/calcNSF)

# Row $A$

## Definition

Row  $A$  is the space spanned by the rows of matrix  $A$ .

We can show that

- $\dim(\text{Row}(A)) = \dim(\text{Col}(A))$
- a basis for Row  $A$  is the pivot rows of  $A$

Note that  $\text{Row}(A) = \text{Col}(A^T)$ , but in general Row  $A$  and Col  $A$  are not related to each other

## Example 3

Describe the  $\text{Null}(A)$  in terms of an orthogonal subspace.

A vector  $\vec{x}$  is in  $\text{Null } A$  if and only if

1.  $A\vec{x} =$

2. This means that  $\vec{x}$  is  to each row of  $A$ .

3.  $\text{Row } A$  is  to  $\text{Null } A$ .

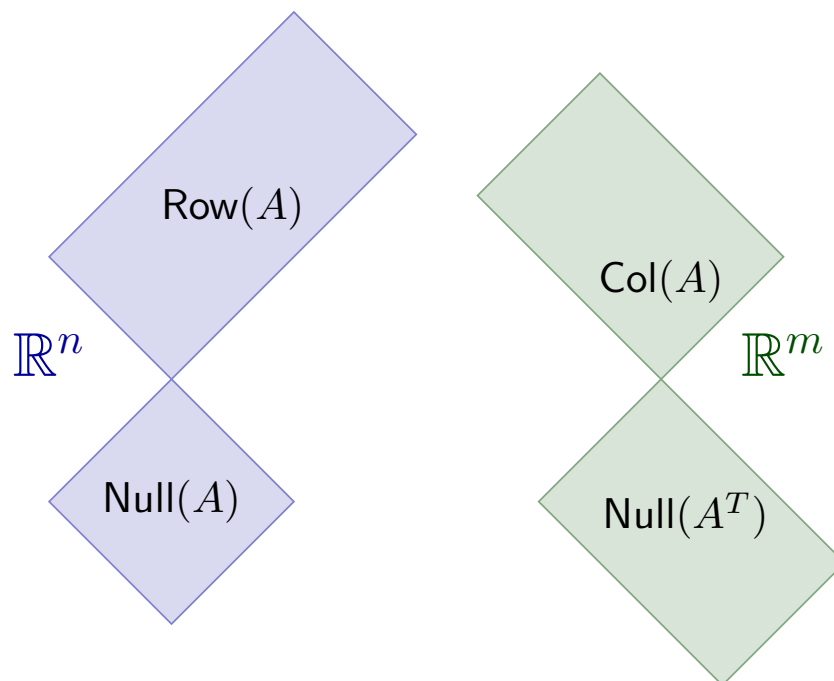
4. The dimension of  $\text{Row } A$  plus the dimension of  $\text{Null } A$  equals



### Theorem (The Four Subspaces)

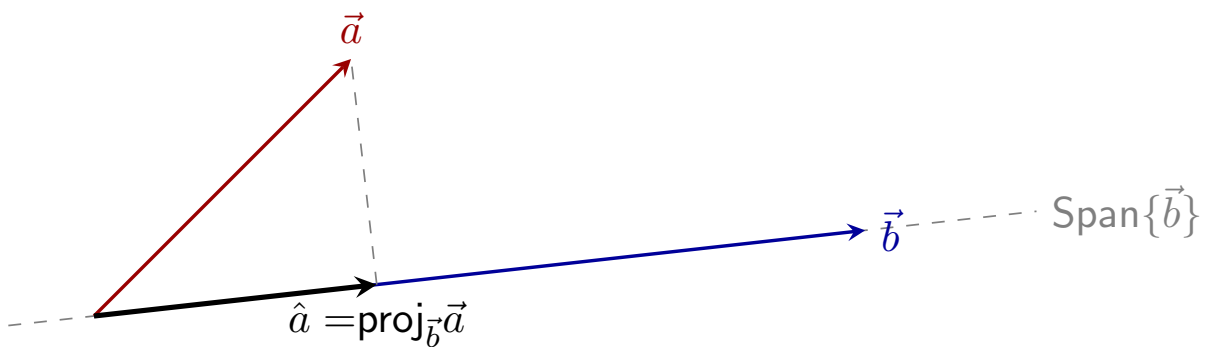
For any  $A \in \mathbb{R}^{m \times n}$ , the orthogonal complement of  $\text{Row } A$  is  $\text{Null } A$ , and the orthogonal complement of  $\text{Col } A$  is  $\text{Null } A^T$ .

The idea behind this theorem is described in the diagram below.



## Looking Ahead - Projections

Suppose we want to find the closed vector in  $\text{Span}\{\vec{b}\}$  to  $\vec{a}$ .



- Later in this Chapter, we will make connections between dot products and **projections**.
- Projections are also used throughout multivariable calculus courses.

# Section 6.2 : Orthogonal Sets

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra

# Topics and Objectives

## Topics

1. Orthogonal Sets of Vectors
2. Orthogonal Bases and Projections.

## Learning Objectives

1. Apply the concepts of orthogonality to
  - a) compute orthogonal projections and distances,
  - b) express a vector as a linear combination of orthogonal vectors,
  - c) characterize bases for subspaces of  $\mathbb{R}^n$ , and
  - d) construct orthonormal bases.

## Motivating Question

What are the special properties of this basis for  $\mathbb{R}^3$ ?

$$\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} / \sqrt{11}, \quad \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} / \sqrt{6}, \quad \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix} / \sqrt{66}$$

## Orthogonal Vector Sets

### Definition

A set of vectors  $\{\vec{u}_1, \dots, \vec{u}_p\}$  are an **orthogonal set** of vectors if for each  $j \neq k$ ,  $\vec{u}_j \perp \vec{u}_k$ .

**Example:** Fill in the missing entries to make  $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$  an orthogonal set of vectors.

$$\vec{u}_1 = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} 0 \\ \phantom{0} \\ \phantom{0} \end{bmatrix}$$

# Linear Independence

## Theorem (Linear Independence for Orthogonal Sets)

Let  $\{\vec{u}_1, \dots, \vec{u}_p\}$  be an orthogonal set of vectors. Then, for scalars  $c_1, \dots, c_p$ ,

$$\|c_1\vec{u}_1 + \dots + c_p\vec{u}_p\|^2 = c_1^2\|\vec{u}_1\|^2 + \dots + c_p^2\|\vec{u}_p\|^2.$$

In particular, if all the vectors  $\vec{u}_r$  are non-zero, the set of vectors  $\{\vec{u}_1, \dots, \vec{u}_p\}$  are linearly independent.

## Orthogonal Bases

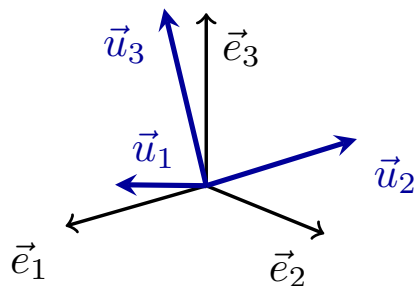
### Theorem (Expansion in Orthogonal Basis)

Let  $\{\vec{u}_1, \dots, \vec{u}_p\}$  be an orthogonal basis for a subspace  $W$  of  $\mathbb{R}^n$ . Then, for any vector  $\vec{w} \in W$ ,

$$\vec{w} = c_1 \vec{u}_1 + \dots + c_p \vec{u}_p.$$

Above, the scalars are  $c_q = \frac{\vec{w} \cdot \vec{u}_q}{\vec{u}_q \cdot \vec{u}_q}$ .

For example, any vector  $\vec{w} \in \mathbb{R}^3$  can be written as a linear combination of  $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ , or some other orthogonal basis  $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ .



## Example

$$\vec{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{u} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad \vec{v} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{s} = \begin{pmatrix} 3 \\ -4 \\ 1 \end{pmatrix}$$

Let  $W$  be the subspace of  $\mathbb{R}^3$  that is orthogonal to  $\vec{x}$ .

- Check that an orthogonal basis for  $W$  is given by  $\vec{u}$  and  $\vec{v}$ .
- Compute the expansion of  $\vec{s}$  in basis  $W$ .



# Projections

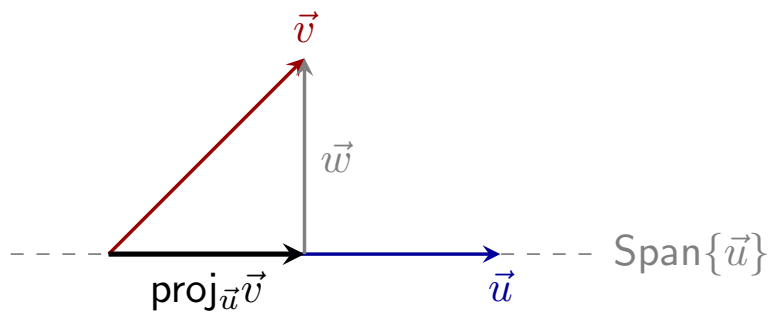
Let  $\vec{u}$  be a non-zero vector, and let  $\vec{v}$  be some other vector. The **orthogonal projection of  $\vec{v}$  onto the direction of  $\vec{u}$**  is the vector in the span of  $\vec{u}$  that is closest to  $\vec{v}$ .

$$\text{proj}_{\vec{u}} \vec{v} = \frac{\vec{v} \cdot \vec{u}}{\vec{u} \cdot \vec{u}} \vec{u}.$$

The vector  $\vec{w} = \vec{v} - \text{proj}_{\vec{u}} \vec{v}$  is orthogonal to  $\vec{u}$ , so that

$$\vec{v} = \text{proj}_{\vec{u}} \vec{v} + \vec{w}$$

$$\|\vec{v}\|^2 = \|\text{proj}_{\vec{u}} \vec{v}\|^2 + \|\vec{w}\|^2$$



## Example

Let  $L$  be spanned by  $\vec{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ .

1. Calculate the projection of  $\vec{y} = (-3, 5, 6, -4)$  onto line  $L$ .
2. How close is  $\vec{y}$  to the line  $L$ ?

## Definition

### Definition (Orthonormal Basis)

An **orthonormal basis** for a subspace  $W$  is an orthogonal basis  $\{\vec{u}_1, \dots, \vec{u}_p\}$  in which every vector  $\vec{u}_q$  has unit length. In this case, for each  $\vec{w} \in W$ ,

$$\vec{w} = (\vec{w} \cdot \vec{u}_1)\vec{u}_1 + \dots + (\vec{w} \cdot \vec{u}_p)\vec{u}_p$$
$$\|\vec{w}\| = \sqrt{(\vec{w} \cdot \vec{u}_1)^2 + \dots + (\vec{w} \cdot \vec{u}_p)^2}$$

## Example

The subspace  $W$  is a subspace of  $\mathbb{R}^3$  perpendicular to  $x = (1, 1, 1)$ . Calculate the missing coefficients in the orthonormal basis for  $W$ .

$$u = \frac{1}{\sqrt{\quad}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad v = \frac{1}{\sqrt{\quad}} \begin{bmatrix} \quad \\ \quad \end{bmatrix}$$

# Orthogonal Matrices

An **orthogonal matrix** is a square matrix whose columns are orthonormal.

## Theorem

An  $m \times n$  matrix  $U$  has orthonormal columns if and only if  $U^T U = I_n$ .

Can  $U$  have orthonormal columns if  $n > m$ ?

# Theorem

## Theorem (Mapping Properties of Orthogonal Matrices)

Assume  $m \times m$  matrix  $U$  has orthonormal columns. Then

1. (Preserves length)  $\|U\vec{x}\| =$

2. (Preserves angles)  $(U\vec{x}) \cdot (U\vec{y}) =$

3. (Preserves orthogonality)

## Example

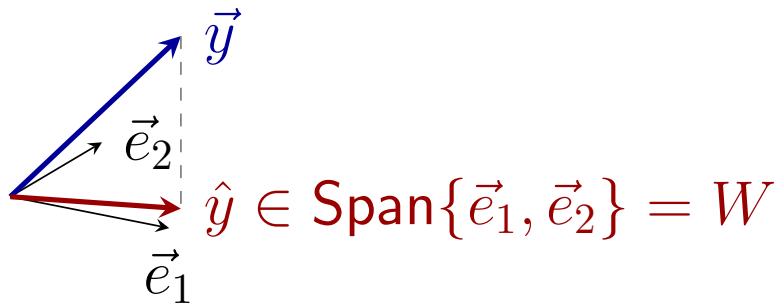
Compute the length of the vector below.

$$\begin{bmatrix} 1/2 & 2/\sqrt{14} \\ 1/2 & 1/\sqrt{14} \\ 1/2 & -3/\sqrt{14} \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ -3 \end{bmatrix}$$

## Section 6.3 : Orthogonal Projections

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



Vectors  $\vec{e}_1$  and  $\vec{e}_2$  form an orthonormal basis for subspace  $W$ .

Vector  $\vec{y}$  is not in  $W$ .

The orthogonal projection of  $\vec{y}$  onto  $W = \text{Span}\{\vec{e}_1, \vec{e}_2\}$  is  $\hat{y}$ .



# Topics and Objectives

## Topics

1. Orthogonal projections and their basic properties
2. Best approximations

## Learning Objectives

1. Apply concepts of orthogonality and projections to
  - a) compute orthogonal projections and distances,
  - b) express a vector as a linear combination of orthogonal vectors,
  - c) construct vector approximations using projections,
  - d) characterize bases for subspaces of  $\mathbb{R}^n$ , and
  - e) construct orthonormal bases.

**Motivating Question** For the matrix  $A$  and vector  $\vec{b}$ , which vector  $\hat{b}$  in column space of  $A$ , is closest to  $\vec{b}$ ?

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -4 & -2 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

## Example 1

Let  $\vec{u}_1, \dots, \vec{u}_5$  be an orthonormal basis for  $\mathbb{R}^5$ . Let  $W = \text{Span}\{\vec{u}_1, \vec{u}_2\}$ . For a vector  $\vec{y} \in \mathbb{R}^5$ , write  $\vec{y} = \hat{y} + w^\perp$ , where  $\hat{y} \in W$  and  $w^\perp \in W^\perp$ .

## Orthogonal Decomposition Theorem

### Theorem

Let  $W$  be a subspace of  $\mathbb{R}^n$ . Then, each vector  $\vec{y} \in \mathbb{R}^n$  has the **unique** decomposition

$$\vec{y} = \hat{y} + w^\perp, \quad \hat{y} \in W, \quad w^\perp \in W^\perp.$$

And, if  $\vec{u}_1, \dots, \vec{u}_p$  is any orthogonal basis for  $W$ ,

$$\hat{y} = \frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 + \dots + \frac{\vec{y} \cdot \vec{u}_p}{\vec{u}_p \cdot \vec{u}_p} \vec{u}_p.$$

We say that  $\hat{y}$  is the **orthogonal projection of  $\vec{y}$  onto  $W$** .

If time permits, we will explain some of this theorem on the next slide.

## Explanation (if time permits)

We can write

$$\hat{y} =$$

Then,  $w^\perp = \vec{y} - \hat{y}$  is in  $W^\perp$  because

## Example 2a

$$\vec{y} = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix}, \quad \vec{u}_1 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad \vec{u}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Construct the decomposition  $\vec{y} = \hat{y} + w^\perp$ , where  $\hat{y}$  is the orthogonal projection of  $\vec{y}$  onto the subspace  $W = \text{Span}\{\vec{u}_1, \vec{u}_2\}$ .

## Best Approximation Theorem

### Theorem

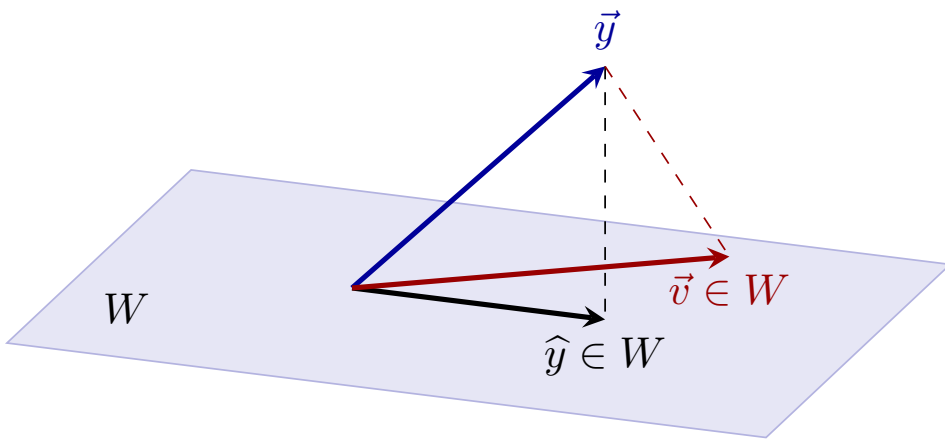
Let  $W$  be a subspace of  $\mathbb{R}^n$ ,  $\vec{y} \in \mathbb{R}^n$ , and  $\hat{y}$  is the orthogonal projection of  $\vec{y}$  onto  $W$ . Then for **any**  $\vec{w} \neq \hat{y} \in W$ , we have

$$\|\vec{y} - \hat{y}\| < \|\vec{y} - \vec{w}\|$$

That is,  $\hat{y}$  is the unique vector in  $W$  that is closest to  $\vec{y}$ .

## Proof (if time permits)

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



## Example 2b

$$\vec{y} = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix}, \quad \vec{u}_1 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad \vec{u}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

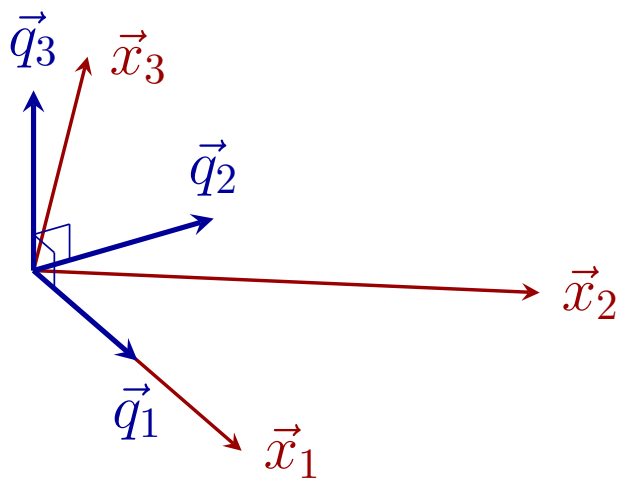
What is the distance between  $\vec{y}$  and subspace  $W = \text{Span}\{\vec{u}_1, \vec{u}_2\}$ ? Note that these vectors are the same vectors that we used in Example 2a.



## Section 6.4 : The Gram-Schmidt Process

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



Vectors  $\vec{x}_1, \vec{x}_2, \vec{x}_3$  are given linearly independent vectors. We wish to construct an orthonormal basis  $\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}$  for the space that they span.

# Topics and Objectives

## Topics

1. Gram Schmidt Process
2. The  $QR$  decomposition of matrices and its properties

## Learning Objectives

1. Apply the iterative Gram Schmidt Process, and the QR decomposition, to construct an orthogonal basis.
2. Compute the  $QR$  factorization of a matrix.

**Motivating Question** The vectors below span a subspace  $W$  of  $\mathbb{R}^4$ . Identify an orthogonal basis for  $W$ .

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

## Example

The vectors below span a subspace  $W$  of  $\mathbb{R}^4$ . Construct an orthogonal basis for  $W$ .

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

## The Gram-Schmidt Process

Given a basis  $\{\vec{x}_1, \dots, \vec{x}_p\}$  for a subspace  $W$  of  $\mathbb{R}^n$ , iteratively define

$$\vec{v}_1 = \vec{x}_1$$

$$\vec{v}_2 = \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1$$

$$\vec{v}_3 = \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2$$

$$\vdots$$

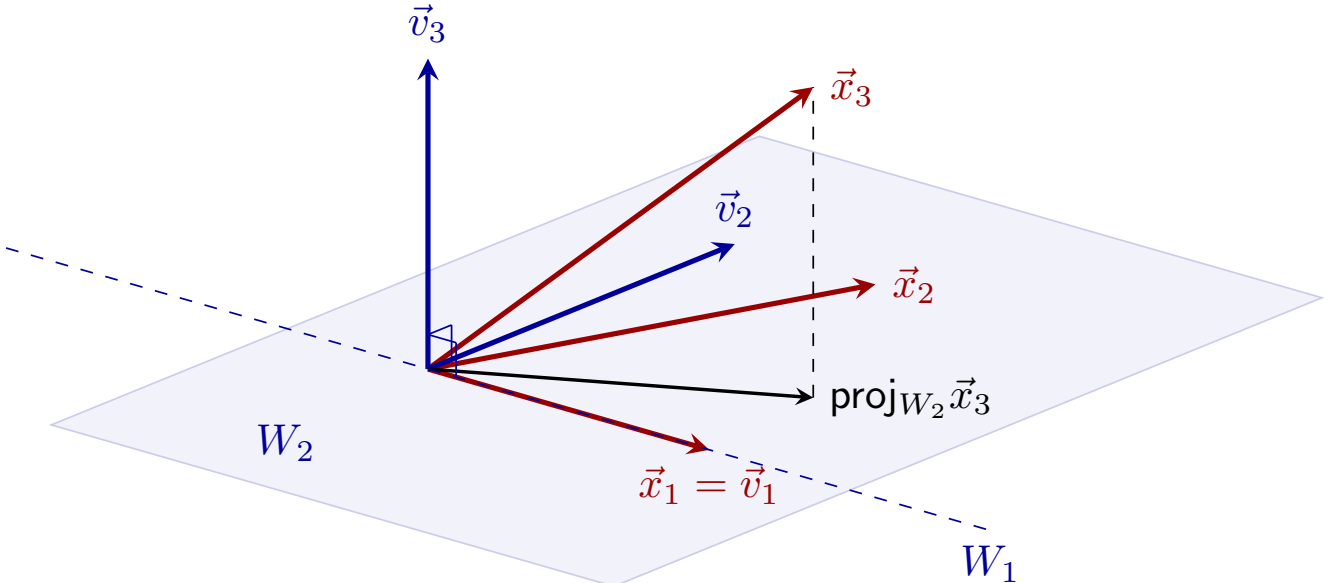
$$\vec{v}_p = \vec{x}_p - \frac{\vec{x}_p \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \dots - \frac{\vec{x}_p \cdot \vec{v}_{p-1}}{\vec{v}_{p-1} \cdot \vec{v}_{p-1}} \vec{v}_{p-1}$$

Then,  $\{\vec{v}_1, \dots, \vec{v}_p\}$  is an orthogonal basis for  $W$ .

# Proof

## Geometric Interpretation

Suppose  $\vec{x}_1, \vec{x}_2, \vec{x}_3$  are linearly independent vectors in  $\mathbb{R}^3$ . We wish to construct an orthogonal basis for the space that they span.



We construct vectors  $\vec{v}_1, \vec{v}_2, \vec{v}_3$ , which form our **orthogonal** basis.

$$W_1 = \text{Span}\{\vec{v}_1\}, W_2 = \text{Span}\{\vec{v}_1, \vec{v}_2\}.$$

## Orthonormal Bases

### Definition

A set of vectors form an **orthonormal basis** if the vectors are mutually orthogonal and have unit length.

### Example

The two vectors below form an orthogonal basis for a subspace  $W$ . Obtain an orthonormal basis for  $W$ .

$$\vec{v}_1 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}.$$

# QR Factorization

## Theorem

Any  $m \times n$  matrix  $A$  with linearly independent columns has the **QR factorization**

$$A = QR$$

where

1.  $Q$  is  $m \times n$ , its columns are an orthonormal basis for  $\text{Col } A$ .
2.  $R$  is  $n \times n$ , upper triangular, with positive entries on its diagonal, and the length of the  $j^{\text{th}}$  column of  $R$  is equal to the length of the  $j^{\text{th}}$  column of  $A$ .

In the interest of time:

- we will not consider the case where  $A$  has linearly dependent columns
- students are not expected to know the conditions for which  $A$  has a QR factorization



# Proof

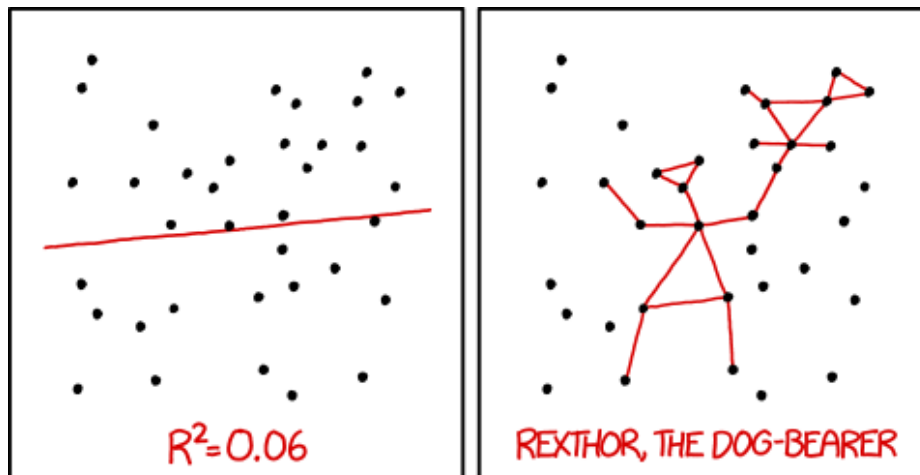
## Example

Construct the  $QR$  decomposition for  $A = \begin{bmatrix} 3 & -2 \\ 2 & 3 \\ 0 & 1 \end{bmatrix}$ .

## Section 6.5 : Least-Squares Problems

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER  
TO GUESS THE DIRECTION OF THE CORRELATION FROM THE  
SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

<https://xkcd.com/1725>

# Topics and Objectives

## Topics

1. Least Squares Problems
2. Different methods to solve Least Squares Problems

## Learning Objectives

1. Compute general solutions, and least squares errors, to least squares problems using the normal equations and the  $QR$  decomposition.

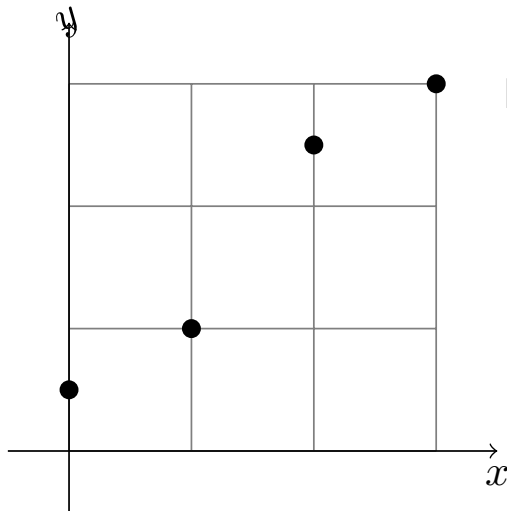
**Motivating Question** A series of measurements are corrupted by random errors. How can the dominant trend be extracted from the measurements with random error?

# Inconsistent Systems

Suppose we want to construct a line of the form

$$y = mx + b$$

that best fits the data below.



From the data, we can construct the system:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 2.5 \\ 3 \end{bmatrix}$$

Can we 'solve' this inconsistent system?

# The Least Squares Solution to a Linear System

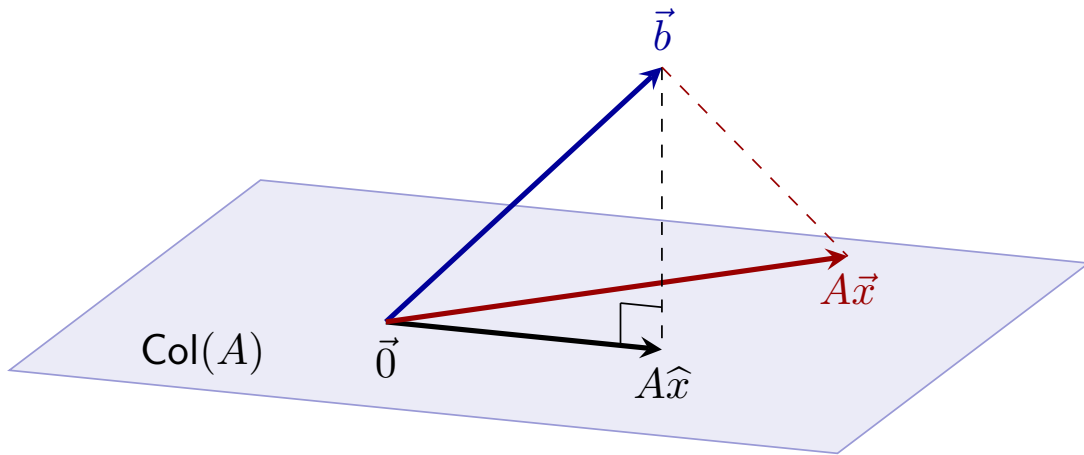
## Definition: Least Squares Solution

Let  $A$  be a  $m \times n$  matrix. A **least squares solution** to  $A\vec{x} = \vec{b}$  is the solution  $\hat{x}$  for which

$$\|\vec{b} - A\hat{x}\| \leq \|\vec{b} - A\vec{x}\|$$

for all  $\vec{x} \in \mathbb{R}^n$ .

## A Geometric Interpretation



The vector  $\vec{b}$  is closer to  $A\hat{x}$  than to  $A\vec{x}$  for all other  $\vec{x} \in \text{Col} A$ .

1. If  $\vec{b} \in \text{Col} A$ , then  $\hat{x}$  is ...
2. Seek  $\hat{x}$  so that  $A\hat{x}$  is as close to  $\vec{b}$  as possible. That is,  $\hat{x}$  should solve  $A\hat{x} = \hat{b}$  where  $\hat{b}$  is ...

# The Normal Equations

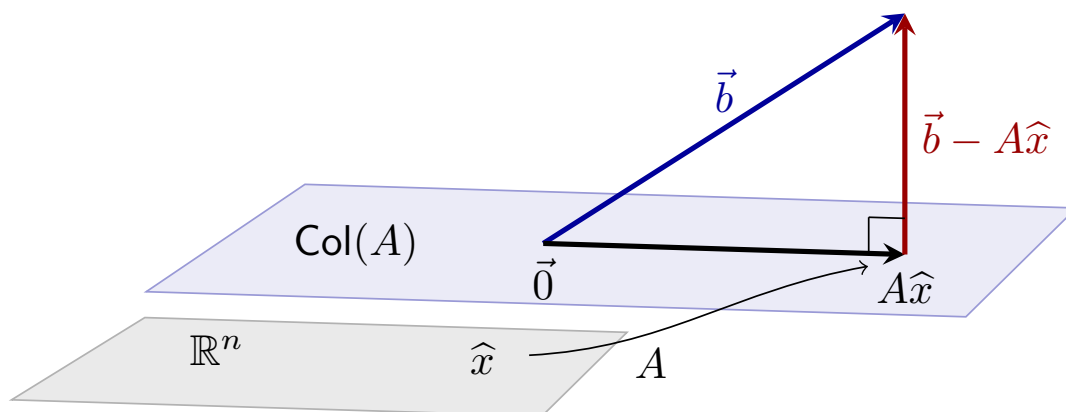
## Theorem (Normal Equations for Least Squares)

The least squares solutions to  $A\vec{x} = \vec{b}$  coincide with the solutions to

$$\underbrace{A^T A \vec{x} = A^T \vec{b}}_{\text{Normal Equations}}$$



## Derivation



The least-squares solution  $\hat{x}$  is in  $\mathbb{R}^n$ .

1.  $\hat{x}$  is the least squares solution, is equivalent to  $\vec{b} - A\hat{x}$  is orthogonal to   $A$ .
2. A vector  $\vec{v}$  is in  $\text{Null } A^T$  if and only if   $\vec{v} = \vec{0}$ .
3. So we obtain the Normal Equations:

## Example

Compute the least squares solution to  $A\vec{x} = \vec{b}$ , where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

**Solution:**

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} =$$

$$A^T \vec{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} =$$

The normal equations  $A^T A \vec{x} = A^T \vec{b}$  become:

# Theorem

## Theorem (Unique Solutions for Least Squares)

Let  $A$  be any  $m \times n$  matrix. These statements are equivalent.

1. The equation  $A\vec{x} = \vec{b}$  has a unique least-squares solution for each  $\vec{b} \in \mathbb{R}^m$ .
2. The columns of  $A$  are linearly independent.
3. The matrix  $A^T A$  is invertible.

And, if these statements hold, the least square solution is

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}.$$

Useful heuristic:  $A^T A$  plays the role of 'length-squared' of the matrix  $A$ . (See the sections on symmetric matrices and singular value decomposition.)

## Example

Compute the least squares solution to  $A\vec{x} = \vec{b}$ , where

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

Hint: the columns of  $A$  are orthogonal.



### Theorem (Least Squares and $QR$ )

Let  $m \times n$  matrix  $A$  have a  $QR$  decomposition. Then for each  $\vec{b} \in \mathbb{R}^m$  the equation  $A\vec{x} = \vec{b}$  has the unique least squares solution

$$R\hat{x} = Q^T\vec{b}.$$

(Remember,  $R$  is upper triangular, so the equation above is solved by back-substitution.)

**Example 3.** Compute the least squares solution to  $A\vec{x} = \vec{b}$ , where

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

**Solution.** The  $QR$  decomposition of  $A$  is

$$A = QR = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$$



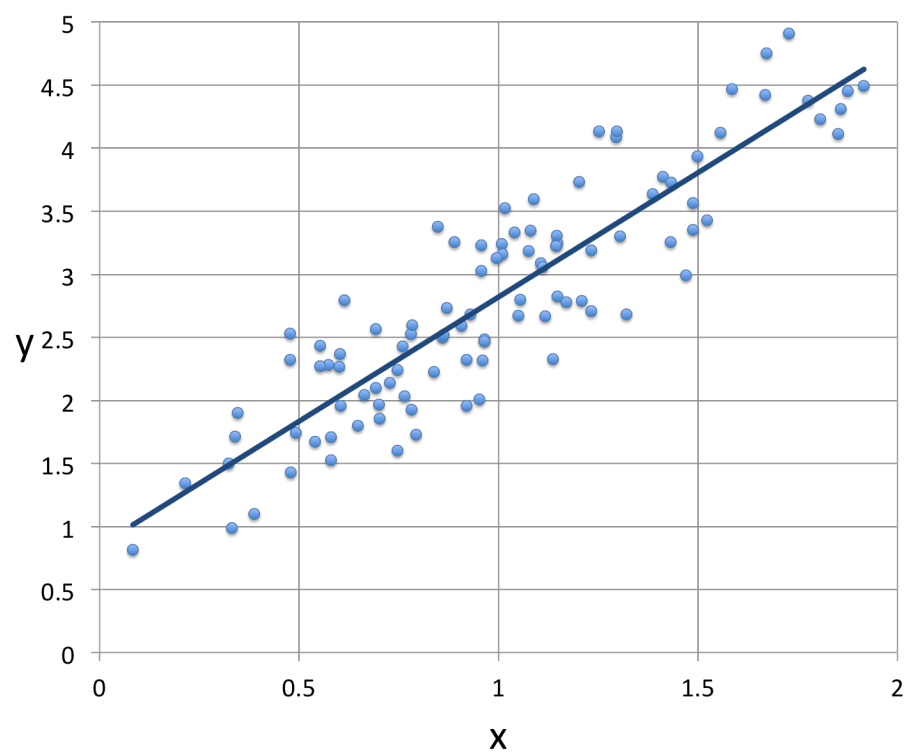
$$Q^T \vec{b} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} -6 \\ 4 \end{bmatrix}$$

And then we solve by backwards substitution  $R\vec{x} = Q^T \vec{b}$

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -6 \\ -6 \\ 4 \end{bmatrix}$$

## Chapter 6 : Orthogonality and Least Squares

### 6.6 : Applications to Linear Models



# Topics and Objectives

## Topics

1. Least Squares Lines
2. Linear and more complicated models

## Learning Objectives

For the topics covered in this section, students are expected to be able to do the following.

1. Apply least-squares and multiple regression to construct a linear model from a set of data points.
2. Apply least-squares to fit polynomials and other curves to data.

## Motivating Question

Compute the equation of the line  $y = \beta_0 + \beta_1 x$  that best fits the data

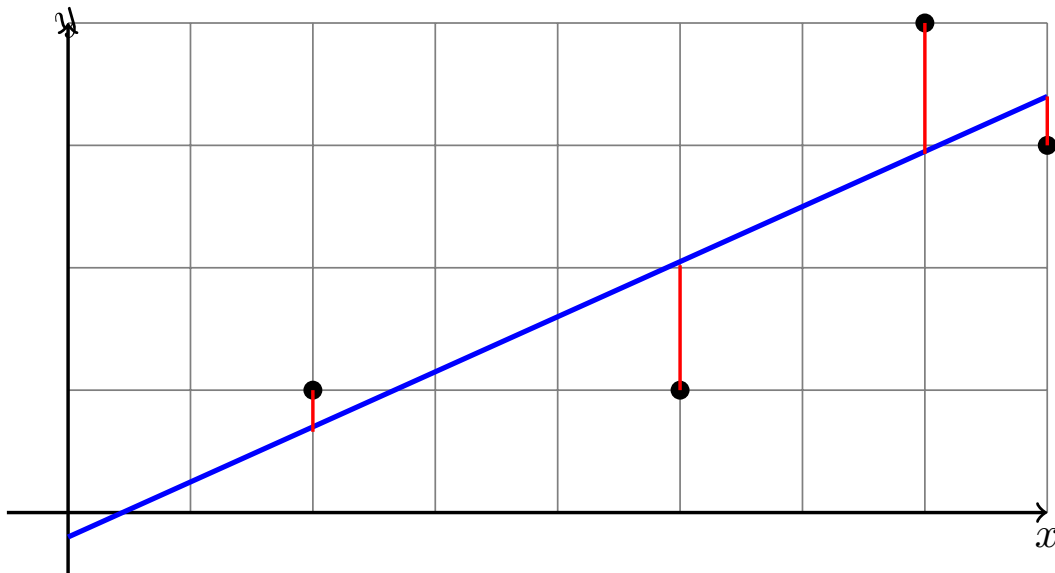
$x$	2	5	7	8
$y$	1	1	4	3

# The Least Squares Line

Graph below gives an approximate linear relationship between  $x$  and  $y$ .

1. Black circles are data.
2. Blue line is the **least squares** line.
3. Lengths of red lines are the \_\_\_\_\_.

The least squares line minimizes the sum of squares of the \_\_\_\_\_.



**Example 1** Compute the least squares line  $y = \beta_0 + \beta_1 x$  that best fits the data

$x$	2	5	7	8
$y$	1	1	4	3

We want to solve

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 4 \\ 3 \end{bmatrix}$$

This is a least-squares problem :  $X\vec{\beta} = \vec{y}$ .

The normal equations are

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$
$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 59 \\ 9 \\ 59 \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

So the least-squares solution is given by

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

$$y = \beta_0 + \beta_1 x = \frac{-5}{21} + \frac{19}{42}x$$

As we may have guessed,  $\beta_0$  is negative, and  $\beta_1$  is positive.

## Least Squares Fitting for Other Curves

We can consider least squares fitting for the form

$$y = c_0 + c_1 f_1(x) + c_2 f_2(x) + \cdots + c_k f_k(x).$$

If functions  $f_i$  are known, this is a linear problem in the  $c_i$  variables.

### Example

Consider the data in the table below.

$x$	-1	0	0	1
$y$	2	1	0	6

Determine the coefficients  $c_1$  and  $c_2$  for the curve  $y = c_1 x + c_2 x^2$  that best fits the data.

## WolframAlpha and Mathematica Syntax

Least squares problems can be computed with WolframAlpha, Mathematica, and many other software.

### WolframAlpha

`linear fit  $\{\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_n, y_n\}\}$`

### Mathematica

`LeastSquares[ $\{\{x_1, x_1, y_1\}, \{x_2, x_2, y_2\}, \dots, \{x_n, x_n, y_n\}\}$ ]`

Almost any spreadsheet program does this as a function as well.