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.

proof

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

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

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

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

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

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

$$\phi_B(\lambda) = \det(B-\lambda I)$$

Similar Matrices

Definition

Two $n \times n$ matrices A and B are **similar** if there is a invertible 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,

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

$$\phi_A = \lambda^2 \qquad \phi_B = \lambda^2$$

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

A matrix is **diagonal** if the only non-zero elements, if any, are on the main diagonal.

The following are all diagonal matrices.

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

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,

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

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

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

$$\begin{pmatrix} a_1 & & & 0 \\ & a_2 & & \\ & & \ddots & \\ 0 & & & a_n \end{pmatrix}$$

But what if A is not diagonal?

① A is similar to diagonal : $A = P \cdot \overset{\shortparallel}{D} \cdot P^{-1}$

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

$$= P \cdot D \cdot \underbrace{(P^{-1} \cdot P)}_{= I} \cdot D \cdot P^{-1} = P \cdot D \cdot D \cdot P^{-1}$$

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

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

$$= P \cdot \begin{pmatrix} a_1^k & & & 0 \\ & a_2^k & & \\ & & \ddots & \\ 0 & & & a_n^k \end{pmatrix} P^{-1}$$

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a+b \\ 0 & 1 \end{pmatrix}$$

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

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

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

Q: When can we diagonalize A ?

Q: How?

$A \in \mathbb{R}^{n \times n}$, $\lambda_1, \lambda_2, \cdots, \lambda_n$: eigenvalues

$v_1 \quad v_2 \quad v_n$: eigenvectors

$$\left.\begin{aligned} A v_1 &= \lambda_1 v_1 \\ A v_2 &= \lambda_2 v_2 \\ &\vdots \\ A v_n &= \lambda_n v_n \end{aligned}\right\} \quad n \text{ matrix eqn} \Rightarrow 1 \text{ Equation.}$$

$$A \cdot \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \\ | & | & & | \end{bmatrix}$$

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

$$\boxed{AP = PD}$$

If P is invertible

$$\Longrightarrow A = PDP^{-1}.$$

Diagonalization

$P = (v_1 \; - - - \; v_n)$

P 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

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

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

Q: When do we have n lin. indep. eigenvectors?

Example 1

Diagonalize if possible.

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

① Eigenvalues :

$$\phi(\lambda) = \det(A - \lambda I) = \det\begin{pmatrix} 2-\lambda & 6 \\ 0 & -1-\lambda \end{pmatrix}$$

$$= \lambda^2 - (2+(-1))\lambda + (2\cdot(-1) - 6\cdot 0)$$

$$= \lambda^2 - \lambda - 2 = 0$$

$$\lambda = 2, -1$$

② <u>Eigenvectors</u>

$\lambda = 2$: $\quad v_1 \overset{\neq 0}{\in} E_2 = \mathrm{Nul}(A - 2I)$

$$A - 2I = \begin{pmatrix} 0 & 6 \\ 0 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 \\ \hline 0 & 0 \end{pmatrix}$$

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \boxed{y=0}$$

$\lambda = -1$: $\quad v_2 \in E_{-1} = \mathrm{Nul}(A - (-1)\cdot I)$

$$A + I = \begin{pmatrix} 3 & 6 \\ 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 \\ \hline 0 & 0 \end{pmatrix} \qquad \begin{aligned} x+2y&=0 \\ x &= -2y \end{aligned}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2y \\ y \end{bmatrix} = y \underbrace{\begin{bmatrix} -2 \\ 1 \end{bmatrix}}, \quad v_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

$$\begin{tikzcd} \lambda = & 2, \arrow[d, dash] & -1 \arrow[d, dash] & \\ & \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \begin{bmatrix} -2 \\ 1 \end{bmatrix} & \arrow[l] \text{lin. indep.} \end{tikzcd}$$

$\Rightarrow$ A is diagonalizable

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

Note Suppose λ_1, λ_2 : distinct eigenvalues

$$\begin{tikzcd} \lambda_1, \arrow[d, dash] & \lambda_2 \arrow[d, dash] \\ v_1 \neq 0 & v_2 \boxed{\neq 0} \end{tikzcd}$$

$\Rightarrow$ v_1, v_2 are linearly indep.

Proof Assume $a v_1 + b v_2 = 0$ (Goal: $a, b = 0$)

- if a or b are zero, both should be zero.
- suppose $a, b \neq 0$.

$$0 = A(a v_1 + b v_2) = a(\boxed{A v_1}) + b \boxed{A v_2}$$

$$a \lambda_1 v_1 + b \lambda_2 v_2 = 0$$

$$-\,\underline{\big|\, a \lambda_1 v_1 + b \lambda_1 v_2 = 0}$$

$$0 + \underset{\neq 0}{b}\,\underset{\neq 0}{\underline{(\lambda_2 - \lambda_1)}} \cdot v_2 = 0$$

$\Rightarrow$ $v_2 = 0$: contradiction.

Distinct Eigenvalues

Theorem

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

Why does this theorem hold?

Thm

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

$v_1 \quad v_2 \quad \ldots \quad v_n \quad \Rightarrow \quad \boxed{\{v_1, \ldots, v_n\} : \text{lin. indep.}}$

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

Recall

$A \in \mathbb{R}^{n \times n}$ is diagonalizable

$\underset{\text{definition}}{\Longleftrightarrow}$ There exists an invertible matrix P and a diagonal matrix D such that

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

Suppose $\lambda_1, \lambda_2, \cdots, \lambda_n$ are eigenvalues w/ eigenvectors $v_1, v_2, \cdots, v_n$, then

$$A \underbrace{[v_1 \ v_2 \cdots v_n]}_{P} = [\lambda_1 v_1 \ \ \lambda_2 v_2 \ \cdots \ \lambda_n v_n]$$

$$= \underbrace{[v_1 \cdots v_n]}_{P} \cdot \underbrace{\begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}}_{D}$$

$AP = PD$

Q: Is P invertible?

This is true when $\{v_1, \cdots, v_n\}$ is linearly indep.

① If $\lambda_1, \cdots, \lambda_n$ are distinct, then $\{v_1, \cdots, v_n\}$ are linearly indep. $\Longrightarrow$ A is diagonalizable.

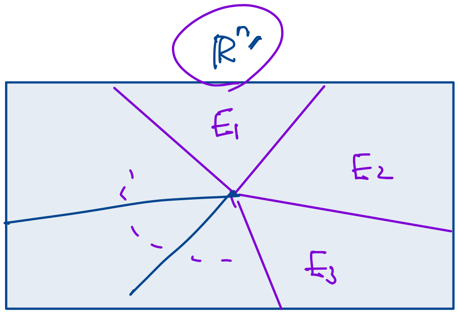
Today's Question: What if $\lambda_1, \cdots, \lambda_n$ are NOT distinct

If $\sum_i \dim(E_i) = n$

then we can choose

n linearly indep. eigenvectors.

$\Rightarrow$ A is diagonalizable.



E_i : eigenspaces

$E_i = \mathrm{Nul}\,(A - \lambda_i I)$

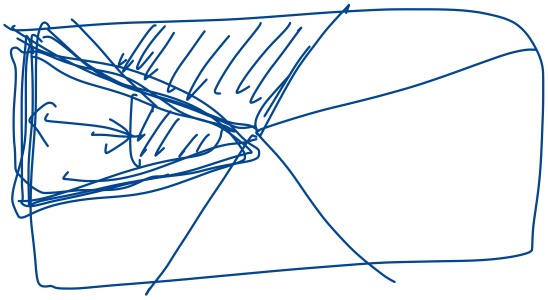
Non-Distinct Eigenvalues

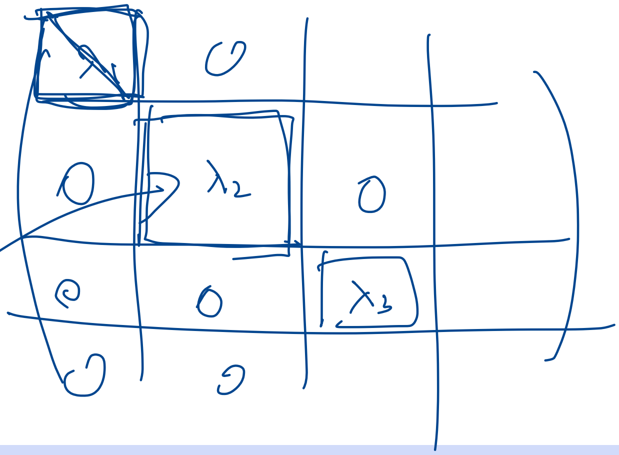
Theorem. Suppose

- A is $n \times n$
- A has distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, $k \leq n$
- a_i = algebraic multiplicity of λ_i
- d_i = dimension of λ_i eigenspace ("geometric multiplicity")

Then

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





Example 2

Diagonalize if possible.

$$\boxed{\begin{pmatrix}3 & 1\\ 0 & 3\end{pmatrix}^k = \left(3\cdot\begin{pmatrix}1 & \frac{1}{3}\\ 0 & 1\end{pmatrix}\right)^k = 3^k\cdot\begin{pmatrix}1 & \frac{1}{3}\\ 0 & 1\end{pmatrix}^k}$$

$$\| \qquad \begin{pmatrix}1 & \frac{k}{3}\\ 0 & 1\end{pmatrix}$$

① Eigenvalue : $\phi(\lambda) = \lambda^2 - (3+3)\lambda + (3^2 - 1\cdot 0)$

$= \lambda^2 - 6\lambda + 9 = (\lambda-3)^{②}$

$\lambda = 3$. with alg. multi ②

② Eigenspace E_3 = Nul $(A - 3I)$

$A - 3I = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}$ $\quad y = 0$. $\quad \begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}x\\ 0\end{bmatrix} = \underset{\sim}{x}\,\underset{\sim\sim}{\begin{bmatrix}1\\ 0\end{bmatrix}}$

$\dim(E_3) = 1$ = Geom. Multi. $\lneq 2$

A is NOT diagonalizable.

Example 3

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

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

$E_1 = \text{Nul}(A - I) : \quad \Rightarrow \quad G. M = 1$

$$A - I = \begin{pmatrix} 6 & 4 & 16 \\ 2 & 4 & 8 \\ -2 & -2 & -6 \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & 2 & 8 \\ 1 & 2 & 4 \\ 1 & 1 & 3 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 2 & 4 \\ 0 & -4 & -4 \\ 0 & -1 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$x_1 + 2x_3 = 0$

$x_2 + x_3 = 0$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_3 \\ -x_3 \\ x_3 \end{bmatrix} = x_3 \cdot \underline{\begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix}}$$

$E_3 = \text{Nul}(A - 3I) \quad \Rightarrow \quad \text{Geom. multi} = 2$

$$A - 3I = \begin{pmatrix} 4 & 4 & 16 \\ 2 & 2 & 8 \\ -2 & -2 & -8 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$x_1 + x_2 + 4x_3 = 0$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_2 - 4x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \cdot \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}$$

$$P = \begin{array}{r} \\ \lambda=1 \nearrow \end{array} \left[\begin{array}{ccc} \Leftarrow v_1 & \Leftarrow v_2 & \Leftarrow v_3 \\ -2 & -1 & -4 \\ -1 & 1 & 0 \\ 1 & \underbrace{0 & 1}_{\lambda=3} \end{array}\right] \qquad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

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

$$\phi(\lambda) = (\lambda-1)(\lambda-3)^2$$

- $\mathcal{B} = \{ \underbrace{v_1, v_2, v_3}\} \leftarrow$ a basis

$$[Av_1]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$Av_1 = \overset{\checkmark}{\lambda_1} v_1 + \overset{\checkmark}{0} \cdot v_2 + \overset{\checkmark}{0} \cdot v_3$$

$$[Av_2]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, \qquad [Av_3]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

Additional Example (if time permits)

Note that

$$\underbrace{1 \quad 1}\ \underbrace{\ \ 2}\quad 3 \quad 5 \quad 8 \quad 13 \quad 21 \quad \dots \overset{k}{\downarrow}$$

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

generates a well-known sequence of numbers.

$$x_1 = \begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} = \begin{bmatrix}1\\ 1+1\end{bmatrix} = \begin{bmatrix}1\\ 2\end{bmatrix} \qquad x_2 = \begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}\cdot\begin{bmatrix}1\\ 2\end{bmatrix} = \begin{bmatrix}2\\ 1+2\end{bmatrix} = \begin{bmatrix}2\\ 3\end{bmatrix}$$

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

$$x_3 = \begin{bmatrix}3\\ 5\end{bmatrix} \qquad x_4 = \begin{bmatrix}5\\ 8\end{bmatrix} \qquad x_5 = \begin{bmatrix}8\\ 13\end{bmatrix} \qquad x_6 = \begin{bmatrix}13\\ 21\end{bmatrix} \quad \dots\dots$$

$$x_k = \underbrace{\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}}^{k}\cdot\begin{bmatrix}1\\ 1\end{bmatrix} = \begin{bmatrix}\bigcirc\\ \bigcirc\end{bmatrix}$$

<u>Exercise</u>.

$$\phi(\lambda) = \underline{\lambda^2 - \lambda - 1 = 0}$$

$$\lambda = \frac{1 \pm \sqrt{5}}{2}$$

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:

If $x \in \mathbb{R}$, $x^2 \geqslant 0$.

$$x^2 + 1 = 0$$

The roots of this equation are:

$$x^2 = -1$$

$$x = \pm\sqrt{-1}$$

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

The set of complex numbers $= \mathbb{C}$

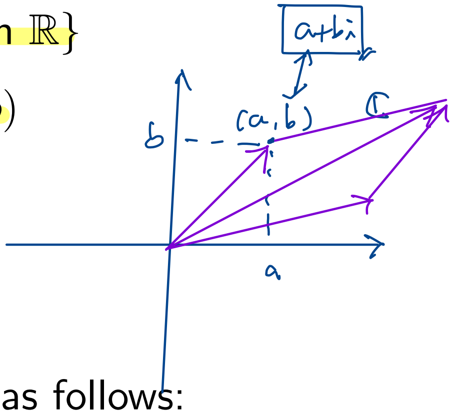
$= \{ a + bi : a, b \in \mathbb{R} \}$

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 $\mathbb{R}^2$: $\quad a + bi \leftrightarrow (a, b)$



vector addition

Q: geometric meaning?

We can add and multiply complex numbers as follows:

$$(2-3i) + (-1+i) = \big(2+(-1)\big) + \big((-3)+1\big)\cdot i = 1 - 2i$$

Componentwise

$$(2-3i)(-1+i) = \underbrace{2(-1)} + \underbrace{2\cdot i} + \underbrace{(-3i)\cdot(-1)} + \underbrace{(-3\overbrace{\boxed{i})\cdot\boxed{i}}^{-1}}$$

$$= \underbrace{-2} + \underbrace{2i} + \underbrace{3i} + \underbrace{3}.$$

$$= 1 + 5i$$

$i = \sqrt{-1}$

$i^2 = -1$

Complex Conjugate, Absolute Value, Polar Form

real part

We can **conjugate** complex numbers: $\overline{a+bi} = \underline{a \; - \; bi}$

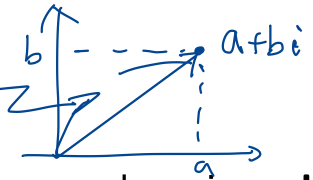
imaginary part.

$z = a+bi$, $w = c+di \in \mathbb{C}$

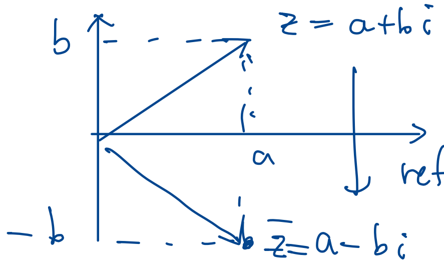
- $\overline{(\overline{z})} = z$
- $\overline{z+w} = \overline{z} + \overline{w}$
- $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$
- If $\overline{z} = z$ then $z \in \mathbb{R}$
- $z \cdot \overline{z} = (a+bi)\cdot(a-bi) = a^2-(bi)^2$
  $= a^2+b^2 \geqslant 0$ (If $z \cdot \overline{z} = 0$ then $z=0$)

The **absolute value** of a complex number: $|a+bi| = \underline{\sqrt{z \cdot \overline{z}} = \sqrt{a^2+b^2}}$

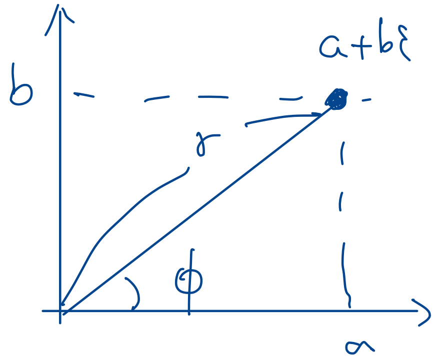
length $= \sqrt{a^2+b^2}$
$= |z|$



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



reflection with respect to x-axis



$a = r \cdot \cos\phi$

$b = r \cdot \sin\phi$

$a+bi = r\cos\phi + r\sin\phi \cdot i$

$= r \cdot (\cos\phi + i\sin\phi)$.

Notation $z = a+bi$. $\mathrm{Re}(z) = a$, $\mathrm{Im}(z) = b$

Complex Conjugate Properties

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

- $\overline{(x+y)} = \overline{x} + \overline{y}$
- $\overline{A\vec{v}} = A\overline{\vec{v}}$ $A \in \mathbb{R}^{n\times n}$ $v = (v_1, \dots, v_n)$ $v_1, v_2, \dots, v_n \in \mathbb{C}$
- $\mathrm{Im}(x\overline{x}) = 0$. ✓ (∵ $x \cdot \overline{x} = a^2 + b^2$ if $x = a+bi$)

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

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

Exercice : try $x = a+bi$, $y = c+di$

$$\overline{A\vec{v}} = \overline{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}} = \begin{bmatrix} \overline{w_1} \\ \overline{w_2} \\ \vdots \\ \overline{w_n} \end{bmatrix} = \overset{A}{\overset{\shortparallel}{\overline{A}}} \cdot \overline{v} = \begin{bmatrix} \overline{a_{11}} & \cdots & \\ \vdots & \overline{a_{ij}} & \\ \vdots & & \overline{a_{nn}} \end{bmatrix} \cdot \begin{bmatrix} \overline{v_1} \\ \vdots \\ \overline{v_n} \end{bmatrix}$$

$\uparrow$

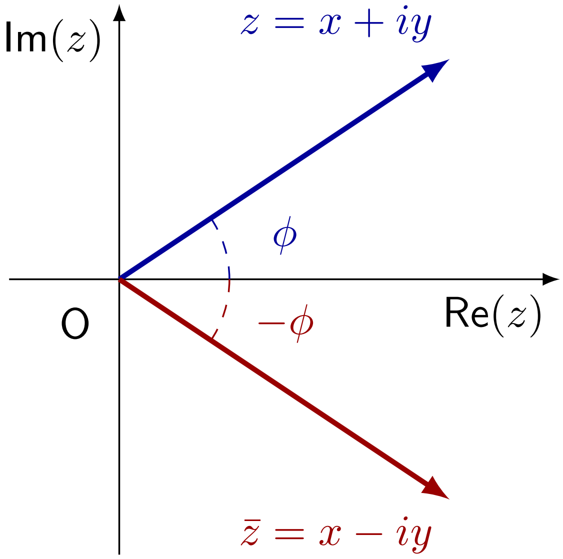
$a_{ij} \in \mathbb{R}$

$\overline{a_{ij}} = a_{ij}$

$$\overline{(A \cdot \vec{v})} = A \cdot \overline{v}$$

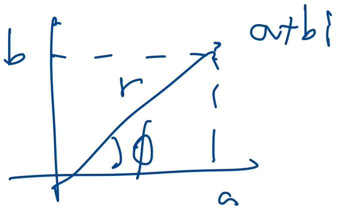
Polar Form and the Complex Conjugate

Conjugation reflects points across the real axis.



$\phi \in \mathbb{R}$

$$\boxed{e^{i\phi} = \cos\phi + i\sin\phi}.$$

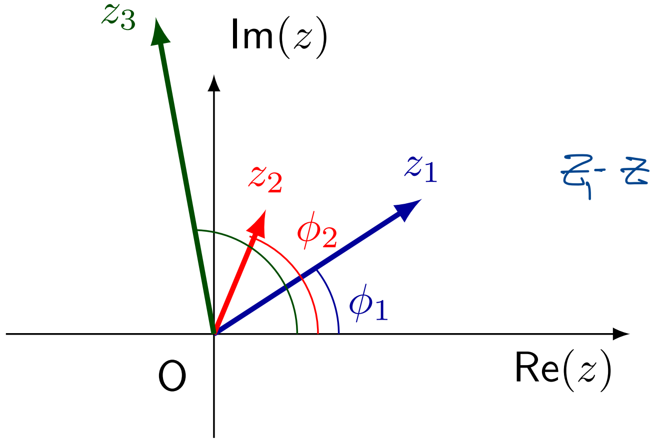


$$z = a+bi = r\cdot(\underline{\cos\phi + i\sin\phi})$$
$$= \underset{\smile}{r}\cdot e^{i\phi} = |z|\cdot e^{i\phi}.$$

Euler's Formula : Geometric meaning of multiplication.

Suppose z_1 has angle ϕ_1, and z_2 has angle ϕ_2.

$$z_1 = |z_1|\cdot e^{i\phi_1}$$
$$z_2 = |z_2|\, e^{i\phi_2}$$
$$z_1\cdot z_2 = \underbrace{(|z_1|\cdot|z_2|)}_{\text{length}}\cdot e^{i\cdot\underbrace{(\phi_1+\phi_2)}_{\text{angle}}}$$



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

$$\boxed{z = |z|\,\mathrm{e}^{i\phi}}$$

The product $z_1 z_2$ is:

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

Recall $A \in \mathbb{R}^{n\times \underline{n}} \rightsquigarrow \phi_A(\lambda) = \underline{\det(A-\lambda I)} = 0$

a polynomial of λ (degree $\underline{n}$)

Roots of $\phi_A(\lambda)=0$ = Eigenvalues.

$(x-3)(x+1) = \underline{x^2-2x-3} = 0$ $\quad$ $x = 3, -1$

Complex Numbers and Polynomials

Theorem: Fundamental Theorem of Algebra

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

Roots are $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$

$$\underline{\phi_A(\lambda)} = \underline{(\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda)} = 0 .$$

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 λ is an eigenvalue of real matrix A with eigenvector $\vec{v}$, then $\bar{\lambda}$ is an eigenvalue of A with eigenvector $\bar{\vec{v}}$.

$$\phi_A(\lambda) = \det(\underline{\underline{A}}-\lambda I) = \underset{\uparrow}{a_n}\cdot\lambda^n + \underset{\uparrow}{a_{n-1}}\cdot\lambda^{n-1} + \dots + \underset{\uparrow}{a_1}\cdot\lambda + \underset{\uparrow}{a_0}$$

real numbers $\quad$ " $\quad$ "

Suppose $z \in \mathbb{C}$ is a root of $\phi_A(\lambda) = 0$

$\phi_A(z) = 0$ $\qquad$ $\overline{a_n\cdot z^n} + \overline{a_{n-1} z^{n-1}} + \dots + \overline{a_1 z} + \overline{a_0} = \overline{0}$

$\overline{a_n}\cdot(\overline{z^n}) + \overline{a_{n-1}}\cdot(\overline{z^{n-1}}) + \dots + \overline{a_1}\cdot\overline{z} + \overline{a_0} = 0$ $\qquad$ $a_i \in \mathbb{R}$

$a_n\cdot(\overline{z})^n + a_{n-1}(\overline{z})^{n-1} + \dots + a_1\cdot\overline{z} + a_0 = 0$

$\phi_A(\bar{z}) = 0$ That is, $\bar{z}$ is a root.

<u>Recall</u> $\mathbb{C} = \{ a + bi : a, b \in \mathbb{R} \}$

$z = a + bi$ $Re(z) = a$, $Im(z) = b$

$\bar{z} = a - bi$ (Conjugate)

$|z| = \sqrt{a^2 + b^2} = \sqrt{z \cdot \bar{z}}$

$\overline{z + w} = \bar{z} + \bar{w}$, $\overline{z \cdot w} = \bar{z} \cdot \bar{w}$

$A \in \mathbb{R}^{n \times n}$, $\overline{A \cdot v} = \underset{\sim}{\bar{A}} \cdot \bar{v} = A \cdot \bar{v}$

$v \in \mathbb{C}^n$

Suppose $A \in \mathbb{R}^{n \times n}$, if $\overset{\text{eigenvalue}}{z}$ is a root of $\phi_A(\lambda) = \overset{= 0}{\det(A - \lambda I)}$

then $\bar{z}$ is also a root of $\phi_A(\lambda) = 0$.

Furthermore, if $v \neq 0$ is an eigenvector for z

$\Rightarrow \overline{\underset{\sim}{A \cdot v}} = \overline{z \cdot v}$

$A \cdot \bar{v} = \bar{z} \cdot \bar{v}$

$\Rightarrow \bar{v}$ is an eigenvector for $\bar{z}$.

Example

$A \in \mathbb{R}^{7\times 7}$

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

real

$\overline{4+i} = 4-i$ $\overline{-4-i} = -4+i$ $\overline{i} = -i$

$A \in \mathbb{R}^{7\times 7} \Rightarrow$ 7 eigenvalues with multiplicities
at most 7.

$\Rightarrow$ All eigenvalues are $-2,\ 4 \pm i,\ -4 \pm i,\ \pm i$.

Question $\phi_A(\lambda) = c\cdot(\lambda-(-2))(\lambda-(4+i))(\lambda-(4-i))\cdots\overline{(\lambda+i)(\lambda-i)}$

$= c\cdot(\lambda+2)(\lambda^2-8\lambda+17)(\lambda^2+8\lambda+16)(\lambda^2+1)$

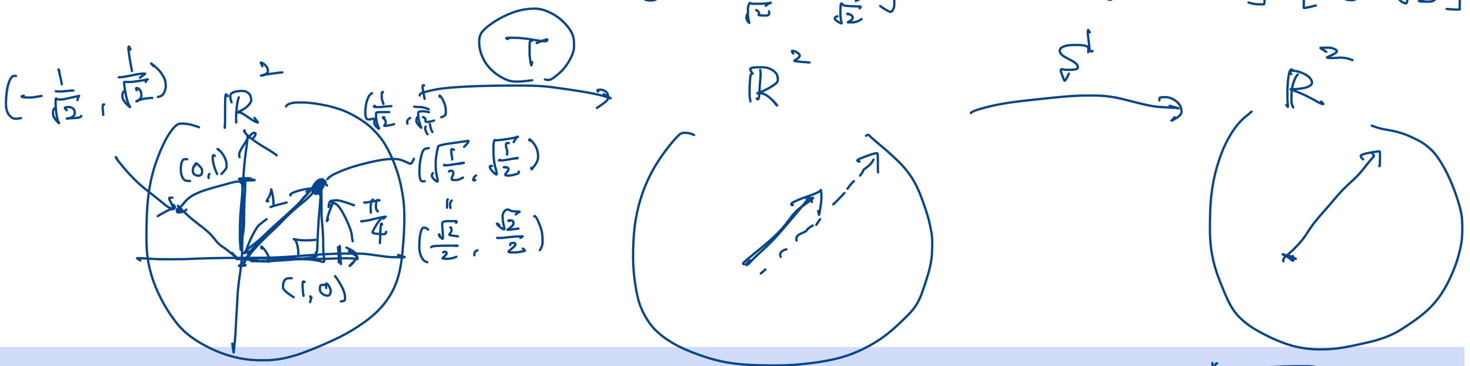
$\uparrow$ from cofactor expansion. $c = (-1)^7 = -1$ (Execise).

Question Is A diagonalizable? Yes.

7 distinct eigenvalues $\Rightarrow$ 7 linearly indep. eigenvector

$\Rightarrow$ A is diagonalizable.

$$T = \begin{bmatrix} Te_1 & Te_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \qquad S = \begin{bmatrix} Se_1 & Se_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$



Example

$$\vec{x} \longmapsto T \cdot \vec{x} \longmapsto \boxed{S \cdot T} \, \vec{x}$$

$$A = $$

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.

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

$$\phi_A(\lambda) = \lambda^2 - \underbrace{(1+1)}_{tr(A)}\lambda + \underset{1\cdot 1-(-1)\cdot 1}{\overset{\det(A)}{2}} = \overbrace{\lambda^2 - 2\lambda + \underset{1+1}{2}} \overset{?}{=} 0$$

tr(A) = Sum of diagonals

$$(\lambda - 1)^{2} = -1$$

$$\lambda - 1 = i \text{ or } -i$$

$$\lambda = 1+i \text{ or } 1-i$$

$$\begin{bmatrix} -i & -1 \\ 0 & 0 \end{bmatrix}$$

$$A - (1+i)I = \begin{bmatrix} 1-(1+i) & -1 \\ 1 & 1-(1+i) \end{bmatrix} = \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix}$$

$-i\,x - y = 0 \qquad y = -i x$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -i x \end{bmatrix} = x \cdot \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

$\begin{bmatrix} 1 \\ -i \end{bmatrix}$ eigenvector corresponding to $\lambda = 1+i$

$\Rightarrow \overline{\begin{bmatrix} 1 \\ -i \end{bmatrix}} = \begin{bmatrix} 1 \\ i \end{bmatrix}$ eigenvector " $\lambda = 1-i$

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

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.

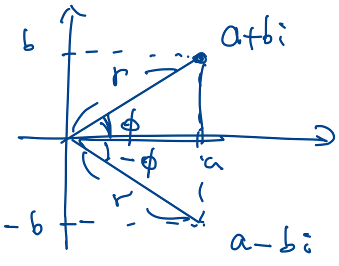
$$\phi_C(\lambda) = \lambda^2 - tr(C)\cdot\lambda + det\, C$$

$$= \lambda^2 - 2a\lambda + (\underset{\sim}{a^2} + \underset{\sim}{b^2}) = 0$$

$$(\underset{\sim}{\lambda - a})^2 = (\lambda^2 - 2a\lambda + a^2) = \boxed{-1}\cdot\underset{\sim}{b}^2$$

$$\lambda - a = b\cdot i \quad \text{or} \quad -b\cdot i$$

$$\lambda = a \pm bi.$$



$$a+bi = r\cdot e^{i\phi}$$

$$a-bi = r\cdot e^{-i\phi}$$

$$\begin{cases} r = |a+bi| = \sqrt{a^2+b^2} \\ \tan\phi = \dfrac{b}{a} \end{cases}$$

Example

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

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

① $\phi_A(\lambda) = \det(A - \lambda I) = \lambda^2 - (1+3)\lambda + (1\cdot 3 - (-2)\cdot 1)$

$= \lambda^2 - 4\lambda + \underset{4+1}{\underline{\underline{5}}} = 0$

$\underbrace{(\lambda-2)}^{2} = \underline{\underline{-1}}$

$\lambda - 2 = i$ or $-i$

$\lambda = 2 \pm i$

② $\lambda = 2+i$: $A - (2+i)I = \begin{bmatrix} 1-(2+i) & -2 \\ 1 & 3-(2+i) \end{bmatrix}$

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

$1\cdot x + \underline{(1-i)}y = 0$

$x = (i-1)\cdot y$

$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (i-1)\cdot y \\ y \end{bmatrix} = y \cdot \begin{bmatrix} i-1 \\ 1 \end{bmatrix}$

③ $\lambda = 2 - i$, $v = \overline{\begin{bmatrix} i - 1 \\ 1 \end{bmatrix}} = \begin{bmatrix} -i - 1 \\ 1 \end{bmatrix}$.

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

$$\vec{u}\cdot\vec{v} = \underset{\substack{\uparrow \\ \text{row vector} \\ 1\times n}}{\vec{u}^T}\,\overset{\substack{\text{column vector} \\ n\times 1 \\ \downarrow}}{\vec{v}} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

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

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

$$u \cdot v = \begin{bmatrix} -1 & 3 & k & 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \\ 1 \\ -3 \end{bmatrix}$$

$$= \boxed{-1\cdot 4} + \cancel{3\cdot 2} + k\cdot 1 + \cancel{2(-3)} = 0$$

$$k = 4$$

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) $\vec{u} \cdot \vec{w} = \vec{w} \cdot \vec{u}$
2. (Linear in each vector) $(\vec{v} + \vec{w}) \cdot \vec{u} = \vec{v} \cdot \vec{u} + \vec{w} \cdot \vec{u}$
3. (Scalars) $(c\vec{u}) \cdot \vec{w} = c \cdot (\vec{u} \cdot \vec{w}) = \vec{u} \cdot (c\vec{w})$
4. (Positivity) $\vec{u} \cdot \vec{u} \geq 0$, and the dot product equals $u_1^2 + \cdots + u_n^2$

1. $\vec{u} \cdot \vec{w} = \vec{u}^T \cdot \vec{w} = (\vec{u}^T \cdot w)^T = w^T \cdot (u^T)^T = w^T \cdot u$

$(A \cdot B)^T = B^T \cdot A^T \qquad (A^T)^T = A \qquad = \vec{w} \cdot \vec{u}$

4. $\vec{u} = (u_1, u_2, \cdots, u_n) \qquad \vec{u} \cdot \vec{u} = u_1^2 + u_2^2 + \cdots + u_n^2$

if $u_1, \cdots, u_n \in \mathbb{R}$

$\vec{u} \cdot \vec{u} \geq 0$

and $\vec{u} \cdot \vec{u} = 0 \Rightarrow u_1 = 0, u_2 = 0, \cdots$

$\Rightarrow \vec{u} = 0$

The Length of a Vector

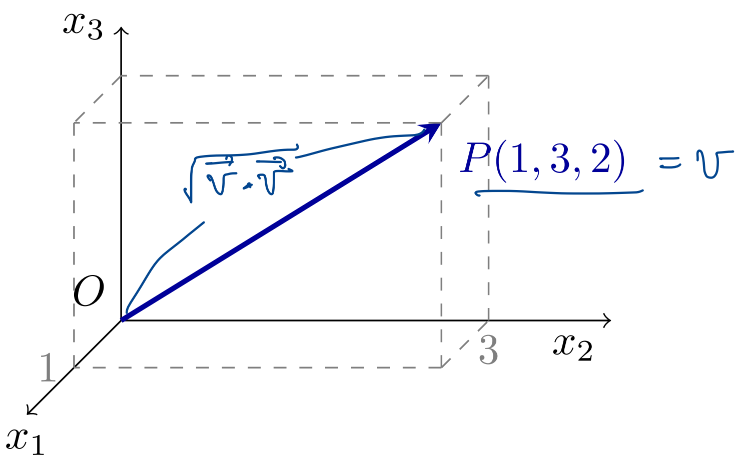
Definition

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 $\underline{\vec{u} \cdot \vec{v} = -1}$. Compute the value of $\|\vec{u} + \vec{v}\|$.

$$\|\vec{u}+\vec{v}\|^2 = (\vec{u}+\vec{v}) \cdot (\vec{u}+\vec{v}) = \underline{u \cdot u} + \underline{u \cdot v} + \underline{v \cdot u} + \underline{v \cdot v}$$

$$= \underline{\|u\|^2} + 2\,\underline{u \cdot v} + \underline{\|v\|^2}$$

$$= 5^2 + 2 \cdot (-1) + (\sqrt{3})^2$$

$$= 25 - 2 + 3 = \underline{\underline{26}}.$$

$$\|u+v\| = \sqrt{26}.$$

$$v \cdot \underline{v} \overset{def}{=} v^T \cdot v = [v_1 \ \cdots \ v_n] \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = v_1^2 + v_2^2 + \cdots + v_n^2$$

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.

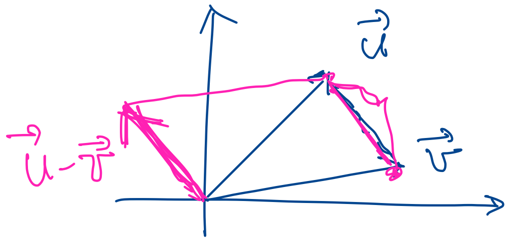
$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \vec{y} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \vec{v} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

Ex

$$\underline{\vec{v}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\|v\| = \sqrt{\vec{v} \cdot \vec{v}} = \sqrt{1^2 + 2^2} = \boxed{\sqrt{5}}.$$

$$\left\| \frac{1}{\sqrt{5}} \cdot \vec{v} \right\| = \sqrt{\left(\frac{1}{\sqrt{5}} \cdot \vec{v}\right)\left(\frac{1}{\sqrt{5}} \vec{v}\right)} = \frac{1}{\sqrt{5}} \cdot \underbrace{\sqrt{\vec{v} \cdot \vec{v}}} = 1.$$



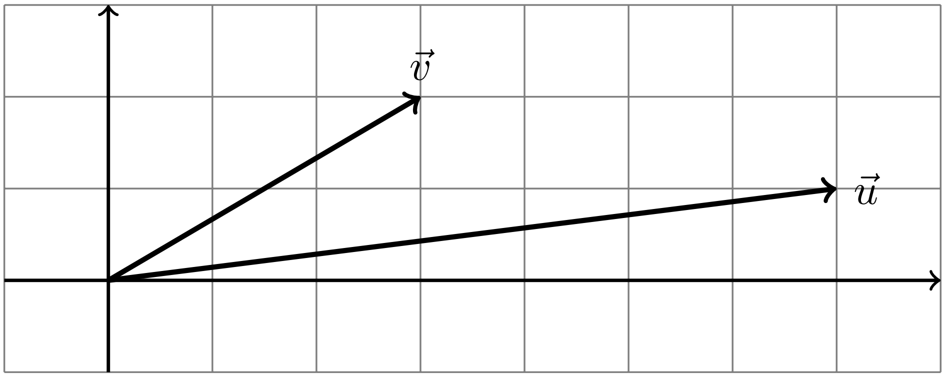
Distance in $\mathbb{R}^n$

Definition

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

$$\boxed{\|\vec{u}-\vec{v}\| = \sqrt{(\vec{u}-\vec{v})\cdot(\vec{u}-\vec{v})}}$$

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



$$\vec{u}-\vec{v} = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$\|\vec{u}-\vec{v}\| = \sqrt{4^2+(-1)^2} = \sqrt{17}.$$

The Cauchy-Schwarz Inequality

Theorem: Cauchy-Bunyakovsky–Schwarz Inequality

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

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

$|\vec{u} \cdot \vec{v}| = \|\vec{u}\| \cdot \|\vec{v}\|$

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

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

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

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

$n=2.$ $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

$|u \cdot v|^2 = (u_1 \cdot v_1 + u_2 \cdot v_2)^2$

$\|u\|^2 = (u_1^2 + u_2^2)$ $\|v\|^2 = (v_1^2 + v_2^2)$

$\boxed{\|u\|^2 \cdot \|v\|^2 - |u \cdot v|^2} = \underline{(u_1^2 + u_2^2) \cdot (v_1^2 + v_2^2)} - (u_1 v_1 + u_2 v_2)^2$

$$= \left(\cancel{u_1^2 v_1^2} + \underline{\underline{u_1^2 v_2^2}} + \underline{u_2^2 v_1^2} + \cancel{u_2^2 v_2^2}\right)$$

$$- \left(\cancel{u_1^2 v_1^2} + \underline{2} \cdot \underline{u_1} u_2 v_1 \underline{v_2} + \cancel{u_2^2 v_2^2}\right)$$

$$= (u_1 \cdot v_2)^2 - 2 \cdot (u_1 v_2) \cdot (u_2 v_1) + (u_2 v_1)^2$$

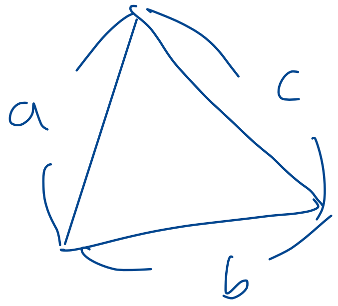
$$= (u_1 v_2 - u_2 v_1)^2 \boxed{\geq 0}$$

$$\|u\|^2 \|v\|^2 \geq |u \cdot v|^2$$

$$\|u\| \, \|v\| \geq |u \cdot v| .$$

Equality holds $\iff$ $u_1 v_2 = u_2 v_1$

$\frac{v_2}{v_1} = \frac{u_2}{u_1}$ i.e. $\vec{u}, \vec{v}$ are parallel.

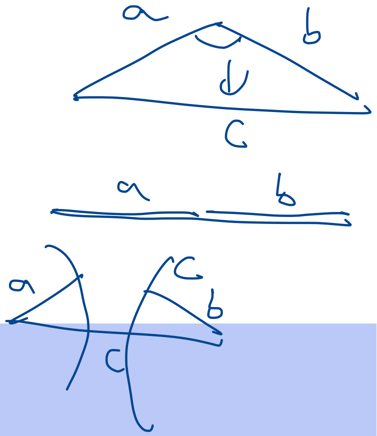


$c > a , b$

$\Rightarrow \quad a+b \boxed{>} c$

If $\underline{a+b} = c \quad \Rightarrow$

$a+b < c$



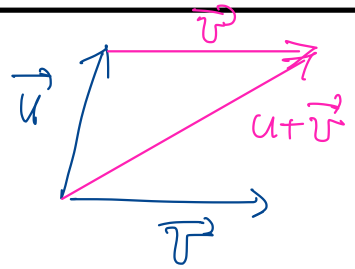
The Triangle Inequality

Theorem: Triangle Inequality

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

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

Proof:



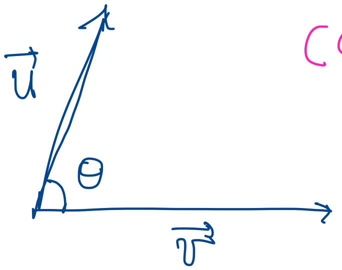
$$\begin{aligned} \|\vec{u}+\vec{v}\|^2 &= (\vec{u}+\vec{v})\cdot(\vec{u}+\vec{v}) \\ &= \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\vec{u}\cdot\vec{v} \\ &\leq \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\|\vec{u}\|\|\vec{v}\| \\ &= \underline{(\|\vec{u}\| + \|\vec{v}\|)^2} \end{aligned}$$

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

(C-S)

$a^2 + b^2 + 2\cdot a\cdot b = (a+b)^2$

$-\|u\|\cdot\|v\| \le u\cdot v \le \|u\|\cdot\|v\|$



(C-S) (C-S)

$$-1 \le \cos\theta = \frac{\vec{u}\cdot\vec{v}}{\|u\|\cdot\|v\|} \le 1$$

Angles

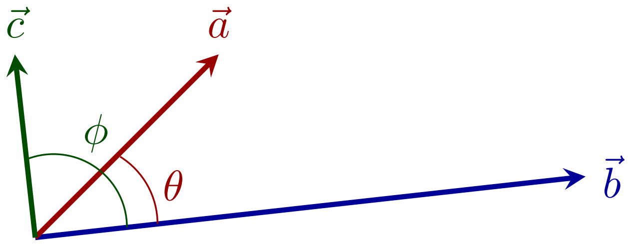
$\cos\theta = 0$ or $\|a\|\cdot\|b\| = 0$

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 zero vectors, or
- $\vec{a}$ and $\vec{b}$ are perpendicular.

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:

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

$\uparrow$

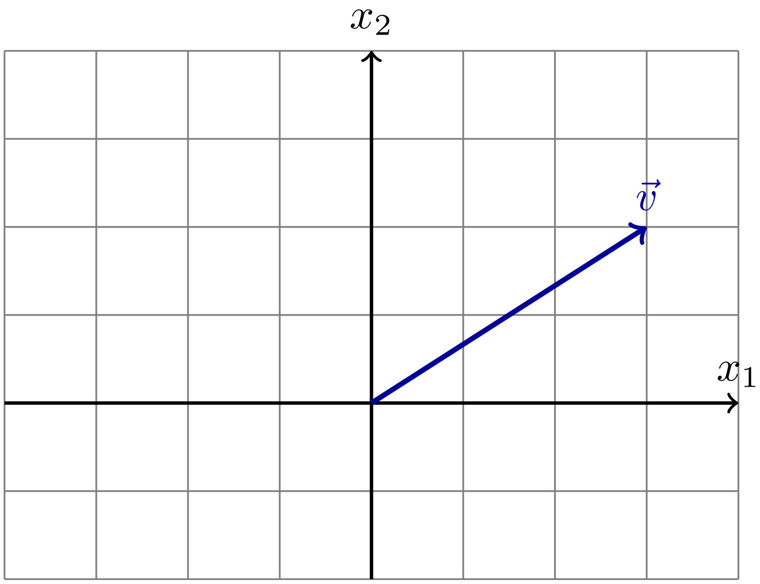
Pythagorean.

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.

$$\|\vec{u} + \vec{w}\|^2 = \|\vec{u}\|^2 + \|\vec{w}\|^2 + \underbrace{2 \cdot \vec{u} \cdot \vec{w}}_{\overset{\parallel}{0}}$$

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 compliment** of W, or $W^\perp$ or 'W perp.'

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

subspace.

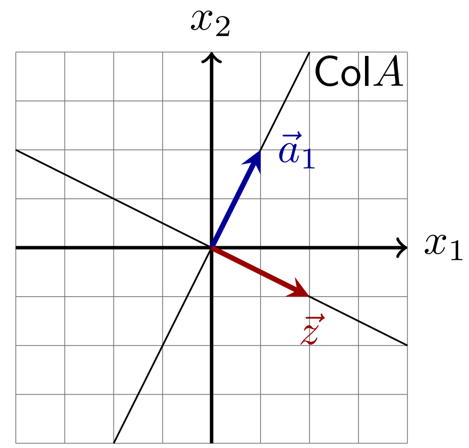
Recall $\vec{u} \cdot \vec{v} = [u_1 \ \cdots \ u_n] \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \cdots + u_n v_n$

- $\vec{u}$ is orthogonal to $\vec{v}$ if $\vec{u} \cdot \vec{v} = \underset{\sim}{0}$
- $\vec{u}$ is orthogonal to a subspace W if $\vec{u} \perp \vec{w}$ for all $\vec{w}$ in W.

Example

lin. dep.

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



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

$$\parallel$$

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} : \begin{bmatrix} x \\ y \end{bmatrix} \cdot \vec{w} \overset{=0}{\underset{\text{for all}}{\frown}} \vec{w} \in \text{Col}(A) \right\}$$

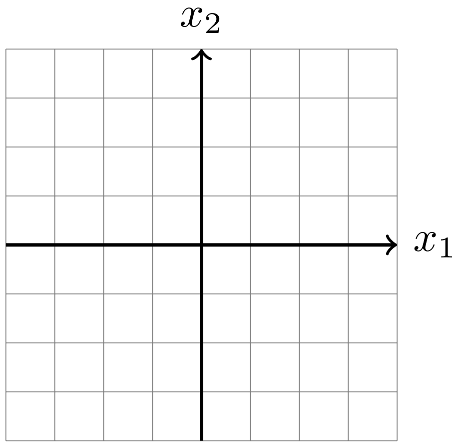
$$= \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : c\overbrace{\begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix}} = 0 \ \text{for all } c \in \mathbb{R} \right\}$$

$$= \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : \underline{x + 2y = 0} \right\}$$

$$= \text{Null}\left(\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \right)$$

$$= \text{Null}\,(A^T)$$

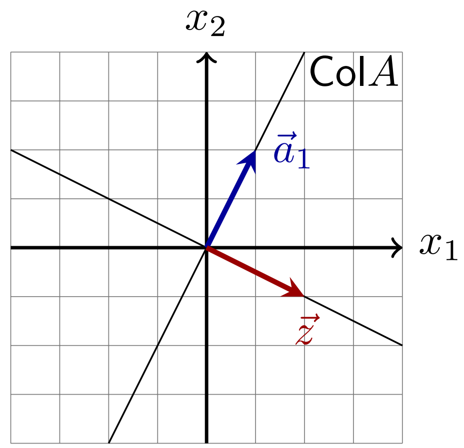
Sketch Null$A$ and Null$A^\perp$ on the grid below.



Example

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

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



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

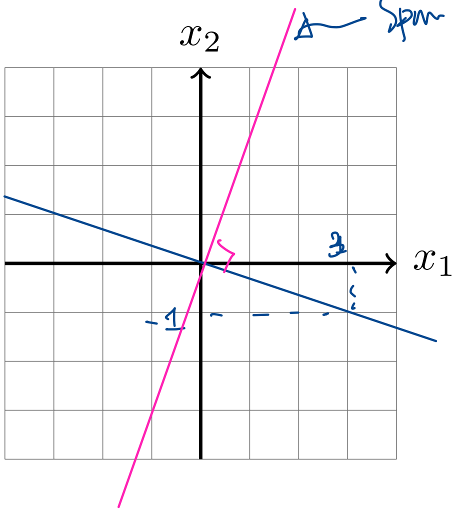
Null (A)

$= \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : \underline{A \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 0} \right\}$

$= \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \right\}$

$= \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : \underline{\boxed{x} + 3y = 0} \right\}$

$\underline{y = -\frac{1}{3} x}$



$= \left\{ c \begin{bmatrix} 3 \\ -1 \end{bmatrix} : c \in \mathbb{R} \right\} \leftarrow \left(\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -3y \\ y \end{bmatrix} = -y \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right)$

$= \text{Span} \left(\begin{bmatrix} 3 \\ -1 \end{bmatrix} \right)$

$\text{Null}(A)^{\perp}$

$= \text{Span} \left\{ \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right\}^{\perp}$

$= \text{Null} \left(\begin{bmatrix} 3 & -1 \end{bmatrix} \right)$

$= \underline{\text{Span} \left(\begin{bmatrix} 1 \\ 3 \end{bmatrix} \right)}$

$= \text{Span} \left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right)$

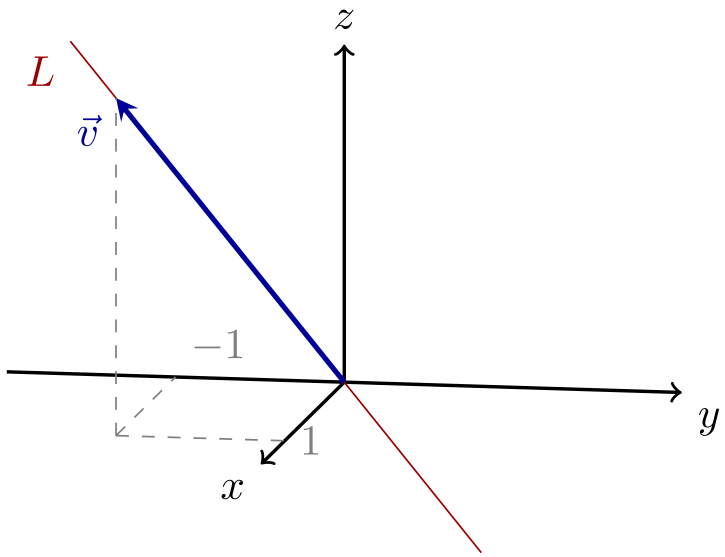
$= \text{Col} \left(\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \right)$

$= \text{Col} \left(A^T \right)$

$\text{Col}(A)^{\perp} = \text{Null}(A^T)$, $\text{Null}(A)^{\perp} = \underbrace{\text{Col}(A^T)}_{\uparrow \text{ spanned by rows of } A} = \text{Row}(A)$

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

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

$$\text{Col}(A)^{\perp} = \text{Null}(A^T)$$

$$\text{Null}(A)^{\perp} = \text{Col}(A^T) = \text{Row}(A)$$

$$A \in \mathbb{R}^{m\times n} \Rightarrow \dim(\text{Null}(A)) + \overbrace{\dim(\text{Col}(A))}^{= \text{Rank}(A)} = n$$

$$A^T \in \mathbb{R}^{n\times m} \Rightarrow \dim(\text{Null}(A^T)) + \dim(\underset{\text{Row}(A)}{\underline{\text{Col}(A^T)}}) = m$$

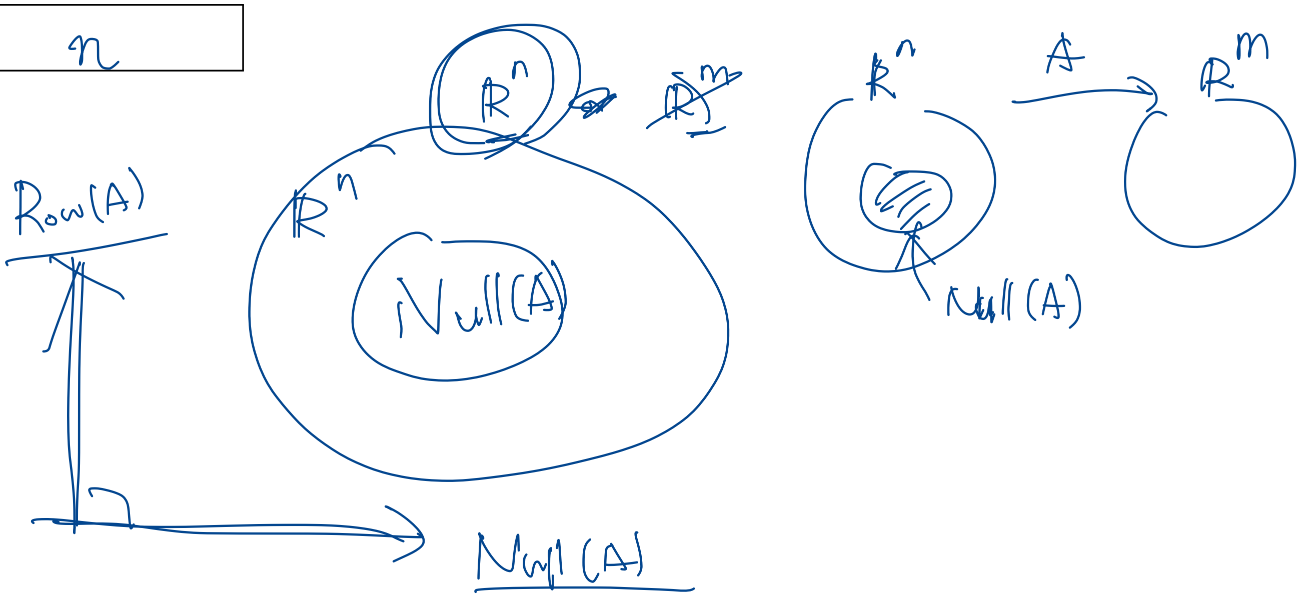
Example 3

$A \in \mathbb{R}^{m \times n}$

Describe the Null(A) in terms of an orthogonal subspace.

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

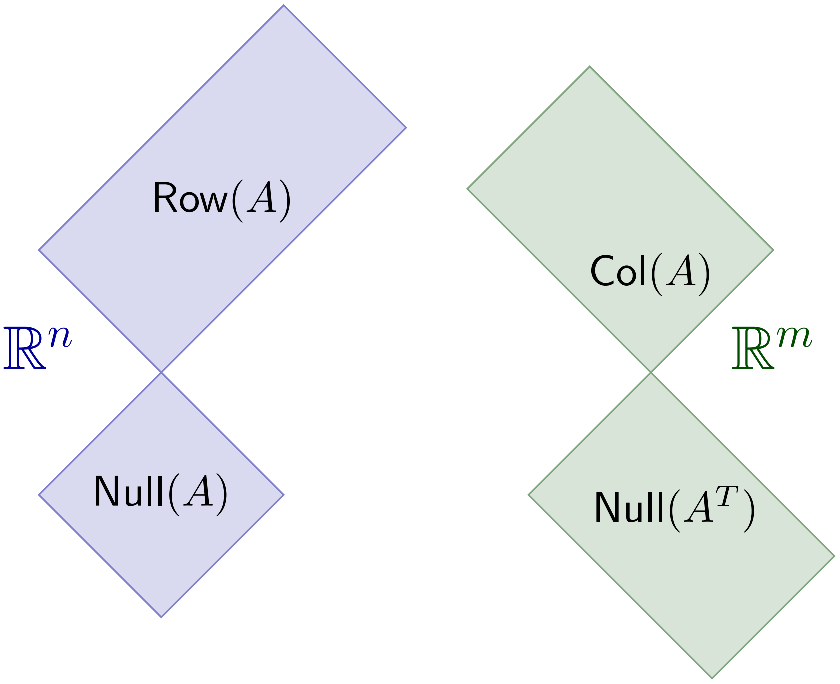
1. $A\vec{x} = \vec{0} \iff \begin{bmatrix} - \vec{a}_1 - \\ - \vec{a}_2 - \\ \vdots \\ - \vec{a}_m - \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow$
$\vec{a}_1 \cdot \vec{x} = 0$
$\vec{a}_2 \cdot \vec{x} = 0$
$\vdots$
$\vec{a}_m \cdot \vec{x} = 0$
2. This means that $\vec{x}$ is $\boxed{\text{orthogonal}}$ to each row of A.
3. Row A is $\boxed{\text{orthogonal}}$ to Null A.
$\text{Row}(A) = (\text{Null}(A))^{\perp}$
4. The dimension of Row A plus the dimension of Null A equals $\boxed{n}$



Theorem (The Four Subspaces)

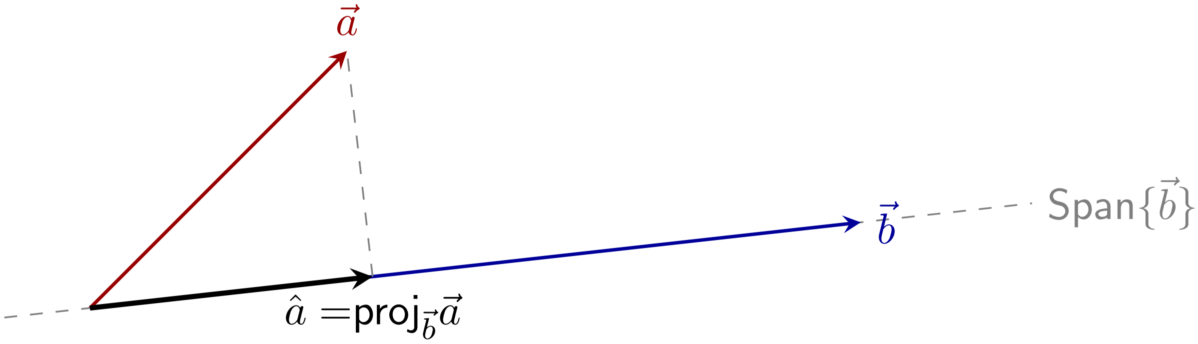
For any $A \in \mathbb{R}^{m\times n}$, the orthogonal complement of $\overset{\overset{\text{Col}(A^T)}{\shortparallel}}{\text{Row}\,A}$ is $\text{Null}\,A$, and the orthogonal complement of $\text{Col}\,A$ is $\text{Null}\,A^T$.

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



Looking Ahead - Projections

Suppose we want to find the closed vector in 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 $\{\vec{u}_1, \ldots, \vec{u}_p\}$ 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.

$$\vec{u}_1 = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} -2 \\ 0 \\ 8 \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$\vec{u}_1 \cdot \vec{u}_2 = 0$

$\vec{u}_2 \cdot \vec{u}_3 = 0$

$$\parallel$$

$4 \cdot (-2) + 0 \cdot 0 + 1 \cdot \square = 0$

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 $c_1, \dots, c_p$,

$$\left\|c_1\vec{u}_1 + \cdots + c_p\vec{u}_p\right\|^2 \overset{\text{Pythagorean}}{=} c_1^2\|\vec{u}_1\|^2 + \cdots + c_p^2\|\vec{u}_p\|^2.$$

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 set $\underset{\not\Leftarrow}{\boxed{\Rightarrow}}$ Linearly Indep.

Proof

$$c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_p\vec{u}_p = \vec{0}$$

Need : $\boxed{c_1 = c_2 = \cdots = c_p = 0}$

$$\boxed{\vec{u}_1} \cdot \left(c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_p\vec{u}_p\right) = 0$$

$$\underbrace{\boxed{c_1 \cdot \vec{u}_1 \cdot \vec{u}_1}}_{\|u_1\|^2} + c_2 \underbrace{\vec{u}_1 \cdot \vec{u}_2}_{=0} + \cdots + c_p \underbrace{\vec{u}_1 \cdot \vec{u}_p}_{=0} \boxed{=0}$$

$$c_1 \cdot \underbrace{\|\vec{u}_1\|^2}_{\neq 0} = 0 \Rightarrow c_1 = 0$$

Orthogonal Bases

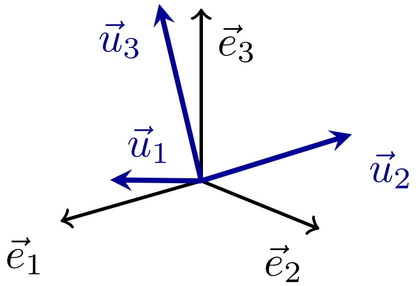
Theorem (Expansion in Orthogonal Basis)

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

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

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

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 $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$.



To find c_q,

$$\vec{u}_q \cdot \vec{w} = \vec{u}_q \cdot \left(c_1 \vec{u}_1 + \cdots + c_q \vec{u}_q + \cdots + c_p \vec{u}_p \right)$$

$$= c_q \cdot \underbrace{\vec{u}_q \cdot \vec{u}_q}_{} \quad \Rightarrow \quad c_q = \frac{\vec{u}_q \cdot \vec{w}}{\vec{u}_q \cdot \vec{u}_q}.$$

orthogonal ⇒ lin. Indep.

Example

$\in W$ $\in W$ $\in W$

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

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

a) Check that an orthogonal basis for W is given by $\vec{u}$ and $\vec{v}$. ✓

b) Compute the expansion of $\vec{s}$ in basis W.

$$\overset{\in W}{\begin{bmatrix}x\\y\\z\end{bmatrix}} \cdot \begin{bmatrix}1\\1\\1\end{bmatrix} = x+y+z = 0 \leftarrow \dim = 2$$

orthogonal

$$\vec{s} = a \cdot \vec{u} + b \cdot \vec{v}$$

Find a, b.

$$a = \frac{\vec{u} \cdot \vec{s}}{\vec{u} \cdot \vec{u}}, \quad b = \frac{\vec{v} \cdot \vec{s}}{\vec{v} \cdot \vec{v}}$$

Recall $\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_p\}$: an orthogonal set ($\vec{u}_i \cdot \vec{u}_j = 0 \quad \forall i \neq j$)

Suppose $\vec{\omega} = c_1 \vec{u}_1 + c_2 \vec{u}_2 + \cdots + c_p \cdot \vec{u}_p$

(i) $c_q = \dfrac{\vec{\omega} \cdot \vec{u}_q}{\vec{u}_q \cdot \vec{u}_q}$

(ii) $\|\omega\|^2 = \|c_1 \vec{u}_1\|^2 + \|c_2 \vec{u}_2\|^2 + \cdots + \|c_p \vec{u}_p\|^2$

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

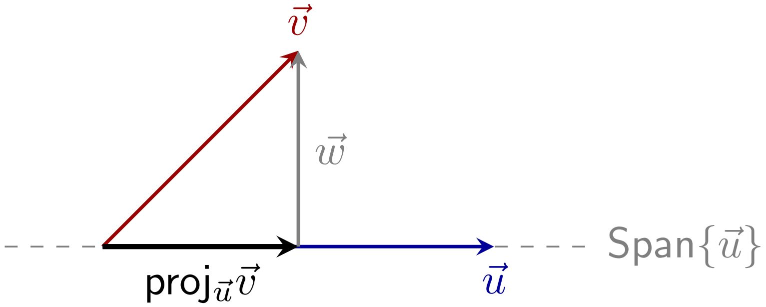
Projections

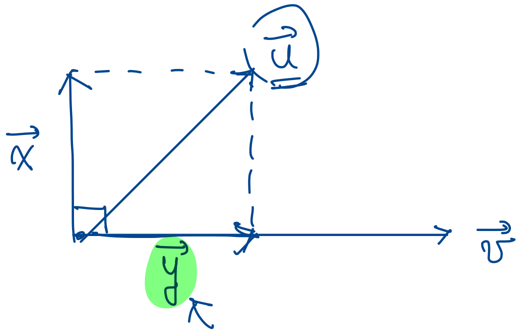
Let $\vec{u}$ be a non-zero vector, and let $\vec{v}$ be some other vector. The **orthogonal projection of $\vec{v}$ onto the direction of $\vec{u}$** is the vector in the span of $\vec{u}$ that is closest to $\vec{v}$.

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

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

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





Projection of $\vec{u}$ onto $\vec{v}$

① $\vec{u} = \vec{y} + \vec{x}$

② $\vec{y} = c\,\vec{v}$ $\Big\}$ ⇒ Find $\vec{y}$

③ $\vec{x}\cdot\vec{v} = 0$

$$\vec{v}\cdot\vec{u} = (c\cdot\vec{v} + \vec{x})\cdot\vec{v}$$

$$\vec{v}\cdot\vec{u} = c\cdot\vec{v}\cdot\vec{v}$$

$$\therefore\ c = \frac{\vec{v}\cdot\vec{u}}{\vec{v}\cdot\vec{v}}$$

Projections

$$\Rightarrow \boxed{\vec{y} = \frac{\vec{v}\cdot\vec{u}}{\vec{v}\cdot\vec{v}}\cdot\vec{v}}$$

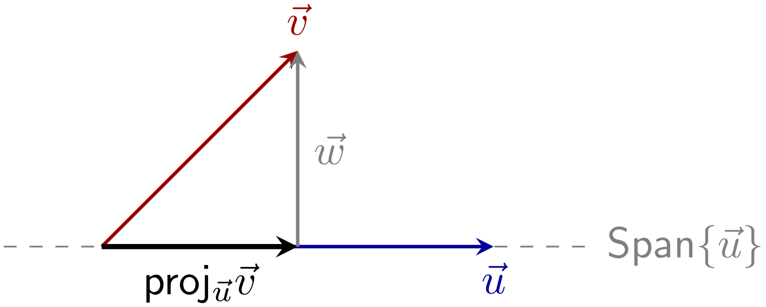
Let $\vec{u}$ be a non-zero vector, and let $\vec{v}$ be some other vector. The **orthogonal projection of $\vec{v}$ onto the direction of $\vec{u}$** is the vector in the span of $\vec{u}$ that is closest to $\vec{v}$.

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

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

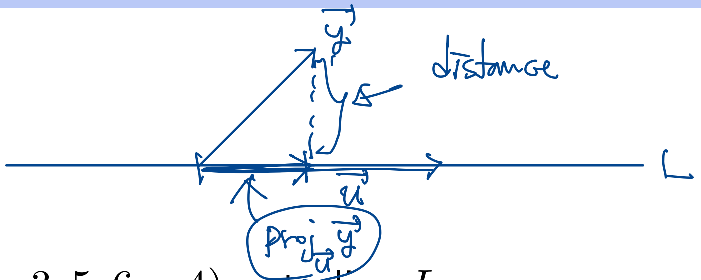
$$\vec{v} = \text{proj}_{\vec{u}}\vec{v} + \vec{w}$$

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



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?

1. $\text{proj}_{\vec{u}}\vec{y} = \frac{\vec{y}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\cdot\vec{u} = \frac{4}{4}\cdot\vec{u} = \vec{u}$.

$\vec{y} - \vec{u} = \begin{bmatrix} -4 \\ 4 \\ 5 \\ -5 \end{bmatrix}$

2. distance between $\vec{y}$ and L

$= \| \vec{y} - \text{proj}_{\vec{u}}\vec{y} \| = \| \vec{y} - \vec{u} \|$

$= \sqrt{(-4)^2 + 4^2 + 5^2 + (-5)^2} = \sqrt{82}$

$\{\vec{u}_1, \ldots, \vec{u}_p\}$:

- orthogonal : $\vec{u}_i \cdot \vec{u}_j = 0 \quad \forall i \neq j$
- basis of W : $\begin{cases} \text{lin. indep.} \\ \text{spans } W \end{cases}$
- orthonormal $\|u_i\| = 1$

Definition

Definition (Orthonormal Basis)

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

$$\vec{w} = (\vec{w} \cdot \vec{u}_1)\vec{u}_1 + \cdots + (\vec{w} \cdot \vec{u}_p)\vec{u}_p$$

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

For every $\vec{w} \in W$,

$$\vec{w} = c_1 \vec{u}_1 + c_2 \vec{u}_2 + \cdots + c_p \vec{u}_p$$

Orthogonal ⇒ $c_q = \dfrac{\vec{u}_q \cdot \vec{w}}{\boxed{u_q \cdot u_q}} = \vec{u}_q \cdot \vec{w}$ ↑ orthonormal

$$u_q \cdot u_q = \|u_q\|^2 = 1$$

$$\|\vec{w}\|^2 = \|c_1 \vec{u}_1\|^2 + \cdots + \|c_p u_p\|^2$$

$$= c_1^2 + \cdots + c_p^2$$

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

Example

If $\begin{bmatrix}x\\y\\z\end{bmatrix} \in W$, $\begin{bmatrix}x\\y\\z\end{bmatrix}\cdot\begin{bmatrix}1\\1\\1\end{bmatrix} = \underline{x+y+z=0}$.

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.

$$u = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix} \qquad v \overset{\in W}{=} \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\1\end{bmatrix}$$

① $W = \text{Null}([1 \ 1 \ 1])$ $\quad \dim(W) = 2$.

② $u \cdot v = 0$ $\quad \begin{cases} x+y+z=0 \\ \underline{x-z=0} \end{cases}$ $\quad x = 1 \quad z = 1 \quad y = -2$

③ $\vec{x}^{\neq 0} \xrightarrow[\text{normalize}]{} \frac{1}{\|\vec{x}\|}\cdot x$: unit vector.

Orthogonal Matrices

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

orthogonal w/ length 1.

Theorem

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

Can U have $\overset{\text{orthogonal}}{\text{orthonormal}}$ columns if $\boxed{n > m}$?

① $\vec{U} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$

$\vec{u}_i \in \boxed{\mathbb{R}^m}$

$\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_n\}$ orthonormal $\Rightarrow$ linearly indep. $\Rightarrow$ $\boxed{n \leq m}$

② $A \in \mathbb{R}^{m \times n}$ $\quad \vec{x} \in \mathbb{R}^n$ $\quad \vec{y} \in \mathbb{R}^m$

$A \cdot \vec{x} \in \mathbb{R}^m$

$$\overset{\in \mathbb{R}^m}{\vec{y}} \cdot \underbrace{(A \cdot \vec{x})}_{\in \mathbb{R}^m} = \overset{\in \mathbb{R}^n}{(A^T \cdot \vec{y})} \cdot \overset{\in \mathbb{R}^n}{\vec{x}}$$

Theorem

$U^T \cdot U = I$

Theorem (Mapping Properties of Orthogonal Matrices)

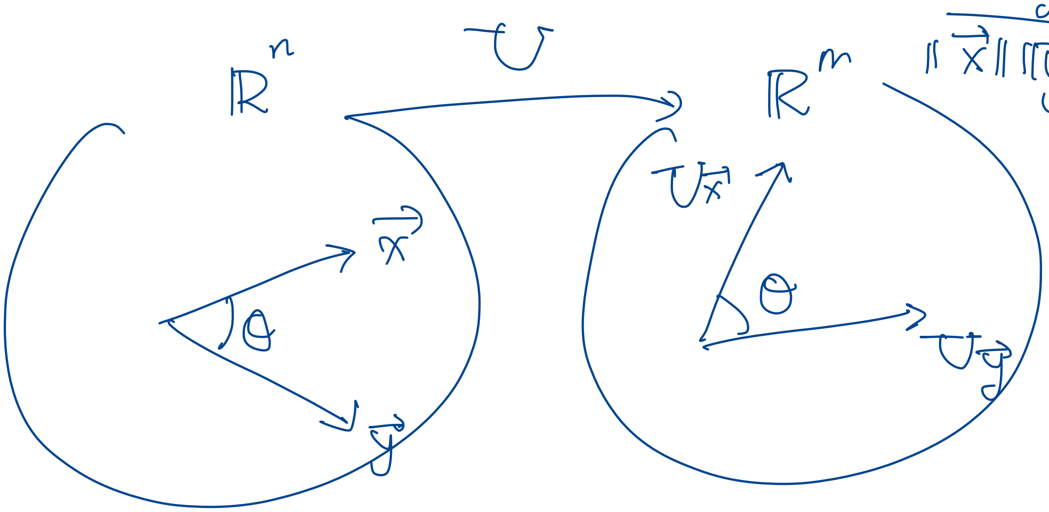
Assume $m \times m$ matrix U has orthonormal columns. Then

1. (Preserves length) $\|U\vec{x}\| = \boxed{\|\vec{x}\|}$
2. (Preserves angles) $(U\vec{x}) \cdot (U\vec{y}) = \boxed{\vec{x} \cdot \vec{y}}$
3. (Preserves orthogonality) $\vec{x} \cdot \vec{y} = 0 \iff Ux \cdot Uy = 0$

$$(U \cdot \vec{x}) \cdot (\boxed{U}\vec{y}) = (\underbrace{U^T \cdot U}_{\overset{\shortparallel}{I}} \cdot \vec{x}) \cdot \vec{y} = \vec{x} \cdot \vec{y}$$

$$\|U\vec{x}\|^2 = (U \cdot \vec{x}) \cdot (U \cdot \vec{x}) = \vec{x} \cdot \vec{x} = \|x\|^2$$

$$\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|} = \cos\theta$$



Example

Compute the length of the vector below.

$$\left\| \underbrace{\begin{bmatrix} 1/2 & 2/\sqrt{14} \\ 1/2 & 1/\sqrt{14} \\ 1/2 & -3/\sqrt{14} \\ 1/2 & 0 \end{bmatrix}}_{\substack{\shortparallel \\ U}} \begin{bmatrix} \sqrt{2} \\ -3 \end{bmatrix} \right\| = \left\| \begin{bmatrix} \sqrt{2} \\ -3 \end{bmatrix} \right\| = \sqrt{11}$$

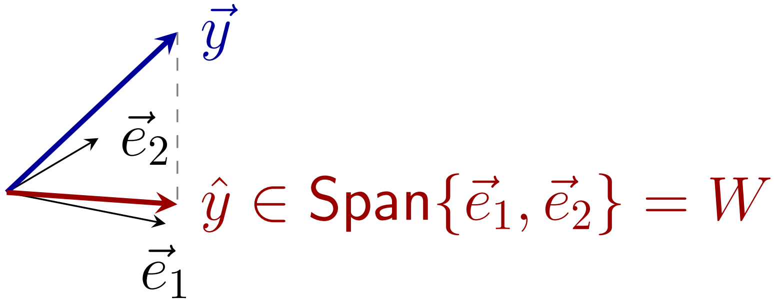
U has orthonormal columns

$$\underset{\substack{\downarrow \\ \text{length } 1}}{\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}} \cdot \underset{\substack{\downarrow \\ \text{length } 1}}{\begin{bmatrix} \frac{2}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ -\frac{3}{\sqrt{14}} \\ 0 \end{bmatrix}} = 0$$

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 =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 $\widehat{b}$ in column space of A, is closest to $\vec{b}$?

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

Example 1

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

$$\vec{y} = \underbrace{\overset{\in W}{\underset{\underset{(\vec{y}\cdot\vec{u}_1)}{\|}}{c_1} \cdot \vec{u}_1 + \underset{\underset{(\vec{y}\cdot\vec{u}_2)}{\|}}{c_2} \cdot \vec{u}_2}}_{\underset{\widehat{y}}{\|}} + \underbrace{\overset{\in W^{\perp}}{\cdots + \underset{\underset{(\vec{y}\cdot\vec{u}_5)}{\|}}{c_5}\, \vec{u}_5}}_{\underset{w^{\perp}}{\|}}$$

$\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$ orthogonal basis for W

For $\vec{y} \in W$, $\vec{y} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_p\vec{u}_p$

$c_q = \dfrac{\vec{y}\cdot\vec{u}_q}{\vec{u}_q\cdot\vec{u}_q}$ for $q = 1, \ldots, p$

If $\mathcal{B}$ is orthonormal, $\|\vec{u}_q\| = 1$, $\vec{u}_q\cdot\vec{u}_q = 1$

Orthogonal Decomposition Theorem

$c_q = \vec{y}\cdot\vec{u}_q$

$W^\perp = \{ z \in \mathbb{R}^n : \underline{z \perp W} \}$

$z\cdot w = 0$ $\forall w \in W$

Theorem

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

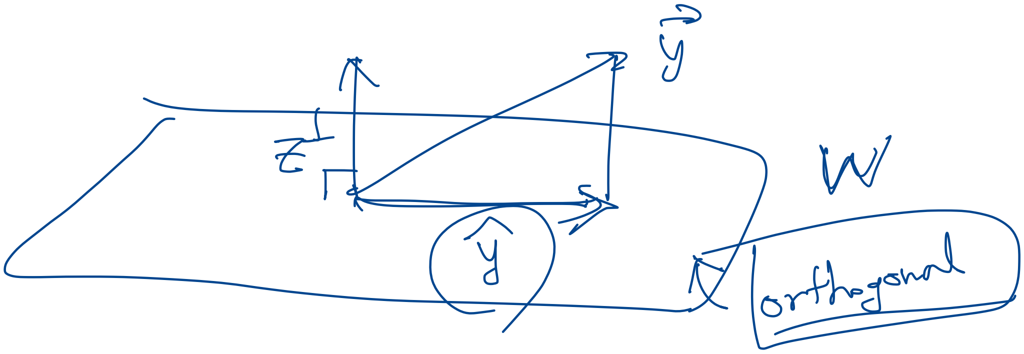
$$\vec{y} = \widehat{y} + w^\perp, \quad \widehat{y} \in W, \quad w^\perp \in W^\perp.$$

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

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

We say that $\widehat{y}$ 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

$$\widehat{y} =$$

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

Example 2a

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

$\vec{u}_1 \cdot \vec{u}_2 = 0$

$\{\vec{u}_1, \vec{u}_2\}$ is an orthogonal basis

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

$$\hat{y} = \frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\cdot\vec{u}_1 + \frac{\vec{y}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\cdot\vec{u}_2$$

$$= \frac{8}{8}\vec{u}_1 + \frac{3}{1}\vec{u}_2 = \vec{u}_1 + 3\vec{u}_2$$

$$= \begin{bmatrix}2\\2\\3\end{bmatrix} = \text{proj}_W(\vec{y})$$

$\vec{y}\cdot\vec{u}_1 = 8$

$\vec{u}_1\cdot\vec{u}_1 = 2^2 + 2^2 = 8$

$\vec{y}\cdot\vec{u}_2 = 3$

$\vec{u}_2\cdot\vec{u}_2 = 1.$

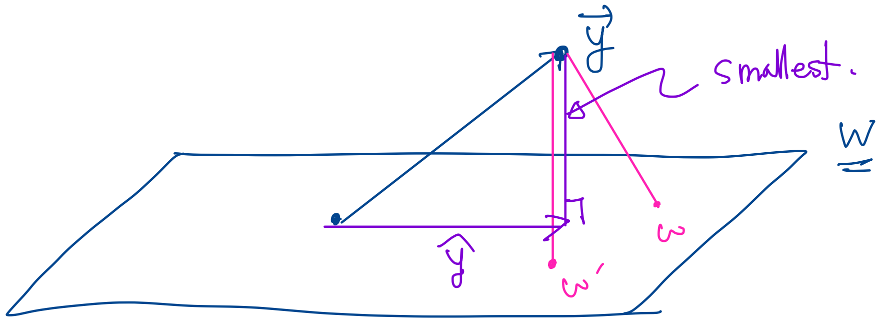
$$w^{\perp} = \vec{y} - \hat{y} = \begin{bmatrix}4\\0\\3\end{bmatrix} - \begin{bmatrix}2\\2\\3\end{bmatrix} = \begin{bmatrix}2\\-2\\0\end{bmatrix}$$

Check

$$w^{\perp} = \begin{bmatrix}2\\-2\\0\end{bmatrix} \perp W \ ?$$

$w^{\perp}\cdot\vec{u}_1 = 0$

$w^{\perp}\cdot\vec{u}_2 = 0$



Q: WANT TO MEASURE Distance between $\vec{y}$ and W.

Best Approximation Theorem

$= $ minimum of $\text{dist}(\vec{y}, \vec{w})$ among $\vec{w} \in W$

Theorem

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

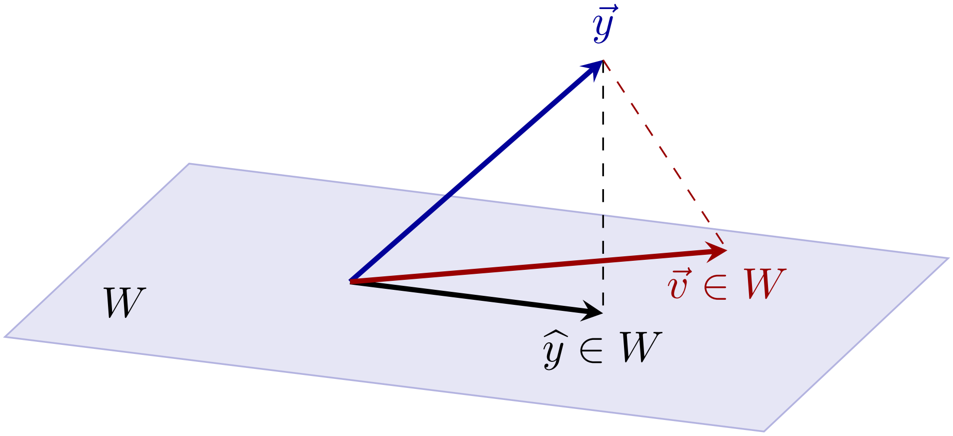
$$\text{dist}(\vec{y}, W) = \|\vec{y} - \widehat{y}\| < \|\vec{y} - \vec{w}\|$$

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

(W is convex)

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 $W = \mathrm{Span}\{\vec{u}_1, \vec{u}_2\}$? Note that these vectors are the same vectors that we used in Example 2a.

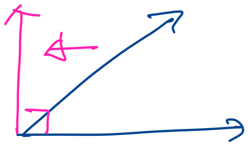
$$\hat{y} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ -2 \\ 0 \end{bmatrix}$$

$$\vec{y} - \hat{y} = \boxed{w^{\perp}}$$

$$\mathrm{dist}(\vec{y}, W) = \| \vec{y} - \hat{y} \|$$

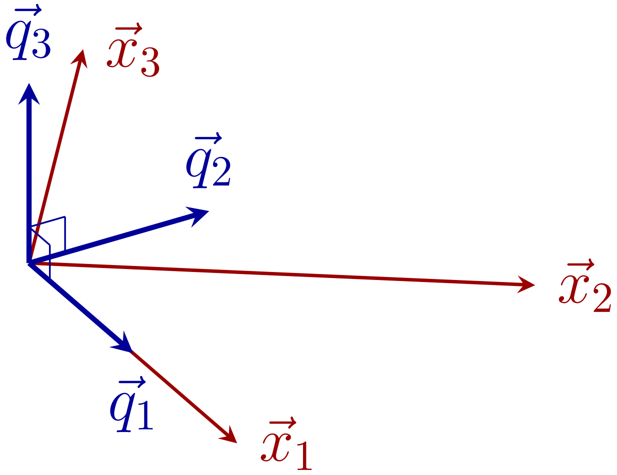
$$= \sqrt{2^2 + (-2)^2 + 0^2} = \sqrt{8}.$$



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.

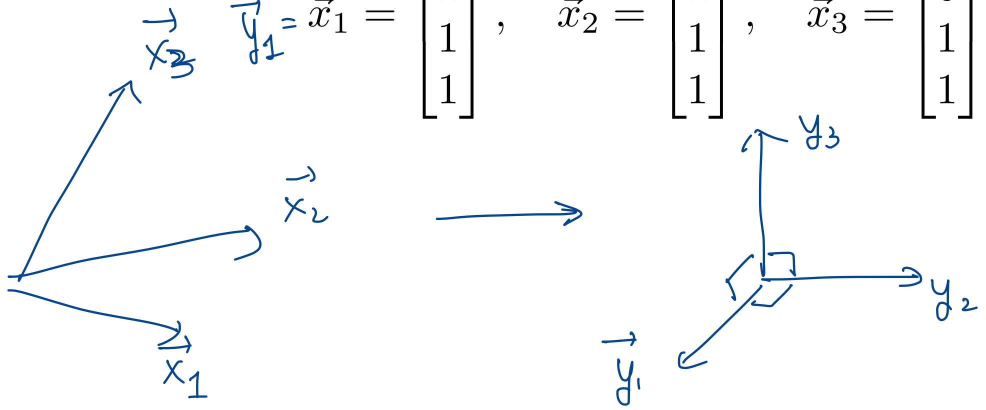
Motivating Question The vectors below span a subspace W of $\mathbb{R}^4$. Identify an orthogonal basis for W.

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

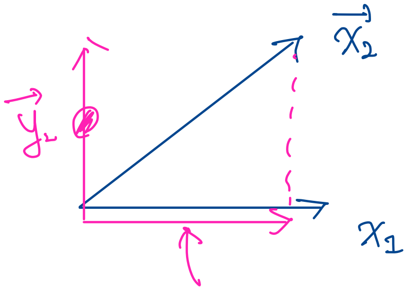
Example

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

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



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

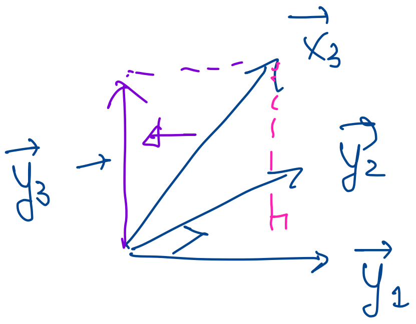


$x_1 = y_1$

$\vec{x}_2 = \vec{y}_2 + \text{proj}_{\vec{y}_1}(\vec{x}_2)$

$$\vec{y}_2 = \vec{x}_2 - \text{Proj}_{y_1}(\vec{x}_2) = \vec{x}_2 - \frac{\overset{3}{\vec{x}_2 \cdot \vec{y}_1}}{\underset{= 4}{\vec{y}_1 \cdot y_1}}\vec{y}_1$$

$$= \begin{bmatrix}0\\1\\1\\1\end{bmatrix} - \begin{bmatrix}3/4\\3/4\\3/4\\3/4\end{bmatrix} = \begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix}$$



$W = \text{Span}\{y_1, y_2\}$

$\vec{y}_3 = \vec{x}_3 - \text{proj}_W(x_3)$

$$\text{proj}_{W}(\vec{x}_3) = \frac{\vec{x}_3\cdot\vec{y}_1}{\vec{y}_1\cdot\vec{y}_1}\vec{y}_1 + \frac{\vec{x}_3\cdot\vec{y}_2}{\vec{y}_2\cdot\vec{y}_2}\vec{y}_2 \qquad \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{4}{6} = \frac{2}{3}$$

$$= \frac{2}{4}\vec{y}_1 + \frac{\boxed{1/2}}{\underline{\left(\left(\frac{3}{4}\right)^2 + 3\cdot\left(\frac{1}{4}\right)^2\right)}}\vec{y}_2$$

$$= \frac{1}{2}\vec{y}_1 + \frac{2}{3}\vec{y}_2$$

$$= \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} + \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$\vec{y}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{3} \\ 2/3 \\ 2/3 \end{bmatrix}$$

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}, \begin{bmatrix} 0 \\ -\frac{1}{3} \\ 2/3 \\ 2/3 \end{bmatrix} \right\}$$

The Gram-Schmidt Process

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

$$\begin{aligned}
\vec{v}_1 &= \vec{x}_1 \\
\vec{v}_2 &= \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 \quad = \operatorname{proj}_{\vec{v}_1}(\vec{x}_2) \\
\vec{v}_3 &= \vec{x}_3 - \left(\frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2\right) = \operatorname{proj}_{\operatorname{Span}\{\vec{v}_1, \vec{v}_2\}}(\vec{x}_3) \\
&\vdots \\
\vec{v}_p &= \vec{x}_p - \frac{\vec{x}_p \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \cdots - \frac{\vec{x}_p \cdot \vec{v}_{p-1}}{\vec{v}_{p-1} \cdot \vec{v}_{p-1}}\vec{v}_{p-1}
\end{aligned}$$

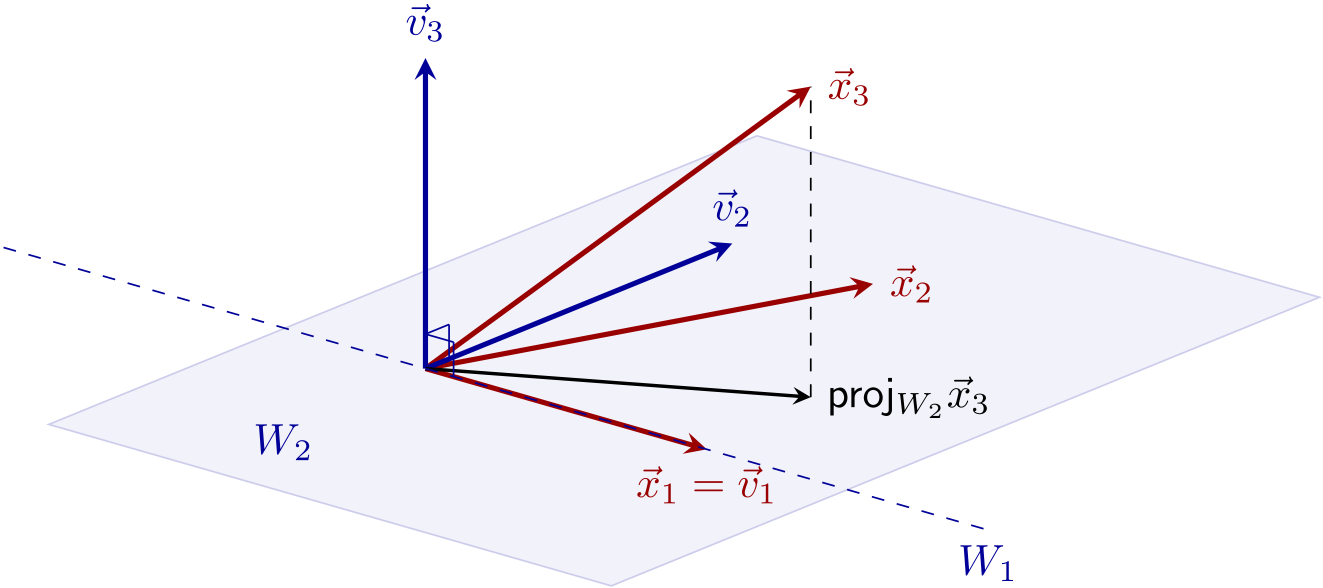
Then, $\{\vec{v}_1, \dots, \vec{v}_p\}$ is an orthogonal basis for W.

$\left\{ \frac{\vec{v}_1}{\|\vec{v}_1\|}, \frac{\vec{v}_2}{\|\vec{v}_2\|}, \dots \frac{\vec{v}_p}{\|\vec{v}_p\|} \right\}$ orthonormal basis

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.

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

$$\|\vec{v}_2\| = \sqrt{3^2+2^2+0^2} = \sqrt{13} \qquad \|\vec{v}_2\| = \sqrt{(-2)^2+3^2+1^2} = \sqrt{14}$$

$$\left\{ \frac{1}{\sqrt{13}} \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \ \frac{1}{\sqrt{14}} \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} \right\} \quad \text{is orthonormal.}$$

$\mathcal{B} = \{ \vec{x}_1, \vec{x}_2, \dots, \vec{x}_p \}$ basis for W

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

$\vec{y}_2 = \vec{x}_2 - \text{proj}_{\vec{y}_1}(\vec{x}_2)$

$\vec{y}_3 = \vec{x}_3 - \text{proj}_{\text{Span}\{\vec{y}_1, \vec{y}_2\}}(\vec{x}_3)$

$\vdots$

$\vec{y}_p = \vec{x}_p - \text{proj}_{\text{Span}\{\vec{y}_1, \dots, \vec{y}_{p-1}\}}(\vec{x}_p)$

$\tilde{\mathcal{B}} = \{ y_1, \dots, y_p \}$ orthogonal.

$u_q = y_q / \| y_q \|$

$\mathcal{C} = \{ u_1, \dots, u_q \}$ orthonormal.

Gram – Schmidts Process .

$\{x_1, x_2, \cdots, x_p\}$ $\left\{ \begin{array}{l} \text{linearly independent} \\ \text{a basis for } W = \operatorname{Span}\{x_1, \cdots, x_p\} \end{array} \right.$

$$y_1 = x_1$$

$$y_2 = x_2 - \operatorname{proj}_{y_1}(x_2) = x_2 - \frac{x_2 \cdot y_1}{y_1 \cdot y_1} y_1$$

$$y_3 = x_3 - \operatorname{proj}_{\operatorname{Span}\{y_1, y_2\}}(x_3) = x_3 - \left(\frac{x_3 \cdot y_1}{y_1 \cdot y_1} y_1 + \frac{x_3 \cdot y_2}{y_2 \cdot y_2} y_2 \right)$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$y_p = x_p - \operatorname{proj}_{\operatorname{Span}\{y_1, y_2, \cdots, y_{p-1}\}}(x_p)$$

$\{y_1, \cdots, y_p\}$ orthogonal

$u_q = \dfrac{y_q}{\|y_q\|} \Rightarrow \{u_1, \cdots, u_p\}$ orthonormal.

$$\vec{x}_1 = (x_1 \cdot u_1)\,\vec{u}_1 + \cancel{(x_1 \cdot u_2)\,u_2 + \cdots + (x_1 \cdot u_p)\,u_p}$$

$$\vec{x}_2 = (x_2 \cdot u_1)\,\vec{u}_1 + (x_2 \cdot u_2)\,\vec{u}_2 \cancel{+ \cdots + (x_2 \cdot u_p)\,u_p}$$

$$\vec{x}_3 = (x_3 \cdot u_1)\,\vec{u}_1 + (x_3 \cdot u_2)\,\vec{u}_2 + (x_3 \cdot u_3) \cdot \vec{u}_3$$

$$\vdots$$

$$\vec{x}_p = (x_p \cdot u_1) \cdot \vec{u}_1 + (x_p \cdot u_2) \cdot \vec{u}_2 + \cdots + (x_p \cdot u_p)\,\vec{u}_p .$$

$$\vec{A} = \overset{\in \mathbb{R}^{m \times n}}{\begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \end{bmatrix}} = \underset{\substack{\uparrow \\ \text{matrix with} \\ \text{orthonormal columns}}}{\begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{bmatrix}} \cdot \overset{\substack{\text{upper triangular} \\ \downarrow}}{\begin{bmatrix} (x_1 \cdot u_1) & (x_2 \cdot u_1) & (x_3 \cdot u_1) & & (x_p \cdot u_1) \\ 0 & (x_2 \cdot u_2) & (x_3 \cdot u_2) & & (x_p \cdot u_2) \\ 0 & 0 & (x_3 \cdot u_3) & \cdots & \vdots \\ \vdots & \vdots & 0 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & (x_p \cdot u_p) \end{bmatrix}}$$

$\vec{x}_q \in \mathbb{R}^m$ $\quad (p = n)$

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 $\operatorname{Col} A$.
2. R is $n \times n$, upper triangular, with positive entries on its diagonal, and the length of the j^{th} column of R is equal to the length of the j^{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 $A = \underbrace{\begin{array}{cc} \overrightarrow{x_1} & \overrightarrow{x_2} \\ \| & \| \\ 3 & -2 \\ 2 & 3 \\ 0 & 1 \end{array}}_{\text{orthogonal}}$.

$x_1 \cdot x_2 = 0$

$\|\overrightarrow{x_1}\| = \sqrt{3^2+2^2+0^2} = \sqrt{13}$

$\|\overrightarrow{x_2}\| = \sqrt{(-2)^2+3^2+1^2} = \sqrt{14}$

$u_1 = \frac{1}{\sqrt{13}} \cdot \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$ $\qquad u_2 = \frac{1}{\sqrt{14}} \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$

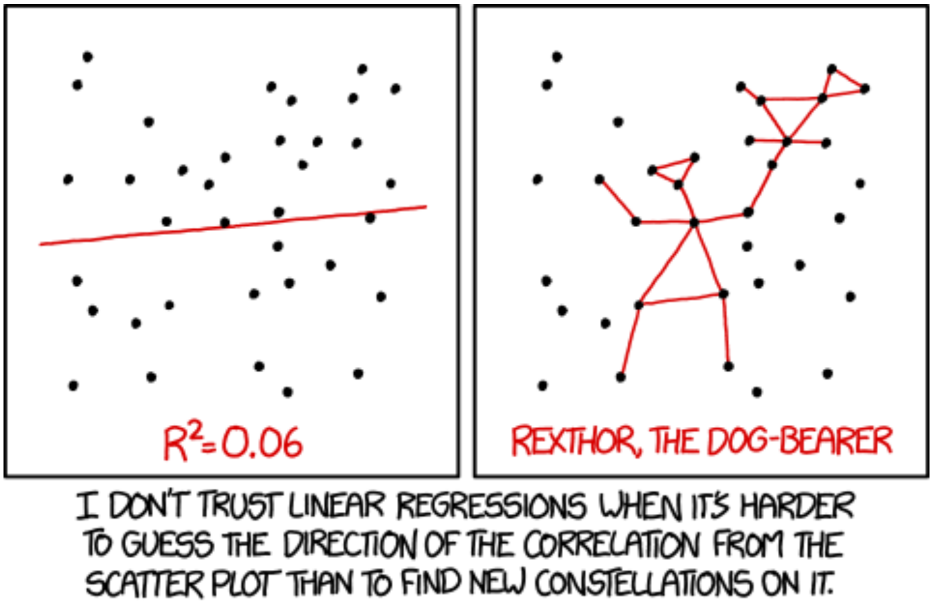
$$Q = [u_1 \quad u_2] = \begin{bmatrix} \frac{3}{\sqrt{13}} & -\frac{2}{\sqrt{14}} \\ \frac{2}{\sqrt{13}} & \frac{3}{\sqrt{14}} \\ 0 & \frac{1}{\sqrt{14}} \end{bmatrix}$$

$$R = \begin{bmatrix} (x_1 \cdot u_1) & \underline{(x_2 \cdot u_1)} \\ 0 & (x_2 \cdot u_2) \end{bmatrix} = \begin{bmatrix} \sqrt{13} = \|\overrightarrow{x_1}\| & 0 \\ 0 & \underset{\|\overrightarrow{x_2}\|}{\underset{\|}{\sqrt{14}}} \end{bmatrix}$$

Section 6.5 : Least-Squares Problems

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



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

$$y = mx + b \Rightarrow$$

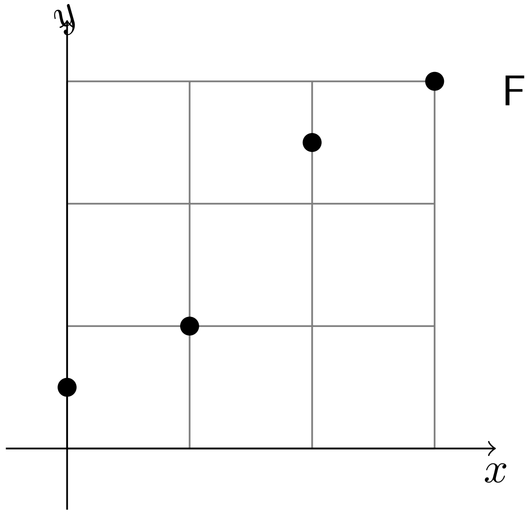
$$0.5 = m \cdot 0 + b \cdot 1$$

that best fits the data below.

$$1 = m \cdot 1 + b \cdot 1$$

$$2.5 = m \cdot 2 + b \cdot 1$$

$$3 = m \cdot 3 + b \cdot 1$$



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?

$A\vec{x} = \vec{b}$ inconsistent.

Consistent means $\exists\ \vec{x}_0$ such that $\boxed{A\vec{x}_0 = b}$

$\min \|A\vec{x} - \vec{b}\| \boxed{= 0}$

$\hat{x}$ is a least square solution if

$$\min_{\vec{x}} \| A\vec{x} - b \| = \| A\hat{x} - \vec{b} \|$$

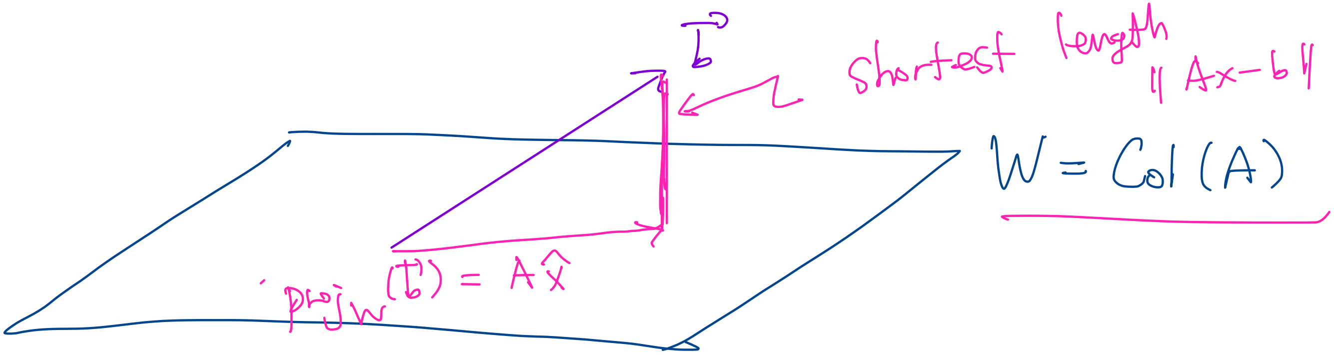
The Least Squares Solution to a Linear System

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 $\widehat{x}$ for which

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

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



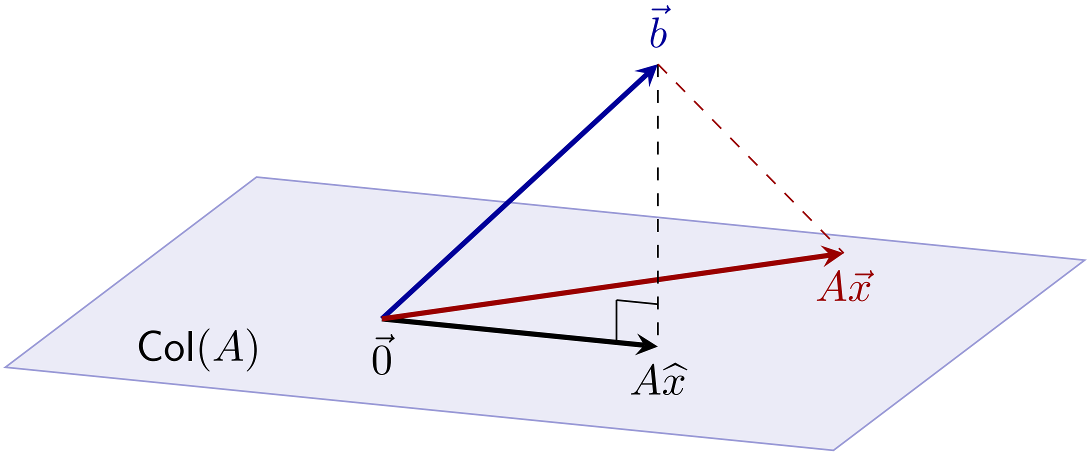
Consistent $\Longleftrightarrow$ $\vec{b} \in W = Col(A)$

$A\vec{x} = \vec{b}$.

How to find $\hat{x}$?

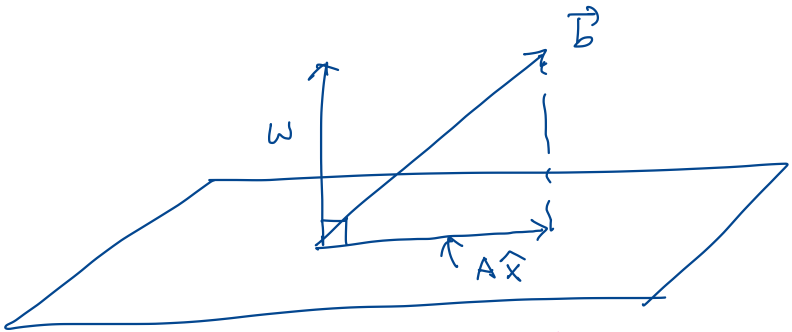
Solve $A\vec{x} = proj_{Col(A)}(\vec{b})$ ← consistent.

A Geometric Interpretation



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

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



Col(A)

$A\hat{x} \in Col(A)$

$w \in Col(A)^{\perp} = Nul(A^T)$

$A^T \cdot w = 0$

$\vec{b} = A\hat{x} + w$

$A^T \cdot \vec{b} = A^T \cdot A \cdot \hat{x} + \underbrace{A^T \cdot w}_{0}$

Always consistent

$\boxed{A^T \cdot A \cdot \hat{x} = A^T \vec{b}}$

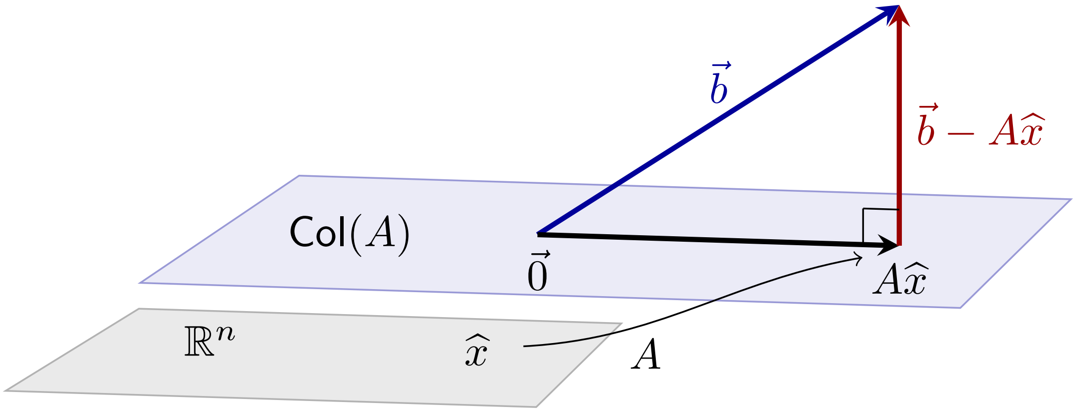
The Normal Equations

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

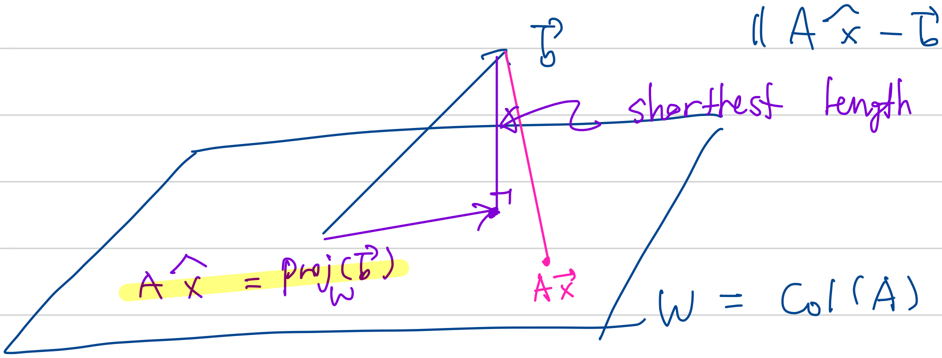
$A\vec{x} = \vec{b}$

this holds if consistent.

- $\vec{x}_0$ is a solution if $\|A\vec{x}_0 - \vec{b}\| \underset{\sim}{=} \underset{\sim}{0} = \min_{\vec{x}} \|A\vec{x} - \vec{b}\|$

- $\hat{x}$ is a least-squares solution if

$$\|A\hat{x} - \vec{b}\| = \min_{\vec{x}} \|A\vec{x} - \vec{b}\|$$



(i) Solution of $A\vec{x} = \text{proj}_W(\vec{b})$

= Least squares solution of $Ax = b$.

$\vec{b} = A\hat{x} + w$, $\quad w \in W^\perp = \text{Col}(A)^\perp = \text{Nul}(A^T)$

$A^T\vec{b} = A^TA\hat{x} + \underbrace{A^Tw}_{=0}$ $\quad \underline{A^T \cdot w = 0}$

(ii) $A^TA\hat{x} = A^T \cdot \vec{b}$: Normal Equation.

↑

Note : is always $\underline{\underline{\text{consistent}}}$. (Explain later)

↑

$\underbrace{A^TA\vec{x} = A^T\vec{b}}$ is $\underline{\text{consistent}}$ if $\boxed{A^T}\underline{\vec{b}} \in \text{Col}(A^TA)$

$$\text{Col}(A^TA) \;\|\; \text{Col}(A^T)$$

Remark (i) $A^T \cdot A$ is square ($A \in \mathbb{R}^{m\times n}$, $A^TA \in \mathbb{R}^{n\times n}$)

(ii) A^TA is symmetric (B is symmetric if $B = B^T$)

($\because$ $(A^TA)^T = A^T \cdot (A^T)^T = A^T \cdot A$)

Example

(iii) $tr(A^TA)$ = sum of diagonals ≥ 0

(length of A = $\sqrt{tr(A^T \cdot A)}$)

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

$A^TA\vec{x} = A^T\vec{b}$

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

Solution:

$$A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4^2+0^2+1^2 & 4\cdot 0+0\cdot 2+1\cdot 1 \\ 4\cdot 0+0\cdot 2+1\cdot 1 & 0^2+2^2+1^2 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 4\cdot 2 + 1\cdot 11 \\ 0\cdot 2+2\cdot 0+1\cdot 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}}_{} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{17\cdot 5 - 1} \cdot \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \cdots$$

$tr(A^T \cdot A)$ = sum of squares of entries in A.

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

Assume A has linearly independent columns.

$\Rightarrow B = A^T \cdot A$ is invertible.

Proof WANT: $T_B(\vec{x}) = B \cdot \vec{x}$ is 1 - 1

$B \cdot \vec{x} = 0$ implies $\vec{x} = 0$

$B \cdot \vec{x} = 0$ has the only trivial solution.

Suppose $B\vec{x} = A^TA \cdot \vec{x} = 0$

$$0 = \vec{x} \cdot (A^T \cdot A \vec{x}) = (\underline{A \cdot \vec{x}}) \cdot (\underline{A\vec{x}}) = \| A\vec{x} \|^2$$

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

$\Rightarrow A\vec{x} = 0 \Rightarrow \vec{x} = 0 \Rightarrow A^TA$ is invertible. ▨

$$\begin{tikzcd}
A^TA \cdot \vec{x} = 0 \arrow[r, phantom, "\Longleftrightarrow"] \arrow[d, Leftrightarrow] & \vec{x} \in \underbrace{Nul(A^TA)} \arrow[d, phantom, "\|"] \\
A\vec{x} = 0 \arrow[r, phantom, "\Longleftrightarrow"] & \vec{x} \in \underbrace{Nul(A)}
\end{tikzcd}$$

$Nul(A^TA) = Nul(A)$

$Nul(A^TA)^{\perp} = Nul(A)^{\perp}$

$* \; Col(A^TA) = Col(A^T)$

$A\vec{x} = \vec{b}$

$A^T \cdot A \vec{x} = A^T \cdot \vec{b}$

If A has linearly independent columns.

$\Rightarrow$ $A^T \cdot A$ is invertible

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

Theorem

$A = QR \Rightarrow R\hat{x} = Q^T \vec{b}$

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

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

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

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.)

A has linearly indep. columns

↳ A^TA is invertible $\Rightarrow A^TA\vec{x} = A^T\vec{b}$

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

$A = [x_1, ---- . x_n]$

Gram - Schmidts $\Rightarrow \underline{Q = [u_1 \; --- \; u_n]}$ orthonormal

$\Downarrow$

$A = Q\underline{R}$ upper-triangula

$Q^TQ = I$

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

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

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

$$A\vec{x} = \vec{b}$$

$$QR\vec{x} = \vec{b}$$

$$\underbrace{Q^TQ}R\vec{x} = Q^T\vec{b}$$

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

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

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

Solution. The QR decomposition of A is

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

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

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

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

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

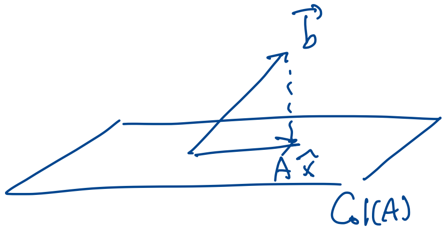
$2x_3 = 4 \quad \therefore \quad x_3 = 2$ ✓

$2x_2 + 3\underset{\overset{\uparrow}{2}}{x_3} = -6 \qquad x_2 = -6$

$x_1 = ?$

Recall $\quad A\vec{x} = \vec{b}$ $\qquad \hat{x}$ is least squares solution if

$$\|A\hat{x} - \vec{b}\| = \min_{\vec{x}} \|A\vec{x} - \vec{b}\|$$



(i) $\quad A\hat{x} = \text{Proj}_{\text{Col}(A)}(\vec{b})$

(ii) $\quad A^TA\hat{x} = A^T\vec{b}$ $\quad$ (Normal Equation)

If A has linearly indep. columns

(iii) $\quad A^TA$ is invertible, $\quad \hat{x} = (A^TA)^{-1}\cdot A^T\cdot\vec{b}$ (unique solution)

(iv) $\quad A = \underset{\sim}{Q}R$ $\qquad A\vec{x} = \vec{b} \Rightarrow \underset{\sim}{R}\vec{x} = Q^T\vec{b}$

$Q^TQ = I$ $\qquad$ upper triangular.

Example

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

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

$\vec{x}_1 =$ $\qquad = \vec{x}_2$

$\vec{x}_1\cdot\vec{x}_2 = -6-2+1+7$

$= 0$

Hint: the columns of A are orthogonal. $\Rightarrow$ linearly indep.

Normal Equ

$$A^T\cdot A = \begin{bmatrix}1 & 1 & 1 & 1\\ -6 & -2 & 1 & 7\end{bmatrix}\begin{bmatrix}1 & -6\\ 1 & -2\\ 1 & 1\\ 1 & 7\end{bmatrix} = \begin{bmatrix}4 & 0\\ 0 & 90\end{bmatrix}$$

$$A^T\cdot b = \begin{bmatrix}1 & 1 & 1 & 1\\ -6 & -2 & 1 & 7\end{bmatrix}\begin{bmatrix}-1\\ 2\\ 1\\ 6\end{bmatrix} = \begin{bmatrix}8\\ 45\end{bmatrix}$$

$$\begin{bmatrix}4 & 0\\ 0 & 90\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}8\\ 45\end{bmatrix}$$

$4x = 8$

$90\cdot y = 45$ $\quad\Rightarrow\quad$ $\boxed{x = 2 \quad y = \frac{1}{2}}$

QR Decomposition.

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix} = [\vec{x}_1 , \vec{x}_2]$$

$$u_1 = \frac{\vec{x}_1}{\|\vec{x}_1\|} = \frac{1}{2} \cdot \vec{x}_1$$

$$u_2 = \frac{\vec{x}_2}{\|\vec{x}_2\|} = \frac{1}{3\sqrt{10}} \cdot \vec{x}_2$$

$$\{\vec{u}_1 , \vec{u}_2\} \quad \text{orthonormal}$$

$$Q = \underbrace{[\vec{u}_1 \quad \vec{u}_2]}_{}$$

$$\underline{A = Q \cdot R}$$

$$R = \begin{bmatrix} x_1 \cdot u_1 & x_2 \cdot u_1 \overset{}{=} 0 \\ 0 & x_2 \cdot u_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \|x_1\| \overset{}{=} 2 & 0 \\ 0 & \|x_2\| \end{bmatrix}}_{= 3\sqrt{10}}$$

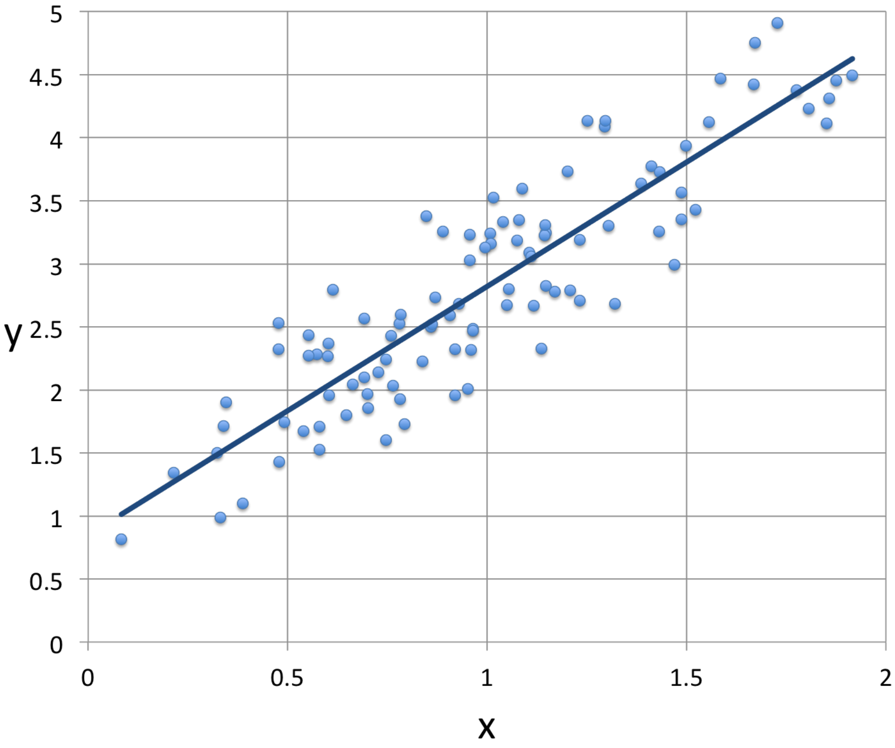
$$x_1 \cdot u_1 = x_1 \cdot \frac{x_1}{\|x_1\|} = \frac{\|x_1\|^2}{\|x_1\|} = \|x_1\|$$

$$A\vec{x} = \vec{b} \quad \Rightarrow \quad R\vec{x} = \underset{\sim}{Q^T} \vec{b}$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 3\sqrt{10} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{6}{3\sqrt{10}} & -\frac{2}{3\sqrt{10}} & \frac{1}{3\sqrt{10}} & \frac{7}{3\sqrt{10}} \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

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

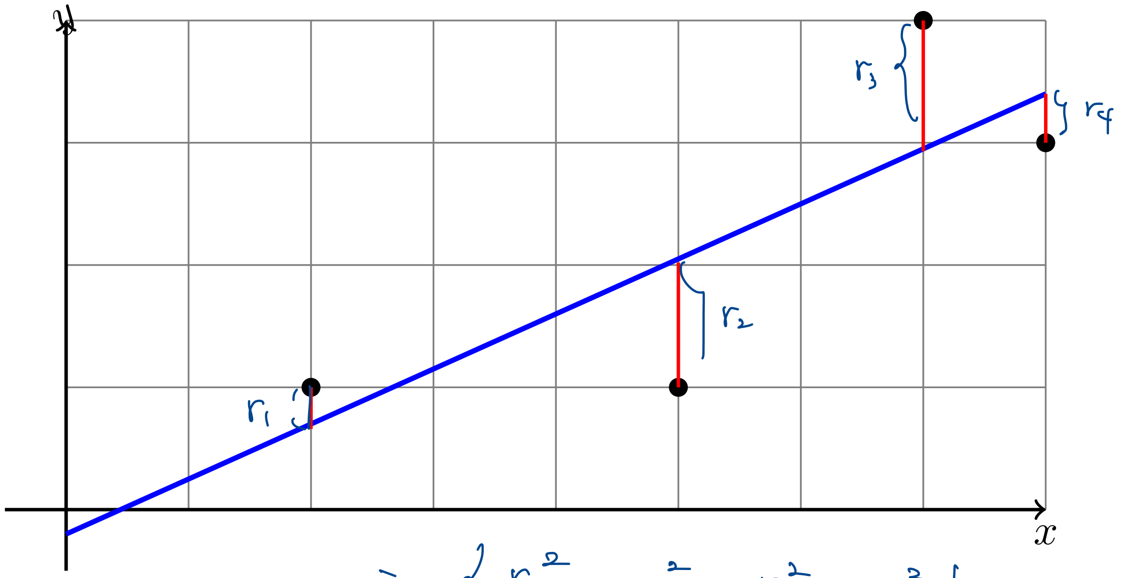
x	2	5	7	8
y	1	1	4	3

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 difference between data on line.

The least squares line minimizes the sum of squares of the errors.



$$\min \{ r_1^2 + r_2^2 + r_3^2 + r_4^2 \}$$

$$= \min \| Ax - b \|^2$$

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

$$1 = \beta_0 + \beta_1 \cdot 2$$
$$1 = \beta_0 + \beta_1 \cdot 5$$
$$4 = \beta_0 + \beta_1 \cdot 7$$
$$3 = \beta_0 + \beta_1 \cdot 8$$

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 : $X\vec{\beta} = \vec{y}$.

$$\hat{\beta} = ?$$

$$X^T \cdot X \cdot \hat{\beta} = X^T \cdot \vec{y}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \cdot \hat{\beta} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 4 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

The normal equations are

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 & \\ & & & & \end{bmatrix} \begin{bmatrix} 1 & \\ 1 & \\ 1 & \\ 1 & \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

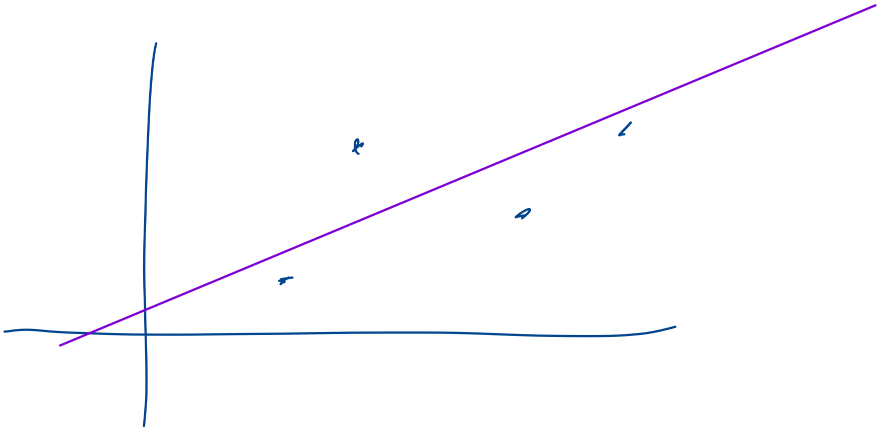
$$X^T\vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & \\ & & & & \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

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, β_0 is negative, and β_1 is positive.



Ex) $f_1(x) = x$, $f_2(x) = x^2$, $f_3(x) = e^x$ ...

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 $\underline{y = c_1 x + c_2 x^2}$ that best fits the data.

① — $\underline{2 = -c_1 + c_2}$

$$\left.\begin{array}{l} 1 = c_1 \cdot 0 \quad c_2 \cdot 0 \\ 0 = c_1 \cdot 0 \quad c_2 \cdot 0 \end{array}\right\}$$

② — $\underline{6 = c_1 + c_2}$

$8 = 2 \cdot c_2$ ∴ $c_2 = 4$

$c_1 = 2$

$$\begin{bmatrix} -1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 6 \end{bmatrix}$$

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

$$\begin{bmatrix} -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix} \Rightarrow \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

WolframAlpha and Mathematica Syntax

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

WolframAlpha

$$\texttt{linear fit } \{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}\}$$

Mathematica

$$\texttt{LeastSquares}[\{\{x_1, x_1, y_1\}, \{x_2, x_2, y_2\}, \ldots, \{x_n, x_n, y_n\}\}]$$

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