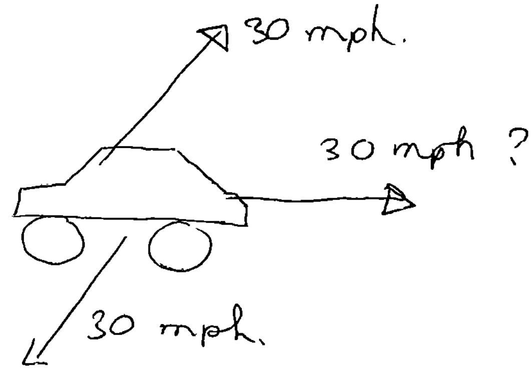
MATH 349: Elementary Linear Algebra

Chapter 1: Vectors. Introduction to the geometry and algebra of $\mathbb{R}^n$.

Motivation: In physics, many measurable quantities can be described by a single real number (e.g. temperature, length, mass, etc.).

Other important quantities have both a magnitude and a direction (e.g. speed, acceleration, force, etc.)

Example: Speed. "A car moves at 30 mph" does not describe the movement of the car fully.



Need to specify a direction...

Scalar quantities: temperature, length, mass, etc.

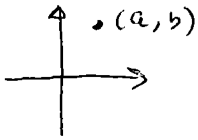
Vector quantities: speed, acceleration, force, etc.

A vector has:
1) A magnitude
2) An orientation
3) (Not always) An origin. [If no origin is specified, we generally assume the origin is the origin of the Cartesian plane]

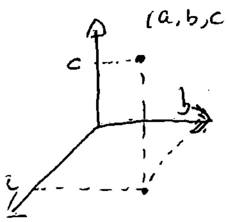
We represent vectors using arrows.

Notation: Euclidean space

$\mathbb{R}$ = the set of real numbers $(1, 2, \sqrt{2}, \pi, e, \text{etc.})$



$\mathbb{R}^2 = \{(a,b) : a, b \in \mathbb{R}\}$ plane, pair of ordered coordinates.



$\mathbb{R}^3 = \{(a,b,c) : a,b,c \in \mathbb{R}\}$ 3 dimensional space

(Read "r two", "r three")

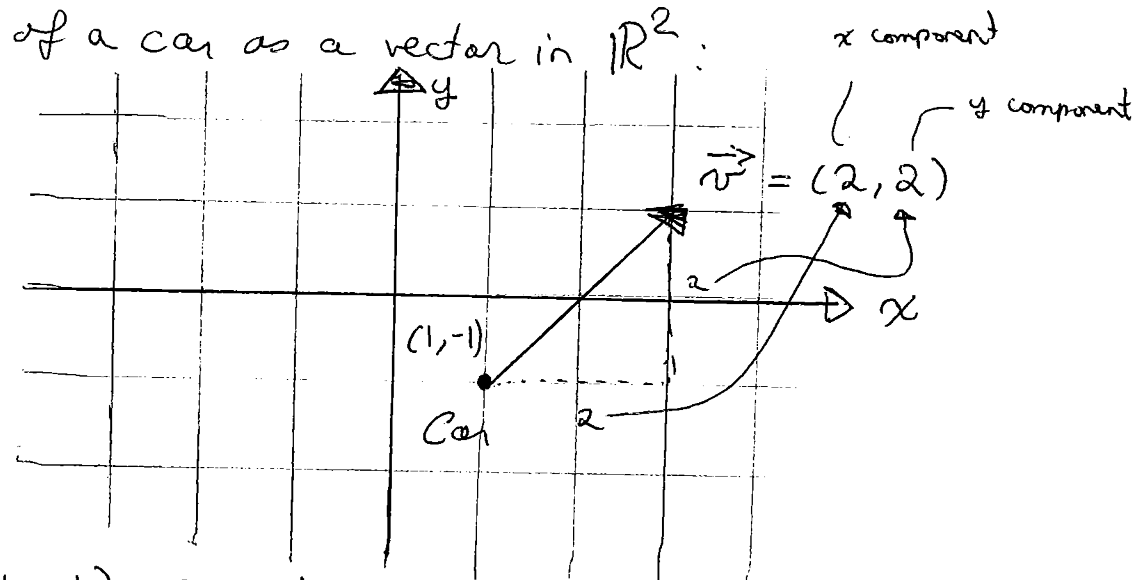
More generally, given a natural number n, we define

$$\mathbb{R}^n = \{(a_1, a_2, \dots, a_n) : a_1, a_2, \dots, a_n \in \mathbb{R}\}$$ n-tuple

The n-dimensional Euclidean space.

We naturally repesent vectors using points in $\mathbb{R}^n$.

Example: speed of a car as a vector in $\mathbb{R}^2$:



Car is at $(1,-1)$. • Speed in NE direction.

• Magnitude of speed = length of vector

$= \sqrt{2^2 + 2^2} = \sqrt{8}$

Notation: Vectors are often denoted by bold letters (see book), or by adding an arrow above a letter (e.g. $\vec{v}, \vec{w}, \vec{x}$, etc.)

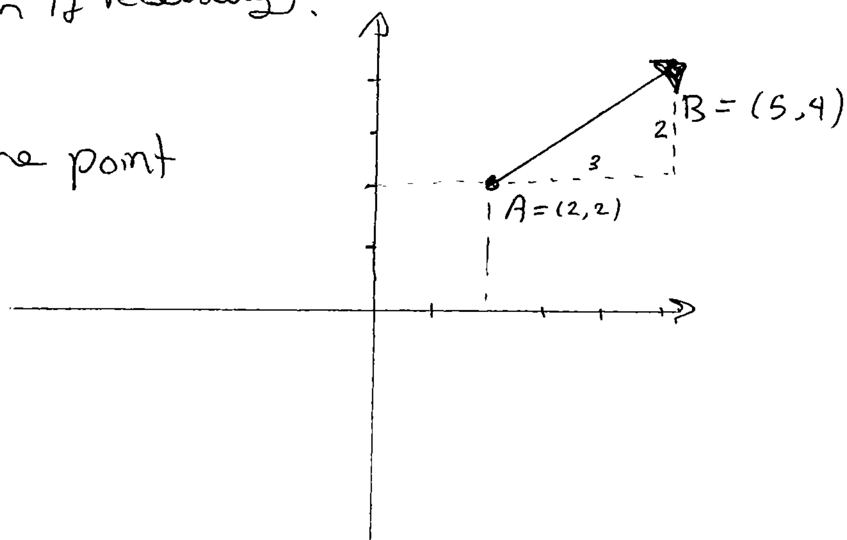
In mathematics, we generally just write $v \in \mathbb{R}^2$ ("v belongs to $\mathbb{R}^2$")

Clearly, every point in $\mathbb{R}^2$ defines a vector (origin $= (0,0)$ or specify).

Conversely, every vector in the plane can be associated to a point in $\mathbb{R}^2$ (translate to origin if necessary).

$\vec{AB} = (3,2)$ (origin $= A$)

Vector from one point to another



Notation: $\vec{AB}$

If $A = (x_1, y_1)$
$B = (x_2, y_2)$ then $\vec{AB} = (x_2 - x_1, y_2 - y_1)$.

<u>Notation and terminology:</u>

- If $v = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ is a vector in $\mathbb{R}^n$, then $v_1, v_2, \dots, v_n$ are called the "components" of v.

<u>Ex:</u> $v = \vec{AB} = (3,2)$. x-component of $v = 3$
y-component of $v = 2$.

- Other notations for vectors: Some people like to use square brackets
$$v = [3,2].$$

- Can use <u>row</u> or <u>column</u> vectors $v = (3,2)$ or $v = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$
$v = [3,2]$ or $v = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

* For now, we will just use row vectors. You will understand later why column vectors are useful. *

- The "zero vector": $\vec{0}_{1\times n} = \vec{0} = \underbrace{(0,0,\dots,0)}_{n \text{ components}}$ (Can use $\vec{0}$ if the dimension is clear from problem/context)

Vectors have a very natural <u>algebraic structure</u>:

- <u>Addition</u>: $\vec{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$

  $\vec{w} = (w_1, w_2, \dots, w_n) \in \mathbb{R}^n$.

$$\vec{v} + \vec{w} := (v_1 + w_1, v_2 + w_2, \dots, v_n + w_n) \in \mathbb{R}^n.$$

- <u>Scalar multiplication</u>: $\vec{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$

  $\lambda \in \mathbb{R}$

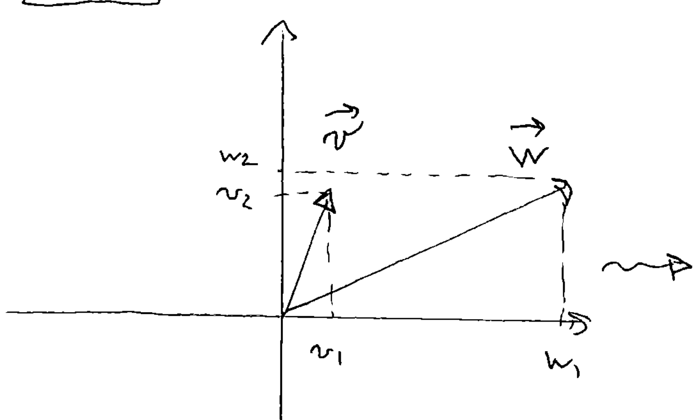
$$\lambda \vec{v} := (\lambda v_1, \lambda v_2, \dots, \lambda v_n) \in \mathbb{R}^n.$$

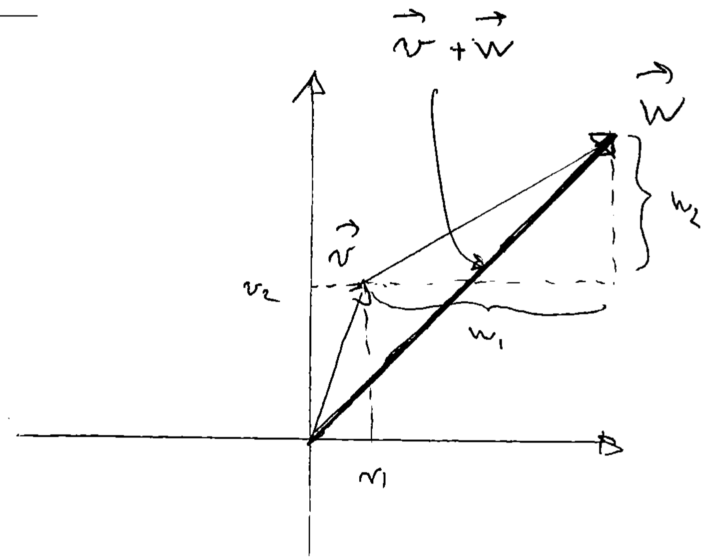
<u>Remark</u>: No obvious/natural notion of multiplication in general (more about that later).

<u>Geometric interpretation when $n = 2$</u>

$\boxed{n=2}$

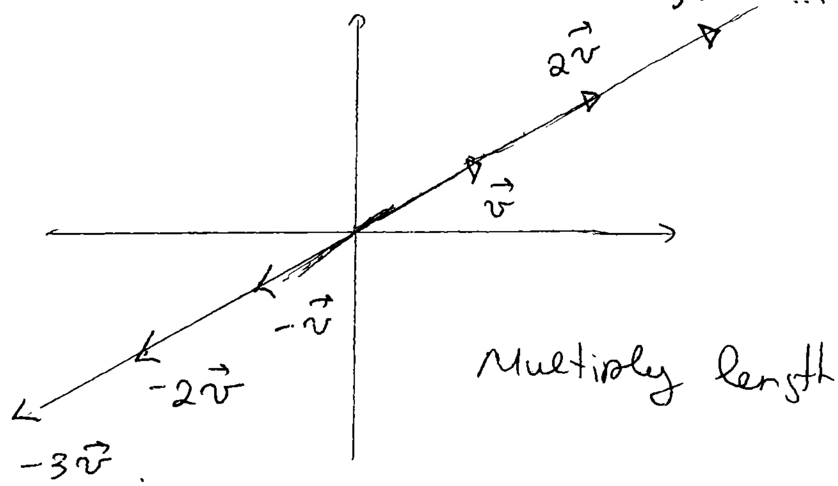
<u>Addition</u>





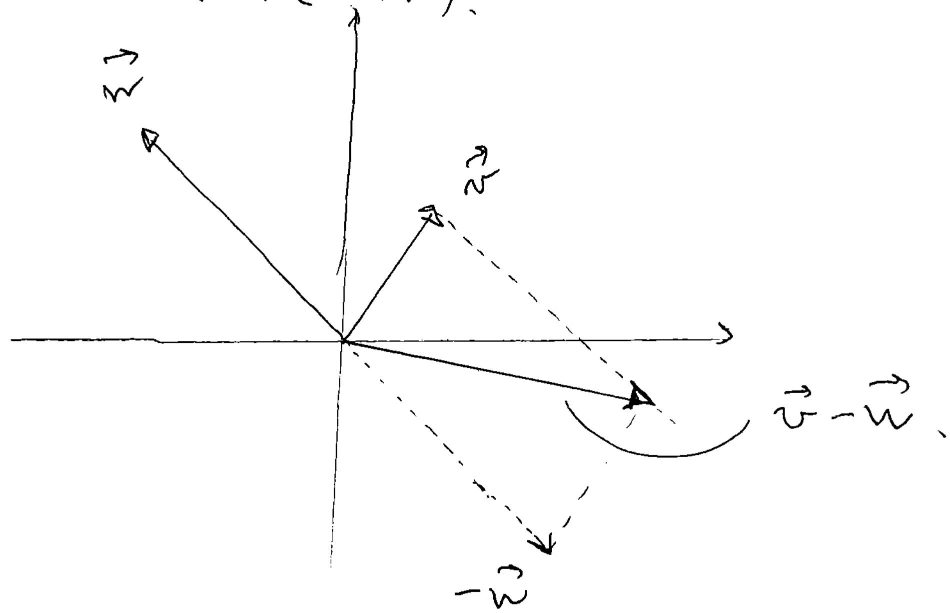
Translate one vector at the end of the other.

Scalar multiplication



Multiply length by λ.

Remark: $\vec{V} - \vec{W} = \vec{V} + (-\vec{W})$.



* See also "The parallelogram rule" in book.

** Similar geometric interpretation for $n=3$.

Example: (See Example 1.2 in book).

Example: (Book Example 1.3). $\vec{v} = (-2, 4)$.

$$2\vec{v} = (2\times(-2), 2\times 4) = (-4, 8)$$
$$\frac{1}{2}\vec{v} = \left(\frac{1}{2}(-2), \frac{1}{2}\cdot 4\right) = (-1, 2)$$

$$-2\vec{v} = (-2(-2), -2\times 4) = (4, -8)$$

(See geometric interpretation in book)

Example: $\vec{u} = (1,-1,1,-1,\dots,1,-1) \in \mathbb{R}^{42}$

$\vec{v} = (-1,1,-1,1,\dots,-1,1) \in \mathbb{R}^{42}$

Then $\vec{u}+\vec{v} = \vec{0}_{1\times 42}$

$\vec{u}-\vec{v} = (2,-2,2,-2,\dots,2,-2) \in \mathbb{R}^{42}$.

Theorem (Algebraic properties of vectors in $\mathbb{R}^n$)

Let $u, v, w \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$. Then:

A) Addition satisfies:

① $u+v = v+u$

② $(u+v)+w = u+(v+w)$

③ $u+0 = u$ (here $0 = \vec{0}$ of the same dimension as u).

④ $u+(-u) = 0$

B) Scalar multiplication satisfies:

⑤ $\lambda(u+v) = \lambda u + \lambda v$

⑥ $(\lambda+\mu)u = \lambda u + \mu u$

⑦ $\lambda(\mu u) = (\lambda\mu)u$

⑧ $1u = u$ (here 1 is the real number "one")

Proof: All trivial. We will show ① (Please make sure you know how to prove the other properties.)

① Let $u = (u_1,\dots,u_n)$ & $v = (v_1,\dots,v_n)$, then by definition:

$u+v = (u_1+v_1,\dots,u_n+v_n)$. By commutativity of the addition in $\mathbb{R}$, it follows that $u+v = (u_1+v_1,\dots,u_n+v_n) = (v_1+u_1,\dots,v_n+u_n) = v+u$.

Definition: A vector v is said to be a <u>linear combination</u> of the
(Important) vectors $v_1, v_2, \dots, v_K$ if there exist scalars $\lambda_1, \dots, \lambda_K \in \mathbb{R}$ such that:

$$\boxed{v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_K v_K}$$

Example: (Example 1.6 in book) $v = \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix}$ is a linear combination of

$v_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, $v_2 = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}$, $v_3 = \begin{pmatrix} 5 \\ -4 \\ 0 \end{pmatrix}$ since:

$$v = 3v_1 + 2v_2 - v_3 \quad \text{(check!)}$$

Example: Let $e_1 = (1, 0)$, $e_2 = (0, 1)$. Then any vector $v = (x, y) \in \mathbb{R}^2$

can be written as $v = x e_1 + y e_2$, i.e., every vector in $\mathbb{R}^2$ is a linear combinaison of the two "elementary" vectors e_1, e_2.
(This is an important idea the will be explored in detail later).

Question to think about: Given two fixed vectors $v_1, v_2 \in \mathbb{R}^2$, can we always write any given vector $v = (x, y) \in \mathbb{R}^2$ as a linear combination of v_1 and v_2?

For example, if $v_1 = (1, 2)$ and $v_2 = (1, 3)$, can we always find λ_1, λ_2 s.t. $(x, y) = \lambda_1 v_1 + \lambda_2 v_2$? Can you solve for λ_1, λ_2?

Dot product, length, angle, and projection:

The notions of length and angle between vectors can naturally be expressed using the "dot product" (inner product):

Definition: (Dot product / Inner product)

Let $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{R}^n$ and $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$. The dot product (a.k.a. inner product) between u and v, denoted by $u \cdot v$ or $\langle u, v \rangle$, is given by:

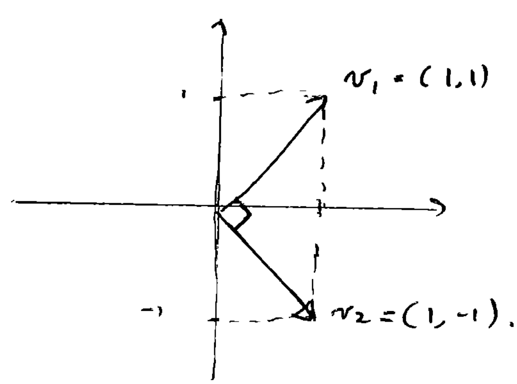
$$\boxed{\begin{aligned} u \cdot v &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \\ &= \sum_{j=1}^{n} u_j v_j . \end{aligned}}$$

Example: Let 1) $u = (1,0,1)^T$, $v = (-1,2,3)^T$. Then

$$u \cdot v = 1 \cdot (-1) + 0 \cdot 2 + 1 \cdot 3 = 2.$$

2) Let $v_1 = (1,1)^T$, $v_2 = (1,-1) \in \mathbb{R}^2$. Then

$$v_1 \cdot v_2 = 1 \cdot 1 + 1 \cdot (-1) = 0.$$



* Note: v_1 and v_2 are perpendicular.

Theorem: (Properties of the dot product)

Let $u, v, w \in \mathbb{R}^n$ and let $\lambda \in \mathbb{R}$. Then:

1). a) $u \cdot (v+w) = u \cdot v + u \cdot w$
b) $u \cdot (\lambda v) = \lambda (u \cdot v)$ } (linearity)

2) $u \cdot v = v \cdot u$ (symmetry)

3) $u \cdot u \geq 0$ and $u \cdot u = 0$ iff $u = 0$ (positive definiteness)

"if and only if"

Remark: The dot product is a mapping that takes two vectors of the same dimension, and returns a scalar:

$$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$
$$(u, v) \mapsto \langle u, v \rangle$$

According to (1), (*) $\langle u, \lambda v + \mu w \rangle = \lambda \langle u, v \rangle + \mu \langle u, w \rangle$ $(\lambda, \mu \in \mathbb{R})$.

Using (2), we also have:

$$(**) \quad \langle \lambda u + \mu v, w \rangle = \lambda \langle u, w \rangle + \mu \langle v, w \rangle.$$

(*) and (**) $\Leftrightarrow$ "Bilinearity".

We say that the dot product is a bilinear, symmetric, positive definite form.
↳ ("form" means that it maps into scalars).

Proof:

Proof of 1a): Let $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$, $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, $w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$. Then by definition:

$$u \cdot (v + w) = \sum_{i=1}^{n} u_i (v_i + w_i) \underset{\substack{\text{standard} \\ \text{distributivity} \\ \text{in } \mathbb{R}}}{=} \sum_{i=1}^{n} (u_i v_i + u_i w_i)$$

$$= \sum_{i=1}^{n} u_i v_i + \sum_{i=1}^{n} u_i w_i$$

$$\overset{\text{By def}^{n}}{=} u \cdot v + u \cdot v.$$

Proof of 3):

$$u \cdot u = \sum_{i=1}^{n} u_i^2 \geq 0 \quad \text{since } u_i^2 \geq 0 \text{ for all } i.$$

If $u \cdot u = 0$, then $\sum_{i=1}^{n} u_i^2 = 0 \implies u_i^2 = 0$ for all i

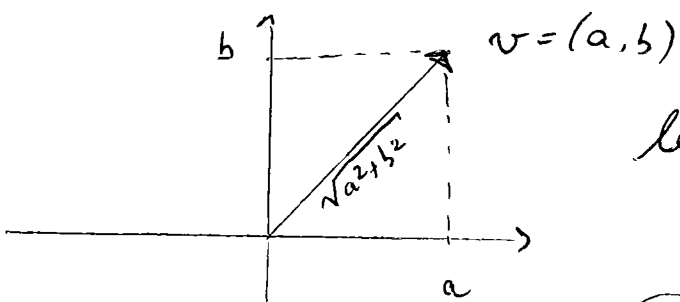
$\implies u_i = 0$ for all i

$\implies u = 0$.

(Ex): Prove the rest of the theorem.

The length of a vector:

In $\mathbb{R}^2$:



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

Remark: $\boxed{\text{length} = \sqrt{v \cdot v}}$

Definition: The norm (or length) of a vector $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$, denoted by $\|v\|$, is defined by:

$$\|v\| := \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

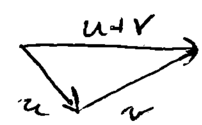
The norm generalizes the notion of "length" to arbitrary dimensions.

Remark: The second part of 3) in the previous theorem says that a vector has zero length iff it is the 0 vector.
The first part implies $\|v\| \geq 0$.

Example: $v = (3,4)^T$. $\|v\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$.

Theorem: Let $u, v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Then

1) $\|v\| = 0$ iff $v = 0$
2) $\|\lambda v\| = |\lambda| \cdot \|v\|$.
3) $\|u + v\| \leq \|u\| + \|v\|$. (Triangle inequality)



Proof: We already know (1) holds (see remark above)

(2) $\|\lambda v\| = \sqrt{(\lambda v) \cdot (\lambda v)} = \sqrt{\lambda^2 (v \cdot v)} = |\lambda| \sqrt{v \cdot v} = |\lambda| \, \|v\|$.

To prove (3), we need the Cauchy-Schwarz inequality.

Theorem: (Cauchy-Schwarz inequality)
For all vectors $u, v \in \mathbb{R}^n$, we have:

$$|u \cdot v| \leq \|u\| \cdot \|v\|.$$

(Moreover, equality holds iff $u = \lambda v$ for some $\lambda \in \mathbb{R}$)

In other words:

$$\left| \sum_{i=1}^{n} u_i v_i \right| \leq \left(\sum_{i=1}^{n} u_i^2 \right)^{1/2} \cdot \left(\sum_{i=1}^{n} v_i^2 \right)^{1/2}.$$

<u>Proof of Cauchy - Schwarz:</u>

For $t \in \mathbb{R}$, we have:

$$0 \leq \| u + t v \|^2 = (u + tv) \cdot (u + tv)$$

$$= u \cdot u + t\, u \cdot v + t\, v \cdot u + t^2\, v \cdot v$$

$$= t^2 \|v\|^2 + 2t\, u \cdot v + \|u\|^2.$$

[Recall that a polynomial $ax^2 + bx + c$ is ≥ 0 iff $b^2 - 4ac \leq 0$ ...]

This is called the "discriminant" of the polynomial.

thus,

$$4(u \cdot v)^2 - 4\|u\|^2 \cdot \|v\|^2 \leq 0.$$

It follows that

$$|u \cdot v| \leq \|u\| \cdot \|v\|.$$

∎.

We can now prove the triangle inequality:

$$\|u + v\|^2 = (u+v) \cdot (u+v)$$

$$= u \cdot u + 2\, u \cdot v + v \cdot v$$

$$\leq \|u\|^2 + 2\, |u \cdot v| + \|v\|^2$$

$$\leq \|u\|^2 + 2\, \|u\| \cdot \|v\| + \|v\|^2 \quad \text{by Cauchy-Schwarz}$$

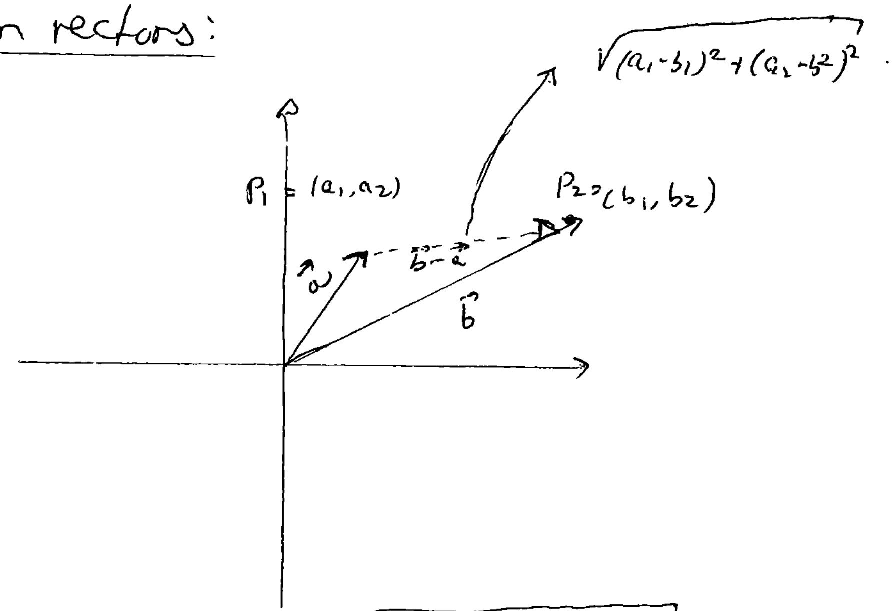
$$= (\|u\| + \|v\|)^2.$$

[Now take square roots]

∎ 12.

Distance between vectors:

Recall:



Distance between P_1 and P_2: $\sqrt{(a_1-b_1)^2+(a_2-b_2)^2}$

$$= \|\vec{b}-\vec{a}\| = \|\vec{a}-\vec{b}\|.$$

Definition. The distance between $u, v \in \mathbb{R}^n$, denoted by $d(u,v)$, is given by:

$$d(u,v) := \|u-v\|. \qquad \left(= \sqrt{\sum_{i=1}^{n}(u_i-v_i)^2}\right)$$

Angles between vectors:

We have already seen that the dot product can be used to compute length (norm) and distance. We will now show how to compute the angle between vectors. In particular, we can use the dot product to verify if vectors are orthogonal (perpendicular). [very important]

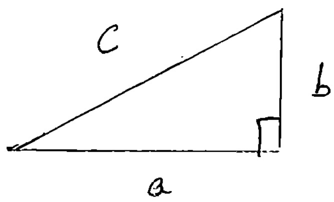
The law of cosines:

Recall: $\theta = 90^\circ = \pi/2$ $\qquad\qquad$ θ arbitrary

Remark: The law of cosines shows that $c^2 = a^2+b^2 \Leftrightarrow \theta = \pi/2$

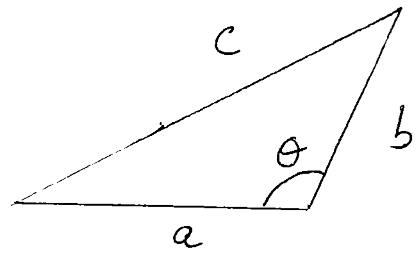
$(\Leftarrow)$ = Pythagoras

$(\Rightarrow)$ = Converse of Pythagoras



$$c^2 = a^2+b^2$$

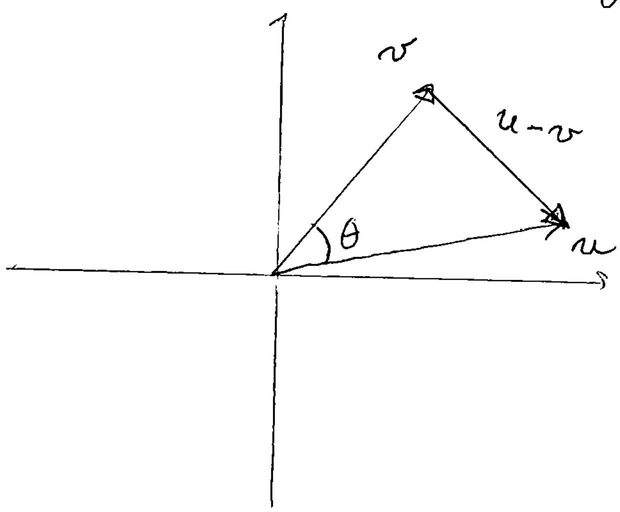
Pythagoras



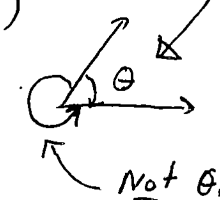
$$c^2 = a^2+b^2 - 2ab\cos\theta$$

Law of cosines

Let us reformulate the law of cosines using vectors:



(Note: We assume θ is the smallest angle between u and v)



Law of cosines: (vector form) $\quad \|u-v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\cdot\|v\|\cdot\cos\theta$

Now, we know $\|u-v\|^2 = (u-v)\cdot(u-v)$

$$= u\cdot u - 2u\cdot v + v\cdot v$$
$$= \|u\|^2 + \|v\|^2 - 2u\cdot v.$$

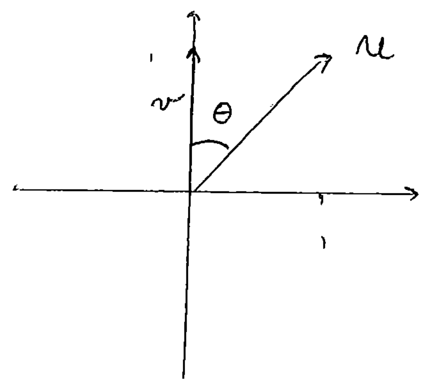
Combining the last two equations, we obtain:

$$\|u\|^2 + \|v\|^2 - 2\|u\|\cdot\|v\|\cdot\cos\theta = \|u\|^2 + \|v\|^2 - 2u\cdot v.$$

Thus,

$$\boxed{u \cdot v = \|u\| \cdot \|v\| \cdot \cos\theta} \qquad (u, v \in \mathbb{R}^2).$$

<u>Example</u>: Let $u = (1,1)^T$ and $v = (0,1)^T$. What is the angle between u and v?



(Clearly, $\theta = \pi/4$. Let's verify that...)

$$u \cdot v = u_1 v_1 + u_2 v_2 = 1 \cdot 0 + 1 \cdot 1 = 1.$$

$$\|u\| = \sqrt{u_1^2 + u_2^2} = \sqrt{1^2 + 1^2} = \sqrt{2}$$

$$\|v\| = \sqrt{v_1^2 + v_2^2} = \sqrt{1^2 + 0^2} = 1$$

$$\Rightarrow \cos\theta = \frac{u \cdot v}{\|u\| \cdot \|v\|} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}.$$

$$\Rightarrow \theta = \arccos\left(\frac{\sqrt{2}}{2}\right) = \pi/4 \quad \checkmark$$

<u>Remark</u>: $\theta = \pi/2 \iff u \cdot v = 0$. Thus, the inner product provides an easy way to check for orthogonality.

We naturally generalize the notion of angle between vectors to $\mathbb{R}^n$:

<u>Definition</u>: For nonzero vectors $u, v \in \mathbb{R}^n$, we define the angle θ between u and v by:

$$\boxed{\cos\theta := \frac{u \cdot v}{\|u\| \cdot \|v\|}.}$$

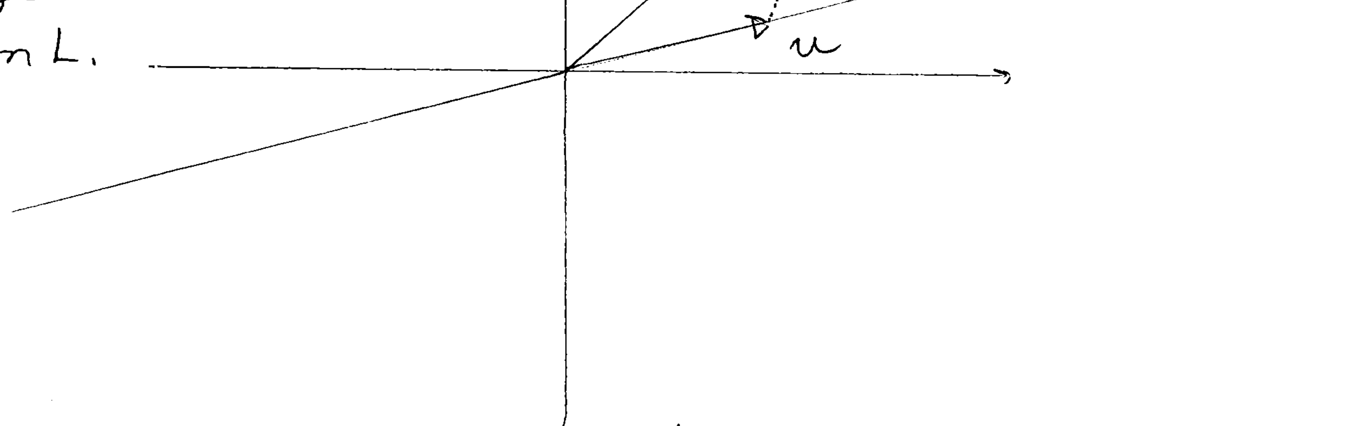
We say that two nonzero vectors $u, v \in \mathbb{R}^n$ are <u>orthogonal</u> if $\theta = \pi/2$, i.e., if $u \cdot v = 0$.

<u>Orthogonal projections</u>:

<u>Idea</u>: Given a line, find the best approximation of a vector by vectors on that line.

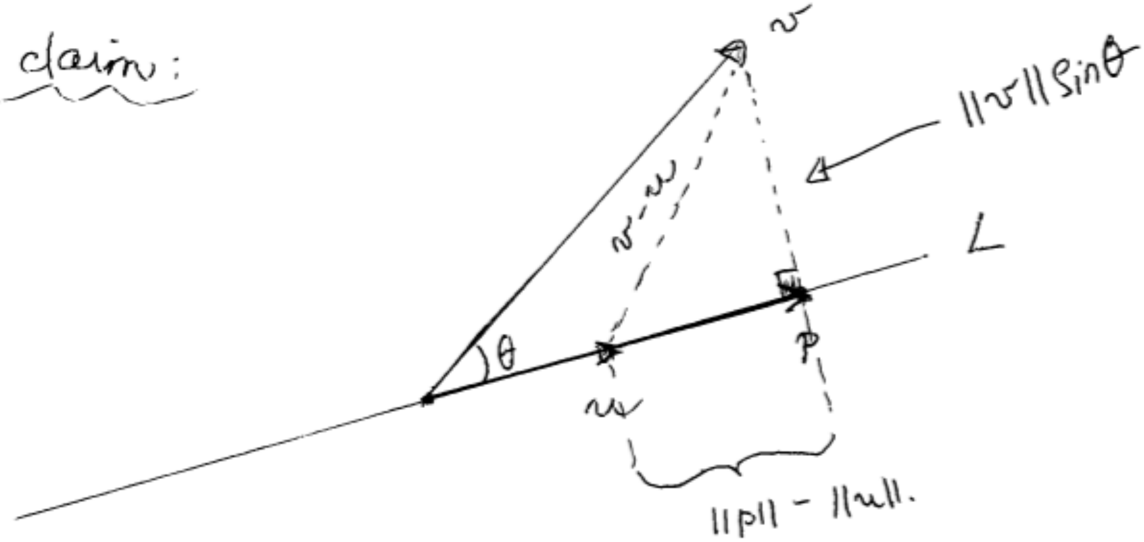
"Find u on L as close as possible to v"

* Minimize $\|v - u\|$ for u on L.



<u>Claim</u>: $\|u - v\|$ is minimal when u and $v-u$ are orthogonal.

Proof of claim:



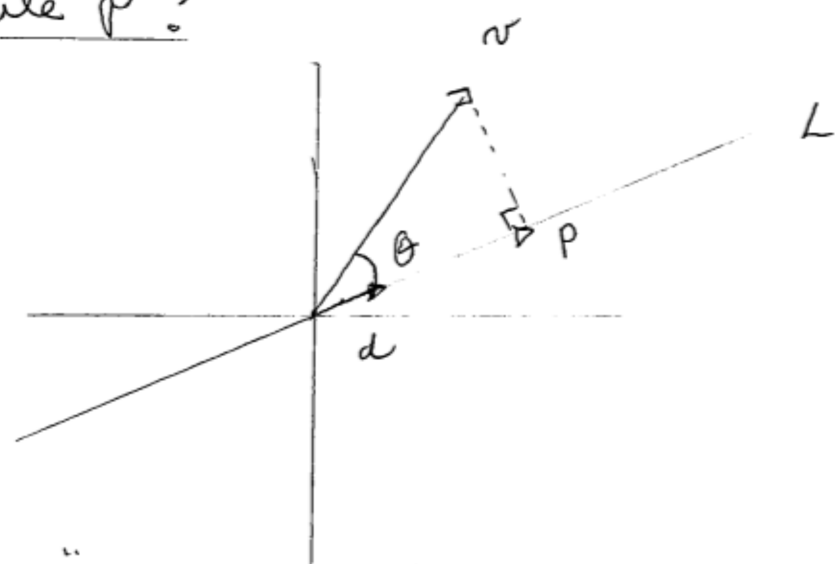
By Pythagoras' theorem: $\|v-u\|^2 = \left(\|p\| - \|u\|\right)^2 + \underbrace{\|v\|^2\sin^2\theta}$

fixed (depends only on v)

fixed (depends only on v)

Clearly, $\|v-u\|$ is minimized when $\|u\| = \|p\|$, i.e., when $u = p$. That proves the claim.

How do we compute p?



Let $d \in \mathbb{R}^2$ be a direction vector for L. We have:

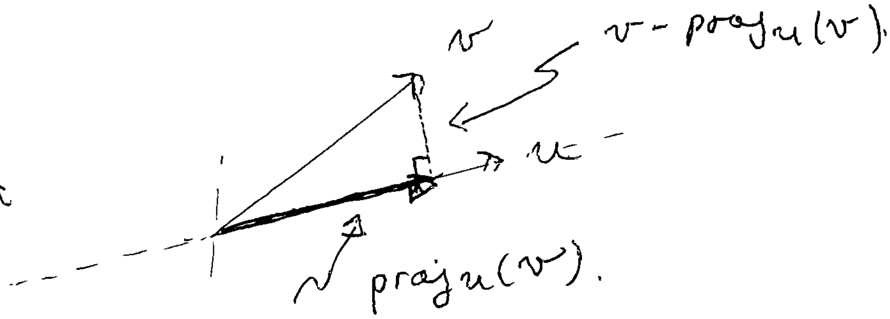
$$p = \|v\| \cos\theta \cdot \frac{d}{\|d\|} = \left(\frac{v \cdot d}{d \cdot d}\right) d.$$

Definition: Let $u, v \in \mathbb{R}^n$ with $u \neq 0$. The *projection* of v onto u, denoted $\mathrm{proj}_u(v)$ is defined by:

$$\boxed{\mathrm{proj}_u(v) := \left(\frac{u \cdot v}{u \cdot u}\right) u}$$

Geometrically:

Best approximation of v by vectors on straight line "spanned" by u.



As we saw, $\mathrm{proj}_u(v)$ and $v - \mathrm{proj}_u(v)$ are orthogonal.

Chapter 2: Systems of linear equations

Numerous problems in science and engineering involve solving one or multiple systems of linear equations.

Examples: 1) Solving systems of (not necessarily linear) equations;

2) Optimization problems where the goal is to find the min/max of a function subject to constraints

3) Solving differential equations, simulation of complex systems.

4) Chemical equations balancing, electrical circuit analysis

5) Clustering objects into categories to detect patterns.

What are "linear" equations?

Definition: A linear equation in n variables $x_1, x_2, \dots, x_n$ is an equation that can be written in the form:

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = b$$

for some scalars (real numbers) $a_1, \dots, a_n, b$.

Examples:

① $n=2$: $ax+by=c$ ⇝ line in Cartesian plane

② $n=3$: $ax+by+cz=d$ ⇝ plane in $\mathbb{R}^3$.

③ $3r+\sqrt{2}\,s+\pi t+\log 2 \cdot u$ is linear in r, s, t, u.

④ x^2+1 is NOT a linear equation.

Definition: • A system of linear equations is a finite set of linear equations, each with the same variables.

• A solution of a linear system is a vector that is simultaneously a solution of each equation in the system.

Example:

$$\begin{aligned} 2x - y &= 3 \\ x + 3y &= 5 \end{aligned}$$

has $(2,1)$ as a solution (i.e., $x=2, y=1$ is a solution of both equations)

A system of equations with real coefficients has either:

1) A unique solution
2) Infinitely many solutions
3) No solutions.

A common approach to solving linear systems is to transform the system to an equivalent system that is easier to solve.

We will generally aim to reduce a system to a "triangular system" as in the example below.

Example: Solve the system:

$$\begin{cases} x - y - z = 2 & (1) \\ \quad\; y + 3z = 5 & (2) \\ \qquad\;\; 5z = 10 & (3) \end{cases}$$

Start with last equation and work backward: (Back Substitution)

SOLVE (3) $5z = 10 \Rightarrow \boxed{z = 2}$

SUBSTITUTE IN (1), (2): $\begin{cases} x - y = 4 & (1') \\ \quad\; y = -1 & (2') \end{cases}$

SOLVE (2') $\boxed{y = -1}$

SUBSTITUTE IN (1') $\boxed{x = 3}$

unique solution $= (3, -1, 2)$

Triangular systems can be solved immediately using "back substitution".

A general linear system in n variables:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \qquad \ddots \qquad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases}$$

is generally written in matrix form:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right)$$

The "augmented matrix" of the system

(arrays of numbers, forget variables)

Definition: • A matrix is a (rectangular) array of numbers (scalars).

• A matrix with n rows and m columns is said to be a "n by m matrix".

Notation: • Matrices are generally denoted by capital letters (e.g. A, B, C, M, N; etc.)

• The set of $n \times m$ matrices is often denoted by $M_{n,m}(\mathbb{R})$, or $M_{n,m}$, or $\mathbb{R}^{n\times m}$

Example:
$$\left.\begin{array}{r} 2x+3y+z=5 \\ 3x+4y+z=2 \\ x \qquad +z=4 \end{array}\right\} \overset{\text{identified with}}{\longleftrightarrow} \left(\begin{array}{ccc|c} 2 & 3 & 1 & 5 \\ 3 & 4 & 1 & 2 \\ 1 & 0 & 1 & 4 \end{array}\right) \in M_{3,4}$$

Transforming linear systems:

We begin with an example to illustrate the technique we are going to use.

Example: Solve the system

$$\begin{cases} x - y - z = 2 \\ 3x - 3y + 2z = 16 \\ 2x - y + z = 9 \end{cases}$$

We want to transform the system to a triangular system, without changing the solutions.

Matrix Form | **"Equations form"**

$$\begin{aligned} x - y - z &= 2 \\ 3x - 3y + 2z &= 16 \\ 2x - y + z &= 9 \end{aligned} \qquad \left(\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{array}\right)$$

$\xrightarrow{R_2 - 3R_1}$

$$\begin{aligned} x - y - z &= 2 \\ 5z &= 10 \\ 2x - y + z &= 9 \end{aligned} \qquad \xrightarrow{R_2 - 3R_1} \left(\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 2 & -1 & 1 & 9 \end{array}\right)$$

$\xrightarrow{R_3 - 2R_1}$

$$\begin{aligned} x - y - z &= 2 \\ 5z &= 10 \\ y + 3z &= 5 \end{aligned} \qquad \xrightarrow{R_3 - 2R_1} \left(\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 0 & 1 & 3 & 5 \end{array}\right)$$

$\xrightarrow{R_3 \leftrightarrow R_2}$

$$\begin{aligned} x - y - z &= 2 \\ y + 3z &= 5 \\ 5z &= 10 \end{aligned} \qquad \xrightarrow{R_3 \leftrightarrow R_2} \left(\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 5 & 10 \end{array}\right)$$

The system can now be easily solved by back substitution.
(Ex.)

We now formalize the reasoning from the previous example.

Direct methods for solving linear systems:

Definition: A matrix is in <u>row echelon form</u> if it satisfies the following properties:

1. Any rows consisting entirely of zeros are at the bottom.
2. In each nonzero row, the first nonzero entry (called the leading entry) is in a column to the left of any leading entries below it.

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 5 & 6 & 7 \\ 0 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

are in echelon form.

The matrices $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 0 & 7 & 0 \end{pmatrix}$ are NOT in echelon form

In solving a linear system, it will not always be possible to achieve a triangular form as in the previous examples. However, as we will see, we can always achieve an echelon form.

To reduce a matrix to an echelon form, we use row operations:

	Notation
① Interchange to rows	$R_i \leftrightarrow R_j$
② Multiply a row by a nonzero constant	$R_i \to c \cdot R_i \quad (c \in \mathbb{R} \setminus \{0\})$
③ Add a multiple of a row to another row	$R_i \to R_i + cR_j$

Theorem: Applying elementary row operations to a linear system does not change the solutions of the system. Moreover, every matrix can be reduced to an echelon form by performing finitely many row operations.

The process of reducing a matrix to an echelon form is called Gaussian elimination.

(Carl Friedrich Gauss, 1777-1855).

Gaussian elimination:

① Write the augmented matrix corresponding to the linear system of equations.

② Reduce the matrix to echelon form using elementary row operations.

③ Use back substitution to solve the resulting equivalent system.

Example: Solve:

$$\begin{cases} 2x_2 + 3x_3 = 8 \\ 2x_1 + 3x_2 + x_3 = 5 \\ x_1 - x_2 - 2x_3 = -5 \end{cases}$$

We use Gaussian elimination. The "augmented matrix" of the system is:

$$\left(\begin{array}{ccc|c} 0 & 2 & 3 & 8 \\ 2 & 3 & 1 & 5 \\ 1 & -1 & -2 & -5 \end{array}\right)$$

$$\xrightarrow{R_1 \leftrightarrow R_3} \left(\begin{array}{ccc|c} 1 & -1 & -2 & -5 \\ 2 & 3 & 1 & 5 \\ 0 & 2 & 3 & 8 \end{array}\right)$$

[Use first row to "remove" the other nonzero entries in the first column]

remove the "2" in first column

$$\xrightarrow{R_2 - 2R_1} \left(\begin{array}{ccc|c} 1 & -1 & -2 & -5 \\ 0 & 5 & 5 & 15 \\ 0 & 2 & 3 & 8 \end{array}\right)$$

[Now, use the (2,2) entry to remove the "2" in the third line]

$$\xrightarrow{R_2 \to \frac{1}{5}R_2} \left(\begin{array}{ccc|c} 1 & -1 & -2 & -5 \\ 0 & 1 & 1 & 3 \\ 0 & 2 & 3 & 8 \end{array}\right)$$

(This step was not really necessary, but makes our life easier. We could directly perform $R_3 \to R_3 - \frac{2}{5}R_2$)

$$\xrightarrow{R_3 \to R_3 - 2R_2} \left(\begin{array}{ccc|c} 1 & -1 & -2 & -5 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

Matrix is now in echelon form!

Now, we work backwards to compute the solution of the system:

- R_3 directly gives: $\boxed{x_3 = 2}$.
- Substituting in R_2, we obtain: $x_2 + x_3 = 3$

$$\Leftrightarrow x_2 + 2 = 3$$

$$\Rightarrow \boxed{x_2 = 1}$$

- Substituting in R_1, we obtain: $x_1 - x_2 - 2x_3 = -5$

$$\Leftrightarrow x_1 = x_2 + 2x_3 - 5$$

$$= 1 + 2 \cdot 2 - 5$$

$$= 0.$$

So $\boxed{x_1 = 0}$

The system therefore has a unique solution:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}.$$

In the previous example, the system had a unique solution.

this is NOT always the case.

The following example illustrates what happens when a system has multiple (infinitely many) solutions.

<u>Example</u>: Solve:

$$\begin{cases} w - x - y + 2z = 1 \\ 2w - 2x - y + 3z = 3 \\ -w + x - y = -3 \end{cases}$$

The associated augmented matrix is:

$$\left(\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 2 & -2 & -1 & 3 & 3 \\ -1 & 1 & -1 & 0 & -3 \end{array}\right)$$

$$\xrightarrow{R_2 - 2R_1} \left(\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ -1 & 1 & -1 & 0 & -3 \end{array}\right) \xrightarrow{R_3 \to R_3 + R_1} \left(\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & -2 & 2 & -2 \end{array}\right)$$

WE ARE NOT DONE YET. NOT ECHELON FORM!

$$\xrightarrow{R_3 \to R_3 + 2R_2} \left(\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Echelon form ✓.

Now, observe that the coefficients of all the variables are zero.... What does that mean?

The original system is equivalent to:

$$\begin{cases} w - x - y + 2z = 1 \\ \qquad\quad y - z = 1 \end{cases}$$

This system has infinitely many solution. Why? How do we compute them?

Use back substitution.

Using R_2: $\boxed{y = 1 + z}$ $(*)$

Substitute in R_1: $w - x - (1+z) + 2z = 1$

$$\Leftrightarrow w - x - 1 - z + 2z = 1$$

$$\Leftrightarrow \boxed{w = 2 + x - z} \qquad (**)$$

Now, for ANY choice of x and z, $(*)$ and $(**)$ provide a solution to the system!

If we assign parameters $x = s$, $z = t$ to x and z, we can write the solution as:

$$\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2+s-t \\ s \\ 1+t \\ t \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

Remark: In the previous example: suppose the echelon form had been:
$$\left(\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{array}\right).$$

What would that mean?

Last row: $0 = 5$ IMPOSSIBLE. Therefore, the system would have NO SOLUTIONS!

Definition: The rank of a matrix is the number of nonzero rows in its echelon form.

Notation: rank of A $=: \operatorname{rank}(A)$

In the previous example, the echelon form of the system has 1 row of zeros. Thus the rank of the corresponding matrix is 2. Note that we needed 2 "free variables" s, t to write down the general solution of the system. This is not a coincidence:

Theorem: (The rank theorem):

Let A be the matrix of a linear system with n variables. If the system is consistent (i.e. has at least one solution), then

$$\underset{\text{needed to express the solution space}}{\text{number of free variables}} = n - \operatorname{rank}(A).$$

So far: to solve a linear system:

① Reduce the augmented matrix to echelon form.

② Back substitution.

The process is called <u>Gaussian elimination</u>.

We now introduce a variant of Gaussian elimination that simplifies the back substitution phase

<u>Gauss-Jordan elimination</u>:

<u>Definition</u>: A matrix is in <u>reduced row echelon form</u> if:

① It is in row echelon form;

② The leading entry in each nonzero row is 1

③ Each column containing a leading 1 has zeros everywhere else.

<u>Example</u>: $\begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}$ is in reduced row echelon form.

$\begin{pmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}$ is in echelon form, but NOT in reduced echelon form (How would you fix that using elementary row operations?)

Gauss-Jordan elimination:

1. Write the augmented matrix of the linear system.
2. Use elementary row operations to reduce the matrix to reduced row echelon form.
3. If the system has at least a solution, solve it directly (easy)

How to reduce to reduced row echelon form using row operations? The idea is similar to what we did before, except that we scale the leading entry in each row, and "use that leading entry to put zeros in the whole column." We illustrate that process in the next example.

Example: Solve the system:

$$\begin{cases} 2x_1 - 2x_2 + 4x_3 = 6 \\ x_1 + 2x_2 - x_3 = -3 \\ 2x_2 - x_3 = 1 \end{cases}$$

Augmented matrix: $\left(\begin{array}{ccc|c} 2 & -2 & 4 & 6 \\ 1 & 2 & -1 & -3 \\ 0 & 2 & -1 & 1 \end{array}\right)$

$R_1 \to \frac{1}{2}R_1$ $\longrightarrow$ $\left(\begin{array}{ccc|c} 1 & -1 & 2 & 3 \\ 1 & 2 & -1 & -3 \\ 0 & 2 & -1 & 1 \end{array}\right)$ We scale the (1,1) entry.

$R_2 \to R_2 - R_1$ $\longrightarrow$ $\left(\begin{array}{ccc|c} 1 & -1 & 2 & 3 \\ 0 & 3 & -3 & -6 \\ 0 & 2 & -1 & 1 \end{array}\right)$ We use the (1,1) entry to put zeros below it in column 1.

$\boxed{\text{DONE WITH COLUMN 1. WE NOW WORK ON COLUMN 2.}}$

$R_2 \to \frac{1}{3}R_2$ $\longrightarrow$ $\left(\begin{array}{ccc|c} 1 & -1 & 2 & 3 \\ 0 & 1 & -1 & -2 \\ 0 & 2 & -1 & 1 \end{array}\right)$ We scale the leading entry in row 2.

$R_1 \to R_1 + R_2$ $\longrightarrow$ $\left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 2 & -1 & 1 \end{array}\right)$ Use the (2,2) entry to "put zero" in (1,2) entry.

$R_3 \to R_3 - 2R_2$ $\longrightarrow$ $\left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & 5 \end{array}\right)$ Use the (2,2) entry to put a zero in (3,2) entry.

$\boxed{\text{DONE WITH COLUMN 2. MOVE TO COLUMN 3}}$

(3,3) entry is scaled ✓. Use it to put zeros in (2,3) and (1,3).

$R_2 \to R_2 + R_3$ $\longrightarrow$ $\left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{array}\right)$ Put zero in (2,3)

$R_1 \to R_1 - R_3$
$$\longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & -4 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{array}\right)$$

Put zero in $(1,3)$.

Now, read the solution !! $\begin{cases} x_1 = -4 \\ x_2 = 3 \\ x_3 = 5 \end{cases}$

Example: Find the intersection of the planes:

$$\begin{cases} x + 2y - z = 3 \\ 2x + 3y + z = 1. \end{cases}$$

Let us use Gauss - Jordan:

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 2 & 3 & 1 & 1 \end{array}\right) \xrightarrow{R_2 - 2R_1} \left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & -1 & 3 & -5 \end{array}\right)$$

$R_2 \to -R_2$
$$\longrightarrow \left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 1 & -3 & 5 \end{array}\right) \xrightarrow{R_1 - 2R_2} \left(\begin{array}{ccc|c} 1 & 0 & 5 & -7 \\ 0 & 1 & -3 & 5 \end{array}\right)$$

The system is thus equivalent to:

$$\begin{cases} x + \quad 5z = -7 \\ \quad y - 3z = 5 \end{cases} \qquad (z \text{ free}).$$

Let us set $z := t$.

The general solution of the system can be written:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -7 \\ 5 \\ 0 \end{pmatrix} + t \begin{pmatrix} -5 \\ 3 \\ 1 \end{pmatrix}.$$

(See figure 2.2 in Book).

<u>Remark</u>: # of free variables $= 1 = \underbrace{3}_{\#\text{ of variables}} - \underbrace{2}_{\text{rank}(A)}$.

(Consistent with rank theorem).

<u>Definition</u>: A linear system is said to be <u>homogeneous</u> if the constant term in each equation is zero.

<u>Example</u>: $\begin{cases} 2x + 3y - z = 0 \\ -x + 5y + 2z = 0 \end{cases}$

Note that $\mathbf{0} = 0_{n\times 1}$ is always a solution of a homogeneous system. Thus, a homogeneous system is <u>always</u> consistent.

<u>Theorem</u>: A homogeneous system with m equations and n variables with $m < n$ always has infinitely many solutions.

<u>Proof</u>: Clearly, the matrix of the system has rank at most m. By the rank theorem.

$$\#\text{ of free variables} = n - \text{rank}(A) \geq n - m > 0,$$

Spanning sets and linear independence

We saw before that the solution of a linear system can be written as a linear combination of vectors:

$$\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \underbrace{\begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix}}_{\text{translation}} + \underbrace{s\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \\ 1 \end{pmatrix}}_{\text{hyperplane in } \mathbb{R}^4.} \qquad (s, t \in \mathbb{R})$$

Note that

$$\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \\ 1 \end{pmatrix} + u\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

describes the same set of solutions since

$$\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

i.e. it is a <u>linear combination</u> of the other two vectors.

The vector $u \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ is thus redondant.

A fundamental problem in linear algebra is to determine if a vector is a linear combination of other vectors and, if that is the case, compute the coefficients.

This problem is closely connected to solving linear systems.

<u>Example</u>: Is $v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ a linear combination of $v_1 = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$ and $v_2 = \begin{pmatrix} -1 \\ 1 \\ -3 \end{pmatrix}$?

<u>Sol^n</u>: Equivalent to solving

$$v = x v_1 + y v_2$$ (if there is a solution at all)

i.e.,

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} + y \begin{pmatrix} -1 \\ 1 \\ -3 \end{pmatrix}$$

$\Leftrightarrow$ solving the system $\left(\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 3 & -3 & 3 \end{array}\right)$.

We use Gauss-Jordan:

$$\left(\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 3 & -3 & 3 \end{array}\right) \xrightarrow{R_3 - 3R_1} \left(\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right) \xrightarrow{R_1 \to R_1 + R_2} \left(\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right)$$

i.e., $x = 3$, $y = 2$ <u>unique</u> solution.

Thus, $v = 3v_1 + 2v_2$ IS a linear combination of v_1, v_2

Important remark:

A linear system $[A|b]$ with $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$

has a solution iff b is a linear combination of the columns of A. Indeed, if the system is

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ c_1 & c_2 & \cdots & c_n & b \\ | & | & & | & | \end{array}\right), \qquad (c_1, \dots, c_n \text{ are vectors.})$$

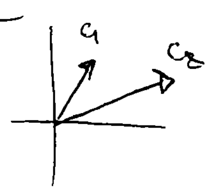
then $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is a solution iff $b = x_1 c_1 + x_2 c_2 + \dots + x_n c_n$.

Remark: Consider a 2×2 system

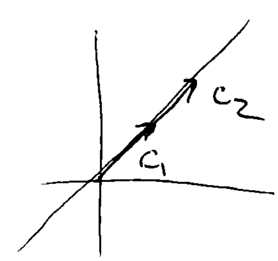
$$\begin{cases} a_{11} x + a_{12} y = b_1 \\ a_{21} x + a_{22} y = b_2 \end{cases}$$

i.e. $\left(\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{array}\right)$.

The two columns $c_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$ and $c_2 = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$ either "generate" the whole plane



or a line



- In the first case, any vector b is a linear combination of e_1 and e_2. (So the system always has a unique solution!)
- In the second case, the system has a solution only if b belongs to the line. (So no solution, or infinitely many solutions)

<u>Definition</u>: (Spanning set) Let $S = \{v_1, \dots, v_K\}$ be a set of vectors in $\mathbb{R}^n$. The set of all linear combinations of $v_1, \dots, v_K$ is called the <u>span</u> of $v_1, \dots, v_K$:

$$\begin{aligned} \operatorname{span}(S) &= \operatorname{span}(v_1, \dots, v_K) \\ &= \{\lambda_1 v_1 + \dots + \lambda_K v_K : \lambda_1, \lambda_2, \dots, \lambda_K \in \mathbb{R}\} \end{aligned}$$

If $\operatorname{span}(S) = \mathbb{R}^n$, S is said to be a <u>spanning set</u> for $\mathbb{R}^n$.

<u>Examples</u>:

① Clearly $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is a spanning set for $\mathbb{R}^2$.

② $\mathbb{R}^2 = \operatorname{span}\left(\begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix} \right)$.

Proof: We need to show that an arbitrary vector $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$ is a linear combination of $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$.

This is equivalent to solving the system:

$$\left(\begin{array}{cc|c} 2 & 1 & a \\ -1 & 3 & b \end{array} \right). \quad \text{for a given pair } a, b \in \mathbb{R}.$$

We use Gauss-Jordan:

$$\left(\begin{array}{cc|c} 2 & 1 & a \\ -1 & 3 & b \end{array} \right) \xrightarrow{R_1 \to \frac{1}{2} R_1} \left(\begin{array}{cc|c} 1 & 1/2 & a/2 \\ -1 & 3 & b \end{array} \right)$$

$$\xrightarrow{R_2 \to R_2 + R_1} \left(\begin{array}{cc|c} 1 & 1/2 & a/2 \\ 0 & 7/2 & a/2 + b \end{array} \right)$$

$$\xrightarrow{R_2 \to R_2 \cdot \frac{2}{7}} \left(\begin{array}{cc|c} 1 & 1/2 & a/2 \\ 0 & 1 & a/7 + \frac{2b}{7} \end{array} \right)$$

$$\xrightarrow{R_1 \to R_1 - 1/2 R_2} \left(\begin{array}{cc|c} 1 & 0 & a/2 - \frac{1}{2}(a/7 + \frac{2b}{7}) \\ 0 & 1 & a/7 + 2b/7 \end{array} \right) = \left(\begin{array}{cc|c} 1 & 0 & (3a-b) \cdot \frac{1}{7} \\ 0 & 1 & (a+2b) \cdot \frac{1}{7} \end{array} \right)$$

UNIQUE SOLUTION!

Therefore, the two vectors span $\mathbb{R}^2$.

Linear independence: $(\lambda, \mu \in \mathbb{R})$

Clearly, $\mathrm{span}(v, w, \lambda v + \mu w) = \mathrm{span}(v, w)$ ...

How do we determine if a spanning set is "minimal"?

<u>Definition:</u> (Linear independence) A set of vectors $v_1, \dots, v_k$ is <u>linearly dependent</u> if there are scalars $\lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{R}$ at least one of which is not zero, s.t.

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k = 0$$

the zero vector. $\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

A set of vectors that is not linearly dependent is said to be <u>linearly independent</u>.

<u>Remark:</u> Another way of defining linear independence is to say that if

$$\lambda_1 v_1 + \dots + \lambda_k v_k = 0,$$

then $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$.

Theorem: A set of vectors $v_1, \dots, v_K \in \mathbb{R}^n$ is linearly dependent iff at least one of the vectors can be expressed as a linear combination of the other vectors.

Proof:

($\Leftarrow$) Suppose one vector is a linear comb. of the others, say:

$$v_1 = \lambda_2 v_2 + \lambda_3 v_3 + \dots + \lambda_K v_K.$$

Then $v_1 - \lambda_2 v_2 - \lambda_3 v_3 - \dots - \lambda_K v_K = 0$

and not all coefficients are 0 (since the coefficient of v_1 is 1).

thus $v_1, \dots, v_K$ are linearly dependent.

($\Rightarrow$) Conversly, suppose $v_1, \dots, v_K$ are lin. dep. Then there exists $\lambda_1, \dots, \lambda_K \in \mathbb{R}$ (not all zero) s.t. $\lambda_1 v_1 + \dots + \lambda_K v_K$.

WLOG, assume $\lambda_1 \neq 0$. Then

$$v_1 = -\left(\frac{\lambda_2}{\lambda_1}\right) v_2 - \left(\frac{\lambda_3}{\lambda_1}\right) v_3 - \dots - \left(\frac{\lambda_n}{\lambda_1}\right) v_n$$ and

so v_1 is a lin. comb. of the other vectors. $\boxtimes$

Examples:

① Any set of vectors containing $0 = 0_{n\times 1}$ is linearly dependent. (Trivial). (Check!)

② Let $v_1 := \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $v_2 := \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, $v_3 := \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

Then $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0_{3\times 1}$ if and only if

$$\begin{cases} \lambda_1 + \lambda_3 = 0 \\ \lambda_1 + \lambda_2 = 0 \\ \lambda_2 + \lambda_3 = 0 \end{cases}$$

Let's solve this system:

NOTE: The columns of the matrix are v_1, v_2, v_3 ...

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right) \xrightarrow{R_2 \to R_2 - R_1} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right)$$

$$\xrightarrow{R_3 \to R_3 - R_2} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 2 & 0 \end{array}\right) \quad \text{(Echelon form)}$$

$$\xrightarrow{R_3 \to \frac{1}{2}R_3} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \xrightarrow[R_2 \to R_2 + R_3]{R_1 \to R_1 - R_3} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

Thus, $\lambda_1 = \lambda_2 = \lambda_3 = 0$ and so v_1, v_2, v_3 are linearly independent!

The following theorem summarizes the procedure used in the previous example to test for linear independence.

Theorem: Let $v_1, \ldots, v_m \in \mathbb{R}^{n \times 1}$ be (column) vectors and let A be the $n \times m$ matrix having these vectors as its columns, i.e.,

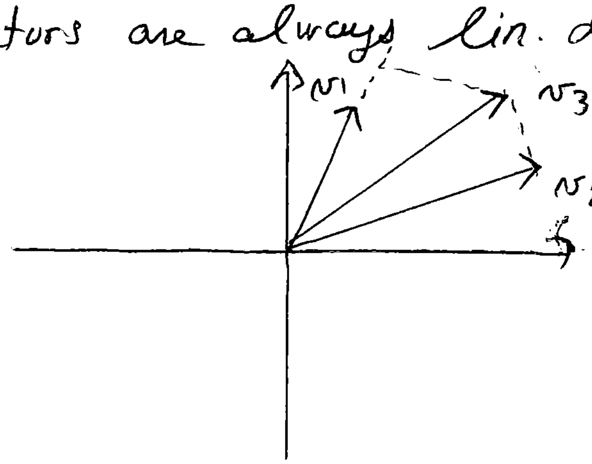
$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix}.$$

then $v_1, \ldots, v_m$ are lin. dependent iff the homogeneous system $[A \mid 0_{n\times 1}]$ has a nontrivial (i.e., $\neq 0_{n\times 1}$) solution.

Equivalently, the vectors $v_1, \ldots, v_m$ are lin. ind. iff $0_{n\times 1}$ is the only solution of the system $[A \mid 0_{n\times 1}]$.

Proof: (Exercise).

In $\mathbb{R}^2$, 3 vectors are always lin. dependent:



$$v_3 = \lambda_1 v_1 + \lambda_2 v_2$$

for some λ_1, λ_2 not all zero ...

Similarly, in $\mathbb{R}^3$, 4 vectors are always lin. dependent.

The following theorem provides a formal proof of this idea.

<u>Theorem</u>: Any set of m vectors in $\mathbb{R}^n$ is lin. dependent if $m > n$.

<u>Proof</u>: The vectors are lin. dep. iff the system

$$[A \mid 0_{n\times 1}]$$ has a nontrivial solution,

where $A = [v_1, v_2, \ldots, v_m]$.

Recall that a homogeneous system with less rows than columns has infinitely many solutions (we proved this result before using the rank theorem). Thus $[A \mid 0_{n\times1}]$ has a nontrivial solution and so the vectors are lin. dep.

$\boxed{M}$

Chapter 3: Matrices.

So far, we have used matrices to conveniently record the coefficients of linear systems. In this chapter, you will see that matrices are much deeper algebraic objects. Matrices naturally "act" or "operate" on vectors. They can be seen as functions that map each vector to another vector in a "linear" way. These "linear functions" can be composed, which leads to a natural way to multiply matrices.

We will explore these concepts in the next few lectures

Recall that a <u>matrix</u> is a rectangular array of numbers (called the entries of the matrix).

- The <u>size</u> (or dimension) of a matrix is a description of its number of rows and columns. A $m \times n$ (read "m by n") matrix has m rows and n columns.

Example:

① $\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$ is a 2×2 matrix. When $m=n$, the matrix is said to be a square matrix.

② We often denote the (i,j) th entry of a matrix A by a_{ij}. Thus

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

represents a general $m\times n$ matrix.

Notation: Sometimes, we write $A=(a_{ij})$ or $A=(a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ to make clear that a_{ij} is the entry of A in the i^{th} row and j^{th} column.

③ The entries $a_{11}, a_{22}, \ldots$ of a matrix A are called the diagonal entries of A. For example:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

$1, 5, 9$ are the diagonal entries of A.

④ The identity matrix $I = I_n = I_{n\times n}$ is the $n\times n$ matrix with 1s on the diagonal, and 0s elsewhere.

Ex: $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

⑤ If $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ is a $m \times n$ matrix, its transpose, denoted A^T (read "A transpose") is the $n \times m$ matrix

$$A^T := (a_{ji}).$$

Note: The rows of A^T are the columns of A.

Ex: $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, $A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

⑥ Two matrices A and B are equal (denoted $A = B$) if they have the same size, and the same entries.

⑦ A matrix A is symmetric if $A = A^T$.

Ex: $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}$ is symmetric.

Note: A matrix is symmetric $\Leftrightarrow$ its entries are symmetric with respect to its diagonal.

⑧ A $1 \times n$ matrix $A = (a_1, a_2, \dots, a_n)$ is called a row vector.

A $m \times 1$ matrix $A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$ is called a column vector.

We now define <u>algebraic operations</u> on matrices.

<u>Definition</u>: (Addition and scalar multiplication)

Let $A=(a_{ij}), B=(b_{ij}) \in \mathbb{R}^{m\times n}$ be two matrices of the same size, and let $\lambda \in \mathbb{R}$. We define:

$$A+B := (a_{ij}+b_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}} \qquad (\text{SUM})$$

$$\lambda A := (\lambda a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}} \qquad (\text{SCALAR MULTIPLICATION})$$

<u>Note</u>: Like vectors, matrices are added "entrywise". Same for multiplication by a scalar.

We now define a notion of multiplication for matrices.

<u>IMPORTANT</u>: The usual notion of multiplication of matrices is <u>NOT</u> entrywise multiplication.

<u>Definition</u>: (Matrix product)

Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times r}$. The product $C=AB$ is the $m\times r$ matrix with entries:

$$c_{ij} := a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

Remarks:

① The matrices A and B need not be the same size.

$$\begin{array}{ccccc} A & & B & = & AB \\ m \times \boxed{n} & & \boxed{n} \times r & & m \times r \end{array}$$

HAVE TO BE THE SAME

Note: AB is not defined if the # of columns of A is different from the # of rows of rows of B

② The (i,j) entry of AB is the inner product (dot product) of the i^{th} row of A with the j^{th} column of B:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & & \ddots & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1r} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2r} \\ \vdots & & & & \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{nr} \end{pmatrix}$$

Example: Any linear system:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

can be written as:

$$\boxed{Ax = b},$$

where $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$, $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$

$m \times n$ matrix, $n \times 1$ matrix (column vector), $n \times 1$ matrix (column vector)

Remark: ① If $c_1, \ldots, c_n$ are the columns of A, i.e., $A = \begin{pmatrix} | & | & & | \\ c_1 & c_2 & \cdots & c_n \\ | & | & & | \end{pmatrix}$,

then $Ax = x_1 c_1 + x_2 c_2 + \cdots + x_n c_n$ (linear combination of the columns)

② Note that $I_n x = x \quad \forall x \in \mathbb{R}_n$.

③ If we can find a matrix B such that $BA = I_n$, then mutiplying by B on both sides of $Ax = b$, we obtain:

$$BAx = Bb,$$

and so $$I_n x = Bb$$

$$\Leftrightarrow \boxed{x = Bb}$$ which provides the solution of the linear system $Ax = b$.

When such a matrix exists, we call it the <u>inverse</u> of A, and denote it by A^{-1}. We will learn how to compute inverses later.

(4) If A, B are square matrices, say $A, B \in \mathbb{R}^{n\times n}$, then AB and BA are defined, and are both $n\times n$ matrices (check!).

In general, $AB \neq BA$. Verify that statement by computing the product of a few 2×2 matrices.

For that reason, the matrix product is said to be <u>non-commutative</u>

<u>Example:</u> (Example 3.7 in book)

Ann and Bert are shopping for fruits.

	Apples	Grapefruit	Oranges
Ann	6	3	10
Bert	4	8	5

Amount of fruits they buy.

	Store #1	Store #2
Apple	\$0.10	\$0.15
Grapefruit	\$0.40	\$0.30
Orange	\$0.10	\$0.20

Price of fruits at two different stores.

Let $D = \begin{pmatrix} 6 & 3 & 10 \\ 4 & 8 & 5 \end{pmatrix}$, $\quad P = \begin{pmatrix} 0.10 & 0.15 \\ 0.40 & 0.30 \\ 0.10 & 0.20 \end{pmatrix}$

"Demand" matrix $\qquad$ "Price matrix".

Then $DP = \begin{pmatrix} 2.80 & 3.80 \\ 4.10 & 4.00 \end{pmatrix}$ $\quad$ (i, j) entry = amount of money spent by Ann $(i = 1)$ and Bert $(i=2)$ at store j.

<u>Matrix algebra:</u>

Exept for the non-commutativity of the product ($AB \neq BA$ in general), addition and multiplication of matrices have properties that are similar to real numbers:

<u>Theorem:</u> (Algebraic properties of matrix addition and scalar multiplication)

Let $A, B, C \in \mathbb{R}^{m \times n}$, and $\lambda, \mu \in \mathbb{R}$. Then:

1. $A + B = B + A$ (Commutativity of addition)
2. $A + (B + C) = (A + B) + C$ (Associativity of addition)
3. $A + O_{m \times n} = A$
4. $A + (-A) = O_{m \times n}$
5. $\lambda(A + B) = \lambda A + \lambda B$
6. $(\lambda + \mu) A = \lambda A + \mu A$
7. $\lambda(\mu A) = (\lambda \mu) A$
8. $1 A = A$ (here 1 is the real number $1 \in \mathbb{R}$)

The proof of these properties is trivial.

Theorem: (Properties of matrix multiplication).

Let A, B, C be matrices (whose size are such that the indicated operations can be performed) and let $\lambda \in \mathbb{R}$.

Then:

① $A(BC) = (AB)C$

② $A(B+C) = AB + AC$

③ $(A+B)C = AC + BC$

④ $\lambda(AB) = (\lambda A)B = A(\lambda B)$

⑤ $I_m A = A = A I_n$ if $A \in \mathbb{R}^{m \times n}$.

identity matrices of dimensions m and n.

Matrix powers:

Let k be an integer and $A \in \mathbb{R}^{n \times n}$. We define

$$A^1 := A \quad \text{and} \quad A^k := \underbrace{A \cdot A \cdot \ldots \cdot A}_{k \text{ times}}.$$

__Example:__

Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Then

$$A^2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}.$$

$$A^3 = A \cdot A^2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}.$$

In general, $A^n = \begin{pmatrix} 2^{n-1} & 2^{n-1} \\ 2^{n-1} & 2^{n-1} \end{pmatrix}$ $(n \geq 1)$.

(Can be proved using mathematical induction)

__Example:__ Let A, B be square matrices of the same size.

Is it true that $(A+B)^2 = A^2 + 2AB + B^2$?

__Ans:__ We compute: $(A+B)^2 = (A+B)(A+B)$

$$= A^2 + AB + BA + B^2$$

Thus $(A+B)^2 = A^2 + 2AB + B^2 \iff 2AB = AB + BA$

$\iff AB = BA$

$\iff A$ & B commute.

The identity is therefore not true in general!

Theorem (properties of the transpose)

Let A, B be matrices (whose sizes are such that the indicated operations can be performed) and let $\lambda \in \mathbb{R}$. Then:

① $(A^T)^T = A$

② $(A+B)^T = A^T + B^T$

③ $(\lambda A)^T = \lambda A^T$

④ $(AB)^T = B^T A^T$

⑤ $(A^k)^T = (A^T)^k$ for all integers $k = 1, 2, 3, \ldots$

Proof: ①, ②, ③ are clear. Let us prove ④. Suppose $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times r}$, we have:

$$[(AB)^T]_{ij} = (AB)_{ji} = \sum_{l=1}^{n} a_{jl} b_{li}$$

On the other hand,

$$(B^T A^T)_{ij} = \sum_{l=1}^{n} (B^T)_{il} (A^T)_{lj}$$

$$= \sum_{l=1}^{n} b_{li} a_{jl} = \sum_{l=1}^{n} a_{jl} b_{li}.$$

Since the two quantities are equal, we have $(AB)^T = B^T A^T$.

Finally, we use ④ to prove ⑤.

$$(A^k)^T = (A \cdot A^{k-1})^T = (A^{k-1})^T A^T$$

$$= (A A^{k-2})^T \cdot A^T$$

$$= (A^{k-2})^T A^T A^T = (A^{k-2})^T (A^T)^2$$

$$= \cdots$$

$$= (A^T)^k$$

(Use mathematical induction for a rigorous proof).

The inverse of a matrix:

Suppose $a, b \in \mathbb{R}$. To solve the equation:

$$ax = b,$$

we can simply divide by a:

$$x = b/a.$$

Equivalently, we multiply by $a^{-1} = \frac{1}{a}$ on both side of the equation:

$$a^{-1} a x = a^{-1} b$$

$$1 x = a^{-1} b$$

$$\boxed{x = a^{-1} b}.$$

We know that we can write a linear system using the matrix product:

$$Ax = b,$$

where $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^n$.

Suppose there exists $B \in \mathbb{R}^{n \times n}$ s.t. $BA = I_n$. Then we can solve the linear system as we would do with real numbers: $BAx = Bb \Rightarrow Ix = \boxed{x = Bb}$.

Definition: If $A \in \mathbb{R}^{n \times n}$, an <u>inverse</u> of A is an $n \times n$ matrix B with the property that:

$$\begin{cases} AB = I_n \\ BA = I_n \end{cases}$$

If A has an inverse, we say that A is invertible.

Example: Let $A = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}$ and let $B = \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix}$

Then $AB = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$BA = \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Therefore B is an inverse of A!

<u>Note</u>: In particular, we can solve $Ax = b$ for ANY b using the inverse of A!

Example: A matrix may not have an inverse:

Consider the matrix $B = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$. Suppose there exists $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ s.t. $BA = I_2$.

Then $BA = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+2c & b+2d \\ 2a+4c & 2b+4d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Thus, a, b, c, d satisfy the linear system:

$$\begin{cases} a + 2c = 1 & (1) \\ b + 2d = 0 & (2) \\ 2a + 4c = 0 & (3) \\ 2b + 4d = 1. & (4) \end{cases}$$

This system clearly has NO solutions ((1),(3) and (2),(4) contradict themselves)

Therefore, A doesn't have an inverse.

An inverse may not exist. However, if it exists, it is UNIQUE.

Theorem: If $A \in \mathbb{R}^{n\times n}$ is invertible, then its inverse is unique.

Proof: Suppose $BA = AB = I_n$, $CA = AC = I_n$, i.e., both B and C are inverses of A.

We will prove that $B = C$.

Indeed, $B = B(A\overset{I_n}{\overbrace{C}}) = (BA)C = I_n \cdot C = C$.

Therefore, the inverse is unique if it exists. ∎

<u>Notation</u>: If $A \in \mathbb{R}^{n \times n}$ is invertible, we denote its (unique) inverse by A^{-1}.

<u>Remark</u>: We NEVER denote the inverse of A by $\frac{1}{A}$.

(For example, how would we interpret $\frac{BC}{A}$? $A^{-1}BC$? BCA^{-1}? Remember, the matrix product is non-commutative!)

<u>Notation</u>: The set of invertible $n \times n$ matrices with real entries is often denoted by $GL_n(\mathbb{R})$.

"GL" stands for "general linear". $GL_n(\mathbb{R})$ is called the "general linear group".

<u>Theorem</u>: If $A \in \mathbb{R}^{n \times n}$ is invertible, then the system of linear equations $Ax = b$ has a UNIQUE solution $x = A^{-1}b$ for any $b \in \mathbb{R}^n$.

Remark: Once you know a matrix A is invertible, you can:

1. Guarantee that every system $Ax = b$ has a unique solution.

2. Compute the solution for multiple vectors $b \in \mathbb{R}^n$ easily (after the inverse has been computed).

Remark: A matrix $A \in \mathbb{R}^{n \times n}$ is invertible iff the span of its columns is $\mathbb{R}^n$.

Proof of the theorem:

Suppose A is invertible and $b \in \mathbb{R}^n$. Let $x = A^{-1}b$. Then:

$$Ax = A(A^{-1}b) = I_n b = b.$$

Therefore, $x = A^{-1}b$ is a solution of the system.

Suppose the solution is not unique, i.e., there exists $y \in \mathbb{R}^n$ s.t. $Ay = b$.

Then, multiplying by A^{-1} on both sides of the equation:

$$A^{-1}Ay = A^{-1}b \Rightarrow y = A^{-1}b = x.$$

Thus $y = x$ and so the solution is unique. $\square$

Naturally, given a matrix A, we would like to know if it is invertible and, if it is, compute its inverse. This problem has an easy solution for 2×2 matrices.

<u>Theorem</u>: A 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad-bc \neq 0$, in which case:

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

The expression $ad-bc$ is called the <u>determinant</u> of A (and is denoted by $\det A$).

<u>Proof</u>:

Suppose $\det A = ad-bc \neq 0$. Then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} = (\det A)\cdot I_n$$

Similarly, $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} = (\det A)\, I_n$.

Therefore, if $\det A \neq 0$, A is invertible and its inverse is as described.

Now, suppose $\det A = 0$.

we consider two cases: $b \neq 0$ and $b = 0$.

(1) Suppose $b \neq 0$. Then $ad - bc = 0$

$$\Rightarrow c = \frac{ad}{b}.$$

As a consequence, the matrix A can be written as:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ \frac{ad}{b} & d \end{pmatrix} = \begin{pmatrix} \frac{a}{b} \cdot b & b \\ \frac{a}{b} \cdot d & d \end{pmatrix}$$

Note that the first column of A is a multiple of the second column. As a consequence, $Ax = b$ has a solution only if b belongs to the line

$$\left\{ \lambda \begin{pmatrix} b \\ d \end{pmatrix} : \lambda \in \mathbb{R} \right\}.$$

The matrix A is therefore not invertible (if it were, the system $Ax = b$ would have a unique solution for every $b \in \mathbb{R}^n$...)

(2) Now suppose $b = 0$. In that case

$$ad - bc = 0 \Rightarrow ad = 0 \Rightarrow a = 0 \text{ or } b = a$$

If $a=0$, the matrix A reduces to:

$$A = \begin{pmatrix} 0 & 0 \\ c & d \end{pmatrix}.$$

Clearly the two columns of A are linearly dependent and so A is not invertible

If $d=0$ instead,

$$A = \begin{pmatrix} a & 0 \\ d & 0 \end{pmatrix}$$

then the span of the columns is still a line

$$\left\{ \lambda \begin{pmatrix} a \\ d \end{pmatrix} : \lambda \in \mathbb{R} \right\}$$

and so A is NOT invertible.

That concludes the proof of the theorem.

$\square$

Example: Find the inverses of $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 12 & -15 \\ 4 & -5 \end{pmatrix}$ if they exist.

Sol^n: We have $\det A = 1 \cdot 4 - 2 \cdot 3 = -2 \neq 0$. so A is invertible

and $A^{-1} = \frac{1}{-2} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}$.

On the other hand, $\det B = 12\cdot(-5) - 4(-15) = 0$.

Therefore, B is not invertible

<u>Example</u>: Suppose $u_1 \neq u_2$. Then

$$A = \begin{pmatrix} 1 & 1 \\ u_1 & u_2 \end{pmatrix} \text{ is invertible.}$$

Indeed, $\det A = u_2 - u_1 \neq 0$ by construction.

<u>Example</u>: Solve the linear system using a matrix inverse.

$$\begin{cases} x + 2y = 3 \\ 3x + 4y = -2 \end{cases}$$

<u>Sol$^{\underline{n}}$</u>: The system is equivalent to $Ax = b$, where

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

we have $\det A = -2 \neq 0$ so A is invertible and

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix} \quad \text{(computed in previous example).}$$

Therefore, the system has a unique solution:

$$x = A^{-1} b = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \end{pmatrix} = \begin{pmatrix} -8 \\ 11/2 \end{pmatrix}.$$

The following theorem records some of the important properties of invertible matrices.

<u>Theorem</u>: Let $A, B \in \mathbb{R}^{n\times n}$. Then:

① If A is invertible, then A^{-1} is invertible and

$$(A^{-1})^{-1} = A.$$

② If A is invertible and $c \in \mathbb{R}\setminus\{0\}$, then cA is invertible, and

$$(cA)^{-1} = \frac{1}{c}A^{-1}.$$

③ If A and B are invertible, then AB is invertible, and

$$(AB)^{-1} = B^{-1}A^{-1}.$$

④ If A is invertible, then A^T is invertible and

$$(A^T)^{-1} = (A^{-1})^T.$$

⑤ If A is invertible, then A^n is invertible for all nonnegative integers n and:

$$(A^n)^{-1} = (A^{-1})^n.$$

Proof:

① B is an inverse of A^{-1} iff $BA^{-1} = A^{-1}B = I_n$.

Clearly $B = A$ satisfies this property.

② We verify

$$(cA)\left(\frac{1}{c}A^{-1}\right) = AA^{-1} = I_n$$

$$\left(\frac{1}{c}A^{-1}\right)(cA) = A^{-1}A = I_n.$$

Therefore $(cA)^{-1} = \frac{1}{c}A^{-1}$.

③ AB is invertible iff we can find a matrix $C \in \mathbb{R}^{n\times n}$ s.t.

$$C(AB) = (AB)C = I_n.$$

Now: $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB$

$= B^{-1}B$

$= I_n.$

Similarly, $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1}$

$= AA^{-1}$

$= I_n.$

therefore $(AB)^{-1} = B^{-1}A^{-1}$.

④ Let $B = A^{-1}$.

Then $(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise.} \end{cases}$

Now
$$\left[A^T (A^{-1})^T\right]_{ij} = \left[A^T B^T\right]_{ij}$$
$$= \sum_{k=0}^{n} (A^T)_{ik} (B^T)_{kj}$$
$$= \sum_{k=0}^{n} a_{ki} b_{jk}$$
$$= \sum_{k=0}^{n} a_{ik} b_{kj} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise.} \end{cases}$$

Thus $A^T (A^{-1})^T = I_n$.

Similarly, we show that $(A^{-1})^T A^T = I_n$.

⑤ By induction on n. The result clearly holds for $n=1$.

Suppose $(A^k)^{-1} = (A^{-1})^k$ for some $k \geq 1$. Then

$A^{k+1} = A \cdot A^k$ is invertible by ③ and

$$(A^{k+1})^{-1} = (A^k)^{-1} \cdot A^{-1} \overset{\text{induction hypothesis}}{=} (A^{-1})^k \cdot A^{-1} = (A^{-1})^{k+1}.$$

This completes the induction. $\square$

Subspace, dimension and rank:

We have seen before that the solution of a linear system is "generated" by a set of vectors in $\mathbb{R}^n$ (when there are infinitely many solutions). For example:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + t \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + s \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \qquad (t, s \in \mathbb{R})$$

In that particular case, the set of solutions is a plane in $\mathbb{R}^3$. The plane $t\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ is naturally identified with $\mathbb{R}^2$. Note that if v_1, v_2 belong to that plane, then

$$\lambda_1 v_1 + \lambda_2 v_2 \quad \text{also belong to the plane.}$$

We therefore think of $V = \left\{ t\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} : t, s \in \mathbb{R} \right\}$ as a <u>subspace</u> of $\mathbb{R}^3$ (of dimension 2).

Similarly, we see a line in $\mathbb{R}^n$ that goes through the origin as a one dimensional subspace of $\mathbb{R}^n$.

Formally:

<u>Definition</u>: A <u>subspace</u> of $\mathbb{R}^n$ is any collection S of vectors in $\mathbb{R}^n$ such that:
1. $0_{n\times 1} \in S$
2. $u, v \in S \Rightarrow u + v \in S$
3. $u \in S \Rightarrow \lambda u \in S$ for all $\lambda \in \mathbb{R}$.

Remark: Equivalently, $S \subset \mathbb{R}^n$ is a subspace if

① $0_{n\times 1} \in S$

② $u, v \in S \Rightarrow \lambda u + \mu v \in S$ for all $\lambda, \mu \in \mathbb{R}$.

Remark: A subspace always contain the origin. A "translated subspace" such as

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + t\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} : \ t, s \in \mathbb{R} \right\}$$

is called an affine subspace.

(Formally, S is an affine subspace of $\mathbb{R}^n$ if there exists $u \in \mathbb{R}^n$ and a subspace $V \subset \mathbb{R}^n$ s.t.

$$s \in S \iff s = u + v \text{ for some } v \in V.)$$

Examples: ① The "line" $L = \left\{ \lambda \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} : \lambda \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^4$.

Proof: Clearly $0_{4\times 1} = 0 \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \in L$. Now, let $u, v \in L$ and let $\lambda, \mu \in \mathbb{R}$. Since $u, v \in L$, there exist $\lambda_1, \lambda_2 \in \mathbb{R}$ s.t. $u = \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$ and $v = \lambda_2 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$. It follows that

$$\lambda u + \mu v = \lambda \cdot \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} + \mu \lambda_2 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \underbrace{(\lambda \cdot \lambda_1 + \mu \lambda_2)}_{\text{scalar.}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

We conclude that $\lambda u + \mu v \in L$ and so L is a subspace of $\mathbb{R}^4$.

More generally:

<u>Theorem</u>: Let $v_1, \ldots, v_k \in \mathbb{R}^n$. Then $S = \operatorname{span}(v_1, \ldots, v_k)$ is a subspace of $\mathbb{R}^n$.

<u>Proof</u>: Clearly $0 \in S$. Now, let $a, b \in S$. Then there exists $\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_k \in S$ s.t.

$$a = \lambda_1 v_1 + \cdots + \lambda_k v_k$$

$$b = \mu_1 v_1 + \cdots + \mu_k v_k.$$

It follows that

$$\lambda a + \mu b = \lambda(\lambda_1 v_1 + \cdots + \lambda_k v_k) + \mu(\mu_1 v_1 + \cdots + \mu_k v_k)$$

$$= (\lambda \lambda_1 + \mu \mu_1) v_1 + \cdots + (\lambda \lambda_k + \mu \mu_k) v_k \in S.$$

$\boxtimes$

<u>Example</u>: The set of vectors $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ s.t. $\begin{matrix} x = 3y \\ z = -2y \end{matrix}$ is a subspace of $\mathbb{R}^3$.

Indeed, this set of vectors is the same as:

$$\left\{ \begin{pmatrix} 3t \\ t \\ -2t \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ t \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} : t \in \mathbb{R} \right\}$$

which is a (one-dimensional) subspace of $\mathbb{R}^3$.

Example: Determine whether the set of vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ where $y = x^2$ is a subspace of $\mathbb{R}^2$.

Let $S = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : y = x^2, x \in \mathbb{R} \right\}$.

Clearly $0_{2\times 1} \in S$. For S to be a subspace, we need to verify if

$u, v \in S \Rightarrow u + v \in S$

$u \in S \Rightarrow \lambda v \in S \quad \forall \lambda \in \mathbb{R}$.

It is not hard to verify that these properties fail in general. For example: $\begin{pmatrix} 1 \\ 1 \end{pmatrix} \in S$, but $\begin{pmatrix} 2 \\ 2 \end{pmatrix} \notin S$.

__Subspace associated to matrices:__

Def^{n}: Let $A \in \mathbb{R}^{m \times n}$. Then

① The __row space__ of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A. (Denoted $\text{row}\, A$)

② The __column space__ of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A. (Notation: $\text{col}\, A$)

__Remark:__ $Ax = b$ has a solution iff $b \in \text{col}\, A$.

__Theorem:__ If B is obtained from A by performing finitely many elementary row operations, then

$$\text{row}(B) = \text{row}(A)$$

Theorem: Let $A \in \mathbb{R}^{m\times n}$ and let N be the set of solutions of the linear system $Ax = 0_{m\times 1}$, i.e.,

$$N = \{x \in \mathbb{R}^n : Ax = 0_{m\times 1}\}.$$

Then N is a subspace of $\mathbb{R}^n$ called the null space (or kernel) of A.

Notation: $A = \text{null}(A)$ or $A = \text{Ker}\, A$.

Proof: ① $A 0_{n\times 1} = 0_{m\times 1}$ so $0_{n\times 1} \in N$.

② Suppose $u, v \in N$. Then $Au = Av = 0_{m\times 1}$.

therefore $A(u+v) = Au + Av = 0_{m\times 1} + 0_{m\times 1} = 0_{m\times 1}$.

thus $u+v \in N$.

③ If $u \in N$, then $Au = 0_{m\times 1}$ and so $A(\lambda u) = \lambda Au$

$= \lambda \cdot 0_{m\times 1}$

$= 0_{m\times 1}$

Therefore $\lambda u \in N$.

It follows that N is a vector subspace of $\mathbb{R}^n$.

Using the previous theorem, we can finally prove that a linear system has either:

1- No solution
2- A unique solution
3- Infinitely many solutions.

Proof: If the system has either a unique solution or no solution, then we are done. Now, suppose the system has at least two solutions, say

$$A x_1 = b \qquad \text{with} \quad x_1 - x_2 \neq 0.$$
$$A x_2 = b.$$

Then $A(x_1 - x_2) = 0$ and so $x_1 - x_2 \in \text{null}(A)$.

Set $x_0 := x_1 - x_2$. Since null(A) is a subspace, then $\lambda x_0 \in \text{null}(A)$ for all $\lambda \in \mathbb{R}$. Moreover:

$$A(x_1 + \lambda x_0) = b + 0 = b.$$

Therefore $x_1 + \lambda x_0$ is a solution of the system for every $\lambda \in \mathbb{R}$ and so the system has infinitely many solutions.. $\square$.

Basis:

We now introduce the fundamental concept of a basis of a vector subspace.

Motivation: In $\mathbb{R}^2$, the vectors $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ can be used to decompose any vector in a unique way:

$$\begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = x e_1 + y e_2.$$

Similarly, if $S \subset \mathbb{R}^3$ is a plane that goes through the origin, say:

$$S = \left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\},$$

then every vector $v \in S$ can be written as a linear combination $v = \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Moreover, the combination is unique since $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ are linearly independent. Indeed, if

$$v = \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \mu_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

then $(\lambda_1 - \mu_1) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (\lambda_2 - \mu_2) \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 0_{3\times 1}$

$\Rightarrow$ $\lambda_1 - \mu_1 = 0$, $\lambda_2 - \mu_2 = 0$ since $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ are lin. independent.

Therefore $\lambda_1 = \mu_1$ & $\lambda_2 = \mu_2$.

We can see the tuple (λ_1, λ_2) as "local coordinates" for the point v in the plane S.

The idea of a "basis" captures the idea of finding vectors in a subspace that can be used to decompose any vector in that subspace in a unique way.

Definition: Let S be a subspace of $\mathbb{R}^n$. A basis of S is a set of vectors $\{v_1, \dots, v_k\}$ such that:

① $S = \text{Span}(v_1, \dots, v_k)$

② The vectors $v_1, \dots, v_k$ are linearly independent.

Example:

① Clearly $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ is a basis of $\mathbb{R}^2$.

More generally, any pair of linearly independent vectors in $\mathbb{R}^2$ is a basis of $\mathbb{R}^2$.

② Let $S = \text{span}(u, v, w)$, where

$$u = \begin{pmatrix} 3 \\ -1 \\ 5 \end{pmatrix}, \quad v = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \quad w = \begin{pmatrix} 0 \\ -5 \\ 1 \end{pmatrix}.$$

Find a basis for S.

Clearly u, v, w span S. However, they may not be linearly independent... In fact, $w = 2u - 3v$.

Therefore $\text{Span}(u, v, w) = \text{span}(u, v)$

Clearly $u \neq \lambda v$. Thus u, v are lin. ind.

Conclusion: $\{u, v\}$ is a basis of S.

The following theorem shows that the notion of basis captures the idea of decomposing vectors in a subspace in a unique way.

<u>Theorem</u>: Let $\mathcal{B} = \{v_1, \dots, v_K\}$ be a basis of a subspace $S \subset \mathbb{R}^n$. Then every vector in S can be written as a linear combination of $v_1, \dots, v_K$. Moreover, that linear combination is unique.

<u>Proof</u>: Let $v \in S$. Since $\operatorname{span} \mathcal{B} = S$, there exist $\lambda_1, \dots, \lambda_K \in \mathbb{R}$ s.t. $v = \lambda_1 v_1 + \dots + \lambda_K v_K$. Suppose there exists another linear combination s.t. $v = \mu_1 v_1 + \dots + \mu_K v_K$.

Then $0 = v - v = (\lambda_1 - \mu_1) v_1 + \dots + (\lambda_K - \mu_K) v_K$.

Since $v_1, \dots, v_K$ are lin. independent, we conclude that

$$\lambda_1 - \mu_1 = \dots = \lambda_K - \mu_K = 0$$

i.e. $\lambda_1 = \mu_1, \lambda_2 = \mu_2, \dots, \lambda_K = \mu_K$.

The combination is therefore unique.

$\square$.

<u>Example</u>: Find a basis for row A, where

$$A = \begin{pmatrix} 1 & 1 & 3 & 1 & 6 \\ 2 & -1 & 0 & 1 & -1 \\ -3 & 2 & 1 & -2 & 1 \\ 4 & 1 & 6 & 1 & 3 \end{pmatrix}.$$

Solⁿ: The reduced row echelon form of A is:

$$R = \begin{pmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We know that row A = row R (Recall row A is the span of the rows of A)

Clearly, row R is spanned by the nonzero rows of R.

It is not hard to check that $u = (1,0,1,0,-1)$, $v = (0,1,2,0,3)$, $w = (0,0,0,1,4)$ are lin. independent. Indeed,

$$\lambda_1 u + \lambda_2 v + \lambda_3 w = 0 \iff \begin{aligned} \lambda_1 \quad\quad\quad &= 0 \\ \lambda_2 \quad\quad &= 0 \\ \lambda_1 + 2\lambda_2 \quad\quad &= 0 \\ \lambda_3 &= 0 \\ -\lambda_1 + 3\lambda_2 + 4\lambda_3 &= 0 \end{aligned}$$

$$\iff \lambda_1 = \lambda_2 = \lambda_3 = 0$$

Therefore row A = span(u, v, w). "3 dimensional" subspace of $\mathbb{R}^5$.

A basis provides "local coordinates" for vectors in a subspace

$$(\lambda_1, \dots, \lambda_k) \longleftrightarrow \lambda_1 v_1 + \cdots + \lambda_k v_k.$$

If a basis has k vectors, we naturally identify the subspace with $\mathbb{R}^k$ and think of the subspace as being "k-dimensional".

The following theorem shows that this notion is well-defined.

Theorem (The basis theorem) (Note: plural of "basis" = bases)

Let S be a subspace of $\mathbb{R}^n$. Then any two bases for S have the same number of vectors.

Proof: Let $\mathcal{B} = \{u_1, \dots, u_r\}$, $\mathcal{C} = \{v_1, \dots, v_s\}$ be two bases of S. We need to show $r = s$. Suppose $r \neq s$. Then $r < s$ or $r > s$.

Suppose $r < s$ (which we may assume without loss of generality. Otherwise, interchange $\mathcal{B}$ and $\mathcal{C}$).

We will prove that this forces $\mathcal{C}$ to be a linearly dependent set of vectors. To this end, let

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_s v_s = 0. \qquad (1)$$

Since $\mathcal{B}$ is a basis of S, we can find a_{ij} s.t.

$$\begin{aligned} v_1 &= a_{11} u_1 + a_{12} u_2 + \dots + a_{1r} u_r \\ v_2 &= a_{21} u_1 + a_{22} u_2 + \dots + a_{2r} u_r \\ &\vdots \\ v_s &= a_{s1} u_1 + a_{s2} u_2 + \dots + a_{sr} u_r. \end{aligned} \qquad (2)$$

Substituting in (1), we obtain:

$$\lambda_1 (a_{11} u_1 + \dots + a_{1r} u_r) + \dots + \lambda_s (a_{s1} u_1 + \dots + a_{sr} u_r) = 0$$

Regrouping, we have:

$$\lambda_1 a_{11} + \lambda_2 a_{21} + \dots + \lambda_s a_{s1} = 0$$
$$\lambda_1 a_{12} + \lambda_2 a_{22} + \dots + \lambda_s a_{s2} = 0$$
$$\vdots \qquad \ddots \qquad \vdots$$
$$\lambda_1 a_{1r} + \lambda_2 a_{2r} + \dots + \lambda_s a_{sr} = 0 .$$

This is a linear homogeneous system of r equations and s unknowns with $r < s$. We proved before that such a system has infinitely many solutions (rank theorem) Each non-zero solution produces a non-trivial lin. comb of the v_is that equals 0. This contradicts the fact that $v_1, \dots, v_s$ are lin. independent. We conclude that $r = s$

$\square$.

Definition: If S is a subspace of $\mathbb{R}^n$, then the number of vectors in a basis for S is called the <u>dimension</u> of S, denoted $\dim S$.

Remarks:

1) Note that: $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$ is a basis of $\mathbb{R}^n$. Therefore $\dim \mathbb{R}^n = n$.

2) If $v_1, \dots, v_k$ are linearly independent, then

$$\dim \operatorname{span}(v_1, \dots, v_k) = k .$$

(This is because $v_1, \dots, v_k$ is a basis of $\operatorname{span}(v_1, \dots, v_k) \dots$).

③ $S = \{0_{n\times 1}\}$ is a subspace of $\mathbb{R}^n$. However, $\{0\}$ does not have a basis. We set $\dim\{0\} = 0$ as a convention.

Recall that we defined the rank of a matrix as the "number of non-zero rows in its echelon form".

In fact, $\boxed{\text{rank } A = \dim \text{row}(A)}$.

To prove this result, first note that $\dim \text{row}(A) = \dim \text{row}(R)$ where R is the echelon form of A. (In fact, $\text{row}(A) = \text{row}(R)$ as we saw before...)

Now, clearly $\dim \text{row } A \leq$ # of nonzero rows of R.

Using the "staircase pattern" of R, it is not hard to see that the nonzero rows of R are linearly independent. We conclude that $\text{rank } A = \dim \text{row}(A)$.

In fact, even more is true:

<u>Theorem</u>: The row and column space of a matrix A have the same dimension, i.e.,

$$\boxed{\dim \text{row } A = \dim \text{col } A}$$

Proof: Let $A \in \mathbb{R}^{m\times n}$ and suppose $r = \operatorname{rank} A$.

Let $v_1, \dots, v_r$ be a basis of row A. We claim that $Av_1, \dots, Av_r$ are linearly independent. Indeed, suppose

$$\lambda_1 A v_1 + \lambda_2 A v_2 + \cdots + \lambda_r A v_r = 0.$$

Then $A(\underbrace{\lambda_1 v_1 + \cdots + \lambda_r v_r}_{\text{Clearly, this vectors belong to row } A.}) = 0.$

Let $v := \lambda_1 v_1 + \cdots + \lambda_r v_r$. Since $Av = 0$, then v is orthogonal to every row of A

$$\begin{pmatrix} - & r_1 & - \\ - & r_2 & - \\ & \vdots & \\ - & r_m & - \end{pmatrix} \begin{pmatrix} | \\ v \\ | \end{pmatrix} = \begin{pmatrix} \langle r_1, v\rangle \\ \vdots \\ \langle r_m, v\rangle \end{pmatrix} = 0.$$

Thus v is orthogonal to every vector in row A.

But v belongs to row A. Therefore v is orthogonal to itself, i.e., $\langle v, v\rangle = \|v\|^2 = 0$. This proves $v = 0$, i.e.,

$$\lambda_1 v_1 + \cdots + \lambda_r v_r = 0.$$

Since $v_1, \dots, v_r$ are linearly independent, we conclude that $\lambda_1 = \cdots = \lambda_r = 0$. Thus, $Av_1, \dots, Av_r$ are linearly independent.

We conclude that $\dim \operatorname{col} A \geq r = \dim \operatorname{row} A$.

Conversely,

$$r = \dim \operatorname{row} A = \dim \operatorname{col} A^T \overset{\text{previous result}}{\geq} \dim \operatorname{row} A^T = \dim \operatorname{col} A$$

Therefore $\dim \operatorname{row} A \geq \dim \operatorname{col} A$

This concludes the proof. $\square$.

Definition: • The rank of $A \in \mathbb{R}^{m\times n}$, denoted rank A, is the dimension of its row (or column) space:

$$\operatorname{rank} A := \dim \operatorname{row} A = \dim \operatorname{col} A.$$

• The nullity of $A \in \mathbb{R}^{m\times n}$, denoted nullity$(A)$, is the dimension of its null space:

$$\operatorname{nullity}(A) = \dim \operatorname{null}(A).$$

In other words, nullity (A) is the dimension of the space of solutions of $Ax = 0_{m\times 1}$.

Recall that $v_1, \ldots, v_n \in \mathbb{R}^m$ are linearly independent iff

$$Ax = 0_{n\times 1} \Leftrightarrow x = 0_{m\times 1}$$

where $A = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}_{m\times n}$

Note that in that case:

$$\text{rank } A = n$$
$$\text{nullity } A = 0.$$

In particular $\text{rank } A + \text{nullity } A = n$.

More generally, we have the following important theorem.

<u>Theorem</u>: (Rank-nullity theorem)

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

$$\boxed{\text{rank } A + \text{nullity } A = n.}$$

<u>Proof</u>: Let R denote the reduced row echelon form of A. and let $r = \text{rank } A$. Then R has r leading 1s, so there are r variables and $n-r$ free variables in the solution of the linear system $AX = 0_{m \times 1}$. Then $\dim \text{null } A = n-r$ and $\text{rank } A + \text{nullity } A = r + n - r = n$.

$\square$.

<u>Example</u>: Find the nullity of

$$M = \begin{pmatrix} 2 & 3 \\ 1 & 5 \\ 4 & 7 \\ 3 & 6 \end{pmatrix}.$$

Solⁿ: $M \in \mathbb{R}^{4\times 2}$ so $n = 2$. Clearly rank $M = 2$ since the two columns of M are linearly independent. Since

$$2 = \text{rank } A + \text{nullity } A$$

we conclude that nullity $A = 0$.

<u>Important remark:</u>

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

$$\begin{aligned}
\text{rank } A = n &\iff \text{Col } A = \mathbb{R}^n \\
&\iff \text{nullity } A = 0 \\
&\iff \text{null } A = \{0\} \\
&\iff 0_{n\times 1} \text{ is the only solution of } Ax = 0_{m\times 1} \\
&\iff \text{The columns of } A \text{ are linearly independent.}
\end{aligned}$$

<u>Example:</u> Find the nullity of

$$M = \begin{pmatrix} 2 & 1 & -2 & -1 \\ 4 & 4 & -3 & 1 \\ 2 & 7 & 1 & 8 \end{pmatrix}.$$

$$M \to \begin{pmatrix} 2 & 1 & -2 & -1 \\ 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{aligned} &\Rightarrow \text{rank } M = 2 \\ &\Rightarrow \text{nullity } M = 4 - 2 = 2. \end{aligned}$$

We have already observed many relations between rank, nullity, etc. The following theorem characterizes when a linear system has a unique solution for every b.

Theorem: (Fundamental theorem of invertible matrices)

Let $A \in \mathbb{R}^{n\times n}$. Then the following are equivalent:

a) A is invertible (i.e. $\exists B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I_n$)

b) $Ax = b$ has a unique solution for every $b \in \mathbb{R}^n$.

c) $Ax = 0_{n\times 1}$ iff $x = 0_{n\times 1}$.

d) The reduced row echelon form of A is I_n

e) A is a product of elementary matrices.

f) rank $A = n$ (i.e. A has full rank.)

g) nullity $A = 0$ (equivalently, null $A = \{0\}$)

h) The columns of A are linearly independent.

i) col $A = \mathbb{R}^n$

j) The columns of A form a basis of $\mathbb{R}^n$.

k) The rows of A are linearly independent

l) row $A = \mathbb{R}^n$

m) The rows of A form a basis of $\mathbb{R}^n$.

Example: Show that the vectors

$$\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix}, \text{ and } \begin{pmatrix}4\\9\\7\end{pmatrix}$$

form a basis of $\mathbb{R}^3$.

Sol^n: By the fundamental theorem, the vectors form a basis of $\mathbb{R}^3$ iff the matrix

$$A = \begin{pmatrix}1 & -1 & 4\\2 & 0 & 9\\3 & 1 & 7\end{pmatrix}$$ has rank 3.

The row echelon form of A is: $\begin{pmatrix}1 & -1 & 4\\0 & 2 & 1\\0 & 0 & 7\end{pmatrix}$

and so $\operatorname{rank} A = 3$. Thus, the three vectors form a basis of $\mathbb{R}^3$.

Section 3.6: Linear transformations.

So far, we saw:

① Linear systems can be identified with matrices

② Matrices can be better understood in terms of their row and column vectors. In turn, this approach provides deep insights to understand linear systems.

We will now look at matrices in a different way.

A matrix $A \in \mathbb{R}^{m\times n}$ naturally "operates" on $\mathbb{R}^n$, i.e., each matrix defines a function from $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$f(x) = Ax \qquad (x \in \mathbb{R}^n).$$

That function satisfies:

$$① \ f(x+y) = f(x) + f(y) \qquad \forall\, x, y \in \mathbb{R}^n$$

$$② \ f(\lambda x) = \lambda f(x) \qquad \forall\, \lambda \in \mathbb{R},\ x \in \mathbb{R}^n.$$

Such functions (or mappings) are called "linear".

Definition: A function $T: \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation if

$$① \ T(u+v) = T(u) + T(v) \qquad \forall\, u, v \in \mathbb{R}^n$$

$$② \ T(\lambda v) = \lambda T(v) \qquad \forall\, \lambda \in \mathbb{R},\ v \in \mathbb{R}^n.$$

Example: Let $T: \mathbb{R}^2 \to \mathbb{R}^3$ be defined by:

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 2x - y \\ 3x + 4y \end{pmatrix}.$$

Then $T\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) = T\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ 2x_1 + 2x_2 - y_1 - y_2 \\ 3x_1 + 3x_2 + 4y_1 + 4y_2 \end{pmatrix}$

$$= \begin{pmatrix} x_1 \\ 2x_1 - y_1 \\ 3x_1 + 4y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ 2x_2 - y_2 \\ 3x_2 + 4y_2 \end{pmatrix} = T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}.$$

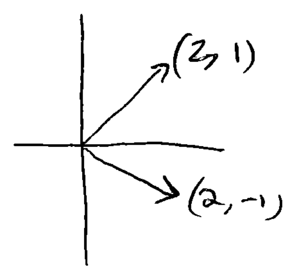
<u>Also</u>: $$T\left(\lambda\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) = T\begin{pmatrix} \lambda x_1 \\ \lambda y_1 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ 2\lambda x_1 - \lambda y_1 \\ 3\lambda x_1 + 4\lambda y_1 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ 2x_1 - y_1 \\ 3x_1 + 4y_1 \end{pmatrix}$$
$$= \lambda T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}.$$

Therefore T is linear.

<u>Remark</u>: $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$!

<u>Example</u>: Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the function that maps every vector $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ to its reflection with respect to the x-axis, i.e.,

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}$$



Then ① $$T\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) = \begin{pmatrix} x_1 + x_2 \\ -(y_1 + y_2) \end{pmatrix} = T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

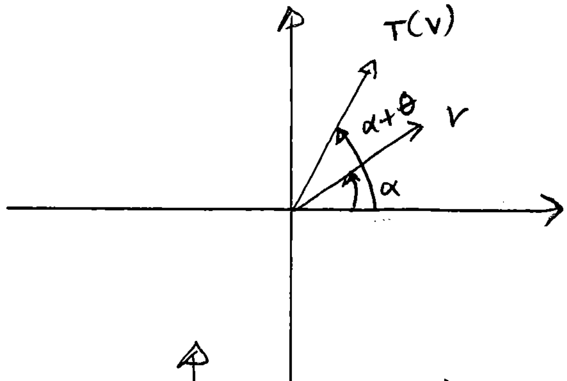
② $$T\left(\lambda\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) = \begin{pmatrix} \lambda x_1 \\ -\lambda y_1 \end{pmatrix} = \lambda T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}.$$

Thus T is linear!

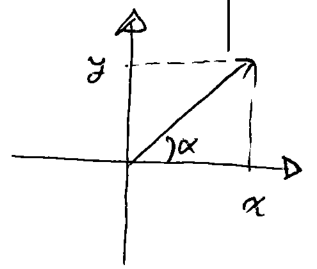
<u>Remark</u>: $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$!

Note: If $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, then $A^2 = I_2$ ...
$\Leftrightarrow$ Reflection of reflection = Identity..

Example: Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the "rotation by θ" mapping, i.e.,:



If $v = \begin{pmatrix} x \\ y \end{pmatrix}$



then $x = \|v\| \cos \alpha$
$y = \|v\| \sin \alpha$.

Therefore, if $\begin{pmatrix} x' \\ y' \end{pmatrix} = T\begin{pmatrix} x \\ y \end{pmatrix}$, then $x' = \|v\| \cos(\alpha + \theta)$
$y' = \|v\| \sin(\alpha+\theta)$

Recall: $\cos(a+b) = \cos a \cos b - \sin a \sin b$
$\sin(a+b) = \sin a \cos b + \sin b \cos a$.

Thus $x' = \|v\| (\cos\alpha \cos\theta - \sin\alpha \sin\theta)$

$= x \cos\theta - y \sin\theta$

$y' = \|v\| (\sin\alpha \cos\theta + \sin\theta \cos\alpha)$

$= y \cos\theta + x \sin\theta$

We conclude: $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$

Thus, T is linear!

Remarks ① If $T_1(x) = Ax$, $T_2(x) = Bx$, then $T_1(T_2 x) = ABx$.

$$\begin{tikzcd} \mathbb{R}^m \arrow[r, "T_2"] \arrow[rr, bend right, "T_1 \circ T_2"'] & \mathbb{R}^n \arrow[r, "T_1"] & \mathbb{R}^p \end{tikzcd}$$

Thus, the matrix product corresponds to composing the corresponding linear transformation.

② Linear systems are in correspondance with matrices. Thus, each linear system can be identified with a linear transformation.

We saw many examples of linear transformations. In each Example, we could find a matrix A s.t. $Tx = Ax$. This is not a coincidence.

Theorem: Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a matrix $A \in \mathbb{R}^{m \times n}$ such that

$$T(x) = Ax.$$

More precisely, $T(x) = Ax$ where $A = \begin{pmatrix} | & | & & | \\ Te_1 & Te_2 & \cdots & Te_n \\ | & | & & | \end{pmatrix}$.

(Here, $\{e_1, \dots, e_n\}$ is the standard basis of $\mathbb{R}^n$).

Proof: Let $x \in \mathbb{R}^n$. Then $x = x_1 e_1 + \dots + x_n e_n$. Since T is linear, $T(x) = x_1 T(e_1) + \dots + x_n T(e_n) = Ax$. $\square$

Example: Let T be the transformation $T: \mathbb{R}^2 \to \mathbb{R}^2$ obtained by

① first reflecting v w.r.t. the x axis and

② rotating v by $45°$.

Then $T = T_2 \circ T_1$ where T_1 = reflection

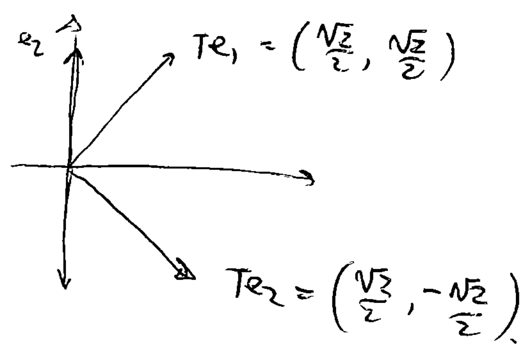
T_2 = rotation by $\pi/4$.

we know that $$T_1 x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

$$T_2 x = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Thus $$T = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{pmatrix}$$

Note that $$T = \begin{pmatrix} | & | \\ Te_1 & Te_2 \\ | & | \end{pmatrix}$$



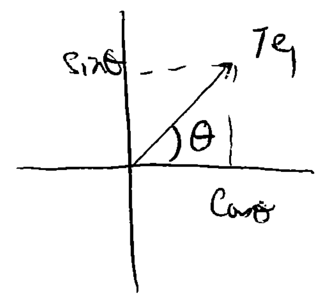
Example: A linear transformation $R: \mathbb{R}^n \to \mathbb{R}^n$ is invertible if there exists $S: \mathbb{R}^n \to \mathbb{R}^n$ s.t. $R(S(x)) = S(R(x)) = x$.

In other words, $R \circ S = Id$ and $S \circ R = Id$.

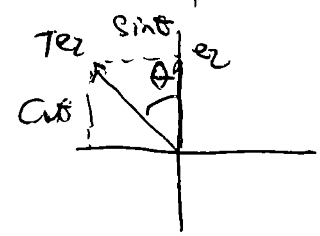
Clearly, R is invertible $\iff$ R^{-1} exists.

Let us now revisit the rotation example.

Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation by θ. Clearly T is linear. Also:



$$Te_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$



$$Te_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

Thus $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

Similarly, we can find the matrix of a rotation in $\mathbb{R}^3$ by looking at its effect on $e_1, e_2, e_3 \ldots$

Chapter 4: Eigenvalues and eigenvectors.

Eigenvalues and eigenvectors are a fundamental topics in linear algebra, and have numerous applications.

Def: Let $A \in \mathbb{R}^{n\times n}$. A scalar $\lambda \in \mathbb{R}$ is said to be an eigenvalue of A if $Av = \lambda v$ for some vector $v \neq 0_{n\times n}$. The vector v is said to be an eigenvector associated to λ.

Example: Let $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. Then

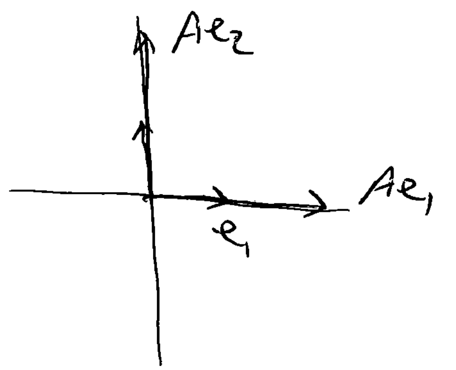
$$A\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}4\\4\end{pmatrix} = 4\begin{pmatrix}1\\1\end{pmatrix}.$$

Thus 4 is an eigenvalue of A w/ associated eigenvector $\begin{pmatrix}1\\1\end{pmatrix}$.

Motivation: Suppose $A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ is a diagonal matrix.

Then $Ae_1 = ae_1$ and $Ae_2 = be_2$, i.e.,

the action of A on the basis $\{e_1, e_2\}$ is to "rescale the basis":



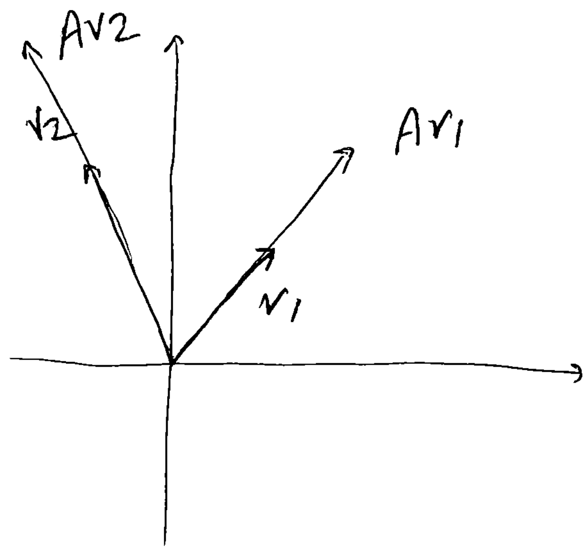
More generally $A\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}ax\\by\end{pmatrix}$ so the transformation A "rescales" vectors.

Now, suppose A is an arbitrary 2×2 matrix and let

$$Av_1 = \lambda_1 v_1$$
$$Av_2 = \lambda_2 v_2$$

with v_1, v_2 linearly independent.

Then A acts as a diagonal matrix when we look at it in the $\{v_1, v_2\}$ basis:



So, in some sense, A is equivalent to the transformation

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Example: Let $A = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}$. Find all the eigenvectors associated to $\lambda = 5$.

Sol^n: We have

$$Av = 5v \iff (A - 5I)v = 0.$$

$$\iff \begin{pmatrix} -4 & 2 \\ 4 & -2 \end{pmatrix} v = 0$$

(Thus v is such an eigenvector iff $v \in$ nullspace of $A - \lambda I$).

Now, $\left(\begin{array}{cc|c} -4 & 2 & 0 \\ 4 & -2 & 0 \end{array}\right) \xrightarrow{R_2 \to R_2 + R_1} \left(\begin{array}{cc|c} -4 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right) \xrightarrow{R_1 \to -\frac{1}{4} R_1} \left(\begin{array}{cc|c} 1 & -1/2 & 0 \\ 0 & 0 & 0 \end{array}\right)$

$* \; x_1 - 1/2\, x_2 = 0 \; *$

Thus $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} t/2 \\ t \end{pmatrix} = t \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$ $(t \in \mathbb{R})$ $\quad x_2 = t$, $x_1 = t/2$

$= s \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $(s \in \mathbb{R})$.

Remark: It is not surprising that if $Av = \lambda v$, then

$$A(\mu v) = \mu Av = \mu \lambda v = \lambda(\mu v)$$

i.e., if v is an eigenvector associated to λ, then μv is also an eigenvector associated to λ for any $\mu \in \mathbb{R}$.

More generally,

If λ is an eigenvalue of A, then the set of eigenvectors associated to λ is $\text{null}(A - \lambda I)$.

Indeed
$$\begin{aligned} Av = \lambda v &\iff (A - \lambda I)v = 0 \\ &\iff v \in \text{null}(A - \lambda I). \end{aligned}$$

Thus, the set of eigenvectors associated to λ forms a subspace.

Def: Let $A \in \mathbb{R}^{n\times n}$ and let λ be an eigenvalue of A. The eigenvectors associated to λ form a subspace of $\mathbb{R}^n$ called the eigenspace of λ, denoted E_λ.

Exercise: Show that $\lambda = 6$ is an eigenvalue of

$$A = \begin{pmatrix} 7 & 1 & -2 \\ -3 & 3 & 6 \\ 2 & 2 & 2 \end{pmatrix}.$$

Find a basis for E_6.

Example: We saw that $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is the reflection w.r.t. the x-axis.

Clearly
$$A e_1 = e_1$$
$$A e_2 = -e_2.$$

Thus $\lambda = 1$ and $\lambda = -1$ are eigenvalues of A.

It is not hard to verify that
$$E_{-1} = \operatorname{span}\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$
$$E_1 = \operatorname{span}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right).$$

thus A has a basis of eigenvectors.

A fundamental observation:

Recall that
$$A \text{ is invertible} \iff \operatorname{null} A = \{0\}$$
$$\iff \det A \neq 0$$

Thus λ is an eigenvalue of A $\iff$ $Av = \lambda v$ for some $v \neq 0$

$\iff$ $(A - \lambda I)\, v = 0$ for some $v \neq 0$

$\iff$ $\exists\, v \neq 0$ s.t. $v \in \operatorname{null}(A - \lambda I)$

$\iff$ $\operatorname{null}(A - \lambda I) \neq \{0\}$

$\iff$ $\det(A - \lambda I) = 0$.

Therefore, finding the eigenvalues of A is equivalent to finding the roots of the polynomial

$$p(\lambda) = \det(A - \lambda I)$$

<u>Note</u>: p is a degree 2 polynomial!

<u>Example</u>: Find the eigenvalues of $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$.

we have $\det(A - \lambda I) = \det \begin{pmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix}$

$= (3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8.$

$\Rightarrow \lambda = 4$ and $\lambda = 2$ are the only eigenvalues of A.

How do we find eigenvectors?

$v \in E_4 \longleftarrow$ eigenspace associated to $\lambda = 4$

$\iff \quad (A - 4I)v = 0$

$\iff \quad v \in \text{null}(A - 4I)$

we can easily compute the nullspace (i.e. solve $(A-4I)x = 0$)

$$A - 4I = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{so we solve} \quad \left(\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array}\right)$$

$$\left(\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array}\right) \xrightarrow{R_2 \to R_2 + R_1} \left(\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

So $v = \begin{pmatrix} x \\ y \end{pmatrix}$ is an eigenvector iff $-x + y = 0$ i.e. if $x = y$.

Thus $v = \begin{pmatrix} t \\ t \end{pmatrix} = t\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and so

$$E_4 = \text{span}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right).$$

Determinants:

For 2×2 matrices $A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we saw that the determinant $\det A = ad-bc$ plays an important role to decide whether A is invertible or not. We now generalize this idea to $n\times n$ matrices.

<u>Remark</u>: Another notation for $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

<u>3×3 case</u>: Let $A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Then we define:

$$\det A = a_{11}\cdot\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\cdot\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \qquad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \qquad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

<u>Remark</u>: ① The 2×2 determinants in the expression of $\det A$ are obtained by deleting rows and columns in the original matrix.

② The signs in front of the 2×2 determinants alternate.

Example:

Let $A = \begin{pmatrix} 5 & -3 & 2 \\ 1 & 0 & 2 \\ 2 & -1 & 3 \end{pmatrix}$. Then:

$$\det A = 5 \begin{vmatrix} 0 & 2 \\ -1 & 3 \end{vmatrix} - (-3) \begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} + 2 \begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix}$$

$$= 5(0\times 3 - 2\times(-1)) + 3(1\times 3 - 2\times 2) + 2(1\times -1 - 0\times 2)$$

$$= 5\times 2 + 3\times(-1) + 2(-1)$$

$$= 5.$$

Let $A \in \mathbb{R}^{n\times n}$. We denote by A_{ij} the $(n-1)\times(n-1)$ submatrix of A obtained by deleting the i^{th} row and j^{th} column of A.

A_{ij} is called the (i,j) minor of A.

Using that notation we have for a 3×3 matrix A:

$$\det A = a_{11}\det A_{11} - a_{12}\det A_{12} + a_{13}\det A_{13}.$$

$$= \sum_{j=1}^{3} (-1)^{1+j} a_{ij} \det A_{1j}.$$

$n\times n$ determinant:

Let $A \in \mathbb{R}^{n\times n}$. We define:

$$\boxed{\det A = a_{11}\det A_{11} - a_{12}\det A_{12} + \cdots + (-1)^{1+n} a_{1n}\det A_{1n}}$$

$$= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j}.$$

<u>Remark</u>: ① As in the 3×3 case, the determinant is computed by moving along the first row, alternating signs, and computing determinants obtained by deleting one row and one column.

② The definition is recursive in the sense that the same procedure has to be performed to compute $\det A_{ij}$ until only 2×2 matrices remain.

<u>Example</u>: Let $A = \begin{pmatrix} 2 & -3 & 0 & 1 \\ 5 & 4 & 2 & 0 \\ 1 & -1 & 0 & 3 \\ -2 & 1 & 0 & 0 \end{pmatrix}$. Then:

$$\det A = 2\begin{vmatrix} 4 & 2 & 0 \\ -1 & 0 & 3 \\ 1 & 0 & 0 \end{vmatrix} - (-3)\begin{vmatrix} 5 & 2 & 0 \\ 1 & 0 & 3 \\ -2 & 0 & 0 \end{vmatrix} + 0\cdot\begin{vmatrix} 5 & 4 & 0 \\ 1 & -1 & 3 \\ -2 & 1 & 0 \end{vmatrix}$$

$$- 1 \times \begin{vmatrix} 5 & 4 & 2 \\ 1 & -1 & 0 \\ -2 & 1 & 0 \end{vmatrix}.$$

Now:

$$\begin{vmatrix} 4 & 2 & 0 \\ -1 & 0 & 3 \\ 1 & 0 & 0 \end{vmatrix} = 4 \cdot \begin{vmatrix} 0 & 3 \\ 0 & 0 \end{vmatrix} - 2 \begin{vmatrix} -1 & 3 \\ 1 & 0 \end{vmatrix} = 4 \cdot 0 - 2(-3) = 6$$

$$\begin{vmatrix} 5 & 2 & 0 \\ 1 & 0 & 3 \\ -2 & 0 & 0 \end{vmatrix} = 5 \begin{vmatrix} 0 & 3 \\ 0 & 0 \end{vmatrix} - 2 \begin{vmatrix} 1 & 3 \\ -2 & 0 \end{vmatrix} = 5 \cdot 0 - 2(6) = -12$$

$$\begin{vmatrix} 5 & 4 & 2 \\ 1 & -1 & 0 \\ -2 & 1 & 0 \end{vmatrix} = 5 \begin{vmatrix} -1 & 0 \\ 1 & 0 \end{vmatrix} - 4 \begin{vmatrix} 1 & 0 \\ -2 & 0 \end{vmatrix} + 2 \begin{vmatrix} 1 & -1 \\ -2 & 1 \end{vmatrix}$$

$$= 5 \cdot 0 - 4 \cdot 0 + 2(1-2) = -2$$

Putting everything together:

$$\det A = 2 \cdot 6 - 3 \cdot 12 - 1 \cdot (-2)$$

$$= 12 - 36 + 2 = -22$$

<u>A "better" way of computing determinants:</u>

Let $A \in \mathbb{R}^{n \times n}$. The (i,j) <u>cofactor</u> of A is defined to be:

$$\boxed{C_{ij} := (-1)^{i+j} \det A_{ij}}.$$

With this notation, we have:

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij}.$$

Remark: The sign $(-1)^{i+j}$ in the definition of C_{ij} can be computed using a "checkerboard" pattern:

$$\begin{matrix} + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

For example the sign in $C_{3,4}$ is -1.

The following theorem shows that determinants can be expanded along any row or column.

Theorem: (Laplace expansion theorem)

Let $A \in \mathbb{R}^{n \times n}$ $(n \geq 2)$. Then for any $1 \leq i \leq n$,

$$\det A = a_{i1} C_{i1} + a_{i2} C_{i2} + \cdots + a_{in} C_{in}.$$

(Cofactor expansion along i^{th} row).

Similarly, for any $1 \leq j \leq n$,

$$\det A = a_{1j} C_{1j} + a_{2j} C_{2j} + \cdots + a_{nj} C_{nj}.$$

(Cofactor expansion along i^{th} column).

With that theorem in mind, let us revisit the previous example:

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

<u>Idea</u>: Expand along a row or column with many zeros...

Here, column 3 is a good choice:

$$\det A = -2 \begin{vmatrix} 2 & -3 & 1 \\ 1 & -1 & 3 \\ -2 & 1 & 0 \end{vmatrix}$$ (much better than before...)

Let us row expand along the third row (since it contains 1 zero):

$$= -2\left[-2\begin{vmatrix} -3 & 1 \\ -1 & 3 \end{vmatrix} - 1 \cdot \begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix}\right]$$

$$= -2\left[-2(-9+1) - 1(6-1)\right]$$

$$= -2\left[+16-5\right] = -22$$

<u>Exercise</u>: Try other rows and columns. For example: <u>using 4^{th} column</u>:

$$\det A = -\begin{vmatrix} 5 & 4 & 2 \\ 1 & -1 & 0 \\ -2 & 1 & 0 \end{vmatrix} - \begin{vmatrix} 2 & -3 & 0 \\ 5 & 4 & 2 \\ -2 & 1 & 0 \end{vmatrix}$$

$$\begin{matrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{matrix}$$

the signs are given by the "checkerboard pattern"

Exercise If $A = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}$ is diagonal, then

$$\det A = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{nn} = \prod_{i=1}^{n} a_{ii}.$$

Properties of the determinant: Let $A = \begin{pmatrix} | & | & & | \\ c_1 & c_2 & \cdots & c_n \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{n\times n}$.

We can see the determinant $\det A$ as a function of the rows of A, i.e.,

$$\det A = \det(c_1, \ldots, c_n),$$

and examine what happens when we perform operations on the columns.

Theorem: Let $v_1, \ldots, v_n \in \mathbb{R}^n$ and let $\det(v_1, \ldots, v_n) = \det\begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$.
$w_1, \ldots, w_n \in \mathbb{R}^n$

Then ① $\det(v_1, \ldots, v_{i-1}, \lambda v_i, v_{i+1}, \ldots, v_n) = \lambda \det(v_1, \ldots, v_n)$

(Multiplying a column by λ multiplies the det by λ)

② $\det(v_1, \ldots, v_{i-1}, \overset{\longleftrightarrow}{v_{i+1}, v_i}, v_{i+2}, \ldots, v_n)$

$= -\det(v_1, \ldots, v_n)$ (Switching columns change the sign)

③ $\det(v_1, \ldots, v_{i-1}, v_i + w_i, v_{i+1}, \ldots, v_n)$
$= \det(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n) + \det(v_1, \ldots, v_{i-1}, w_i, v_{i+1}, \ldots, v_n)$

Theorem: $\det A = \det A^T$.

Proof: Use the Laplace expansion formula.

As a consequence, if we see $\det A$ as a function of the rows of A, i.e. $\det A = \det(r_1, \dots, r_n)$ where

$$A = \begin{pmatrix} - & r_1 & - \\ & \vdots & \\ - & r_n & - \end{pmatrix},$$

then properties ①, ②, ③ of the previous theorem hold.

Remark: Properties ①, ③ of the previous theorem mean that det is "linear in each of its arguments", i.e., it's a linear function in each row/column of A.

When a function has property ②, we say that it is alternating.

The theorem can thus be summarized by saying that det is a multilinear alternating function in the rows or columns of A.

(In fact, it is the unique such function that satisfies $\det I_n = 1$).

Consequences of the previous results: (See Thm 4.3 in book)

① If $A \in \mathbb{R}^{n\times n}$ has two equal rows or columns, then $\det A = 0$

(This is because if $\tilde{A}$ is the matrix obtained by permuting the two equal row or columns that are equal, then

$$\det A = \det \tilde{A} = -\det A, \text{ so } \det A = 0.)$$

② If B is obtained by adding to a row (column) of A a multiple of another row/column of A, then

$$\det B = \det A.$$

(In other words, we perform $R_i \to R_i + k R_j$ or $C_i \to C_i + k C_j$)

③ $\det(\lambda A) = \lambda^n \det A$

Proof:

$$\det B = \det(C_1, \dots, C_{i-1}, C_i + k C_j, C_{i+1}, \dots, C_n)$$

$$= \det(C_1, \dots, C_{i-1}, C_i, C_{i+1}, \dots, C_n)$$

$$+ k \det(C_1, \dots, C_{i-1}, C_j, C_{i+1}, \dots, C_n)$$

$$= \det A + 0$$

The second det is 0 since C_j is repeated twice...

$\square$

An interesting consequence of ⓐ is that we can compute $\det A$ by transforming it to its row echelon form...

<u>WARNING</u>: The operation $R_i \to R_i + k R_j$ does not change det.

HOWEVER, 1) permuting rows changes the sign of det

2) Multiplying a row by k multiplies det by k.

So we need to keep track of these two operations to get the right value of det..

<u>Example</u>: Compute $\det A$ where $A = \begin{pmatrix} 0 & 2 & -4 & 5 \\ 3 & 0 & -3 & 6 \\ 2 & 4 & 5 & 7 \\ 5 & -1 & -3 & 1 \end{pmatrix}$.

Since the matrix is 4×4 and has not many zeros, we will reduce it to an echelon form:

$$\det A = \begin{vmatrix} 0 & 2 & -4 & 5 \\ 3 & 0 & -3 & 6 \\ 2 & 4 & 5 & 7 \\ 5 & -1 & -3 & 1 \end{vmatrix} \overset{R_1 \leftrightarrow R_2}{=} - \begin{vmatrix} 3 & 0 & -3 & 6 \\ 0 & 2 & -4 & 5 \\ 2 & 4 & 5 & 7 \\ 5 & -1 & -3 & 1 \end{vmatrix}$$

$$\overset{R_1 \to \frac{1}{3} R_1}{=} -3 \cdot \begin{vmatrix} 1 & 0 & -1 & 6 \\ 0 & 2 & -4 & 5 \\ 2 & 4 & 5 & 7 \\ 5 & -1 & -3 & 1 \end{vmatrix} \overset{\substack{R_3 \to R_3 - 2R_1 \\ R_4 \to R_4 - 5R_1}}{=} -3 \cdot \begin{vmatrix} 1 & 0 & -1 & 2 \\ 0 & 2 & -4 & 5 \\ 0 & 4 & 7 & 3 \\ 0 & -1 & 2 & -9 \end{vmatrix}$$

$$\overset{R_2 \leftrightarrow R_4}{=} -(-3)\begin{vmatrix} 1 & 0 & -1 & 2 \\ 0 & -1 & 2 & -9 \\ 0 & 4 & 7 & 3 \\ 0 & 2 & -4 & 5 \end{vmatrix} \overset{\substack{R_3 \to R_3 + 4R_2 \\ R_4 \to R_4 + 2R_2}}{=} 3 \cdot \begin{vmatrix} 1 & 0 & -1 & 2 \\ 0 & -1 & 2 & -9 \\ 0 & 0 & 15 & -33 \\ 0 & 0 & 0 & -13 \end{vmatrix}.$$

Now, the matrix is in row echelon form.

<u>Claim</u>: The determinant of a triangular matrix is the product of its diagonal elements, i.e.

if $A = \begin{pmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}$, then $\det A = a_{11} \times \dots \times a_{nn}$.

To prove the claim, just expand the determinant along the first column:

$$\det A = a_{11} \times \begin{vmatrix} a_{22} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{vmatrix}.$$

etc... (Use induction to prove the claim rigorously)

thus $\det A = 3 \times 1 \times -1 \times 15 \times -13 = 585$.

Determinant of matrix products:

Theorem:

Let $A, B \in \mathbb{R}^{n\times n}$, then

$$\boxed{\det(AB) = (\det A)(\det B)}$$

We saw before that a 2×2 matrix is invertible iff its det is nonzero. We can now prove that the same holds for $n\times n$ matrices.

Theorem: Let $A \in \mathbb{R}^{n\times n}$. Then A is invertible if and only if $\det A \neq 0$.

Proof: ($\Rightarrow$) Suppose A is invertible. Then $A \cdot A^{-1} = I_n$.

Thus $\det I_n = 1 = (\det A)(\det A^{-1})$. (This is impossible if $\det A = 0$...)

Thus $\det A \neq 0$.

($\Leftarrow$) Suppose $\det A \neq 0$ and let R be a reduced row echelon form of A. It is not hard to see that $\det R \neq 0$.

Thus R has no zero row and so is invertible.

$\square$.

Remark: In the previous proof, we in fact showed that if A is invertible, then $\boxed{\det A^{-1} = \frac{1}{\det A}}$ ...

For 2×2 matrices, recall that we had a nice formula for the inverse:

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

A similar formula exists for $n\times n$ matrices (but is quite complicated). We will obtain it soon. For now, let us observe that if $A \in \mathbb{R}^{n\times n}$, then:

$$\begin{aligned} Av = \lambda v &\iff \operatorname{null}(A-\lambda I) \neq \{0\} \\ &\iff A-\lambda I \text{ is not invertible} \\ &\iff \det(A-\lambda I) = 0. \end{aligned}$$

Thus the eigenvalues of A are the roots of the polynomial

$$p(\lambda) = \det(A-\lambda I).$$

this is a degree n polynomial so it has exactly n roots (maybe some are complex).

We now examine how to obtain an explicit formula for the inverse of a matrix (and in turn, for the solution of a linear system).

Recall that $A \in \mathbb{R}^{n\times n}$ is invertible $\iff$ $Ax=b$ has a unique sol$^{\underline{n}}$ for all $b \in \mathbb{R}^n$

$\iff \det A \neq 0.$

For 2x2 matrices, we know that

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \qquad \text{if } \det A \neq 0.$$

We will now find a similar formula for $n\times n$ matrices that are invertible.

We first discuss Cramer's rule which provides a closed-form expression for the solution of $Ax=b$.

<u>Notation</u>: Given $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$, we denote by $A_i(b)$ the matrix obtained from A by replacing the i^{th} column of A by b, i.e.,

$$A = \begin{pmatrix} | & & | & | & | & & | \\ c_1 & \cdots & c_{i-1} & c_i & c_{i+1} & \cdots & c_n \\ | & & | & | & | & & | \end{pmatrix} \rightsquigarrow A_i(b) = \begin{pmatrix} | & & | & | & | & & | \\ c_1 & \cdots & c_{i-1} & b & c_{i+1} & \cdots & c_n \\ | & & | & | & | & & | \end{pmatrix}$$

Theorem: (Cramer's rule)

Let $A \in \mathbb{R}^{n \times n}$ be invertible and let $b \in \mathbb{R}^n$. Then the unique solution $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ of $Ax = b$ is given by:

$$\boxed{x_i = \frac{\det A_i(b)}{\det A}} \qquad (i = 1, \ldots, n).$$

Remark: ① Computing each coordinate of the solution x involves computing a determinant.

② Cramer's theorem is very useful to compute the solution of small systems. However, computing large determinants is expensive. Other methods (e.g. Gaussian elimination) are generally used to solve large systems.

③ Cramer's rule provides an explicit solution to a linear system, and is a fundamental theoretical tool.

Proof of Cramer's rule:

There are many known proofs of Cramer's rule. See e.g.

Brunetti, M., "Old and New Proofs of Cramer's Rule", 2014.

We will look at a proof from N. Trudi (1811-1884).
(See book for a different proof).

Let $A = \begin{pmatrix} | & & | \\ c_1 & \cdots & c_n \\ | & & | \end{pmatrix} \in \mathbb{R}^{n \times n}$. Suppose $Ax = b \left(\Leftrightarrow \sum_{i=1}^{n} x_i c_i = b. \right)$

Then $x_i \cdot \det A = \det \begin{pmatrix} | & & | & | & | & & | \\ c_1 & \cdots & c_{i-1} & x_i c_i & c_{i+1} & \cdots & c_n \\ | & & | & | & | & & | \end{pmatrix}$

(since multiplying row by λ mutiplies det by λ).

Now adding a multiple of a column to another column does not change the determinant.

Thus $x_i \det A = \det \begin{pmatrix} | & & | & | & | & & | \\ c_1 & \cdots & c_{i-1} & \boxed{x_1 c_1 + \cdots + x_n c_n} & c_{i+1} & \cdots & c_n \\ | & & | & | & | & & | \end{pmatrix}$

we add $x_1 c_1, x_2 c_2, \ldots, x_{i-1} c_{i-1}, x_{i+1} c_{i+1}, \ldots, x_n c_n$ to the ith column.

But $b = x_1 c_1 + \cdots + x_n c_n$. Thus

$$x_i \det A = \det A_i(b).$$

$\square$.

Example: Use Cramer's rule to solve the system:

$$\begin{cases} x_1 + 2x_2 = 2 \\ -x_1 + 4x_2 = 1 \end{cases}$$

we have $A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$ $b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Now, $\det A = 1\times 4 - 2(-1) = 6 \neq 0$ (So Cramer's rule can be applied...)

We have:

$$x_1 = \frac{\det A_1(b)}{\det A} = \frac{1}{6}\begin{vmatrix} 2 & 2 \\ 1 & 4 \end{vmatrix} = \frac{1}{6}(8-2) = 1.$$

$$x_2 = \frac{\det A_2(b)}{\det A} = \frac{1}{6}\begin{vmatrix} 1 & 2 \\ -1 & 1 \end{vmatrix} = \frac{1}{6}(1-(-2)) = 1/2.$$

So $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$ is the unique solution of the system.

Exercise: Verify that we get the same solution using Gaussian elimination.

We can now use Cramer's formula to obtain a formula for the inverse of a matrix.

Recall that $C_{ij} = (-1)^{i+j}A_{ij}$ is the (i,j)-th cofactor of A.

Let $C = (C_{ij})_{i,j=1}^{n}$ denote the <u>cofactor matrix</u> of A.

The matrix C^T is called the <u>adjoint</u> (or adjugate) of A and is denoted $\operatorname{adj} A$.

<u>Theorem</u>: Let $A \in \mathbb{R}^{n\times n}$ be invertible. Then:

$$\boxed{A^{-1} = \frac{1}{\det A}\operatorname{adj} A}$$

Remark: $\operatorname{adj} A = \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & & & \vdots \\ \vdots & & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{pmatrix} = C^T.$

Example: 2×2 case: Let $A \in \mathbb{R}^{2\times 2}$ be invertible, say

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

Then $C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$ and so

$$\operatorname{adj} A = C^T = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Example: Use the adjoint to compute the inverse of

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 2 & 4 \\ 1 & 3 & -3 \end{pmatrix}.$$

We compute $\det A = -2$ and:

$C_{11} = +\begin{vmatrix} 2 & 4 \\ 3 & -3 \end{vmatrix} = -18$, $C_{12} = -\begin{vmatrix} 2 & 4 \\ 1 & -3 \end{vmatrix} = 10$, $C_{13} = +\begin{vmatrix} 2 & 2 \\ 1 & 3 \end{vmatrix} = 4$

$C_{21} = -\begin{vmatrix} 2 & -1 \\ 3 & -3 \end{vmatrix} = 3$, $C_{22} = +\begin{vmatrix} 1 & -1 \\ 1 & -3 \end{vmatrix} = -2$, $C_{23} = -\begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix} = -1$

$C_{31} = +\begin{vmatrix} 2 & -1 \\ 2 & 4 \end{vmatrix} = 10$, $C_{32} = -\begin{vmatrix} 1 & -1 \\ 2 & 4 \end{vmatrix} = -6$, $C_{33} = +\begin{vmatrix} 1 & 2 \\ 2 & 2 \end{vmatrix} = -2$.

Thus $C = \begin{pmatrix} -18 & 10 & 4 \\ 3 & -2 & -1 \\ 10 & -6 & -2 \end{pmatrix}$. $\operatorname{adj} A = C^T = \begin{pmatrix} -18 & 3 & 10 \\ 10 & -2 & -6 \\ 4 & -1 & -2 \end{pmatrix}$

Therefore: $$A^{-1} = \frac{1}{\det A} \operatorname{adj} A = -\frac{1}{2} \begin{pmatrix} -18 & 3 & 10 \\ 10 & -2 & -6 \\ 4 & -1 & -2 \end{pmatrix} = \begin{pmatrix} 9 & -3/2 & -5 \\ -5 & 1 & 3 \\ -2 & 1/2 & 1 \end{pmatrix}.$$

Remark: The adjoint method is an important theoretical tool, but is not the fastest way to inverse a matrix.

A better approach to invert matrices is to use Gaussian elimination.

Idea: Let $A \in \mathbb{R}^{n \times n}$ and consider the "augmented matrix"

$$A' = (A \mid I_n).$$

① Perform row operations on A' until the left part of the augmented matrix becomes the identity I_n.

② The right part of the augmented matrix is now A^{-1}.

(See example 3.30 in book).

Proof of the adjoint formula:

$$B \text{ is the inverse of } A \iff AB = I_n.$$

Idea: Use Cramer's rule to compute each column of B.

$$\begin{pmatrix} & & \\ & A & \\ & & \end{pmatrix}\begin{pmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

$$\iff \begin{pmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_n \\ | & | & & | \end{pmatrix}$$

So B satisfies $Ab_j = e_j$.

Note that $b_{ij} = i^{th}$ component of b_j.

Using Cramer's rule:

$$b_{ij} = \frac{\det A_i(e_j)}{\det A}$$

But

$$\det A_i(e_j) = \det \begin{array}{c} \begin{pmatrix} a_{11} & a_{12} & 0 & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & & \vdots \\ & & 0 & & \\ & & 1 & \cdots & \\ a_{n1} & a_{n2} & 0 & & \\ & & 0 & \cdots & a_{nn} \end{pmatrix} \\ \uparrow \\ i\text{th column} \end{array} \begin{array}{l} \\ \\ \\ j^{th} \text{ row} \\ \\ \\ \\ \\ \end{array}$$

Expanding the determinant along the i^{th} column, we obtain:

$$\det A_i(e_j) = (-1)^{i+j} A_{ji}.$$

So $b_{ij} = \dfrac{C_{ji}}{\det A}$ and $B = \dfrac{C^T}{\det A} = \dfrac{\operatorname{adj} A}{\det A}$.

$\square$.

Invertibility, determinants, and eigenvalues:

Theorem (Fundamental theorem of invertible matrices)

Let $A \in \mathbb{R}^{n\times n}$. Then the following are equivalent:

a) A is invertible.

b) $Ax = b$ has a unique solution for all $b \in \mathbb{R}^n$.

c) $Ax = 0_{n\times 1}$ has only the trivial solution $x = 0_{n\times 1}$.

d) The reduced row echelon form of A is I_n.

e) $\operatorname{rank}(A) = n$.

f) $\operatorname{nullity}(A) = 0$.

g) The columns of A are linearly independent.

h) $\operatorname{col} A = \mathbb{R}^n$

i) The columns of A form a basis of $\mathbb{R}^n$.

j) The rows of A are linearly independent

k) $\operatorname{row} A = \mathbb{R}^n$

l) The rows of A form a basis of $\mathbb{R}^n$

m) $\det A \neq 0$

n) 0 is not an eigenvalue of A.

4.4. Similarity and diagonalization:

Change of basis.

Let $\{b_1, \dots, b_n\}$ be a basis of $\mathbb{R}^n$ (for example, $b_i = e_i$ as usual).

Every vector $x \in \mathbb{R}^n$ can be decomposed as.

$$x = x_1 b_1 + \dots + x_n b_n \quad \text{for some } x_i \in \mathbb{R}.$$

What if we want to expand x in another basis? [For example, we could try to construct a basis of eigenvectors of a given matrix A, and represent x in that basis.]

Suppose $\{b_1', \dots, b_n'\}$ is another basis of $\mathbb{R}^n$. Then

$$x = x_1' b_1' + \dots + x_n' b_n'.$$

What is the relation between x_i and x_i'? $P = (p_{ji})$.

Suppose
$$\begin{aligned} b_1' &= p_{11} b_1 + \dots + p_{n1} b_n \\ &\vdots \\ b_n' &= p_{1n} b_1 + \dots + p_{nn} b_n \end{aligned} \quad \text{, i.e., } \begin{pmatrix} b_1' \\ \vdots \\ b_n' \end{pmatrix} = P \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Then
$$\begin{aligned} x &= x_1'(p_{11} b_1 + \dots + p_{1n} b_n) + x_2'(p_{21} b_1 + \dots + p_{2n} b_n) + \dots + x_n'(p_{n1} b_1 + \dots + p_{nn} b_n) \\ &= (p_{11} x_1' + p_{21} x_2' + \dots + p_{n1} x_n') b_1 + \dots + (p_{1n} x_1' + \dots + p_{nn} x_n') b_n \\ &= P \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}. \end{aligned}$$

In other words,

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = P \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}$$

P = change of coordinate matrix

or $$\begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} = P^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

(also called the transition matrix.)

i.e., If $\mathcal{B} = \{b_1, \dots, b_n\}$, $\mathcal{B}' = \{b_1', \dots, b_n'\}$, then

$$[x]_{\mathcal{B}} = P[x]_{\mathcal{B}'} \quad \text{AND} \quad [x]_{\mathcal{B}'} = P^{-1}[x]_{\mathcal{B}}$$

x expressed in the basis $\mathcal{B}$

x expressed in the basis $\mathcal{B}'$.

Now, let $T: \mathbb{R}^n \to \mathbb{R}^n$. We saw that T can be written using a matrix.

Idea: Use a "convenient" basis to write T.

If $x = x_1 b_1 + \dots + x_n b_n$, i.e., $x = [x]_{\mathcal{B}}$, then

$T(x) = x_1 T(b_1) + \dots + x_n T(b_n)$, i.e.

$$T(x) = \begin{pmatrix} | & & | \\ T(b_1) & \cdots & T(b_n) \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} | & & | \\ T(b_1) & \cdots & T(b_n) \\ | & & | \end{pmatrix} [x]_{\mathcal{B}}.$$

We write:

$$M(T)_{\mathcal{B}} = \begin{pmatrix} | & & | \\ T(b_1) & \cdots & T(b_n) \\ | & & | \end{pmatrix}$$ "The matrix of T in the basis $\mathcal{B}$".

Now, we could write T in any basis. If $\mathcal{B}' = \{b_1', \dots, b_n'\}$ is another basis of $\mathbb{R}^n$, what is the relation between $M(T)_{\mathcal{B}}$ and $M(T)_{\mathcal{B}'}$?

It's not very complicated. We have:

$$[T(x)]_{\mathcal{B}} = M(T)_{\mathcal{B}} [x]_{\mathcal{B}}$$

$$[T(x)]_{\mathcal{B}'} = M(T)_{\mathcal{B}'} [x]_{\mathcal{B}'}.$$

But $[T(x)]_{\mathcal{B}} = P[T(x)]_{\mathcal{B}'}$.

So:
$$\begin{aligned} P[T(x)]_{\mathcal{B}'} &= M(T)_{\mathcal{B}} [x]_{\mathcal{B}} \\ &= P M(T)_{\mathcal{B}'} [x]_{\mathcal{B}'} \\ &= P M(T)_{\mathcal{B}'} P^{-1} [x]_{\mathcal{B}}. \end{aligned}$$

Since this is true for all $[x]_{\mathcal{B}}$, we conclude that:

$$M(T)_{\mathcal{B}} = P M(T)_{\mathcal{B}'} P^{-1}.$$

Equivalently,

$$M(T)_{B'} = P^{-1} M(T)_B P.$$

<u>Def</u>: Two matrices $A, B \in \mathbb{R}^{n\times n}$ are said to be <u>similar</u>, if $B = P^{-1}AP$ for some invertible matrix P.

<u>Remark</u>: ① By the above argument, two similar matrices can be seen as representing the same linear transformation in two different basis of $\mathbb{R}^n$.

② If $B = P^{-1}AP$, then $\det B = \det A$.

③ If $B = P^{-1}AP$, then the eigenvalues of B are the same as the eigenvalues of A. Indeed:

$$\begin{aligned}\det(B-\lambda I) &= \det(P^{-1}AP - \lambda I)\\ &= \det(P^{-1}(A-\lambda I)P)\\ &= \det(A-\lambda I).\end{aligned}$$

Important idea:

Suppose $A \in \mathbb{R}^{n\times n}$ has eigenvalues $\lambda_1, \dots, \lambda_n$ and eigenvectors $v_1, \dots, v_n \in \mathbb{R}^n$. Suppose furthermore that the eigenvectors $\{v_1, \dots, v_n\}$ form a basis of $\mathbb{R}^n$ (not always true)

Let $\mathcal{B} = \{v_1, \dots, v_n\}$. Then

$$A[x]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} [x]_{\mathcal{B}}.$$

Thus, there exists $P \in \mathbb{R}^{n\times n}$ invertible such that

$$P^{-1} A P = D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Def: A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if it is similar to a diagonal matrix D, i.e., if there exists an invertible matrix $P \in \mathbb{R}^{n\times n}$ s.t. $D = P^{-1}AP$.

Theorem: A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if its eigenvectors form a basis of $\mathbb{R}^n$.

Proof: Suppose A has a basis $\{p_1, \dots, p_n\}$ of eigenvectors.

Let $P = \begin{pmatrix} | & & | \\ p_1 & \cdots & p_n \\ | & & | \end{pmatrix}$. Then P is invertible since its columns are lin. independent.

Now $AP = \begin{pmatrix} | & & | \\ Ap_1 & \cdots & Ap_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \lambda_1 p_1 & \cdots & \lambda_n p_n \\ | & & | \end{pmatrix}$

$$= P \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

So $P^{-1}AP = D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$.

Conversely, suppose $P^{-1}AP = D$ for some invertible matrix P and some diagonal matrix D.

Then $AP = \begin{pmatrix} | & & | \\ Ap_1 & \cdots & Ap_n \\ | & & | \end{pmatrix} = DP = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \begin{pmatrix} | & & | \\ p_1 & \cdots & p_n \\ | & & | \end{pmatrix}$

$$= \begin{pmatrix} | & & | \\ d_1 p_1 & \cdots & d_n p_n \\ | & & | \end{pmatrix}.$$

It follows that $Ap_i = \lambda_i p_i$ $(i = 1, \dots, n)$. Also, the p_i's are linearly independent since they are the columns of an invertible matrix P. Thus, they form a basis of $\mathbb{R}^n$.

Example: Let $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{pmatrix}$.

Exercise: $\lambda_1 = \lambda_2 = 1$ $\qquad E_1 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$

$\lambda_3 = 2.$ $\qquad E_2 = \text{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \right\}$

$\Rightarrow$ A is NOT diagonalizable.

Example: Let $A = \begin{pmatrix} -1 & 0 & 1 \\ 3 & 0 & -3 \\ 1 & 0 & -1 \end{pmatrix}$.

$\lambda_1 = \lambda_2 = 0, \quad \lambda_3 = -2.$

$E_0 = \text{span}\left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$

$E_{-2} = \text{span}\left\{ \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix} \right\}$

$\Rightarrow$ The eigenvectors of A form a basis of $\mathbb{R}^3$.

$\Rightarrow$ A is diagonalizable.

Let $P = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 3 \\ 0 & 1 & 1 \end{pmatrix}$.

Then $P^{-1}AP = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}$ (check!)

To decide if a matrix is diagonalizable, we need to verify if its eigenvectors generate $\mathbb{R}^n$. The following results are useful to determine if it's the case.

Def:
- The <u>algebraic multiplicity</u> of an eigenvalue λ of A matrix A is the multiplicity of λ as a root of the polynomial $p(\lambda) = \det(A - \lambda I)$.
- The <u>geometric multiplicity</u> of λ is $\dim E_\lambda$.

Example: If $p(\lambda) = \det(A - \lambda I) = \lambda^3 (\lambda - 1)^2 (\lambda - 2)$, then

$\lambda = 0$ has algebraic multiplicity 3

$\lambda = 1$ — — — — — — 2

$\lambda = 2$ — — — — — — 1.

Theorem: Eigenvectors corresponding to different eigenvalues are linearly independent.

Proof: We proceed by contradiction. Let $\lambda_1, \ldots, \lambda_m$ be distinct eigenvalues of A with associated eigenvectors $v_1, \ldots, v_m$. We need to show $v_1, \ldots, v_m$ are linearly independent. Suppose they are dependent. (We will see that this leads to a contradiction).

Let v_{K+1} be the "first" vector among $v_1, \dots, v_m$ that can be written as a linear combination of the others, i.e.,

$$v_{K+1} = c_1 v_1 + \dots + c_K v_K.$$

(Note that $v_1, \dots, v_K$ are lin. independent since v_{K+1} is the first such vector.)

Now, applying A on both sides of the equation, we get:

$$\lambda_{K+1} v_{K+1} = c_1 \lambda_1 v_1 + \dots + c_K \lambda_K v_K.$$

On the other hand, multiplying by λ_{K+1} on both sides, we get:

$$\lambda_{K+1} v_{K+1} = c_1 \lambda_{K+1} v_1 + \dots + c_K \lambda_{K+1} v_K,$$

thus:

$$c_1 (\lambda_1 - \lambda_{K+1}) v_1 + \dots + c_K (\lambda_K - \lambda_{K+1}) v_K = 0.$$

By lin. independence of $v_1, \dots, v_K$, we conclude that

$$c_1 (\lambda_1 - \lambda_{K+1}) = \dots = c_K (\lambda_K - \lambda_{K+1}) = 0$$

Now, $\lambda_i - \lambda_{K+1} \neq 0$ for $i = 1, \dots, K$ since $\lambda_1, \dots, \lambda_{K+1}$ are distinct! Thus $c_1 = c_2 = \dots = c_K = 0$.

We conclude that $v_{K+1} = 0$. This is a contradiction since v_{K+1} is an eigenvector! Therefore $v_1, \dots, v_m$ are independent □

Important consequence:

If all the eigenvalues of A are distinct, then A is diagonalisable.

[This is true since the n eigenvectors of A are lin. independent by the previous theorem. They thus form a basis of $\mathbb{R}^n$!]

Recall that by the fundamental theorem of algebra, a polynomial of degree n has exactly n roots (counting multiplicity). [Some roots may be complex..]

Thus the sum of the ALGEBRAIC multiplicities of the eigenvalues of $A \in \mathbb{R}^{n \times n}$ is exactly n.

For A to be diagonalizable, the geometric multiplicity of the eigenvalues of A needs to be large enough (in fact, it has to be maximal for each eigenvalue as we we see).

<u>Lemma</u>: Let $A \in \mathbb{R}^{n\times n}$. Then the geometric multiplicity of each eigenvalue λ of A is less than or equal to its algebraic multiplicity.

In other words, $\dim E_\lambda \leq$ algebraic multiplicity of λ.

<u>Proof</u>: See Poole Lemma 4.26.

As a consequence, for A to be diagonalizable, the geometric multiplicity of each eigenvalue of A needs to be equal to its algebraic multiplicity!
(i.e, the "number of eigenvectors" associated to each eigenvalue of A has to be "MAXIMAL").

<u>Theorem</u>: Let $A \in \mathbb{R}^{n\times n}$ have distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Then the following are equivalent.

1. A is diagonalizable.
2. The union of bases of $E_{\lambda_1}, \dots, E_{\lambda_k}$ forms a basis of $\mathbb{R}^n$.
3. The geometric multiplicity of each eigenvalue of A equals its algebraic multiplicity.

Some applications of diagonalization:

① Computing powers and the exponential of a matrix:

Suppose $P^{-1}AP = D$ where D is diagonal, i.e.,

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

Then $A^2 = PD\underbrace{P^{-1}P}_{I}DP^{-1} = PD^2P^{-1}$.

More generally, $A^K = PD^KP^{-1}$ for any $K \geq 1$!

Note: If $D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}$, then $D^K = \begin{pmatrix} d_1^K & & \\ & \ddots & \\ & & d_n^K \end{pmatrix}$.

(very easy to compute)

Example: Let $A = \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}$. Compute A^{10}.

We have $\lambda_1 = -1$, $\lambda_2 = 2$ with corresponding eigenvectors $v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Letting $P = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}$, we therefore have: $P^{-1}AP = D = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$.

Thus $A^{10} = P\begin{pmatrix} (-1)^{10} & 0 \\ 0 & 2^{10} \end{pmatrix}P^{-1}$.

More generally, you can verify that:

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

$$= \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} (-1)^n & 0 \\ 0 & 2^n \end{pmatrix} \begin{pmatrix} 2/3 & -1/3 \\ 1/3 & 1/3 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{2(-1)^n + 2^n}{3} & \frac{(-1)^{n+1} + 2^n}{3} \\ \frac{2(-1)^{n+1} + 2^{n+1}}{3} & \frac{(-1)^{n+2} + 2^{n+1}}{3} \end{pmatrix}.$$

<u>Matrix exponential</u>:

We define $e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = \sum_{i=0}^{\infty} \frac{A^i}{i!}$

<u>Note</u>: Here, $A^k = \underbrace{A \cdots A}_{k \text{ times}}$ (matrix product)

One can show that the matrix e^A is well-defined (i.e., the entries of $I + \cdots + \frac{A^n}{n!}$ converge as $n \to \infty$)

We can compute e^A easily if A is diagonalizable:

$$A = PDP^{-1} \Rightarrow A^k = PD^kP^{-1}$$

$$\Rightarrow e^A = P\left(I + D + \frac{D^2}{2!} + \cdots\right)P^{-1} = Pe^DP^{-1}$$

where $e^D = \begin{pmatrix} e^{d_1} & & \\ & \ddots & \\ & & e^{d_n} \end{pmatrix}$, i.e., take exponential of diagonal!

② *Linear dynamical systems:*

Some models (e.g. population evolution) involve systems of the form

$$x_{k+1} = A x_k \qquad k \geq 0 \qquad (x_k \in \mathbb{R}^n)$$

we can determine the long term behavior of x_k easily when A is diagonalizable:

$$A = PDP^{-1}, \qquad D \text{ diagonal}$$

x_0 given.

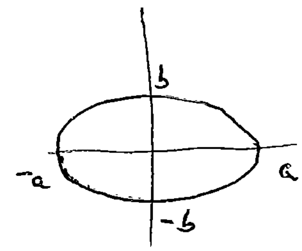
$$x_1 = A x_0$$
$$x_2 = A x_1 = A^2 x_0$$
$$\vdots$$
$$x_k = A^k x_0$$

Now: $A^k = P D^k P^{-1} = P\begin{pmatrix} d_1^k & & \\ & \ddots & \\ & & d_n^k \end{pmatrix} P^{-1}$.

If, for example, $|d_i| < 1$ for all i, then $x_k \to 0_{n\times 1}$ as $n \to \infty$.

③ Diagonalization of quadratic forms:

Recall: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1 \rightsquigarrow$ Ellipse.



What about $5x^2 - 4xy + 2y^2 = 30$?

Note that
$$\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$
$$= \begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} ax+by \\ bx+cy \end{pmatrix}$$
$$= ax^2 + bxy + bxy + cy^2$$
$$= ax^2 + 2bxy + cy^2.$$
← general degree 2 quadratic polynomial in two variables

So $5x^2 - 4xy + 2y^2 = \begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$

Now, for $A = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}$, we have $\lambda_1 = 1$, $\lambda_2 = 6$

So $A = P\begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}P^{-1}$ for some matrix P.

When A is symmetric, we will see that $P^{-1} = P^T$ (next lecture). Thus

$$A = P\begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}P^T.$$

Now
$$(x \ y) A \begin{pmatrix} x \\ y \end{pmatrix} = (x \ y) P \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} P^T \begin{pmatrix} x \\ y \end{pmatrix}$$

$$= \left[P^T\begin{pmatrix} x \\ y \end{pmatrix}\right]^T \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} \left[P^T \begin{pmatrix} x \\ y \end{pmatrix}\right]$$

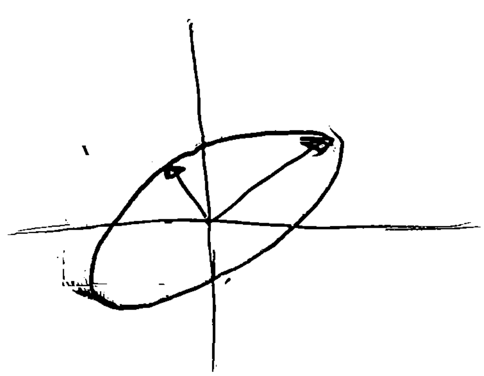
Let $\begin{pmatrix} x' \\ y' \end{pmatrix} = P^T\begin{pmatrix} x \\ y \end{pmatrix}$. Then

$$= (x' \ y')\begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix}.$$

$$= (x')^2 + 6(y')^2$$

So $5x^2 - 4xy + 2y^2 = 30 \iff (x')^2 + 6(y')^2 = 30$

ellipse with the eigenvectors of A as principal axes.



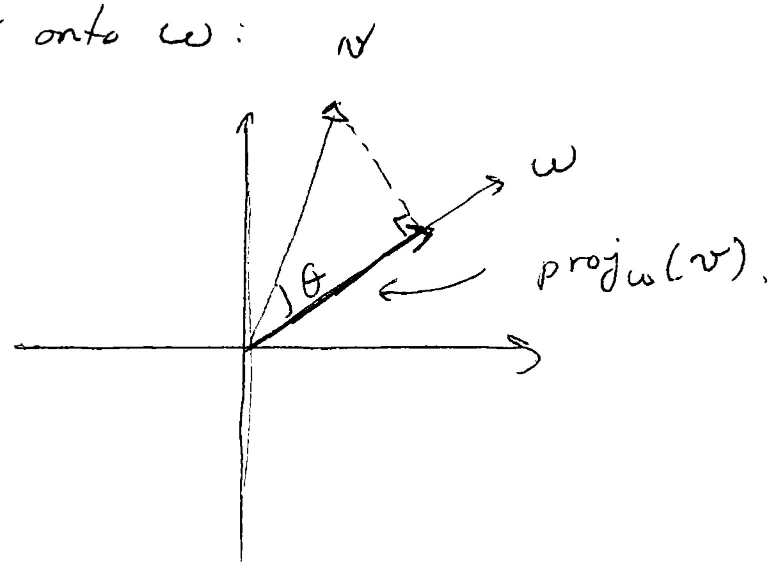
Orthogonality:

Recall that for $v, w \in \mathbb{R}^n$,

$$v \cdot w = \langle v, w \rangle = \sum_{i=1}^{n} v_i w_i = \|v\| \cdot \|w\| \cdot \cos\theta$$

"angle" between v and w.

Projection of v onto w:



We saw before that $\operatorname{proj}_w(v) = \|v\| \cos\theta \cdot \dfrac{w}{\|w\|}$

$$= \frac{v \cdot w}{\|w\|} \cdot \frac{w}{\|w\|}$$

$$= \frac{v \cdot w}{\|w\|^2} \cdot w.$$

The standard basis $\{e_1, \dots, e_n\}$ is orthonormal, i.e.,

$$\begin{cases} e_i \cdot e_j = 0 \quad \forall\, i \neq j & \leftarrow \text{orthogonal} \\ \|e_i\| = 1 \quad \forall\, i. & \leftarrow \text{unit length (normalized)} \end{cases}$$

As we will see, orthonormal bases are very useful.

Example: Let $v_1 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, $v_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.

It is not hard to verify that $\{v_1, v_2, v_3\}$ form an orthogonal basis of $\mathbb{R}^3$.

Remark: Intuitively, a vector cannot be written as a linear combination of vectors that are orthogonal to it.. This idea is formalized in the next theorem.

Theorem: If $\{v_1, \dots, v_K\}$ is an orthogonal set of nonzero vectors, then these vectors are linearly independent.

Proof: Suppose $\lambda_1 v_1 + \dots + \lambda_K v_K = 0_{n\times 1}$

Taking the inner product with v_i, we obtain:

$$\lambda_1 v_i \cdot v_1 + \dots + \lambda_K v_i \cdot v_K = v_i \cdot 0_{n\times 1} = 0_{n\times 1}$$

Now $v_i \cdot v_j = \begin{cases} 0 & \text{if } i \neq j \\ \|v_i\|^2 & \text{if } i = j. \end{cases}$

Thus $\lambda_i \|v_i\|^2 = 0 \implies \lambda_i = 0$ since $\|v_i\| \neq 0$. $\square$

Def^n: An orthogonal (orthonormal) basis for a subspace W of $\mathbb{R}^n$ is a basis that forms an orthogonal (orthonormal) set.

Example: Find an orthogonal basis for the subspace

$$W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : x - y + 2z = 0 \right\} \subseteq \mathbb{R}^3.$$

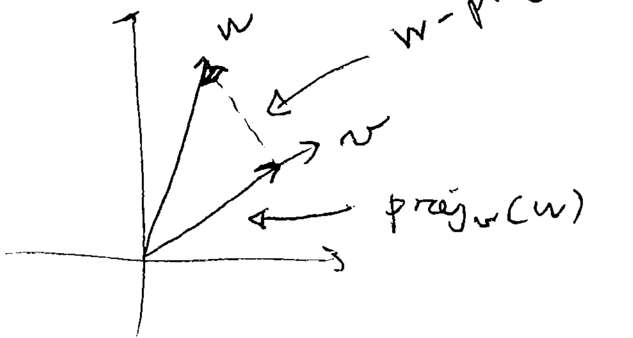
we have $v \in W \iff v = \begin{pmatrix} y - 2z \\ y \\ z \end{pmatrix}$ for some $y, z \in \mathbb{R}$.

$$\therefore \quad = y\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + z\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}.$$

thus $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \right\}$ is a basis of W.

However, $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} = -2 \neq 0.$

Idea:



So v and $w - \mathrm{proj}_v(w)$ are orthogonal...

Let $v = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ $w = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$.

Then $\operatorname{proj}_v(w) = \dfrac{v \cdot w}{\|v\|^2} v$

$= \dfrac{-2}{2} v = -v$

So $w - \operatorname{proj}_v(w) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} =: w'$

Clearly $v \cdot w' = 0$ and $\{v, w'\}$ form a basis of W.

One reason why orthogonal bases are important is because it's easy to find the coordinates of vectors in such bases

Indeed, if $\{v_1, \ldots, v_K\}$ is an orthogonal basis of a subspace $W \subseteq \mathbb{R}^n$, then for any $v \in W$, there exists $\lambda_1, \ldots, \lambda_K$ s.t.

$$v = \lambda_1 v_1 + \cdots + \lambda_K v_K.$$

Now, compute $v \cdot v_i$:

$$v \cdot v_i = \lambda_i \|v_i\|^2.$$ (using the orthogonality of the basis)

Thus λ_i = i^{th} coordinate of v in the basis $\{v_1, \dots, v_k\}$ $= \boxed{\dfrac{v \cdot v_i}{\|v_i\|^2}.}$

In particular, if $\{v_1, \dots, v_k\}$ is an ortho<u>normal</u> basis of W, then $\boxed{\lambda_i = v \cdot v_i}$.

<u>Example</u>: Let $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : x - y + 2z = 0 \right\}$ as in the previous example. Find the coordinates of $\begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} \in W$ in the basis $\mathcal{B} = \left\{ \underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}}_{v_1}, \underbrace{\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}}_{v_2} \right\}$, i.e., find λ_1, λ_2 s.t.

$$\begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} = \lambda_1 v_1 + \lambda_2 v_2.$$

we know:
$$\lambda_1 = \begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} \cdot \frac{v_1}{\|v_1\|^2} = \frac{\begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}}{2}$$

$$= 6/2 = 3$$

$$\lambda_2 = \begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} \cdot \frac{v_2}{\|v_2\|^2} = \frac{\begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}}{3} = -3/3 = -1$$

thus: $\begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} = 3v_1 - v_2$.

Indeed: $3v_1 - v_2 = 3\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix}$. ✓

Orthogonal matrices:

Def: An orthogonal matrix matrix is a matrix whose columns form an orthonormal set.

$$\begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \qquad \{v_1, \dots, v_n\} \text{ orthonormal.}$$

Remark: when changing basis into an orthonormal basis, the passage matrix P is orthogonal.

Note: If $A = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$, then

$$A^T A = \begin{pmatrix} - v_1 - \\ \vdots \\ - v_n - \end{pmatrix} \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} = \begin{pmatrix} v_1 \cdot v_1 & v_1 \cdot v_2 & \cdots & v_1 \cdot v_n \\ v_2 \cdot v_1 & v_2 \cdot v_2 & \cdots & \\ \vdots & & \ddots & \\ v_n \cdot v_1 & & & v_n \cdot v_n \end{pmatrix}$$

thus $\boxed{A \text{ is orthogonal} \iff A^T A = I_n}$ ← Often used as a definition of orthogonal matrices.

Remark: $Q^T Q = I \iff Q^T = Q^{-1}$.

Thus, the inverse of an orthogonal matrix is just its transpose.

Examples: $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

(permute axes) (rotation)

are orthogonal.

Theorem: Let $Q \in \mathbb{R}^{n\times n}$. TFAE:

1. Q is orthogonal.
2. $\|Qx\| = \|x\| \quad \forall x \in \mathbb{R}^n$ (i.e., Q is an isometry)
3. $\langle Qx, Qy\rangle = \langle x, y\rangle \quad \forall x, y \in \mathbb{R}^n$ (Angle preserving)

Proof: we show (1) ⇒ (3) ⇒ (2) ⇒ (3) ⇒ (1)

(1) ⇒ (3). If $Q^T Q = I$, then

$$\langle Qx, Qy\rangle = (y^T Q^T)(Qx) = y^T Q^T Q x$$
$$= y^T x = \langle x, y\rangle$$

(3) ⇒ (2) clear by setting $x = y$.

(2) ⇒ (3) we have $\langle x, y\rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right) \quad \forall x, y \in \mathbb{R}^n$

Thus, if $\|Qx\| = \|x\|$, then

$$\langle Qx, Qx \rangle = \frac{1}{4}\left(\| Q(x+y) \|^2 - \| Q(x-y) \|^2 \right)$$

$$= \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 \right)$$

$$= \langle x, y \rangle$$

That shows $(2) \Rightarrow (3)$.

Finally, if (3) holds, and $Q = \begin{pmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{pmatrix}$, then

$$\langle q_i, q_j \rangle = \langle Qe_i, Qe_j \rangle = \langle e_i, e_j \rangle$$

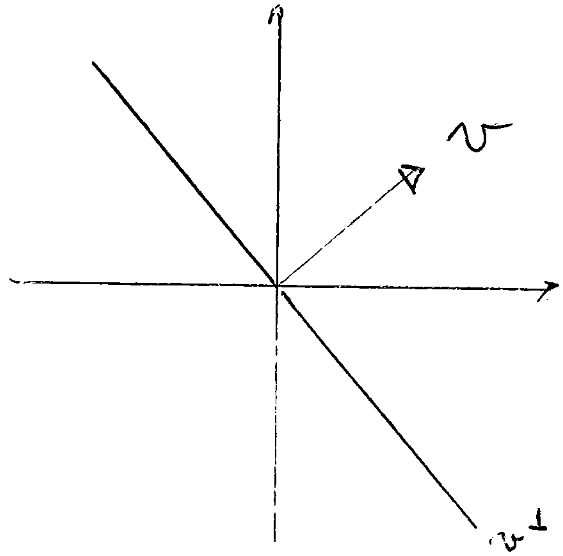
$$= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Thus the columns of Q are orthonormal and so Q is orthogonal.

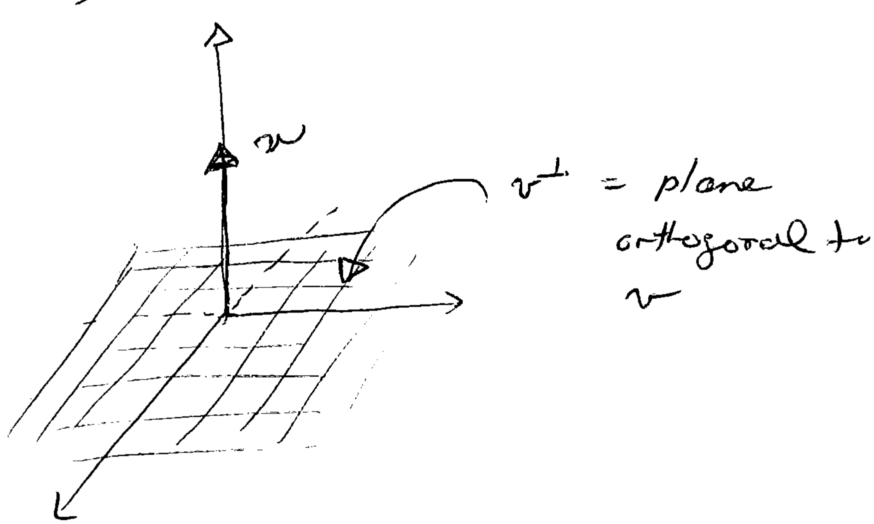
$\square$.

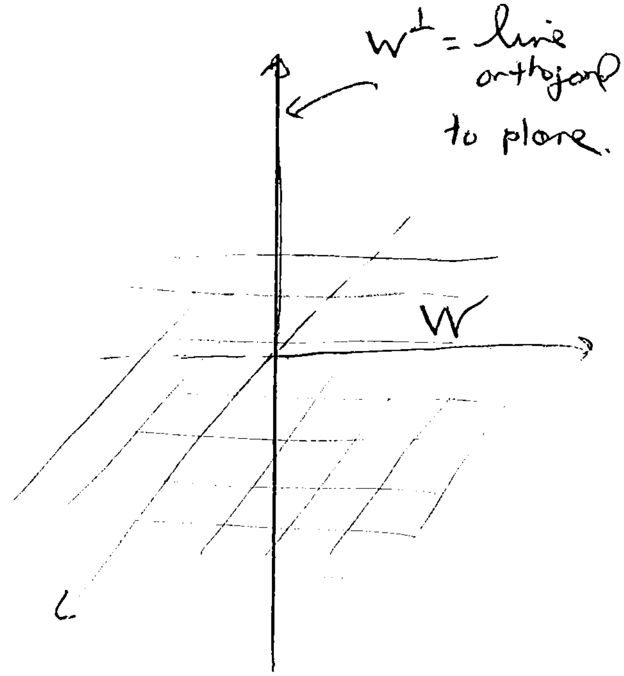
Orthogonal complements and projections:

Given a vector $v \in \mathbb{R}^2$, we can compute a line orthogonal to v:



Similarly, in $\mathbb{R}^3$:





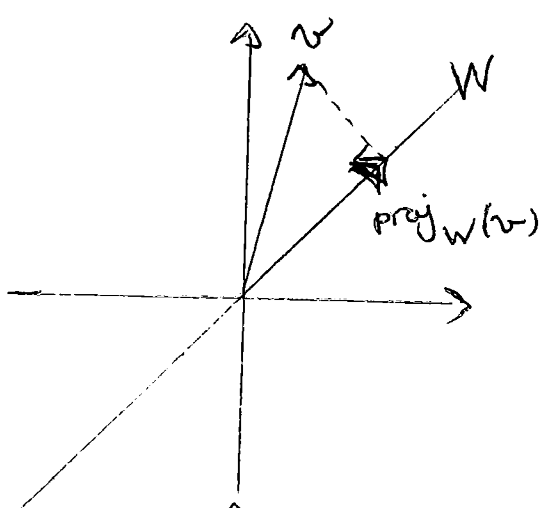
Def^{n}: Let W be a subspace of $\mathbb{R}^n$. We say that $v \in \mathbb{R}^n$ is <u>orthogonal to W</u> if v is orthogonal to every vector $w \in W$. We define the <u>orthogonal complement</u> of W to be the set of vectors orthogonal to W:

$$W^\perp = \{v \in \mathbb{R}^n : v \cdot w = 0 \ \forall w \in W\}.$$

Remark: $W^\perp$ is a subspace of $\mathbb{R}^n$.

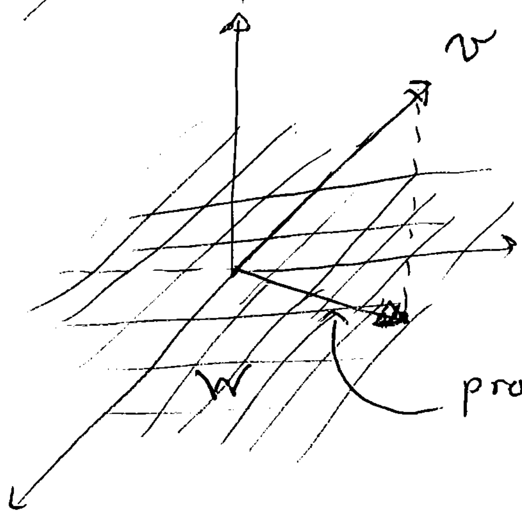
We now generalize the concept of projection to subspaces of $\mathbb{R}^n$.

Idea: In $\mathbb{R}^2$:



Projection of v on line W = best approximation of v by vectors in W.

In $\mathbb{R}^3$



$\mathrm{proj}_W(v)$ = best approximation of v by a vector in the plane W.

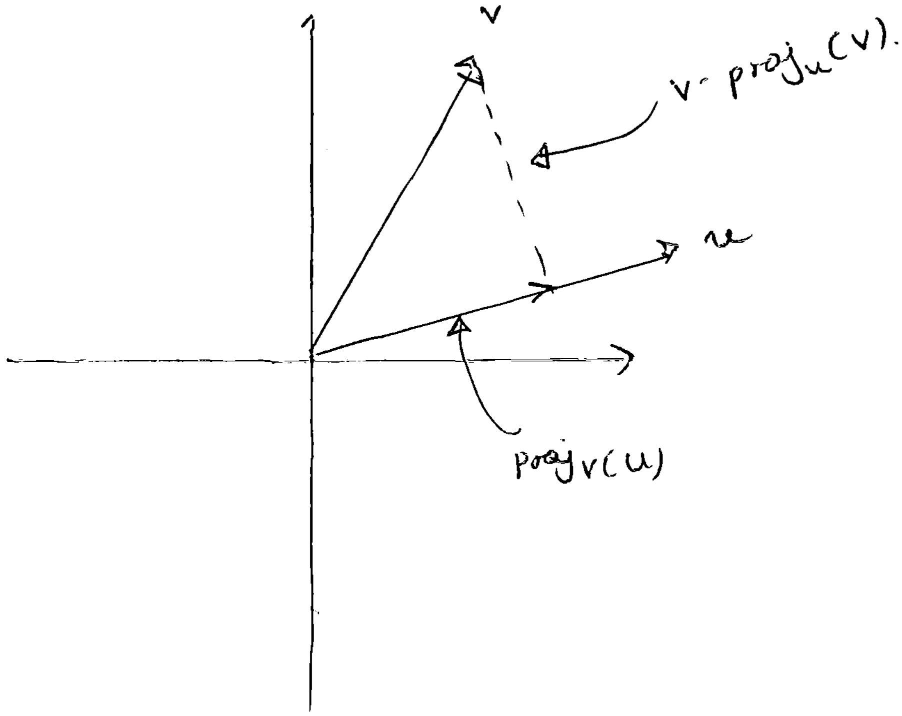
More generally, if $W \subseteq \mathbb{R}^n$ is a subspace, and $v \in \mathbb{R}^n$, then we want to define $\mathrm{proj}_W(v)$ to be a vector such that

$$\|v - \mathrm{proj}_W(v)\| \leq \|v - w\| \quad \forall w \in W.$$

Does such a vector $\mathrm{proj}_W(v)$ exist? Is it unique?

The Gram-Schmidt orthonormalization process.

Recall that for $u, v \in \mathbb{R}^n$, $u, v \neq 0$:



So u and $v - \operatorname{proj}_u(v) = v - \dfrac{u \cdot v}{\|u\|^2} u$ are orthogonal.

Indeed
$$u \cdot (v - \operatorname{proj}_u(v)) = u \cdot \left(v - \frac{u \cdot v}{\|u\|^2} u \right)$$
$$= u \cdot v - \frac{u \cdot v}{\|u\|^2} \underbrace{u \cdot u}_{= \|u\|^2}$$
$$= u \cdot v - u \cdot v$$
$$= 0.$$

Also, note that $v - \dfrac{u \cdot v}{\|u\|^2} u \neq 0$ if $u \not\parallel v$. So the above technique can always be used to "orthogonalize" two vectors.

11/9.

If we want an orthonormal set instead:

$$\frac{u}{\|u\|}, \quad \frac{v - \mathrm{proj}_u(v)}{\|v - \mathrm{proj}_u(v)\|}.$$

The Gram-Schmidt process generalizes this idea.

Gram-Schmidt:

Input: $\{x_1, \ldots, x_K\} \subseteq \mathbb{R}^n$ a basis of a subspace W of $\mathbb{R}^n$.

Output: $\{v_1, \ldots, v_K\} \subseteq \mathbb{R}^n$ an orthonormal basis of W.

Remark: The idea of Gram-Schmidt is to "orthonormalize" a set of vectors without changing their span. This is accomplished by forming linear combinations of $\{x_1, \ldots, x_K\}$.

The Gram-Schmidt process:

$$u_1 := x_1 \qquad v_1 := \frac{u_1}{\|u_1\|}.$$

$$u_2 := x_2 - \mathrm{proj}_{u_1}(x_2) \qquad v_2 := \frac{u_2}{\|u_2\|}.$$

$$u_3 := x_3 - \mathrm{proj}_{u_1}(x_3) - \mathrm{proj}_{u_2}(x_3) \qquad v_3 := \frac{u_3}{\|u_3\|}.$$

$$\vdots$$

$$u_K := x_K - \mathrm{proj}_{u_1}(x_K) - \cdots - \mathrm{proj}_{u_{K-1}}(x_3) \qquad v_K := \frac{u_K}{\|u_K\|}.$$

Recall: $\mathrm{proj}_u(x) = \dfrac{x \cdot u}{\|u\|^2}\, u = x \cdot \left(\dfrac{u}{\|u\|}\right) \cdot \dfrac{u}{\|u\|}$.

As a result, we can rewrite the algorithm in the following way:

$$u_1 := x_1 \qquad v_1 := \frac{u_1}{\|u_1\|}.$$

$$u_2 := x_2 - (x_2 \cdot v_1) v_1 \qquad v_2 = \frac{u_2}{\|u_2\|}$$

$$u_3 := x_3 - (x_3 \cdot v_2) v_2 - (x_3 \cdot v_1) v_1 \qquad v_3 := \frac{u_3}{\|u_3\|}.$$

$$\vdots$$

$$u_k = x_k - (x_k \cdot v_1) v_1 - \cdots - (x_k \cdot v_{k-1}) v_{k-1} \qquad v_k := \frac{u_k}{\|u_k\|}.$$

Example: Construct an orthonormal basis of

$$\mathrm{span}\Bigg(\underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{x_1}, \underbrace{\begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix}}_{x_2}, \underbrace{\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{x_3} \Bigg).$$

Let us first verify if $\{x_1, x_2, x_3\}$ are linearly independent. Note that

$$\det \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = -2 \neq 0$$

Thus the "bottom parts" of $\{x_1, x_2, x_3\}$ are linearly independent. It follows easily that $\{x_1, x_2, x_3\}$ are linearly independent.

we now apply the Gram-Schmidt process:

$$u_1 = x_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad v_1 = \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

$$u_2 = x_2 - (x_2 \cdot v_1) v_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix} - \left[\begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right] \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix} - \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\|u_2\|^2 = \frac{1}{4} + \frac{1}{4} + 1 + 1 = 5/2.$$

$$= \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -1/2 \\ -1 \\ 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ -2 \\ 2 \end{pmatrix}$$

$$v_2 = \frac{u_2}{\|u_2\|} = \frac{\sqrt{2}}{\sqrt{5}} \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ -2 \\ 2 \end{pmatrix} = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ -1 \\ -2 \\ 2 \end{pmatrix}$$

$$u_3 = x_3 - (x_3 \cdot v_1) v_1 - (x_3 \cdot v_2) v_2$$

$$x_3 \cdot v_1 = (0,1,1,1) \cdot \frac{1}{\sqrt{2}} (1,1,0,0) = \frac{1}{\sqrt{2}}$$

$$x_3 \cdot v_2 = (0,1,1,1) \cdot \frac{1}{\sqrt{10}} (1,-1,-2,2) = -\frac{1}{\sqrt{10}}$$

So: $$u_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \frac{1}{\sqrt{10}} \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ -1 \\ -2 \\ 2 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \frac{1}{10} \begin{pmatrix} 1 \\ -1 \\ -2 \\ 2 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 2/5 \\ 4/5 \\ 6/5 \end{pmatrix} = 2/5 \begin{pmatrix} -1 \\ 1 \\ 2 \\ 3 \end{pmatrix}$$

$$\|u_3\|^2 = \frac{4}{25}(1+1+4+9) = \frac{4 \cdot 15}{25} = \frac{12}{5} \qquad v_3 = \frac{2}{5} \frac{\sqrt{5}}{\sqrt{12}} \begin{pmatrix} -1 \\ 1 \\ 2 \\ 3 \end{pmatrix} = \frac{1}{\sqrt{15}} \begin{pmatrix} -1 \\ 1 \\ 2 \\ 3 \end{pmatrix}$$

Conclusion:

$$v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\\0\end{pmatrix} \; ; \; v_2 = \frac{1}{\sqrt{10}}\begin{pmatrix}1\\-1\\-2\\2\end{pmatrix} \; ; \; v_3 = \frac{1}{\sqrt{15}}\begin{pmatrix}-1\\1\\2\\3\end{pmatrix}$$

is an orthonormal basis of the subspace.

Remark: Using Gram-Schmidt, we can convert any basis into an orthonormal basis.

Theorem: Let W be a subspace of $\mathbb{R}^n$. Then

1. $n = \dim W + \dim W^{\perp}$.
2. $\mathbb{R}^n = W \oplus W^{\perp}$ (Direct sum).
3. $W^{\perp\perp} = W$.

Here $\mathbb{R}^n = W \oplus W^{\perp}$ means that every $v \in \mathbb{R}^n$ can be uniquely written as $v = w_1 + w_2$ where $w_1 \in W$ and $w_2 \in W^{\perp}$.

Def: Let $W \subseteq \mathbb{R}^n$ be a subspace and let $\{u_1, \dots, u_k\}$ be an orthogonal basis of W. For any $v \in \mathbb{R}^n$, we define the orthogonal projection of v onto W by:

$$\mathrm{proj}_W(v) = \frac{v \cdot u_1}{\|u_1\|^2} u_1 + \dots + \frac{v \cdot u_k}{\|u_k\|} u_k.$$

Remark: $\text{proj}_W(v) = \text{proj}_{u_1}(v) + \text{proj}_{u_2}(v) + \cdots + \text{proj}_{u_k}(v)$

If $\{u_1, \ldots, u_k\}$ is orthonormal, then

$$\text{proj}_W(v) = (v \cdot u_1)u_1 + \cdots + (v \cdot u_k)u_k.$$

Important properties of orthogonal projections.

① If $v \in W$, then $\text{proj}_W(v) = v$

② If $v \in \mathbb{R}^n$, then $\text{proj}_{W^\perp}(v) = v - \text{proj}_W(v)$

Proof: Let $\{u_1, \ldots, u_k\}$ be an orthonormal basis of W.

We can complete the basis to obtain an orthonormal basis of $\mathbb{R}^n$:

$$\{\underbrace{u_1, \ldots, u_k}_{\text{basis of } W}, u_{k+1}, \ldots, u_n\}.$$

i)

If $v \in W$, then $v = \lambda_1 u_1 + \cdots + \lambda_k u_k$ and

$$\lambda_i = v \cdot u_i \qquad (i = 1, \ldots, k)$$

thus $v = (v \cdot u_1)u_1 + \cdots + (v \cdot u_k)u_k = \text{proj}_W(v)$.

2) Claim: $\{u_{k+1}, \ldots, u_n\}$ forms an orthonormal basis of $W^\perp$.

Clearly $\text{span}\{u_{k+1}, \ldots, u_n\} \subseteq W^\perp$.

Now, If $v \in W^\perp$, then

$$v = \lambda_1 u_1 + \cdots + \lambda_K u_K + \lambda_{K+1} u_{K+1} + \cdots + \lambda_n u_n$$

(Since $\{u_1, \ldots, u_n\}$ is a basis of $\mathbb{R}^n$).

But $\lambda_i = v \cdot u_i$ so $\lambda_1 = \cdots = \lambda_K = 0$.

It follows that every $v \in W^\perp$ can be written as a lin. combination of $\{u_{K+1}, \ldots, u_n\}$.

Now, given $v = \lambda_1 u_1 + \cdots + \lambda_n u_n \in \mathbb{R}^n$,

$$\begin{aligned} \operatorname{proj}_{W^\perp}(v) &= \lambda_{K+1} u_{K+1} + \cdots + \lambda_n u_n \\ &= v - \operatorname{proj}_W(v). \end{aligned}$$

$\square$.

We can now prove the above theorem.

<u>Proof of the theorem:</u>

2. Given $v \in \mathbb{R}^n$, we have

$$v = \underset{\substack{\cap \\ W}}{\operatorname{proj}_W(v)} + \underset{\substack{\cap \\ W^\perp}}{(v - \operatorname{proj}_W(v))}.$$

So every v has the desired decomposition "$W + W^\perp$". We now show this decomposition is unique.

Suppose $v = w_1 + w_1^\perp = w_2 + w_2^\perp$ with $w_1, w_2 \in W$
$w_1^\perp, w_2^\perp \in W^\perp$.

Then $w_1 - w_2 = w_2^\perp - w_1^\perp$.

$\Rightarrow w_1 - w_2 \in W \cap W^\perp = \{0\}$
$w_2^\perp - w_1^\perp \in W \cap W^\perp = \{0\}$.

Thus $w_1 = w_2$ and $w_1^\perp = w_2^\perp$.

1. We need to show $n = \dim W + \dim W^\perp$.

Let $\{u_1, \ldots, u_k\}$ be a basis of W $\quad \dim W = k$

$\{v_1, \ldots, v_l\}$ - - - $W^\perp$. $\quad \dim W^\perp = l$.

Every $v \in \mathbb{R}^n$ can be uniquely written as

$$\lambda_1 u_1 + \cdots + \lambda_k u_k + \mu_1 v_1 + \cdots + \mu_l v_l.$$

It follows that $\{u_1, \ldots, u_k, v_1, \ldots, v_l\}$ is an orthonormal basis of $\mathbb{R}^n$.
Thus $n = k + l$.

3. Let $w \in W$ and $w^\perp \in W^\perp$. Then $w \cdot w^\perp = 0$. Thus $w \in (W^\perp)^\perp$ (since $w \cdot w^\perp = 0 \quad \forall w^\perp \in W^\perp$)

Conversely, suppose $v \in (W^\perp)^\perp$ does not belong to W.

Write $v = \underset{W}{\underset{\in}{w}} + \underset{W^\perp}{\underset{\in}{w^\perp}}$ (unique decomposition)

Since $v \in (W^\perp)^\perp$, $v \cdot w^\perp = 0$. But

$$0 = v \cdot w^\perp = (w + w^\perp) \cdot w^\perp = w^\perp \cdot w^\perp = \|w^\perp\|^2.$$

we conclude $w^\perp = 0$ so $v \in W$.

$\square$

Using the above result, we can revisit the rank-nullity theorem.

Recall that for $A \in \mathbb{R}^{m \times n}$,

$$n = \text{rank } A + \text{nullity } A.$$

In fact, we can say much more...

Theorem: Let $A \in \mathbb{R}^{m \times n}$. Then:

1) $(\text{row } A)^\perp = \text{null}(A)$

2) $(\text{col } A)^\perp = \text{null}(A^T)$.

Proof:

1) Let $A = \begin{pmatrix} - r_1 - \\ - r_2 - \\ \vdots \\ - r_m - \end{pmatrix}$. Note that $Av = \begin{pmatrix} r_1 \cdot v \\ r_2 \cdot v \\ \vdots \\ r_m \cdot v \end{pmatrix}$.

So $v \perp \text{row } A \iff Av = 0 \iff v \in \text{null } A$.

Similarly, if $A = \begin{pmatrix} | & & | \\ c_1 & \cdots & c_n \\ | & & | \end{pmatrix}$, then

$$A^T v = \begin{pmatrix} c_1 \cdot v \\ \vdots \\ c_n \cdot v \end{pmatrix} \quad \text{so} \quad v \perp \operatorname{col} A \iff v \in \operatorname{null}(A^T)$$

$\square$.

Now let $A \in \mathbb{R}^{m \times n}$. Taking $W = \operatorname{row} A$, we get

$$\mathbb{R}^n = \operatorname{row}(A) \oplus \operatorname{null}(A).$$

So every vector in $\mathbb{R}^n$ has a unique decomposition

$$v = v_r + v_n$$

with $v_r \in \operatorname{row}(A)$ and $v_n \in \operatorname{null}(A)$

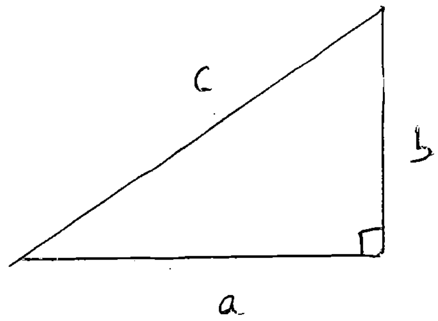
In particular
$$\begin{aligned} n &= \dim \operatorname{row}(A) + \dim \operatorname{null}(A) \\ &= \operatorname{rank}(A) + \operatorname{nullity}(A) \end{aligned}$$

Similarly, $\mathbb{R}^m = \operatorname{col}(A) \oplus \operatorname{null}(A^T)$.

So $m = \operatorname{rank} A + \operatorname{nullity}(A^T)$.

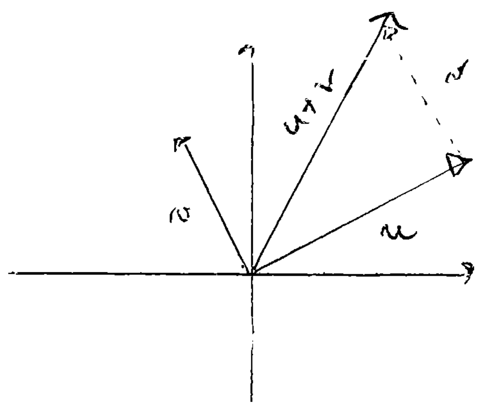
Pythagoras theorem revisited:

Recall:



$$c^2 = a^2 + b^2.$$

In vector form:



$$\|u+v\|^2 = \|u\|^2 + \|v\|^2$$

(Note $u \perp v$ here...)

More generally, we have the following theorem:

Theorem (Pythagoras theorem)

Let $W \subseteq \mathbb{R}^n$ be a subspace and let $v \in \mathbb{R}^n$. Write

$$v = w + w^\perp, \text{ where } w \in W \text{ and } w^\perp \in W^\perp.$$

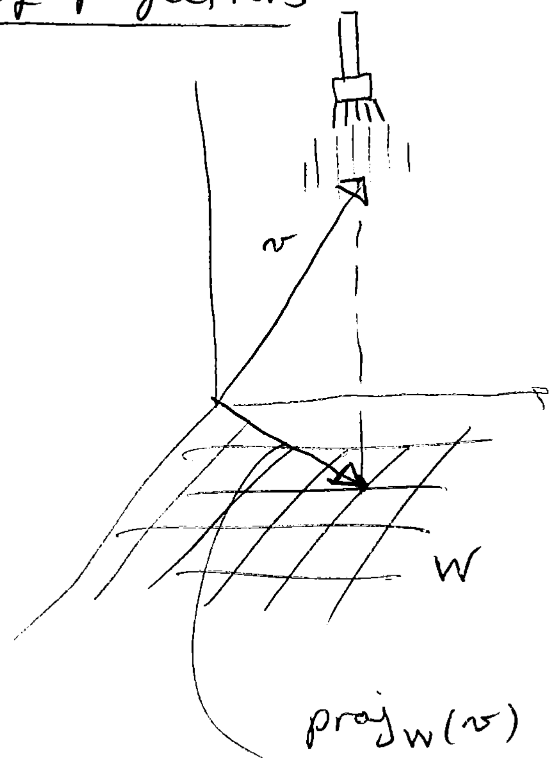
Then $\|v\|^2 = \|w\|^2 + \|w^\perp\|^2.$

Proof: We have
$$\begin{aligned}\|v\|^2 &= v \cdot v \\ &= (w + w^\perp)\cdot(w + w^\perp) \\ &= w\cdot w + \underbrace{w\cdot w^\perp}_{0} + \underbrace{w^\perp\cdot w}_{0} + w^\perp\cdot w^\perp \\ &= \|w\|^2 + \|w^\perp\|^2.\end{aligned}$$

$\square$

Characterization of projectors:

Intuition:



$\text{proj}_W(v)$ = "shade of v on W".

$\text{proj}_W(v)$ is the "best approximation" of v by vectors in W.

The following theorem makes this precise.

Theorem: Let W be a subspace of $\mathbb{R}^n$ and let $v \in \mathbb{R}^n$. Then

$$\underbrace{\|v - \text{proj}_W(v)\|}_{\text{distance between } v \text{ and } \text{proj}_W(v)} \leq \underbrace{\|v - w\|}_{\text{Distance between } v \text{ and } w.} \quad \forall w \in W.$$

Proof: Let $\{u_1, \dots, u_K\}$, $\{u_{K+1}, \dots, u_n\}$ be orthonormal bases of W and $W^\perp$ respectively. Then

$$v = \underbrace{\lambda_1 u_1 + \cdots + \lambda_K u_K}_{v_1 \in W} + \underbrace{\lambda_{K+1} u_{K+1} + \cdots + \lambda_n u_n}_{v_2 \in W^\perp}.$$

for some scalars $\lambda_1, \dots, \lambda_n$. By Pythagoras theorem,

Now. let $w = \mu_1 u_1 + \cdots + \mu_K u_K \in W$.

Then $$\|v - w\|^2 = \|(\lambda_1 - \mu_1)u_1 + \cdots + (\lambda_K - \mu_K)u_K\|^2$$

$$+ \|\lambda_{K+1} u_{K+1} + \cdots + \lambda_n u_n\|^2 \quad \text{(Pythagoras)}$$

$$= \overbrace{(\lambda_1 - \mu_1)^2 + \cdots + (\lambda_K - \mu_K)^2}^{\text{Minimized when } \mu_i = \lambda_i \ (i = 1, \dots, K)}$$

in which case $w = \text{proj}_W(v)$

$$+ \underbrace{\lambda_{K+1}^2 + \cdots + \lambda_n^2}_{\text{No control on that term}}$$

Thus $\|v - \text{proj}_W(v)\| \leq \|v - w\| \quad \forall w \in W.$

$\square$

Example: Compute the projection of $v = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ onto the plane $x+y+z=0$.

Sol^{n}: Let $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : x+y+z=0 \right\}$.

Then $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ -x-y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$.

Thus $W = \text{span}\left(\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right)$. (Moreover, the two vectors are clearly independent).

Let $x_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, $x_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$. Then $x_1 \cdot x_2 = 1 \neq 0$.

Let's orthonormalize the basis:

$$u_1 = x_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad v_1 = \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

$$u_2 = x_2 - (x_2 \cdot v_1) v_1$$

$$= \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} - \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1 \\ -1/2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

$$v_2 = \frac{u_2}{\|u_2\|} = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

So $v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, $v_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$ form an orthonormal basis of the plane!

Recall: $\mathbb{R}^n = W \oplus W^\perp$ $\quad v = \lambda_1 u_1 + \cdots + \lambda_k u_k + \lambda_{k+1} u_{k+1} + \cdots + \lambda_n u_n$

$\{u_1, \ldots, u_k\}$ $\{u_{k+1}, \ldots, u_n\}$ orthonormal bases

Then $\operatorname{proj}_W(v) = \lambda_1 u_1 + \cdots + \lambda_k u_k.$

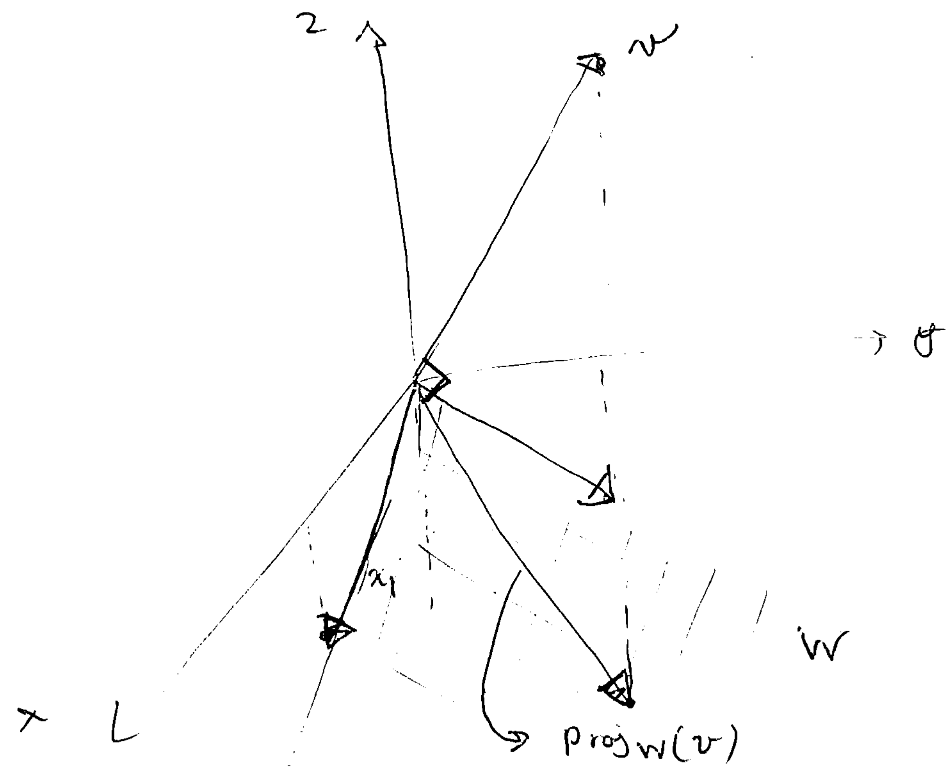
Thus
$$\operatorname{proj}_W(v) = (v \cdot v_1) v_1 + (v \cdot v_2) v_2$$

$$= \left[\begin{pmatrix}1\\2\\2\end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix}\right] \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix} + \left[\begin{pmatrix}1\\2\\2\end{pmatrix} \cdot \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\-1\end{pmatrix}\right] \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\-1\end{pmatrix}$$

$$= \frac{-1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix} + \frac{1}{\sqrt{6}}\,\frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\-1\end{pmatrix}$$

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

$$= \begin{pmatrix}-2/3\\1/3\\1/3\end{pmatrix} = \frac{1}{3}\begin{pmatrix}-2\\1\\1\end{pmatrix}.$$



Least squares approximation:

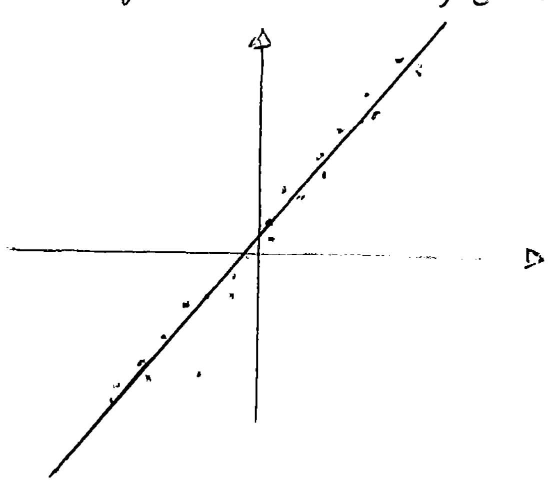
In many applications, data are used to infer a mathematical relationship among variables.

Examples:

- Model the growth of a tree as a function of temperature, moisture, etc.
- Detect relationships between financial instruments.
- Discover products/movies a consumer may like based on previous purchases. (See "The Netflix prize")
- Predict someone's opinion on a given topic given his/her social media posts.

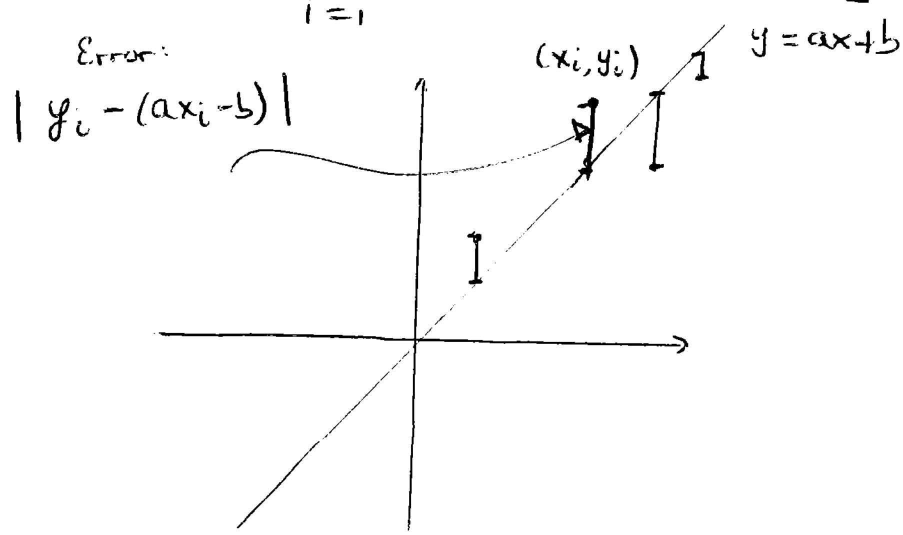
Basic case: Measurements (x_1, y_1), (x_2, y_2), $\vdots$, (x_n, y_n). Model $y_i \approx a x_i + b$.

What are the "best" parameters a, b?



A <u>natural approach</u>: Choose a, b to minimize the total error:

$$L(a,b) = \sum_{i=1}^{n} \left[y_i - (a x_i - b) \right]^2$$



In other words, we want to solve

$$\min_{a,b \in \mathbb{R}} \sum_{i=1}^{n} \left[y_i - (a x_i - b) \right]^2 .$$

How can we solve that problem?

<u>Reformulation</u>: Let $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$.

Note that $A \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a + b x_1 \\ \vdots \\ a + b x_n \end{pmatrix}$.

IF (x_i, y_i) have a <u>perfect</u> linear relationship (i.e., $y_i = a x_i + b$ for all i), then $y = A \begin{pmatrix} a \\ b \end{pmatrix}$. This happens precisely when $y \in \operatorname{col} A$ ... (not very likely to happen in practice...)

If $y \notin \operatorname{col} A$, then we can reformulate our problem as

$$\min_{z \in \mathbb{R}^2} \|y - Az\|^2 \quad \text{where } z = \begin{pmatrix} a \\ b \end{pmatrix}.$$

How can we solve this problem?

Recall: If $W \subseteq \mathbb{R}^n$ is a subspace and $v \notin W$,

we saw that $\operatorname{proj}_W(v)$ is the best approximation

of v be a vector in W, in the sense that:

$$\|v - \operatorname{proj}_W(v)\| \leq \|v - w\| \quad \forall\, w \in W$$

(i.e., $\operatorname{proj}_W(v)$ is the vector in W that is the closest to v).

Now, the above problem is equivalent to find the best approximation of y be vectors in $\operatorname{col} A$ since

$$\operatorname{col} A = \left\{ \lambda_1 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \lambda_1, \lambda_2 \in \mathbb{R} \right\}$$

Thus, the minimum is achieved when $Az = \operatorname{proj}_{\operatorname{col} A} y$!

So, to solve $\min_{z \in \mathbb{R}^2} \|y - Az\|$, we could

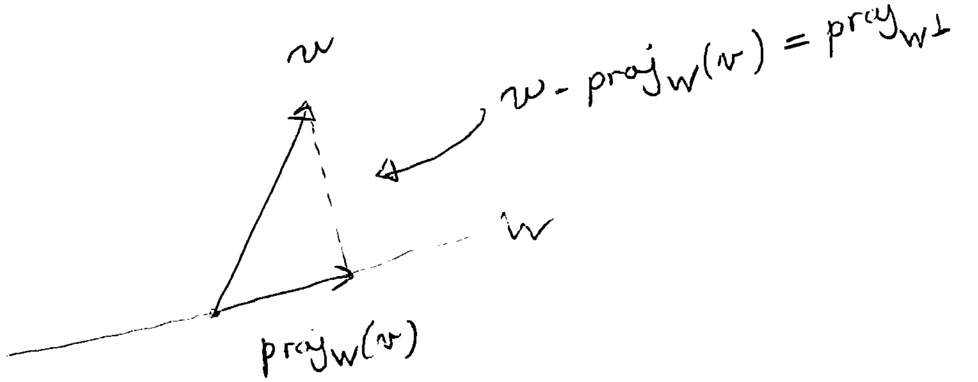
(1) Compute $\text{proj}_{\text{col}A}\, y =: \tilde{y}$.

(2) Solve the system $\tilde{y} = A\tilde{z}$ for $\tilde{z}$.

The resulting $\tilde{z}$ is the solution to our problem.

It works, but there is an easier way to proceed:

Recall:



So $y - \tilde{y} = y - A\tilde{z} = y - \text{proj}_{\text{col}A}\, y$

$$= \text{proj}_{(\text{col}A)^\perp}(y)$$

Recall we saw before that $(\text{col}A)^\perp = \text{null}(A^T)$.

Thus $y - A\tilde{z} = \text{proj}_{\text{null}(A^T)}(y)$.

Thus
$$A^T(y - A\tilde{z}) = 0 \quad \text{since } y - A\tilde{z} \in \text{null}(A^T).$$

Equivalently, $\boxed{A^T A \tilde{z} = A^T y}$ Normal equations for $\tilde{z}$.

So we have proved that if $\tilde{z}$ is the solution of the problem $\min_{z \in \mathbb{R}^2} \|y - Az\|$, then $\tilde{z}$ satisfies the normal equations.

Definition: Let $A \in \mathbb{R}^{m \times n}$ and let $b \in \mathbb{R}^m$. A vector $\tilde{x} \in \mathbb{R}^n$ is said to be a <u>least squares solution</u> of the system $Ax = b$ if
$$\|b - A\tilde{x}\| \leq \|b - Ax\| \quad \forall x \in \mathbb{R}^n.$$

The above reasoning is completely general and shows that a least squares solution of $Ax = b$ satisfies
$$\boxed{A^T A \tilde{x} = A^T b}$$

Conversely, we can work backwards and show that if $\tilde{x}$ satisfies the normal equations, then $\tilde{x}$ is a least squares solution of $Ax=b$.

We have thus established the following theorem:

Theorem: (The least squares theorem):

Let $A \in \mathbb{R}^{m \times n}$ and let $b \in \mathbb{R}^m$. Then $Ax=b$ always has at least one least squares solution $\tilde{x}$. Moreover,

1) $\tilde{x}$ is a least squares solution of $Ax=b$ if and only if $\tilde{x}$ satisfies the normal equations

$$\boxed{A^T A \tilde{x} = A^T b}.$$

2) A has linearly independent columns if and only if $\tilde{x}$ is unique. That happens if and only if A^TA is invertible, in which case the unique least squares solution of $Ax=b$ is given by:

$$\boxed{\tilde{x} = (A^TA)^{-1} A^T b}$$

Moral: We want to solve $Ax=b$ to find relations between variables using data.

- If $b \in \text{col} A$, then the system has a solution (proceed as usual)

- If $b \notin \text{col} A$, the system has no solution. How close to b can Ax be? Solve the system in the "least squares sense":

$$A^T A \tilde{x} = A^T b.$$

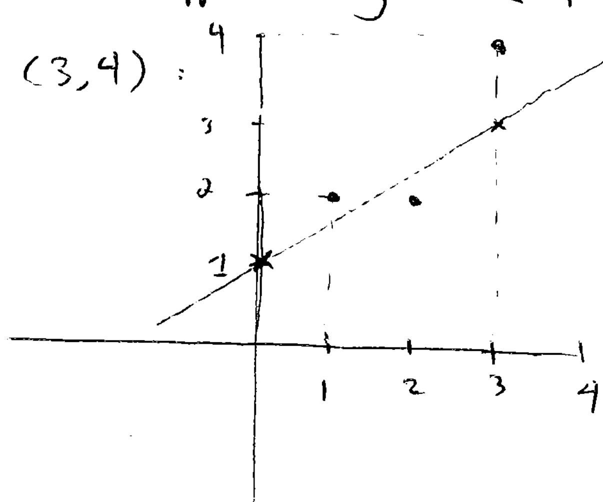
Let $\tilde{b} = A\tilde{x}$. How good is the approximation? Compute:

$$\| b - \tilde{b} \|.$$

If $\tilde{b} \approx b$, then $\tilde{x}$ is an acceptable solution of $Ax=b$ (taking uncertainties into account). If $\|\tilde{b} - b\|$ is "large", there is probably a deeper modeling problem...

Example: Find the best approximating line for the points:

$(1,2), (2,2), (3,4)$:



$y = 2/3\, x + 1.$

(see next page)

Let $b = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$ and $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$.

We want to solve $Ax = b$: $\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$

Note: rank A = dim col A = 2 (the two columns are clearly lin. ind.)

Does $b \in$ col A? Many ways to check:

① Compute det $\begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 3 & 4 \end{pmatrix}$.

② Row reduce $\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 3 & 4 \end{array}\right)$.

③ Solve the least squares system $A^T A \tilde{x} = A^T b$ and compute $A\tilde{x} - b$ ($= 0$ iff $b \in$ col A).

Let's solve the least squares system:

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}.$$

$$A^T b = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 8 \\ 18 \end{pmatrix}.$$

We need to solve:

$$\begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 8 \\ 18 \end{pmatrix}.$$

Note: $\det A^TA = 3\times 14 - 6\times 6 = 6 \neq 0$ so the is a unique least squares solution. (In fact, we already knew that since the columns of A are linearly independent.).

The solution is (check!) $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2/3 \\ 1 \end{pmatrix}$.

Thus $y = 2/3\,x + 1$ is our best approximating line!

What is the total error made by this approximation?

$$\text{Total error (squared)} = \|b - A\tilde{x}\|^2 = \sum_{i=1}^{3} \left[b_i - (ax_i + b)\right]^2.$$

$$A\tilde{x} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 2/3 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/3 \\ 8/3 \\ 11/3 \end{pmatrix} = \begin{pmatrix} 1.66.. \\ 2.66.. \\ 3.66.. \end{pmatrix}$$

$$\|b - A\tilde{x}\|^2 = \left\| \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix} - A\tilde{x} \right\|^2 = \left\| \begin{pmatrix} 1/3 \\ -2/3 \\ 1/3 \end{pmatrix} \right\|^2 = \frac{1}{9} + \frac{4}{9} + \frac{1}{9} = 6/9 = 2/3$$

Diagonalization of symmetric matrices:

Recall: A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if

$$P^{-1}AP = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$ for some invertible $P \in \mathbb{R}^{n\times n}$.

A is diagonalizable $\iff$ $\dim E_\lambda$ = algebraic multiplicity of λ for each eigenvalue λ of A.

In that case, the is a basis of $\mathbb{R}^n$ $\{v_1, \dots, v_n\}$ consisting of eigenvectors of A.

eigenvalues of A

If $P = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$ and $D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$,

then $P^{-1}AP = D$. Equivalently, $A = PDP^{-1}$.

When A is symmetric ($A^T = A$), we can say much more.

Def: A matrix $A \in \mathbb{R}^{n\times n}$ is orthogonally diagonalizable if $Q^TAQ = D$ for some orthogonal matrix Q (i.e., $Q^TQ = I$) and some diagonal matrix D.

Note: If Q is orthogonal, then $Q^{-1} = Q^T$.

Thus $Q^T A Q = Q^{-1} A Q$.

So a matrix A is orthogonally diagonalizable iff the matrix "P" in $P^{-1}AP$ can be chosen to be orthogonal.

Example: Let $A = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}$.

$$p(\lambda) = \det(A - \lambda I) = \begin{pmatrix} 1-\lambda & 2 \\ 2 & -2-\lambda \end{pmatrix} = (1-\lambda)(-2-\lambda) - 4$$
$$= -2 - \lambda + 2\lambda + \lambda^2 - 4$$
$$= \lambda^2 + \lambda - 6$$
$$= (\lambda+3)(\lambda-2)$$

So the eigenvalues of A are $\lambda_1 = 2$, $\lambda_2 = -3$.

It is not hard to verify that $E_{\lambda_1} = \text{span}\left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right)$

$E_{\lambda_2} = \text{span}\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right)$.

Let $u_1 := \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $u_2 := \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

Then $u_1 \cdot u_2 = 0$! Let $v_1 := \dfrac{u_1}{\|u_1\|} = \begin{pmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}$

$v_2 := \dfrac{u_2}{\|u_2\|} = \begin{pmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{pmatrix}$.

Then $\{v_1, v_2\}$ is an orthonormal basis of $\mathbb{R}^2$.

Moreover $E_{\lambda_1} = \text{span}(u_1) = \text{span}(v_1)$

$$E_{\lambda_2} = \text{span}(u_2) = \text{span}(v_2)$$

Let $P = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix} = \begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & -2/\sqrt{5} \end{pmatrix}$.

Then $P^T P = \begin{pmatrix} v_1 \cdot v_1 & v_1 \cdot v_2 \\ v_1 \cdot v_2 & v_2 \cdot v_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$.

Thus P is orthogonal and

$$P^{-1} A P = P^T A P = \begin{pmatrix} 2 & 0 \\ 0 & -3 \end{pmatrix}.$$

In the previous example, the (symmetric) matrix A had orthonormal eigenvectors and, as a result, was orthogonally diagonalizable. This is not a coincidence as we will see.

<u>Theorem</u>: If $A \in \mathbb{R}^{n \times n}$ is symmetric, then its eigenvalues are real.

Recall: If $z = a + bi \in \mathbb{C}$, then $\bar{z} = a - bi$ and $z\bar{z} = |z|^2 = a^2 + b^2$.

For $A \in \mathbb{C}^{m \times n}$, we define $\bar{A} = (\bar{a}_{ij})$.

If $v \in \mathbb{C}^{n \times 1}$, $v_j = a_j + i b_j$, then $\bar{v}^T v = \sum_{j=1}^{n} a_j^2 + b_j^2$

Proof: Let λ be an eigenvalue of A with corresponding eigenvector v, i.e.,

$$A v = \lambda v.$$

Then $\underbrace{A\bar{v} = \overline{Av}}_{\text{since } \bar{A}=A} = \overline{\lambda v} = \bar{\lambda}\,\bar{v}$

Thus $A\bar{v} = \bar{\lambda}\bar{v}$. Taking transpose and using $A^T = A$,

$$\bar{v}^T A = \bar{\lambda}\,\bar{v}^T.$$

Therefore: $\bar{\lambda}\,\underbrace{\bar{v}^T v}_{\substack{\text{real number..}\\ \neq 0 \text{ since } v \neq 0_{n\times 1}}} = \bar{v}^T A v = \bar{v}^T \lambda v = \lambda\,\bar{v}^T v.$

Since v is an eigenvector, $v \neq 0_{n\times 1}$ and

$$\bar{v}^T v = \sum_{j=1}^{n} a_j^2 + b_j^2 \neq 0.$$

Thus $\bar{\lambda} = \lambda$, i.e., $\lambda \in \mathbb{R}$!

$\square$.

We know that the eigenvalues of a symmetric matrix are real

Can we say anything about the eigenvectors ?

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and let v_1, v_2 be two eigenvectors of A corresponding to distinct eigenvalues. Then v_1 and v_2 are orthogonal.

Proof: Suppose
$$Av_1 = \lambda_1 v_1 \qquad \text{for } \lambda_1 \neq \lambda_2.$$
$$Av_2 = \lambda_2 v_2$$

Note: $x \cdot y = x^T y$ for $x, y \in \mathbb{R}^n$

Then:
$$\lambda_1 v_1 \cdot v_2 = (\lambda_1 v_1) \cdot v_2 = (Av_1) \cdot v_2 = (Av_1)^T v_2$$
$$= v_1^T A^T v_2$$
$$= v_1^T A v_2 \qquad (\text{since } A^T = A)$$
$$= v_1^T \lambda_2 v_2$$
$$= \lambda_2 v_1 \cdot v_2.$$

We conclude: $\lambda_1 v_1 \cdot v_2 = \lambda_2 v_1 \cdot v_2$. Since $\lambda_1 \neq \lambda_2$, the only way we can have equality is if $v_1 \cdot v_2 = 0$.

$\square$

Theorem (The spectral theorem)

Let $A \in \mathbb{R}^{n\times n}$. Then A is symmetric if and only if A is orthogonally diagonalizable.

Proof:

($\Leftarrow$) Suppose A is orthogonally diagonalizable, i.e.,

$$Q^T A Q = D \qquad \left(\begin{array}{l} Q^T Q = I \\ D \text{ diagonal} \end{array}\right)$$

Then $A = Q D Q^T$. Thus $A^T = (Q^T)^T D^T Q^T$

$$= Q D Q^T = A.$$

So $A^T = A$.

($\Rightarrow$) Now, suppose $A^T = A$. Let λ_1 be an eigenvalue

(Sketch) of A with corresponding eigenvector v_1, i.e.,

$$A v_1 = \lambda_1 v_1. \quad \left(\begin{array}{c} \text{we can assume } \|v_1\| = 1, \\ \text{otherwise we rescale } v_1 \end{array}\right)$$

Let $v_2, \ldots, v_n$ be vectors (not necessarily eigenvectors) such that $\{v_1, \ldots, v_n\}$ forms an orthonormal basis of $\mathbb{R}^n$.

Let $Q = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}$. Then $Q^T Q = I$ by construction.

Also $\quad AQ = \begin{pmatrix} | & & | \\ Av_1 & \cdots & Av_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \lambda_1 v_1 & Av_2 & \cdots & Av_n \\ | & | & & | \end{pmatrix}$

Recall: $v_2, \ldots, v_n$ may not be eigenvectors...

Now: $$Q^T A Q = \begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{pmatrix} \begin{pmatrix} | & | & & | \\ \lambda_1 v_1 & Av_2 & \cdots & Av_n \\ | & | & & | \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 & v_1^T A v_2 & \cdots & v_1^T A v_n \\ 0 & & & \\ \vdots & & \boxed{*} & \\ 0 & & & \end{pmatrix}$$

$(n-1)\times(n-1)$

we compute $$v_1^T A v_2 = v_1 \cdot (Av_2) = (Av_2)\cdot v_1$$
$$= (Av_2)^T v_1$$
$$= v_2^T A v_1$$
$$= v_2^T \lambda_1 v_1$$
$$= 0 \qquad \text{since } v_1 \cdot v_2 = 0$$

Similarly, $v_1^T A v_i = 0 \quad (i = 2, \ldots, n)$. Thus

$$Q^T A Q = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boxed{B} & \\ 0 & & & \end{pmatrix} \quad \text{where } \in \mathbb{R}^{(n-1)\times(n-1)}$$

and $B^T = B$. Repeat the same process... (Use induction for a rigorous proof...) $\square$

In practice, how do we diagonalize symmetric matrices?

According to the spectral theorem, every symmetric matrix is diagonalizable. Thus

$$\dim E_{\lambda_i} = \text{algebraic multiplicity of } \lambda_i \quad \forall i.$$

We proceed as follows: Let $A \in \mathbb{R}^{n\times n}$ be symmetric.

1. Find the eigenvalues $\lambda_1, \dots, \lambda_k$ of A.

2. For each i, find a basis of E_{λ_i}.

3. Use Gram-Schmidt to orthonormalize each basis found in 2,

4. The (orthonormal) basic vectors together form an orthonormal basis $\{v_1, \dots, v_n\}$ of $\mathbb{R}^n$.

   Let $Q = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$. Then

$$Q^T A Q = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Example: Let $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$

The eigenvalues of A are $\lambda_1 = 4$, $\lambda_2 = 1$. and

$$E_4 = \operatorname{span}\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right)$$

$$E_1 = \operatorname{span}\left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}\right).$$

We need to orthonormalize the bases of the eigenspaces.

$\underline{E_4}$ $\quad b_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. $\checkmark$

$\underline{E_1}$: Let $x_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$, $x_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$. We apply Gram-Schmidt:

$$v_1 = \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$u_2 = x_2 - \langle x_2, v_1 \rangle v_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1 \\ -1/2 \end{pmatrix}$$

$$v_2 = \frac{u_2}{\|u_2\|} = \sqrt{\frac{2}{3}}\begin{pmatrix} -1/2 \\ 1 \\ -1/2 \end{pmatrix} = \frac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}.$$

$\{v_1, v_2\}$ is an orthonormal basis of E_1. (Check!)

It follows that $\{b_1, v_1, v_2\}$ forms an orthonormal basis of $\mathbb{R}^3$.

Let $Q = \begin{pmatrix} | & | & | \\ b_1 & v_1 & v_2 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \end{pmatrix}$.

Then $Q^T A Q = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ (Check!).

An introduction to vector spaces:

All the results we have seen so far concern vectors in $\mathbb{R}^n$. However, they hold in much more generality.

- Real numbers (scalars), can be replaced by elements in a <u>field</u>.
- vectors in $\mathbb{R}^n$ can be replaced by general abstract vectors

Fields: A field is an algebraic structure that mimics properties of the real numbers

Let G be a set with a "multiplication" defined on it:

$$* : G \times G \to G$$
$$(a,b) \mapsto a*b$$

we say that $(G,*)$ is a <u>group</u> if

(1) $a*(b*c) = (a*b)*c \quad \forall a,b,c \in G$

(2) $\exists\, e \in G$ s.t. $e*g = g*e = g \quad \forall g \in G$

e is called the identity element (or neutral element)

(3) For every $a \in G$, there exists $b \in G$ s.t.

$$a*b = b*a = e.$$

we denote $b = a^{-1}$.

If $a * b = b * a \quad \forall a, b \in G$, we say that the group is Abelian (or commutative).

integers addition

<u>Examples:</u> $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$

$(\mathbb{R}, +)$, $(\mathbb{C}, +)$

$(\mathbb{N}, +)$ is NOT a group since $-x \notin \mathbb{N}$ for all $x \in \mathbb{N}$...

special linear group.

The set $SL_n(\mathbb{R}) := \{A \in \mathbb{R}^{n \times n} : \det A = 1\}$ with matrix multiplication is a group.

Proof. Clearly if $A, B \in SL_n(\mathbb{R})$, then $\det(AB) = \det A \cdot \det B = 1$.

Thus $AB \in SL_n(\mathbb{R})$ if $A, B \in SL_n(\mathbb{R})$.

(1) is clear.

(2) $I_n \in SL_n(\mathbb{R})$ satisfies $I_n \cdot A = A \cdot I_n = A$. ✓

(3) Every $A \in SL_n(\mathbb{R})$ is invertible (since $\det A \neq 0$).

Moreover $\det(A \cdot A^{-1}) = \underset{=1}{\det A} \cdot \det A^{-1} = \det I = 1$

So $\det(A^{-1}) = 1$. Therefore $A^{-1} \in SL_n(\mathbb{R})$.

$\square$.

Let R be equiped with two laws : $+$, $\cdot$

$$+ : R \times R \to R$$

$$\cdot : R \times R \to R$$

We say that R is a (unitary) ring if:

(1) $(R, +)$ is a Abelian group (with identity 0).

(2) The multiplication satisfies.

a) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$

b) $a \cdot (b + c) = a \cdot b + a \cdot c$

$(a+b) \cdot c = a \cdot c + b \cdot c$

(3) There exists an element denoted $1 \in R$ s.t.

$$1 \cdot a = a \quad \forall a \in R.$$

The ring is <u>commutative</u> if $a b = b \cdot a \ \forall a, b \in R$

<u>Note</u>: we are <u>NOT</u> assuming elements of R have multiplicative inverses.

A <u>field</u> $(K, +, \cdot)$ is a commutative ring such that every $k \in K \setminus \{0\}$ has a multiplicative inverse, i.e.,

if $k \neq 0$, then there exists k' s.t. $k \cdot k' = 1$.

Examples: $(\mathbb{Q}, +, \cdot)$ is a field.

$(\mathbb{R}, +, \cdot)$ - - -

$(\mathbb{C}, +, \cdot)$ - - -

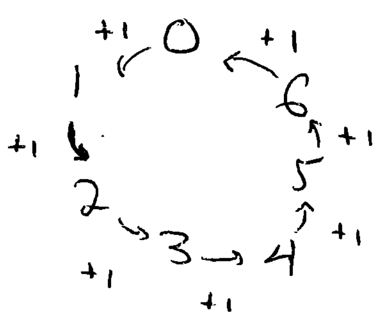
$(\mathbb{Z}, +, \cdot)$ is a ring but NOT a field ($\frac{1}{x} \in \mathbb{Z}$ in general...)

Other fields?

Let p be a prime number. We let $\mathbb{Z}_p = \{0, 1, \dots, p-1\}$

where addition and multiplication are defined "modulo p".

Example: $\mathbb{Z}_7 = \{0, 1, 2, 3, 4, 5, 6\}$



$6 + 1 = 0$

$6 + 2 = 1$

$\vdots$

etc.

$2 \cdot 4 = 8 = 1$

When we add/multiply two numbers, we ignore multiples of p.

One can show that $\mathbb{Z}_p$ is a field if p is prime!

Example: Table of multiplication

$\mathbb{Z}_2 = \{0, 1\}$

	0	1
0	0	0
1	0	1

$\mathbb{Z}_3 = \{0, 1, 2\}$

	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

$1 \cdot 1 = 1$ so $1^{-1} = 1$

$2 \cdot 2 = 1$ so $2^{-1} = 2$.

Vector space: Let K be a field

A set V is said to be a K vector space (or a vector space over K) if two operations are defined on V:

1) An addition of vectors $+ : V \times V \to V$

2) A scalar multiplication $\cdot : K \times V \to V$

(Note: $+$ and $\cdot$ are NOT the operations on K).

and these operations satisfy:

(1) $u + (v + w) = (u + v) + w \quad \forall u, v, w \in V$

(2) $u + v = v + u \quad \forall u, v \in V$

(3) $\exists\, 0 \in V$ s.t. $0 + v = v + 0 = v \quad \forall v \in V$

(4) $\forall v \exists v$, $\exists -v \in V$ s.t. $v + (-v) = 0$

(5) $\lambda(\mu v) = (\lambda\mu) v \quad \forall \lambda, \mu \in K, \forall v \in V$

(6) $1 v = v \quad \forall v \in V$ (1 is the multiplicative identity of K)

(7) $\lambda(u+v) = \lambda u + \lambda v \quad \forall \lambda \in K, \forall u, v \in V$

(8) $(\lambda+\mu) v = \lambda v + \mu v \quad \forall \lambda, \mu \in K$
$\forall v \in V$.

Examples:

① $\mathbb{R}^n$ is a vector space over $\mathbb{R}$.
$\mathbb{Z}_p^n$ is a vector space over $\mathbb{Z}_p$

② Let $\mathcal{P}_2 := \{ a_0 + a_1 x + a_2 x^2 : a_0, a_1, a_2 \}$,

the set of polynomials of degree at most 2.

$$p(x) = a_0 + a_1 x + a_2 x^2$$

$$q(x) = b_0 + b_1 x + b_2 x^2.$$

$$\begin{cases} p(x) + q(x) = (a_0+b_0) + (a_1+b_1) x + (a_2+b_2) x^2 \\ \lambda p(x) = \lambda a_0 + \lambda a_1 x + \lambda a_2 x^2 \end{cases}$$

$\mathcal{P}_2$ is a vector space over $\mathbb{R}$.

③ The set of continuous functions on an interval $[a,b]$.

④ The set $\mathbb{R}^{m\times n}$ of $m\times n$ matrices is a vector space over $\mathbb{R}$.

⑤ The set of integrable functions $f: \mathbb{R} \to \mathbb{R}$ s.t.

$$\int_{-\infty}^{\infty} f^2(x)\,dx < \infty.$$

is a vector space over $\mathbb{R}$.

All the ideas discussed during the semester generalize to abstract vector spaces!

A subset $W \subset V$ is a subspace of V if. W is a vector space with respect to the same addition and scalar multiplication as in V.

A basis of a vector space V is a set of linearly independent vectors $S \subseteq V$ s.t. every $v \in V$ can be written as a finite linear combinations of elements of S.

Theorem: Every vector space has a basis.

Every basis has the same number of elements. That number is called the dimension of V.

Note: $\dim V$ can be infinite.

Example: Let $\mathbb{R}[x] = \{$set of polynomials with real coefficients (and arbitrary degree)$\}$.

$\mathbb{R}[x]$ is a vector space.

$\{1, x, x^2, x^3, \dots\}$ is a basis of $\mathbb{R}[x]$.

We have $\dim \mathbb{R}[x] = \infty$.

Example: Let $V = \mathbb{R}^{m\times n}$, the vector space of $m\times n$ matrices.

Let $E_{i,j} = \begin{array}{c c} & \begin{array}{c c c c} & & j & \end{array} \\ i & \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \cdots & & 0 \\ 0 \cdots 0 & 1 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots 0 \end{pmatrix} \end{array}$ entries $= 0$ except $(i,j)^{th}$ entry $= 1$.

Every matrix $A \in \mathbb{R}^{m\times n}$ is equal to $\sum_{i,j} a_{ij} E_{i,j}$.

Not hard to show that $\{E_{ij} : 1 \le i \le m, 1 \le j \le n\}$ is linearly independent. Thus $\{E_{ij}\}$ is a basis of $\mathbb{R}^{m \times n}$.

$$\dim \mathbb{R}^{m \times n} = m \cdot n.$$