Lecture 3D   Thurs 7/6

Today:

- conceptual connection between column and null spaces
- determinant
- pagerank, steady state of state transition system
- Def. eigenvalues and eigenvectors

Let's start with the column space.

Recall: $\mathrm{Col}(A) = \mathrm{span}\{\text{columns of } A\}$

The question we will ask is: what is the set of all possible outputs of A, i.e. $\{A\vec{x} : \vec{x} \in \mathbb{R}^n\}$?

all outputs   all inputs

- the set of all possible outputs of a function is also called the "range" of the function.
  - the set of inputs is called the "domain"
  - you may or may not remember these things from algebra

Ex. $f(x) = 2x$ $f(x) = 2$

domain: $\mathbb{R}$ domain: $\mathbb{R}$

range: $\mathbb{R}$ range: $\{2\}$

Q. So what is the range of a matrix A?

Consider column view of matrix-vector multiplication:

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

input   output

$$\begin{bmatrix} \vec{a_1} \dots \vec{a_n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{a_1} + \dots + x_n\vec{a_n} = \vec{y}$$

So $\vec{y}$ is a linear combination of the columns of A.

- we want to know what all the possible $\vec{y}$'s are.
- If we look across all possible $\vec{x}$'s (all possible inputs), we will get all possible linear combinations of the columns of A as the set of possible outputs.
- in other words, $\underline{\text{Range}(A)} = \text{span}\{\text{cols of } A\} = \underline{\text{Col}(A)}$

how convenient.

- So that's another interpretation of the column space.

Consider a wide matrix

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

output $\in \mathbb{R}^2$

input $\in \mathbb{R}^3$

domain $= \mathbb{R}^3$

Range $(A) = \text{Col}(A) = \mathbb{R}^2$

(in this case)

So A maps from $\mathbb{R}^3$ to $\mathbb{R}^2$.

- $\mathbb{R}^3$ is a "bigger" space than $\mathbb{R}^2$! (larger dimension)
- A is "squishing" a bigger space into a smaller space
  $\Rightarrow$ information is being lost!
  - there are multiple $\vec{x}$'s that map to the same $\vec{y}$
  - for example, there are infinite solutions to $A\vec{x} = 0$: many vectors are mapped to zero.
- in other words, A having a nontrivial nullspace can be interpreted as A being a function that loses information.

- we also said that non-invertible square matrices lose information, and it turns out their nullspaces are always nontrivial too.

In summary, the idea I'm getting at is that the size of the nullspace corresponds to some notion of losing some of the information from the input space.

- Notice that the dimension of the input space equals the number of columns in A
- A matrix that doesn't lose information would have an output space with dimension = number of columns (can't create more information than the input)

Let us formalize this with a handy theorem (that I won't prove).

The **Rank-Nullity Theorem:**

For a matrix A with m columns,

$$\dim(\mathrm{Col}(A)) + \dim(N(A)) = m.$$

The null space EXACTLY "makes up" the dimensionality that the output space Col (A) failed to preserve!

- *isn't that nice.*

Ok I understand if all those conceptual thoughts were confusing. I just think it's very cool so I'm sharing my intuition and understanding with you. You don't have to understand fully (honestly I don't).

Another way to think about the null space:

- Consider $\vec{x_0} \in \mathcal{N}(A)$. So $A\vec{x_0} = \vec{0}$
- Then take some solution $\vec{x}^*$ to $A\vec{x} = \vec{b}$.
- Consider $A(\vec{x}^* + \vec{x_0}) = A\vec{x}^* + A\vec{x_0} = \vec{b} + \vec{0} = \vec{b}$

  $\Rightarrow \vec{x}^* + \vec{x_0}$ is a solution to $A\vec{x} = \vec{b}$!

You can add any null space vector to any solution and get another solution!

- it's like the vectors in the null space are "invisible" to the matrix A!

ok. maybe I've convinced you that the null space is the coolest space.

We can return to more concrete things now.

Why was the theorem called "rank-nullity"?

Def The **rank** of a matrix A is the dimension of its column space.

(you don't need to know this but "nullity" = $\dim(\mathcal{N}(A))$)

The rank of a matrix is something people talk about a lot, so let's do some quick examples:

Ex. $\text{rank}\left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\right) = 2$

Ex. $\text{rank}\left(\begin{bmatrix} 0 & 0 \\ -1 & 2 \end{bmatrix}\right) = 1$

Ex. $\text{rank}\left(\begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 5 \end{bmatrix}\right) = 2$

Ex. $\text{rank}\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & -1 \end{bmatrix}\right) = 2$

Another way to view rank is as the number of linearly independent columns in A.

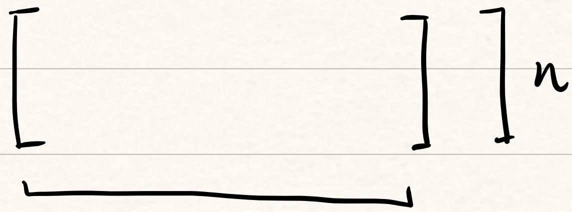
- because that's how many vectors are in a basis for $Col(A)$.

[not in scope, but rank really is an important concept, and as another example, people often talk about "low rank" matrices because if your matrix represents some data you have (e.g. a set of images), then the rank of the matrix actually tells you "how much information" is in the dataset. If there's a lot of redundancy ("low rank") there isn't a lot of information.]

Q. What is the maximum rank of an $n \times m$ matrix?

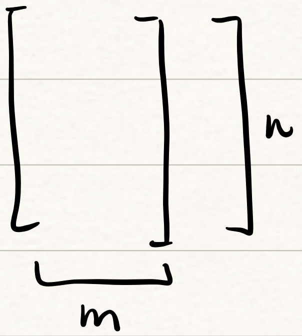
Consider wide and tall cases:

- Wide:



At most, the columns span $\mathbb{R}^n$!

- Tall:



At most, the columns span $\mathbb{R}^m$!

$\Rightarrow$ max rank of $n \times m$ is $\min(n, m)$.

$\boxed{\text{Def}}$ An $n \times m$ matrix is called full rank if its rank is equal to $\min(n, m)$.

- a square matrix that is full rank is invertible!

Let's move on.

We're going to talk about another way to characterize a matrix, called its determinant, written det (A).

- recall: inverse of a 2x2 matrix:

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

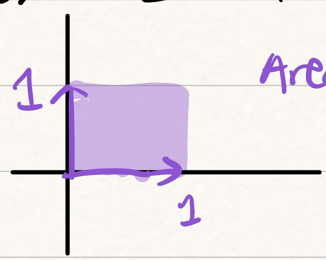
↳ called the determinant

- if det (A) = 0, A is not invertible!

(for all square A of any size)

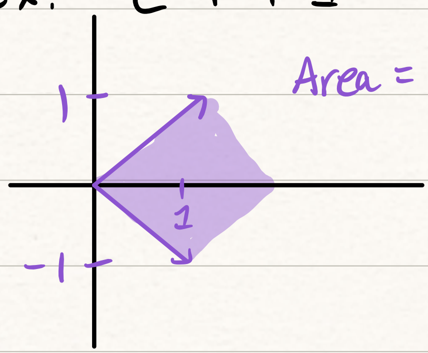
- turns out: geometrically, the determinant is equal to the area of the parallelogram formed by the columns of A

Ex. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$



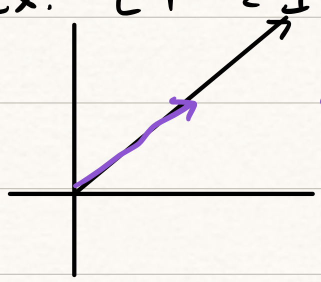
Area = 1 $\qquad ad-bc = 1$ ✓

Ex. $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$



Area = $(\sqrt{2})^2 = 2$ $\qquad ad-bc = 1\cdot 1-(-1)\cdot 1 = 2$

Ex. $\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$



Area = 0 $\qquad ad-bc = 2-2 = 0$

Notably, linearly dependent columns ⇒ det = 0

⇒ not invertible.

In higher dimensions, det = volume (3D) or "hyper-volume" (4D+) formed by column vectors.

- did you know a 3D parallelogram is called a parallelepiped?

We said Col (A) is the set of all possible outputs of A.

- For the last example, Col (A) = span $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ and dim (col (A)) = 1 even though the input dim is 2
  - squishes 2D space into 1D
- this squishing is exactly captured by the zero determinant!
  - if there is squishing, the area/volume will always be 0!

You will see how to compute det for larger matrices in discussion.

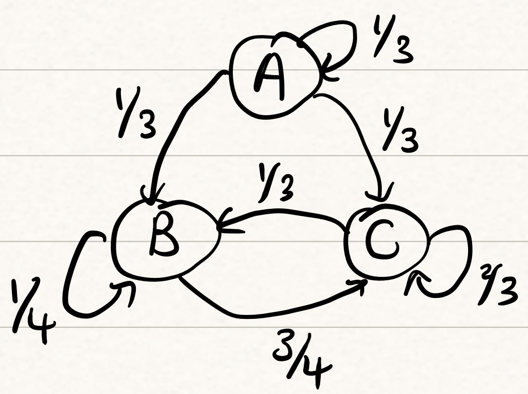
Determinants are useful because it is often an easy way to tell if a matrix is invertible.

- it's just a scalar to compute
- they will also come up to calculate eigenvalues!

We are going to motivate eigenvalues/vectors with a particular example of a state transition system called Page Rank.

- Page Rank is the original Google Search algorithm for ranking/ordering search results
- it was the key to Google's product (at first)
- there's a paper called "The $25B eigenvector: the linear algebra behind Google"

Let me explain the model:



Each circle represents a particular website. Arrows represent hyperlinks between them.

- fractions represent probability people will follow a given link.

Model: internet users randomly move around to these different websites based on these link probabilities.

Q. In the long term, what fraction of users will be at each website?

Imagine you start with some $\vec{x}[0]$.

- you compute $\vec{x}[1] = Q\vec{x}[0]$.
- keep going until $\vec{x}[1000]$.
- Will we converge to some "steady state"?

Ex. Notice above that A loses $\frac{2}{3}$ of its users at every time step, and never gets new users...

- predict : $x_A[\infty] = ?$ probably zero.

"Steady state" would mean advancing time wouldn't change the state : $\underbrace{\vec{x}[t+1] = Q\vec{x}[t] = \vec{x}[t]}$

let's consider this equation

$$Q\vec{x}_s = \vec{x}_s$$

$$Q\vec{x}_s - \vec{x}_s = \vec{0}$$

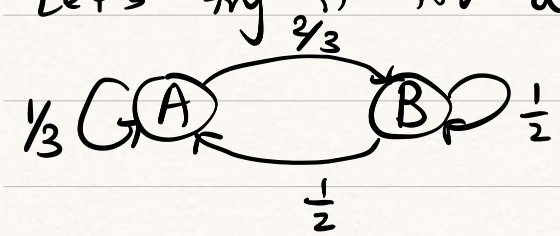
$$Q\vec{x}_s - \underline{I}\vec{x}_s = 0$$

$$(Q - I)\vec{x}_s = 0$$

adding I doesn't change anything, but lets us factor out $\vec{x}_s$ nicely.

Notice: The solutions to this equation are exactly the <u>nullspace</u> of the matrix <u>$Q-I$</u>

Let's try it for a small example



$$Q = \begin{bmatrix} & ? & \end{bmatrix}$$

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

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

Solve for nullspace: $(Q - I)\vec{x}_s = \vec{0}$

$$\left[\begin{array}{cc|c} -2/3 & 1/2 & 0 \\ 2/3 & -1/2 & 0 \end{array}\right] \xrightarrow{R_1 + R_2 \to R_2} \left[\begin{array}{cc|c} -2/3 & 1/2 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

$$-\frac{2}{3}x_1 + \frac{1}{2}x_2 = 0$$
$$\frac{1}{2}x_2 = \frac{2}{3}x_1$$
$$x_2 = \frac{4}{3}x_1$$

$\Rightarrow$ solution set is $\quad \text{span}\left\{\begin{bmatrix} 1 \\ \frac{4}{3} \end{bmatrix}\right\} = \mathcal{N}(Q-I)$

Check: $Q\vec{x_s} = \vec{x_s}$

$$\begin{bmatrix} 1/3 & 1/2 \\ 2/3 & 1/2 \end{bmatrix}\begin{bmatrix} 1 \\ \frac{4}{3} \end{bmatrix} = \begin{bmatrix} 1/3 + 2/3 \\ 2/3 + 2/3 \end{bmatrix} = \begin{bmatrix} 1 \\ 4/3 \end{bmatrix} \checkmark$$

So if we start at $\vec{x}[0] = \vec{x}_{ss}$, the state will never change!

- but how do we know we will converge to $\vec{x}_{ss}$?

 we will answer this next time...

Solving the equation $Q\vec{x} = \vec{x}$ is the key to Page Rank!

- this is a special case of an eigenvalue-eigenvector problem. let's define.

$\boxed{\text{Def}}$ For an $n \times n$ matrix A, if

$$A\vec{v} = \lambda\vec{v} \quad \text{for some } \vec{v} \in \mathbb{R}^n,\ \lambda \in \mathbb{R},$$

then $\vec{v}$ is called an <u>eigenvector</u> of A with <u>eigenvalue</u> λ.

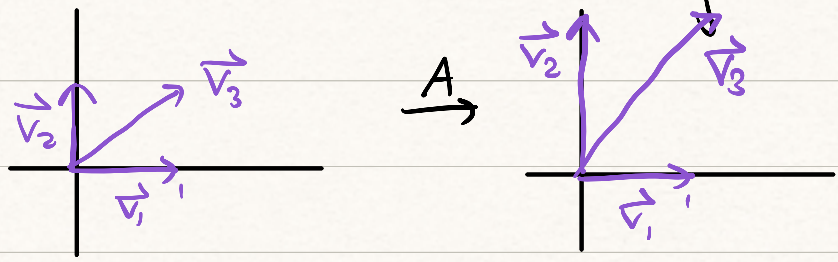
- steady state = eigenvector with eigenvalue 1

Seems like another arbitrary definition... what does it mean?

- The matrix A probably transforms most vectors to some vector pointing in another direction
- But! if $\vec{v}$ is an eigenvector of A, then A just stretches or shrinks $\vec{v}$! still points in same direction

Ex. $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

- what does A do visually? stretches x_2 component $2x$



stretches space vertically

Q. What are the eigenvectors of A?

Which vectors are only stretched/shrunk by A but will still point in the same direction?

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ (and scalar multiples of these)

$\lambda = 1$ $\quad$ $\lambda = 2$

What are their eigenvalues?

- notice that $\vec{v}_3$ changes direction

Maybe it still seems like an arbitrary definition... but honestly eigenvalues/vectors are probably the MOST important linear algebra thing you learn in this class, certainly in Module 1.

They appear EVERYWHERE in every field/problem.

Tomorrow I'll try to explain a little why that is.