UNIT II : Lecture 1

Vectors are easier to understand when they are described in terms of orthogonal bases. In addition, the four fundamental subspaces are orthogonal to each other in pairs.

If A is a rectangular matrix, $Ax=b$ is often unsolvable. The matrix A^TA will help us find a vector $\hat{x}$ that comes as close as possible to solving $Ax=b$

Orthogonal vectors and subspaces

In this lecture we learn what it means for vectors, bases and subspaces to be *orthogonal*. The symbol for this is $\perp$.

The "big picture" of this course is that the row space of a matrix' is orthogonal to its nullspace, and its column space is orthogonal to its left nullspace.

row space	column space
dimension r	dimension r
$\perp$	$\perp$
nullspace	left nullspace $N(A^T)$
dimension $n-r$	dimension $m-r$

Orthogonal vectors

Orthogonal is just another word for *perpendicular*. Two vectors are *orthogonal* if the angle between them is 90 degrees. If two vectors are orthogonal, they form a right triangle whose hypotenuse is the sum of the vectors. Thus, we can use the Pythagorean theorem to prove that *the dot product* $\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}$ *is zero* exactly when $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. (The length squared $||\mathbf{x}||^2$ equals $\mathbf{x}^T\mathbf{x}$.)

Note that all vectors are orthogonal to the zero vector.

Orthogonal subspaces

Subspace S is *orthogonal* to subspace T means: every vector in S is orthogonal to every vector in T. The blackboard is not orthogonal to the floor; two vectors in the line where the blackboard meets the floor aren't orthogonal to each other.

In the plane, the space containing only the zero vector and any line through the origin are orthogonal subspaces. A line through the origin and the whole plane are never orthogonal subspaces. Two lines through the origin are orthogonal subspaces if they meet at right angles.

Nullspace is perpendicular to row space

The row space of a matrix is orthogonal to the nullspace, because $A\mathbf{x} = \mathbf{0}$ means the dot product of $\mathbf{x}$ with each row of A is 0. But then the product of $\mathbf{x}$ with any combination of rows of A must be 0.

The column space is orthogonal to the left nullspace of A because the row space of A^T is perpendicular to the nullspace of A^T.

In some sense, the row space and the nullspace of a matrix subdivide $\mathbb{R}^n$ into two perpendicular subspaces. For $A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix}$, the row space has dimension 1 and basis $\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$ and the nullspace has dimension 2 and is the plane through the origin perpendicular to the vector $\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$.

Not only is the nullspace orthogonal to the row space, their dimensions add up to the dimension of the whole space. We say that the nullspace and the row space are *orthogonal complements* in $\mathbb{R}^n$. The nullspace contains all the vectors that are perpendicular to the row space, and vice versa.

We could say that this is part two of the fundamental theorem of linear algebra. Part one gives the dimensions of the four subspaces, part two says those subspaces come in orthogonal pairs, and part three will be about orthogonal bases for these subspaces.

$N(A^TA) = N(A)$

Due to measurement error, $A\mathbf{x} = \mathbf{b}$ is often unsolvable if $m > n$. Our next challenge is to find the best possible solution in this case. The matrix A^TA plays a key role in this effort: the central equation is $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.

We know that A^TA is square ($n \times n$) and symmetric. When is it invertible?

Suppose $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix}$. Then:

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 30 \end{bmatrix}$$

is invertible. A^TA is not always invertible. In fact:

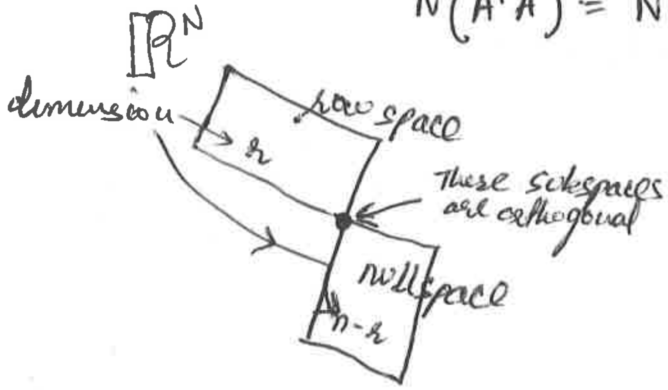
$$\begin{aligned} N(A^TA) &= N(A) \\ \text{rank of } A^TA &= \text{rank of } A. \end{aligned}$$

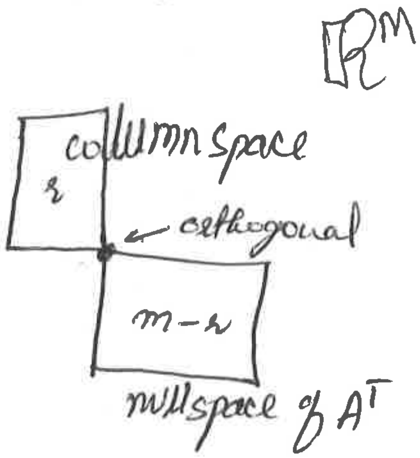
We conclude that A^TA is invertible exactly when A has independent columns.

Big Picture

null space $\perp$ row space
$N(A^TA) = N(A)$

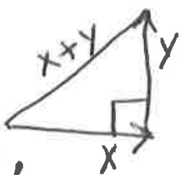
} Subspaces are at an angle of 90°.





Orthogonal vectors

Pythagoras



$x^Ty = 0$ if $\vec{x} \perp \vec{y}$

$\|x\|^2 + \|y\|^2 = \|x+y\|^2$ (when $x \perp y$ or for a right triangle)

$x^Tx + y^Ty = (x+y)^T(x+y)$ [true only for right angle)

$\cancel{x^Tx} + \cancel{y^Ty} = \cancel{x^Tx} + \cancel{y^Ty} + x^Ty + y^Tx$

$x^Ty + y^Tx = 0$ or $x^Ty + x^Ty = 0$

$0 = 0$

$2x^Ty = 0$

$\boxed{x^Ty = 0}$ → $\boxed{\text{For orthogonal vectors}}$

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

$x \perp y$

$\|x\|^2 = 14$ $\|y\|^2 = 5$

$x+y = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}$

$\|x+y\|^2 = 19$

Subspace S is orthogonal to subspace T means:
Every vector in S is orthogonal to every vector in T

Rowspace is orthogonal to nullspace.
Why? $Ax=0$ when x is in nullspace of A.

$$\begin{bmatrix} \text{row 1 of } A \\ \text{row 2 of } A \\ \text{row } m \text{ of } A \end{bmatrix} \begin{bmatrix} X \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

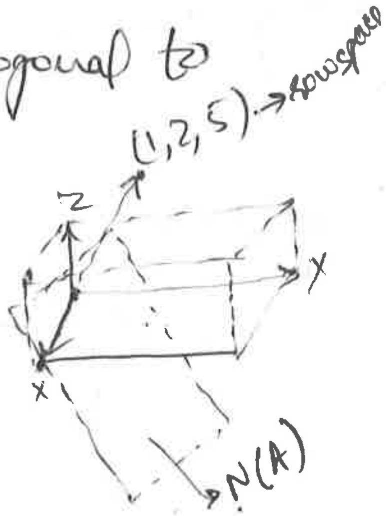
Now we can see that $(\overrightarrow{\text{row 1 of } A}) \cdot (\vec{X}) = 0$

$(\overrightarrow{\text{row 2 of } A}) \cdot (\vec{X}) = 0$

Thus rowspace of A is orthogonal to $N(A)$. To show this, we have to prove that every vector in rowspace of A is $\perp$ to ~~N(A)~~ "X" in $N(A)$.
OR we have to check:

$$\left.\begin{array}{l} (\text{row }1)^T x = 0 \longrightarrow c_1(\text{row}1)^T x = 0 \\ (\text{row }2)^T x = 0 \longrightarrow c_2(\text{row}2)^T x = 0 \end{array}\right\} (c_1 \text{row}1 + c_2 \text{row}2 + \cdots)^T x = 0$$

In the same way, we can show columnspace is orthogonal to nullspace of A^T.



Suppose $A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix}$ $\underbrace{\dim(A^T) = 1}_{\substack{\text{dim of rowspace} = 1 \\ \text{b/c Row2} = 2\,\text{Row1}}}$

$n=3$

rank $=1$

$\dim N(A) =$

for $N(A)$ $A^T x = 0$ → $\dim N(A) = 2$ as $\dim(A^T) = 1$

$$\begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\dim(\text{rowspace}) + \dim(\text{nullspace}) = 3$

Dim $N(A)=2$ means its a plane in 3D where every vector in this plane will be orthogonal to one-dimension rowspace through (1,2,5) or $N(A) = \begin{bmatrix} 1 & 2 & 5 \\ \cancel{2} & \cancel{4} & \cancel{10} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$ Remove R_2 from A as it doesn't have any effect

$\Rightarrow \boxed{x_1 + 2x_2 + 5x_3 = 0}$ this is plane

Nullspace & rowspace are orthogonal complements in R^n
This word "compliments" means that: Nullspace contains all vectors that are $\perp$ to rowspace. Also the dimension of $N(A)$ & rowspace of A should add up to "n"

MAIN Problem of Chapter Coming: $Ax = b$, I would like to solve $Ax = b$ when it is not possible or solve $Ax = b$ when there is no solution. This means "b" is not in column space of "A".

This can happen when $m > n$ (A is rectangular). Rank $< m$
(# of unknowns; A then)
& lots of R.H.S "b" will have ~~free~~ no solution.

If A is $m \times n$ matrix = rectangular matrix
Consider A^TA: this A^TA will be
i) Square ($n \times n$)
ii) Symmetric B/c $(A^TA)^T = A^TA^{TT} = A^TA$
iii) Is it invertible? If not what is it nullspace?

when $Ax = b$ is not solvable then
$A^TAx = A^Tb$ will be a good equation

OR $\boxed{A^TA\hat{x} = A^Tb}$ we hope $\hat{x}$ is solution

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix}$

$m = 3$
$n = 2$
$r = 2$ (rank is 2) $\longrightarrow$ 2 columns are independent

$Ax = b \Longrightarrow x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

We can not solve $Ax = b$. We can only solve for those b's which are in the column space of A.

Now $A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} = \underbrace{\begin{bmatrix} 3 & 8 \\ 8 & 30 \end{bmatrix}}_{\text{It is invertible.}}$

If $A^T A = \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix}}_{\text{rank 1}} \underbrace{\begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix}}_{\text{rank 1}} = \begin{bmatrix} 3 & 9 \\ 9 & 27 \end{bmatrix}$ Now $A^T A$ is not invertible.

Keypoint

$$\left.\begin{aligned} N(A^T A) &= N(A) \\ \text{rank of } A^T A &= \text{rank}(A) \end{aligned}\right\} \quad A^T A \text{ is invertible exactly if "A" has independent columns.}$$

Recitation videos

We often want to find the line (or plane or hyperplane) that best fits our data. This amounts to finding the best possible approximation to some unsolvable system of linear equations $Ax=b$. The algebra of finding these best fit solutions begins with the projection of a vector onto a subspace.

Unit II: Lecture 2

Projections onto subspaces

Projections

If we have a vector **b** and a line determined by a vector **a**, how do we find the point on the line that is closest to **b**?

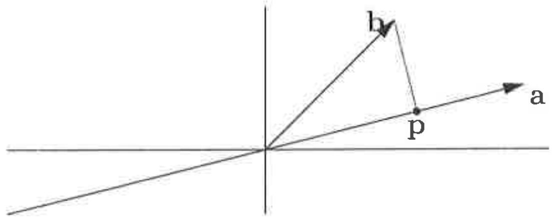


Figure 1: The point closest to **b** on the line determined by **a**.

We can see from Figure 1 that this closest point **p** is at the intersection formed by a line through **b** that is orthogonal to **a**. If we think of **p** as an approximation of **b**, then the length of $\mathbf{e} = \mathbf{b} - \mathbf{p}$ is the error in that approximation.

We could try to find **p** using trigonometry or calculus, but it's easier to use linear algebra. Since **p** lies on the line through **a**, we know $\mathbf{p} = x\mathbf{a}$ for some number x. We also know that **a** is perpendicular to $\mathbf{e} = \mathbf{b} - x\mathbf{a}$:

$$\begin{aligned} \mathbf{a}^T(\mathbf{b} - x\mathbf{a}) &= 0 \\ x\mathbf{a}^T\mathbf{a} &= \mathbf{a}^T\mathbf{b} \\ x &= \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}, \end{aligned}$$

and $\mathbf{p} = \mathbf{a}x = \mathbf{a}\dfrac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$. Doubling **b** doubles **p**. Doubling **a** does not affect **p**.

Projection matrix

We'd like to write this projection in terms of a *projection matrix* P: $\mathbf{p} = P\mathbf{b}$.

$$\mathbf{p} = x\mathbf{a} = \frac{\mathbf{a}\mathbf{a}^T\mathbf{a}}{\mathbf{a}^T\mathbf{a}},$$

so the matrix is:

$$P = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}.$$

Note that $\mathbf{a}\mathbf{a}^T$ is a three by three matrix, not a number; matrix multiplication is not commutative.

The column space of P is spanned by **a** because for any **b**, $P\mathbf{b}$ lies on the line determined by **a**. The rank of P is 1. P is symmetric. $P^2\mathbf{b} = P\mathbf{b}$ because the projection of a vector already on the line through $\mathbf{a}$ is just that vector. In general, projection matrices have the properties:

$$P^T = P \quad \text{and} \quad P^2 = P.$$

Why project?

As we know, the equation $A\mathbf{x} = \mathbf{b}$ may have no solution. The vector $A\mathbf{x}$ is always in the column space of A, and $\mathbf{b}$ is unlikely to be in the column space. So, we project $\mathbf{b}$ onto a vector $\mathbf{p}$ in the column space of A and solve $A\hat{\mathbf{x}} = \mathbf{p}$.

Projection in higher dimensions

In $\mathbb{R}^3$, how do we project a vector $\mathbf{b}$ onto the closest point $\mathbf{p}$ in a plane?

If $\mathbf{a}_1$ and $\mathbf{a}_2$ form a basis for the plane, then that plane is the column space of the matrix $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$.

We know that $\mathbf{p} = \hat{x}_1\mathbf{a}_1 + \hat{x}_2\mathbf{a}_2 = A\hat{\mathbf{x}}$. We want to find $\hat{\mathbf{x}}$. There are many ways to show that $\mathbf{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the plane we're projecting onto, after which we can use the fact that $\mathbf{e}$ is perpendicular to $\mathbf{a}_1$ and $\mathbf{a}_2$:

$$\mathbf{a}_1^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \text{and} \quad \mathbf{a}_2^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0.$$

In matrix form, $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$. When we were projecting onto a line, A only had one column and so this equation looked like: $a^T(\mathbf{b} - x\mathbf{a}) = \mathbf{0}$.

Note that $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ is in the nullspace of A^T and so is in the left nullspace of A. We know that everything in the left nullspace of A is perpendicular to the column space of A, so this is another confirmation that our calculations are correct.

We can rewrite the equation $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ as:

$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}.$$

When projecting onto a line, A^TA was just a number; now it is a square matrix. So instead of dividing by $\mathbf{a}^T\mathbf{a}$ we now have to multiply by $(A^TA)^{-1}$

In n dimensions,

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ \mathbf{p} = A\hat{\mathbf{x}} &= A(A^TA)^{-1}A^T\mathbf{b} \\ P &= A(A^TA)^{-1}A^T. \end{aligned}$$

It's tempting to try to simplify these expressions, but if A isn't a square matrix we can't say that $(A^TA)^{-1} = A^{-1}(A^T)^{-1}$. If A does happen to be a square, invertible matrix then its column space is the whole space and contains $\mathbf{b}$. In this case P is the identity, as we find when we simplify. It is still true that:

$$P^T = P \quad \text{and} \quad P^2 = P.$$

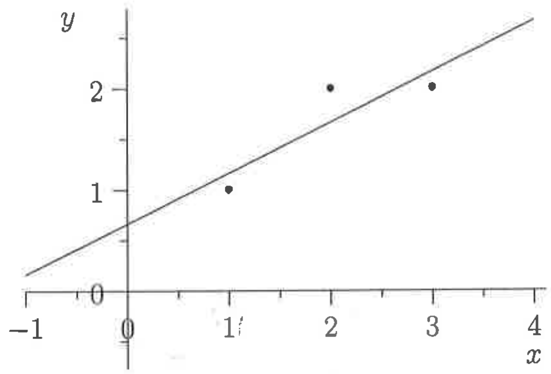


Figure 2: Three points and a line close to them.

Least Squares

Suppose we're given a collection of data points (t, b):

$$\{(1,1),(2,2),(3,2)\}$$

and we want to find the closest line $b = C + Dt$ to that collection. If the line went through all three points, we'd have:

$$\begin{aligned} C + D &= 1 \\ C + 2D &= 2 \\ C + 3D &= 2, \end{aligned}$$

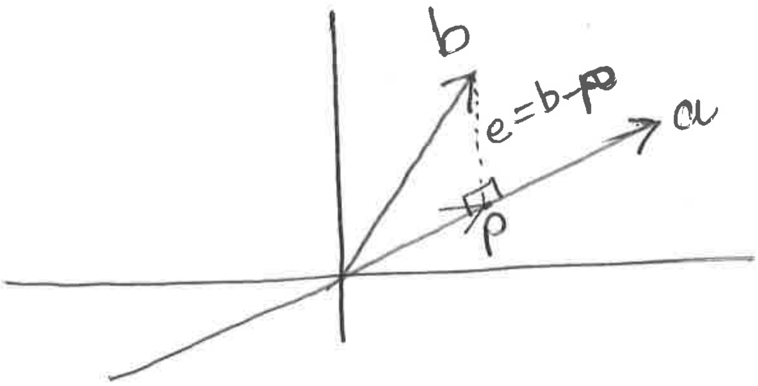
which is equivalent to:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}}_{\mathbf{b}}.$$

In our example the line does not go through all three points, so this equation is not solvable. Instead we'll solve:

$$A^T A \hat{\mathbf{x}} = A^T \mathbf{b}.$$

18.06SC Linear Algebra
Fall 2011



a = 1D - subspace
e = error
$\vec{p}$ = projection of $\vec{b}$ on $\vec{a}$

Since "$\vec{p}$" is in 1D - subspace $\vec{a}$. $\Rightarrow \bar{p} = x a$ i.e "x" is some multiple

since $e \perp a$

$$\Rightarrow a^T(b-p) = 0$$
$$a^T(b-xa) = 0$$
$$a^Tb - a^Txa = 0$$
or
$$x\, a^Ta = a^Tb$$
$$x = \frac{a^Tb}{a^Ta} \quad \& \quad p = ax$$

$$p = a\frac{a^Tb}{a^Ta}$$

i) If I double "b", the projection $p = a\frac{a^Tb}{a^Ta}$ will be doubled
ii) If I double "a", then projection "p" will be same. Hence we can see the projection p can be written as projection Matrix P.

$$\Rightarrow \quad \boxed{P = \frac{\overbrace{a a^T}^{\text{matrix}}}{\underbrace{a^T a}_{\text{scalar}}}} \quad \text{Projection Matrix}$$

Properties of this Matrix

whenever I multiply any vector "b" by any matrix "A", I land inside the column space of matrix "A".

i) $C(P)$ = line through $\vec{a}$

ii) $rank(P) = 1$ (we know that aa^T is column x row matrix thats a rank-one matrix & that column is the basis for column space)

Two main properties

iii) P is symmetric or $P^T = P$

iv) $P^2 = P$ (If I project $\vec{b}$ on $\vec{a}$ twice, then I will be on the same point on $\vec{a}$.

i.e $\quad p = Pb$

$$p^2 = P(Pb)$$

$$\boxed{p^2 = Pb}$$

← uptill here we have projected on a line →

Why Projection:

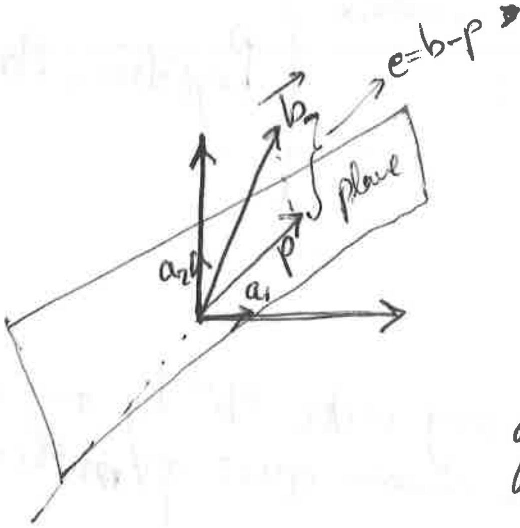
because $Ax = b$ may have no solution.

So whats the issue? Ax is in the column space of A but b is not in the column space of A. so I change "b". Thus solve

solve $A\hat{x} = P$ → projection of b onto column space

So we must find P as close as possible to "b"

In 3D



We have a plane. $\vec{b}$ is outside plane. $\vec{p}$ is the projection of $\vec{b}$ in plane. To identify plane, we have its basis $\vec{a_1}$ & $\vec{a_2}$ (they don't need to $\perp$).

This plane is the column space of $A = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix}$

"e" is perpendicular to plane

Now $\vec{p} = x_1 a_1 + x_2 a_2$ (some multiple of basis of column vectors

$= \hat{x}_1 a_1 + \hat{x}_2 a_2$

$\boxed{p = A\hat{x}}$ find the right combination of columns so that error "$e = b - p$" is $\perp$ to plane

Now $p = A\hat{x}$ Find $\hat{x}$

Key: $b - A\hat{x}$ is perp. to plane

As $(b - A\hat{x}) \perp \vec{a_1}$ & $(b - A\hat{x}) \perp \vec{a_2}$

$\Rightarrow a_1^T(b - A\hat{x}) = 0$ & $a_2^T(b - A\hat{x}) = 0$

OR $\underset{2\times1}{\begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix}} \underset{1\times1}{(b - A\hat{x})} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

OR

$$A^T(b - A\hat{x}) = 0$$

On Projection over line

$a^T(b - xa) = 0$

On Projection over plane

$A^T(b - A\hat{x}) = 0$

Now $A^T\underbrace{(b - A\hat{x})}_{\vec{e}} = 0$

① As $A^Te = 0$, then e is in $N(A^T)$. This means $e \perp C(A)$ [see the orthogonality of four fundamental subspaces]

Rewrite: $A^T(b - A\hat{x}) = 0$

$$\boxed{A^TA\hat{x} = A^Tb}$$

* In one dimension subspace, we had $x\, a^Ta = a^Tb$

Now $\hat{x} = (A^TA)^{-1}A^Tb$

$P = A\hat{x} = A(A^TA)^{-1}A^Tb$

in 1-D it was $\frac{aa^T}{a^Ta}$. Now see what happened in N-D

Projection matrix $P = A(A^TA)^{-1}A^T$

Now see, if we expand $A(A^TA)^{-1}A^T$ to $\underbrace{AA^{-1}}_{I}\underbrace{A^{T^{-1}}A^T}_{I}$, it will be Identity matrix, so what happend.
If A is square matrix, invertible, full rank, this means it column space is $\mathbb{R}^N$. Thus if we want to project vector b in $\mathbb{R}^N$, the projection matrix will be identity. But if A is not square matrix & not full rank, then we can not split apart $A(A^TA)^{-1}A^T$ to $AA^{-1}A^{T^{-1}}A^T$. Because there A^{-1} doesn't exist.

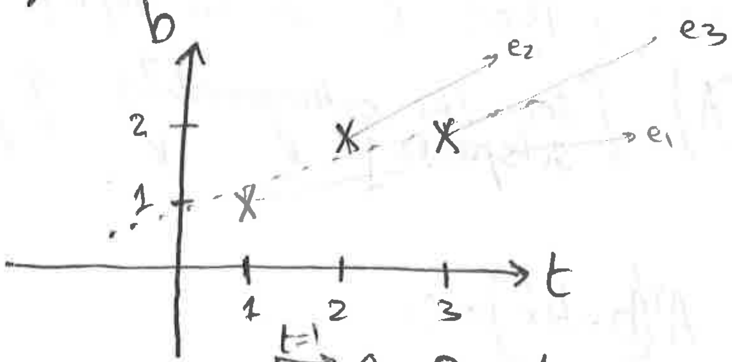
Properties of $P = A(A^TA)^{-1}A^T$:

① $P^T = P$

② $P^2 = P$

Application : Least Squares

Fitting by a line :
$(1,1), (2,2), (3,2)$



Now $b = C + Dt$ $\longrightarrow$

$$\begin{cases} t=1: & C + D = 1 \\ t=2: & C + 2D = 2 \\ t=3: & C + 3D = 2 \end{cases} \xrightarrow[\text{matrix form}]{\text{In}} \underbrace{\underset{A}{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}} \underset{x}{\begin{bmatrix} C \\ D \end{bmatrix}} = \underset{b}{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}}}_{\text{No solution}}$$

Since $Ax = b$ has no solution but when I multiply both sides by A^T, then we get the solution.

$$A^T(Ax) = A^T b$$

Recitation Videos ??

Unit2 lecture 3

Linear regression is commonly used to fit a line to a collection of data. The method of least squares can be viewed as finding the projection of a vector. Linear algebra provides a powerful and efficient description of linear regression in terms of the matrix A^TA.

Projection matrices and least squares

Projections

Last lecture, we learned that $P = A(A^TA)^{-1}A^T$ is the matrix that projects a vector $\mathbf{b}$ onto the space spanned by the columns of A. If $\mathbf{b}$ is perpendicular to the column space, then it's in the left nullspace $N(A^T)$ of A and $P\mathbf{b} = \mathbf{0}$. If $\mathbf{b}$ is in the column space then $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x}$, and $P\mathbf{b} = \mathbf{b}$.

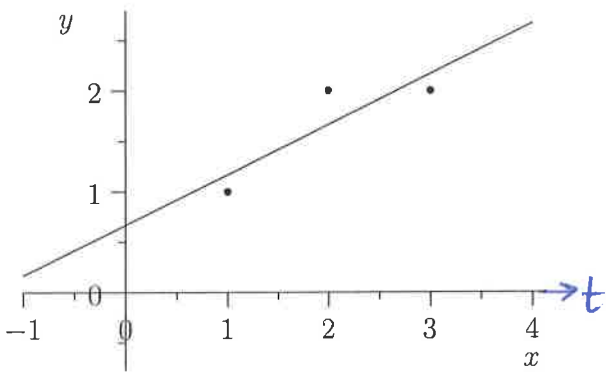
A typical vector will have a component $\mathbf{p}$ in the column space and a component $\mathbf{e}$ perpendicular to the column space (in the left nullspace); its projection is just the component in the column space.

The matrix projecting $\mathbf{b}$ onto $N(A^T)$ is $I - P$:

$$\begin{aligned} \mathbf{e} &= \mathbf{b} - \mathbf{p} \\ \mathbf{e} &= (I - P)\mathbf{b}. \end{aligned}$$

Naturally, $I - P$ has all the properties of a projection matrix.

Least squares



$y = C + Dt$

Figure 1: Three points and a line close to them.

We want to find the closest line $b = C + Dt$ to the points $(1, 1)$, $(2, 2)$, and $(3, 2)$. The process we're going to use is called *linear regression*; this technique is most useful if none of the data points are *outliers*.

By "closest" line we mean one that minimizes the error represented by the distance from the points to the line. We measure that error by adding up the squares of these distances. In other words, we want to minimize $||A\mathbf{x} - \mathbf{b}||^2 = ||\mathbf{e}||^2$.

$C = 2/3$

$D = 1/2$

$b = 2/3 + 1/2\, t$

If the line went through all three points, we'd have:

$$\begin{aligned} C+D &= 1 \\ C+2D &= 2 \\ C+3D &= 2, \end{aligned}$$

but this system is unsolvable. It's equivalent to $A\mathbf{x} = \mathbf{b}$, where:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} C \\ D \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

There are two ways of viewing this. In the space of the line we're trying to find, e_1, e_2 and e_3 are the vertical distances from the data points to the line. The components p_1, p_2 and p_3 are the values of $C + Dt$ near each data point; $\mathbf{p} \approx \mathbf{b}$.

In the other view we have a vector $\mathbf{b}$ in $\mathbb{R}^3$, its projection $\mathbf{p}$ onto the column space of A, and its projection $\mathbf{e}$ onto $N(A^T)$.

We will now find $\hat{\mathbf{x}} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$ and $\mathbf{p}$. We know:

$$\begin{aligned} A^TA\hat{\mathbf{x}} &= A^T\mathbf{b} \\ \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} &= \begin{bmatrix} 5 \\ 11 \end{bmatrix}. \end{aligned}$$

From this we get the *normal equations*:

$$\begin{aligned} 3\hat{C} + 6\hat{D} &= 5 \\ 6\hat{C} + 14\hat{D} &= 11. \end{aligned}$$

We solve these to find $\hat{D} = 1/2$ and $\hat{C} = 2/3$.

We could also have used calculus to find the minimum of the following function of two variables:

$$e_1^2 + e_2^2 + e_3^2 = (C + D - 1)^2 + (C + 2D - 2)^2 + (C + 3D - 2)^2.$$

Either way, we end up solving a system of linear equations to find that the closest line to our points is $b = \frac{2}{3} + \frac{1}{2}t$.

This gives us:

i	p_i	e_i
1	7/6	−1/6
2	5/3	1/3
3	13/6	−1/6

or $\mathbf{p} = \begin{bmatrix} 7/6 \\ 5/3 \\ 13/6 \end{bmatrix}$ and $\mathbf{e} = \begin{bmatrix} -1/6 \\ 2/6 \\ -1/6 \end{bmatrix}$. Note that $\mathbf{p}$ and $\mathbf{e}$ are orthogonal, and also that $\mathbf{e}$ is perpendicular to the columns of A.

The matrix $A^T A$

We've been assuming that the matrix $A^T A$ is invertible. Is this justified?

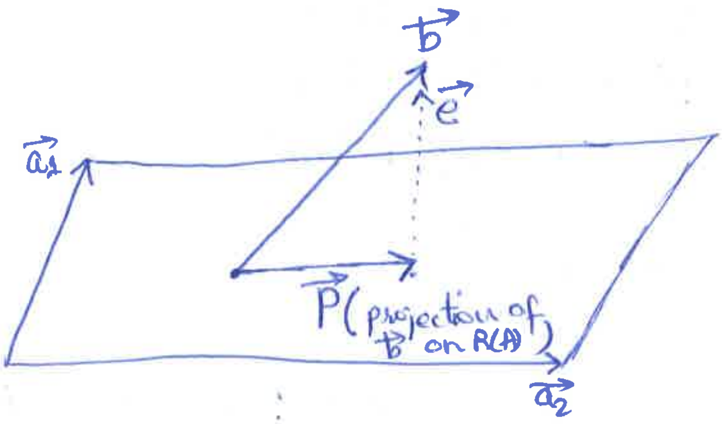
If A has independent columns, then $A^T A$ is invertible.

To prove this we assume that $A^T A\mathbf{x} = \mathbf{0}$, then show that it must be true that $\mathbf{x} = \mathbf{0}$:

$$\begin{aligned} A^T A\mathbf{x} &= \mathbf{0} \\ \mathbf{x}^T A^T A\mathbf{x} &= \mathbf{x}^T\mathbf{0} \\ (A\mathbf{x})^T(A\mathbf{x}) &= \mathbf{0} \\ A\mathbf{x} &= \mathbf{0}. \end{aligned}$$

Since A has independent columns, $A\mathbf{x} = \mathbf{0}$ only when $\mathbf{x} = \mathbf{0}$.

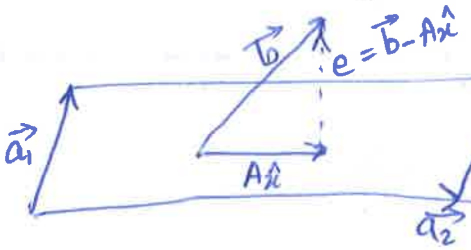
As long as the columns of A are independent, we can use linear regression to find approximate solutions to unsolvable systems of linear equations. The columns of A are guaranteed to be independent if they are *orthonormal*, i.e. if they are perpendicular unit vectors like $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, or like $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ and $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$.



$A = \begin{bmatrix} | & | \\ \vec{a_1} & \vec{a_2} \\ | & | \end{bmatrix}$

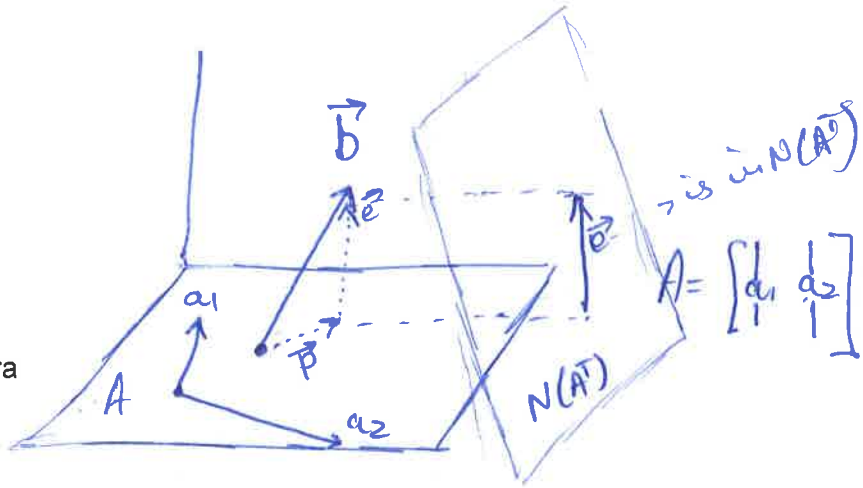
$\vec{P} = \hat{x}_1 \vec{a_1} + \hat{x}_2 \vec{a_2}$.
(linear combinatio[n] of basis vector)

$\vec{P} = A\hat{x}$



$\vec{a_1} \cdot (b - A\hat{x}) = 0$
$\vec{a_2} \cdot (b - A\hat{x}) = 0$

$\Rightarrow a_1 \cdot (b - A\hat{x}) + a_2 \cdot (b - A\hat{x}) = 0$
$a_1^T (b - A\hat{x}) + a_2^T (b - A\hat{x}) = 0$.
$A^T (b - A\hat{x}) = 0$ Normal equation



Proj Matrix

$$P = A(A^TA)^{-1}A^T$$

If b in column space $\quad Pb = b$

If $b \perp$ columnspace $\quad Pb = 0$

If $b \perp$ colomnspace(A) then I am in $N(A^T)$.

$$\Rightarrow Pb = A(A^TA^{-1})^{-1}\underbrace{A^Tb}_{0}$$

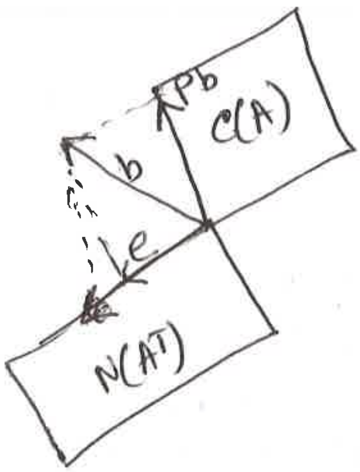
If b is in $C(A)$ then $b = Ax$

$$\Rightarrow Pb = A(A^TA)^{-1}A^Tb$$

$$Pb = A(A^TA)^{-1}A^T\underbrace{Ax}_{b}$$

$$Pb = \underbrace{A(A^TA)^{-1}A^T}_{P}b$$

$$Pb = Pb$$



Projection of b on C(A)

Now $Pb + e = b$

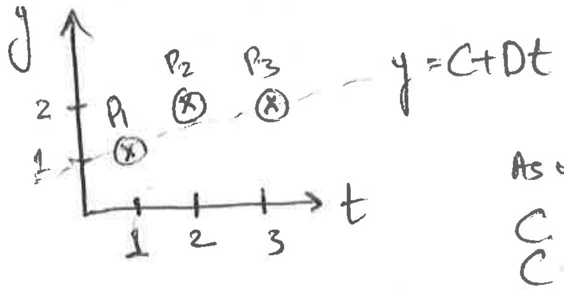
Projection of "b" on $N(A^T)$

$$e = b - Pb$$

$$e = b(I - P)$$

$$e = \underbrace{(I-P)}_{\text{projection onto } \perp \text{ space}}b$$

Properties of I-P : ① If P is symmetric, $I-P$ is symmetric
② If $P=P^2$ then $(I-P)=(I-P)^2$
③ If P is projection, then $I-P$ is projection



How to pick the best line? So that overall error is minimum.

As $y=C+Dt$

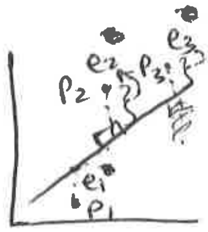
$$\left.\begin{aligned} C+D &= 1 \\ C+2D &= 2 \\ C+3D &= 2 \end{aligned}\right\} \text{No Solution. Find the best Approx. sol}$$

$$A=\begin{bmatrix}1 & 1\\ 1 & 2\\ 1 & 3\end{bmatrix} \quad x=\begin{bmatrix}C\\ D\end{bmatrix} \quad b=\begin{bmatrix}1\\ 2\\ 2\end{bmatrix}$$

Basis for $C(A) = \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\2\\3\end{pmatrix}$ but it doesn't contain $\begin{pmatrix}1\\2\\2\end{pmatrix}$.

· Minimize $Ax-b \longrightarrow e$ (error vector)

OR Minimize $\|Ax-b\|^2$



where $\|Ax-b\| = \|e\|^2$

$\|e\| = \|e_1\|^2 + \|e_2\|^2 + \|e_3\|^2$

we are doing regression here. we can say linear regression.

Let P_1, P_2, P_3 lie on desired line. They will be new b_1, b_2, b_3.

Find $\hat{x} = \begin{bmatrix}\hat{C}\\ \hat{D}\end{bmatrix}$, P?

As $\boxed{A^TA\hat{x} = A^Tb} \longrightarrow$ Most important equation in statistics. Eq in estimation

$$A^TA = \begin{bmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 2\\ 1 & 3\end{bmatrix} = \underbrace{\begin{bmatrix}3 & 6\\ 6 & 14\end{bmatrix}}$$

I expect it to be
① symmetric
② invertible
③ positive SD.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 3 & 2 \end{array}\right] = \left[\begin{array}{cc|c} 3 & 6 & 5 \\ 6 & 14 & 11 \end{array}\right]$$

$A^T A \; b$

$\Rightarrow \; 3C + 6D = 5$ —— (i) $\rightarrow$ derivative of error w.r.t C

$6C + 14D = 11$ —— (ii) $\rightarrow$ derivative of error w.r.t D

In $A^T A x = A^T b \longrightarrow$ normal equations

In Minimize $\|Ax + b\|^2 = e_1^2 + e_2^2 + e_3^2$

$= (C+D-1)^2 + (C+2D-2)^2 + (C+3D-2)^2$ } from calculus

We can $\frac{d}{dC} = 0$, $\frac{d}{dD} = 0$ & we will get (i) & (ii)

Using (i) $\quad 2D = 1$

$\boxed{D = 1/2} \quad \boxed{C = \frac{2}{3}}$

Best line is $\frac{2}{3} + \frac{1}{2}t$

$\Rightarrow \quad y = \frac{2}{3} + \frac{1}{2}t$

$\Rightarrow P_1 = 7/6 \quad \left(\frac{2}{3} + \frac{1}{2}\right) \quad e_1 = -1/6$

$P_2 = 5/3 \quad \left(\frac{2}{3} + 1\right) \quad e_2 = +2/6$

$P_3 = 13/6 \quad \left(\frac{2}{3} + \frac{1}{2} \times 3\right) \quad e_3 = -1/6$

$e = \begin{bmatrix} 1/6 \\ 2/6 \\ -1/6 \end{bmatrix}$

Now $b = P + e$

$$\underset{b}{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}} = \underset{P}{\begin{bmatrix} 7/6 \\ 5/3 \\ 13/6 \end{bmatrix}} + \underset{e}{\begin{bmatrix} -1/6 \\ 2/6 \\ -1/6 \end{bmatrix}}$$

1. P & e are orthogonal
2. e is $\perp$ to C(A)

Key Eq $\quad A^TA\hat{x} = A^Tb$

And $\quad P = A\hat{x}$

If A has independent column. then A^TA is invertible.

Using above fact we solved $A^TA\hat{x} = A^Tb$. ~~If A^TA is not invertible then~~

<u>To Prove</u>: If A has indep columns then A^TA is invertible

Suppose $A^TAx = 0$ then "x" must be 0 ~~then A^TA is not inverti~~ (i.e the matrix is invertible when its nullspace has $\vec{0}$ vector only)

<u>Trick</u> $\quad A^TAx = 0 \quad$ take dot product of this eq with x

<u>Idea</u> $\Rightarrow \quad x^TA^TAx = 0$

$\Rightarrow (Ax)^T(Ax) = 0$

Let $y = Ax$

$\Rightarrow Ax = 0$

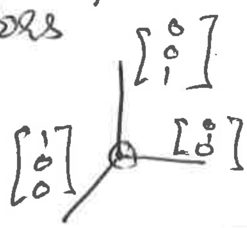
Now we will use our hypotheses,

if A has independent columns, then "x" must be zero

$$0 = (Ax)^T(Ax) \xrightarrow{\text{(square)}} Ax = 0 \xrightarrow{\text{This means } A \text{ has indep columns}} x = 0$$

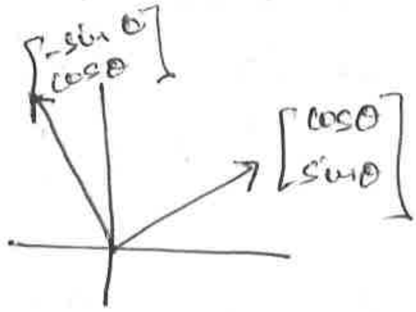
Columns are definitely independent if they are perpendicular unit vectors

{for instance $\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}$}.



~~If we are dealing with~~ $\boxed{\underbrace{\text{perpendicular unit vectors}}_{\text{ortho-normal vectors}}}$

Recitation Video



Unit 2: Lecture #4

MANY CALCULATIONS becomes simpler when performed using orthonormal vectors or orthogonal matrices. In this session, we will learn a procedure for converting any basis to an orthonormal one.

Orthogonal matrices and Gram-Schmidt

In this lecture we finish introducing orthogonality. Using an orthonormal basis or a matrix with orthonormal columns makes calculations much easier. The Gram-Schmidt process starts with any basis and produces an orthonormal basis that spans the same space as the original basis.

Orthonormal vectors

The vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_n$ are *orthonormal* if:

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

In other words, they all have (normal) length 1 and are perpendicular (ortho) to each other. Orthonormal vectors are always independent.

Orthonormal matrix

If the columns of $Q = \begin{bmatrix} \mathbf{q}_1 & \ldots & \mathbf{q}_n \end{bmatrix}$ are orthonormal, then $Q^TQ = I$ is the identity.

Matrices with orthonormal columns are a new class of important matrices to add to those on our list: triangular, diagonal, permutation, symmetric, reduced row echelon, and projection matrices. We'll call them "orthonormal matrices".

A square orthonormal matrix Q is called an *orthogonal matrix*. If Q is square, then $Q^TQ = I$ tells us that $Q^T = Q^{-1}$.

For example, if $Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ then $Q^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. Both Q and Q^T are orthogonal matrices, and their product is the identity.

The matrix $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal. The matrix $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ is not, but we can adjust that matrix to get the orthogonal matrix $Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. We can use the same tactic to find some larger orthogonal matrices called *Hadamard matrices*:

$$Q = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

An example of a rectangular matrix with orthonormal columns is:

$$Q = \frac{1}{3}\begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}.$$

We can extend this to a (square) orthogonal matrix:

$$\frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}.$$

These examples are particularly nice because they don't include complicated square roots.

Orthonormal columns are good

Suppose Q has orthonormal columns. The matrix that projects onto the column space of Q is:

$$P = Q^T(Q^TQ)^{-1}Q^T.$$

If the columns of Q are orthonormal, then $Q^TQ = I$ and $P = QQ^T$. If Q is square, then $P = I$ because the columns of Q span the entire space.

Many equations become trivial when using a matrix with orthonormal columns. If our basis is orthonormal, the projection component $\hat{x}_i$ is just $\mathbf{q}_i^T\mathbf{b}$ because $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$ becomes $\hat{\mathbf{x}} = Q^T\mathbf{b}$.

Gram-Schmidt

With elimination, our goal was "make the matrix triangular". Now our goal is "make the matrix orthonormal".

We start with two independent vectors $\mathbf{a}$ and $\mathbf{b}$ and want to find orthonormal vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ that span the same plane. We start by finding orthogonal vectors $\mathbf{A}$ and $\mathbf{B}$ that span the same space as $\mathbf{a}$ and $\mathbf{b}$. Then the unit vectors $\mathbf{q}_1 = \frac{\mathbf{A}}{||\mathbf{A}||}$ and $\mathbf{q}_2 = \frac{\mathbf{B}}{||\mathbf{B}||}$ form the desired orthonormal basis.

Let $\mathbf{A} = \mathbf{a}$. We get a vector orthogonal to $\mathbf{A}$ in the space spanned by $\mathbf{a}$ and $\mathbf{b}$ by projecting $\mathbf{b}$ onto $\mathbf{a}$ and letting $\mathbf{B} = \mathbf{b} - \mathbf{p}$. ($\mathbf{B}$ is what we previously called $\mathbf{e}$.)

$$\mathbf{B} = \mathbf{b} - \frac{\mathbf{A}^T\mathbf{b}}{\mathbf{A}^T\mathbf{A}}\mathbf{A}.$$

If we multiply both sides of this equation by $\mathbf{A}^T$, we see that $\mathbf{A}^T\mathbf{B} = 0$.

What if we had started with three independent vectors, $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$? Then we'd find a vector $\mathbf{C}$ orthogonal to both $\mathbf{A}$ and $\mathbf{B}$ by subtracting from $\mathbf{c}$ its components in the $\mathbf{A}$ and $\mathbf{B}$ directions:

$$\mathbf{C} = \mathbf{c} - \frac{\mathbf{A}^T\mathbf{c}}{\mathbf{A}^T\mathbf{A}}\mathbf{A} - \frac{\mathbf{B}^T\mathbf{c}}{\mathbf{B}^T\mathbf{B}}\mathbf{B}.$$

For example, suppose $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$. Then $\mathbf{A} = \mathbf{a}$ and:

$$\begin{aligned} \mathbf{B} &= \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{\mathbf{A}^T\mathbf{b}}{\mathbf{A}^T\mathbf{A}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Normalizing, we get:

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix}.$$

The column space of Q is the plane spanned by $\mathbf{a}$ and $\mathbf{b}$.

When we studied elimination, we wrote the process in terms of matrices and found $A = LU$. A similar equation $A = QR$ relates our starting matrix A to the result Q of the Gram-Schmidt process. Where L was lower triangular, R is upper triangular.

Suppose $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$. Then:

$$\begin{array}{ccc} A & & Q \qquad\qquad R \\ \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} & = & \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1^T\mathbf{q}_1 & \mathbf{a}_2^T\mathbf{q}_1 \\ \mathbf{a}_1^T\mathbf{q}_2 & \mathbf{a}_2^T\mathbf{q}_2 \end{bmatrix} \end{array}.$$

If R is upper triangular, then it should be true that $\mathbf{a}_1^T\mathbf{q}_2 = 0$. This must be true because we chose $\mathbf{q}_1$ to be a unit vector in the direction of $\mathbf{a}_1$. All the later $\mathbf{q}_i$ were chosen to be perpendicular to the earlier ones.

Notice that $R = Q^T A$. This makes sense; $Q^T Q = I$.

LA lecture 17

Orthogonal basis $q_1, \dots q_n$

Orthogonal matrix Q [if Q is square]

Gram-Schmidt $A \longrightarrow Q$

Orthonormal vectors

$$q_i^T q_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

ortho: for 0 $i \neq j$

normal: normalised, unit length i.e 1 $i = j$: That's why we say orthonormal b/c $q_1, \dots q_n$ are unit normalised vectors.

PART 1

I want $q_i^T q_N$ to be orthonormal

$$Q = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix}$$

doesn't have to be square

$$Q^T Q = \begin{bmatrix} - q_1^T - \\ \vdots \\ - q_n^T - \end{bmatrix} \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & & & 1 \end{bmatrix}$$

$$Q^T \quad Q \quad = \quad I$$

If Q is square then $Q^T Q = I$ tells us $Q^T = Q^{-1}$

Example

i) perm $Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ $QQ^T = I$ (we can check it)

ii) $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ here $QQ^T = I$

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{bmatrix}$$

iii) $Q = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \xrightarrow{\text{to make it orthogonal}} Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

iv) $\underbrace{Q = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}}_{\text{worlds favourite orthogonal matrix}} \xrightarrow[\text{unit vectors}]{\text{to make its orthogonal columns}} \underbrace{Q = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}}_{\text{Adamhar}}$

Question

Why do we want orthogonal matrices? What calculation make it easy?

Now take an example of rectangular Q:

$$Q = \begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix} \xrightarrow{\text{make it normalised}} Q = \frac{1}{3} \begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}$$

They are orthonormal column vectors & they span 2D-space. If we want to make Q square, we need to add another vector that should be orthonormal as well to q_1 & q_2. This will be done by Gram-Schmidt. thus

$$Q = \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}$$

Ques: What formula become easier when we have Q.

Ans: Suppose Q has orthonormal columns.

I wanna project onto its column space

$$P = Q(Q^TQ)^{-1}Q^T = QQ^T$$

(when Q is orthonormal then $(Q^TQ)^{-1} = I$, we don't need to calculate inverse)

when we have got orthonormal basis in C(Q).

Properties of Projection Matrix:

i) It symmetric

ii) If you project & project again, we doesn't change

i.e $\underbrace{(QQ^T)(QQ^T)}_{I} = (QQ^T)$

$$P = Q(Q^TQ)^{-1}Q^T = QQ^T \begin{cases} = I \text{ if } Q \text{ is square} \end{cases}$$

It is the matrix that projects a vector b onto the space spanned by columns of Q.

$A^TA\hat{x} = A^Tb$

Now A is Q

$Q^TQ\hat{x} = Q^Tb$

$\Rightarrow$ As $Q^TQ = I$

$\Rightarrow$ $\boxed{\hat{x} = Q^Tb}$

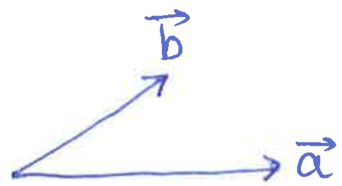
OR $\hat{x}_i = q_i^T b$

2nd Half of Lecture:

We will start with different independent vectors.

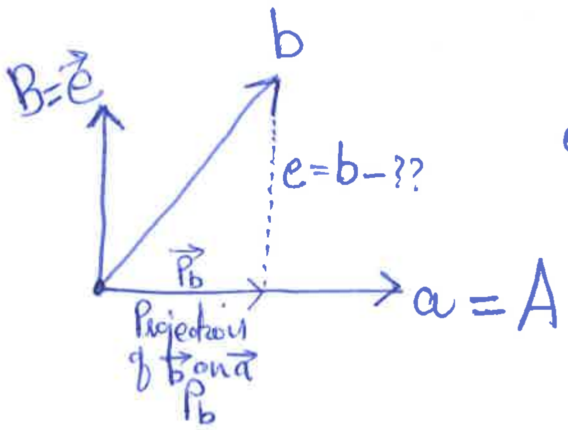
Gram-Schmidt: Make our matrix orthonormal

Start $\vec{a}, \vec{b}$ where $\vec{a}$ is independent of $\vec{b}$.



~~Vector~~

$$\text{Vector } \vec{a}, \vec{b} \longrightarrow \underbrace{\underset{\text{orthogonal}}{A, B}}_{\substack{\text{Gram's Part} \\ A \perp B}} \xrightarrow{} \underbrace{\overset{\text{orthonormal}}{q_1 = \frac{A}{\|A\|} \quad q_2 = \frac{B}{\|B\|}}}_{\text{Schmidt Part}}$$



$$e = B = b - P_b$$

$$e = B = b - \underset{\text{of } A}{\text{some multiple}}$$

$$\Rightarrow \boxed{B = b - \underbrace{\frac{A^T b}{A^T A} A}_{\text{Projection of } b \text{ on } A}} \quad \text{Gram's formula}$$

Now $B \perp A$: Checking it :

$$A^T B = 0$$

$$A^T \left(b - \frac{A^T b}{A^T A} A \right) = 0$$

$$A^T b - A^T A \frac{A^T b}{A^T A} = 0$$

$$\boxed{0 = 0}$$

Now this was the story with two vectors $\vec{a}, \vec{b}$. If we introduce $\vec{c}$ (third independent vector), then we will have:

$$\text{Vector } \vec{a}, \vec{b}, \vec{c} \longrightarrow \overset{\text{orthogonal}}{A, B, C} \longrightarrow \overset{\text{orthonormal}}{q_1 = \frac{A}{\|A\|}, \ q_2 = \frac{B}{\|B\|}, \ q_3 = \frac{C}{\|C\|}}$$

How to find $\vec{C}$? , we have $\vec{A}, \vec{B}$.

$C = ??$

$C = c -$ its component in A, B direction

$$C = \underbrace{c - \frac{A^T c}{A^T A} A}_{c - \text{its projection in direction of } A} - \underbrace{\frac{B^T c}{B^T B} B}_{\text{projection of } c \text{ in direction of } B}$$

Now $C \perp A$

$C \perp B$

$$\underset{\text{orthonormal}}{C} = \frac{C}{\|C\|}$$

Example

I will give you two vectors. You will give me orthonormal basis.

Graham Schmidt

$$a = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

a, b is not orthonormal.

Initialise

$$\boxed{A = a}$$

$B = b -$ some multiple of A

$$B = \underset{b}{\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}} - \frac{A^T b}{A^T A} \underset{A}{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} \Rightarrow B = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\boxed{B = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}}$$

Its correct b/c $A \perp B$ i.e $A^TB = 0$

$$\Rightarrow q_1 = \frac{A}{\|A\|}, \quad q_2 = \frac{B}{\|B\|}$$

$$Q = \begin{bmatrix} q_1 & q_2 \end{bmatrix}$$

$$Q = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 1/\sqrt{2} \end{bmatrix}$$ Orthonormal matrix from GramSchmidt

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \xrightarrow[\text{Schmidt}]{\text{Graham}} Q = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{we ended with } Q}$$

we started with A

Column-Spaces of A & Q

How the column space of A & Q are related. C(A) is a plane in 3D space so as C(Q). what is the relation b/w planes ~~f~~ formed by C(A) & C(Q). They both are same ^col^ plane. b/c ~~"||"~~ OR its the same column-space. [B/c all ^col^ vectors in Q are the linear combination of col vectors in A i.e

$$A = \begin{bmatrix} | & | \\ \vec{a} & \vec{b} \\ | & | \end{bmatrix} \qquad Q = \begin{bmatrix} | & | \\ \vec{A} & \vec{B} \\ | & | \end{bmatrix}$$

$\vec{a} = \vec{A}$

$\vec{B} = \vec{b} -$ some multiple of $\vec{A}$].

The original C(A) had perfectly good basis $\begin{pmatrix}1\\1\\1\end{pmatrix}\begin{pmatrix}1\\0\\2\end{pmatrix}$ but they are not as good as $\begin{pmatrix}\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{pmatrix}, \begin{pmatrix}0\\-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}$ b/c ~~they~~ it ~~are~~ is orthonormal & will simplify normal equation $Q(Q^TQ)^{-1}Q^T$.

Elimination in Matrix Language $A = LU$

Now $A = QR$ (expression of Gram Schmidt)

Let $A = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \end{bmatrix} \begin{bmatrix} & \\ 0 & \end{bmatrix}$$

$$A \quad = \quad Q \quad R$$

R → upper triangle

$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \end{bmatrix} \begin{bmatrix} a_1^T q_1 & a_2^T q_1 \\ a_1^T q_2 & a_2^T q_2 \end{bmatrix}$$

why $a_1^T q_2 = 0$? Key for R being Upper ~~tour~~ triangle.

$\boxed{a_1^T q_2 = 0 \text{ or } a^T q_2 = 0}$ b/c we constructed later q's i.e q_2, q_3 to be perpendicular to earlier q's i.e q_1, q_2. Thats why we get a triangular matrix.

Result is extremely satisfactory: If $A = QR$

And I have a matrix(A) with independent columns, the Gram Schmidt produces a matrix (Q) with orthonormal columns And the connection b/w those is a triangular matrix (R). That connection of triangular matrix should be reviewed in a book.

$\boxed{\text{Recitation Video}}$

Unit 2: Lecture 5

The determinant of a matrix is a single number which encodes a lot of 'information' about the matrix. Three simple properties completely describe the determinant. In this lecture, we also list seven more properties like det AB = (det A)(det B) that can be derived from 1st three.

Properties of determinants

Determinants

Now halfway through the course, we leave behind rectangular matrices and focus on square ones. Our next big topics are determinants and eigenvalues.

The *determinant* is a number associated with any square matrix; we'll write it as $\det A$ or $|A|$. The determinant encodes a lot of information about the matrix; the matrix is invertible exactly when the determinant is non-zero.

Properties

Rather than start with a big formula, we'll list the properties of the determinant. We already know that $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$; these properties will give us a formula for the determinant of square matrices of all sizes.

1. $\det I = 1$

2. If you exchange two rows of a matrix, you reverse the sign of its determinant from positive to negative or from negative to positive.

3. (a) If we multiply one row of a matrix by t, the determinant is multiplied by t: $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

   (b) The determinant behaves like a linear function on the rows of the matrix:
$$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}.$$

Property 1 tells us that $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$. Property 2 tells us that $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$. The determinant of a permutation matrix P is 1 or -1 depending on whether P exchanges an even or odd number of rows.

From these three properties we can deduce many others:

4. If two rows of a matrix are equal, its determinant is zero.

   This is because of property 2, the exchange rule. On the one hand, exchanging the two identical rows does not change the determinant. On the other hand, exchanging the two rows changes the sign of the determinant. Therefore the determinant must be 0.

5. If $i \neq j$, subtracting t times row i from row j doesn't change the determinant.

In two dimensions, this argument looks like:

$$\begin{aligned}
\begin{vmatrix} a & b \\ c-ta & d-tb \end{vmatrix} &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - \begin{vmatrix} a & b \\ ta & tb \end{vmatrix} && \text{property 3(b)} \\
&= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - t\begin{vmatrix} a & b \\ a & b \end{vmatrix} && \text{property 3(a)} \\
&= \begin{vmatrix} a & b \\ c & d \end{vmatrix} && \text{property 4.}
\end{aligned}$$

The proof for higher dimensional matrices is similar.

6. If A has a row that is all zeros, then $\det A = 0$.

   We get this from property 3 (a) by letting $t = 0$.

7. The determinant of a triangular matrix is the product of the diagonal entries (pivots) $d_1, d_2, ..., d_n$.

   Property 5 tells us that the determinant of the triangular matrix won't change if we use elimination to convert it to a diagonal matrix with the entries d_i on its diagonal. Then property 3 (a) tells us that the determinant of this diagonal matrix is the product $d_1 d_2 \cdots d_n$ times the determinant of the identity matrix. Property 1 completes the argument.

   Note that we cannot use elimination to get a diagonal matrix if one of the d_i is zero. In that case elimination will give us a row of zeros and property 6 gives us the conclusion we want.

8. $\det A = 0$ exactly when A is singular.

   If A is singular, then we can use elimination to get a row of zeros, and property 6 tells us that the determinant is zero.

   If A is not singular, then elimination produces a full set of pivots $d_1, d_2, ..., d_n$ and the determinant is $d_1 d_2 \cdots d_n \neq 0$ (with minus signs from row exchanges).

We now have a very practical formula for the determinant of a non-singular matrix. In fact, the way computers find the determinants of large matrices is to first perform elimination (keeping track of whether the number of row exchanges is odd or even) and then multiply the pivots:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \longrightarrow \begin{bmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{bmatrix}, \text{ if } a \neq 0, \text{ so}$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a(d - \frac{c}{a}b) = ad - bc.$$

9. $\det AB = (\det A)(\det B)$

   This is very useful. Although the determinant of a sum does not equal the sum of the determinants, it is true that the determinant of a product equals the product of the determinants.

For example:

$$\det A^{-1} = \frac{1}{\det A},$$

because $A^{-1}A = 1$. (Note that if A is singular then A^{-1} does not exist and $\det A^{-1}$ is undefined.) Also, $\det A^2 = (\det A)^2$ and $\det 2A = 2^n \det A$ (applying property 3 to each row of the matrix). This reminds us of volume – if we double the length, width and height of a three dimensional box, we increase its volume by a multiple of $2^3 = 8$.

10. $\det A^T = \det A$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc.$$

This lets us translate properties (2, 3, 4, 5, 6) involving rows into statements about columns. For instance, if a column of a matrix is all zeros then the determinant of that matrix is zero.

To see why $|A^T| = |A|$, use elimination to write $A = LU$. The statement becomes $|U^TL^T| = |LU|$. Rule 9 then tells us $|U^T||L^T| = |L||U|$.

Matrix L is a lower triangular matrix with 1's on the diagonal, so rule 5 tells us that $|L| = |L^T| = 1$. Because U is upper triangular, rule 5 tells us that $|U| = |U^T|$. Therefore $|U^T||L^T| = |L||U|$ and $|A^T| = |A|$.

We have one loose end to worry about. Rule 2 told us that a row exchange changes the sign of the determinant. If it's possible to do seven row exchanges and get the same matrix you would by doing ten row exchanges, then we could prove that the determinant equals its negative. To complete the proof that the determinant is well defined by properties 1, 2 and 3 we'd need to show that the result of an odd number of row exchanges (odd permutation) can never be the same as the result of an even number of row exchanges (even permutation).

Properties of det A:

1. det $I = 1$
2. Exchange rows: Reverse sign of det. [i.e det $P = 1$ (# of exchange was even) $= -1$ (# of exchange is odd)] permutation matrix
3.

Example ① $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$ $\quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

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

Property 3a $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$

Property 3b $\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$

These properties are about linear combination

The determinant ~~be~~ is linear for each row (3b) separately.

④ ~~Two~~ 2 equal rows → det $= 0$

Exchange those rows → we get same matrix but sign changes ∴ det $= 0$

(5) Subtract $l \times$ row i from row k

DET Doesn't change (in $A=LU$ det of A is equal to det of U)

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \longrightarrow \begin{vmatrix} a & b \\ c-la & d-lb \end{vmatrix} = \underbrace{\begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a & b \\ -la & -lb \end{vmatrix}}_{\text{Property 3b}}$$

$$= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - l \underbrace{\begin{vmatrix} a & b \\ a & b \end{vmatrix}}_{\substack{\det = 0 \\ \text{Property (4)}}}$$

(6) Row of zeros $\longrightarrow \det A = 0$

$$5\begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} = \begin{vmatrix} 5(0) & 5(0) \\ c & d \end{vmatrix}$$

$$5 \det A = \det A$$

Thus $\det A = 0$

(7) $$U = \begin{bmatrix} d_1 & * & * & * \\ 0 & d_2 & * & * \\ 0 & 0 & \ddots & * \\ 0 & 0 & 0 & d_n \end{bmatrix}$$ (U = upper triangle matrix)

$\det U = (d_1)(d_2)\dots(d_n)$

= product of pivots (if there is no row-exchange) (problem of + sign, -ve sign due to row exchange)

Proof

$$|U| = \begin{vmatrix} d_1 & 0 & 0 & 0 & \cdots \\ 0 & d_2 & 0 & 0 & \cdots \\ 0 & 0 & d_3 & & \\ & & & \ddots & \\ & & & & d_n \end{vmatrix}$$ By elimination

$\boxed{\text{Rules 5, 3a, 1 are used}}$

$$|U| = d_1 d_2 d_3 \dots d_n \begin{vmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \end{vmatrix} \rightarrow \det = 1$$

$|U| = d_1 d_2 \dots d_n$

8 $\det A = 0$ exactly when A is singular $\rightarrow$ row of zeros

$\det A \neq 0$ when A is invertible $\xrightarrow{\text{we go to}} U \xrightarrow{\text{then}} D \to d_1 d_2 \ldots d_n$

Example

1st pivot

$$\begin{matrix} a & b \\ c & d \end{matrix} \longrightarrow \begin{matrix} a & b \\ 0 & d - \frac{c}{a} b \end{matrix} \longrightarrow ad - bc$$

9 $\det AB = (\det A)(\det B)$ // $\det A^{-1} = ??$

$\det A^{-1} = ?? = 1/\det A$

As $A^{-1}A = I \longrightarrow \det(A^{-1}A) = \det(I)$

$\Rightarrow \det(A^{-1}) \det(A) = 1$

$$\boxed{\det(A^{-1}) = \frac{1}{\det(A)}}$$

If $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ then $A^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}$

Also $\det A^2 = \det(A \cdot A)$
$= \det(A) \det(A)$
$= (\det A)^2$

$\det 2A = 2^n \det(A)$

multiply every row in A by 2

If $A = m \times n$

⑩ $\det A^T = \det A$

Prove #10 $|A^T| = |A|$

$$\begin{array}{ccc} |A^T| & = & |A| \\ \downarrow & & \downarrow \\ |U^T L^T| & = & |LU| \end{array}$$

$A \xrightarrow{\text{factored}} LU$

$A^T \longrightarrow U^T L^T$

$$\boxed{|U^T||L^T| = |L||U|}$$

$L \rightarrow$ lower triangular matrix

$|L| =$ product of diagonals

Recitation Video

Unit2 Lecture6 :

One way to compute the determinant is by elimination. In this lecture, we derive two related formulas for the determinant using the properties from last lecture.

Determinant formulas and cofactors

Now that we know the properties of the determinant, it's time to learn some (rather messy) formulas for computing it.

Formula for the determinant

We know that the determinant has the following three properties:

1. $\det I = 1$
2. Exchanging rows reverses the sign of the determinant.
3. The determinant is linear in each row separately.

Last class we listed seven consequences of these properties. We can use these ten properties to find a formula for the determinant of a 2 by 2 matrix:

$$\begin{aligned}
\begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} \\
&= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} \\
&= 0 + ad + (-cb) + 0 \\
&= ad - bc.
\end{aligned}$$

By applying property 3 to separate the individual entries of each row we could get a formula for any other square matrix. However, for a 3 by 3 matrix we'll have to add the determinants of twenty seven different matrices! Many of those determinants are zero. The non-zero pieces are:

$$\begin{aligned}
\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} \\
&\quad + \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{33} - a_{12}a_{21}a_{33} \\
&\quad + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.
\end{aligned}$$

Each of the non-zero pieces has one entry from each row in each column, as in a permutation matrix. Since the determinant of a permutation matrix is either 1 or -1, we can again use property 3 to find the determinants of each of these summands and obtain our formula.

One way to remember this formula is that the positive terms are products of entries going down and to the right in our original matrix, and the negative terms are products going down and to the left. This rule of thumb doesn't work for matrices larger than 3 by 3.

The number of parts with non-zero determinants was 2 in the 2 by 2 case, 6 in the 3 by 3 case, and will be $24 = 4!$ in the 4 by 4 case. This is because there are n ways to choose an element from the first row (i.e. a value for α), after which there are only $n-1$ ways to choose an element from the second row that avoids a zero determinant. Then there are $n-2$ choices from the third row, $n-3$ from the fourth, and so on.

The big formula for computing the determinant of any square matrix is:

$$\det A = \sum_{n! \text{ terms}} \pm a_{1\alpha}a_{2\beta}a_{3\gamma}...a_{n\omega}$$

where $(\alpha, \beta, \gamma, ...\omega)$ is some permutation of $(1, 2, 3, ..., n)$. If we test this on the identity matrix, we find that all the terms are zero except the one corresponding to the trivial permutation $\alpha = 1$, $\beta = 2$, ..., $\omega = n$. This agrees with the first property: $\det I = 1$. It's possible to check all the other properties as well, but we won't do that here.

Applying the method of elimination and multiplying the diagonal entries of the result (the pivots) is another good way to find the determinant of a matrix.

Example

In a matrix with many zero entries, many terms in the formula are zero. We can compute the determinant of:

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

by choosing a non-zero entry from each row and column, multiplying those entries, giving the product the appropriate sign, then adding the results.

The permutation corresponding to the diagonal running from a_{14} to a_{41} is $(4, 3, 2, 1)$. This contributes 1 to the determinant of the matrix; the contribution is positive because it takes two row exchanges to convert the permutation $(4, 3, 2, 1)$ to the identity $(1, 2, 3, 4)$.

Another non-zero term of $\sum \pm a_{1\alpha}a_{2\beta}a_{3\gamma}a_{4\omega}$ comes from the permutation $(3, 2, 1, 4)$. This contributes -1 to the sum, because one exchange (of the first and third rows) leads to the identity.

These are the only two non-zero terms in the sum, so the determinant is 0. We can confirm this by noting that row 1 minus row 2 plus row 3 minus row 4 equals zero.

Cofactor formula

The cofactor formula rewrites the big formula for the determinant of an n by n matrix in terms of the determinants of smaller matrices.

In the 3×3 case, the formula looks like:

$$\begin{aligned} \det A &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(-a_{21}a_{33} + a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix} \end{aligned}$$

This comes from grouping all the multiples of a_{ij} in the big formula. Each element is multiplied by the *cofactors* in the parentheses following it. Note that each cofactor is (plus or minus) the determinant of a two by two matrix. That determinant is made up of products of elements in the rows and columns NOT containing a_{1j}.

In general, the cofactor C_{ij} of a_{ij} can be found by looking at all the terms in the big formula that contain a_{ij}. C_{ij} equals $(-1)^{i+j}$ times the determinant of the $n-1$ by $n-1$ square matrix obtained by removing row i and column j. (C_{ij} is positive if $i+j$ is even and negative if $i+j$ is odd.)

For $n \times n$ matrices, the cofactor formula is:

$$\boxed{\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}.}$$

Applying this to a 2×2 matrix gives us:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad + b(-c).$$

Tridiagonal matrix

A *tridiagonal matrix* is one for which the only non-zero entries lie on or adjacent to the diagonal. For example, the 4×4 tridiagonal matrix of 1's is:

$$A_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

What is the determinant of an $n \times n$ tridiagonal matrix of 1's?

$$|A_1| = 1, |A_2| = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 0, |A_3| = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = -1$$

$$|A_4| = 1 \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = |A_3| - 1|A_2| = -1$$

In fact, $|A_n| = |A_{n-1}| - |A_{n-2}|$. We get a sequence which repeats every six terms:

$$|A_1| = 1, |A_2| = 0, |A_3| = -1, |A_4| = -1, |A_5| = 0, |A_6| = 1, |A_7| = 1.$$

Unit II Lecture 7 :

Now we start to use the determinant. Understanding the cofactor formula allows us to show that $A^{-1} = (1/\det A) C^T$, where C is the matrix of cofactors of A. Combining this formula with the equation $x = A^{-1}b$ gives us Cramer's Rule for solving $Ax = b$. Also, the absolute value of determinant give the volume of a box.

Cramer's rule, inverse matrix, and volume

We know a formula for and some properties of the determinant. Now we see how the determinant can be used.

Formula for A^{-1}

We know:

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

Can we get a formula for the inverse of a 3 by 3 or n by n matrix? We expect that $\frac{1}{\det A}$ will be involved, as it is in the 2 by 2 example, and by looking at the cofactor matrix $\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$ we might guess that cofactors will be involved.

In fact:

$$A^{-1} = \frac{1}{\det A} C^T$$

where C is the matrix of cofactors – please notice the transpose! Cofactors of row one of A go into column 1 of A^{-1}, and then we divide by the determinant.

The determinant of A involves products with n terms and the cofactor matrix involves products of $n - 1$ terms. A and $\frac{1}{\det A} C^T$ might cancel each other. This is much easier to see from our formula for the determinant than when using Gauss-Jordan elimination.

To more formally verify the formula, we'll check that $AC^T = (\det A)I$.

$$AC^T = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix}.$$

The entry in the first row and first column of the product matrix is:

$$\sum_{j=1}^{n} a_{1j} C_{j1} = \det A.$$

(This is just the cofactor formula for the determinant.) This happens for every entry on the diagonal of AC^T.

To finish proving that $AC^T = (\det A)I$, we just need to check that the off-diagonal entries of AC^T are zero. In the two by two case, multiplying the entries in row 1 of A by the entries in column 2 of C^T gives $a(-b) + b(a) = 0$. This is the determinant of $A_s = \begin{bmatrix} a & b \\ a & b \end{bmatrix}$. In higher dimensions, the product of the first row of A and the last column of C^T equals the determinant of a matrix whose first and last rows are identical. This happens with all the off diagonal matrices, which confirms that $A^{-1} = \frac{1}{\det A} C^T$.

This formula helps us answer questions about how the inverse changes when the matrix changes.

Cramer's Rule for $\mathbf{x} = A^{-1}\mathbf{b}$

We know that if $A\mathbf{x} = \mathbf{b}$ and A is nonsingular, then $\mathbf{x} = A^{-1}\mathbf{b}$. Applying the formula $A^{-1} = C^T/\det A$ gives us:

$$\mathbf{x} = \frac{1}{\det A} C^T \mathbf{b}.$$

Cramer's rule gives us another way of looking at this equation. To derive this rule we break $\mathbf{x}$ down into its components. Because the i'th component of $C^T\mathbf{b}$ is a sum of cofactors times some number, it is the determinant of some matrix B_j.

$$x_j = \frac{\det B_j}{\det A},$$

where B_j is the matrix created by starting with A and then replacing column j with $\mathbf{b}$, so:

$$B_1 = \begin{bmatrix} \mathbf{b} & \begin{array}{c}\text{last n-1}\\ \text{columns}\\ \text{of } A\end{array} \end{bmatrix} \quad \text{and}$$

$$B_n = \begin{bmatrix} \begin{array}{c}\text{first n-1}\\ \text{columns}\\ \text{of } A\end{array} & \mathbf{b} \end{bmatrix}.$$

This agrees with our formula $x_1 = \frac{\det B_1}{\det A}$. When taking the determinant of B_1 we get a sum whose first term is b_1 times the cofactor C_{11} of A.

Computing inverses using Cramer's rule is usually less efficient than using elimination.

$|\det A|$ = volume of box

Claim: $|\det A|$ is the volume of the box (*parallelepiped*) whose edges are the column vectors of A. (We could equally well use the row vectors, forming a different box with the same volume.)

If $A = I$, then the box is a unit cube and its volume is 1. Because this agrees with our claim, we can conclude that the volume obeys determinant property 1.

If $A = Q$ is an orthogonal matrix then the box is a unit cube in a different orientation with volume $1 = |\det Q|$. (Because Q is an orthogonal matrix, $Q^T Q = I$ and so $\det Q = \pm 1$.)

Swapping two columns of A does not change the volume of the box or (remembering that $\det A = \det A^T$) the absolute value of the determinant (property 2). If we show that the volume of the box also obeys property 3 we'll have proven $|\det A|$ equals the volume of the box.

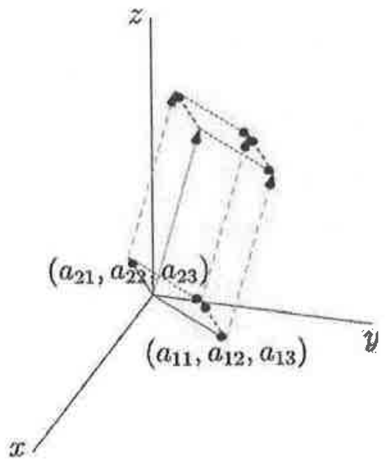


Figure 1: The box whose edges are the column vectors of A.

If we double the length of one column of A, we double the volume of the box formed by its columns. Volume satisfies property 3(a).

Property 3(b) says that the determinant is linear in the rows of the matrix:

$$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}.$$

Figure 2 illustrates why this should be true.

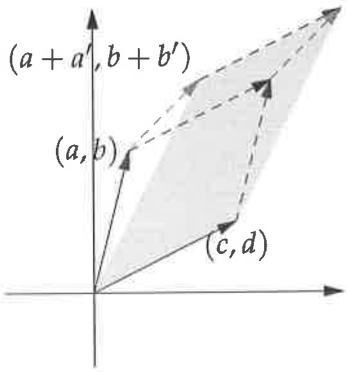


Figure 2: Volume obeys property 3(b).

Although it's not needed for our proof, we can also see that determinants obey property 4. If two edges of a box are equal, the box flattens out and has no volume.

Important note: If you know the coordinates for the corners of a box, then computing the volume of the box is as easy as calculating a determinant. In particular, the area of a parallelogram with edges $\begin{bmatrix} a \\ b \end{bmatrix}$ and $\begin{bmatrix} c \\ d \end{bmatrix}$ is $ad - bc$. The area of a triangle with edges $\begin{bmatrix} a \\ b \end{bmatrix}$ and $\begin{bmatrix} c \\ d \end{bmatrix}$ is half the area of that parallelogram, or $\frac{1}{2}(ad - bc)$. The area of a triangle with vertices at (x_1, y_1), (x_2, y_2) and (x_3, y_3) is:

$$\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

Unit 2 Lecture 6:

Formula for det A (n! terms)
Cofactor formula
Tridiagonal matrices

① det I = 1
② Sign Reverse with row-exchange
③ det is linear in each row separately.

Derivation of determinants: 2x2, 3x3 (missed)

$$\det A = \sum_{n!\ \text{terms}} \pm\, a_{1\alpha}\, a_{2\beta}\, a_{3\gamma} \cdots a_{n\omega}$$

$(\alpha, \beta, \gamma, \dots, \omega) = $ perm of $(1, 2, \dots, n)$

* A matrix is singular if we can find something in its nullspace.

Cofactor of $a_{ij} = C_{ij}$

$\pm \det \begin{pmatrix} n-1 \text{ matrix} \\ \text{with row } i \text{ col } j \text{ erased} \end{pmatrix}$

$+$: $i+j$ even
$-$: $i+j$ odd

$$\begin{vmatrix} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{vmatrix}$$

$$\boxed{\begin{array}{l} \text{Cofactor formula (along row 1)} \\ \det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} \end{array}}$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad + b(-c) = ad - bc$$

Ex

$$A_4 = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix}$$

tridiagonal determinant

$|A_1| = 1$ $\quad |A_2| = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 0$ $\quad |A_3| = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = -1$

$|A_4| = 1 \cdot |A_3| - 1\,|A_2|$

$|A_n| = |A_{n-1}| - |A_{n-2}|$

$|A_4| = -1$

Recitation $\Rightarrow$

Unit 2 Lecture 7

1. Formula of A^{-1}
2. Cramers Rule for $x = A^{-1}b$
3. $|\det A| =$ volume of box

Cofactor formula

$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$ for row i

---

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

$$A^{-1} = \frac{1}{\det A} C^T$$

Matrix of cofactors

Check $AC^T = (\det A)I$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} C_{11} & & C_{n1} \\ C_{12} & & \vdots \\ \vdots & & \vdots \\ C_{1n} & & C_{nn} \end{bmatrix} = \begin{bmatrix} \det A & & \text{All zeros} \\ & \det A & \\ \text{All zeros} & & \det A \end{bmatrix}$$

The reason I get zero is that we are taking ~~determinant~~ & cofactors of ~~zero~~ singular matrix.

2nd Application

CRAMER'S RULE

$$Ax = b$$
$$x = A^{-1} b$$
$$x = \frac{1}{\det A} C^T b$$

$$x_1 = \frac{\det B_1}{\det|A|}$$
$$x_2 = \frac{\det B_2}{\det A}$$
$$x_j = \frac{\det B_j}{\det A}$$

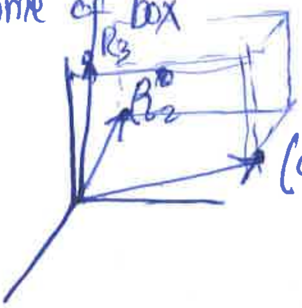
$B_1 = \begin{bmatrix} b & \begin{matrix} n-1 \\ \text{columns of} \\ A \end{matrix} \end{bmatrix}$ = A with col1 replaced by b

B_j = A with col j replaced by b.

$\boxed{\text{Mathematics for a million} \quad \text{~~Man~~ Book}}$

Say $A = 3 \times 3$

$|\det A|$ = volume of box



(a_{11}, a_{12}, a_{13})

$\boxed{\text{each row of } A \text{ is an edge of a box}}$

$A = I \longrightarrow$ Its a cube

Suppose $A = Q$ (orthogonal matrix) (columns are $\perp$ unit vectors) $\longrightarrow$ Its a rotated cube

i.e $Q^T Q = I$

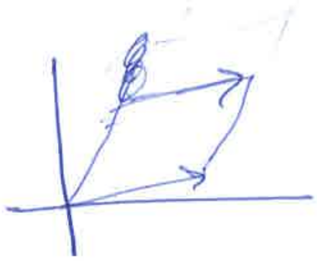
why $\det|Q| = \pm 1$?

$$\det(Q^T Q) = \det I$$
$$\det Q^T \det Q = \det I$$
$$|Q^T|\,|Q| = 1$$
$$|Q|\,|Q| = 1 \Rightarrow \boxed{|Q|^2 = 1} \text{ Ans}$$

* If we double an edge of Q, then volume of box/cube will increase & it will become rectangle. The determinant will be doubled.

$|\det A|$ = volume of box

$\begin{matrix} a & b \\ c & d \end{matrix}$



Area of parallelogram $= \det\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$= ad - bc$ for //gm

$= \frac{1}{2}(ad-bc)$ for triangle

Unit II : lecture # 8

If the product Ax points in the same direction as the vector X, we say that X is an eigenvector of A. Eigenvalues and eigenvectors describe what happens when a matrix is multiplied by a vector. In this session, we learn how to find the eigenvalues and eigenvectors of a matrix.

Eigenvalues and eigenvectors

The subject of eigenvalues and eigenvectors will take up most of the rest of the course. We will again be working with square matrices. Eigenvalues are special numbers associated with a matrix and eigenvectors are special vectors.

Eigenvectors and eigenvalues

A matrix A *acts on* vectors $\mathbf{x}$ like a function does, with input $\mathbf{x}$ and output $A\mathbf{x}$. *Eigenvectors* are vectors for which $A\mathbf{x}$ is parallel to $\mathbf{x}$. In other words:

$$A\mathbf{x} = \lambda\mathbf{x}.$$

In this equation, $\mathbf{x}$ is an eigenvector of A and λ is an *eigenvalue* of A.

Eigenvalue 0

If the eigenvalue λ equals 0 then $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$. Vectors with eigenvalue 0 make up the nullspace of A; if A is singular, then $\lambda = 0$ is an eigenvalue of A.

Examples

Suppose P is the matrix of a projection onto a plane. For any $\mathbf{x}$ in the plane $P\mathbf{x} = \mathbf{x}$, so $\mathbf{x}$ is an eigenvector with eigenvalue 1. A vector $\mathbf{x}$ perpendicular to the plane has $P\mathbf{x} = \mathbf{0}$, so this is an eigenvector with eigenvalue $\lambda = 0$. The eigenvectors of P span the whole space (but this is not true for every matrix).

The matrix $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has an eigenvector $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ with eigenvalue 1 and another eigenvector $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ with eigenvalue -1. These eigenvectors span the space. They are perpendicular because $B = B^T$ (as we will prove).

$\det(A - \lambda I) = 0$

An n by n matrix will have n eigenvalues, and their sum will be the sum of the diagonal entries of the matrix: $a_{11} + a_{22} + \cdots + a_{nn}$. This sum is the *trace* of the matrix. For a two by two matrix, if we know one eigenvalue we can use this fact to find the second.

Can we solve $A\mathbf{x} = \lambda\mathbf{x}$ for the eigenvalues and eigenvectors of A? Both λ and $\mathbf{x}$ are unknown; we need to be clever to solve this problem:

$$\begin{aligned} A\mathbf{x} &= \lambda\mathbf{x} \\ (A - \lambda I)\mathbf{x} &= \mathbf{0} \end{aligned}$$

In order for λ to be an eigenvector, $A - \lambda I$ must be singular. In other words, $\det(A - \lambda I) = 0$. We can solve this *characteristic equation* for λ to get n solutions.

If we're lucky, the solutions are distinct. If not, we have one or more *repeated eigenvalues*.

Once we've found an eigenvalue λ, we can use elimination to find the nullspace of $A - \lambda I$. The vectors in that nullspace are eigenvectors of A with eigenvalue λ.

Calculating eigenvalues and eigenvectors

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

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} \\ &= (3-\lambda)^2 - 1 \\ &= \lambda^2 - 6\lambda + 8. \end{aligned}$$

Note that the coefficient 6 is the trace (sum of diagonal entries) and 8 is the determinant of A. In general, the eigenvalues of a two by two matrix are the solutions to:

$$\lambda^2 - \operatorname{trace}(A) \cdot \lambda + \det A = 0.$$

Just as the trace is the sum of the eigenvalues of a matrix, the product of the eigenvalues of any matrix equals its determinant.

For $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$, the eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = 2$. We find the eigenvector $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for $\lambda_1 = 4$ in the nullspace of $A - \lambda_1 I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$.

x_2 will be in the nullspace of $A - 2I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. The nullspace is an entire line; x_2 could be any vector on that line. A natural choice is $x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Note that these eigenvectors are the same as those of $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Adding $3I$ to the matrix $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ added 3 to each of its eigenvalues and did not change its eigenvectors, because $A\mathbf{x} = (B + 3I)\mathbf{x} = \lambda\mathbf{x} + 3\mathbf{x} = (\lambda + 3)\mathbf{x}$.

A caution

Similarly, if $A\mathbf{x} = \lambda\mathbf{x}$ and $B\mathbf{x} = \alpha\mathbf{x}$, $(A + B)\mathbf{x} = (\lambda + \alpha)\mathbf{x}$. It would be nice if the eigenvalues of a matrix sum were always the sums of the eigenvalues, but this is only true if A and B have the same eigenvectors. The eigenvalues of the product AB aren't usually equal to the products $\lambda(A)\lambda(b)$, either.

Complex eigenvalues

The matrix $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ rotates every vector in the plane by 90°. It has trace $0 = \lambda_1 + \lambda_2$ and determinant $1 = \lambda_1 \cdot \lambda_2$. Its only real eigenvector is the zero vector; any other vector's direction changes when it is multiplied by Q. How will this affect our eigenvalue calculation?

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} \\ &= \lambda^2 + 1. \end{aligned}$$

$\det(A - \lambda I) = 0$ has solutions $\lambda_1 = i$ and $\lambda_2 = -i$. If a matrix has a complex eigenvalue $a + bi$ then the *complex conjugate* $a - bi$ is also an eigenvalue of that matrix.

Symmetric matrices have real eigenvalues. For *antisymmetric* matrices like Q, for which $A^T = -A$, all eigenvalues are imaginary ($\lambda = bi$).

Triangular matrices and repeated eigenvalues

For triangular matrices such as $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$, the eigenvalues are exactly the entries on the diagonal. In this case, the eigenvalues are 3 and 3:

$$\begin{aligned} \det(A - \lambda \det I) &= \begin{vmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{vmatrix} \\ &= (3-\lambda)(3-\lambda) \quad \Big(= (a_{11} - \lambda)(a_{22} - \lambda) \Big) \\ &= 0, \end{aligned}$$

so $\lambda_1 = 3$ and $\lambda_2 = 3$. To find the eigenvectors, solve:

$$(A - \lambda I)\mathbf{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0}$$

to get $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. There is no independent eigenvector $\mathbf{x}_2$.

LA - Lect 21

EigenValues – EigenVectors

$\det[A - \lambda I] = 0$

$\text{TRACE} = \lambda_1 + \lambda_2 + \cdots + \lambda_N$

MATRICES ARE SQUARE

Q

Whats an EigenVector ?

In calculus, we have seen $f(x)$. "x" goes in & $f(x)$ comes out.
In linear Algebra, we have seen Ax, A = matrix, X = vector. "X" goes in & outcome is Ax. We are interested in those vectors that have the same direction as the input "X" to Ax. So those vectors in Ax that are parallel to "X" are eigen vectors.

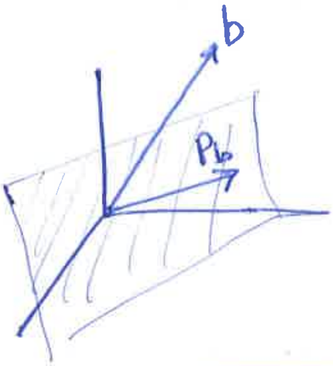
So $\underbrace{Ax}_{\text{EigenVectors}}$ parallel to X

i.e $\boxed{Ax = \lambda X}$ {λ can be -ve, zero or even some imaginary number}

eigenvalue

If A is singular {i.e it takes "X" to zero} the $\lambda = 0$ is an eigenvalue.

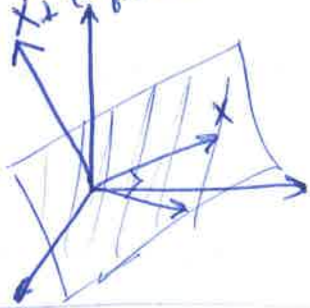
$\boxed{\text{"}\lambda\text{" \& "X" are unknown}}$



what are the x's & λ's for projection matrix (P)?

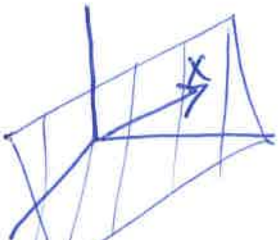
Ans: Any X in plane: $PX = P$
$\Rightarrow \lambda = 1$ when $PX = P$

we have whole plane of eigenvectors. Are there any other eigenvector? Ans: Yes. Two of them



in plane & one EV pointing outside plane.

Let say b is not in plane highlighted above. we have projection matrix P which ~~Pb~~ will project b in plane, highlighted plane. Now, ~~what~~ how can get the output vector in the same direction as input vector (X)?



Ans: If $\vec{X}$ is in the ~~plane of~~ col. space of Projection matrix. OR if $\vec{X}$ is in the plane of $\vec{Px}$.

Thus Any $X \perp$ plane $PX = 0$
OR $PX = (0)X$ $\quad \lambda = 0$

Another Example

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

what vector should be multiplied to it & we get the resultant in same direction?

A = permutation matrix

Let $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, then $AX = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \lambda = 1$

input $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ has the same direction as AX (output). Thus $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is eigenvector.

Since we have 2x2 matrix, thus we can have 2nd eigenvector.

$$X = \begin{bmatrix} -1 \\ +1 \end{bmatrix} \Rightarrow AX = \begin{bmatrix} +1 \\ -1 \end{bmatrix} \Rightarrow \boxed{\begin{matrix} AX = (-1)X \\ \lambda = -1 \end{matrix}}$$

Thus $\left.\begin{matrix} AX = X \\ AX = -X \end{matrix}\right\}$ in our examples. $\left\}\begin{matrix}\text{here two eigenvectors are} \\ \perp.\end{matrix}\right.$

<u>Fact</u> N×N matrix will have "N" Eigenvalues.

<u>FACT</u> Sum of λ's (eigenvalues) $= a_{11} + a_{22} + \cdots + a_{nn}$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{bmatrix}$$

<u>Q</u>

How to Solve $AX = \lambda X$? $\boxed{X, \lambda \text{ are variables}}$

Rewrite $(AX - \lambda X) = 0$

$(A - \lambda I)X = 0$

Since $X \neq 0, \Rightarrow A - \lambda I = 0$

Matrix $A - \lambda I$ should be singular. We know that det(sing) = 0

$\Rightarrow |A - \lambda I| = 0$ (Characteristic/Eigen Value equation) (Find λ First).

$\boxed{\text{Repeated } \lambda\text{'s are trouble}}$

After λ, we will use elimination to find X.

<u>Example</u> $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ → Symmetric Matrix, → Diagonals are constant } Eigenvalues will be nice ~~sqmme~~ real numbers.

$$\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 - 1$$

$$= \lambda^2 - \underbrace{6}_{\text{trace}}\lambda + 8 \rightarrow \text{det of matrix } A$$

$= \lambda^2 - (\text{trace})\lambda + \det$ (In 2x2 case)

$\Rightarrow \lambda^2 - 6\lambda + 8 = 0 \qquad \lambda = 4, \ \lambda = 2$

Eigenvalues

$\lambda_1 = 4 \qquad \lambda_2 = 2$

Eigenvectors: They are guys in null space of $A - \lambda x$, or when we make A singular by shifting it by λx.

EVector $\lambda_1 = 4 \quad x_1 = ??$

$$A - \lambda I = A - 4I = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} - 4\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A - 4I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \longrightarrow$$ singular Matrix (There will be someone in its nullspace other than $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$)

for x_1: whats the vector in nullspace of $A-4I$ that goes in & comes out with same direction?

$$(A-4I)x_1$$
$$= \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} x_1 \qquad \text{If } x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

when $\lambda_2 = 2$

$$A - 2I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \longrightarrow$$ singular Matrix

Here whole line is in $N(A-2I)$. Thus we need basis for that line. We will use special solution i.e $x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ will be another eigen vector

Check

$$\underset{A}{\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}} \underset{X_{\text{Eigen Vector}}}{\begin{bmatrix} -1 \\ 1 \end{bmatrix}} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \underset{\text{Eigen Value}}{2} \underset{\text{Eigen Vector}}{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}$$

Elimination

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} [x] = 0 \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$\boxed{x_1 + x_2 = 0} \rightarrow ??$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \longrightarrow x_1 + x_2 = 0$$

Set $x_1 = -1 \Rightarrow x_2 = 1$ } special sol

$$\Rightarrow X_{EV} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$\Rightarrow$ $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ $\lambda_1 = 4$ $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\lambda_2 = 2$ $x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ Let $A = A_1$

Previously we solved

$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $\lambda_1 = 1$ $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\lambda_2 = -1$ $x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ Let $A = A_2$ here

How can we relate the above two matrices A_1 & A_2??

$A_1 = A_2 + 3I$ & we can see that their eigenvectors are same but their eigen values are just shifted by 3.

Suppose if we have $Ax = \lambda x$

Then $(A+3I)x = \lambda x + 3x$

$= (\lambda+3)x$

Modifying A by I changes eigenvalues but not eigen vector

Not So great (A+B, AB are not so great)

Suppose If $Ax = \lambda x$, B has eigenvalues α

then $Bx = \alpha x \longrightarrow$ eq(i)

$\Rightarrow$ $(A+B)x \neq (\lambda+\alpha)x$ B/c it is not necessary that eigen vectors of A are necessarily eigen vectors of B.

Thus we can modify eq(i)

$By = \alpha y$ (Because B is not multiple of I)

Example Let take an example of rotational matrix. It will rotate every vector by 90°.

$$Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{(Also orthogonal matrix.)}$$

A Priori we know sum of two eigenvalues $= 0$ ($\because$ trace $= 0$)

$$\text{trace} = 0 + 0 = \lambda_1 + \lambda_2$$

$$\det = 1 = \lambda_1 \lambda_2$$ (See previous example where we have seen that product of 2 E.V. is equal to determinant $x^2 - (\text{sum of roots})x + \text{product of roots} = 0$)

(in rotation matrix)

det of matrix for which evalue is sought.

Some thing will go wrong here?

What vector will go in & come out parallel to itself after rotation..

So in Rotation matrix, what vector can go in & come out in the same direction as they went inside? What will be the eigenvector? we have trouble here.

We have to solve

$$\underbrace{\lambda_1 + \lambda_2 = 0 \qquad \lambda_1 \lambda_2 = 1}_{\text{How?}}$$

$$\det(Q - \lambda I) = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1$$

$$\lambda^2 + 1 = 0 \qquad \lambda^2 = -1$$

$$\underbrace{\lambda_1 = i \qquad \lambda_2 = -i}_{\text{Not real numbers even with real matrix } Q}$$

λ_1, λ_2 are complex conjugate.

Even worst case

Suppose
$$A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$$
If the matrix is tridiagonal, you can read its eigenvalue. They are on diagonal

$$\left.\begin{matrix} \lambda_1 = 3 \\ \lambda_2 = 3 \end{matrix}\right\} \text{Apriori}$$

$$\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{vmatrix} = (3-\lambda)(3-\lambda) \Rightarrow \lambda_1 = 3, \ \lambda_2 = 3$$

Eigenvector:

$$(A-\lambda I)x=0$$

$$\underbrace{(A-3I)}\,x=0 \Rightarrow \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \end{bmatrix}=\begin{bmatrix} 0 \end{bmatrix}$$

$$\Rightarrow \quad x_1=\begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ with } \lambda_1=3$$

x_2 = independent eigenvector doesn't exist. No 2nd independent eigenvector

Thus due to x_2, we can say that "A" is degenerate.

Recitation ??

If $Av=\lambda v$; $\lambda \rightarrow$ eigenvalue of eigenvector v
$v \rightarrow$ eigenvector of A

then i) $A^2v = A(Av) = A(\lambda v) = \lambda(Av) = \lambda(\lambda v) = \lambda^2 v$

ii) $A^{-1}v \underset{(\lambda\neq 0)}{=} A^{-1}\left(\dfrac{Av}{\lambda}\right) = A^{-1}A\left(\dfrac{v}{\lambda}\right) = I\left(\dfrac{v}{\lambda}\right) = \dfrac{1}{\lambda}v$

iii) $(A^{-1}-I)v = A^{-1}v - v = \dfrac{v}{\lambda} v - v$
$= (\lambda^{-1}-1)v$

If we know the eigenvalues & eigenvector of A, then we can evaluate eigenvalues & eigenvector of A^2v, $A^{-1}v$, $A^{-1}-I$.

Unit II : lecture 9

If A has "n" independent eigenvectors, we can write $A = S\Lambda S^{-1}$, where Λ is a diagonal matrix containing the eigenvalues of A. This allows us to easily compute powers of A which in turns allows us to solve difference equations $U_{K+1} = AU_K$.

Diagonalization and powers of A

We know how to find eigenvalues and eigenvectors. In this lecture we learn to *diagonalize* any matrix that has n independent eigenvectors and see how diagonalization simplifies calculations. The lecture concludes by using eigenvalues and eigenvectors to solve *difference equations*.

Diagonalizing a matrix $S^{-1}AS = \Lambda$

If A has n linearly independent eigenvectors, we can put those vectors in the columns of a (square, invertible) matrix S. Then

$$\begin{aligned} AS &= A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{x}_1 & \lambda_2\mathbf{x}_2 & \cdots & \lambda_n\mathbf{x}_n \end{bmatrix} \\ &= S \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} = S\Lambda. \end{aligned}$$

Note that Λ is a diagonal matrix whose non-zero entries are the eigenvalues of A. Because the columns of S are independent, S^{-1} exists and we can multiply both sides of $AS = S\Lambda$ by S^{-1}:

$$S^{-1}AS = \Lambda.$$

Equivalently, $A = S\Lambda S^{-1}$.

Powers of A

What are the eigenvalues and eigenvectors of A^2?

$$\begin{aligned} \text{If } A\mathbf{x} &= \lambda\mathbf{x}, \\ \text{then } A^2\mathbf{x} &= \lambda A\mathbf{x} = \lambda^2\mathbf{x}. \end{aligned}$$

The eigenvalues of A^2 are the squares of the eigenvalues of A. The eigenvectors of A^2 are the same as the eigenvectors of A. If we write $A = S\Lambda S^{-1}$ then:

$$A^2 = S\Lambda S^{-1}S\Lambda S^{-1} = S\Lambda^2 S^{-1}.$$

Similarly, $A^k = S\Lambda^k S^{-1}$ tells us that raising the eigenvalues of A to the kth power gives us the eigenvalues of A^k, and that the eigenvectors of A^k are the same as those of A.

Theorem: If A has n independent eigenvectors with eigenvalues λ_i, then $A^k \to 0$ as $k \to \infty$ if and only if all $|\lambda_i| < 1$.

A is guaranteed to have n independent eigenvectors (and be *diagonalizable*) if all its eigenvalues are different. Most matrices do have distinct eigenvalues.

Repeated eigenvalues

If A has repeated eigenvalues, it may or may not have n independent eigenvectors. For example, the eigenvalues of the identity matrix are all 1, but that matrix still has n independent eigenvectors.

If A is the triangular matrix $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ its eigenvalues are 2 and 2. Its eigenvectors are in the nullspace of $A - \lambda I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ which is spanned by $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. This particular A does not have two independent eigenvectors.

Difference equations $\mathbf{u}_{k+1} = A\mathbf{u}_k$

Start with a given vector $\mathbf{u}_0$. We can create a sequence of vectors in which each new vector is A times the previous vector: $\mathbf{u}_{k+1} = A\mathbf{u}_k$. $\mathbf{u}_{k+1} = A\mathbf{u}_k$ is a *first order difference equation*, and $\mathbf{u}_k = A^k\mathbf{u}_0$ is a solution to this system.

We get a more satisfying solution if we write $\mathbf{u}_0$ as a combination of eigenvectors of A:

$$\mathbf{u}_0 = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n = S\mathbf{c}.$$

Then:

$$A\mathbf{u}_0 = c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 + \cdots + c_n\lambda_n\mathbf{x}_n$$

and:

$$\mathbf{u}_k = A^k\mathbf{u}_0 = c_1\lambda_1^k\mathbf{x}_1 + c_2\lambda_2^k\mathbf{x}_2 + \cdots + c_n\lambda_n^k\mathbf{x}_n = \Lambda^k S\mathbf{c}.$$

Fibonacci sequence

The Fibonacci sequence is $0, 1, 1, 2, 3, 5, 8, 13, ...$ In general, $F_{k+2} = F_{k+1} + F_k$. If we could understand this in terms of matrices, the eigenvalues of the matrices would tell us how fast the numbers in the sequence are increasing.

$\mathbf{u}_{k+1} = A\mathbf{u}_k$ was a first order system. $F_{k+2} = F_{k+1} + F_k$ is a second order scalar equation, but we can convert it to first order linear system by using a clever trick. If $\mathbf{u}_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$, then:

$$\begin{aligned} F_{k+2} &= F_{k+1} + F_k && (1) \\ F_{k+1} &= F_{k+1}. && (2) \end{aligned}$$

is equivalent to the first order system $\mathbf{u}_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{u}_k$.

What are the eigenvalues and eigenvectors of $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$? Because A is symmetric, its eigenvalues will be real and its eigenvectors will be orthogonal.

Because A is a two by two matrix we know its eigenvalues sum to 1 (the trace) and their product is -1 (the determinant).

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - \lambda - 1$$

Setting this to zero we find $\lambda = \frac{1 \pm \sqrt{1+4}}{2}$; i.e. $\lambda_1 = \frac{1}{2}(1+\sqrt{5}) \approx 1.618$ and $\lambda_2 = \frac{1}{2}(1-\sqrt{5}) \approx -.618$. The growth rate of the F_k is controlled by λ_1, the only eigenvalue with absolute value greater than 1. This tells us that for large k, $F_k \approx c_1 \left(\frac{1+\sqrt{5}}{2}\right)^k$ for some constant c_1. (Remember $\mathbf{u}_k = A^k \mathbf{u}_0 = c_1 \lambda_1^k x_1 + c_2 \lambda_2^k x_2$, and here λ_2^k goes to zero since $|\lambda_2| < 1$.)

To find the eigenvectors of A note that:

$$(A - \lambda I)\mathbf{x} = \begin{bmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{bmatrix} \mathbf{x}$$

equals $\mathbf{0}$ when $\mathbf{x} = \begin{bmatrix} \lambda \\ 1 \end{bmatrix}$, so $\mathbf{x}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$.

Finally, $\mathbf{u}_0 = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2$ tells us that $c_1 = -c_2 = \frac{1}{\sqrt{5}}$.

Because $\begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \mathbf{u}_k = c_1 \lambda_1^k x_1 + c_2 \lambda_2^k x_2$, we get:

$$F_k = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^k - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^k.$$

Using eigenvalues and eigenvectors, we have found a *closed form expression* for the Fibonacci numbers.

Summary: When a sequence evolves over time according to the rules of a first order system, the eigenvalues of the matrix of that system determine the long term behavior of the series. To get an exact formula for the series we find the eigenvectors of the matrix and then solve for the coefficients $c_1, c_2, \ldots$

LA: Lec 22

Diagonalizing a matrix $S^{-1}AS = \Lambda$

Powers of A / equation $U_{k+1} = AU_k$

$A - \lambda I$ singular

$Ax = \lambda x$ → eigenvector

λ: Eigen value

Suppose we have "n" linearly independent eigenvectors of A. Put them in column of S. S is eigenvector matrix.

$$AS = A\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \cdots & \lambda_n x_n \end{bmatrix}$$

$Ax_1 = \lambda_1 x_1$
$Ax_2 = \lambda_2 x_2$

$$= \underbrace{\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & \\ 0 & 0 & & \lambda_N \end{bmatrix}}_{\Lambda}$$

diagonal eigen value matrix

$$= S\Lambda$$

As $AS = S\Lambda$

$\boxed{S^{-1}AS = \Lambda}$ OR $ASS^{-1} = S\Lambda S^{-1}$

$\Downarrow$

$S^{-1}AS = \underbrace{S^{-1}S}_{I}\Lambda$

$\boxed{A = S\Lambda S^{-1}}$

We have seen that ~~[struck]~~ square matrices might have dependent eigenvectors in last lecture. ~~therefore we~~ Here we have to take S^{-1}, therefore we assume that "S" has independent eigenvectors which makes S^{-1} possible.

Now If $Ax = \lambda x$, Multiply by A

$$A^2 x = \lambda A x$$
$$A^2 x = \lambda(\lambda x)$$
$$\boxed{A^2 x = \lambda^2 x}$$

$\rightarrow$ conclusion: A^2 will have eigenvalues λ^2. It has same eigenvector as of A.

We can also see the above fact from $A = S\Lambda S^{-1}$

If $A = S\Lambda S^{-1}$

$$A^2 = (S\Lambda \underbrace{S^{-1})(S}_{I}\Lambda S^{-1})$$

$$\boxed{A^2 = S\Lambda^2 S^{-1}}$$

Again, S is same mean eigenvectors for A^2 are same as of A. And eigenvalues that are stored in diagonal of Λ^2 are the square of eigenvalues of A.

Now $$\boxed{A^K = S\Lambda^K S^{-1}}$$

Its telling us that eigenvalues of A^K are λ^K. But eigenvectors are same to the original A & λ.

Eigenvalues tells us ~~the~~ about powers of matrix. It was not possible before.

Theorem

$A^K \rightarrow 0$ as $K \rightarrow \infty$

if all $|\lambda_i| < 1$

why?? In $S\Lambda^K S^{-1}$, S & S^{-1} are not changing. The only thing that is variable is our Λ^K which contains eigenvalues. So if all eigenvalues have magnitude <1, the raising the power of Λ^K will make them smaller & smaller.

To work with diagonalization, we must ~~has~~ have "n" independent eigenvectors in S otherwise S^{-1} does not exist in $S\Lambda S^{-1} = A$.

Which matrices ~~are~~ can be diagonalize?

FACT:

A is sure to have "n" independent eigenvector (and be diagonalizable)

if all the λ's are different (No repeated λ's)

eig(rand(10,10))

10 distinct eigenvalues

(Proof in book)

* Repeated eigenvalues // may or may not have "n" independent vectors

* Suppose $A = I_{10\times10}$. Its Eigenvalues are 1 & Evector are taken as each column. here A is already diagonal matrix

* Suppose A is triangular matrix.

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$$ (Troubling case)

Its Eigenvalues = 2, 2

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

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

$$\boxed{\lambda = 2}$$

For EVectors: $A - 2I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ & we will look for nullspace...

Nullspace is 1D. Only Eigenvector $X_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Thus A is not diagonalisable because we cannot form S with "n" independent eigenvectors. Here n=2.

Thus there are some matrices that are not covered by diagonalization but majority would.

In above example with $\lambda = 2$, the algebraic multiplicity is "2" b/c $(2-\lambda)^2 = 0 \Rightarrow \lambda = 2, 2$ but geometric multiplicity is 1 as we have only one-eigenvector.

Equation $U_{k+1} = A U_k$, first order difference equation

Start with given vector u_0

$U_1 = A U_0$, $U_2 = A^2 U_0$ $\boxed{U_k = A^k u_0}$

To solve (really solve): write

$$U_0 = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

($x_1, x_2, \cdots x_n$ are eigen vector)

Lets suppose eigenvectors are normalised to unit vectors

$$\Rightarrow A U_0 = c_1 \lambda_1 x_1 + c_2 \lambda_2 x_2 + \cdots + c_n \lambda_n x_n$$

$$\Rightarrow A^{100} U_0 = c_1 \lambda_1^{100} x_1 + c_2 \lambda_2^{100} x_2 + \cdots + c_n \lambda_n^{100} x_n$$

$$= \Lambda^{100} S c$$

$$\Rightarrow U_{100} = \Lambda^{100} S c$$

Fibonacci example: $0, 1, 1, 2, 3, 5, 8, 13, \ldots$ $\boxed{F_{100} = ?}$

Q: How fast fibonacci numbers are growing?
The answer lies in eigenvalues

$$F_{k+2} = F_{k+1} + F_k$$ (Its not a system, Its a 2nd order equation)

Trick

To Turn it into 2×2 system ~~of~~ ~~[illegible]~~ ~~equation~~

Let $U_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$

$$\Rightarrow \left.\begin{matrix} F_{k+2} = F_{k+1} + F_k \\ F_{k+1} = F_{k+1} \end{matrix}\right] \text{our system}$$

Adding another equation

~~[illegible]~~

$$\Rightarrow U_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} \Rightarrow U_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$$

forming system into vector

Changing 2nd order scalar problem to 1st order system

$$\Rightarrow U_{K+1} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}_{A} \underbrace{U_K}_{\begin{bmatrix} F_K \\ F_{K+1} \end{bmatrix}}$$ Lets find its eigen values & eigenvectors.

Now $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ Its symmetric. Its eigenvalues are real & eigenvectors are orthogonal

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix}$$

Apriori $\lambda_1 + \lambda_2 = 1$

$\lambda_1 \cdot \lambda_2 = \det(A) = -1$

$$= \lambda^2 - \lambda - 1 \quad (\text{Compare it with } F_{K+1} - F_K - 1 = 0)$$

$$\Rightarrow \lambda^2 - \lambda - 1 = 0$$

$$\lambda = \frac{1 \pm \sqrt{1+4}}{2}$$

$$\boxed{\lambda_1 = \frac{1+\sqrt{5}}{2}} \quad \boxed{\lambda_2 = \frac{1-\sqrt{5}}{2}}$$

$$\underbrace{\boxed{\lambda_1 = \tfrac{1}{2}(1+\sqrt{5})}}_{\approx 1.618} \quad \underbrace{\boxed{\lambda_2 = \tfrac{1}{2}(1-\sqrt{5})}}_{\approx -.618}$$

$\left.\begin{matrix} \lambda_1 > 1 \\ \lambda_2 < 1 \end{matrix}\right\}$ A is diagonalizable

How fast fibonacci numbers are increasing ?

$$F_{100} \approx \underbrace{\underbrace{C_1 \left(\frac{1+\sqrt{5}}{2}\right)^{100}}_{\text{Its heavy}} + \underbrace{C_2(-0.618)^{100}}_{\text{Its nothing as } -0.618 < 1}}_{\text{This shows system is evolving}}$$

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

For $\lambda_1 = \left(\frac{1+\sqrt{5}}{2}\right)$

$\lambda_2 = \left(\frac{1-\sqrt{5}}{2}\right)$

Since X_1 & X_2 are in $N(A - \lambda I)$ $\Rightarrow (A - \lambda I) X_1 = 0$

$\boxed{X_1, X_2 = \text{eigen vectors}}$ $\quad (A - \lambda I) X_2 = 0$

For x_1: $(A-\lambda I)x_1 = \begin{bmatrix} 1-\lambda_1 & 1 \\ 1 & -\lambda_1 \end{bmatrix}\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

For x_2: $(A-\lambda I)x_2 = \begin{bmatrix} 1-\lambda_2 & 1 \\ 1 & -\lambda_2 \end{bmatrix}\begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$x_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$ $\quad x_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$ (This trick is available in 2x2)

$\Rightarrow \quad u_0 = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

For c_1, c_2

$$c_1 x_1 + c_2 x_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

When things are evolving in time by 1st order system, eigen value will tell whether system is blowing or not.

Recitation?

Unit II : Lecture 10

We can copy Taylor's series for e^x to define e^{At} for a matrix A. If A is diagonalizable, we can use Λ to find the exact value of e^{At}. This allows to solve systems of differential equations $du/dt = Au$ the same way we solved equations like $dy/dt = ky$.

Differential equations and e^{At}

The system of equations below describes how the values of variables u_1 and u_2 affect each other over time:

$$\frac{du_1}{dt} = -u_1 + 2u_2$$
$$\frac{du_2}{dt} = u_1 - 2u_2.$$

Just as we applied linear algebra to solve a difference equation, we can use it to solve this differential equation. For example, the initial condition $u_1 = 1$, $u_2 = 0$ can be written $\mathbf{u}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Differential equations $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$

By looking at the equations above, we might guess that over time u_1 will decrease. We can get the same sort of information more safely by looking at the eigenvalues of the matrix $A = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix}$ of our system $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$. Because A is singular and its trace is -3 we know that its eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = -3$. The solution will turn out to include e^{-3t} and e^{0t}. As t increases, e^{-3t} vanishes and $e^{0t} = 1$ remains constant. Eigenvalues equal to zero have eigenvectors that are *steady state* solutions.

$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an eigenvector for which $A\mathbf{x}_1 = 0\mathbf{x}_1$. To find an eigenvector corresponding to $\lambda_2 = -3$ we solve $(A - \lambda_2 I)\mathbf{x}_2 = \mathbf{0}$:

$$\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \mathbf{x}_2 = 0 \quad \text{so} \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and we can check that $A\mathbf{x}_2 = -3\mathbf{x}_2$. The general solution to this system of differential equations will be:

$$\mathbf{u}(t) = c_1 e^{\lambda_1 t}\mathbf{x}_1 + c_2 e^{\lambda_2 t}\mathbf{x}_2.$$

Is $e^{\lambda_1 t}\mathbf{x}_1$ really a solution to $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$? To find out, plug in $\mathbf{u} = e^{\lambda_1 t}\mathbf{x}_1$:

$$\frac{d\mathbf{u}}{dt} = \lambda_1 e^{\lambda_1 t}\mathbf{x}_1,$$

which agrees with:

$$A\mathbf{u} = e^{\lambda_1 t} A\mathbf{x}_1 = \lambda_1 e^{\lambda_1 t}\mathbf{x}_1.$$

The two "pure" terms $e^{\lambda_1 t}\mathbf{x}_1$ and $e^{\lambda_2 t}\mathbf{x}_2$ are analogous to the terms $\lambda_i^k \mathbf{x}_i$ we saw in the solution $c_1\lambda_1^k\mathbf{x}_1 + c_2\lambda_2^k\mathbf{x}_2 + \cdots + c_n\lambda_n^k\mathbf{x}_n$ to the difference equation $\mathbf{u}_{k+1} = A\mathbf{u}_k$.

Plugging in the values of the eigenvectors, we get:

$$\mathbf{u}(t) = c_1 e^{\lambda_1 t}\mathbf{x}_1 + c_2 e^{\lambda_2 t}\mathbf{x}_2 = c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

We know $\mathbf{u}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, so at $t = 0$:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

$c_1 = c_2 = 1/3$ and $\mathbf{u}(t) = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{3}e^{-3t}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

This tells us that the system starts with $u_1 = 1$ and $u_2 = 0$ but that as t approaches infinity, u_1 decays to $2/3$ and u_2 increases to $1/3$. This might describe stuff moving from u_1 to u_2.

The steady state of this system is $\mathbf{u}(\infty) = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$.

Stability

Not all systems have a steady state. The eigenvalues of A will tell us what sort of solutions to expect:

1. Stability: $\mathbf{u}(t) \to 0$ when $\mathrm{Re}(\lambda) < 0$.
2. Steady state: One eigenvalue is 0 and all other eigenvalues have negative real part.
3. Blow up: if $\mathrm{Re}(\lambda) > 0$ for any eigenvalue λ.

If a two by two matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has two eigenvalues with negative real part, its trace $a + d$ is negative. The converse is not true: $\begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}$ has negative trace but one of its eigenvalues is 1 and e^{1t} blows up. If A has a positive determinant and negative trace then the corresponding solutions must be stable.

Applying S

The final step of our solution to the system $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$ was to solve:

$$c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

In matrix form:

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

or $S\mathbf{c} = \mathbf{u}(0)$, where S is the eigenvector matrix. The components of $\mathbf{c}$ determine the contribution from each pure exponential solution, based on the initial conditions of the system.

In the equation $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$, the matrix A couples the pure solutions. We set $\mathbf{u} = S\mathbf{v}$, where S is the matrix of eigenvectors of A, to get:

$$S\frac{d\mathbf{v}}{dt} = AS\mathbf{v}$$

or:

$$\frac{d\mathbf{v}}{dt} = S^{-1}AS\mathbf{v} = \Lambda\mathbf{v}.$$

This diagonalizes the system: $\frac{dv_i}{dt} = \lambda_i v_i$. The general solution is then:

$$\begin{aligned} \mathbf{v}(t) &= e^{\Lambda t}\mathbf{v}(0), \quad \text{and} \\ \mathbf{u}(t) &= Se^{\Lambda t}S^{-1}\mathbf{v}(0) = e^{At}\mathbf{u}(0). \end{aligned}$$

Matrix exponential e^{At}

What does e^{At} mean if A is a matrix? We know that for a real number x,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots.$$

We can use the same formula to define e^{At}:

$$e^{At} = I + At + \frac{(At)^2}{2} + \frac{(At)^3}{6} + \cdots.$$

Similarly, if the eigenvalues of At are small, we can use the geometric series $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$ to estimate $(I - At)^{-1} = I + At + (At)^2 + (At)^3 + \cdots$.

We've said that $e^{At} = Se^{\Lambda t}S^{-1}$. If A has n independent eigenvectors we can prove this from the definition of e^{At} by using the formula $A = S\Lambda S^{-1}$:

$$\begin{aligned} e^{At} &= I + At + \frac{(At)^2}{2} + \frac{(At)^3}{6} + \cdots \\ &= SS^{-1} + S\Lambda S^{-1}t + \frac{S\Lambda^2 S^{-1}}{2}t^2 + \frac{S\Lambda^3 S^{-1}}{6}t^3 + \cdots \\ &= Se^{\Lambda t}S^{-1}. \end{aligned}$$

It's impractical to add up infinitely many matrices. Fortunately, there is an easier way to compute $e^{\Lambda t}$. Remember that:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$

When we plug this in to our formula for e^{At} we find that:

$$e^{\Lambda t} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & e^{\lambda_n t} \end{bmatrix}.$$

This is another way to see the relationship between the stability of $\mathbf{u}(t) = Se^{\Lambda t}S^{-1}\mathbf{v}(0)$ and the eigenvalues of A.

Second order

We can change the second order equation $y'' + by' + ky = 0$ into a two by two first order system using a method similar to the one we used to find a formula for the Fibonacci numbers. If $u = \begin{bmatrix} y' \\ y \end{bmatrix}$, then

$$u' = \begin{bmatrix} y'' \\ y' \end{bmatrix} = \begin{bmatrix} -b & -k \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y' \\ y \end{bmatrix}.$$

We could use the methods we just learned to solve this system, and that would give us a solution to the second order scalar equation we started with.

If we start with a kth order equation we get a k by k matrix with coefficients of the equation in the first row and 1's on a diagonal below that; the rest of the entries are 0.

Linear Algebra Lecture 23

Differential Eqns $\frac{du}{dt} = Au$

Exponential e^{At} of a matrix

Example

$$\frac{du_1}{dt} = -u_1 + 2u_2$$

$$\frac{du_2}{dt} = u_1 - 2u_2$$

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

Sol $u(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ initial condition

$A = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix}$ A is singular, hence $\lambda = 0, \underbrace{-3}_{\text{sum of trace}}$

$$|A - \lambda I| = \begin{vmatrix} -1-\lambda & 2 \\ 1 & -2-\lambda \end{vmatrix}$$

$$\Rightarrow \lambda^2 + 3\lambda = 0$$

$$\lambda(\lambda+3) = 0$$

$$\boxed{\lambda = 0 \quad \lambda = -3}$$

For eigen vectors: i) $\lambda_1 = 0$

$x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ $\boxed{Ax_1 = 0x_1}$

ii) $\lambda_2 = -3$

$x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ $A-(-3)I = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$ $\boxed{Ax_2 = -3x_2}$

Solution: $u(t) = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2$ (general solution of differential equation)

Check: $\frac{du}{dt} = Au$ Plug in $e^{\lambda_1 t} x_1$

$$\frac{d}{dt} e^{\lambda_1 t} x_1 = A(e^{\lambda_1 t} x_1) \Rightarrow \lambda_1 \cancel{e^{\lambda_1 t}} x_1 = A \cancel{e^{\lambda_1 t}} x_1 \Rightarrow \boxed{\lambda_1 x_1 = Ax_1}$$

for C_1: Use solution $\lambda_1 = 0$ $\lambda_2 = -3$

$$u(t) = C_1 \cdot 1 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} + C_2 e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Use $u(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

At $t=0$

$$c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 e^0 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

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

$$2c_1 + c_2 = 1$$
$$c_1 - c_2 = 0$$

$c_1 = 1/3$, $c_2 = 1/3$

$$\text{Sol} = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{3} e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$\to 0$ when $t \to \infty$

Steady State $u(\infty) = 1/3 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

① When do we get stability i.e $u(t) \to 0$ Ans: when eigenvalues are negative

we Need $e^{\lambda t} \to 0$. i.e $\lambda < 0$

If λ is imaginary i.e $\lambda = -3+6i$. Does this imaginary part play role.

$$\left| e^{(-3+6i)t} \right| = e^{-3t} \quad b/c \quad \underbrace{\left| e^{6it} \right| = 1}_{\text{This complex number is running around unit circle.}}$$

As $e^{i\theta} = \cos\theta + i\sin\theta$

$|e^{i\theta}| = \sqrt{\cos^2\theta + \sin^2\theta}$

$|e^{i\theta}| = 1$

we care about real part of λ only

② Steady State:

when would we have steady state ?

when $\lambda_1 = 0$ and other $\text{Real}(\lambda) < 0$

③ we blow up if any $\text{Real}(\lambda) > 0$

If we change sign of A, then eigenvalues will change sign i.e $\lambda = 0, 3$ & system will blow off b/c $\lambda_2 = 3$ & $e^{3t} = \infty$ when $t \to \infty$

2x2 Stability

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad Re\,\lambda_1 < 0 \Big| Re\,\lambda_2 < 0$$

* trace $a + d = \lambda_1 + \lambda_2 < 0$ (for stability)

Negative trace is not enough b/c trace could be < 0 but system could still blow up for instance $\begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}$ $\underline{e^{+t}, e^{-2t}}$ $\longrightarrow e^{t} \to \infty$ when $t \to \infty$

trace $(\lambda_1 + \lambda_2) \longrightarrow -ve$

Det $(\lambda_1 \lambda_2) > 0 \longrightarrow +ve$

when $\lambda_1 + \lambda_2 < 0$

when $t = 0$ $\quad c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ we were finding c_1 & c_2

we can write as follows: $\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}}_{S:\ \text{matrix of eigenvectors}} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Coming back to $\frac{du}{dt} = Au$ $\qquad$ Set $u = Sv$

$S =$ eigen vector matrix

$$S\frac{dv}{dt} = ASv$$

$$\frac{dv}{dt} = S^{-1}ASv$$

$$\frac{dv}{dt} = \Lambda v \quad \Bigg| \quad \frac{dv_1}{dt} = \lambda_1 v_1 \quad \vdots \quad \frac{dv_n}{dt} = \lambda_n v_n$$

} i.e they are system of equation ~~not~~ doesn't depend on each other.

$$\Rightarrow v(t) = e^{\Lambda t} v(0) \qquad u(t) = Se^{\Lambda t}S^{-1}u(0) = e^{At}u(0)$$

$$e^{At} = S e^{\Lambda t} S^{-1}$$

Λ = diagonal matrix

whats the exponential of Matrix ?

e^{At} : Matrix Exponential

$$e^{At} = I + At + \frac{(At)^2}{2} + \frac{(At)^3}{6} + \cdots + \frac{(At)^n}{n!}$$ Taylor Series

$$e^x = \sum_0^\infty \frac{x^n}{n!}$$ one beautiful Taylor series

$$\frac{1}{1-x} = \sum_0^\infty x^n$$ (Geometric series) Another beautiful Taylor Series

$$(I - At)^{-1} = I + At + (At)^2 + (At)^3 + \cdots$$ (Reasonable if t is small)

$(I - At)^{-1}$ converges if $|\lambda(At)| < 1$

Now, we hope to compute e^{At} from $S e^{\Lambda t} S^{-1}$

$$e^{At} = I + At + \frac{(At)^2}{2} + \cdots$$

$$e^{At} = I + \underbrace{S\Lambda S^{-1}}_{A} t + \frac{S\Lambda^2 S^{-1}}{2!} t^2 + \frac{S\Lambda^3 S^{-1}}{3!} t^3$$ As $A = S\Lambda S$, $A^2 = S\Lambda^2 S$

$$= SS^{-1} + S\Lambda S^{-1} t + \frac{S\Lambda^2 S^{-1}}{2} t^2 + \cdots$$

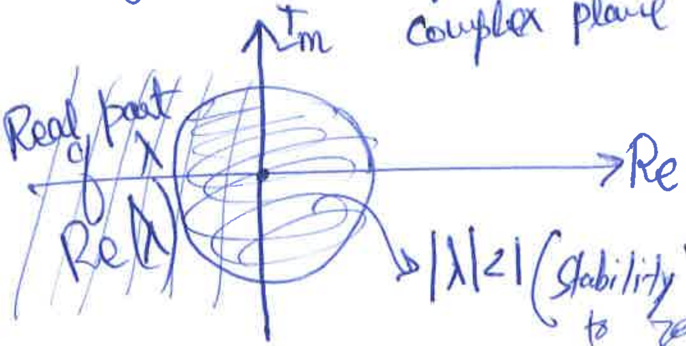
$$= S\left(I + \Lambda t + \frac{\Lambda^2 t^2}{2} + \cdots\right)S^{-1}$$

$$= S e^{\Lambda t} S^{-1}$$ (This formula does not work when S does not have independent eigenvectors & thus they can't be diagonalised)

What is $e^{\Lambda t}$? Λ = diagonal matrix $\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

test $e^{\Lambda t} = \begin{bmatrix} e^{\lambda_1 t} & & & 0 \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_3 t} \end{bmatrix}$

when does $e^{\Lambda t}$ goes to zero? It goes to zero when $Re(\lambda) < 0$



$|\lambda| < 1$ (stability when power of A goes to zero)

Final Example 2nd order equation

$y'' + by' + Ky = 0$ How to change one 2nd order equation to 2x2 1st order eq.

$u = \begin{bmatrix} y' \\ y \end{bmatrix} \quad u' = \begin{bmatrix} y'' \\ y' \end{bmatrix} = \begin{bmatrix} -b & -K \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y' \\ y \end{bmatrix}$

$\Rightarrow \left. \begin{array}{l} y'' = -by' - Ky \\ y' = y' \end{array} \right\}$

$\begin{bmatrix} 1 & & & \\ 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$ 5^{th} order to 5x5 1st order.

The parallel lecture with powers of matrix, we can now do exponential of matrix

Rotation?

Unit II: Lecture 11

Like differential equations, Markov matrices describe changes over time. Once again, the eigenvalues & eigenvectors describe the long term behaviour of the system. In this session, we also learnt about Fourier Series, which describe periodic functions as points in an infinite dimensional vector space.

Markov matrices; Fourier series

In this lecture we look at Markov matrices and Fourier series – two applications of eigenvalues and projections.

Eigenvalues of A^T

The eigenvalues of A and the eigenvalues of A^T are the same:

$$(A - \lambda I)^T = A^T - \lambda I,$$

so property 10 of determinants tells us that $\det(A - \lambda I) = \det(A^T - \lambda I)$. If λ is an eigenvalue of A then $\det(A^T - \lambda I) = 0$ and λ is also an eigenvalue of A^T.

Markov matrices

A matrix like:

$$A = \begin{bmatrix} .1 & .01 & .3 \\ .2 & .99 & .3 \\ .7 & 0 & .4 \end{bmatrix}$$

in which all entries are non-negative and each column adds to 1 is called a *Markov matrix*. These requirements come from Markov matrices' use in probability. Squaring or raising a Markov matrix to a power gives us another Markov matrix.

When dealing with systems of differential equations, eigenvectors with the eigenvalue 0 represented steady states. Here we're dealing with powers of matrices and get a steady state when $\lambda = 1$ is an eigenvalue.

The constraint that the columns add to 1 guarantees that 1 is an eigenvalue. All other eigenvalues will be less than 1. Remember that (if A has n independent eigenvectors) the solution to $\mathbf{u}_k = A^k \mathbf{u}_0$ is $\mathbf{u}_k = c_1\lambda_1^k \mathbf{x}_1 + c_2\lambda_2^k \mathbf{x}_2 + \cdots + c_n\lambda_n^k \mathbf{x}_n$. If $\lambda_1 = 1$ and all others eigenvalues are less than one the system approaches the steady state $c_1\mathbf{x}_1$. This is the $\mathbf{x}_1$ component of $\mathbf{u}_0$.

Why does the fact that the columns sum to 1 guarantee that 1 is an eigenvalue? If 1 is an eigenvalue of A, then:

$$A - 1I = \begin{bmatrix} -.9 & .01 & .3 \\ .2 & -.01 & .3 \\ .7 & 0 & -.6 \end{bmatrix}$$

should be singular. Since we've subtracted 1 from each diagonal entry, the sum of the entries in each column of $A - I$ is zero. But then the sum of the rows of $A - I$ must be the zero row, and so $A - I$ is singular. The eigenvector $\mathbf{x}_1$ is in the nullspace of $A - I$ and has eigenvalue 1. It's not very hard to find $\mathbf{x}_1 = \begin{bmatrix} .6 \\ 33 \\ .7 \end{bmatrix}$.

We're studying the equation $\mathbf{u}_{k+1} = A\mathbf{u}_k$ where A is a Markov matrix. For example u_1 might be the population of (number of people in) Massachusetts and u_2 might be the population of California. A might describe what fraction of the population moves from state to state, or the probability of a single person moving. We can't have negative numbers of people, so the entries of A will always be positive. We want to account for all the people in our model, so the columns of A add to $1 = 100\%$.

For example:

$$\begin{bmatrix} u_{\text{Cal}} \\ u_{\text{Mass}} \end{bmatrix}_{t=k+1} = \begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix} \begin{bmatrix} u_{\text{Cal}} \\ u_{\text{Mass}} \end{bmatrix}_{t=k}$$

assumes that there's a 90% chance that a person in California will stay in California and only a 10% chance that she or he will move, while there's a 20% percent chance that a Massachusetts resident will move to California. If our initial conditions are $\begin{bmatrix} u_{\text{Cal}} \\ u_{\text{Mass}} \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 1000 \end{bmatrix}$, then after one move $\mathbf{u}_1 = A\mathbf{u}_0$ is:

$$\begin{bmatrix} u_{\text{Cal}} \\ u_{\text{Mass}} \end{bmatrix}_1 = \begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix} \begin{bmatrix} 0 \\ 1000 \end{bmatrix} = \begin{bmatrix} 200 \\ 800 \end{bmatrix}.$$

For the next few values of k, the Massachusetts population will decrease and the California population will increase while the total population remains constant at 1000.

To understand the long term behavior of this system we'll need the eigenvectors and eigenvalues of $\begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix}$. We know that one eigenvalue is $\lambda_1 = 1$. Because the trace $.9 + .8 = 1.7$ is the sum of the eigenvalues, we see that $\lambda_2 = .7$.

Next we calculate the eigenvectors:

$$A - \lambda_1 I = \begin{bmatrix} -.1 & .2 \\ .1 & -.2 \end{bmatrix} \mathbf{x}_1 = \mathbf{0},$$

so we choose $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. The eigenvalue 1 corresponds to the steady state solution, and $\lambda_2 = .7 < 1$, so the system approaches a limit in which $2/3$ of 1000 people live in California and $1/3$ of 1000 people are in Massachusetts. This will be the limit from any starting vector $\mathbf{u}_0$.

To know how the population is distributed after a finite number of steps we look for an eigenvector corresponding to $\lambda_2 = .7$:

$$A - \lambda_2 I = \begin{bmatrix} .2 & .2 \\ .1 & .1 \end{bmatrix} \mathbf{x}_1 = \mathbf{0},$$

so let $\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

From what we learned about difference equations we know that:

$$\mathbf{u}_k = c_1 1^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2(.7)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

When $k = 0$ we have:

$$\mathbf{u}_0 = \begin{bmatrix} 0 \\ 1000 \end{bmatrix} = c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

so $c_1 = \frac{1000}{3}$ and $c_2 = \frac{2000}{3}$.

In some applications Markov matrices are defined differently – their rows add to 1 rather than their columns. In this case, the calculations are the transpose of everything we've done here.

Fourier series and projections

Expansion with an orthonormal basis

If we have an orthonormal basis $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n$ then we can write any vector $\mathbf{v}$ as $\mathbf{v} = x_1\mathbf{q}_1 + x_2\mathbf{q}_2 + \cdots + x_n\mathbf{q}_n$, where:

$$\mathbf{q}_i^T\mathbf{v} = x_1\mathbf{q}_i^T\mathbf{q}_1 + x_2\mathbf{q}_i^T\mathbf{q}_2 + \cdots + x_n\mathbf{q}_i^T\mathbf{q}_n = x_i.$$

Since $\mathbf{q}_i^T\mathbf{q}_j = 0$ unless $i = j$, this equation gives $x_i = \mathbf{q}_i^T\mathbf{v}$.

In terms of matrices, $\begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{v}$, or $Q\mathbf{x} = \mathbf{v}$. So $\mathbf{x} = Q^{-1}\mathbf{v}$. Because the q_i form an orthonormal basis, $Q^{-1} = Q^T$ and $\mathbf{x} = Q^T\mathbf{v}$. This is another way to see that $x_i = \mathbf{q}_i^T\mathbf{v}$.

Fourier series

The key idea above was that the basis of vectors $\mathbf{q}_i$ was orthonormal. Fourier series are built on this idea. We can describe a function $f(x)$ in terms of trigonometric functions:

$$f(x) = a_0 + a_1\cos x + b_1\sin x + a_2\cos 2x + b_2\sin 2x + \cdots.$$

This *Fourier series* is an infinite sum and the previous example was finite, but the two are related by the fact that the cosines and sines in the Fourier series are orthogonal.

We're now working in an infinite dimensional vector space. The vectors in this space are functions and the (orthogonal) basis vectors are 1, $\cos x$, $\sin x$, $\cos 2x$, $\sin 2x$, ...

What does "orthogonal" mean in this context? How do we compute a dot product or *inner product* in this vector space? For vectors in $\mathbb{R}^n$ the inner product is $\mathbf{v}^T\mathbf{w} = v_1w_1 + v_2w_2 + \cdots + v_nw_n$. Functions are described by a continuum of values $f(x)$ rather than by a discrete collection of components v_i. The best parallel to the vector dot product is:

$$f^T g = \int_0^{2\pi} f(x)g(x)\,dx.$$

We integrate from 0 to 2π because Fourier series are periodic:

$$f(x) = f(x+2\pi).$$

The inner product of two basis vectors is zero, as desired. For example,

$$\int_0^{2\pi} \sin x \cos x\,dx = \frac{1}{2}(\sin x)^2\Big|_0^{2\pi} = 0.$$

How do we find a_0, a_1, etc. to find the coordinates or *Fourier coefficients* of a function in this space? The constant term a_0 is the average value of the function. Because we're working with an orthonormal basis, we can use the inner product to find the coefficients a_i.

$$\begin{aligned}
\int_0^{2\pi} f(x)\cos x\,dx &= \int_0^{2\pi} (a_0 + a_1\cos x + b_1 \sin x + a_2\cos 2x + \cdots)\cos x\,dx \\
&= 0 + \int_0^{2\pi} a_1 \cos^2 x\,dx + 0 + 0 + \cdots \\
&= a_1\pi.
\end{aligned}$$

We conclude that $a_1 = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x\,dx$. We can use the same technique to find any of the values a_i.

LA: Lecture 24

Markov Matrices
Steady State : $\lambda = 1$ (in power case)
Fourier Series & Projections

Applications of EigenValues

A = markov matrix $= \begin{bmatrix} .1 & .01 & .3 \\ .2 & .99 & .3 \\ .7 & 0 & .4 \end{bmatrix}$

Powers of A will also be markov Matrix

Two Properties of A:

1. All entries ≥ 0
2. All columns add to 1. (It guarantees that one eigenvalue of $A = 1$)

Keypoints for A
1. $\lambda = 1$ is an eigenvalue
2. All other eigenvalues are in magnitude < 1
   ie $|\lambda_i| < 1$

We remember $U_K = A^K U_0 = \underbrace{c_1 \lambda_1^K x_1 + c_2 \lambda_2^K x_2 + \cdots + c_N \lambda_N^K x_N}_{\text{This requires complete set of } \lambda_i}$

when $\lambda_1 = 1$, & $\lambda_i < 0$ then as we go forward in time $c_2 \lambda_2^K x_2 + \cdots + c_N \lambda_N^K x_N \approx 0$, And

Steady state = $c x_1$ As $\lambda_1 = 1$.

① $\lambda = 1$ is an e.val / eigenvector $X > 0$ } at steady ~~state~~

$$A - 1I = \begin{bmatrix} -0.9 & .01 & .3 \\ 0.2 & -.01 & .3 \\ 0.7 & 0 & -.6 \end{bmatrix}$$

It should be singular As $\lambda = 1$ is an eigenvalue

All columns of $A-I$ add to zero. $\Rightarrow$ $A - I$ is singular

This means that columns are dependent (as columns add to zero) $R_1 + R_2 + R_3 = [0\ 0\ 0]$ (Rows are dependent)

Its singular $(A - 1I)$ because $(1,1,1)$ is in $N(A^T)$. So who is in $N(A)$? Its eigenvector X_1 which is in $N(A)$. And X_1 is steady state.

What can you tell me about?

i) eigenvalues of A

ii) eigenvalues of A^T

Ans: They are same! why? Because ~~For F~~

for Eigen Value of A

$$\det(A - \lambda I) = 0$$

$$\det(A^T - (\lambda I)^T) = 0$$

$$\det(A^T - \lambda I) = 0 = \det(A^T - \lambda I)$$

thus $(A - 1I)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.9 & .01 & .3 \\ 0.2 & -.01 & .3 \\ 0.7 & 0 & -.6 \end{bmatrix} \underbrace{\begin{bmatrix} 0.6 \\ 0.33 \\ 0.7 \end{bmatrix}}_{\text{Eigen Vector}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Eigen Vector, all entries are +ve

<u>To Solve</u>

$$U_{K+1} = A U_K \qquad A \text{ is Markov}$$

$$\begin{bmatrix} U_{cal} \\ U_{mass} \end{bmatrix}_{t=K+1} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} U_{cal} \\ U_{mass} \end{bmatrix}_{t=K}$$

0.9 people in California stay in Cal at time K

$$\begin{bmatrix} U_{cal} \\ U_{mass} \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 1000 \end{bmatrix}$$

$$\begin{bmatrix} U_{cal} \\ U_{mass} \end{bmatrix}_1 = \begin{bmatrix} 200 \\ 800 \end{bmatrix} \begin{matrix} \to \text{in Cal} \\ \to \text{in Mass} \end{matrix}$$

$$\begin{bmatrix} U_{cal} \\ U_{mass} \end{bmatrix}_2 = \begin{bmatrix} >200 \\ <800 \end{bmatrix}$$

what will happen after 100 multiplications? Now we will find eigenvalues & eigenvectors.

$$A = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \qquad \lambda_1 = 1 \; \left(\begin{matrix}\text{Markov} \\ \text{Matrix}\end{matrix}\right)$$

$$\lambda_2 = \text{trace} - \lambda_1 = 0.7 = 0.9 + 0.8 - 1$$

$$(\lambda_1, \lambda_2) = (1, 0.7)$$

$\lambda_1 = 1$ Eigenvector X_1 (Steady State sol)

Its in $N(A - I)$:

$$\begin{bmatrix} -0.1 & 0.2 \\ 0.1 & -0.2 \end{bmatrix} \underbrace{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}_{X_1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\lambda_2 = 0.7$ Eigenvector X_2

$$\begin{bmatrix} .2 & .2 \\ .1 & .1 \end{bmatrix} \underbrace{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}_{X_2} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$U_K = C_1 \, 1^K \begin{bmatrix} 2 \\ 1 \end{bmatrix} + C_2 (.7)^K \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$U_0 = \begin{bmatrix} 0 \\ 1000 \end{bmatrix} = C_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + C_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\left.\begin{matrix} 2C_1 - C_2 = 0 \\ C_1 + C_2 = 1000 \end{matrix}\right\} \quad C_1 = \frac{1000}{3} \qquad C_2 = \frac{2000}{3}$$

$\Rightarrow$ Sol : $\frac{1000}{3}(1)^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{2000}{3}(0.7)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

Projection with Orthonormal basis

Any v $\rightarrow$ $q_1, \dots, q_n$ in $\mathbb{R}^n$.

$$V = x_1 q_1 + x_2 q_2 + \cdots + x_n q_n \quad \text{——} (i)$$

The word projection can be replaced by expansion because I am expanding a vector V into its basis.

whats the formula for x_1 ?

$$q_1^T V = x_1 q_1^T q_1 + 0 \text{——} + 0$$

$$\boxed{q_1^T V = x_1}$$

In matrix language, eq (i) is

$$\underbrace{\begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{x} = V$$

$$Qx = V$$

$$x = Q^{-1} V$$

As Q has orthonormal columns

Q^{-1} Q^T

$$x = Q^T V$$

$$\Rightarrow x_1 = q_1^T V$$
$$x_2 = q_2^T V$$
$$\vdots$$
$$x_n = q_n^T V$$

Key ingredient: q's are orthonormal.

Fourier Series:

We want to expand $f(x)$ as:

$$f(x) = a_0 + a_1\cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots$$

(This equation has infinite terms but orthogonal components are possible with sin, cos)

Joseph Fourier realised that he can work in functional space instead of vector space. Instead of orthogonal basis vectors for $\mathbb{R}^N$, I can have constant, cos x, sin x, cos 2x, sin 2x, ... infinitely many terms. Now we are moving from finite dimensional space to infinite dimensional space. Now vectors will be functions and basis vector is also function. we have infinite basis vectors.

i.e basis = 1, cos x, sin x, cos 2x, sin 2x, ...

The reason Fourier is succesful is that these functions are orthogonal.

We know the notion of dot product in vector space, what is the meaning of dot product in functional space $f(x)$? How do I compute ~~f(x)~~ dot product for functions?

For Vectors

$$V^T W = v_1 w_1 + \cdots + v_N w_N$$

functions f, g (they are continuous)

$$f^T g = \int f(x) \cdot g(x) dx$$

(Inner product of functions)

When we were having vectors, we were adding. But now we are in continuous spectrum, thus we should integrate. What about limits, sin, cos are periodic thus limits

$$f^T g = \int_0^{2\pi} f(x) g(x)\, dx$$

$f(x) = f(x+2\pi)$ periodic function

Checking orthogonality:

$$\int_0^{2\pi} \sin x \cos x \, dx$$

$$= \frac{1}{2} (\sin x)^2 \Big|_0^{2\pi}$$

$$= 0$$

Now, what is a_1? How much cosine is in function $f(x)$?

$a_1 = ??$ Take the inner product with $\cos x$

$$\int_0^{2\pi} f(x) \cos x \, dx = a_1 \int_0^{2\pi} (\cos x)^2 \, dx$$

$$= a_1 \pi$$

$$\Rightarrow \boxed{a_1 = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos x \, dx}$$

Coefficients of Fourier series is an expansion in an orthonormal basis.