Linear Algebra

vector - list of numbers $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

matrix - "spreadsheet of numbers" $\begin{bmatrix} x_{11} & x & \cdots \\ x & x & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$

Linear algebra allows us to do parallel computation & simple expression of complex relationships

↳ indeed, conceptualize problems in high dim.

Section 1: vectors

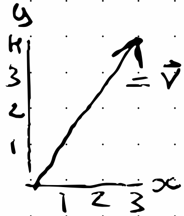
$\mathbb{R}^2$

↳ ordered list of numbers $\begin{bmatrix} 3 \\ 4 \end{bmatrix} \neq \begin{bmatrix} 4 \\ 3 \end{bmatrix}$

length = dimensionality

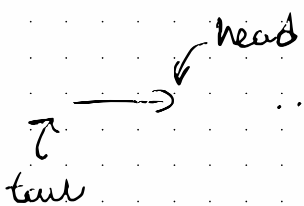
$\begin{bmatrix} a \\ b \\ c \end{bmatrix}^T = [a \ b \ c] = \mathbf{v} = \vec{v} = \vec{v} = v$ ← italics

"column vec" "row vec"



$= \vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ ↳ line w/ length & direction

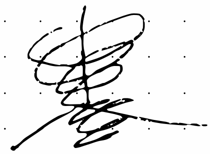
NB: vectors needn't start from origin



... not a coordinate!

Note: vectors don't need to only contain numbers.

$$\begin{bmatrix} \sin(x) \\ \cos(x) \\ x \end{bmatrix} \longrightarrow$$



→ **vector addition & subtraction**

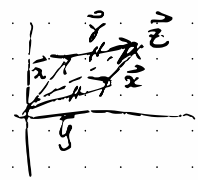
↳ dimensions must match.

↳ add or subtract elementwise

$\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $\vec{y} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ $\vec{x} + \vec{y} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} = \vec{z}$

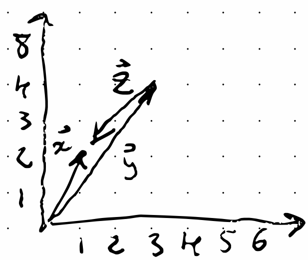
note that order doesn't matter.

↳



subtract "head-head"

$$\vec{x} - \vec{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \vec{z}$$



→ **vector scalar multiplication**

↳ scalar = number (1×1 vector/matrix)

↳ tends to be denoted by greek letters.

$$\lambda\vec{v} = 7\begin{bmatrix}-1\\0\\1\end{bmatrix} = \begin{bmatrix}-7\\0\\7\end{bmatrix}$$

stretching/shrinking by this value.
($\lambda > 1$) ($0 > \lambda > 1$)

↳ reverse when $\lambda < 0$

NB: Will not change the angle of the vector.

Dot product

One way to multiply vectors ⇒ used everywhere.

notation: $a \cdot b$, $\langle a, b \rangle$, $a^T b$, $\sum_{i=1}^{N} a_i b_i$

shows exactly how it's done.

$$\begin{bmatrix}1\\2\\3\end{bmatrix} \cdot \begin{bmatrix}1\\2\\3\end{bmatrix} = 1 + 4 + 9$$
$$= 14$$

must have same dimensions

Vector length

$$\|v\| = \sqrt{v^T v}$$

first time I've ever seen it defined like this

each value squared, makes sense.

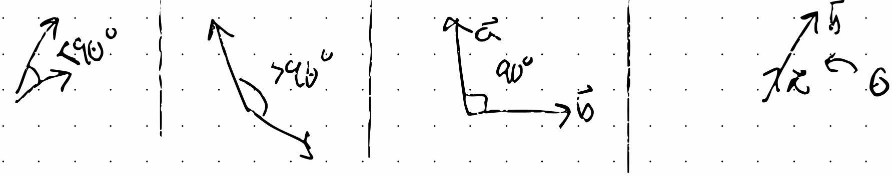
Dot product geometry

$a \cdot b$

geometric interpretation: $\alpha = |a||b|\cos(\theta_{ab})$

$\therefore \theta_{ab} = \cos^{-1}\left(\frac{a \cdot b}{|a|\,|b|}\right)$

$\alpha > 0$	$\alpha < 0$	$\alpha = 0$	$\alpha = \lvert\vec{a}\rvert\lvert\vec{b}\rvert$



$\sum_{i=1}^{N} a_i b_i = \cos(\theta_{ab})\|a\|\|b\|$

Law of cosines: $|\vec{a} - \vec{b}|^2 = |\vec{a}|^2 + |\vec{b}|^2 - 2|\vec{a}||\vec{b}|\cos(\theta_{\vec{a}\vec{b}})$

↳ Pythagorean theorem when $\cos(\theta) = 0$

↳ generalization when $\theta \neq 90°$

$|\vec{a} - \vec{b}| = (\vec{a} - \vec{b})^T(\vec{a} - \vec{b})$ $\boxed{a^Tb = b^Ta}$

$= \vec{a}^T\vec{a} + b^Tb - a^Tb - b^Ta$

$= a^Ta + b^Tb - 2a^Tb$

$\therefore \boxed{a^Tb = |a||b|\cos(\theta_{ab})}$

Is the dot product communative?

$$a^T b = \sum_{i=1}^{N} a_i b_i \qquad b^T A = \sum_{i=1}^{N} b_i a_i$$

$$\therefore a^T b = b^T a$$

↳ Since multiplication & addition are communicative, so is the dot product.

Outer product

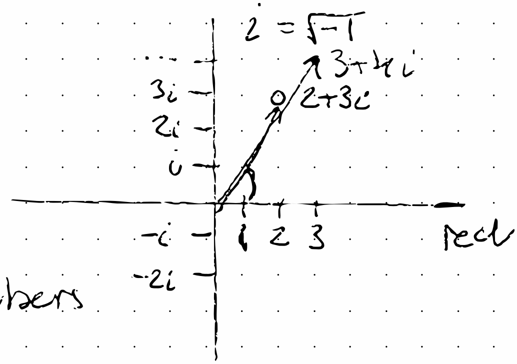
$\vec{x}, \vec{b}$ $N\times1$ $N\times1$ $\vec{x}^T b \rightarrow$ scalar $1\times N$ $N\times1$ 1×1

outer product

$\vec{x}\,\vec{b}^T \rightarrow M$(atrix) $N\times1$ $1\times N$ $N\times N$

Complex vectors

↳ each point is its own 2d vector



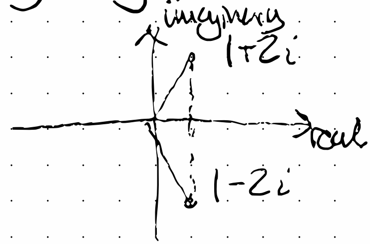
↳ multiplying complex numbers

$$(2+3i)(3+4i) = 2\times3 + 3\times3i + 2\times4i - 1(3\times4)$$
$$= 6 + 9i + 8i - 12$$
$$= -6 + 17i$$

↑ using this, we can include $\mathbb{C}$ in vector notation

Hermetian transpose

complex conj: flip sign of imaginary part



↑ complex conj.

↳ thus, hermetian transpose transposes the vec & changes imaginary components sign.

e.g.

$$\begin{bmatrix} 1\,3i \\ -2i \\ 4 \\ -5 \end{bmatrix}^H = \begin{bmatrix} 1-3i & 2i & 4 & -5 \end{bmatrix}$$

· Why is this important?

↳ what is norm$(3+4i)$? ⑤ how?

w/ $z^T z$ where $z = [3\ 4i]$? No.

$$z^T z = -7 + 24i$$

$$z^H z = [3\ -4i]\begin{bmatrix} 3 \\ 4i \end{bmatrix} = 9 + \cancel{12i - 12i} - 16(-1)$$
$$= 25$$

$\therefore \sqrt{z^H z} = 5$ ✓

Creating unit vectors

↳ vector such that its magnitude is 1

$\mu\vec{v}$ s.t. $\|\mu\vec{v}\| = 1$

$\mu = \frac{1}{\|\vec{v}\|}$

useful because it can be used to create basis vectors.

↳ cannot be done w/ $\vec{0}$ because $\frac{1}{\|\vec{0}\|}$ is undefined.

$\|\mu 0\| = 1$

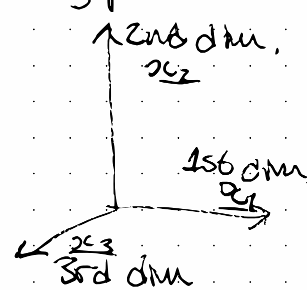
you also can't make 1 out of 0

Dimensions in Linear Algebra

↳ number of elements in a vector

↳ dimensions are not made redundant by prev. dim.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \Rightarrow$$



→ Field - a set on which $\{+, -, \times, \div\}$ can be used

↳ $\mathbb{R}$ - real numbers $\mathbb{C}$ - complex numbers

$\mathbb{Z}$ - integers

Why is this not a field? belongs to $\mathbb{R}$?

↳ when discussing $\mathbb{R}^n$ we are talking of a field w/ n dimensions of real numbers.

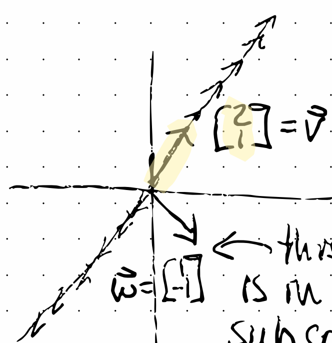
ie $\boxed{\vec{v} \in \mathbb{R}^n}$

vector is in $\mathbb{R}^n$

<u>subspaces</u>

↳ all vectors that can be created using $\lambda \vec{v}$, where $\lambda \in \mathbb{R}$.

ie line created by vector



← this vector is in another subspace.

$\frac{1}{2}\vec{v} = \begin{bmatrix} 1 \\ .5 \end{bmatrix} \neq \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ ✓ x

there is no way to mult. $\vec{v}$ to get $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

NB: $\lambda \vec{v} + \beta \vec{w}$ ← spans all of $\mathbb{R}^2$

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

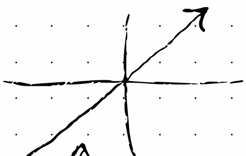
Formal definition:

$$\underbrace{\forall \vec{v},\vec{w}\in V}_{\text{for any 2 vecs in same subsp.}},\ \underbrace{\forall \lambda,\alpha\in\mathbb{R}}_{\text{for any 2 scalars in }\mathbb{R}};\ \underbrace{\lambda\vec{v}+\alpha\vec{w}\in V}_{\text{linear combos of }\vec{v},\vec{w},\lambda,\alpha\text{ will also be in subsp. }V}$$

Subsets

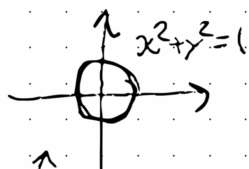
↳ set of points that satisfy some conditions

↳ arbitrary

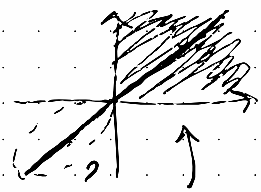


all points on this line?

↳ subset & subsp.

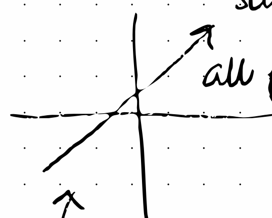


also not subsp. but subset



subspace? not closed

subset ✓



all points on this line?

× subsp. (doesn't cont. $\vec{0}$)

✓ subset

Span

↳ span of set of vectors is all vectors that can be made by linear combinations of these vectors.

Linear Independance

↳ none of the vectors in a set can be expressed as a linear combo of the other.

formally: $\underbrace{\sum_{i=1}^{N} \lambda_i \vec{v}_i = 0, \lambda \in \mathbb{R}, \lambda \neq 0}$

↳ if this condition is met then linearly dependant.

Basis

↳ Ways to express coordinates within a space

$S = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ ← unit vectors, span $\mathbb{R}^2$, orthogonal

but any set of vectors can be used. as long as they span the space.

Note: basis must be lin. indep. or else there will be ∞ ways to describe one point.

Some bases can also be chosen to compress data more efficiently.

↳ super important for stats, ml, dim red. & compression.

Section 2: Matrices

Note: expressed as bold letter.

e.g. $\mathbf{A} = \begin{bmatrix} 1 & 6 & 0 \\ 7 & 2 & 4 \\ 4 & 1 & 1 \end{bmatrix}$

indexing

a_{ij} = ith row, jth column of ~~the~~ matrix A. $a_{1,2} = 6$

NB: everything in lin algebra is 'row' then 'column'

NB: block matrices as shorthand. $\begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$

$\mathbf{D} = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$

$\mathbf{1} = [1]$

$\mathbf{0} = [0]$

$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 1 & 1 & 3 & 0 \\ 1 & 1 & 0 & 4 \end{bmatrix}$

diagonal, elements from $a_{1,1}$ to $a_{n,n}$

↳ $\neg a_{n,n}$ = off-diagonal.

→ Shapes of matrices

once again, rows then columns

$$\boxed{\underset{m \times n}{A}}$$

→ m rows, n columns

→ <u>dimensionality</u>

depends on definition $\mathbb{R}^{MN}$, $\mathbb{R}^{M \times N}$, etc.

NB $C(M) \in \mathbb{R}^M$, $R(M) \in \mathbb{R}^N$

Where $C(M)$ is column space - span of column vectors

$R(M)$ is row space - span of row vectors.

↳ Matrices are flexible concepts but can be cumbersome to think about.

→ <u>Types of matrices</u>

$M^T = M$: symmetric, eg $\begin{bmatrix} 1 & 5 \\ 5 & -1 \end{bmatrix}^T = \begin{bmatrix} 1 & 5 \\ 5 & -1 \end{bmatrix}$

$I = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \ddots \end{bmatrix}$ ← only 1 on diagonal "identity"

$\mathbf{0} = [0]$ ← empty matrix "zero matrix"

"Diagonal" $= \begin{bmatrix} \alpha & & 0 \\ & \beta & \\ 0 & & c \ddots \end{bmatrix}$ only ≠0 values on diagonal

L / R = "triangular matrix" $=$ R $\begin{bmatrix} \diagdown & \\ 0 & \end{bmatrix}$ or L $\begin{bmatrix} & 0 \\ \diagdown & \end{bmatrix}$

Augmented matrix → 2 matrices concatenated together width-ways

$$A|B = \left[\begin{array}{cc|ccc} a_{11} & a_{12} & b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & b_{21} & b_{22} & b_{23} \end{array}\right]$$

→ Matrix addition & subtraction

→ only valid when matrices have same shape
↳ add element-wise

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 4 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 7 & 2 \end{bmatrix}$$

$A+B = B+A$

↳ commutative, order doesn't matter
↳ subtraction works in the same fashion
↳ also associative, $(A+B)+C = A+(B+C)$

→ Matrix scalar multiplication

$$\delta \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \delta a & \delta b \\ \delta c & \delta d \end{bmatrix}$$

also commutative & associative

→ Transpose

"flipping a matrix". I made note of this earlier

→ Complex Matrices

↳ matrix that contains complex numbers.
instead of transposing these we take the Hermitian transpose as w/ vectors.

$$\begin{bmatrix} 1 & -1+5i & 0 \\ -1 & -2 & -4 \\ 6i & -4 & 5-2i \end{bmatrix}^H = \begin{bmatrix} 1 & -1-5i & 0 \\ -1 & -2 & -4 \\ -6i & -4 & 5+2i \end{bmatrix}^T$$

$$= \begin{bmatrix} 1 & -1 & -6i \\ -1-5i & -2 & -4 \\ 0 & -4 & 5+2i \end{bmatrix}$$

Diagonal & Trace

$$\text{diag}\left(\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$ ← select only diagonal & make vector

$$\text{Trace} = \sum_{i=1}^{N} \text{diag}(M)_i \,, \quad \text{trace}\left(\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right) = 3$$

↳ only defined for sq. matrices.

Section 3: matrix multiplication

- Matrix multiplication is not commutative

$$AB \neq BA$$

why?

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = a_{11} b_{11} \cdot a_{21} b_{12} + \dots$$

$$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = b_{11} a_{11} + b_{21} a_{12}$$

! Should be row/col. instead of col/row but the logic is still sound

these operations can result in different outcomes.

→ When can be multiplied

When column vectors of A are the same dimension as row vectors of B.

$$M_A, \boxed{N_A \times M_B}, N_B$$

returns shape M_A, N_B

ie. $5, \underbrace{2 \times 2}, 5 = 5 \times 5$

$2, \underbrace{5 \times 5} \times 3 = 2 \times 3$

$2, 3 \times 5, 2 = \cancel{\emptyset}$ doesn't work.

→ Diagonal matrix multiplication

↳ D. matrix has properties that make A easy to multiply

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\begin{bmatrix} a & & 0 \\ & b & \\ 0 & & c \end{bmatrix} = \begin{bmatrix} 1a & 2b & 3c \\ 4a & 5b & 6c \\ 7a & 8b & 9c \end{bmatrix}$$

↳
$$\begin{bmatrix} a & & 0 \\ & b & \\ 0 & & c \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1a & 2a & 3a \\ 4b & 5b & 6b \\ 7c & 8c & 9c \end{bmatrix}$$

→ Matrix order of operations

$(AB)^T = B^T A^T$ ← How to remember?

$(LIVE)^T = (E^T V^T I^T L^T)$

→ Matrix-vector multiplication

↳ same as mult w/ matrices except one matrix has shape including 1

NB: result is always a vector!

↳ dimension could be different depending on order

- if matrix is symmetric then order doesn't matter

$S\vec{w} = \vec{v}$

$\vec{w}^T S^T = \vec{v}^T \rightarrow \therefore \vec{w}^T S = \vec{v}^T \quad (S^T = S)$

→ ## 2D transformation matrices

Say we have a vector & we multiply some matrix to it.

$$\begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

$\uparrow$ function (stretch + rotate) $\quad \uparrow$ input $\quad \uparrow$ output

↳ rotation matrix

$$\lambda \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

← expresses any rotation & stretch.

NB: rotation on its own will not stretch because $\sqrt{\cos^2(\theta) + \sin^2(\theta)} = 1$

→ ## Additive & multiplicative matrix identities

remember, I is $n \times n$ matrix w/ a_{ij} where $i=j \Rightarrow 1$ else 0

$\boxed{AI = IA = A}$ however $A + I \neq A$

- Multiplicative Identity.

$AO = OA \neq A$ however $\boxed{A + O = A}$

- Additive Identity

→ Additive & multiplicative symmetric matrices

↳ symmetric matrices have special properties
but how do we make them?

→ Additive symmetric $((A + A^T) \times \frac{1}{2}) = S$

$$\frac{1}{2}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} a & c \\ b & d \end{bmatrix}\right) = \frac{1}{2} \times \begin{bmatrix} 2a & b+c \\ b+c & 2d \end{bmatrix}$$

normalization

· only works w/ square matrices else addition fails.

→ Multiplicative symmetric $(A^T A) = S$

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{bmatrix} = S$$

also works w/ rectangular matrix.

↳ used for cov. matrix.

Note that $A^T A$ is symmetric.

$$(A^T A)^T = A^T A^{TT}$$
$$= A^T A \text{ (symmetric)}$$

→ Hadamard Multiplication

↳ elementwise multiplication. ↙ only works if shape is the same.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \odot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2 & b^2 \\ c^2 & d^2 \end{bmatrix}$$

→ Linear operation also.

→ Multiplication of symmetric matrices

↳ does multiplication of multiple S also create S?

$$\begin{bmatrix} a & c \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ f & g \end{bmatrix} = \begin{bmatrix} ae+cf & af+cg \\ ce+af & cf+ag \end{bmatrix}$$

↳ seems not!

↳ However $a=d$ & $e=g$ then result is symmetric.

→ Frobenius Dot prod.

1. compute Hadamard mult.
2. sum over all i, j. ie $\sum_{i=1}^{N} \sum_{j=1}^{M} a_{ij} b_{ij}$

↑ similar to dot product.

also works by $\text{trace}(A^T B)$

↳ can also use to calculate norm

$\text{norm}(A) = \sqrt{\langle A, A \rangle}$

→ Matrix division

- Hadamard division $A \oslash B = \begin{bmatrix} a_{11}/b_{11} & a_{12}/b_{12} & \cdots \\ \vdots & \ddots & \end{bmatrix}$

constraint, B cannot contain 0

Section 4: Matrix Rank

↳ Amount of information found in a matrix.

indicated by r, scalar that characterizes a matrix.

↳ $\in \mathbb{R}$, $r > 0$

note that if you have a matrix in $\mathbb{R}^3$ then you can get lower rank if a vector inside is lin. dep.

↳ Also cannot be higher than min(M,N)
if M < N then rank cannot be N

→ computing rank

↳ count linearly independant columns

↳ "row echelon form" & count pivots

↳ SVD & count >0 singular values

↳ compute eigendecomp. & count >0 eigenvalues difficulties of calculating matrix rank

↳ computer rounding error for eigenvalues

↳ measuring error for emperical data.

→ **Rank of added & multiplied matrices**

→ Addition

$$\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$$

↳ Linear pooling of information

Applies to square & rectangular matrices.

→ Multiplication

$$\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$$

→ **Rank of A^TA & AA^T**

$$\text{rank}(A) = \text{rank}(A^TA) = \text{rank}(AA^T) = \text{rank}(A^T)$$

↳ Transposing a matrix doesn't change its rank

multiplying matrices of the same rank will produce a matrix of that rank

→ note that if you have A of full rank (rectangular) A^TA & AA^T will also be full rank

→ <u>Making a matrix full rank by shifting</u>

↳ real data is often rank deficient.

$$\tilde{A} = A + \lambda I$$

How to select this param.? can be difficult.

add as little information as possible.

Section 5: matrix spaces

$C(A)$

- column space: vector space spanned by columns on a matrix
- row space: vector space spanned by rows in a matrix

↳ $R(A) = C(A^T)$

- Nullspace: $N(A)$ set of vectors such that $Av = 0$ & $v \neq 0$

eg $\begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = 0$ ← what values of α, β give 0? (non trivial)

Note that nullspace is related to rank & linear dep.

↳ there is {} nullspace for full rank matrix

ie no points where applying the transformation causes it to land on 0

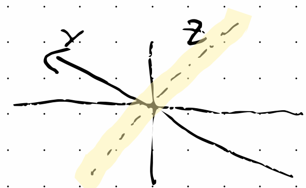
↳ "left nullspace" $N(A^T)$ is the nullsp. for the row sp.

→ Orthogonal matrix spaces

"col sp" "left Nsp" "row sp" "Nsp"

$$C(A) \perp N(A^T) \quad , \quad C(A^T) \perp N(A)$$

Oh shit this is true!



$\alpha y \cdot \beta z = 0$

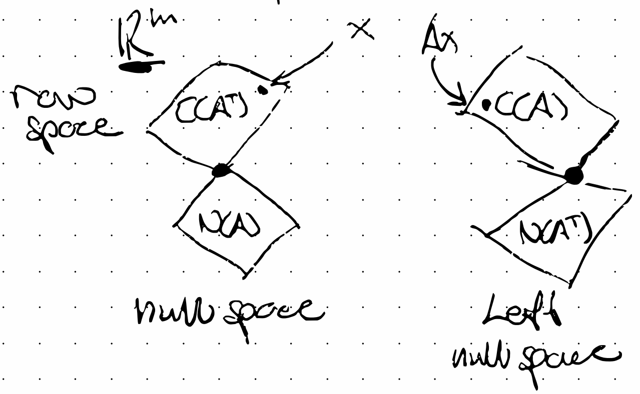
no matter of magnitude, always parallel

$Ax = y$

$\therefore N(A) = z$

Dimensions of column/row/null spaces

A
$n \times m$



$$\boxed{\begin{array}{l} C(A^T) \perp N(A) \\ C(A) \perp N(A^T) \end{array}}$$

- $d(C(A^T)) + d(N(A)) = n$
- $d(C(A)) + d(N(A^T)) = m$

spans whole ambient dimension

Example

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 3 & -3 \end{bmatrix}$$

$c_3 = -2c_1 + c_2$

$\therefore$ c_3 is lin. dep

$N(A) = \boxed{Ax = 0}$

$$x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ -3 \end{bmatrix} = 0$$

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

$\rightarrow C(A) = \left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$
$\in \mathbb{R}^2$, dim 2

$\rightarrow C(A^T) = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}^T, \begin{bmatrix} 3 \\ 3 \\ -3 \end{bmatrix}^T \right\}$
$\in \mathbb{R}^3$, dim 2

$\rightarrow N(A^T) = \{ \}$
$\in \mathbb{R}^2$, dim 0

$\rightarrow N(A) = \left\{ \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \right\}$
$\in \mathbb{R}^3$, dim 1

More on $Ax = b$ & $Ax = 0$

↳ fundamental to applications of LA.

$\rightarrow \boxed{Ax = b}$? solve for x.

known, known

→ is there a solution?

when $b \in C(A)$

If $b \notin C(A)$. How do we best estimate it?

$A\hat{x} = \hat{b}$ such that $\|\hat{b} - b\|$ is min. (OLS)

$\rightarrow \boxed{Ax = 0}$ people care about $(A - \lambda I)x = 0$

↳ PCA, QED, SVD, FDA, etc.

eigen val. eigenvec.

Section 6: solving systems of eqtns.

→ Gaussian elimmination

1. convert sys of eqtns to matrix form

e.g. $2x + 3y = 5$ $\Rightarrow$ $\underset{A}{\begin{bmatrix} 2 & 3 \\ -1 & 7 \end{bmatrix}} \underset{x}{\begin{bmatrix} x \\ y \end{bmatrix}} = \underset{b}{\begin{bmatrix} 5 \\ 7 \end{bmatrix}}$

$-x + 7y = 7$

2. convert A to U or L, remember to aug A w/ b

$$\left[\begin{array}{cc|c} 2 & 3 & 5 \\ -1 & 7 & 7 \end{array}\right] \nearrow \begin{bmatrix} 1 & 3/2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5/2 \\ 19/17 \end{bmatrix}$$

$$\searrow \begin{bmatrix} 1 & 0 \\ 0.148 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0.824 \\ 1 \end{bmatrix}$$

Note: $A = LU$ done w/ scipy.linalg.lu()

3. Solve equations

if A is a singular matrix, no solution exists

either $1 = 0$ or row of zeroes

↳ rows are linearly dep.

→ Echelon form

leading nonzero terms are pivots →

$$\begin{bmatrix} \boxed{a} & b & c & d & e \\ 0 & \boxed{f} & g & h & i \\ 0 & 0 & 0 & \boxed{j} & k \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

cannot swap these rows

n pivots = 3 = rank

All 0 below & to left of values

How to get there, also gaussian elimination

↳ not really used practically lol

→ Row reduced Echelon form

from echelon form

↳ set all pivots to 1

↳ 0 above pivots as well as below

Note you lose a lot of information but each operation done can be expressed as a matrix mult back to original matrix.

↳ Also not high application value.

→ Matrix spaces after row reduction

→ $\text{rank}(A) = \text{rank}(\text{rref}(A))$

→ $C(A^T) = C(\text{rref}(A^T))$ ← vectors in $C(A^T)$ will always be linearly independent from linear transformations

→ column space can change after row red.

→ Notes from G. Strang

↳ RRef(A) can take 4 forms

- $r = n = m$ $\quad \text{rref}(A) = I$ $\quad$ 1 solution
- $r = n < m$ $\quad \text{rref}(A) = \begin{bmatrix} I \\ 0 \end{bmatrix}$ $\quad$ 0 or 1 solution
- $r\ m < n$ $\quad \text{rref}(A) = \begin{bmatrix} I & F \end{bmatrix}$ $\quad$ ∞ solutions
- $r < m,\ r < n$ $\quad \text{rref}(A) = \begin{bmatrix} I & F \\ 0 & \end{bmatrix}$ $\quad$ 0 or ∞ solutions

(Lecture 8 of 18.06)

where F are free variables re values in x_p (x-particular)

Section: 7 Matrix Determinants

→ Concepts & applications

- Notation: $\det(A)$ or $|A|$ where A is a matrix & not a vector
- Only square matrix has determinant.
- Scalar, reflects entire matrix
  ↳ 0 if A has linear dependent cols.

computers have error so det(A) isn't great at discerning singular matrices outside of classrooms.

$$\det(A_{(2\times2)}) = |ad - cb|$$ ✓

$\det(A_{(3\times3)})$ = use a computer :)

determinants are used when calculating eigen values.

Section 8: Matrix Inverse

consider $3x = 1$, how to find x?

$$3^{-1}(x) = 3^{-1}(3)$$
$$x = 3^{-1}$$
$$= \frac{1}{3}$$

inverse of 3 undoes stretching by 3.

Matrix inverse follows a similar idea.

try $0x = 1$

$0^{-1}0x = 0^{-1}1$ $\times$ $\quad 0^{-1} = \frac{1}{0}$ ← undefined

its not possible to inverse out of 0

$$Ax = b$$ ← Note how we add A^{-1} on the same side of both eqtns

$$\underbrace{A^{-1}A}_{I}x = A^{-1}b$$ ← only works when $\det(A) \neq 0$ (square & full rank)

In practice: difficult to calculate for large matrices. To be avoided.

→ inverse of 2x2 matrices

- Always check if invertible first.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \rightarrow \begin{bmatrix} d & b \\ c & a \end{bmatrix} \rightarrow \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \rightarrow \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \frac{1}{ad-cb}$$

swap diagonal

reverse sign of off-diag.

divide by det(A)

→ MCA algorithm for matrix inverse

M → compute minors

C → " cofactors

A → adjugate matrix $(M \circ C)^T$

divide each element by determinant.

→ Compute inverse via row reduction

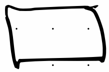
↳ uses rref

$[A \mid I] \rightarrow$ perform rref A $\rightarrow [I \mid A^{-1}]$

works because A^{-1} are the operations that undo A to I.

→ Left inverse & right inverse

consider a tall rectangular matrix:

$A =$  , $A^TA =$ 

sq. matrix w/ full rank

used in OLS $(A^TA)^{-1}A^TA = I$

$\underbrace{(A^TA)^{-1}A^T}$ left inverse only if A is full column rank

What about wide A?

$A =$  ← can never have full column rank, no left inverse

But right inverse!

$$AA^T(AA^T)^{-1} = I$$

$\underbrace{A^T(AA^T)^{-1}}$ right inverse

Note: you can't put l/r inverses on opposite side & create valid equations.

→ Inverse is unique

Ass A is invertible, $B = A^{-1}$, C distinct inverse of A

$\uparrow$ $\therefore B \neq C$

→ $AB = I$

$AC = I$

↑ $B = C$, creates contradiction

$\therefore$ inverse is unique

→ Pseudo inverse

↳ reduced rank matrix or

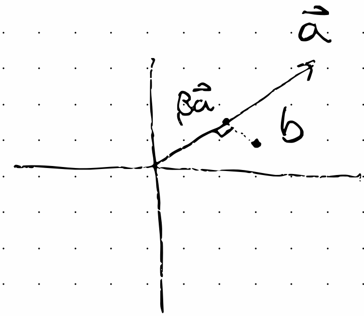
Denoted by A^{*} or $A^{\ddagger}$

- $AA^{\ddagger} \neq I$, however, $AA^{\ddagger}$ will have larger numbers on the diagonal.

Note: $AA^{\ddagger} \neq A^{\ddagger}A$

Section 9: Projections & orthogonalization

→ Projections in $\mathbb{R}^2$



What scalar do we multiply by $\vec{a}$ to reduce $(b-\beta\vec{a})$

↳ parallel ∴ $\vec{a}^T\cdot(b-\beta\vec{a})=0$

∴ $\boxed{\beta = \frac{\vec{a}^T b}{\vec{a}^T a}}$ ← division works because a^Ta is a scalar

↳ projects b onto $\vec{a}$

→ Projections in $\mathbb{R}^n$

↳ rework $\vec{a}^T(b-\beta\vec{a})=0$ into matrix form:

$A^T(b-Ax)=0$ ∴ $\boxed{(A^TA)^{-1}A^Tb = x}$ ← weights

← left inverse of A

If A is wide matrix then this method will break down.

How to think about this further...

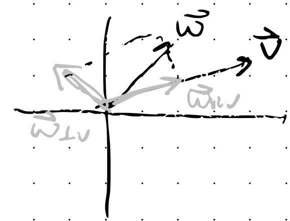
$$\underbrace{A^TA}x = \underbrace{A^Tb}$$
$$Cx = d$$

"d is in C(C)"

"x shows how we combine columns"

→ <u>Orthogonal & parallel vector components</u>

decompose w into components ⊥ & ∥ to v.



ie $\vec{w} = \vec{w}_{\parallel \vec{v}} + \vec{w}_{\perp \vec{v}}$

formula for $\vec{w}_{\parallel v}$ is in last section

$$\boxed{\vec{w}_{\parallel \vec{v}} = \frac{\vec{w}^T\vec{v}}{\vec{v}^T\vec{v}}\vec{v}}$$

$$\vec{w}_{\perp v} = \boxed{\vec{w} - \vec{w}_{\parallel \vec{v}}} = \vec{w} - \frac{\vec{w}^T\vec{v}}{\vec{v}^Tv}\vec{v}$$

example: $w\begin{bmatrix}4\\0\end{bmatrix}$ $v\begin{bmatrix}2\\3\end{bmatrix}$

$$\vec{v}_{\parallel \vec{w}} = \frac{\begin{bmatrix}4 & 0\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}}{\begin{bmatrix}4 & 0\end{bmatrix}\begin{bmatrix}4\\0\end{bmatrix}}\begin{bmatrix}4\\0\end{bmatrix} = \frac{8}{16}\begin{bmatrix}4\\0\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$$

$$\vec{v}_{\perp \vec{w}} = \begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}2\\0\end{bmatrix} = \begin{bmatrix}0\\3\end{bmatrix}$$

→ Orthogonal Matrices Q

↳ All columns are pairwise orthogonal
↳ each column has a magnitude of 1.

$$\boxed{Q^TQ = I} = Q^{-1}Q$$

← why? dot products of cols in Q with other columns = 1 or 0

Note that if Q is not square then only one of Q^TQ or QQ^T will make I

note that Q^T will not necessarily be Q either

→ $\frac{1}{3}\begin{bmatrix} 2 & 1 \\ -2 & 2 \\ 1 & 2 \end{bmatrix} = A$ $\quad A^TA = I_2 \quad AA^T \neq Q$

$A^T \neq Q$

→ Graham Schmidt & QR decomposition

↳ use orthogonalization technique to turn any matrix into an orthonormal matrix.

$\begin{bmatrix} | & | & | \\ c_1 & c_2 & c_3 \\ | & | & | \end{bmatrix}$

1. c_1 unit vector
2. $c_2 \perp c_1$
3. $c_3 \perp \{c_1, c_2\}$
4. etc

⋮

- lose information
- cannot retrieve info

→ QR decomposition ($A = QR$)

↳ R retains information lost by making Q

↳ How to get R?

$$Q^T A = \underbrace{Q^T Q}_{I} R \Rightarrow \boxed{Q^T A = R}$$

Graham schmidt needs to be done first.

(or use pc. lol)

→ Matrix Inverse via QR decomp

↳ $A = QR \rightarrow A^{-1} = (QR)^{-1}$

$$= R^{-1} Q^{-1}$$

$$= R^{-1} Q^T$$

upper triangular

easy inverse

→ Section 10: OLS & fitting in Stats

world is complex → use simplified model to explain the world.

"How do we find optimal parameters?"

<u>Step 1</u>. Find equation to express your data

ie $\text{height} = \beta_1 \text{sex} + \beta_2 \text{parents heights} + \beta_3 \text{nutrition}$

<u>Step 2</u>: Map data into model eqtns.

sex	p.heig	nutri	-	height

↖ each observation is an eqtn.

<u>Step 3</u>: convert into system of equations

$$Ax = b \Rightarrow y = X\beta$$

<u>Step 4</u>: solve for x

<u>Step 5</u>: evaluate model & infer

→ Section 11: eigendecomposition

↳ What are eigenvectors & eigenvalues?

Uses: differential eqtns
find axes of geometric shapes
image compression
Stats / data science

→ defined ONLY for $m \times m$ matrices.

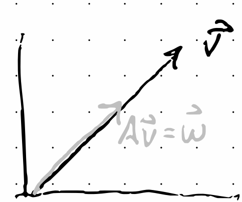
→ extracts eigenvalues & eigenvectors

↳ 1 for each col./row

What are these things?

↳ Matrix A can be described as
- rotation
- stretch

eigenvector doesn't change basis after A is applied. eigenvalue is the amount vector stretches by.

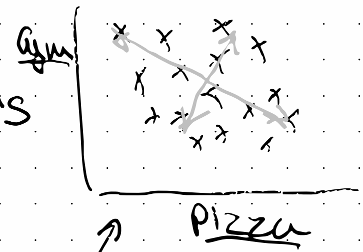


$$\boxed{A\vec{v} = \lambda\vec{v}}$$

Evec Eval

Statistical Picture

↳ plot gymgoers/pizzagoers



What vectors best describe this system?

↳ two ⊥ vectors w/ max variance

(PCA)

→ <u>Finding eigenvalues</u>

↳ from previous lectures →

$$Av = \lambda v$$

$$Av - \lambda v = 0$$

$$\therefore \boxed{(A - \lambda I)v = 0}$$

Search for vector that lies in the nullspace of A shifted by eigenvalues on the diagonal

↳ if $(A - \lambda I)$ is singular then its det is 0

$$|A - \lambda I| = 0 \leftarrow \text{solve for } \lambda$$

if $A - \lambda I$ is not singular then the nullspace is spanned by the zero vector.

solution for λ is a set of ε-values

ie $\left| \begin{bmatrix} 1 & 5 \\ 2 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$

$\begin{vmatrix} 1-\lambda & 5 \\ 2 & 4-\lambda \end{vmatrix} = 0 \Rightarrow (1-\lambda)(4-\lambda) - 10 = 0$

$\therefore \lambda = \{6, -1\}$

↳ Note that an $n \times n$ matrix will have n eigen values due to fundamental theorem of algebra

Shortcut: $\lambda^2 - tr(A)\lambda + \det$

→ <u>Finding ε-vectors</u>

ε-vectors are encrypted in lock. decrypt w/ λ (ε-val.)

↳ $v \in N(A - \lambda I)$

example $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, $\lambda^2 - 2\lambda - 3 = 0$

$= (\lambda - 3)(\lambda + 1)$

$\therefore \lambda = \{3, -1\}$

Step 2: Shift A by λ

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

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

or any scaled multiple thereof.

combine to make ε-~~matrix~~ vec matrix?

ε-vec matrix $= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = V$

ε-val matrix $\begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} = D$

symmetric & full rank

can also be scaled so norm is 1.

→ More examples by hand

$$\begin{bmatrix} 3 & 1 \\ 4 & 6 \end{bmatrix} \qquad \lambda^2 - 9\lambda + 14 = 0 \qquad \begin{matrix} 1 & 14 \\ 2 & 7 \end{matrix}$$

$$\begin{bmatrix} 3-\lambda & 1 \\ 4 & 6-\lambda \end{bmatrix} \qquad (3-\lambda)(6-\lambda) - 4 = 0$$

$$\lambda^2 - 9\lambda + 18 - 4 = 0$$

$$(\lambda - 2)(\lambda - 7) = 0$$

$$\therefore \lambda = \{7, 2\}$$

$$\begin{bmatrix} 3-7 & 1 \\ 4 & 6-7 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = 0 = \begin{bmatrix} -4 & 1 \\ 4 & -1 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = 0$$

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$\begin{bmatrix} 3-2 & 1 \\ 4 & 6-2 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = 0 = \begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = 0$$

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\therefore U = \begin{bmatrix} 1 & 1 \\ 4 & -1 \end{bmatrix}$$

$$D = \begin{bmatrix} 7 & 0 \\ 0 & 2 \end{bmatrix}$$

Diagonalisation

$m \times m$ matrix $\Rightarrow$ columns $1, 2, \dots, m$

→ E-vals

→ E-vec

$\underbrace{1 \; 2 \cdots m}_{\text{combine to make matrices}}$

$$VD = \begin{bmatrix} | & | & | & \\ v_1 & v_2 & v & \cdots \\ | & | & | & \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \end{bmatrix} = \begin{bmatrix} | & | & \\ \lambda_1 v_1 & \lambda_2 v_2 & \cdots \\ | & | & \end{bmatrix}$$

$$\boxed{AV = V\Lambda}$$

Λ is a diagonal matrix of E-vals

V is matrix of E-vecs

A is original matrix

thus... $\boxed{A = V\Lambda V^{-1}}$

V must have an inverse (full rank)

easy to work w/ diagonal matrices.

Matrix Powers w/ diagonalisation

↳ note similarity between Λ & rref(A)

System is easy to solve.

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

$$\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 8 & 9 \end{bmatrix}$$

↳ okay, simple enough

$$\begin{bmatrix} -2 & 2 & -3 \\ -4 & 1 & -6 \\ -1 & 2 & 0 \end{bmatrix}^4 = ?$$

← tedious to solve by hand

$$A = V \Lambda V^{-1}$$

$$A^4 = V \Lambda \overbrace{V^{-1} V}^{I} \Lambda \overbrace{V^{-1} V}^{I} \Lambda \overbrace{V^{-1} V}^{I} \Lambda V^{-1}$$

$$= V \Lambda^4 V^{-1}$$

↑ easy to compute for diagonal matrix

Question...

if eigendecomp $(A) = V \Lambda V^{-1}$
what is eigendecomp $(A^n) = V \Lambda^n V^{-1}$

why? $Ax = \lambda x \rightarrow A^2 x = \lambda^2 x$

$\downarrow \times A$

$$AAx = A\lambda x = \lambda \underbrace{Ax}_{\lambda x} = \lambda^2 x$$

$\mathcal{E}$-vecs w/ distinct $\mathcal{E}$-vals

↳ to be demonstrated: different λ are associated w/ different v.

$\lambda_1 \neq \lambda_2$, $\beta_1 v_1 = \beta_2 v_2$ or $\vec{V}^T \vec{\beta} = 0$

→ premult by A: $A\beta_1 v_1 + A\beta_2 v_2 = 0$

→ because $Av = \lambda v$: $\lambda_1 \beta_1 v_1 + \lambda_2 \beta_2 v_2 = 0$

→ mult ① by λ_1: $\ominus \; \lambda_1 \beta_1 v_1 + \cancel{\lambda_2} \lambda_1 \beta_2 v_2 = 0$

$$(\lambda_2 - \lambda_1)\beta_2 v_2 = 0$$

cannot be 0 ↗ $\beta_2 = 0$ ↖ cannot be 0 if $\beta_2 = 0$ then β_1 must also be 0

↳ if only way to create $\vec{0}$ is by $\vec{\beta} = \vec{0}$ then vectors are lin. indep.

↳ makes no sense

→ E-vecs of ~~dis~~ repeated E-vals

↳ Ass $\lambda_1 = \lambda_2$; $\beta_1 v_1 = \beta_2 v_2$

could be 0

from last section: $(\lambda_2 - \lambda_1)\beta_2 v_2 = 0$

can't be 0

absolutely 0

no contradictions.

$\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix} \rightarrow \lambda^2 - 6\lambda + 9 = 0 \rightarrow (\lambda - 3)^2 = 0$

$\lambda = \{3, 3\}$

$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ← any other vector I can use?

β must be 0, all vecs are in same subspace.

What are the possibilities?

↳ $\mathcal{E}$-vector, $\mathcal{E}$-plane (& above)

<u>Eigendecomposition of symmetric matrices</u>

symmetric, $A = A^T$ $\qquad \lambda_1 \neq \lambda_2$

$$\lambda_1 v_1^T v_2 = (Av_1)^T v_2 = v_1^T A^T v_2 = v_1^T \lambda_2 v_2$$
$$= \lambda_2 v_1^T v_2$$

$$\therefore (\lambda_1 - \lambda_2) \underbrace{v_1^T v_2}_{\text{is } 0} = 0$$

$\uparrow$
$!= 0$

(as long as $\lambda_1 \neq \lambda_2$ & A is sym)

example

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \rightarrow \lambda^2 - 6\lambda + 8 = 0$$
$$(\lambda - 2)(\lambda - 4) = 0$$
$$\therefore \lambda = \{2, 4\} \quad \begin{cases} v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{cases}$$

↳ $v_1 \perp v_2$

implication of orthogonal E-vecs.

$$\underset{V}{\begin{bmatrix} & \\ & \end{bmatrix}}\underset{V^T}{\begin{bmatrix} & \\ & \end{bmatrix}} = \begin{bmatrix} \alpha & & 0 \\ & \beta & \\ 0 & & \ddots \end{bmatrix}$$

pure rotation matrix.

↳ if norm = 1 then $VV^T = I$

then $VV^T = VV^{-1}$

important because symmetric matrices are ubiquitous in applied mathematics.

↳ ie cov. matrices

Eigendecomp. on singular matrices

$$\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \rightarrow \lambda^2 - 3\lambda + 0 = 0$$

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

$$\therefore \lambda = \{0, 3\}$$

$\lambda = 0$ is always solution for singular matrix.

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

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

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

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

E-vecs don't have to be $\vec{0}$

Generalised eigendecomposition

$$Av = \lambda v$$
$$Av = \lambda I v$$
$$Av = \lambda R v$$

R = right size to make eqtn valid.

(diagonalisation)

what is this?

$$\underbrace{R^{-1}A}_{=C} v = \lambda v \qquad \therefore \quad Cv = \lambda v$$

where $R^{-1}A$ = "$\frac{A}{R}$" maximises difference between A & R.

↳ signal & noise?

What if R is not invertible?

$$Av - \lambda R v = 0$$

$$\underbrace{(A - \lambda R)}_{\text{singular}} v = 0$$

singular, find vector(s) in the nullspace.

Section 12: Singular Value Decomposition (SVD)

↳ also works on rectangular matrices.

remember $A^TA \neq AA^T$ when $n \neq m$

also $C(A) = C(AA^T)$
$C(A^T) = C(A^TA^{TT}) = C(A^TA)$

 Hmmm

general idea: $A = \overset{m\times m}{U}\ \overset{m\times n}{\Sigma}\ \overset{n\times n}{V^T}$

orth basis for $C(A^T)$

orth. basis for $C(A)$

singular values (diagonals)

$$A^TA = (U\Sigma V^T)^T U\Sigma U^T \quad AA^T = U\Sigma U^T$$
$$= V\Sigma U^TU\Sigma U^T$$
$$= V\Sigma^2U^T = V\Lambda U^T$$

eigenvalues rooted

↳ Note that $AV = U\Sigma$ & $U^TA = \Sigma V^T$

→ SVD & four subspaces

↳ subspaces: Column, Row, Null, Left-Null

↳ SVD provides bases for these spaces

$$\underset{7\times3}{[A]} = \underset{7\times7}{[U]}\underset{7\times3}{[\Sigma]}\underset{3\times3}{[V^T]}$$

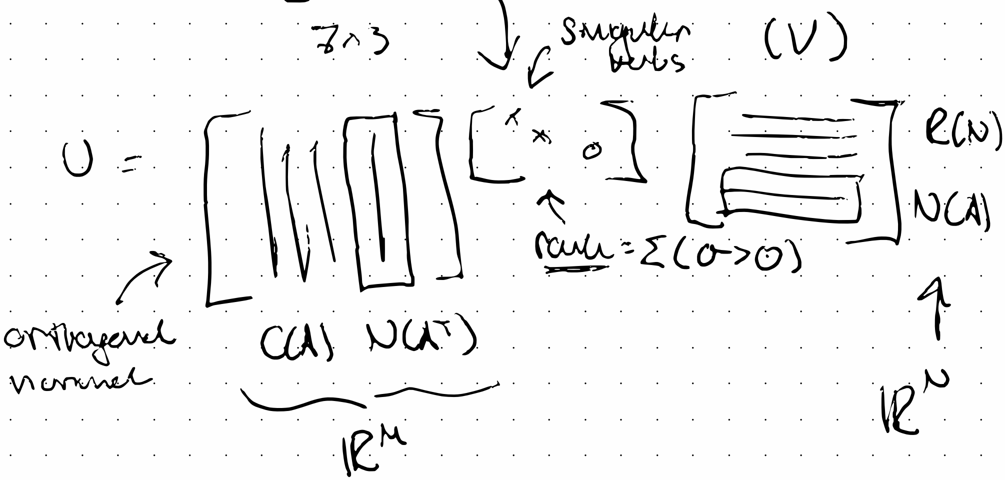
if A is wide
U is small &
V^T is big

r = 2

↳ 2 singular vals

$$\Sigma = \underset{7\times3}{\begin{bmatrix} \alpha & & 0 \\ & \beta & \\ & & 0 \\ 0 & & \end{bmatrix}}$$

N(A)



→ Magnitude of each col is 1.

→ Spectral theory of matrices

↳ rainbows → spectrum

matrix ← lots of information

↳ Meaning?

↳ Break matrix into component parts.

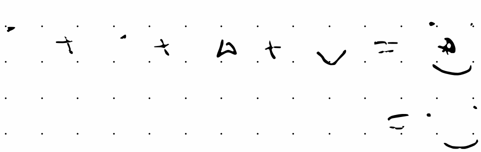
What does SVD show?

↳ important directions in a matrix

↳ how important each direction is is defined by singular values.

→ SVD for low rank approx.

↳ layered version



not important information

△ is removed but we can still understand picture

↳ only use subset of cols from SVD to create lower level picture

→ Convert singular values to perc variance

↳ singular vals are scale dep

$$\begin{bmatrix} 3 & 8 \\ 1 & 9 \end{bmatrix} = A \qquad \Sigma = \begin{bmatrix} 12.4 & 0 \\ 0 & 1.54 \end{bmatrix}$$

$$\Sigma_{A\times 100} = \begin{bmatrix} 124 & 0 \\ 0 & 15.4 \end{bmatrix}$$

encodes importance of directions

$$\sum_{i=1}^{N} \sigma_i = \text{total importance (variance)}$$

$$\therefore \text{importance} = \frac{\sigma_i}{\sum_{i=1}^{N} \sigma_i}$$

(plot scree plot)

→ SVD, matrix inv. pseudo inv.

$$A = U \Sigma V^T$$

$$A^{-1} = V^T \Sigma^{-1} U^{-1}$$

$$= V \Sigma^{-1} U^T$$

reciprocal of diag elements

if A isn't full rank, 0 or stay same

→ <u>Condition number of a matrix</u>

$$C.N = \frac{\sigma_{max}}{\sigma_{min}} = \kappa$$? what does this mean?

"kappa" singular? $\frac{\sigma_{max}}{\sigma}$

well cond. when large & ill conditioned when small

is an indicator of large scale dominance of a feature in a matrix

↳ Also "stability"

Section 12: Quadratic form & definiteness

↳ comes math (geom), programming & stats

consider $\underset{M \times M}{S} \times \underset{M \times 1}{w} = \underset{M \times 1}{Sw} \Rightarrow \boxed{\underset{1 \times 1}{w^T S w}}$

{ quadratic form

↳ "energy in S at each coord in w"

example:

$$[3\ {-1}]\begin{bmatrix} 2 & 4 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 3 \\ -1 \end{bmatrix} = [3\ {-1}]\begin{bmatrix} 2 \\ -3 \end{bmatrix} = 9$$

generally...

$$\underset{S}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \quad \underset{w}{\begin{bmatrix} x \\ y \end{bmatrix}} \Rightarrow w^T S w = ax^2 + (b+c)xy + dy^2$$

↳ +ve values describe how "definite" a matrix is.

what if we have different vectors?

	w^TSw
$w_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$	9
$w_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$	19
$w_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$	2

↳ back to general $\underset{S}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \underset{w}{\begin{bmatrix} x \\ y \end{bmatrix}}$

If we vary x & y

$$w^TSw = \underbrace{ax^2}_{\to I\alpha} + (b+c)xy + \underbrace{dy^2}_{\to I\alpha}$$

could balance

if S is symmetric $[w^TSw]^T = w^TSw$

↳ is S is larger than 2x2 then higher order quadratics will be generated

→ Quad. form in geometry

↳ think of quad form as function

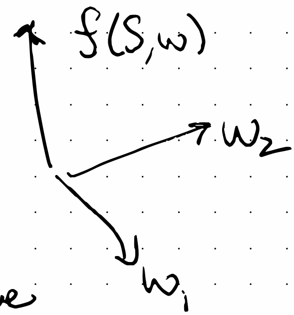
$$f(S, w_{ij}) = \text{scalar}$$

ith & jth elements in vec w

w can be larger than 2

creates a picture, a surface.

use many points of w_1, w_2 to make surface.



example

→ thus, we get an idea of energy

↳ if surface is always >0 then +ve def.

& vice versa.

→ <u>Normalized Quad form</u>

consider $\underset{w}{\arg\max}\left\{ w^T S w \right\}$

$$S = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$$

in which direction does the energy increase the most?

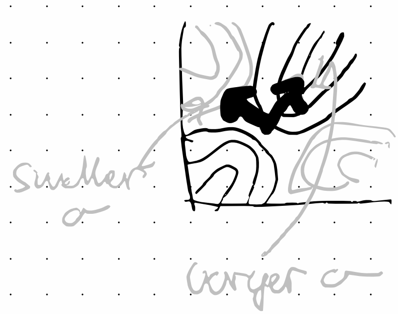
↳ normalize by magnitude of vec

ie $\underset{w}{\arg\max}\left\{ \frac{w^T S w}{w^T w} \right\}$ proving my coords, not interesting for us

→ <u>Eigenvectors & quad form surface</u>

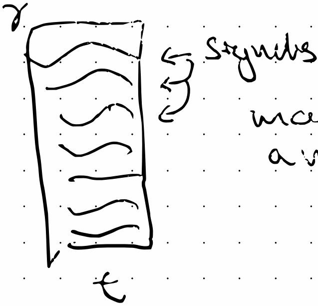
So how do E-vecs tie into this?

↳ they point towards ridges & valleys



→ Application of norm. quad. form & PCA

↳ if matrix is cov matrix then eigendecomp is PCA



makes a matrix → A → pairwise lin interactions is described by AA^T

↳ sq., sym., full rank

thus, AA^T is a cov matrix.

↳ ~~we~~ find linear combo that max cov in dataset

ie $\lambda = \arg\max_{w} \left\{ \frac{w^T S w}{w^T w} \right\}$

eigvecs point to ridge & valleys

why is this? turn to find out!

if we think of set of vectors then

$$\Lambda = (W^T W)^{-1} W^T S W$$

matrix of input vectors
(sq, full rank)

$$\therefore \Lambda = W^{-1} W^{-T} W^T S W$$

$$= W^{-1} S W$$

$$\boxed{W \Lambda = S W}$$

matrix version of E-val eqtn.
$(AV = V\Lambda)$

↳ thus we have n # λ in Λ (diag)

... he's not explaining this very well...

I hope he explains Λ better in his next course.

→ <u>Quad form of gen. GD</u>

↳ what if we replace implicit I for B?

$$\lambda = \arg\max_{w} \left\{ \frac{w^T A w}{w^T I w} \right\}$$

put B here

... notes end here because teaching quality is bad.