Essence of Linear Algebra (3B1B)

Chp 2 – Linear combinations, span, and basis vectors.

In the 2D coordinate system, we have 2 basis vectors:

- $\hat{\imath}$ = x-direction
- $\hat{\jmath}$ = y-direction

These define our xy-coordinate system.

Any 2D vector can be written as a linear combination of $\hat{\imath}$ and $\hat{\jmath}$.

. $\begin{bmatrix} a \\ b \end{bmatrix} = a\hat{\imath} + b\hat{\imath}$ (a, b: scalars — they scale the basis vectors!)

Span = Set of all linear combinations of vectors $\vec{v}$ and $\vec{w}$.

- i.e. If we let a, b vary over all $\mathbb{R}$, where can $a\vec{v} + b\vec{w}$ reach?
- For any two non-zero, non-parallel vectors $\vec{v}$ and $\vec{w}$ $\Rightarrow$ span = all R^2
- But if $\vec{v} = \lambda \vec{w}$ (i.e. parallel): span is just a line!

Span of two 3D vectors $\Rightarrow$ PLANE

Add a third non-parallel vector $\rightsquigarrow$ we are translating the plane in the dir. of the 3^{rd} vector $\Rightarrow$ all R^3!

If an additional vector does NOT add to the span of another set of vector(s)

$\Rightarrow$ New vector is linearly dependent on the others $\Rightarrow$ can be expressed as a linear combination of the others!

$\Rightarrow$ Correspondingly, if new vector adds to span $\Rightarrow$ it is linearly independent

> " The basis of a vector space is a set of linearly-dependent vectors that span the full space. "

Chp 3 — Linear transformations and matrices.

A linear transformation is a function that converts some input vector to some output vector.

$\Rightarrow$ LINEAR: ① Lines remain lines ; ② Origin remains fixed

$\downarrow$

Grid lines remain parallel and evenly-spaced!

We can define a linear transformation as some matrix showing where each basis vector ends.

↳ E.g. 2D transformation → matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, where:

Transformed $\hat{\imath}$: $\begin{bmatrix} a \\ c \end{bmatrix}$, transformed $\hat{\jmath}$ = $\begin{bmatrix} b \\ d \end{bmatrix}$

E.g Take the vector $\vec{v} = -1\hat{\imath} + 2\hat{\jmath}$, and transformation $\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix}$

$\therefore$ (Transformed $\vec{v}$ = transformed $\hat{\imath}$ + transformed $\hat{\jmath}$)

$\therefore$ Transformed $\vec{v} = -1\begin{pmatrix} 1 \\ -2 \end{pmatrix} + 2\begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$

Generalizing, for some transformation $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$:

- Transformed $\hat{\imath}$: $\begin{pmatrix} a \\ c \end{pmatrix}$, transformed $\hat{\jmath}$: $\begin{pmatrix} b \\ d \end{pmatrix}$
- $\left.\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} &= x\begin{pmatrix} a \\ c \end{pmatrix} + y\begin{pmatrix} b \\ d \end{pmatrix} \\ &= \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix} \end{aligned}\right\}$ Adding the scaled versions of the basis vectors.

NOTE: If the transformed basis vectors are linearly dependent, then the linear transformation squishes all space into one dimension.

> A matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ represents a linear transformation, such that $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$ are the transformed basis vectors, and $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = x\begin{pmatrix} a \\ c \end{pmatrix} + y\begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}$.

Chp 4 — Matrix multiplication as composition:

Consider applying two transformations back-to-back.

- The new transformation is still a linear transformation
- Called the composition of the two linear transformations

⇒ Can be expressed as a single matrix!

Say we apply the transformation $\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$, then the transformation $\begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}$

⇒ Result of: $\underbrace{\begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \underbrace{\begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}}\begin{pmatrix} x \\ y \end{pmatrix}$

product

Basis vectors:

$\hat{\imath} \Rightarrow \begin{pmatrix} a_1 \\ a_3 \end{pmatrix} \Rightarrow \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}\begin{pmatrix} a_1 \\ a_3 \end{pmatrix} = a_1\begin{pmatrix} b_1 \\ b_3 \end{pmatrix} + a_3\begin{pmatrix} b_2 \\ b_4 \end{pmatrix} = \begin{pmatrix} a_1b_1 + a_3b_2 \\ a_1b_3 + a_3b_4 \end{pmatrix}$

$\hat{\jmath} \Rightarrow \begin{pmatrix} a_2 \\ a_4 \end{pmatrix} \Rightarrow \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}\begin{pmatrix} a_2 \\ a_4 \end{pmatrix} = a_2\begin{pmatrix} b_1 \\ b_3 \end{pmatrix} + a_4\begin{pmatrix} b_2 \\ b_4 \end{pmatrix} = \begin{pmatrix} a_2b_1 + a_4b_2 \\ a_2b_3 + a_4b_4 \end{pmatrix}$

Overall.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}$$

1st column: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{pmatrix} e \\ g \end{pmatrix}$ — applying 1st matrix to 1st col of 2nd

2nd col: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{pmatrix} f \\ h \end{pmatrix}$ — applying 1st matrix to 2nd col of 2nd

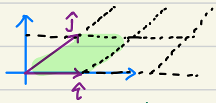
> Multiplying two matrices $M_1 \times M_2$ is equivalent to transforming $\begin{pmatrix} x \\ y \end{pmatrix}$ with M_2, then transforming the result with M_1.

Chp 6 — The determinant:

Consider how some transformations stretch space, while others squeeze space

- E.g $\begin{bmatrix}3&0\\0&2\end{bmatrix}$: $\hat{\imath}$ scaled by 3x, $\hat{\jmath}$ scaled by 2x
  - ⇒ Linear transformation has scaled area by 6x

- E.g. Shear $\begin{bmatrix}1&1\\0&1\end{bmatrix}$:



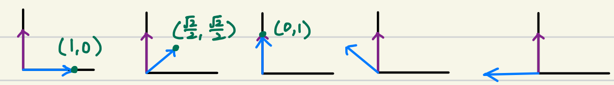
  - ⇒ Area of unit square is still 1.

Scaling factor is constant for a given transformation — all squares, and hence all areas, are scaled by the same amount.
↳ because all areas can be represented as many infinitely-small squares!

Determinant = "area scaling factor" of a transformation.

- If determinant = 0: transformation "squishes" area into a line!
- If determinant < 0: "invert the orientation of space" (space flips!)
  - ↳ Recall $\hat{\imath}$ starts on the right of $\hat{\jmath}$.
    Orientation is flipped ⇔ $\hat{\jmath}$ is on the right of $\hat{\imath}$.
  - ↳ Absolute value of determinant is still the area scaling factor!

(Intuition: As $\hat{\imath}$ rotates anti-clockwise, det goes from $1 \to 0 \to -1$)

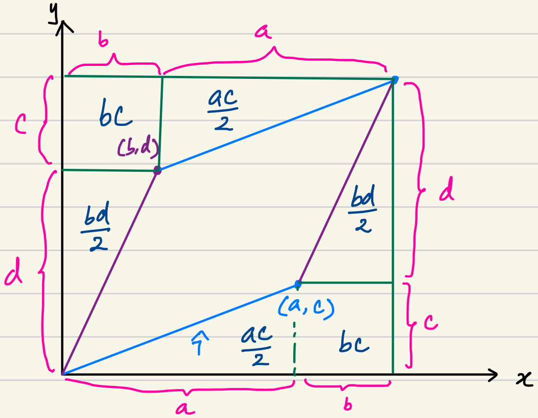


det: +1 det: $+\frac{\sqrt{2}}{2}$ det = 0 det: $-\frac{\sqrt{2}}{2}$ det = −1

Correspondingly, det (3D transformation) = volume of parallelipiped from unit cube.
↳ det(M) < 0 ⇒ $\hat{\imath}, \hat{\jmath}, \hat{k}$ no longer follow right-hand rule

Computing the determinant:

Consider $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$: $\hat{\imath}$ goes to $\binom{a}{c}$, $\hat{\jmath}$ goes to $\binom{b}{d}$



Area of parallelogram

$= (a+b)(c+d) - ac - bd - 2bc$

$= ac + ad + bc + bd - ac - bd - 2bc$

$= ad - bc$

$\therefore \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$

Very intuitively, $\det(M_1 M_2) = \det(M_1) \cdot \det(M_2)$

$\Rightarrow$ We are just applying back-to-back transformations!

Chp 7 — Inverse matrices, column space, and null space.

Matrices are great for solving a linear system of equations

$$\begin{aligned} 2x + 5y + 3z &= -3 \\ 4x + 0y + 8z &= 0 \\ 1x + 3y + 0z &= 2 \end{aligned} \iff \underbrace{\begin{bmatrix} 2 & 5 & 3 \\ 4 & 0 & 8 \\ 1 & 3 & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}}_{\vec{v}}$$

$\Downarrow$

Represented by $A\vec{x} = \vec{v}$:

$\Rightarrow$ We are looking for a vector $\vec{x}$, that after transforming w/ matrix A, gives us vector $\vec{v}$!

If $\det(A) \neq 0$. Only ONE vector $\vec{x}$ such that $A\vec{x} = \vec{v}$ $(\vec{x} = A^{-1}\vec{v})$

(i.e. start from $\vec{v}$, and "rewind" the transformation A $\rightsquigarrow$ 1-to-1')

$\Rightarrow A^{-1}A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$

$\Downarrow$

Corresponds to 2 eqts, 2 unknowns, have one unique set of solutions

If $\det(A) = 0$. Transformation $\vec{A}$ squishes $\vec{x}$ into a smaller dimension

$\Rightarrow$ No inverse! Functions can only map one-to-one, cannot do line $\to$ 2D plane.

↳ We only have a solution when output vector $\vec{v}$ lives in the lower-dimensional space.

↳ Intuitively, a transformation of 3D $\to$ 1D line feels more restrictive than 3D $\to$ 1D line, yet both are $\det(A) = 0$.

↳ **Rank**: Number of dimensions in OUTPUT in a transformation.

- E.g. Rank 1: 2D $\to$ 1D, 3D $\to$ 1D, etc.

Column space = Set of all possible outputs $A\vec{v}$, given a matrix A

↳ Each column in A = where one basis vector lands

↳ The SPAN of transformed basis vectors = COLUMN SPACE!

↳ the columns!

Using this new terminology:

- Solution to $A\vec{x} = \vec{v}$ exists if $\vec{v}$ lives in the <u>column space</u> of A
- Full Rank: # of dimensions in column space = # of columns in A
- $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is always in the column space ⇝ origin is never moved!
  - ↳ In full rank transformations, only $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ ends up at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$
  - ↳ But in non-full rank transformations, entire lines/planes can be squished to $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$!
- Null space / Kernel: Set of all space that gets transformed into $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ by A.
  - ↳ When $\vec{v} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, the null space is the solution to $A\vec{x} = \vec{v}$.

Chp 8 — Nonsquare matrices as transformations between dimensions:

Consider a 2D → 3D transformation: (3 × 2 matrix)

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

$\underbrace{\quad}_{\text{transformed } \hat{\imath}}$ $\underbrace{\quad}_{\text{transformed } \hat{\jmath}}$

2 colums = input is 2D (2 basis vectors)
3 rows = output is 3D

- The COLUMN SPACE is a <u>plane</u>
  (all output vectors exist on some plane)

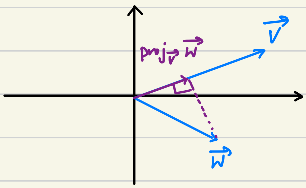
⇒ Matrix is FULL RANK
(output column space = input col. space = $\mathbb{R}^2$)

Chp 9 — Dot products and duality

Numerically, dot product of same length vectors = pairing up terms + summing

↳ e.g $\begin{bmatrix}1\\2\\3\end{bmatrix} \cdot \begin{bmatrix}4\\5\\6\end{bmatrix} = (1)(4) + (2)(5) + (3)(6) = 32$

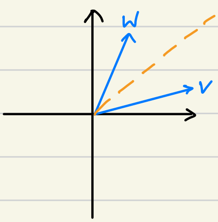
Geometrically: $\underbrace{\begin{bmatrix}4\\1\end{bmatrix}}_{\vec{v}} \cdot \underbrace{\begin{bmatrix}2\\-1\end{bmatrix}}_{\vec{w}}$.



Dot product = (length of $\vec{v}$) × (length of projection of $\vec{w}$ on $\vec{v}$)

- Perpendicular: $\|proj_{\vec{v}} \vec{w}\| = 0 \rightsquigarrow$ Dot $= 0$
- Diff. directions: $proj_{\vec{v}} \vec{w}$ in opposite direction to $\vec{v} \Rightarrow$ Dot < 0

Why is the dot product commutative? $(a \cdot b = b \cdot a)$



① If $\|\vec{w}\| = \|\vec{v}\|$, then we can draw line of symmetry

$\Rightarrow \|proj_{\vec{w}} \vec{v}\| = \|proj_{\vec{v}} \vec{w}\|$

② Then, if we scale $\vec{v}$ to $2\vec{v}$:

- $(2\vec{v} \cdot \vec{w}) = 2 \cdot \|\vec{v}\| \cdot \underline{\|proj_{\vec{v}} \vec{w}\|} = 2(\vec{v} \cdot \vec{w})$

(scaling $\vec{v}$ doesn't change $\vec{w}$ projection)

- $(\vec{w} \cdot 2\vec{v}) = \|\vec{w}\| \cdot \underline{2 \cdot \|proj_{\vec{w}} \vec{v}\|} = 2(\vec{v} \cdot \vec{w})$

scaling $\vec{v}$ by 2 scales its projection by 2!

Consider linear transformations from $R^{\geq 2}$ to R^1 (number line).

- Property: Line of evenly-spaced dots in input space → still evenly spaced on number line after transformation
- Each transformed basis vector is a single number!

DUALITY in Mathematics
A vector: both an arrow in space, and encoding a transformation that collapses all space into a number line along that arrow

$$\text{Example: } L(\vec{v}) = \begin{bmatrix} \overset{\text{transformed } \hat{\imath}}{1} & \overset{\text{transformed } \hat{\jmath}}{-2} \end{bmatrix}$$

- Apply this transformation to $\begin{bmatrix} 4 \\ 3 \end{bmatrix} = 4\hat{\imath} + 3\hat{\jmath}$

Result: $4(1) + 3(-2) = \underline{-2}$

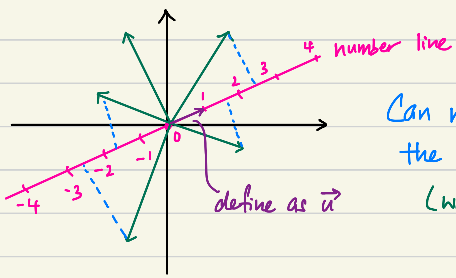
(this looks just like the dot product!)

$$\overset{\text{matrix}}{\begin{bmatrix} 1 & -2 \end{bmatrix}} \overset{\text{vector}}{\begin{bmatrix} 4 \\ 3 \end{bmatrix}} = (4)(1) + (3)(-2)$$

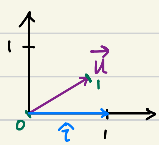
↳ Notice the association between 1×2 matrices $[x \ \ y]$ and 2D vectors $\begin{bmatrix} x \\ y \end{bmatrix}$

Can go back at forth!

Some connection between linear transformations taking vectors to numbers and vectors themselves... ??



Can map any 2D vector to some number on the number line ⇝ this is a linear transformation!
(what is the 1x2 matrix defining this?)



$\|\vec{u}\| = \|\hat{\imath}\| = 1$ ⇝ symmetric!!
(so $\|proj_{\vec{u}}\, \hat{\imath}\| = \|proj_{\hat{\imath}}\, \vec{u}\| = u_x$)
Similarly, $\|proj_{\vec{u}}\, \hat{\jmath}\| = \|proj_{\hat{\jmath}}\, \vec{u}\| = u_y$!

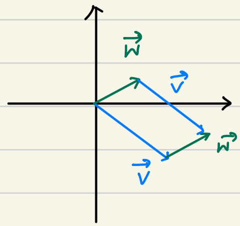
Hence, for any arbitrary vector $\begin{bmatrix} x \\ y \end{bmatrix} = x\hat{\imath} + y\hat{\jmath}$

- TRANSFORMATION by $[u_x \ \ u_y]$ = $[u_x \ \ u_y] \begin{bmatrix} x \\ y \end{bmatrix} = (u_x)(x) + (u_y)(y)$
  ($\hat{\imath}$ goes to u_x, $\hat{\jmath}$ goes to u_y)
- DOT PRODUCT: $\begin{bmatrix} u_x \\ u_y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = (u_x)(x) + (u_y)(y)$
  (projecting $\begin{bmatrix} x \\ y \end{bmatrix}$ on $\vec{u}$)

EQUAL!

⇒ Hence, for non-unit $\vec{u}$, dot product = taking projection on unit $\vec{u}$, then SCALING by $\|\vec{u}\|$.

Chp 10 — Cross product:



$\vec{v} \times \vec{w}$ = Area spanned by parallelogram

- If $\vec{v}$ is on the RIGHT of $\vec{w}$

  $\Rightarrow \vec{v} \times \vec{w} > 0$ (recall $\hat{\imath} \times \hat{\jmath} = 1$)

- If $\vec{v}$ is on the LEFT of $\vec{w}$

  $\Rightarrow \vec{v} \times \vec{w} < 0$ ($\hat{\jmath} \times \hat{\imath} = -1$)

Order matters!!

Consider $\vec{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$: $\vec{v} \times \vec{w} = \det\left(\begin{array}{cc} \vec{v} & \vec{w} \\ \multicolumn{2}{c}{\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}} \end{array} \right)$

- $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$ transforms $\hat{\imath}$ to $\vec{v}$, and $\hat{\jmath}$ to $\vec{w}$

$\Rightarrow$ We end up scaling the unit square $\rightarrow$ the parallelogram we want!

But wait.. cross product a VECTOR.

- Length = area of parallelogram
- Direction = perpendicular to parallelogram + right-hand rule!

Computing the cross product:

$$\det\left(\begin{bmatrix} \hat{\imath} & v_1 & w_1 \\ \hat{\jmath} & v_2 & w_2 \\ \hat{k} & v_3 & w_3 \end{bmatrix} \right)$$

why does this work ??

Chp 11 — Cross products in the light of linear transformations.

Recall duality — any 2D-to-1D transformation is associated w/ some vector.
(conducting transformation $\Leftrightarrow$ taking dot product w/ vector)

Plan for cross product: ① Define 3D → 1D transformation in terms of $\vec{v}$ and $\vec{w}$
② Find its dual vector
③ Show the dual vector is $\vec{v} \times \vec{w}$

Recall $\vec{v} \times \vec{w}$ in 2D gives area of parallelogram.

Consider $f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \det\left(\begin{bmatrix} x & v_1 & w_1 \\ y & v_2 & w_2 \\ z & v_3 & w_3 \end{bmatrix}\right)$

- $\vec{v}$ and $\vec{w}$ are fixed
- Given some input vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightsquigarrow f$ transforms it into a <u>number</u>.
- For any $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, the function calculates the volume of the parallelapiped spanned by $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, $\vec{v}$, and $\vec{w}$.

By duality, we know there is some 1×3 vector $[a\ b\ c]$ describing the transformation
$\Rightarrow$ We want some: $\underbrace{\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{p_1 x + p_2 y + p_3 z} = \underbrace{\det\left(\begin{bmatrix} x & v_1 & w_1 \\ y & v_2 & w_2 \\ z & v_3 & w_3 \end{bmatrix}\right)}_{x(v_2 w_3 - v_3 w_2) + y(v_3 w_1 - v_1 w_3) + z(v_1 w_2 - v_2 w_1)}$

Numerically, $\vec{v} \times \vec{w}$ is the vector that, when dotted with any vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$, gives $\det\left(\begin{bmatrix} x & v_1 & z_1 \\ y & v_2 & z_2 \\ z & v_3 & z_3 \end{bmatrix}\right)$ — i.e. the volume spanned by $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$, $\vec{v}$, and $\vec{w}$.

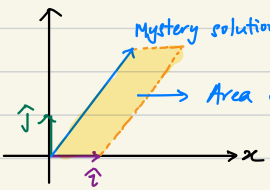
Geometrically: Recall that $\vec{p} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \text{proj}_{\vec{p}}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) \cdot \|\vec{p}\|$

- The volume of the parallelepiped spanned by $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$, $\vec{v}$, and $\vec{w}$
  = (area of parallelogram formed by $\vec{v}$ and $\vec{w}$)(proj. of $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ in a vector $\perp$ to $\vec{v}$ and $\vec{w}$)
  = (area of parallelogram)($\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ · some vector perpendicular to $\vec{v}$ and $\vec{w}$)
  = $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cdot \vec{p}$, where $\|\vec{p}\|$ = area of parallelogram & $\vec{p} \perp \vec{v}$ and w)

Chp 12 — Cramer's rule, explained geometrically

Consider: $3x + 2y = -4$
$-1x + 2y = -2$
$\iff \begin{bmatrix} 3 & 2 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$

"Which input vector $\begin{bmatrix} x \\ y \end{bmatrix}$ will land on $\begin{bmatrix} -4 \\ -2 \end{bmatrix}$ after being transformed by $\begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}$?"



→ Area of the parallelogram formed by $\hat{i}$ and $\begin{bmatrix} x \\ y \end{bmatrix} = y$

Correspondingly, area of ▱ formed by $\hat{j}$ and $\begin{bmatrix} x \\ y \end{bmatrix} = x$

In 3D, we get the same result: $z = \det\left(\begin{bmatrix} 1 & 0 & x \\ 0 & 1 & y \\ 0 & 0 & z \end{bmatrix}\right)$ } Signed volume of parallelapiped $= z$

keep $\hat{i}$ keep $\hat{j}$ use $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$

Recall that for a given matrix transformation, all areas are scaled by $\det(A)$.

$\therefore y = \dfrac{\text{Area of new parallelogram}}{\det(A)}$ ⇝ $\det\left(\begin{bmatrix} 3 & -4 \\ -1 & -2 \end{bmatrix}\right)$

new $\hat{i}$ output vector

Chp 13 – Change of Basis

A coordinate system requires a set of basis vectors, e.g. $\hat{\imath}$ and $\hat{\jmath}$.

- To translate some vector from NEW system → OLD system.
  - ⇒ Multiply by matrix containing new basis vectors
  - ⇒ Begin in say, $\begin{bmatrix}1\\2\end{bmatrix}$ in old system → linear transformation → what $\begin{bmatrix}1\\2\end{bmatrix}$ on NEW system looks like on our system.

- To go from OLD system → NEW system:
  - ⇒ Take the inverse matrix!

$\therefore A\begin{bmatrix}x_{new}\\y_{new}\end{bmatrix} = \begin{bmatrix}x_0\\y_0\end{bmatrix}$; $\begin{bmatrix}x_{new}\\y_{new}\end{bmatrix} = A^{-1}\begin{bmatrix}x_0\\y_0\end{bmatrix}$

To translate a matrix in a new system:

Given vector $\vec{v}$ and transformation matrix A,

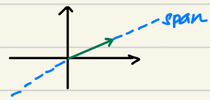
$$A^{-1} M A \vec{v}$$

- $\vec{v}$: take vector in new system
- A: ① transform to our system
- M: ② perform transformation in our system
- A^{-1}: ③ transform back to new system

$A^{-1} M A$ = transformation matrix in new system!

> $A M^{-1} A$ represents a kind of mathematical empathy, where.
> M is the transformation as WE see it,
> and A^{-1} and A converts it to someone else's perspective!

Chp 14 — Eigenvectors and eigenvalues



Consider some linear transformation $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$:

- Recall the span of a vector: everything it can reach via SCALING (i.e. infinite line connecting origin → vector)
- Most vectors will be knocked off their span by the transformation
- But some special vectors remain on their span!
  - ⇒ Effect of transformation is a simple scaling
  - ⇒ E.g. Any vector along the x-axis here. since $\hat{i}$ is transformed to $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$ — Same span!
  - (By linearity, ALL vectors along this span are linearly scaled by the same factor!
    - → factor of scaling is the eigenvalue

(*) Called the eigenvectors of the transformation!

One application is finding the axis of rotation of some rotation → this is simply the eigenvector, with eigenvalue 1.
(rotations don't stretch/squish)

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

transformation of $\vec{v}$ by A = scaling A by λ!

(We need to solve for $\vec{v}$ and λ, for given A) — identity: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

⇒ Note that $\lambda \vec{v} = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \vec{v} = (\lambda I)\vec{v}$

∴ $A\vec{v} = (\lambda I)\vec{v} \Rightarrow \underbrace{(A - \lambda I)\vec{v}} = \vec{0}$ (for non-zero $\vec{v}$)

For the matrix multiplication of $(A - \lambda I)$ by non-zero $\vec{v}$ to result in $\vec{0}$, $(A - \lambda I)$ must squish space into a lower dimension

∴ $\det(A - \lambda I) = 0$ ⇝ find λ to make $\det(A - I\lambda) = 0$

E.g. $A = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}$: $A - \lambda I = \begin{bmatrix} 2-\lambda & 2 \\ 1 & 3-\lambda \end{bmatrix}$

∴ $\det(A - \lambda I) = (2-\lambda)(3-\lambda) - (2)(1) = 0$

∴ $6 - 2\lambda - 3\lambda + \lambda^2 - 2 = 0$

∴ $\lambda^2 - 5\lambda + 4 = 0$

∴ $\lambda = 1$ OR $\lambda = 4$ ⇝ Eigenvalues: 1 and 4

E.g. $A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$:

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

$\therefore \det\left(\begin{bmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}\right) = 0$

$(3-\lambda)(2-\lambda) - (1)(0) = 0$

$\therefore \lambda = 2$ OR $\lambda = 3$ $\rightsquigarrow$ eigenvalues

Now consider $\lambda = 2$: $\begin{bmatrix} 3-2 & 1 \\ 0 & 2-2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \vec{0}$

$\left.\begin{aligned} (3-2)x + 1y &= 0 \\ 0x + 0y &= 0 \end{aligned}\right\}$ $1x + 1y = 0$

$\therefore y = -x$ (all vectors on line $y = -x$ are eigenvectors!) with corresponding eigenvalue $= 2$.

Consider rotation $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ w/ no eigenvectors:

$\det\left(\begin{bmatrix} 0-\lambda & -1 \\ 1 & 0-\lambda \end{bmatrix}\right) = 0$

$(0-\lambda)(0-\lambda) - (1)(-1) = 0$

$\lambda^2 + 1 = 0$ $\rightsquigarrow$ no real solutions for λ $\Rightarrow$ no eigenvectors!

NOTE: Each eigenvalue can be associated with ≥ 1 lines of eigenvectors, e.g. Matrix $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ only has one eigenvalue (2), but it scales everything by 2 $\rightarrow$ everything is an eigenvector!!

Eigenbasis: What if both basis vectors are eigenvectors?

↳ Diagonal matrix, e.g. $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ $\rightarrow$ all values other than diagonal $= 0$.

$\Rightarrow$ All basis vectors are eigenvectors, and each value = eigenvalue!

<u>Application</u>: If we have sufficient eigenvectors to span the whole space, we can use change of basis to make the eigenvectors our basis vectors!

E.g. To use eigenvectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ as basis: Use $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$

$\Rightarrow$ Apply same transformation, but from perspective of new system!

Benefit: Output matrix $A^{-1}MA$ is always <u>diagonal</u> $\Rightarrow$ basis vectors are scaled

↳ Eigenbasis = Basis where all vectors are eigenvectors.

e.g. to calculate $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}^{100}$, first change basis $\rightarrow$ compute $\rightarrow$ change back!

Chp 15 — Trick for computing eigenvalues.

For a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we usual find eigenvalues λ using.

$$A\vec{v} = (\lambda I)\vec{v}$$

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

$$\therefore \det(A - \lambda I) = 0 \rightsquigarrow \det\left(\begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix}\right) = 0$$

$$(a-\lambda)(d-\lambda) + (b)(c) = 0$$

But we also know.

Sum of eigenvalues: $\lambda_1 + \lambda_2 = tr\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + d \rightsquigarrow \text{mean} = \frac{\lambda_1 + \lambda_2}{2}$

Product of eigenvalues: $\lambda_1 \lambda_2 = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$

Eigenvalues $\lambda_1, \lambda_2 = \text{mean} \pm x$ (must be equidistant from mean!)

$$\left.\begin{aligned} \therefore (m+x)(m-x) &= p \\ m^2 - x^2 &= p \\ x^2 &= m^2 - p \\ x &= \pm\sqrt{m^2 - p} \end{aligned}\right\} \quad \lambda = m \pm \sqrt{m^2 - p}, \quad m = \frac{a+d}{2}, \quad p = ad - bc$$

Chp 1 — Deep Learnin notation

Weights: entry (i,j) = from ith neuron $\rightarrow$ jth neuron in next layer

Activations (neurons) of a layer

$$\sigma\left(\begin{bmatrix} w_{0,0} & w_{0,1} & w_{0,2} & \cdots & w_{0,m} \\ w_{1,0} & w_{1,1} & w_{1,2} & \cdots & w_{1,m} \\ \vdots & . & \vdots & \ddots & \\ w_{n,0} & w_{n,1} & w_{n,2} & & w_{n,m} \end{bmatrix}^T \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ a_2^{(0)} \\ \vdots \\ a_n^{(0)} \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix} \right)$$

$n \times m$ matrix — $1 \times n$ vector — $1 \times m$ bias

$a^{(1)} = \sigma(W^T \sigma^{(0)} + b)$ $\rightsquigarrow$ we often use <u>ReLU</u> now!