Linear Algebra

I'll keep it as brief as is possible. Pl. watch 3b1b if you need a refresher on intuition.
Discovery precedes Analysis.

Vectors, fundamentally, are abstract entities. We begin with a Vector space, characterised by scalar multiplication, vector addition, & the vector inner product. These 3 alone are fundamental; These obey a set of axioms

RULES FOR VECTOR SPACES

1. $|x\rangle + |y\rangle = |y\rangle + |x\rangle$ [Commutativity in +]
2. $|x\rangle + (|y\rangle + |z\rangle) = (|x\rangle + |y\rangle) + |z\rangle$ [Associativity in +]
3. $(ab)|x\rangle = a(b|x\rangle)$ [Associativity in x]
4. $a(|x\rangle + |y\rangle) = a|x\rangle + a|y\rangle$ [Distributivity]
5. $(a+b)|x\rangle = a|x\rangle + b|x\rangle$ [Distributivity]
6. $|0\rangle + |x\rangle = |x\rangle$ [Existence of Null vector]
7. $|x\rangle + (-|x\rangle) = |0\rangle$ [Existence of Additive inverse]
8. $1|x\rangle = |x\rangle$ [Existence of Unit Scalar]

RULES FOR INNER PRODUCTS

$\langle x|x\rangle \geq 0$

$\langle x|y\rangle = \overline{\langle y|x\rangle}$

$\langle \alpha x|y\rangle = \alpha\langle x|y\rangle$

$\langle x|\beta y\rangle = \beta^*\langle x|y\rangle$

$\langle x|(|a\rangle + |b\rangle) = \langle x|a\rangle + \langle x|b\rangle$

The above are for rigour, & should be obvious. Now, Let's build.

- A Vector $|z\rangle$ is said to be **Linearly Dependent** on $|x\rangle$ & $|y\rangle$ if $\exists\, \alpha, \beta \in \mathbb{C}$ s.t

$$\alpha|x\rangle + \beta|y\rangle = |z\rangle \quad \text{or,} \quad \exists\, \alpha, \beta, \gamma \in \mathbb{C} \text{ s.t } \alpha|x\rangle + \beta|y\rangle + \gamma|z\rangle = 0$$

- A set of vectors is called a **basis** if every vector in a Vector space V is Linearly dependent on this set of vectors. We say that is set "spans" V.

- If $\langle e_i|e_j\rangle = 0$, $|e_i\rangle$ cannot be expressed in terms of $|e_j\rangle$. They are orthogonal to each other. Thus by adding vectors orthogonal to each other, we can construct a basis. To rigorously show that there exists **an orthonormal basis**: 1. Take a basis $\{|v_1\rangle, |v_2\rangle \ldots\}$

2. $|e_1\rangle = \dfrac{|v_1\rangle}{\langle v_1|v_1\rangle}$ $\quad |w_2\rangle = |v_2\rangle - \langle e_1|v_2\rangle|e_1\rangle$ $\quad |w_i\rangle = |v_i\rangle - \sum_j \langle e_j|v_i\rangle|e_j\rangle$

$|e_2\rangle = \dfrac{|w_2\rangle}{\langle w_2|w_2\rangle}$ $\quad |e_i\rangle = \dfrac{|w_i\rangle}{\||w_i\rangle\|}$

This process is known as **Gram Schmidt Orthonormalisation**

- Now we posses a basis $\{|e_1\rangle, |e_2\rangle \ldots\}$ s.t $\langle e_i|e_j\rangle = \delta_{ij}$

Let any vector $|x\rangle \in V$ be expressed as $\sum_i a_i|e_i\rangle$

Now, $\langle e_j|x\rangle = \langle e_j|\sum a_i|e_i\rangle = \underline{\underline{a_j}}$

• The pivotal observation here is that $\langle e_i | v \rangle$ is the $\boxed{\text{component of } |v\rangle}$ along $|e_i\rangle$. Thus, $|v\rangle = \sum_i \langle e_i | v\rangle |e_i\rangle$

• Now, any abstract vector can be represented as a column of scalars.

$$|v\rangle = \begin{bmatrix} \langle e_1 | v\rangle \\ \langle e_2 | v \rangle \\ \vdots \end{bmatrix}$$

Further, <u>Linear operators</u> on abstract vectors canbe constructed as matrices. For operator A, in basis of $|e_i\rangle$, $A_{ij} = \langle e_i | A | e_j \rangle$

$$|v\rangle = \sum \langle e_i|v\rangle |e_i\rangle = \sum |e_i\rangle\langle e_i|v\rangle = \underbrace{\left(\sum |e_i\rangle\langle e_i|\right)}_{\boxed{\text{Identity operator.}}} |v\rangle$$

At this point, we must define the outer product & matrix vector multiplicati

where $\hat{j}$ lands

$$\begin{bmatrix} | & | & | & | & | \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ \vdots \end{bmatrix} \begin{matrix} \leftarrow \hat{i} \text{ component} \\ \leftarrow \hat{j} \text{ component} \end{matrix} \qquad A|b\rangle = \sum_i \begin{bmatrix} A_{1i} \\ A_{2i} \\ \vdots \end{bmatrix} b_i$$

where $\hat{i}$ lands

I will now sacrifice generality & define $\langle e_i | =$ conjugate transpose of $|e_i\rangle$

$$|e_i\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \leftarrow \text{at } i^{th} \text{ component} \qquad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} \leftarrow \text{only } \hat{i} = \begin{bmatrix} 0 & - & \\ | & 0 & - \\ | & | & 1 \end{bmatrix}$$

where $\hat{i}$ lands

1 is at (4,4)

Thus $\sum |e_i\rangle\langle e_i| = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \ddots \end{bmatrix}$

I've established general rules of Linear Algebra. Now, I will delve into the comfortable domain of $v \in \mathbb{R}^n$

Any vector $\vec{v}$ can be seen as an arrow in Space. $M\vec{v}$ is as defined earlier.

$\underset{m \times n}{M} \underset{n \times l}{N} = \underset{m \times l}{L}$ Take each column of N & transform it with M. You get L

Now, each Linear Transfomation represented by M, has $\boxed{\text{4 fundamental Subspaces.}}$ Firstly, A '$\boxed{\text{Vector Subspace}}$' refers to a set of vectors $S \subseteq V$ such that, if $\vec{v}_1, \vec{v}_2 \in S$, then $\alpha\vec{v}_1 + \beta\vec{v}_2 \in S \ \forall \alpha, \beta \in \mathbb{R}$

Since $\forall \vec{v} \in S$, $\alpha\vec{v} \in S \Rightarrow -\vec{v} \in S \Rightarrow \vec{v} + (-\vec{v}) = \vec{0} \in S \Rightarrow$ All subspaces must contain the origin.

Now, the first of the 4 great subspaces is the $\boxed{\text{Column space}}$, the space Spanned by all the Columns of the matrix.

Since any vector transformed by M becomes a linear combination of its columns, the Column space represents the 'Output Space' post the transformation.

The Dimension of the column space is known as $\boxed{\text{Rank}}$.

$\{\alpha \vec{v}_1 + \beta \vec{v}_2 + \dots \mid \alpha, \beta \dots \in \mathbb{R}\}$ is the column space, & it is quite obviously a valid vector subspace.

Now, 2 questions arise: 1. given a vector $\vec{v}$, can you tell if it belongs to the Column space? 2. How to find the Rank?

For 1. If $v \in Col(A)$ then v is linearly dependent on the columns of A.

$\Rightarrow \sum \alpha_i \vec{a}_i = \vec{v}$ for some choice of α_is.

Rephrasing, $\sum \alpha_i \vec{a}_i + \beta \vec{v} = 0$ has a solution which is non zero set.

Further rephrasing, the Augmented Matrix $[A|\vec{v}]$ satisfies $[A|\vec{v}]\vec{x} = \vec{0}$ For $\vec{x} \neq \vec{0}$.

For 2. The Rank is simply the number of independent vectors in the columns of A. $\Rightarrow$ 1. Solves this issue.

$\Rightarrow$ Both 1 & 2 simplify to $\boxed{\text{solving } A\vec{x} = \vec{b}}$

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 12 \\ 2 \end{bmatrix}$$

$$\begin{aligned} x + 2y + z &= 2 \\ 3x + 8y + z &= 12 \\ 4y + z &= 2 \end{aligned}$$

$(2) - 3 \times (1)$

$$\Rightarrow \left[\begin{array}{ccc|c} \boxed{1} & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 4 & 1 & 2 \end{array}\right] \xrightarrow{(3) - 2\times(2)} \left[\begin{array}{ccc|c} \boxed{1} & 2 & 1 & 2 \\ 0 & \boxed{2} & -2 & 6 \\ 0 & 0 & \boxed{5} & -10 \end{array}\right]$$

Now, we have z & Can back substitute. It is solved.

This Upper Triangular Form is called the Row echelon form.

Back substitution is too much work so we can take it further, to the $\boxed{\text{Reduced Row echelon form.}}$

$$\left[\begin{array}{ccc|c} \boxed{1} & 2 & 1 & 2 \\ 0 & \boxed{1} & -1 & 3 \\ 0 & 0 & \boxed{1} & -2 \end{array}\right] \xrightarrow[(1)-(3)]{(2)+(3)} \left[\begin{array}{ccc|c} \boxed{1} & 2 & 0 & 4 \\ 0 & \boxed{1} & 0 & 1 \\ 0 & 0 & \boxed{1} & -2 \end{array}\right] \xrightarrow{(1)-(2)\cdot 2} \left[\begin{array}{ccc|c} \boxed{1} & 0 & 0 & 2 \\ 0 & \boxed{1} & 0 & 1 \\ 0 & 0 & \boxed{1} & -2 \end{array}\right]$$

we are now able read off the solution $x = 2, y = 1 \ \& \ z = -2$

Now, we must generalise this procedure.

1. Find the Left most non zero Column & bring $\boxed{1}$ to the top. Swap rows if needed. Make all other entries zero.
2. Pretend that row containing your $\boxed{1}$ does not exist. Repeat 1, 2 till no rows are left.
3. Destroy columns upward to reach Reduced REF.

Now, we know this procedure, known as Gaussian Elimination, to solve any set of equations. However, we will NOT get one unique solution always.

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

- A is $n \times n$
  - $\to |A| = 0$
    - $\to \vec{b} \in col(A) \Rightarrow \infty$ solns (1.1.1)
    - $\to \vec{b} \notin col(A) \Rightarrow$ No solns (1.1.2)
  - $\to |A| \neq 0 \longrightarrow A^{-1}\vec{b}$ is the soln (1.2)

- A is $m \times n$, $m > n$, more eqns than unknowns, overspecified
  - $\to Rank(A) = n$
    - $\to b \in col(A) \Rightarrow$ one soln (2.1.1)
    - $\to b \notin col(A) \Rightarrow$ No solns (2.1.2)
  - $\to Rank(A) < n$
    - $\to b \in col(A) \Rightarrow \infty$ solns (2.2.1)
    - $\to b \notin col(A) \Rightarrow$ No solns (2.2.2)

- A is $m \times n$, $m < n$, underspecified
  - $\to b \in col(A) \to \infty$ solns (3.1)
  - $\to b \notin col(A) \to$ No solns (3.2)

This much is intuitive. That intuition needs to be applied on RREF.
Let me present a radically new question: Identify among the given vectors, those which are $\boxed{\text{linearly dependent}}$ on $\begin{bmatrix}-1\\0\\1\end{bmatrix}$ & $\begin{bmatrix}2\\1\\-1\end{bmatrix}$: $\begin{bmatrix}1\\1\\0\end{bmatrix}\begin{bmatrix}5\\2\\-3\end{bmatrix}\begin{bmatrix}8\\5\\4\end{bmatrix}\begin{bmatrix}-3\\0\\3\end{bmatrix}\begin{bmatrix}-1\\1\\1\end{bmatrix}\begin{bmatrix}-2\\3\\3\end{bmatrix}$

Now:

$$\left[\begin{array}{cccccccc|c} -1 & 2 & 1 & 5 & 8 & -3 & -1 & -2 & 0 \\ 0 & 1 & 1 & 2 & 5 & 0 & 1 & 3 & 0 \\ 1 & -1 & 0 & -3 & 4 & 3 & 1 & 3 & 0 \end{array}\right]$$

$$\underset{3\times 8}{A}\ \vec{x} = \vec{0}$$

$$\left[\begin{array}{cccccccc|c} \boxed{1} & -2 & -1 & -5 & -8 & 3 & 1 & 2 & 0 \\ 0 & 1 & 1 & 2 & 5 & 0 & 1 & 3 & 0 \\ 0 & 1 & 1 & 2 & 12 & 0 & 0 & 1 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccccccc|c} \boxed{1} & -2 & -1 & -5 & -8 & 3 & 1 & 2 & 0 \\ 0 & \boxed{1} & 1 & 2 & 5 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & \boxed{7} & 0 & -1 & -2 & 0 \end{array}\right]$$

This is a little hard to interpret right off the bat, so let me build.
Imagine we just had $\begin{matrix}-1 & 2 & 8\\0 & 1 & 5\\1 & -1 & 4\end{matrix} \to \left[\begin{array}{ccc|c}\boxed{1} & -2 & -8 & 0\\0 & \boxed{1} & 5 & 0\\0 & 0 & \boxed{7} & 0\end{array}\right]$ if we put it in RREF, we will find $x=y=z=0$ is the only solution. $\begin{bmatrix}8\\5\\4\end{bmatrix}$ is thus independent of the other 2 vectors.

Similarly, the other 2 with non zero entries on the 3rd row are independent.

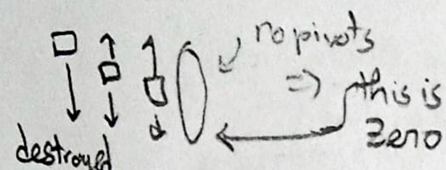
Further, $\begin{bmatrix}1\\1\\0\end{bmatrix}\begin{bmatrix}5\\2\\-3\end{bmatrix}$ & $\begin{bmatrix}-3\\0\\3\end{bmatrix}$ are linearly dependent on the given 2 vectors.

This is because, in $\left[\begin{array}{ccc|c}\boxed{1} & -2 & -1 & 0\\0 & \boxed{1} & 1 & 0\\0 & 0 & 0 & 0\end{array}\right] \to$ Put in RREF $\Rightarrow \begin{matrix}x = f(z)\\ y = g(z)\\ 0 = 0\end{matrix} \Rightarrow \begin{bmatrix}f(z)\\g(z)\\z\end{bmatrix}$ is a soln.

Consider general $A\vec{x}=\vec{0}$ A is $m \times n$. The matrix is put in RREF.
Note that there are 'n' variables.

- If there are $\leq n$ pivots, each pivoted variable can be written as a function of non pivoted 'free' variables. <u>The number of non pivoted columns is the dimension of the null space.</u>

- <u>If a column is unpivoted, it is necessarily linearly dependent on the pivoted columns.</u> This is fairly intuitive. Why is it true? Consider the set of pivoted columns + one unpivoted column. Clearly, there exists a non trivial null space, as the unpivoted column <u>causes an unpivoted row.</u>



Since the null space is non trivial, there exists linear dependence.

Now, in the set of pivoted columns, there are no unpivoted variables, thus, there is no null space & the set is independent. QED.

- Now, $\boxed{\text{no of pivoted columns} = \text{no of independent vectors} = \text{Rank of } A,}$ the dimension of $Col(A)$

no of non pivoted columns = **Nullity**, the dimension of $NS(A)$

Since no of pivoted columns + no of non pivoted columns = total no of colums,

We have:

$$\Rightarrow \boxed{\text{Rank} + \text{Nullity} = n, \text{ the dimension of the Input Space}}$$

'The Rank Nullity Theorem'

Having made these 3 profound observations, in the question that sparked this discussion, We can directly say that the columns eligible for a 3rd pivot are the independent ones. In their absence, we would have a fully dependent set.

So far, we have established that

- By studying the RREF of $A\vec{x}=\vec{0}$, we can find
  1. Independent vectors
  2. Rank
  3. Nullity.

  1, 2 ⇒ We know $Col(A)$
  It is the Span of pivoted vectors.

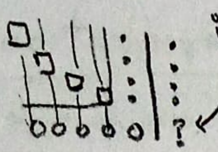
The Null space of A is given the vector $\begin{bmatrix} f(\text{non pivoted}) \\ \vdots \\ \text{non pivot} \end{bmatrix}$ eg: $\begin{bmatrix} f(w,u) \\ f(w,u) \\ w \\ u \end{bmatrix}$

The eg has a 2D null space.

The $Col(A)$ came with a basis for free. To find a $\boxed{\text{basis for the null space}}$ we must do a little more work. Just put all variables except one as 0. We will get $\begin{bmatrix} \text{stuff} \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} \text{stuff} \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \dots$ & those are clearly = no of free variales & linearly independent ⇒ Valid Basis.

We have seen all there is to see about $A\vec{x} = \vec{0}$. But we are yet to tackle the more general $\boxed{A\vec{x} = \vec{b}}$, for which I listed some cases. With our newfound familiarity for RREF let's solve this problem.

<u>Case 1.1</u>: $|A| = 0 \Rightarrow$ There is a pivotless column. Since it is square, there is a pivotless row.

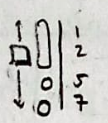


if this value is non zero, there are <u>no</u> solutions. (1.1.2)

if it <u>is</u> zero, Take the free variables to the RHS & have ∞ solns (1.1.1)

<u>Case 1.2</u>: Simple; Basic case.

<u>Case 2.1</u>: All columns are pivoted. There are some null rows at the bottom. (Since pivots destroy the rest of the column) If $0 =$ non zero exists, no soln. (2.1.2) If all are $0 = 0$, Unique soln is taken with the n pivots.

<u>Case 2.2</u>: There are pivotless columns.  As usual, the free variables go to the RHS.

Note that in the columns to the right of the right most pivot, below the last pivot, we must necessarily have zeroes. Else, we get a new pivot.

(2.2.1) ∞ solns if $0 = 0$ rows exist. (2.2.2) $\underset{x \neq 0}{0 = x}$, no solns.

<u>Case 3</u>: If there are m pivots, all of the output space is mappable. Take free variables to RHS. ∞ solns. (3.1) If there are $< m$ pivots, there are zero rows. If $0 = 0$, ∞ solns (3.1) if $0 \neq 0$, (3.2) no solns.

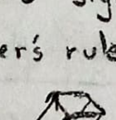
It's super intuitive so, practically we can solve any case with RREF without worrying about which case we are in. The only things to remember are to check the zero rows for no solns, & to take free variables to RHS for ∞ solns.

That concludes 'solving $A\vec{x} = \vec{b}$'.

Almost. There's this beautiful rule to solve this in the case of square matrices:

$\boxed{\text{Cramer's Rule.}}$ $A = [x_1, x_2 \dots]$ $A\vec{y} = \vec{b}$ Then, $y_i = \dfrac{\det(x_1, \dots \vec{b} \dots)}{\det(A)}$ ← at column i

The way to understand this 3b1b style is pretty, but not necessarily helpful on the fly. Cramer's rule, however, Is super helpful, so don't ignore it.

The idea is this: $\det(\hat{\imath}\ \hat{\jmath}\ \vec{y}) = y_3$  & $\hat{\imath}\ \hat{\jmath}\ \vec{y}$ transforms into $\vec{x}_1\ \vec{x}_2\ \vec{b}$.

Since all volumes scale by $\det(A)$, vol before $\times \det(A) =$ vol after

$\Rightarrow y_3 \cdot \det(A) = \det(x_1\ x_2\ y)$ Generalising this, we have

$$\boxed{y_i = \frac{\det(x_1, x_2 \dots \vec{x}_{i-1}\ \vec{b}\ \vec{x}_{i+1} \dots)}{\det(A)}}$$

We are simply watching a volume transform to get coordinates.

We've covered substantial ground. Most of the fundamental stuff has been well established. Before I go back to the 4 fundamental subspaces, some minor remarks: (only about square matrices)

- When RREF is established, we do 'row operations.'

Swapping, adding & multiplying rows are all representable by left multiplication by a matrix. This is one of the times the row way of looking at matrix multiplication shines.

$\begin{bmatrix} 1 & \cdot & \cdot \\ 0 & \cdot & \cdot \\ 0 & \cdot & \cdot \end{bmatrix}\begin{bmatrix} \equiv \end{bmatrix}$ The first row is just $\vec{r}$. each value is scaled by $\begin{smallmatrix}1\\0\\0\end{smallmatrix}$, resulting in a matrix. Next you do this to row 2 & so on, & add all the matrices.

$\begin{matrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{matrix}$ means row1 → row1 + row2, row2 → row2 + row3

Given a Matrix M, we essentially doing $[\cdots R_4 R_3 R_2 R_1]M$

For square matrices, $RM = I$, so, R is M^{-1}. If we do all these row operations on I, as we do it on M, we will $\boxed{\text{get } M^{-1}}$ directly. That is, $[M|I]$ in RREF is $[I|M^{-1}]$

- Another popular way to get A^{-1} is by finding the matrix adjoint.(A) which is computed as follows:

$M_{ij} = \det(A \text{ without row } i \text{ or column } j)\cdot(-1)^{i+j}$ $\quad M^T = \mathrm{adj}(A)$

$\boxed{A^{-1} = \dfrac{\mathrm{adj}(A)}{|A|}}$ I cannot be bothered to prove this. We have a better, more logical way already.

- For square matrices of rank $< n-1$, Since 2 columns are dependent on others, whatever may be deleted, M_{ij} has a dependent column, so $\mathrm{adj}(A) = 0$ matrix.

For square matrices of rank $= n-1$, the [null space of the adjoint = column space of A]*, so if $\boxed{\mathrm{adj}(A)\,\vec{b} \neq 0}$, $A\vec{x} = \vec{b}$ has no solns for sure. If $\mathrm{adj}(A)\,\vec{b} = 0$, $\mathrm{adj}(A) \neq 0$, then $\vec{b}$ has infinite solns.

* This is provable. I put out a stackxchng post, but ended up answering it myself. My proofs a little convoluted & not worth getting into.

- The Following are special types of square matrices:

$A = A^T$ Symmetric	$A = -A^T$ Skew symmetric ← Diagonal is zeroes
$A = A^*$ Hermition	$A = -A^*$ Skew Hermition ← Diagonal is Imaginary/zero.

Diagonal is Real.

any $M = M + \frac{M^T}{2} - \frac{M^T}{2} = \underbrace{\frac{1}{2}(M + M^T)}_{\text{Symmetric}} + \underbrace{\frac{1}{2}(M - M^T)}_{\text{Skew symmetric}}$ Similar thing for Hermition.

$\boxed{A^TA = I \text{ Orthogonal} \qquad A^*A = I \text{ Unitary} \qquad A^*A = AA^* \text{ Normal}}$

$\langle Ax, y\rangle = \langle x, A^*y\rangle$ because $y^*Ax = (A^*y)^*x$

Now, $\langle Ax, Ay\rangle = \langle x, A^*Ay\rangle$ ⇒ Unitary & real Orthogonal matrices preseves inner product, lengths of all vectors.

PS: Just Realised QM notation & Mathematician's notation are flipped. To be fair, QM had y^* pre done.

$A^*A = I$

a_{ij} in A^*A is column j in A times row i in A^*.

$\Rightarrow \quad a_{ij} = a_i^* a_j \quad$ where a_i is the i^{th} column of A

$\Rightarrow \langle a_j, a_i \rangle = \delta_{ij} \Rightarrow$ All the columns of a unitary matrix are orthogonal to each other & are of unit norm

Since $I^* = I$,

$A^*A = I \Rightarrow AA^* = I = A^*A \Rightarrow$ Unitary matrices are Normal

Hermitan matrices have only real eigenvalues. Eigenvectors with distinct eigenvalues are othogonal to each other

$H v_i = \lambda_i v_i \qquad v_j^* H v_i = \lambda_i v_j^* v_i$

$\qquad v_i^* H^* v_j = \lambda_i^* v_i^* v_j$

$\qquad = v_i^* \lambda_j v_j \qquad \Rightarrow \quad v_i^* v_j (\lambda_j - \lambda_i^*) = 0$

when $i = j \quad v_i^* v_i \neq 0 \Rightarrow \lambda_i = \lambda_i^*$

$i \neq j \quad v_i^* v_j = 0 \quad$ as $\lambda_i \neq \lambda_j$

Skew Hermitian matrices have purely imaginary or zero eigenvalues. Eigenvectors of distinct eigenvalue are orthogonal

$v_j^* H v_i = \lambda_i v_j^* v_i$

$-v_i^* H v_j = \lambda_i^* v_i^* v_j \quad \Rightarrow \quad (\lambda_i^* + \lambda_j) v_i^* v_j = 0$

Unitary matrices have eigenvalues of unit norm. Eigenvectors of distinct eigenvalue are orthogonal

$U^* U = I \qquad U v_i = \lambda_i v_i$ Take norm. $\|Uv_i\| = \|v_i\|$ as Unitary matrices preserve norm.

$\Rightarrow \|v_i\| = |\lambda| \|v_i\| \Rightarrow |\lambda| = 1$

$U v_j = \lambda_j v_j \quad \langle U v_i, U v_j \rangle = v_j^* v_i = \lambda_j^* v_j^* \lambda_i v_i \Rightarrow v_j^* v_i = 0$ or $\lambda_j^* \lambda_i = 1$ ← NOT possible.

$\lambda_j \lambda_j^* \lambda_i = \lambda_j \Rightarrow \lambda_i = \lambda_j$

Unitary matrices have a determinant of unit norm.

This is obvious since all column vectors are of unit norm. However, it's also a consequence of the previous result, since $\boxed{\det(A) = \prod \lambda_i}$

$$Av = \lambda v \Rightarrow (A - \lambda I)v = 0$$

"characteristic equation" $\rightarrow \det(A - \lambda I) = 0$

$$\begin{vmatrix} a-\lambda & c \\ b & d-\lambda \end{vmatrix} = ad - \lambda(a+d) + \lambda^2 - bc \Rightarrow \boxed{\lambda^2 - \lambda \operatorname{Tr}(A) + \det(A) = 0}$$

The solutions to the characteristic equation are the n eigenvalues.

$$\det(A - \lambda I) = (-1)^k (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n) \quad \text{(Factorised)}$$

Any power of (-1) will yield same zeroes. But this is the coefficient of λ^n. By simply expanding the determinant, we know that $k = n$.

$$\Rightarrow \det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$$

Put $\lambda = 0$ $\boxed{\det(A) = \prod \lambda_i}$

Now that the topic of the characteristic equation has come up, I'd like to note that every matrix satisfies its own characteristic polynomial, the $\boxed{\text{Cayley Hamilton Theorem}}$

Eg: $A^2 - A\,tr(A) + \det(A) I = 0$ matrix (if A is 2×2)

Also note: $Tr(A+B) = Tr(A) + Tr(B)$ $\quad \det(AB) = \det(A)\det(B)$

$\boxed{Tr(AB) = Tr(BA)}$

$$\sum (AB)_{ii} = \sum_i \sum_j A_{ij} B_{ji} \qquad \sum (BA)_{ii} = \sum_i \sum_j B_{ij} A_{ji}$$

Just swap the names 'i' & 'j'. The order of summation doesn't matter. The 2 are the same.

Also, this means Trace is invariant with change of basis. $Tr(P^{-1}\underline{AP})$
$= Tr(APP^{-1}) = \underline{\underline{Tr(A)}}$

At this point, I must acknowledge that this detour is no longer 'minor'
Regardless, we will finish this world tour before turning back.

SUMMARY

Hermitian → Diagonal is Real → Eigenvalues are real → Eigenvectors are orthogonal (if $\lambda_i \neq \lambda_j$)
→ Always Diagonalisable with a Unitary Matrix (Spectral Theorem)

Unitary → All columns & all rows are of unit norm → All eigenvalues are of unit norm ($e^{i\phi}$)
→ Lengths & Inner products are preserved under the transformation
→ Has determinant of unit norm. ($e^{i\phi}$)

Among these, Only the Spectral theorem has not been proven yet.

Schur's Theorem: Any square matrix is similar to an upper triangular matrix. $A = U^* T U$
(Two matrices A & B are called similar if $\exists P$ s.t $B = P^{-1}AP$)
(Two matrices are called similar if ∃ P s.t ... only in different bases)

(Essentially, they're the same transformation, only in different bases)

This is a cool result that comes along with a procedure. Consider a 3×3 matrix A. Let $P(\lambda)$ be its characteristic polynomial. $P(\lambda)$ will have 3 roots $\in \mathbb{C}$. Pick any one of these. Let v_1 be the eigenvector associated with the eigenvalue. ($\|v_1\| = 1$)

Pick any 2 vectors in the space linearly independent to v_1, & use Gram Schmidt Orthonormalisation to form a basis.

$$U = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$$ This matrix is certainly unitary; for one, it's columns are orthogonal & of unit norm.

Secondly, $$U^*U = \begin{bmatrix} -\bar{v}_1^T- \\ -\bar{v}_2^T- \\ -\bar{v}_3^T- \end{bmatrix} \begin{bmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ precisely because $M_{ij} = \bar{v}_i^T v_j$

Now, $$AU = A\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} | & | & | \\ \lambda v_1 & Av_2 & Av_3 \\ | & | & | \end{bmatrix}$$

and $$U^*AU = \begin{bmatrix} -\bar{v}_1^T- \\ -v_2^T- \\ -v_3^T- \end{bmatrix} \begin{bmatrix} | & | & | \\ \lambda v_1 & Av_2 & Av_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} \lambda & ? & ? \\ 0 & \multicolumn{2}{c}{\boxed{B}} \\ 0 & & \end{bmatrix}$$

Now Find a matrix $P_{2\times 2}$ to achieve the same effect on B.

$$P^*BP = \begin{bmatrix} \lambda_2 & ? \\ 0 & \lambda_3 \end{bmatrix}$$ Let $U_2 = \begin{bmatrix} 1 & 0\ 0 \\ 0 & \boxed{P} \\ 0 & \end{bmatrix}$ ← Unitary coz columns ⊥ & unit norm

$$[U^*AU]U_2 = \begin{bmatrix} \lambda & ? & ? \\ 0 & BP & \\ 0 & & \end{bmatrix}$$ & $$U_2^*U^*AUU_2 = \begin{bmatrix} \lambda & ? & ? \\ 0 & \lambda_2 & ? \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

This procedure can be extended to any $n\times n$ matrix. Thus, any square matrix is upper triangularisable.

Beyond this, it's very easy to see why the spectral theorem is true:

Let A be hermitian. $(A = A^*)$

$$U^*AU = T \,;\quad T^* = U^*A^*U = U^*AU = T$$

T^* is lower triangular.

If $T = T^*$, then T is purely diagonal.

$\Rightarrow$ Any Hermitian matrix is Unitarily Diagonalisable.

Resuming Discussion on the 4 spaces. The 2 spaces we have come to be intimately familiar with are the column space & the null space

Consider:
$$A = \begin{bmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \end{bmatrix}$$ The column space is $\text{span}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

The null space is $x - y + 2z = 0 \Rightarrow \begin{bmatrix} y-2z \\ y \\ z \end{bmatrix} \Rightarrow \text{Span}\left(\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right)$

(Normal to plane is $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$) as eqn is $[x\ y\ z]\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = 0$

Now consider $A^T = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & -2 \end{bmatrix}$ Column space is span$\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$

Null space is $x - y = 0 \Rightarrow$ span$\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$

By some miracle, the column space of A^T is orthogonal to the Null space of A, and the null space of A^T is orthogonal to the Column space of A.

This is, in fact, always true. An $m \times n$ matrix has an n D input space & an m D output space. The rank of the matrix lives in the output space, & has dimension r (say). The null space of A lives in the input space & has dim $n-r$.

The remaining r dimensional space orthogonal to $N(A)$ is $Col(A^T)$. The $m-r$ D space $\perp$ to $Col(A)$ is $N(A^T)$.

The dimensions check out, but <u>why</u> is this true?

if $A^T x = 0$, $\quad x^T A = 0$

Given any vector $b = Av$, $\quad x^T b = x^T A v = 0$

$\Rightarrow$ Every vector in the $\boxed{\text{Left Null space}}$ is orthogonal to every vector in $Col(A)$

Similarly, $Ax = 0$

$b = A^T v \qquad b^T x = v^T A x = 0$

$\Rightarrow$ Every vector in the $\boxed{\text{Row Space}}$ is orthogonal to every vector in $N(A)$

In both these cases, when you see it as $x^T A$, you can see that rows are being linearly combined to get the resultant row vector. Because the row vector needs to be transposed, the A^T business exists, otherwise you can say

Left null space $\perp$ Right 'column' space

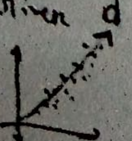
Left 'row' space $\perp$ Right Null space.

Algebraically, the connection is clear, but it's lacking geometrically & intuitively.

No, I can't think of anything. Let's wrap this up.

Typically mathematicians don't use non mathematical problems as motivation, they just go crazy & explore things for fun. I'll do something unconventional & talk about the problem of Regression.

Given d data points, we wish to fit a curve to it, let's say, a line.



To achieve this we must minimise $\frac{1}{d}\sum_{i=1}^{d}(y_i - f(x_i))^2 = J$ 'cost fn'

If we differentiate & set to 0, $\frac{\partial J}{\partial \theta_i} = \frac{2}{d}\sum (f(x_i) - y_i)\frac{\partial f(x_i)}{\partial \theta_i}$

Typically we would run gradient descent on this, but this is an exactly solvable problem. Consider a simple example:

$(2, y_1)$ $\quad f(x) = mx$

$(3, y_2)$ $\quad \frac{\partial J}{\partial m} = \left[(m\cdot 2 - y_1)2 + (m\cdot 3 - y_2)\cdot 3 + (m\cdot 4 - y_3)\cdot 4\right]\frac{2}{3} = 0$

$(4, y_3)$

$$m(2^2 + 3^2 + 4^2) = 2y_1 + 3y_2 + 4y_3$$

$$m = \frac{2y_1 + 3y_2 + 4y_3}{2^2 + 3^2 + 4^2} = \frac{x^T y}{x^T x} \quad \text{where } x = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \ \& \ y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Note that this observation holds purely because we use a Linear function through the origin.

$$\frac{\partial J}{\partial \theta_i} = \frac{2}{n}\sum (f(x_i) - y_i)\frac{\partial f(x)}{\partial \theta_i}\Big|_{x = x_i}, \quad \text{if } f(x) = \theta_1 x \text{, then,}$$

$$\frac{\partial J}{\partial \theta_1} = \frac{2}{n}\sum(\theta_1 x_i - y_i)x_i = 0 \quad \sum \theta_1 x_i^2 - x_i y_i = 0 \quad \boxed{\theta_1 = \frac{\sum x_i y_i}{\sum x_i^2}}$$

$\boxed{\theta_1 = \frac{x^T y}{x^T x}}$ That is, θ_1 = projection of y along the unit vector along x.

Now, what if we had multiple inputs? Area of house, no of bedrooms etc predicting y = House price

$$d\left[\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots \\ & & & \end{bmatrix}\right]_{n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix}_{d\times 1}$$

$d \times n$

$$f(x) = \theta_1 x^1 + \theta_2 x^2 + \dots \theta_n x^n$$

$$\frac{\partial J}{\partial \theta_k} = \sum_i^d \left(\sum_j^n \theta_j x_{ij} - y_i\right) x_{ik} = 0$$

Let x^i denote the ith column & x_i the ith row

$$\Rightarrow \sum_i (\theta^T x_i)\, x_{ik} = \sum_i y_i x_{ik}$$

$$\begin{bmatrix} \theta^T x_1 \\ \theta^T x_2 \\ \vdots \end{bmatrix}^T_{1\times d} x^k_{d\times 1} = (x^k)^T y$$

$$(\theta^T_{1\times n} X^T_{n\times d})\, x^k_{d\times 1} = (x^k)^T_{1\times d}\, y_{d\times 1}$$

Since this holds for every k, We can say $\theta^T_{1\times n} X^T_{n\times d} \begin{bmatrix} x^1 & x^2 & x^3 & \cdots \end{bmatrix}_{d\times n} \stackrel{?}{=} \begin{bmatrix} x_{11} & \cdots \\ x_{12} & x_2 \\ \vdots & \end{bmatrix}_{n\times d} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_{d\times 1}$

So, $X^T X\theta = X^T y$ $\qquad [\theta^T X^T x^1 \ \ \theta^T X^T x^2 \dots] \stackrel{?}{=} \begin{bmatrix} (x^1)^T y \\ (x^2)^T y \\ \vdots \end{bmatrix}$

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

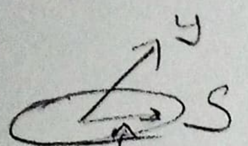
That one last compilation step was exhausting. I tried to extend both the $d\times 1$ x^ks into $d\times n$ X matrices & hoped the columns would individually form results, which they do, only, one was flipped & needed a transpose.

How is this related to Linear Algebra in any way? Well since in 1D, we saw that y was projected onto x, it's entirely possible there's a greater reason for it. We are looking for a scalar θ to scale a vector x, so that it most resembles y. It's natural that this scalar is y^Tx/x^Tx.

Similarly we want a scalar f each x^i to resemble y.

$$\Rightarrow x^{1T}x^1\,\theta_1 = x_1^Ty \Rightarrow X^TX\theta = X^Ty$$

Magic, isn't it? We are just projecting y onto each column of matrix X & using it as corresponding scalars. This is "Projecting onto a Subspace"



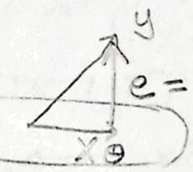
projection of y on S $= X[(X^TX)^{-1}X^Ty]$ note that given any y, the

Projection matrix $\boxed{\mathbb{P} = X(X^TX)^{-1}X^T}$ will grant the projection of y on X.

We have arrived at $\mathbb{P}$ in 2 ways already, firstly by least squares minimisation & secondly by getting scalars for each column of X. Here's a third way:

The projection of y on col(X) is Some linear combination of the columns of X.

$\Rightarrow X\theta$ = projection of y on col(X)



We want θ. Clearly, e is orthogonal to the subspace col(X) [Otherwise the projection can be made better by making it orthogonal]

Thus $e \in$ Left Null space that is, $X^Te = 0$

$$\Rightarrow X^T(y - X\theta) = 0 \qquad \boxed{X^Ty = X^TX\theta}$$

Boom. QED.

All this reinforces what we know, but I like to think of it as $\theta = x^Ty/x^Tx$ being extended. It's so intuitive & elegant.

(Oh by the way, you can totally skip gradient descent & literally finish your linear regression code with one line: $\theta = (X^TX)^{-1}X^Ty$, but the computation of $(\underset{n\times d}{X^T}\underset{d\times n}{X})^{-1}$ is infeasibly long when n is large: $O(n^3)$. Typically, this so called "Normal Equation method" is used if you have less than 10^5 features in your data.)

(Oh, by the way X^TX is guaranteed to be an invertible matrix provided the colums of X are linearly independent, & is full rank. Why? If $Xv=0$, then $X^TXv = 0 \Rightarrow$ it's uninvertible $\Rightarrow$ X is full rank. Further this means that if $d < n$, we are in trouble, because there can only be d independent columns in a d D space. $\Rightarrow$ if you have more data than features & you don't have redundant colums, you're good.)

(Oh, the proof seems to be incomplete, I only showed that if not full rank, not invertible, but the original statement was that if it's full rank it's GUARANTEED to be invertible. Much stronger. Why? X is one to one, as it's full rank. So is X^T. So is X^TX. X^TX is square & 1-1. It's invertible.)

(One more side note: $\mathbb{P}^T = \mathbb{P}$, $\mathbb{P}^2 = \mathbb{P}$. Further, given M, if $M^2 = M$ & $M^T = M$, then $M = \mathbb{P}$)

Actually there are several more remarks I wish to make on regression, but I'll save it for the time we do a very ML handling of the Topic.

Where next? Well, there is one loose thread. That's the connection between eigenvalues & diagonalisation. Last time I didn't really talk much about it.

[Side note: To determine v : $Av = \lambda v$, note that $(A - \lambda I)v = 0 \Rightarrow v \in$ Null space $(A - \lambda I)$

Solve for λ using $\det(A - \lambda I) = 0$, then find a basis for $N(A - \lambda_i I)$. That's all.]

- Any $n \times n$ matrix with n independent eigenvectors is diagonalisable.

Consider $S = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$ (eigenvector matrix)

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

$$\Rightarrow AS = S\Delta \qquad \boxed{A = S \Delta S^{-1}}$$

- Given a set of eigenvectors $\{v_1, v_2 \cdots v_n\}$ with $\lambda_1 \neq \lambda_2 \neq \cdots \neq \lambda_n$, then the set is linearly independent.

Really, we just need to show if v_1 has λ_1 & v_2, λ_2 then v_1 & v_2 cannot be linearly dependent. Since that hols for every pair, the property is obvious. To be rigorous,

if $c_1 v_1 + c_2 v_2 = 0 \longrightarrow \lambda_2 c_1 v_1 + c_2 \lambda_2 v_2 = 0$ -①

$A(c_1 v_1 + c_2 v_2) = 0$

$c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 = 0$ -② $\Rightarrow$ ②-① $c_1(\lambda_1 - \lambda_2) v_1 = 0$

must be 0. $\neq 0$ $\quad \neq 0$

$c_1 = 0$, $c_2 v_2 = 0$ $\quad v_2 \neq 0$

$\Rightarrow c_2 = 0$

$\Rightarrow$ Not linearly dependent.

- Given a square matrix is diagonalisable, it must have n independent eigenvectors.

Given: $A = S \Delta S^{-1}$

$\Rightarrow AS = S\Delta$ $\quad$ If column i of S is v_i, column i of AS is Av_i and column i of $S\Delta$ is $v_i \lambda_i$

$\Rightarrow Av_i = \lambda_i v_i \Rightarrow v_i$ is an eigenvector of A, & λ_i is it's eigenvalue.

If the v_is are linearly dependent, S does not have an inverse.

Observation: Δ is necessarily the list of eigenvalues.

- Given $A = S\Delta S^{-1}$, $A^k = S \Delta^k S^{-1}$

Observe that if $Av = \lambda v$, $A^2 v = A(Av) = A(\lambda v) = \lambda(Av) = \lambda^2 v$

$\Rightarrow A^k v = \lambda^k v$. A^k has the same eigenvector set as A, only with exponentiated λs.

Given $\Delta = S^{-1}AS$, $\Delta^k = (S^{-1}AS)(S^{-1}AS)(S^{-1}AS) \cdots$

$= S^{-1} A^k S$ $\quad$ QED.