Definitions and facts

Linear Algebra :

- Let x be some vector, and X be a matrix. Unless explicit, their norms are given by

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \quad , \text{ and}$$

$$\|X\| = \left(\operatorname{trace}(X'X)\right)^{1/2}$$

- Trace is invariant under cyclical permutations

$$tr(ABCD) = tr(BCDA) = tr(CDAB) = tr(DABC).$$

- Cauchy - Schwarz : $E\| \underset{k\times 1}{a} \ \underset{1\times l}{b'} \| \leq \left(E\|a\|^2 \cdot E\|b\|^2\right)^{1/2}$.

- $A_{k\times l} \ b_{l\times 1} = b_1 \ col_1(A) + \dots + b_l \ col_l(A)$

- Matrix A is full column rank if

$$A b = 0 \quad \text{iff} \quad b = 0.$$

- If B is any $l\times k$ matrix, then

$$\operatorname{rank}(\underset{k\times l}{A} B) \leq \min(\operatorname{rank}(A), \operatorname{rank}(B))$$

- Sylvester's rank inequality: if A is $m\times n$, and B is $n\times k$, then

$$\operatorname{rank}(A) + \operatorname{rank}(B) - n \leq \operatorname{rank}(AB)$$

- Let A, B be symmetric pos def matrices. Then
  a) $I - A$ is pos def iff $A^{-1} - I$ is pos def.
  b) $B - A$ is pos def iff $A^{-1} - B^{-1}$ is pos def.

Probability and Asymptotics

- We say a sequence of vectors $X_n = o(a_n)$ iff

$$\lim_{n\to\infty} \left\| \frac{X_n}{a_n} \right\| = 0 .$$

- We say a sequence of vectors $X_n = O(a_n)$ iff $\exists\, M < \infty$ such that

$$\|X_n\| \leq M \cdot a_n , \quad \forall\, n \in \mathbb{N}.$$

- We say a sequence of random vectors $X_n = o_p(a_n)$ iff

$$\lim_{n\to\infty} P\left(\left\| \frac{X_n}{a_n} \right\| > \varepsilon \right) = 0 , \text{ for all } \varepsilon > 0 .$$

- We say a sequence of random vectors $X_n = O_p(a_n)$ iff for all $\varepsilon > 0$, $\exists\, M_\varepsilon < \infty$ such that

$$\lim_{n\to\infty} P\left(\left\| \frac{X_n}{a_n} \right\| > M_\varepsilon \right) < \varepsilon .$$

- Some implications are :
  a) $o_p(1) + o_p(1) = o_p(1)$.
  b) $O_p(1) + o_p(1) = O_p(1)$.
  c) $O_p(1) \cdot o_p(1) = o_p(1)$.
  d) $o_p(1) \cdot O(1) = o_p(1)$.
  e) $o(1)$ sequence is also $o_p(1)$.

- Let $X_n \xrightarrow[d]{} X$, for some random vector X, and $A_n \xrightarrow[p]{} a$, for some constant a. Then, there is joint convergence in distribution, i.e,

$$(X_n, A_n) \xrightarrow[d]{} (X, a).$$

- Continuous Mapping Theorem: let $\{X_n\}$ be a sequence of random vectors in $\mathbb{R}^d$ such that $X_n \xrightarrow[d]{} X$. Also, let $g(\cdot): \mathbb{R}^d \to \mathbb{R}^k$ be a continuous function on a set G such that $P\{X \in G\} = 1$ (i.e. almost everywhere). Then

$$g(X_n) \xrightarrow[d]{} g(X) \quad \text{as} \quad n \to \infty.$$

- Slutsky's Theorem : let $X_n \xrightarrow{d} X$ and $A_n \xrightarrow[p]{} a$, where a is a constant. Then,

  1) $X_n + A_n \xrightarrow{d} X + a$
  2) $A_n X_n \xrightarrow{d} aX$

  A trivial implication is that if all variables converge in probability to constants, then

  3) $A_n X_n + B_n \xrightarrow[p]{} aX + b$.

- Weak Law of Large Numbers (iid) : let $\{X_i\}_1^n$ be a sequence of iid random vectors such that $E\|X_i\| < \infty$. Then

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[p]{} E X_i .$$

- Central Limit Theorem (iid) : let $\{X_i\}_{i=1}^{n}$ be an iid sequence of random variables. Suppose $Var(X_i)$ is finite and bounded away from zero. Then

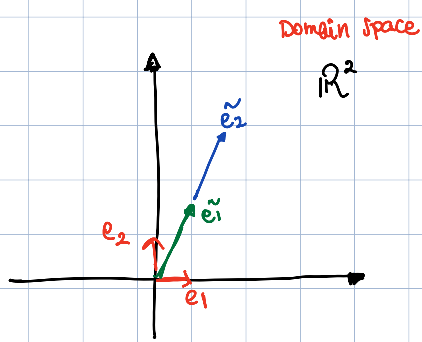
$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - EX_i}{Var(X_i)} \xrightarrow[d]{} N(0,1)$$

- To obtain a multivariate version of the CLTs we can apply the Cramer - Wald device, i.e. if it holds for any linear combination $\lambda' X$, $\lambda \neq 0$, then it must hold for the random vector.

Linear System of Equations

a) Underidentification : $Ax = b$

Consider $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Since it has only one independent column so it has $\text{rank}(A) = 1$. Notice that $A : \underbrace{\mathbb{R}^2}_{\text{domain}} \to \underbrace{\mathbb{R}^2}_{\text{codomain}}$ (where it can come out)



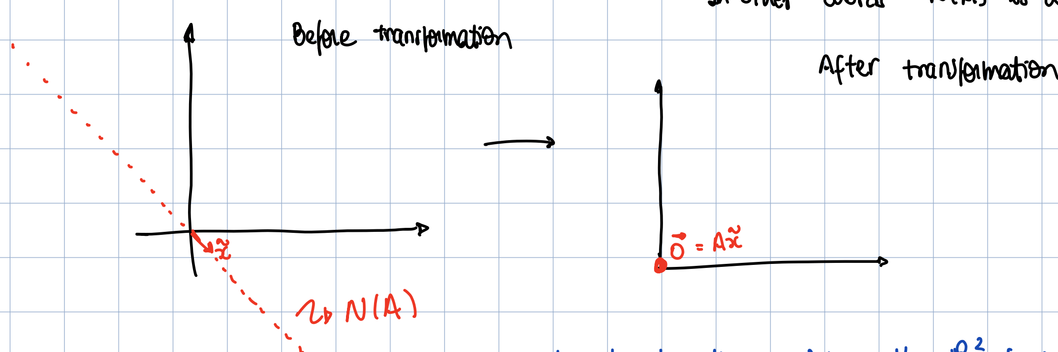
Finding the null space:

$A\tilde{x} = 0$

$\tilde{x}_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \tilde{x}_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 0$

$\tilde{x} = \begin{bmatrix} 1 \\ -1/2 \end{bmatrix}$

Notice that any other scaling of $\tilde{x}$ satisfies the homogeneous equation. In other words $N(A)$ is a subspace.



The transformation squishes the $\mathbb{R}^2$ space into a range given by the rank of A.

Theorem.- (Rank Nullity) Let $A : V \to W$ be a linear transformation. Then

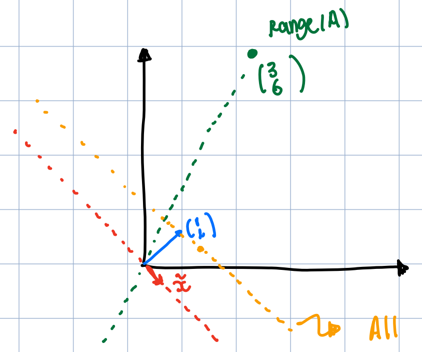
$$\text{Rank}(A) + \dim(N(A)) = \dim V$$

So in this case $\operatorname{rank}(A) = 2 - \underbrace{1}_{N(A)}$. $N(A)$ is a line, so it has dimension 1.

Therefore, the rank is a measure of how much space gets squished!

For example, the zero matrix map every vector to 0, so it's something that squishes more. In that case rank = 0.

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$



All of these vectors are $\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \underbrace{\tilde{x}}_{\text{for } \tilde{x} \in N(A)}$ and all of them are mapped to $\begin{pmatrix} 3 \\ 6 \end{pmatrix}$.

The problem of finding $A x = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$ is that it is not a bijective mapping, i.e. from $\begin{pmatrix} 3 \\ 6 \end{pmatrix}$ we don't have information about which vector in the orange line was the one that got transformed. That is essentially why inversion of A in the standard way it's not gonna work.

Def.- (Moore-Penrose inverse) The MP inverse of a matrix A (denoted by A^+) is a generalized inverse that satisfies

Reflexive Generalized Inv { Generalized Inv → (1) $A A^+ A = A$

(2) $A^+ A A^+ = A^+$

(3) $A^+ A$ is symmetric

(4) $A A^+$ is symmetric

⊛ The benefit of this is that it's unique!

Lemma.- A general solution to the homogenous system of linear equations $Ax = 0$ is

$$x = (I_m - A^+ A) q$$

where q is an arbitrary vector.

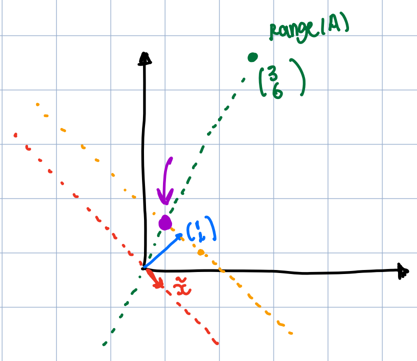
proof:

$$A(I_m - A^+ A) q = (A - \underbrace{A A^+ A}_{= A \text{ by } (1)}) q = 0.$$

In our example we got that the red line is the null space of A. Any generalized inverse of A should return a vector in the orange line.

$$A^+ = \begin{bmatrix} 0.04 & 0.08 \\ 0.08 & 0.16 \end{bmatrix}$$

And $A^+ b = \begin{bmatrix} 0.6 \\ 1.2 \end{bmatrix}$ which returns the following vector



And we know that any other $\tilde{x}$ in $N(A)$ that we sum still satisfies the equation.

Lemma.- $Ax = b$ has solution iff $\operatorname{rank}(A) = \operatorname{rank}([A \ b])$

intuition of the proof: Notice that $A = \underbrace{[\underbrace{a_1}_{\ell \times 1} \ \underbrace{a_2}_{\ell \times 1} \ \dots \ \underbrace{a_k}_{\ell \times 1}]}_{\ell \times k}$, $x = \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}}_{k \times 1}$

Then $Ax = x_1 \ a_1 + x_2 \ a_2 + \dots + x_k \ a_k$.

$\operatorname{Rank}(A) = \operatorname{Rank}([A \ b])$ means that b is linearly dependent i.e. can be written as a linear combination of the colums of A which is literally the same as saying that $Ax = b$ has a solution.

Lemma.- $Ax = b$ has a solution iff $AA^+b = b$.

proof: $(\Rightarrow)$ suppose x is a solution of $Ax = b$. Then by MP inverse property

$$\underbrace{AA^+A}_{AA^+A \ = \ A} x = b$$

$$AA^+ \underbrace{Ax}_{b} = b \Rightarrow AA^+b = b.$$

$(\Leftarrow)$ suppose $AA^+b = b$. Set $\tilde{x} = A^+b$. Then $A\tilde{x} = AA^+b = b$ so that it is a solution.

Lemma.- If $Ax = b$ has a solution, then it takes the following form

$$x = A^+b + (I_m - A^+A)q$$ where q is an arbitrary vector.

Then notice that we can always get a solution in the least squares problem even with under identification. Now we will see two special cases of this general solution.

b) <u>overidentification</u>:

$$\underset{\ell \times k}{A} \underset{k \times 1}{x} = \underset{\ell \times 1}{b}$$

- A has full column rank K
- $\ell > K$

($Z'X \beta = Z'Y$ maybe now it looks more familiar)

Consider some positive definite and symmetric matrix W.

$$\underset{\ell\times\ell}{W^{1/2}} \underset{\ell\times k}{A} \underset{k\times 1}{x} = \underset{\ell\times\ell}{W^{1/2}} \underset{\ell\times 1}{b}$$

$$A' W^{1/2} W^{1/2} A x = A' W^{1/2} W^{1/2} b$$

$$\underbrace{\underset{k\times\ell}{A'} \underset{\ell\times\ell}{W} \underset{\ell\times k}{A}}_{\text{now this has full rank!}} x = A' W b$$

$$x = \underbrace{(A'WA)^{-1} A'W}_{A^*} b$$

We'll see what conditions we need so that the generalized inverse A^* is a Moore Penrose inverse.

(1) $A A^* A = A \underbrace{(A'WA)^{-1} A'WA}_{I} = A$ ✓

(2) $A^* A A^* = (A'WA)^{-1} \underbrace{A'WA (A'WA)^{-1}}_{I} A'W = A^*$ ✓

(3) To be symmetric $(A'WA)^{-1} A'WA = A'WA (A'WA)^{-1} = I$ ✓
$\Rightarrow A^+ A = A' A^+$

(4) To be symmetric $A(A'WA)^{-1} A'W = W'A (A'WA)^{-1} A'$ ✓
which is only satisfied if $W = I$.

Therefore, the MP inverse of A is

$$A^+ = (A'A)^{-1} A' .$$