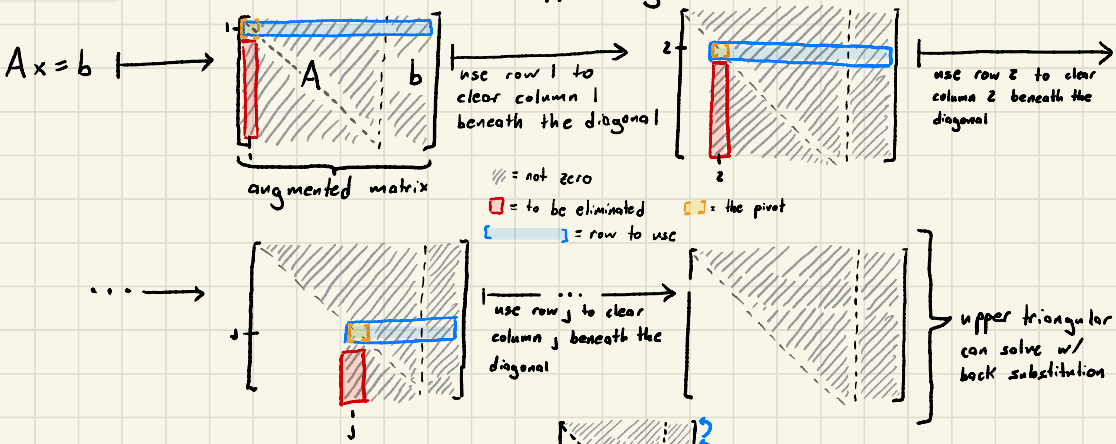
Week 4 - Tuesday - 10/18/2022

- **Logistics:**
  - HW 3 posted, due Thursday on gradescope
  - Midterm dates posted, next Wed & Thurs
  - Midterm sign up posted. Sign up for a slot this week
  - HW 4 will post on Thursday

- **Goals:**
  - Numerical Considerations for reduction:
    1. pivoting & stability
    2. time cost
  - Fundamentals:
    1. the 4 fundamental subspaces
    2. the fundamental thm of linear algebra
    3. application: least squares problems * if we make it

- **Numerics for Linear System Solving:** given $Ax=b$ we currently solve by...

1. **row-reduce:** use rows to cancel columns until upper triangular



2. **back-sub:** solve backwards bottom to top.

Questions:
1. can this algorithm fail? ← answered last week
2. is it stable?
3. how expensive/efficient is it?

} today

2. <u>Robustness & Stability</u>: does reduction + back-sub work for all nonsingular systems?
how do rounding errors during calculation propagate?

• <u>Ex</u>:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 6 \\ 3 & -1 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 2 & 2 & 6 & 2 \\ 3 & -1 & 4 & 5 \end{array}\right] \xrightarrow[\substack{-2(i)+(ii) \\ -3(i)+(iii)}]{} \left[\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 0 & \boxed{0} & 2 & 0 \\ 0 & \boxed{-4} & -2 & 2 \end{array}\right]$$

our pivot = 0, so I can't reduce further using row 2.

notice, not singular, row (ii) is not all 0's, try rearranging

• swap rows (ii) and (iii)

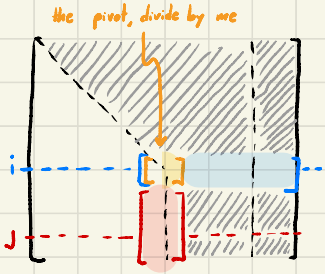
$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & -4 & -2 & 2 \end{array}\right] \xrightarrow{\text{pivoting}} \left[\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 0 & -4 & -2 & 2 \\ 0 & 0 & 2 & 0 \end{array}\right]$$

upper triangular, reduction is done proceed w/ back-sub.

• <u>Pivoting</u>: if at some stage of reduction (stage i) the pivot, $a_{ii} = 0$ then look down the col. beneath row i to try and find a_{ji} $(j > i)$, $a_{ji} \neq 0$, then swap rows j & i

• <u>Stability</u>: at stage i when we combine row i w/ row j use multiplier $\alpha = -\dfrac{a_{ji}}{a_{ii}}$ ← pivot

row j = $\alpha \cdot$ row i + row j



unstable/lose digits of accuracy if the pivot a_{ii} is very small...

• <u>Partial Pivoting</u>: at stage i of reduction find row $j \geq i$ that maximizes $|a_{ji}|$, swap rows j and i to maximize the pivot. Then reduce.

- see Matlab demo in HW 3 for demonstration.
- partial pivoting ensures robustness & stability...

3. Efficiency: how many operations does our algorithm require?

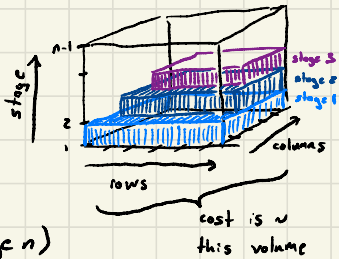
- Cost/Efficiency of: # of operations, solving $Ax=b$, A $n \times n$

  - Row-reduction: given A $n \times n$
    the cost of red.
    is $\propto n^3$ ($\Theta(n^3)$)
    expensive/slow

    - why n^3? (n stages, at j^{th} stage need to operate on $(n-j-1)$ rows & $(n-j)$ columns per row

      $$\text{cost} \simeq \sum_{j=1}^{n-1} (n-j)(n-j-1) \propto n^3 \text{ for large } n)$$

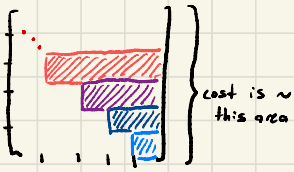
      proportional to



  - Back-Sub: cost $\propto n^2$ ($\Theta(n^2)$)

    (n stages, at j^{th} stage do $\sim 2(j-1)$ operations,

    $$\text{cost} \simeq \sum_{j=1}^{n} 2(j-1) \propto n^2 \text{ for large } n)$$

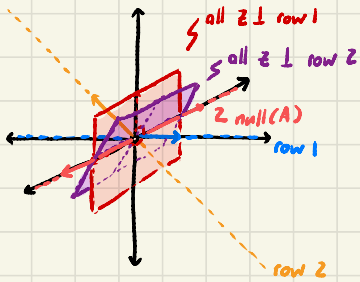


- moral: row-reduction is much more expensive than back-sub.
- See Matlab demo...

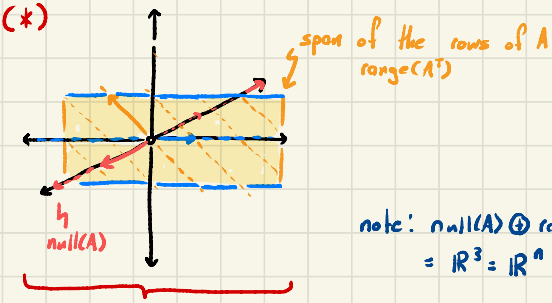
Fundamentals:

- recall from Thursday last week

- **Ex:** $A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \end{bmatrix}$ rows $\in \mathbb{R}^3$



(*)



note: $\text{null}(A) \oplus \text{range}(A^T) = \mathbb{R}^3 = \mathbb{R}^n$

null(A) is all vectors in $\mathbb{R}^3$ ⊥ to the span of the rows of A

- suggests a deep observation:

  the null(A) is all vectors ⊥ to the span of the rows of A

this idea leads to the fundamental theorem of linear algebra...

- **4 Fundamental Subspaces:**

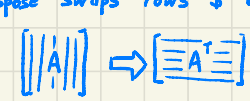
given A, $m \times n$ $\longrightarrow$ $\begin{bmatrix} A \end{bmatrix}$ (m × n)

1. range(A) = all lin. comb. of the col. of A .... $\subseteq \mathbb{R}^m$
2. null(A) = all x s.t. $Ax = 0$ ← lin. dep. of col's .. $\subseteq \mathbb{R}^n$
3. range(A^T) = all lin. comb. of the rows of A .. $\subseteq \mathbb{R}^n$
4. null(A^T) = x s.t. $A^T x = 0$ ← lin. dep. of rows .. $\subseteq \mathbb{R}^m$

- **Def:** given A, A^T (A transpose) is the matrix whose rows are the columns of A...

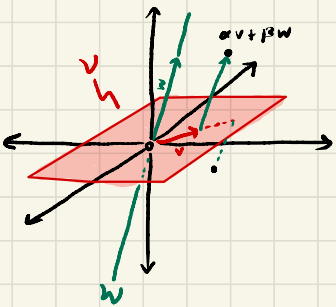
$$[A^T]_{ij} = a_{ji}$$

transpose swaps rows & columns



- **Def:** the direct sum of $\mathcal{V}$ and $\mathcal{W}$

$$\mathcal{V} \oplus \mathcal{W} = \{ \text{all } \alpha v + \beta w \text{ where } v \in \mathcal{V},\ w \in \mathcal{W} \}$$



if $\mathcal{V} = \text{span}\{v_1, v_2, \dots v_n\}$, $\mathcal{W} = \text{span}\{w_1, w_2, \dots w_m\}$ $\Big\}$ $\mathcal{V} \oplus \mathcal{W} = \text{span}\{v_1, \dots v_n, w_1, \dots w_m\}$

$\mathcal{V} \oplus \mathcal{W} = \mathbb{R}^3$

- go back to figure (*)... what is true about null(A) and range(A^T)?

- **Fund. Thm part I:**

$\mathbb{R}^m$ = the set of all vectors w/ m real entries = $\underbrace{\text{null}(A^T)}_{\text{lin. dep. rows}} \oplus \underbrace{\text{range}(A)}_{\text{span of col's}}$

(m = # rows)

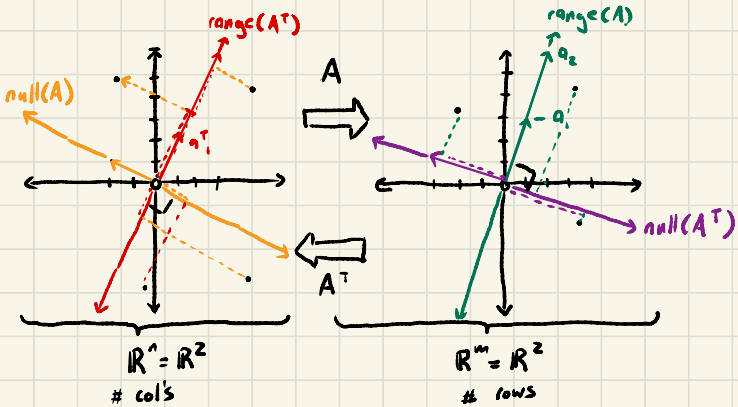
$\mathbb{R}^n$ = the set of all vectors w/ n real entries = $\underbrace{\text{null}(A)}_{\text{lin. dep. col's}} \oplus \underbrace{\text{range}(A^T)}_{\text{span of rows}}$

(n = # col's)

- Ex:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

1. $\text{range}(A) = \text{span}\left\{\begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}2\\6\end{bmatrix}\right\} \subseteq \mathbb{R}^2$ ← outputs of Ax
   $= \text{span}\left\{\begin{bmatrix}1\\3\end{bmatrix}\right\}$
2. $\text{null}(A) = \left\{x, \quad x_1\begin{bmatrix}1\\3\end{bmatrix} + x_2\begin{bmatrix}2\\6\end{bmatrix} = 0\right\}$, ← $x = \begin{bmatrix}2\\-1\end{bmatrix} \rightarrow \text{null}(A) = \text{span}\left\{\begin{bmatrix}2\\-1\end{bmatrix}\right\}$
3. $\text{range}(A^T) = \text{span}\left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}3\\6\end{bmatrix}\right\} \subseteq \mathbb{R}^2$ ← outputs of A^Tx
   $= \text{span}\left\{\begin{bmatrix}1\\2\end{bmatrix}\right\}$
4. $\text{null}(A^T) = \left\{x, \quad x_1\begin{bmatrix}1\\2\end{bmatrix} + x_2\begin{bmatrix}3\\6\end{bmatrix} = 0\right\}$, ← $x = \begin{bmatrix}3\\-1\end{bmatrix} \rightarrow \text{null}(A^T) = \text{span}\left\{\begin{bmatrix}3\\-1\end{bmatrix}\right\}$



$\mathbb{R}^n = \text{range}(A^T) \oplus \text{null}(A)$ $\qquad$ $\mathbb{R}^m = \text{range}(A) \oplus \text{null}(A^T)$

- **Def:** subspaces $\mathcal{V}, \mathcal{W}$ are <u>orthogonal</u>, if $v \in \mathcal{V}$, $w \in \mathcal{W}$ are orthogonal.
  ... any $v \in \mathcal{V}$ is $\perp$ any $w \in \mathcal{W}$

- **Def:** <u>orthogonal complements</u>: $\mathcal{V}$ and $\mathcal{W}$ are orthogonal complements if
  $\mathcal{V} = \{$all vectors $v \perp$ all vectors $w \in \mathcal{W}\}$

- **Fund. Thm II:** null(A) and range(A^T) are $\perp$ complements
  null(A^T) and range(A) are $\perp$ complements

- Let's try to prove some of the Fundamental Thm...

  - <u>notice</u>: all claims regarding null(A) and range(A^T) extend to null(A^T) and range(A)

  - <u>Claim</u>: if $z \in \text{null}(A)$ and $b \in \text{range}(A^T)$ then $\underbrace{z \text{ is } \perp \text{ to } b}_{b^T z = 0}$...

    $Az = 0$; $b = A^T x$ for some x then

    compute: $b^T z = (A^T x)^T z = (x^T A) z = x^T \overbrace{(Az)}^{=0} = 0$ (see notes)

    b is $\perp$ z. ✓

  - <u>Claim</u>: if $\underbrace{z \text{ is } \perp \text{ to any}}_{b^T z = 0}$ $b \in \text{range}(A^T)$ then $z \in \text{null}(A)$ ($b = A^T x$ for some x then $Az = 0$)

    (null(A) is the $\perp$ complement to range(A^T))

    we know that $b^T z = 0$ so, $(A^T x)^T z = 0$

    $$x^T (Az) = 0$$

    $$\left[— x^T — \right] \begin{bmatrix} | \\ Az \\ | \end{bmatrix} = 0$$

    holds for <u>any</u> x, try:

    $x = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = e_1$ then $x^T(Az) = [Az]_1 = 0$

    $x = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = e_2$ then $x^T(Az) = [Az]_2 = 0$

    $\vdots$

    $\therefore Az = 0.$ ⟵ holds for all i... ⟶ $x = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = e_i$ (ith) then $x^T(Az) = [Az]_i = 0$

  - <u>Challenge</u>: if b is $\perp$ to any $z \in \text{null}(A)$ then $b \in \text{range}(A^T)$

  - <u>Challenge</u>: how are the rank-nullity & fundamental thm's related? does one imply the other?

- <u>Questions for next unit</u>:

  1. can we compute the dimensions of the subspaces?
  2. can we find a basis for the subspaces?
  3. can we represent A in terms of its subspaces?

- now, let's look ahead to see why the fundamental thm matters...

- **Example Application**: Least Squares Problems... Solving $Ax \approx b$

  - Suppose: given A $m \times n$ and b $m \times 1$
    usually seek x s.t. $Ax = b$
    what if $b \notin \text{range}(A)$?
    then... ??

    often occurs when overconstrained...
    more eqn than unknowns
    $m > n$

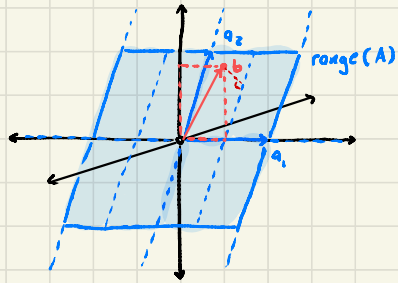
this is the standard setting when fitting data to a model:

$$\text{data} = Ax + \text{measurement error}$$

so $\text{data} \notin \text{range of } A$

or,

$$\text{data} \approx Ax \quad \text{but w/ modeling error}$$



so, no x exists s.t. $Ax = b$, let's find x s.t. Ax is as close to b as possible, that is x that minimizes the discrepancy:

$$\text{discrepancy} = Ax - b \qquad \text{want } Ax \approx b \text{ so}$$

LS problem: find x which minimizes $\|Ax - b\|$

equivalent to:

find x which minimizes $\|Ax - b\|^2 = \sum_{i=1}^{m} \overbrace{(Ax-b)_i}^{\text{disc. in } i^{th} \text{ entry}}{}^2$

$= \text{sum of squares of discrepancy.}$

- **Fact**: the LS problem has a soln for all A & b ← a soln always exists

  (note: the soln may not be unique)

- What about uniqueness? suppose x_* to the LS problem

  $z \in \text{null}(A)$, $z \neq 0$

  $$\|A(x_* + z) - b\|^2 = \|Ax_* + \underbrace{Az}_{0} - b\|^2 = \|Ax_* - b\|^2$$

  then $x_* + z$ is also a soln, so soln's are not unique.

... using LS does not solve the uniqueness...

- LS problems are the <u>most</u> widely solved opt. problems...

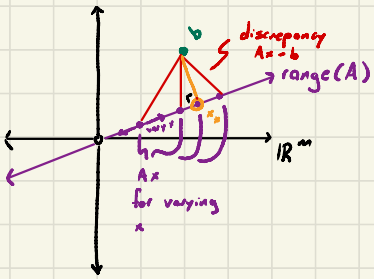
  1. data is often noisy, noise ~ Gaussian dist, prob $\propto \exp(-\|\text{discrepancy}\|^2)$
  2. LS problems are "easy" to solve at large scale
  3. LS problems admit many methods/approaches
  4. LS problems are easily generalized, adaptible

- ok, so how do we solve them?

- <u>A geometric soln for LS problems</u>:

  - given A $m \times n$, b $m \times 1$ find: $x_* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left(\|Ax - b\|^2 \right)$

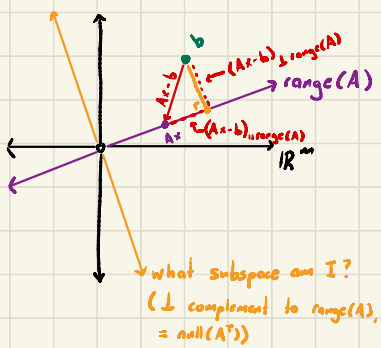
  - a picture to build intuition...

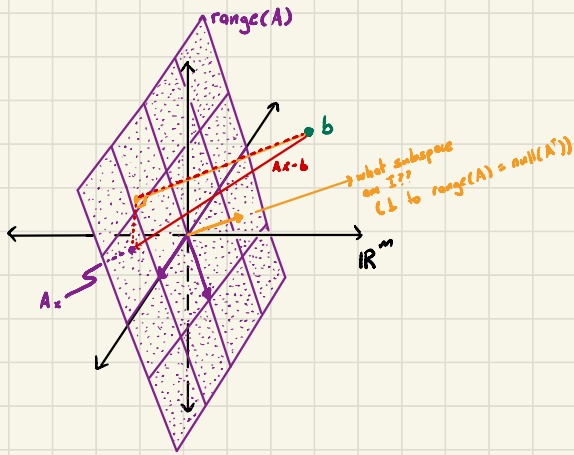


- try a series of x values

- <u>notice</u>: x_* is chosen such that the discrepancy is $\perp$ to the range(A).

- why?





- $b \in \mathbb{R}^m$ so $Ax - b \in \mathbb{R}^m$ $\exists$ can decompose the discrepancy into a component in range(A) and null(A^T)... (Fund. Thm part I: $\mathbb{R}^m$ = range(A) $\oplus$ null(A^T))

$$Ax - b = (Ax-b)_{\parallel \text{range}(A)} + (Ax-b)_{\parallel \text{null}(A^T)}$$

• moreover $(Ax-b)_{\parallel range(A)}$ is $\perp$ $(Ax-b)_{\parallel null(A^T)}$ (Fund. Thm. part II)

so, by pythagorus:

$$(*) \qquad \|Ax-b\|^2 = \underbrace{\|(Ax-b)_{\parallel range(A)}\|^2}_{\text{depends on } x} + \underbrace{\|(Ax-b)_{\parallel null(A^T)}\|^2}_{\text{same for all } x}$$

so, minimizing $\|Ax-b\|^2$ is the same as minimizing $\|(Ax-b)_{\parallel range(A)}\|^2$

(**) • <u>Fact</u>: for any A, b $\exists$ an x_* s.t. $(Ax-b)_{\parallel range(A)} = 0$

$\therefore$ by (*) $\|Ax-b\|^2 \geq \|(Ax-b)_{\parallel null(A^T)}\|^2$ for all x

and by (**) at x_* s.t. $(Ax-b)_{\parallel range(A)} = 0$, $\|Ax-b\|^2 = \|(Ax-b)_{\parallel null(A^T)}\|^2$

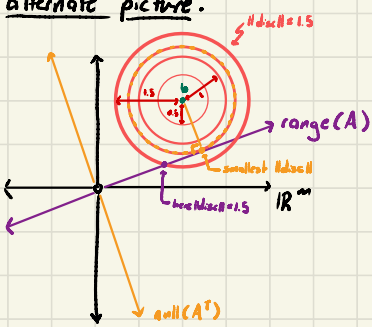
so if x_* solves the LS problem then $(Ax_*-b)_{\parallel range(A)} = 0$

so:

$$\text{discrepancy} = Ax_* - b \in null(A^T)$$

$\therefore$ (Fund. Thm) discrepancy is $\perp$ range(A).

• <u>alternate picture</u>:



at soln x_* the discrepancy must be $\perp$ to range(A), since it is a radial vector to the circle/sphere centered at b of smallest radius intersecting range(A). The circle must be tangent to the range there, so the radial vector is $\perp$ the range.

$\therefore$ discrepancy $\in null(A^T)$ (Fund Thm)

• <u>Consequence</u>: at soln x_*, discrepancy $\in null(A^T)$

so...

$$A^T(\text{discrepancy}) = A^T(Ax_* - b) = 0$$

rearrange:

$$\underbrace{A^T(Ax_*) = A^T b.}_{\text{the normal eqn's}} \longleftarrow \text{but this is a linear system!!}$$

• Can solve all LS problems by solving a linear system!