Recall from last lecture:

Theorem (Projection Theorem):

Let $X \subseteq \mathbb{R}^n$ be a subspace of $\mathbb{R}^n$.
Then for any $y \in \mathbb{R}^n$, there exists a unique $x^* \in X$ such that

$$\|y - x^*\| \leq \|y - x\| \quad \forall x \in X.$$

The point $x^* \in X$ is uniquely characterized by $(y - x^*) \in X^\perp$. We call x^* the projection of y onto X.

explicit formula:

- $$x^* = \underbrace{A(A^TA)^{-1}A^T}_{= AA^\dagger =: P_X} y$$

viewed as a constrained optimization problem:

- x^* is optimizer of
$$\min_x \|y - x\| \quad \text{s.t.} \quad x \in X \subseteq \mathbb{R}^n$$

Today:

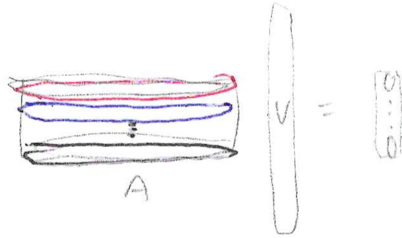
Theorem: Consider an underdetermined system $Ax = b$ where $A \in \mathbb{R}^{m \times n}$ has linearly independent ROWS. Then the unique minimum norm solution is characterized by $x^* \in \text{range}(A^T)$. Using this information alone, the explicit formula

$$x^* = A^T(AA^T)^{-1} b.$$

can be derived

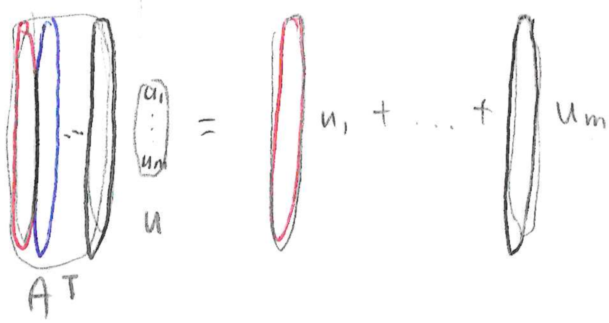
First, observe: $\quad A \in \mathbb{R}^{m \times n}, \quad m \leq n$

$$\text{null}(A) = \{ v \in \mathbb{R}^n : Av = 0 \}$$



null(A) is set of vectors v that are orthogonal to all $\underbrace{\text{rows of } A}$

= columns of A^T. (by def. of A^T)



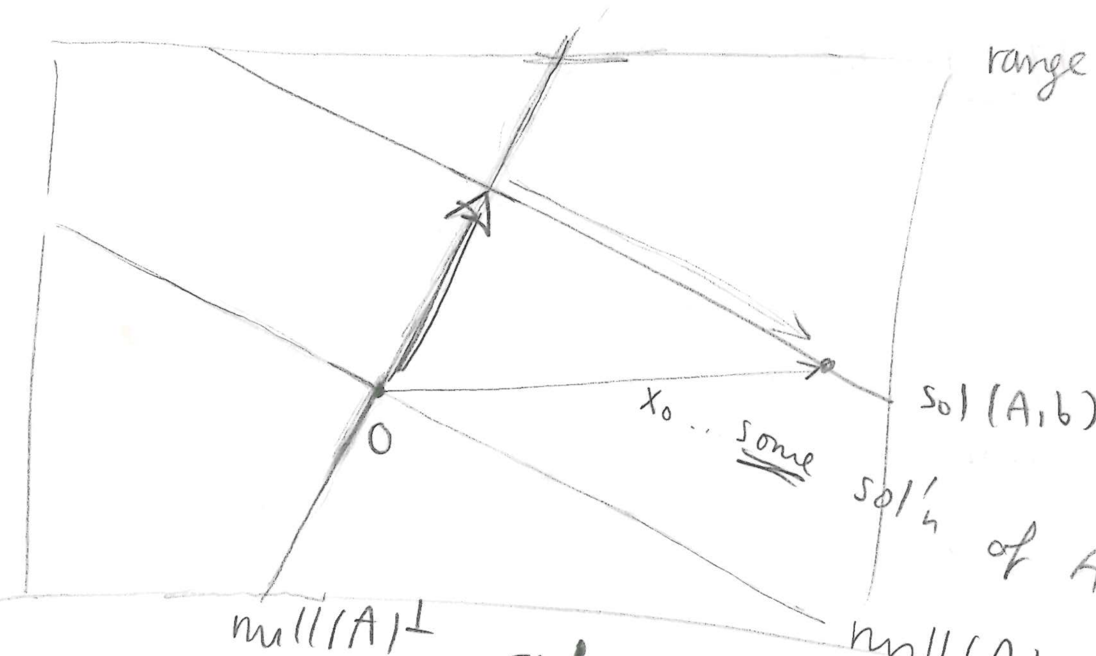
$\in$ range(A^T)

$\Rightarrow$ vectors in null(A) are orthogonal to vectors in range(A^T).

$$\text{null}(A)^\perp = \text{range}(A^T)$$
$$(\text{range}(A^T)^\perp = \text{null}(A))$$

$\mathrm{range}(A^T) \overset{\perp}{\oplus} \mathrm{null}(A) = \mathbb{R}$

remove that portion



x_0 ... <u>some</u> sol'n of $Ax=b$, i.e. $Ax_0=b$

$\mathrm{null}(A)^{\perp} = \mathrm{range}(A^T)$!

PROOF #1 : (Proof by picture)

Project x_0 onto $\mathrm{range}(A^T)$ by using the formula
(*) $P_X = A(A^TA)^{-1}A^T$ for projecting onto $\mathrm{range}(A)$.
That is, replace all appearances of A w/ A^T in (*) and vice versa.

x^* ... optimal sol of $Ax=b$

$= A^T(AA^T)^{-1}\underbrace{Ax_0}_{=b}$

Sol'n now stated in terms of b and independent of choice of particular solution! ☺

$$\boxed{x^* = A^T(AA^T)^{-1}b}$$

PROOF #2 (more formal, using the Projection Theorem)

Recall that for an underdetermined system $Ax=b$, once we have A SOLUTION (one of infinitely many) x_0, we can generate all the other solutions by adding elements of **null(A)** to that solution x_0.

$$A(\underset{\substack{\uparrow \\ \text{in null}(A)}}{x_0 + h}) = \underbrace{Ax_0}_{=b} + \underbrace{Ah}_{=0} = b \Rightarrow x_0 + h \text{ is sol'n also.}$$

Denote x^* the min. norm sol'n of $Ax=b$ (to be sought)

Write $x^* = \underset{\substack{\uparrow \\ \text{some "anker"}}}{x_0} - y^*$ with $y^* \in \text{null}(A)$.

Since x^* is min. norm sol'n, it has to hold that

$$\|x_0 - y^*\| \leq \|x_0 - y\| \quad \forall y \in \text{null}(A) = X$$

This fits description of <u>Projection Theorem</u> precisely.

Thus $y^* \in \text{null}(A)$ is <u>uniquely characterized</u> by

$$x^* = x_0 - y^* \in \text{null}(A)^\perp = \underline{\text{range}(A^T)}$$

Now we know that the optimal $x^* \in \text{range}(A^T)$, we substitute

$$x^* = A^T u^*$$

and plug it into $A\underbrace{x}_{A^T u} = b$ as follows:

$$(AA^T) u = b \Rightarrow \underset{\substack{\uparrow \\ \text{unique}}}{u} = (AA^T)^{-1} b \Rightarrow \underline{x^* = A^T(AA^T)^{-1} b}$$

back of envelope proof portion

A having linearly independent rows means full row rank and thus AA^T (square matrix) is exactly invertible! ▨

Conjugate Gradient Method

Particularly useful for solving quadratic problems

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T Q x - b^T x \quad , \quad Q \succ 0 ,$$

or equivalently, the linear system $Qx = b$
since $(\nabla f)(x) = 0 \iff Qx - b = 0$.

For large n, $Qx = b$ can be surprisingly hard to solve, so having an iterative method is desired.

"Standard" Gradient Descent with exact line search produces directions that are necessarily orthogonal (HW3/P1), which leads to a zig-zag path that takes unnecessarily long to approach x^*, especially towards later iterations.

The Conjugate Gradient Method guarantees convergence to x^* in n steps!

Def. Given an $n \times n$ symmetric matrix Q, we call a set of n non-zero vectors $\{d_1, \dots, d_n\} \subset \mathbb{R}^n$ Q-conjugate if

$$\underbrace{d_i^T Q d_j}_{= \langle d_i, d_j \rangle_Q} = 0 \quad \forall_{\substack{i \neq j \\ i,j \in \{1, \dots, n\}}}.$$

inner product induced by Q.

Proposition: If $Q \succ 0$ and $d_1, \dots, d_n$ are Q-conjugate, then $d_1, \dots, d_n$ are linearly independent!

Proof: HW 6, P4.

Main idea of Conjugate Gradient Method:

With $d_0, d_1, \dots, d_{n-1}$ Q-conjugate directions,
(by the proposition, this is a basis of $\mathbb{R}^n$!),

The solution x^* of the quadratic optimization problem can be expressed as

$$x^* = \sum_{i=0}^{n-1} \alpha_i d_i$$

for some suitable α_i

Applying $d_i^T Q$ from the left onto $x^* = \sum_{i=0}^{n-1} \alpha_i d_i$ yields

$$d_i^T Q x^* = \alpha_i d_i^T Q d_i \quad (d_i^T Q d_j = 0 \text{ if } i \neq j)$$

$$\Rightarrow \alpha_i = \frac{d_i^T Q x^*}{d_i^T Q d_i} \underset{Qx^* = b}{=} \frac{d_i^T b}{d_i^T Q d_i}$$

← now independent of x^*! NICE

Thus
$$x^* = \sum_{i=0}^{n-1} \alpha_i d_i = \sum_{i=0}^{n-1} \frac{d_i^T b}{d_i^T Q d_i} d_i .$$

The above expansion for x^* can be considered to be the result of an iterative process of n steps where the ith step adds "$+ \alpha_i d_i$":

$\underline{i=0}$: $\alpha_0 d_0$

$\underline{i=1}$: $\alpha_0 d_0 + \alpha_1 d_1$

$\vdots$

$i = n-1$: $x^* = \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{n-1} d_{n-1}$

The following result generalizes the above observations by incorporating an initialization $x^0 \neq 0$. The derivation is analogous and left as an exercise (HW 6 / P5).

Theorem (Conjugate Direction Theorem)

Let $d_0, d_1, \ldots, d_{n-1}$ be a set of nonzero Q-conjugate vectors ($Q \succ 0$). For any $x^0 \in \mathbb{R}^n$, the sequence generated via

$$x^{k+1} = x^k + \alpha_k d_k$$

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k} \quad \text{where } g_k = Qx^k - b, \quad \left(\text{gradient of } f(x) = \frac{1}{2}x^T Q x - b^T x\right)$$

converges to the unique solution x^* of $Qx = b$ after n steps, i.e. $x^n = x^*$.