Linear Algebra
Week 9

Felix Breuer
G-04

The four fundamental subspaces cont'd

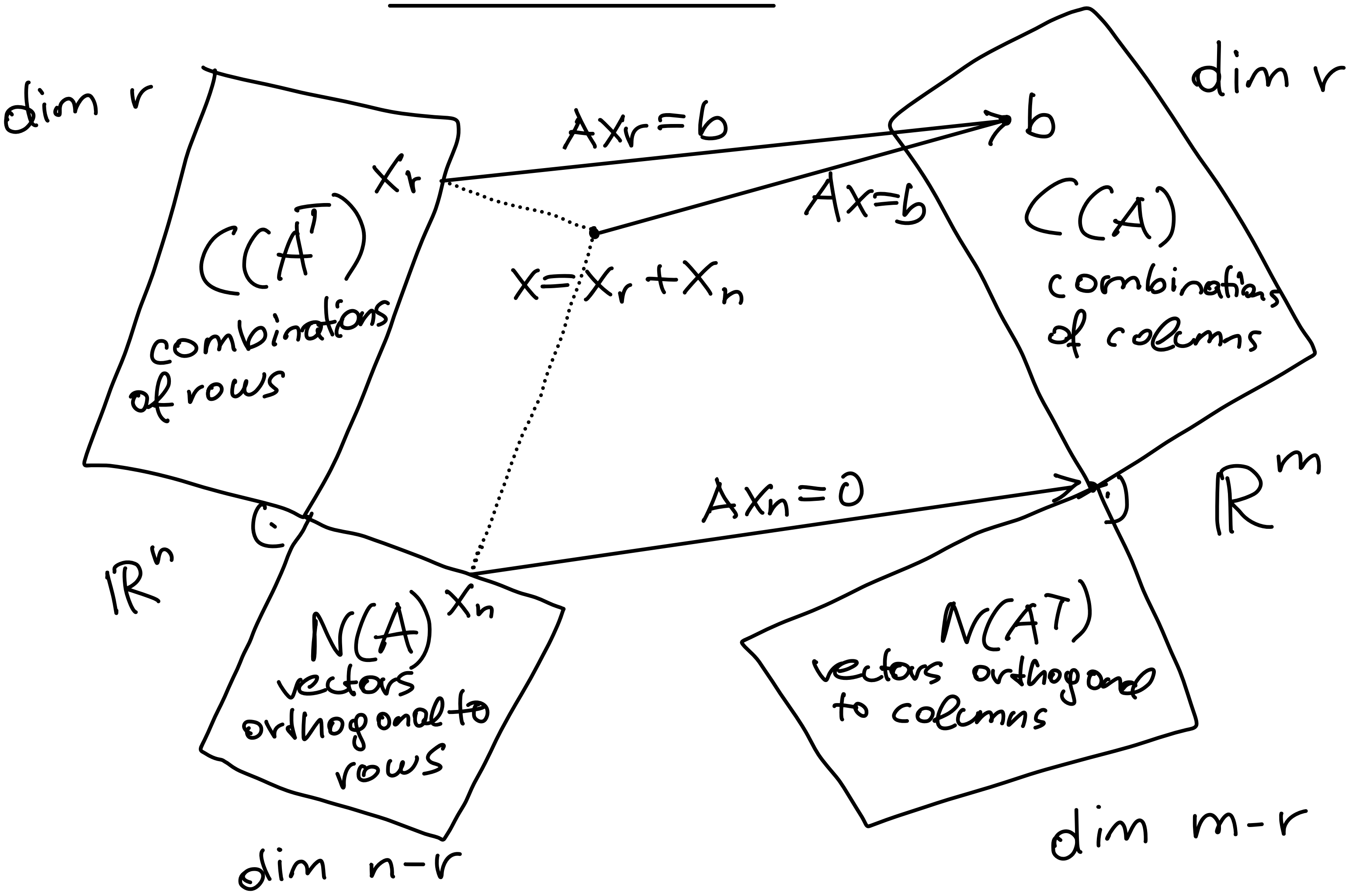
$A \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A) = r$

$$C(A) = N(A^T)^{\perp}$$
$$C(A^T) = N(A)^{\perp}$$

$\dim C(A) = r \qquad \dim N(A^T) = m - r$

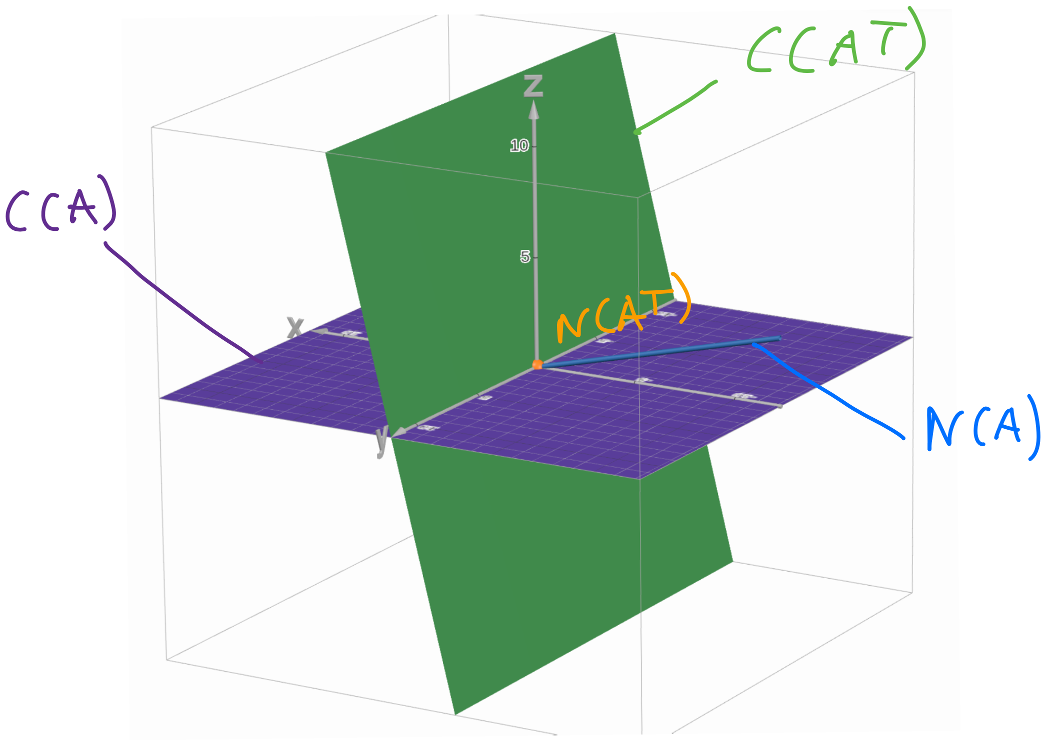
$\dim C(A^T) = r \qquad \dim N(A) = n - r$

"The big picture"



Example

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 5 & 8 \end{bmatrix}$$



https://www.desmos.com/3d/d8fb99479d

$m = \#\text{rows} = 2 = \dim C(A) + \dim N(A^T) = 2 + 0$

$n = \#\text{columns} = 3 = \dim C(A^T) + \dim N(A) = 2 + 1$

subspace	basis	dimension
$C(A)$	$\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right\} \subseteq \mathbb{R}^2$	2
$N(A)$	$\left\{ \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^3$	1
$C(A^T)$	$\left\{ \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} \right\} \subseteq \mathbb{R}^3$	2
$N(A^T)$	$\cancel{\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}} \subseteq \mathbb{R}^2$ corrected: $\{\} = \emptyset$	0

Extra Example (not discussed in session)

1. Let $A \in \mathbb{R}^{4\times 7}$ have rank 3.
What are the dimensions of:
$C(A)$, $N(A)$, $C(A^T)$, $N(A^T)$?

Answer:

$\operatorname{rank} A = \dim C(A) = \dim C(A^T) = 3$

$\dim N(A) = 7 - 3 = 4$

$\dim N(A^T) = 4 - 3 = 1$

Just knowing the rank and dimensions of A tells us a lot of information!
→ e.g. regarding uniqueness and existence of solutions to $Ax=b$

Exercise

2. Let $A \in \mathbb{R}^{n\times n}$ be symmetric.
Claim: If $Ax=0$ and $Az=5z$, x and z are orthogonal.

We can use that $N(A)$ and $C(A^T)$ are orthogonal.

$Ax=0$, hence $x \in N(A)$.

$Az = A^T z = 5z$, hence $5z \in C(A^T)$.

$\Rightarrow x \cdot 5z = 0 \quad (N(A)^\perp = C(A^T))$

$\Leftrightarrow 5(x \cdot z) = 0$

$\Leftrightarrow x \cdot z = 0$

Thus per definition x and z are orthogonal.

A related and famous theorem is the following:

Rank-Nullity-Theorem

$A : \mathbb{R}^n \to \mathbb{R}^m$

$$\dim C(A) + \dim N(A) = \dim \operatorname{domain}(A) = n$$

(pg 31)

A similar result as the following was proved in the lecture.
We didn't do this proof in the exercise session but if you found the proof at the bottom of page 31 confusing maybe this is helpful.

Claim: $C(A^T) \perp N(A)$

Let $x \in N(A)$. Then per definition $Ax = 0$:

dot product

$$Ax = 0 \iff \begin{bmatrix} - & a_1 & - \\ & \vdots & \\ - & a_m & - \end{bmatrix} x = 0 \iff \begin{matrix} a_1^T \cdot x = 0 \\ \vdots \\ a_m^T \cdot x = 0 \end{matrix} \quad (1)$$

$\Rightarrow$ x is orthogonal to any row of A (def orthogonal)

Let $y \in R(A) = C(A^T)$.
Then $y = \alpha_1 a_1^T + \dots + \alpha_m a_m^T$ for some $\alpha_1, \dots, \alpha_m \in \mathbb{R}$.

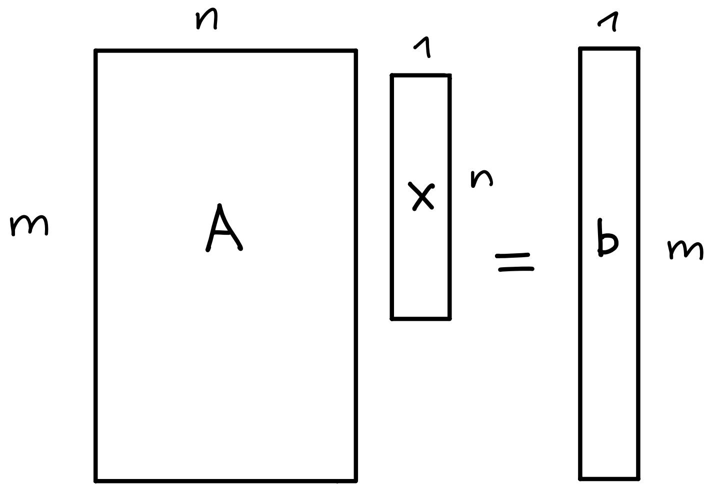
$$\begin{aligned} x \cdot y &= x \cdot (\alpha_1 a_1^T + \dots + \alpha_m a_m^T) \\ &= (x \cdot \alpha_1 a_1^T) + \dots + (x \cdot \alpha_m a_m^T) \quad \text{(distributivity dot product)} \\ &= \alpha_1 (x \cdot a_1^T) + \dots + \alpha_m (x \cdot a_m^T) \\ &= \alpha_1 0 + \dots + \alpha_m 0 \quad (1) \\ &= 0 \end{aligned}$$

$\Rightarrow$ x is orthogonal to any element in $R(A) = C(A^T)$
(which is a linear combination of the rows of A)

An arbitrary $x \in N(A)$ is orthogonal to all elements of $C(A)^T$. Hence $C(A^T) \perp N(A)$ per definition.

Least Squares

Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. $m >> n$



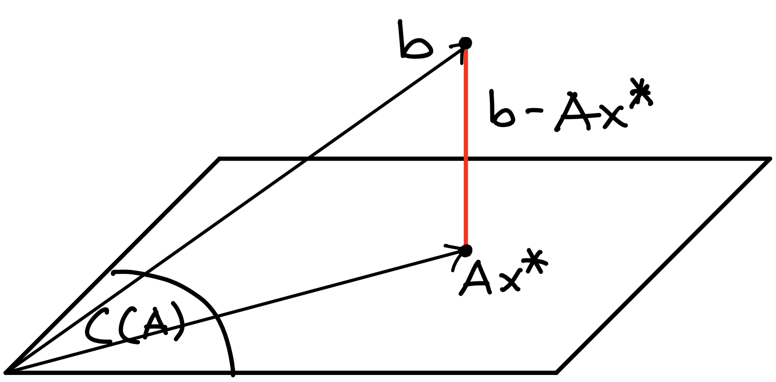
"wish list" (A) where not all wishes (b) can be fulfilled (by x)

When the number of equations (m) is much larger than the number of unknowns (n) usually $Ax = b$ has no solution. The least squares method allows us to find an approximate solution:

$$x^* = \underset{x \in \mathbb{R}^n}{\arg\min} \|Ax - b\|_2^2$$

argmin gives us the vector x that minimizes the expression, not the minimum itself.

But how do we find x^*? A geometric derivation of the normal equations:



$$b - Ax^* \perp C(A)$$
$$\Rightarrow b - Ax^* \in C(A)^\perp$$
$$\Rightarrow b - Ax^* \in N(A^T)$$
$$\Rightarrow A^T(b - Ax^*) = 0$$
$$\Leftrightarrow \underbrace{A^T A x^* = A^T b}_{\text{the normal equations}}$$

Exercise (HS20)

Given are data points:

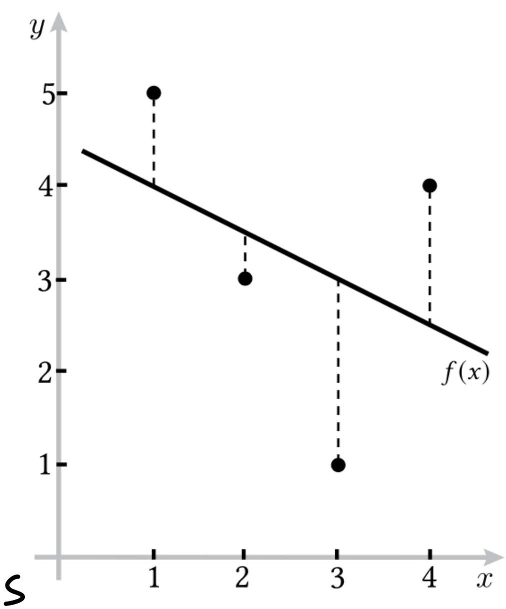
$(x_1, y_1) = (1, 5)$
$(x_2, y_2) = (2, 3)$
$(x_3, y_3) = (3, 1)$
$(x_4, y_4) = (4, 4)$

Find coefficients of the polynomial $f(x) = a_0 + a_1 x$ such that vertical distances between $f(x)$ and the given points is minimized in the least squares sense and compute

$$x^* = \underset{a_0, a_1 \in \mathbb{R}}{\operatorname{argmin}} \sum_{j=1}^{4} (f(x_j) - y_j)^2$$

Solution:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 1 \\ 4 \end{bmatrix}$$



Solve with normal equations

Side note:

$A^T A$'s entries: dot products of columns

$A A^T$'s entries: dot products of rows of A → symmetric, can save steps

What does this actually minimize?

$$(Ax-b)_k = \left(\sum_{j=0}^{n-1} a_{kj} x_j\right) - b_k$$

$$\|Ax-b\|^2 = \langle Ax-b, Ax-b \rangle = \underbrace{\sum_{k=0}^{m-1} \left(\left(\sum_{j=0}^{n-1} a_{kj} x_j\right) - b_k\right)^2}_{\text{sum of squared function}}$$

Some remarks:

- crucial part: finding A
  → practice. sometimes the main task is understanding what you're supposed to do, so getting familiar with how things might be worded is good. Sometimes there's a picture in the exam, e.g. of non linear function → not useful for solving the task usually.
- As long as objective function is linear in the parameters we want to find, we can use least squares
- Slides from last year that show application of least squares:

https://igl.ethz.ch/teaching/linear-algebra/la2022/notes/22_11_30+12_02.pdf

Projections

The projection of $b \in \mathbb{R}^m$ on a subspace $S \subseteq \mathbb{R}^n$ is

$$\text{proj}_S(b) = \underset{p \in S}{\text{argmin}} \; \|b - p\|$$

There's many fundamental and interesting questions about projections: e.g. Why is such a projection orthogonal? which might have been covered in the lecture.
I encourage you to use the lecture notes and question the given definitions: What do they capture and why/how?

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$:

The vector $x^* \in \mathbb{R}^n$ that fulfills the normal equations leads to the projection of b onto $C(A)$. Assuming $\text{rank}(A) = n$ we get:
(using $\text{rank}(A) = n \Rightarrow \text{rank}(A^TA) = n$ (exercise sheet 7.2))

$$A^TAx^* = A^Tb \iff Ax^* = \underbrace{A(A^TA)^{-1}A^Tb}_{\text{projection of } b \text{ onto } C(A)}$$

$$\frac{a^Tb}{a^Ta}a$$

$$= \left.\frac{aa^T}{a^Ta} b \right\} \text{projection of } b \text{ onto subspace spanned by } a$$

We can confirm this with the geometric meaning of the dot product discussed in week 2 and derive the formula with a similar argument to the one used earlier to derive the normal equations (this can also be found in the lecture notes).

Orthonormality

We call a set of vectors $\{q_1, \dots, q_n\}$ **orthonormal** if each vector in the set has length one and is orthogonal to all others in the set. We can express this as:

$$q_i^T q_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} = \delta_{ij}$$

- Can you give an example of such a set?
- Is any such set linearly independent?

An example of such a set is the canonical basis of $\mathbb{R}^3$,
$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Why are orthonormal vectors useful?
$\longrightarrow$ They make projections easier $\rightarrow$ insert in normal equations
(there's also other answers to this)

We call a matrix $Q \in \mathbb{R}^{n \times n}$ **orthogonal** if $Q^T Q = I$.
Properties
- the columns and rows of Q form orthonormal bases of $\mathbb{R}^n$
- they preserve lengths: $\|Qx\| = \|x\|$

$\longrightarrow$ What kinds of matrices are orthogonal?
Rotations, reflections

The Gram-Schmidt Algorithm

input: $\{a_1, \dots, a_n\}$, linearly independent

output: $\{u_1, \dots, u_n\}$, orthonormal

$\mathrm{span}\{u_1, \dots, u_k\} = \mathrm{span}\{a_1, \dots, a_k\}$ for all $1 \leq k \leq n$

Pseudocode

$$u_1 = \frac{a_1}{\|a_1\|}$$

for $k = 2, \dots, n$

scalar product

$$\tilde{u}_k = a_k - \underbrace{\sum_{i=1}^{k-1} \langle a_k, u_i \rangle u_i}_{\text{projection of } a_k \text{ onto } \mathrm{span}(u_1, \dots, u_{k-1})}$$

$$u_k = \frac{\tilde{u}_k}{\|\tilde{u}_k\|}$$

Demo

https://www.desmos.com/3d/ac00d3e14b

References:

Last years course

https://github.com/mitmath/1806