Linear Algebra
Week 11

Felix Breuer
G-04

Various remarks:

I recommend this for any proofs!

1) Writing induction proofs
   - clearly state the statement you are proving
   - make sure the quantifiers in the induction hypothesis are correct

Example boilerplate:

For all $n \in \mathbb{N}$ let $P(n)$ denote the statement ...

Base case ($n=0$)

...

Induction hypothesis

Assume there is some $n \in \mathbb{N}$ such that $P(n)$ holds.

Induction step ($n \to n+1$)

... (prove the implication $P(n) \Rightarrow P(n+1)$ for an arbitrary $n \in \mathbb{N}$. This n may specifically be the one from the base case, we can't make any assumptions on it if we want to prove $P(n)$ for all $n \in \mathbb{N}$)

Hence per the principle of induction $P(n)$ holds for all $n \in \mathbb{N}$.

2) Computing the $A = QR$ decomposition

- When computing Q via the Gram-Schmidt Algorithm, process the columns of A in their order!

  Column swaps affect the output of Gram-Schmidt.

- It is possible to extend the QR decomposition to allow column swaps giving us $AP = QR$ but we do not allow this with the definition from the lecture!

3 useful properties of the Pseudoinverse

1) If A is invertible, $A^+ = A^{-1}$

The formula for full column rank (or row rank) applies:

$$A^+ = (A^T A)^{-1} A^T \underset{\substack{\downarrow \\ \text{this step only works as } A \text{ is invertible}}}{=} A^{-1} (A^T)^{-1} A^T = A^{-1}$$

2) If $Ax = b$ has no solution, $x^+ = A^+ b$ is closest to a solution

$Ax^+ = AA^+ b$ is the projection of b onto $C(A)$.

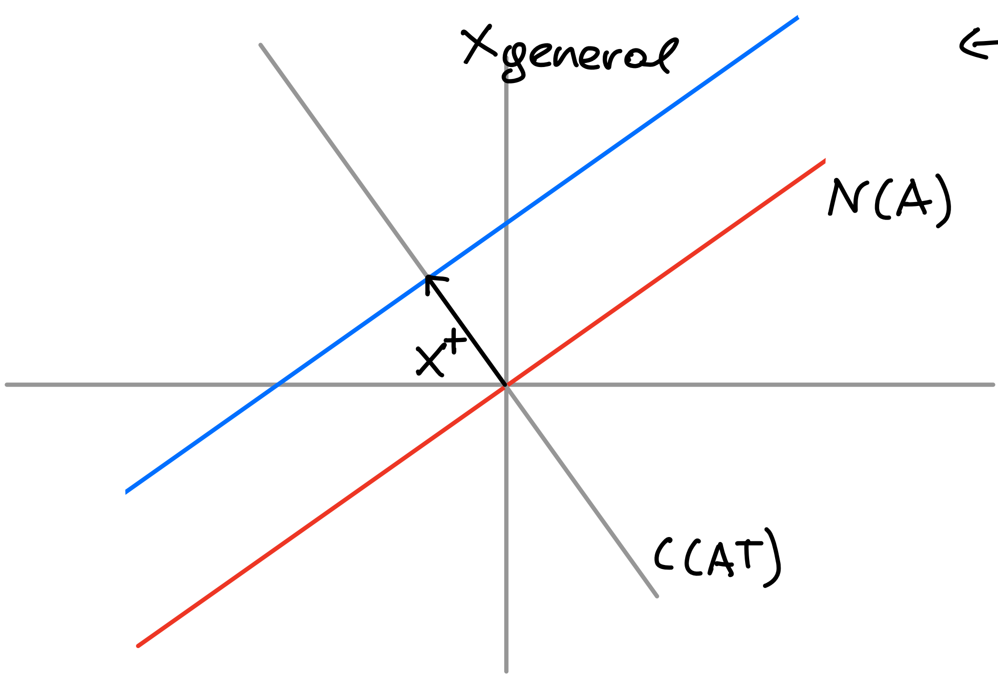
In particular, x^+ also satisfies the normal equations and solves the least squares problem $x^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Ax - b\|^2$

3) If $Ax=b$ has infinitely many solutions, $x^+=A^+b$ is the solution of minimal norm

The set of solutions to $Ax=b$ is given by

$x_{general} = \{x_p + x_H \mid x_H \in N(A)\}$.

x^+ is exactly that solution in $x_{general}$ which has $x_H=0$, $x^+ \in C(A^T)$. The following sketch illustrates this:



← this sketch may be helpful for understanding the proof of prop. 4.5.5.

$x^+=A^+b$ also gives the minimum norm solution for the normal equations $A^TAx=A^Tb$.

Exercise 9.1

a) Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $\operatorname{rank}(A) = \operatorname{rank}(B) = n$

We first prove $\operatorname{rank}(AB) = n$ so we can apply Prop 4.5.9. on AB:

<u>Claim</u> $\operatorname{rank}(AB) = n$

We show $C(AB) = C(A)$ by proving $C(A) \subseteq C(AB)$ and $C(AB) \subseteq C(A)$:

($\subseteq$) Let $y \in C(AB)$, then $y = ABx$, $x \in \mathbb{R}^p$.
$ABx \in C(A)$
Hence $C(AB) \subseteq C(A)$.

($\supseteq$) Let $y \in C(A)$
$\Rightarrow y = Ax$, $x \in \mathbb{R}^n$
B is surjective, hence there is some $z \in \mathbb{R}^p$ such that $x = Bz$
$\Rightarrow y = ABz \in C(AB)$
Hence $C(A) \subseteq C(AB)$.

Now we get $\operatorname{rank}(AB) = \dim C(AB) = \dim C(A) = \operatorname{rank}(A) = n$.
Finally, we can apply Prop. 4.5.9. on $M = AB$:

$$M^+ = (AB)^+ = B^+ A^+$$

Remark:
In general, $\operatorname{rank}(AB) \leq \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}$. The reason that $\operatorname{rank}(AB) = \operatorname{rank}(A) = \operatorname{rank}(B)$ holds here is that A has full column rank (it is injective) and B has full row rank (it is surjective). An alternative approach is to prove $N(AB) = N(B)$ and use the rank nullity theorem.

b) Claim: $(A^+)^T = (A^T)^+$

We consider a full rank decomposition $A = ST$ of A where $S \in \mathbb{R}^{m \times r}$, $T \in \mathbb{R}^{r \times n}$, $\operatorname{rank}(A) = \operatorname{rank}(S) = \operatorname{rank}(T) = r$.

$$(A^+)^T \overset{4.5.9.}{=} (T^+ S^+)^T = (S^+)^T (T^+)^T$$

$$(A^T)^+ = (\underbrace{T^T S^T}_{\text{this is a full rank decomp. of } A^T})^+ \overset{4.5.9.}{=} (S^T)^+ (T^T)^+$$

Now proving $(A^T)^+ = (A^+)^T$ for 1) A with full column rank
2) A with full row rank

suffices to conclude the claim: S has full column rank, T full row rank.

1) $(A^+)^T = ((A^TA)^{-1}A^T)^T$ (A has full column rank)

$= (A^T)^T((A^TA)^{-1})^T$

$= (A^T)^T((A^TA)^T)^{-1}$

$= (A^T)^T(A^T(A^T)^T)^{-1}$

$= (A^T)^+$ (A^T has full row rank)

2) $(A^+)^T = (A^T(AA^T)^{-1})^T$ (A has full row rank)

$= ((AA^T)^{-1})^T(A^T)^T$

$= ((AA^T)^T)^{-1}(A^T)^T$

$= ((A^T)^TA^T)^{-1}(A^T)^T$

$= (A^T)^+$ (A has full column rank)

Hence $(A^+)^T = (S^+)^T(T^+)^T \overset{1,2}{=} (S^T)^+(T^T)^+ = (A^T)^+$

c) <u>Claim</u> : AA^+ is symmetric and projection matrix onto $C(A)$

Let $A = CR$ be a CR decomposition of A.

We get $AA^+ = CR(CR)^+ = CRR^+C^+ = CC^+$

$$= C(C^TC)^{-1}C^T$$

$$(C(C^TC)^{-1}C^T)^T = (C^T)^T((C^TC)^{-1})^T C^T = C((C^TC)^T)^{-1}C^T$$

$= C(C^TC)^{-1}C^T$, hence AA^+ is symmetric.

Per Theorem 4.2.6., this is the projection matrix onto $C(C) = C(A)$.

9.2. $f: C(A^T) \longrightarrow C(A)$, $x \mapsto Ax$

<u>Claim</u> : f is bijective

Alternative to master solution : Per definition we prove that f is injective and surjective.

<u>injectivity</u>:

Let $x, y \in C(A^T)$ such that $Ax = Ay$. Then $A(x-y) = 0$, $(x-y) \in N(A)$. Per assumption and as $C(A^T)$ is closed under addition, $(x-y) \in C(A^T)$. As $N(A) \cap C(A^T) = \{0\}$, $x-y=0$, $x=y$.

Per definition, f is injective.

<u>surjectivity</u>:

Prop 4.3.2 applied on A^T gives us $C(A) = C(AA^T) = f(C(A^T))$.

Per definition, f is surjective.

Injective and surjective linear maps

Recall: Let X, Y be vector spaces over some field $\mathbb{F}$:

$f: X \to Y$ is a linear map if:

1. $f(x_1 + x_2) = f(x_1) + f(x_2)$ for any $x_1, x_2 \in X$
2. $f(\alpha x_1) = \alpha f(x_1)$ for any $\alpha \in \mathbb{F}, x_1 \in X$

<u>injective:</u> no information lost

$x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$ or $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$

<u>surjective:</u> all elements of codomain reached

$f(X) = Y$, for any $y \in Y$ there is $x \in X$ such that $f(x) = y$

<u>bijective:</u> injective and surjective $\Leftrightarrow f^{-1}$ exists

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

injective if $\text{rank}(A) = n$ (full column rank)
surjective if $\text{rank}(A) = m$ (full row rank)
bijective if square, $\text{rank}(A) = n = m$

<u>Comments:</u>

- function composition of linear maps corresponds to matrix multiplication: cf $(f \circ g)(x)$, ABx
- Example: Coordinate vector regarding some basis

Consider $v = 5 + 6x - 3x^2 \in P_2$. The coordinate representation of v regarding the basis $\mathcal{B} = \{1, x, x^2\}$ is given by

$[v]_{\mathcal{B}} = \begin{bmatrix} 5 \\ 6 \\ -3 \end{bmatrix}$ $\rightarrow$ tells us how to combine basis vectors to get v.

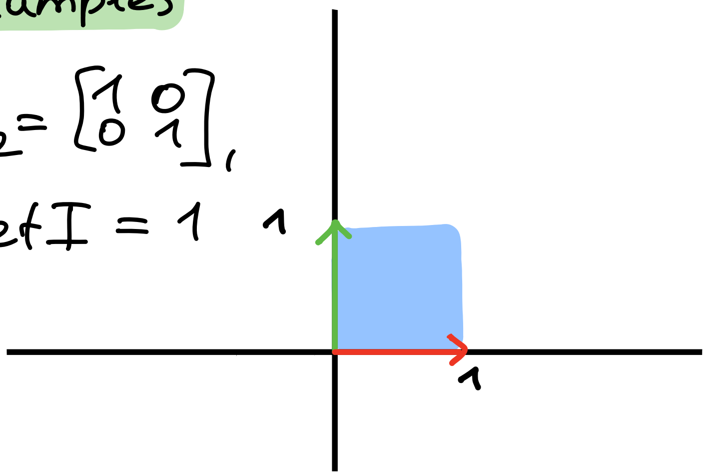
The determinant

We consider $\det A = D(a_1, \dots, a_n)$ where $a_1, \dots, a_n \in \mathbb{R}^n$ are the columns of $A \in \mathbb{R}^{n \times n}$ as a function $\mathbb{R}^{n \times n} \longrightarrow \mathbb{R}$ that gives us the oriented volume of the n-dimensional parallelogram spanned by $a_1, \dots, a_n$.

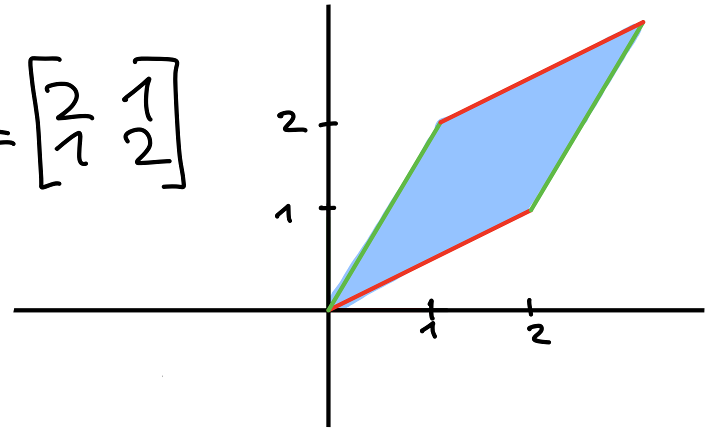
Examples

$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,

$\det I = 1$



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



Most important properties

1) Linear in each column / row
2) Swapping columns / rows changes sign
3) $\det A$ doesn't change if we add multiple of one column (row) to another
4) $\det A \neq 0 \iff A$ is invertible
5) $\det A = \det A^T$
6) $\det(AB) = \det(A)\det(B)$
7) $\det(\alpha A) = \alpha^n \det(A)$ $\longrightarrow$ n times linearity per column / row
8) determinant of triangular matrix is product of diagonal entries
   $\searrow$ in particular of diagonal matrix

Computing the determinant

1) Laplace-Expansion → benefits from rows / columns that have many zeros

expand

k^{th} row: $\det A = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det A[k,j]$

ℓ^{th} column: $\det A = \sum_{j=1}^{n} (-1)^{j+\ell} a_{j\ell} \det A[j,\ell]$

general formula for 2x2 matrices:

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

expansion along first row:

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

2) Gauss-Elimination → used in practice

$$\det A = (-1)^{\underbrace{r}_{=\#\text{row swaps}}} \underbrace{\prod_{i=1}^{n} r_{ii}}_{\text{product of pivots in REF(A)}}$$

References:
Last years course for some definitions
Sergey Treil, Linear Algebra Done Wrong, https://www.math.brown.edu/streil/papers/LADW/LADW_2021_01-11.pdf