Linear Algebra I

Fall 2023

General description of the Gram-Schmidt alg.

Given: Basis $\mathcal{B} = (b_1, \dots, b_m)$ of U.

Want construct ONB $\mathcal{F} = (f_1, \dots, f_m)$.

Step 1: Set $f_1 = \hat{b}_1 = \frac{1}{\|b_1\|} b_1$.

$U_1 = \operatorname{span}\{f_1\} = \operatorname{span}\{b_1\}$.

Step ℓ: Have ON $(f_1, \dots, f_{\ell-1})$, $U_{\ell-1} = \operatorname{span}\{f_1, \dots, f_{\ell-1}\}$

$2 \leq \ell \leq m$

$= \operatorname{span}\{b_1, \dots, b_{\ell-1}\}$.

(i.e. $b_\ell \notin U_{\ell-1}$)

We can write $b_\ell = \underset{\in U_{\ell-1}}{(b_\ell)_{\parallel}} + \underset{\in U_{\ell-1}^{\perp}}{(b_\ell)_{\perp}}$

Set $w_\ell = (b_\ell)_\perp = b_\ell - (b_\ell \cdot f_1) f_1 - \dots - (b_\ell \cdot f_{\ell-1}) f_{\ell-1}$

$f_\ell = \hat{w}_\ell = \frac{1}{\|w_\ell\|} w_\ell$.

$\Rightarrow$ $(f_1, \dots, f_\ell)$ ON, $U_\ell = \operatorname{span}\{f_1, \dots, f_\ell\}$

$= \operatorname{span}\{b_1, \dots, b_{\ell-1}, f_\ell\}$

$= \operatorname{span}\{b_1, \dots, b_\ell\}$.

After m steps $U_m = \operatorname{span}\{f_1, \dots, f_m\}$
$$= \operatorname{span}\{b_1, \dots, b_m\} = U.$$

$\Rightarrow \quad F = (f_1, \dots, f_m)$ ONB of U.

Theorem 12.8 Every subspace has an ONB.

Proof: Every subspace has a basis (Thm. 10.4).
Using GSA we get an ONB. $\square$

Corollary 12.9 Let $U \subset \mathbb{R}^n$ be a subspace.
For all $x \in \mathbb{R}^n$ there exist unique $x_{\parallel} \in U$ and $x_{\perp} \in U^{\perp}$ with

$$x = x_{\parallel} + x_{\perp}.$$

Proof: Existence: By Thm. 12.6 there exists an ONB $(f_1, \dots, f_m)$ of U. And by Lemma 12.5 we get $x_{\parallel}$ and $x_{\perp}$.

Uniqueness: Let $x = x_{\parallel} + x_{\perp} = y_{\parallel} + y_{\perp}$ for $x_{\parallel}, y_{\parallel} \in U$, $x_{\perp}, y_{\perp} \in U^{\perp}$.

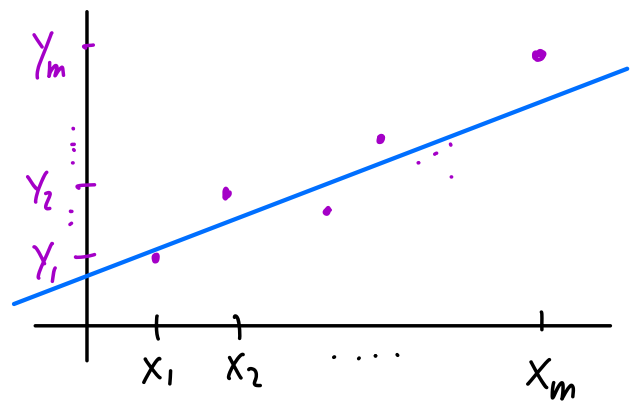
$\Rightarrow \quad U \ni x_{\parallel} - y_{\parallel} = x_{\perp} - y_{\perp} \in U^{\perp}$

$$\Rightarrow \quad X_\parallel - Y_\parallel = X_\perp - Y_\perp \underset{\substack{\uparrow \\ \text{Lemma 12.5}}}{=} 0 \Rightarrow X_\parallel = Y_\parallel \text{ and } X_\perp = Y_\perp$$

$\square$

§ 13 Orthogonal projection & Least squares

Motivation: Assume you measure some data $(x_1, y_1), \dots, (x_m, y_m)$



and you want to find a line which interpolates these points in the best possible way.

If all points would lie on a line $\ell(x) = ax + b$ then they would satisfy

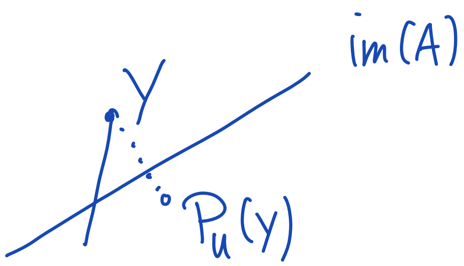
$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \\ &\vdots \\ ax_m + b &= y_m \end{aligned} \quad \rightsquigarrow \quad \underbrace{\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}}_{A} \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}}_{y}$$

$$\Leftrightarrow \quad A \begin{pmatrix} a \\ b \end{pmatrix} = y. \quad (*)$$

But if they are not on one line (like in the picture), then the linear system $(*)$ has no solutions because $y \notin \operatorname{im}(A)$.

But in the picture we see that there might be a "best possible" line.

Main idea: Project y onto the image of A.
$\rightsquigarrow$ obtain a linear system we can solve for $\binom{a}{b}$.



$Ax = P_u(y)$ has a solution x.

Will see: x can be obtained by solving the normal equation

$$A^T A x = A^T y$$

<u>Definition 13.1</u> Let $U \subset \mathbb{R}^n$ be a subspace.
The map

$$P_U : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$
$$x \longmapsto x_{\parallel}$$

from Cor. 12.9

is the <u>orthogonal projection</u> onto U.

Notice: This generalizes the P_u for $u \in \mathbb{R}^n$, $u \neq 0$ we defined before by setting $U = \mathrm{span}\{u\}$.

Proposition 13.2 Let $U \subset \mathbb{R}^n$ be a subspace

i) P_U is a linear map.

ii) $P_U^2 = P_U$.

iii) $\operatorname{Ker}(P_U) = U^\perp$ and $\operatorname{im}(P_U) = U$.

iv) If $(f_1, \dots, f_m)$ is an ONB of U then

$$P_U(x) = (x \cdot f_1) f_1 + \dots + (x \cdot f_m) f_m$$

for all $x \in \mathbb{R}^n$.

Proof: iv) is exactlly Lemma 12.5 iv).

i), ii), iii) are direct consequences of iv). (Check!).

+Lemma 12.5 $\square$

Proposition 13.3 Let $U \subset \mathbb{R}^n$ be a subspace and $x \in \mathbb{R}^n$.
Then for all $u \in U$ we have

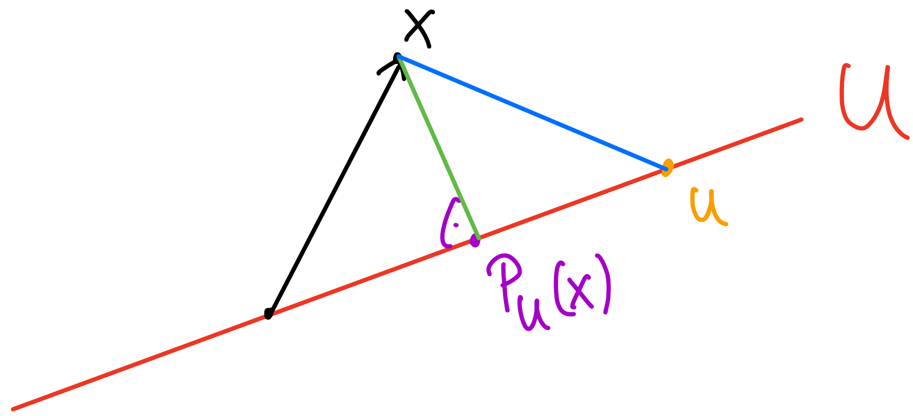
$$\| x - P_U(x) \| \leq \| x - u \|.$$

We just have "=" in the case $u = P_U(x)$.

In other words: If x is outside of U, then $P_U(x) \in U$ is the closest point to x which is in U.

Proof: We will not give a proof, but the statement should be clear by considering the following picture.

See lecture notes



$$\|x - P_U(x)\| \leq \|x - u\| .$$

Definition 13.4 The *transpose* of a matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ is the matrix $A^T = (a_{ji}) \in \mathbb{R}^{n \times m}$.

Example:
$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in \mathbb{R}^{2 \times 3}$$
$$A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \in \mathbb{R}^{3 \times 2}.$$

Proposition 13.5 i) For $A, B \in \mathbb{R}^{m \times n}$ and $\lambda \in \mathbb{R}$ we have
$$(A+B)^T = A^T + B^T, \qquad (\lambda A)^T = \lambda A^T.$$
ii) For $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times \ell}$ we have
$$(AB)^T = B^T A^T \in \mathbb{R}^{\ell \times m}$$

iii) For $x, y \in \mathbb{R}^n$ we have $x \cdot y = x^T y$.

Proof: Can be checked by direct calculation. $\square$

For $A \in \mathbb{R}^{m\times n}$ we can define a linear map $F: x \mapsto Ax$.
We write $\operatorname{im}(A) := \operatorname{Ker}(F)$ and $\operatorname{Ker}(A) := \operatorname{Ker}(F)$.

As mentioned in the motivation at the begining, we are intererted in projecting onto the image of A. For this we will first try to understand its orthogonal complement.

Proposition 13.6 For all $A \in \mathbb{R}^{m\times n}$ we have

$$\operatorname{im}(A)^{\perp} = \operatorname{Ker}(A^T).$$

Proof: Let $x \in \mathbb{R}^n$. Then we have

$$
\begin{aligned}
x \in (\operatorname{im} A)^{\perp} &\iff y \cdot x = 0 \quad \underline{\forall y \in \operatorname{im}(A)} \qquad y = Av \text{ for some } v \in \mathbb{R}^n \\
&\iff (Av) \cdot x = 0 \quad \forall v \in \mathbb{R}^n \\
&\overset{\text{Prop. 13.5 iii)}}{\iff} (Av)^T x = 0 \quad \forall v \in \mathbb{R}^n \\
&\overset{\text{Prop. 13.5 ii)}}{\iff} v^T A^T x = 0 \quad \forall v \in \mathbb{R}^n \\
&\iff v \cdot A^T x = 0 \quad \forall v \in \mathbb{R}^n \\
&\iff A^T x \in \underline{(\mathbb{R}^n)^{\perp}} = \{0\} \\
&\iff A^T x = 0 \\
&\iff x \in \operatorname{Ker}(A^T).
\end{aligned}
$$

$\square$

Corollary 13.7 Let $A \in \mathbb{R}^{m \times n}$.

i) We have $\operatorname{Ker}(A^T A) = \operatorname{Ker}(A)$.

ii) The following statements are equivalent

$$\operatorname{Ker}(A) = \{0\} \iff A^T A \in \mathbb{R}^{n \times n} \text{ is invertible.}$$

Proof: i) We have

$$x \in \operatorname{Ker}(A^T A) \iff A^T A x = 0 \iff \underbrace{Ax}_{\in \operatorname{im}(A)} \in \operatorname{Ker}(A^T) \overset{\text{Prop 13.6}}{=} \operatorname{im}(A)^\perp$$

$$\iff Ax \in \operatorname{im}(A) \cap \operatorname{im}(A)^\perp \underset{\text{Lemma 12.5}}{=} \{0\}$$

$$\iff Ax = 0$$

$$\iff x \in \operatorname{Ker}(A).$$

ii)

$$\operatorname{Ker}(A) = \{0\} \iff \operatorname{Ker}(A^T A) = \{0\}$$

$$\underset{\text{Thm. 8.7}}{\iff} A^T A \text{ is invertible} \qquad \square$$

$A^T A \in \mathbb{R}^{m \times m}$

$\operatorname{Ker}(A^T A) = \{0\} \Rightarrow$ $\mathbb{R}^m \to \mathbb{R}^m$, $x \mapsto A^T A x$ is injective

$\iff$ bijective

$n=m$

We can now use our results to answer the question in the motivation.

Least squares method

Problem: Given a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ and $b \in \mathbb{R}^m$.

Find $x \in \mathbb{R}^n$ that minimizes

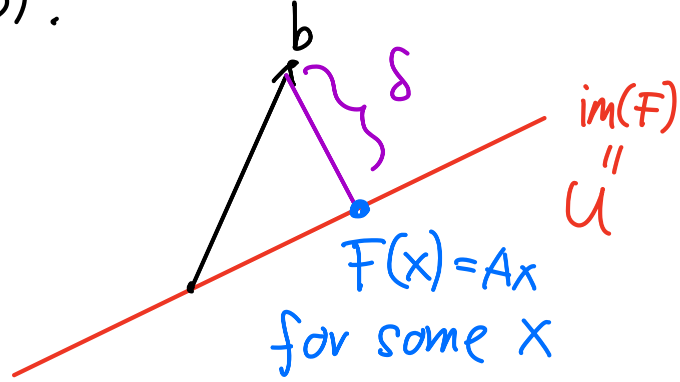
$$\delta = \| F(x) - b \| .$$

("Least squares" because if $F(x) = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$, then $\| F(x) - b \| = \sqrt{(y_1 - b_1)^2 + \dots + (y_m - b_m)^2}$.

We want to minimize the sum of squares of the differences $y_i - b_i$.)

Notice: Minimal δ is $0 \iff b \in \mathrm{im}(F)$, i.e. $F(x) = b$ has a solution.

By Proposition 13.3 the minimal δ is given in the case $F(x) = P_{\mathrm{im}(F)}(b)$.



Write $[F] = A$, then we want to find $x \in \mathbb{R}^n$ such that $Ax = P_{im(F)}(b)$

$$\iff Ax - b \in (\operatorname{im} A)^{\perp} = \operatorname{Ker}(A^T)$$

$$\iff A^T(Ax - b) = 0$$

$$\iff \boxed{A^T A x = A^T b}$$

Normal equation.

Therefore if $\operatorname{Ker}(A) = \{0\}$ (i.e. the columns of A are lin. indep.)

then (by Corollary 13.7)

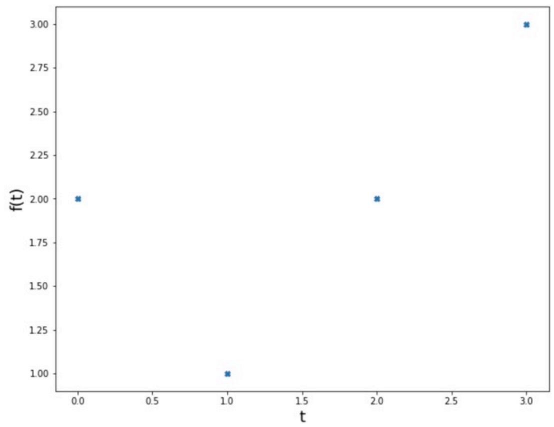
$A^T A$ is invertible and we get the unique solution to our problem by

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

Example 44 (Polynom interpolation)

Find the best possible quadratic polynomial

$$f(t) = a_0 + a_1 t + a_2 t^2$$ to fit the data points $(0,2)$, $(1,1)$, $(2,2)$, $(3,3)$.



We first translate this problem into linear algebra:

We want to minimize

$$(f(0)-2)^2 + (f(1)-1)^2 + (f(2)-2)^2 + (f(3)-3)^2.$$

So we define the linear map

$$F: \mathbb{R}^3 \longrightarrow \mathbb{R}^4$$

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \longmapsto \begin{pmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 0 + a_2 0 \\ a_0 + a_1 \cdot 1 + a_2 \cdot 1^2 \\ a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 \\ a_0 + a_1 \cdot 3 + a_2 \cdot 3^2 \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}}_{x} = Ax$$

We want to find $x \in \mathbb{R}^3$ such that $\|Ax-b\|$ with $b = \begin{pmatrix} 2 \\ 1 \\ 2 \\ 3 \end{pmatrix}$ is minimal.

$\Rightarrow$ We need to solve the normal equation

$$\underline{A^T A} x = \underline{A^T b}$$

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix}, \quad A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} = \underline{\begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{pmatrix}}$$

$$A^T b = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 2 \\ 3 \end{pmatrix} = \underline{\begin{pmatrix} 8 \\ 14 \\ 36 \end{pmatrix}}$$

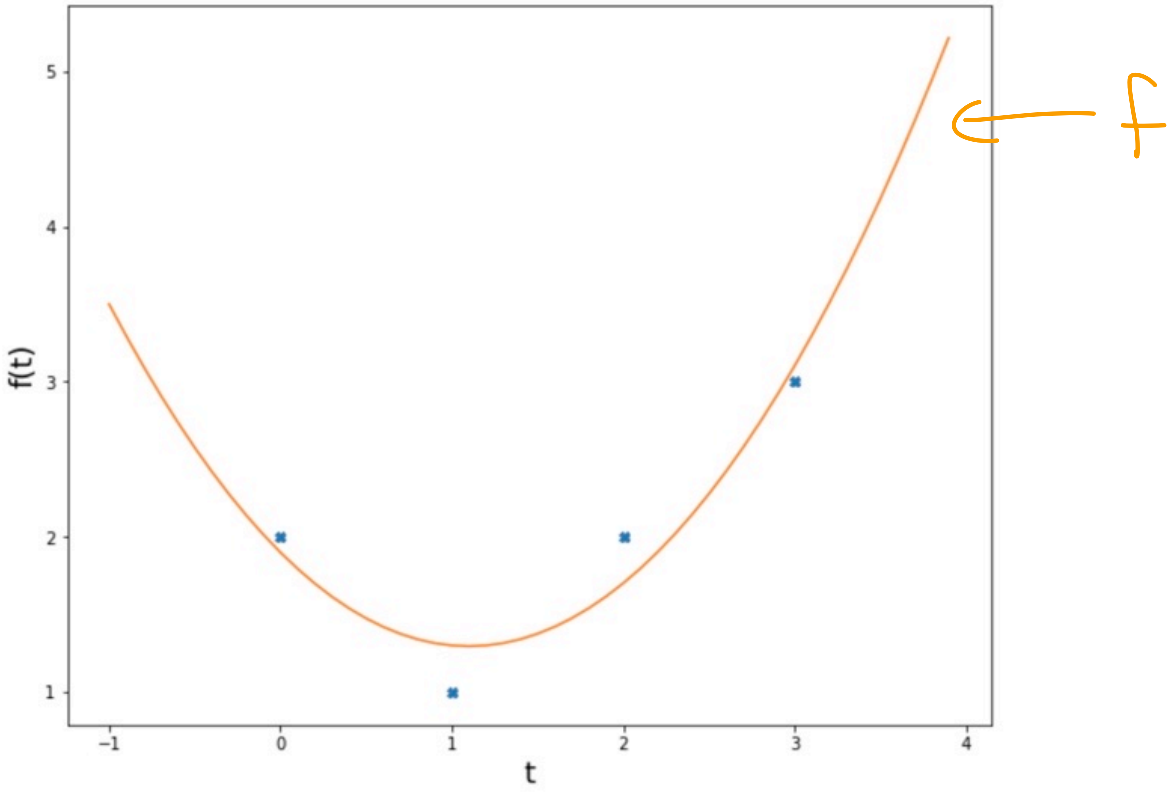
$\Rightarrow$ Want to solve

$$\underline{\begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{pmatrix}} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \underline{\begin{pmatrix} 8 \\ 14 \\ 36 \end{pmatrix}}$$

Unique solution is

$$a_0 = \frac{19}{10}$$
$$a_1 = -\frac{11}{10}$$
$$a_2 = \frac{1}{2}$$

$\Rightarrow$ Best fit polynomial is $f(t) = \frac{19}{10} - \frac{11}{10}t + \frac{1}{2}t^2$



Notice: This works for arbitrary polynomials (i.e. in particular for lines).

Always have a unique solution if the columns of A are lin. indep.

In applications this is usually the case since usually $\underbrace{\#\text{data points}}_{\substack{=m \\ =\text{rows of } A}} > \underbrace{\text{degree of polynomial}}_{\substack{=n-1 \\ =\text{columns of } A - 1}}$

If you read this then you finished reading all the content for Linear algebra I. Congratulations! ☺

Hope to see you in LA II!