Warm up

- $S = \left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} \right\} \subseteq \mathbb{R}^4$

  Is S orthogonal? Is it orthonormal? Is it a basis?

  ↳ yes! ↳ no! ↳ is not a basis of $\mathbb{R}^4$

  ↳ but is a basis of $W := \mathrm{Span}\left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} \right\}$

- $A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{m \times n}$

  Is A an orthogonal matrix?

  Is $f(x) := Ax$ injective? surjective? bijective?

  ↳ yes ↳ no ↳ no

- True / False:
  - Every orthogonal set of vectors is independent. False
  - Every orthonormal set of vectors is independent. True
  - Every independent set is orthogonal. False
  - An orthogonal projection onto a linear subspace is a linear map. True

→ see HW 12 for proofs.

Recall If $W \subseteq \mathbb{R}^n$ subspace and $x \in \mathbb{R}^n$

$\Rightarrow \exists$ unique $w \in W$ and $\tilde{w} \in W^\perp$ s.t. $x = w + \tilde{w}$

$\mathrm{proj}_W(x) := w$ the orthogonal projection of x to the subspace W.

Formula for computation: $\mathrm{proj}_W(x) := \frac{x \cdot b_1}{b_1 \cdot b_1} b_1 + \dots + \frac{x \cdot b_k}{b_k \cdot b_k} \cdot b_k$

where $B = \{b_1, \dots, b_k\}$ is an orthog. basis of W.

Thm. 6.8 Let W be a subspace of $\mathbb{R}^n$ and $\{u_1, \dots, u_k\}$ an ONB of W. Let $U = (u_1, \dots, u_k) \in \mathbb{R}^{n \times k}$. Then:

$$\mathrm{proj}_W(x) = (x \cdot u_1) u_1 + \dots + (x \cdot u_k) u_k = U U^T x$$

Proof Since $\{u_1, \dots, u_k\}$ is orthonormal: $1 = \|u_j\|^2 = u_j \cdot u_j$.

This justifies the first equality.

Next, observe that

$$U U^T x = U(U^T x) = U \cdot \left(\underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix}}_{\text{matrix}} \underbrace{\begin{pmatrix} x \end{pmatrix}}_{\text{vector}} \right) = U \cdot \begin{pmatrix} u_1 \cdot x \\ \vdots \\ u_k \cdot x \end{pmatrix}$$

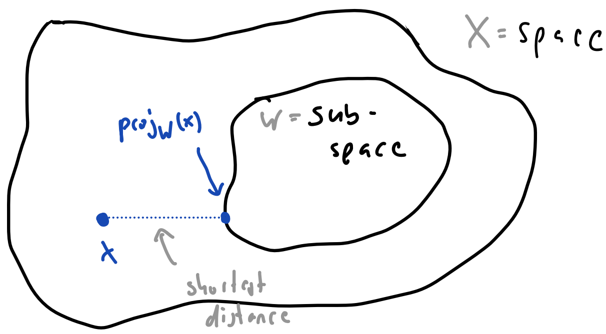
$$= (u_1 \cdot x) u_1 + \dots + (u_k \cdot x) u_k \qquad \square$$

Hence, orthogonal projections onto a subspace are easy to compute, if we have an orthogonal or orthonormal basis for the subspace $\rightarrow$ Motivation for Section 6.4 (next section).

Orthogonal projection as closest point

Often in math (outside of linear algebra) when a function is referred to as "a projection of a space onto a subspace" one means that we map a point x to the closest point in the subspace.

Schematically:



Orthogonal projections in $\mathbb{R}^n$ as defined have this property as well

(Compare the picture at the beginning of section 6.2).

Thm. 6.9 Let W be a subspace of $\mathbb{R}^n$. Let $x \in \mathbb{R}^n$.

Then $\| x - proj_W(x) \| \leq \| x - w' \| \quad \forall w' \in W.$

Moreover, the closest point is unique.

In other words: among all the $w' \in W$, $proj_W(x)$ who also is an element of W is the closest to x

Proof. Let $w' \in W$:

$$\|x - w'\|^2 = \|\underbrace{x - w}_{\in W^\perp} + \underbrace{w - w'}_{\in W}\|^2 \quad \text{where } w = proj_W(x)$$

Recall $x = w + \tilde{w}$

$$\overset{\text{Thm. 6.1}}{=} \|x - w\|^2 + \underbrace{\|w - w'\|^2}_{\geq 0} \geq \|x - w\|^2$$

$$\Rightarrow \|x - w'\|^{\cancel{2}} \geq \|x - w\|^{\cancel{2}}$$

$$\Rightarrow \|x - w'\| \geq \|x - proj_W(x)\|$$

Uniqueness: assume that $\|x - w\| = \|x - w'\|$ for some $w' \in W$.

$$\Rightarrow \|x - w\|^2 = \|x - w'\|^2 \Rightarrow \|w - w'\|^2 = 0 \Rightarrow w = w'.$$

As in above computation

$\square$

Justified by Thm. 6.9, we define:

Def. For $W \subseteq \mathbb{R}^n$ subspace and $x \in \mathbb{R}^n$: $dist(x, W) = \|x - proj_W(x)\|$

Ex. Compute the distance between $x = e_1$ and W from the previous example.

$$proj_W(e_1) = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 1/3 \\ 1/3 \end{pmatrix}$$

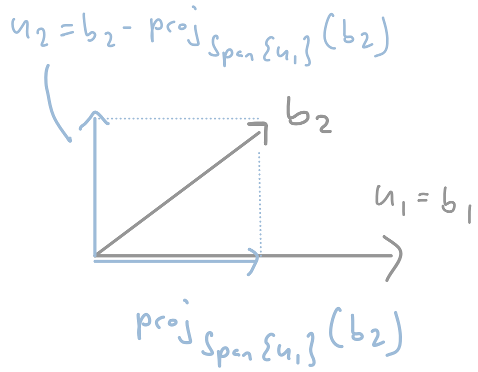
$$\Rightarrow dist(e_1, W) = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 2/3 \\ 1/3 \\ 1/3 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1/3 \\ -1/3 \\ -1/3 \end{pmatrix} \right\| = \frac{\sqrt{3}}{3}$$

6.4 The Gram-Schmidt process

The Gram-Schmidt process is an algorithm that transforms any basis of $\mathbb{R}^n$ or of $W \subseteq \mathbb{R}^n$ into an orthogonal basis.

Thm. 6.10 (Gram-Schmidt)

Let W be a subspace of $\mathbb{R}^n$ and $\mathcal{B} = \{b_1, \dots, b_k\}$ a basis for W. We define:



$$u_1 := b_1$$
$$u_2 := b_2 - \frac{b_2 \cdot u_1}{u_1 \cdot u_1} u_1$$
$$u_3 := b_3 - \frac{b_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{b_3 \cdot u_2}{u_2 \cdot u_2} \cdot u_2$$
$$\vdots$$
$$u_k := b_k - \frac{b_k \cdot u_1}{u_1 \cdot u_1} u_1 \quad \dots \quad - \frac{b_k \cdot u_{k-1}}{u_{k-1} \cdot u_{k-1}} \cdot u_{k-1}$$

in other words: $u_i := b_i - \text{proj}_{\text{Span}\{u_1, \dots u_{i-1}\}}(b_i)$

Then: (1) $\{u_1, \dots, u_k\}$ are orthogonal

(2) $\text{Span}\{u_1, \dots, u_i\} = \text{Span}\{b_1, \dots, b_i\} \quad \forall i$

In particular, $\{u_1, \dots, u_k\}$ is an orthogonal basis of W.

Proof For every $k \in \{1, \dots k\}$ define $W_k := \operatorname{Span}\{b_1, \dots, b_k\}$

Claim $\{u_1 \dots u_k\}$ is an orthog. basis of W_k.

Induction: $k=1$ $\{u_1\}$ is orthogonal and $u_1 = b_1$

$\Rightarrow \operatorname{Span}\{u_1\} = W_1$ and $\{u_1\}$ an orthog. basis

$k-1 \rightsquigarrow k$ Assume $\{u_1, \dots, u_{k-1}\}$ is an orthog. basis of W_{k-1}

$\Rightarrow u_k = b_k - \operatorname{proj}_{W_{k-1}}(b_k) \Rightarrow u_k \in W_{k-1}^{\perp}$

$\Rightarrow \{u_1, \dots, u_{k-1}, u_k\}$ is orthogonal

$u_k \neq 0 \Rightarrow \{u_1, \dots, u_{k-1}, u_k\}$ lin indep.

because $u_k = 0$ would imply $b_k = \operatorname{proj}_{W_{k-1}}(b_k) \in W_{k-1}$

$\Rightarrow b_1 \dots b_k$ are not indep $\lightning$

$u_k = \underbrace{b_k}_{\in W_k} - \underbrace{\operatorname{proj}_{W_{k-1}}(b_k)}_{\in W_{k-1} \subset W_k} \in W_k \Rightarrow \{u_1, \dots, u_{k-1}, u_k\}$ is a basis of W_k.

So the claim is proven and the case $k=k$ proves the Thm. $\square$

Cor. 6.11 Every non-zero subspace of $\mathbb{R}^n$ has an ONB.

Proof We know: every subspace of $\mathbb{R}^n$ has a basis

$\underset{\text{Thm 6.10}}{\Rightarrow}$ it has an orthogonal basis $\underset{\text{normalize}}{\Rightarrow}$ it has an ONB. $\square$

$\underline{\text{Ex.}}$ Find an orthogonal basis for the subspace W of $\mathbb{R}^4$ that is spanned by

$$\mathcal{B} = \left\{ \underbrace{\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}}_{=b_1}, \underbrace{\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}}_{=b_2}, \underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}}_{=b_3} \right\}$$

Gram-Schmidt:

$$u_1 = b_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} \implies u_1 \cdot u_1 = \|u_1\|^2 = 3$$

$$u_2 = b_2 - \frac{b_2 \cdot u_1}{u_1 \cdot u_1} u_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \frac{\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}}{3} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \\ 0 \end{pmatrix}$$

$$\implies \|u_2\|^2 = u_2 \cdot u_2 = \frac{2}{3}$$

$$u_3 := b_3 - \frac{b_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{b_3 \cdot u_2}{u_2 \cdot u_2} \cdot u_2 = \dots$$

$$= \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix} - 0 \cdot \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} - \frac{3}{2} \begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \\ 2 \end{pmatrix}$$

Thm 6.10

$$\implies \{u_1, u_2, u_3\} = \left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \\ 0 \end{pmatrix} \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \\ 2 \end{pmatrix} \right\}$$ is an orthogonal basis of W.

Q: Is the orthogonal basis we found also orthonormal?

→ No! But we can find an ONB for W by normalizing the vectors u_1, u_2, u_3 of our orthogonal basis:

We have already computed $\|u_1\|$ and $\|u_2\|$ during Gram-Schmidt.

$\|u_3\| = u_3 \cdot u_3 = 3$

$$v_1 = \frac{1}{\|u_1\|} u_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$$

$$v_2 = \frac{1}{\|u_2\|} u_2 = \frac{\sqrt{3}}{\sqrt{2}} \begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}$$

$$v_3 = \frac{1}{\|u_2\|} u_2 = \frac{\sqrt{2}}{3} \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \\ 2 \end{pmatrix}$$

$$\Rightarrow \{v_1, v_2, v_3\} = \left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \frac{\sqrt{2}}{3} \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \\ 2 \end{pmatrix} \right\}$$

is an ONB of W.

QR-factorization of matrices

Motivation: Numerical purposes.

<u>Thm. 6.11</u> Let $A \in \mathbb{R}^{m \times n}$ have linearly indep. columns.

Then, one can write A as a product of matrices

$A = QR$ such that:

- $Q \in \mathbb{R}^{m \times n}$ and its columns are an ONB of $Col(A)$
- $R \in \mathbb{R}^{n \times n}$ is upper triangular and all diagonal entries are positive.

<u>Proof.</u> Apply Gram-Schmidt to the columns $A_1, \dots, A_n$ of A and normalize the output

$\Rightarrow \{\tilde{q}_1, \dots, \tilde{q}_n\}$ ONB for $Col(A)$

Next we adjust the signs
- If $A_j \cdot \tilde{q}_j \geq 0$ then $q_j = \tilde{q}_j$
- If $A_j \cdot \tilde{q}_j < 0$ then $q_j = -\tilde{q}_j$

$\Rightarrow \{q_1, \dots, q_n\}$ ONB for $Col(A)$ and $A_j q_j \geq 0 \; \forall j$

Define $Q = (q_1 \dots q_n) \in \mathbb{R}^{m \times n}$

columns of the matrix Q

Let $R = \begin{pmatrix} r_{11} & \cdot & \cdots & r_{1n} \\ 0 & r_{22} & & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \vdots & & r_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n}$

$\begin{array}{ccc} R_1 & R_2 & R_n \end{array}$

$QR = (QR_1, \dots, QR_n)$ ⇝ the QR_j are the columns of QR.

$QR_j = r_{1j} \cdot q_1 + \dots + r_{jj} q_j$ $(*)$ for each j : this sum has j terms

We want $A_j = QR_j$ since we want $A = QR$.

Is this possible ?

→ Yes since indeed $A_j \in \operatorname{Span}\{q_1 \dots q_j\}$ by Gram-Schmidt

What are the coefficients r_{ij} in the linear combination $(*)$?

Since the q_i are orthonormal, by Thm 6.5 : $r_{ij} = A_j \cdot q_i$

So $R := \begin{pmatrix} A_1 \cdot q_1 & A_2 \cdot q_1 & \cdots & A_n \cdot q_1 \\ 0 & A_2 \cdot q_2 & & A_n \cdot q_2 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_n \cdot q_n \end{pmatrix} \in \mathbb{R}^{n \times n}$

We are almost done : we have $QR = A$ and $r_{ii} = A_i \cdot q_i \geq 0$

But we need > 0. So, can $A_k \cdot q_k$ be $= 0$?

⇝ No ! Because if $A_k \cdot q_k = 0$ then by $(*)$ we have

$A_k \in \operatorname{Span}\{q_1 \dots q_{k-1}\} = \operatorname{Span}\{A_1 \dots A_{k-1}\}$

This is a contradiction to $\{A_1 \dots A_n\}$ independent. □

Rmk Notice that the proof has given us a method to find the QR-factorization of a matrix A:

- Gram-Schmidt for columns of A
- normalize
- adjust signs

$\to \{q_1 \dots q_n\}$ ONB for $\mathrm{Span}\{A_1 \dots A_k\}$

$$\to \quad Q := (q_1 \dots q_n) \quad \text{and} \quad R := \begin{pmatrix} A_1\cdot q_1 & A_2\cdot q_1 & \cdots & A_n\cdot q_1 \\ & A_2\cdot q_2 & & A_n\cdot q_2 \\ & & \ddots & \vdots \\ & & & A_n\cdot q_n \end{pmatrix}$$

Ex. $A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ $\quad$ **Q**: Find a QR-factorization for A.

Gram-Schmidt: see previous example

$$\Rightarrow \quad u_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \\ -4 \end{pmatrix}$$

Normalize see previous example

$$\tilde{q}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{q}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{q}_2 = \frac{\sqrt{2}}{3}\begin{pmatrix} 1/2 \\ 0 \\ -1/2 \\ 2 \end{pmatrix}$$

Signs

$$A_1 \cdot \tilde{q}_1 = \begin{pmatrix}1\\-1\\1\\0\end{pmatrix} \cdot \frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\\0\end{pmatrix} = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\\0\end{pmatrix}\cdot\begin{pmatrix}1\\-1\\1\\0\end{pmatrix} = \frac{3}{\sqrt{3}} = \underbrace{\sqrt{3}}_{\geq 0} \Rightarrow q_1 := \tilde{q}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\\0\end{pmatrix}$$

$$A_2 \cdot \tilde{q}_2 = \begin{pmatrix}1\\0\\1\\0\end{pmatrix} \cdot \frac{1}{\sqrt{6}}\begin{pmatrix}1\\2\\1\\0\end{pmatrix} = \dots = \underbrace{\frac{\sqrt{2}}{\sqrt{3}}}_{\geq 0} \Rightarrow q_2 := \tilde{q}_2 = \frac{1}{2}\begin{pmatrix}1\\2\\1\\0\end{pmatrix}$$

$$A_3 \cdot \tilde{q}_3 = \begin{pmatrix}1\\1\\0\\2\end{pmatrix} \cdot \frac{\sqrt{2}}{3}\begin{pmatrix}1/2\\0\\-1/2\\2\end{pmatrix} = \dots = \frac{\sqrt{2}}{3}\,\frac{9}{2} = \underbrace{\frac{3}{\sqrt{2}}}_{\geq 0} \Rightarrow q_3 := \tilde{q}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix}1/2\\0\\-1/2\\2\end{pmatrix}$$

(So this example happens to have no changes of signs.)

Q and R

$$Q = (q_1, q_2, q_3) = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{2} & \frac{1}{2\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & 1 & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{2} & -\frac{1}{2\sqrt{3}} \\ 0 & 0 & \frac{2}{\sqrt{3}} \end{pmatrix}$$

$$R = \begin{pmatrix} A_1 \cdot q_1 & A_2 \cdot q_1 & A_3 \cdot q_1 \\ 0 & A_2 \cdot q_2 & A_3 \cdot q_2 \\ 0 & 0 & A_3 \cdot q_3 \end{pmatrix} = \begin{pmatrix} \sqrt{3} & \frac{3}{\sqrt{2}} & 0 \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & \frac{3}{\sqrt{6}} \\ 0 & 0 & \frac{3}{\sqrt{2}} \end{pmatrix}$$

From above: $A_1 \cdot q_1 = \sqrt{3}$, $A_2 \cdot q_2 = \frac{\sqrt{2}}{\sqrt{3}}$, $A_3 \cdot q_3 = \frac{3}{\sqrt{2}}$

And we can compute: $A_2 \cdot q_1 = \begin{pmatrix}1\\0\\1\\0\end{pmatrix} \cdot \frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\\0\end{pmatrix} = \frac{2}{\sqrt{3}}$, $A_3 \cdot q_1 = \begin{pmatrix}1\\1\\0\\2\end{pmatrix} \cdot \frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\\0\end{pmatrix} = 0$

$$A_3 \cdot q_2 = \begin{pmatrix}1\\1\\0\\2\end{pmatrix} \cdot \frac{1}{\sqrt{6}}\begin{pmatrix}1\\2\\1\\0\end{pmatrix} = \frac{3}{\sqrt{6}}$$

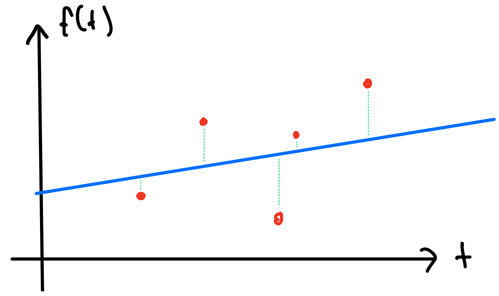
6.5 Least square problems

A main topic of this course has been finding solutions $x \in \mathbb{R}^n$ for linear systems $Ax = b$ $(A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m)$.

Sometimes, such systems do not have solutions.

In particular, if m is (much) larger than n, i.e. if the system has more equations than unknowns.

E.g.



Given: 5 data points

Wanted: an affine fct. describing the data.

Problem: 5 data points but only 2 parameters

↳ straight line

⇝ description is only possible approximately!

⇝ Q: What is a good approximation? How can we find it?

Will there be a unique best approximation?

Our version of approximation will be the one that minimizes the sum of the vertical distances of the data points to the approximation line.

In Section 6.5 we build a theory; in Section 6.6. we apply it to examples as shown above.

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. We know that it can happen that $Ax = b$ has no solution, i.e., $Ax - b = 0$ has no solution.

In that case, we can still ask:

Is there an $x \in \mathbb{R}^n$ so that $Ax - b$ becomes as small as possible?

$Ax - b$ is a vector, so what is small? $\rightsquigarrow$ $\underbrace{\|Ax - b\|}_{\text{length of the vector}}$ is small.

<u>Def.</u> Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}$. A least-square solution of the equation $Ax = b$ is an $x^* \in \mathbb{R}^n$ such that

$$\|Ax^* - b\| \leq \|Ax - b\| \quad \forall x \in \mathbb{R}^n.$$

We call $\|Ax^* - b\|$ the least square error.

Least-square problems can be solved using orthogonal projections:

We know that:
- $\{Ax : x \in \mathbb{R}^n\} = \mathrm{Ran}(A) = \mathrm{Col}(A)$
- the point in $\mathrm{Col}(A)$ that is closest to b is $\mathrm{proj}_{\mathrm{Col}(A)}(b)$.

  i.e. $\|\mathrm{proj}_{\mathrm{Col}(A)}(b) - b\| \leq \|\underbrace{y}_{= Ax} - b\| \quad \forall y \in \mathrm{Col}(A)$

Hence we have proven:

<u>Lemma 6.12</u> An $x^* \in \mathbb{R}^n$ is a least-square solution for $Ax = b$ if and only if $Ax^* = \mathrm{proj}_{\mathrm{Col}(A)}(b)$

Notice: In cases where we are interested in least square solutions, A is not invertible (most times not even square). So we cannot trivially find x^* from $Ax^* = \mathrm{proj}_{Col(A)}(b)$

Thm 6.12 $x^* \in \mathbb{R}^n$ is a least square solution of $Ax = b$ if and only if x^* is a solution for $A^T A x = A^T b$

Moreover, a least square solution always exists.

Proof. $\boxed{\Rightarrow}$ Let x^* be a least square sol. of $Ax = b$.

By Lem. 6.12 $\Rightarrow$ $Ax^* = \mathrm{proj}_{Col(A)}(b)$

$$b - Ax^* = b - \mathrm{proj}_{Col(A)}(b) \underset{\substack{\text{Thm 6.7} \\ \text{(orthogonality of proj.)}}}{\in} Col(A)^\perp \underset{\text{Thm 6.3}}{=} Ker(A^T)$$

$$\Rightarrow A^T(b - Ax^*) = 0 \Rightarrow A^T b = A^T A x^*$$

$\boxed{\Leftarrow}$ Now assume $A^T b = A^T A x^*$

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

$$\Rightarrow b - Ax^* \in Ker(A^T) = Col(A)^\perp$$

$$b = \underbrace{b - Ax^*}_{\in Col(A)^\perp} + \underbrace{Ax^*}_{\in Col(A)}$$

$$\underset{\text{Thm 6.7}}{\Rightarrow} Ax^* = \mathrm{proj}_{Col(A)} b \overset{\text{Lem. 6.12}}{\Rightarrow} x^* \text{ is a least square sol.}$$

<u>Existence</u> : $\tilde{b} := \text{proj}_{\text{Col}(A)} b \in \text{Col}(A) = \text{Ran}(A)$

$\Rightarrow \exists x^* \in \mathbb{R}^n$ s.t. $Ax^* = \tilde{b} = \text{proj}_{\text{Col}(A)} b$ $\square$

Definition of Ran(A)

<u>Ex.</u> $A = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}$ $b = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}$ Q : Find all least square solutions.

(Recall : A does not need to be square! Actually it usually has more rows than columns, i.e., "more data points then parameters")

x^* is a least square solution of the equation $\underset{\text{matrix}}{A} x = \underset{\text{vector}}{b}$

$\Leftrightarrow$ x^* is a (actual) solution of the system $\underbrace{(A^T A)}_{\text{matrix}} x = \underbrace{A^T b}_{\text{vector}}$

Thm. 6|2

$$A^T A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix}, \quad A^T b = \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}$$

So we need to find the solution space of $\begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6 \\ -2 \end{pmatrix}$ (*)

Since $\begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix}$ is invertible, (*) has a unique solution.

We can guess or compute it : $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -2/5 \end{pmatrix}$

$\Rightarrow$ $x^* = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -2/5 \end{pmatrix}$ is the only least square solution of the equation $Ax = b$.

Rmk $A^T A$ is always square but not always invertible.

If $A^T A$ is not invertible, then there exist several least square sol. for A:

Thm. 6.13 Let $A \in \mathbb{R}^{m \times n}$. The following are equivalent:

(i) For every $b \in \mathbb{R}^m$, the equation $Ax = b$ has a unique least square solution.

(ii) $A^T A$ invertible

(iii) The columns of A are lin. indep.

Proof. $\rightarrow$ HW13

Hint: (i) $\Leftrightarrow$ (ii) : use Thm. 6.12

(i) $\Rightarrow$ (iii) : a brief computation

(iii) $\Rightarrow$ (ii) : a brief computation

We can also find least square solutions using QR-decomposition:

Thm 6.14 Let $A \in \mathbb{R}^{m \times n}$ be a matrix with lin. indep. columns and let $A = QR$ be its QR-fact. Then for each $b \in \mathbb{R}^m$ the unique least square solution of $Ax = b$ is: $x^* = R^{-1}Q^T b$.

Proof By Thm 6.13: The least square solution x^* is unique.

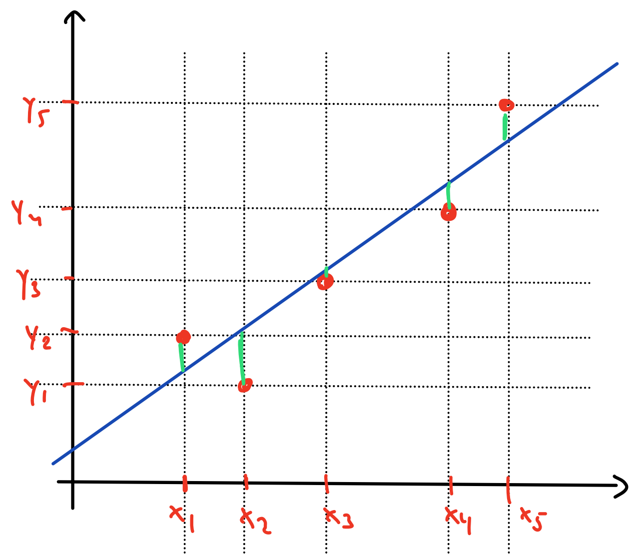
Also: $A(R^{-1}Q^T b) = QRR^{-1}Q^T b = QQ^T b \underset{\text{Thm 6.8}}{=} \text{proj}_{\text{Col}(A)} b$

$\overset{(*)}{\Rightarrow} \quad x^* = R^{-1}Q^T b$ $\square$

6.6. Applications to linear models

For some $x_1, \dots, x_m \in \mathbb{R}$,
we have measurements
$y_1, \dots, y_m \in \mathbb{R}$.

E.g. x_i = points in time
y_i = water temperature



We want: an affine fct. $f(x) = ax + b$ whose graph approximates the points (x_i, y_i).
I.e. a affine fct. $f(x) = ax + b$ that is as close as possible to satisfying $f(x_i) = y_i \ \forall i$.

→ have to quantify what we mean by "as close as possible"
→ many different choices possible!
→ very popular: closest with respect to least square error.

This means: We choose f (i.e. we choose a and b) so that it minimizes the sum of vertical errors:

$$|f(x_1) - y_1|^2 + |f(x_2) - y_2|^2 + \dots + |f(x_m) - y_m|^2$$

→ see the | in pic.

Translate this into a linear algebra problem:

We want to find $a, b \in \mathbb{R}$ (or $\binom{a}{b} \in \mathbb{R}^2$) so that

$$|f(x_1) - y_1|^2 + |f(x_2) - y_2|^2 + \dots + |f(x_m) - y_m|^2 =: (\text{error})^2$$

becomes as small as possible.

Recall: $f(x_i) - y_i = a x_i + b - y_i$

Define $A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix} \in \mathbb{R}^{m \times 2}$ $\Rightarrow A \cdot \underbrace{\binom{a}{b}}_{=: v} = \begin{pmatrix} a x_1 + b \\ \vdots \\ a x_m + b \end{pmatrix}$

Define $y := \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m$.

$\Rightarrow$ error $= \| A v - y \|$

So finding $a, b \in \mathbb{R}$ with minimal error means finding the least square solution v^* for $Av = y$.

Ex. Approximate the data $(1, 0), (2, 3), (3, 3), (4, 2)$ linearly with minimal least square error.

$\Rightarrow$ We want to find a least square solution for $Ax = y$ where

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 0 \\ 3 \\ 3 \\ 2 \end{pmatrix}$$

So we want to find the solutions of $A^T A \begin{pmatrix} a \\ b \end{pmatrix} = A^T y$

$$A^T A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}$$

$$A^T y = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 3 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 23 \\ 8 \end{pmatrix}$$

$\Rightarrow \det(A^T A) \neq 0 \Rightarrow A^T A$ invertible $\Rightarrow A^T A \begin{pmatrix} a \\ b \end{pmatrix} = y$ has unique sol.

Compute or guess: $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3/5 \\ 1/2 \end{pmatrix}$

$\Rightarrow f(x) = ax + b = \frac{3}{5} x + \frac{1}{2}$

In other words, the line that best approximates the data $(1,0), (2,3), (3,3), (4,2)$ in the sense of least squares is the graph of the function $f(x) = \frac{3}{5} x + \frac{1}{2}$

$$L = \left\{ \begin{pmatrix} 0 \\ \underset{\nwarrow b}{1/2} \end{pmatrix} + t \begin{pmatrix} 1 \\ \underset{\uparrow a}{3/5} \end{pmatrix} : t \in \mathbb{R} \right\}$$