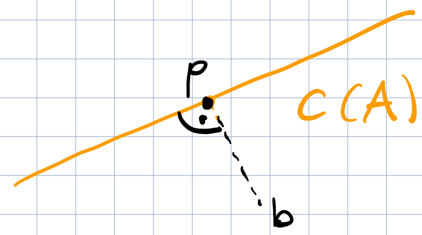
Woche 9

Erinnerung: $\overbrace{Ax=b \text{ hat keine Lösung}}^{\text{d.h. } b \notin C(A)}$: finde das nächstgelegene lösbare System $Ax=p$, d.h. $\|b-p\|$ kleinstmöglich. p heisst <u>Projektion</u> von b auf $C(A)$.



Lemma 5.2.3:

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}; \quad b \in \mathbb{R}^m$$

$S = \operatorname{span}(a_1, \ldots, a_n) = C(A)$.

Die <u>Projektion</u> $\operatorname{proj}_S(b)$ von b auf S ist gegeben durch

$$\operatorname{proj}_S(b) = A\hat{x},$$

wobei $\hat{x}$ eine beliebige Lösung der Normalgleichungen

$$A^TAx = A^Tb$$

ist.

Explizite Formel für $\operatorname{proj}_S(b)$? Ja, wenn A^TA invertierbar und $\hat{x}$ somit eindeutig ist.

Lemma 5.2.4. A^TA ist invertierbar gdw. A linear unabh. Spalten hat.

Beweis: A^TA invertierbar $\Leftrightarrow$ A^TA hat linear unabh. Spalten

$\Leftrightarrow N(A^TA) = \{0\}$

Lemma 5.1.11 $\quad$ $A^TAx = 0$ hat nur die triviale Lösung

$\Leftrightarrow N(A) = \{0\}$

$\Leftrightarrow$ Die Spalten von A sind linear unabhängig.

Theorem 5.2.6 Sei S Unterraum von $\mathbb{R}^m$ und A eine Matrix, deren Spalten eine <u>Basis</u> von S sind. Die Projektion von $b \in \mathbb{R}^m$ auf S ist

$$\text{proj}_S(b) = Pb,$$

wobei $P = A(A^TA)^{-1}A^T$ die Projektionsmatrix ist.

Beweis: $\text{proj}_S(b) = A\hat{x}$, $\quad A^TA\hat{x} = A^Tb$

$$\Updownarrow$$

$$\hat{x} = (A^TA)^{-1}A^Tb \quad \checkmark$$

Es gilt: $P = A(A^TA)^{-1}A^T \overset{\text{Lemma 3.9}}{=} \underbrace{AA^{-1}}_{I} \cdot \underbrace{(A^T)^{-1}A^T}_{I} = I$

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

Unsinn! Gilt nur, wenn A (und damit auch A^T) invertierbar sind.

- A, A^T sind im allgemeinen nicht quadratisch und können dann gar nicht invertierbar sein.

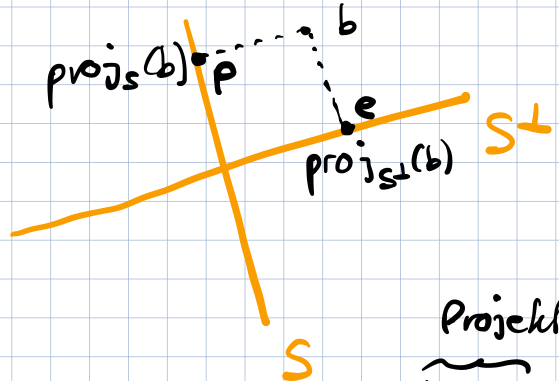
- Falls A quadratisch und invertierbar ist, gilt $S = C(A) = \mathbb{R}^m$ und tatsächlich $P = I$ (b ist dann bereits die Projektion auf $\mathbb{R}^m$).
- P ist die Matrix einer linearen Transformation $T_P(b) = Pb$ (Projektion auf S)

Sanity check: Es sollte gelten: $\overbrace{\mathrm{proj}_S(\mathrm{proj}_S(b))}^{P^2 b} = \underbrace{\mathrm{proj}_S(b)}_{Pb}$.

Tatsächlich gilt: $P^2 = P$:

$$P^2 = (A(A^TA)^{-1}A^T)(A(A^TA)^{-1}A^T)$$
$$= A\underbrace{(A^TA)^{-1}(A^TA)}_{I}(A^TA)^{-1}A^T = \underbrace{A(A^TA)^{-1}A^T}_{P}$$

- Sei $S^\perp$ das orthogonale Komplement von S.



Dann gilt: $\mathrm{proj}_{S^\perp}(b) = \overbrace{(I - P)}^{\text{Projektionsmatrix auf } S^\perp} b$

Beweis: $b = p + e$ mit eindeutigen $p \in S$, $e \in S^\perp$.

Haben gezeigt: $p = \mathrm{proj}_S(b)$ (Beweis von Lemma 5.2.3)

Das gleiche Argument zeigt: $e = \mathrm{proj}_{S^\perp}(b)$.

Also gilt: $\overbrace{\mathrm{proj}_{S^\perp}(b)}^{e} = b - \overbrace{\mathrm{proj}_S(b)}^{p}$

$= Ib - Pb = (I-P)b.$

- Sanity check: $(I-P)^2 = I-P$? ✓

$\Vert$ binomische Formel

$$\underbrace{I^2}_{I} - \underbrace{2IP}_{2P} + \underbrace{P^2}_{P} = I - P.$$

Clicker:

$PA = P$? $\quad \boxed{P}\,\boxed{A} = \boxed{P}$? nein!

$PA = A$ (jede Spalte von A ist in $C(A)$)

$Pv = 0$ ($v = p + e$, $\underbrace{p}_{0 = \mathrm{proj}_{C(A)}(v) = Pv} \in C(A)$, $e \in C(A)^\perp$

$Av = 0$ (Falsche Dimensionen:

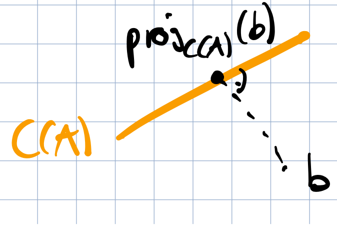
A ist $m \times n$, $v \in \mathbb{R}^m$ ↯) $\quad n \neq m$

Es stimmt: $v^T A = 0$ (wegen $v \in C(A)^\perp$)

Least Squares (Methode der kleinsten Quadrate, 5.3)

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2, \quad A\ m \times n,\ b \in \mathbb{R}^m$$

$$= \min_{p \in C(A)} \|p - b\|^2 = \underbrace{\|\mathrm{proj}_{C(A)}(b) - b\|^2}_{\text{Abstand von } b \text{ zu } C(A).}$$

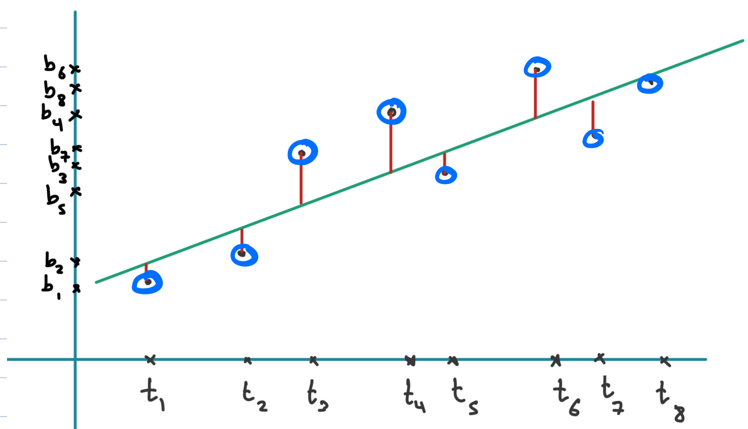


↓ Finde $\hat{x}$, das $Ax=b$ „am besten“ löst
(Abstand von $A\hat{x}$ und b ist kleinstmöglich).

Unzählige Anwendung in Fällen, wo $Ax=b$ nur ungefähr lösbar ist.

Falls A linear unabh. Spalten hat, ist $\hat{x}=(A^TA)^{-1}A^Tb$ die Lösung (einzige Lösung der Normalgleichungen; $A\hat{x} = \text{proj}_{C(A)}(b)$.)

Anwendung: Gerade, die Datenpunkte „erklärt“



Lineare Regression:

- Datenpunkte $(t_1, b_1), (t_2, b_2), \ldots, (t_m, b_m)$
  (z.B. Messungen eines Werts b zu bestimmten Zeiten t).
- Falls b linear von t abhängt (wir wissen aber nicht wie), sollten wir Konstanten α_0, α_1 suchen,

so dass
$$b_k \approx \alpha_0 + \alpha_1 t_k \quad \text{für } k = 1 \ldots m$$
"=" ist nicht zu erwarten, aber "≈" sollte möglichst gut sein.

- Finde α_0, α_1, so dass die Summe der quadratischen Fehler minimiert wird!

$$\min_{\alpha_0, \alpha_1} \sum_{k=1}^{m} \left(b_k - (\alpha_0 + \alpha_1 t_k)\right)^2$$

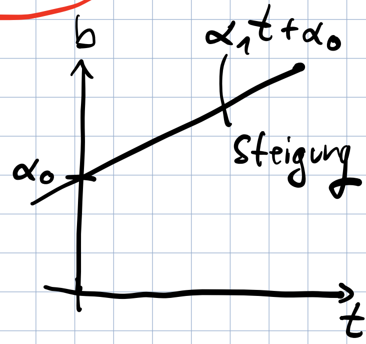
k-te Eintrag von $A\begin{bmatrix}\alpha_0\\\alpha_1\end{bmatrix}$

k-te Eintrag von b

Matrix-Vektor-Notation:

$$\min_{\alpha_0, \alpha_1} \left\| A \begin{bmatrix}\alpha_0\\\alpha_1\end{bmatrix} - b \right\|^2,$$

$$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$



Falls A linear unabh. Spalten hat, ist die Lösung

$$\begin{bmatrix}\alpha_0\\\alpha_1\end{bmatrix} = (A^T A)^{-1} A^T b$$

$$= \underbrace{\begin{bmatrix} m & \sum_{k=1}^{m} t_k \\ \sum_{k=1}^{m} t_k & \sum_{k=1}^{m} t_k^2 \end{bmatrix}}_{A^T A}^{-1} \underbrace{\begin{bmatrix} \sum_{k=1}^{m} b_k \\ \sum_{k=1}^{m} b_k t_k \end{bmatrix}}_{A^T b}$$

In allen vernünftigen Fällen hat A lin. unabh. Spalten.

Einzige Möglichkeit für lin. abh. Spalten ist, dass

$\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{bmatrix}$ skalares Vielfaches von $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ ist, d.h.

alle t_k sind gleich. Sobald wir zwei verschiedene t_k's haben, sind die Spalten linear unabhängig.

Bem. 5.3.3: Falls die Spalten von A orthogonal sind (d.h. $\sum_{k=1}^{m} t_k = 0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{bmatrix}$), ist $A^T A$ eine Diagonalmatrix und leicht zu invertieren:

$$\begin{bmatrix} m & 0 \\ 0 & \sum_{k=1}^{m} t_k^2 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{m} & 0 \\ 0 & \frac{1}{\sum_{k=1}^{m} t_k^2} \end{bmatrix}.$$

Wir haben dann die explizite Lösung

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{m} \sum_{k=1}^{m} b_k \\ \sum_{k=1}^{m} t_k b_k \,/\, \sum_{k=1}^{m} t_k^2 \end{bmatrix}.$$

Orthogonalität hilft bei Projektionen (mehr dazu später).

Verallgemeinerung: kubische, quartische, ...

Falls wir an eine quadratische Abhängigkeit glauben, d.h. $b_k \approx \alpha_0 + \alpha_1 t_k + \alpha_2 t_k^2$, können wir eine Parabel finden, die die Datenpunkte erklärt:

$$\min_{\alpha_0, \alpha_1, \alpha_2} \left\| A \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} - b \right\|^2,$$

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

↑ Spalten linear unabhängig, sobald es drei verschiedene t_k's gibt.