Matrix: $A^T A$

↳ $A \in \mathbb{R}^{n \times m}$

↳ $A^T A \in \mathbb{R}^{m \times m}$

1) $A^T A$ ist symmetrisch, denn:

$$(A^T A)^T = A^T (A^T)^T = A^T A$$

2) Es ist $\operatorname{Kern}(A) = \operatorname{Kern}(A^T A)$

Sei $x \in \operatorname{Kern}(A)$. Dann ist $A^T A x = A^T \cdot 0 = 0 \quad \Rightarrow x \in \operatorname{Kern}(A^T A)$

Sei $x \in \operatorname{Kern}(A^T A)$. Dann ist

$$\begin{aligned} 0 = A^T A x &= x^T A^T A x \\ &= (Ax)^T (Ax) \\ &= \|Ax\|_2^2 \\ &\Rightarrow Ax = 0, \text{ also } x \in \operatorname{Kern}(A) \end{aligned}$$

3) $A^T A$ ist invertierbar $\Leftrightarrow$ $\operatorname{Kern}(A) = \{0\}$

4) $A^T A$ ist immer mindestens positiv semi-definit.

Sei $x \in \mathbb{R}^m \setminus \{0\}$. Dann ist $x^T A^T A x = (Ax)^T (Ax) = \|Ax\|_2^2 \geq 0$.

Ausgleichsrechnung

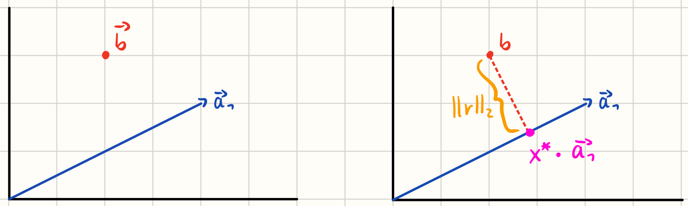
Motivation: · überbestimmtes (unlösbares) System $A \cdot x = b$, $A \in \mathbb{R}^{n \times m}$

↳ $n > m$ und $rg(A) = m$

· $\exists b \in \mathbb{R}^n$, das nicht im Spaltenraum von $A \in \mathbb{R}^{n \times m}$ liegt

↳ Erinnerung: $Bild(A) \subseteq \mathbb{R}^n$ und $\dim(Bild(A)) = rg(A)$

Idee: Bestimme x^*, sodass $\|r\|_2 = \|Ax - b\|_2$ minimiert wird



$x^* \hat{=}$ Wie viel $\vec{a}_1$ brauche ich, um Projektionspunkt zu erreichen?

· x^* muss so gewählt werden, dass Residuum r orthogonal zu den Spalten von A ist.

$$\vec{a}_i^{\,T} (Ax - b) = 0 \quad \text{für alle } 1 \leq i \leq m$$

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

$$\Leftrightarrow A^T A x = A^T b$$

Löse also anstatt $Ax = b$ das System $A^T A x = A^T b$

· existiert immer eine Lösung?

· ist die Lösung eindeutig?

Definitheit von Matrizen

Sei $A \in \mathbb{R}^{n\times n}$. Dann unterscheiden wir folgende Begriffe:

positiv (semi) definit $\iff$ $\forall x \neq 0 \;:\; x^T A x > 0 \quad (\geq 0)$

negativ (semi) definit $\iff$ $\forall x \neq 0 \;:\; x^T A x < 0 \quad (\leq 0)$

indefinit $\iff$ $\exists x, y \;:\; x^T A x > 0$ und $y^T A y < 0$

Ist A symmetrisch, so gilt:

positiv (semi) definit: [n+h] • alle EW sind > 0 (≥ 0)
[n+h] • alle führenden Hauptminoren sind > 0 (für psd zu kompliziert)
[n] • alle Diagonalelemente sind > 0 (≥ 0) (auch ohne Symmetrie)

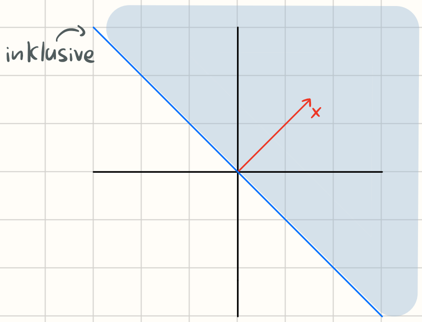
negativ (semi) definit: [n+h] • alle EW sind < 0 (≤ 0)
[n+h] • führenden Hauptminoren sind abwechselnd > 0 und < 0, wobei $\det(a_{1,1}) < 0$ sein muss (für nsd zu kompliziert)
[n] • alle Diagonalelemente sind < 0 (≤ 0) (auch ohne Symmetrie)

indefinit (sonst): [n+h] • es gibt positive und negative EW
[h] • es gibt positive und negative Diagonalelemente (auch ohne Symmetrie)

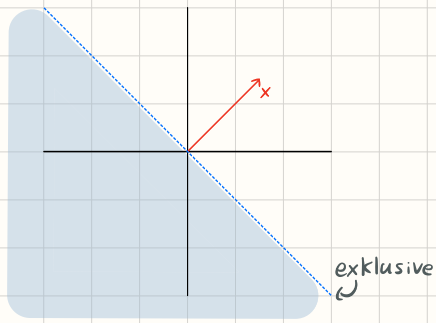
n ⇝ notwendige Bedingung
h ⇝ hinreichende Bedingung

Warum gilt das mit den Diagonalelementen?
Sei z.B. A positiv definit. Dann gilt insb. für $e_i \in \mathbb{R}^n$ $\quad 0 < e_i^T A e_i = e_i^T \vec{a}_i = a_{i,i}$

Geometrische Interpretation
Für $x \in \mathbb{R}^n$ gilt: $x^T A x = x^T \overbrace{(Ax)}^{=:y} = x^T y = \langle x, y \rangle_2 = \overbrace{\|x\| \cdot \|y\|}^{\geq 0} \cdot \cos(\gamma)$
↳ die Definitheit hängt also vom Winkel zw. x und Ax ab



A ist positiv semidefinit
⇝ Bereich für Ax, x bel.



A ist negativ definit
⇝ Bereich für Ax, x bel.

Es gilt: A ist positiv/negativ definit $\Rightarrow$ $\text{Kern}(A) = \{0\}$

↳ A pd $\Rightarrow$ $x^T A x > 0$ für alle $x = 0$ $\Rightarrow$ $Ax \neq 0$ für alle $x \neq 0$ $\Rightarrow$ $\text{Kern}(A) = \{0\}$

↳ A nd analog...

Beispiel:

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 4 & -1 & -2 \\ 6 & 1 & 2 \end{bmatrix}$$

ist indefinit, da $e_1^T A e_1 = a_{1,1} > 0$ und $e_2^T A e_2 = a_{2,2} < 0$

$$B = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Wir untersuchen die führenden Hauptminoren:

$\det(4) > 0$ $\quad \det\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} = 7 > 0$ $\quad \det(B) = 0$

Da B sym. ist, folgt, dass B positiv semidefinit ist.

Cholesky-Zerlegung

Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch und positiv (semi-) definit.
Dann heißt $A = L \cdot L^T$ für eine Δ-Matrix L die Choleskyzerlegung von A.

Gibt es so eine Zerlegung auch, wenn $A \neq A^T$?
↳ Nein, denn $(LL^T)^T = LL^T$ ist symmetrisch

Gibt es so eine Zerlegung auch für Matrizen, die nicht p(s)d sind?
↳ Nein, denn $x^T A x = x^T L L^T x = (L^T x)^T (L^T x) = \|L^T x\| \geq 0$ für alle $x \in \mathbb{R}^n \setminus \{0\}$

Algorithmus zur Berechnung von L

$$\ell_{i,j} = \begin{cases} \sqrt{a_{i,i} - \sum\limits_{k=1}^{i-1} \ell_{i,k}^2} & \text{falls } i = j \\ \frac{1}{\ell_{j,j}} \left(a_{i,j} - \sum\limits_{k=1}^{j-1} \ell_{i,k} \ell_{j,k} \right) & \text{sonst} \end{cases}$$

spielt diesen Algorithmus unbedingt mal für eine 4x4-Matrix selbst durch!

Bsp.: $A = \begin{bmatrix} 1 & -1 & 3 \\ -1 & 5 & -7 \\ 3 & -7 & 14 \end{bmatrix}$ $L = \begin{bmatrix} 1 & & \\ -1 & 2 & \\ 3 & -2 & 1 \end{bmatrix}$ $L = \begin{bmatrix} 1. & & \\ 2. & 3. & \\ 4. & 5. & 6. \end{bmatrix}$

Berechnungsreihenfolge

$$\ell_{1,1} = \sqrt{a_{1,1}} = 1$$

$$\ell_{2,1} = \frac{1}{\ell_{1,1}} \cdot a_{2,1} = \frac{1}{1} \cdot (-1) = -1$$

$$\ell_{2,2} = \sqrt{a_{2,2} - \ell_{2,1}^2} = \sqrt{5 - (-1)^2} = \sqrt{4} = 2$$

$$\ell_{3,1} = \frac{1}{\ell_{1,1}} \cdot a_{3,1} = \frac{1}{1} \cdot 3 = 3$$

$$\ell_{3,2} = \frac{1}{\ell_{2,2}} \cdot (a_{3,2} - \ell_{2,1} \cdot \ell_{3,1}) = \frac{1}{2} (-7 - (-1) \cdot 3) = -2$$

$$\ell_{3,3} = \sqrt{a_{3,3} - \ell_{3,1}^2 - \ell_{3,2}^2} = \sqrt{14 - 3^2 - (-2)^2} = 1$$

L hat positive Diagonaleinträge $\Leftrightarrow$ A ist positiv definit
- per Konstruktion ist L^T eine Matrix in ZSF.

1) sind Diagonalelemente in ZSF alle $\neq 0$, dann $rg(L) = rg(L^T) = n$
↳ $x^T A x = x^T L L^T x = \|L^T x\|_2^2 > 0$ für alle $x \neq 0$, da L^T invertierbar $\Rightarrow$ A ist pd

2) gibt es 0en auf Hauptdiagonalen, dann ist bekanntlich $rg(L) = rg(L^T) < n$
↳ $\exists x \in \mathbb{R}^n \setminus \{0\} : L^T x = 0 \Rightarrow Ax = LL^T x = L \cdot 0 = 0$
und damit A nicht pd.

Daraus erhalten wir eine erste <u>wichtige</u> Anwendung

A ist positiv definit $\Leftrightarrow$ Algorithmus läuft durch und L hat nur positive Diagonalelemente

Wo kann der Algorithmus scheitern, falls A nicht pd ist?

1) Bei Berechnung der Diagonalelemente wird Ausdruck unter der Wurzel negativ
   ↳ ist genau dann der Fall wenn $A^{(i)} = (a_{k,t})_{1 \le k,t \le i}$ nicht psd ist.

2) Bei Berechnung der anderen Einträge ist $l_{i,i} = 0$
   ↳ ist A pd, dann passiert das nicht

Also: Algo. aus der VL kann bei psd-Matrizen scheitern
↳ für pd Matrix funktioniert er immer

Wie verwenden wir Cholesky für das Lösen von $Ax = b$?

1) Berechne L mit $A = LL^T$ $\quad O(n^3)$ wie bei Gauss

2) Löse $Ax = b \Leftrightarrow LL^T x = b$ in 2 Schritten
   - Zuerst $Ly = b$ durch Vorwärtseinsetzen $\quad O(n^2)$
   - Dann $L^T x = y$ durch Rückwärtseinsetzen $\quad O(n^2)$

3) wir können LL^T für verschiedene rechte Seiten nutzen

Lineare Regression

Gegeben seien Punkte (x_i, y_i), $1 \leq i \leq n$

Dann suchen wir eine Funktion f, die die Punkte approximiert

↳ $f(x_i) \approx y_i$, wobei Approximation „optimal" sein soll

Beispiel:

Betrachte die Punkte $(1,1)$, $(2,2)$, $(3,2)$

Wir suchen $\beta_0, \beta_1 \in \mathbb{R}$, sodass $f: \mathbb{R} \to \mathbb{R}$ $f(x) = \beta_1 \cdot x + \beta_0$ die Punkte $\overbrace{\text{(optimal)}}^{?}$ approximiert.

Perfekt wäre $f(x_i) = y_i$, also als LGS dann

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \quad \underset{A}{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}} \cdot \underset{\beta}{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}} = \underset{y}{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}} \quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Dieses System hat keine Lösung, denn... ?

Wir wollen also (β_0, β_1) so wählen, dass $|y_i - f(x_i)|$ minimal wird

Also wollen wir $r := y - A\beta$ bzw. $\|r\| = \|y - A\beta\|$ minimieren

↳ Normalengleichung!

Wir lösen also $A^T A \beta = A^T y$ und erhalten

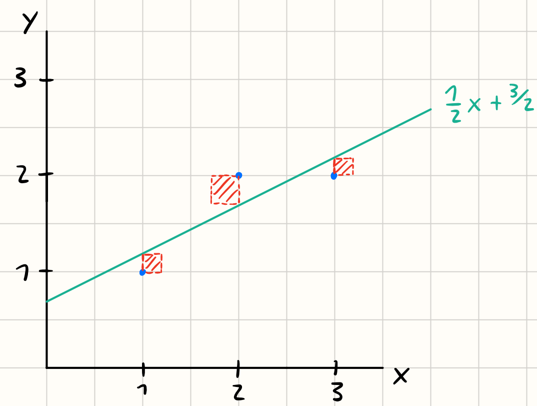
$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix} \quad \Rightarrow \quad \begin{aligned} \beta_0 &= 2/3 \\ \beta_1 &= 1/2 \end{aligned}$$

hat genau eine Lösung, da Spalten von A lin. unabh.

↳ warum das i.A. so ist, werden wir uns im Kapitel zu Interpolation angucken

Welchen Fehler minimieren wir?



... denn wir minimieren $\|r\|_2 = \|y - A\beta\|_2$

also $\sqrt{\sum_{i=1}^{3} \Big(y_i - \underbrace{(\beta_1 x_i + \beta_0 \cdot 1)}_{f(x_i)}\Big)^2}$

und jeder Summand entspricht genau einem der roten Quadrate in der Grafik links.

(da die Wurzel monoton ist, spielt sie hier keine Rolle)

Wir hätten auch eine andere Funktion f wählen können!

(1) $f(x) = \beta_2 x^2 + \beta_1 x + \beta_0$

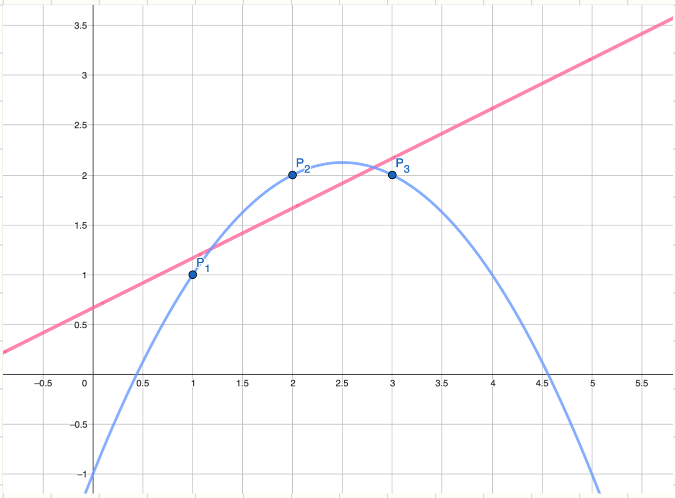
$$\rightsquigarrow \quad A = \begin{bmatrix} 1 & x_1^1 & x_1^2 \\ 1 & x_2^1 & x_2^2 \\ 1 & x_3^1 & x_3^2 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Für unsere Punkte erhalten wir z.B.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

- hat dieses System schon eine (eindeutige) Lösung?
  ↳ brauchen wir hier $A^T A \beta = A^T y$?
  Was passiert, wenn wir das betrachten?

Graphische Darstellung der erhaltenen Funktionen



$f(x) = \frac{1}{2}x + \frac{2}{3}$

$f(x) = -\frac{1}{2}x^2 + \frac{5}{2}x - 1$

Kann man auch etwas anderes als Polynome betrachten?
Ja, aber man muss darauf achten, dass die Bilder der Stützstellen unter den gewählten Basisvektoren lin. unabh. sind!
↳ sonst verliert man die Eindeutigkeit der Lösung (klar?)

Bei unserem Bsp. ginge z.B. auch problemlos $f(x) = \beta_1 \cdot e^x + \beta_0 \cdot e^{2x}$, da

$$\begin{bmatrix} e^1 \\ e^2 \\ e^3 \end{bmatrix} \text{ und } \begin{bmatrix} e^2 \\ e^4 \\ e^6 \end{bmatrix} \text{ lin. unabh. sind.}$$