§8 Singulärwertzerlegung:

Was tun für eine Matrix m×n?

Bem: Für $\underline{A}$ quadratisch und diagonalisierbar (normale Matrix):

es gibt eine unitäre Matrix $\underline{S}$

$$\underline{A} = \underline{S}\,\underline{\Lambda}\,\underline{S}^H = \lambda_1 \underline{P}_1 + \lambda_2 \underline{P}_2 + \dots + \lambda_n \underline{P}_n$$

orthogonale Projektionen !

Summe von Projektoren

Basis der EV → Eigenbasis

$\underline{P}_n$ = Matrizen von Rang 1

s_n .. Eigenvektoren

$$= \lambda_1 s_1 s_1^H + \lambda_2 s_2 s_2^H + \dots + \lambda_n s_n s_n^H$$

Theorem [SVD = singular value decomposition]

gilt für beliebige Matrix !

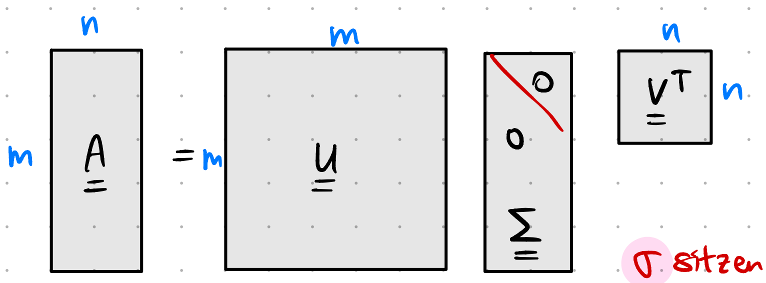
$\underline{A} \in \mathbb{R}^{m\times n}$ beliebig. Dann existieren !

$\underline{U} \in \mathbb{R}^{m\times m}$, $\underline{V} \in \mathbb{R}^{n\times n}$ beide orthogonal sodass

$$\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T$$

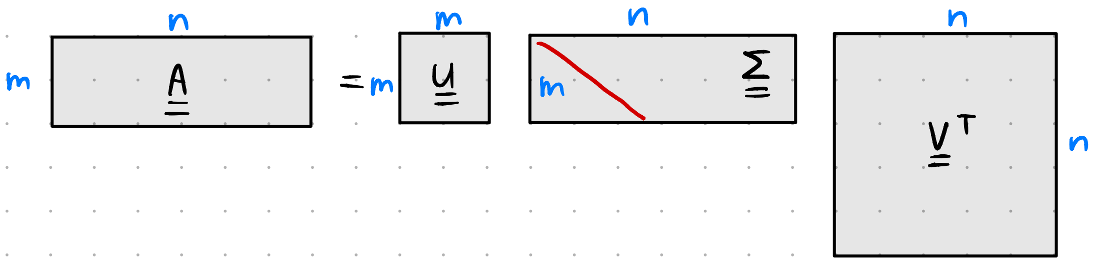
mit $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_p) \in \mathbb{R}^{m\times n}$, $p = \min\{m,n\}$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ Singulärwerte von $\underline{A}$



$\underline{U}$ & $\underline{V}$ sind orthogonal !

σ sitzen auf der Hauptdiagonalen von Σ



$$\underline{\underline{A}} = \underline{\underline{U}}\;\underline{\underline{\Sigma}}\;\underline{\underline{V}}^T$$

Beweis:

1) $\underline{\underline{A}} \in \mathbb{R}^{m\times n} \Rightarrow \underline{\underline{A}}^T\underline{\underline{A}} \in \mathbb{R}^{n\times n}$ symmetrisch

Spektralsatz $\Rightarrow \underline{\underline{A}}^T\underline{\underline{A}}$ hat n paarweise orthonormale $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_r$ EV

entsprechend zu $\lambda_1, \lambda_2, \dots, \lambda_r \in \mathbb{R}\setminus\{0\}$ EW von $\underline{\underline{A}}^T\underline{\underline{A}}$

mit $r = \text{Rang}(\underline{\underline{A}}^T\underline{\underline{A}}) = \text{Rang}(\underline{\underline{A}})$

↳ $\underline{\underline{A}}\underline{v} = 0$ für die EV $\underline{v}$ die zum EW 0 entsprechen.

2) EW für $j = 1, 2, \dots, r$: $\underline{\underline{A}}^T\underline{\underline{A}}\underline{v}_j = \lambda_j \underline{v}_j \Rightarrow$

$\underline{\underline{A}}\,|$

$\Rightarrow \underline{\underline{A}}\,\underline{\underline{A}}^T(\underline{\underline{A}}\underline{v}_j) = \lambda_j(\underline{\underline{A}}\underline{v}_j) \Rightarrow \underline{\underline{A}}\underline{v}_j$ EV der Matrix $\underline{\underline{A}}\,\underline{\underline{A}}^T$ zum EW λ_j

$\underline{\underline{A}}\,\underline{\underline{A}}^T$ symmetrisch $\underset{\text{Spektralsatz}}{\Rightarrow}$ EV orthogonal zueinander $\Rightarrow$

$\Rightarrow \underline{\underline{A}}\underline{v}_j$ sind paarweise orthogonal zueinander

3) Berechne dessen Norm:

$$\|\underline{\underline{A}}\underline{v}_j\|_2^2 = (\underline{\underline{A}}\underline{v}_j)^T(\underline{\underline{A}}\underline{v}_j) = \underline{v}_j^T\underbrace{\underline{\underline{A}}^T\underline{\underline{A}}}\underline{v}_j = \underline{v}_j^T(\underbrace{(\underline{\underline{A}}^T\underline{\underline{A}})\underline{v}_j}_{=\,\lambda_j\underline{v}_j \text{ da } \underline{v}_j \text{ EW von } \underline{\underline{A}}^T\underline{\underline{A}}}) = \underline{v}_j^T\lambda_j\underline{v}_j = \lambda_j\underline{v}_j^T\underline{v}_j =$$

$$= \lambda_j\|\underline{v}_j\|_2^2 = \lambda_j$$

Somit $\lambda_j = \|\underline{\underline{A}}\underline{v}_j\|_2^2 > 0$ da $\underline{v}_j \notin \text{Kern}(\underline{\underline{A}})$

↓ wir haben den EW 0 eliminiert.

Notiere $\sigma_j = \|\underline{\underline{A}}\underline{v}_j\|_2 = \sqrt{\lambda_j} > 0$

4) Notiere den normierten EV $\underline{\underline{A}}\,\underline{\underline{A}}^T$ mit $\underline{u}$

$$\underline{u}_j = \frac{1}{\sigma_j}\underline{\underline{A}}\,\underline{v}_j$$

und berechne $\underline{u}_i^T \underline{\underline{A}}\,\underline{v}_j = \left(\frac{1}{\sigma_i}\underline{\underline{A}}\,\underline{v}_i\right)^T \underline{\underline{A}}\,\underline{v}_j = \frac{1}{\sigma_i}\underline{v}_i^T \underbrace{\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{v}_j}_{\overset{\shortparallel}{\lambda_j \underline{v}_j}}$

$$= \frac{1}{\sigma_i}\underline{v}_i^T \lambda_j \underline{v}_j = \frac{\lambda_j}{\sigma_i}\underline{v}_i^T\underline{v}_j = \begin{cases} 0 & \text{für } i \neq j \\ \frac{\lambda_i}{\sigma_i} & \text{für } i = j \end{cases}$$

$v_j \perp v_i$ für $i \neq j$

$$v_i^T v_j = \begin{cases} 0 & \text{für } i \neq j \\ 1 & \text{für } i = j \end{cases}$$

$$\frac{\lambda_i}{\sigma_i} = \frac{\sigma_i^2}{\sigma_i} = \sigma_i \quad \text{da } \sigma_i = \sqrt{\lambda_i}$$

Somit: $\underline{u}_i^T \underline{\underline{A}}\,\underline{v}_j = \begin{cases} 0 & \text{für } i \neq j \\ \sigma_i & \text{für } i = j \end{cases}$

<u>Somit:</u>

$$\underbrace{\begin{array}{c|c} & m \\ \hline r & \begin{bmatrix} \underline{u}_1^T \\ \underline{u}_2^T \\ \vdots \\ \underline{u}_r^T \end{bmatrix} \end{array}}_{\hat{\underline{\underline{U}}}^T} \underline{\underline{A}}^{m\times n} \underbrace{\begin{array}{c|c} & r \\ \hline n & \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_r \end{bmatrix} \end{array}}_{\hat{\underline{\underline{V}}}} = \begin{array}{c|c} & r \\ \hline r & \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix} \end{array}$$

weil wir alle $\lambda = 0$ eliminiert haben, sind alle $\sigma > 0$

$$\underset{r \times m}{\hat{\underline{\underline{U}}}^T}\ \underset{m \times n}{\underline{\underline{A}}}\ \underset{n \times r}{\hat{\underline{\underline{V}}}} = \underset{r \times r}{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}}$$

Ziel:

$$\underbrace{\underline{\underline{U}}^T}_{m\times m}\;\underbrace{\underline{\underline{A}}}_{m\times n}\;\underbrace{\underline{\underline{V}}}_{n\times n} = \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \\ & 0 & \end{bmatrix}}_{m\times n}$$

5) $\underline{u}_1, \dots, \underline{u}_r \in \mathbb{R}^m$ paarweise orthogonal.

wir erweitern diese Menge zu einer ONB in $\mathbb{R}^m$

$\underline{u}_1, \dots, \underline{u}_r, \underline{u}_{r+1}, \dots, \underline{u}_m \in \mathbb{R}^m$ ONB

$\underbrace{\underline{v}_1, \dots, \underline{v}_r}_{\perp \operatorname{Ker}(\underline{\underline{A}})}, \underbrace{\underline{v}_{r+1}, \dots, \underline{v}_n}_{\in \operatorname{Ker}(\underline{\underline{A}})} \in \mathbb{R}^n$ ONB

$u_i^T A v_j$

$$\underline{u}_i^T \underline{\underline{A}}\, \underline{v}_j = \begin{cases} \sigma_i & \text{falls } i = j \leq r \\ 0, & \text{sonst} \end{cases}$$

$\underline{\underline{A}}\,\underline{v}_{r+1} = \underline{0}, \dots, \underline{\underline{A}}\,\underline{v}_n = \underline{0}$ ← da kein u existiert

$\Rightarrow \underline{\underline{U}}^T \underline{\underline{A}}\, \underline{\underline{V}} = \underline{\underline{\Sigma}} \Rightarrow \underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T$

$\underline{\underline{U}}$ | $\qquad \underline{\underline{V}}^T$ |

$$\underline{\underline{A}} = \underbrace{\begin{bmatrix} \underline{u}_1 & \underline{u}_2 & \dots & \underline{u}_r & \underline{u}_{r+1} & \dots & \underline{u}_m \end{bmatrix}}_{m \times m} \begin{bmatrix} \begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{matrix} & 0 \\ 0 & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{bmatrix} \underbrace{\begin{bmatrix} \underline{v}_1^T \\ \underline{v}_2^T \\ \vdots \\ \underline{v}_r^T \\ \underline{v}_{r+1}^T \\ \vdots \\ \underline{v}_n^T \end{bmatrix}}_{n \times n}$$

$\underline{v}_{r+1}, \dots, \underline{v}_n$ ONB in $\operatorname{Ker}(\underline{\underline{A}})$

$\underline{v}_1, \dots, \underline{v}_r$ = rechte Singularvektoren = EV von $\underline{\underline{A}}^T\underline{\underline{A}}$

$\underline{u}_1, \dots, \underline{u}_r$ = linke Singularvektoren = EV von $\underline{\underline{A}}\,\underline{\underline{A}}^T$

im Beweis: $\underline{u}_1^T \underline{\underline{A}}\, \underline{v}_1 = \sigma_1$

$$\underline{u}_1 = \frac{1}{\sigma_1}\underline{\underline{A}}\,\underline{v}_1 \Rightarrow \underline{\underline{A}}\,\underline{v}_1 = \sigma_1 \underline{u}_1$$

$$\dots \quad \underline{\underline{A}}\,\underline{v}_r = \sigma_r \underline{u}_r$$

$$\underline{\underline{A}}\,\underline{v}_{r+1} = 0, \dots, \underline{\underline{A}}\,\underline{v}_n = 0$$

$$\left.\begin{array}{l} \Rightarrow \underline{u}_1, \dots, \underline{u}_r \in \text{Bild}(\underline{\underline{A}}),\ \underline{r \text{ orthogonal}} \\ \dim \text{Bild}(\underline{\underline{A}}) = r \end{array}\right\} \Rightarrow \text{ONB in Bild}(\underline{\underline{A}})$$

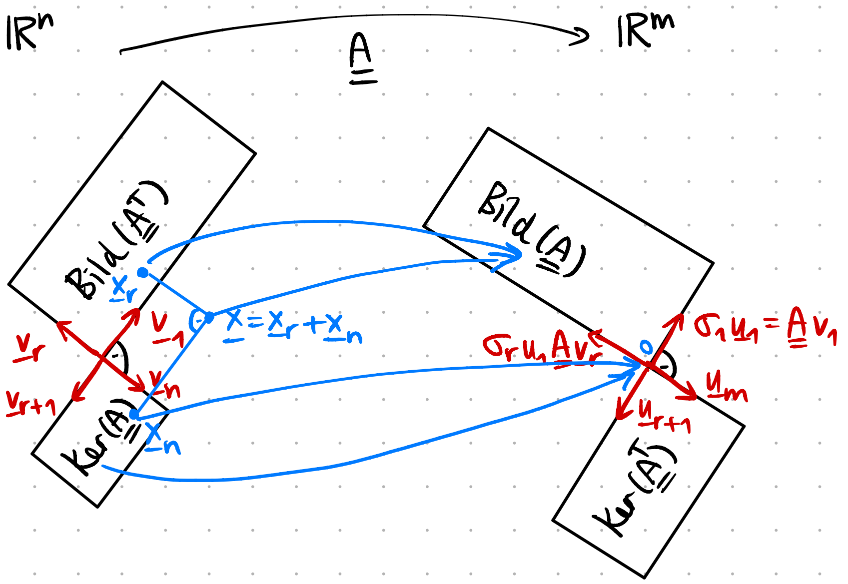
$$\left.\begin{array}{l} \underline{v}_{r+1}, \dots, \underline{v}_n \ (n-r), \text{ orthogonal in Ker}(\underline{\underline{A}}) \\ \text{DimKer}(\underline{\underline{A}}) = n-r \end{array}\right\} \Rightarrow$$

$$\Rightarrow \text{Ker}(\underline{\underline{A}}) = \text{span}\{\underline{v}_{r+1}, \dots, \underline{v}_n\} \text{ ONB in Ker}(\underline{\underline{A}})$$

$$\rightarrow \text{Dim}(\underline{\underline{A}}) \overset{!}{=} \text{Dim}(\text{Bild}(\underline{\underline{A}})) + \text{Dim}(\text{Kern}(\underline{\underline{A}})) \overset{!}{=} n$$

Bem Fundamentalsatz der lin. Algebra:

$$\underline{\underline{A}} \in \mathbb{R}^{m \times n}, \quad \mathcal{A}: \mathbb{R}^n \to \mathbb{R}^m, \quad \mathcal{A}(\underline{x}) = \underline{\underline{A}}\,\underline{x}$$



Bsp: $\underline{\underline{A}} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$ Ziel: SVD

$\underline{\underline{A}}^T\underline{\underline{A}} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$ berechne EW von $\underline{\underline{A}}^T\underline{\underline{A}}$:

$$\begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} = (5-\lambda)^2 - 9 = \lambda^2 - 10\lambda + 25 - 9 = (\lambda-2)(\lambda-8) \stackrel{!}{=} 0$$

$$\Rightarrow \text{EW } 2, 8 \Rightarrow \sigma_1^2 = 8, \quad \sigma_2^2 = 2 \Rightarrow \sigma_1 = 2\sqrt{2} \quad \sigma_2 = \sqrt{2}$$

EV zu 8: $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

EW zu 2: $\frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

daraus $\underline{\underline{V}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \Rightarrow \underline{\underline{V}}^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

Vektoren $\underline{u}$ bauen:

$\underline{\underline{A}}\,\underline{v}_1 = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix} \xRightarrow{\text{Normiere}} \underline{u}_1 = \frac{1}{\sigma_1} \underline{\underline{A}}\,\underline{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$\underline{\underline{A}}\,\underline{v}_2 = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix} \Rightarrow \underline{u}_2 = \frac{1}{\sigma_2} \underline{\underline{A}}\,\underline{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Somit ist die SVD von $\underline{\underline{A}}$:

$$\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

<u>Bsp 2:</u> $\underline{\underline{A}} = \begin{bmatrix} 1 & 1 \\ 7 & 7 \end{bmatrix} \Rightarrow \underline{\underline{A}}^T \underline{\underline{A}} = \begin{bmatrix} 50 & 50 \\ 50 & 50 \end{bmatrix}$

EW von $\underline{\underline{A}}^T\underline{\underline{A}}$: $\det \begin{bmatrix} 50-\lambda & 50 \\ 50 & 50-\lambda \end{bmatrix} = (50-\lambda)^2 - 50^2 = \lambda(\lambda - 100)$

$$\Rightarrow \sigma_1^2 = 100, \quad \sigma_2^2 = 0 \quad \Rightarrow \sigma_1 = 10, \quad \sigma_2 = 0$$

EV zu EW 100: $\begin{bmatrix} -50 & 50 \\ 50 & -50 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underline{0} \Leftrightarrow \begin{bmatrix} -50 & -50 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underline{0}$

freie Variable

$x_2 = a$

$-50x_1 + 50a = 0 \quad \Rightarrow \quad a\begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \underline{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$x_1 = a$

erweitere $\underline{v}_1$ zu ONB in $\mathbb{R}^2$

$$\underline{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \perp \underline{v}_1$$

$$\underline{u}_1 = \frac{1}{\sigma_1} \underline{\underline{A}}\underline{v}_1 = \frac{1}{50}\begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

erweitere zu ONB in $\mathbb{R}^2 \Rightarrow \underline{u}_2 = \frac{1}{\sqrt{50}}\begin{bmatrix} -7 \\ 1 \end{bmatrix}$

$$\underline{\underline{A}} = \frac{1}{\sqrt{50}}\begin{bmatrix} 1 & -7 \\ 7 & 1 \end{bmatrix}\begin{bmatrix} 10 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$A^{2\times 3}$

$\underbrace{\underline{\underline{A}}\,\underline{\underline{A}}^T}_{2\times 3 \;\; 3\times 2} = 2\times 2$

egal welches nehmen.

$\underbrace{\underline{\underline{A}}^T\underline{\underline{A}}}_{3\times 2 \;\; 2\times 3} = 3\times 3$

⇐ kleinere Matrix nehmen (in der Regel einfacher)

aber!! es kann auch sein, dass die grössere Matrix schöner ist!

Bem: Reduzierte SVD

$$\underline{A} = [\underline{u}_1, \dots \underline{u}_r] \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix} \begin{bmatrix} \underline{v}_1^T \\ \dots \\ \underline{v}_r^T \end{bmatrix}$$

§ 8.2 Anwendungen der SVD:

1) Speicherreduktion

$\underline{A}$ braucht $m \times n$ Plätze

red. SVD: $\left.\begin{array}{l} \underline{u}_1, \dots, \underline{u}_r : rm \\ \underline{v}_1, \dots, \underline{v}_r : rn \\ \sigma_1, \dots, \sigma_r : r \end{array}\right\} \Rightarrow r(m+n+1)$ Plätze $\ll mn$

vor allem wenn r klein

2) Berechnung der Norm einer Matrix:

$$\|\underline{A}\|_2 = \max_{\|x\|_2 = 1} \|\underline{A}\underline{x}\|_2$$

$\underline{U}$ orthogonal

$$\|\underline{A}\underline{x}\|_2^2 = \|U\Sigma U^T x\|_2^2 = \|\Sigma V^T x\|_2^2 = \sigma_1^2 |\underbrace{\underline{v}_1^T \underline{x}}_{y_1}|^2 + \sigma_2^2 |\underbrace{\underline{v}_2^T \underline{x}}_{y_2}|^2 + \dots + \sigma_r^2 |\underbrace{\underline{v}_r^T \underline{x}}_{y_r}|$$

$y_1, \dots, y_r$ = die ersten r Komponenten von $\underline{y} = \underline{V}^T \underline{x}$

$$\|\underline{A}\|_2 = \max_{\|\underline{x}\|_2 = 1} \|\underline{A}\,\underline{x}\|_2 = \max_{\|\underline{x}\|_2 = 1} \|\Sigma V^T x\|_2$$

$$\underline{y} = V^T x \Leftrightarrow \underline{x} = \underline{V}\underline{y}$$

$$= \max_{\|y\|_2 = 1} \|\underline{\Sigma}\,\underline{y}\|_2$$

da $\underline{V}$ orthogonal
(verändert 2-Norm nicht)

$$\|\Sigma y\|_2^2 = \sigma_1^2|y_1|^2 + \sigma_2^2|y_2|^2 + \ldots + \sigma_r^2|y_r|^2$$

$$= \sigma_1\Big[|y_1|^2 + \underbrace{\left(\tfrac{\sigma_2}{\sigma_1}\right)^2}_{\leq 1}|y_2|^2 + \ldots + \underbrace{\left(\tfrac{\sigma_r}{\sigma_1}\right)^2}_{\leq 1}|y_r|^2\Big] \leq \sigma_1^2\|\underline{y}\|_2^2 \Rightarrow$$

$$\max_{\|\underline{y}\|_2=1} \|\underline{\underline{\Sigma}}\,\underline{y}\|_2 = \sigma_1 \text{ erreicht für } \underline{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Rightarrow \|\underline{\underline{A}}\|_2 = \sigma_1$$

Bem: $$\max_{\underline{x}\in\mathbb{R}^n} \frac{\|\underline{\underline{A}}\,\underline{x}\|_2}{\|\underline{x}\|_2} = \sigma_1 \text{ erreicht für } \underline{x} = \underline{\underline{V}}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underline{v}_1$$

$$\max_{\underline{x}\in\mathbb{R}^n} \frac{\|\underline{\underline{A}}\,\underline{x}\|_2}{\|\underline{x}\|_2} = \max_{\boxed{\substack{\underline{x}\in\mathbb{R}^n \\ \underline{v}_1^T\underline{x}=0}}} \frac{\|\underline{\underline{A}}\,\underline{x}\|_2}{\|\underline{x}\|_2} = \sigma_2 \text{ erreicht für } \underline{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\underline{x} = \underline{\underline{V}}\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \underline{v}_2$$

und so weiter!

3) Datenkompression

Erinnerung: $\underline{\underline{A}}$ quadratisch, symmetrisch: $\Rightarrow$ reelle EW $\Rightarrow$ wir können EW $\lambda_1 \geq \lambda_2 > \ldots \geq \lambda_n$ anordnen.

$$\underline{\underline{A}} = \sum_{j=1}^{n} \lambda_j \underline{s}_j \hat{\underline{s}}_j^T = \underline{\underline{S}}\,\underline{\underline{\Lambda}}\,\underline{\underline{S}}^{-1}$$

sobald EW λ_j imaginär $\Rightarrow$ geht das nicht mehr.

SVD: $\underline{\underline{A}}$ beliebig:

$$\underline{\underline{A}} = \sigma_1 \underline{u}_1\underline{v}_1^T + \sigma_2\underline{u}_2\underline{v}_2^T + \ldots + \sigma_r\underline{u}_r\underline{v}_r^T$$

= Summe von Matrizen von Rang 1

$\{\underline{u}_1, \ldots, \underline{u}_r\}$, $\{\underline{v}_1, \ldots, \underline{v}_r\}$ biorthogonale Basen

Rang $\underline{\underline{A}} = r \Rightarrow$ Man kann daraus eine Approximation (Kompression) von $\underline{\underline{A}}$

mit Matrizen von niedrigerem Rang machen

$$\underline{A} = \sigma_1 \underline{u}_1 \underline{v}_1^T + \dots + \sigma_r \underline{u}_r \underline{v}_r^T \approx \sum_{j=1}^{k} \sigma_j \underline{u}_j \underline{v}_j^T \quad \text{mit } k \ll r$$

<u>Bsp:</u> $\underline{A} = \begin{bmatrix} 2 & 100 \\ 0 & 1 \end{bmatrix} \quad \lambda_1 = 2, \ \lambda_2 = 1$

$$\underline{A} = \underline{S}\,\underline{\Lambda}\,\underline{S}^{-1} = \underbrace{2}_{\lambda_1}\underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\underline{s}_1}\underbrace{\begin{bmatrix} 1 & 100 \end{bmatrix}}_{\hat{\underline{s}}_1^T} + \underbrace{1}_{\lambda_2}\underbrace{\begin{bmatrix} 1 \\ -\frac{1}{100} \end{bmatrix}}_{\underline{s}_2}\underbrace{\begin{bmatrix} 0 & -100 \end{bmatrix}}_{\hat{\underline{s}}_2^T}$$

Vorschlag zur Approximation:

$$\underline{A} - \lambda_1 \underline{s}_1 \hat{\underline{s}}_1^T = \begin{bmatrix} 0 & -100 \\ 0 & 1 \end{bmatrix} \quad ☹$$

$$\underline{A} - \lambda_2 \underline{s}_1 \hat{\underline{s}}_2^T = \begin{bmatrix} 2 & 200 \\ 0 & 0 \end{bmatrix} \quad ☹$$

nicht klein

$\Rightarrow$ Kompression via Diagonalisierung bringt nichts. ↯

Besser mit SVD:

$$\underline{A} = \underbrace{100.025}_{\sigma_1}\underbrace{\begin{bmatrix} 0.9995 \\ 0.001 \end{bmatrix}}_{\underline{u}_1}\underbrace{\begin{bmatrix} 0.01999 & 0.9998 \end{bmatrix}}_{\underline{v}_1^T} + \underbrace{0.02}_{\sigma_2}\underbrace{\begin{bmatrix} 0.01 \\ 0.9995 \end{bmatrix}}_{\underline{u}_2}\underbrace{\begin{bmatrix} 0.9998 & 0.01999 \end{bmatrix}}_{\underline{v}_2^T}$$

Vorschlag zur Approximation: $\underline{A} \approx \sigma_1 \underline{u}_1 \underline{v}_1^T$

$$\text{Fehler} \quad \underline{A} - \sigma_1 \underline{u}_1 \underline{v}_1^T = \sigma_2 \underline{u}_2 \underline{v}_2^T = \begin{bmatrix} 2 \times 10^{-4} & 0 \\ -2 \times 10^{-2} & 9 \times 10^{-4} \end{bmatrix} \quad ☺$$

kleine Fehler

<u>Theorem:</u> [Eckart-Young]

$\underline{A}$ $m \times n$. Für jedes $k \le \text{Rang}(\underline{A})$ gibt die abgebrochene SVD

$$\underline{A}_k = \sum_{j=1}^{k} \sigma_j \underline{u}_j \underline{v}_j^H$$

die beste Approximation von Rang k an $\underline{A}$ im Sinne von:

$$\|\underline{\underline{A}} - \underline{\underline{A}}_k\| = \min_{\underline{\underline{X}} \text{ mit Rang}(\underline{\underline{X}}) \le k} \|\underline{\underline{A}} - \underline{\underline{X}}\| = \sigma_{k+1}$$

Bem:

$\|\cdot\|$ = Euklidische Norm, Frobenius Norm, Nuclear Norm.

Spektralnorm:

$$\|\underline{\underline{A}}\|_2 = \max \frac{\|\underline{\underline{A}}x\|_2}{\|x\|_2} = \sigma_1$$

Frobenius Norm: $\|\underline{\underline{A}}\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2} = \sqrt{\sum_{i,j} |A_{ij}|^2} = \text{spur}(\underline{\underline{A}}^T \underline{\underline{A}})$

Nuclear Norm: $\|\underline{\underline{A}}\|_N = \sigma_1 + \sigma_2 + \dots + \sigma_r$

Bem: $\underline{\underline{Q}}$ unitär / orthogonal. $\underline{\underline{Q}} \in \mathbb{R}^{n\times n} / \mathbb{C}^{n\times n}$

$\|\underline{\underline{Q}}\|_2 = 1$, $\|\underline{\underline{Q}}\|_k = \sqrt{n}$, $\|\underline{\underline{Q}}\|_N = n$.

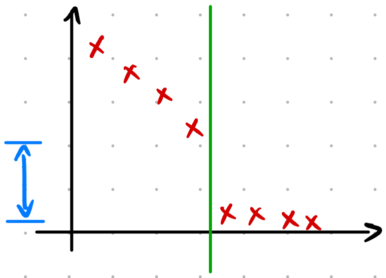
Beweis: baut auf der SVD / Fundamentalsatz der linearen Algebra.

4) Numerischer Rang einer Matrix

$$\sigma_1 \ge \sigma_2 > \dots \ge \sigma_k \gg \sigma_{k+1} \ge \dots \ge \sigma_r > 0$$

k: numerischer Rang

r: exakter, mathematischer Rang von $\underline{\underline{A}}$



5) Pseudoinverse

$\underline{\underline{A}}$ $m\times n$ $\qquad \underline{\underline{A}}^+ = \underbrace{\underline{\underline{V}}}_{n\times n} \underline{\underline{\Sigma}}^+ \underbrace{\underline{\underline{U}}^T}_{m\times m}$ $n\times m$

$$\text{mit } \Sigma^+ = \begin{bmatrix} \sigma_1^{-1} & & & & \\ & \sigma_2^{-1} & & 0 & \\ & & \ddots & & \\ & & & \sigma_r^{-1} & \\ & 0 & & 0 & \ddots & 0 \end{bmatrix} \; n\times m$$

$Q^T Q = I$

$Q^{-1} = Q^T$

orth. matrix: $A^{-1} = A$

$$A^{-1} = A$$

$$\begin{tikzcd} \mathbb{R}^m \arrow[r, bend left, "A"] & \mathbb{R}^n \arrow[l, bend left, "A^+"] \end{tikzcd}$$

A^+ Pseudoinverse

$$A = USV^T$$

$$A^+ = (USV^T)^+$$

$$= (V^T)^+ S^+ U^+$$

$$= V S^+ U^T$$

$A^+ : \mathbb{R}^n \to \mathbb{R}^m$ als lineare Abbildung

gegeben: $b \in \mathbb{R}^n$, $b \mapsto A^+ b = V \Sigma^+ U^T b$

↳ Lösung des Ausgleichsproblems mittels SVD.

⊛ Operator ⟺ Abbildung

6) Polarzerlegung $A \in \mathbb{R}^{n \times n}$

$$A = QH$$

Q: orthogonal, H: symmetrisch positiv definit

$$A = U\Sigma V^H = \underbrace{UV^H}_{Q \text{ orthogonal}} \underbrace{V \Sigma V^H}_{H \text{ symmetrisch pos. def.}} = QH$$

$V^H V$

Bem 1) $A^T A = H^2$

2) $A = KQ$

Q: orthogonal

K: sym. pos. def.

3) Mechanik: Q = Drehung, H = Streckung

↳ EV von H zeigen in den Richtungen der Streckung

↳ EW ($= \sigma$'s) sind die Faktoren der Streckung

wenn λ alg. Vielf. > 1 (z.B. 3)

σ auch 3x aufschreiben in Σ ?

↳ ja

F.S. Anwendungen der SVD: **V14** 18.12.20

7) Ausgleichsrechnung (Methode der kleinsten Quadrate):

$$A x = b \quad \overset{A^T |}{\Rightarrow} \quad \underbrace{A^T A}_{} x = A^T b$$

$$\left.\begin{array}{r} \text{quadratisch, symmetrisch} \\ \text{Falls Rang}(A) = n \end{array}\right\} \Rightarrow A^T A \text{ invertierbar} \Rightarrow \text{mit LU-Zerlegung}$$

$\Rightarrow$ Cholesky-Zerlegung (weil symmetrisch) pos. definit

$\Rightarrow x$

(falls A orthogonal $(A^T A = A A^T = I)$
$\Rightarrow A^T A x = A^T b \Rightarrow x = A^T b$ (hier nicht der Fall).)

besser mit QR-Zerlegung: Stabilität, Struktur von A

Falls $\text{Rang}(A) = r < n$: via SVD lösen:

$$A = \begin{array}{c} \\ \left[\begin{array}{c|c} \overset{r}{U_1} & \overset{m-r}{U_2} \end{array}\right] \end{array} \left[\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array}\right] \begin{array}{c} \left[\begin{array}{c} \\ \hline V_2^T \end{array}\right] \end{array} \begin{array}{l} r \\ \}\, n-r \end{array}$$

U_1: bilden Basis im Bild (A)

V_2^T: bilden Basis im Kern (A)

$$\Sigma_r = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix} \quad \text{mit } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$$

Ziel: $\|A x - b\|_2^2 = \min!$

$$\|Ax - b\|_2^2 = \|U \Sigma V^T x - b\|_2^2 \overset{U \text{ orth.}}{=} \|\Sigma V^T x - U^T b\|_2^2 =$$

$$= \left\| \left[\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} V_1^T \\ \hline V_2^T \end{array}\right] x - \left[\begin{array}{c} U_1^T \\ \hline U_2^T \end{array}\right] b \right\|_2^2 =$$

$$= \left\| \left[\begin{array}{c} \Sigma_r V_1^T x \\ \hline 0 \end{array}\right] - \left[\begin{array}{c} U_1^T b \\ \hline U_2^T b \end{array}\right] \right\|_2^2 = \min$$

Das ist Minimal $\|\underline{\underline{U}}_2^T \underline{b}\|_2^2$ für $\underline{x}$ so dass $\underline{\underline{\Sigma}}_r \underline{\underline{V}}_1^T \underline{x} = \underline{\underline{U}}_1^T \underline{b}$

oder alternativ: notiere $\underline{y} = \underline{\underline{V}}^T \underline{x}$ $\Rightarrow$

?

$\Rightarrow$ $\underline{\underline{\Sigma}}_r \underline{y} = \underline{\underline{U}}^T \underline{b}$ wobei nur die ersten Komponenten von $\underline{y}$ zu bestimmen sind.

$$\underline{x} = \underline{\underline{V}}_1 \underline{\underline{\Sigma}}_r^{-1} \underline{\underline{U}}_1^T \underline{b}$$

$$r \underbrace{\boxed{\quad \underline{\underline{V}}_1^T \quad}}_{n} \Rightarrow \underline{\underline{V}}_1^T \underline{\underline{V}}_1 = \underline{\underline{I}}_r$$

→ das wird in Standardsoftware verwendet!

8) Totales lineares Ausgleichsproblem:

total = Messfehler auch in den Messorten $\underline{t_i}$

$\Rightarrow$ Messfehler befinden sich in $\underline{b}$ und $\underline{\underline{A}}$

$\underline{\underline{A}}\,\underline{x} = \underline{b}$

Ziel: Finde $\hat{\underline{\underline{A}}}$, $\hat{\underline{b}}$, $\hat{\underline{x}}$ so dass $\hat{\underline{\underline{A}}}\hat{\underline{x}} = \hat{\underline{b}}$

und $\hat{\underline{\underline{C}}} = [\hat{\underline{\underline{A}}}\ \hat{\underline{b}}]$ so "nah" wie möglich an

$\underline{\underline{C}} = [\underline{\underline{A}}\ \underline{b}]$

$\|\underline{\underline{C}} - \hat{\underline{\underline{C}}}\| = \min!$

typischerweise: Frobeniusnorm

$$\begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix} \begin{bmatrix} EV_1 & EV_2 & EV_3 \end{bmatrix}$$

möglich

$\sigma_2 = \sigma_3 \Rightarrow$

$\Rightarrow$ stimmt das ???

gleiche EW

→ Ordnung der EV nicht eindeutig

und $\hat{\underline{\underline{A}}}\hat{\underline{x}} = \hat{\underline{b}}$ lösbar, d.h. $\hat{\underline{b}} \in \text{Bild}(\hat{\underline{\underline{A}}})$

d.h. falls $\text{Rang}(\underline{\underline{A}}) = n$ möchte dass $\text{Rang}(\hat{\underline{\underline{C}}}) = n$

Lösung: niedrigrangapproximation an $\underline{\underline{C}}$

$$\underline{\underline{C}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T = \sum_{j=1}^{n+1} \sigma_j \underline{u}_j \underline{v}_j^T$$

Definiere $\hat{\underline{\underline{C}}} = \sum_{j=1}^{n} \sigma_j \underline{u}_j \underline{v}_j^T$

wissen (aus Eckart-Young-Theorem): $\hat{\underline{\underline{C}}}$ ist die beste Approximation an $\underline{\underline{C}}$ in der Menge der Rang n Matrizen.

Ausserdem: $\underline{v}_j$ sind paarweise orthogonal $\Rightarrow$ $\hat{\underline{\underline{C}}}\,\underline{v}_{n+1} = \sum_{j=1}^{n} \sigma_j \underline{u}_j \underbrace{\underline{v}_j^T \underline{v}_{n+1}}_{=0} = 0$

$\Leftrightarrow \underline{v}_{n+1} \in \operatorname{Ker} \hat{\underline{\underline{C}}}$

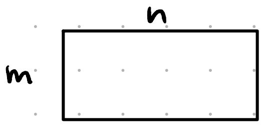
Andererseits: $\hat{\underline{\underline{A}}}\hat{\underline{x}} = \hat{\underline{b}} \Leftrightarrow \hat{\underline{\underline{A}}}\underline{x} - \hat{\underline{b}} = 0 \Leftrightarrow [\,\hat{\underline{\underline{A}}} \quad \hat{\underline{b}}\,]\begin{bmatrix}\underline{x}\\ -1\end{bmatrix} = 0 \Leftrightarrow$

$\hat{\underline{\underline{C}}}\begin{bmatrix}\hat{\underline{x}}\\ -1\end{bmatrix} = 0$ $\Rightarrow$

$\Rightarrow \begin{bmatrix}x\\ -1\end{bmatrix} = k \cdot \underline{v}_{n+1}$

Falls $v_{n+1,n+1} \neq 0 \Rightarrow k = -\frac{1}{v_{n+1,n+1}} \Rightarrow \hat{\underline{x}} = -\frac{1}{v_{n+1,n+1}}\begin{bmatrix}v_{1n+1}\\ \vdots\\ v_{nn+1}\end{bmatrix}$

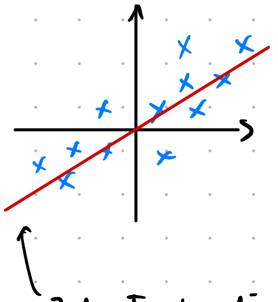
9) PCA = principal component analysis



n Versuche, m Messungen (Messwerte)

Mittelwert "−" ⇒ jede Spalte hat Mittelwert 0

$$\underline{A} = [\underline{a}_1, \underline{a}_2, \dots, \underline{a}_n]$$



$$\underline{a}_1, \underline{a}_2, \dots, \underline{a}_n \in \mathbb{R}^m$$

$$\sum_{j=1}^{m} a_{ij} = 0 \quad \text{für } i = 1, 2, \dots, n$$

Ziel: Finde diese Richtung = wo entlang liegen die Punkte?

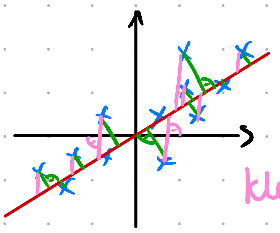
Welche Gerade / Welcher niedrigdimensionaler Raum liegt am nächsten an den Messungen im Sinne der kleinsten Quadrate?

Jede Messung wird auf der Gerade (niedrig dim. Raum) projiziert:

⇒ die Summe der Entfernungen soll minimiert werden

≠ Ausgleichsrechnung (LGS)

= "orthogonale Ausgleichsrechnung

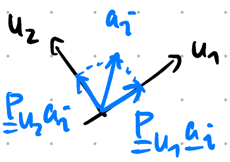


PCA

kleinste Quadrate = Ausgleichsrechnung

$$\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T \qquad \underline{\underline{U}} = [\overbrace{\underline{u}_1 \;\; \underline{u}_2}^{r \text{ bilden eine Basis im Bildraum}} \mid \dots]$$

$$\underline{a}_i = \underline{\underline{P}}_{\underline{u}_1}\underline{a}_i + \underline{\underline{P}}_{\underline{u}_2}\underline{a}_i$$



$$\Rightarrow \|\underline{a}_i\|_2^2 = |\underline{a}_i^T\underline{u}_1|^2 + |\underline{a}_i^T\underline{u}_2|^2$$

$$\Rightarrow \underbrace{\sum_{i=1}^{n} \|\underline{a}_i\|_2^2}_{\text{Fest (gegeben).}} = \underbrace{\sum_{i=1}^{n} |\underline{a}_i^T\underline{u}_1|}_{\sum_{i=1}^{n} \underline{u}_1^T\underline{a}_i\,\underline{a}_i^T\underline{u}_1 = \underline{u}_1^T(\underline{\underline{A}}\,\underline{\underline{A}}^T)\underline{u}_1 = \text{Max}} + \underbrace{\sum_{i=1}^{n} |\underline{a}_i^T\underline{u}_2|^2}_{\min}$$

Somit gibt $\underline{u}_1$ = der erste linke Singularvektor die Richtung der besten Gerade (im oben beschriebenen Sinne)

$\underline{u}_2$ = der zweite linke Singularvektor, orthogonal auf $\underline{u}_1$

$\Rightarrow$ die Richtung in der die Daten am wenigsten variieren

$\Rightarrow$ $\underline{u}_1$ = principal component = die wichtigste Richtung

$\underline{u}_2 \perp \underline{u}_1$: und zeigt in die am wenigsten wichtige Richtung.

Bem: In der Statistik:

Varianz: $\operatorname{diag}(\underline{\underline{A}}\,\underline{\underline{A}}^T)$

Kovarianz: $\operatorname{nichtdiag}(\underline{\underline{A}}\,\underline{\underline{A}}^T)$

Kovarianzmatrix $S = \frac{1}{n-1}\underline{\underline{A}}\,\underline{\underline{A}}^T$

(n = Anzahl Versuche)

$\sqrt{\text{EW von } \underline{\underline{A}}\,\underline{\underline{A}}^T}$: σ_1, σ_2

EV von $\underline{\underline{A}}\,\underline{\underline{A}}^T$: $\underline{u}_1, \underline{u}_2$

10) lineare Ausgleichsrechnung mit linearen Nebenbedingungen

Ziel: Finde $\underline{x} \in \mathbb{R}^n$ bei gegebenem $\underline{\underline{A}} \in \mathbb{R}^{m \times n}$, $\underline{b} \in \mathbb{R}^m$ mit $m \geq n = \operatorname{Rang}(\underline{\underline{A}})$ so dass

$$\begin{cases} \|\underline{\underline{A}}\,\underline{x} - \underline{b}\|_2^2 = \text{minimal} \\ \underline{\underline{C}}\,\underline{x} = \underline{d} \end{cases}$$

$$\left.\begin{array}{l} \text{mit } \underline{\underline{C}} \in \mathbb{R}^{p \times n} \\ \operatorname{Rang} \underline{\underline{C}} = p \\ \text{und } \underline{d} \in \mathbb{R}^p \end{array}\right\} \text{gegeben} \qquad \text{mit } p < n$$

Methode 1: Lagrange-Multiplikatoren: $\underline{m} \in \mathbb{R}^p$ als zusätzliche Unbekannte

$$L(\underline{x}, \underline{m}) = \frac{1}{2}\|\underline{\underline{A}}\,\underline{x} - \underline{b}\|_2^2 + \underline{m}^T(\underline{\underline{C}}\,\underline{x} - \underline{d})$$

$$\min_{\underline{x} \in \mathbb{R}^n} \max_{\underline{m} \in \mathbb{R}^p} L(\underline{x}, \underline{m}) \qquad \text{Sattelpunktproblem}$$

(minimiert nach x und maximiert nach m)

$$\begin{cases} \dfrac{dL}{dx}(\underline{x},\underline{m}) = 0 \\ \dfrac{dL}{dm}(\underline{x},\underline{m}) = 0 \end{cases} \qquad \begin{cases} \underline{\underline{A}}^T(\underline{\underline{A}}\,\underline{x}-\underline{b}) + \underline{\underline{C}}^T\underline{m} = 0 \\ \underline{\underline{C}}\,\underline{x} - \underline{d} = 0 \end{cases}$$

$$\Leftrightarrow \begin{bmatrix} \underline{\underline{A}}^T\underline{\underline{A}} & \underline{\underline{C}}^T \\ \underline{\underline{C}} & 0 \end{bmatrix} \begin{bmatrix} \underline{x} \\ \underline{m} \end{bmatrix} = \begin{bmatrix} \underline{\underline{A}}^T\underline{b} \\ \underline{d} \end{bmatrix}$$

lösen mit Block-LR-Zerlegung

Nachteil: A^TA unter Umständen schlecht konditioniert

Vorteil: Wenn $\underline{\underline{A}}$, $\underline{\underline{C}}$ Struktur $\Rightarrow$ kann man das nutzen!

Methode 2: mit SVD ist besser für Kondition

(Nachteil: man verliert die Struktur von $\underline{\underline{A}}$ und $\underline{\underline{C}}$)

Kondition:

Struktur:

?

$\rightarrow$ 11:47 schauen

$$\underline{\underline{C}} = \underline{\underline{U}}\,[\underline{\underline{\Sigma}}\ \underline{\underline{0}}] \begin{array}{c} \begin{bmatrix} \underline{\underline{V}}_1^T \\ \hline \underline{\underline{V}}_2^T \end{bmatrix} \end{array} \begin{array}{l} \}\, p \\ \}\, n-p \end{array}$$

$\operatorname{Ker} \underline{\underline{C}} = \operatorname{Bild} \underline{\underline{V}}_2$

Trick: $\underline{x}_0 = \underline{\underline{V}}_1 \underline{\underline{\Sigma}}_1^{-1} \underline{\underline{U}}^T \underline{d}$

Dann suche Lösungen $\underline{x} = \underline{x}_0 + \underline{\underline{V}}_2 \underline{y}$ mit $\underline{y} \in \mathbb{R}^{n-p}$

da $\operatorname{Bild} \underline{\underline{V}}_2 = \operatorname{Ker} \underline{\underline{C}} \Rightarrow \underline{\underline{C}}\,\underline{x} = \underline{d}$ ist automatisch erfüllt.

$$\|\underline{\underline{A}}\,\underline{x} - \underline{b}\|_2^2 = \|\underline{\underline{A}}\,\underline{x}_0 + \underline{\underline{A}}\,\underline{\underline{V}}_2 \underline{y} - \underline{b}\|_2^2 = \|\underbrace{(\underline{\underline{A}}\,\underline{\underline{V}}_2)}_{\text{Matrix}}\ \underbrace{\underline{y}}_{\text{zu finden}} + \underbrace{(\underline{\underline{A}}\,\underline{x}_0 - \underline{b})}_{\text{fest}}\|_2^2$$

Das ist ein lin. Ausgleichsproblem mit Matrix $\underline{\underline{A}}\,\underline{\underline{V}}_2$ und rechte Seite $\underline{\underline{A}}\,\underline{x}_0 - \underline{b}$.