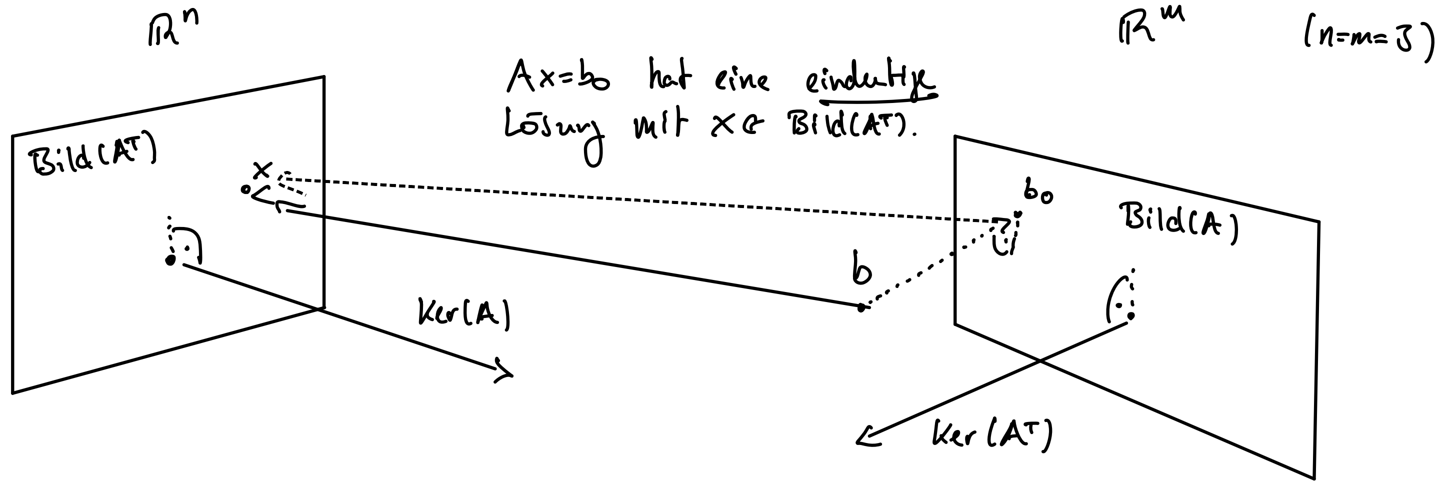
Vorlesung 2: Lineare Algebra con't.

$A \in \mathbb{R}^{m \times n}$, $\quad A = (a_{ij}) \quad 1 \le i \le m, \quad 1 \le j \le n$

$$= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

$A: \mathbb{R}^n \to \mathbb{R}^m, \quad x \mapsto Ax. \quad , \quad x \in \mathbb{R}^n, \quad Ax \in \mathbb{R}^m$



$\text{Bild}(A^T) = \{ A^T y \mid y \in \mathbb{R}^m \}$

Die Pseudoinverse A^+: $\quad A^+ b = x, \quad A^+ \in \mathbb{R}^{n \times m}$

mit $Ax = b_0$, $\quad b_0 = \underset{y \in \text{Bild}(A)}{\operatorname{argmin}} \| b - y \|$

Proposition 1.7 Rang von A

Sei $A \in \mathbb{R}^{m \times n}$.

1) Falls $r(A) = n$, so gilt: $\quad A^+ = (A^T A)^{-1} A^T$ und $A^+ A = 1_n$

2) Falls $r(A) = m$, so gilt: $\quad A^+ = A^T (A A^T)^{-1}$ und $A A^+ = 1_m$.

Beweis

Sei $A^+ b = x$ für $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$! Dann gilt:

$$Ax = b_0 = \underset{y \in \text{Bild}(A)}{\operatorname{argmin}} \|b_0 - b\|$$

Nach Lemma 1.4: $A^T b = A^T b_0$.

$$\Rightarrow \quad A^T A x = A^T b$$

Da $r(A) = n$ und $A^T A \in \mathbb{R}^{n \times n}$, $\rightarrow$ $r(A^T A) = n$ $\rightarrow$ $A^T A$ ist invertierbar.

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

$$\Rightarrow \quad A^+ = (A^T A)^{-1} A^T$$

Teil (2) in der Übung.

Als nächstes wollen wir eine wichtige Basiswahl für $\mathbb{R}^n$ und $\mathbb{R}^m$ thematisieren. Sei $A \in \mathbb{R}^{m \times n}$

Die Matrix $A^T A \in \mathbb{R}^{n \times n}$ ist symmetrisch und positiv-semidefinit.

($B \in \mathbb{R}^{n \times n}$ ist symmetrisch positiv-semidefinit, falls:

- $v^T B v \geq 0$ für alle $v \in \mathbb{R}^n$
- B hat nur nicht-negative Eigenwerte)

Beachte: $v^T A^T A v = (Av)^T (Av) = w_1^2 + \ldots + w_m^2 \geq 0$, $w = Av$.

$$= \langle Av, Av \rangle$$

D.h. $A^T A$ hat eine orthogonale Basis $\{v_1, \ldots, v_n\}$ aus Eigenvektoren, d.h. $\langle v_i, v_j \rangle = \delta_{ij}$. (weil $A^T A$ symmetrisch ist $\rightarrow$ "Spektraltheorem").

Sei $A^T A v_i = \lambda_i v_i$. Dann ist $\lambda_i \geq 0$.

Wir nehmen an, dass $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_n = 0$, $r = r(A)$.

Wir setzen nun:

$$u_i := \frac{1}{\sqrt{\lambda_i}} A v_i \in \mathbb{R}^m, \quad 1 \le i \le r.$$

Es gilt dann: $\langle u_i, u_j \rangle = u_i^T u_j = \frac{1}{\sqrt{\lambda_i \lambda_j}} v_i^T A^T A v_j$

$$= \frac{1}{\sqrt{\lambda_i \lambda_j}} v_i^T (\lambda_j v_j)$$

$$= \frac{\lambda_j}{\sqrt{\lambda_i \lambda_j}} v_i^T v_j = \delta_{ij}$$

$\Rightarrow$ $\{u_1, \dots, u_r\}$ bilden eine orthonormale Basis von $\mathrm{Bild}(A)$.

$\{v_1, \dots, v_r\}$ bilden eine orthonormale Basis von $\ker(A)^\perp = \mathrm{Bild}(A^T)$.

Es gilt $A v_i = \sigma_i u_i, \quad \sigma_i = \sqrt{\lambda_i}, \quad 1 \le i \le r$

$A v_j = 0 \qquad r+1 \le j \le n.$

Daher:

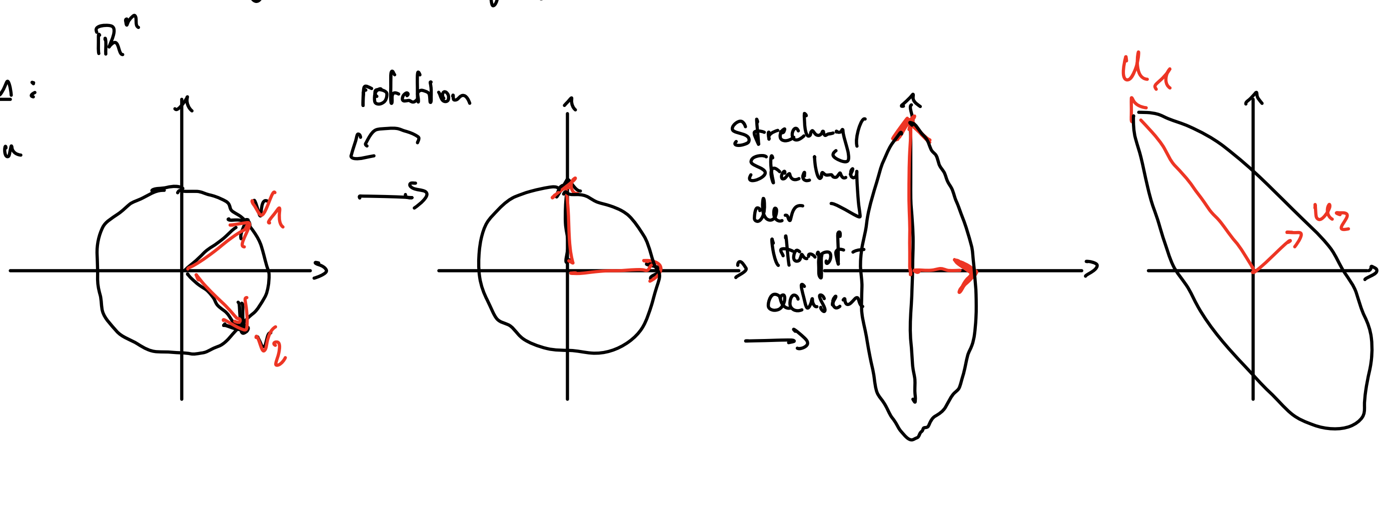
$$A = U \Sigma V^T, \qquad (*)$$

$U = [u_1, \dots, u_r] \in \mathbb{R}^{m \times r}$, $V = [v_1, \dots, v_r] \in \mathbb{R}^{n \times r}$, $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r \times r}$.

$(*)$ heißt Singulärwertzerlegung von A. Die σ_i heißen Singulärwerte.

Cartoon:

$A: \mathbb{R}^n \to \mathbb{R}^n$



Theorem 1.8

Sei $A \in \mathbb{R}^{m\times n}$ und $r = r(A)$. Dann existieren Matrizen $U = [u_1, \dots, u_r] \in \mathbb{R}^{m\times r}$ und $V = [v_1, \dots, v_r] \in \mathbb{R}^{n\times r}$ mit $U^T U = V^T V = 1_r$.
und eindeutig bestimmte Werte $\sigma_1 \geq \dots \geq \sigma_r > 0$

$$A = U \Sigma V^T, \qquad \Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$$

Es gilt: $\operatorname{Bild}(A) = \operatorname{Bild}(U)$ und $\operatorname{Bild}(A^T) = \operatorname{Bild}(V)$.

Falls die σ_i paarweise verschieden sind, sind die u_i und v_i bis auf Vorzeichen eindeutig bestimmt.

Beweis

Existenz und dass $\operatorname{Bild}(A) = \operatorname{Bild}(U)$ und $\operatorname{Bild}(A^T) = \operatorname{Bild}(V)$ folgen aus der obigen Diskussion.

(1) Eindeutigkeit der Singulärwerte:

Angenommen
$$A = U \Sigma V^T = \tilde{U} \tilde{\Sigma} \tilde{V}^T, \qquad \tilde{\Sigma} = \operatorname{diag}(\tilde{\sigma}_1, \dots, \tilde{\sigma}_r)$$
$$\tilde{U} = [\tilde{u}_1, \dots, \tilde{u}_r]$$
$$\tilde{V} = [\tilde{v}_1, \dots, \tilde{v}_r]$$

Dann:

$$AA^T = U \Sigma V^T V \Sigma U^T$$
$$= U \Sigma^2 U^T \quad , \text{ da } V^T V = 1_r.$$

und

$$AA^T = \tilde{U} \tilde{\Sigma}^2 \tilde{U}^T \quad \text{aus dem gleichen Grund.}$$

D.h. $$AA^T u_i = U \Sigma^2 U^T u_i = U \Sigma e_i \quad , \quad e_i = (0, \dots, \underset{i}{1}, 0, \dots, 0)$$
$$= \sigma_i^2 U e_i$$
$$= \sigma_i^2 u_i$$

und

$$AA^T \tilde{u}_i = \tilde{\sigma}_i^2 \tilde{u}_i.$$

Da $r(AA^T) = r$: $\{\sigma_1^2, \dots, \sigma_r^2\}$ sind die nicht-null Eigenwerte von AA^T

$\{\tilde\sigma_1^2, \dots, \tilde\sigma_r^2\}$

Da Eigenwerte eindeutig bestimmt sind haben wir:

$$\{\sigma_1^2, \dots, \sigma_r^2\} = \{\tilde\sigma_1^2, \dots, \tilde\sigma_r^2\}$$

Da $\sigma_1 \geq \dots \geq \sigma_r$ und $\tilde\sigma_1 \geq \dots \geq \tilde\sigma_r$ folgt: $\sigma_i = \tilde\sigma_i \Rightarrow \Sigma = \tilde\Sigma$

$\rightarrow$ Singulärwerte sind eindeutig bestimmt!

2) <u>Eindeutigkeit der u_i und v_j.</u>

Falls $\sigma_1 > \dots > \sigma_r > 0$, dann gilt auch: $\sigma_1^2 > \dots > \sigma_r^2 > 0$.

Da u_i Eigenvektor von AA^T zum Eigenwert σ_i^2 ist und σ_i^2 ein <u>einfacher</u> Eigenwert von AA^T ist, ist u_i <u>eindeutig</u> bis auf Vorzeichen.

Für die v_j wiederholen wir das Argument mit A^TA. $\square$

<u>Lemma 1.9</u>

Sei $A \in \mathbb{R}^{m \times n}$ und $A = U\Sigma V^T$ <u>eine/die</u> SVD ("Singulärwertzerlegung") von A wie in Thm 1.8.

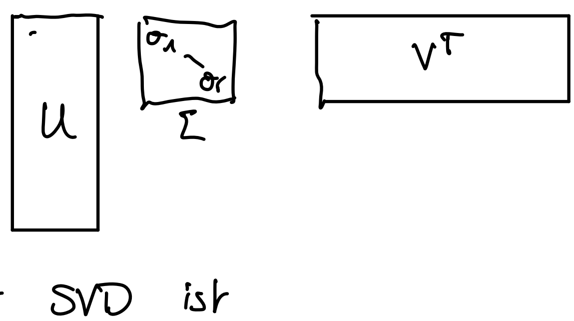
Dann gilt:

$$A^+ = V\Sigma^{-1}U^T$$

<u>Beweis</u>

Übung.

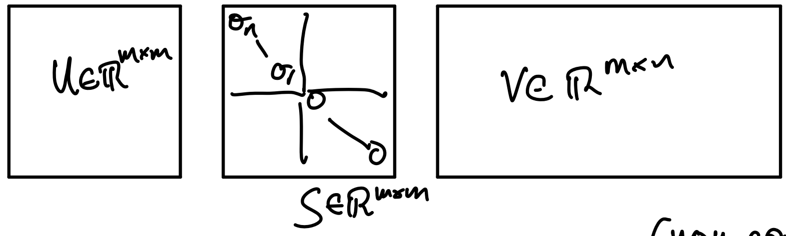
Die SVD ("Singular value decomposition") von Thm. 1.8 lässt wie folgt darstellen:

$A = U\Sigma V^T =$ 

(compact SVD)

Eine alternative Definition der SVD ist

$m \leq n$

$A = U S V^T =$ 

Der Unterschied zwischen compact und non-compact (non-compact SVD) SVD ist, dass wir in der non-compact Version in U eine orth. Basis vom ganzen $\mathbb{R}^m$ haben, und nicht nur von Bild(A) (und entsprechend für V).

Beide Alternativen haben ihre Berechtigung. Die Art wie man über die SVD denken sollte, ist nicht ob compact oder non-compact, sondern dass in beiden Fällen die SVD eine Zerlegung ist, die grundlegende Eigenschaften und Informationen über A sichtbar macht.