Vorlesung 1: Lineare Algebra

Viele mathematische Methoden in den Datenwissenschaften basieren auf Linearer Algebra. In dieser Vorlesung verwenden wir folgende Notation:

$$A = (a_{ij}) \in \mathbb{R}^{m \times n} / \mathbb{C}^{m \times n}, \quad \begin{array}{l} 1 \leq i \leq m \\ 1 \leq j \leq n \end{array}$$

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

ist eine $m \times n$-Matrix mit reellen oder komplexen Einträgen a_{ij}.

Die Spaltenvektoren sind

$$a_j = (a_{ij})_{i=1}^{m} \in \mathbb{R}^m .$$

Eine Matrix $A \in \mathbb{R}^{m \times n}$ kann als eine geordnete Liste von n Vektoren im $\mathbb{R}^m$ betrachtet werden.

Für $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ haben:

$$Ax = x_1 \cdot a_1 + \ldots + a_n \cdot x_n \in \mathbb{R}^n$$

eine Linearkombination der Spaltenvektoren von A.

Andere Interpretationen von A,

1) Liste von n Vektoren im $\mathbb{R}^m$
2) Liste von m Vektoren im $\mathbb{R}^n$
3) Lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$, $x \mapsto Ax$.
4) Lineare Abbildung $\mathbb{R}^m \to \mathbb{R}^n$, $y \mapsto A^T y$
5) Bilineare Abbildung $\mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $(x, y) \mapsto y^T A x$.

All diese Punkte lassen sich am besten durch 4 Unterräume verstehen (2 von $\mathbb{R}^n$ und 2 von $\mathbb{R}^m$).

<u>Definition 1.1</u> (4 Unterräume)

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

1) $\mathrm{Bild}(A) := \{ Ax \mid x \in \mathbb{R}^n \}$
2) $\mathrm{Bild}(A^T) := \{ A^T y \mid y \in \mathbb{R}^m \}$
3) $\ker(A) := \{ x \in \mathbb{R}^n \mid Ax = 0 \}$
4) $\ker(A^T) := \{ y \in \mathbb{R}^m \mid A^T y = 0 \}$

Wir geben den $\mathbb{R}$-Vektorräumen $\mathbb{R}^n$ und $\mathbb{R}^m$ die Struktur eines Euklidischen Raumes, indem wir die positiv-definite Form

$$\langle a, b \rangle := a^T b \qquad a, b \in \mathbb{R}^n \text{ oder } \mathbb{R}^m$$

definieren:

$$\measuredangle(a, b) = \arccos \frac{\langle a, b \rangle}{\sqrt{\langle a, a \rangle \langle b, b \rangle}}$$

Wir sagen, dass a und b <u>orthogonal</u> sind, falls $\langle a, b \rangle = 0$, $a \perp b$.

<u>Beispiel</u> $a = (1, 0)^T$, $b = (0, 1)^T$

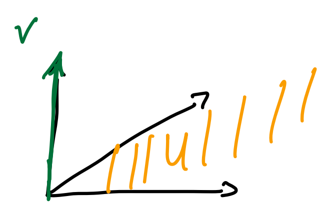
$a^T b = 1 \cdot 0 + 0 \cdot 1 = 0$.

Für $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m \times n}$ ist $A^T y = (\langle a_i, y \rangle)_{i=1}^{n}$

<u>Definition 1.2</u>

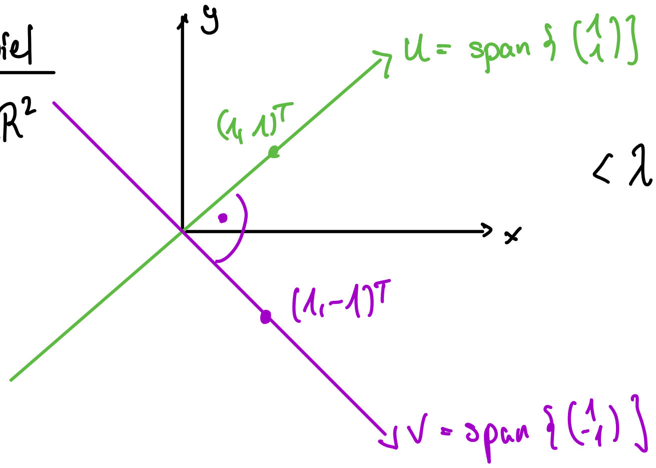
Seien $U, V \subset \mathbb{R}^n$ Unterräume. Wir schreiben $U \perp V$, falls für alle $u \in U$ und $v \in V$ gilt: $\langle u, v \rangle = 0$.

<u>Beispiel</u>: $\mathbb{R}^3$

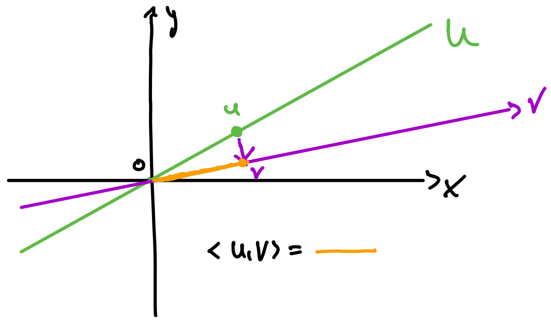


Beispiel

$\mathbb{R}^2$



$$\langle \lambda \tbinom{1}{1}, \mu \tbinom{1}{-1} \rangle = \lambda\mu \langle \tbinom{1}{1}, \tbinom{1}{-1} \rangle$$
$$= \lambda\mu\,(1 + (-1)) = 0.$$



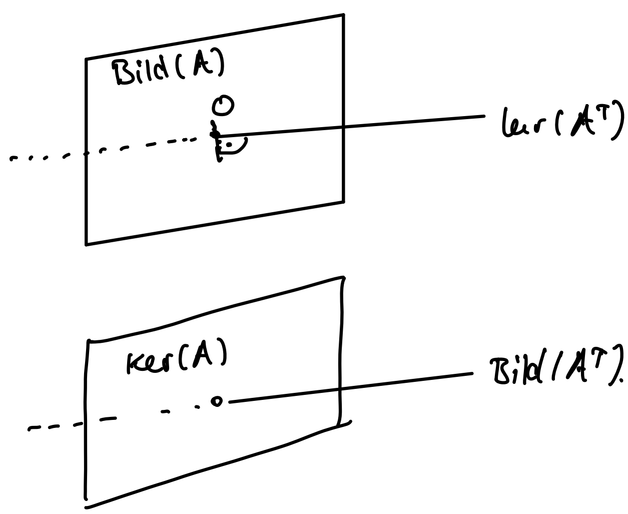
$\langle u, u \rangle = 1$

Theorem 1.3

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

1) $\text{Bild}(A) \oplus \ker(A^T) = \mathbb{R}^m$
2) $\text{Bild}(A) \perp \ker(A^T)$
3) $\text{Bild}(A^T) \oplus \ker(A) = \mathbb{R}^n$
4) $\text{Bild}(A^T) \perp \ker(A)$.

Beispiel: $\mathbb{R}^3$



Beweis

Sei $r(A)$ der Rang von A, $r(A) = \dim \text{Bild}(A) = \dim \text{Bild}(A^T)$

Wir wissen auch: $\dim \text{Bild}(A^T) + \dim \ker(A^T) = m$

$\rightarrow \dim \text{Bild}(A) + \dim \ker(A^T) = m = \dim \mathbb{R}^m$

$\rightarrow \text{Bild}(A) + \ker(A^T) = \mathbb{R}^m$

Zweitens: für $y \in \ker(A^T)$ und $\underset{= Az}{x} \in \text{Bild}(A)$ gilt:

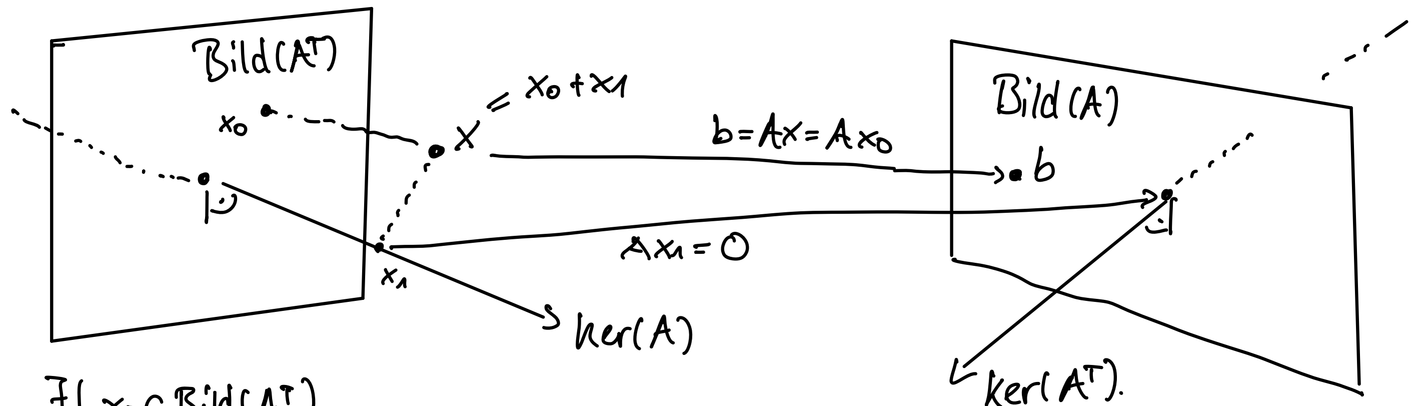
$$\langle y, x\rangle = \langle y, Az\rangle = y^T A z = (\underbrace{A^T y}_{0})^T z = 0.$$

$\Rightarrow \quad \ker(A^T) \perp \mathrm{Bild}(A)$. Das zeigt (1) + (2).

(3) + (4) folgen entsprechend. $\square$.

Wir wollen die Lösung des GLS $Ax = b$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ verstehen im Kontext von Theorem 1.3.

Zunächst nehmen wir an, dass $b \in \mathrm{Bild}(A)$.



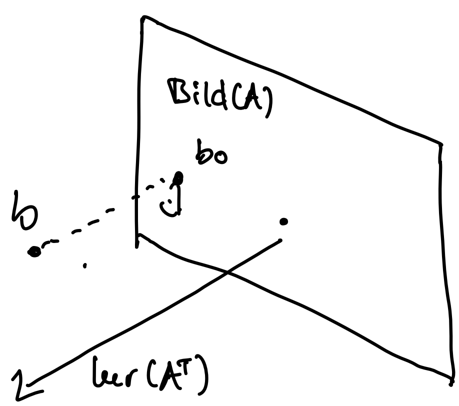
$\exists!\ x_0 \in \mathrm{Bild}(A^T)$
$\exists!\ x_1 \in \ker(A)$ $: \ x = x_0 + x_1$

$$\mathbb{R}^n \xrightarrow{\quad A \quad} \mathbb{R}^m$$

Wenn $b \notin \mathrm{Bild}(A)$, ist $Ax = b$ nicht lösbar. Wir können stattdessen den Punkt $b_0 \in \mathrm{Bild}(A)$ finden, der den Abstand

$$\|b - b_0\| = \sqrt{\langle b - b_0, b - b_0\rangle}$$

minimiert.



$\rightarrow \ b - b_0 =: c \in \ker(A^T)$

$b = b_0 + c$, $\quad b_0 \in \mathrm{Bild}(A)$
$\qquad\qquad\quad\ \ c \in \ker(A^T)$.

Lemma 1.4

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

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

ist eindeutig bestimmt durch

1) die nach Theorem 1.3 eindeutige Zerlegung

$$b = b_0 + c, \qquad b_0 \in \text{Bild}(A), \ c \in \ker(A^T).$$

2) $A^T b = A^T b_0$.

Beweis

eindeutige Zerlegung nach Thm 1.3.

Zunächst: (2) impliziert (1), weil für $b = b_1 + c$, $b_1 \in \text{Bild}(A)$, $c \in \ker(A^T)$

gilt: $A^T b = A^T (b_1 + c) = A^T b_1 = A^T b_0$ und $b_1 - b_0 \in \ker(A^T)$ impliziert $b_1 = b_0$.

Zu 2) Sei $b_0 = A x_0$. Wir definieren

$$\phi(x) = \|b - Ax\| = \sqrt{\langle b - Ax, b - Ax \rangle}, \quad x = (x_1, \dots, x_n)$$

Wir wissen

$$\nabla \phi(x_0) = 0 \overset{\text{Übung}}{\Rightarrow} 0 = A^T (A x_0 - b)$$

$$\Rightarrow A^T A x_0 = A^T b \Rightarrow A^T b_0 = A^T b$$

$\square$.

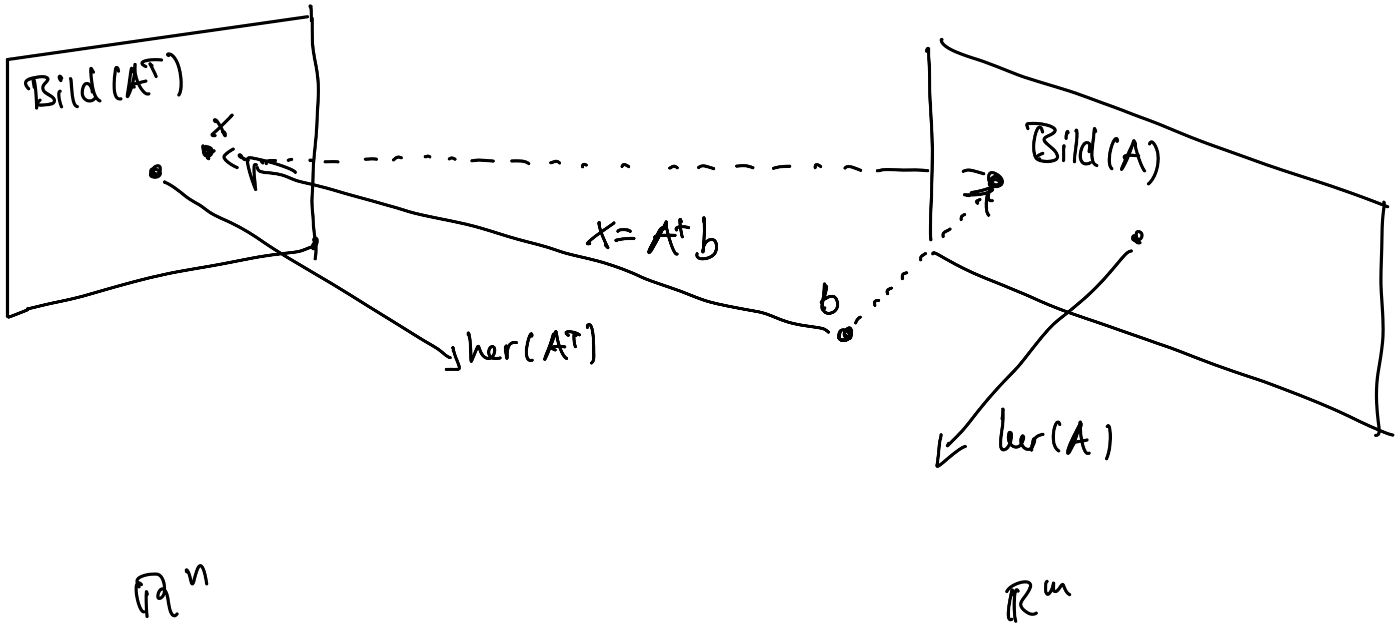
Definition 1.5

Sei $A \in \mathbb{R}^{m \times n}$. Die Pseudo-Inverse $A^+ \in \mathbb{R}^{n \times m}$ ist die Matrix mit

$$A^+ b = x \qquad , b \in \mathbb{R}^m,$$

wobei:

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



<u>Korollar 1.6</u>

1) Falls A invertierbar ist, gilt $A^+ = A^{-1}$.

2) AA^+ ist die orthogonale Projektion auf $\operatorname{Bild}(A)$.