Beispiel 4.1. Das Ohmsche Gesetz:

$$U = RI$$

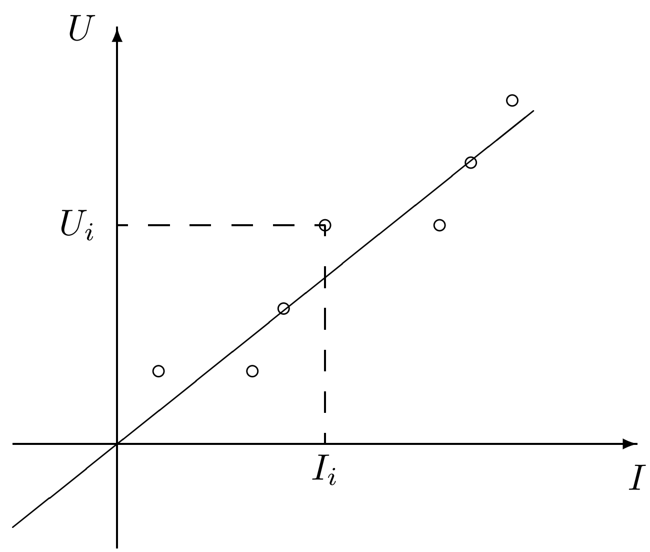
Eine Meßreihe von Daten:

$$(U_i, I_i) \quad \text{(Spannung, Stromstärke)}, \quad i = 1, \dots, m.$$

Aufgabe: man bestimme aus diesen Meßdaten den Widerstand R im Stromkreis. Theoretisch:

$$U_i = RI_i, \quad i = 1, \dots, m.$$

Aber Daten sind mit *Fehlern* behaftet.



Man kann hierzu versuchen, die durch eine Wahl von R bedingten Residuen $U_i - RI_i$ zu quadrieren, aufzusummieren und dasjenige R zu suchen, das diesen Ausdruck minimiert:

$$f(R) := \sum_{i=1}^{m} (RI_i - U_i)^2 = \min .$$

Da f eine quadratische Funktion ist, kann nur ein Extremum vorliegen, das durch die Nullstelle der Ableitung gegeben ist:

$$0 = f'(R) = \sum_{i=1}^{m} 2(RI_i - U_i)I_i = 2R\Big(\sum_{i=1}^{m} I_i^2\Big) - 2\sum_{i=1}^{m} U_i I_i.$$

Hier ergibt sich diese Nullstelle R^* als

$$R^* = \Big(\sum_{i=1}^{m} U_i I_i\Big)\Big/\Big(\sum_{i=1}^{m} I_i^2\Big). \qquad \triangle$$

Beispiel 4.2.

In der Fourieranalyse wird eine T-periodische Funktion f durch eine Linearkombination der T-periodischen trigonometrischen Polynome

$$1,\ \cos(ct),\ \sin(ct),\ \cos(2ct),\ \sin(2ct), \ldots,\ \cos(Nct),\ \sin(Nct)$$

mit $c := \frac{2\pi}{T}$ in der Form

$$g_N(t) = \frac{1}{2}a_0 + \sum_{k=1}^{N} \left(a_k \cos(kct) + b_k \sin(kct)\right),$$

approximiert.

Annahme: nicht f, sondern nur eine Reihe von Meßdaten

$$b_i \approx f(t_i), \quad 0 \le t_1 < t_2 < \ldots < t_m \le T,$$

ist bekannt, wobei $m > 2N + 1$.

Ansatz zur Bestimmung der Koeffizienten $a_0, a_1, b_1, a_2, b_2, \ldots, a_N, b_N$:

$$\sum_{i=1}^{m} \left(g_N(t_i) - b_i\right)^2 = \min. \qquad \triangle$$

Das allgemeine lineare Ausgleichsproblem

Das Minimierungsproblem

$$\sum_{i=1}^{m} (y(t_i; x_1, \ldots, x_n) - b_i)^2 = \sum_{i=1}^{m} (a_{i,1}x_1 + \ldots + a_{i,n}x_n - b_i)^2 = \min$$

wird in Matrixform dargestellt. Setzt man

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

nimmt das Minimierungsproblem eine kompakte Form an:

$$\|Ax - b\|_2^2 = \min_{x \in \mathbb{R}^n} .$$

Also:

Zu gegebenem $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^n$, bestimme $x^* \in \mathbb{R}^n$, für das

$$\|Ax^* - b\|_2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2$$

gilt.

Man vermutet, daß die Meßdaten

t	0	1	2	3
y	3	2.14	1.86	1.72

einer Gesetzmäßigkeit der Form

$$y = f(t) = \alpha \frac{1}{1+t} + \beta$$

mit noch zu bestimmenden Parametern $\alpha, \beta \in \mathbb{R}$ gehorchen. Das zugehörige lineare Ausgleichsproblem hat die Gestalt (4.14), mit

$$x = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 \\ \frac{1}{2} & 1 \\ \frac{1}{3} & 1 \\ \frac{1}{4} & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 2.14 \\ 1.86 \\ 1.72 \end{pmatrix}. \qquad \triangle$$

Die Lösung von (4.14) läßt sich auf die Lösung des linearen Gleichungssystems

$$A^TAx = A^Tb$$

reduzieren, das häufig als *Normalgleichungen* bezeichnet wird. Beachte: für $A \in \mathbb{R}^{m\times n}$ ist die Matrix $A^TA \in \mathbb{R}^{n\times n}$ stets *quadratisch*.

Kernaussage:

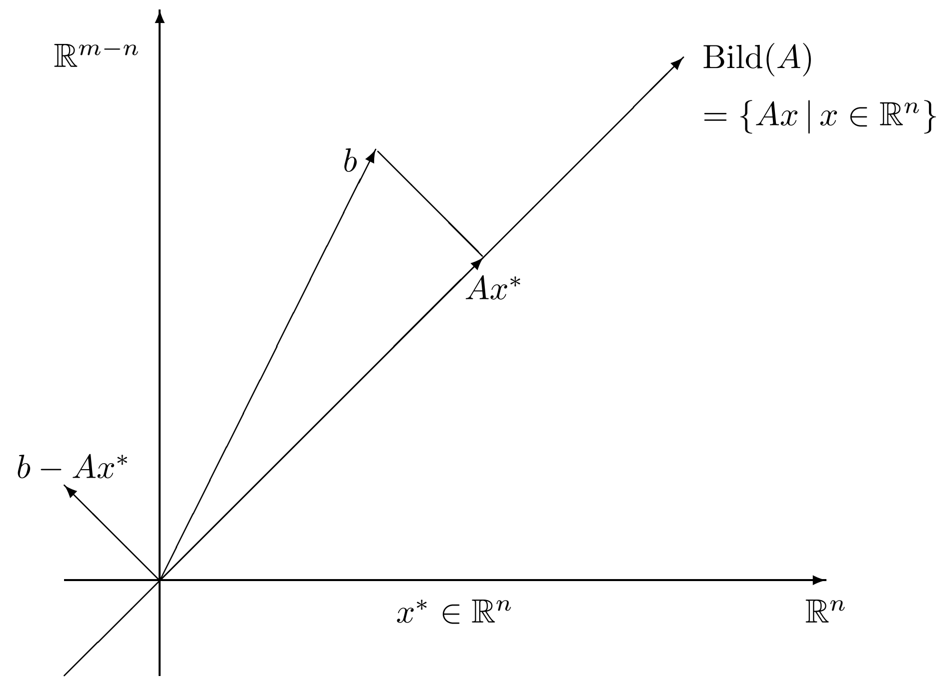
Satz 4.5. $x^* \in \mathbb{R}^n$ ist genau dann Lösung des linearen Ausgleichsproblems (4.14), wenn x^* Lösung der Normalgleichungen

$$A^TAx = A^Tb$$

ist. Das System der Normalgleichungen hat stets mindestens eine Lösung. Sie ist genau dann *eindeutig*, wenn $\text{Rang}(A) = n$ gilt.

Anschaulich ist klar, daß die Differenz $b - Ax$ gerade *senkrecht* auf dem Bildraum Bild$(A) = \{Ax \,|\, x \in \mathbb{R}^n\}$ stehen muß, damit der Abstand $\|Ax - b\|_2$ minimal ist. Also gilt:

$$\|Ax - b\|_2 = \min \iff Ax - b \perp \text{Bild}(A)\,,$$



In Beispiel 3.32 wurde bereits folgende Tatsache gezeigt:

Bemerkung 4.6. Falls $A \in \mathbb{R}^{m \times n}$ vollen (Spalten-)Rang n hat, so ist die Matrix $A^TA \in \mathbb{R}^{n \times n}$ symmetrisch positiv definit.

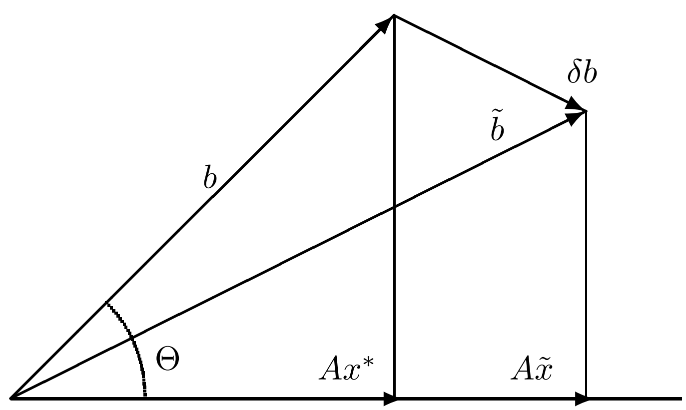
Annahme: *Wir beschränken uns in den Abschnitten 4.3 und 4.4 auf den Fall, daß A vollen Spaltenrang hat:* $\mathrm{Rang}(A) = n$.

Der Fall $\mathrm{Rang}(A) < n$ wird in Abschnitt 4.7 diskutiert.

$$\kappa_2(A) := \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \Big/ \min_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

Ausgleichsproblem mit Störungen:

$$\|A\tilde{x} - \tilde{b}\|_2 = \min .$$



Satz 4.7. Für die Kondition des linearen Ausgleichsproblems bezüglich Störungen in b gilt

$$\frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2} \leq \frac{\kappa_2(A)}{\cos \Theta} \frac{\|\tilde{b} - b\|_2}{\|b\|_2}.$$

Beispiel 4.8.

$$A := \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad b := \begin{pmatrix} 0.01 \\ 1 \\ 0 \end{pmatrix}.$$

Man kann einfach nachrechnen, daß $\kappa_2(A) \approx 2.62$ und

$$x^* = (A^T A)^{-1} A^T b = \begin{pmatrix} 0.01 \\ 0 \end{pmatrix}$$

gilt. Für $\tilde{b} = (0.01, 1, 0.01)^T$ erhält man

$$\tilde{x} = (A^T A)^{-1} A^T \tilde{b} = \begin{pmatrix} 0 \\ 0.01 \end{pmatrix}.$$

Daraus folgt

$$\frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2} \approx 100 \frac{\|\tilde{b} - b\|_2}{\|b\|_2},$$

also eine schlechte Kondition. Es gilt

$$\cos \Theta = \frac{\|Ax^*\|_2}{\|b\|_2} = 0.01. \qquad \triangle$$

Satz 4.9. Für die Kondition des linearen Ausgleichsproblems bezüglich Störungen in A gilt

$$\frac{\|\tilde{x}-x^*\|_2}{\|x^*\|_2} \leq \left(\kappa_2(A)+\kappa_2(A)^2 \tan\Theta\right) \frac{\|\tilde{A}-A\|_2}{\|A\|_2}.$$

Lösung der Normalgleichungen

Die Matrix A^TA ist symmetrisch positiv definit.
Folglich ergibt sich die Methode:

- Berechne A^TA, A^Tb.
- Berechne die Cholesky-Zerlegung

$$LDL^T = A^TA$$

  von A^TA.
- Löse

$$Ly = A^Tb, \quad L^Tx = D^{-1}y$$

  durch Vorwärts- bzw. Rückwärtseinsetzen.

Nachteile dieser Vorgehensweise

- Die Berechnung von A^TA ist für große m aufwendig und birgt die Gefahr von Genauigkeitsverlust durch Auslöschungseffekte. Die Einträge von A^TA sind also mit (möglicherweise erheblichen relativen) Fehlern behaftet.

- Bei der Lösung des Systems $A^TAx = A^Tb$ über das Cholesky-Verfahren werden die Rundungsfehler in A^TA und A^Tb mit

$$\kappa_2(A^TA)$$

verstärkt. Es gilt

$$\kappa_2(A^TA) = \kappa_2(A)^2.$$

Folglich wird die Rundungsfehlerverstärkung durch $\kappa_2(A)^2$ beschrieben.

$$A = \begin{pmatrix} \sqrt{3} & \sqrt{3} \\ \delta & 0 \\ 0 & \delta \end{pmatrix}, \quad b = \begin{pmatrix} 2\sqrt{3} \\ \delta \\ \delta \end{pmatrix}, \quad 0 < \delta \ll 1.$$

Das lineare Ausgleichsproblem $\|Ax - b\|_2 = \min$ hat die Lösung $x^* = (1,1)^T$ (für alle $\delta > 0$). Außerdem gilt $\Theta = 0$.
Daher wird die Kondition dieses Problems durch $\kappa_2(A)$ beschrieben.

Man rechnet einfach nach, daß

$$\kappa_2(A) \approx \frac{\sqrt{6}}{\delta}$$

gilt. Ein stabiles Verfahren sollte ein Resultat $\tilde{x}$ liefern, mit

$$\frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2} \lessapprox \kappa_2(A)\,\mathrm{eps}\,.$$

Die Lösung dieses Problems über die Normalgleichungen und das Cholesky-Verfahren auf einer Maschine mit eps $\approx 10^{-16}$ ergibt:

$$\delta = 10^{-4} \quad : \quad \frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2} \approx 2 * 10^{-8} \approx \frac{1}{3}\kappa_2(A)^2 \, \mathrm{eps}$$

$$\delta = 10^{-6} \quad : \quad \frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2} \approx 2 * 10^{-4} \approx \frac{1}{3}\kappa_2(A)^2 \, \mathrm{eps}\,.$$

Satz 4.13. Sei $A \in \mathbb{R}^{m\times n}$ mit Rang$(A) = n$ und $b \in \mathbb{R}^m$. Sei $Q \in R^{m\times m}$ eine orthogonale Matrix und $\tilde{R} \in \mathbb{R}^{n\times n}$ eine obere Dreiecksmatrix, so daß

$$QA = R := \begin{pmatrix} \tilde{R} \\ \emptyset \end{pmatrix} \begin{matrix} \} \; n \\ \} \; m-n \end{matrix} \quad .$$

Dann ist die Matrix $\tilde{R}$ regulär. Schreibt man

$$Qb = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \begin{matrix} \} \; n \\ \} \; m-n \end{matrix} ,$$

dann ist $x^* = \tilde{R}^{-1} b_1$ die Lösung des linearen Ausgleichsproblems (4.14). Die Norm $\|Ax^* - b\|_2$ ist gerade durch $\|b_2\|_2$ gegeben.

Grundidee:

$$\|Ax - b\|_2^2 = \|QAx - Qb\|_2^2 = \|Rx - Qb\|_2^2 = \|\tilde{R}x - b_1\|_2^2 + \|b_2\|_2^2.$$

Aus Satz 4.13 ergibt sich nun folgende Methode:

- Bestimme von A die QR-Zerlegung

$$QA = \begin{pmatrix} \tilde{R} \\ \emptyset \end{pmatrix} \qquad (\tilde{R} \in \mathbb{R}^{n \times n}),$$

  z.B. mittels Givens-Rotationen oder Householder-Spiegelungen und berechne $Qb = \binom{b_1}{b_2}$.
- Löse

$$\tilde{R}x = b_1$$

  mittels Rückwärtseinsetzen.

Die Norm des Residuums $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 = \|Ax^* - b\|_2$ ist gerade durch $\|b_2\|_2$ gegeben.

Sei

$$A = \begin{pmatrix} 3 & 7 \\ 0 & 12 \\ 4 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 10 \\ 1 \\ 5 \end{pmatrix},$$

d.h. $m = 3$, $n = 2$. Man bestimme die Lösung $x^* \in \mathbb{R}^2$ von

$$\|Ax - b\|_2 = \min .$$

- Annullierung von $a_{3,1}$:

$$A^{(2)} = G_{1,3}A = \begin{pmatrix} 5 & 5 \\ 0 & 12 \\ 0 & -5 \end{pmatrix}, \qquad b^{(2)} = G_{1,3}b = \begin{pmatrix} 10 \\ 1 \\ -5 \end{pmatrix}.$$

  (In der Praxis werden die Transformationen $G_{1,3}A$ und $G_{1,3}b$ ausgeführt, *ohne* daß $G_{1,3}$ explizit berechnet wird.)

- Annullierung von $a_{3,2}^{(2)}$:

$$A^{(3)} = G_{2,3}A^{(2)} = \begin{pmatrix} 5 & 5 \\ 0 & 13 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \tilde{R} \\ \emptyset \end{pmatrix}, \quad b^{(3)} = G_{2,3}b^{(2)} = \begin{pmatrix} 10 \\ \frac{37}{13} \\ -\frac{55}{13} \end{pmatrix}.$$

Lösung von

$$\begin{pmatrix} 5 & 5 \\ 0 & 13 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 10 \\ \frac{37}{13} \end{pmatrix}$$

durch Rückwärtseinsetzen:

$$x^* = \left(\frac{301}{169}, \frac{37}{169}\right)^T.$$

Als Norm des Residuums ergibt sich:

$$\|b_2\|_2 = \frac{55}{13}. \qquad \triangle$$

Wegen Satz 3.41 gilt

$$\kappa_2(A) = \kappa_2(\tilde{R}),$$

d.h., das Quadrieren der Kondition, das bei den Normalgleichungen auftritt, wird vermieden. Außerdem ist die Berechnung der QR-Zerlegung über Givens- oder Householder-Transformationen ein sehr stabiles Verfahren, wobei die Fehlerverstärkung durch $\kappa_2(A)$ (und nicht $\kappa_2(A)^2$) beschrieben wird.

Beispiel 4.16

Wir nehmen A und b wie in Beispiel 4.12. Die Methode über die QR-Zerlegung von A, auf einer Maschine mit eps $\approx 10^{-16}$, ergibt

$$\delta = 10^{-4} \quad : \quad \frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2} \approx 2.2 * 10^{-16},$$

$$\delta = 10^{-6} \quad : \quad \frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2} \approx 1.6 * 10^{-16}.$$

Wegen der sehr guten Stabilität dieser Methode sind diese Resultate viel besser als die Resultate in Beispiel 4.12. $\triangle$

Gegeben seien Daten $(t_1, y_1), \ldots, (t_m, y_m)$ mit

t_i: feste (deterministische) Meßpunkte,

y_i: Realisierungen von *Zufallsvariablen* Y_i.

Lineare Regression basiert auf dem Ansatz

$$Y_i = \sum_{k=1}^{n} a_k(t_i) x_k + F_i, \qquad i = 1, \ldots, m,$$

$a_k(t)$: geeignete Ansatzfunktionen,

F_i: Meßfehler (Zufallsvariablen).

Ziel: eine *Schätzung* $\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n)^T$ für den unbekannten Parametersatz $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ bestimmen.
Einen solchen Schätzer liefert die lineare Ausgleichsrechnung.
Sei nämlich $\widehat{x}$ die Lösung des linearen Ausgleichsproblems

$$\|Ax - y\|_2^2 \to \min .$$

Dann ist $\widehat{x}$ ebenfalls eine Zufallsvariable.

Annahmen:

- F_i *unabhängig, identisch verteilt* mit Erwartungswert $\mathbb{E}(F_i) = 0$
- *Varianz-Kovarianzmatrix*

$$V(F) := \mathbb{E}(FF^T) = \Big(\mathbb{E}(F_i F_j)\Big)_{i,j=1}^{m} = \sigma^2 I$$

Dann gilt

$$\mathbb{E}(\widehat{x}) = x, \qquad V(\widehat{x}) = \mathbb{E}\Big((\widehat{x} - x)(\widehat{x} - x)^T\Big) = \sigma^2 (A^T A)^{-1}.$$

Der Schätzer ist **erwartungstreu** und hat **minimale Varianz**.

Man spricht von einem *Best Linear Unbiased Estimator* (BLUE).

Annahmen:

- F_i unabhängig, identisch verteilt mit Erwartungswert 0
- Varianz-Kovarianzmatrix $V(F) = \sigma^2 I$
- F_i *normalverteilt*

$\Rightarrow$: Y_i normalverteilt, $\mathbb{E}(Y) = Ax$, $V(Y) = \sigma^2 I$. Dichtefunktion von Y_i:

$$f_i(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-(Ax)_i}{\sigma}\right)^2}.$$

Für die Meßreihe $y_1, \ldots, y_m$ ist die Likelihood-Funktion definiert durch

$$L(x; y_1, \ldots, y_m) := \prod_{i=1}^{m} f_i(y_i) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} e^{-\frac{1}{2\sigma^2}\|y-Ax\|_2^2}.$$

Ein Parameterwert $\tilde{x}$ heißt *Maximum-Likelihood-Schätzwert*, wenn

$$L(\tilde{x}; y_1, \ldots, y_m) \geq L(x; y_1, \ldots, y_m) \quad \text{für alle} \quad x \in \mathbb{R}^n$$

Der Maximum-Likelihood-Schätzer ist gerade der Schätzer $\hat{x}$ aus dem linearen Ausgleichsproblem, weil

$$\|y - A\tilde{x}\|_2 = \min_{x\in\mathbb{R}^n} \|y - Ax\|_2.$$

Gegeben: ein Vektorraum V über $\mathbb{R}$ mit einem Skalarprodukt $\langle \cdot, \cdot \rangle$.

Durch $\|v\| := \langle v, v \rangle^{\frac{1}{2}}$ wird eine Norm auf V definiert.

Aufgabe. Sei $U \subset V$ ein n-dimensionaler Teilraum von V.
Zu $v \in V$ bestimme $u^* \in U$, für das

$$\|u^* - v\| = \min_{u \in U} \|u - v\|$$

gilt.

Im Falle des linearen Ausgleichsproblems ist $U = \text{Bild}(A)$, $A \in \mathbb{R}^{m \times n}$, $V = \mathbb{R}^m$, $\langle u, v \rangle = \sum_{j=1}^{m} u_j v_j$, D.h. $\| \cdot \| = \| \cdot \|_2$.

Bemerkung 4.18. Weil U ein endlich-dimensionaler Teilraum ist, existiert ein Element in U mit minimalem Abstand zu v, d.h. es existiert $u^* \in U$, für das $\|u^* - v\| = \min_{u \in U} \|u - v\|$ gilt. $\triangle$

Satz 4.20. Unter den Bedingungen von Aufgabe 4.17 existiert ein eindeutiges $u^* \in U$, das

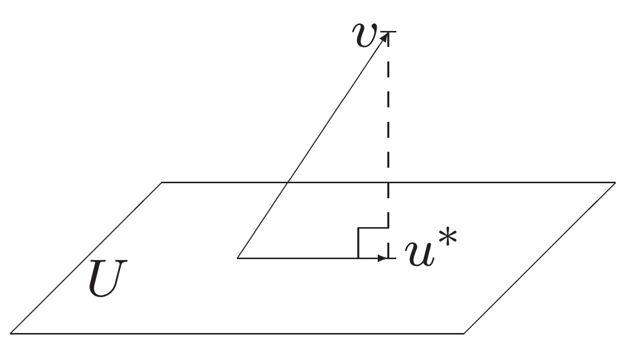
$$\|u^* - v\| = \min_{u \in U} \|u - v\|$$

erfüllt. Ferner gilt das genau dann, wenn

$$\langle u^* - v, u \rangle = 0 \quad \forall u \in U,$$

d.h. , $u^* - v$ senkrecht (bzgl. $\langle \cdot, \cdot \rangle$) zu U ist.
u^* ist somit die **orthogonale Projektion** (bzgl. $\langle \cdot, \cdot \rangle$) von v auf U.

Die Lösung der Aufgabe 4.17 ist also die *orthogonale Projektion* (bzgl. $\langle \cdot, \cdot \rangle$) *von* v *auf* den Unterraum U.



Zu $v \in V$ existiert ein eindeutiges $P_U(v) \in U$, so daß $v - P_U(v) \perp U$, d.h. , $\langle v - P_U(v), u\rangle = 0 \quad \forall \;\; u \in U$.

Mit $P_U : V \to U$ ist also eine wohldefinierte Abbildung gegeben.

(i) Die Abbildung $P_U : V \to U$ ist linear.

(ii) P_U ist ein *Projektor*, d.h. $P_U(u) = u$ für alle $u \in U$ ($P_U^2 = P_U$).

(iii) Die Abbildung P_U ist symmetrisch,

d.h. $\langle P_U(v), w\rangle = \langle v, P_U(w)\rangle, \quad \forall \;\; v, w \in V$.

(iv) P_U is beschränkt und zwar gilt $\|P_U\| = \sup_{\|v\|=1} \|P_U(v)\| = 1$.

Wie kann man $P_U(v)$ *berechnen*?

Sei $\{\phi_1, \ldots, \phi_n\}$ eine *Basis* für U. Dann hat $\hat{u} = P_U(v)$ eine eindeutige Darstellung

$$P_U(v) = \sum_{j=1}^{n} c_j \phi_j$$

mit gewissen Koeffizienten $c_j = c_j(v)$. Es gilt

$$0 \;=\; \langle v - P_U(v), \phi_k \rangle = \langle v, \phi_k \rangle - \sum_{j=1}^{n} c_j \langle \phi_j, \phi_k \rangle, \qquad k = 1, \ldots, n.$$

Definiert man die *Gram-Matrix* $\mathbf{G} := \Big(\langle \phi_k, \phi_j \rangle \Big)_{j,k=1}^{n}$ und die Vektoren $\mathbf{c} = (c_1, \ldots, c_n)^T$, $\mathbf{v} = (\langle v, \phi_1 \rangle, \ldots, \langle v, \phi_n \rangle)^T$ so ergibt sich

$$\mathbf{Gc} = \mathbf{v}$$

Die Berechnung einer orthogonalen Projektion läuft also im allgemeinen auf die Lösung eines symmetrisch positiv definiten Gleichungssystems hinaus.

Wir definieren die Lösungsmenge

$$L(b) := \{x \in \mathbb{R}^n \mid x \text{ ist Lösung des linearen Ausgleichproblems}\}$$

Lemma 4.24. Die Lösungsmenge $L(b)$ hat folgende Eigenschaften:

(i) Es existiert ein **eindeutiges** $x^* \in \mathbb{R}^n$, so daß $x^* = L(b) \cap \text{Kern}(A)^\perp$, wobei $\text{Kern}(A)^\perp := \{z \in \mathbb{R}^n \mid y^T z = 0 \ , \forall \ y \in \text{Kern}(A)\}$.

(ii) Für alle $x \in L(b) \setminus \{x^*\}$ gilt $\|x\|_2 > \|x^*\|_2$, d.h., x^* hat die kleinste Euklidische Norm in $L(b)$.

Folgerung 4.25. Sei $b \in \mathbb{R}^m, \ A \in \mathbb{R}^{m \times n}$. Die Aufgabe

$$\begin{cases} \text{bestimme } x^* \text{ mit minimaler Euklidischer Norm,} \\ \text{für das } \ \|Ax^* - b\|_2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2 \ \text{ gilt,} \end{cases}$$

hat eine eindeutige Lösung.

Satz 4.27. Zu jeder Matrix $A \in \mathbb{R}^{m \times n}$ existieren orthogonale Matrizen $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ und eine Diagonalmatrix

$$\Sigma := \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m \times n}, \quad p = \min\{m, n\} \ ,$$

mit

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0,$$

so daß

$$U^T A V = \Sigma.$$

Die σ_i heißen *Singulärwerte* von A (singular values). Die Spalten der Matrizen U, V nennt man die *Links*- bzw.*Rechtssingulärvektoren*.

Satz 4.28. Sei $U^T AV = \Sigma$ eine Singulärwertzerlegung von $A \in \mathbb{R}^{m\times n}$ mit Singulärwerten

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0,\ p = \min\{m,n\}.$$

Definiere $A^+ \in \mathbb{R}^{n\times m}$ durch

$$A^+ = V\Sigma^+ U^T \quad \text{mit}$$
$$\Sigma^+ = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0) \in \mathbb{R}^{n\times m}\ .$$

Dann ist $A^+ b = x^*$ die Lösung des allgemeinen linearen Ausgleichproblems.
A^+ heißt Pseudoinverse von A.

Lemma 4.29. Sei $U^T A V = \Sigma$ eine Singulärwertzerlegung von A mit Singulärwerten $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0$, $p = \min\{m, n\}$. Die Spalten der Matrizen U und V werden mit u_i bzw. v_i notiert. Dann gilt:

(i) $Av_i = \sigma_i u_i, \quad A^T u_i = \sigma_i v_i, \quad i = 1, \ldots, p.$

(ii) $\mathrm{Rang}(A) = r$.

(iii) $\mathrm{Bild}(A) = \mathrm{span}\{u_1, \ldots, u_r\}, \quad \mathrm{Kern}(A) = \mathrm{span}\{v_{r+1}, \ldots, v_n\}.$

(iv) $\|A\|_2 = \sigma_1$.

(v) Sei $\kappa_2^*(A) := \|A\|_2 \|A^+\|_2 = \frac{\sigma_1}{\sigma_r}$. Falls $\mathrm{Rang}(A) = n \leq m$, so gilt

$$\kappa_2^*(A) = \kappa_2(A) = \frac{\max_{\|x\|_2=1} \|Ax\|_2}{\min_{\|x\|_2=1} \|Ax\|_2}$$

(vi) $\{\, \sigma_i \mid i = 1, \ldots, r \} = \{\, \sqrt{\lambda_i(A^T A)} \mid i = 1, \ldots, n \,\} \setminus \{0\}$.

Abb. 4.6. Orthogonale Basis in $\mathbb{R}^n$ und $\mathbb{R}^m$

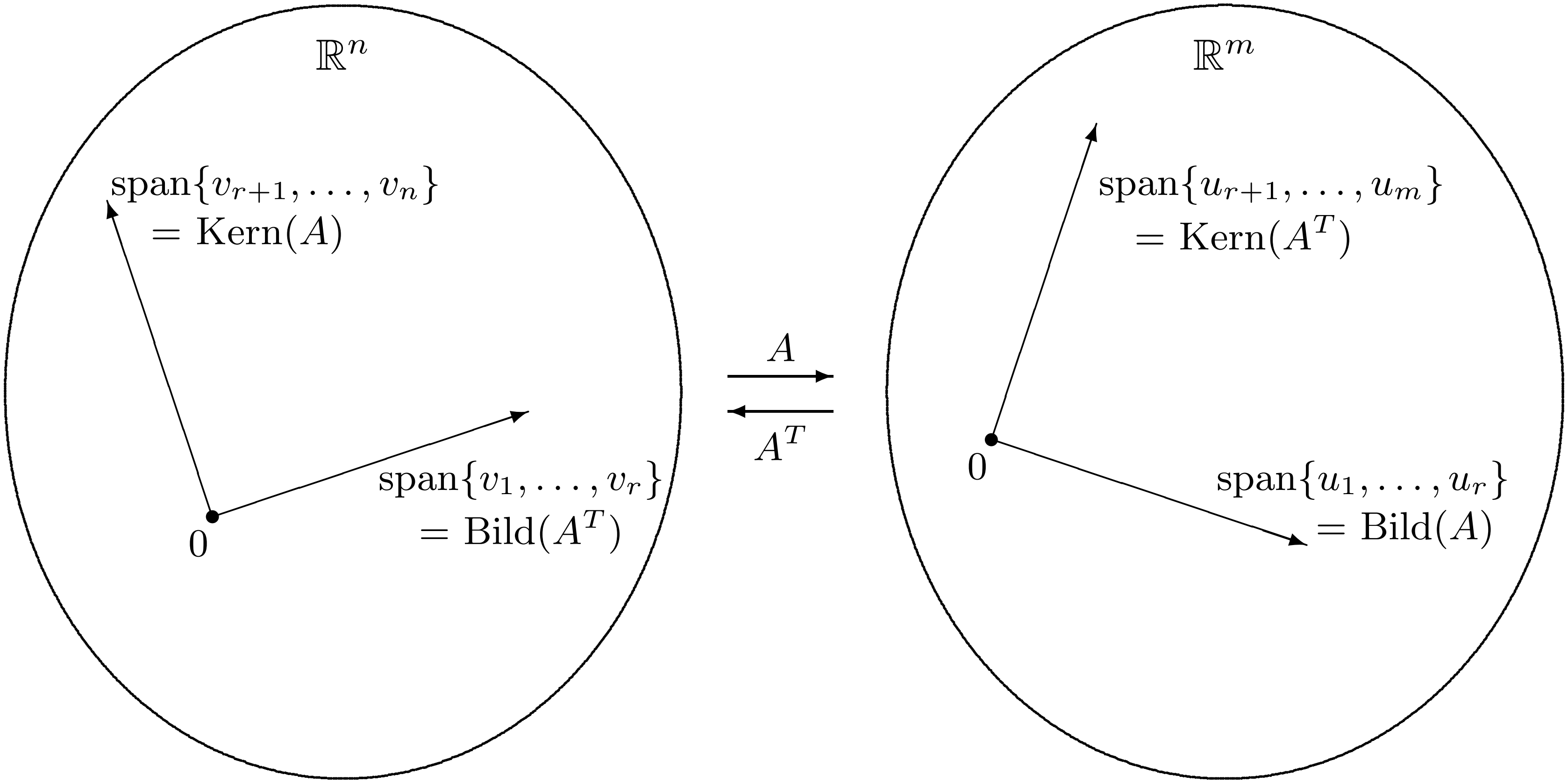
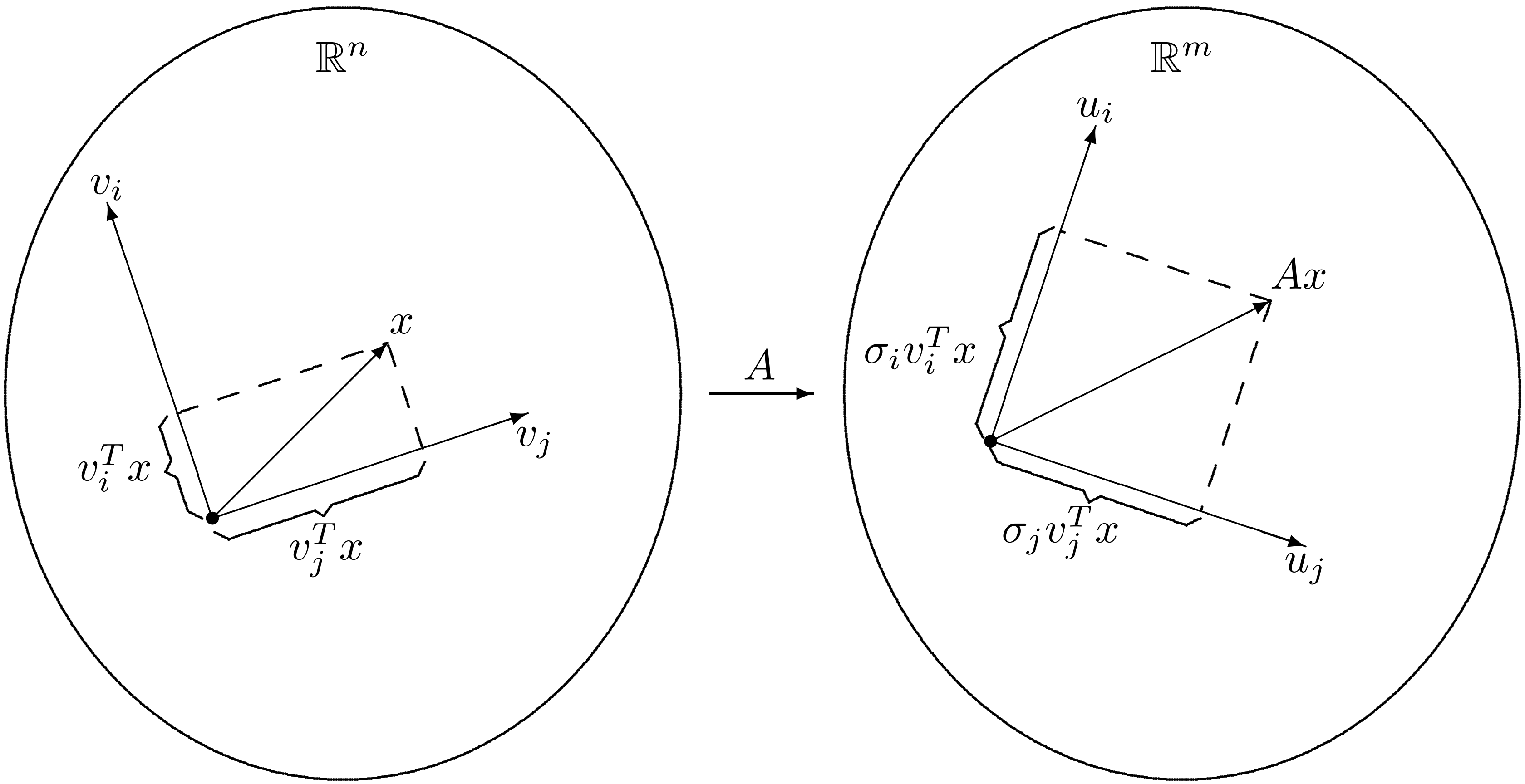


Abb. 4.7. Geometrische Interpretation der Singulärwertzerlegung



Lemma 4.30. Sei $A \in \mathbb{R}^{m\times n}$, und seien $Q_1 \in \mathbb{R}^{m\times m}$, $Q_2 \in \mathbb{R}^{n\times n}$ orthogonale Matrizen. Dann haben A und $Q_1 A Q_2$ die gleichen Singulärwerte.

Transformation auf Bidiagonalgestalt; Beispiel:

Eine Householder-Transformation Q_1, so daß

$$Q_1 A = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix} = \begin{pmatrix} * & v_1^T \\ \emptyset & * \end{pmatrix}$$

Sei $\tilde{Q}_1 \in \mathbb{R}^{3\times 3}$ eine Householder-Transformation, so daß

$\tilde{Q}_1 v_1 = (*\ \ 0\ \ 0)^T$. Mit $\hat{Q}_1 := \begin{pmatrix} 1 & \emptyset \\ \emptyset & \tilde{Q}_1 \end{pmatrix}$ erhält man

$$Q_1 A \hat{Q}_1 = \begin{pmatrix} * & v_1^T \\ \emptyset & * \end{pmatrix} \begin{pmatrix} 1 & \emptyset \\ \emptyset & \tilde{Q}_1 \end{pmatrix} = \begin{pmatrix} * & * & 0 & 0 \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix} .$$

Auf ähnliche Weise können Nulleinträge erzeugt werden in der 2. Spalte, 2. Zeile, 3. Spalte und 4. Spalte:

$$Q_1 A \hat{Q}_1 \ \rightarrow\ Q_2 Q_1 A \hat{Q}_1 = \begin{pmatrix} * & * & 0 & 0 \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix} \ \rightarrow\ Q_2 Q_1 A \hat{Q}_1 \hat{Q}_2 = \begin{pmatrix} * & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & 0 & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix}$$

$$\rightarrow\ Q_3 Q_2 Q_1 A \hat{Q}_1 \hat{Q}_2 = \begin{pmatrix} * & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \end{pmatrix} \rightarrow Q_4 Q_3 Q_2 Q_1 A \hat{Q}_1 \hat{Q}_2 = \begin{pmatrix} * & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

Bemerkung 4.31. Der Aufwand zur Berechnung der oberen Bidiagonalmatrix B in 4.65 beträgt $mn^2 + \mathcal{O}(mn)$ Operationen. $\triangle$

Die Matrix A und die sich ergebende Bidiagonalmatrix B haben die gleichen Singulärwerte.

Die Singulärwerte der Matrix A sind die Wurzeln der Eigenwerte der *Tridiagonalmatrix* $B^T B$.

Für die Berechnung der Eigenwerte dieser Matrix werden im allgemeinen sehr viel weniger arithmetische Operationen benötigt als für die Berechnung der Eigenwerte der (vollbesetzten) Matrix $A^T A$.

Lemma 4.32. Sei $U^T AV = \Sigma$ eine Singulärwertzerlegung von A mit Singulärwerten $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0$, $p = \min\{m,n\}$. Für $0 \leq k \leq p-1$ gilt:

$$\min\{\, \|A - B\|_2 \mid B \in \mathbb{R}^{m\times n},\ \mathrm{Rang}(B) \leq k \,\} = \sigma_{k+1}\ .$$

Folgerung 4.33. Für A und $\tilde{A} = A + \Delta A \in \mathbb{R}^{m\times n}$ mit Singulärwerten $\sigma_1 \geq \ldots \geq \sigma_p$ bzw. $\tilde{\sigma}_1 \geq \ldots \geq \tilde{\sigma}_p$, $p = \min\{m,n\}$, gilt

$$\frac{|\sigma_k - \tilde{\sigma}_k|}{|\sigma_1|} \leq \frac{\|\Delta A\|_2}{\|A\|_2}\ , \qquad \text{für} \quad k = 1,\ldots,p.$$

In diesem Sinne ist das Problem der Singulärwertbestimmung **gut konditioniert**.

Sei $\tilde{A} = (\tilde{a}_{i,j})$ eine mit Rundungsfehlern behaftete Annäherung von A, wobei $\tilde{a}_{i,j} = a_{i,j}(1 + \epsilon_{i,j})$ mit $|\epsilon_{i,j}| \leq \text{eps}$.
Mit $E := (a_{i,j}\epsilon_{i,j}) \in \mathbb{R}^{m\times n}$ ergibt sich

$$\|A - \tilde{A}\|_2 = \|E\|_2 \leq \sqrt{m}\|E\|_\infty \leq \sqrt{m}\|A\|_\infty \,\text{eps} \leq \sqrt{mn}\|A\|_2 \,\text{eps} \ .$$

Deswegen definieren wir:

$$\mathcal{B}_{\tilde{A}}(\text{eps}) := \{\, C \in \mathbb{R}^{m\times n} \mid \frac{\|\tilde{A} - C\|_2}{\|\tilde{A}\|_2} \leq \sqrt{mn}\,\text{eps}\,\} \ .$$

Der **numerische Rang** $\text{Rang}_{\text{num}}(\tilde{A})$ der Matrix $\tilde{A}$ ist:

$$\text{Rang}_{\text{num}}(\tilde{A}) := \min\{\,\text{Rang}(B) \mid B \in \mathcal{B}_{\tilde{A}}(\text{eps}) \quad \} \ .$$

Dieser numerische Rang hängt von der Maschinengenauigkeit eps ab.

Bestimmung des numerischen Ranges:

Seien $\tilde{\sigma}_1 \geq \ldots \geq \tilde{\sigma}_p$ die Singulärwerte der Matrix $\tilde{A}$. Es gilt

$$\min_{\text{Rang}(B)\leq k} \frac{\|\tilde{A}-B\|_2}{\|\tilde{A}\|_2} = \frac{\tilde{\sigma}_{k+1}}{\tilde{\sigma}_1}\ , \quad 1 \leq k < p\ .$$

Die Umgebung $\mathcal{B}_{\tilde{A}}(\text{eps})$ enthält also eine Matrix B mit $\text{Rang}(B) = k$ genau dann, wenn $\tilde{\sigma}_{k+1}/\tilde{\sigma}_1 \leq \sqrt{mn}$ eps.

Der numerische Rang der Matrix $\tilde{A}$ ist deshalb:

$$\text{Rang}_{\text{num}}(\tilde{A}) = \min\{\, 1 \leq k \leq p \mid \tilde{\sigma}_{k+1} \leq \tilde{\sigma}_1 \sqrt{mn}\ \text{eps}\,\}.$$

Wir betrachten die Matrizen

$$A_1 = \begin{pmatrix} \frac{1}{10} & \frac{1}{3} & 0 \\ \frac{2}{10} & \frac{2}{3} & 3 \\ \frac{3}{10} & \frac{3}{3} & 0 \\ \frac{4}{10} & \frac{4}{3} & 7 \end{pmatrix}, \qquad A_2 = A_1 + 10\ \text{eps} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \end{pmatrix},$$

wobei eps $\approx 2 * 10^{-16}$. Es gilt $\text{Rang}(A_1) = 2$, $\text{Rang}(A_2) = 3$.
Die *berechneten* Singulärwerte dieser Matrizen sind

$$7.776,\ 1.082,\ 1.731 * 10^{-16} \quad \text{bzw.} \quad 7.776,\ 1.082,\ 2.001 * 10^{-15}.$$

In beiden Fällen sind die drei berechneten Singulärwerte strikt positiv, jedoch gilt

$$\text{Rang}_{\text{num}}(A_1) = \text{Rang}_{\text{num}}(A_2) = 2.$$

In der Praxis würde man hieraus schließen, daß beide Matrizen den Rang 2 haben. $\triangle$