§ 9 Fehleranalyse linearer Gleichungssysteme

Aufgabenstellung 9.1:
(siehe auch Bemerkung 6.37 im WS 2015/16)
Gegeben sei folgendes lineare Gleichungssystem über $\mathbb{R}$ mit n Variablen $x_1, \dots, x_n$ und n Gleichungen:

(9.1.1) $\quad a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n = b_1 \quad ,$

$\vdots$

(9.1.i) $\quad a_{i1} \cdot x_1 + \dots + a_{in} \cdot x_n = b_i \; ,$

$\vdots$

(9.1 n) $\quad a_{n1} \cdot x_1 + \dots + a_{nn} \cdot x_n = b_n \; .$

Mit $A := (a_{ij})_{1 \leq i, j \leq n} \in Mat_{n \times n}(\mathbb{R})$ und $b := (b_1, \dots, b_n)^T \in \mathbb{R}^n$, $x := (x_1, \dots, x_n)^T \in \mathbb{R}^n$ können wir auch schreiben:

(9.1) $$A \cdot x = b \; .$$

Wir nehmen im folgenden an: $\quad \det A \neq 0.$
Dann gibt es genau eine Lösung von (9.1), nämlich

(9.1') $$x = A^{-1} \cdot b.$$

Zur praktischen Lösung von (9.1) liegt folgender Ansatz nahe:

Eliminiere x_1 mittels (9.1.1) aus (9.1.2) – (9.1.n),
eliminiere x_2 mittels (9.1.2) aus (9.1.3) – (9.1.n),
$\vdots$
eliminiere x_{n-1} mittels (9.1.n-1) aus (9.1.n) ,
berechne x_n aus (9.1.n),
berechne x_{n-1} aus (9.1.n-1),
$\vdots$
berechne x_1 aus (9.1.1).

Dieses <u>Gaußsche Eliminationsverfahren</u> funktioniert immer unter der Voraussetzung $\det A \neq 0$, wobei zusätzlich zwischen einzelnen Eliminationsschritten eventuell noch gewisse Gleichungen zu vertauschen sind.

Im folgenden soll das Gaußsche Eliminations= verfahren dadurch analysiert werden, daß A als Produkt einfacherer Matrizen geschrieben wird.

<u>Definition 9.2:</u>

i) Eine Matrix $L = (l_{ij})_{1 \leq i, j \leq n} \in Mat_{n \times n}(\mathbb{R})$ heißt eine <u>untere Dreiecksmatrix</u>, wenn für alle i, j mit $1 \leq i < j \leq n$ gilt: $l_{ij} = 0$.

L hat dann also folgende Gestalt:

$$L = \begin{pmatrix} * & & & 0 \\ & * & & \\ & & \ddots & \\ * & & & * \end{pmatrix} .$$

ii) Eine Matrix $R = (r_{ij})_{1 \leq i, j \leq n} \in Mat_{n \times n}(\mathbb{R})$ heißt eine <u>obere Dreiecksmatrix</u>, wenn für alle i, j mit $1 \leq j < i \leq n$ gilt: $r_{ij} = 0$.

R hat dann folgende Gestalt:

$$R = \begin{pmatrix} * & & & * \\ & * & & \\ & & \ddots & \\ 0 & & & * \end{pmatrix} .$$

iii) Eine Matrix $P \in Mat_{n \times n}(\mathbb{R})$ heißt eine <u>Permutationsmatrix</u>, wenn P aus der Einheitsmatrix entsteht, indem gewisse Zeilen- oder Spalten- permutiert werden.

<u>Beispiele für Permutationsmatrizen:</u>

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

<u>Bemerkungen 9.3:</u>

i) Das Produkt zweier unterer (bzw. oberer) Dreiecksmatrizen ist wieder eine untere (bzw. obere) Dreiecksmatrix.
Für gegebene Koeffizienten $l_{ij} \in \mathbb{R}$ mit $1 \leq j < i \leq n$ setzen wir

(9.3 a) $$L_j := \begin{pmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ & 0 & l_{j+1,j} & \ddots & & \\ & & \vdots & 0 & \ddots & \\ & & l_{nj} & & & 1 \end{pmatrix} \qquad \text{für } 1 \leq j \leq n-1.$$

Dann folgt:

(9.3 b) $$L_1 \cdots L_{n-1} = \begin{pmatrix} 1 & & & & 0 \\ l_{21} & 1 & & & \\ \vdots & & \ddots & & \\ l_{n1} & l_{n2} & \dots & l_{n,n-1} & 1 \end{pmatrix}.$$

Auf der linken Seite in (9.3 b) darf dabei die Reihenfolge der Faktoren nicht vertauscht werden!

ii) Eine (untere oder obere) Dreiecksmatrix ist genau dann invertierbar, wenn alle Koeffizienten der Hauptdiagonalen von 0 verschieden sind; denn das Produkt dieser Koeffizienten ist die Determinante. Gegebenenfalls ist die zugehörige inverse Matrix wieder eine untere bzw. obere Dreiecksmatrix.

Speziell folgt für L_j wie in (9.3a):

$$(9.3c) \qquad L_j^{-1} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ 0 & & -l_{j+1\,j} & \ddots & & \\ & & \vdots & & \ddots & \\ & & -l_{nj} & 0 & & 1 \end{pmatrix} \quad \text{für } 1 \leq j \leq n-1.$$

<u>Bemerkung 9.4, Durchführung des Gaußschen Eliminationsverfahrens mit Hilfe von Dreiecksmatrizen:</u>

Wir nehmen hier der Einfachheit halber an: Bei der Durchführung des Gaußschen Eliminationsverfahrens brauchen in (9.1) keine Zeilen (mehr) vertauscht zu werden.

Für $1 \leq k \leq n-1$ erhalten wir nach $k-1$ Eliminationsschritten eine Matrix der Gestalt

$$(A^{(k-1)}, b^{(k-1)}) = \left(\begin{array}{cccc|c} a_{11}^{(k-1)} & & * & * & \\ & \ddots & & & \\ 0 & & a_{k-1\,k-1}^{(k-1)} & & b^{(k-1)} \\ & & & & \\ & & 0 & B^{(k-1)} & \end{array}\right),$$

ausgehend von $A^{(0)} = A$, $b^{(0)} = b$.

Dabei gilt nun:

$$B^{(k-1)} = (a_{ij}^{(k-1)})_{k \le i,j \le n} \in Mat_{(n-k+1) \times (n-k+1)}(\mathbb{R}),$$
$$b^{(k-1)} \in \mathbb{R}^n .$$

Sodann setzen wir

(9.4a) $\quad l_{ik} := a_{ik}^{(k-1)} / a_{kk}^{(k-1)} \quad$ für $k < i \le n$.

Mit L_k wie in (9.3a) folgt dann - zusammen mit (9.3c):

$$(A^{(k)}, b^{(k)}) := L_k^{-1} \cdot (A^{(k-1)}, b^{(k-1)})$$

$$= \left(\begin{array}{ccc|c} a_{11}^{(k)} & * & & \\ & \ddots & & \\ 0 & a_{kk}^{(k)} & * & b^{(k)} \\ 0 & & B^{(k)} & \end{array} \right)$$

für passende Koeffizienten $a_{ij}^{(k)} \in \mathbb{R}$,
$B^{(k)} \in Mat_{(n-k) \times (n-k)}(\mathbb{R})$, $b^{(k)} \in \mathbb{R}^n$.

Am Ende erhalten wir eine obere Dreiecksmatrix

$$R := A^{(n-1)} = \begin{pmatrix} a_{11}^{(n)} & \cdots & a_{1n}^{(n)} \\ & \ddots & \vdots \\ 0 & & a_{nn}^{(n)} \end{pmatrix} .$$

Dabei gilt - laut Durchführung der Eliminationsschritte:

$$L_{n-1}^{-1} \cdot L_{n-2}^{-1} \cdots L_2^{-1} \cdot L_1^{-1} \cdot A = R .$$

Nach Bemerkung 9.3 ist mit den Matrizen $L_1, \dots, L_{n-1}$ auch

$$L := (L_{n-1}^{-1} \cdots L_1^{-1})^{-1} = L_1 \cdots L_{n-1}$$

eine untere Dreiecksmatrix, und es folgt:

$$A = L \cdot R.$$

Berücksichtigen wir auch den Fall, dass vor bzw. während des Eliminationsverfahrens eventuell noch Zeilen zu vertauschen sind, so erhalten wir allgemein:

<u>Satz und Definition 9.5:</u>

Zu der gegebenen $n \times n$-Matrix A über $\mathbb{R}$ mit $\det A \neq 0$ gibt es eine Permutationsmatrix P, eine untere Dreiecksmatrix L und eine obere Dreiecksmatrix R, so dass gilt:

(9.5) $$P \cdot A = L \cdot R.$$

Die Produktdarstellung durch die Matrix $L \cdot R$ heißt eine <u>Dreieckszerlegung</u> der Matrix $P \cdot A$.

<u>Man beachte dabei:</u>

$P \cdot A$ entsteht aus A, indem die <u>Zeilen</u> der Matrix A permutiert werden.

Bemerkung 9.6:

Ist speziell eine Dreieckszerlegung der Gestalt

$$A = L \cdot R$$

gegeben, so kann die Lösung eines linearen Gleichungssystems $A \cdot x = b$ wie in (9.1) zurückgeführt werden auf die beiden - einfacher zu lösenden - Dreieckssysteme

(9.6a) $\quad L \cdot y = b,$

(9.6b) $\quad R \cdot x = y.$

Sind (9.6a) und (9.6b) gelöst, so folgt nämlich:

(9.6c) $\quad A \cdot x = L \cdot (R \cdot x) = L \cdot y = b.$

Beispiel 9.7:

Wir lösen folgendes lineare Gleichungssystem:

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \\ 3 & 1 & 6 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 2 \end{pmatrix}.$$

Wir setzen

$$A := \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \\ 3 & 1 & 6 \end{pmatrix}, \quad L_1 := \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix}$$

und erhalten:

$$L_1^{-1} \cdot A = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \\ 3 & 1 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & -2 & 3 \end{pmatrix} =: A^{(1)}.$$

Mit $L_2 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}$ folgt dann weiter:

$$L_2^{-1} \cdot A^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & -2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix} =: R.$$

Mit $L := L_1 \cdot L_2 = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix}$ folgt nun:

$$A = L_1 \cdot A^{(1)} = L_1 \cdot L_2 \cdot R = L \cdot R.$$

Wir erhalten also die folgenden beiden Dreieckssysteme:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

Die Lösung des ersten linearen Gleichungssystems ist gegeben durch:

$$y_1 = 4,$$
$$y_2 = 7 - 2y_1 = 7 - 8 = -1,$$
$$y_3 = 2 - 3y_1 - 2y_2 = 2 - 12 + 2 = -8.$$

Das zweite lineare Gleichungssystem lautet nun also:

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ -8 \end{pmatrix}.$$

Die Lösung ist gegeben durch:

$$x_3 = -8,$$
$$x_2 = -1 \cdot (-1 - x_3) = 1 + x_3 = -7,$$
$$x_1 = 4 - x_2 - x_3 = 19.$$

<u>Probe:</u>

$$x_1 + x_2 + x_3 = 4,$$
$$2x_1 + x_2 + 3x_3 = 38 - 7 - 24 = 7,$$
$$3x_1 + x_2 + 6x_3 = 57 - 7 - 48 = 2.$$

Fragestellung 9.8:

Sei $A \in Mat_{n \times n}(\mathbb{R})$ invertierbar und $b \in \mathbb{R}^n \setminus \{0\}$. Wie stark kann die Lösung $\tilde{x} \in \mathbb{R}^n$ eines gestörten linearen Gleichungssystems

(9.8) $$(A + \Delta A) \cdot \tilde{x} = b + \Delta b$$

die Lösung $x \in \mathbb{R}^n \setminus \{0\}$ von

(9.1) $$A \cdot x = b$$

verfälschen?

Um diese Fragestellung zu behandeln, müssen wir eine Möglichkeit haben, die „Größe" eines Vektors und einer Matrix durch eine reelle Zahl zu messen. - Dies geschieht nun mittels einer Norm.

Definition 9.9, siehe auch Aufgabe Nr. 22:

Eine Abbildung $\| \; \| : \mathbb{R}^n \to \mathbb{R}$ heißt eine Norm auf $\mathbb{R}^n$, wenn gilt:

(N1) Es ist $\|x\| \geq 0$ für alle $x \in \mathbb{R}^n$; ferner gilt folgende Äquivalenz:
$$\|x\| = 0 \Leftrightarrow x = 0.$$

(N2) Für alle $x \in \mathbb{R}^n$ und alle $\lambda \in \mathbb{R}$ gilt:
$$\|\lambda \cdot x\| = |\lambda| \cdot \|x\|.$$

(N3) Für alle $x, y \in \mathbb{R}^n$ gilt die Dreiecksungleichung:
$$\|x + y\| \leq \|x\| + \|y\|.$$

Beispiele 9.10:

Es sei jeweils $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$.

i) $\|x\|_1 := \sum_{i=1}^{n} |x_i|$ (Summennorm)

ii) $\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2}$ (Euklidische Norm)

iii) $\|x\|_p := \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$ für $p \in \mathbb{N}$ (l_p-Norm)

iv) $\|x\|_\infty := \|x\|_{sup} := \max\limits_{1 \le i \le n} |x_i|$ (Maximumnorm)

Bemerkungen 9.11:

i) Je zwei Normen $\| \;\|$ und $\| \;\|'$ auf $\mathbb{R}^n$ sind in folgendem Sinne äquivalent:
Es gibt - von $\| \;\|$, $\| \;\|'$ und n abhängige - positive reelle Zahlen c_1, c_2, so dass gilt:

(9.11) $\quad \forall x \in \mathbb{R}^n: \quad c_1 \cdot \|x\| \le \|x\|' \le c_2 \cdot \|x\|.$

ii) Definition 9.9 und die obigen Beispiele können für $m, n \in \mathbb{N}$ direkt auf $Mat_{m \times n}(\mathbb{R})$ übertragen werden, weil $Mat_{m \times n}(\mathbb{R})$ mit $\mathbb{R}^{m \cdot n}$ identifiziert werden kann.

Definition 9.12:

Sei $\| \;\|: \mathbb{R}^n \to \mathbb{R}$ eine Norm. Dann ist die zugehörige Matrixnorm $\| \;\|: Mat_{n \times n}(\mathbb{R}) \to \mathbb{R}$ definiert durch

(9.12) $$\|A\| := \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|A \cdot x\|}{\|x\|} = \max_{\substack{x \in \mathbb{R}^n \\ \|x\| = 1}} \|A \cdot x\| .$$

<u>Bemerkungen 9.13:</u>

i) In der rechten Seite in (9.12) wird das Maximum aus Stetigkeitsgründen angenommen.
Die rechte Gleichung in (9.12) gilt wegen

$$\|A\cdot(\lambda\cdot x)\| = \|\lambda\cdot(A\cdot x)\| = |\lambda|\cdot\|A\cdot x\|$$

für alle $x\in\mathbb{R}^n$ (mit $\|x\|=1$) und alle $\lambda\in\mathbb{R}$.

ii) Die Normeigenschaften der Matrixnorm folgen direkt aus den Normeigenschaften der gegebenen Norm.

iii) Die zugehörige Matrixnorm wird manchmal auch mit „lub" bezeichnet (das bedeutet: "least-upper-bound-Norm");
Schreibweise: $\|A\| = \mathrm{lub}(A)$.

iv) Für eine gegebene Norm $\|\ \|: \mathbb{R}^n\to\mathbb{R}$ ist die zugehörige Matrixnorm $\|\ \|$ die kleinste Norm auf $\mathrm{Mat}_{n\times n}(\mathbb{R})$, so dass für alle $A\in\mathrm{Mat}_{n\times n}(\mathbb{R})$ und alle $x\in\mathbb{R}^n$ gilt:

(9.13) $$\|A\cdot x\| \le \|A\|\cdot\|x\|.$$

Dies folgt direkt aus (9.12).

<u>Satz und Definition 9.14:</u>

Die zur Maximumnorm $\|\ \|_\infty: \mathbb{R}^n\to\mathbb{R}$ gehörige Matrixnorm $\|\ \|_\infty: \mathrm{Mat}_{n\times n}(\mathbb{R})\to\mathbb{R}$ ist die <u>Zeilensummennorm</u>, gegeben durch

(9.14) $$\|A\|_\infty = \max_{1\le i\le n}\left(\sum_{j=1}^{n}|a_{ij}|\right)$$

für $A=(a_{ij})_{1\le i,j\le n}$.

Beweis:

Wir erhalten mittels (9.12):

$$\|A\|_\infty = \max_{\substack{x \in \mathbb{R}^n \\ \|x\|_\infty = 1}} \|A \cdot x\|_\infty$$

$$= \max_{\substack{x \in \mathbb{R}^n \\ \|x\|_\infty = 1}} \left(\max_{1 \le i \le n} \left| \sum_{j=1}^{n} a_{ij} \cdot x_j \right| \right) \quad \left(\text{mit } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)$$

$$= \max_{1 \le i \le n} \left(\sum_{j=1}^{n} a_{ij} \cdot \operatorname{sign}(a_{ij}) \right) \quad (\text{für optimales } i \text{ setze hier } x_j := \operatorname{sign}(a_{ij}) \text{ für } 1 \le j \le n)$$

$$= \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|.$$

$\square$

Im folgenden sei $\| \ \| : \mathbb{R}^n \to \mathbb{R}$ eine Norm; die zugehörige Matrixnorm werde ebenfalls mit $\| \ \|$ bezeichnet.

Definition 9.15:

Die Kondition einer invertierbaren Matrix $A \in \mathrm{Mat}_{n \times n}(\mathbb{R})$ ist gegeben durch

(9.15) $\quad \operatorname{cond}(A) := \|A\| \cdot \|A^{-1}\|.$

Lemma 9.16:

i) Für die n-reihige Einheitsmatrix I_n gilt:

(9.16a) $\quad \|I_n\| = 1.$

ii) Für alle $B, C \in \mathrm{Mat}_{n \times n}(\mathbb{R})$ gilt:

(9.16b) $\quad \|B \cdot C\| \le \|B\| \cdot \|C\|.$

iii) Ist $A \in \mathrm{Mat}_{n \times n}(\mathbb{R})$ invertierbar, so gilt:

(9.16c) $\quad \operatorname{cond}(A) \ge 1.$

Beweis:

i) (9.12) impliziert direkt:

$$\|I_n\| = \max_{\substack{x\in\mathbb{R}^n\\ \|x\|=1}} \|I_n\cdot x\| = \max_{\substack{x\in\mathbb{R}^n\\ \|x\|=1}} \|x\| = 1 \quad .$$

ii) ist trivial für $C=0$. Sei nun $C\neq 0$. Dann ist

$$T := \{x\in\mathbb{R}^n : \quad \|C\cdot x\| > 0\} \neq \emptyset .$$

(9.12) liefert weiter:

$$\|B\cdot C\| = \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|(B\cdot C)\cdot x\|}{\|x\|}$$

$$= \max_{x\in T}\left(\frac{\|B\cdot(C\cdot x)\|}{\|C\cdot x\|}\cdot\frac{\|C\cdot x\|}{\|x\|}\right)$$

$$\leq \max_{y\in\mathbb{R}^n\setminus\{0\}} \frac{\|B\cdot y\|}{\|y\|} \cdot \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|C\cdot x\|}{\|x\|}$$

$$= \|B\|\cdot\|C\| .$$

iii) Aus i) und ii) folgt:

$$\mathrm{cond}(A) = \|A\|\cdot\|A^{-1}\| \geq \|A\cdot A^{-1}\| = \|I_n\| = 1.$$

$\square$

Lemma 9.17:

Sei $F\in \mathrm{Mat}_{n\times n}(\mathbb{R})$ mit $\|F\|<1$. Dann ist die Matrix I_n+F invertierbar, und es gilt:

$$(9.17) \qquad \|(I_n+F)^{-1}\| \leq \frac{1}{1-\|F\|} .$$

Beweis:

Aus (9.13) folgt für alle $x\in\mathbb{R}^n\setminus\{0\}$:

$$\|(I_n+F)\cdot x\| = \|x+F\cdot x\| \geq \|x\| - \|F\cdot x\|$$

$$\geq \|x\| - \|F\|\cdot\|x\| = (1-\|F\|)\cdot\|x\| > 0 ;$$

also ist I_n+F invertierbar.

Lemma 9.16 ii) liefert nun für $B := F$ und $C := (I_n + F)^{-1}$:

$$1 = \|(I_n + F)\cdot C\| \geq \|C\| - \|F\cdot C\| \geq \|C\| - \|F\|\cdot\|C\|$$
$$= \|C\|\cdot(1 - \|F\|) > 0.$$

Damit folgt auch:

$$\|(I_n + F)^{-1}\| = \|C\| \leq \frac{1}{1 - \|F\|}.$$

$\square$

Warnung:

Die Lemmata 9.16 und 9.17 gelten nicht für beliebige Matrix-Normen.

Nun können wir zeigen:

Satz 9.18:

Neben den in der Fragestellung 9.8 getroffenen Konventionen gelte zusätzlich:

(9.18 a) $\quad \|\Delta A\|\cdot\|A^{-1}\| < 1.$

Dann ist die gestörte Gleichung (9.8) eindeutig lösbar, und für den relativen Fehler $\frac{\|x - \tilde{x}\|}{\|x\|}$ gilt:

(9.18) $$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \frac{\operatorname{cond}(A)}{1 - \|\Delta A\|\cdot\|A^{-1}\|}\cdot\left(\frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|}\right).$$

Beweis:

Nach Voraussetzung ist

$$\|A^{-1}\cdot\Delta A\| \leq \|A^{-1}\|\cdot\|\Delta A\| < 1.$$

Nach Lemma 9.17 ist somit die Matrix $I_n + A^{-1}\cdot\Delta A$ invertierbar, und es gilt:

$$(9.18b) \qquad \|(I_n + A^{-1}\cdot\Delta A)^{-1}\| \le \frac{1}{1-\|A^{-1}\cdot\Delta A\|} \le \frac{1}{1-\|A^{-1}\|\cdot\|\Delta A\|}\,.$$

Weiter ist auch $A+\Delta A = A\cdot(I_n + A^{-1}\cdot\Delta A)$ invertierbar; also gibt es genau ein $\tilde{x}\in\mathbb{R}^n$, das (9.8) löst.

Wir setzen nun $\Delta x := \tilde{x} - x$ und erhalten:

$$(A+\Delta A)\cdot(x+\Delta x) = b + \Delta b$$

$$\Rightarrow \Delta A\cdot x + (A+\Delta A)\cdot\Delta x = \Delta b \qquad (\text{nach } (9.1))$$

$$\Rightarrow (A+\Delta A)\cdot\Delta x = \Delta b - \Delta A\cdot x$$

$$\Rightarrow (I_n + A^{-1}\cdot\Delta A)\cdot\Delta x = A^{-1}\cdot(\Delta b - \Delta A\cdot x)$$

$$\Rightarrow \Delta x = (I_n + A^{-1}\cdot\Delta A)^{-1}\cdot A^{-1}\cdot(\Delta b - \Delta A\cdot x)\,.$$

Also folgt mittels (9.18b) und wiederholter Anwendung von (9.13):

$$(9.18c) \qquad \|\Delta x\| \le \frac{\|A^{-1}\|}{1-\|A^{-1}\|\cdot\|\Delta A\|}\cdot\left(\|\Delta b\| + \|\Delta A\|\cdot\|x\|\right).$$

Weiter gilt wegen $\|b\| = \|A\cdot x\| \le \|A\|\cdot\|x\|$:

$$\|\Delta b\| + \|\Delta A\|\cdot\|x\| = \|b\|\cdot\frac{\|\Delta b\|}{\|b\|} + \|A\|\cdot\|x\|\cdot\frac{\|\Delta A\|}{\|A\|}$$

$$\le \|A\|\cdot\|x\|\cdot\left(\frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|}\right).$$

Zusammen mit (9.18c) folgt

$$\frac{\|\Delta x\|}{\|x\|} \le \frac{\|A^{-1}\|\cdot\|A\|}{1-\|A^{-1}\|\cdot\|\Delta A\|}\cdot\left(\frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|}\right)$$

wie gewünscht. $\square$

Bemerkung 9.19:

(9.18) ist eine gute Abschätzung, falls cond(A) „nicht wesentlich" größer als 1 ist.

Im Gegensatz dazu studieren wir folgendes

Beispiel 9.20:

Sei $A := \begin{pmatrix} 1 & 0 \\ 1000 & 1 \end{pmatrix}$.

Dann ist $A^{-1} = \begin{pmatrix} 1 & 0 \\ -1000 & 1 \end{pmatrix}$.

Wir betrachten ein zugehöriges lineares Gleichungssystem

$$x_1 = b_1$$
$$\wedge \quad 1000\,x_1 + x_2 = b_2 .$$

Eine kleine Änderung von b_1 bewirkt – bei gleich bleibender Matrix A – natürlich die gleiche Änderung von x_1. Diese kann aber eine große Änderung von x_2 bewirken.

Mit der durch $\| \ \|_\infty$ induzierten Kondition cond = cond_∞ gilt nach Satz 9.14:

$$\|A\| = \|A^{-1}\| = 1000 + 1 = 1001$$

und damit

$$\mathrm{cond}_\infty(A) = \|A\| \cdot \|A^{-1}\| = 1001^2 > 10^6 .$$

Bemerkung 9.21, Rundungsfehleranalyse beim Gaußschen Eliminationsverfahren

Sobald die Gleitpunktoperationen nicht mehr exakt durchgeführt werden können, sollte während der Durchführung der Eliminations= schritte folgendes beachtet werden:

i) Die Matrizen $A^{(k-1)}$ sollten – für $1 \le k < n$ – equilibriert sein; das heißt:

$$\sum_{j=1}^{n} |a_{ij}^{(k-1)}| \approx \sum_{j=1}^{n} |a_{\nu j}^{(k-1)}| \text{ für alle } i, \nu \quad .$$

Das kann durch geeignete Zeilen-Multipli= kationen erreicht werden.

ii) Sodann sollte das Pivotelement $a_{kk}^{(k-1)}$, durch das im k-ten Eliminationsschritt dividiert wird, dem Betrage nach möglichst groß sein; das heißt:

$$|a_{ik}^{(k-1)}| \le |a_{kk}^{(k-1)}| \quad \text{für } k < i \le n$$

und folglich

$$l_{ik} := a_{ik}^{(k-1)} / a_{kk}^{(k-1)} \in [-1, 1] \text{ für } k < i \le n.$$

Das kann durch eine Zeilen-Vertauschung erreicht werden.

Bemerkung 9.22, Das lineare Ausgleichsproblem:
In der Praxis sind oft die Werte gewisser Größen $x_1, \ldots, x_n \in \mathbb{R}$ zu bestimmen, die nicht direkt gemessen oder berechnet werden können. Statt dessen kann man aber andere Größen $y_1, \ldots, y_m \in \mathbb{R}$ ermitteln, die – im einfachsten Fall – linear von $x_1, \ldots, x_n$ abhängen; das heißt, es gibt eine Matrix $A \in \mathrm{Mat}_{m \times n}(\mathbb{R})$ mit:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \tag{9.22}$$

Die Matrix A sei dabei ebenfalls bekannt. Damit $x_1, \ldots, x_n$ eindeutig bestimmt sind, muss $m \geq n$ sein. Für $m > n$ ist (9.22) im allgemeinen ein überbestimmtes lineares Gleichungssystem, das gewöhnlich keine Lösung besitzt, weil $y_1, \ldots, y_m$ mit unvermeidbaren Messfehlern behaftet sind. – Weil aber die Informationen mit der Zahl der durchgeführten Experimente wachsen, ist es eher unzweckmäßig, $m = n$ statt $m \geq n$ zu fordern. Statt dessen wird versucht, eine geeignet erscheinende Ersatzaufgabe zu lösen.

Die Behandlung einer der beiden folgenden Aufgaben liegt nahe, wenn $a_1, \ldots, a_m$ die Zeilenvektoren der Matrix A beschreiben:

i) Suche $x_1, \dots, x_n \in \mathbb{R}$ mit

(9.22 a) $$\left\| \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} - A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_1 = \sum_{k=1}^{m} \left| y_k - a_k \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right| \overset{!}{=} \text{Min}.$$

ii) Suche $x_1, \dots, x_n \in \mathbb{R}$ mit

(9.22 b) $$\left\| \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} - A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_2^2 = \sum_{k=1}^{m} \left(y_k - a_k \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right)^2 \overset{!}{=} \text{Min}.$$

(9.22 b) heißt ein <u>lineares Ausgleichsproblem</u>.
Es mag eher auf der Hand liegen, (9.22 a) statt (9.22 b) zu betrachten. Die Lösung von (9.22 b) ist aber nicht nur einfacher, sondern auch wahrscheinlichkeitstheoretisch eher gerechtfertigt.

<u>Satz 9.23:</u>

Es seien $m, n \in \mathbb{N}$ mit $m \geq n$, und es seien $A \in \mathrm{Mat}_{m \times n}(\mathbb{R})$ und $y = (y_1, \dots, y_m)^T \in \mathbb{R}^m$ gegeben. Dann besitzt das lineare Ausgleichsproblem

$$\|y - A \cdot x\|_2^2 \overset{!}{=} \text{Min}$$

mindestens eine Lösung $x_0 \in \mathbb{R}^n$.
Ist $\mathbb{L}$ die Menge aller Löungen dieses linearen Ausgleichsproblems, so gilt:

(9.23) $$\mathbb{L} = \{x \in \mathbb{R}^n \mid A \cdot x = A \cdot x_0\} = \{x \in \mathbb{R}^n \mid A^T \cdot A \cdot x = A^T \cdot y\}$$

<u>Beweis:</u>

Es sei

$$U_0 := \{A \cdot x \mid x \in \mathbb{R}^n\};$$

das ist der von den <u>Spalten</u> von A aufgespannte Unterraum von $\mathbb{R}^m$.

Zu dem vorgegebenen $y \in \mathbb{R}^m$ gibt es genau ein $u \in U_0$ und genau ein $w \in U_0^{\perp}$ mit

(9.23 a) $\quad y = u + w.$

Laut Definition von U_0 gibt es – mindestens – ein $x_0 \in \mathbb{R}^n$ mit $A \cdot x_0 = u$.

Für dieses x_0 setzen wir nun

$$\mathbb{L}_1 := \{x \in \mathbb{R}^n \mid A \cdot x = A \cdot x_0\}, \; \mathbb{L}_2 := \{x \in \mathbb{R}^n \mid A^T \cdot A \cdot x = A^T \cdot y\}.$$

Zu zeigen ist: $\mathbb{L} = \mathbb{L}_1 = \mathbb{L}_2$.

Dazu zeigen wir zunächst: $A^T \cdot w = 0$.

Dazu genügt es, zu zeigen:

$$\forall v \in \mathbb{R}^n: \qquad v^T \cdot A^T \cdot w = 0.$$

Wegen $v^T \cdot A^T \cdot w \in \mathbb{R}$ erhalten wir:

$$v^T \cdot A^T \cdot w = (v^T \cdot A^T \cdot w)^T = w^T \cdot A \cdot v = \langle w, A \cdot v \rangle = 0.$$

Dabei gilt die letzte Gleichung wegen $A \cdot v \in U_0$ und $w \in U_0^{\perp}$.

Aus der Beziehung $A^T \cdot w = 0$ folgt nun mittels (9.23a):

$$A^T \cdot A \cdot x_0 = A^T \cdot u = A^T \cdot (y - w) = A^T \cdot y.$$

Es ist also $x_0 \in \mathbb{L}_2$ – und damit auch $\mathbb{L}_1 \subseteq \mathbb{L}_2$.

Nun sei $x \in \mathbb{L}_2$. Ist dann $v \in \mathbb{R}^n$ beliebig, so folgt mit $z := A \cdot v - A \cdot x$:

$$z^T \cdot (y - A \cdot x) = (v^T \cdot A^T - x^T \cdot A^T) \cdot (y - A \cdot x)$$

$$= (v^T - x^T) \cdot A^T \cdot (y - A \cdot x) = (v^T - x^T) \cdot (A^T \cdot y - A^T \cdot A \cdot x) = 0;$$

denn es ist $x \in \mathbb{L}_2$. Damit folgt weiter:

$$\|y - A \cdot v\|_2^2 = \|(y - A \cdot x) + (A \cdot x - A \cdot v)\|_2^2 = \|(y - A \cdot x) - z\|_2^2$$

$$= \|y - A \cdot x\|_2^2 + \|z\|_2^2 \geq \|y - A \cdot x\|_2^2.$$

Dabei gilt Gleichheit genau im Falle $z = 0$.

Das bedeutet: v löst das Minimierungsproblem (9.22) genau dann, wenn $A \cdot v = A \cdot x$ ist.
Da $x \in \mathbb{L}_2$ beliebig gewählt wurde, folgt insbesondere: $\mathbb{L}_2 \subseteq \mathbb{L}$.
Wenden wir die letzte Überlegung speziell an für $x = x_0$, so folgt auch: $\mathbb{L} = \mathbb{L}_1$.
Damit ist alles bewiesen. $\square$

Bemerkung 9.24:

Die n Gleichungen des linearen Gleichungssystems

(9.24) $\qquad A^T \cdot A \cdot x = A^T \cdot y$

heißen auch die Normalgleichungen für x.

Beispiel 9.25:

Wir lösen die Normalgleichungen für das überbestimmte lineare Gleichungssystem

$x_1 - x_2 = 0$, $x_1 + x_2 = 0$, $5x_1 + x_2 = 12$.

Hier ist

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 5 & 1 \end{pmatrix}, \quad A^T = \begin{pmatrix} 1 & 1 & 5 \\ -1 & 1 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 0 \\ 0 \\ 12 \end{pmatrix}.$$

Damit folgt weiter:

$$A^T \cdot A = \begin{pmatrix} 27 & 5 \\ 5 & 3 \end{pmatrix}, \quad A^T \cdot y = \begin{pmatrix} 60 \\ 12 \end{pmatrix}.$$

Die Normalgleichungen lauten nun also:

$$27x_1 + 5x_2 = 60 \quad \wedge \quad 5x_1 + 3x_2 = 12.$$

Als Lösung erhalten wir - etwa mittels der Cramerschen Regel:

$$x_1 = 2\frac{1}{7}\ , \quad x_2 = \frac{3}{7}\ .$$

In diesem Zusammenhang ist schließlich noch folgender Satz von Interesse, der etwa mit Hilfe des Orthogonalisierungsverfahrens von Gram-Schmidt (siehe Satz 7.18 im WS 2015/16) bewiesen werden kann:

<u>Satz 9.26, Die QR-Zerlegung:</u>
Seien $m, n \in \mathbb{N}$ mit $m \geq n$, und sei $A \in Mat_{m \times n}(\mathbb{R})$. Dann gibt es eine Matrix $Q \in Mat_{m \times n}(\mathbb{R})$, deren Spalten paarweise orthogonal zueinander sind, und eine obere Dreiecksmatrix $R \in Mat_{n \times n}(\mathbb{R})$, so dass gilt:

(9.26) $\qquad A = Q \cdot R.$