Kapitel 8
Normierte Räume

8.1 Vektor- und Matrixnormen

In diesem vorbereitenden Abschnitt geht es um die Verallgemeinerung des Betrages reeller Zahlen zu *Normen* von Vektoren und Matrizen.

Definition 8.1. Es sei V ein linearer Raum über $\mathbb{R}$. Eine Abbildung

$$\|\cdot\| : V \to \mathbb{R}$$

heißt *Norm*, falls für alle $x, y \in V$ und $\alpha \in \mathbb{R}$ gilt

$$\|x\| \geq 0 , \qquad \|x\| = 0 \Leftrightarrow x = 0 , \tag{8.1}$$
$$\|\alpha x\| = |\alpha| \, \|x\| \qquad \text{(Homogenität)} , \tag{8.2}$$
$$\|x+y\| \leq \|x\| + \|y\| \qquad \text{(Dreiecksungleichung)} . \tag{8.3}$$

Ein linearer Raum, der mit einer Norm versehen ist, heißt *normierter Raum*.

Beispiele:

Der Betrag

$$|x| = x \operatorname{sgn}(x) \text{ auf } V = \mathbb{R} .$$

Die Euklidische Norm

$$\|x\|_2 = (x_1^2 + x_2^2)^{\frac{1}{2}} \text{ auf } V = \mathbb{R}^2 .$$

Die Maximumsnorm

$$\|f\|_\infty = \max_{x \in [a,b]} |f(x)| \text{ auf } V = C[a,b] .$$

Die Maximumsnorm

$$\|x\|_\infty = \max_{i=1,\ldots,n} |x_i| , \qquad x = (x_i)_{i=1}^n \in \mathbb{R}^n \text{ auf } V = \mathbb{R}^n .$$

Die ℓ^p–Norm

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} , \qquad x = (x_i)_{i=1}^n \in \mathbb{R}^n$$

mit $1 \leq p < \infty$ auf $V = \mathbb{R}^n$.

Anmerkung 8.2. Der Nachweis der Eigenschaften (8.1), (8.2), (8.3) für die obigen Beispiele ist einfach. Nur die Dreiecksungleichung für $\|\cdot\|_p$ macht Schwierigkeiten. Sie lautet ausgeschrieben

$$\left(\sum_{i=1}^{n}|x_i+y_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}$$

und heißt *Minkowskische Ungleichung*. Einen Beweis findet man z.B. im Funktionalanalysis–Lehrbuch von Werner [40], Seite 12.

Anmerkung 8.3.

- Es gilt für jedes fest gewählte $x \in \mathbb{R}^n$

$$\lim_{p\to\infty}\|x\|_p = \|x\|_\infty \,.$$

- Im Falle von $p = 2$ erhält man die Euklidische Norm

$$\|x\|_2 = \left(\sum_{i=1}^{n}x_i^2\right)^{\frac{1}{2}} .$$

- Mittels der Bijektion $\varphi : \mathbb{R}^{n,n} \to \mathbb{R}^{n^2}$, definiert durch

$$\varphi\left((a_{ij})_{i,j=1}^{n}\right) = (x_i)_{i=1}^{n^2} \,,$$
$$x_{(i-1)n+j} = a_{ij} \,, \qquad i,j = 1,\ldots,n \,,$$

läßt sich der lineare Raum der $n \times n$–Matrizen $\mathbb{R}^{n,n}$ mit $\mathbb{R}^{n^2}$ identifizieren. Jede Vektornorm auf $\mathbb{R}^{n^2}$ liefert also eine Matrixnorm auf $\mathbb{R}^{n,n}$. Ein Beispiel ist die sogenannte *Frobenius–Norm*

$$\|A\|_F = \left(\sum_{i,j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}} \qquad A = (a_{ij})_{i,j=1}^{n} \in \mathbb{R}^{n,n} \,.$$

Wir wollen den Konvergenzbegriff von $\mathbb{R}$ auf normierte Räume verallgemeinern.

Definition 8.4. Es sei V ein normierter Raum mit der Norm $\|\cdot\|$ und $(x^{(\nu)})_{\nu\in\mathbb{N}}$ eine Folge aus V. Diese Folge heißt *Cauchy–Folge*, falls es zu jedem $\varepsilon > 0$ ein $\nu_0 \in \mathbb{N}$ gibt, so daß

$$\|x^{(\nu)} - x^{(\mu)}\| \leq \varepsilon \qquad \forall \nu,\mu \geq \nu_0 \,.$$

Die Folge heißt *konvergent* gegen $x \in V$, also

$$x^{(\nu)} \to x \,, \qquad \nu \to \infty \,,$$

falls

$$\|x - x^{(\nu)}\| \to 0 \,, \qquad \nu \to \infty \,.$$

V heißt *vollständig*, falls alle Cauchy–Folgen in V konvergent sind. Ein vollständiger normierter linearer Raum heißt *Banach–Raum*.

Wir haben schon gesehen, daß man auf $\mathbb{R}^n$ oder $\mathbb{R}^{n,n}$ viele verschiedene Normen definieren kann. Es stellt sich die Frage, ob es Folgen gibt, die bezüglich der einen Norm konvergieren und bezüglich der anderen nicht. Der folgende Satz besagt, daß so etwas nicht sein kann.

Satz 8.5. *Sei V endlichdimensional, und es seien $\|\cdot\|$ sowie $|||\cdot|||$ zwei verschiedene Normen auf V. Dann gibt es Konstanten c, C, die nicht von x abhängen, mit der Eigenschaft*

$$c\|x\| \leq |||x||| \leq C\|x\| \qquad \forall x \in V \,.$$

Beweis. Der Beweis beruht auf der Kompaktheit der Einheitskugel in $\mathbb{R}^n$. Wir verweisen wieder auf Werner [40], Seite 26.

Anmerkung 8.6. Für unendlichdimensionale Räume, wie z.B. $V = C[a,b]$, ist dieser Satz falsch!

Als Folgerung aus Satz 8.5 wollen wir die Vollständigkeit von $\mathbb{R}^n$ und $\mathbb{R}^{n,n}$ zeigen.

Satz 8.7. *Die linearen Räume $\mathbb{R}^n$ und $\mathbb{R}^{n,n}$ sind vollständig bezüglich jeder beliebigen Norm $\|\cdot\|_{\mathbb{R}^n}$ und $\|\cdot\|_{\mathbb{R}^{n,n}}$.*

Beweis. Es sei $(x^{(\nu)})_{\nu\in\mathbb{N}}$ eine Cauchy–Folge in $\mathbb{R}^n$. Nach Satz 8.5 gibt es c, C unabhängig von x, so daß

$$c\|x\|_{\mathbb{R}^n} \le \|x\|_\infty \le C\|x\|_{\mathbb{R}^n} , \qquad \forall x \in \mathbb{R}^n .$$

Also ist $(x^{(\nu)})_{\nu\in\mathbb{N}}$ auch Cauchy–Folge bezüglich $\|\cdot\|_\infty$. Dann ist aber auch die Folge von Komponenten $(x_i^{(\nu)})_{\nu\in\mathbb{N}}$ für jedes fest gewählte $i = 1,\dots,n$, eine Cauchy–Folge in $\mathbb{R}$. Da $\mathbb{R}$ vollständig ist, gibt es ein $x_i \in \mathbb{R}$ mit

$$x_i^{(\nu)} \to x_i \qquad \nu \to \infty .$$

Führt man diesen Schluß für alle $i = 1,\dots,n$ durch, erhält man einen Vektor $x = (x_i)_{i=1}^n$ mit der Eigenschaft

$$\|x - x^{(\nu)}\|_\infty \to 0 \qquad \nu \to \infty .$$

Die Konvergenz bezüglich $\|\cdot\|_{\mathbb{R}^n}$ folgt aus

$$\|x - x^{(\nu)}\|_{\mathbb{R}^n} \le \frac{1}{c}\|x - x^{(\nu)}\|_\infty \to 0 \qquad \nu \to \infty .$$

Da sich $\mathbb{R}^{n,n}$ (Matrizen) mit $\mathbb{R}^{n^2}$ (Vektoren) identifizieren lassen, sind wir fertig. □

Wir wollen uns nun speziell den Matrizen zuwenden. Auf $\mathbb{R}^{n,n}$ ist durch

$$A \cdot B = C , \qquad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} ,$$

die Multiplikation von Matrizen erklärt. Wir wollen eine Klasse von Matrixnormen betrachten, die mit der Matrix–Multiplikation verträglich ist.

Definition 8.8. Es sei $\|\cdot\|$ eine Norm auf $\mathbb{R}^n$. Dann ist durch

$$\|A\| = \sup_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{\|Ax\|}{\|x\|} , \qquad A \in \mathbb{R}^{n,n} , \tag{8.4}$$

die *zugehörige Matrixnorm* definiert.

Man überzeuge sich davon, daß durch (8.4) überhaupt eine Norm definiert ist!

Anmerkung 8.9. • Es gilt

$$\|Ax\| \le \|A\| \, \|x\| .$$

- Es existiert ein $x^* \in \mathbb{R}^n$ mit $\|x^*\| = 1$ und $\|Ax^*\| = \|A\|$.
- Matrixnormen der Form (8.4) sind mit der Matrixmultiplikation verträglich, d.h. es gilt

$$\|AB\| \le \|A\| \, \|B\| \qquad \forall A, B \in \mathbb{R}^{n,n} , \qquad \text{(Submultiplikativität)} .$$

- Die Norm der Einheitsmatrix I ist

$$\|I\| = 1 .$$

Satz 8.10 (Zeilensummennorm). *Die Matrixnorm*

$$\|A\|_\infty = \max_{i=1,\dots,n} \sum_{j=1}^{n} |a_{ij}|$$

gehört zur Maximumsnorm $\|\cdot\|_\infty$ auf $\mathbb{R}^n$.

Beweis. Für alle $x \in \mathbb{R}^n$ gilt

$$\begin{aligned}\|Ax\|_\infty &= \max_{i=1,\dots,n}\left|\sum_{j=1}^n a_{ij}x_j\right| \\ &\le \left(\max_{i=1,\dots,n}\sum_{j=1}^n |a_{ij}|\right)\max_{j=1,\dots,n}|x_j| = \|A\|_\infty\|x\|_\infty\ .\end{aligned}$$

Es folgt

$$\sup_{x\neq 0}\frac{\|Ax\|_\infty}{\|x\|_\infty} \le \|A\|_\infty\ .$$

Wir konstruieren nun ein $x^* \in \mathbb{R}^n$ mit der Eigenschaft

$$\|A\|_\infty\|x^*\|_\infty = \|Ax^*\|_\infty\ .$$

Sei dazu i_0 so gewählt, daß

$$\|A\|_\infty = \sum_{j=1}^n |a_{i_0 j}|\ .$$

Dann sei

$$x_j^* = \ \mathrm{sgn}(a_{i_0 j}) \qquad j = 1,\dots,n\ .$$

Es folgt offenbar $\|x^*\|_\infty = 1$ und

$$\|A\|_\infty\|x^*\|_\infty \ge \|Ax^*\|_\infty \ge \left|\sum_{j=1}^n a_{i_0 j}x_j^*\right| = \sum_{j=1}^n |a_{i_0 j}| = \|A\|_\infty\|x^*\|_\infty\ .$$

□

Kapitel 9
Lineare Gleichungssysteme

9.1 Kondition

Wir betrachten das lineare Gleichungssystem

$$Ax = b\ . \tag{9.1}$$

Dabei seien die *Koeffizientenmatrix* $A = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n,n}$ sowie die *rechte Seite* $b = (b_i)_{i=1}^n \in \mathbb{R}$ gegeben und die Lösung $x = (x_i)_{i=1}^n \in \mathbb{R}$ gesucht.

Satz 9.1. *Die Koeffizientenmatrix A sei regulär, d.h.*

$$Ax \neq 0 \qquad \forall x \in \mathbb{R}^n\ , \quad x \neq 0\ .$$

Dann ist (9.1) *eindeutig lösbar mit der Lösung* $x = A^{-1}b$.

Im folgenden sei $A \in \mathbb{R}^{n,n}$ regulär und $b \neq 0$, so daß eine eindeutig bestimmte Lösung $x \neq 0$ existiert. Außerdem wird mit $\|\cdot\|$ durchweg eine Vektornorm *und* die zugehörige Matrixnorm bezeichnet.

Wir wollen nun die Auswirkungen von Eingabefehlern in den Daten A, b auf die Lösung x untersuchen. Dabei spielt die sogenannte Kondition von A eine zentrale Rolle.

Definition 9.2. Sei $A \in \mathbb{R}^{n,n}$. Ist A eine reguläre Matrix, so heißt

$$\kappa(A) = \|A\|\|A^{-1}\|$$

Kondition von A. Ist A singulär, so wird $\kappa(A) = \infty$ gesetzt.

Anmerkung 9.3. Es gibt eine MATLAB–Funktion `cond`.

Es gilt

$$\kappa(A) = \|A\|\|A^{-1}\| \geq \|A^{-1}A\| = \|I\| = 1\ .$$

Insbesondere ist $\kappa(I) = 1$. Eine weitere wichtige Eigenschaft ist die Submultiplikativität der Kondition. Sind A und B regulär, so folgt aus der Submultiplikativität der Matrix-Norm (vgl. die Bemerkung zu Definition 8.8)

$$\|AB\| \leq \|A\|\|B\|\ , \qquad \|(AB)^{-1}\| = \|B^{-1}A^{-1}\| \leq \|A^{-1}\|\|B^{-1}\|$$

und daraus unmittelbar

$$\kappa(AB) \leq \kappa(A)\kappa(B)\ .$$

Als erstes betrachten wir Störungen der rechten Seite b.

Satz 9.4. *Sei x die Lösung von* (9.1) *und* $\tilde{x}$ *die Lösung des gestörten Systems*

$$A\tilde{x} = \tilde{b} \tag{9.2}$$

mit beliebigem $\tilde{b} \in \mathbb{R}^n$. *Dann gilt*

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \kappa(A)\ \frac{\|b - \tilde{b}\|}{\|b\|}\ . \tag{9.3}$$

Es existieren rechte Seiten b, $\tilde{b} \in \mathbb{R}^n$, so daß in (9.3) *Gleichheit vorliegt.*

Beweis. Aus der Definition der zur Vektornorm $\|\cdot\|$ gehörigen Matrixnorm folgt

$$\|x-\tilde{x}\| = \|A^{-1}b - A^{-1}\tilde{b}\| \le \|A^{-1}\|\,\|b-\tilde{b}\|\,, \qquad \|b\| = \|Ax\| \le \|A\|\|x\|. \tag{9.4}$$

Einsetzen ergibt die Abschätzung

$$\frac{\|x-\tilde{x}\|}{\|x\|} \le \frac{\|Ax\|}{\|x\|}\|A^{-1}\|\frac{\|b-\tilde{b}\|}{\|b\|} \le \|A\|\,\|A^{-1}\|\frac{\|b-\tilde{b}\|}{\|b\|}$$

und damit (9.3).

Nach der Bemerkung im Anschluß an die Definition 8.8 der Matrixnorm existieren x^* und b^* mit den Eigenschaften $\|x^*\| = \|b^*\| = 1$, $\|Ax^*\| = \|A\|$ und $\|A^{-1}b^*\| = \|A^{-1}\|$. Damit gilt bei Wahl von $b = Ax^*$, $\tilde{b} = b - \varepsilon b^*$ und $\varepsilon > 0$ das Gleichheitszeichen in den beiden Abschätzungen (9.4). Einsetzen liefert das Gleichheitszeichen in (9.3). □

Die Kondition $\kappa(A)$ ist also eine obere Schranke für die Verstärkung des relativen Fehlers in der rechten Seite b. Eine kleinere Schranke, die *gleichmäßig für alle* $b \in \mathbb{R}^n$ besteht, gibt es nicht.

In welchem Zusammenhang steht dieses Resultat mit unserem Konditionsbegriff aus Abschnitt 6 ? Um diese Frage zu beantworten, bemerken wir zunächst, daß sich die Definition 6.11 der relativen Kondition $\kappa_{\text{rel}}(\xi_0)$ der Auswertung einer Funktion f an einer Stelle ξ_0 direkt auf Funktionen $f : \mathbb{R}^n \to \mathbb{R}^m$ erweitern läßt[1]. Dazu hat man in (6.14) nur die Beträge durch die jeweiligen Vektornormen zu ersetzen. Nun betrachten wir die Funktion $f : \mathbb{R}^n \to \mathbb{R}^n$, definiert durch

$$f(\xi) = A^{-1}\xi\,.$$

Die Auswertung von f an der Stelle $b \in \mathbb{R}^n$ ist offenbar gleichbedeutend mit der Lösung des linearen Gleichungssystems (9.1). Satz 9.4 besagt nun nichts weiter als

$$\kappa_{\text{rel}}(b) \le \kappa(A) \qquad\qquad \forall\, b \in \mathbb{R}^n\,.$$

Wir betrachten als nächstes die Auswirkung von Störungen der Koeffizientenmatrix A auf die Lösung x von (9.1). Diesmal müssen wir einen längeren Anlauf nehmen. Der Grund liegt darin, daß der entsprechende Lösungsoperator $f : \{X \in \mathbb{R}^{n,n} \mid X \text{ regulär}\} \to \mathbb{R}^n$, definiert durch

$$x = f(X) = X^{-1}b\,, \qquad X \in \mathbb{R}^{n,n}\,,$$

nun nichtlinear ist. Insbesondere ist nicht zu erwarten, daß ein gestörtes System

$$\tilde{A}\tilde{x} = b$$

mit *beliebig* gestörter Matrix $\tilde{A} \in \mathbb{R}^{n,n}$ überhaupt eine Lösung $\tilde{x}$ hat.

Lemma 9.5. *Es sei $C \in \mathbb{R}^{n,n}$ und $\|C\| < 1$. Dann ist $I - C$ regulär, und es gilt*

$$(I-C)^{-1} = I + \sum_{k=1}^{\infty} C^k \qquad \textit{(Neumannsche Reihe).}$$

Beweis. Erinnerung: $\mathbb{R}^{n,n}$ versehen mit $\|\cdot\|$ ist vollständig! Wir zeigen zunächst, daß die Partialsummen

$$S_n = \sum_{k=0}^{n} C^k$$

eine Cauchy–Folge in $\mathbb{R}^{n,n}$ bilden. Dies folgt aus

$$\|S_n - S_m\| \le \left\| \sum_{k=n}^{m} C^k \right\| \le \sum_{k=n}^{m} \|C^k\| \le \sum_{k=n}^{m} \|C\|^k \to 0 \qquad n,m \to \infty\,,$$

[1] Wir bezeichnen die unabhängige Variable ab hier mit ξ, um Konflikte mit der Lösung x von (9.1) zu vermeiden.

denn die geometrische Reihe $\sum_{k=0}^{n} \|C\|^k$ konvergiert in $\mathbb{R}$. Aus der Vollständigkeit von $\mathbb{R}^{n,n}$ folgt die Konvergenz von S_n. Es gibt also ein $S \in \mathbb{R}^{n,n}$, so daß

$$\|S_n - S\| \to 0 \qquad n \to \infty .$$

Nun folgt

$$S_n(I-C) = S_n - S_nC = I + S_n - S_{n+1} \to I \qquad n \to \infty .$$

Andererseits gilt

$$S_n(I-C) \to S(I-C) \qquad n \to \infty .$$

Eindeutigkeit des Grenzwertes liefert

$$S(I-C) = I .$$

□

Satz 9.6. *Sei x die Lösung von* (9.1) *und* $\tilde{A} \in \mathbb{R}^{n,n}$ *eine gestörte Koeffizientenmatrix mit der Eigenschaft*

$$\frac{\|A-\tilde{A}\|}{\|A\|} < \frac{1}{\kappa(A)} . \tag{9.5}$$

Dann existiert eine eindeutig bestimmte Lösung $\tilde{x}$ *des gestörten Systems*

$$\tilde{A}\tilde{x} = b , \tag{9.6}$$

und es gilt

$$\frac{\|x-\tilde{x}\|}{\|x\|} \le \kappa(A) \, \frac{\|A-\tilde{A}\|}{\|A\|} + o(\|A-\tilde{A}\|) . \tag{9.7}$$

Es existieren Koeffizientenmatrizen $A, \tilde{A} \in \mathbb{R}^{n,n}$, *so daß in* (9.7) *Gleichheit vorliegt.*

Beweis. Wir setzen $C = A^{-1}(A-\tilde{A})$. Aus der Voraussetzung (9.5) folgt dann

$$\|C\| \le \|A^{-1}\| \, \|A-\tilde{A}\| < 1 . \tag{9.8}$$

Auf Grund von Lemma 9.5 ist daher $I-C$ regulär. Wegen

$$\tilde{A} = A(I-C)$$

ist auch $\tilde{A}$ als Produkt zweier regulärer Matrizen regulär.

Wir kommen zum Nachweis von (9.7). Es gilt

$$\tilde{A}^{-1} = (A(I-C))^{-1} = (I-C)^{-1}A^{-1} = \left(I + C + \sum_{k=2}^{\infty} C^k \right) A^{-1} = A^{-1} + CA^{-1} + C^2(I-C)^{-1}A^{-1} . \tag{9.9}$$

Wegen (9.8) ist

$$\|(I-C)^{-1}\| \le \sum_{k=0}^{\infty} \|C\|^k = \frac{1}{1-\|C\|} \le \frac{1}{1-\|A^{-1}\| \, \|A-\tilde{A}\|} . \tag{9.10}$$

Durch Einsetzen der Beziehung (9.9) und der Abschätzungen (9.8) und (9.10) folgt

$$\begin{aligned}
\|x-\tilde{x}\| &= \|A^{-1}b - \tilde{A}^{-1}b\| \\
&= \|A^{-1}b - A^{-1}b - CA^{-1}b - C^2(I-C)^{-1}A^{-1}b\| \\
&\le \left(\|C\| + \|C\|^2 \|(I-C)^{-1}\| \right) \|x\| \\
&\le \left(\|A^{-1}\| + \frac{\|A^{-1}\|^2 \|A-\tilde{A}\|}{1-\|A^{-1}\| \, \|A-\tilde{A}\|} \right) \|A-\tilde{A}\| \, \|x\| \\
&= \left(\|A\| \, \|A^{-1}\| + \frac{\|A\| \, \|A^{-1}\|^2 \, \|A-\tilde{A}\|}{1-\|A^{-1}\| \, \|A-\tilde{A}\|} \right) \frac{\|A-\tilde{A}\|}{\|A\|} \|x\| .
\end{aligned}$$

Damit ist (9.7) beweisen.

Bei Wahl von $A = I$ und $\tilde{A} = (1+\varepsilon)I$ ist offenbar $\kappa(A) = 1$ und $\|A - \tilde{A}\|/\|A\| = \varepsilon$. Für beliebiges $b \in \mathbb{R}^n$, $b \neq 0$, ist dann $x = b$ und $\tilde{x} = (1+\varepsilon)^{-1} b$. Einsetzen liefert

$$\frac{\|x - \tilde{x}\|}{\|x\|} = \frac{\varepsilon}{1+\varepsilon} = \varepsilon + o(\varepsilon) = \kappa(A)\, \frac{\|A - \tilde{A}\|}{\|A\|} + o(\|A - \tilde{A}\|)\,.$$

□

Schließlich notieren wir noch ohne Beweis ein Resultat zur Auswirkung von Störungen von Koeffizientenmatrix *und* rechter Seite auf die Lösung x von (9.1).

Satz 9.7. *Sei x die Lösung von* (9.1)*, $\tilde{b} \in \mathbb{R}^n$ und $\tilde{A} \in \mathbb{R}^{n,n}$ eine gestörte Koeffizientenmatrix mit der Eigenschaft*

$$\frac{\|A - \tilde{A}\|}{\|A\|} < \frac{1}{\kappa(A)}\,.$$

Dann existiert eine eindeutig bestimmte Lösung $\tilde{x}$ des gestörten Systems

$$\tilde{A}\tilde{x} = \tilde{b}\,, \tag{9.11}$$

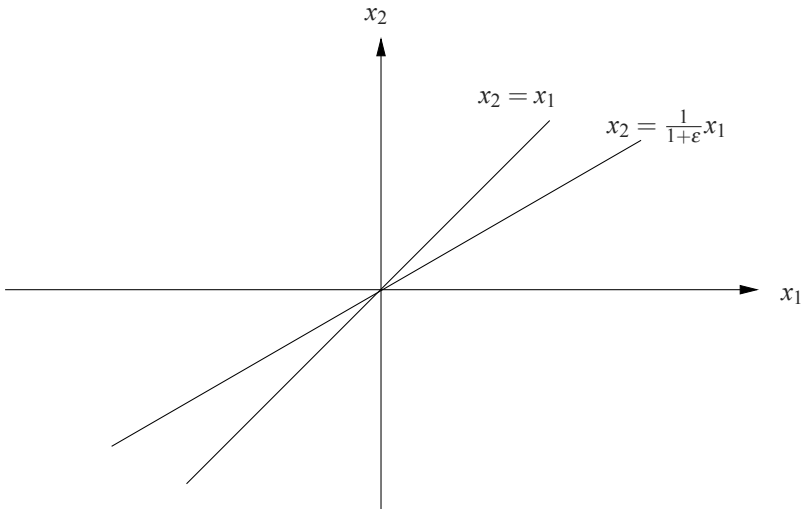
und es gilt

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \kappa(A)\, \left(\frac{\|b - \tilde{b}\|}{\|b\|} + \frac{\|A - \tilde{A}\|}{\|A\|} \right) + o(\|A - \tilde{A}\| + \|b - \tilde{b}\|)\,. \tag{9.12}$$

Es existieren rechte Seiten b, $\tilde{b} \in \mathbb{R}^n$ und Koeffizientenmatrizen A, $\tilde{A} \in \mathbb{R}^{n,n}$, so daß in (9.12) *Gleichheit vorliegt.*

Es kann sein, daß zwar A regulär ist, aber $\tilde{A} = \mathrm{rd}(A)$ nicht.

Beispiel 9.8 (Schleifender Schnitt).



Den eindeutig bestimmten Schnittpunkt $x = 0$ berechnet man aus $Ax = 0$ mit

$$A = \begin{pmatrix} -1 & 1 \\ -1 & 1+\varepsilon \end{pmatrix}\,.$$

Die Inverse von A ist

$$A^{-1} = (-\varepsilon)^{-1} \begin{pmatrix} 1+\varepsilon & -1 \\ 1 & -1 \end{pmatrix}\,.$$

Bei der Berechnung der Kondition verwenden wir die Zeilensummennorm und erhalten

$$\kappa_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty = \frac{(2+\varepsilon)^2}{\varepsilon} \to \infty\,, \qquad \varepsilon \to 0\,.$$

Ist $\varepsilon <$ eps (Maschinengenauigkeit), so folgt

$$\mathrm{rd}(A) = \tilde{A}\,, \qquad \tilde{A} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}\,.$$

Die Regularitätsbedingung an $\|A - \tilde{A}\| < \|A^{-1}\|^{-1}$ aus Satz 9.6 und Satz 9.7 ist dann nicht erfüllt. Offenbar ist $\tilde{A}$ singulär. Die Gleitkomma-Approximation

$$\tilde{A}\tilde{x} = 0\,, \qquad \tilde{A} = \mathrm{rd}(A)\,,$$

hat damit *keine* eindeutig bestimmte Lösung.

Die Kondition $\kappa(A)$ lässt sich als *Quantifizierung der Regularität* von A interpretieren.

Satz 9.9. *Für alle regulären Matrizen $A \in \mathbb{R}^{n,n}$ gilt*

$$\inf\left\{ \frac{\|A-B\|}{\|A\|} \;\middle|\; B \in \mathbb{R}^{n,n} \textit{singulär} \right\} \geq \frac{1}{\kappa(A)}\,.$$

Beweis. Ist $\|A-B\|/\|A\| < 1/\kappa(A)$, so muss B nach Satz 9.6 regulär sein. Also gilt $\|A-B\|/\|A\| \geq 1/\kappa(A)$ für alle singulären $B \in \mathbb{R}^{n,n}$. Diese Abschätzung vererbt sich auf das Infimum. □

Im Falle der Zeilensummennorm $\|\cdot\|_\infty$ kann man sogar zeigen, daß es eine singuläre Matrix B mit der Eigenschaft

$$\frac{\|A-B\|_\infty}{\|A\|_\infty} = \frac{1}{\kappa(A)}$$

gibt! In nächster Umgebung einer schlecht konditionierten Matrix befinden sich also singuläre Matrizen. In diesem Sinne sind schlecht konditionierte Matrizen „fast singulär“.

Beispiel 9.10 (Schleifender Schnitt). Wir betrachten wieder die Matrix

$$A = \begin{pmatrix} -1 & 1 \\ -1 & 1+\varepsilon \end{pmatrix}$$

mit der Kondition $\kappa_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty = \frac{(2+\varepsilon)^2}{\varepsilon}$. Für die singuläre Matrix B,

$$B = \begin{pmatrix} -1+\frac{\varepsilon}{2+\varepsilon} & 1 \\ -1-\frac{\varepsilon}{2+\varepsilon} & 1+\varepsilon \end{pmatrix}\,,$$

erhält man gerade

$$\frac{\|A-B\|_\infty}{\|A\|_\infty} = \frac{\varepsilon}{(2+\varepsilon)^2} = \frac{1}{\kappa(A)}\,.$$

9.2 Der Gaußsche Algorithmus

9.2.1 Motivation

Grundidee des Gaußschen Algorithmus (nach Gauß: Theoria Motus etc. 1809) ist die sukzessive Elimination der Unbekannten.

Beispiel 9.11.

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 5, \\ 2x_1 + 5x_2 + 8x_3 &= -1, \\ 3x_1 + 6x_2 + 10x_3 &= 0. \end{aligned}$$

Auflösen nach x_1:

$$x_1 = 5 - 4x_2 - 7x_3 \ . \tag{9.13}$$

Elimination von x_1 durch Einsetzen in die verbliebenen Gleichungen:

$$\begin{aligned} -3x_2 - \ 6x_3 &= -11, \\ -6x_2 - 11x_3 &= -15. \end{aligned}$$

Auflösen nach x_2:

$$-3x_2 = -11 + 6x_3 \tag{9.14}$$

und Elimination von x_2 aus der zweiten Gleichung ergibt

$$x_3 = 7 \ . \tag{9.15}$$

Einsetzen in (9.14):

$$x_2 = -\frac{31}{3} \ .$$

Einsetzen in (9.13):

$$x_1 = -\frac{8}{3} \ .$$

Einen anderen Blickwinkel erlaubt die folgende Matrixschreibweise des Gaußschen Algorithmus

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \\ 0 \end{pmatrix}$$

$$A \qquad x \quad = \quad b \ .$$

Sukzessive Elimination lässt sich auf der erweiterten Matrix

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 5 \\ 2 & 5 & 8 & -1 \\ 3 & 6 & 10 & 0 \end{array}\right)$$

durchführen: Eliminieren von x_1:

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 5 \\ 2 & 5 & 8 & -1 \\ 3 & 6 & 10 & 0 \end{array}\right) \begin{matrix} \\ -2 * 1. \text{ Zeile} \\ -3 * 1. \text{ Zeile} \end{matrix} \rightarrow \left(\begin{array}{ccc|c} 1 & 4 & 7 & 5 \\ 0 & -3 & -6 & -11 \\ 0 & -6 & -11 & -15 \end{array}\right) \ .$$

Eliminieren von x_2:

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 5 \\ 0 & -3 & -6 & -11 \\ 0 & -6 & -11 & -15 \end{array}\right) \begin{matrix} \\ \\ -2 * 2. \text{ Zeile} \end{matrix} \rightarrow \left(\begin{array}{ccc|c} 1 & 4 & 7 & 5 \\ 0 & -3 & -6 & -11 \\ 0 & 0 & 1 & 7 \end{array}\right) \ .$$

Als Ergebnis erhalten wir das gestaffelte Gleichungssystem:

$$\begin{pmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ -11 \\ 7 \end{pmatrix}$$

$$R \qquad x \quad = \quad z \ .$$

R obere Dreiecksmatrix; sukzessives Einsetzen liefert x

Erklärung (Beobachtung). Aus den im Eliminationsprozess auftretenden Faktoren bilde man die untere Dreiecksmatrix

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \ .$$

Dann gilt (nachrechnen!)

$$A = LR.$$

Zufall? Wir werden sehen.

9.2.2 Gaußscher Algorithmus und LR–Zerlegung

Wir betrachten das lineare Gleichungssystem

$$Ax = b \tag{9.16}$$

mit

$$A = (a_{ij})_{i,j=1}^{n} \in \mathbb{R}^{n,n}\,, \qquad \text{regulär}\,,$$
$$b = (b_i)_{i=1}^{n} \in \mathbb{R}\,.$$

Der Koeffizient a_{11} heißt 0Pivotelement (frz. *Drehpunkt*). Unter der Voraussetzung

$$a_{11} \neq 0$$

erhalten wir nach dem 1. Eliminationsschritt die erweiterte Matrix

$$(A^{(1)}|b^{(1)}) = \left(\begin{array}{cccc|c} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & a_{n2}^{(1)} & \cdots & a_{nn}^{(1)} & b_n^{(1)} \end{array}\right).$$

Dabei ist

$$\begin{aligned} a_{1j}^{(1)} &= a_{1j}\,, \quad j = 1,\ldots,n\,, \quad b_1^{(1)} = b_1\,, \\ a_{ij}^{(1)} &= a_{ij} - \frac{a_{i1}}{a_{11}} a_{1j}\,, \quad i,j = 2,\ldots,n\,, \\ b_i^{(1)} &= b_i - \frac{a_{i1}}{a_{11}} b_1\,, \quad i = 2,\ldots,n\,. \end{aligned}$$

Das Pivotelement für den 2. Eliminationsschritt ist $a_{22}^{(1)}$. Wir haben

$$a_{22}^{(1)} \neq 0$$

vorauszusetzen, um die Unbekannte x_2 aus den Gleichungen 3 bis n zu eliminieren. Unter der Voraussetzung nichtverschwindender Pivotelemente

$$a_{kk}^{(k-1)} \neq 0\,, \qquad k = 2,\ldots,n-1\,,$$

werden durch weitere Eliminationsschritte die Gleichungssysteme

$$A^{(k)}x = b^{(k)}\,, \qquad k = 1,\ldots,n-1\,,$$

erzeugt. Die Berechnung erfolgt analog zum 1. Eliminationsschritt. Das führt auf folgenden Algorithmus:

Algorithmus 9.12 (Gaußsche Elimination).

Für $k = 1,...,n-1$ berechne

{

Für $i = k+1,..,n$ berechne

{

$\ell_{ik} = \dfrac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}$

$b_i^{(k)} = b_i^{(k-1)} - \ell_{ik} b_k^{(k-1)}$

$a_{ik}^{(k)} = 0$

Für $j = k+1,...,n$ berechne

{

$a_{ij}^{(k)} = a_{ij}^{(k-1)} - \ell_{ik} a_{kj}^{(k-1)}$

}

}

}

Dabei ist $a_{ij}^{(0)} = a_{ij}$ und $b_i^{(0)} = b_i$ gesetzt.

Ergebnis ist das Dreieckssystem

$$\begin{pmatrix} a_{11}^{(n-1)} & a_{12}^{(n-1)} & \cdots & a_{1n}^{(n-1)} \\ 0 & a_{22}^{(n-1)} & \cdots & a_{2n}^{(n-1)} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn}^{(n-1)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} b_1^{(n-1)} \\ b_2^{(n-1)} \\ \vdots \\ b_n^{(n-1)} \end{pmatrix}$$

Dieses System läßt sich einfach auflösen.

Algorithmus 9.13 (Rückwärtssubstitution).

$$x_n = \frac{1}{a_{nn}^{(n-1)}} b_n^{(n-1)}.$$

Für $i = n-1, \ldots, 1$ berechne

$$x_i = \frac{1}{a_{ii}^{(n-1)}} \left(b_i^{(n-1)} - \sum_{j=i+1}^{n} a_{ij}^{(n-1)} x_j \right)$$

(Warum ist $a_{nn}^{(n-1)} \neq 0$?)

Erklärung (Bemerkungen zur Implementierung).

- Man kann die alten Elemente

$$a_{ij}^{(k-1)}, \qquad i,j = k+1, \ldots, n$$

mit den neuen Elementen

$$a_{ij}^{(k)}, \qquad i,j = k+1, \ldots, n$$

überschreiben.

- Man kann statt der erzeugten Nullen in der k-ten Spalte die Eliminationsfaktoren

$$\ell_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \qquad i = k+1, \ldots, n,$$

abspeichern.

Wir wollen nun den *Aufwand* zur Lösung des linearen Gleichungssystems (9.16) mit dem Gaußschen Algorithmus abschätzen. Wir wählen

Aufwandsmaß = Anzahl der Multiplikationen.

Anmerkung 9.14. • Die Wahl der *Anzahl der Multiplikationen* als Aufwandsmaß hat historische Gründe. Frühere FPUs (floating-point processing units) konnten viel schneller addieren als multiplizieren. Bei modernen FPUs ist der Unterschied dagegen zu vernachlässigen, so daß eigentlich die Anzahl der Elementaroperationen ein geeigneteres Aufwandsmaß wäre. An der asymptotischen Ordnung ändert das aber nichts.

• Auch hier ist der ermittelte Aufwand bei weitem nicht proportional zur Rechenzeit und gestattet nur qualitative Aussagen über das Verhalten für große n. Die eigentlich viel interessantere *Rechenzeit* kann selbst bei gleicher Anzahl von Elementaroperationen je nach Implementierung um einen Faktor 100 oder mehr unterschiedlich sein!

Wir zählen zunächst die Multiplikationen von Algorithmus 9.12 (Gaußsche Elimination). Dann wird aus jeder for–Schleife eine Summation, und wir erhalten

$$\begin{aligned}
&\sum_{k=1}^{n-1}\sum_{i=k+1}^{n}\left(1+\sum_{j=k+1}^{n}1\right)\\
&=\sum_{k=1}^{n-1}\sum_{i=k+1}^{n}(n-k+1)=\sum_{k=1}^{n-1}(n-k+1)(n-k) \qquad\qquad (\text{setze } j=n-k)\\
&=\sum_{j=1}^{n-1}(j+1)j=\frac{1}{3}(n^3-n)\,.
\end{aligned}$$

Die letzte Identität bestätigt man durch vollständige Induktion. Der zusätzliche Aufwand für die Berechnung von $b^{(n-1)}$ ist

$$\sum_{k=1}^{n-1}\sum_{i=k+1}^{n}1=\sum_{k=1}^{n-1}n-k=\sum_{k=1}^{n-1}k=\frac{1}{2}(n^2-n)\,.$$

Der Aufwand für Algorithmus 9.13 (Rückwärtssubstitution) ist

$$\sum_{i=1}^{n}\left(1+\sum_{j=i+1}^{n}1\right)=\sum_{i=1}^{n}(n-i+1)=\sum_{j=1}^{n}j=\frac{1}{2}(n^2+n)\,.$$

Damit erhalten wir als Gesamtaufwand für die Lösung von (9.16) mit dem Gaußschen Algorithmus

$$\begin{aligned}
\frac{1}{3}(n^3-n)+\frac{1}{2}(n^2-n)+\frac{1}{2}(n^2+n)&=\frac{1}{3}n^3+n^2-\frac{1}{3}n\\
&=\frac{1}{3}n^3+\mathcal{O}(n^2)\,.
\end{aligned}$$

Anmerkung 9.15. Man vergleiche den polynomiellen Aufwand des Gaußschen Algorithmus mit dem exponentiellen Aufwand der Cramerschen Regel (*Lineare Algebra II*).

Der Aufwand läßt sich reduzieren, wenn A gewisse Struktureigenschaften hat, etwa tridiagonal oder symmetrisch und positiv definit ist. Wir verweisen wieder auf weiterführende Vorlesungen oder z.B. auf das Lehrbuch von Deuflhard und Hohmann [10], Kapitel 1.

Als nächstes wollen wir unserer Vermutung aus dem einführenden Abschnitt 9.2 nachgehen.

Wir definieren $G_k \in \mathbb{R}^{n,n}$ durch

$$(G_k)_{ij}=\begin{cases}\ell_{ik}=\frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} & i=k+1,\ldots,n,\ j=k\\ 0 & \text{sonst}\,,\end{cases}$$

und $I \in \mathbb{R}^{n,n}$ sei die Einheitsmatrix.

Lemma 9.16. *Es gilt*

$$A^{(k)} = (I - G_k)A^{(k-1)}$$
$$b^{(k)} = (I - G_k)b^{(k-1)} \qquad k = 1, \ldots, n-1 \ ,$$

wobei wieder $A^{(0)} = A$ und $b^{(0)} = b$ gesetzt ist.

Beweis. Ausrechnen liefert mit der Definition von G_k

$$\begin{aligned}\left(G_k A^{(k-1)}\right)_{ij} &= \sum_{l=1}^{n} (G_k)_{il} \left(A^{(k-1)}\right)_{lj} \\ &= (G_k)_{ik} \left(A^{(k-1)}\right)_{kj} \\ &= \begin{cases} 0 & i \le k \ , \ j = 1, \ldots, n \\ \ell_{ik} a_{kj}^{(k-1)} & i \ge k+1 \ , \ j = 1, \ldots, n \end{cases} .\end{aligned}$$

Also ist

$$\left((I - G_k)A^{(k-1)}\right)_{ij} = \begin{cases} a_{ij}^{(k-1)} & i \le k \ , \ j = 1, \ldots, n \\ a_{ij}^{(k-1)} - \ell_{ik} a_{kj}^{(k-1)} & i \ge k+1 \ , \ j = 1, \ldots, n \end{cases}$$

□

Satz 9.17. *Ist der Gaußsche Algorithmus für $A \in \mathbb{R}^{n,n}$ durchführbar (d.h. erhält man Pivotelemente $a_{kk}^{(k-1)} \neq 0$) und ergeben sich dabei die Eliminationsmatrizen $G_1, \ldots, G_{n-1}$, so gilt*

$$A = LR$$

mit

$$L = I + \sum_{k=1}^{n-1} G_k = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \ell_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \ell_{n1} & \cdots & \ell_{n,n-1} & 1 \end{pmatrix} , \quad R = \prod_{k=1}^{n-1} (I - G_{n-k})A = \begin{pmatrix} a_{11}^{(n-1)} & \cdots & \cdots & a_{1n}^{(n-1)} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn}^{(n-1)} \end{pmatrix} .$$

Beweis. Nach Lemma 9.16 ist

$$\begin{aligned} R = A^{(n-1)} &= (I - G_{n-1})A^{(n-2)} \\ &= (I - G_{n-1})(I - G_{n-2})A^{(n-3)} \ . \\ &= (I - G_{n-1}) \cdots (I - G_1)A \ . \end{aligned} \tag{9.17}$$

Weiter gilt

$$\begin{aligned} (G_k G_k)_{ij} &= \sum_{l=1}^{n} (G_k)_{il} (G_k)_{lj} \\ &= (G_k)_{ik} (G_k)_{kj} = (G_k)_{ik} \cdot 0 = 0 \qquad \forall i, j = 1, \ldots, n \ , \end{aligned} \tag{9.18}$$

also $(I + G_k)(I - G_k) = I - G_k G_k = I$ und somit

$$(I - G_k)^{-1} = I + G_k \ .$$

Aus (9.17) ergibt sich daher

$$A = (I + G_1) \cdots (I + G_{n-1})R \ .$$

Ähnlich wie (9.18) folgt

$$G_k \cdot G_l = 0 \qquad \forall l \ge k \ .$$

Daher ist

$$\begin{aligned}(I+G_1)\cdots(I+G_{n-1}) &= (I+G_2)\cdots(I+G_{n-1})+(G_1+\underbrace{G_1G_2}_{=0})(I+G_3)\cdots(I+G_{n-1})\\ &= (I+G_2)\cdots(I+G_{n-1})+G_1\\ &= I+G_1+G_2+\cdots+G_{n-1}=L\,.\end{aligned} \tag{9.19}$$

□

Der Gaußsche Algorithmus liefert also (falls durchführbar!) eine Faktorisierung

$$A = LR$$

in Dreiecksmatrizen $L,R\in\mathbb{R}^{n,n}$. Kennt man eine solche *LR–Zerlegung* von A, so kann man das Gleichungssystem $Ax=b$ in zwei Schritten lösen, nämlich

$$\begin{aligned}\text{Vorwärtssubstitution: } & Lz=b\\ \text{Rückwärtssubstitution: } & Rx=z\,.\end{aligned}$$

Dieses Vorgehen ist besonders empfehlenswert, wenn mehrere Gleichungssysteme mit verschiedenen rechten Seiten zu lösen sind. Der Aufwand ist dann jeweils nur $\mathcal{O}(n^2)$.

9.2.3 Die Stabilität des Gaußschen Algorithmus

Nach Satz 9.17 lässt sich der Gaußsche Algorithmus wie folgt formulieren

$$A^{(k)} = (I-G_k)A^{(k-1)}\,,\quad A^{(0)}=A\,,\qquad b^{(k)}=(I-G_k)b^{(k-1)}\,,\quad b^{(0)}=b\,, \tag{9.20}$$

$$x = R^{-1}z\,,\qquad R=A^{(n-1)}\,,\quad z=b^{(n-1)}\,. \tag{9.21}$$

Bei der praktischen Durchführung in Gleitkommaarithmetik treten Rundungsfehler auf. Vereinfachend nehmen wir an, daß die exakten Eliminationsmatrizen verwendet werden und nur nach jedem Eliminationsschritt gerundet wird. Dies führt auf den folgenden approximativen Algorithmus

$$\tilde{A}^{(k)} = \mathrm{rd}\left((I-G_k)\tilde{A}^{(k-1)}\right)\,,\quad \tilde{A}^{(0)}=A\,,\qquad \tilde{b}^{(k)}=\mathrm{rd}\left((I-G_k)\tilde{b}^{(k-1)}\right)\,,\quad \tilde{b}^{(0)}=b\,, \tag{9.22}$$

$$\tilde{x} = \tilde{R}^{-1}\tilde{z}\,,\qquad \tilde{R}=\tilde{A}^{(n-1)}\,,\quad \tilde{z}=\tilde{b}^{(n-1)}\,. \tag{9.23}$$

Dabei ist die Rundung komponentenweise zu verstehen, also

$$\mathrm{rd}(B) = (\mathrm{rd}(b_{ij}))_{i,j=1}^n\,,\qquad \mathrm{rd}(b)=(\mathrm{rd}(b_i))_{i=1}^n\,.$$

Man bestätigt leicht

$$\frac{\|B-\mathrm{rd}(B)\|_\infty}{\|B\|_\infty}\le eps\,,\quad \forall B\in\mathbb{R}^{n,n}\,,\qquad \frac{\|b-\mathrm{rd}(b)\|_\infty}{\|b\|_\infty}\le eps\,,\quad \forall b\in\mathbb{R}^n\,. \tag{9.24}$$

Wir betrachten daher von nun an die Maximums-Norm $\|\cdot\|=\|\cdot\|_\infty$.

Zunächst untersuchen wir die Fehlerfortpflanzung im Eliminationsprozess.

Lemma 9.18. *Es gilt*

$$\frac{\|R-\tilde{R}\|_\infty}{\|R\|_\infty}\le \sigma_E\ eps+o(eps)\,,\qquad \frac{\|z-\tilde{z}\|_\infty}{\|z\|_\infty}\le \sigma_E\ eps+o(eps)\,,$$

wobei

$$\sigma_E=\sum_{j=1}^{n-1}\prod_{k=j+1}^{n-1}\kappa_k = 1+\kappa_{n-1}(1+\kappa_{n-2}(1+\ldots\kappa_3(1+\kappa_2))\ldots))\,,\qquad \kappa_k=\kappa(I-G_k).$$

gesetzt ist.

Beweis. Wir zeigen nur die erste der beiden Abschätzungen. Der Nachweis der zweiten erfolgt analog. Die Grundidee besteht darin, eine Rekursion für die Fehler

$$e_k = \frac{\|A^{(k)} - \tilde{A}^{(k)}\|_\infty}{\|A^{(k)}\|_\infty}\,, \quad k = 1,\ldots,n-1\,, \qquad e_0 = 0\,,$$

herzuleiten. Zunächst bemerken wir

$$\|A^{(k-1)}\|_\infty = \|(I-G_k)^{-1}A^{(k)}\|_\infty \le \|(I-G_k)^{-1}\|_\infty \|A^{(k)}\|_\infty\,.$$

Mit der Dreiecksungleichung folgt nun aus (9.24) und dieser Abschätzung

$$\begin{aligned} e_k &\le \frac{\|A^{(k)} - (I-G_k)\tilde{A}^{(k-1)}\|_\infty}{\|A^{(k)}\|_\infty} + eps\,\frac{\|(I-G_k)\tilde{A}^{(k-1)}\|_\infty}{\|A^{(k)}\|_\infty} \\ &\le \frac{\|I-G_k\|_\infty \|A^{(k-1)} - \tilde{A}^{(k-1)}\|_\infty}{\|A^{(k)}\|_\infty} + eps\,\frac{\|(I-G_k)(A^{(k-1)} - \tilde{A}^{(k-1)})\|_\infty}{\|A^{(k)}\|_\infty} + eps \\ &\le \kappa_k(1+eps)e_{k-1} + eps\,. \end{aligned}$$

Die Anwendung von Lemma 7.7 mit $\alpha_k = \kappa_k(1+eps)$ und $\beta_k = eps$ liefert nun

$$\frac{\|R-\tilde{R}\|_\infty}{\|R\|_\infty} = e_{n-1} \le eps \sum_{j=1}^{n-1}\prod_{k=j+1}^{n-1}\kappa_k(1+eps) \le eps \sum_{j=1}^{n-1}\prod_{k=j+1}^{n-1}\kappa_k + o(eps)$$

und damit die Behauptung. □

Nun können wir die angestrebte Fehlerabschätzung formulieren.

Satz 9.19. *Unter der Voraussetzung*

$$\frac{\|R-\tilde{R}\|}{\|R\|} < \frac{1}{\kappa(R)}$$

gilt

$$\frac{\|x-\tilde{x}\|}{\|x\|} \le \sigma_G\, eps + o(eps)\,,$$

wobei

$$\sigma_G = 2\kappa(A)\sigma_K\,\sigma_E\,, \qquad \sigma_K = \prod_{k=1}^{n-1}\kappa_k\,,$$

gesetzt ist.

Beweis. Aus Satz 9.7 folgt unmittelbar

$$\frac{\|x-\tilde{x}\|}{\|x\|} \le \kappa(R)\left(\frac{\|R-\tilde{R}\|}{\|R\|} + \frac{\|z-\tilde{z}\|}{\|z\|}\right) + o(\|R-\tilde{R}\|)\,.$$

Nun gilt

$$\frac{\|R-\tilde{R}\|}{\|R\|} + \frac{\|z-\tilde{z}\|}{\|z\|} \le 2\sigma_E\, eps + o(eps)$$

wegen Lemma 9.18. Außerdem ist

$$\kappa(R) = \kappa((I-G_{n-1})\cdots(I-G_1)A) \le \kappa(A)\prod_{k=1}^{n-1}\kappa_k\,.$$

Damit ist alles gezeigt. □

Nach Satz 9.19 ist die Kondition κ_k der Eliminationsmatrizen $I-G_k$ entscheidend für die Stabilität des Gaußschen Algorithmus. Am besten wäre $\kappa_k = 1$. Der folgende Satz dämpft unsere Hoffnungen.

Satz 9.20. *Es gilt*

$$\kappa(I-G_k) = \|I-G_k\|_\infty \|(I-G_k)^{-1}\|_\infty = \max_{i=k+1,\dots,n} (1+|\ell_{ik}|)^2 \,, \qquad \ell_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} \,.$$

Insbesondere ist $\kappa(I-G_k) = 1$ *genau dann, wenn* $a_{ik}^{(k-1)} = 0 \; \forall i = k+1,\dots,n$ *vorliegt.*

Beweis. Offenbar ist

$$\|I-G_k\|_\infty = \max_{i=k+1,\dots,n} (1+|\ell_{ik}|) \,.$$

Aus $(I-G_k)^{-1} = I+G_k$ folgt weiter

$$\|(I-G_k)^{-1}\|_\infty = \|I+G_k\|_\infty = \max_{i=k+1,\dots,n} (1+|\ell_{ik}|) \,.$$

und damit die Behauptung. □

Nach Satz 9.20 liegt $\kappa_k = 1$ genau dann vor, wenn es nichts zu elimineren gibt. Im nicht-trivialen Fall ist $\kappa_k > 1$. Zumindest sollte man aber $\kappa_k \gg 1$ vermeiden, wie folgendes Beispiel zeigt.

Beispiel 9.21 (2×2*–Matrix).* Wir betrachten das lineare Gleichungssystem

$$\begin{aligned} 10^{-4}x_1 + x_2 &= 1 \\ x_1 + x_2 &= 2 \end{aligned}$$

mit der exakten Lösung

$$x_1 = 1 + \frac{1}{9999} \,, \qquad x_2 = 1 - \frac{1}{9999} \,.$$

Es ist

$$A^{-1} = (10^{-4}-1)^{-1} \begin{pmatrix} 1 & -1 \\ -1 & 10^{-4} \end{pmatrix}$$

die Inverse der Koeffizientenmatrix

$$A = \begin{pmatrix} 10^{-4} & 1 \\ 1 & 1 \end{pmatrix} \,.$$

Wir können daraus die Kondition

$$\kappa(A) = \|A\|_\infty \|A^{-1}\|_\infty = 2 \cdot 2(1-10^{-4})^{-1} \approx 4$$

ablesen. Unser Problem ist also gut konditioniert!

Wir nehmen an, daß uns *3 gültige Stellen* zur Verfügung stehen. Dann ist

$$\mathrm{rd}(x_1) = 1 \,, \qquad \mathrm{rd}(x_2) = 1$$

die gerundete, exakte Lösung. Der Gaußsche Algorithmus liefert

$$\left(\begin{array}{cc|c} 10^{-4} & 1 & 1 \\ 1 & 1 & 2 \end{array}\right) \to \left(\begin{array}{cc|c} 10^{-4} & 1 & 1 \\ 0 & -\mathrm{rd}(9999) & -\mathrm{rd}(9998) \end{array}\right)$$

$$= \left(\begin{array}{cc|c} 10^{-4} & 1 & 1 \\ 0 & -10000 & -10000 \end{array}\right)$$

und damit

$$\tilde{x}_1 = 0 \,, \qquad \tilde{x}_2 = 1 \,.$$

Nach Zeilentausch ist $a_{11} = 1 \gg 10^{-4}$. Man erhält

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 10^{-4} & 1 & 1 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & \mathrm{rd}(0.9999) & \mathrm{rd}(0.9998) \end{array}\right)$$

$$= \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 1 \end{array}\right)$$

und damit die gerundete exakte Lösung

$$\tilde{x}_1 = 1\,, \qquad \tilde{x}_2 = 1\,!$$

Beispiel 9.22 (Wilkinson–Matrix). Als zweites Beispiel betrachten wir die Matrix

$$W = (w_{ij})_{i,j=1}^{n}\,, \qquad w_{ij} = \begin{cases} 1 & \text{falls } i = j \text{ oder } j = n \\ -1 & \text{falls } i > j \\ 0 & \text{sonst} \end{cases}$$

Die Abbildung 9.1 zeigt, daß zwar die Kondition von W nur moderat mit n wächst, die Kondition von R aber explodiert: Für $n = 50$ gilt beispielsweise $\kappa(W) = 22.3$, aber $\kappa(R) = 7.5 \cdot 10^{14}$! Der Gaußsche Algorithmus 9.12 erweist sich als instabil.

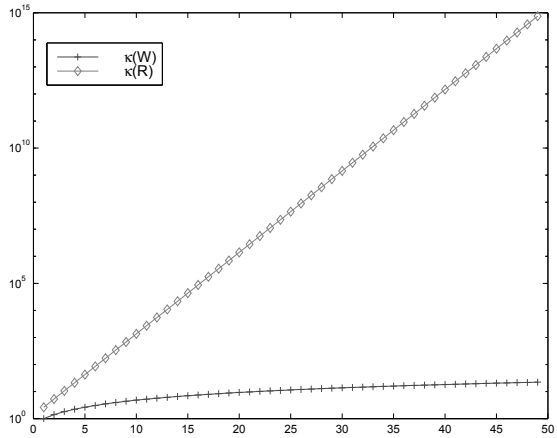


Abb. 9.1 Kondition der Wilkinson-Matrix W und ihres Faktors R.

9.2.4 Gaußscher Algorithmus mit Spaltenpivotsuche

Um die Durchführbarkeit des Gaußschen Algorithmus zu gewährleisten, wollen wir nun vor jedem Eliminationsschritt durch Zeilentausch sicherstellen, daß für das (eventuell neue) Pivotelement $a_{kk}^{(k-1)} \neq 0$ gilt. (Ist das im Falle regulärer Matrizen immer möglich?) Aus Stabilitätsgründen wählen wir $a_{kk}^{(k-1)}$ betragsmäßig möglichst groß, um entsprechend Satz 9.20 für möglichst gut konditionierte Eliminationsmatrizen $I - G_k$ zu sorgen.

Diese Überlegungen führen auf die folgende praktikable Variante des „naiven“ Verfahrens 9.12.

Algorithmus 9.23 (Gauß-Elimination mit Spaltenpivoting).

Für $k = 1, \ldots, n-1$ berechne
{
  $k_0 = k$
  Für $i = k+1, \ldots, n$ berechne
    {
      Falls $\left|a_{ik}^{(k-1)}\right| > \left|a_{k_0,k}^{(k-1)}\right|$, setze $k_0 := i$
    }
  Vertausche die k–te Zeile mit der k_0–ten Zeile
  k–ter Eliminationsschritt wie in Algorithmus 9.12.
}

Anmerkung 9.24. • Der Zusatzaufwand für Spaltenpivoting besteht in $\mathcal{O}(n^2)$ Vergleichen und Vertauschungen.

- Aufgrund unserer Stabilitätsanalyse sollte $\max_{i=k+1,\ldots,n} |\ell_{ik}|$ möglichst klein sein. Daher sollte man besser k_0 so wählen, daß gilt

$$\max_{i=k,\ldots,n} \frac{\left|a_{ik}^{(k-1)}\right|}{\left|a_{k_0k}^{(k-1)}\right|} \leq \max_{i=k,\ldots,n} \frac{\left|a_{ik}^{(k-1)}\right|}{\left|a_{jk}^{(k-1)}\right|} \qquad \forall j = k,\ldots,n\ .$$

Dieses Vorgehen ist jedoch aufwendiger. Meist verwendet man stattdessen eine geschickte Skalierung des Gleichungssystems. Wir verweisen auf Deuflhard und Hohmann [10], S. 13.

Aufgrund der Wahl der Pivotelemente gilt

$$|\ell_{ik}| = \left|\frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}\right| \leq 1\ , \qquad i = k+1,\ldots,n\ , \quad k = 1,\ldots,n-1\ .$$

Daraus folgt

$$\sigma_E \leq 2^{n-1} - 1\ , \qquad \sigma_K \leq 2^{n-1}\ ,$$

und man gewinnt aus Satz 9.19 die Stabilitätsabschätzung

$$\frac{\|x-\tilde{x}\|_\infty}{\|x\|_\infty} \leq 2\kappa(A)2^{n-1}(2^{n-1}-1)\ eps + o(eps)\ .$$

Also ist für große n damit zu rechnen, daß der Gauß-Algorithmus instabil ist, selbst dann, wenn Spaltenpivoting durchgeführt wird. Für Matrizen mit Spezialstruktur (symmetrisch positiv definite, diagonaldominante, zufällige, Bandmatrizen, etc.) können deutlich bessere Resultate gezeigt werden.

Eine allgemein anwendbare Methode zur Verbesserung der Stabilität ist die *Nachiteration* (siehe [10]):

(1) zerlege $A = LR$
(2) löse $Lz = b$
(3) löse $R\tilde{x} = z$
(4) berechne $r = b - A\tilde{x}$
(5) löse $L\tilde{z_r} = r$
(6) löse $R\tilde{d} = \tilde{z_r}$
(7) setze $x = \tilde{x} + \tilde{d}$

Das Analogon zu Satz 9.17 ist nun

Satz 9.25. *Die Gaußsche Elimination mit Spaltenpivotsuche liefert eine Zerlegung*

$$LR = PA$$

mit unterer Dreiecksmatrix L, oberer Dreiecksmatrix R und einer Permutationsmatrix P. PA unterscheidet sich von A also nur durch Vertauschung der Zeilen.

Beweis. Die Vertauschung von k–ter Zeile mit k_0–ter Zeile wird durch die Permutationsmatrix P_k repräsentiert. So bewirkt $P_kA^{(k-1)}$ gerade die Vertauschung von k–ter und k_0–ter Zeile. Dann liefert der Eliminationsprozess Matrizen G_k mit

$$(I - G_{n-1})P_{n-1}\cdots(I-G_1)P_1A = R\ .$$

Wegen $P_k^{-1} = P_k$ und $(I-G_k)^{-1} = (I+G_k)$ ist

$$A = P_1(I+G_1)\ldots P_{n-1}(I+G_{n-1})R\ . \tag{9.25}$$

Wir setzen

$$Q_k = P_{n-1}\cdots P_k\ , \quad k = 1,\ldots,n-1\ , \quad Q_n = I\ .$$

Wegen $P_kP_k = I$ gilt dann

$$\begin{aligned} Q_k P_k (I+G_k) &= Q_{k+1} P_k P_k (I+G_k) \\ &= (I + Q_{k+1} G_k Q_{k+1}^{-1}) Q_{k+1} \\ &= (I + Q_{k+1} G_k) Q_{k+1} \,, \end{aligned}$$

denn

$$G_k Q_{k+1}^{-1} = G_k P_{k+1} \cdots P_{n-1} = G_k \,.$$

Multiplikation von rechts mit P_l bewirkt nämlich die Vertauschung der Spalten l und $l_0 \geq l$, und die sind beide Null für $l > k$.

Setzt man $P = Q_1 = P_{n-1} \cdots P_1$, so folgt aus (9.25) schließlich

$$\begin{aligned} P P_1 (I+G_1) \cdots P_{n-1}(I+G_{n-1}) &= (I+Q_2 G_1) Q_2 P_2 (I+G_2) \cdots P_{n-1}(I+G_{n-1}) \\ &= (I+Q_2 G_1)(I+Q_3 G_2) \cdots (I+G_{n-1}) \\ &= I + \sum_{m=1}^{n-1} Q_{k+1} G_k = L \,, \end{aligned} \tag{9.26}$$

denn die Matrizen $Q_{k+1} G_k$ sind von der gleichen Bauart wie G_k, und man kann wie in (9.19) schließen. Aus (9.25) und (9.26) folgt die Behauptung. □

Übrigens liefert der MATLAB–Befehl

```
[L,R,P]=lu(A)
```

eine *LR*–Zerlegung von A mit Pivoting. Ausprobieren!

9.3 Aufgaben

Aufgabe 9.1 *Zeigen Sie für $p = 1$ und $p = 2$, dass $\|x\|_p = (\sum_{i=1} |x_i|^p)^{\frac{1}{p}}$ eine Norm auf $\mathbb{R}^n$ ist. Für $p = 2$ können Sie die Cauchy-Schwarzsche Ungleichung*

$$|\langle x,y \rangle| \leq \|x\|_2 \|y\|_2 \qquad \forall x,y \in \mathbb{R}^n$$

verwenden.

Aufgabe 9.2 *1. Zeigen Sie, dass auf dem $\mathbb{R}^n$ die Normen $\|\cdot\|_1$, $\|\cdot\|_2$ und $\|\cdot\|_\infty$ äquivalent sind, d.h., dass Konstanten $C_n(\alpha,\beta) > 0$, $c_n(\alpha,\beta) > 0$ existieren mit*

$$1/c_n(\alpha,\beta)\, \|x\|_\alpha \,\leq\, \|x\|_\beta \,\leq\, C_n(\alpha,\beta)\, \|x\|_\alpha \,.$$

Dabei sei $\alpha,\beta \in \{1,2,\infty\}$. Geben Sie die kleinstmöglichen Konstanten $C_n(\alpha,\beta)$, $c_n(\alpha,\beta)$ für $\alpha,\beta \in \{1,2,\infty\}$ explizit an.
Hinweis: Zeigen Sie zuerst, dass $2|ab| \leq a^2 + b^2$ für $a,b \in \mathbb{R}$ gilt.

2. Was geschieht mit den Konstanten $C_n(\alpha,\beta)$, $c_n(\alpha,\beta)$ für den Fall $n \to \infty$?

Aufgabe 9.3 *Zeichnen Sie jeweils die Einheitskugel*

$$B_{1,\alpha}(0) := \{x \,|\, x \in \mathbb{R}^2, \|x\|_\alpha \leq 1\}$$

für $\alpha \in \{1,2,4,\infty\}$, d.h. die Einheitskugeln bezüglich der Normen $\|\cdot\|_1$, $\|\cdot\|_2$,$\|\cdot\|_4$ und $\|\cdot\|_\infty$.

Aufgabe 9.4 *Zu $x \in \mathbb{R}$ sei die Matrix $A(x) = \begin{pmatrix} 1 & 2 \\ x & 4 \end{pmatrix} \in \mathbb{R}^{2,2}$ gegeben.*

1. *Für welche $x \in \mathbb{R}$ ist $A(x)$ invertierbar?*
2. *Berechnen Sie $A(x)^{-1}$ und $\kappa_\infty(A(x))$.*
3. *Was geschieht mit $\kappa_\infty(A(x))$ in der Nähe der x_0, für die $A(x_0)$ nicht invertierbar ist?*
4. *Geben Sie ein x an, für das $\kappa_\infty(A(x)) < 10$ gilt.*

Aufgabe 9.5 *Die sogenannte Hilbert-Matrix $H \in \mathbb{R}^{n,n}$ sei durch $h_{ij} = 1/(i+j-1)$, $i = 1,\ldots,n$, gegeben. Wir betrachten für $x, b \in \mathbb{R}^n$ das lineare Gleichungssystem*

$$Hx = b\,. \tag{9.27}$$

Wählt man als rechte Seite $b \in \mathbb{R}^n$ mit

$$b_i = \begin{cases} 1, & i = 1 \\ 0, & i > 1 \end{cases}\,.$$

so ist die exakte Lösung $x \in \mathbb{R}^n$ durch

$$x_i = (-1)^{i+1}\, i \binom{n+i-1}{n-1} \binom{n}{n-i}$$

gegeben.

Zeigen Sie für $n = 4$, dass x wie oben angegeben tatsächlich das lineare Gleichungssystem (9.27) löst. Implementieren Sie ein `MATLAB` *Programm, das für gegebenes n die zugehörige Hilbertmatrix H erzeugt. Schreiben Sie weiter ein Programm, das mit Hilfe des „Backslash"-Operators „\" das lineare Gleichungssystem (9.27) löst. Die numerisch erhaltene Lösung sei $\hat{x}$. Bestimmen Sie die relative und absolute maximale Abweichung*

$$\frac{\|x-\hat{x}\|_\infty}{\|x\|_\infty} \qquad \textit{bzw.} \qquad \|x\|_\infty$$

von $\hat{x}$ experimentell für $n = 1,2,3,\ldots,13$, und visualisieren Sie Ihre Ergebnisse. Was fällt Ihnen auf?

Aufgabe 9.6 *1. Es sei $\|\cdot\|$ eine Norm auf $\mathbb{R}^n$. Zeigen Sie, dass für $A \in \mathbb{R}^{n,n}$*

$$\|A\| := \sup_{\substack{x \neq 0 \\ x \in \mathbb{R}^n}} \frac{\|Ax\|}{\|x\|} = \sup_{\substack{\|x\|=1 \\ x \in \mathbb{R}^n}} \|Ax\|\,.$$

Zeigen Sie weiter, dass auf diese Weise tatsächlich eine Norm auf $\mathbb{R}^{n,n}$ definiert ist.

2. Bestimmen Sie ohne Zuhilfenahme des `MATLAB` *Befehls* `cond` *die relative Kondition der Hilbertmatrix H für $n = 3$ bezüglich der Maximums–Norm $\|\cdot\|_\infty$.*
Hinweis: Um H und H^{-1} zu bestimmen, können Sie `MATLAB` *verwenden.*

Gegeben sei die *Tridiagonalmatrix*

$$A = \begin{pmatrix} a_1 & b_1 & 0 & \ldots & 0 \\ c_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & \ldots & 0 & c_{n-1} & a_n \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Wir nehmen an, dass A invertierbar ist und betrachten das lineare Gleichungssystem

$$Ax = b$$

für ein festes $b \in \mathbb{R}^n$.

Aufgabe 9.7 *1. Zeigen Sie, dass in der Gauß-Elimination (Alg. 10.1 im Skript)*

$$A_{ij}^{(k)} = 0 \qquad \textit{für alle } k \in \mathbb{N} \textit{ und } i > j+1$$

gilt.

2. Geben Sie einen modifizierten Gauß-Algorithmus an, der die spezielle Struktur des Problems benutzt. Insbesondere sollen Multiplikationen, von denen a priori klar ist, dass sie Null ergeben, gar nicht erst ausgeführt werden. Wieviele Multiplikationen braucht Ihr Algorithmus für ein Gleichungssystem der Größe $n \times n$?