Wissenschaftliches Rechnen

Lineare Gleichungssysteme

wr@isg.cs.uni-magdeburg.de

SoSe 2019

1 Motivation: Computertomographie

Ziel der Computertomographie ist es, die Dichteverteilung innerhalb eines Objektes zu bestimmen, ohne das Objekt zu zerschneiden. Das Objekt wird dazu mit Strahlen $\{S_k\}_{k \in K}$ durchleuchtet (siehe Abbildung 1) und für jeden Strahl S_k wird der Intensitätsverlust bestimmt, welcher beim Durchdringen des Objektes auftritt. Aus den Intensitätsverlusten lässt sich dann die Dichteverteilung im Objekt rekonstruieren, siehe Abb. 3.

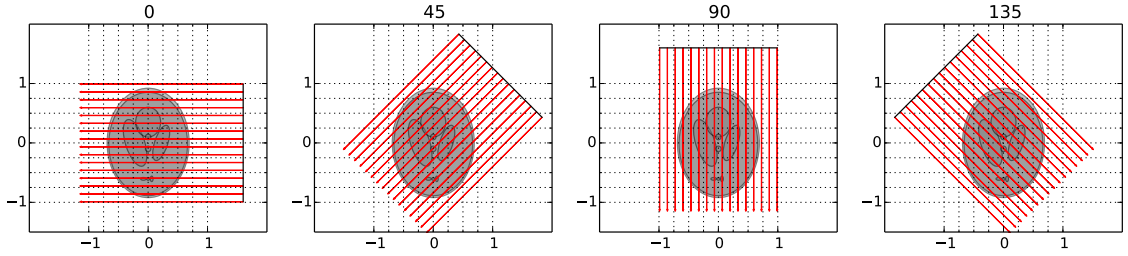

Abbildung 1: Bei der Computertomographie werden eine Reihe paralleler Strahlen aus verschiedenen Winkeln ausgesendet und es wird der Intensitätsverlust der Strahlen gemessen, welcher beim Durchdringen des Objektes auftritt.

Grundlage der Rekonstruktion ist die folgende Gleichung, welche die Dichte $\varrho(x)$ im Volumen und den Intensitätsverlust $I_{\text{in}}^{(k)}/I_{\text{out}}^{(k)}$ in Zusammenhang setzt:

$$\log\left(\frac{I_{\text{in}}^{(k)}}{I_{\text{out}}^{(k)}}\right) = \int_{S_k} \varrho(s)\,\mathrm{d}s. \tag{1}$$

Die rechte Seite der Gleichung beschreibt den Intensitätsverlust, welcher durch die Absorption durch das Material auftritt, durch welchen sich der Strahl S_k bewegt. Da der infinitesimale Verlust proportional zur Dichte des Materials ist, kann der gesamte Intensitätsverlust durch ein Integral bestimmt werden. Die Intensitäten $I_{\text{in}}^{(k)}$ und $I_{\text{out}}^{(k)}$ am Anfang und Ende des k-ten Strahles können in einem Tomographen gemessen werden. Diese ist eine grundlegende Voraussetzung für die Rekonstruktion der Dichte $\varrho(x)$ im Objekt.

Gleichung 1 bekommt eine besonders einfache Form, wenn angenommen wird, dass die Dichte innerhalb des Objektes stückweise konstant ist, siehe Abb. 2. Ein Strahl S_k lässt sich dann in Segmente $\gamma_1 \dots \gamma_n$ der Länge $|\gamma_i|$ unterteilen, über welchen die Dichte $\varrho(x)$ einen konstanten Wert ϱ_i hat. Für Gleichung 1 gilt dann

$$\log\left(\frac{I_{\text{in}}^{(k)}}{I_{\text{out}}^{(k)}}\right) = \int_{S_k} \varrho(s)\,\mathrm{d}s = \sum_{i=1}^{n} \int_{\gamma_i} \varrho(s)\,\mathrm{d}s = \sum_{i=1}^{n} |\gamma_i|\,\varrho_i. \tag{2}$$

Diese Vereinfachung wird bei der Computertomographie ausgenutzt. Um die Dichteverteilung innerhalb des Objektes näherungsweise zu bestimmen, wird ein reguläres $n \times n$ Gitter über das Objekt gelegt und für jede Zelle des Gitters angenommen, dass die Dichte darin konstant ist. Bezeichnen wir mit Q_{ij} die Zelle in der i-ten Zeile und j-ten Spalte des Gitters, ϱ_{ij} die Dichte in Zelle Q_{ij} und $L_{ij}^{(k)}$ die Länge des

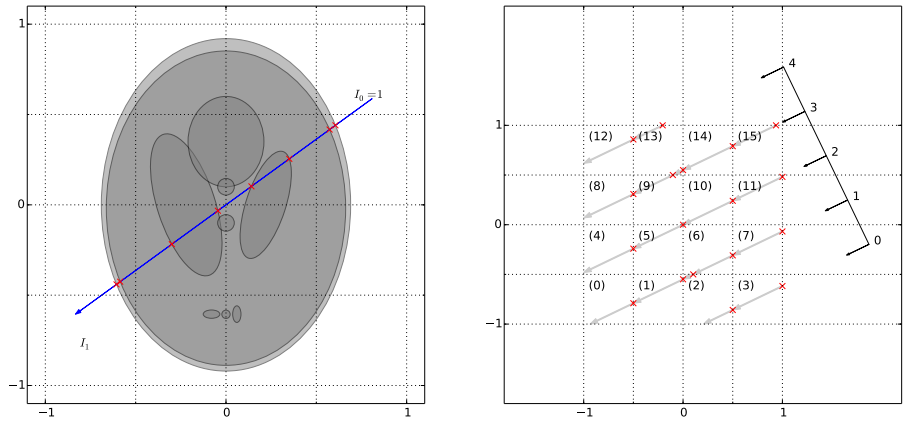

Abbildung 2: Links: Objekt mit einer Dichteverteilung, welche stückweise konstant ist. Rechts: Um numerisch die Dichteverteilung bestimmen zu können, wir ein reguläres Gitter über das Objekt gelegt und angenommen, dass die Dichte in jeder Zelle konstant ist.

Schnittes des Strahls S_k mit den Zellen Q_{ij}, siehe Abb. 2, so erhalten wir von Gleichung 2:

$$V_k = \log\left(\frac{I_0}{I_1}\right) = \sum_{i,j=1}^{n} \int_{Q_{ij} \cap S_k} \varrho_{ij} \, \mathrm{d}s = \sum_{i,j} L_{ij}^{(k)} \varrho_{ij}. \tag{3}$$

Gleichung 3 ist ein lineares Gleichungssystem mit einer Gleichung für jeden Strahl und den Dichten in den Zellen als Unbekannten. In Matrixnotation lässt sich dieses Gleichungssystem folgendermaßen schreiben:

$$\underbrace{\begin{pmatrix} L_{11}^{(1)} & L_{12}^{(1)} & \cdots & L_{1n}^{(1)} & L_{21}^{(1)} & L_{22}^{(1)} & \cdots & L_{nn}^{(1)} \\ L_{11}^{(2)} & L_{12}^{(2)} & \cdots & L_{1n}^{(2)} & L_{21}^{(2)} & L_{22}^{(2)} & \cdots & L_{nn}^{(2)} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ L_{11}^{(K)} & L_{12}^{(K)} & \cdots & L_{1n}^{(K)} & L_{21}^{(K)} & L_{22}^{(K)} & \cdots & L_{nn}^{(K)} \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} \varrho_{11} \\ \varrho_{12} \\ \vdots \\ \varrho_{1n} \\ \varrho_{21} \\ \varrho_{22} \\ \vdots \\ \varrho_{nn} \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} V_1 \\ V_2 \\ \\ \vdots \\ \\ V_K \end{pmatrix}}_{b}, \tag{4}$$

d.h. $Ax = b$, mit $A \in \mathbb{R}^{K \times n^2}$, $x \in \mathbb{R}^{n^2}$ und $b \in \mathbb{R}^K$. In der Regel sind sinnvolle Ergebnisse nur zu erwarten, wenn die Anzahl der Strahlen K (d.h. die Anzahl der Gleichungen) größer ist als die Anzahl n^2 der Zellen im Gitter, und damit der zu bestimmenden Dichten. Es handelt sich dann um ein überbestimmtes Gleichungssystem, welches mit Methoden der Ausgleichsrechnung gelöst werden kann. Die Lösung solcher Gleichungssysteme ist Teil der Ausgleichrechnung, welche im nächsten Abschnitt diskutiert werden wird. Zunächst betrachten wir jedoch noch einmal näher, was lineare Gleichungssysteme sind. In Abschnitt 8 wird anschließend untersucht, wie diese numerisch zuverlässig mit Gauß-Elimination gelöst werden können, wenn sie nicht überbestimmt sind.

2 Definition

Unter einem *linearen Gleichungssystem* versteht man eine Menge von linearen Gleichungen, welche von den gleichen Variablen abhängen:

$$\begin{aligned} a_{11}\, x_1 + a_{12}\, x_2 + \ldots + a_{1n}\, x_n &= b_1 \\ a_{21}\, x_1 + a_{22}\, x_2 + \ldots + a_{1n}\, x_n &= b_2 \\ &\vdots \\ a_{m1}\, x_1 + a_{m2}\, x_2 + \ldots + a_{mn}\, x_n &= b_m \end{aligned} \tag{5}$$

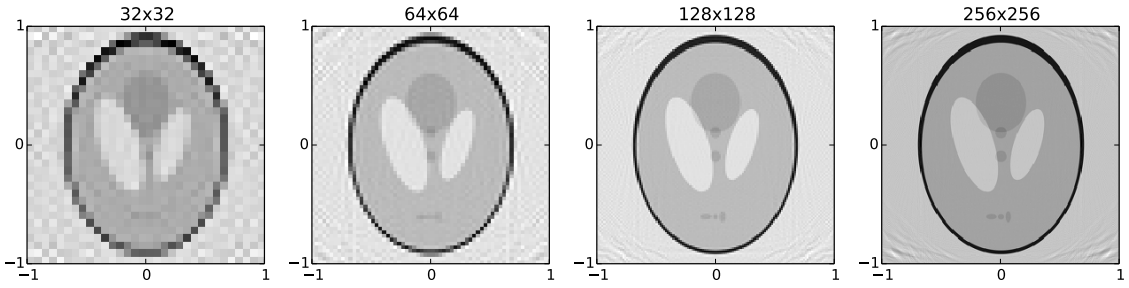

Abbildung 3: Tomographische Bilder bei welchen die Helligkeit die rekonstruierten Dichten darstellt.

Viele technische Anwendungen können auf die Lösung von linearen Gleichungssystem zurückgeführt werden. Die dabei entstehenden Systeme bestehen oft aus sehr vielen Gleichungen und sind mit den Methoden der klassischen linearen Algebra nicht effizient lösbar. Als Teilbereich der Numerik hat sich daher die numerische lineare Algebra entwickelt, welche insbesondere Methoden und Techniken zur computergestützten Lösung großer linearer Gleichungssysteme betrachtet.

3 Matrix-Darstellung eines linearen Gleichungssystems

Lineare Gleichungssysteme lassen sich mit Matrizen und Vektoren darstellen. Dazu werden die Koeffizienten a_{ij} des Gleichungssystems zu einer Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \tag{6}$$

und die Unbekannten x_j und rechten Seiten b_i zu Vektoren

$$x = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T \quad b = \begin{pmatrix} b_1 & b_2 & \dots & b_m \end{pmatrix}^T \tag{7}$$

zusammengefasst. Das System von Gleichungen (5) lässt sich dann als $Ax = b$ schreiben:

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{= \begin{pmatrix} a_{11}\,x_1 + a_{12}\,x_2 + \ldots + a_{1n}\,x_n \\ a_{21}\,x_1 + a_{22}\,x_2 + \ldots + a_{2n}\,x_n \\ \vdots \\ a_{m1}\,x_1 + a_{m2}\,x_2 + \ldots + a_{mn}\,x_n \end{pmatrix}} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{8}$$

und durch ausmultiplizieren der Matrix-Vektor Gleichung kann man sehen, dass diese tatsächlich zum ursprünglichen Gleichungssystem äquivalent ist.

Wir werden ein lineares Gleichungssystem oft wie folgt darstellen:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right) \tag{9}$$

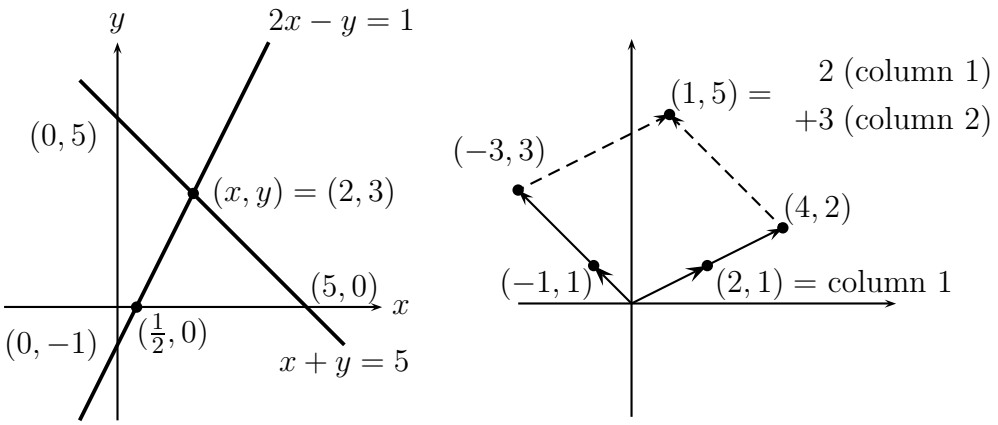

Abbildung 4: Zeilenweise (links) und spaltenweise (rechts) Interpretation des Gleichungssystem in Gl. 10.

4 Geometrische Interpretation von Gleichungssystemen

Gleichungssysteme haben zwei geometrische Interpretationen: entlang der Zeilen und entlang der Spalten. Um diese zu erläutern, betrachten wir ein einfaches System mit 2 Unbekannten:[1]

$$2x - y = 1 \tag{10a}$$
$$x + y = 5. \tag{10b}$$

Zeilen-Interpretation Wenn wir ein Gleichungssystem mit n Unbekannten zeilenweise interpretieren, dann entspricht jede Zeile einer $(n-1)$-dimensionalen Hyperebene in $\mathbb{R}^n$ und die Lösung des Gleichungssystems ist der Schnittpunkt der Ebenen. Für das System in Gl. 10 mit zwei Unbekannten entsprechen die Zeilen also 1-dimensionalen Hyperebenen im $\mathbb{R}^2$, d.h. Geraden. Durch auflösen nach y erhalten wir explizite Ausdrücke für die Geraden:

$$y_1 = 2x - 1 \tag{11a}$$
$$y_2 = -x + 5. \tag{11b}$$

und diese sind graphisch in Abb. 4 (links) dargestellt.

Spalten-Interpretation Wenn wir ein Gleichungssystem spaltenweise interpretieren, dann fassen wir den Vektor auf der rechten Seite als Linearkombination der Spalten auf der linken Seite auf und die Unbekannten sind die zu bestimmenden linearen Gewichte. Gl. 10 können wir also wie folgt interpretieren:

$$x \begin{pmatrix} 2 \\ 1 \end{pmatrix} + y \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \end{pmatrix} \tag{12}$$

und dies ist graphisch in Abb. 4 (rechts) dargestellt.

5 Lineare Gleichungssysteme in Diagonalform

Das bezüglich der Lösung einfachste linearen Gleichungssystem liegt vor, wenn das System durch eine quadratische Diagonalmatrix gegeben ist:

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{13}$$

[1] Das Beispiel ist aus G. Strang, *Linear Algebra and Its Applications*. Thomson, Brooks/Cole, 2006.

Das lineare Gleichungssystem besteht dann aus n unabhängigen Gleichungen und die Unbekannte x_j in the j-ten Zeile hängt nur von a_{jj} und b_j in der gleiche Zeile ab. Für $a_{ii} \neq 0$ ergibt sich also:

$$x_i = b_i / a_{ii}. \tag{14}$$

Im Falle $a_{ii} = 0$ haben wir

$$0\, x_i = b_i. \tag{15}$$

Wenn $b_i = 0$ dann kann die Unbekannte x_i beliebige Werte annehmen, da $0\, x_i = 0$ für beliebige $x_i \in \mathbb{R}$ erfüllt ist. Im Falle $b_i \neq 0$ existiert keine Lösung.

6 Lineare Gleichungssysteme in Dreiecksform

Ein Gleichungssystem ist in *unterer Dreicksform*, wenn es folgende Form hat:

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{16}$$

Diese System sind von besonderer Bedeutung, da eine Lösung besonders effizient gefunden werden kann. Betrachten wir die erste Zeile so kann die Lösung direkt bestimmt werden:

$$x_1 = \frac{b_1}{a_{11}}. \tag{17}$$

Nach der Berechnung von x_1 ist in der zweiten Zeile nur noch x_2 unbekannt und wir können wieder direkt für die Variable lösen:

$$x_2 = \frac{b_2 - a_{21} x_1}{a_{22}} \tag{18}$$

Dies kann Zeile für Zeile fortgesetzt werden bis man in der letzten Zeile angekommen ist:

$$x_n = \frac{b_n - a_{n1} x_1 - \ldots - a_{n,n-1} x_{n-1}}{a_{nn}}. \tag{19}$$

Die Lösung wird also von oben nach unten schrittweise berechnet und man spricht daher von *Vorwärtseinsetzen*. Die allgemeine Rechenvorschrift für das Verfahren lautet:

$$x_i = \frac{b_i - \sum_{j=1}^{i-1} a_{ij} x_j}{a_{ii}} \qquad (i = 1, \ldots, n) \tag{20}$$

Im Falle einer oberen Dreiecksmatrix

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{21}$$

kann ähnlich vorgegangen werden. Die Lösung wird schrittweise von unten nach oben berechnet werden. Man spricht daher von *Rückwärtseinsetzen*:

$$\begin{aligned} x_n &= \frac{b_n}{a_{nn}} \\ x_{n-1} &= \frac{b_{n-1} - a_{n-1,n} x_n}{a_{n-1,n-1}} \\ &\vdots \\ x_1 &= \frac{b_1 - a_{12} x_2 - \ldots - a_{1,n} x_n}{a_{11}}, \end{aligned} \tag{22}$$

oder allgemein:

$$x_i = \frac{b_i - \sum_{j=i+1}^{n} a_{ij} x_j}{a_{ii}} \qquad (i = n, \ldots, 1). \tag{23}$$

7 Beispiel: Schnitt von Gerade und Ebene

Wir wollen den Schnittpunkt einer Ebene und einer Gerade im Raum bestimmen. Dazu nehmen wir an, dass die Ebene E und die Gerade G in Parameterform gegeben sind:

$$E : x = \begin{pmatrix} -3 \\ 1 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix}, \quad \text{mit } \alpha, \beta \in \mathbb{R} \tag{24}$$

$$G : x = \begin{pmatrix} 2 \\ -3 \\ 2 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}, \quad \text{mit } \mu \in \mathbb{R} \tag{25}$$

Ein Schnittpunkt von Gerade und Ebene ist ein Punkt, welcher sowohl die Ebenengleichung als auch die Geradengleichung erfüllt. Durch Gleichsetzen der beiden Gleichungen erhalten wir ein Gleichungssystem in den Unbekannten α, β und μ. Dieses ist linear, da die Unbekannten nur als gewichtete Summanden auftreten:

$$-3 + \alpha = 2 + \mu \tag{26a}$$

$$1 - \alpha - \beta = -3 - 2\mu \tag{26b}$$

$$1 - \alpha + 2\beta = 2 + 3\mu \tag{26c}$$

Durch das Addieren von Vielfachen dieser Gleichungen bzw. das Austauschen von Gleichungen ändert sich die Lösungsmenge des Gleichungssystems nicht. Eine Lösungsstrategie besteht nun darin, durch diese Operation drei neue Gleichungen zu erhalten, die eine Dreiecksstruktur haben: Die dritte Gleichung hängt nur von einer Variable (z.B. μ) ab, die zweite Gleichung von zwei Variable (z.B. μ und α) usw. Die Formalisierung dieses Verfahrens heißt *Gaußsche Eliminationsmethode*. Wir haben bereits in Abschnitt (6) gesehen, dass sich ein System mit oberer Dreiecksstruktur effizient durch Rückwärtseinsetzen lösen lässt. Die Gaußsche Eliminationsmethode zusammen mit Rückwärtseinsetzen ergibt also einen Algorithmus zur Lösung linearer Gleichungssysteme.

Die Darstellung von Gleichungssystemen in Matrixform ermöglicht eine übersichtliche Durchführung der Gauß-Elimination, welche wir in Abschnitt 8 formell einführen werden. Addieren von Gleichungen entspricht dann dem Addieren von Zeilen der Matrix und von Elementen des Vektors auf der rechten Seite. Darüber hinaus, und für unsere Zwecke natürlich noch viel wichtiger, ist die Matrixdarstellung die Voraussetzung für eine effiziente Umsetzung des Verfahrens auf dem Computer. Gl. 26 hat dann die Form:

$$\begin{pmatrix} 1 & 0 & -1 \\ -1 & -1 & 2 \\ -1 & 2 & -3 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \mu \end{pmatrix} = \begin{pmatrix} 5 \\ -4 \\ 1 \end{pmatrix} \tag{27}$$

und noch kompakter lässt sich das System wie folgt darstellen:

$$\left(\begin{array}{ccc|c} 1 & 0 & -1 & 5 \\ -1 & -1 & 2 & -4 \\ -1 & 2 & -3 & 1 \end{array}\right) \tag{28}$$

Um unser Beispiel in Dreiecksform zu überführen, müssen wir alle Elemente in der erste Spalte, bis auf das erste, eliminieren. Wenn wir also die erste mit der zweiten Zeile addieren und die erste mit der dritten, so erhalten wir das folgende, äquivalente Gleichungssystem:

$$\left(\begin{array}{ccc|c} 1 & 0 & -1 & 5 \\ 0 & -1 & 1 & 1 \\ 0 & 2 & -4 & 6 \end{array}\right) \tag{29}$$

wobei wir auch die rechte Seite transformiert haben. Das Ziel der Dreiecksform ist schon fast erreicht. Das Element A_{32} der Matrix kann Null gesetzt werden, in dem wir das doppelte der zweiten Zeile mit der dritten addieren. Dann erhalten wir:

$$\left(\begin{array}{ccc|c} 1 & 0 & -1 & 5 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -2 & 8 \end{array}\right) \tag{30}$$

Das umgeformte System lässt sich jetzt einfach durch Rückwärtseinsetzen lösen. Aus der letzten Zeile entnehmen wir das $-2\mu = 8$ und damit $\mu = -4$. Einsetzen dieses Wertes in der zweiten Zeile liefert $-\beta - 4 = 1$, so dass $\beta = -5$. Analog erhalten wir $\alpha = 1$. Den Schnittpunkt erhält man nun indem man entweder α und β in die Ebenengleichung oder μ in die Geradengleichung einsetzt. Explizit erhalten wir also für den Schnittpunkt mithilfe der Ebenengleichung:

$$x = \begin{pmatrix} -3 \\ 1 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} - 5 \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ 5 \\ -10 \end{pmatrix} \tag{31a}$$

und mit der Geradengleichung erhalten wir:

$$x = \begin{pmatrix} 2 \\ -3 \\ 2 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ -3 \\ 2 \end{pmatrix} - 4 \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ 5 \\ -10 \end{pmatrix}. \tag{31b}$$

In jedem Schritt des Algorithmus haben wir zunächst ein Diagonalelement, das sogenannte Pivotelement, gewählt und mit ihm die Spalte darunter ëliminiert“. Dies kann allerdings nicht funktionieren, wenn das Pivotelement Null ist. In diesem Fall tauscht man einfach die Zeile des Pivotelements mit einer Zeile weiter unten in der Matrix, sodass auf der Diagonale ein Element ungleich Null steht und fährt wie gewohnt fort.

Es kann passieren, dass am Ende des Verfahrens ein System der Form

$$\left(\begin{array}{ccc|c} x_{00} & x_{01} & x_{02} & a \\ 0 & x_{11} & x_{12} & b \\ 0 & 0 & 0 & c \end{array}\right) \quad , \quad c \neq 0 \tag{32a}$$

oder der Form

$$\left(\begin{array}{ccc|c} x_{00} & x_{01} & x_{02} & a \\ 0 & x_{11} & x_{12} & b \\ 0 & 0 & 0 & 0 \end{array}\right) \tag{32b}$$

entsteht. Dies entspricht den degenerierten Konfigurationen von Ebene und Gerade: die Gerade ist parallel zur Ebene in einer Distanz ungleich Null, so dass es keinen Schnittpunkt gibt, oder in einer Distanz Null, so dass jeder Punkt auf der Gerade eine Lösung ist. Gleichung 32a entspricht dem ersten Fall, dass heißt es gibt keine Lösung. Dies folgt algebraisch, da die Gleichung einen Widerspruch enthält, was auch für allgemeine Gleichungssysteme anzeigt, dass es keine Lösung gibt. Gleichung 32b entspricht unendlich vielen Lösungen. Dies hat zur Folge, dass man für die letzte Variable einen beliebigen Wert angeben kann. Für jeden dieser Werte erhält man einen Lösungsvektor. Enthält die Matrix k dieser Null-Zeilen, so erhält man einen $k-$dimensionalen Lösungsraum.

8 Gaußsche Eliminationsmethode mit Pivoting

8.1 Gaußsche Eliminationsmethode

Sei $Ax = b$ ein lineares Gleichungssystem. Bei der Gaußschen Eliminationsmethode wird die Matrix A schrittweise auf eine obere Dreiecksmatrix R reduziert. Anschließend erhält man die Lösung des Gleichungssystems mit Rückwärtseinsetzen. Bezeichnet $A^{(1)} = A$ und $b^{(1)} = b$, so erhalten wir in jedem Schritt, $1 < k \leq n$, eine Matrix

$$A^{(k)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & \cdots & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & \cdots & \cdots & a_{2n}^{(2)} \\ & & \ddots & & & \vdots \\ & & & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ & & & \vdots & & \vdots \\ & & & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} \end{pmatrix}, \tag{33}$$

deren erste bis $(k-1)$-te Spalte unterhalb der Diagonalen Null ist, sowie eine entsprechende rechte Seite:

$$b^{(k)} = \left(b_1^{(k)}, \ldots, b_n^{(k)}\right)^T \tag{34}$$

Der Schritt von $k \to k+1$ ist dabei gegeben durch:

$$l_{ik} = a_{ik}^{(k)} \Big/ a_{kk}^{(k)} \qquad i = k+1, \ldots, n \tag{35}$$

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - l_{ik}\, a_{kj}^{(k)} \qquad i,j = k+1, \ldots, n \tag{36}$$

$$b_i^{(k+1)} = b_i^{(k)} - l_{ik}\, b_k^{(k)} \qquad i = k+1, \ldots, n \tag{37}$$

Dabei müssen vor jedem Schritt gegebenenfalls Zeilen so vertauscht werden, dass das Pivotelement $a_{kk}^{(k)}$ ungleich Null ist.

8.2 Notwendigkeit von Pivoting

Im Folgenden wollen wir uns verdeutlichen, dass eine ungünstige Auswahl des Pivotelementes zu numerischen Problemen führen kann. Wir nehmen dabei an, dass die Berechnungen im Gleitkommazahl-Format $\mathbb{G}(10,3,5)$ mit drei Ziffern in der Mantisse erfolgen, d.h.

$$\mathbb{G} \ni x = \pm m_1\,.\,m_2\,m_3 \times 10^k \tag{38}$$

wobei m_1, m_2, m_3 für die drei Ziffern stehen; zum Beispiel

$$x = 0.312 \;\to\; \hat{x} = +3.12 \times 10^{-1} \tag{39}$$

und die Abbildung $G : \mathbb{R} \to \mathbb{G}$ von den reellen Zahlen $\mathbb{R}$ in die Gleitkommazahlen $\mathbb{G} \equiv \mathbb{G}(10,3,5)$ erfolgt entsprechend den normalen Rundungsregeln.

Betrachtet werden soll das folgende Gleichungssystem:

$$\begin{array}{rcrcl} 0.0001\,x & + & 1.00\,y & = & 1.00 \\ 1.00\,x & + & 1.00\,y & = & 2.00 \end{array} \tag{40}$$

welches als exakte Lösung

$$x = \frac{10000}{9999} \approx 1.0001, \quad y = \frac{9998}{9999} \approx 0.9999 \tag{41}$$

besitzt. In $\mathbb{G}(10,3,5)$ erhalten wir also durch die notwendige Rundung für das Ergebnis:

$$x = 1.0001 = 1.0001 \times 10^0 \;\to\; \hat{x} = 1.00 \times 10^0 \tag{42a}$$

$$y = 0.9999 = 9.9999 \times 10^{-1} \;\to\; \hat{y} = 1.00 \times 10^0. \tag{42b}$$

In Matrixform sieht das Gleichungssystem wie folgt aus:

$$Ax = \begin{pmatrix} 0.0001 & 1.00 \\ 1.00 & 1.00 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.00 \\ 2.00 \end{pmatrix} = b \tag{43}$$

und da alle Werte exakt in unserer Gleitkomma-Darstellung abgebildet werden können haben wir $\hat{A} = G(A) = A$ und es ist keine Rundung bei den Eingangsdaten notwendig und diese können exact in $\mathbb{G}(10,3,5)$ dargestellt werden. Führen wir nun Gauß-Elimination durch, so können wir das Matrixelement A_{21} Null setzen, wenn die erste und zweite Zeile–geeignet skaliert–addiert werden. Der Skalierungsfaktor ergibt sich aus dem Pivotelement A_{11}:

$$l_{21} = \frac{A_{21}}{A_{11}} = \frac{1.0}{0.0001} = 10000.00 \;\to\; \hat{l}_{21} = 1.00 \times 10^4. \tag{44}$$

Die Elemente, welche durch die Addition modifiziert werden müssen, sind (hier mit einen Strich angegeben):

$$A'_{22} = \hat{A}_{22} - \hat{l}_{21} \cdot \hat{A}_{12} \tag{45a}$$
$$= 1.00 - 10000.00 \cdot 1.00 \tag{45b}$$
$$= -9.999 \times 10^3 \tag{45c}$$

und die Repräsentation im Gleitkommazahl-Format ist demzufolge

$$\rightarrow \hat{A}'_{22} = -1.00 \times 10^4. \tag{45d}$$

Analog erhalten wir für die rechte Seite des Gleichungssystems

$$b'_2 = \hat{b}_2 - \hat{l}_{21} \cdot \hat{b}_1 \tag{45e}$$
$$= 2.00 - 10000.00 \cdot 1.00 \tag{45f}$$
$$= -9.998 \times 10^3 \tag{45g}$$
$$\rightarrow \hat{b}'_2 = -1.00 \times 10^4 \tag{45h}$$

Das Gleichungssystem ist in oberer Dreiecksform in $\mathbb{G}(10, 3, 5)$ also durch

$$\hat{A} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.0 \times 10^{-4} & 1.0 \times 10^{0} \\ 0.0 \times 10^{0} & -1.0 \times 10^{4} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.0 \times 10^{0} \\ -1.0 \times 10^{4} \end{pmatrix} \tag{46}$$

gegeben. Die Lösung erhalten wir durch Rückwärtseinsetzen:

$$\hat{y} = 1.0 \times 10^0 = 1.00, \tag{47}$$
$$\hat{x} = 0.0 \times 10^0 = 0. \tag{48}$$

Der relative Fehler für x bezüglich der bestmöglichen Lösung in $\mathbb{G}(10, 3, 5)$ ist damit:

$$E_{\mathrm{r}}(\hat{x}) = \frac{|1.0 - 0.0}{|1.0|} \tag{49}$$

und damit sehr groß.

Wir wollen nun betrachten was passiert, wenn wir vor der Durchführung der Gauß-Elimination die Zeilen vertauschen. Das heißt, wir betrachten das System:

$$\begin{pmatrix} 1.00 & 1.00 \\ 0.0001 & 1.00 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2.00 \\ 1.00 \end{pmatrix} \tag{50}$$

was weiterhin die gleiche Lösung wie Gl. 40 besitzt. Gauß-Elimination ergibt dann:

$$l_{21} = \frac{a_{21}}{a_{11}} = \frac{1. \times 10^{-4}}{1.00} \rightarrow \hat{l}_{21} = 1.00 \times 10^{-4} \tag{51a}$$
$$A'_{22} = A_{22} - l_{21} \cdot A_{12} \tag{51b}$$
$$= 1.00 - 0.0001 \cdot 1.00 \tag{51c}$$
$$= 0.9999 \tag{51d}$$
$$\rightarrow \hat{A}'_{22} = 1.00 \times 10^0 \tag{51e}$$
$$b'_2 = b_2 - l_{21} \cdot b_1 \tag{51f}$$
$$= 1.00 - 0.0001 \cdot 2.00 \tag{51g}$$
$$= 0.9998 \tag{51h}$$
$$\rightarrow \hat{b}'_2 = 1.00 \times 10^0 \tag{51i}$$

In oberer Dreiecksform in $\mathbb{G}(10, 3, 5)$ erhalten wir also

$$\hat{A} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.00 & 1.00 \\ 0 & 1.00 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2.00 \\ 1.00 \end{pmatrix}. \tag{52}$$

Durch Rückwärtseinsetzen erhalten wir also $\hat{x} = 1.0$ und $\hat{y} = 1.00$ und das Ergebnis ist gleich der Darstellung des exakten Ergebnisses in unserer Gleitkommazahl-Format in Gl. 42 ist.

Warum hat die einfache Vertauschung der Zeilen so fundamentale Folgen? Wenn wir $A_{11} = 0.0001$ als Pivot verwenden, dann erhalten wir für die Zwischenergebnisse in Gl. 45, welche groß sind im Vergleich zu den Ergebniswerten. Die Zwischenergebnisse in Gl. 45 müssen gerundet werden, um sie in der verwendeten Gleitkomma-Darstellung abbilden zu können. Der relative Fehler bei dieser Rundung ist klein, $|10000.0 - 9999.0/9999.0| \approx 0.0001$ aber der absolute Fehler, $10000.0 - 9999.0 = 1.0$, ist, in Relation zum Endergebnis, groß. Dieser kleine relative Fehler bei der Rundung der "großen" Zwischenergebnisse führt später zu dem katastrophalen Fehler beim "kleinen" Endergebnis. Dieser Effekt ist ähnlich zur Auslöschung, wo auch der Unterschied in der Größenordnung von Zwischenergebnissen und Endergebnis zu katastrophalen Folgen führt. Wenn wir $A_{11} = 1.0$ als Pivot verwenden, dann erhalten wir Zwischenergebnisse in Gl. 51 welche die gleiche Größenordnung haben wie das Endergebnis. Damit findet keine Verstärkung von Rundungsfehlern statt und das Ergebnis im Gleitkommazahl-Format besitzt einen kleinen relative Fehler.

Im Allgemeinen gilt, dass das Pivot immer so groß wie möglich gewählt werden sollte, um numerische Probleme und insbesondere die Verstärkung von Rundungsfehlern von Zwischenergebnissen zu vermeiden. Die Suche nach einem geeigneten Pivotelement heißt *Pivotsuche*. In Pseudo-Code kann die Gauss-Elimination mit Pivoting wie folgt beschrieben werden:

a) Wähle im Eliminationsschritt $A^{(k)} \to A^{(k+1)}$ ein $p \in \{k, \ldots, n\}$, so dass

$$\left|a_{pk}^{(k)}\right| \geq \left|a_{jk}^{(k)}\right| \text{ für alle } j \in \{k, \ldots, n\}$$

b) Vertausche die Zeilen k und p und führe den Eliminationsschritt aus.

Neben dem hier beschriebenen Zeilen-Pivoting existiert auch Spalten und totales Pivoting. Wir werden diese nicht näher betrachten.