

5.4 Numerische Lineare Algebra

Wir unterscheiden hier zwischen direkten (exakten) Verfahren und Iterationsverfahren. Beginnen wollen wir mit dem klassischen direkten Verfahren, dem Gauss-Verfahren. Ziel ist die Lösung eines Gleichungssystems

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$$

mit $\mathbf{A} \in \mathbb{R}^{(m,n)}$, $\mathbf{x} \in \mathbb{R}^n$ (die Variablen) und $\mathbf{b} \in \mathbb{R}^m$. Wir haben also m Gleichungen in n Unbekannten. Man versucht, die Matrix durch Zeilenumformungen in "Zeilenstufenform" zu bringen. Eine Matrix $\mathbf{A} = (a_{i,j})$ heißt in Zeilenstufenform, wenn es Zahlen $1 \leq j_1 < j_2 < \dots < j_k \leq n$ gibt mit $a_{i,j} = 0$ für $j < j_i$, und es gilt $a_{i,j_i} \neq 0$. Ferner gilt $a_{i,j} = 0$ für $i > k$. Man kann ein lineares Gleichungssystem leicht lösen, wenn \mathbf{A} in Zeilenstufenform gegeben ist. Das Gauss-Verfahren konstruiert eine Folge von Matrizen $\mathbf{L}_i, \mathbf{P}_i$ ($i = 1, \dots, m$) so, dass $\mathbf{P}_i = \mathbf{I}$ ist, oder \mathbf{P}_i ist eine Permutationsmatrix, die man aus \mathbf{I} durch Vertauschung zweier Zeilen erhält. Ferner sind die \mathbf{L}_i quadratische untere Dreiecksmatrizen der Größe $m \times m$ mit Diagonaleinträgen 1, in der alle Spalten mit einer Ausnahme die Einheitsvektoren sind. Man kann die \mathbf{P}_i und die \mathbf{L}_i vertauschen, so dass man erreicht:

$$\mathbf{A} = \mathbf{P} \cdot \mathbf{L} \cdot \mathbf{U}$$

und \mathbf{L} ist eine untere Dreiecksmatrix mit Diagonaleinträgen 1 und \mathbf{U} hat Zeilenstufenform. Man kann diese sogenannte LU -Zerlegung wie folgt finden:

- Es sei $\mathbf{A} \in \mathbb{R}^{(m,n)}$. Wir initialisieren drei Matrizen $\mathbf{P} = \mathbf{L} = \mathbf{I}_m$ und $\mathbf{U} = \mathbf{A}$.
- Setze $k = 1$.
- Suche in \mathbf{U} einen Eintrag $a_{i,j_i} \neq 0$ mit $i \geq k$ so, dass $a_{i',j'} = 0$ für alle Paare (i', j') mit $i' \geq k$ und $j' < j_i$ (in der Regel ist das der Eintrag an der Position (k, k)). Sollte es keinen solchen Eintrag geben, STOP. Sollte $k < i$ gelten, vertausche Zeile i und k in \mathbf{U} sowie die entsprechenden Spalten in \mathbf{P} . Ferner vertauschen wir in diesem Fall in \mathbf{L} die ersten $k - 1$ Einträge der Zeilen k und i .
- Addiere nun $\alpha_{i,k}$ mal die k -te Zeile zu den i 'ten Zeilen ($i > k$), so dass alle Einträge in den Positionen (i, j_k) mit $i > k$ Null werden. Die Einträge (i, k) von \mathbf{L} werden $-\alpha_{i,k}$.
- Erhöhe k um 1 und fahre so fort.

Am Ende gilt, wie oben schon erwähnt,

$$\mathbf{A} = \mathbf{P} \cdot \mathbf{L} \cdot \mathbf{U}$$

und \mathbf{L} ist eine untere Dreiecksmatrix mit Diagonaleinträgen 1 und \mathbf{U} hat Zeilenstufenform.

Im dritten Schritt oben sind Zeilenvertauschungen manchmal zwingend erforderlich, wenn z.B. der erste Eintrag in der ersten Zeile von \mathbf{A} Null ist. Man macht manchmal auch Zeilenvertauschungen von Zeile i und k (mit $i > k$), die eigentlich nicht notwendig sind, damit anschließend $|a_{k,j_k}| \geq |a_{i,j_k}|$ für $i > k$ gilt, wir also in Spalte j_k das betragsmäßig größte mögliche Pivotelement wählen.

Beispiel 5.8.

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 4 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & -7 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 2 & 4 & 1 \\ 2 & 4 & 1 & 1 \end{pmatrix}.$$

Bemerkung 5.9. Wenn man konkret $\mathbf{P} \cdot \mathbf{L} \cdot \mathbf{U} \cdot \mathbf{x} = \mathbf{b}$ lösen will, berechnet man zunächst ein \mathbf{y} mit $\mathbf{L}\mathbf{y} = \mathbf{P}^T\mathbf{b}$ (beachte dass $\mathbf{P}^T = \mathbf{P}^{-1}$). Das geht einfach, weil \mathbf{L} eine untere Dreiecksmatrix ist. Anschließend löst man $\mathbf{U}\mathbf{x} = \mathbf{y}$, was ebenfalls einfach ist.

Bemerkung 5.10 (Nachiteration). Aus numerischen Gründen (Rundungsfehler) erhält man beim Lösen manchmal keine ganz exakte Lösung von $\mathbf{A}\mathbf{x} = \mathbf{b}$ (wir gehen jetzt einmal davon aus, dass es genau eine Lösung gibt). Die Lösung, die man erhält, sei $\mathbf{x}^{(0)}$. Wir berechnen den Defekt $\mathbf{d}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$. Gilt $\mathbf{d}^{(0)} = \mathbf{0}$, so sind wir fertig, wir haben mit $\mathbf{x}^{(0)}$ die exakte Lösung gefunden. Andernfalls lösen wir $\mathbf{A}\mathbf{y}^{(0)} = \mathbf{d}^{(0)}$ und setzen dann $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \mathbf{y}^{(0)}$. Der Defekt von $\mathbf{x}^{(1)}$ sollte nun kleiner sein als der von $\mathbf{x}^{(0)}$ (im Idealfall $\mathbf{0}$). Gegebenenfalls wird nochmals nachiteriert.

Es gibt noch eine weitere wichtige Zerlegung, die in dieser Form allerdings nur für positiv definite (also insbesondere nur für symmetrische) Matrizen funktioniert. Es wird nachher, wenn wir über die Kondition von Matrizen sprechen, deutlich, warum dies eine numerisch günstige Zerlegung ist.

Satz 5.11 (Cholesky-Zerlegung). *Zu jeder symmetrischen positiv definiten Matrix \mathbf{A} gibt es genau eine untere Dreiecksmatrix \mathbf{G} , deren Einträge auf der Diagonalen positiv sind und für die gilt*

$$\mathbf{G} \cdot \mathbf{G}^T = \mathbf{A}.$$

Ist \mathbf{B} positiv semidefinit, so gibt es eine untere Dreiecksmatrix \mathbf{H} , deren Einträge auf der Diagonalen nichtnegativ sind und für die gilt

$$\mathbf{H} \cdot \mathbf{H}^T = \mathbf{B}.$$

Die Matrix \mathbf{H} ist im semidefiniten Fall nicht eindeutig bestimmt.

Beispiel 5.12. Hier ist ein Beispiel einer Cholesky-Faktorisierung:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -2 \\ 2 & 5 & -2 \\ -2 & -2 & 9 \end{pmatrix}.$$

Bemerkung 5.13. Man kann die Cholesky-Faktorisierung leicht ausrechnen: Dazu sei $\mathbf{G} = (g_{i,j})_{i,j=1,\dots,n}$ und $\mathbf{A} = (a_{i,j})_{i,j=1,\dots,n}$. Es gilt

$$g_{j,j} = \sqrt{a_{j,j} - \sum_{k=1}^{j-1} g_{j,k}^2},$$

$$g_{i,j} = \frac{1}{g_{j,j}} \left(a_{i,j} - \sum_{k=1}^{j-1} g_{i,k} g_{j,k} \right), \quad i = j+1, \dots, n.$$

In der ersten dieser beiden Gleichungen betrachtet man einfach das Skalarprodukt der j -ten Zeile von \mathbf{G} mit sich selbst, in den anderen Gleichungen das Skalarprodukt der i -ten Zeile mit der j -ten Zeile. Es sei bemerkt, dass man mit diesem Ansatz auch entscheiden kann, ob eine Matrix positiv definit ist: Wenn man nämlich eine Matrix \mathbf{G} findet, dann ist die Matrix \mathbf{A} positiv definit, andernfalls nicht! Im nicht positiv definiten Fall würde man im Laufe der Berechnungen irgendwann die Situation haben, dass man eine Wurzel aus einer negativen Zahl zieht (dann ist \mathbf{A} mit Sicherheit nicht positiv semidefinit) oder aber in der Matrix \mathbf{G} einen Diagonaleintrag 0 erhält, in welchem Fall die Matrix \mathbf{A} eventuell nur positiv semidefinit ist.

Die letzte Zerlegung, mit der wir uns beschäftigen, ist die QR -Zerlegung. Sie funktioniert (ähnlich wie die Cholesky-Zerlegung) nicht über beliebigen Körpern, sondern nur im reellen (und in leicht abgewandelter Form auch über den komplexen Zahlen). Im Gegensatz existiert die LU -Zerlegung (die ja eigentlich nur das Gauß-Verfahren ist) über beliebigen Körpern.

Satz 5.14 (QR -Zerlegung). Sei $\mathbf{A} \in \mathbb{R}^{(m,n)}$ mit $m \geq n$ und $\text{Rang}(\mathbf{A}) = n$. Die Spalten von \mathbf{A} seien $\mathbf{a}_1, \dots, \mathbf{a}_n$ (Vektoren in \mathbb{R}^m). Dann sind die Vektoren

$$\mathbf{u}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{u}_j, \mathbf{a}_i \rangle}{\|\mathbf{u}_j\|} \cdot \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|}$$

orthogonal. Normierung liefert Vektoren

$$\mathbf{q}_i := \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}.$$

Sei \mathbf{Q} die Matrix, deren Spalten die \mathbf{q}_i sind. Ferner sei $\mathbf{R} = (r_{i,j})_{i,j=1,\dots,n}$ mit

$$r_{i,j} = \begin{cases} \frac{\langle \mathbf{u}_i, \mathbf{a}_j \rangle}{\|\mathbf{u}_i\|} & \text{wenn } i \leq j \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt

$$\mathbf{A} = \mathbf{Q} \cdot \mathbf{R}.$$

Das QR -Verfahren wird bei Matrizen mit "schlechter Kondition" (wird später definiert) angewendet. Man kann damit das sogenannte Ausgleichsproblem lösen:

Satz 5.15. Sei $\mathbf{A} = \mathbf{Q} \cdot \mathbf{R}$ die QR -Zerlegung einer $m \times n$ -Matrix \mathbf{A} vom Rang n . Sei \mathbf{x}^* die eindeutige Lösung von $\mathbf{R}\mathbf{x} = \mathbf{Q}^T\mathbf{b}$. Dann löst \mathbf{x}^* das Problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|.$$

Insbesondere ist \mathbf{x}^* Lösung des linearen Gleichungssystems $\mathbf{A}\mathbf{x} = \mathbf{b}$, sofern dieses lösbar ist.

Bemerkung 5.16. Zur Lösung von $\mathbf{A}\mathbf{x} = \mathbf{b}$ mit $\mathbf{A} \in \mathbb{R}^{(m,n)}$ und $m \leq n$, $\text{Rang}(\mathbf{A}) = m$, bildet man die QR -Zerlegung von $\mathbf{A}^T = \mathbf{Q} \cdot \mathbf{R}$. Dann löst man zunächst $\mathbf{R}^T\mathbf{y} = \mathbf{b}$. Für dieses (eindeutig bestimmte) $\mathbf{y} \in \mathbb{R}^m$ gilt, dass $\mathbf{Q} \cdot \mathbf{y}$ die Lösung kleinster Norm von $\mathbf{A}\mathbf{x} = \mathbf{b}$ ist.

Wir wollen uns nun überlegen, warum es überhaupt so viele verschiedene Verfahren zur Lösung linearer Gleichungssysteme gibt. Dazu definieren wir die folgende Matrixnorm $\|\mathbf{A}\|_2$ (Spektralnorm)

$$\|\mathbf{A}\|_2 := \max_{\mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2.$$

Dabei bezeichnet

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}, \mathbf{x} \in \mathbb{R}^n$$

die übliche Norm.

Weitere Normen auf \mathbb{R}^n wären

- $\|\mathbf{x}\|_1 = \left\| \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\|_1 = \sum_{i=1}^n |x_i|;$
- $\|\mathbf{x}\|_\infty = \left\| \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\|_\infty = \max_{1 \leq i \leq n} |x_i|.$

Auf $\mathbb{R}^{(n,n)}$, $\mathbf{A} = (a_{ij})_{i,j=1,\dots,n}$ definieren wir zugehörige Matrixnormen:

- $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$ (Spaltensummennorm);

- $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$ (Zeilensummennorm).

Es gilt:

Satz 5.17. Sei $\|\cdot\|$ eine der drei Normen $\|\cdot\|_1$, $\|\cdot\|_2$ oder $\|\cdot\|_\infty$ (jeweils für Vektoren als auch Matrizen). Die Normen $\|\mathbf{A}\|$ sind submultiplikativ, d.h.

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|.$$

Ferner gilt

$$\|\mathbf{A} \cdot \mathbf{x}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{x}\|.$$

Bemerkung 5.18. Man kann zeigen, dass die Matrixnorm $\|\cdot\|_2$ die Wurzel des größten Eigenwertes von $\mathbf{A} \cdot \mathbf{A}^T$ ist. Das bedeutet insbesondere, dass die Kondition einer orthogonalen Matrix 1 ist:

Definition 5.19 (Kondition). Die Kondition einer invertierbaren Matrix ist

$$\text{cond}(\mathbf{A}) := \|\mathbf{A}\|_2 \cdot \|\mathbf{A}^{-1}\|_2.$$

Bemerkung 5.20. Man kann die Kondition auch bezüglich der Normen $\|\cdot\|_1$ und $\|\cdot\|_\infty$ definieren.

Der Grund, warum die Kondition einer Matrix im Zusammenhang mit dem Lösen linearer Gleichungssysteme eine Rolle spielt, ist folgender: Sei $\tilde{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$ und $\tilde{\mathbf{b}} = \mathbf{b} + \Delta\mathbf{b}$. Dabei soll $\Delta\mathbf{b}$ klein sein. Sei $\mathbf{A}\mathbf{x} = \mathbf{b}$, dann ist $\mathbf{A} \cdot \Delta\mathbf{x} = \Delta\mathbf{b}$. Sei \mathbf{A} invertierbar. Dann gilt

$$\begin{aligned} \|\Delta\mathbf{x}\| &= \|\mathbf{A}^{-1} \cdot \Delta\mathbf{b}\| \leq \|\mathbf{A}^{-1}\| \cdot \|\Delta\mathbf{b}\| \\ \|\mathbf{b}\| &= \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{x}\|. \end{aligned}$$

Das zeigt

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\| \cdot \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}.$$

Es kann bei großer Kondition also passieren, dass relativ kleine Änderungen der rechten Seite eines Gleichungssystems (Rundungs- oder Messfehler) eine große Änderung der Lösung bewirken.

Beispiel 5.21. Sei

$$\mathbf{A} = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \end{pmatrix}$$

für ein kleines $\epsilon > 0$. Dann ist die Kondition $1/\epsilon$, also sehr groß. Es gilt

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} = \mathbf{A}^{-1} \cdot \begin{pmatrix} \Delta b_1 \\ \Delta b_2 \end{pmatrix} = \begin{pmatrix} (1/\epsilon)\Delta b_1 \\ \Delta b_2 \end{pmatrix}.$$

Das folgende Beispiel zeigt, dass sich die Kondition bei Gauß-Elimination in der Tat stark ändern kann: Wenn wir auf

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 1 \\ -1 & 1 & 0 & 0 & \cdots & 1 \\ -1 & -1 & 1 & 0 & \cdots & 1 \\ \vdots & & & \ddots & \ddots & \vdots \\ \vdots & & & & 1 & 1 \\ -1 & -1 & -1 & \cdots & -1 & 1 \end{pmatrix},$$

dessen Kondition $\mathcal{O}(n)$ ist, Gauß-Elimination anwenden, erhalten wir (wenn wir keine Pivotsuche machen, sondern in jeder Spalte jeweils das “oberste” Element als Pivotelement nehmen)

$$\begin{pmatrix} 1 & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 2 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 2^{n-1} \end{pmatrix},$$

eine Matrix, deren Kondition $\mathcal{O}(2^{n-1})$ ist.

Im Fall von Matrizen mit schlechter Kondition sollte man die QR -Zerlegung verwenden, weil sich die Kondition von \mathbf{R} gegenüber der von \mathbf{A} nicht verschlechtert (Kondition orthogonaler Matrizen ist 1). Deshalb ist auch die Cholesky-Zerlegung numerisch gut.

Wir wollen jetzt abschließend zwei Iterationsverfahren zur Lösung linearer Gleichungssysteme angeben. Diese beiden Verfahren funktionieren für strikt diagonaldominante Matrizen:

Definition 5.22. Eine quadratische Matrix $\mathbf{A} = (a_{i,j})_{i,j=1,\dots,n}$ heißt **strikt diagonal-dominant** falls $|a_{j,j}| > \sum_{i=1, i \neq j}^n |a_{j,i}|$ für alle $j = 1, \dots, n$ gilt.

Bemerkung 5.23. Man kann zeigen (*Übung*), dass jede strikt diagonaldominante Matrix invertierbar ist.

Jacobi-Verfahren

Sei $\mathbf{A} \in \mathbb{R}^{(n,n)}$ und $\mathbf{A} = \mathbf{L} + \mathbf{U} + \mathbf{D}$, wobei \mathbf{D} eine Diagonalmatrix ist. Ferner sei \mathbf{L} eine untere und \mathbf{U} eine obere Dreiecksmatrix jeweils mit Diagonaleinträgen 0. Das Jacobi-Verfahren berechnet iterativ eine Folge von Vektoren $\mathbf{x}^{(i)}$:

- Sei $\mathbf{x}^{(0)} \in \mathbb{R}^n$, $k := 0$.
- $\mathbf{x}^{(k+1)} := \mathbf{D}^{-1}(\mathbf{b} - (\mathbf{L} + \mathbf{U})\mathbf{x}^{(k)})$.
- Setze $k := k + 1$ und wiederhole obige Iteration, bis eine Abbruchbedingung erfüllt ist.

Übliche Abbruchbedingungen sind, dass der Defekt $\|\mathbf{Ax} - \mathbf{b}\|$ oder aber die relative Verbesserung $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|/\|\mathbf{x}^{(k)}\|$ unterhalb einer gewissen Schranke liegen. Man kann auch die absolute Änderungen $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|$ betrachten oder aber einfach nach einer gewissen Anzahl Iterationsschritte aufhören.

Gauss-Seidel Verfahren

Sei $\mathbf{A} \in \mathbb{R}^{(n,n)}$ und $\mathbf{A} = \mathbf{L} + \mathbf{U} + \mathbf{D}$, wobei \mathbf{D} eine Diagonalmatrix ist. Ferner sei \mathbf{L} eine untere und \mathbf{U} eine obere Dreiecksmatrix jeweils mit Diagonaleinträgen 0. Das Gauss-Seidel-Verfahren berechnet iterativ eine Folge von Vektoren $\mathbf{x}^{(i)}$:

- Sei $\mathbf{x}^{(0)} \in \mathbb{R}^n$, $k := 0$
- $\mathbf{x}^{(k+1)} := (\mathbf{D} + \mathbf{L})^{-1}(\mathbf{b} - \mathbf{U}\mathbf{x}^{(k)})$.
- Setze $k := k + 1$ und wiederhole obige Iteration, bis eine Abbruchbedingung erfüllt ist.

Für beide Verfahren gilt:

Satz 5.24. *Sowohl das Jacobi-Verfahren als auch das Gauss-Seidel-Verfahren konvergieren für strikt diagonaldominante Matrizen gegen die Lösung des Gleichungssystems (beachte dass \mathbf{A} im Fall strikt diagonaldominanter Matrizen invertierbar ist) $\mathbf{Ax} = \mathbf{b}$.*