17. VL Einführung in die Num. Mathematik, 10.6.09 ①

Die iterative Lösung linearer Gleichungssysteme

Neben der schon beschriebenen direkten Lösung linearer Gleichungssysteme durch den Gaußschen Algorithmus oder durch bestimmte Matrix-Faktorisierungen ist es oft sinnvoll, lineare Gleichungssysteme

$$A\vec{x} = \vec{b} \qquad (5.11)$$

mit der regulären Matrix A vom Typ n×n und $\vec{b} \in \mathbb{R}^n$ iterativ zu lösen (o.B.d.A sei $a_{kk} \neq 0, k=1,\dots,n$). Zerlegt man A mit der regulären Matrix B in der Form $A = B + (A-B)$ dann gilt für (5.11)

$$A\vec{x} = \vec{b} \Longleftrightarrow B\vec{x} = (B-A)\vec{x} + \vec{b} \Longleftrightarrow \vec{x} = (E - B^{-1}A)\vec{x} + B^{-1}\vec{b}.$$

Wählt man B als leicht invertierbare Matrix, dann ergibt sich im Fall der Konvergenz der Fixpunktiteration

$$\vec{x}_k = (E - B^{-1}A)\vec{x}_{k-1} + B^{-1}\vec{b}, \quad k = 1,2,\dots \qquad (5.12)$$

bei Wahl irgendeiner Startnäherung $\vec{x}_0 \in \mathbb{R}^n$ mit dem Grenzwert $\vec{x} = \lim\limits_{k\to\infty} \vec{x}_k$ die Lösung des lin. Gleichungssystems (5.11). Die Lösung ist ein Fixpunkt der Abb. $\vec{F}: \mathbb{R}^n \to \mathbb{R}^n$

$$\vec{F}(\vec{x}) = (E - B^{-1}A)\vec{x} + B^{-1}\vec{b} \qquad (5.13)$$

Die Matrix $S = (E - B^{-1}A)$ heißt Iterationsmatrix.
Konvergenz liegt dann vor, wenn $\lim_{k\to\infty} \|\vec{x}-\vec{x}_k\| = 0$ ist.
Mit $\vec{x}$ und $\delta\vec{x}_k = \vec{x}-\vec{x}_k$ folgt

$$\delta\vec{x}_k = (E-B^{-1}A)\,\delta\vec{x}_{k-1} = (E-B^{-1}A)^k\,\delta\vec{x}_0 = S^k\,\delta\vec{x}_0 ,$$

also gilt für irgendeine Vektornorm und eine dadurch induzierte Matrixnorm

$$\|\delta\vec{x}_k\| \le \|S^k\|\,\|\delta\vec{x}_0\| . \qquad (5.14)$$

Damit konvergiert das Lösungsverfahren, wenn $\lim_{k\to\infty} S^k = \Theta$ bzw. $\lim_{k\to\infty} \|S^k\| = 0$ gilt.

Hilfreich zur Konvergenzuntersuchung ist der

Satz 5.10

Sei S eine Matrix vom Typ $n\times n$. S habe den Spektralradius $r(S)<1$. Dann sind folgende Aussagen äquivalent:

a) Der Spektralradius $r(S)$ von S ist kleiner als 1.
b) $S^k \to \Theta$ für $k\to\infty$.
c) Es gibt eine Vektornorm, so dass sich für die induzierte Matrixnorm $\|S\|<1$ ergibt.
d) $S-\lambda E$ ist f.a. λ mit $|\lambda|\ge 1$ regulär.

Beweis

a) $\rightarrow$ b)

Betrachten die verallgemeinerte Jordansche Normalform $S = T^{-1}JT$ mit einer regulären Matrix T und J mit den Jordan-Blöcken J_i:

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{pmatrix}, \quad J_i = \begin{pmatrix} \lambda_i & \varepsilon & & \\ & \lambda_i & \varepsilon & \\ & & \ddots & \ddots \\ & & & \lambda_i \;\; \varepsilon \\ & & & & \lambda_i \end{pmatrix}$$

wobei $0 < \varepsilon < 1 - |\lambda_i|$ für $i = 1, \ldots, r$ gewählt wurde

für die Eigenwerte $\lambda_1, \ldots, \lambda_r$ von S. Es gilt $\|S^k\| = \|TJ^kT^{-1}\|$. Die Potenzen von J enthalten für wachsendes k immer größere Potenzen von λ_i, sodass wegen $|\lambda_i| < 1$ f.a. alle Eigenwerte $\|S^k\|$ gegen null geht.

a) $\Rightarrow$ c)

Mit der Zeilensummennorm gilt wegen der Voraussetzung zum Spektralradius von S, der gleich dem von J ist, und der Wahl von ε

$$\|J\|_\infty = \max_{i=1,\ldots,r} \|J_i\|_\infty < 1 .$$

Durch

$$\|\vec{x}\|_T := \|T\vec{x}\|_\infty \quad (\vec{x} \in \mathbb{R}^n)$$

ist eine Norm auf dem $\mathbb{R}^n$ erklärt (Nachweis als Übung). Für die durch $\|\cdot\|_T$ induzierte Matrixnorm gilt $\|S\|_T < 1$, denn es gilt

$$\|S\vec{x}\|_T = \|TS\vec{x}\|_\infty = \|JT\vec{x}\|_\infty \le \|J\|_\infty \|T\vec{x}\|_\infty = \|J\|_\infty \|\vec{x}\|_T$$

und damit $\dfrac{\|S\vec{x}\|_T}{\|\vec{x}\|_T} \le \|J\|_\infty < 1$ f.a. $\vec{x} \ne \Theta$.

Der Nachweis der anderen Implikationen sollte als Übung gezeigt werden (s. auch Plato). $\square$

Als Folgerung des Satzes 5.10 erhält man das folgende Konvergenzkriterium.

Satz 5.11

Seien A und B reguläre $(n \times n)$-Matrizen. Die Iteration (5.12) konvergiert f.a. Startwerte $\vec{x}_0$ genau dann gegen die eindeutig bestimmte Lösung $\vec{x}$ von $A\vec{x} = \vec{b}$, wenn der Spektralradius $r = r(S)$ der Iterationsmatrix $S = (E - B^{-1}A)$ kleiner als 1 ist. Ist S diagonalisierbar, dann gilt

$$\|\vec{x}_k - \vec{x}\| \leq C r^k \quad , \quad C = \text{const.} \in \mathbb{R} . \tag{5.15}$$

Für die weitere Betrachtung konkreter Verfahren stellen wir die quadratische Matrix $A = (a_{ij})$ als Summe der unteren Dreiecksmatrix $L = (l_{ij})$, der Diagonalmatrix $D = (d_{ij})$ und der oberen Dreiecksmatrix $U = (u_{ij})$

$$A = L + D + U \tag{5.16}$$

mit

$$l_{ij} = \begin{cases} a_{ij} & i > j \\ 0 & i \leq j \end{cases}, \quad u_{ij} = \begin{cases} 0 & i \geq j \\ a_{ij} & i < j \end{cases}, \quad d_{ij} = \begin{cases} a_{ij} & i = j \\ 0 & i \neq j \end{cases}$$

dar. Bei Iterationsverfahren der Form (5.12) ist für den Aufwand natürlich die einfache Invertierbarkeit von B entscheidend. Das wird bei den nun zu diskutierenden Verfahren auch berücksichtigt.

Jacobi-Verfahren oder Gesamtschrittverfahren

Die Wahl von $B = D$ ergibt die Iterationsmatrix

$$S = E - B^{-1}A = -D^{-1}(L+U) = \begin{pmatrix} 0 & \frac{a_{12}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} \\ \frac{a_{21}}{a_{22}} & 0 & & \\ \vdots & & \ddots & \\ \frac{a_{n1}}{a_{nn}} & & & 0 \end{pmatrix} \quad (5.17)$$

Das Verfahren (5.12) mit der durch die Wahl von $B = D$ definierten Iterationsmatrix (5.17) heißt Jacobi-Verfahren oder Gesamtschrittverfahren.

Zur besseren Darstellung von Details der Iterationsverfahren setzen wir den Iterationsindex k nach oben in Klammern, also

$$\vec{x}^{(k)} = \vec{x}_k \in \mathbb{R}^n,$$

und die Komponenten von $\vec{x}^{(k)}$ bezeichnen wir durch $x_j^{(k)}$, $j = 1, \ldots, n$. Damit ergibt sich für die Jacobi-Verfahren koordinatenweise

$$x_j^{(k)} = \frac{1}{a_{jj}}\Big(b_j - \sum_{\substack{i=1 \\ i \neq j}}^{n} a_{ji} x_i^{(k-1)}\Big), \quad j = 1, \ldots, n,$$

$k = 1, 2, \ldots$

Zur Konvergenz des Jacobi-Verfahrens gilt der

Satz 5.11

Sei A eine strikt diagonal dominante $(n \times n)$-Matrix. Dann ist der Spektralradius kleiner als 1 und das Verfahren konvergiert.

Beweis

$$S = -D^{-1}(L+U)$$

Zeilensummen von S

$$\sum_{j=1}^{n} s_{ij} = \frac{1}{a_{ii}} \sum_{\substack{j=1\\ j\neq i}}^{n} a_{ij} \,,$$

aufgrund der strikten Diagonaldominanz ist

$$\sum_{\substack{j=1\\ j\neq i}}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| < 1 \rightarrow \|S\|_\infty < 1 \rightarrow r(S) < 1 .$$

□

Bei der numerischen Lösung von elliptischen Randwertproblemen treten oft Matrizen auf, die nicht strikt diagonal dominant sind, aber die folgende etwas schwächere Eigenschaft besitzen.

Definition 5.12

a) Eine Matrix vom Typ $(n \times n)$ heißt schwach diagonal dominant, wenn gilt

$$|a_{ii}| \geq \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|$$ und es gibt einen Index l mit

$$|a_{ll}| > \sum_{\substack{j=1\\ j\neq l}}^{n} |a_{lj}| .$$

b) Eine $(n \times n)$-Matrix $A = (a_{ij})$ heißt irreduzibel, wenn für alle $i,j \in \{1,2,\dots,n\}$ entweder $a_{ij} \neq 0$ ist oder eine Indexfolge $i_1,\dots,i_s \in \{1,2,\dots,n\}$ existiert, so dass $a_{ii_1} a_{i_1 i_2} \cdots a_{i_s j} \neq 0$ ist. Andernfalls

**) heißt A reduzibel.

Satz 5.12
Für eine irreduzible, schwach diagonal dominante Matrix A ist das Jacobi-Verfahren konvergent.
Beweis ist recht aufwendig (s. z.B. Schwarz)

Gauß-Seidel-Iterationsverfahren oder Einzelschrittverfahren

Wählt man ausgehend von der Matrixzerlegung (5.16) $B = L + D$, dann heißt das Iterationsverfahren (5.17) Gauß-Seidel-Verfahren oder Einzelschrittverfahren, d.h. es ergibt sich

$$\vec{x}^{(k)} = \underbrace{(E - B^{-1}A)}_{S}\vec{x}^{(k-1)} + B^{-1}\vec{b} = (L+D)^{-1}(-U\vec{x}^{(k-1)} + \vec{b}) \quad (5.18)$$
$$k = 1, 2, \dots$$

Die Matrix $B = L + D$ ist eine reguläre untere Dreiecksmatrix und damit leicht zu invertieren, was aber keine Arbeit bedeuten wird, wie wir etwas später sehen werden.

Satz 5.13
Das Gauß-Seidel-Verfahren konvergiert für strikt diagonal dominante Matrizen A für beliebige Startiterationen $\vec{x}^{(0)} \in \mathbb{R}^n$.

Beweis Es ist

$$S = E - B^{-1}A = -(L+D)^{-1}U, \quad \lambda\vec{v} = S\vec{v} = (E - B^{-1}A)\vec{v} = -(L+D)^{-1}U\vec{v}, \quad \vec{v} \neq \Theta,$$

für eine EW λ mit dem EV $\vec{v}$,

bzw. $\lambda(L+D)\vec{v} = -U\vec{v}$; $|v_k| = \max_{1 \le i \le n}(|v_i|) > 0$;

Wir betrachten die k-te Zeile:

$$\lambda\left(a_{kk}v_k + \sum_{j=1}^{n} a_{kj}v_j\right) = -\sum_{j=1}^{n} a_{kj}v_j$$

$$\rightsquigarrow \lambda\Big(1+\underbrace{\sum_{\substack{j=1\\j<k}}^{n}\frac{a_{kj}v_j}{a_{kk}v_k}}_{\alpha}\Big) = -\underbrace{\sum_{\substack{j=1\\j>k}}^{n}\frac{a_{kj}v_j}{a_{kk}v_k}}_{\beta}, \quad \left|\frac{v_j}{v_k}\right|\le 1,$$

$$\Leftrightarrow \quad \lambda(1+\alpha) = \beta, \quad |\alpha|<1, |\beta|<1$$

aufgrund der strikten Diagonaldominanz von A.

$$\Leftrightarrow \quad \lambda = \frac{\beta}{1+\alpha} \rightsquigarrow |\lambda| = \frac{|\beta|}{|1+\alpha|} \le \frac{|\beta|}{1-|\alpha|}.$$

Aus der strengen Diagonaldominanz folgt schließlich

$$|\alpha|+|\beta| = \left|\sum_{\substack{j=1\\j<k}}^{n}\frac{a_{kj}v_j}{a_{kk}v_k}\right| + \left|\sum_{\substack{j=1\\j>k}}^{n}\frac{a_{kj}v_j}{a_{kk}v_k}\right| \le \sum_{\substack{j=1\\j\neq k}}^{n}\left|\frac{a_{kj}}{a_{kk}}\right| < 1,$$

woraus

$$1 > \frac{|\beta|}{1-|\alpha|} \ge |\lambda|$$

für alle EW λ folgt $(r(S)<1)$. $\square$

Bemerkung 5.14

Ebenso wie beim Jacobi-Verfahren kann man die Voraussetzung der strikten Diagonaldominanz von A abschwächen.
Das Gauß-Seidel-Verfahren ist für irreduzible, schwach diagonal dominante Matrizen A konvergent.

Wenn man das Gauß-Seidel-Verfahren (5.18) in der äquivalenten Form

$$\vec{x}^{(k)} = D^{-1}\left(-L\vec{x}^{k} - U\vec{x}^{(k-1)} + \vec{b}\right), \quad k=1,2,\ldots \tag{5.19}$$

aufschreibt, erkennt man bei der koordinatenweisen Berechnung der neuen Iteration

$$x_j^{(k)} = \frac{1}{a_{jj}}\left(b_j - \sum_{i=1}^{j-1} a_{ji} x_i^{(k)} - \sum_{i=j+1}^{n} a_{ji} x_i^{(k-1)}\right), \quad j=1,2,\ldots,n; \; k=1,2,\ldots \tag{5.20}$$

zwar, dass auf beiden Seiten der Formeln $\vec{x}^{(k)}$ vorkommt. Allerdings benötigt man zur Berechnung der j-ten Komponente von $\vec{x}^{(k)}$ nur die Komponenten $x_1^{(k)}, \ldots, x_{j-1}^{(k)}$ der neuen Iteration. Diese kennt man aber bereits. Damit kann man die Formel (5.20) für $j=1,2,\ldots,n$ sukzessiv zum Update der Koordinaten von $\vec{x}$ anwenden. Man hat also mit (5.20) eine explizite Berechnungsvorschrift und braucht damit $B = L+D$ nicht wirklich zu invertieren!

<u>Verallgemeinerung des Gauß-Seidel-Verfahrens</u>

Wählt man $S = S_\omega = (D+\omega L)^{-1}[(1-\omega)D - \omega U] =$
$= E - \omega(D+\omega L)^{-1}A$, bzw. $B = \omega(D+\omega L)^{-1}$,
dann hat man mit

$$\vec{x}^{(k)} = S_\omega \vec{x}^{(k-1)} + B^{-1}\vec{b}, \quad k=1,2,\ldots, \quad \omega \in]0,2[$$

das Gauß-Seidel-Verfahren mit Relaxation erklärt. Für $\omega > 1$ spricht man vom sukzessiven Überrelaxationsverfahren und SOR-Verfahren genannt. Das SOR-Verfahren konvergiert in allen Fällen, in denen das Gauß-Seidel-Verfahren ($\omega = 1$) konvergiert. Allerdings kann man in vielen Fällen mit einer Wahl von $\omega > 1$ eine schnellere Konvergenz als mit dem Gauß-Seidel-Verfahren erreichen.

*) c) → d)

Annahme: $S-\lambda E$ singulär, d.h.
es ex. $\vec{x} \neq \Theta$ mit $(S-\lambda E)\vec{x} = \Theta$,
daraus folgt $S\vec{x} = \lambda\vec{x} \Longleftrightarrow \|S\vec{x}\|_T = |\lambda|\,\|\vec{x}\|_T$

$$\Longleftrightarrow \quad \frac{\|S\vec{x}\|_T}{\|\vec{x}\|_T} = |\lambda| \geq 1,$$

andererseits gilt $1 > \|S\|_T \geq \dfrac{\|S\vec{x}\|_T}{\|\vec{x}\|_T}$, d.h.
es ergibt sich ein Widerspruch
und die Annahme war falsch.
Damit ist $S-\lambda E$ regulär für $|\lambda| \geq 1$. $\square$

**) Bemerkung

1) Man kann entscheiden, ob eine Matrix reduzibel oder irreduzibel ist, indem man für $A=(a_{ij})_{i,j=1,\dots,n}$ einen Graphen mit n Knoten konstruiert, in dem eine gerichtete Kante von Knoten i zum Knoten j existiert, wenn $a_{ij} \neq 0$ ist.
Kann man, in diesem Graphen ausgehend von einem Knoten alle anderen Knoten auf einem gerichteten Weg (Folge von gerichteten Kanten) erreichen, ist A irreduzibel, andernfalls reduzibel.

2) Ein weiteres Kriterium zur Entscheidung, ob A vom Typ $n \times n$ irreduzibel oder reduzibel ist, ist folgendes:

Hilfssatz 5.12

Die $(n\times n)$-Matrix A ist irreduzibel, falls es keine Permutationsmatrix P vom Typ $n\times n$ gibt, so dass bei gleichzeitiger Zeilen- und Spaltenpermutation

$$P^T A P = \begin{pmatrix} F & O \\ G & H \end{pmatrix}$$

gilt, wobei F und H quadratische Matrizen sind und O eine Nullmatrix ist, andernfalls ist A reduzibel.

Beispiel

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}, \text{ mit } P^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

ergibt sich

$$P^T A P = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}, \text{ also ist } A \text{ reduzibel.}$$

Das Gleichungssystem $A\vec{x} = \vec{b}$ ist äquivalent zu

$$P^T A P \vec{\tilde{x}} = \vec{\tilde{b}}$$

mit $\vec{\tilde{x}} = P^T \vec{x}$ und $\vec{\tilde{b}} = P^T \vec{b}$, d.h. man kann die Lösung in 2 kleinere Teilaufgaben zerlegen (deshalb reduzibel = zerlegbar).