Iterationsverfahren für lineare Systeme mit Anwendungen

Peter Albrecht, SS 1986

Kapitel I: GRUNDLEGENDE BEGRIFFE UND ERGEBNISSE ÜBER MATRIZEN

1	Eigenwerte	1
2	Matrixnormen	9
3	Fehlerbetrachtungen	13
4	Schranken für die Eigenwerte	17
5	Spezielle Matrizen	22
5.1.	Gerichtete Graphen	22
5.2.	Zerlegbare und schwach-zyklische Matrizen	24
5.3.	Zyklische Reduktion eines Systems	28
5.4.	Nicht-negative Matrizen	29

Kapitel II: ITERATIVE METHODEN FÜR LINEARE SYSTEME

6	Einleitung	33
7	Drei klassische Verfahren	36
7.1.	Die Jacobi-Iteration	36
7.2.	Die Gauß-Seidel-Iteration	37
7.3.	Die Überrelaxation	39
8	Block-Iterationen	41
8.1.	Das Block-Jacobi-Verfahren	42
8.2.	Das Block-Gauß-Seidel-Verfahren und d. Block-SOR	44
9	Konvergenz	46
9.1.	Konvergenzsätze	46
9.2.	Anwendungen	48
10	Vergleichstheoreme und hinreichende Konvergenzbedingungen	50
10.1.	Nicht-negative Jacobi-Matrizen	50
10.2.	Das Zeilen- und Spaltensummenkriterium	53
10.3.	Hermite'sche Matrizen	55
11	Konsistent geordnete Systeme	58
11.1.	Eine Beziehung zwischen den Eigenwerten von J und R_ω	58
11.2.	Eine Klasse konsistent (r.g.)-geordneter Matrizen	62
11.3.	Über den Einfluß von Umordnungen auf die Konvergenz der Iterationsverfahren	67
12	Bemerkungen zur Diskretisierung der partiellen Differentialgleichung (6.3.), $u_{xx} + u_{yy} = 0$	72

Literatur 79

Index 75

Kapitel I

Grundlegende Begriffe und Ergebnisse über Matrizen

Im folgenden werden einige Grundbegriffe und Ergebnisse der Theorie der Matrizen zusammengefaßt, soweit sie im Rahmen dieser Vorlesung benötigt werden.

Es werden die folgenden Bezeichnungen benutzt:

$\mathbb{C}(m,n)$: Raum der komplexwertigen (m,n)-Matrizen
$\mathbb{R}(m,n)$: Raum der reellwertigen (m,n)-Matrizen
$I \in \mathbb{R}(n,n)$: Einheitsmatrix; $\mathcal{O}$: Nullmatrix

A^T: transponierte Matrix von $A=(a_{ij})$; $A^T:=(a_{ji})$
$\overline{A}$: konjugiert komplexe Matrix von $A=(a_{ij})$; $\overline{A}:=(\overline{a}_{ij})$
A^*: adjungierte Matrix von $A=(a_{ij})$; $A^*:=\overline{A}^T$; $(A^*)^*=A$
A^{-1}: inverse Matrix von $A \in \mathbb{C}(n,n)$; $A^{-1}A=AA^{-1}=I$
$|A|$: Betragsmatrix; $|A|=(|a_{ij}|)$.

§ 1 Eigenwerte

Definition 1.1.

Sei $A \in \mathbb{C}(n,n)$ und $\lambda \in \mathbb{C}$, $x \in \mathbb{C}^n$

$$Ax=\lambda x \; ; \quad x \neq \mathcal{O}$$

dann heißt λ Eigenwert von A und x ist der zugehörige Eigenvektor.

Die Eigenwerte von A sind Nullstellen des charakteristischen Polynoms der Matrix $A \in \mathbb{C}(n,n)$:

$$\det(\lambda I - A) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_{n-1}\lambda + c_n \qquad (1.1)$$

mit $c_1 = -\sum_{i=1}^{n} a_{ii}$ und $c_n = (-1)^n \det A$.

$\mathrm{Sp}A := \sum_{i=1}^{n} a_{ii}$ heißt Spur von A.

Nach dem Satz von Viëta gilt für die Eigenwerte λ_j, $j=1(1)n$:

$$\begin{aligned} \lambda_1 + \lambda_2 + \ldots + \lambda_n &= -c_1 &&= \mathrm{Sp}\ A \\ \lambda_1 \cdot \lambda_2 \ldots \lambda_n &= (-1)^n c_n &&= \det A \end{aligned} \qquad (1.2)$$

Verschiedene Matrizen können dasselbe charakteristische Polynom haben; insbesondere hat die Frobenius-Matrix

$$F = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -c_n \\ 1 & 0 & 0 & \ldots & 0 & -c_{n-1} \\ 0 & 1 & 0 & \ldots & 0 & -c_{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 & -c_1 \end{pmatrix}$$

das charakteristische Polynom (1.1)

Satz 1.1.

Sei $A,T \in \mathbb{C}(n,n)$ und T regulär, dann haben $T^{-1}AT$ und A dasselbe charakteristische Polynom und folglich dieselben Eigenwerte.

Die Transformation $T^{-1}AT$ heißt Ähnlichkeitstransformation, $T^{-1}AT$ und A heißen ähnlich.

Beweis von Satz 1.1.

$$\begin{aligned}\det (T^{-1}AT-\lambda I) &= \det (T^{-1}(A-\lambda I)T) \\ &= \det T^{-1} \cdot \det (A-\lambda I) \cdot \det T \\ &= \det (A-\lambda I) \quad \triangleleft\end{aligned}$$

Definition 1.2.

Eine Matrix $A \in \mathbb{C}(n,n)$ heißt hermitesch (bzw. symmetrisch, falls $A \in \mathbb{R}(n,n)$) wenn gilt

$$A = A^*$$

Eine hermitesche Matrix heißt positiv definit, wenn für beliebiges $x \in \mathbb{C}(n,n)$, $x \neq 0$: $\bar{x}^T Ax > 0$ ist.

Satz 1.2.

Die Eigenwerte einer hermiteschen Matrix A und die quadratische Form $Q(x)=\bar{x}^T Ax$, $x \in \mathbb{C}^n$, sind reell.

Beweis:

$Q(x)=\bar{x}^T Ax=\bar{x}^T\bar{A}^T x$, weil A hermitesch

$=x^T\overline{Ax}$, weil für alle $x,y,A: x^TAy=y^TA^Tx$

$=\overline{\bar{x}^TAx} = \overline{Q(x)}$

Daher ist Q(x) reell. Ist λ Eigenwert und y zugehöriger Eigenvektor von A, so ist: $Q(y)=\lambda|y|^2$. Folglich ist auch λ reell. ◁

Der Beweis zeigt, daß die Eigenwerte einer positiv definiten Matrix positiv sind. Allgemeiner gilt:

Satz 1.3.

Eine hermitesche Matrix ist genau dann positiv definit, wenn ihre Eigenwerte positiv sind.

Satz 1.4.

Bei beliebigem $A\in\mathbb{C}(n,n)$ haben AA^* und A^*A dieselben Eigenwerte λ_j, $j=1(1)n$; diese sind reell und es gilt $\lambda_j\geq 0$.

Beweis:

Ist $\mu=0$ Eigenwert von AA^*, so offenbar auch von A^*A. Sei $\mu\neq 0$ Eigenwert von AA^* und $x\neq 0$ zugehöriger Eigenvektor:

$$AA^*x=\mu x$$

Multiplikation von links mit A^* ergibt mit $A^*x=y$:

$$A^*Ay=\mu y$$

μ ist somit auch Eigenwert von A^*A.

Weiter gilt mit $x^* = \bar{x}^T$:

$$\sum_{j=1}^{n} |y_j|^2 = y^*y = x^*AA^*x = \mu x^*x = \mu \sum_{j=1}^{n} |x_j|^2$$

Somit ist μ Eigenwert von A^*A und $\mu > 0$.

Satz 1.5.

Ist λ Eigenwert von A, so ist λ^m Eigenwert von A^m für jedes $m \in \mathbb{N}$. Das gilt auch für $-m \in \mathbb{N}$, wenn $\det A \neq 0$ ist.

Beweis folgt durch Induktion aus

$$Ax = \lambda x \quad \text{und} \quad A^{-1}x = \lambda^{-1}x. \qquad \triangleleft$$

Satz 1.6.

Jede Matrix $A \in \mathbb{C}(n,n)$ läßt sich durch Ähnlichkeitstransformation auf ihre Jordansche Normalform transformieren:

$$N := T^{-1}AT = \begin{bmatrix} B_1 & & & \sigma \\ & B_2 & & \\ & & \ddots & \\ \sigma & & & B_p \end{bmatrix} \qquad (1.3)$$

mit Jordanzellen $B_\ell \in \mathbb{C}(m_\ell, m_\ell)$, $\sum_{\ell=1}^{p} m_\ell = n$, der Form:

$$B_\ell := \begin{bmatrix} \lambda_\ell & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_\ell & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_\ell & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda_\ell \end{bmatrix} \qquad (1.4)$$

wobei λ_ℓ, $\ell = 1(1)p$, die (nicht notwendig verschiedenen) Eigenwerte von A sind.

Den Beweis findet man in der einschlägigen Literatur, z.B. in [2].

Definition 1.3.

Die charakteristischen Polynome der Jordanzellen B_ℓ, heißen (zu dem Eigenwert λ_ℓ gehörige) Elementarteiler von A.

Die zu einem λ gehörigen Elementarteiler von A sind offensichtlich genau dann linear, wenn die zugehörigen Jordanzellen die Dimension 1 haben. Sind alle Elementarteiler von A linear, so hat die Normalform Diagonalgestalt.

Beispiel:

Die Jordansche Normalform einer Matrix A mit den Eigenwerten λ_1, λ_2 und λ_3 sei:

$$N = \begin{bmatrix} \lambda_1 & 0 & & & & & \\ & \lambda_1 & 0 & & & & \\ & & \lambda_2 & 1 & & & \\ & & 0 & \lambda_2 & 0 & & \\ & & & & \lambda_2 & 0 & \\ & & & & & \lambda_3 & 1 & 0 \\ & & & & & 0 & \lambda_3 & 1 \\ & & & & & 0 & 0 & \lambda_3 \end{bmatrix}$$

Zu λ_1 gehören nur lineare Elementarteiler: $(\lambda-\lambda_1)$, zu λ_2 gehören die Elementarteiler $(\lambda-\lambda_2)^2$ und $(\lambda-\lambda_2)$ und zu λ_3 der Elementarteiler $(\lambda-\lambda_3)^3$.

Satz 1.7.

Für beliebiges $r\in\mathbb{N}$ mit $r\geq \max_{1\leq\ell\leq p}(m_\ell-1)$ gilt:

$$N^r=\begin{bmatrix} B_1^r & & \sigma \\ & B_2^r & \\ \sigma & & \ddots \\ & & & B_p^r \end{bmatrix} \quad \text{mit } B_\ell^r=\begin{bmatrix} \lambda_\ell^r & \binom{r}{1}\cdot\lambda_\ell^{r-1} & \dots & \binom{r}{m_\ell-1}\cdot\lambda_\ell^{r-(m_\ell-1)} \\ 0 & \lambda_\ell^r & \dots & \binom{r}{m_\ell-2}\cdot\lambda_\ell^{r-(m_\ell-2)} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \binom{r}{1}\cdot\lambda_\ell^{r-1} \\ 0 & 0 & \dots & \lambda_\ell^r \end{bmatrix} \tag{1.5}$$

Beweis: Sei $D\in\mathbb{R}(m,m)$ mit $d_{i,i+1}=1$, $i=1(1)m-1$, und $d_{ik}=0$ für $k\neq i+1$. Dann hat $F:=D^r$, $0\leq r\leq m-1$, die Elemente $f_{i,i+r}=1$, $i=1(1)m-r$, und $f_{ik}=0$ für $k\neq i+r$, und es ist $D^r=\sigma$ für $r\geq m$.

Daher gilt für $N=\lambda I+D$ und $r\geq m$:

$$N^r=(\lambda I+D)^r=\sum_{j=0}^{m-1}\binom{r}{j}\lambda^{r-j}\,D^j \qquad \triangleleft$$

Aus (1.5) folgt

a) $\lim_{r\to\infty} N^r=\sigma$ genau dann, wenn $|\lambda_j|<1$, $j=1(1)n$

b) $\sup_{r\in\mathbb{N}} |p_{ij}(r)|\leq$ const., $N^r=(p_{ij}(r))$, genau dann, wenn die zu den Eigenwerten vom Betrag 1 gehörigen Elementarteiler linear sind und $|\lambda_j|\leq 1$, $j=1(1)n$.

Aus $A^r=TN^rT^{-1}$ folgt daher der folgende Satz, der für spätere Betrachtungen von großer Bedeutung ist:

Satz 1.8.

Seien λ_j, $j=1(1)n$, die Eigenwerte von $A\in\mathbb{C}(n,n)$ und bezeichnet $a_{ij}(r)$ die Elemente von A^r, so gilt

a) $\lim_{r\to\infty} A^r=\mathcal{O}$ genau dann, wenn $|\lambda_j|<1$, $j=1(1)n$

b) $\sup_{r\in\mathbb{N}}|a_{ij}(r)|\leq$ const.[1] genau dann, wenn $|\lambda_j|\leq 1$, $j=1(1)n$, und wenn zu den Eigenwerten vom Betrag 1 nur lineare Elementarteiler gehören.

[1] mit der im nächsten Abschnitt eingeführten Matrixnorm ist dies der Relation: $\|A^r\| \leq$ const. für alle $r\in\mathbb{N}$ äquivalent.

§ 2 Matrixnormen

Im Raum $\mathbb{C}(n,n)$ der quadratischen, n-reihigen Matrizen wird nun eine Norm definiert:

Definition 2.1.

Die Zahl $\|A\| \geq 0$ heißt Norm von $A \in \mathbb{C}(n,n)$, wenn gilt:

$$\text{(a)} \quad \|A\| = 0 \text{ genau dann, wenn } A = 0 \tag{2.1}$$

$$\text{(b)} \quad \|A+B\| \leq \|A\| + \|B\| \text{ für alle } A,B \in \mathbb{C}(n,n) \tag{2.2}$$

$$\text{(c)} \quad \|cA\| = |c| \, \|A\|; \quad c \in \mathbb{C} \tag{2.3}$$

Die Matrizen aus $\mathbb{C}(n,n)$ können auch als lineare Operatoren von $\mathbb{C}^n$ in $\mathbb{C}^n$ interpretiert werden; es ist daher sinnvoll, die Norm $\|A\|$ zugleich als Operatornorm behandeln zu können. Dazu definieren wir:

Definition 2.2.

Eine Matrixnorm heißt mit einer Vektornorm verträglich (consistent), wenn für beliebige $A \in \mathbb{C}(n,n)$ und $x \in \mathbb{C}^n$ gilt:

$$\|Ax\| \leq \|A\| \cdot \|x\| ; \tag{2.4}$$

sie heißt der Vektornorm zugeordnet (subordinate) oder Operatornorm, wenn es außerdem zu jedem $A \in \mathbb{C}(n,n)$ ein $y \in \mathbb{C}^n$, $y \neq 0$, gibt, so daß gilt: $\|Ay\| = \|A\| \cdot \|y\|$ (2.5)

In der Interpretation von A als Operator im normierten Raum $\mathbb{C}^n$ besagt die Verträglichkeit, daß der Operator beschränkt ist; Relation (2.5) besagt, daß $\|A\|$ Norm des Operators ist, d.h. kleinste obere Schranke von $\frac{\|Ax\|}{\|x\|}$, $x \neq 0$. Diese Schranke wird bei beliebiger Vektornorm angenommen, d.h. die durch

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \tag{2.6}$$

definierte Matrixnorm ist der gewählten Vektornorm zugeordnet.

Für zugeordnete Matrizen gilt:

$$\|I\| = 1 \qquad \text{und} \qquad \|A^m\| \leq \|A\|^m, \; m \in \mathbb{N} \tag{2.7}$$

und aus $\|ABy\| = \|AB\| \cdot \|y\| \leq \|A\| \cdot \|B\| \cdot \|y\|$ folgt für beliebige $A, B \in \mathbb{C}(n,n)$:

$$\text{(d)} \qquad \|AB\| \leq \|A\| \cdot \|B\| \tag{2.8}$$

Oft werden die Relationen (a)-(d) zur Definition der Matrixnorm angegeben.

Häufig benutzte Matrixnormen sind:

1. Die Zeilensummennorm:

$$\|A\|_\infty := \max_i \sum_{j=1}^{n} |a_{ij}| \tag{2.9}$$

Zur Übung zeige man, daß $\|A\|_\infty$ der Vektornorm

$$\|x\|_\infty := \max_j |x_j| \tag{2.10}$$

zugeordnet ist.

2. Die Spaltensummennorm:

$$\|A\|_1 := \max_j \sum_{i=1}^{n} |a_{ij}| \tag{2.11}$$

die der Vektornorm $\|x\|_1 := \sum_{j=1}^{n} |x_j|$ zugeordnet ist. (2.12)

3. Die Spektralnorm:

$$\|A\|_S := \max_i |\sqrt{\mu_i}| \quad ; \quad \mu_i: \text{Eigenwerte von } A^* A \tag{2.13}$$

ist der euklidischen Vektornorm $\|x\|_E := \sqrt{\sum_{i=1}^{n} |x_i|^2}$ (2.14)

zugeordnet (s. den folgenden Satz 2.1.)

4. Die Quadratsummennorm:

$$\|A\|_E := \left| \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2} \right| = \left| \sqrt{Sp(A^* A)} \right| \tag{2.15}$$

ist mit der euklidischen Vektornorm verträglich, ihr aber nicht zugeordnet, dennoch gilt: $\|AB\|_E \leq \|A\|_E \|B\|_E$

Satz 2.1.

Es gilt:

$$\|A\|_S = \max_{x \neq 0} \frac{\|Ax\|_E}{\|x\|_E} \tag{2.16}$$

Beweis: Sei μ_1 größter Eigenwert von $A^* A$. Für alle $x \neq 0$ gilt:

$$\|A\|_S^2 - \frac{\|Ax\|_E^2}{\|x\|_E^2} = \mu_1 - \frac{x^* A^* Ax}{x^* x} = \frac{x^* (\mu_1 I - A^* A) x}{x^* x} \geq 0$$

denn nach Satz 1.4. sind die Eigenwerte von $(\mu_1 I - A^* A)$ größer gleich Null. Das Gleichheitszeichen gilt, wenn x der zu μ_1 gehörige Eigenvektor von $A^* A$ ist.

Daher kann auch (2.16) zur Definition des Spektralradius benutzt werden.

Satz 2.2.

Für $A \in \mathbb{C}(n,n)$ gilt: $\|A\|_S \leq \|A\|_E \leq \sqrt{n}\, \|A\|_S$ (2.17)

Beweis: Aus (2.15) folgt: $\|A\|_E^2 = Sp(AA^*)$

$$\|A\|_E^2 = Sp(AA^*) = \sum_{j=1}^{n} \mu_j \qquad \text{(vgl. (1.2))}$$

wobei μ_j die Eigenwerte von AA^* sind. Da diese nichtnegativ sind, gilt mit $\mu_1 \geq \mu_j$, $j=2(1)n$:

$$\mu_1 \leq \sum_{j=1}^{n} \mu_j \leq n\mu_1$$

Im folgenden Abschnitt verwenden wir Matrixnormen zur Gewinnung von Fehlerabschätzungen bei linearen Systemen.

§ 3 Fehlerbetrachtungen

Betrachten wir einleitend das folgende Beispiel von T.S. Wilson[1]:

$$\begin{aligned} 10x_1 + 7x_2 + 8x_3 + 7x_4 &= 32 \\ 7x_1 + 5x_2 + 6x_3 + 5x_4 &= 23 \\ 8x_1 + 6x_2 + 10x_3 + 9x_4 &= 33 \\ 7x_1 + 5x_2 + 9x_3 + 10x_4 &= 31 \end{aligned} \qquad (3.1)$$

Substituiert man hier die Näherung

$$x_1 = 9{,}2 \; ; \quad x_2 = -12{,}6 \; ; \quad x_3 = 4{,}5 \; ; \quad x_4 = -1{,}1 \qquad (3.2)$$

so ergibt sich auf der linken Seite von (3.1)

$$b_1 = 32{,}1 \; ; \quad b_2 = 22{,}9 \; ; \quad b_3 = 33{,}1 \; ; \quad b_4 = 30{,}9$$

und man könnte meinen, (3.2) sei eine gute Näherung für die Lösung von (3.1).

Mit der völlig anderen "Lösung"

$$x_1 = 1{,}82 \; ; \quad x_2 = -0{,}36 \; ; \quad x_3 = 1{,}35 \; ; \quad x_4 = 0{,}79 \qquad (3.3)$$

erhält man jedoch

$$b_1 = 32{,}01 \; ; \quad b_2 = 22{,}99 \; ; \quad b_3 = 33{,}01 \; ; \quad b_4 = 30{,}99$$

Offenbar ist (3.3) eine bessere Näherung als (3.2), aber sie ist noch weit entfernt von der exakten Lösung: $x_1 = x_2 = x_3 = x_4 = 1$.

---

[1] s. Todd,J.: Survey of Numerical Analysis, McGraw-Hill, 1962, S. 242.

Es zeigt sich also, daß der Defekt $\Delta b := A\bar{x} - b$ kein brauchbares Kriterium für die Güte der Approximation $\bar{x}$ ist und es stellt sich die Frage, wann mit Phänomenen der oben aufgezeigten Art zu rechnen ist.

Für die Determinante des Systems (3.1) gilt: $\det A = 1$. Ersetzt man z.B. das Element $a_{11} = 10$ von A durch 9,96, so wird $\det A = 0$; in der Kleinheit von $\det A$ ist eine der Ursachen für die Empfindlichkeit zu sehen, mit der der Lösungsvektor x auf kleine Änderungen des Vektors b reagiert, sie ist aber kein hinreichendes Kriterium für das im Beispiel gezeigte Verhalten.

Betrachten wir den durch einen Fehler Δb von b hervorgerufenen Fehler Δx:

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

$$A\Delta x = \Delta b$$

$$\|\Delta x\| \leq \|A^{-1}\| \, \|\Delta b\| \,, \quad \text{bei verträglichen Normen.}$$

Aus $\|b\| = \|Ax\| \leq \|A\| \cdot \|x\|$ folgt $\|x\| \geq \|b\| \cdot \|A\|^{-1}$ und für den relativen Fehler von x erhält man die Abschätzung:

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

Damit ist der folgende Satz bewiesen:

Satz 3.1.

Ist A regulär und hat b den Fehler Δb, so gilt für den Fehler Δx der Lösung x von $Ax = b$:

$$\frac{\|\Delta x\|}{\|x\|} \leq K(A) \frac{\|\Delta b\|}{\|b\|} \quad \text{mit } K(A) := \|A\| \cdot \|A^{-1}\| \tag{3.4}$$

Definition 3.1.

Die Zahl

$$K(A) := \|A\| \cdot \|A^{-1}\| \; ; \; A: \text{regulär} \tag{3.5}$$

heißt Konditionszahl der Matrix A; insbesondere nennen wir $K_s(A) := \|A\|_s \cdot \|A^{-1}\|_s$ spektrale Konditionszahl. Das Problem Ax=b heißt schlecht konditioniert, wenn K(A) groß ist.

Im obigen Beispiel (3.1) ist $K_s(A) \approx 2984$.

Neben (3.5) werden oft noch andere Konditionszahlen angegeben, z.B. von Turing:

$$N(A) = \frac{1}{n} \|A\|_E \cdot \|A^{-1}\|_E \tag{3.6}$$

und von Todd:

$$P(A) = \frac{\max |\lambda_i|}{\min |\lambda_i|} \; ; \quad \lambda_i: \text{Eigenwerte von } A \tag{3.7}$$

Für reguläre hermitesche Matrizen ist $P(A) = K_s(A)$.

Satz 3.2.

Für beliebiges, reguläres A gilt: $K_s(A) \leq K_s(A^* A)$ (3.8)

Beweis: Bezeichne $\mu_{max}(A^*A)$ bzw. $\mu_{min}(A^*A)$ den größten bzw. kleinsten Eigenwert von A^*A, dann ist

$$\|A\|_s = \sqrt{\mu_{max}} \text{ und } \|A^{-1}\|_s = \sqrt{\mu_{min}^{-1}}$$

und $$\|A^* A\|_s = \mu_{max}, \; \|(A^*A)^{-1}\|_s = \mu_{min}^{-1}$$

Daher:

$$K_s(A) = \sqrt{\frac{\mu_{max}}{\mu_{min}}} \leq \frac{\mu_{max}}{\mu_{min}} = K_s(A^* A)$$

Das Problem $A^*Ax=b$ ist somit i.allg. schlechter konditioniert als $Ax=b$. Es empfiehlt sich daher i.allg. nicht, das Problem $Ax=b$ durch Miltiplikation mit A^* zu symmetrisieren.

Ähnlich wie in (3.4) läßt sich eine Schranke für den relativen Fehler von x auch für den Fall angeben, daß (statt der Elemente b der rechten Seite) die Matrix A einen Fehler ΔA hat:

Satz 3.3.

Sei ΔA der Fehler der regulären Matrix A, sei $\|A^{-1}\| \cdot \|\Delta A\| < 1$ und sei die verwendete Matrixnorm eine Operatornorm; dann gilt für den Fehler Δx der Lösung des Systems $Ax=b$:

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{K(A)}{1-K(A)\frac{\|\Delta A\|}{\|A\|}} \frac{\|\Delta A\|}{\|A\|} \tag{3.9}$$

Beweis: Sei $G=(I+A^{-1}\Delta A)^{-1}$. Wegen $\|A^{-1}\| \cdot \|\Delta A\| < 1$ existiert G und man kann zeigen (vgl. Satz 4.4.), daß gilt:

$$\|G\| \leq \frac{1}{1-\|A^{-1}\| \cdot \|\Delta A\|} \tag{3.10}$$

Wegen $GA^{-1}=(A+\Delta A)^{-1}$ existiert dann auch $(A+\Delta A)^{-1}$.
Aus $(A+\Delta A)(x+\Delta x)=b$ folgt

$$\begin{aligned}
(A+\Delta A)\Delta x &= -\Delta A\, x \\
\Delta x &= -(A+\Delta A)^{-1} \cdot \Delta A x \\
&= -[A(I+A^{-1}\Delta A)]^{-1}\, \Delta A x \\
&= -G\, A^{-1}\, \Delta A x \\
\|\Delta x\| &\leq \|G\| \cdot \|A^{-1}\| \cdot \|\Delta A\| \cdot \|x\|
\end{aligned}$$

Mit (3.10) ist

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{\|A^{-1}\| \cdot \|\Delta A\|}{1-\|A^{-1}\| \cdot \|\Delta A\|} = \frac{\|A^{-1}\| \cdot \|A\| \frac{\|\Delta A\|}{\|A\|}}{1-\|A^{-1}\| \cdot \|A\| \cdot \frac{\|\Delta A\|}{\|A\|}}$$

§ 4 Schranken für die Eigenwerte

Definition 4.1.

Seien λ_j, $j=1(1)n$, die Eigenwerte von $A\in\mathbb{C}(n,n)$; dann heißt die Zahl

$$\rho(A)=\max_{1\le j\le n}|\lambda_j| \tag{4.1}$$

Spektralradius von A.

Satz 4.1.

Für jedes reguläre $A\in\mathbb{C}(n,n)$ gilt bei beliebiger, konsistenter Matrixnorm:

$$\frac{1}{\|A^{-1}\|}\le|\lambda_j|\le\|A\|\quad,\quad j=1(1)n \tag{4.2}$$

Beweis: Sei x ein zum Eigenwert λ_j gehöriger Eigenvektor, dann folgt aus $Ax=\lambda_j x$: $\|\lambda_j x\|=\|Ax\|\le\|A\|\cdot\|x\|$. Daher $|\lambda_j|\le\|A\|$. Ebenso folgt aus $A^{-1}x=\lambda_j^{-1}x$: $|\lambda_j^{-1}|\le\|A^{-1}\|$ und daher $\|A^{-1}\|^{-1}\le|\lambda_j|$.

Satz 4.2.

Zu jeder Matrix A und beliebigem $\varepsilon>0$ gibt es eine zugeordnete Matrixnorm, so daß gilt

$$\rho(A)\le\|A\|\le\rho(A)+\varepsilon \tag{4.3}$$

Den Beweis findet man in der angegebenen Literatur. Die in Satz 4.2. verwendete Operatornorm hängt von der Matrix A und von ε ab. (4.3) zeigt, daß $\rho(A)$ das Infinum von $\|A\|$ ist, genommen über alle Operatornormen:

$$\rho(A)=\inf_{\{\|\cdot\|\}}\|A\|$$

Mit den Normen (2.9), (2.11), (2.15) und (2.13) ergeben sich die Abschätzungen:

Korollar 4.2.1.

$$\rho(A) \leq \max_i \sum_{j=1}^{n} |a_{ij}| \tag{4.4}$$

$$\rho(A) \leq \max_j \sum_{i=1}^{n} |a_{ij}| \tag{4.5}$$

$$\rho(A) \leq \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2} \tag{4.6}$$

$$\frac{1}{\|A^{-1}\|_s} = \sqrt{\mu_{min}(A^* A)} \leq |\lambda_j(A)| \leq \sqrt{\mu_{max}(A^* A)} = \|A\|_s \tag{4.7}$$

wobei $\mu_{min}(A^*A)$ und $\mu_{max}(A^*A)$ den betragskleinsten bzw. betragsgrößten Eigenwert von A^*A bezeichnet.

(4.4) und (4.5) ergibt, angewandt auf die Frobeniusmatrix F, die folgenden, bekannten Abschätzungen für die Nullstellen λ_j des Polynoms (1.1):

$$\begin{aligned} |\lambda_j| &\leq \max\{1+|c_1|,\ 1+|c_2|,\ldots,1+|c_{n-1}|,\ |c_n|\} \\ |\lambda_j| &\leq \max\{1, \sum_{j=1}^{n} |c_j|\} \end{aligned} \tag{4.8}$$

(4.4)und (4.5) auf $D^{-1}AD$ angewandt, ergibt mit

$$D := \begin{pmatrix} d_1 & & \sigma \\ & d_2 & \\ \sigma & & \ddots \\ & & & d_n \end{pmatrix} \quad ; \quad d_j > 0,\ j=1(1)n,$$

Korollar 4.2.2.

Für beliebige positive Zahlen d_j, $j=1(1)n$, gilt

$$\rho(A)\leq\min(m_1,m_2) \tag{4.9}$$

$$\text{mit } m_1:=\max_i \{d_i^{-1}\sum_{j=1}^{n} d_j|a_{ij}|\} = \|D^{-1}AD\|_\infty$$

$$m_2:=\max_j \{d_j\sum_{i=1}^{n} d_i^{-1}|a_{ij}|\}= \|D^{-1}AD\|_1$$

Mittels Korollar 4.2.2. lassen sich manchmal genauere Schranken für den Spektralradius finden.

Beispiel:

$$A=\begin{pmatrix}3 & 0 & 1\\ 3 & 2 & 3\\ 2 & 0 & 2\end{pmatrix} \quad ; \quad \lambda_1=1;\ \lambda_2=2;\ \lambda_3=4$$

(4.4) und (4.5) ergeben: $\rho(A)\leq 8$. Mit $d_1=d_3=1$, $d_2=3$ erhält man $m_1=4$, $m_2=6$, somit $\rho(A)\leq 4$.

Ein einfaches Kriterium zur Abschätzung der Eigenwerte einer Matrix A liefert der folgende Satz von Gerschgorin:

Satz 4.3.

Sei $A=(a_{ij})\in\mathbb{C}(n,n)$ und $r_i:=\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|$, $i=1(1)n$.

Dann liegen alle Eigenwerte von A in der Vereinigung der folgenden Kreise in der komplexen Ebene:

$$|\lambda-a_{ii}|\leq r_i\ , \quad i=1(1)n. \tag{4.10}$$

Beweis: Sei x der zum Eigenwert λ gehörige Eigenvektor mit $\max_{1\leq j\leq n} |x_j| = x_i = 1$. Aus $Ax=\lambda x$ folgt dann

$$(\lambda - a_{ii})x_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij}x_j, \quad i=1(1)n$$

Daher liegt jeder Eigenwert λ in einer Kreisscheibe

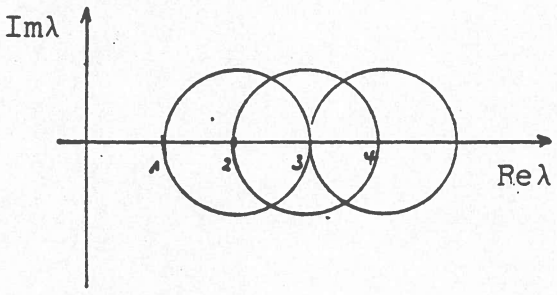
$$|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \cdot |x_j| \leq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| = r_i \ ,$$

Somit liegen alle Eigenwerte in der Vereinigung solcher Kreisscheiben.

Hinweis: Da A^T und A dieselben Eigenwerte haben, kann r_i durch die Spaltensummen: $\sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}|$ ersetzt werden.

Beispiel:

Die Eigenwerte der Matrix $A=\begin{pmatrix} 3 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & 4 \end{pmatrix}$ liegen in der Vereinigung der Kreisscheiben $|\lambda-2|\leq 1; |\lambda-3|\leq 1; |\lambda-4|\leq 1$



Ohne Beweis führen wir noch die folgende Verschärfung des Satzes 4.3. für unzerlegbare Matrizen an, die auf Seite 25 definiert werden:

Korollar 4.3.

Ist A unzerlegbar und liegt ein Eigenwert λ auf dem Rande der Vereinigung aller Kreisscheiben $|\lambda - a_{ii}| \leq r_i$, so gehen alle Kreise $|\lambda - a_{ii}| = r_i$, $i=1(1)n$, durch λ.

Für spätere Zwecke beweisen wir abschließend noch:

Satz 4.4.

Sei $A \in \mathbb{C}(n,n)$ und $\rho(A)<1$. Dann existiert $(I-A)^{-1}$ und es gilt

$$(I-A)^{-1}=I+A+A^2+\ldots$$

Beweis: Ist λ Eigenwert von A, so ist $(1-\lambda)$ Eigenwert von $(I-A)$. Wegen $|\lambda|<1$ ist $(I-A)$ regulär.

Aus $I=(I-A)(I+A+\ldots+A^k)+A^{k+1}$ folgt

$$(I-A)^{-1}=I+A+\ldots+A^k+(I-A)^{-1}\cdot A^{k+1}$$

Somit gilt für alle $k \in \mathbb{N}$: $\|(I-A)^{-1}-I-A-\ldots-A^k\| \leq \|(I-A)^{-1}\| \cdot \|A^{k+1}\|$

Aus Satz 1.8. folgt $\lim_{k\to\infty} \|A^k\| = 0$.

§ 5 Spezielle Matrizen

5.1. Gerichtete Graphen

Die Theorie der Graphen hat in der Vergangenheit Anwendung auf viele Bereiche der Mathematik gefunden. Auch gewisse Eigenschaften von Matrizen lassen sich aus einem zugeordneten Graphen ablesen.

Dazu ordnen wir auf die folgende Weise einer Matrix $A \in \mathbb{C}(n,n)$ einen gerichteten Graphen zu:

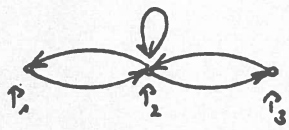
Von n Knoten P_i, i=1(1)n, verbinden wir P_i mit P_j durch einen gerichteten Weg, wenn $a_{ij} \neq 0$ ist.

Definition 5.1.

Der so gewonnene endliche, gerichtete Graph heißt Graph von A und wird mit G(A) bezeichnet. G(A) heißt stark zusammenhängend, wenn es für zwei beliebige Knoten (P_i, P_j) $i \neq j$, stets wenigstens einen gerichteten Weg von P_i nach P_j gibt. Unter der Länge eines gerichteten Weges verstehen wir die Anzahl seiner Kanten.

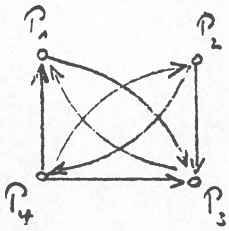
Beispiele:

$$A_1 = \begin{pmatrix} 0 & 2 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$ hat den Graphen:



$G(A_1)$ ist stark zusammenhängend.

$$A_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$ hat den Graphen:



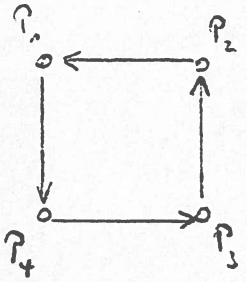
$G(A_2)$ ist nicht stark zusammenhängend, denn es existiert kein gerichteter Weg von P_3 nach P_2. Der kleinste gerichtete Weg von P_2 nach P_1 hat die Länge 2.

Definition 5.2.

Sei $G(A)$ stark zusammenhängend und sei p der größte gemeinsame Teiler der Längen aller geschlossenen, gerichteten Wege in $G(A)$; dann heißt $G(A)$ zyklisch vom Index p, wenn $p>1$ ist, anderenfalls primitiv.

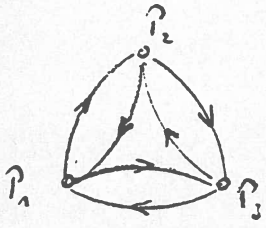
Beispiele:

$$A_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$ hat den Graphen:



$G(A_3)$ ist zyklisch vom Index 4.

$$A_4 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$ hat den Graphen:



$G(A_4)$ ist primitiv, denn er enthält geschlossene, gerichtete Wege der Länge 2 und 3.

Im nächsten Abschnitt werden wir sehen, welche Eigenschaften einer Matrix sich aus ihrem Graphen ablesen lassen.

5.2. Zerlegbare und schwach-zyklische Matrizen

Durch Multiplikation mit Permutationsmatrizen können die Zeilen oder Spalten einer Matrix umgeordnet werden. Permutationsmatrizen enthalten in jeder Zeile und in jeder Spalte genau ein Element 1 und sonst nur Nullen und werden im folgenden mit P bezeichnet.

Beispiel:

Multipliziert man eine Matrix $A \in \mathbb{C}(4,4)$ von links mit der Permutationsmatrix:

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

so hat die Matrix $B := PA$ dieselben Zeilen wie A, jedoch gemäß $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$ permutiert.
Ebenso permutiert man die Spalten von A gemäß $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$ durch Multiplikation von rechts mit P^T, $C := AP^T$. ◁

Die Transformation PAP^T bewirkt daher eine gleichzeitige Permutation der Zeilen und Spalten von A.

Aus $PIP^T = I$ folgt $P^T = P^{-1}$ (5.1)

PAP^T ist daher eine Ähnlichkeitstransformation.

Es ist leicht einzusehen, daß die Transformation PAP^T den Graphen G(A) invariant läßt und nur die Numerierung seiner Knoten permutiert. Daher lassen sich aus dem Graphen G(A) diejenigen Eigenschaften der Matrix A ablesen, die bezüglich der Transformation PAP^T invariant sind, insbesondere die Eigenschaften "zerlegbar" und "schwach-zyklisch", die im folgenden definiert werden.

Definition 5.3.

Eine Matrix $A \in \mathbb{C}(n,n)$, $n \geq 2$, heißt zerlegbar, wenn es eine Permutationsmatrix P gibt, so daß gilt:

$$PAP^T = \begin{pmatrix} A_{11} & \sigma \\ A_{21} & A_{22} \end{pmatrix} \tag{5.2}$$

mit $A_{11} \in \mathbb{C}(r,r)$, $A_{21} \in \mathbb{C}(n-r,r)$, $A_{22} \in \mathbb{C}(n-r,n-r)$, $1 \leq r < n$;

anderenfalls heißt sie unzerlegbar.

Ist A zerlegbar, so kann Ax=b auf zwei Systeme niedrigerer Ordnung reduziert werden:

$$A_{11}x_1 = b_1 \; ; \quad A_{22}x_2 + A_{21}x_1 = b_2 \; ,$$

Ob A zerlegbar ist oder nicht, erkennt man am Graphen G(A); es gilt nämlich der folgende Satz:

Satz 5.1.

Eine Matrix $A \in \mathbb{C}(n,n)$ ist genau dann unzerlegbar, wenn G(A) stark zusammenhängend ist.

Der Beweis wird als Übung empfohlen.

Beispiel: Die Matrix A_2 auf Seite 23 ist reduzierbar, wie der zugehörige Graph zeigt. Durch die Permutation $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$ ihrer Zeilen und Spalten, d.h. durch die Transformation PAP^T mit

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$ geht sie über in $B = PA_2P^T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$

Der Graph G(B) ist mit G(A) identisch bis auf die Numerierung seiner Knoten, die gemäß $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$ permutiert ist.

Definition 5.4.

Eine Matrix $A_B \in \mathbb{C}(n,n)$ [1] heißt schwach-zyklisch vom Index p, p>1, wenn A durch Ähnlichkeitstransformation mit einer Permutationsmatrix P auf die Normalform schwach-zyklischer Matrizen:

$$PA_BP^T = \begin{pmatrix} \sigma & \sigma & \dots & \sigma & A_{1p} \\ A_{21} & \sigma & \dots & \sigma & \sigma \\ \sigma & A_{32} & \dots & \sigma & \sigma \\ \dots & \dots & \dots & \dots & \dots \\ \sigma & \sigma & \dots & A_{p,p-1} & \sigma \end{pmatrix} \qquad (5.4)$$

transformiert werden kann, wobei die Nullmatrizen in der Hauptdiagonalen quadratisch sind.

Hinweis: Eine Matrix kann, abhängig von der gewählten Aufteilung in Blöcke, schwach-zyklisch mit verschiedenem Index sein. Die folgende Matrix A_1, beispielsweise, ist schwach-zyklisch vom Index 2, 3 oder 4, je nachdem welche Blockunterteilung gewählt wird:

$$A_1 = \left(\begin{array}{c|c|cc} 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad ; \text{ (Blockbildung für Index 3)}$$

Es gilt der folgende Satz [2]:

Satz 5.2.

$A \in \mathbb{C}(n,n)$ ist genau dann schwach-zyklisch vom Index p, wenn der Graph G(A) zyklisch vom Index p ist oder wenn er durch Zusammenlegung von Knoten und Hinzufügen von Wegen zyklisch vom Index p (und somit auch stark zusammenhängend) gemacht werden kann.

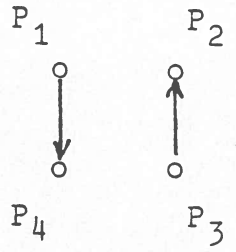
---

[1] im folgenden bedeutet der Index B eine Unterteilung der Matrix in Blöcke; er wird immer dann verwendet, wenn eine Block-Struktur von A hervorgehoben werden soll.

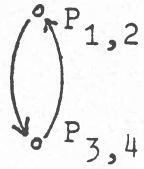
[2] aus P. Albrecht: Numerische Analysis, Livro Técnico Científico, Rio de Janeiro, 1973

Beispiele:

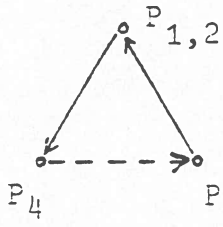
1. Die Matrix $A_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ hat den Graphen:



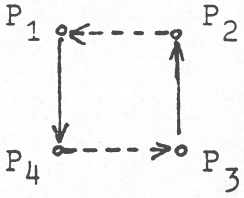
a) Durch Zusammenlegung der Knoten P_1 mit P_2 sowie P_3 mit P_4 ergibt sich ein zyklischer Graph vom Index 2.



b) Durch Zusammenlegung von P_1 mit P_2 und durch Hinzufügen eines gerichteten Weges von P_4 nach P_3 ergibt sich ein zyklischer Graph vom Index 3.

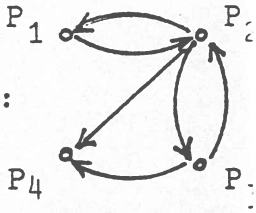


c) Durch Hinzufügen gerichteter Wege von P_2 nach P_1 und von P_4 nach P_3 ergibt sich der nebenstehende Graph.



A_1 ist daher, je nach Block-Struktur, zyklisch vom Index 2, 3 oder 4.

2. Die (zerlegbare) Matrix $A_2 = \begin{pmatrix} 0 & x & 0 & 0 \\ x & 0 & x & x \\ 0 & x & 0 & x \\ 0 & 0 & 0 & 0 \end{pmatrix}$ hat den Graphen:



$G(A_2)$ kann nicht zyklisch gemacht werden, denn Zusammenlegung von Punkten und Hinzufügen von Wegen ergibt höchstens primitive Graphen.

5.3. Zyklische Reduktion eines Systems

Zyklische Matrizen sind von großer Bedeutung bei der iterativen Behandlung linearer Gleichungssysteme (s. Teil II, § 5), jedoch ist auch die folgende Anwendung zur Reduktion der Ordnung eines linearen Systems von Interesse:

Gegeben sei ein lineares System der Form

$$\begin{pmatrix} I_{11} & -A_{12} \\ -A_{21} & I_{22} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}; \quad \begin{matrix} I_{11} \varepsilon \mathbb{R}(r,r);\ I_{22} \varepsilon \mathbb{R}(n-r,n-r) \\ x_1 \varepsilon \mathbb{R}^r;\ x_2 \varepsilon \mathbb{R}^{n-r} \end{matrix}$$

(5.4) läßt sich schreiben in der Form

$$(I-A_B)x=b \quad \text{mit} \quad A_B = \begin{pmatrix} \sigma & A_{12} \\ A_{21} & \sigma \end{pmatrix} \tag{5.5}$$

A_B ist schwach-zyklisch vom Index 2.

Multiplikation von links mit $(I+A_B)$ ergibt

$$\begin{pmatrix} I_{11}-A_{12}\cdot A_{21} & \sigma \\ \sigma & I_{22}-A_{21}\cdot A_{12} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1+A_{12}\cdot b_2 \\ b_2+A_{21}\cdot A_1 \end{pmatrix} \tag{5.6}$$

Dieses System ist zerlegbar:

$$(I_{11}-A_{12}\ A_{21})x_1 = b_1+A_{12}\ b_2 \tag{5.7}$$

$$(I_{22}-A_{21}\ A_{12})x_2 = b_2+A_{21}\ b_1 \tag{5.8}$$

Ist $x_1 \varepsilon \mathbb{R}^r$ Lösung von (5.7), so ergibt sich $x_2 \varepsilon \mathbb{R}^{n-r}$ aus (5.4) durch einfache Multiplikation:

$$x_2 = A_{21}\cdot x_1 + b_2$$

Damit ist die Lösung eines Systems n-ter Ordnung zurückgeführt auf die Lösung eines Systems der Ordnung r. Dieses Verfahren heißt zyklische Reduktion des Systems (5.5). Es beruht darauf, daß A_B^p zerlegbar ist, wenn A_B schwach-zyklisch vom Index p ist.

5.4. Nicht-negative Matrizen

Definition 5.5.

$A=(a_{ij})\in\mathbb{R}(n,n)$ heißt nicht-negativ, wenn für alle i und j $a_{ij}\geq 0$ ist. Wir schreiben $A\geq 0$ bzw. $A\geq B$, wenn gilt: $a_{ij}\geq 0$ bzw. $a_{ij}\geq b_{ij}$.

Hinweis: Es kann $A\geq B$ sein, ohne daß $A>B$ oder $A=B$ gilt (z.B. ist $I\geq\mathcal{O}$ jedoch $I\neq\mathcal{O}$ und $I\not>\mathcal{O}$).

Die Theorie nicht-negativer Matrizen wurde von Perron (1907) und Frobenius (1912) entwickelt; die wichtigsten Ergebnisse sind im folgenden Satz zusammengefaßt, dessen Beweis man in der angegebenen Literatur findet:

Satz 5.3.

Sei $A=(a_{ij})\in\mathbb{R}(n,n)$ nicht-negativ und unzerlegbar, dann gilt:

a) $\lambda_1:=\rho(A)>0$ ist einfacher Eigenwert von A,

b) Alle Eigenwerte λ_j von A mit $|\lambda_j|=\rho(A)$, $j=1(1)r$, sind verschieden und Lösung von $\lambda^r-\lambda_1^r=0$.

c) Zum Eigenwert λ_1 existiert ein Eigenvektor x_1 mit positiven Komponenten: $x_1>\mathcal{O}$.

d) Es gibt keinen von x_1 linear unabhängigen, nicht-negativen Eigenvektor.

e) $\rho(A)$ wächst, wenn irgendein a_{ij} vergrößert wird.

Hiermit läßt sich nun für nicht-negative, unzerlegbare Matrizen der folgende Einschließungssatz für den Spektralradius beweisen:

Satz 5.4.

Sei $A=(a_{ij})\in\mathbb{R}(n,n)$ nicht-negativ und unzerlegbar, dann gilt entweder:

$$\sum_{j=1}^{n} a_{ij}=\rho(A) \text{ für alle } i=1(1)n \tag{5.9}$$

oder

$$\min_{1\le i\le n}\sum_{j=1}^{n} a_{ij}<\rho(A)<\max_{1\le i\le n}\sum_{j=1}^{n} a_{ij} \tag{5.10}$$

Beweis:

a) Sei $\sum_{j=1}^{n} a_{ij}=a$ für alle $i=1(1)n$. Dann folgt aus (4.4): $\rho(A)\le a$. Andererseits hat A den Eigenvektor x mit den Komponenten $x_i=1$, $i=1(1)n$, und a ist zugehöriger Eigenwert. Nach Definition des Spektralradius gilt daher $\rho(A)\ge a$. Somit ist $\rho(A)=a$.

b) Wenn die Summen $\sum_{j=1}^{n} a_{ij}$, $i=1(1)n$, nicht alle gleich sind, betrachten wir die Matrix $B\ge\mathcal{O}$, $B\neq\mathcal{O}$, die sich ergibt, wenn die Elemente von A so vergrößert werden, daß für alle $i=1(1)n$ gilt:

$$\sum_{j=1}^{n} b_{ij}=b=\max_{1\le i\le n}\sum_{j=1}^{n} a_{ij}$$

Gemäß a) ist dann $\rho(B)=b$ und aus Satz 5.3.(e) folgt: $\rho(A)<b$.

Ebenso findet man durch Verkleinern der Elemente von A eine unzerlegbare, nicht-negative Matrix C, $C\neq\mathcal{O}$

mit $\sum_{j=1}^{n} c_{ij}=c=\min_{1\le i\le n}\sum_{j=1}^{n} a_{ij}$ und $\rho(A)>c$. ◁

Hinweis: Da A^T und A dieselben Eigenwerte haben, können die Zeilensummen in (5.9) und (5.10) durch die Spaltensummen ersetzt werden.

Weiter läßt sich zeigen:

Satz 5.5.

Sei $A\in\mathbb{R}(n,n)$ und $B\in\mathbb{C}(n,n)$ mit $|B|\leq A$; dann gilt:

$$\rho(B)\leq\rho(A). \tag{5.11}$$

Für $A=|B|$ folgt insbesondere:

$$\rho(B)\leq\rho(|B|) \tag{5.12}$$

Beweis: Sei $\varepsilon>0$ beliebig und $A_1:=(\rho(A)+\varepsilon)^{-1}A$ und $B_1:=(\rho(A)+\varepsilon)^{-1}B$

Dann folgt aus $\rho(cA)=c\rho(A)$, $c\in\mathbb{R}^+$:

$$\rho(A_1)<1 \quad\text{und}\quad |B_1|^r\leq A_1^r \quad\text{für alle } r\in\mathbb{N}$$

Nach Satz 1.8. ist $\lim\limits_{r\to\infty} A_1^r=\mathcal{O}$, daher auch $\lim\limits_{r\to\infty} B_1^r=\mathcal{O}$ und somit $\rho(B_1)<1$.

Aus $\rho(B)=(\rho(A)+\varepsilon)\rho(B_1)<\rho(A)+\varepsilon$ für jedes $\varepsilon>0$ folgt $\rho(B)\leq\rho(A)$. ◁

Bedenkt man, daß jede zerlegbare, nicht-negative Matrix als Grenzwert einer Folge $\{A_n\}$ unzerlegbarer, nicht-negativer Matrizen darstellbar ist, so erhält man die folgende Verallgemeinerung von Satz 5.3., bei der nicht mehr vorausgesetzt wird, daß A unzerlegbar ist:

Satz 5.6.

Sei $A=(a_{ij})\in\mathbb{R}(n,n)$ nicht-negativ; dann gilt:

a) $\lambda_1:=\rho(A)\geq 0$ ist Eigenwert von A,

b) zum Eigenwert λ_1 existiert ein Eigenvektor x_1 mit nicht-negativen Komponenten: $x_1\geq\mathcal{O}$.

c) $\rho(A)$ nimmt nicht ab, wenn irgendein a_{ij} vergrößert wird.

Für das Kapitel II. benötigen wir noch das folgende Ergebnis:

Satz 5.7.

Seien $B,C\in\mathbb{R}(n,n)$, B regulär, $B^{-1}\geq\mathcal{O}$, $C\geq\mathcal{O}$ und $S=B-C$.

Existiert S^{-1} und ist es nicht-negativ, so gilt

$$\rho(B^{-1}C)=\frac{\rho(S^{-1}C)}{1+\rho(S^{-1}C)}<1 \qquad (5.13)$$

Beweis:

a) Zunächst wird bewiesen: Ist x Eigenvektor von $G:=S^{-1}C$ und λ zugehöriger Eigenwert, so ist x auch Eigenvektor von $B^{-1}C$ und $\mu=\lambda(1+\lambda)^{-1}$ zugehöriger Eigenwert; ist umgekehrt μ Eigenwert von $B^{-1}C$ so ist $\lambda=\mu(1-\mu)^{-1}$ Eigenwert von G.

Es ist: $B^{-1}C=(S+C)^{-1}C=[S(I+S^{-1}C)]^{-1}C=(I+G)^{-1}G$ (5.14)

Ist $\lambda\neq-1$ Eigenwert von G: $Gx=\lambda x$, so folgt $(I+G)x=(1+\lambda)x$ und mit (5.14): $B^{-1}Cx=(I+G)^{-1}Gx=\lambda(1+\lambda)^{-1}x$.

x ist daher Eigenvektor von $B^{-1}C$ und $\mu:=\lambda(1+\lambda)^{-1}$ zugehöriger Eigenwert. Ist umgekehrt x Eigenvektor von $B^{-1}C$ und μ zugehöriger Eigenwert, so folgt:

$$B^{-1}Cx=(I+G)^{-1}Gx=\mu x$$

d.h. $$Gx=\mu(I+G)x$$

Daher ist $\mu\neq1$ und $Gx=\mu(1-\mu)^{-1}x$

x ist daher Eigenvektor von G und $\lambda:=\mu(1-\mu)^{-1}$ zugehöriger Eigenwert.

b) Sei x nicht-negativer Eigenvektor zu $\mu_1=\rho(B^{-1}C)$, dann ist x auch Eigenvektor von $G:=S^{-1}C$ und der zugehörige Eigenwert λ muß positiv sein. Aus der Monotonie von $\lambda(1+\lambda^{-1})$ für $\lambda\geq0$ und Satz 5.6.(a) folgt, daß $\lambda=\rho(S^{-1}C)$ sein muß, d.h.

$$\rho(B^{-1}C)=\mu_1=\frac{\lambda}{1+\lambda}=\frac{\rho(S^{-1}C)}{1+\rho(S^{-1}C)} \qquad \triangleleft$$

Kapitel II

Iterative Methoden für lineare Systeme

§ 6 Einleitung

Das Problem $f(x)=0$, $f:\mathbb{R}\to\mathbb{R}$, läßt sich bekanntlich iterativ lösen, indem man es in die Form $x=\phi(x)$ überführt (was auf unendlich viele Arten möglich ist) und gemäß $x^{(k+1)}=\phi(x^{(k)})$ mit einem Startwert x_0 iteriert.

Völlig analog geht das System $Ax=b$ durch die Zerlegung
$A=M-N$, $A,M,N\in\mathbb{C}(n,n)$, A,M: regulär

über in
$$Mx=Nx+b \tag{6.1}$$

und kann, ausgehend von einem Startvektor x_0, iterativ gemäß

$$Mx^{(k+1)}=Nx^{(k)}+b \;;\quad k=0,1,\ldots \tag{6.2}$$

behandelt werden. Es gibt unendlich viele Zerlegungen von A und jede definiert ein iteratives Verfahren der Form (6.2), dessen Konvergenz vom Spektralradius der Matrix $M^{-1}N$ abhängt.

Definition 6.1.

Die Matrix $H:=M^{-1}N$ hießt Iterationsmatrix.

Iterative Verfahren sind besonders bei großem n nützlich; große Systeme treten in der Praxis insbesondere bei der numerischen Lösung partieller Differentialgleichungen durch Diskretisierung auf. Hierzu betrachten wir das folgende einfache Beispiel:

Gegeben sei die elliptische partielle Differentialgleichung

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0 \tag{6.3}$$

Gesucht ist die Lösung im Quadrat $Q: 0<x<1,\ 0<y<1$, wenn u auf dem Rand von Q bekannt ist.

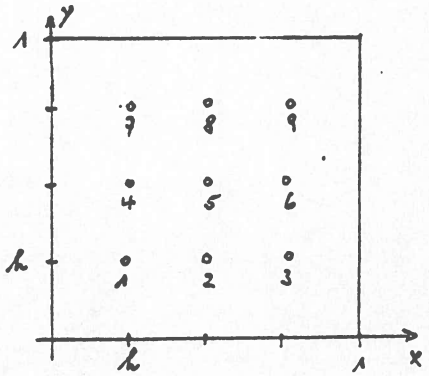


Fig. 6.1.

Diskretisiert man (6.3) in den Punkten $P_{ij}=(x_i,y_j)$, $x_i=hi$, $y_j=hj$, $h=\frac{1}{n+1}$, mittels endlicher Differenzen, so ergibt sich das folgende lineare System für die n^2 Näherungen u_{ij} an den inneren Punkten P_{ij}, $i=1(1)n$, $j=1(1)n$:

$$(u_{i+1,j}-2u_{i,j}+u_{i-1,j})h^{-2}+(u_{i,j+1}-2u_{i,j}+u_{i,j-1})h^{-2}=0 \qquad (6.4)$$

d.h. $$u_{i+1,j}+u_{i-1,j}+u_{i,j+1}+u_{i,j-1}-4u_{ij}=0$$

Ordnet man die Punkte P_{ij} zeilenweise (vgl. Fig. 6.1.), d.h. setzt man: $x_\ell=u_{ij}$, $\ell=i+n(j-1)$, so geht (6.4) über in:

$$x_{\ell+1}+x_{\ell-1}+x_{\ell+n}+x_{\ell-n}-4x_\ell=0\ ; \quad \ell=1(1)n^2 \qquad (6.5)$$

wenn alle auftretenden Indizes in der Menge $\{1,2,\ldots,n^2\}$ enthalten sin die Terme, bei denen das nicht der Fall ist, gehören zu einem Randpunkt, sind also bekannt und bilden die rechte Seite des Systems (6.5).

Für n=3 lautet dieses:

$$\begin{pmatrix} 4 & -1 & 0 & -1 & 0 & 0 & & & \\ -1 & 4 & -1 & 0 & -1 & 0 & & \mathcal{O} & \\ 0 & -1 & 4 & 0 & 0 & -1 & & & \\ -1 & 0 & 0 & 4 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 4 & 0 & 0 & -1 \\ & & & -1 & 0 & 0 & 4 & -1 & 0 \\ & \mathcal{O} & & 0 & -1 & 0 & -1 & 4 & -1 \\ & & & 0 & 0 & -1 & 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_9 \end{pmatrix} = \begin{pmatrix} (u_{01}+u_{10}) \\ u_{20} \\ (u_{30}+u_{41}) \\ u_{02} \\ 0 \\ u_{42} \\ (u_{03}+u_{14}) \\ u_{24} \\ (u_{34}+u_{43}) \end{pmatrix} \qquad (6.6)$$

Bei feinerer Diskretisierung, d.h. größerem n, wächst die Zahl der Matrixelemente sehr schnell; für n=20 sind es bereits 160 000, von denen allerdings weniger als 2 000 von Null verschieden sind.

In der Praxis wird A daher gar nicht im Rechner gespeichert, vielmehr ordnet man jedem inneren Punkt P_{ij} des Quadrats Q einen Speicherplatz zu und belegt ihn mit einem k-ten Näherungswert $u_{ij}^{(k)}$. Gemäß (6.4) kann man dann einen neuen Iterationswert $u_{ij}^{(k+1)}$ als arithmetisches Mittel der k-ten Näherungen in den Nachbarpunkten berechnen:

$$u_{ij}^{(k+1)} = \frac{1}{4}\left(u_{i+1,j}^{(k)} + u_{i-1,j}^{(k)} + u_{i,j+1}^{(k)} + u_{i,j-1}^{(k)}\right)$$

Dieses Vorgehen entspricht der Auflösung des Systems (6.6) nach den Diagonalelementen und anschließender Iteration. Hierbei kann man a) bei der Berechnung der $x_\ell^{(k+1)} := u_{ij}^{(k+1)}$ nur die Werte $x_\ell^{(k)} := u_{ij}^{(k)}$ aus dem k-ten Iterationsschritt benutzen ("Gesamtschritt-Verfahren") oder b) die im (k+1)-ten Iterationsschritt gewonnenen Werte $x_\ell^{(k+1)} := u_{ij}^{(k+1)}$ sofort bei der Berechnung nachfolgender Nährungen $x_m^{(k+1)}$, $m>\ell$, desselben Iterationsschrittes verwenden ("Einzelschritt-Verfahren"). Wir werden später sehen, daß im vorliegenden Fall die Methode b) schneller konvergiert als a).

§ 7 Drei klassische Verfahren

Sei $A=D-E-F$ mit $\det D \neq 0$ und

$$D=\begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}; \quad E=\begin{pmatrix} 0 & & 0 \\ -a_{21} & \ddots & \\ \vdots & \ddots & \ddots \\ -a_{n1} & \cdots & -a_{n,n-1} \quad 0 \end{pmatrix}; \quad F=\begin{pmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ & \ddots & \ddots & \vdots \\ & & \ddots & -a_{n-1,n} \\ 0 & & & 0 \end{pmatrix}$$

$$L:=D^{-1}E; \quad U:=D^{-1}F \tag{7.1}$$

Aus den Zerlegungen: $A=D-(E+F)$ (7.2)

$$A=(D-E)-F \tag{7.3}$$

$$A=\frac{1}{\omega}(D-\omega E)-\frac{1}{\omega}(\omega F+(1-\omega)D) \tag{7.4}$$

ergeben sich die Verfahren von Jacobi und Gauss-Seidel sowie das Überrelaxationsverfahren, die im folgenden vorgestellt werden.

7.1. Die Jacobi Iteration

Die Zerlegung (7.2) ergibt das Iterationsverfahren von Jacobi (Gesamtschritt-Verfahren):

$$x^{(k+1)}=Jx^{(k)}+D^{-1}b; \quad J:=L+U \tag{7.5a}$$

Ausgeschrieben lautet es:

$$x_i^{(k+1)}=-\sum_{j=1}^{i-1}\frac{a_{ij}}{a_{ii}}x_j^{(k)}-\sum_{j=i+1}^{n}\frac{a_{ij}}{a_{ii}}x_j^{(k)}+\frac{b_i}{a_{ii}}; \quad i=1(1)n \tag{7.5b}$$

d.h. das System wird nach den Diagonalelementen aufgelöst und iteriert. Konvergenzaussagen für dieses Verfahren findet man in den Sätzen 9.1., 9.2., 10.3.

Definition 7.1.

Die Matrix J=L+U heißt Jacobi-Matrix von A.

Beispiel

Für

$$\begin{aligned} 3x_1 + x_2 + x_3 &= 5 \\ x_1 + 2x_2 &= 3 \\ x_1 + \tfrac{1}{2}x_2 + 2x_3 &= 6 \end{aligned} \qquad (7.6)$$

lautet die Jacobi-Iteration:

$$\begin{aligned} x_1^{(k+1)} &= -\tfrac{1}{3}x_2^{(k)} - \tfrac{1}{3}x_3^{(k)} + \tfrac{5}{3} \\ x_2^{(k+1)} &= -\tfrac{1}{2}x_1^{(k)} + \tfrac{3}{2} \\ x_3^{(k+1)} &= -\tfrac{1}{2}x_1^{(k)} - \tfrac{1}{4}x_2^{(k)} + 3 \end{aligned}$$

und die Jacobi-Matrix ist:

$$J = \begin{pmatrix} 0 & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{4} & 0 \end{pmatrix} \quad ; \quad \rho(J) = \tfrac{1}{4}(1+\sqrt{\tfrac{7}{3}}) \approx 0{,}63 \qquad (7.7)$$

7.2. Die Gauss-Seidel-Iteration

Die Zerlegung (7.3) ergibt das Iterationsverfahren von Gauss-Seidel (Einzelschritt-Verfahren):

$$x^{(k+1)} = Lx^{(k+1)} + Ux^{(k)} + D^{-1}b \qquad (7.8a)$$

$$x^{(k+1)} = (I-L)^{-1}\,Ux^{(k)} + (D-E)^{-1}b \qquad (7.8b)$$

(Die Inversen existieren wegen $\det(I-L)=1$, $\det D \neq 0$)

Ausgeschrieben lautet es:

$$x_i^{(k+1)} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}} \; ; \; i=1(1)n \qquad (7.8c)$$

Wie man sieht, unterscheidet sich das Gauss-Seidel-Verfahren vom Jacobi-Verfahren nur darin, daß bereits berechnete Komponenten von $x_\ell^{(k+1)}$ zur Berechnung der nachfolgenden Komponenten $x_m^{(k+1)}$, $m>\ell$, benutzt werden (siehe Beispiel). Daher erübrigt sich die beim Jacobi-Verfahren erforderliche Speicherung von $x^{(k)}$ und $x^{(k+1)}$; stattdessen wird $x^{(k)}$ im Laufe der Rechnung mit $x^{(k+1)}$ überschrieben und somit nur ein Vektor gespeichert. Unter bestimmten Voraussetzungen konvergiert das Gauss-Seidel-Verfahren schneller als das Jacobi-Verfahren; man beachte hierzu die Sätze 10.1. und 11.1. (Korollar).

Definition 7.2.

Die Matrix $R_1=(I-L)^{-1}U$ heißt Gauss-Seidel-Matrix von A.

Beispiel:

Für das System (7.6) lautet die Gauss-Seidel-Iteration:

$$\begin{aligned} x_1^{(k+1)} &= -\tfrac{1}{3} x_2^{(k)} - \tfrac{1}{3} x_3^{(k)} + \tfrac{5}{3} \\ x_2^{(k+1)} &= -\tfrac{1}{2} x_1^{(k+1)} + \tfrac{3}{2} \\ x_3^{(k+1)} &= -\tfrac{1}{2} x_1^{(k+1)} - \tfrac{1}{4} x_2^{(k+1)} + 3 \end{aligned} \qquad (7.9)$$

und durch sukzessive Substitution der Werte $x_1^{(k+1)}$ und $x_2^{(k+1)}$ findet man die Gauss-Seidel-Matrix:

$$R_1 = \begin{pmatrix} 0 & -\frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{8} & \frac{1}{8} \end{pmatrix} \; ; \quad \rho(R_1) = \tfrac{7}{24} \approx 0{,}29 \qquad (7.10)$$

Hinweis: R_1 wird nicht etwa explizit zur Berechnung der Werte $x^{(k)}$ ausgerechnet!

7.3. Die Überrelaxation

Analog dem Problem $f(x)=0$, das auf $x=x+\omega f(x)$ mit beliebigem Parameter ω zurückgeführt und dann iterativ gelöst werden kann, läßt sich auch das lineare System $b-Ax=0$ auf

$$x=x+\omega D^{-1}(b-Ax) \tag{7.11}$$

mit beliebigem Parameter ω zurückführen und mit Einzelschritt-Iterationen lösen. Das entspricht der Zerlegung (7.4) von A und man erhält so das Überrelaxationsverfahren:

$$\begin{aligned} x^{(k+1)} &= x^{(k)}+\omega D^{-1}(b-Dx^{(k)}+Ex^{(k+1)}+Fx^{(k)}) \\ &= (1-\omega)x^{(k)}+\omega(Lx^{(k+1)}+Ux^{(k)}+D^{-1}b) \end{aligned} \tag{7.12a}$$

Das Überrelaxationsverfahren kann somit auch als gewichtetes Mittel (7.12a) von $x^{(k)}$ und der Gauss-Seidel-Näherung $\tilde{x}^{(k+1)}$ interpretiert werden. Für $\omega=1$ geht es über in das Gauss-Seidel-Verfahren.

Mit $\det(I-\omega L)^{-1}=1$ und $R_\omega=(I-\omega L)^{-1}(\omega U+(1-\omega)I)$ folgt aus (7.12a):

$$x^{(k+1)}=R_\omega x^{(k)}+\omega(D-\omega E)^{-1}b \tag{7.12b}$$

Ausgeschrieben lautet es:

$$x_i^{(k+1)}=(1-\omega)x_i^{(k)}+\omega\left(-\sum_{j=1}^{i-1}\frac{a_{ij}}{a_{ii}}x_j^{(k+1)}-\sum_{j=i+1}^{n}\frac{a_{ij}}{a_{ii}}x_j^{(k)}+\frac{b_i}{a_{ii}}\right) \tag{7.12c}$$

$$i=1(1)n$$

In Satz 9.3. wird gezeigt werden, daß $0<\omega<2$ notwendig für die Konvergenz des Verfahrens ist. In vielen Fällen wird $\omega>1$ gewählt (s. Satz 10.2. und 11.3.), d.h. die "Korrektur" in (7.11) ist größer als der Defekt $D^{-1}(b-Ax)$; daher der Name "Überrelaxation".

Allgemeine Konvergenzaussagen für das Verfahren finden sich in den Sätzen 9.1., 9.2., 9.3., 10.2., 10.5., 11.3.

Definition 7.3.

Die Matrix $R_\omega=(I-\omega L)^{-1}(\omega U+(1-\omega)I)$ heißt Überrelaxationsmatrix von A.

Satz 7.1.

Hat $L\in\mathbb{C}(n,n)$ untere Dreiecksgestalt, dann ist

$$(I-\omega L)^{-1}=I+\omega L+(\omega L)^2+\ldots+(\omega L)^{n-1} \tag{7.13}$$

und

$$\det(I-\omega L)^{-1}=1 \tag{7.14}$$

Der Beweis folgt aus

$$(I-\omega L)(I+\omega L+\ldots+(\omega L)^{n-1})=I-\omega^n L^n$$

mit $L^n=\sigma$.

Beispiel

Angewandt auf das System (7.6) lautet die Überrelaxation:

$$\begin{aligned}
x_1^{(k+1)}&=(1-\omega)x_1^{(k)}+\omega(-\tfrac{1}{3}\,x_2^{(k)}-\tfrac{1}{3}\,x_3^{(k)}+\tfrac{5}{3})\\
x_2^{(k+1)}&=(1-\omega)x_2^{(k)}+\omega(-\tfrac{1}{2}\,x_1^{(k+1)}+\tfrac{3}{2})\\
x_3^{(k+1)}&=(1-\omega)x_3^{(k)}+\omega(-\tfrac{1}{2}\,x_1^{(k+1)}-\tfrac{1}{4}\,x_2^{(k+1)}+3)
\end{aligned} \tag{7.15}$$

Man vergleiche (7.15) mit der Gauss-Seidel-Iteration (7.9).

§ 8 Block-Iterationen

Wir denken uns im folgenden die Matrix $A \in \mathbb{C}(n,n)$ in Blöcke unterteilt:

$$A_B = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \cdot & \cdot & \cdots & \cdot \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{pmatrix}$$

wobei die A_{ii}, $i=1(1)N$, quadratisch sind: $A_{ii} \in \mathbb{C}(m_i,m_i)$ mit $\sum_{i=1}^{N} m_i = n$. Eine solche Blockunterteilung von A wird häufig durch die Form von A nahegelegt. In Beispiel (6.6), zum Beispiel, ist sie Folge der Numerierung der Punkte P_{ij}: Den Punkten P_{ij}, $i=1(1)N$, j: fest (d.h. jeder Zeile in Fig. 6.1.) entspricht ein Diagonalblock A_{ii}, $i=1(1)N$. In diesem Zusammenhang werden die zugehörigen Block-Methoden auch "Zeilen-Methoden" (line-methods) genannt.

Wendet man nun die im vorigen Abschnitt dargestellten Iterationsverfahren formal an auf das System

$$\sum_{j=1}^{N} A_{ij}x_j = b_i; \quad b_i, x_i \in \mathbb{R}^{m_i}; \quad i=1(1)N$$

so ergeben sich neue Verfahren der Gestalt (6.2). Mit den Blockmatrizen:

$$D_B = \begin{pmatrix} A_{11} & & \sigma \\ & A_{12} & \\ \sigma & & \ddots \\ & & & A_{1N} \end{pmatrix}; \quad E_B = \begin{pmatrix} \sigma & & & \\ -A_{21} & \ddots & & \sigma \\ \vdots & \ddots & \ddots & \\ -A_{N1} & \cdots & -A_{N,N-1} & \sigma \end{pmatrix}; \quad F_B = \begin{pmatrix} \sigma & -A_{12} & \cdots & -A_{1N} \\ & \ddots & \ddots & \vdots \\ & & \ddots & -A_{N-1,N} \\ \sigma & & & \sigma \end{pmatrix}$$

(8.1)

haben $L_B := D_B^{-1}E_B$ und $U_B := D_B^{-1}F_B$ untere bzw. obere (Block-) Dreiecksgestalt, und wie in § 7 erhält man mit den Zerlegungen (7.2-4) das Block-Jacobi- bzw. Block-Gauss-Seidel-Verfahren

sowie die Block-Überrelaxation, die im folgenden behandelt werden.

In jedem Iterationsschritt sind jetzt N lineare Systeme (z.B. mit nicht-iterativen Verfahren) zu lösen, denn die Inversion von D_B erfolgt jetzt nicht mehr durch einfache Division durch die Diagonalelemente, wie bei den entsprechenden Verfahren in § 7.

Im allgemeinen ist es ratsam, die Inversen A_{ii}^{-1} vor Beginn der Iteration zu berechnen und für die Lösung der N Systeme bereitzustellen. Häufig (vgl. z.B. (6.6)) bestehen die Matrizen A_{ii} aus einer Hauptdiagonalen und einer oberen und unteren Nebendiagonalen (Tridiagonalmatrizen), und sind daher besonders einfach zu behandeln[1].

Vorteile der Block-Iteration:

1. Sie wird angewandt, wenn das System so groß ist, daß es nicht im Rechner gespeichert werden kann.
2. In bestimmten Fällen konvergiert die Block-Iteration asymptotisch schneller als die einfache Iteration (man beachte die Definition 9.1., Satz 10.6. und die Beispiele (8.3) und (8.6) auf Seite 43/44).

8.1. Das Block-Jacobi-Verfahren

Mit der Zerlegung $M=D_B$ und $N=E_B+F_B$ ergibt (6.2) das Block-Jacobi-Verfahren:

$$x^{(k+1)}=J_B x^{(k)}+D_B^{-1}b \; ; \quad J_B:=L_B+U_B \tag{8.2}$$

Definition 8.1.

Die Matrix $J_B=L_B+U_B$ heißt Block-Jacobi-Matrix von A.

---

[1] siehe z.B. Varga |4|, S. 195

Beispiele:

1. Sei

$$A_B := \left(\begin{array}{cc:cc:c} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ \hdashline 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ \hdashline 0 & 0 & 0 & -1 & 1 \end{array}\right) \quad ; \quad b = \begin{pmatrix} 3 \\ -5 \\ 9 \\ -6 \\ 2 \end{pmatrix} \tag{8.3}$$

Mit der angegebenen Blockunterteilung lautet $D_B x^{(k+1)} = (E_B + F_B) x^{(k)} + b$ ausgeschrieben:

$$\begin{aligned} 2x_1^{(k+1} - x_2^{(k+1)} &= 3 \\ -x_1^{(k+1)} + 2x_2^{(k+1)} &= x_3^{(k)} - 5 \\ 2x_3^{(k+1)} - x_4^{(k+1)} &= x_2^{(k)} + 9 \\ -x_3^{(k+1)} + 2x_4^{(k+1)} &= x_5^{(k)} - 6 \\ x_5^{(k+1)} &= x_4^{(k)} + 2 \end{aligned} \tag{8.4}$$

Aus $D_B^{-1} = \frac{1}{3}$ $\left(\begin{array}{cc:cc:c} 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 3 \end{array}\right)$ folgt $J_B = \left(\begin{array}{cc:cc:c} 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 & 0 \\ \hdashline 0 & \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ \hdashline 0 & 0 & 0 & 1 & 0 \end{array}\right)$

Es gilt:

$$\rho(J_B) = \frac{1}{3} \sqrt[4]{18} \approx 0{,}687 < \rho(J) \approx 0{,}866 \tag{8.5}$$

Hieraus folgt, wie im folgenden Paragraphen gezeigt werden wird, daß in diesem Beispiel die Block-Jacobi-Iteration bei hinreichend vielen Iterationsschritten schneller konvergiert als das einfache Jacobi-Verfahren.

2. Eine Blockunterteilung des Systems (7.6) gemäß

$$A_B = \left(\begin{array}{cc:c} 3 & 1 & 1 \\ 1 & 2 & 0 \\ \hdashline 1 & \frac{1}{2} & 2 \end{array}\right) \tag{8.6}$$

ergibt

$$D_B^{-1} = \frac{1}{10}\left(\begin{array}{cc:c} 4 & -2 & 0 \\ -2 & 6 & 0 \\ \hdashline 0 & 0 & 5 \end{array}\right) \quad ; \quad J_B = \frac{1}{20}\left(\begin{array}{cc:c} 0 & 0 & -8 \\ 0 & 0 & 4 \\ \hdashline -10 & -5 & 0 \end{array}\right)$$

mit $\rho(J_B)=\sqrt{\frac{3}{20}} \approx 0{,}387 < \rho(J)=\frac{1}{4}\left(1+\sqrt{\frac{7}{3}}\right) \approx 0{,}632$ (8.7)

J_B ist schwach-zyklisch (vom Index 2), J dagegen nicht.

8.2. Das Block-Gauss-Seidel-Verfahren und die Block-Überrelaxation

Mit der Zerlegung $M=D_B-E_B$ und $N=F_B$ ergibt (6.2) das Block-Gauss-Seidel-Verfahren:

$$D_B x^{(k+1)} = E_B x^{(k+1)} + F_B x^{(k)} + b \tag{8.8a}$$

d.h. $$x^{(k+1)} = R_{1B} x^{(k)} + (D_B - E_B)^{-1} b; \quad R_{1B} = (I - L_B)^{-1} U_B \tag{8.8b}$$

Definition 8.2.

Die Matrix $R_{1B}=(I-L_B)^{-1}U_B$ heißt Block-Gauss-Seidel-Matrix von A.

Für die Blockunterteilung (8.6) lautet sie:

$$R_{1B} = \frac{1}{20}\begin{pmatrix} 0 & 0 & -8 \\ 0 & 0 & 4 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{mit} \quad \rho(R_{1B}) = \frac{3}{20} = 0{,}15 \tag{8.9}$$

Die Zerlegung $M=\frac{1}{\omega}(D_B-\omega E_B)$ und $N=\frac{1}{\omega}(\omega F_B+(1-\omega)D_B)$ ergibt schließlich die Block-Überrelaxation

$$x^{(k+1)}=(1-\omega)x^{(k)}+\omega(L_B x^{(k+1)}+U_B x^{(k)}+D_B^{-1}b) \qquad (8.10a)$$

d.h.
$$x^{(k+1)}=R_{\omega B}x^{(k)}+\omega(D_B-\omega E_B)^{-1}b \qquad (8.10b)$$

$$R_{\omega B}=(I-\omega L_B)^{-1}(\omega U_B+(1-\omega)I)$$

Definition 8.3.

Die Matrix $R_{\omega B}=(I-\omega L_B)^{-1}(\omega U_B+(1-\omega)I)$ heißt Block-Überrelaxationsmatrix.

Es sei nochmals ausdrücklich darauf hingewiesen, daß die Iterationsmatrizen J, R_1 und R_ω bzw. J_B, R_{1B} und $R_{\omega B}$ hier nur ausgerechnet werden, um später Aussagen über ihre Struktur (schwach-zyklisch, nicht-negativ, etc.) und ihren Spektralradius machen zu können; für die praktische Anwendung der Verfahren werden sie nicht gebraucht und nicht berechnet, wie z.B. auch die Gleichungen (7.5b), (7.8c), (7.12c) und (8.4) zeigen.

§ 9 Konvergenz

9.1. Konvergenzsätze

Alle bisher beschriebenen Verfahren haben die Form

$$x^{(k+1)}=Tx^{(k)};\quad x^{(o)}=a;\quad k=o,1,... \tag{9.1}$$

mit $\quad Tx:=Hx+c;\quad H\varepsilon\mathbb{C}(n,n);\quad c\varepsilon\mathbb{C}^n.$

Für beliebige $x_1,x_2\varepsilon\mathbb{C}^n$ und verträgliche Normen gilt:

$$\|Tx_1-Tx_2\| = \| Hx_1-Hx_2\| \leq \| H \| \cdot \| x_1-x_2 \|$$

Der Operator T ist daher kontrahierend in $\mathbb{C}^n$, wenn $\|H\|<1$ ist. Aus dem Banachschen Fixpunktsatz ergibt sich daher die folgende hinreichende Konvergenzbedingung und Fehlerabschätzung:

Satz 9.1.

Die Iteration (9.1) konvergiert für jedes $x^{(o)}\varepsilon\mathbb{C}^n$ gegen die Lösung u von x=Tx, wenn gilt:

$$P:=\|H\|<1\ ;$$

der Fehler $(x^{(k+1)}-u)$ nach (k+1) Iterationen kann dann durch:

$$\|x^{(k+1)}-u\| \leq \frac{P}{1-P}\,\|x^{(k+1)}-x^{(k)}\| \tag{9.2}$$

abgeschätzt werden.

Eine notwendige und hinreichende Konvergenzbedingung enthält der folgende Satz:

Satz 9.2.

Die Iteration (9.1) konvergiert für jedes $x^{(o)}$ genau dann, wenn für den Spektralradius der Iterationsmatrix H gilt:

$$\rho(H)<1 \tag{9.3}$$

Beweis:

Ist u Lösung von $x=Tx$ und $x^{(k)}$ k-te Approximation von u, so gilt: $u=Hu+c$ und $x^{(k+1)}=Hx^{(k)}+c$.

Hieraus folgt für den Fehler $e^{(k)}=x^{(k)}-u$ die Gleichung: $e^{(k+1)}=He^{(k)}$ und durch Induktion:

$$e^{(k)}= H^k\, e^{(o)} \tag{9.4}$$

bei beliebigem Anfangsfehler $e^{(o)}=x^{(o)}-u$. Die Behauptung folgt nun aus Satz 1.8. ◁

Bezeichnet $B:=S^{-1}HS$ die Jordansche Normalform von H, so gilt $H^k=SB^kS^{-1}$, wobei die Elemente von B^k gemäß (1.5) von den Eigenwerten von H abhängen. Man kann zeigen, daß für hinreichend großes k gilt:

$$\|H_1^k\| < \| H_2^k\|$$

wenn $\rho(H_1)<\rho(H_2)<1$ ist.

Im Hinblick auf (9.4) definiert man daher:

Definition 9.1.

Seien H_1 und H_2 Iterationsmatrizen; dann sagt man, die Iteration $x^{(k+1)}=Hx^{(k)}+c$ mit $H=H_1$ konvergiere asymptotisch schneller als die Iteration mit $H=H_2$ wenn gilt: $\rho(H_1)<\rho(H_2)<1$.

Hinweis:

Die Norm der Matrix H^k konvergiert für $\rho(H)<1$ mit $k\to\infty$ gegen Null, jedoch keineswegs monoton! Daher kann es vorkommen, daß nach k=m Iterationen $\|H_2\|^m<\| H_1\|^m$ ist obwohl $\rho(H_1)<\rho(H_2)$ gilt. Deswegen kann eine Iteration mit der Matrix H_2 nach m Iterationsschritten eine bessere Approximation von u ergeben als die Iteration mit H_1. Die Iteration mit H_1 konvergiert lediglich ("asymptotisch") für $k\to\infty$ schneller.

Beispiel:

Sei $H_1 = \begin{pmatrix} 0{,}90 & 4 \\ 0 & 0{,}90 \end{pmatrix}$; $H_2 = \begin{pmatrix} 0{,}90 & 0 \\ 0 & 0{,}95 \end{pmatrix}$

Es ist $\rho(H_1)=0{,}90$; $\rho(H_2)=0{,}95$ und

$$H_1^k = \begin{pmatrix} (0{,}90)^k & 4k(0{,}90)^{k-1} \\ 0 & (0{,}90)^k \end{pmatrix}; \quad H_2^k = \begin{pmatrix} (0{,}90)^k & 0 \\ 0 & (0{,}95)^k \end{pmatrix}$$

$\|H_1^k\|_\infty = (0{,}90)^{k-1}(4k+0{,}90)$; $\|H_2^k\|_\infty = (0{,}95)^k$

Obwohl $\rho(H_1)<\rho(H_2)$ gilt, ist $\|H_1^k\|_\infty > \|H_2^k\|_\infty$ für alle $k<115$.

9.2. Anwendungen

Wir sind nun in der Lage, die Beispiele in den Paragraphen 7 und 8 zu interpretieren:
Für das System (7.6) ergab sich $\rho(J)\approx 0{,}63$ und $\rho(R_1)\approx 0{,}29$ (vgl. (7.7) und (7.10)). In diesem Fall konvergiert daher die Gauss-Seidel-Iteration asymptotisch schneller als die Jacobi-Iteration.

Die Blockunterteilung (8.6) desselben Systems führte auf die Iterationsmatrizen J_B und R_{1B} mit $\rho(J_B)\approx 0{,}39$ und $\rho(R_{1B})=0{,}15$ (vgl. (8.7) und (8.9)). In diesem Fall ergibt die Block-Iteration eine weitere Verbesserung der asymptotischen Konvergenz.

Aus diesen Beispielen darf nicht geschlossen werden, daß das Jacobi-Verfahren grundsätzlich schlechter konvergiert als das Gauss-Seidel-Verfahren, wie die folgenden Beispiele von Collatz zeigen:

Beispiele:

1. Für $A_1=\begin{pmatrix}1 & 2 & -2\\ 1 & 1 & 1\\ 2 & 2 & 1\end{pmatrix}$ erhält man $J=\begin{pmatrix}0 & -2 & 2\\ -1 & 0 & -1\\ -2 & -2 & 0\end{pmatrix}$; $R_1=\begin{pmatrix}0 & -2 & 2\\ 0 & 2 & -3\\ 0 & 0 & 2\end{pmatrix}$

Es ist $\rho(J)=0$, d.h. die Jacobi-Iteration konvergiert
und $\rho(R_1)=2$, d.h. die Gauss-Seidel-Iteration divergiert.

2. Für $A_2=\begin{pmatrix}1 & -\frac{1}{2} & \frac{1}{2}\\ 1 & 1 & 1\\ -\frac{1}{2} & -\frac{1}{2} & 1\end{pmatrix}$ erhält man $J=\begin{pmatrix}0 & \frac{1}{2} & -\frac{1}{2}\\ -1 & 0 & -1\\ \frac{1}{2} & \frac{1}{2} & 0\end{pmatrix}$; $R_1=\begin{pmatrix}0 & \frac{1}{2} & -\frac{1}{2}\\ 0 & -\frac{1}{2} & -\frac{3}{2}\\ 0 & 0 & -\frac{1}{2}\end{pmatrix}$

mit $\rho(J)=\frac{\sqrt{5}}{2}$ (Divergenz der Jacobi-Iteration)
und $\rho(R_1)=\frac{1}{2}$ (Konvergenz der Gauss-Seidel-Iteration)

Im folgenden Paragraphen 10 werden Kriterien angegeben, die in wichtigen Spezialfällen zu entscheiden gestatten, welche der Methoden asymptotisch schneller konvergiert.

Mit Satz 9.2. läßt sich abschließend zeigen, daß $o<\omega<2$ notwendig für die Konvergenz der Überrelaxation ist:

Satz 9.3.

Sei $R_\omega=(I-\omega L)^{-1}(\omega U+(1-\omega)I)$; L,U wie in (7.1); dann ist für jedes reelle ω: $\rho(R_\omega)\geq|\omega-1|$.

Beweis:

Für die Eigenwerte $\lambda_i(i=1(1)n)$ von R_ω gilt (1.2):

$$\prod_{i=1}^{n}\lambda_i=\det R_\omega=\det(I-\omega L)^{-1}\cdot\det(\omega U+(1-\omega)I)$$

Da L und U obere bzw. untere Dreiecksmatrizen sind, folgt:

$$\prod_{i=1}^{n}\lambda_i=(1-\omega)^n\leq\rho^n(R_\omega)\quad\text{wegen}\quad|\lambda_i|\leq\rho(R_\omega)$$

Somit gilt: $\rho(R_\omega)\geq|\omega-1|$. ◁

In einigen wichtigen Fällen (vgl. Satz 10.2. und 11.3.) ist auch die Wahl $o<\omega<1$ ("Unterrelaxation") nicht empfehlenswert.

§ 10 Vergleichstheoreme und hinreichende Konvergenzbedingungen

Die Beispiele auf Seite 49 machen deutlich, daß es nicht möglich ist, ein Iterationsverfahren anzugeben, welches für alle Systemmatrizen A "das beste" ist; für spezielle Klassen von Matrizen A jedoch lassen sich die dargestellten Methoden untereinander vergleichen und hinreichende Konvergenzbedingungen angeben, die leichter zu überprüfen sind als die Forderung (9.3).

In diesem Paragraphen behandeln wir die folgenden Spezialfälle:

10.1. Die Jacobi-Matrix (oder Block-Jacobi-Matrix) von A ist nicht-negativ $J \geq 0$.

10.2. Die Matrix A erfüllt das Zeilensummen- oder Spaltensummenkriterium.

10.3. Die Matrix A ist hermitesch.

10.1. Nicht-negative Jacobi-Matrizen

Bei nicht-negativer Jacobi-Matrix gilt das folgende Vergleichskriterium von Stein-Rosenberg, welches einen Vergleich der Konvergenz des Jacobi-Verfahrens mit dem Gauss-Seidel-Verfahren ermöglicht:

Satz 10.1.

Sei $R_1=(I-L)^{-1}U$ die Gauss-Seidel-Matrix und $J=L+U$ die Jacobi-Matrix von A. J sei unzerlegbar[1] und nicht-negativ mit $L\neq\mathcal{O}$. Dann gilt eine der folgenden Aussagen:

(1) $\rho(R_1)<\rho(J)<1$
(2) $\rho(R_1)=\rho(J)=1$
(3) $\rho(R_1)>\rho(J)>1$

Beweis:

Sei $\mu=\rho(J)$, dann existiert nach Satz 5.3. genau ein $x>\mathcal{O}$ mit $Jx=\mu x, \mu>0$. Hieraus folgt:

$$Ux+(1-\mu)Lx=\mu x-\mu Lx$$

und durch Multiplikation mit $(I-L)^{-1}$ erhält man:

$$R_1x+(1-\mu)(I-L)^{-1}Lx=\mu x; \quad x>\mathcal{O}; \quad R_1\geq\mathcal{O} \tag{10.1}$$

Für $\mu=1$ ergibt sich $R_1x=\mu x$ mit $x>\mathcal{O}$. (10.2)

Nach Satz 5.3.(d) ist x der einzige, nicht-negative Vektor der (10.2) erfüllt, nach Satz 5.3.(b) muß daher $\rho(R_1)=\mu$ gelten. Damit ist (2) gezeigt.

Für $\mu\leq 1$ ist wegen $L\neq\mathcal{O}$: $R_1x\neq\mu x$, $x>\mathcal{O}$ und $R_1+(1-\mu)(I-L)^{-1}L\geq R_1$.

Nach Satz 5.6. folgt hieraus in Verbindung mit (10.1):

$$\rho(R_1)\leq\mu$$

Nun ist aber $\rho(R_1)\neq\mu$; wäre nämlich $\rho(R_1)=\mu>0$, so gäbe es ein $y\geq\mathcal{O}$, $y\neq\mathcal{O}$ mit

$$R_1y=\mu y; \quad y\neq x$$

Daher
$$R_1y+(1-\mu)(I-L)^{-1}Ly\geq\mu y$$
$$Uy+(1-\mu)Ly\geq(I-L)\mu y$$
$$Jy\geq\mu y \quad \text{mit} \quad Jy\neq\mu y$$

Dann müßte es eine Matrix $K\neq\mathcal{O}$ geben mit $\mathcal{O}\leq K\leq J$ und $Ky=\mu y$, in Widerspruch zu Satz 5.3.(e)/5.6.(c). Damit ist (1) bewiesen. Der Beweis von (3) ergibt sich durch Umkehrung der Ungleichheitszeichen. ◁

[1] Man kann zeigen, daß der Satz auch für zerlegbares J gilt.

Der folgende Satz zeigt, daß es nicht sinnvoll ist, den Relaxationsparameter $\omega<1$ zu wählen, wenn J nicht-negativ und unzerlegbar ist:

Satz 10.2.

Seien J und R_ω die Jacobi- bzw. Überrelaxationsmatrizen von A (oder die entsprechenden Block-Matrizen). Ist J nicht-negativ und unzerlegbar und gilt $\rho(J)<1$, dann

(1) konvergiert die Überrelaxation für jedes $0<\omega\leq 1$ und

(2) $\rho(R_{\omega_2})<\rho(R_{\omega_1})<1$, wenn $0<\omega_1<\omega_2\leq 1$.

Beweis:

Mit $B_\omega := \frac{1}{\omega}(I-\omega L)$ und $C_\omega := \frac{1}{\omega}(\omega U+(1-\omega)I)$ ist

$$R_\omega = B_\omega^{-1} C_\omega \quad \text{und} \quad S := B_\omega - C_\omega = I - J$$

Für $0<\omega\leq 1$ ist $C_\omega \geq \mathcal{O}$ und nach Satz 4.4. sind B_ω und S regulär mit $B_\omega^{-1} \geq \mathcal{O}$ und $S^{-1} \geq \mathcal{O}$.

Aus Satz 5.7. und $B_\omega^{-1} C_\omega = R_\omega$ folgt jetzt:

$$\rho(R_\omega) = \frac{\rho(S^{-1}C_\omega)}{1+\rho(S^{-1}C_\omega)} < 1$$

Damit ist (1) bewiesen; (2) folgt aus der noch zu beweisenden Beziehung $\rho(S^{-1}C_{\omega_1})>\rho(S^{-1}C_{\omega_2})$. (10.3)

$S^{-1}=I+J+J^2+\ldots$ ist nicht-negativ und unzerlegbar, weil J es ist. Für $0<\omega<1$ ist dann auch $S^{-1}C_\omega$ nicht-negativ und unzerlegbar, denn $C_\omega := U+(\frac{1}{\omega}-1)I=(c_{ij}) \geq \mathcal{O}$ mit $c_{ii}\neq 0$. (10.3) folgt jetzt aus Satz 5.3(e).

10.2. Das Zeilen- und Spaltensummenkriterium

Definition 10.1.

Eine Matrix $A \in \mathbb{C}(n,n)$ erfüllt das

(a) Zeilensummenkriterium, wenn gilt:

$$r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| < |a_{ii}|, \quad i=1(1)n , \tag{10.4a}$$

(b) Spaltensummenkriterium, wenn gilt:

$$r_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| < |a_{jj}|, \quad j=1(1)n , \tag{10.4b}$$

(c) schwache Zeilensummenkriterium, wenn gilt:

$$r_i \leq |a_{ii}|, \ i=1(1)n \tag{10.5a}$$

und $r_i < |a_{ii}|$ für mindestens ein $i=i_o$,

(d) schwache Spaltensummenkriterium, wenn gilt:

$$r_j \leq |a_{jj}|, \ j=1(1)n \tag{10.5b}$$

und $r_j < |a_{jj}|$ für mindestens ein $j=j_o$.

Satz 10.3.

Erfüllt $A \in \mathbb{C}(n,n)$

(a) das Zeilen- oder Spaltensummenkriterium oder

(b) das schwache Zeilen- oder Spaltensummenkriterium und ist A zudem unzerlegbar

so konvergieren das Jacobi- und Gauss-Seidel-Verfahren.

Beweis:

Im Falle (a) folgt aus dem Satz 4.3. von Gerschgorin[1] und im Falle (b) aus Satz 5.4.[1]: $\rho(|J|)<1$.
Aus Satz 5.5. folgt dann: $\rho(J)\leq\rho(|J|)<1$. Damit ist die Konvergenz des Jacobi-Verfahrens gezeigt.
Aus der oberen Dreiecksgestalt von L folgt für $R_1=(I-L)^{-1}U$:

$$|R_1|=|(I+L+\ldots+L^{n-1})U|\leq(I+|L|+\ldots+|L^{n-1}|)|U|$$

$$\leq(I-|L|)^{-1}|U|:=\hat{R}_1$$

$|J|$ bzw. $\hat{R}_1$ ist Jacobi- bzw. Gauss-Seidel-Matrix der nicht-negativen Matrix $|A|$. Aus den Sätzen 5.5., 5.6. und 10.1. folgt jetzt:

$$\rho(R_1)\leq\rho(|R_1|)\leq\rho(\hat{R}_1)\leq\rho(|J|)<1\,. \qquad \triangleleft$$

Im Falle (a) kann der Iterationsfehler des Jacobi-Verfahrens gemäß (9.2) abgeschätzt werden mit

$$P:=\|J\|_\infty=\max_{0\leq i\leq n}\left|\frac{r_i}{a_{ii}}\right|<1 \quad \text{oder} \quad P:=\|J\|_1 \qquad (10.6)$$

und der Fehler des Gauss-Seidel-Verfahrens mit $P:=\|R_1\|_\infty$ oder $P:=\|R_1\|_1$. (vergleiche Collatz | |, S. 189), je nach Wahl der Vektornorm.

[1] Für das Spaltensummenkriterium beachte man die Hinweise auf den Seiten 20 und 30.

10.3. Hermitesche Matrizen

Für hermitesche Matrizen gilt der folgende Satz von Ostrowski:

Satz 10.4.

Sei $A\in\mathbb{C}(n,n)$ hermitesch und positiv definit und sei $A=B-C-C^*$ mit positiv definitem B und $\det(B-\omega C)\neq 0$ für $0<\omega<2$. Dann konvergiert die Iteration

$$x^{(k+1)}=(B-\omega C)^{-1}(\omega C^*+(1-\omega)B)x^{(k)}+\omega(B-\omega C)^{-1}b \qquad (10.7)$$

für jedes $\omega\in(0,2)$.

Beweis:[1)]

Sei $S_\omega:=(B-\omega C)^{-1}(\omega C^*+(1-\omega)B)$; zu zeigen ist: $\rho(S_\omega)<1$. Wegen

$$\det(S_\omega-\lambda I)=\det(B-\omega C)^{-1}\det(\omega C^*+(1-\omega)B-\lambda(B-\omega C))$$

sind alle Eigenwerte von S_ω Wurzeln von

$$\det(\omega C^*+(1-\omega)B-\lambda(B-\omega C))=0$$

sei $x\neq\sigma$ Eigenvektor von S_ω und λ zugehöriger Eigenwert, dann gilt:

$$\lambda(B-\omega C)x=(\omega C^*+(1-\omega)B)x$$

Multiplikation mit $\frac{2}{\omega}$ und Addition bzw. Subtraktion von $\sigma=(A-B+C+C^*)$ in den Klammern ergibt:

$$\lambda\left[B(\tfrac{2}{\omega}-1)+A+(C^*-C)\right]x=\left[B(\tfrac{2}{\omega}-1)-A+(C^*-C)\right]x \qquad (10.8)$$

Wegen $(C^*-C)^*=-(C^*-C)$ ist $\bar{x}^T(C^*-C)x$ rein imaginär für beliebiges $x\neq\sigma$.

Nach Multiplikation von links mit $\bar{x}^T$ ergibt (10.8)

$$\lambda\left[(\tfrac{2}{\omega}-1)b+a+ic\right]=\left[(\tfrac{2}{\omega}-1)b-a+ic\right]$$

mit $a:=\bar{x}^TAx$; $b:=\bar{x}^TBx$; $ic:=\bar{x}^T(C^*-C)x$; c: reell. Für $\omega\in(0,2)$ gilt somit für alle Eigenwerte von λ von S_ω:

$$|\lambda|=\frac{|(\tfrac{2}{\omega}-1)b+ic-a|}{|(\tfrac{2}{\omega}-1)b+ic+a|}<1 \quad \text{wegen } (\tfrac{2}{\omega}-1)>0;\ a>0;\ b>0. \qquad \triangleleft$$

---

1) nach J. Albrecht

Für B=D, C=E, $\hat{C}$=F geht die Iteration (10.7.) über in das Überrelaxationsverfahren; wegen $a_{ii}>0$ folgt daher aus Satz 10.4.:

Satz 10.5.

Ist A hermitesch und positiv definit , so konvergiert die Überrelaxation für jedes $\omega\varepsilon(0,2)$.

Die nach Satz 9.3. notwendige Voraussetzung $\omega\varepsilon(0,2)$ für die Konvergenz der Überrelaxation ist somit bei positiv definiter Systemmatrix A auch hinreichend.

In gleicher Weise lassen sich aus Satz 10.4. auch hinreichende Bedingungen für die Konvergenz der Block-Überrelaxation ableiten.

Durch Multiplikation mit A^* kann das System Ax=b stets in ein solches mit hermitescher positiv definiter Systemmatrix $\hat{A}=A^*A$ (vgl. Satz 1.4.) überführt werden. Das ist jedoch nicht empfehlenswert, weil das neue System im allgemeinen erheblich schlechter konditioniert ist als das alte (vgl. Satz 3.2.).

Ein Spezialfall hermitescher, positiv definiter Matrizen sind die Stieltjes-Matrizen:

Definition 10.2.

$A=(a_{ij})\varepsilon\mathbb{R}(n,n)$ heißt Stieltjes-Matrix, wenn A symmetrisch und positiv definit ist und wenn gilt $a_{ij}\leq 0$ für alle $i\neq j$.

Für diese Matrizen läßt sich entscheiden, ob die Block-Gauss-Seidel-Iteration asymptotisch schneller konvergiert als die Gauss-Seidel-Iteration:

Satz 10.6.

Sei A eine unzerlegbare Stieltjes-Matrix. Existiert dann eine Unterteilung von A in Blöcke derart, daß wenigstens eine der Matrizen A_{ii} außerhalb der Diagonalen von Null verschiedene Elemente besitzt, so konvergiert die Block-Gauss-Seidel-Iteration asymptotisch schneller als die Gauss-Seidel-Iteration.

Der Beweis beruht auf Satz 5.7. und darauf, daß die Inverse einer unzerlegbaren Stieltjes-Matrix positiv ist.

10.4. Beispiel

Betrachten wir das System Ax=b mit der Systemmatrix (8.3):

$$A=\left(\begin{array}{ccc:cc} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ \hdashline 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ \hdashline 0 & 0 & 0 & -1 & 2 \end{array}\right); \quad \text{d.h.} \quad J=\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix} \quad (10.9)$$

Die Systemmatrix A erfüllt das schwache Zeilensummenkriterium, nach Satz 10.3. konvergieren daher sowohl das Jacobi- als auch das Gauss-Seidel-Verfahren.

Die Jacobi-Matrix J von A ist unzerlegbar und nicht-negativ, daher konvergiert — nach dem Satz 10.1. von Stein-Rosenberg — das Gauss-Seidel-Verfahren schneller als das Jacobi-Verfahren.

Die Matrix A ist hermitesch und nach dem Satz von Gerschgorin (Seite 19) in Verbindung mit Korollar 4.3. (Seite 20) positiv definit. Nach Satz 10.5. konvergiert daher die Überrelaxation für alle $\omega\in(0,2)$.

Nach Satz 10.2. erreicht man asymptotisch schnellere Konvergenz, wenn $\omega\geq 1$ gewählt wird. (Mit den Ergebnissen des folgenden Paragraphen kann man sogar zeigen, daß $\omega_o \approx 1{,}333$ eine optimale Wahl von ω ist und hierfür $\rho(R_{\omega_o}) \approx 0{,}333$ gilt).

Die Matrix A ist sogar eine (unzerlegbare) Stieltjes-Matrix. Mit der angegebenen Zerlegung in Blöcke konvergiert die Block-Gauss-Seidel-Iteration daher nach Satz 10.6. schneller als die einfache Gauss-Seidel-Iteration.

§ 11 Konsistent geordnete Systeme

Im Falle nicht-negativer Jacobi-Matrix ermöglichte der Satz von Stein-Rosenberg einen Vergleich der asymptotischen Konvergenz des Jacobi- und Gauss-Seidel-Verfahrens.

In diesem Paragraphen vergleichen wir - unter anderen Voraussetzungen - das Jacobi-Verfahren mit der Überrelaxation, indem wir eine Beziehung zwischen den Eigenwerten der Iterationsmatrizen beweisen. Damit gelingt es, die Aussage des Satzes von Stein-Rosenberg für gewisse Matrizen zu präzisieren und in speziellen Fällen den Überrelaxationsparameter ω optimal zu bestimmen. Auch kann der Einfluß gewisser Umordnungen des Gleichungssystems auf die Konvergenz der Überrelaxation untersucht werden.

11.1. Eine Beziehung zwischen den Eigenwerten von J und R_ω

In diesem Abschnitt behandeln wir die Frage: Unter welchen Voraussetzungen läßt sich ein funktionaler Zusammenhang zwischen den Eigenwerten von $J=L+U$ und $R_\omega=(I-\omega L)^{-1}(\omega U+(1-\omega)I)$ angeben?

Hierzu definieren wir zunächst:

Definition 11.1.

> Die Jacobi-Matrix (oder Block-Jacobi-Matrix) $J=L+U$ heißt konsistent (r,q)-geordnet ($r,q\in\mathbb{N}$ teilerfremd), wenn die Eigenwerte von $\alpha^r L+\alpha^{-q}U$ von $\alpha\neq 0$ unabhängig sind. Wir nennen das System $Ax=b$ konsistent geordnet, wenn die Jacobi-Matrix (oder Block-Jacobi-Matrix) von A konsistent geordnet ist.

Antwort auf die obige Frage gibt nun der folgende Satz:

Satz 11.1.

Sei $J=L+U$ Jacobi-Matrix (oder Block-Jacobi-Matrix) und R_ω Überrelaxations-Matrix (oder Block-Überrelaxationsmatrix) einer Systemmatrix A und sei J konsistent (r,q)-geordnet.

(a) Ist dann $\lambda\neq 0$ Eigenwert von R_ω und μ definiert durch:

$$(\lambda+\omega-1)^p=\omega^p\mu^p\lambda^q;\quad p:=q+r;\quad \omega\in(0,2) \tag{11.1}$$

so ist μ Eigenwert von J.

(b) Ist umgekehrt μ Eigenwert von J und erfüllt eine Zahl λ die Gleichung (11.1), so ist λ Eigenwert von R_ω. Außerdem ist mit μ stets auch $\mu\exp(\frac{k}{p}2\pi i)$, $k=1(1)p-1$, Eigenwert von J.

Beweis:

(a) Sei $\lambda\neq 0$ Eigenwert von $R_\omega\in\mathbb{C}(n,n)$; dann folgt aus $\det(R_\omega-\lambda I)=0$ wegen (7.14):

$$\det(\omega\lambda L+\omega U-(\omega+\lambda-1)I)=0 \tag{11.2}$$

Division durch $(\omega\lambda^{\frac{q}{p}})^n$ ergibt

$$\det(\alpha^r L+\alpha^{-q}U-\mu I)=0 \tag{11.3}$$

wenn $\alpha:=\lambda^{1/p}$ und $\mu:=\dfrac{\omega+\lambda-1}{\omega\lambda^{q/p}}$ (11.4)

gesetzt wird.

Nun ist J aber konsistent (r,q)-geordnet, wegen (11.3) müssen daher $\mu\exp(\frac{k}{p}2\pi i)$, $k=o(1)p-1$, für jedes λ Eigenwerte von L+U sein (dabei ergibt sich der Faktor $\exp(\frac{k}{p}2\pi i)$ aus der p-Deutigkeit von $\lambda^{1/p}$), d.h. es gilt (11.1).

(b) Ist umgekehrt μ Eigenwert von $J=L+U$, so folgt aus $\det(L+U-\mu I)=0$ und der konsistenten Ordnung von J:

$$\det(\alpha^r L+\alpha^{-q}U-\mu I)=0$$

für beliebiges $\alpha\neq 0$; insbesondere ergibt $\alpha=\lambda^{1/p}$ ($\lambda\neq 0$ beliebig):

$$\det(\lambda^{r/p}L+\lambda^{-q/p}U-\mu I)=0$$

d.h.

$$\det(\omega\lambda L+\omega U-\omega\lambda^{q/p}\mu I)=0 \qquad (11.5)$$

Für $\lambda=1$ sieht man, daß mit μ auch $\mu\exp(\frac{k}{p}2\pi i)$, $k=1(1)p-1$, Eigenwert von J ist.

Erfüllt λ die Gleichung (11.1): $\omega\lambda^{q/p}\mu=(\omega+\lambda-1)$
so ergibt (11.5):

$$\det(\omega\lambda L+\omega U-(\omega+\lambda-1)I)=0$$

Das heißt aber (vgl. (11.2)), daß λ Eigenwert von R_ω ist. $\triangleleft$

Für den Spezialfall $\omega=1$ erhält man aus (11.1):

Korollar 11.1.

Ist das System $Ax=b$ konsistent (r,q)-geordnet, so konvergieren die Iterationen von Jacobi und Gauss-Seidel (oder die Block-Iterationen) beide oder sie divergieren beide. Konvergieren sie, so gilt

$$\rho(R_1)=(\rho(J))^{(1+\frac{q}{r})}<1 \qquad (11.6)$$

d.h. die Gauss-Seidel-Iteration konvergiert erheblich schneller.

Damit haben wir eine genauere Aussage über die asymptotische Konvergenz der beiden Verfahren gewonnen, als mit dem Satz von Stein-Rosenberg, allerdings unter ganz anderen Voraussetzungen.

Beispiel:

Da die Ähnlichkeitstransformation $S(\alpha)J\,S^{-1}(\alpha)$ die Eigenwerte von J invariant und damit unabhängig von α läßt, ist das System mit der Systemmatrix (8.3) wegen

$$S(\alpha)J\,S^{-1}(\alpha)=\alpha L+\alpha^{-1}U;\quad S(\alpha)=\begin{pmatrix} 1 & \alpha & & \sigma \\ & \alpha^2 & & \\ & & \alpha^3 & \\ \sigma & & & \alpha^4 \end{pmatrix};\quad \alpha\neq 0$$

konsistent (1,1)-geordnet. Nach Korollar 11.1. folgt aus $\rho(J)\approx 0.866$ (vgl. (8.5)): $\rho(R_1)\approx(0{,}866)^2\approx 0{,}750$ und für die Block-Iteration folgt aus $\rho(J_B)=\frac{1}{3}\sqrt[4]{18}$: $\rho(R_{1B})=\frac{1}{3}\sqrt{2}\approx 0{,}471$. ◁

Es bleibt zu fragen: Wie erkennt man, ob ein System konsistent geordnet ist? Hiermit beschäftigt sich der folgende Abschnitt.

11.2. Eine Klasse konsistent (r,q)-geordneter Matrizen

Im allgemeinen ist es schwierig festzustellen, ob ein System konsistent geordnet ist oder ob es durch Umordnung konsistent geordnet werden kann. Eine Klasse von (in der Praxis oft auftretenden) Matrizen ist jedoch leicht als konsistent geordnet kenntlich, nämlich die im folgenden definierten (r,q)-Band-Matrizen. Eine Klasse von Matrizen, die sich durch simultane Umordnung ihrer Zeilen und Spalten konsistent ordnen lassen, sind, wie wir zeigen werden, die schwach-zyklischen Matrizen (vgl. Abschnitt 5.2.).

J ist offensichtlich konsistent (r,q)-geordnet, wenn es reguläre Matrizen $S(\alpha)$ gibt mit

$$S(\alpha)JS^{-1}(\alpha)=\alpha^{r}L+\alpha^{-q}U;\quad \alpha\neq 0;\quad r,q\in\mathbb{N}\ \text{teilerfremd} \tag{11.7}$$

Als Spezialfall betrachten wir

$$S(\alpha)=\begin{pmatrix} I_0 & & & \mathcal{O} \\ & \alpha I_1 & & \\ & & \alpha^2 I_2 & \\ \mathcal{O} & & \ddots & \\ & & & \alpha^{N-1} I_{N-1} \end{pmatrix};\quad I_j:\text{Einheitsmatrizen } j=o(1)N-1. \tag{11.8}$$

Hiermit ergibt sich

$$S(\alpha)J\,S^{-1}(\alpha)=\begin{pmatrix} \mathcal{O} & \alpha^{-1}U_{12} & \alpha^{-2}U_{13} & \dots & \alpha^{-N+1}U_{1N} \\ \alpha L_{21} & \mathcal{O} & \alpha^{-1}U_{23} & \dots & \alpha^{-N+2}U_{2N} \\ \alpha^2 L_{31} & \alpha L_{32} & \mathcal{O} & \dots & \alpha^{-N+3}U_{3N} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha^{N-1}L_{N1} & \alpha^{N-2}L_{N2} & \alpha^{N-3}L_{N3} & \dots & \mathcal{O} \end{pmatrix} \tag{11.9}$$

und man erkennt, daß J konsistent (r,q)-geordnet ist, wenn gilt: Alle Untermatrizen in (11.9) sind Nullmatrizen, mit Ausnahme der Matrizen in der r-ten unteren Diagonalen (das sind die Matrizen $L_{r+j,j}$, $j=1(1)N-r$) und der Matrizen in der q-ten oberen Diagonalen (d.h. $U_{j,q+j}$, $j=1(1)N-q$).

Derartige Matrizen nennen wir (r,q)-Band-Matrizen.

Beispiele:

1. Die Jacobi-Matrix (10.9) ist konsistent (1,1)-geordnet, die zugehörige Block-Jacobi-Matrix J_B (s. Seite 43) ebenfalls.

2. Die Jacobi-Matrix

$$J = \left(\begin{array}{c:ccc:c} 0 & 1 & 1 & 0 & 0 \\ \hdashline 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ \hdashline 0 & 0 & 1 & 1 & 0 \end{array}\right) \qquad (11.10)$$

ist konsistent (1,1) geordnet, wie die eingezeichnete Zerlegung in Blöcke zeigt. (Dabei braucht J aber keine Block-Iterationsmatrix zu sein!).

3. Die zum System (6.6) gehörige Block-Jacobi-Matrix ist konsistent (1,1)-geordnet, die einfache Jacobi-Matrix jedoch nicht.

Hinweis:

Um Mißverständnisse zu vermeiden, sei ausdrücklich angemerkt, daß J eine Band-Matrix bestehend aus Blöcken sein kann, ohne daß ein Block-Verfahren benutzt wurde. (In solchen Fällen sind allerdings Block-Iteration und einfache Iteration identisch).

Schwach-zyklische Matrizen vom Index p in der Normalform (5.4.) sind konsistent (1,p-1)-geordnet wie man mit (11.9) erkennt; andererseits sind alle (r,q)-Band-Matrizen schwach-zyklisch vom Index (q+r), wie die zugehörigen Graphen in Verbindung mit Satz 5.2. zeigen. Daher können alle schwach-zyklischen Matrizen vom Index p konsistent (1,p-1)-geordnet werden (vgl. Definition 5.4.), d.h. es gilt der folgende Satz:

Satz 11.2.

Ist J schwach-zyklisch vom Index p, so gibt es eine Permutationsmatrix P, so daß PJP^T konsistent (1,p-1)-geordnet ist.

Es kann also mit Hilfe von Satz 5.2. entschieden werden, ob sich eine Matrix auf konsistente Gestalt umordnen läßt.

Beispiele:

1. Die Matrix

$$J_1 = \begin{pmatrix} 0 & 1 & 0 & -4 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 5 \\ -2 & 0 & 1 & 0 \end{pmatrix} \tag{11.11}$$

ist nicht konsistent geordnet. Sie ist aber schwach-zyklisch vom Index 2 und kann daher nach Satz 11.2. konsistent (1,1)-geordnet werden. Tatsächlich ergibt die Ähnlichkeitstransformation mit einer Permutationsmatrix P_1 (d.h. eine simultane Umordnung der Zeilen und Spalten von J) eine konsistent (1,1)geordnete Matrix $\hat{J}_1$:

$$P_1 J_1 P_1^{-1} := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot J_1 \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & -4 \\ 0 & 0 & 1 & 5 \\ 1 & 2 & 0 & 0 \\ -2 & 1 & 0 & 0 \end{pmatrix} := \hat{J}_1$$

2. Die Matrix

$$J_2 = \frac{1}{4} \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{11.12}$$

ist eine (2,3)-Band-Matrix und daher konsistent (2,3)-geordnet. Die simultane Permutation ihrer Zeilen und Spalten gemäß $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 2 & 4 \end{pmatrix}$ ergibt eine konsistent (1,4)-geordnete Matrix.

3. Die zur Matrix

$$A=\begin{pmatrix} 20 & -4 & -4 & -1 & 0 & 0 \\ -4 & 20 & -1 & -4 & 0 & 0 \\ -4 & -1 & 20 & -4 & -4 & -1 \\ -1 & -4 & -4 & 20 & -1 & -4 \\ 0 & 0 & -4 & -1 & 20 & -4 \\ 0 & 0 & -1 & -4 & -4 & 20 \end{pmatrix} \tag{11.13}$$

gehörige Jacobi-Matrix J ist nicht schwach-zyklisch. Zerlegt man jedoch A derart in Blöcke, daß die Diagonale aus (2,2)-Matrizen besteht, so ist die zugehörige Block-Jacobi-Matrix J_B eine (1,1)-Band-Matrix und somit konsistent (1,1)-geordnet.

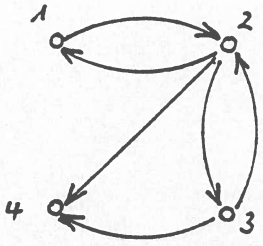
4. Die zum System (6.6) gehörige Jacobi-Matrix J ist nicht konsistent geordnet. Das System kann aber (nach Satz 11.2.) so umgeordnet werden, daß die Jacobi-Matrix konsistent (1,1)-geordnet ist. Zum Beispiel ergibt die Umordnung der Zeilen und Spalten gemäß der Permutation $P: \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 3 & 5 & 7 & 9 & 2 & 4 & 6 & 8 \end{pmatrix}$ die (1,1)-Band-Matrix:

$$\hat{J}:=PJP^T=\frac{1}{4}\begin{pmatrix} & & & & & 1 & 1 & 0 & 0 \\ & & & & & 1 & 0 & 1 & 0 \\ & & 0 & & & 1 & 1 & 1 & 1 \\ & & & & & 0 & 1 & 0 & 1 \\ & & & & & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & & & & \\ 1 & 0 & 1 & 1 & 0 & & 0 & & \\ 0 & 1 & 1 & 0 & 1 & & & & \\ 0 & 0 & 1 & 1 & 1 & & & & \end{pmatrix} \tag{11.14}$$

Die zu (6.6) gehörige Block-Jacobi-Matrix J_B ist bereits eine (1,1)-Band-Matrix und daher konsistent (1,1)-geordnet.

5. Der Graph G der Matrix

$$J=\frac{1}{4}\begin{pmatrix}0&1&0&0\\1&0&1&1\\0&1&0&1\\0&0&0&0\end{pmatrix}\quad;\qquad G:$$



ist weder zyklisch, noch kann er zyklisch gemacht werden (vgl. Beispiel 2, Seite 27). Nach Satz 5.2. ist J daher nicht schwach-zyklisch. Dennoch ist J konsistent (1,1)-geordnet, denn $\det(\alpha L+\alpha^{-1}U-\mu I)=\mu^4-\frac{1}{8}\mu^2$ ist von α unabhängig.

Beispiel 5 macht bereits deutlich, daß die schwach-zyklischen Matrizen nicht die einzigen sind, die durch simultane Umordnung ihrer Zeilen und Spalten konsistent geordnet werden können. Hierzu noch das folgende Beispiel von Broyden[1]:

6. Sei

$$J=\begin{pmatrix}0&-b&b&-b\\-b&0&0&-b\\b&0&0&-b\\-b&-b&-b&0\end{pmatrix};\quad S(\alpha)=\begin{pmatrix}1&0&0&0\\0&\frac{1+\alpha}{2}&\frac{1-\alpha}{2}&0\\0&\frac{1-\alpha}{2}&\frac{1+\alpha}{2}&0\\0&0&0&\alpha\end{pmatrix}.$$

Dann gilt: $S(\alpha)J\,S^{-1}(\alpha)=\alpha L+\alpha^{-1}U$; daher ist J konsistent (1,1)-geordnet. J ist aber nicht schwach-zyklisch.

---

[1] C.G. Broyden: "Some Generalizations of the Theory of Successive Over-Relaxation", Num. Math. 6, 1964, p. 269-284.

11.3. Über den Einfluß von Umordnungen auf die Konvergenz der Iterationsverfahren

Sei $P \in \mathbb{R}(n,n)$ eine beliebige Permutationsmatrix; dann beeinflußt die Umordnung PAP^T einer Systemmatrix $A \in \mathbb{C}(n,n)$ nicht den Spektralradius ihrer Jacobi-Matrix. Hat nämlich A die Jacobi-Matrix $J_1 = L+U = -D^{-1}A+I$, so hat PAP^T die Jacobi-Matrix

$$J_2 = -PD^{-1}P^TPAP^T+I = PJ_1P^{-1} \qquad \text{(wegen } P^T=P^{-1}\text{)};$$

nach Satz 1.1. ist daher $\rho(J_1)=\rho(J_2)$.

Dagegen wird der Spektralradius der Überrelaxationsmatrix R_ω durch Umordnung[1] des Systems verändert. Es ist daher möglich, die Konvergenz der Überrelaxation durch Umordnung des Systems zu verbessern.

Beispiel:

$$\text{Für } A=\begin{pmatrix} 2 & -1 & 0 \\ 0 & 2 & -1 \\ -1 & 0 & 2 \end{pmatrix} \text{ gilt: } J=\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 \end{pmatrix}; \; R_1=\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{4} & 0 \end{pmatrix}$$

mit $\rho(J)=\frac{1}{2}$; $\rho(R_1)=\sqrt{\frac{1}{8}}$.

J ist eine (2,1)-Band-Matrix und schwach-zyklisch von Index 3. Nach Satz 11.2. kann daher das System Ax=b auf konsistent (1,2)-geordnete Form umgeordnet werden.

---

[1] Unter einer "Umordnung" verstehen wir immer eine simultane Umordnung der Zeilen und Spalten von A gemäß PAP^T; sie entspricht einer Permutation der Gleichungen des Systems Ax=b bei entsprechender Permutation der Unbekannten x_k, k=1(1)n.

Mit $P=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}$ ergibt sich ein anderes System $\hat{A}y=\hat{b}$ mit $y=Px$, $\hat{b}=Pb$ und

$$\hat{A}=PAP^T=\begin{pmatrix}2&0&-1\\-1&2&0\\0&-1&2\end{pmatrix};\quad \hat{J}=\begin{pmatrix}0&0&\frac{1}{2}\\\frac{1}{2}&0&0\\0&\frac{1}{2}&0\end{pmatrix};\quad R_1=\begin{pmatrix}0&0&\frac{1}{2}\\0&0&\frac{1}{4}\\0&0&\frac{1}{8}\end{pmatrix}$$

Jetzt ist: $\rho(\hat{J})=\frac{1}{2}$; $\rho(\hat{R}_1)=\frac{1}{8}$.

$\hat{J}$ ist konsistent (1,2)-geordnet. Dieses Beispiel bestätigt Korollar 11.1. und zeigt, daß Umordnungen sich sehr günstig auf die asymptotische Konvergenz des Gauss-Seidel-Verfahrens (und allgemeiner auch der Überrelaxation) auswirken können.
Mit Korollar 11.1. läßt sich auch entscheiden, welche konsistente Ordnung für die asymptotische Konvergenz des Gauss-Seidel-Verfahrens am günstigsten ist, falls mehrere konsistente Ordnungen eines Systems existieren (wie z.B. bei (r,q)-Bandmatrizen mit r+q>2).

Entsprechende Ergebnisse für den allgemeineren Fall des Überrelaxationsverfahrens sind im folgenden Satz zusammengefaßt:

Satz 11.3.

Die Jacobi-Matrix (oder Block-Jacobi-Matrix) J sei konsistent (r,q)-geordnet; es sei $\rho(J)<1$ und die Eigenwerte von J^p, $p:=r+q$, seien reell und nicht-negativ. Dann gilt

(a) für $r\neq 1$: $\min\limits_{\omega}\rho(R_\omega)=\rho(R_1)=(\rho(J))^{(1+\frac{q}{r})}$

(b) für $r=1$: $\min\limits_{\omega}\rho(R_\omega)=\rho(R_{\omega_o})=(\omega_o-1)q<1$, wobei ω_o die einzige Wurzel der Gleichung

$$(\rho(J)\omega_o)^p=p^pq^{-q}(\omega_o-1)\ ;\quad p:=q+r \qquad (11.15)$$

im Intervall $\left[1,\ \frac{q+1}{q}\right]$ ist.

Bemerkungen:

1. Mit Satz 11.1. lassen sich Aussagen über die Konvergenz der Überrelaxation auch unter schwächeren Voraussetzungen machen.
2. Unter den Voraussetzungen des Satzes 11.3. kann der Überrelaxationsparameter optimal bestimmt werden; wie im Falle nicht-negativer Jacobi-Matrizen ist $\omega_o \geq 1$.

Der Beweis von Satz 11.3. beruht auf einer Analyse der Beziehung (11.1). Wir beschränken uns hier auf den Beweis des wichtigsten Falles r=q=1, zu dem insbesondere die symmetrischen, schwach-zyklischen Jacobi-Matrizen vom Index 2 gehören.

Sei also J konsistent (1,1)-geordnet und seien die Eigenwerte von J^2 nicht-negativ; dann hat J (nach Satz 1.5) nur reelle Eigenwerte und nach Satz 11.1(b) ist mit μ auch $-\mu$ Eigenwert. Deshalb ist $\mu=\rho(J)$ Eigenwert von J; aus (11.1) folgt daher:

$$|\lambda(\mu,\omega)| = \left|\tfrac{1}{2}\,\omega^2\mu^2-(\omega-1) \pm \tfrac{1}{2}\,\omega\mu\sqrt{\omega^2\mu^2-4(\omega-1)}\right| \qquad (11.16)$$

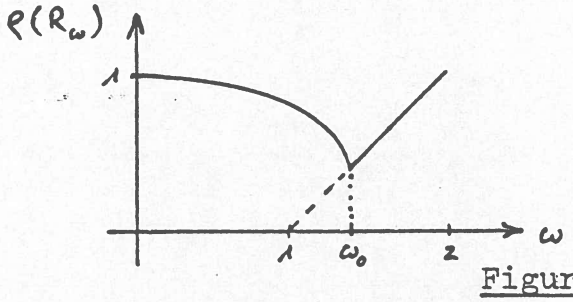
ist Eigenwert von R_ω.

Für $\omega^2\mu^2 \leq 4(\omega-1)$ gilt: $|\lambda|^2=(\omega-1)^2$, d.h. $|\lambda|=|\omega-1|$ für alle Eigenwerte λ von R_ω.

Für $\omega^2\mu^2>4(\omega-1)$ wächst $|\lambda(\mu,\omega)|$ mit μ, daher gilt:

$$\rho(R_\omega)=|\lambda(\mu,\omega)| \qquad \text{für} \quad \mu=\rho(J).$$

$\rho(R_\omega)$ nimmt für wachsendes $\omega\in(0,\omega_o)$ – $\omega_o^2\mu^2=4(\omega_o-1)$ – monoton ab und hat daher den in Figur 11.1. gezeichneten Verlauf.



Figur 11.1.

Insbesondere nimmt $\rho(R_\omega)$ sein Minimum an für $\omega=\omega_o$ mit

$$\omega_o^2\mu^2=4(\omega_o-1), \qquad \text{d.h.} \quad \omega_o=\frac{2}{1+\sqrt{1-\mu^2}} \qquad (11.17)$$

Damit ist (11.15) für r=q=1 bewiesen. ◁

Da $\mu=\rho(J)$ im allgemeinen nur näherungsweise bekannt ist, kann ω_o in der Praxis nur annähernd bestimmt werden. Figur 11.1. macht deutlich, daß es günstiger ist, ω etwas zu groß als etwas zu klein zu wählen.

Beispiel:

Das System

$$\begin{aligned} 2x_1 - x_2 &= 1 \\ 2x_2 - x_3 &= 0 \\ -x_1 + 2x_3 &= 1 \end{aligned} \qquad (11.18)$$

hat die Jacobi-Matrix

$$J = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 \end{pmatrix} \qquad (11.19)$$

J ist eine (2,1)-Band-Matrix, das System (11.18) ist daher konsistent (2,1)-geordnet.
Mit Satz 5.4. findet man $\rho(J)=\frac{1}{2}$ und aus Korollar 11.1. folgt ohne Rechnung: $\rho(R_1)=\sqrt{\frac{1}{8}} \approx 0{,}354$

Da J schwach-zyklisch vom Index J ist, kann das System (11.18) nach Satz 11.2. konsistent (1,2) geordnet werden. Permutiert man seine Zeilen und Spalten gemäß $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, so erhält man das umgeordnete System:

$$\begin{aligned} 2x_1 - x_2 &= 1 \\ -x_1 + 2x_3 &= 1 \\ -x_3 + 2x_2 &= 0 \end{aligned} \qquad (11.20)$$

mit der konsistent (1,2)-geordneten Jacobi-Matrix

$$\hat{J} = \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \quad ; \quad \rho(\hat{J}) = \frac{1}{2} \qquad (11.21)$$

Nach Korollar 11.1. gilt für die zugehörige Gauss-Seidel-Matrix $\hat{R}_1$:

$$\rho(R_1)=\frac{1}{8}=0{,}125$$

Die Gauss-Seidel-Iteration konvergiert also asymptotisch erheblich schneller beim umgeordneten System (11.20) als beim ursprünglichen System (11.18).
Nach Satz 11.1(b) lauten die Eigenwerte von $\hat{J}$:

$$\mu_k = \frac{1}{2} \exp\left(\frac{k}{3} 2\pi i\right), \quad k=0,1,2.$$

J^3 hat daher nur positive Eigenwerte und aus (11.15) folgt für den optimalen Überrelaxationsfaktor $\omega_o \in [1;1,5]$

$$\left(\frac{1}{2}\omega_o\right)^3 = \frac{27}{4}(\omega_o - 1)$$

Hieraus erhält man: $\omega_o \approx 1,0196$, Satz 11.3.(b) ergibt daher ohne weitere Rechnung: $\rho(R_{\omega_o}) \approx 0,0392$.

§ 12 Bemerkungen zur Diskretisierung der partiellen Differentialgleichung (6.3.)

Zum Abschluß sollen die in diesem Kapitel gewonnenen Ergebnisse auf das im § 6 gestellte partielle Differentialgleichungsproblem angewandt werden:

Die Diskretisierung von (6.3) ergab für n=3 das System (6.6), dessen Jacobi- bzw. Block-Jacobi-Matrix gegeben ist durch

$$J=\frac{1}{4}\begin{pmatrix} 0&1&0&1&0&0&0&0&0\\ 1&0&1&0&1&0&0&0&0\\ 0&1&0&0&0&1&0&0&0\\ 1&0&0&0&1&0&1&0&0\\ 0&1&0&1&0&1&0&1&0\\ 0&0&1&0&1&0&0&0&1\\ 0&0&0&1&0&0&0&1&0\\ 0&0&0&0&1&0&1&0&1\\ 0&0&0&0&0&1&0&1&0 \end{pmatrix} \quad \text{bzw.} \quad J_B=\frac{1}{56}\begin{pmatrix} 0&0&0&15&4&1&0&0&0\\ 0&0&0&4&16&4&0&0&0\\ 0&0&0&1&4&15&0&0&0\\ 15&4&1&0&0&0&15&4&1\\ 4&16&4&0&0&0&4&16&4\\ 1&4&15&0&0&0&1&4&15\\ 0&0&0&15&4&1&0&0&0\\ 0&0&0&4&16&4&0&0&0\\ 0&0&0&1&4&15&0&0&0 \end{pmatrix} \tag{11.22}$$

J und J_B sind nicht-negativ und unzerlegbar; aus Satz 5.4. folgt daher:

$$\frac{1}{2}<\rho(J)<1 \quad \text{und} \quad \frac{20}{56}<\rho(J_B)<\frac{48}{56} \tag{11.23}$$

Das Jacobi- und das Block-Jacobi-Verfahren konvergieren daher (nach Satz 9.2.).

Die Gauss-Seidel- und die Block-Gauss-Seidel-Iteration konvergieren asymptotisch schneller als die entsprechenden Jacobi-Iterationen (nach dem Satz von Stein-Rosenberg).

Die Systemmatrix A (s. (6.6), S. 34) ist symmetrisch, hat daher nur reelle Eigenwerte λ_j, $j=1(1)9$ (nach Satz 1.2.). Nach dem Satz von Gerschgorin gilt $\lambda_j \geq 0$. Wäre $\lambda=0$ Eigenwert, so müßten nach Korollar 4.3. alle Kreisscheiben durch $\lambda=0$ gehen; also ist $\lambda_j>0$. A ist daher positiv definit (nach Satz 1.3.) und wegen $a_{ij}\leq 0$ für $i\neq j$ eine Stieltjes Matrix. Aus Satz 10.6. folgt daher:

Die Block-Gauss-Seidel-Iteration konvergiert asymptotisch schneller als die einfache Gauss-Seidel-Iteration.

J_B ist eine (1,1)-Band-Matrix und somit konsistent (1,1)-geordnet. J_B^2 hat nur reelle, nicht-negative Eigenwerte, weil J_B symmetrisch ist (Sätze 1.2. und 1.5.). Nach Satz 11.3. gilt daher $\rho(R_{1B})=(\rho(J_B))^2$

$$\rho(R_{\omega_o B})=(\omega_o-1) \quad \text{mit} \quad \omega_o=\frac{2}{1+\sqrt{1-(\rho(J_B))^2}} \tag{11.24}$$

Aus (11.23) folgt:

$$\frac{2}{1+\sqrt{1-(\frac{20}{50})^2}}\approx 1{,}034 < \omega_o < \frac{2}{1+\sqrt{1-(\frac{43}{56})^2}}\approx 1{,}320$$

Satz 11.3. ergibt jetzt:

Die Block-Überrelaxation konvergiert für $1{,}034<\omega_o<1{,}320$ asymptotisch schneller als die Block-Gauss-Seidel-Iteration.

Eine Berechnung des Spektralradius von J_B ergibt: $\rho(J_B)\approx 0{,}5469$; nach (11.24) ist $\omega_o\approx 1{,}176$.

Die Block-Überrelaxation mit $\omega_o\approx 1{,}176$ konvergiert schneller als jedes andere der hier betrachteten Verfahren.

Satz 11.3. ergibt: $\rho(R_{\omega_o B})\approx 0{,}176$.

Zum Vergleich: $\rho(J) \approx 0{,}707$.

Der Fehler nach k Iterationen hängt gemäß (9.4) von H^k ab. Um eine Vorstellung von der unterschiedlichen Effizienz der Verfahren zu vermitteln wurde zum Vergleich $\|J^k\|_1$, $\|J_B^k\|_1$, $\|R_1^k\|_1$, $\|R_{1B}^k\|_1$ und $\|R_{\omega_o B}^k\|_1$ für k=10 ausgerechnet:

$$\|J^{10}\|_1 \approx 0{,}47\cdot 10^{-1}\ ; \qquad \|J_B^{10}\|_1 \approx 0{,}29\cdot 10^{-2}$$

$$\|R_1^{10}\|_1 \approx 0{,}30\cdot 10^{-2}\ ; \qquad \|R_{1B}^{10}\|_1 \approx 0{,}24\cdot 10^{-4}$$

$$\|R_{\omega_o B}^{10}\|_1 \approx 0{,}33\cdot 10^{-6}$$

Es zeigt sich deutlich die Überlegenheit der Überrelaxation mit optimalem Parameter ω_o.

Literatur:

/1/ Faddejew, D.K. and Faddejewa, M.N.: Num. Meth. d. linearen Algebra, Oldenbourg, 1973.

/2/ Gantmacher, F.R.: The Theory of Matrices, Vol. I, II, Chelsea Publ. Comp., 1964.

/3/ Householder, A.S.: The Theory of Matrices in Numerical Analysis, Blaisdell Publ. Comp., 1964.

/4/ Varga, R.S.: Matrix Iterative Analysis, Prentice-Hall, 1962.

/5/ Wilkinson, J.H.: Rundungsfehler, Heidelberger Taschenbücher Bd. 44, Springer 1969.

/6/ Wilkinson, J.H. und Reinsch, C.: Handbook for automatic computation, Vol. II, Lin. Algebra, Springer, 1971.

/7/ Young, D.M.: Iterative Solution of Large Linear Systems, Academic Press, 1971.

Index

Ähnlichkeitstrans-
formation 3,24,61

Albrecht 26

Bandmatrix 63

Block,-unterteilung 26,14,48,56
,-iteration 41,42
,-Jacobi-Matrix 42
,-Jacobi-Verfahren 42
,-Gauss-Seidel-Matrix 44
,-Gauss-Seidel-Verf. 42
,-matrizen 41
,-Überrelaxation 45
,-Überrelaxationsmatrix 45

Broyden 66

Charakteristisches Polynom 2,3,6
,Nullstellen 2,18

Collatz 48,54

Defekt 14,39

Differentialgleichung
,partielle 33,72

Eigenvektor 1,29,31

Eigenwert 1,3,7,17,29,31,59
-Schranken 17

Einzelschrittverf. 35,37

Elementarteiler 6
,linearer 6-8

Fehler 13,16
,relativer 14,16
,Iterations- 54

Frobenius 29
,-Matrix 2,18

Gauss-Seidel-Iterationsverf. 36,37
,-Matrix 38,59
,Konvergenz 49,51,53,56,60,68

Gerichteter Weg 22
,Länge 22

Gerschgorin, Satz von 19,54

Gesamtschrittverfahren 35,36

Graph, gerichteter 22ff
,einer Matrix 22
,stark zusammenhängender 22,25

Graph,zyklisch vom Index p 23,26
,primitiver 23

Iterationsmatrix 33,47

Jacobi-Iterationsverf. 36
,Matrix 37,59,67
--,nicht-negative 50
--,konsistent(r,q)-geordnete 58,62
-,Konvergenz 51,53,60

Jordanzelle 5

konditioniert, schlecht 15,56

Konditionszahl 15
,spektrale 15
,von Todd 15
,von Turing 15

Konvergenz 46
,-bedingungen 46,49,51,53
,asymptotisch schnellere 47,56,60,72

Matrix
,adjungierte 1
,Betrags- 1
,hermitesche 3,4,15,55,56
,inverse 1
,konjugiert komplexe 1
,nicht-negative 29,31
,positiv definite 3,4,55,56
,schwach-zyklisch vom
Index p 26,28,63
,Spur einer 2
,symmetrische 3,56
,transponierte 1
,unzerlegbare 20,25,29,56
,zerlegbare 25

Matrixnorm 9,10
,verträgliche 9
,der Vektornorm zugeordnete 9,17
,konsistente 9,17

Matrizen,ähnliche 3

Normalform
,Jordansche 5,6,47
,schwach zyklischer Matrizen 26,63

Operator 9
,-norm 9,16,17

Ostrowski, Satz von 55

Permutationsmatrix 24,26,64,67

Perron 29

Quadratzahlnorm 11

Reduktion eines Systems 25
,zyklische 28

Relaxationsparameter 52
,optimaler 69,73

Spaltensummenkriterium 53
,schwaches 53

Spaltensummennorm 10

Spektralnorm 11

Spektralradius 11,17,33,46,67
,Schranken für den 19
,Einschließungssatz für den 30

Stein-Rosenberg, Satz von 50,58,72

Stieltjes-Matrix 56,72

Todd 13,15

Tridiagonalmatrix 42

Turing 15

Überrelaxationsmatrix 40,59,67

Überrelaxationsverfahren 36,39
,Konvergenz 49,52,56,68

Umordnung 67

Unterrelaxation 50

Verfahren,Gesamtschritt- 35,36
,Einzelschritt- 35,37
von Jacobi 36
von Gauss-Seidel 36,37
,Überrelaxations- 36,39

Vieta, Satz von 2

Wilson 13

Zeilensummenkriterium 53
,schwaches 53

Zeilensummennorm 10

Zerlegung einer Matrix 33