

Proseminar Numerik

Leitung: Prof. W. Hofmann

Vortrag von Markus Stürzekarn zum Thema:

Jacobi- und Gauß-Seidel-Verfahren, Jacobi-Relaxationsverfahren

Gegeben sei ein lineares Gleichungssystem in der Form $\mathbf{Ax} = \mathbf{b}$ mit $\mathbf{A} \in \mathbb{C}^{n \times n}$ und $\mathbf{b} \in \mathbb{C}^n$. Die Grundidee der vorgestellten iterativen Methoden zur Lösung des Systems ist eine Zerlegung der Matrix \mathbf{A} nach folgendem Schema:

$$\mathbf{A} = \mathbf{B} + (\mathbf{A} - \mathbf{B}), \quad \mathbf{B} \in \mathbb{C}^{n \times n}$$

Somit ist das LGS äquivalent zu:

$$(\mathbf{B} + (\mathbf{A} - \mathbf{B}))\mathbf{x} = \mathbf{b} \quad \iff \quad \mathbf{B}\mathbf{x} = (\mathbf{B} - \mathbf{A})\mathbf{x} + \mathbf{b}$$

Ist die Matrix \mathbf{B} regulär, so erhält man:

$$\mathbf{x} = \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{x} + \mathbf{B}^{-1}\mathbf{b}$$

und die Lösung des LGS ist Fixpunkt der vektorwertigen Funktion

$$f : \mathbb{C}^n \longrightarrow \mathbb{C}^n, \quad \mathbf{x} \longmapsto f(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{x} + \mathbf{B}^{-1}\mathbf{b}.$$

Nach dem *Verfahren der sukzessiven Iteration* iteriert man also gemäß

$$\mathbf{x}^{m+1} = \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{x}^m + \mathbf{B}^{-1}\mathbf{b}, \quad (1)$$

um das LGS $\mathbf{Ax} = \mathbf{b}$ zu lösen. Aufgrund der Zerlegung der Matrix \mathbf{A} wird dieser Ansatz auch als *Splitting-Methode* bezeichnet.

1. Jacobi-Verfahren

Das Jacobi-Verfahren mit regulärer Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ setzt zur iterativen Lösung eines LGS voraus, dass für die Diagonalelemente von \mathbf{A} gilt:

$$a_{ii} \neq 0, \quad i = 1, \dots, n$$

Gemäß (1) setzen wir $\mathbf{B} := \mathbf{D} := \text{diag}(\mathbf{A})$ mit der regulären Diagonalmatrix \mathbf{D} , die einfach zu invertieren ist. Das führt zum

$$\text{Jacobi-Verfahren: } \mathbf{x}^{m+1} = \underbrace{\mathbf{D}^{-1}(\mathbf{D} - \mathbf{A})}_{=: \mathbf{M}_J} \mathbf{x}^m + \underbrace{\mathbf{D}^{-1} \mathbf{b}}_{=: \mathbf{N}_J} := \mathbf{G}_J(\mathbf{x}^m) \quad (2)$$

Die Matrix \mathbf{M}_J wird *Iterationsmatrix* des Jacobi-Verfahrens genannt. Insgesamt haben wir für die Iterierten x^m folgende Komponentenschreibweise:

$$x_i^{m+1} = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^m \right) \text{ für } i = 1, \dots, n \text{ und } m = 0, 1, 2, \dots \quad (3)$$

Die neue Iterierte x^{m+1} wird also ausschließlich mit der alten Iterierten x^m berechnet, weshalb diese Methode auch als *Gesamtschrittverfahren* bezeichnet wird.

\mathbf{G}_J ist eine Abbildung des \mathbb{C}^n in sich. Die für die Existenz einer Lösung notwendige Selbstabbildungseigenschaft ist also gemäß dem Fixpunktsatz von Banach erfüllt. Damit das Jacobi-Verfahren gegen die Lösung des LGS konvergiert, muss zusätzlich die Kontraktionseigenschaft gelten. Für einander zugeordnete oder zumindestens passende Matrix- und Vektornormen gilt:

$$\begin{aligned} \|\mathbf{G}_J(\mathbf{x}) - \mathbf{G}_J(\mathbf{y})\| &= \|\mathbf{M}_J \mathbf{x} + \mathbf{N}_J \mathbf{b} - (\mathbf{M}_J \mathbf{y} + \mathbf{N}_J \mathbf{b})\| = \|\mathbf{M}_J \mathbf{x} - \mathbf{M}_J \mathbf{y}\| \\ &= \|\mathbf{D}^{-1}(\mathbf{D} - \mathbf{A})(\mathbf{x} - \mathbf{y})\| \leq \|\mathbf{D}^{-1}(\mathbf{D} - \mathbf{A})\| \|\mathbf{x} - \mathbf{y}\| \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \end{aligned}$$

Findet man also Matrixnormen, so dass $\|\mathbf{D}^{-1}(\mathbf{D} - \mathbf{A})\| = \|\mathbf{M}_J\| < 1$ gilt, so ist die Kontraktionseigenschaft erfüllt und das Jacobi-Verfahren konvergiert für jeden beliebigen Startwert gegen die Lösung.

In der Numerikvorlesung wurde das folgende notwendige und gleichzeitig hinreichende, aber schwer nachzuprüfende Kriterium für die Existenz einer Matrixnorm mit $\|\mathbf{M}\| < 1$ für $\mathbf{M} \in \mathbb{C}^{n \times n}$ bewiesen:

Satz 1.1: Es gibt genau dann eine einer Vektornorm zugeordnete Matrixnorm mit $\|\mathbf{M}\| < 1$ für $\mathbf{M} \in \mathbb{C}^{n \times n}$, wenn gilt

$$\rho(\mathbf{M}) < 1,$$

wobei

$$\rho(\mathbf{M}) := \max_i \{ |\lambda_i| : \lambda_i = \text{Eigenwert von } \mathbf{M} \}$$

der betragsmäßig größte Eigenwert von \mathbf{M} ist.

Zum Beweis verweise ich auf das Numerik-Skript von Prof. Hofmann, S.156/157.

$\rho(\mathbf{M})$ wird auch der *Spektralradius* von \mathbf{M} genannt. Je kleiner dieser ist, desto schneller konvergiert ein Verfahren, wenn \mathbf{M} dessen Iterationsmatrix ist.

Im folgenden Satz sind für drei bekannte Matrixnormen einfacher nachzuprüfende Kriterien zusammengefasst, die hinreichend, jedoch nicht notwendig für die Konvergenz sind:

Satz 1.2: Erfüllt die reguläre Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ mit $a_{ii} \neq 0$, $i = 1, \dots, n$ das *starke Zeilensummenkriterium*

$$q_\infty := \max_{i=1, \dots, n} \sum_{\substack{j=1 \\ j \neq i}}^n \frac{|a_{ij}|}{|a_{ii}|} < 1$$

oder das *starke Spaltensummenkriterium*

$$q_1 := \max_{j=1, \dots, n} \sum_{\substack{i=1 \\ i \neq j}}^n \frac{|a_{ij}|}{|a_{ii}|} < 1$$

oder das *Quadratsummenkriterium*

$$q_2 := \sum_{\substack{i, j=1 \\ i \neq j}}^n \left(\frac{|a_{ij}|}{|a_{ii}|} \right)^2 < 1$$

dann konvergiert das Jacobi-Verfahren bei beliebigem Startvektor $\mathbf{x}^0 \in \mathbb{C}^n$ und für beliebige rechte Seite $\mathbf{b} \in \mathbb{C}^n$ gegen $\mathbf{A}^{-1}\mathbf{b}$.

In vielen Anwendungen erfüllt die Matrix \mathbf{A} nicht das starke, sondern nur das schwache Zeilensummenkriterium, was wie folgt definiert ist:

Definition 1.3: Eine Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ erfüllt das *schwache Zeilensummenkriterium*, wenn gilt:

a) $\sum_{\substack{j=1 \\ j \neq i}}^n \frac{|a_{ij}|}{|a_{ii}|} \leq 1$, $i = 1, \dots, n$

b) Es existiert mindestens eine Zeile $i \in \{1, \dots, n\}$ mit

$$\sum_{\substack{j=1 \\ j \neq i}}^n \frac{|a_{ij}|}{|a_{ii}|} < 1.$$

Erfüllt die Matrix \mathbf{A} jedoch gewisse Voraussetzungen, so kann auch für diesen Fall die Konvergenz nachgewiesen werden. Dazu definieren wir die *Zerlegbarkeit einer Matrix*:

Definition 1.4: Eine Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ heißt *zerlegbar*, wenn es nichtleere Teilmengen N_1 und N_2 von $N := \{1, 2, \dots, n\}$ gibt mit:

a) $N_1 \cap N_2 = \emptyset$

b) $N_1 \cup N_2 = N$

c) für jedes $r \in N_1$ und jedes $s \in N_2$ gilt $a_{rs} = 0$

\mathbf{A} heißt *unzerlegbar*, wenn \mathbf{A} nicht zerlegbar ist.

Die Eigenschaft der Unzerlegbarkeit einer Matrix ist gemäß der Definition also ausschließlich von der Besetzungsstruktur dieser abhängig. Die Frage der Konvergenz des Jacobi-Verfahrens im Falle der Erfüllung des schwachen Zeilensummenkriteriums wird mit dieser Definition nun beantwortet durch folgenden Satz:

Satz 1.5: Ist eine Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ unzerlegbar und erfüllt sie das schwache Zeilensummenkriterium, so konvergiert das Jacobi-Verfahren bei beliebigem Startvektor $\mathbf{x}^0 \in \mathbb{C}^n$ und für beliebige rechte Seite $\mathbf{b} \in \mathbb{C}^n$ gegen $\mathbf{A}^{-1}\mathbf{b}$.

Für den Beweis des Satzes benötigen wir u.a. den *Satz von Gerschgorin*, der in der Numerikvorlesung bewiesen wurde und der hier nochmals zitiert wird:

Satz von Gerschgorin: Sei die Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, dann gilt:

a) Alle Eigenwerte liegen in der Vereinigung der Kreise

$$Z_i := \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \right\}$$

b) Alle Eigenwerte liegen in der Vereinigung der Kreise

$$S_j := \left\{ z \in \mathbb{C} : |z - a_{jj}| \leq \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}| \right\}$$

c) Jede Zusammenhangskomponente von $\cup Z_i$ oder $\cup S_j$ enthält genau so viele Eigenwerte, wie Kreise an der Komponente beteiligt sind.

Der Beweis ist im Numerik-Skript von Prof. Hofmann auf S.175/176 zu finden. Es folgt der

Beweis von Satz 1.5: Die Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ erfülle das schwache Zeilensummenkriterium und sei unzerlegbar.

Da alle Diagonalelemente der Iterationsmatrix \mathbf{M}_J des Jacobi-Verfahrens gleich 0 sind und somit nach Gerschgorin alle Z_i und S_j den Mittelpunkt 0 haben, gilt für \mathbf{M}_J :

$$Z_i = \left\{ z \in \mathbb{C} : |z - 0| \leq \sum_{\substack{j=1 \\ j \neq i}}^n \frac{|a_{ij}|}{|a_{ii}|} \leq 1 \right\}, \quad i = 1, \dots, n$$

Außerdem müssen alle Eigenwerte auch in $\{Z_i\} \cap \{S_j\}$ liegen, also folgt, dass alle Eigenwerte von \mathbf{M}_J betragsmäßig auf keinen Fall größer als 1 sein können und somit einen Betrag kleiner gleich 1 aufweisen müssen. Somit bleibt zu zeigen, dass kein Eigenwert von \mathbf{M}_J den Betrag 1 hat, wodurch nach Satz 1.1 die Konvergenz folgen würde, da dann der betragsmäßig größte Eigenwert von \mathbf{M}_J echt kleiner als 1 ist.

Annahme: Es gibt einen Eigenwert $\lambda \in \mathbb{C}$ von \mathbf{M}_J mit $|\lambda| = 1$. Sei v Eigenvektor zum Eigenwert λ mit $\|v\|_\infty = 1$. Wir bilden die beiden Indexmengen

$$N_1 := \{ i \mid i \in N, |x_i| = 1 \} \neq \emptyset \quad \text{und} \quad N_2 := N \setminus N_1,$$

wobei N wie in Definition 1.4 gewählt ist. Wegen $\mathbf{M}_{\mathbf{J}}v = \lambda v$ gilt für alle $i \in N$:

$$-\sum_{\substack{j=1 \\ j \neq i}}^n \frac{a_{ij}}{a_{ii}} v_j = \lambda v_i$$

und somit folgt wegen $|\lambda| = 1$ und $\|v\|_{\infty} = 1$:

$$|v_i| = |\lambda| |v_i| \leq \sum_{\substack{j=1 \\ j \neq i}}^n \frac{|a_{ij}|}{|a_{ii}|} |v_j| \leq \sum_{\substack{j=1 \\ j \neq i}}^n \frac{|a_{ij}|}{|a_{ii}|} \leq 1 \quad (4)$$

Ist $i \in N_1$, so gilt $|v_i| = 1$, für diese i steht also in (4) überall Gleichheit. Da nach Voraussetzung das schwache Zeilensummenkriterium gilt, muss es nach Definition 1.3 ein $k \in N$ geben, so dass

$$\sum_{\substack{j=1 \\ j \neq k}}^n \frac{|a_{kj}|}{|a_{kk}|} < 1$$

gilt, woraus folgt, dass für dieses k auch $|v_k| < 1$ gilt, wie man aus (4) abliest. Daraus folgt $N_2 \neq \emptyset$, denn $k \in N_2$. Weiterhin ist \mathbf{A} per Voraussetzung unzerlegbar, was nun bedeutet:

$$\exists r \in N_1, s \in N_2 \quad \text{mit} \quad a_{rs} \neq 0,$$

denn sonst wäre \mathbf{A} zerlegbar. Für dieses $r \in N_1$ und $s \in N_2$ gilt nun aber wegen $s \in N_2$ und $a_{rs} \neq 0$:

$$\frac{|a_{rs}|}{|a_{rr}|} |v_s| < \frac{|a_{rs}|}{|a_{rr}|},$$

wodurch ein Summand aus (4) echt vergrößert wird und somit dort keine Gleichheit stehen kann. Dies ist aber ein Widerspruch dazu, dass für alle $i \in N_1$, also insbesondere auch für $i = r$, in (4) Gleichheit gilt.

Unter der Voraussetzung, dass \mathbf{A} unzerlegbar ist und dem schwachen Zeilensummenkriterium genügt, kann es also keinen Eigenwert $\lambda \in \mathbb{C}$ von $\mathbf{M}_{\mathbf{J}}$ mit $|\lambda| = 1$ geben.

□

2. Gauß-Seidel-Verfahren

Ausgehend vom Jacobi-Verfahren betrachten wir dessen Komponentenschreibweise (3) und gehen wie folgt vor: Zuerst wird die erste Komponente x_1^{m+1} der neuen Iterierten mittels

$$x_1^{m+1} = \frac{1}{a_{11}} \left(b_1 - \sum_{j=2}^n a_{1j} x_j^m \right)$$

berechnet. Zur Berechnung der zweiten Komponenten x_2^{m+1} setzen wir x_1^{m+1} anstelle von x_1^m in (3) ein und erhalten:

$$x_2^{m+1} = \frac{1}{a_{22}} \left(b_2 - a_{21} x_1^{m+1} - \sum_{j=3}^n a_{2j} x_j^m \right)$$

Die dritte Komponente x_3^{m+1} ergibt sich, indem wir x_1^{m+1} sowie x_2^{m+1} anstelle von x_1^m und x_2^m in (3) einsetzen. Wir erhalten:

$$x_3^{m+1} = \frac{1}{a_{33}} \left(b_3 - \sum_{j=1}^2 a_{3j} x_j^{m+1} - \sum_{j=4}^n a_{3j} x_j^m \right)$$

Wir führen dies fort und somit ergibt sich für die i -te Komponente des *Gauß-Seidel-Verfahrens* folgende Schreibweise:

$$x_i^{m+1} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{m+1} - \sum_{j=i+1}^n a_{ij} x_j^m \right) \text{ für } i = 1, \dots, n \text{ und } m = 0, 1, 2, \dots \quad (5)$$

Dieses Verfahren zieht zur Berechnung der i -ten Komponente der neuen Iterierten x^{m+1} also neben den Komponenten der alten Iterierten x^m die $(i-1)$ bereits bekannten Komponenten der neuen Iterierten x^{m+1} heran. Deshalb wird auch die Bezeichnung *Einzel-schrittverfahren* verwendet. Zur Ermittlung der Matrixschreibweise formen wir (5) um zu:

$$\begin{aligned} \frac{1}{a_{ii}} \left(\sum_{j=1}^{i-1} a_{ij} x_j^{m+1} \right) + x_i^{m+1} &= \frac{1}{a_{ii}} \left(b_i - \sum_{j=i+1}^n a_{ij} x_j^m \right) \\ \Leftrightarrow \frac{1}{a_{ii}} \left(\sum_{j=1}^i a_{ij} x_j^{m+1} \right) &= \frac{1}{a_{ii}} \left(b_i - \sum_{j=i+1}^n a_{ij} x_j^m \right) \\ \Leftrightarrow \sum_{j=1}^i a_{ij} x_j^{m+1} &= - \sum_{j=i+1}^n a_{ij} x_j^m + b_i \end{aligned} \quad (6)$$

Wir zerlegen die Ausgangsmatrix \mathbf{A} des LGS gemäß $\mathbf{A} = \mathbf{L} + \mathbf{R} + \mathbf{D}$, wobei

$$\mathbf{L} = (\ell_{ij})_{i,j=1,\dots,n} \quad \text{mit} \quad \ell_{ij} = \begin{cases} a_{ij}, & i > j \\ 0, & \text{sonst} \end{cases}$$

eine strikte linke untere Dreiecksmatrix,

$$\mathbf{R} = (r_{ij})_{i,j=1,\dots,n} \quad \text{mit} \quad r_{ij} = \begin{cases} a_{ij}, & i < j \\ 0, & \text{sonst} \end{cases}$$

eine strikte rechte obere Dreiecksmatrix ist und \mathbf{D} wie vorher als $\mathbf{D} = \text{diag}(\mathbf{A})$ gewählt wurde. Somit ist (6) die Komponentenschreibweise folgender Gleichung:

$$(\mathbf{D} + \mathbf{L})\mathbf{x}^{m+1} = -\mathbf{R}\mathbf{x}^m + \mathbf{b}$$

Wir formen um und erhalten das

Gauß-Seidel-Verfahren: $\mathbf{x}^{m+1} = \underbrace{-(\mathbf{D} + \mathbf{L})^{-1}\mathbf{R}}_{=: \mathbf{M}_G} \mathbf{x}^m + \underbrace{(\mathbf{D} + \mathbf{L})^{-1}\mathbf{b}}_{=: \mathbf{N}_G} := \mathbf{G}_G(\mathbf{x}^m) \quad (7)$

Analog zum Jacobi-Verfahren sind Kontraktionseigenschaft und somit auch Konvergenz dann gesichert, wenn man Matrixnormen findet, so dass $\|(\mathbf{D} + \mathbf{L})^{-1}\mathbf{R}\| = \|\mathbf{M}_G\| < 1$ gilt. Da jedoch die Matrix $(\mathbf{D} + \mathbf{L})$ nicht so einfach wie zuvor die Matrix \mathbf{D} zu invertieren ist, wäre das Nachprüfen der zu Satz 1.1 analogen Kriterien auch deutlich aufwendiger. Dass jedoch das starke Zeilensummenkriterium auch für die Konvergenz des Gauß-Seidel-Verfahren hinreichend ist, zeigt folgender Satz, der in der Numerikvorlesung bewiesen wurde.

Satz 2.1: Erfüllt die reguläre Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ mit $a_{ii} \neq 0$, $i = 1, \dots, n$ das *starke Zeilensummenkriterium*

$$q_\infty := \max_{i=1, \dots, n} \sum_{\substack{j=1 \\ j \neq i}}^n \frac{|a_{ij}|}{|a_{ii}|} < 1$$

so konvergiert das Gauß-Seidel-Verfahren bei beliebigem Startvektor $\mathbf{x}^0 \in \mathbb{C}^n$ und für beliebige rechte Seite $\mathbf{b} \in \mathbb{C}^n$ gegen $\mathbf{A}^{-1}\mathbf{b}$.

Der Beweis ist ebenfalls im Numerik-Skript von Prof. Hofmann zu finden, siehe S.153/154. Im Allgemeinen konvergiert das Gauß-Seidel-Verfahren öfter und schneller als das Jacobi-Verfahren, da es bereits errechnete Komponenten der nächsten Iterierten, also bessere Näherungen, noch im gleichen Schritt weiterverwendet.

3. Relaxationsverfahren

Unser Ziel ist es in diesem Abschnitt, eine Verbesserung der Konvergenzgeschwindigkeit des Jacobi-Verfahrens zu erreichen. Ausgehend vom allgemeinen Fall (1) formen wir

$$\mathbf{x}^{m+1} = \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{x}^m + \mathbf{B}^{-1}\mathbf{b},$$

durch Ausmultiplizieren und Ausklammern von \mathbf{B}^{-1} um zu:

$$\mathbf{x}^{m+1} = \mathbf{x}^m + \underbrace{\mathbf{B}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}^m)}_{=: \mathbf{r}_m}$$

Man kann also x^{m+1} als Korrektur von x^m unter dem Vektor \mathbf{r}_m verstehen. Die Grundidee der Relaxationsverfahren ist nun eine Gewichtung des Vektors \mathbf{r}_m mit einem $\omega \in \mathbb{R}^+$:

$$\mathbf{x}^{m+1} = \mathbf{x}^m + \omega \mathbf{B}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}^m) \quad (8)$$

Dadurch wird die Lösung des LGS nicht beeinflusst, denn der gesuchte Fixpunkt $\hat{\mathbf{x}}$ von (1) ist ebenfalls Fixpunkt von (8) und somit Lösung des LGS, was man durch Einsetzen nachprüft. Wir formen jetzt (8) um und erhalten:

$$\mathbf{x}^{m+1} = \underbrace{(\mathbf{I} - \omega \mathbf{B}^{-1} \mathbf{A})}_{=: \mathbf{M}(\omega)} \mathbf{x}^m + \underbrace{\omega \mathbf{B}^{-1} \mathbf{b}}_{=: \mathbf{N}(\omega)} \quad (9)$$

Wir suchen nun die bestmögliche Gewichtung von \mathbf{r}_m , das heißt das optimale ω , so dass der Spektralradius der Iterationsmatrix $\mathbf{M}(\omega)$ minimal wird. Wir bestimmen ω also gemäß:

$$\omega = \min_{\alpha \in \mathbb{R}^+} \rho(\mathbf{M}(\alpha))$$

Der Gewichtungsfaktor ω wird auch Relaxationsparameter genannt und (9) wird für $\omega < 1$ als Unterrelaxationsverfahren und für $\omega > 1$ als Überrelaxationsverfahren bezeichnet.

3.1 Jacobi-Relaxationsverfahren

Ersetzen wir in (8) wie im Jacobi-Verfahren die Matrix \mathbf{B} durch \mathbf{D} , so erhalten wir

$$\mathbf{x}^{m+1} = \mathbf{x}^m + \omega \mathbf{D}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}^m), \quad (10)$$

woraus analog zu (9) folgendes Verfahren hervorgeht:

Jacobi-Relaxationsverfahren:
$$\mathbf{x}^{m+1} = \underbrace{(\mathbf{I} - \omega \mathbf{D}^{-1} \mathbf{A})}_{=: \mathbf{M}_J(\omega)} \mathbf{x}^m + \underbrace{\omega \mathbf{D}^{-1} \mathbf{b}}_{=: \mathbf{N}_J(\omega)} \quad (11)$$

Die Komponentenschreibweise des Verfahrens wird durch (10) geliefert:

$$\begin{aligned} x_i^{m+1} &= x_i^m + \frac{\omega}{a_{ii}} \left(b_i - \sum_{j=1}^n a_{ij} x_j^m \right) \\ &= (1 - \omega) x_i^m + \frac{\omega}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^m \right) \quad \text{für } i = 1, \dots, n \text{ und } m = 0, 1, 2, \dots \end{aligned} \quad (12)$$

Der folgende Satz beantwortet die Frage nach dem optimalen Relaxationsparameter ω :

Satz 3.1.1: Die Iterationsmatrix des Jacobi-Verfahrens \mathbf{M}_J habe nur reelle Eigenwerte $\lambda_1 \leq \dots \leq \lambda_n$ mit den zugehörigen linear unabhängigen Eigenvektoren $\mathbf{u}_1, \dots, \mathbf{u}_n$ und es gelte $\rho(\mathbf{M}_J) < 1$. Dann besitzt die Iterationsmatrix $\mathbf{M}_J(\omega)$ des Jacobi-Relaxationsverfahrens die Eigenwerte

$$\mu_i = 1 - \omega + \omega \lambda_i \quad \text{für } i = 1, \dots, n,$$

und es gilt

$$\omega_{opt} =: \omega^* = \min_{\alpha \in \mathbb{R}^+} \rho(\mathbf{M}_J(\alpha)) = \frac{2}{2 - \lambda_1 - \lambda_n}$$

Beweis: Zunächst wird die erste Aussage bewiesen. Mit (11) und

$$\mathbf{D}^{-1}(\mathbf{D} - \mathbf{A})\mathbf{u}_i = -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{R})\mathbf{u}_i = \mathbf{M}_J\mathbf{u}_i = \lambda_i\mathbf{u}_i \quad \text{für } i = 1, \dots, n$$

erhalten wir für die Eigenwerte der Iterationsmatrix $\mathbf{M}_J(\omega)$ des Jacobi-Relaxationsverfahrens:

$$\begin{aligned} \mathbf{M}_J(\omega)\mathbf{u}_i &= (\mathbf{I} - \omega \mathbf{D}^{-1} \mathbf{A})\mathbf{u}_i \\ &= \left(\mathbf{I} - \omega(\mathbf{D}^{-1}(\mathbf{L} + \mathbf{R}) + \mathbf{D}^{-1}\mathbf{D}) \right) \mathbf{u}_i \\ &= \left(\mathbf{I} - \omega \mathbf{I} - \omega \mathbf{D}^{-1}(\mathbf{L} + \mathbf{R}) \right) \mathbf{u}_i \\ &= \left((1 - \omega)\mathbf{I} - \omega \mathbf{D}^{-1}(\mathbf{L} + \mathbf{R}) \right) \mathbf{u}_i \\ &= (1 - \omega + \omega \lambda_i)\mathbf{u}_i \quad \text{für } i = 1, \dots, n \end{aligned}$$

Aufgrund der linearen Unabhängigkeit der Eigenvektoren $\mathbf{u}_1, \dots, \mathbf{u}_n$ existieren keine weiteren Eigenwerte.

Es folgt der Beweis der zweiten Aussage: Gemäß den Voraussetzungen des Satzes und wegen $\omega \in \mathbb{R}^+$ gilt für die Eigenwerte $\mu_i(\omega) = 1 - \omega + \omega\lambda_i$, $i = 1, \dots, n$, der Iterationsmatrix $\mathbf{M}_{\mathbf{J}}(\omega)$ des Jacobi-Relaxationsverfahrens:

$$\mu_1(\omega) \leq \dots \leq \mu_n(\omega) \quad (13)$$

Um den betragsmäßig größten Eigenwert von $\mathbf{M}_{\mathbf{J}}(\omega)$ möglichst klein zu halten, liegt die Idee Nahe, ω so bestimmen, dass kleinster und größter Eigenwert $\mu_1(\omega)$ und $\mu_n(\omega)$ von $\mathbf{M}_{\mathbf{J}}(\omega)$ symmetrisch zum Nullpunkt verteilt sind, sie also den selben Betrag aufweisen. Folgende Bedingung wird also an den optimalen Relaxationsparameter ω^* gestellt:

$$\mu_n(\omega^*) = -\mu_1(\omega^*) \quad (14)$$

Durch (13) und (14) werden folgende Aussagen impliziert:

$$\mu_1(\omega^*) < 0 < \mu_n(\omega^*) \quad (15)$$

$$|\mu_1(\omega^*)| = |\mu_n(\omega^*)| = \rho(\mathbf{M}_{\mathbf{J}}(\omega^*)) \quad (16)$$

Aus (14) erhalten wir:

$$\mu_n(\omega^*) = 1 - \omega^* + \omega^*\lambda_n = -(1 - \omega^* + \omega^*\lambda_1) = -\mu_1(\omega^*)$$

Wir lösen nach ω^* auf und erhalten wie im Satz:

$$\omega^* = \frac{2}{2 - \lambda_1 - \lambda_n} > 0$$

Jetzt müssen wir noch die Optimalität von ω^* nachweisen.

Sei $\omega > \omega^*$. Somit folgt mit (15) für $\mu_1(\omega)$ und $\mu_1(\omega^*)$:

$$\mu_1(\omega) = 1 - \omega \underbrace{(1 - \lambda_1)}_{>0} < 1 - \omega^*(1 - \lambda_1) = \mu_1(\omega^*) < 0$$

Hieraus folgt mit (16):

$$\rho(\mathbf{M}_{\mathbf{J}}(\omega)) \geq |\mu_1(\omega)| > |\mu_1(\omega^*)| = \rho(\mathbf{M}_{\mathbf{J}}(\omega^*))$$

Sei nun $\omega < \omega^*$. Analog zu vorher folgt mit (15) für $\mu_n(\omega)$ und $\mu_n(\omega^*)$:

$$\mu_n(\omega) = 1 - \omega \underbrace{(1 - \lambda_n)}_{>0} > 1 - \omega^*(1 - \lambda_n) = \mu_n(\omega^*) > 0$$

Hieraus folgt analog zu vorher mit (16):

$$\rho(\mathbf{M}_{\mathbf{J}}(\omega)) \geq |\mu_n(\omega)| > |\mu_n(\omega^*)| = \rho(\mathbf{M}_{\mathbf{J}}(\omega^*))$$

Damit ist der Satz bewiesen, denn für jedes beliebige $\omega \neq \omega^*$ gilt

$$\rho(\mathbf{M}_{\mathbf{J}}(\omega)) > \rho(\mathbf{M}_{\mathbf{J}}(\omega^*)),$$

was gleichbedeutend ist mit:

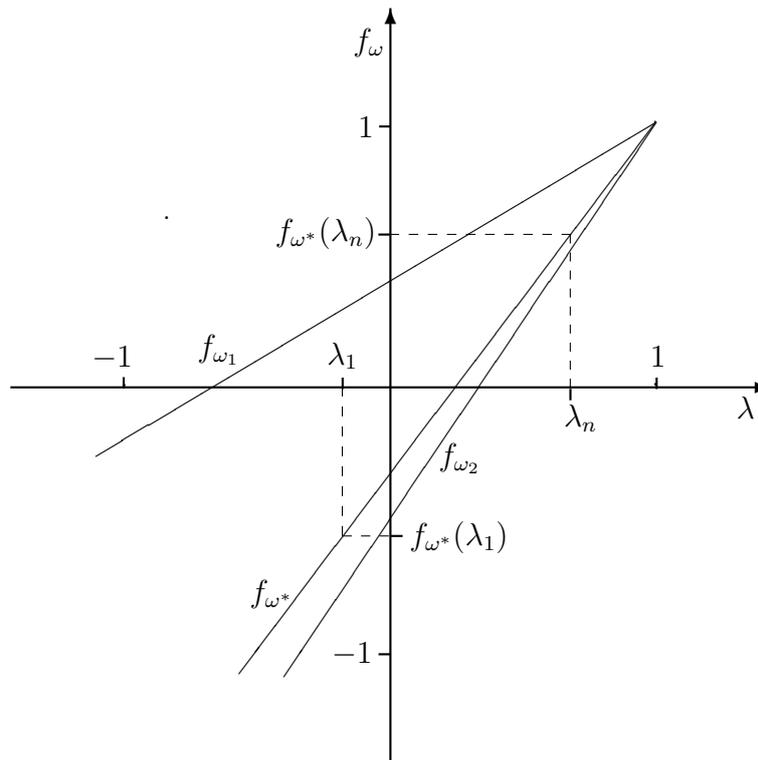
$$\omega^* = \omega_{opt} = \min_{\omega \in \mathbb{R}^+} \rho(\mathbf{M}_{\mathbf{J}}(\omega))$$

□

Wir verdeutlichen noch den Verlauf der Relaxationsfunktion

$$f_\omega : \mathbb{R} \longrightarrow \mathbb{R} \quad , \quad \lambda \longmapsto f_\omega(\lambda) = 1 - \omega + \omega\lambda,$$

für nichtoptimale Parameter ω_1 und ω_2 sowie für den optimalen Parameter ω^* bezüglich gegebener Eigenwerte λ_1 und $\lambda_n \in (-1, 1)$. Für die Relaxationsfunktion gilt per Definition $f_\omega(\lambda_i) = \mu_i(\omega)$, $i = 1, \dots, n$, wodurch man die Wahl von ω^* durch die Bedingung (14) in der Zeichnung ablesen kann.



4. Abschließender Vergleich

Zum Schluss wollen wir die drei vorgestellten Verfahren anhand eines einfachen Beispiels vergleichen. Wir setzen:

$$\mathbf{A} := \begin{pmatrix} 4 & 1 & 2 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{pmatrix} \quad \text{sowie} \quad \mathbf{b} := \begin{pmatrix} 12 \\ 13 \\ 9 \end{pmatrix}$$

\mathbf{A} erfüllt das schwache Zeilensummenkriterium und die Lösung des LGS $\mathbf{Ax} = \mathbf{b}$ ist offensichtlich $\mathbf{x} = (1, 2, 3)^T$.

1) Jacobi-Verfahren:

Die Eigenwerte der Iterationsmatrix \mathbf{M}_J lauten $\lambda_1 \approx -0.921$, $\lambda_2 \approx 0.285$ und $\lambda_3 \approx 0.636$ und somit gilt für den Spektralradius $\rho(\mathbf{M}_J) = |\lambda_1| \approx 0.921$.

2) Gauß-Seidel-Verfahren:

Die Eigenwerte von \mathbf{M}_G lauten hier $\kappa_1 = 0$, $\kappa_2 \approx 0.25$ und $\kappa_3 \approx 0.333$ und für den Spektralradius gilt $\rho(\mathbf{M}_G) = |\kappa_3| \approx 0.333$.

3) Jacobi-Relaxationsverfahren:

Wir erhalten den optimalen Parameter ω^* gemäß Satz 3.1.1 aus den Eigenwerten des Jacobi-Verfahrens:

$$\omega^* = \frac{2}{2 - \lambda_1 - \lambda_3} \approx \frac{2}{2 + 0.921 - 0.636} \approx 0.875$$

Somit gilt für die Eigenwerte von $\mathbf{M}_J(\omega^*)$ nach Satz 3.1.1:

$$\mu_i = 1 - \omega^* + \omega^* \lambda_i, \quad i = 1, 2, 3 \quad \implies \quad \mu_1 = -\mu_3 \approx -0.681, \quad \mu_2 \approx 0.374$$

Hieraus erhält man den Spektralradius von $\mathbf{M}_J(\omega^*)$ als $\rho(\mathbf{M}_J(\omega^*)) = |\mu_1| = |\mu_3| \approx 0.681$.

Es ist also zu erwarten, dass das Gauß-Seidel-Verfahren in diesem Fall am schnellsten konvergiert, da hier der Spektralradius der Iterationsmatrix mit Abstand am kleinsten ist. Auffällig ist, dass die Wahl von ω^* bereits zu einer deutlichen Verkleinerung des Spektralradius geführt hat und somit für das Jacobi-Relaxationsverfahren eine deutlich schnellere Konvergenz als im Jacobi-Verfahren erwartet werden kann, für das wir eher mit langsamer Konvergenz rechnen.

Der Iterationsverlauf der drei Verfahren wird in der folgenden Tabelle für den Startvektor $x^0 = (10, 10, 10)^T$ aufgezeigt:

	Jacobi-Verfahren			Gauß-Seidel-Verf.			Jacobi-Relaxationsv.		
m	x_1^m	x_2^m	x_3^m	x_1^m	x_2^m	x_3^m	x_1^m	x_2^m	x_3^m
0	10	10	10	10	10	10	10	10	10
5	-3.4722	-3.5463	-2.6528	0.9785	1.8258	3.0979	0.0876	0.8770	1.5945
10	4.0059	5.7443	6.6438	1.0000	1.9991	3.0005	1.1603	2.2050	3.1545
15	-0.9840	-0.4692	0.5793	1.0000	2.0000	3.0000	0.9800	1.9760	2.9697
20	2.3131	3.6345	4.6005	1.0000	2.0000	3.0000	1.0035	2.0044	3.0033
50	1.1101	2.1371	3.1342	1.0000	2.0000	3.0000	1.0000	2.0000	3.0000
100	1.0018	2.0022	3.0022	1.0000	2.0000	3.0000	1.0000	2.0000	3.0000
150	1.0000	2.0000	3.0000	1.0000	2.0000	3.0000	1.0000	2.0000	3.0000

Wie zuvor vermutet, konvergiert das Jacobi-Verfahren äußerst langsam, das Jacobi-Relaxationsverfahren schon deutlich schneller und das Gauß-Seidel-Verfahren schließlich am schnellsten. Um eine Genauigkeit von $\|\mathbf{A}^{-1}\mathbf{b} - \mathbf{x}^m\|_\infty < 10^{-6}$ zu erreichen, braucht das Jacobi-Verfahren hier 194 Iterationen, das Gauß-Seidel-Verfahren 17 Iterationen und das Jacobi-Relaxationsverfahren 42 Iterationen.

Literatur

[1] Andreas Meister, Numerik linearer Gleichungssysteme (1999, Vieweg):
Abschnitt 4.1.1 - 4.1.3.1

[2] Wolf Hofmann, Numerische Mathematik (SS 2004 / WS 2005, Universität Hamburg):
§ 11 Iterative Lösung linearer und nichtlinearer Gleichungen

[3] H. Werner, R. Schaback, Numerische Mathematik (2001, Universität Göttingen):
S.150-153, www.num.math.uni-goettingen.de/schaback/teaching/numath.ps