

# **Das SOR-Verfahren (Successive Over-Relaxation)**

Maria Woydich

29. Juni 2010

## **Inhaltsverzeichnis**

<b>1 Herleitung/Rechtfertigung</b>	<b>2</b>
<b>2 Konvergenz</b>	<b>3</b>
<b>3 Analyse im 2-zyklischen Fall</b>	<b>5</b>
<b>4 Beispiel</b>	<b>10</b>
<b>5 Ordnungsverbesserung und praktische Handhabung</b>	<b>11</b>
<b>6 Das SSOR-Verfahren</b>	<b>13</b>
<b>7 Quellenverzeichnis</b>	<b>14</b>

## 1 Herleitung/Rechtfertigung

Wie in den vorherigen Vorträgen betrachten wir das System

$$Ax = b \quad A \in \mathbb{R}^{n \times n} \quad x, b \in \mathbb{R}^n \quad (1)$$

Hierbei soll  $A$  vollen Rang besitzen, d.h.  $Rg(A) = n$ .

Wir betrachten das System für ein beliebiges  $n$  und formen wieder nach  $x_i$  um. Somit ergibt sich:

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

Dies ist eine Fixpunktformulierung des linearen Gleichungssystems (1). Hierfür lautet die Fixpunktiteration:

$$x_i^{(k+1)} = (b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}x_j^{(k)}) / a_{ii} \quad (2)$$

Somit erhalten wir das Gesamtschritt- oder Jacobi-Verfahren. Allerdings greift man hierbei nur auf die vorhergehende Iterierte  $x^{(k)}$  zu. Bei der Berechnung der  $x_i^{(k+1)}$  sind  $x_1^{(k+1)}, \dots, x_{i-1}^{(k+1)}$  bekannt. Daher splitten wir die Summe aus Formel (2) auf und erhalten zur Berechnung der  $x_i^{(k+1)}$  das Einzelschritt- bzw. Gauß-Seidel-Verfahren:

$$x_i^{(k+1)} = (b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)}) / a_{ii} \quad (3)$$

Das SOR-Verfahren ist eine Verfeinerung dieses Einzelschrittverfahrens. Hierbei wird die Iterierte  $x_i^{(k+1)}$  mit Hilfe eines Relaxationsparameters  $\omega$  berechnet:

$$x_i^{(k+1)} = x_i^{(k)} + \omega(\bar{x}_i^{(k+1)} - x_i^{(k)}) \quad (4)$$

wobei  $\bar{x}_i^{(k+1)}$  die sich aus dem Einzelschrittverfahren ergebende Näherung ist, die sich mit Formel (3) berechnet. Für  $\omega < 1$  spricht man von Unterrelaxation und für  $\omega > 1$  von Überrelaxation. Im Fall  $\omega = 1$  gilt:

$$x_i^{(k+1)} = x_i^{(k)} + \bar{x}_i^{(k+1)} - x_i^{(k)} = \bar{x}_i^{(k+1)}$$

Das heißt, man erhält für  $\omega = 1$  das Gauß-Seidel-Verfahren.

Nun formen wir Gleichung (4) um, damit wir das SOR-Verfahren in der Form

$$x^{(k+1)} = Mx^{(k)} + c$$

darstellen können, wobei  $M$  die Iterationsmatrix bezeichnet.

$$\begin{aligned} x_i^{(k+1)} &= (1 - \omega)x_i^{(k)} + \frac{\omega}{a_{ii}}(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)}) \\ \Leftrightarrow a_{ii}x_i^{(k+1)} &= (1 - \omega)a_{ii}x_i^{(k)} + \omega(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)}) \end{aligned} \quad (5)$$

Nun wollen wir das System umformen, um  $x^{(k+1)}$  mit Hilfe der Matrix  $A$  zu berechnen. Dazu zerlegen wir  $A$  in  $A = D - L - R$ , wobei:

$$D = \text{diag}(a_{11} \dots a_{nn})$$

$$L = - \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ a_{21} & 0 & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{pmatrix} \quad R = - \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & 0 & a_{n-1,n} \\ 0 & \cdots & \cdots & 0 \end{pmatrix}$$

Somit lässt sich Gleichung (5) wie folgt umformen:

$$\begin{aligned} Dx^{(k+1)} &= (1 - \omega)Dx^{(k)} + \omega(b + Lx^{(k+1)} + Rx^{(k)}) \\ \Leftrightarrow (D - \omega L)x^{(k+1)} &= (1 - \omega)Dx^{(k)} + \omega b + \omega Rx^{(k)} \end{aligned}$$

Unter der Annahme, dass  $a_{ii} \neq 0$ , hat die Matrix  $D - \omega L$  vollen Rang und ist daher invertierbar. Damit lässt sich das SOR-Verfahren wie folgt schreiben:

$$x^{(k+1)} = (D - \omega L)^{-1}[(1 - \omega)D + \omega R]x^{(k)} + \omega (D - \omega L)^{-1}b \quad (6)$$

Das heißt, für die Iterationsmatrix  $M$  gilt:

$$M_{SOR}(\omega) = (D - \omega L)^{-1}[(1 - \omega)D + \omega R]$$

## 2 Konvergenz

Beim SOR-Verfahren betrachtet man nur den Bereich  $0 < \omega < 2$  beschränken, da das Verfahren ansonsten nicht konvergiert. Dies zeigt der folgende Hilfssatz:

**Lemma 1** (Satz von Kahan)

Für eine beliebige Matrix  $A \in \mathbb{R}^{n \times n}$  mit  $a_{ii} \neq 0 \quad \forall i = 1, \dots, n$ , das heißt mit regulärer Matrix  $D$  gilt

$$\text{spr}(M_{SOR}(\omega)) \geq |\omega - 1| \quad \forall \omega \in \mathbb{R}$$

*Beweis:*

$$\begin{aligned} M_{SOR}(\omega) &= (D - \omega L)^{-1} [(1 - \omega)D + \omega R] \\ &= \left( D \left( I_n - \omega D^{-1}L \right) \right)^{-1} D [(1 - \omega)I_n + \omega D^{-1}R] \\ &= \left( I_n - \omega D^{-1}L \right)^{-1} [(1 - \omega)I_n + \omega D^{-1}R] \end{aligned}$$

$D^{-1}L$  ist wiederum eine untere Dreiecksmatrix mit  $a_{ii} = 0 \quad \forall i = 1, \dots, n$ . Ebenso ist  $D^{-1}R$  eine obere Dreiecksmatrix. Daher lassen sich die Determinanten leicht berechnen:

$$\begin{aligned} \det M_{SOR}(\omega) &= \det \left( \left( I_n - \omega D^{-1}L \right)^{-1} \right) \cdot \det \left( (1 - \omega)I_n + \omega D^{-1}R \right) \\ &= (1 - \omega)^n \end{aligned}$$

Seien  $\lambda_1, \dots, \lambda_n \in \sigma(M_{SOR}(\omega))$ , dann gilt:

$$\det(M_{SOR}(\omega)) = \prod_{i=1}^n \lambda_i$$

Somit ergibt sich für den Spektralradius:

$$\begin{aligned} \operatorname{spr}(M_{SOR}(\omega)) &= \max_{1 \leq i \leq n} |\lambda_i| \\ &\geq \left( \prod_{i=1}^n |\lambda_i| \right)^{\frac{1}{n}} \\ &= |1 - \omega| \end{aligned}$$

□

Der Satz von Kahan zeigt, dass das SOR-Verfahren höchstens für  $\omega \in (0, 2)$  konvergent sein kann, da ansonsten  $\operatorname{spr}(M_{SOR}(\omega)) \geq 1$  wäre.

Für positiv definite Matrizen gilt für diesen Satz auch die Umkehrung:

**Satz 1**

Sei  $A \in \mathbb{R}^{n \times n}$  eine positiv definite Matrix, dann gilt für  $0 < \omega < 2$ :

$$\operatorname{spr}(M_{SOR}(\omega)) < 1$$

*Beweis:* Da  $A$  symmetrisch ist, gilt  $R = L^T$ , das heißt  $A$  lässt sich zerlegen in  $A = D - L - L^T$ . Sei  $\lambda \in \sigma(M_{SOR}(\omega))$  und  $v$  der dazugehörige Eigenvektor. Dann gilt:

$$\begin{aligned} M_{SOR}(\omega)v &= \lambda v \\ \Leftrightarrow (D - \omega L)^{-1} [(1 - \omega)D + \omega R]v &= \lambda v \\ \Leftrightarrow [(1 - \omega)D + \omega R]v &= \lambda(D - \omega L)v \\ \Leftrightarrow \omega(D - L^T)v &= (1 - \lambda)Dv + \lambda\omega Lv \end{aligned}$$

$$\begin{aligned} \omega Av &= \omega(D - L - L^T)v \\ &= \omega(D - L^T)v - \omega Lv \\ &= (1 - \lambda)(Dv - \omega Lv) \end{aligned} \tag{i}$$

$$\begin{aligned} \lambda\omega Av &= \lambda\omega(D - L - L^T)v \\ &= \lambda\omega(D - L^T)v - \lambda\omega Lv \\ &= \lambda\omega(D - L^T)v - (\omega(D - L^T)v - (1 - \lambda)Dv) \\ &= (1 - \lambda)(1 - \omega)Dv + (1 - \lambda)\omega L^T v \end{aligned} \tag{ii}$$

Es gilt:  $v^T Lv = v^T L^T v$ , da  $(v^T Lv)^T = (Lv)^T v = v^T L^T v$  und  $v^T Lv \in \mathbb{R} \Rightarrow v^T Lv = (v^T Lv)^T$ . Daher gilt mit den Gleichungen (i) und (ii):

$$\begin{aligned} \omega v^T Av &= (1 - \lambda)v^T Dv - \omega(1 - \lambda)v^T Lv \\ \lambda\omega v^T Av &= (1 - \lambda)(1 - \omega)v^T Dv + \omega(1 - \lambda)v^T L^T v \\ \Rightarrow (1 + \lambda)\omega v^T Av &= (1 - \lambda)(2 - \omega)v^T Dv \end{aligned}$$

Da mit  $v^T Av > 0$  auch  $v^T Dv > 0$  gilt, was aus der linearen Algebra bekannt ist, folgt mit  $0 < \omega < 2$ :  $\lambda \neq \pm 1 \forall \lambda \in \sigma(A)$ . Durch Umformungen erhält man:

$$\frac{1 + \lambda}{1 - \lambda} = \frac{2 - \omega}{\omega} \frac{v^T Dv}{v^T Av} > 0$$

Wir definieren:  $\mu := \frac{1+\lambda}{1-\lambda}$  Dann gilt, da  $\mu > 0$ :

$$\begin{aligned} |\lambda| &= \left| \frac{\mu - 1}{\mu + 1} \right| \\ |\mu - 1| &< |\mu + 1| \\ \Rightarrow |\lambda| &< 1 \end{aligned}$$

□

### 3 Analyse im 2-zyklischen Fall

**Definition 1** (schwach 2-zyklisch)

Sei  $I$  eine nichtleere, endliche Indexmenge. Eine Matrix  $A \in K^{I \times I}$  heißt schwach 2-zyklisch, wenn es eine Blockstruktur  $\{I_1, I_2\}$  mit nichtleeren Indexteilmengen  $I_1, I_2 \subset I$ ,  $I_1 \cup I_2 = I$  und  $I_1 \cap I_2 = \emptyset$ , gibt, so dass  $a_{ij} = 0$  für  $i, j \in I_1$  oder  $i, j \in I_2$ .

Das bedeutet, dass eine Matrix schwach 2-zyklisch ist, wenn die Diagonalblöcke verschwinden.

**Definition 2**

Das Paar  $\{A, D\}$  heißt schwach 2-zyklisch, wenn  $A-D$  schwach 2-zyklisch ist.

**Definition 3** (2-zyklisch)

$A$  bzw.  $\{A, D\}$  heißen 2-zyklisch, falls die Indexmenge  $I$  angeordnet ist und  $A$  bzw.  $\{A, D\}$  schwach 2-zyklisch bzgl.  $I_1 = \{1, \dots, n_1\}$ ,  $I_2 = \{n_1 + 1, \dots, n\}$  für ein geeignetes  $n_1$  mit  $1 \leq n_1 \leq n - 1$  ist.

Das heißt eine 2-zyklische Matrix hat die Gestalt:

$$A = \left( \begin{array}{c|c} 0 & A_1 \\ \hline A_2 & 0 \end{array} \right)$$

Das Paar  $\{A, D\}$  ist 2-zyklisch, wenn

$$A = \left( \begin{array}{c|c} D_1 & A_1 \\ \hline A_2 & D_2 \end{array} \right) \quad D = \left( \begin{array}{c|c} D_1 & 0 \\ \hline 0 & D_2 \end{array} \right) \quad \text{womit} \quad A - D = \left( \begin{array}{c|c} 0 & A_1 \\ \hline A_2 & 0 \end{array} \right)$$

Bevor wir uns der Analyse des SOR-Verfahrens widmen, untersuchen wir die Eigenschaften einer schwach 2-zyklischen Matrix.

**Lemma 2**

Das Spektrum einer schwach 2-zyklischen Matrix  $B$  mit den Außerdiagonalblöcken  $B_1$  und  $B_2$  ist gegeben durch:

$$\sigma(B) = \pm \sqrt{\sigma(B_1 B_2)} \cup \pm \sqrt{\sigma(B_2 B_1)}$$

Hierbei gilt:  $\pm \sqrt{\sigma(C)} := \{\lambda \in \mathbb{C} \mid \lambda^2 \in \sigma(C)\}$ . Außerdem stimmen die Spektren  $\sigma(B_1 B_2)$  und  $\sigma(B_2 B_1)$  bis auf einen eventuellen Nulleigenwert überein, d.h.

$$\sigma(B_1 B_2) \setminus \{0\} = \sigma(B_2 B_1) \setminus \{0\}$$

*Beweis.* Sei  $\lambda \in \sigma(B)$  und  $v = (v^1 \ v^2)^T$  der dazugehörige Eigenvektor. Dann gilt:

$$Bv = \lambda v \quad \Leftrightarrow \quad \begin{cases} B_1 v^2 = \lambda v^1 \\ B_2 v^1 = \lambda v^2 \end{cases} \quad (i)$$

Setzt man nun beide Gleichungen ineinander ein, erhält man:

$$\begin{aligned} \lambda^2 v^1 &= B_1 B_2 v^1 \\ \lambda^2 v^2 &= B_2 B_1 v^2 \end{aligned}$$

Da  $v \neq 0$ , ist entweder  $v^1 \neq 0$  oder  $v^2 \neq 0$ . Somit ist also  $\lambda^2 \in \sigma(B_1 B_2)$  oder  $\lambda^2 \in \sigma(B_2 B_1)$  und damit gilt:

$$\sigma(B) \subset \pm\sqrt{\sigma(B_1 B_2)} \cup \pm\sqrt{\sigma(B_2 B_1)}$$

Sei nun  $0 \neq \lambda \in \pm\sqrt{\sigma(B_1 B_2)}$ , also  $0 \neq \lambda^2 \in \sigma(B_1 B_2)$ . Für den zugehörigen Eigenvektor  $v^1 \neq 0$  gilt also:

$$\lambda^2 v^1 = B_1 B_2 v^1$$

Wir definieren  $v^2 := \frac{1}{\lambda} B_2 v^1$ . Hierfür gilt:

$$B_1 v^2 = \frac{1}{\lambda} B_1 B_2 v^1 = \frac{1}{\lambda} \lambda^2 v^1 = \lambda v^1 \quad \text{und} \quad \lambda v^2 = B_2 v^1$$

Somit erfüllt  $v^2$  die Eigenschaften aus (i) und  $v = (v^1 \ v^2)^T$  ist somit der Eigenvektor zu  $\lambda \in \sigma(B)$ .

Analog kann man diese Rechnung auch für  $\lambda \in \pm\sqrt{\sigma(B_2 B_1)}$  durchführen. Es bleibt also nur noch der Fall  $0 = \lambda^2 \in \sigma(B_1 B_2) \cup \sigma(B_2 B_1)$  zu betrachten. Gilt diese Aussage, dann muss eine der Matrizen  $B_1, B_2$  einen nichttrivialen Kern besitzen. Wir nehmen o.B.d.A. an, dies sei  $B_1$ , d.h. es gilt:  $B_1 v^2 = 0$  mit  $v^2 \neq 0$ . Wir setzen  $v^1 := 0$ . Somit folgt  $B_2 v^1 = 0$ , womit  $v = (v^1 \ v^2)^T$  der Eigenvektor von  $0 = \lambda \in \sigma(B)$  ist. Also folgt:

$$\pm\sqrt{\sigma(B_1 B_2)} \cup \pm\sqrt{\sigma(B_2 B_1)} \subset \sigma(B)$$

und damit:

$$\sigma(B) = \pm\sqrt{\sigma(B_1 B_2)} \cup \pm\sqrt{\sigma(B_2 B_1)}$$

Nun beweisen wir noch den zweiten Teil des Satzes. Sei  $\lambda \in \sigma(B_1 B_2) \setminus \{0\}$  und  $v$  der dazugehörige Eigenvektor, d.h. es gilt  $B_1 B_2 v = \lambda v$ . Da  $\lambda v \neq 0$ , gilt:  $w := B_2 v \neq 0$ . Nun wird mit  $B_2$  multipliziert und es folgt:

$$B_2 B_1 w = B_2 B_1 B_2 v = B_2 \lambda v = \lambda w,$$

d.h.  $w$  ist der zu  $\lambda \in \sigma(B_2 B_1) \setminus \{0\}$  gehörende Eigenvektor. Somit folgt:

$$\sigma(B_1 B_2) \subset \sigma(B_2 B_1)$$

Analog ergibt sich die umgekehrte Inklusion. □

Mit diesem Lemma ergibt sich, dass wenn  $\lambda$  ein Eigenwert einer schwach 2-zyklischen Matrix ist, dann gilt dies auch für  $-\lambda$ .

Im Folgenden verlangen wir für  $\{E, U\}$  mit  $E := D^{-1}L$ ,  $U := D^{-1}R$ :

$$\text{Die Eigenwerte von } zE + \frac{1}{z}U \text{ sind nicht von } z \in \mathbb{C} \text{ abhängig} \quad (7)$$

Offensichtlich ist  $M_J = E + U$ , wobei  $M_J$  die Iterationsmatrix des Jacobi-Verfahrens ist.

**Lemma 3**

$\{E, U\}$  erfülle Bedingung (7). Dann gilt:

1.  $\sigma(\alpha E + \beta U) = \sigma(\pm\sqrt{\alpha\beta}(E + U)) \quad \forall \alpha, \beta \in \mathbb{C}$
2. Insbesondere haben  $E + U$  und  $-(E + U)$  die gleichen Eigenwerte.
3. Sind alle Eigenwerte von  $E + U$  reell, so gilt:  $\text{spr}(E + U) \in \sigma(E + U)$

*Beweis.* 1. Für den Beweis müssen wir 4 Fälle untersuchen:

- $\alpha\beta \neq 0$ : Wir definieren  $z := \pm\sqrt{\frac{\alpha}{\beta}}$ . Damit gilt:

$$\alpha E + \beta U = \pm\sqrt{\alpha\beta} \left( zE + \frac{1}{z}U \right)$$

Laut Voraussetzung sind die Eigenwerte von  $zE + \frac{1}{z}U$  nicht von  $z$  abhängig, d.h.  $\sigma\left(\pm\sqrt{\alpha\beta} \left( zE + \frac{1}{z}U \right)\right) = \sigma\left(\pm\sqrt{\alpha\beta} (E + U)\right)$ .

- $\alpha = \beta = 0$ : Dieser Fall ist trivial, da  $\sigma(0) = \sigma(0) = \{0\}$ .
- $\alpha = 0, \beta \neq 0$ : Da  $U$  eine strikte Dreiecksmatrix ist, gilt:

$$\sigma(\beta U) = \{0\} = \sigma(0(E + U))$$

- $\alpha \neq 0, \beta = 0$ : Dieser Fall ist analog zum dritten Fall.

2. Für  $\alpha = \beta = 1$  gilt mit Teil 1:

$$\sigma(E + U) = \sigma(\pm 1(E + U))$$

also auch

$$\sigma(E + U) = \sigma(-(E + U))$$

3. Wir definieren:

$$\begin{aligned} \beta_{\min} &:= \min\{\lambda \in \sigma(E + U)\} \\ \beta_{\max} &:= \max\{\lambda \in \sigma(E + U)\} \end{aligned}$$

Mit Lemma 2 gilt  $\beta_{\min} = -\beta_{\max} \leq 0 \leq \beta_{\max}$ . Somit folgt:

$$\begin{aligned} \text{spr}(E + U) &= \max\{|\beta_{\min}|, |\beta_{\max}|\} \\ &= \beta_{\max} \in \sigma(E + U) \end{aligned}$$

□

Im Fall des SOR-Verfahrens stellt sich natürlich die Frage, wie man  $\omega$  wählen muss, um einen möglichst kleinen Spektralradius  $\text{spr}(M_{\text{SOR}}(\omega))$  zu erhalten. Für allgemeine Matrizen ist ein optimales  $\omega$  nur schwer zu finden. Daher muss die Matrix  $A$  einige Voraussetzungen erfüllen, um das optimale  $\omega$  bestimmen zu können.



**Satz 2** (Satz von Young)

Sei für  $A \in \mathbb{R}^{n \times n}$  mit  $a_{ii} \neq 0 \forall i = 1, \dots, n$  die Standardzerlegung gegeben:  $A = D - L - R$ ,  $E := D^{-1}L$ ,  $U := D^{-1}R$ . Die Eigenwerte von  $M_J = E + U$  seien reell mit  $\beta := \text{spr}(M_J) < 1$ .  $D$  und  $I - \omega E$  seien regulär,  $\{E, U\}$  erfülle die Bedingung (7). Dann gilt:

1. Es ist  $\text{spr}(M_{SOR}(\omega)) < 1$  für alle  $\omega \in (0, 2)$ .
2. Es ist

$$\text{spr}(M_{SOR}(\omega)) = \begin{cases} 1 - \omega + \frac{1}{2}\omega^2\beta^2 + \omega\beta\sqrt{1 - \omega + \frac{\omega^2\beta^2}{4}} & \text{für } \omega \in (0, \omega^*] \\ \omega - 1 & \text{für } \omega \in [\omega^*, 2) \end{cases} \quad \begin{matrix} (a) \\ (b) \end{matrix}$$

für

$$\omega^* := \frac{2}{1 + \sqrt{1 - \beta^2}} \in (1, 2).$$

3.  $\omega^*$  ist der optimale Relaxationsparameter im Sinne von

$$\text{spr}(M_{SOR}(\omega^*)) \leq \text{spr}(M_{SOR}(\omega))$$

für alle  $\omega \in (0, 2)$ . Für  $\omega \neq \omega^*$  gilt hierbei die strikte Ungleichung.

4. Für  $\omega \leq \omega^*$  ist  $\text{spr}(M_{SOR}(\omega)) \in \sigma(M_{SOR}(\omega))$  selbst Eigenwert.
5. Für  $\omega \geq \omega^*$  haben alle Eigenwerte von  $M_{SOR}(\omega)$  den Betrag  $|\omega - 1|$

*Beweis.* Als erstes leiten wir einen Zusammenhang zwischen den jeweiligen Eigenwerten von  $M_{SOR}(\omega)$  und  $M_J$  her, da wir diesen Zusammenhang im Verlauf des Beweises benötigen. Sei  $\lambda \in \sigma(M_{SOR}(\omega))$  und  $v$  der zu  $\lambda$  gehörende Eigenvektor, d.h. es gilt  $M_{SOR}(\omega)v = \lambda v$ .

$$\begin{aligned} M_{SOR}(\omega) &= (D - \omega L)^{-1} [(1 - \omega)D + \omega R] \\ \Rightarrow \quad ((1 - \omega)D + \omega R)v &= \lambda(D - \omega L)v \\ &= \lambda D(I - \omega D^{-1}L)v \\ &= \lambda D(I - \omega E)v \\ \Leftrightarrow \quad D \left( (1 - \omega)I + \omega D^{-1}R \right) v &= \lambda D(I - \omega E)v \\ \Leftrightarrow \quad ((1 - \omega)I + \omega U)v &= \lambda(I - \omega E)v \\ \Leftrightarrow \quad (\omega U + \lambda \omega E)v &= (\lambda + \omega - 1)v \end{aligned}$$

d.h.  $\lambda + \omega - 1 \in \sigma(\omega U + \lambda \omega E)$ . Mit Lemma 3.1. gilt  $\sigma(\omega U + \lambda \omega E) = \sigma(\pm \sqrt{\lambda} \omega (E + U))$ . Daraus folgt, dass es einen Eigenwert  $\mu \in \sigma(M_J) = \sigma(E + U)$  gibt, so dass

$$\lambda + \omega - 1 = \pm \sqrt{\lambda} \omega \mu \quad (i)$$

$$\Leftrightarrow (\lambda + \omega - 1)^2 = \lambda \omega^2 \mu^2 \quad (ii)$$

Umstellen nach  $\lambda$  und lösen der quadratischen Gleichung ergibt dann:

$$\lambda = 1 - \omega + \frac{1}{2}\omega^2\mu^2 \pm \omega\mu\sqrt{1 - \omega + \frac{\omega^2\mu^2}{4}} \quad (iii)$$

Da  $M_J$  mit jedem Eigenwert  $\mu$  auch den Eigenwert  $-\mu$  besitzt (Lemma 3.2), gehört jede Lösung  $\mu$  von Formel (ii) zu  $M_J$ . Umgekehrt schließt man daraus, dass zu jedem  $\mu \in \sigma(M_J)$  beide Lösungen  $\lambda$  Gleichung (i) für ein spezielles Vorzeichen in  $\pm\lambda$  erfüllen. Da  $\pm\sqrt{\lambda}\omega\mu \in \sigma(\pm\sqrt{\lambda}\omega(E+U)) = \sigma(\omega U + \lambda\omega E)$ , gilt:  $\lambda \in \sigma(M_{SOR}(\omega))$ . Somit folgt:

$$\lambda \in \sigma(M_{SOR}(\omega)) \Leftrightarrow \mu \in \sigma(M_J) \quad \lambda, \omega \text{ genügen Gleichung (ii)}$$

Sei  $\omega^* \leq \omega < 2$ . Dann gilt durch einfaches Umformen und mit der Definition für  $\omega^*$ :

$$1 - \omega + \frac{\omega^2\beta^2}{4} \leq 0$$

Da  $\beta = \text{spr}(M_J)$ , gilt für alle  $\mu \in \sigma(M_J)$   $-\beta \leq \mu \leq \beta$  und somit:

$$1 - \omega + \frac{\omega^2\mu^2}{4} \leq 0$$

Somit lässt sich jedes  $\lambda$ , das Gleichung (iii) erfüllt, darstellen als:

$$\lambda = \lambda_{Re} \pm i\lambda_{Im} \text{ mit}$$

$$\lambda_{Re} = 1 - \omega + \frac{\omega^2\mu^2}{2} \quad \lambda_{Im} = \omega\mu\sqrt{\omega - 1 - \frac{\omega^2\mu^2}{4}}$$

Nun rechnet man nach (trivial):

$$|\lambda|^2 = (\omega - 1)^2 \Rightarrow |\lambda| = |\omega - 1|,$$

womit 5. und 2.(b) bewiesen wäre.

Sei nun  $0 < \omega < \omega^*$  dann existiert  $\mu \in \sigma(M_J)$ , so dass wieder gilt:

$$1 - \omega + \frac{\omega^2\mu^2}{4} \leq 0$$

Diese Eigenwerte von  $M_J$  erzeugen also wieder  $\lambda \in \sigma(M_{SOR}(\omega))$  mit  $|\lambda| = |\omega - 1|$ . Dieser Wert ist jedoch kleiner als 2.(a). Somit müssen wir nur noch den Fall  $1 - \omega + \frac{\omega^2\mu^2}{4} > 0$  betrachten. 2.(a) ergibt sich, wenn man  $\mu := \beta \in \sigma(M_J)$  wählt. Dies gilt wegen Lemma 3.3. Nun ist nur noch zu zeigen, dass  $|\lambda|$  mit  $\lambda$  aus Gleichung (iii) sein Maximum bei  $\mu = \beta$  annimmt. Es gilt:

$$\lambda \leq 1 - \omega + \frac{1}{2}\omega^2\beta^2 + \omega\beta\sqrt{1 - \omega + \frac{\omega^2\beta^2}{4}}$$

Da  $1 - \omega + \frac{\omega^2\beta^2}{4} > 0$ , gilt  $1 - \omega + \frac{\omega^2\beta^2}{2} > 0$ . Somit wäre auch 2.(a) gezeigt. Da  $\beta \in \sigma(M_J)$  gilt, dass  $\text{spr}(M_{SOR}(\omega)) \in \sigma(M)$ , was 4. beweist. Um Teil 3 zu beweisen, müssen wir zeigen, dass 2.(a) monoton fallend ist.

Zur Abkürzung definieren wir:

$$f(\omega) := \text{spr}(M_{SOR}(\omega))$$

Somit gilt für  $\omega \in (0, \omega^*]$ :

$$\begin{aligned} f(\omega) &= \frac{1}{4} \left( \omega\beta + \sqrt{\beta^2\omega^2 - 4(\omega - 1)} \right)^2 \\ \Rightarrow f'(\omega) &= \frac{1}{2} \left( \omega\beta + \sqrt{\beta^2\omega^2 - 4(\omega - 1)} \right) \left( \beta + \frac{\omega\beta^2 - 2}{\sqrt{\beta^2\omega^2 - 4(\omega - 1)}} \right) \end{aligned}$$

Da der erste Faktor positiv ist, wie schon im 2. Teil bemerkt wurde, müssen wir nur noch zeigen, dass der zweite Faktor einen negativen Wert annimmt. Dazu genügt es nachzuweisen, dass

$$\beta \left( \sqrt{\beta^2 \omega^2 - 4(\omega - 1)} \right) + \omega \beta^2 - 2 < 0 \quad (\text{i})$$

Es gilt:

$$\begin{aligned} \beta^2 \left( \beta^2 \omega^2 - 4(\omega - 1) \right) &= \beta^4 \omega^2 - 4\omega \beta^2 + 4\beta^2 \\ &< \beta^2 \omega^2 - 4\omega \beta^2 + 4 \\ &= \left( \beta^2 \omega - 2 \right)^2 \end{aligned} \quad (\text{ii})$$

Da

$$\omega \beta^2 - 2 \leq \omega^* \beta^2 - 2 = \frac{2}{1 + \sqrt{1 - \beta^2}} \beta^2 - 2 < 0$$

folgt durch Wurzelziehen in (2) die Ungleichung (1). Somit ist  $f'(x) < 0 \forall \omega \in (0, \omega^*)$ , d.h., die Abbildung ist auf  $(0, \omega^*]$  streng monoton fallend, womit bestätigt wäre, dass die Abbildung  $\omega \mapsto \text{spr}(M_{SOR}(\omega))$  in  $\omega = \omega^*$  ihr Minimum annimmt.

Aus der Darstellung von  $\text{spr}(M_{SOR}(\omega))$  und der Tatsache, dass  $\text{spr}(M_{SOR}(\omega))$  streng monoton fallend für  $\omega \in (0, \omega^*]$  ergibt sich, dass  $\text{spr}(M_{SOR}(\omega)) < 1$ , d.h. das SOR-Verfahren konvergiert.  $\square$

## 4 Beispiel

Wir betrachten das Gleichungssystem  $Ax = b$  mit:

$$A = \begin{pmatrix} 0.7 & -0.2 & -0.1 \\ -0.2 & 0.6 & -0.1 \\ -0.1 & -0.1 & 0.9 \end{pmatrix} \quad b = \begin{pmatrix} 20 \\ 40 \\ 0 \end{pmatrix}$$

Hierfür ist:

$$M_J = \begin{pmatrix} 0 & 0.2857 & 0.1429 \\ 1/3 & 0 & 1/6 \\ 1/9 & 1/9 & 0 \end{pmatrix} \quad \beta = 0.3955 \quad \Rightarrow \omega = \omega^* = 1.0425$$

Also:

$$M_{SOR}(\omega) = \begin{pmatrix} -0.0425 & 0.2979 & 0.1489 \\ -0.0148 & 0.0610 & 0.2255 \\ -0.0066 & 0.0416 & 0.0009 \end{pmatrix}$$

Die folgende Tabelle zeigt die Ergebnisse der Iteration:

k	$x_1$	$x_2$	$x_3$	e
1	29.7854	79.8497	12.6993	26.2146
2	54.1947	87.1455	15.8322	1.8053
3	55.7972	87.9367	15.9763	0.2028
4	55.9862	87.9938	15.9987	0.0138
5	55.9985	87.9995	15.9998	0.0015

D.h., dass das SOR-Verfahren nach 5 Iterationsschritten die gewünschte Genauigkeit von  $e = 0.01$  erreicht. Zur Erinnerung: Das Gauß-Seidel-Verfahren gelangte nach 6 Iterationsschritten zu dem gleichen Ergebnis. In diesem Fall ist also der Unterschied zwischen SOR- und Gauß-Seidel-Verfahren nicht so groß, was darauf zurückzuführen ist, dass  $\omega^*$  nahe bei 1 liegt. In Abbildung 1 ist der typische Verlauf der Funktion  $f : \omega \mapsto \text{spr}(M_{SOR}(\omega))$  mit den Parametern aus dem Beispiel aufgezeichnet.

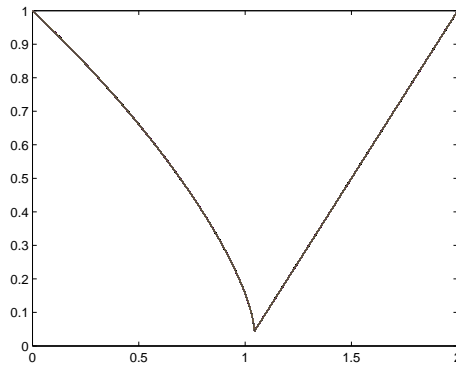


Abbildung 1: Typischer Verlauf der Funktion  $\omega \mapsto \text{spr}(M_{SOR}(\omega))$

## 5 Ordnungsverbesserung und praktische Handhabung

Da man in der Praxis für die meisten Probleme kein Gleichungssystem in der Form  $Ax = b$  vorliegen hat, muss man solche Probleme diskretisieren. Heutzutage ist z.B. das Lösen von Randwertaufgaben ein wichtiges Thema der Numerischen Mathematik. Wir betrachten hierfür das einfachste nichttriviale Beispiel: die Poissongleichung:

$$-\Delta u(x, y) = f(x, y) \quad (x, y) \in \Omega \quad (8)$$

$$u(x, y) = \phi(x, y) \quad (x, y) \in \partial\Omega \quad (9)$$

wobei  $\Delta = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2}$ ,  $\Omega = (0.1) \times (0.1)$ . Wir diskretisieren das Problem, in dem wir  $\Omega$  mit einem Gitter überziehen, wobei jeder Punkt die Darstellung  $x = ih$ ,  $y = ih$  haben muss, mit  $h = \frac{1}{N}$  als Schrittweite. Nun werden die 2. Ableitungen durch den Differenzenquotienten approximiert, wobei dabei immer die jeweils benachbarten Gitterpunkte betrachtet werden. Da jeder Punkt 4 Nachbarpunkte hat, berechnet man die Ableitungen mit Hilfe der 5-Punkte-Regel. Setzt man diese Näherungswerte in die Poisson-Gleichung ein, so erhält man ein lineares Gleichungssystem mit  $N^2$  Unbekannten.

Nun wird der Begriff des Aufwandes eines Iterationsverfahrens definiert. Sei  $s(n)$  die Anzahl der Nichtnullelemente von  $A$ . Für Matrizen  $A$  mit  $Rg(A) = n$ , die aus Diskretisierungen partieller Differenzialgleichungen stammen, gilt:

$$s(n) \leq C_A n,$$

wobei  $C_A$  eine von  $A$  abhängige Konstante ist, die in unserem Fall klein ist, d.h.  $A$  ist dünnbesetzt. Betrachten wir nun das Verfahren  $\Theta$ , so definieren wir  $Aufwand(\Theta, A)$  als die Anzahl der arithmetischen Operationen pro Iterationsschritt. Es gilt:

$$Aufwand(\Theta, A) \leq C_\Theta C_A n$$

Hierbei ist zu beachten, dass  $C_\Theta$  eine verfahrensspezifische Konstante ist, während  $C_A n$  die Schwachbesetztheit von  $A$  kennzeichnet.

Die Konvergenzrate  $spr(M)$  eines Verfahrens  $\Theta$  mit Iterationsmatrix  $M$  liegt normalerweise nahe bei 1, d.h. man kann  $spr(M)$  schreiben als

$$spr(M) = 1 - \eta \quad \eta \text{ klein,}$$

wobei  $\eta$  mit der Schrittweite  $h = \frac{1}{N}$  und mit der von  $\eta$  abhängigen Konstante  $C_\eta$  zusammenhängt:

$$\begin{aligned} \eta &= C_\eta h^\tau + O(h^{2\tau}) \\ \Rightarrow spr(M) &= 1 - C_\eta h^\tau + O(h^{2\tau}) \end{aligned} \quad (10)$$

#### Definition 4

$\tau$  aus (10) wird *Ordnung der Konvergenzrate* genannt. Hat  $\Theta_1$  eine größere Ordnung als  $\Theta_2$ , so ist  $\Theta_1$  bei hinreichend kleinem  $h$  aufwendiger als  $\Theta_2$ .

Gilt  $C_1 := \frac{C_{\Theta_1}}{C_{\eta_1}} < \frac{C_{\Theta_2}}{C_{\eta_2}} =: C_2$ , so ist  $\Theta_2$  um den Faktor  $\frac{C_2}{C_1}$  aufwendiger.

Nun betrachten wir wieder unsere Iterationsverfahren:

Vergleicht man das Jacobi-Verfahren mit dem Gauß-Seidel-Verfahren, so sieht man, dass sich wegen  $spr(M_{GS}) = spr(M_J)^2$  nur der Koeffizient  $C_\eta^{GS} = 2C_\eta^J$  verbessert, während die Ordnung erhalten bleibt. Für das SOR-Verfahren gilt jedoch folgender Satz:

#### Satz 3

Das Jacobi-Verfahren habe die Ordnung  $\tau > 0$ . Es seien die Voraussetzungen von Satz 2 erfüllt. Dann hat das SOR-Verfahren für  $\omega = \omega^*$  die Ordnung  $\frac{\tau}{2}$  und es gilt

$$C_\eta^{SOR} = 2\sqrt{2C_\eta^J}$$

*Beweis.* Nach Satz 2 2.(b) gilt:

$$spr(M_{SOR}(\omega^*)) = \omega^* - 1 = \frac{1 - \sqrt{1 - \beta^2}}{1 + \sqrt{1 - \beta^2}} \quad (i)$$

Da

$$\begin{aligned} 1 - \beta^2 &= 1 - spr(M_J)^2 \\ &= 1 - \left(1 - C_\eta^J h^\tau + O(h^{2\tau})\right)^2 \\ &= 2C_\eta^J h^\tau + O(h^{2\tau}) \end{aligned}$$

hat  $\sqrt{1 - \beta^2}$  die Entwicklung  $\sqrt{2C_\eta^J h^{\frac{\tau}{2}}} + O(h^\tau)$ . Setzt man dies nun in (i) ein, erhält man

$$\text{spr}(M_{SOR}(\omega^*)) = 1 - 2\sqrt{2C_\eta^J h^{\frac{\tau}{2}}} + O(h^\tau)$$

□

In der Praxis ist  $\beta = \text{spr}(M_J)$  unbekannt, so dass man  $\omega^*$  schätzen muss. Dazu geht man wie folgt vor:

Man wählt ein  $\omega \leq \omega^*$ , z.B.  $\omega = 1$ . Dann führt man einige Schritte des SOR-Verfahrens durch und bestimmt aus dem Quotienten von  $\|x^{(m+1)} - x^{(m)}\|_2$  eine Näherung  $\bar{\lambda} = \text{spr}(M_{SOR}(\omega))$ . Dann kann man mit Hilfe von Gleichung (i) aus dem Beweis von Satz 2 eine Näherung von  $\beta \approx \bar{\beta} := \frac{|\bar{\lambda} + \omega - 1|}{\omega\sqrt{\bar{\lambda}}}$  berechnen. Mit  $\bar{\beta}$  lässt sich nun wieder eine neue Näherung von  $\omega^*$  bestimmen. Solange  $\omega \leq \omega^*$  kann man die Näherung von  $\omega^*$  iterieren. Da  $\text{spr}(M_{SOR}(\omega))$  bei  $\omega = \omega^*$  von links eine senkrechte Tangente besitzt, führt jede Abweichung nach links zu einer Konvergenzverschlechterung. Aus diesem Grund wählt man  $\omega^*$  eher zu groß.

## 6 Das SSOR-Verfahren

Das symmetrische SOR-Verfahren verwendet in einer Iteration im Prinzip 2 Schritte des normalen SOR-Verfahrens. Hierbei werden die Gleichungen zuerst in der üblichen Reihenfolge durchlaufen, was als Vorwärtsschritt bezeichnet wird, und danach in umgekehrter Reihenfolge, dem Rückwärtsschritt. Man bezeichnet mit  $x^{(k)}$  die k-te Iterierte. Ist diese gegeben, wird mit  $x^{(k+\frac{1}{2})}$  der Vorwärtsschritt bezeichnet. Mit diesen Bezeichnungen lautet die Iterationsvorschrift für das SSOR-Verfahren:

$$x_i^{(k+\frac{1}{2})} = (1 - \omega) x_i^{(k)} - \frac{\omega}{a_{ii}} \left( \sum_{j=1}^{i-1} a_{ij} x_j^{(k+\frac{1}{2})} + \sum_{j=i+1}^n a_{ij} x_j^{(k)} - b_i \right)$$

für  $i = 1, 2, \dots, n$  und

$$x_i^{(k+1)} = (1 - \omega) x_i^{(k+\frac{1}{2})} - \frac{\omega}{a_{ii}} \left( \sum_{j=i+1}^n a_{ij} x_j^{(k+1)} + \sum_{j=1}^{i-1} a_{ij} x_j^{(k+\frac{1}{2})} - b_i \right)$$

für  $i = n, n-1, \dots, 1$ . Für  $\omega = 1$  sprechen wir wieder vom Symmetrischen Gauß-Seidel-Verfahren.

## 7 Quellenverzeichnis

- Hackbusch, Wolfgang:  
Iterative Lösung großer schwachbesetzter Gleichungssysteme  
Teubner-Verlag, 1991
- Kanzlow, Christian:  
Numerik linearer Gleichungssysteme (Direkte und iterative Verfahren)  
Springer Verlag
- Rannacher, Rolf  
Vorlesungsskriptum Einführung in die numerische Mathematik  
2006
- Neuß, Nicolas  
Vorlesungsskriptum Numerische Lineare Algebra  
Wintersemester 2004\2005