Iterative Solvers for Large Linear Systems
Part II: Classical Splitting Schemes

Andreas Meister

University of Kassel, Department of Analysis and Applied Mathematics

Outline

- Basics of Iterative Methods
- Splitting-schemes
  - Jacobi- u. Gauß-Seidel-scheme
  - Relaxation methods
- Methods for symmetric, positive definite Matrices
  - Method of steepest descent
  - Method of conjugate directions
  - CG-scheme

Outline

- Multigrid Method
  - Smoother, Prolongation, Restriction
  - Twogrid Method and Extension
- Methods for non-singular Matrices
  - GMRES
  - BiCG, CGS and BiCGSTAB
- Preconditioning
  - ILU, IC, GS, SGS, ...

Jacobi method

Procedure: Write $A = D + L + R$ by means of

- $D = diag\{a_{11}, \ldots, a_{nn}\}$
- $L = \begin{pmatrix} 0 & \cdot & \cdot & 0 \\ a_{21} & \cdot & & \cdot \\ \vdots & \ddots & \cdot & \cdot \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{pmatrix}$ and $R = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \cdot & \cdot & \ddots & \vdots \\ \cdot & & \cdot & a_{n-1,n} \\ 0 & \cdot & \cdot & 0 \end{pmatrix}$
- Choose $B_J = D$

$$\Longrightarrow \quad M_J = B_J^{-1}(B_J - A) = -D^{-1}(L+R), \quad N_J = B_J^{-1} = D^{-1}$$

$$\Longrightarrow \quad x_{m+1} = -D^{-1}(L+R)x_m + D^{-1}b$$

$$x_{m+1,i} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j\neq i}^{n} a_{ij} x_{m,j} \right), \quad i = 1, \ldots, n$$

Jacobi method

Procedure: Write $A = D + L + R$ by means of

- $D = diag\{a_{11}, \ldots, a_{nn}\}$
- $L = \begin{pmatrix} 0 & \cdot & \cdot & 0 \\ a_{21} & \cdot & & \cdot \\ \vdots & \ddots & \cdot & \cdot \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{pmatrix}$ and $R = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \cdot & \cdot & \ddots & \vdots \\ \cdot & & \cdot & a_{n-1,n} \\ 0 & \cdot & \cdot & 0 \end{pmatrix}$
- Choose $B_J = D$

$$\Longrightarrow \quad M_J = B_J^{-1}(B_J - A) = -D^{-1}(L + R), \quad N_J = B_J^{-1} = D^{-1}$$

$$\Longrightarrow \quad x_{m+1} = -D^{-1}(L + R)x_m + D^{-1}b$$

$$x_{m+1,i} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j \neq i}^{n} a_{ij} x_{m,j} \right), \quad i = 1, \ldots, n$$

Jacobi method

Procedure: Write $A = D + L + R$ by means of

- $D = diag\{a_{11}, \ldots, a_{nn}\}$
- $L = \begin{pmatrix} 0 & \cdot & \cdot & 0 \\ a_{21} & \cdot & & \cdot \\ \vdots & \ddots & \cdot & \cdot \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{pmatrix}$ and $R = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \cdot & \cdot & \ddots & \vdots \\ \cdot & & \cdot & a_{n-1,n} \\ 0 & \cdot & \cdot & 0 \end{pmatrix}$
- Choose $B_J = D$

  $\Longrightarrow \quad M_J = B_J^{-1}(B_J - A) = -D^{-1}(L + R), \quad N_J = B_J^{-1} = D^{-1}$

  $\Longrightarrow \quad x_{m+1} = -D^{-1}(L + R)x_m + D^{-1}b$
- $$x_{m+1,i} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j \neq i}^{n} a_{ij} x_{m,j} \right), \quad i = 1, \ldots, n$$

Jacobi method

$$x_{m+1,i} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j\neq i}^{n} a_{ij} x_{m,j} \right), \quad i = 1, \ldots, n$$

Properties:

- Calculation of $x_{m+1,i}$ by means of the vector

$$(x_{m,1}, \ldots, x_{m,i-1}, 0, x_{m,i+1}, \ldots, x_{m,n})^T$$

- Independent of the numbering of the unknowns
- Very well suited for parallel computing

Appraisement:

"$\circ$" : Minor assumptions on the matrix A, $(a_{ii} \neq 0 \text{ for } i = 1, \ldots, n)$

"$+$" : Very simple calculation of matrix-vector products $B_J^{-1} x = D^{-1} x$

"$\circ$" : Moderate approximation of A $(B_J \longleftrightarrow A)$

Convergence of the Jacobi method

First idea:

By means of the calculation of

$$\|M\| < 1$$

one directly obtains convergence due to

$$\rho(M) \leq \|M\| < 1.$$

Utilizing

$$M_J = D^{-1}(D - A) = -\begin{pmatrix} 0 & \frac{a_{12}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} \\ \frac{a_{21}}{a_{22}} & 0 & \ddots & \vdots \\ \vdots & & \ddots & \frac{a_{n-1,n}}{a_{n-1,n-1}} \\ \frac{a_{n1}}{a_{nn}} & \cdots & \frac{a_{n,n-1}}{a_{nn}} & 0 \end{pmatrix}$$

we recognize:

Jacobi method

Convergence criteria

If the matrix A with $a_{ii} \neq 0 \quad , i = 1, \ldots, n$ fulfills

$$q_\infty := \max_{i=1,\ldots,n} \sum_{\substack{k=1 \\ k \neq i}}^{n} \frac{|a_{ik}|}{|a_{ii}|} < 1 \qquad \text{Strict row diagonal dominance}$$

or

$$q_1 := \max_{k=1,\ldots,n} \sum_{\substack{i=1 \\ i \neq k}}^{n} \frac{|a_{ik}|}{|a_{ii}|} < 1 \qquad \text{Strict column diagonal dominance}$$

or

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

then the Jacobi scheme will converge to the solution vector $x^\star = A^{-1} b$ independent of the right hand side b as well as the initial guess x_0.

Model problem

$$Ax = b \quad with \quad A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

- Convergence criterion

$$M_J = -D^{-1}(D - A) = \begin{pmatrix} 0 & 4/7 \\ 2/5 & 0 \end{pmatrix} \Longrightarrow q_\infty = \max\left\{\tfrac{4}{7}, \tfrac{2}{5}\right\} = \tfrac{4}{7} < 1$$

$\Longrightarrow$ Jacobi scheme is convergent

- Rate of convergence: Eigenvalues of the iteration matrix M_J

$$0 = \det(M_J - \lambda I) = \lambda^2 - \frac{8}{35} \quad \Longrightarrow \quad \lambda_{1,2} = \pm\sqrt{\frac{8}{35}}$$

$$\rho(M_J) = \sqrt{\frac{8}{35}} \approx 0.478 \approx 0.7^2$$

Expectation:

Approximately half as much iterations as the trivial method to reach the same accuracy

Model problem

$$Ax = b \;\; with \;\; A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

- Convergence criterion

  $$M_J = -D^{-1}(D - A) = \begin{pmatrix} 0 & 4/7 \\ 2/5 & 0 \end{pmatrix} \Longrightarrow q_\infty = \max\left\{\tfrac{4}{7}, \tfrac{2}{5}\right\} = \tfrac{4}{7} < 1$$

  $\Longrightarrow$ Jacobi scheme is convergent

- Rate of convergence: Eigenvalues of the iteration matrix M_J

  $$0 = \det(M_J - \lambda I) = \lambda^2 - \frac{8}{35} \quad \Longrightarrow \quad \lambda_{1,2} = \pm\sqrt{\frac{8}{35}}$$

  $$\rho(M_J) = \sqrt{\frac{8}{35}} \approx 0.478 \approx 0.7^2$$

Expectation:

Approximately half as much iterations as the trivial method to reach the same accuracy

Model problem

$$Ax = b \quad with \quad A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

- Convergence criterion

$$M_J = -D^{-1}(D - A) = \begin{pmatrix} 0 & 4/7 \\ 2/5 & 0 \end{pmatrix} \Longrightarrow q_\infty = \max\left\{\tfrac{4}{7}, \tfrac{2}{5}\right\} = \tfrac{4}{7} < 1$$

$\Longrightarrow$ Jacobi scheme is convergent

- Rate of convergence: Eigenvalues of the iteration matrix M_J

$$0 = \det(M_J - \lambda I) = \lambda^2 - \frac{8}{35} \quad \Longrightarrow \quad \lambda_{1,2} = \pm\sqrt{\frac{8}{35}}$$

$$\rho(M_J) = \sqrt{\frac{8}{35}} \approx 0.478 \approx 0.7^2$$

Expectation:

Approximately half as much iterations as the trivial method to reach the same accuracy

Model problem

$$Ax = b \;\; with \;\; A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

- Convergence criterion

$$M_J = -D^{-1}(D - A) = \begin{pmatrix} 0 & 4/7 \\ 2/5 & 0 \end{pmatrix} \Longrightarrow q_\infty = \max\left\{\tfrac{4}{7}, \tfrac{2}{5}\right\} = \tfrac{4}{7} < 1$$

$\Longrightarrow$ Jacobi scheme is convergent

- Rate of convergence: Eigenvalues of the iteration matrix M_J

$$0 = \det(M_J - \lambda I) = \lambda^2 - \frac{8}{35} \quad \Longrightarrow \quad \lambda_{1,2} = \pm\sqrt{\frac{8}{35}}$$

$$\rho(M_J) = \sqrt{\frac{8}{35}} \approx 0.478 \approx 0.7^2$$

Expectation:

Approximately half as much iterations as the trivial method to reach the same accuracy

Jacobi method

Model problem:

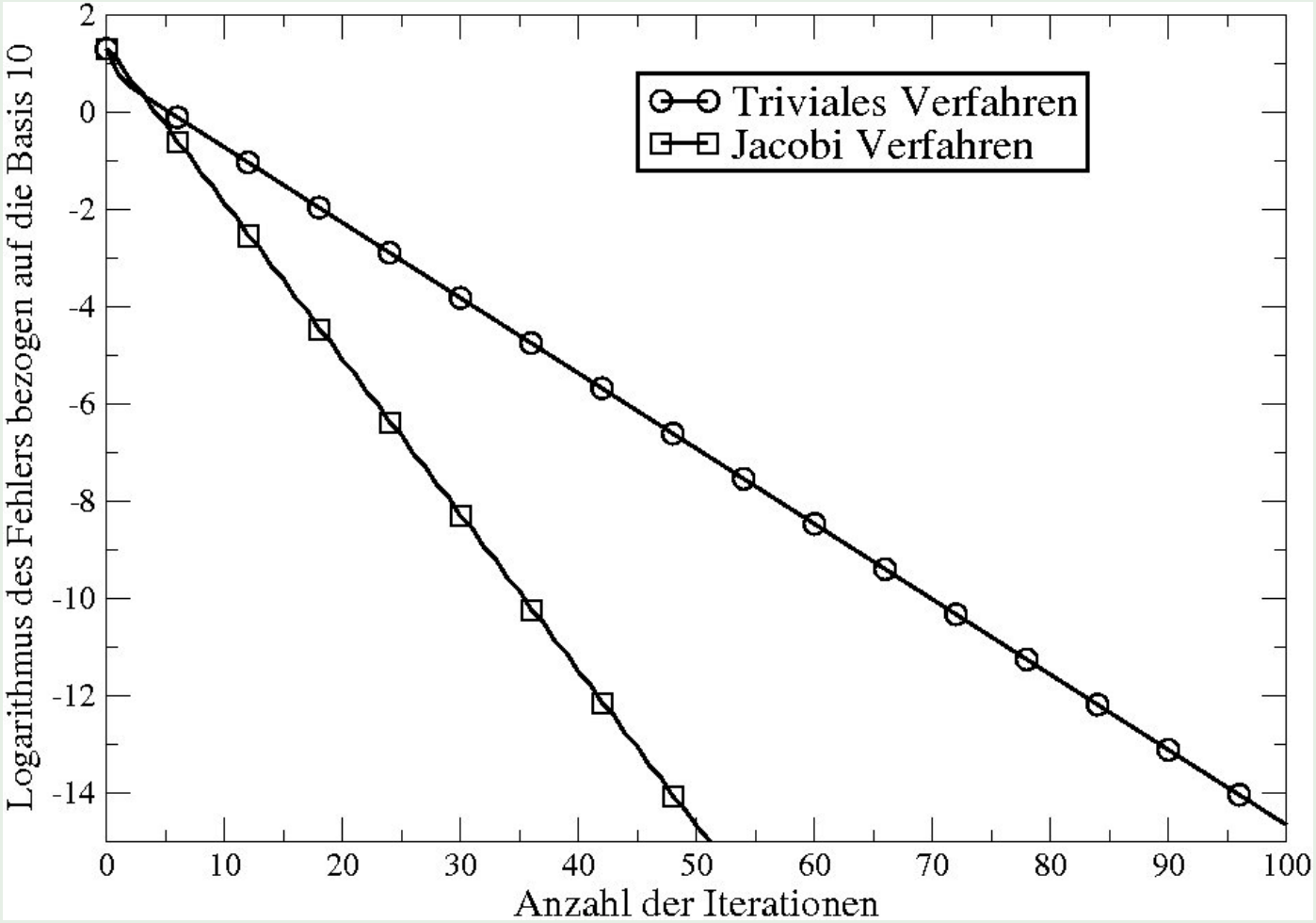


Abbildung: Convergence history $\log_{10} \varepsilon_m$ of the Jacobi method

Jacobi method

Jacobi method				
m	$x_{m,1}$	$x_{m,2}$	$\varepsilon_m := \|x_m - x^*\|_\infty$	$\varepsilon_m/\varepsilon_{m-1}$
0	2.100000e+01	-1.900000e+01	2.000000e+01	
1	-1.042857e+01	9.000000e+00	1.142857e+01	5.714286e-01
2	5.571429e+00	-3.571429e+00	4.571429e+00	4.000000e-01
3	-1.612245e+00	2.828571e+00	2.612245e+00	5.714286e-01
4	2.044898e+00	-4.489796e-02	1.044898e+00	4.000000e-01
5	4.029155e-01	1.417959e+00	5.970845e-01	5.714286e-01
6	1.238834e+00	7.611662e-01	2.388338e-01	4.000000e-01
7	8.635235e-01	1.095534e+00	1.364765e-01	5.714286e-01
8	1.054591e+00	9.454094e-01	5.459059e-02	4.000000e-01
9	9.688054e-01	1.021836e+00	3.119462e-02	5.714286e-01
10	1.012478e+00	9.875222e-01	1.247785e-02	4.000000e-01
11	9.928698e-01	1.004991e+00	7.130199e-03	5.714286e-01
12	1.002852e+00	9.971479e-01	2.852080e-03	4.000000e-01
13	9.983702e-01	1.001141e+00	1.629760e-03	5.714286e-01
14	1.000652e+00	9.993481e-01	6.519039e-04	4.000000e-01
15	9.996275e-01	1.000261e+00	3.725165e-04	5.714286e-01
20	1.000008e+00	9.999922e-01	7.784835e-06	4.000000e-01
25	9.999998e-01	1.000000e+00	2.324102e-07	5.714286e-01
30	1.000000e+00	1.000000e-00	4.856900e-09	4.000000e-01
35	1.000000e-00	1.000000e+00	1.449989e-10	5.714279e-01
40	1.000000e+00	1.000000e-00	3.030243e-12	4.000117e-01
45	1.000000e-00	1.000000e+00	9.037215e-14	5.700280e-01
48	1.000000e+00	1.000000e-00	8.437695e-15	4.086022e-01

Example: 1-D Poisson equation

A central scheme yields the linear system of equations:

$$Ax = b$$

where

$$A = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & & \\ -1 & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & -1 \\ & & & -1 & 2 \end{pmatrix}$$

For $n > 2$ one obtains $q_\infty = q_1 = 1$ and $q_2 \geq 1$.

Furthermore,

$$\left.\begin{array}{l} A \text{ is irreducible} \\ \sum_{j=2}^{n} \frac{|a_{1j}|}{|a_{11}|} = \frac{1}{2} < 1 \end{array}\right\} \Longrightarrow \text{Jacobi scheme is convergent.}$$

Jacobi method

Second idea (Convergence):

- Acceptance of the property $q_\infty = 1$
  in combination with an additional requirement.

Convergence criterion

Let A be irreducible with

$$q_\infty \leq 1.$$

If there exists an index $k \in \{1, \ldots, n\}$ such that

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

then the Jacobi scheme will converge to the solution vector $x^\star = A^{-1}b$ independent of the right hand side b as well as the initial guess x_0.

Jacobi method

Definition:Irreducibility

A matrix is called reducible, if there exists a permutation matrix P such that

$$PAP^T = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}$$

where $\tilde{A}_{ii} \in \mathbb{C}^{n_i \times n_i}, \quad n_1 + n_2 = n.$
Otherwise A is called irreducible.

Irreducibility means:

A matrix is irreducible, if for each pair of indices (k, i) there exists a directed path of length $m + 1$

$$(k, k_1)(k_1, k_2) \dots (k_m, l).$$

Thereby a path (i,j) exists if and only if $a_{ij} \neq 0$.

Jacobi method

Definition:Irreducibility

A matrix is called reducible, if there exists a permutation matrix P such that

$$PAP^T = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}$$

where $\tilde{A}_{ii} \in \mathbb{C}^{n_i \times n_i}, \quad n_1 + n_2 = n.$
Otherwise A is called irreducible.

Irreducibility means:

A matrix is irreducible, if for each pair of indices (k, i) there exists a directed path of length $m + 1$

$$(k, k_1)(k_1, k_2) \dots (k_m, l).$$

Thereby a path (i,j) exists if and only if$a_{ij} \neq 0$.

Jacobi method

Irreducibility means:

A matrix is irreducible, if for each pair of indices (k, l) there exists a directed path of length $m + 1$

$$(k, k_1)(k_1, k_2) \dots (k_m, l).$$

Thereby, a path (i,j) exists if and only if $a_{ij} \neq 0$.

Example:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ reducible} \qquad A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \text{ irreducible}$$

Effect of the irreducibility:

A reduction of the error in the k-th row will successively leads to a reduction of the error in each row $\Longrightarrow$ Convergence

Gauß-Seidel method

Procedure: Write $A = D + L + R$

- Choose $B_{GS} = D + L$

$$\Longrightarrow \quad M_{GS} = (D+L)^{-1}(D+L-A) = -(D+L)^{-1}R,$$

$$N_{GS} = (D+L)^{-1}$$

$$\Longrightarrow \quad x_{m+1} = -(D+L)^{-1}Rx_m + (D+L)^{-1}b$$

- Problem :
  - Calculation of $(D+L)^{-1}$ may be expensive
  - $(D+L)^{-1}$ may be a full matrix
    (PDEs often lead to sparse matrices)
- Solution
  - Component-by-component derivation

Gauß-Seidel method

Procedure: Write $A = D + L + R$

- Choose $B_{GS} = D + L$

  $$\Longrightarrow \quad M_{GS} = (D+L)^{-1}(D+L-A) = -(D+L)^{-1}R,$$

  $$N_{GS} = (D+L)^{-1}$$

  $$\Longrightarrow \quad x_{m+1} = -(D+L)^{-1}Rx_m + (D+L)^{-1}b$$

- Problem :
  - Calculation of $(D+L)^{-1}$ may be expensive
  - $(D+L)^{-1}$ may be a full matrix
    (PDEs often lead to sparse matrices)
- Solution
  - Component-by-component derivation

Gauß-Seidel method

Procedure: Write $A = D + L + R$

- Choose $B_{GS} = D + L$

$$\Longrightarrow \quad M_{GS} = (D+L)^{-1}(D+L-A) = -(D+L)^{-1}R,$$

$$N_{GS} = (D+L)^{-1}$$

$$\Longrightarrow \quad x_{m+1} = -(D+L)^{-1}Rx_m + (D+L)^{-1}b$$

- Problem :
  - Calculation of $(D+L)^{-1}$ may be expensive
  - $(D+L)^{-1}$ may be a full matrix
    (PDEs often lead to sparse matrices)
- Solution
  - Component-by-component derivation

Gauß-Seidel method

Transformation: $x_{m+1} = -(D+L)^{-1}Rx_m + (D+L)^{-1}b$

$$(D+L)x_{m+1} = -Rx_m + b$$

Consider the i-th component:

$$\sum_{j=1}^{i} a_{ij}x_{m+1,j} = -\sum_{j=i+1}^{n} a_{ij}x_{m,j} + b_i$$

$$\Longrightarrow x_{m+1,i} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_{m+1,j} - \sum_{j=i+1}^{n} a_{ij}x_{m,j}\right),\ i = 1,\ldots,n$$

Properties:

- Calculation of $x_{m+1,i}$ by means of $(x_{m+1,1},\ldots,x_{m+1,i-1},0,x_{m,i+1},\ldots,x_{m,n})^T$.
- Dependent on the numbering of the unknowns.
- Horrible for parallel computing

Gauß-Seidel method

Formulation using matrices:

$$x_{m+1} = -(D+L)^{-1} R x_m + (D+L)^{-1} b$$

Pointwise counterpart:

$$x_{m+1,i} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_{m+1,j} - \sum_{j=i+1}^{n} a_{ij} x_{m,j} \right), \quad i = 1, \ldots, n$$

Appraisement:

- $\circ$: Minor assumptions on A
  ($a_{ii} \neq 0 \quad$ für $\quad i = 1, \ldots, n$)
- $+$: Simple calculation of matrix-vector-products by $B_{GS}^{-1} x = (D+L)^{-1} x$
- $+$: Good approximation of A
  ($B_{GS} \longleftrightarrow A$)

Gauß-Seidel method

Convergence criterion

Let A be a square matrix with diagonal elements $a_{ii} \neq 0$. If the auxiliary quantities

$$p_i = \sum_{j=1}^{i-1} \frac{|a_{ij}|}{|a_{ii}|} p_j + \sum_{j=i+1}^{n} \frac{|a_{ij}|}{|a_{ii}|} \quad , i = 1, \ldots, n$$

satisfy

$$p = \max_{i=1,\ldots,n} p_i < 1,$$

then the Gauß-Seidel scheme will converge to the solution vector $x^\star = A^{-1} b$ independent of the right hand side b as well as the initial guess x_0.

Main idea of the proof:

The constraint yields $\|M_{GS}\|_\infty < 1$ that proofs the convergence due to $\rho(M_{GS}) \leq \|M_{GS}\|_\infty$.

Gauß-Seidel method

Example:

$$A = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & & \\ -1 & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & -1 \\ & & & -1 & 2 \end{pmatrix}$$

$$p_1 = \tfrac{1}{2} < 1$$

$$p_i = \tfrac{1}{2}p_{i-1} + \tfrac{1}{2} < 1, \quad i = 2, \ldots, n$$

$$p_n = \tfrac{1}{2}p_{n-1} < 1$$

$\Longrightarrow$ Gauß-Seidel method is convergent.

Modell problem

$$Ax = b \text{ with } A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

- Convergence criterion

$$p_1 = \tfrac{4}{7}, \quad p_2 = \tfrac{2}{5} \cdot \tfrac{4}{7} = \tfrac{8}{35} \quad \Longrightarrow \quad p = \max\{p_1, p_2\} < 1$$

$\Longrightarrow$ Gauß-Seidel method is convergent

- Rate of convergence: Eigenvalues of the iteration matrix M_{GS}

$$0 = \det(M_{GS} - \lambda I) = \det \begin{pmatrix} -\lambda & 4/7 \\ 0 & -8/35 - \lambda \end{pmatrix} \Longrightarrow \lambda_1 = 0, \ \lambda_2 = -\frac{8}{35}$$

$$\rho(M_{GS}) = \frac{8}{35} = \rho(M_J)^2 \approx 0.22857$$

Expectation:

Approximately two times faster convergence compared to the Jacobi scheme

Modell problem

$$Ax = b \text{ with } A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

- Convergence criterion

  $p_1 = \frac{4}{7}, \quad p_2 = \frac{2}{5} \cdot \frac{4}{7} = \frac{8}{35} \quad \Longrightarrow \quad p = \max\{p_1, p_2\} < 1$

  $\Longrightarrow$ Gauß-Seidel method is convergent

- Rate of convergence: Eigenvalues of the iteration matrix M_{GS}

  $$0 = \det(M_{GS} - \lambda I) = \det \begin{pmatrix} -\lambda & 4/7 \\ 0 & -8/35 - \lambda \end{pmatrix} \Longrightarrow \lambda_1 = 0, \ \lambda_2 = -\frac{8}{35}$$

  $$\rho(M_{GS}) = \frac{8}{35} = \rho(M_J)^2 \approx 0.22857$$

Expectation:

Approximately two times faster convergence compared to the Jacobi scheme

Modell problem

$$Ax = b \text{ with } A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

- Convergence criterion

$$p_1 = \tfrac{4}{7}, \quad p_2 = \tfrac{2}{5} \cdot \tfrac{4}{7} = \tfrac{8}{35} \qquad \Longrightarrow \qquad p = \max\{p_1, p_2\} < 1$$

$\Longrightarrow$ Gauß-Seidel method is convergent

- Rate of convergence: Eigenvalues of the iteration matrix M_{GS}

$$0 = \det(M_{GS} - \lambda I) = \det \begin{pmatrix} -\lambda & 4/7 \\ 0 & -8/35 - \lambda \end{pmatrix} \Longrightarrow \lambda_1 = 0, \ \lambda_2 = -\frac{8}{35}$$

$$\rho(M_{GS}) = \frac{8}{35} = \rho(M_J)^2 \approx 0.22857$$

Expectation:

Approximately two times faster convergence compared to the Jacobi scheme

Modell problem

$$Ax = b \text{ with } A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

- Convergence criterion

$$p_1 = \tfrac{4}{7}, \quad p_2 = \tfrac{2}{5} \cdot \tfrac{4}{7} = \tfrac{8}{35} \qquad \Longrightarrow \qquad p = \max\{p_1, p_2\} < 1$$

$\Longrightarrow$ Gauß-Seidel method is convergent

- Rate of convergence: Eigenvalues of the iteration matrix M_{GS}

$$0 = \det(M_{GS} - \lambda I) = \det \begin{pmatrix} -\lambda & 4/7 \\ 0 & -8/35 - \lambda \end{pmatrix} \Longrightarrow \lambda_1 = 0, \ \lambda_2 = -\frac{8}{35}$$

$$\rho(M_{GS}) = \frac{8}{35} = \rho(M_J)^2 \approx 0.22857$$

Expectation:

Approximately two times faster convergence compared to the Jacobi scheme

Gauß-Seidel method

Model problem:

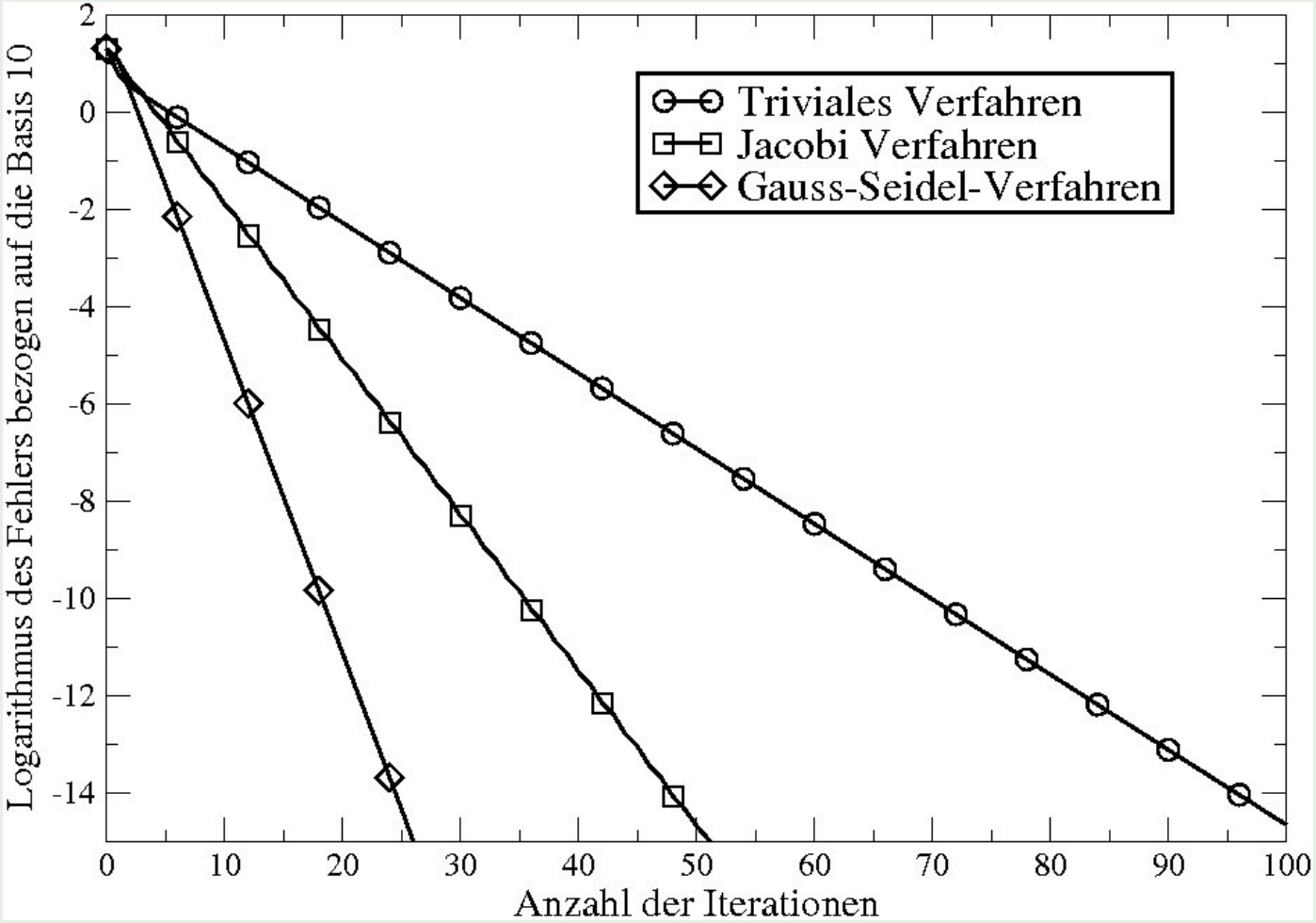


Abbildung: Convergence history $\log_{10} \varepsilon_m$ of the Gauß-Seidel method

Gauß-Seidel method

Gauß-Seidel method				
m	$x_{m,1}$	$x_{m,2}$	$\varepsilon_m := \|x_m - x^*\|_\infty$	$\varepsilon_m/\varepsilon_{m-1}$
0	2.100000e+01	-1.900000e+01	2.000000e+01	
1	-1.042857e+01	-3.571429e+00	1.142857e+01	5.714286e-01
2	-1.612245e+00	-4.489796e-02	2.612245e+00	2.285714e-01
3	4.029155e-01	7.611662e-01	5.970845e-01	2.285714e-01
4	8.635235e-01	9.454094e-01	1.364765e-01	2.285714e-01
5	9.688054e-01	9.875222e-01	3.119462e-02	2.285714e-01
6	9.928698e-01	9.971479e-01	7.130199e-03	2.285714e-01
7	9.983702e-01	9.993481e-01	1.629760e-03	2.285714e-01
8	9.996275e-01	9.998510e-01	3.725165e-04	2.285714e-01
9	9.999149e-01	9.999659e-01	8.514663e-05	2.285714e-01
10	9.999805e-01	9.999922e-01	1.946209e-05	2.285714e-01
11	9.999956e-01	9.999982e-01	4.448477e-06	2.285714e-01
12	9.999990e-01	9.999996e-01	1.016795e-06	2.285714e-01
13	9.999998e-01	9.999999e-01	2.324102e-07	2.285714e-01
14	9.999999e-01	1.000000e-00	5.312234e-08	2.285714e-01
15	1.000000e-00	1.000000e-00	1.214225e-08	2.285714e-01
16	1.000000e-00	1.000000e-00	2.775371e-09	2.285714e-01
17	1.000000e-00	1.000000e-00	6.343704e-10	2.285714e-01
18	1.000000e-00	1.000000e-00	1.449989e-10	2.285713e-01
19	1.000000e-00	1.000000e-00	3.314249e-11	2.285706e-01
20	1.000000e-00	1.000000e-00	7.575385e-12	2.285702e-01
21	1.000000e-00	1.000000e-00	1.731504e-12	2.285698e-01
22	1.000000e-00	1.000000e-00	3.956835e-13	2.285201e-01
23	1.000000e-00	1.000000e-00	9.037215e-14	2.283951e-01
24	1.000000e-00	1.000000e-00	2.065015e-14	2.285012e-01
25	1.000000e-00	1.000000e-00	4.551914e-15	2.204301e-01

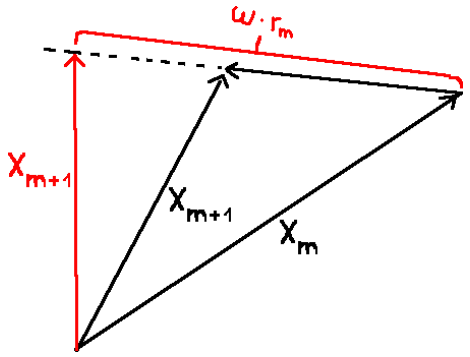
Relaxation method

Utilizing

$$\begin{aligned} x_{m+1} &= B^{-1}(B-A)x_m + B^{-1}b \\ &= x_m + \underbrace{B^{-1}(b-Ax_m)}_{=: r_m \quad \text{(correction vector)}} \end{aligned}$$

Aim of the relaxation

Improving the rate of convergence
by means of weighting the correction vector.



Utilizing $\omega \in \mathbb{R}^+$ (relaxation parameter) leads to

$$x_{m+1} = x_m + \omega r_m = \underbrace{(I - \omega B^{-1}A)}_{= M(\omega)} x_m + \underbrace{\omega B^{-1}}_{= N(\omega)} b$$

Optimality of the relaxation parameter: $\omega_{opt} = \arg \min_{\omega \in \mathbb{R}^+} \rho(M(\omega))$

Gauß-Seidel scheme:

Correction of each component

Jacobi relaxation method

Jacobi scheme:

$$\begin{aligned} x_{m+1} &= D^{-1}(D-A)x_m + D^{-1}b \\ &= x_m + D^{-1}(b - Ax_m) \end{aligned}$$

Jacobi relaxation method:

$$\begin{aligned} x_{m+1} &= x_m + \omega D^{-1}(b - Ax_m) \\ &= (I - \omega D^{-1}A)x_m + \omega D^{-1}b \end{aligned}$$

Jacobi relaxation method

Optimal relaxation parameter

Assumption:

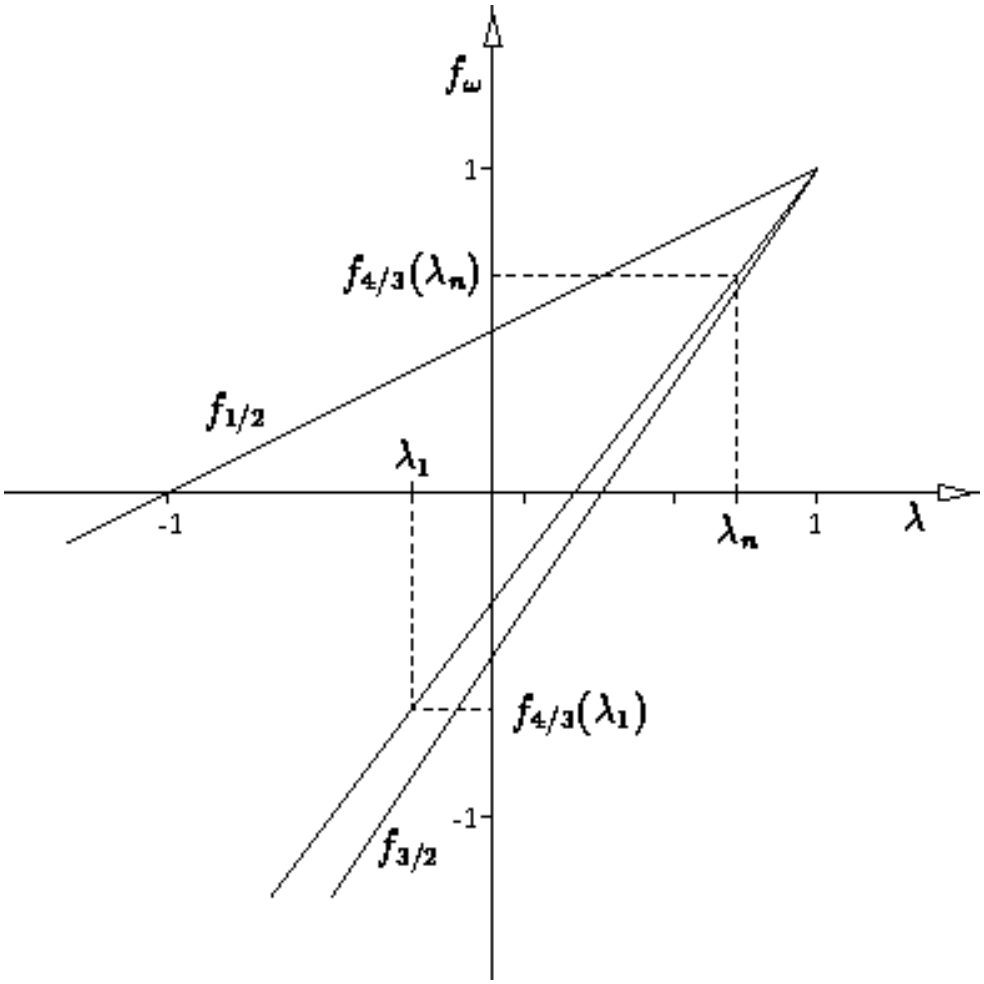
1. $M_J = D^{-1}(D - A)$ possess exclusively real eigenvalues
$$\lambda_1 \leq \cdots \leq \lambda_n.$$
2. Corresp. eigenvectors $u_1, \cdots, u_n$ are linear independent
3. It hold $\rho(M_J) < 1$.

Then:

1. Eigenvalues of the Jacobi relaxation method read:
$$\mu_i = 1 - \omega + \omega\lambda_i$$
2. Optimal relaxation parameter is given by:
$$\omega_{opt} = \frac{2}{2 - \lambda_1 - \lambda_n} > 0$$

Jacobi relaxation method

Idea to obatin $\omega_{opt.}$: $\mu_i = 1 - \omega + \omega\lambda_i, \quad \mu = f(\lambda) = 1 - \omega + \omega\lambda$



- Choose ω such that $\mu_1 = -\mu_n$.
- Improvement is only possible if $\lambda_1 \neq -\lambda_n$.

Jacobi relaxation method

Model problem:

$$A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix} \Longrightarrow M_J = \begin{pmatrix} 0 & 4/7 \\ 2/5 & 0 \end{pmatrix}$$

$$\Longrightarrow \qquad \lambda_{1,2} = \pm\sqrt{\tfrac{8}{35}}$$

$$\Longrightarrow \qquad \text{No improvement}$$

Appraisement

"$\circ$" : Minor assumptions on A

"+" : Very simple calculation of matrix-vector products $\omega D^{-1} x$

"$\circ$" : often no improvement, *MGM*

SOR method (successive overrelaxation)

Basic principle:

Pointwise formulation of the Gauß-Seidel method

$$x_{m+1,i} = x_{m,i} + \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_{m+1,j} - \sum_{j=i}^{n} a_{ij} x_{m,j} \right)$$

Proceeding

Weighting of the correction part

$$x_{m+1,i} = (1 - \omega) x_{m,i} + \frac{\omega}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_{m+1,j} - \sum_{j=i+1}^{n} a_{ij} x_{m,j} \right)$$

SOR method (successive overrelaxation)

Pointwise formulation:

$$x_{m+1,i} = (1-\omega)x_{m,i} + \frac{\omega}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_{m+1,j} - \sum_{j=i+1}^{n} a_{ij}x_{m,j}\right)$$

Matrix formulation:

$$\underbrace{(I+\omega D^{-1}L)}_{=D^{-1}(D+\omega L)}x_{m+1} = \underbrace{\left((1-\omega)I-\omega D^{-1}R\right)}_{=D^{-1}((1-\omega)D-\omega R)}x_m + \omega D^{-1}b$$

$$\Rightarrow \quad (D+\omega L)x_{m+1} = \left((1-\omega)D-\omega R\right)x_m + \omega b$$

$$\Rightarrow \quad x_{m+1} = \underbrace{(D+\omega L)^{-1}\left((1-\omega)D-\omega R\right)}_{=M_{GS}(\omega)}x_m + \underbrace{\omega(D+\omega L)^{-1}}_{N_{GS}(\omega)}b$$

SOR method (successive overrelaxation)

Restrictions on the relaxation parameter:

The method is divergent for all $\omega \notin (0, 2)$.

Reason:

$$\begin{aligned}\prod_{i=1}^{n} \lambda_i &= \det M_{GS}(\omega) \\ &= \det(D + \omega L)^{-1} \cdot \det((1-\omega)D - \omega R) \\ &= \det D^{-1} \cdot \det((1-\omega)D) \\ &= (1-\omega)^n\end{aligned}$$

$$\Longrightarrow \qquad \underbrace{\max_{i=1,\ldots,n} |\lambda_i|}_{=\rho(M_{GS}(\omega))} \geq |1-\omega|$$

SOR method (successive overrelaxation)

Optimal relaxation parameter

Assumptions:

1. A is consistently ordered.
2. Eigenvalues $\lambda_1, \ldots, \lambda_n$ of M_J are real.
3. $\rho := \rho(M_J) < 1$

Then:

1. The method is convergent for all $\omega \in (0, 2)$
2. $\omega_{opt} = \dfrac{2}{1 + \sqrt{1 - \rho^2}}$, $\rho(M_{GS}(\omega_{opt})) = \omega_{opt} - 1$
3. $\lambda_i = \dfrac{\mu_i + \omega - 1}{\omega \mu_i^{1/2}}$ with the eigenvalue μ_i of $M_{GS}(\omega)$

SOR method (successive overrelaxation)

Definition: Consistently ordered

The matrix $A = D + L + R$ with a non-singular diagonal part D is called consistently ordered, if the eigenvalues of

$$C(\alpha) = -(\alpha D^{-1} L + \alpha^{-1} D^{-1} R) \quad , \quad \alpha \in \mathbb{C} \backslash \{0\}$$

are independent of α.

SOR method (successive overrelaxation)

Examples:

- Each 2×2 matrix is consistently ordered.
- Each tridiagonal matrix

$$A = \begin{pmatrix} a_1 & b_1 & & & \\ c_2 & a_2 & b_2 & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & b_{n-1} \\ & & & c_n & a_n \end{pmatrix}$$

  is consistently ordered.
- Each matrix of the form

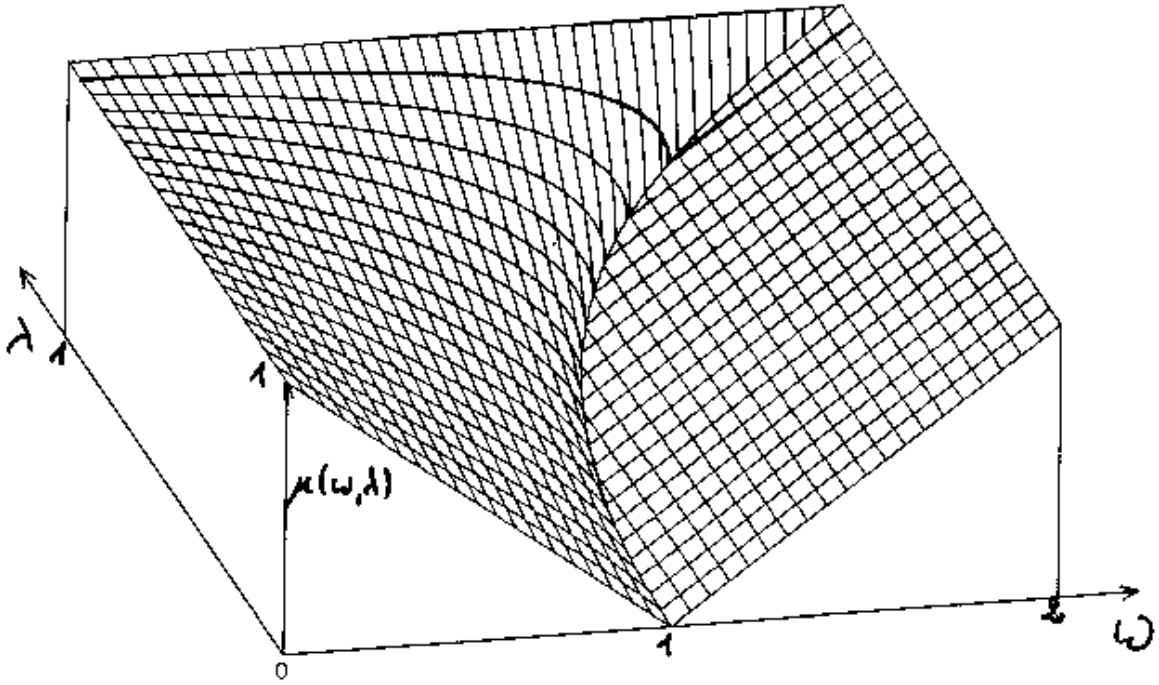
$$A = \begin{pmatrix} I & A_{12} \\ A_{21} & I \end{pmatrix}$$

  is consistently ordered.

SOR method (successive overrelaxation)

"Distribution of the eigenvalues "of the relaxation method

$$\mu(\omega,\lambda) = \begin{cases} \frac{1}{2}\lambda^2\omega^2 - (\omega-1) + \lambda\omega\sqrt{\frac{1}{4}\lambda^2\omega^2 - (\omega-1)} & , \ 0 < \omega < \omega^\star(\lambda) \\ \omega - 1 & , \ \omega^\star(\lambda) \leq \omega < 2 \end{cases}$$

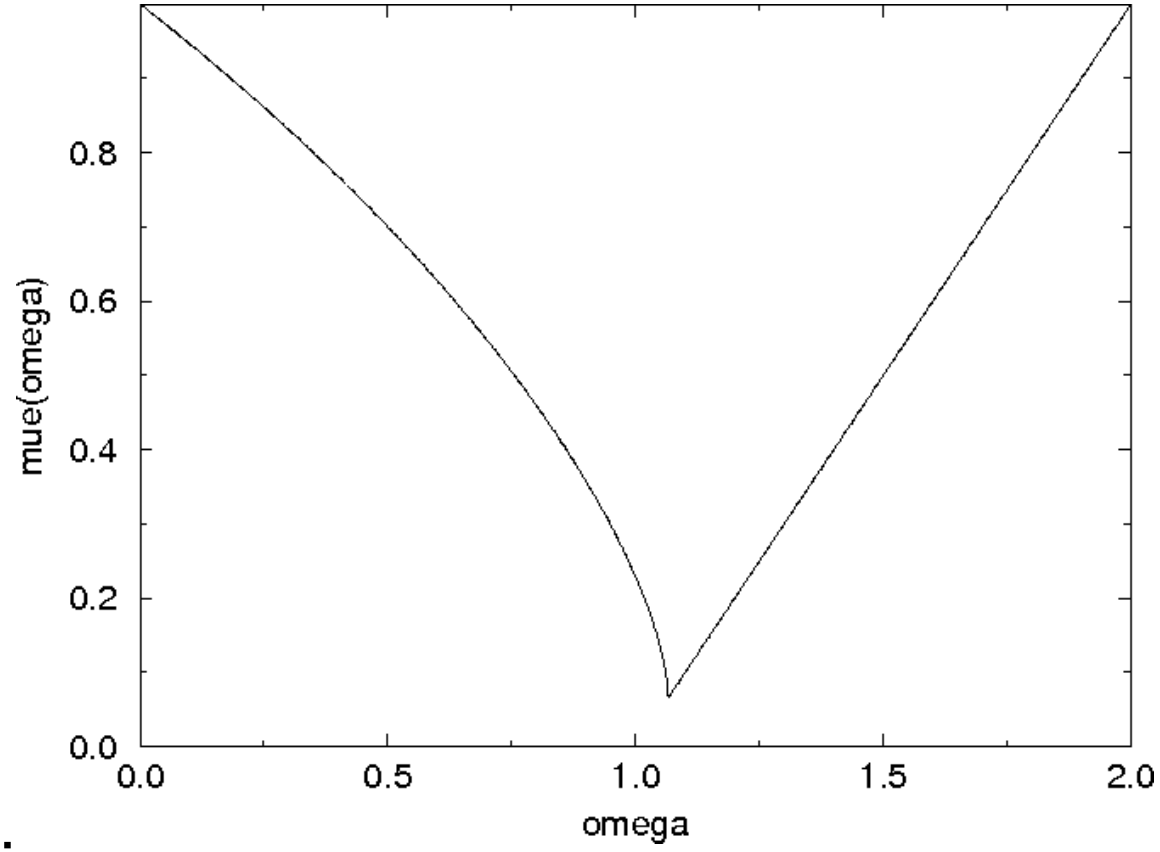


[Bunse, Bunse-Gerstner: Numerische Lineare Algebra]

SOR method (successive overrelaxation)

Relaxation scheme

Spectral radius versus relaxation parameter



Rule of thumb:

- Better to choose ω a bit larger than ω_{opt} instead of a bit smaller than ω_{opt}.

Model problem

$$Ax = b \text{ with } A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

1. A is consistently ordered.
2. The eigenvalues of $M_J = -D^{-1}(L+R)$ are $\lambda_{1,2} = \pm\sqrt{\frac{8}{35}} \in \mathbb{R}$
3. $\rho(M_J) < 1$

$$\Longrightarrow \qquad \omega_{opt} = \frac{2}{1+\sqrt{1-\frac{8}{35}}} \approx 1.06479$$

$$\rho(M_{GS}(\omega_{opt})) = \omega_{opt} - 1 \approx 0.06479 \approx 0.25^2 \approx \rho(M_{GS})^2$$

Expectation:

Approximately half as much iterations as the Gauß-Seidel method to reach the same accuracy

Model problem

$$Ax = b \text{ with } A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

1. A is consistently ordered.
2. The eigenvalues of $M_J = -D^{-1}(L+R)$ are $\lambda_{1,2} = \pm\sqrt{\frac{8}{35}} \in \mathbb{R}$
3. $\rho(M_J) < 1$

$$\Longrightarrow \qquad \omega_{opt} = \frac{2}{1+\sqrt{1-\frac{8}{35}}} \approx 1.06479$$

$$\rho(M_{GS}(\omega_{opt})) = \omega_{opt} - 1 \approx 0.06479 \approx 0.25^2 \approx \rho(M_{GS})^2$$

Expectation:

Approximately half as much iterations as the Gauß-Seidel method to reach the same accuracy

Model problem

$$Ax = b \text{ with } A = \begin{pmatrix} 0.7 & -0.4 \\ -0.2 & 0.5 \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$$

1. A is consistently ordered.
2. The eigenvalues of $M_J = -D^{-1}(L+R)$ are $\lambda_{1,2} = \pm\sqrt{\frac{8}{35}} \in \mathbb{R}$
3. $\rho(M_J) < 1$

$$\Longrightarrow \qquad \omega_{opt} = \frac{2}{1+\sqrt{1-\frac{8}{35}}} \approx 1.06479$$

$$\rho(M_{GS}(\omega_{opt})) = \omega_{opt} - 1 \approx 0.06479 \approx 0.25^2 \approx \rho(M_{GS})^2$$

Expectation:

Approximately half as much iterations as the Gauß-Seidel method to reach the same accuracy

SOR method (successive overrelaxation)

Model problem:

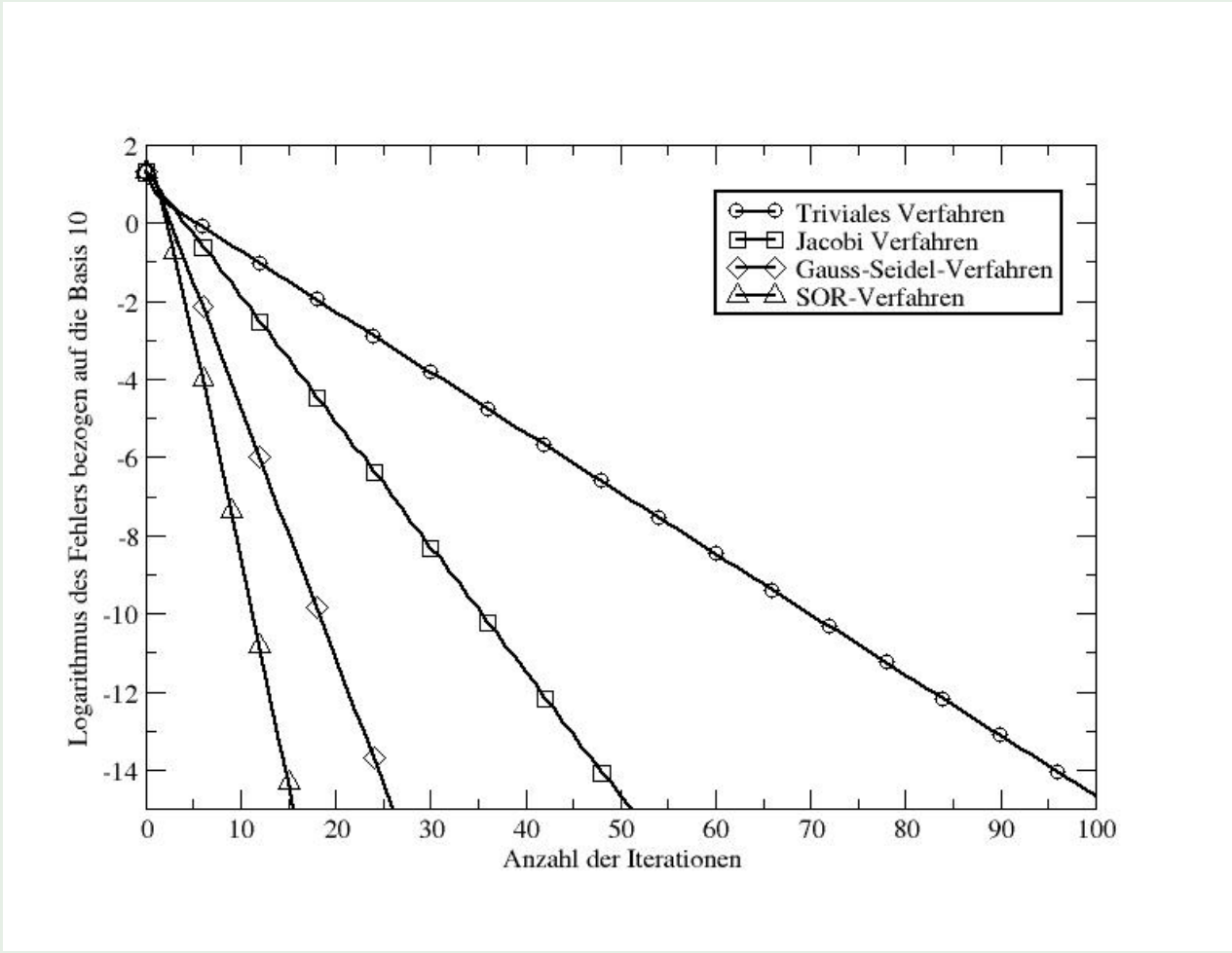


Abbildung: Convergence history $\log_{10} \varepsilon_m$ of the SOR method

SOR method (successive overrelaxation)

SOR method				
m	$x_{m,1}$	$x_{m,2}$	$\varepsilon_m := \|x_m - x^*\|_\infty$	$\varepsilon_m / \varepsilon_{m-1}$
0	2.100000e+01	-1.900000e+01	2.000000e+01	
1	-1.246473e+01	-3.439090e+00	1.346473e+01	6.732366e-01
2	-8.286241e-01	5.087570e-01	1.828624e+00	1.358084e-01
3	8.195743e-01	9.549801e-01	1.804257e-01	9.866748e-02
4	9.842969e-01	9.962285e-01	1.570309e-02	8.703354e-02
5	9.987226e-01	9.997003e-01	1.277401e-03	8.134709e-02
6	9.999004e-01	9.999770e-01	9.960642e-05	7.797587e-02
7	9.999925e-01	9.999983e-01	7.544695e-06	7.574507e-02
8	9.999994e-01	9.999999e-01	5.595127e-07	7.415974e-02
9	1.000000e-00	1.000000e-00	4.083051e-08	7.297514e-02
10	1.000000e-00	1.000000e-00	2.942099e-09	7.205638e-02
11	1.000000e-00	1.000000e-00	2.098393e-10	7.132298e-02
12	1.000000e-00	1.000000e-00	1.484068e-11	7.072406e-02
13	1.000000e-00	1.000000e-00	1.042055e-12	7.021612e-02
14	1.000000e-00	1.000000e-00	7.260859e-14	6.967824e-02
15	1.000000e-00	1.000000e-00	4.884981e-15	6.727829e-02

SOR method (successive overrelaxation)

Appraisement:

"$\circ$" : Minor assumptions on the matrix A

($a_{ii} \neq 0$ für $i = 1, \ldots, n$)

"+" : Simple calculation of matrix-vector products $(D + \omega L)^{-1} x$

"++" : Fast convergence behaviour

Summary:

Splitting methods:

- Easy to derive
- Simple to implement

Rate of convergence:

- Is determined by $\rho(M) = \rho(B^{-1}(B - A))$.
- Rule of thumb: Choose the approximation B as close as possible w.r.t. A to obtain a good convergence behaviour.

Summary:

Typs of Splitting methods

- Trivial scheme: $B = I$ (bad rate of convergence)
- Jacobi method : $B = D$
- Gauß-Seidel method : $B = D + L$
- Relaxation methods
  - Weighting of the correction vector
  - Optimizing $\rho(M(\omega))$