Chapter 3

Detailed Investigation of Classical Iterative Schemes

3.1 General Aspects of Classical Iterative Schemes

Remark 3.1 *Motivation.* This chapter studies the reason for the inefficient behavior, observed in Example 2.5 for SSOR, of classical iterative schemes in detail. This study is performed for the one-dimensional problem (2.1) and the linear system of equations has the form

$$A\mathbf{u} = \mathbf{f}. \tag{3.1}$$

□

Remark 3.2 *General approach.* Classical iterative schemes for the solution of (3.1) were introduced and studied in Numerical Mathematics II. Here, a short review is presented and notations are introduced.

Classical iterative schemes are based on a fixed point iteration for solving the linear system of equations. To this end, decompose the matrix

$$A = M - N, \tag{3.2}$$

where M is a non-singular matrix. Then, one can write system (3.1) in the fixed point form

$$M\mathbf{u} = N\mathbf{u} + \mathbf{f}$$

or

$$\mathbf{u} = M^{-1}N\mathbf{u} + M^{-1}\mathbf{f} =: S\mathbf{u} + M^{-1}\mathbf{f}.$$

Given an initial iterate $\mathbf{u}^{(0)}$, a fixed point iteration can be applied to this equation

$$\mathbf{u}^{(m+1)} = S\mathbf{u}^{(m)} + M^{-1}\mathbf{f}, \quad m = 0, 1, 2, \ldots \tag{3.3}$$

This basic iterative approach might be also damped

$$\mathbf{u}^* = S\mathbf{u}^{(m)} + M^{-1}\mathbf{f}, \quad \mathbf{u}^{(m+1)} = \omega\mathbf{u}^* + (1-\omega)\mathbf{u}^{(m)}, \quad \omega \in \mathbb{R}^+,$$

such that

$$\mathbf{u}^{(m+1)} = \left(\omega S + (1-\omega)I\right)\mathbf{u}^{(m)} + \omega M^{-1}\mathbf{f}. \tag{3.4}$$

□

Remark 3.3 *The residual equation.* Let $\mathbf{u}$ be the solution of (3.1) and $\mathbf{u}^{(m)}$ an approximation computed with (3.3) or (3.4). The error is denoted by

$$\mathbf{e}^{(m)} = \mathbf{u} - \mathbf{u}^{(m)}$$

and the residual by

$$\mathbf{r}^{(m)} = \mathbf{f} - A\mathbf{u}^{(m)}.$$

It is for the fixed point iteration (3.3)

$$\begin{aligned} S\mathbf{e}^{(m)} &= M^{-1}N\mathbf{u} - S\mathbf{u}^{(m)} = M^{-1}N\mathbf{u} - \mathbf{u}^{(m+1)} + M^{-1}\mathbf{f} \\ &= M^{-1}\left(N\mathbf{u} + \mathbf{f}\right) - \mathbf{u}^{(m+1)} = \mathbf{u} - \mathbf{u}^{(m+1)} = \mathbf{e}^{(m+1)}. \end{aligned} \tag{3.5}$$

For both iterations (3.3) and (3.4), the so-called residual equation has the form

$$A\mathbf{e}^{(m)} = A\mathbf{u} - A\mathbf{u}^{(m)} = \mathbf{f} - A\mathbf{u}^{(m)} = \mathbf{r}^{(m)}. \tag{3.6}$$

□

Remark 3.4 *To multigrid methods.* In multigrid methods, the residual equation (3.6) is used for updating the current iterate $\mathbf{u}^{(m)}$. An approximation $\tilde{\mathbf{e}}^{(m)}$ of $\mathbf{e}^{(m)}$ is computed from (3.6) and the new iterate is given by $\mathbf{u}^{(m+1)} = \mathbf{u}^{(m)} + \tilde{\mathbf{e}}^{(m)}$. An advantage of using the residual equation is that, at least close to the solution, $\mathbf{e}^{(m)}$ is small and the zero vector is a good initial guess for an iterative solution of (3.6).
□

Remark 3.5 *To the convergence of classical iteration schemes.* From (3.5), it follows by induction that $\mathbf{e}^{(m)} = S^m\mathbf{e}^{(0)}$, such that

$$\left\|\mathbf{e}^{(m)}\right\| \leq \|S^m\| \left\|\mathbf{e}^{(0)}\right\| \tag{3.7}$$

for each vector norm and its induced matrix norm. The iteration is called convergent if

$$\lim_{m\to\infty} \|S^m\| = 0$$

and $\|S^m\|$ is called contraction number of the fixed point iteration (3.3). It was shown in the course Numerical Mathematics II, Theorem 3.3 in the part on iterative solvers, that the fixed point iteration (3.3) converges for any initial iterate if and only if $\rho(S) < 1$, where $\rho(S) = \max_i |\lambda_i(S)|$ is the spectral radius of S. In connection with iterative schemes, the spectral radius is also called convergence factor. It is loosely speaking the worst factor for the reduction of the error in each step of the iteration.

For each eigenvalue $\lambda_i \in \mathbb{C}$ of a matrix $A \in \mathbb{R}^{n\times n}$ it is $|\lambda_i| \leq \|A\|$, where $\|\cdot\|$ is any induced matrix norm. It follows that $\rho(S) \leq \|S\|$.

Let $M \in \mathbb{N}$ be the smallest natural number for which

$$\frac{\left\|\mathbf{e}^{(M)}\right\|}{\left\|\mathbf{e}^{(0)}\right\|} \leq 10^{-1},$$

i.e., the smallest number of iterations which are needed for reducing the error by the factor 10. This condition is satisfied approximately if

$$\frac{\left\|\mathbf{e}^{(M)}\right\|}{\left\|\mathbf{e}^{(0)}\right\|} \leq \left\|S^M\right\| \approx \rho\left(S^M\right) = \left(\rho(S)\right)^M \approx 10^{-1}.$$

It follows that

$$M \lesssim -\frac{1}{\log_{10}|\rho(S)|}.$$

The number $-\log_{10}|\rho(S)|$ is called rate of convergence. If it is close to zero, i.e., $\rho(S)$ is close to one, then M is large and the convergence is slow. The convergence becomes the faster the closer $\rho(S)$ is to zero. □

3.2 The Jacobi and Damped Jacobi Method

Remark 3.6 *The Jacobi and the damped Jacobi method.* The Jacobi method is given by $M = \text{diag}(A) = D$ in (3.2). A straightforward calculation shows, see also Numerical Mathematics II, that it has the form

$$\mathbf{u}^{(m+1)} = \mathbf{u}^{(m)} + D^{-1}\mathbf{r}^{(m)}.$$

Also the damped Jacobi method was introduced in Numerical Mathematics II

$$\mathbf{u}^{(m+1)} = \mathbf{u}^{(m)} + \omega D^{-1}\mathbf{r}^{(m)}, \quad \omega \in (0, 1]. \tag{3.8}$$

A straightforward calculation shows that it can be written as a basic fixed point iteration (3.3) with $M = \omega^{-1}\text{diag}(A)$. The behavior of the (damped) Jacobi method shall be studied at the one-dimensional model problem (2.1). □

Remark 3.7 *Discrete Fourier modes.* To study the behavior of the (damped) Jacobi method for the one-dimensional model problem, it is sufficient to consider the homogeneous linear system of equations

$$A\mathbf{u} = \mathbf{0} \tag{3.9}$$

and an arbitrary initial iterate $\mathbf{u}^{(0)}$. The solution of the homogeneous system is $\mathbf{u} = \mathbf{0}$. Obviously, the matrix from the finite element discretization can be used without loss of generality.

Let b be a given integrable function in $[0, 1]$ with $b(0) = b(1) = 0$. This function can be expanded in the form

$$b(x) = \sum_{k=1}^{\infty} b_k \sin(k\pi x),$$

where k is the wave number, b_k is the k-th Fourier coefficient, the functions $\sin(k\pi x)$ are called Fourier modes, and the frequency is $1/k$. Small wave numbers characterize long and smooth waves, whereas large wave numbers describe highly oscillating waves, see Figure 3.1.

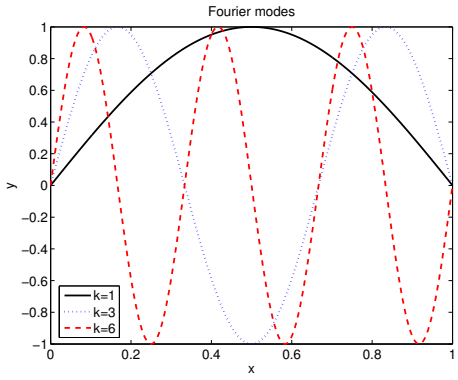

Figure 3.1: Fourier modes.

For the investigation of an iterative method applied to the solution of (3.9), it is of advantage to consider initial iterates (which are equal to the negative of the initial errors) which are discrete, i.e., at $x = j/N$, analogs of the Fourier modes:

$$\mathbf{u}^{(0)} = \left(u_1^{(0)}, \ldots, u_{N-1}^{(0)}\right)^T \quad \text{with} \quad u_j^{(0)} = \sin\left(\frac{jk\pi}{N}\right), \quad j, k = 1, \ldots, N-1. \tag{3.10}$$

Note that these discrete Fourier modes are also the eigenvectors of the matrix A, see (2.7).

The discrete Fourier modes in the lower part of the spectrum $1 \leq k < N/2$ are called low frequency or smooth modes. The modes in the upper part of the spectrum $N/2 \leq k \leq N-1$ are the so-called high frequency or oscillating modes. Note that the classification of the discrete modes depends on the number of nodes N. The discrete analogs of the Fourier modes have different properties on different grids. □

Example 3.8 *Application of the damped Jacobi method for the solution of the model problem.* The damped Jacobi method (3.8) with $\omega = 2/3$ is applied to the solution of the model in the following two situations:

- the number of intervals N is fixed and the wave number k is varied,
- the wave number k is fixed and the number of intervals N is varied.

For each simulation, 100 iterations were performed and the error is measured in the l^∞ vector norm $\|\cdot\|_\infty$. The obtained results are presented in Figures 3.2 and 3.3

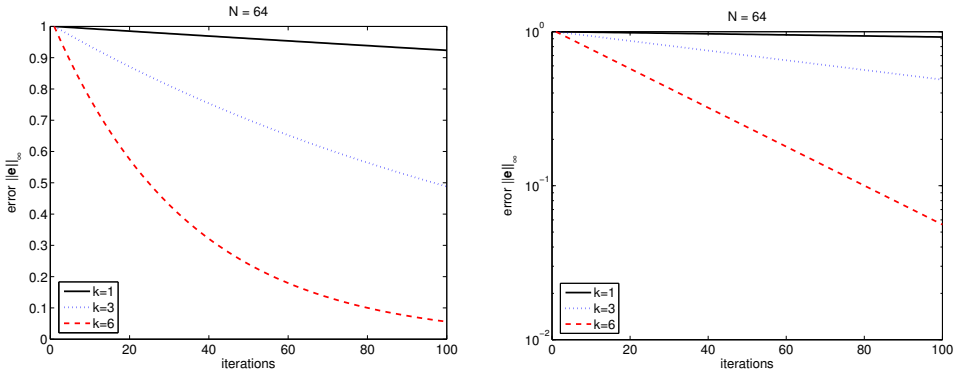

Figure 3.2: Convergence of the damped Jacobi method with $\omega = 2/3$ for initial iterates with different wave numbers on a fixed grid, left linear plot, right semilogarithmic plot.

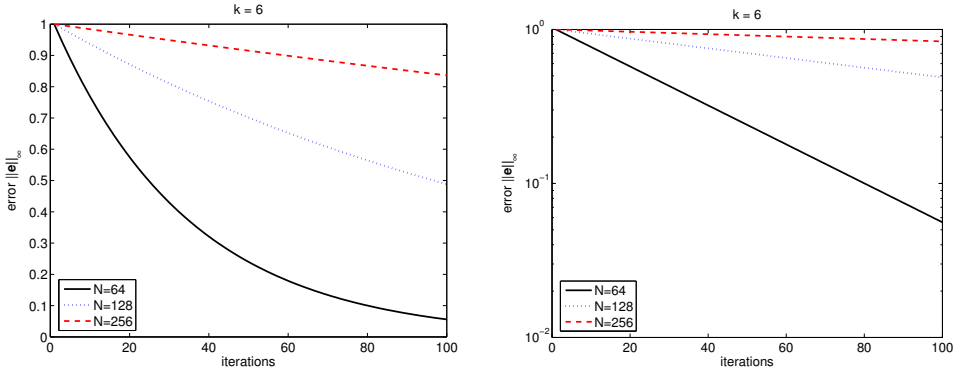

Figure 3.3: Convergence of the damped Jacobi method with $\omega = 2/3$ on different grids for an initial iterate with a fixed wave number, left linear plot, right semilogarithmic plot.

The following observations are of importance:

- On a fixed grid, there is a good damping of the high frequency errors whereas there is almost no damping of the low frequency errors (note that $\mathbf{u}^{(0)} = \mathbf{e}_j^{(0)}$).

- For a fixed wave number, the error is reduced on a coarser grid better than on a finer grid.
- The logarithm of the error decays linearly, i.e., the error itself decays geometrically. Thus, there is a constant $0 < C(k) < 1$ such that

$$\left\|\mathbf{e}^{(n)}\right\|_\infty \leq (C(k))^n \left\|\mathbf{e}^{(0)}\right\|_\infty .$$

In practice, one does not prescribe the number of iterations to be performed but the scheme is applied until the computed solution satisfies a certain criterion with respect to its accuracy. For $k = 6$ and the stopping criterion $\left\|\mathbf{e}^{(m)}\right\|_\infty = \left\|\mathbf{u}^{(m)}\right\|_\infty < 10^{-6}$, the number of iterations are given in Table 3.1. One can see that refining the mesh once, thus halving the mesh width and doubling the number of unknowns, then the number of iterations increases by the factor four.

Table 3.1: Convergence of the damped Jacobi method with $\omega = 2/3$ for the initial iterate with wave number $k = 6$.

N	no. of iterations
16	27
32	116
64	475
128	1908
256	7642
512	30576
1024	122314

□

Remark 3.9 *Analytical considerations of the damped Jacobi method.* The iteration matrix of the damped Jacobi method (3.8) has the form

$$S_{\mathrm{jac},\omega} = I - \omega D^{-1} A = I - \frac{\omega h}{2} A, \tag{3.11}$$

where the diagonal of the finite element system matrix has been inserted. The convergence of the damped Jacobi method is determined by the eigenvalues of the iteration matrix. From the special form of this matrix, one can see that

$$\lambda_k(S_{\mathrm{jac},\omega}) = 1 - \frac{\omega h}{2}\lambda_k(A) = 1 - 2\omega \sin^2\left(\frac{k\pi h}{2}\right), \quad k = 1, \ldots, N-1, \tag{3.12}$$

where (2.6) and $h = 1/N$ have been used. The eigenvectors $\mathbf{v}_k$ of A, see (2.7), are the same as the eigenvectors of $S_{\mathrm{jac},\omega}$, *exercise.* □

Lemma 3.10 Convergence of the damped Jacobi method. *The damped Jacobi method converges for the one-dimensional model problem for all initial iterates if $\omega \in (0, 1]$. The method converges fastest for $\omega = 1$.*

Proof: From Numerical Mathematics II it is known that the method converges for all initial iterates if and only if the spectral radius of the iteration matrix $\rho(S_{\mathrm{jac},\omega})$ is smaller than 1. Since it is

$$0 < \sin^2\left(\frac{k\pi h}{2}\right) = \sin^2\left(\frac{k\pi}{2N}\right) < 1, \quad \text{for } k = 1, \ldots, N-1,$$

it follows from (3.12) that $\lambda_k(S_{\mathrm{jac},\omega}) \in (-1, 1)$ for $k = 1, \ldots, N-1$, and $\omega \in (0, 1]$. Hence it is $\rho(S_{\mathrm{jac},\omega}) < 1$.

It is also known from Numerical Mathematics II that the method converges the faster the smaller $\rho(S_{\mathrm{jac},\omega})$ is, i.e., one has to solve *(exercise)*

$$\min_{\omega\in(0,1]}\max_{k=1,\ldots,N-1}\left|1-2\omega\sin^2\left(\frac{k\pi}{2N}\right)\right|.$$

■

Remark 3.11 *General effect of the damped Jacobi method on the discrete Fourier modes.* For studying the effect of the damped Jacobi methods on the discrete Fourier modes, see (3.10), an arbitrary initial error $\mathbf{e}^{(0)}$ will be represented with respect to the basis $\{\mathbf{w}_1,\ldots,\mathbf{w}_{N-1}\}$, where $\mathbf{w}_1,\ldots,\mathbf{w}_{N-1}$ are the eigenvectors of $S_{\mathrm{jac},\omega}$ and A,

$$\mathbf{e}^{(0)}=\sum_{k=1}^{N-1}c_k\mathbf{w}_k,\quad c_k\in\mathbb{R}.$$

Since the damped Jacobi method can be written in form (3.3), it follows from (3.5) that

$$\mathbf{e}^{(m)}=S_{\mathrm{jac},\omega}^m\mathbf{e}^{(0)}.$$

Using the property of $\mathbf{w}_k$ being an eigenvector of $S_{\mathrm{jac},\omega}$, one obtains

$$\begin{aligned}\mathbf{e}^{(m)} &= \sum_{k=1}^{N-1}c_kS_{\mathrm{jac},\omega}^m\mathbf{w}_k=\sum_{k=1}^{N-1}c_kS_{\mathrm{jac},\omega}^{m-1}\left(S_{\mathrm{jac},\omega}\mathbf{w}_k\right)=\sum_{k=1}^{N-1}c_kS_{\mathrm{jac},\omega}^{m-1}\lambda_k\left(S_{\mathrm{jac},\omega}\right)\mathbf{w}_k\\ &= \ldots=\sum_{k=1}^{N-1}c_k\lambda_k^m\left(S_{\mathrm{jac},\omega}\right)\mathbf{w}_k.\end{aligned}$$

This calculation shows that after m iterations, the initial error with respect to the k-th discrete Fourier mode is reduced by the factor $\lambda_k^m\left(S_{\mathrm{jac},\omega}\right)$. If $|\lambda_k^m\left(S_{\mathrm{jac},\omega}\right)|$ is close to 1, then the reduction will be small. A strong reduction will occur if $|\lambda_k^m\left(S_{\mathrm{jac},\omega}\right)|$ is close to zero. □

Remark 3.12 *Effect on the smooth error modes.* Using (3.12), one finds that

$$\begin{aligned}\lambda_k\left(S_{\mathrm{jac},\omega}\right)\approx 1 &\iff \sin^2\left(\frac{k\pi}{2N}\right)\approx 0 \iff k \text{ small},\\ \lambda_k\left(S_{\mathrm{jac},\omega}\right)\approx -1 &\iff \omega\sin^2\left(\frac{k\pi}{2N}\right)\approx 1 \iff \omega\approx 1 \text{ and } k \text{ close to } N.\end{aligned}$$

For $\lambda_1\left(S_{\mathrm{jac},\omega}\right)$, one finds with a Taylor series expansion

$$\lambda_1\left(S_{\mathrm{jac},\omega}\right)=1-2\omega\sin^2\left(\frac{\pi h}{2}\right)\approx 1-2\omega\frac{h^2\pi^2}{4}=1-\omega\frac{h^2\pi^2}{2}.$$

This eigenvalue is close to 1 for all damping parameters $\omega\in(0,1]$. Hence, there is no choice of the damping parameter which results in an efficient damping of the smooth error modes connected to $\mathbf{w}_1$. In addition, $\lambda_1\left(S_{\mathrm{jac},\omega}\right)$ is the closer to 1 the finer the grid is. It follows that refining the grid results in a worse convergence with respect of the smooth error modes. □

Remark 3.13 *Effect on the oscillating error modes.* The distribution of the eigenvalues of the iteration matrix for $\omega\in\{1,2/3,1/2\}$ and $N=16$ is presented in Figure 3.4. As it was observed in the previous remark, none of the damping parameters gives a method that reduces the smooth error modes efficiently. In can be seen in Figure 3.4, using the damping parameter $\omega=1$, the method does not damp

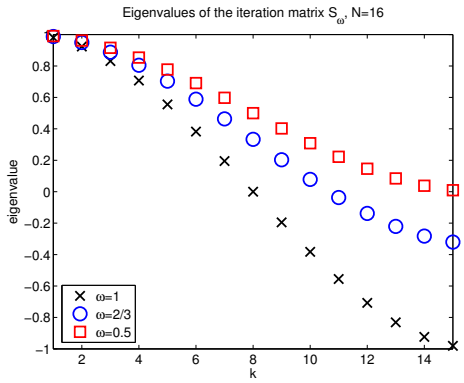

Figure 3.4: Eigenvalues of the iteration matrix $S_{\mathrm{jac},\omega}$ of the damped Jacobi method for different values of the damping parameter, $N = 16$.

efficiently the oscillating error modes neither, but it damps efficiently some intermediate error modes. The situation is much different for the damping parameter $\omega = 1/2$. For this parameter, it can be observed that the oscillating error modes are damped efficiently.

The situation as it occurs for $\omega = 1/2$ is of advantage, since it allows to distinguish clearly between the low and the high frequencies. With the damped Jacobi method and appropriate damping parameters, there is an iterative scheme that damps the high frequencies fast and the low frequencies slowly. Now, one needs another method with complementary properties to combine both methods. The construction of the complementary method is the goal of multigrid methods. □

Example 3.14 *Optimal damping parameter for the oscillating modes.* The damping parameter ω has to be determined such the one finds the smallest interval $[-\bar{\lambda}, \bar{\lambda}]$ with $\lambda_k\left(S_{\mathrm{jac},\omega}\right) \in [-\bar{\lambda}, \bar{\lambda}]$ for $k = N/2, \ldots, N-1$. This goal is achieved with $\omega = 2/3$. In this case it is, using the monotonicity of the sine function,

$$\frac{4}{3} \geq \frac{4}{3}\sin^2\left(\frac{k\pi}{2N}\right) \geq \frac{4}{3}\sin^2\left(\frac{N\pi}{4N}\right) = \frac{4}{3}\cdot\frac{1}{2} = \frac{2}{3}.$$

One gets

$$\max_{k\geq N/2}\left|\lambda_k\left(S_{2/3}\right)\right| = \max_{k\geq N/2}\left|1 - \frac{4}{3}\sin^2\left(\frac{k\pi}{2N}\right)\right| \leq \frac{1}{3},$$

see also Figure 3.4. It follows that the oscillating error modes are reduced in each iteration at least by the factor three. This damping rate for the oscillating error modes is called smoothing rate of the method. As one can see, the smoothing rate is for the damped Jacobi method (with fixed ω) independent of the fineness of the grid. □

Remark 3.15 *On the multigrid idea.* Consider a fixed Fourier mode $\sin(k\pi x)$ and its discrete representation $\sin(jk\pi/N)$, $j = 1, \ldots, N-1$. As already noted at the end of Remark 3.7, the classification of this mode depends on the fineness of the grid:

- If the grid is sufficiently fine, i.e., $k < N/2$, it is a smooth mode and it can be damped only slowly with the damped Jacobi method.
- If the grid is sufficiently coarse, i.e., $N/2 \leq k \leq N-1$, it is an oscillating mode and can be damped quickly with the damped Jacobi method.

From this observation, one can already derive the multigrid idea. On a fine grid, only the oscillating error modes on this grid are damped. The smooth modes on this grid are oscillating on coarser grids and they will be reduced on these grids. □

3.3 The Gauss–Seidel Method and the SOR Method

Remark 3.16 *The Gauss–Seidel Method and the SOR Method.* The Gauss–Seidel method and the SOR (successive over relaxation) method were also already introduced and studied in Numerical Mathematics II. Decompose the system matrix of (3.1) into

$$A = D + L + U,$$

where D is the diagonal, L is the strict lower part, and U is the strict upper part. The Gauss–Seidel method is obtained with $M = D + L$ and $N = -U$ in the fixed point method (3.3)

$$\begin{aligned} \mathbf{u}^{(m+1)} &= \underbrace{-(D+L)^{-1}U}_{S_{\mathrm{GS}}}\mathbf{u}^{(m)} + (D+L)^{-1}\mathbf{f} \\ &= \mathbf{u}^{(m)} + (D+L)^{-1}\left(\mathbf{f} - A\mathbf{u}^{(m)}\right) \quad m = 0, 1, 2, \ldots. \end{aligned}$$

Using $M = \omega^{-1}D + L$ and $N = \omega^{-1}D - (D - U)$ gives the SOR method. This method can be written in the form

$$\begin{aligned} \mathbf{u}^{(m+1)} &= \mathbf{u}^{(m)} + \omega D^{-1}\left(\mathbf{f} - L\mathbf{u}^{(m+1)} - (D+U)\,\mathbf{u}^{(m)}\right) \\ &= \mathbf{u}^{(m)} + \left(\frac{D}{\omega} + L\right)^{-1}\left(\mathbf{f} - A\mathbf{u}^{(m)}\right), \quad m = 0, 1, 2, \ldots. \end{aligned}$$

In the case $\omega = 1$, the Gauss–Seidel method is recovered. Writing the method component-wise

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

one can see that for the computation of $\mathbf{u}^{(m+1)}$ not only the old iterate $\mathbf{u}^{(m)}$ is used, as in the damped Jacobi method, but one uses the already computed components of $\mathbf{u}^{(m+1)}$.

By the last property, one can say that the SOR method is somewhat more advanced than the damped Jacobi method. However, it will turn out that the SOR method shows a similar behavior for the solution of the model problem as the damped Jacobi method. □

Remark 3.17 *Properties of the SOR method.* The properties of the SOR method were studied in Numerical Mathematics II. They will be summarized here.

- Lemma of Kahan[1]. If the SOR method converges for every initial iterates $\mathbf{u}^{(0)} \in \mathbb{R}^n$ then $\omega \in (0, 2)$.
- If $A \in \mathbb{R}^{n\times n}$ a symmetric positive definite matrix. Then the SOR method converges for all initial iterates $\mathbf{u}^{(0)} \in \mathbb{R}^n$ if $\omega \in (0, 2)$.

The rate of convergence depends on ω. It can be shown that for a certain class of matrices, to which also the matrix obtained in the discretization of the model problem belongs to, there is an optimal value $\omega_{\mathrm{opt}} \in (1, 2)$. However, the determination of ω_{opt} is difficult in practice. For the model problem, one finds that ω_{opt} tends to 2 if the grids are refined, cf., Numerical Mathematics II, exercise problem 04/2.

The behavior of the SOR method depends on the numbering of the unknowns, which is in contrast to the damped Jacobi method. There are classes of problems where the efficiency of the SOR method depends essentially on the numbering of the unknowns. □

[1] William M. Kahan, born 1933

Example 3.18 *Application of the Gauss–Seidel method for the solution of the model problem.* The SOR method with $\omega = 1$ is studied in the same way as the damped Jacobi method in Example 3.8. The qualitatively behavior does not change for other values of the relaxation parameter. The numbering of the unknowns in the model problem is from left to right.

Figures 3.5 and 3.6 present the results. One can see that they are qualitatively the same as for the damped Jacobi method.

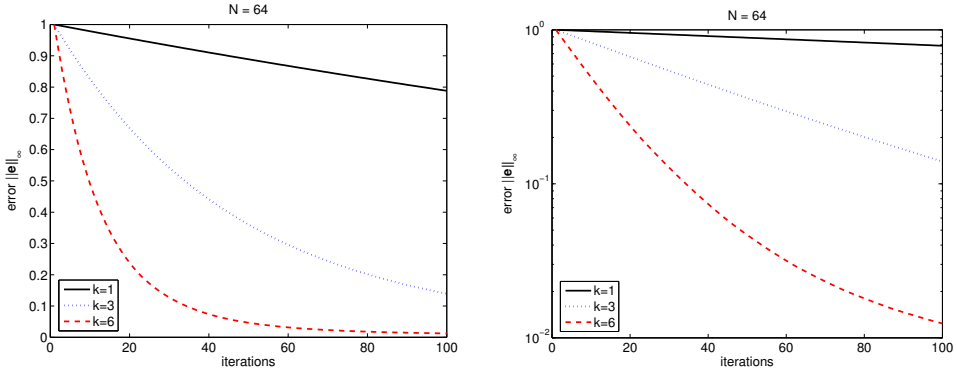

Figure 3.5: Convergence of the SOR method with $\omega = 1$ for initial iterates with different wave numbers on a fixed grid, left linear plot, right semilogarithmic plot.

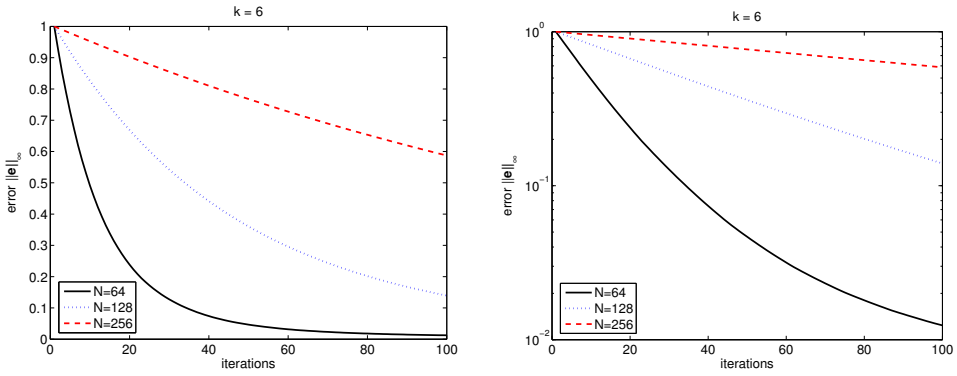

Figure 3.6: Convergence of the SOR method with $\omega = 1$ on different grids for an initial iterate with a fixed wave number, left linear plot, right semilogarithmic plot.

The number of iterations for $k = 6$ and the stopping criterion $\left\|\mathbf{e}^{(m)}\right\|_\infty = \left\|\mathbf{u}^{(m)}\right\|_\infty < 10^{-6}$ is presented in Table 3.2. Like for the Jacobi method, the number increases by the factor of four if the grid is refined once.

Altogether, one can draw for the SOR method the same conclusions as for the damped Jacobi method. □

Lemma 3.19 Some eigenvalues and eigenvectors of S_{GS}. *Let A be the matrix obtained by discretizing the model problem* (2.1) *with the finite element method. Then, some eigenvalues of the iteration matrix of the Gauss–Seidel method are given by*

$$\lambda_k\left(S_{\mathrm{GS}}\right) = \cos^2\left(\frac{k\pi}{N}\right), \quad k = 1, \ldots, N/2,$$

Table 3.2: Convergence of the SOR method with $\omega = 1$ for the initial iterate with wave number $k = 6$.

N	no. of iterations
16	274
32	1034
64	3859
128	14297
256	52595
512	191980
1024	–

and the corresponding eigenvectors are $\mathbf{w}_k = (w_{k,1}, \ldots, w_{k,N-1})^T$ *with*

$$w_{k,j} = (\lambda_k (S_{\mathrm{GS}}))^{j/2} \sin\left(\frac{jk\pi}{N}\right), \quad j = 1, \ldots, N-1.$$

Proof: One has to show that

$$S_{\mathrm{GS}}\mathbf{w}_k = \lambda_k (S_{\mathrm{GS}})\, \mathbf{w}_k, \quad k = 1, \ldots, N-1.$$

Inserting the decomposition of S_{GS} gives

$$-(D+L)^{-1} U\mathbf{w}_k = \lambda_k (S_{\mathrm{GS}})\, \mathbf{w}_k \quad \Longleftrightarrow \quad \lambda_k (S_{\mathrm{GS}})\, (D+L)\, \mathbf{w}_k = -U\mathbf{w}_k.$$

Considering the j-th component and using the special form of the matrices D, L, and U, see Example 2.2, one obtains

$$\lambda_k (S_{\mathrm{GS}}) \left(\frac{2}{h} w_{k,j} - \frac{1}{h} w_{k,j-1}\right) = \frac{1}{h} w_{k,j+1}.$$

Scaling this equation by h and inserting the representation of the k-th eigenvector yields

$$\begin{aligned}
&\lambda_k (S_{\mathrm{GS}}) \left[2 (\lambda_k (S_{\mathrm{GS}}))^{j/2} \sin\left(\frac{jk\pi}{N}\right) - (\lambda_k (S_{\mathrm{GS}}))^{(j-1)/2} \sin\left(\frac{(j-1)k\pi}{N}\right)\right] \\
&= (\lambda_k (S_{\mathrm{GS}}))^{(j+1)/2} \sin\left(\frac{(j+1)k\pi}{N}\right),
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&(\lambda_k (S_{\mathrm{GS}}))^{(j+1)/2} \left[2 (\lambda_k (S_{\mathrm{GS}}))^{1/2} \sin\left(\frac{jk\pi}{N}\right) - \sin\left(\frac{(j-1)k\pi}{N}\right)\right] \\
&= (\lambda_k (S_{\mathrm{GS}}))^{(j+1)/2} \sin\left(\frac{(j+1)k\pi}{N}\right).
\end{aligned}$$

Applying the formula for the eigenvalues, noting that $\cos(k\pi/N) \geq 0$ for $k = 1, \ldots, N/2$, gives

$$\begin{aligned}
&(\lambda_k (S_{\mathrm{GS}}))^{(j+1)/2} \left[2 \cos\left(\frac{k\pi}{N}\right) \sin\left(\frac{jk\pi}{N}\right) - \sin\left(\frac{(j-1)k\pi}{N}\right)\right] \\
&= (\lambda_k (S_{\mathrm{GS}}))^{(j+1)/2} \sin\left(\frac{(j+1)k\pi}{N}\right).
\end{aligned}$$

Using now the relation

$$2 \sin\left(\frac{\alpha+\beta}{2}\right) \cos\left(\frac{\alpha-\beta}{2}\right) = \sin\alpha + \sin\beta$$

with $\alpha = (j+1)k\pi/N$, $\beta = (j-1)k\pi/N$, then one finds that both sides are in fact identical.

For $j = 1$ and $j = N - 1$, one can perform the same calculation with formally introducing

$$w_{k,0} = (\lambda_k (S_{\mathrm{GS}}))^{0/2} \sin\left(\frac{0k\pi}{N}\right) = 0, \quad w_{k,N} = (\lambda_k (S_{\mathrm{GS}}))^{N/2} \sin\left(\frac{Nk\pi}{N}\right) = 0.$$

■

Remark 3.20 *Discussion of Lemma 3.19.* The eigenvalues of S_{GS} are close to one for small k. They are close to zero if k is close to $N/2$. This situation is similar to the Jacobi method without damping.

One can derive, analogosly to the damped Jacobi method in Remark 3.11, the error formula

$$\mathbf{e}^{(m)} = \sum_{k=1}^{N-1} (\lambda_k (S_{\mathrm{GS}}))^m \mathbf{w}_k (S_{\mathrm{GS}}) .$$

Using the eigenvectors $\mathbf{w}_k(S_{\mathrm{GS}})$ of S_{GS} as initial iterates, then it follows from the eigenvalues of S_{GS} that a fast error reduction can be expected only for $k \approx N/2$, whereas for k close to 0, there is only a slow convergence, see Table 3.3. It turns out that the situation for k being close to $N - 1$ is simular as for k being close to 0.

□

Table 3.3: Number of iterations for damping the norm of the error by the factor 100, $N = 64$.

k	$\mathbf{w}_k(S_{\mathrm{GS}})$
1	1895
3	207
6	50
16	6
32	1
60	115
63	1895

3.4 Summary

Remark 3.21 *Summary.* The investigation of classical iterative schemes led to the following important observations:

- Classical iterative schemes might damp highly oscillating discrete error modes very quickly. There is only a slow damping of the smooth discrete error modes.
- A smooth error mode on a given grid is generally on a coarser grid less smooth.

□