Chapter 6
Analysis of Positive Definite Iterations

Abstract This chapter gathers convergence statements about iterations satisfying suitable requirements connected with positive definiteness. Section 6.1 enumerates six cases which are analysed in Section 6.2. In several cases, convergence holds for a suitably damped version of the iteration. Of particular interest are symmetric and positive definite iterations constructed in the previous chapter. In Section 6.3 we analyse traditional symmetric iterative methods: the symmetric Gauss–Seidel iteration and the symmetric SOR method, abbreviated by SSOR. The convergence properties of SSOR are investigated in §§6.3.1–6.3.2, while modifications are described in §§6.3.3–6.3.4. Finally, in §6.3.5, numerical examples illustrate the convergence behaviour.

6.1 Different Cases of Positivity

We distinguish six cases of positivity. Consider any $\Phi \in \mathcal{L}$ and denote the corresponding matrices by

$$M = M[A], \quad N = N[A], \quad W = W[A].$$

- **Case 1**: positive spectrum of NA.

  The weakest condition considered in this chapter is a positive spectrum of NA:

$$\sigma(NA) \subset (0, \infty). \tag{6.1a}$$

- **Case 2**: directly positive definite iterations $\Phi \in \mathcal{L}_{>0}$.

  Positive definiteness appears in two versions. For directly positive definite iterations (cf. Definition 5.14) we have

$$NA > 0\,. \tag{6.1b}$$

Note that in this case no conditions on A are required except for regularity.

W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*,
Applied Mathematical Sciences 95, DOI 10.1007/978-3-319-28483-5_6

- **Case 3**: positive definite iterations $\Phi \in \mathcal{L}_{\text{pos}}$.

The standard situation is the case of an positive definite iteration (cf. Definition 5.8). Application to a positive definite matrix A yields

$$A > 0, \qquad N > 0, \qquad W > 0. \tag{6.1c}$$

Simple conclusions concerning the iteration matrix M are gathered in the next remark.

Remark 6.1. (a) Each condition in (6.1a–c) implies that $\sigma(M) \subset (-\infty, 1)$.
(b) In the case of (6.1b), M is Hermitian and satisfies $M < I$.

The positive definite matrices in (6.1b,c) induce the corresponding vector and matrix norms $\|\cdot\|_X$ for $X \in \{NA, A, N, W\}$ as defined in (C.5a,d).

Remark 6.2. Assume that convergence holds. (a) Then each of the conditions (6.1a–c) implies $\sigma(M) \subset (-1, 1)$.
(b) In the case of (6.1b), the convergence is monotone with respect to the Euclidean norm $\|\cdot\|$ and the norms $\|\cdot\|_{NA}$ and $\|\cdot\|_{(NA)^{-1}}$. The identities $\sigma(M) = \|M\| = \|M\|_{NA} = \|M\|_{(NA)^{-1}}$ hold.
(c) In the case of (6.1c), the convergence is monotone with respect to the norms $\|\cdot\|_A$ and $\|\cdot\|_W$, and $\sigma(M) = \|M\|_A = \|M\|_W$ holds.

Exercise 6.3. Assume (6.1c). The energy scalar product $\langle\cdot,\cdot\rangle_A$ is defined in (C.5b). Prove that M is symmetric with respect to $\langle\cdot,\cdot\rangle_A$, i.e., $\langle Mx, y\rangle_A = \langle x, My\rangle_A$, and that this statement is equivalent to $M = A^{-1}M^{\mathsf{H}}A$.

Let $A > 0$. The symmetry with respect to $\langle\cdot,\cdot\rangle_A$ can be transferred to the usual symmetry by the following similarity transformation: $\hat{M} = \hat{M}^{\mathsf{H}}$ holds for

$$\hat{M} := A^{1/2}MA^{-1/2} = I - A^{1/2}NA^{1/2} = I - A^{1/2}\,W^{-1}A^{1/2}. \tag{6.2a}$$

Similarly, $W > 0$ induces the similarity transformation

$$\check{M} := W^{1/2}MW^{-1/2} = I - W^{-1/2}AW^{-1/2} = I - N^{1/2}AN^{1/2}. \tag{6.2b}$$

The statement of Remark 6.2c can be expressed by

$$\rho(M) = \rho(\hat{M}) = \|\hat{M}\|_2 = \|M\|_A\,, \tag{6.2c}$$
$$\rho(M) = \rho(\check{M}) = \|\check{M}\|_2 = \|M\|_W. \tag{6.2d}$$

The proof follows from (A.6c) and (B.21b).

- **Case 4**: positive definite $W + W^{\mathsf{H}}$.

  The positive definiteness of W can be generalised to

$$W + W^{\mathsf{H}} > A > 0.$$

The weaker condition

$$W + W^{\mathsf{H}} > 0$$

will also be discussed, i.e., the Hermitian part of W is positive definite. An equivalent condition is

$$N + N^{\mathsf{H}} > 0.$$

- **Case 5**: symmetrised iteration $\Phi^{\mathrm{sym}} \in \mathcal{L}_{\mathrm{sym}}$.

  We recall the construction of a symmetric iteration $\Phi^{\mathrm{sym}} = \Phi^* \circ \Phi$ described in §5.4.2. Theorem 5.29 states that $A = A^{\mathsf{H}}$ leads to the matrices

$$\begin{aligned} M^{\mathrm{sym}} &= (I - N^{\mathsf{H}}A)(I - NA) = I - N^{\mathrm{sym}}A, \\ N^{\mathrm{sym}} &= N + N^{\mathsf{H}} - N^{\mathsf{H}}AN, \\ W^{\mathrm{sym}} &= W(W + W^{\mathsf{H}} - A)^{-1}W^{\mathsf{H}} \end{aligned} \tag{6.3}$$

with N and W belonging to Φ, while the Hermitian matrices M^{sym}, N^{sym}, and W^{sym} are associated with Φ^{sym}.

- **Case 6**: perturbed positive definite A.

  A non-Hermitian matrix A may be split into $A = A_0 + \mathrm{i}A_1$ with positive definite $A_0 := \frac{1}{2}(A + A^{\mathsf{H}})$ (cf. (3.27)). If A_1 is small in a suitable sense, A can be regarded as a perturbation of the positive definite matrix A_0.

6.2 Convergence Analysis

6.2.1 Case 1: Positive Spectrum

We assume (6.1a): $\sigma(NA) \subset (0, \infty)$. Sufficient conditions for (6.1a) are given in Lemma 5.18.

In §3.5.1, convergence of the Richardson iteration is investigated under the condition $\sigma(A) \subset (0, \infty)$. Using Proposition 5.44, we can transfer the results in §3.5.1 to NA. The quantities Θ and A in Lemma 3.21 and in Theorems 3.22, 3.23 have to be replaced with ϑ and NA. The matrices corresponding to the damped iteration Φ_ϑ are denoted by $M_\vartheta = I - \vartheta NA$, $N_\vartheta = \vartheta N$, and $W_\vartheta = N_\vartheta^{-1}$.

Lemma 6.4. *Assume that $\sigma(NA) \subset \mathbb{R}$ and denote the extreme eigenvalues of NA by $\lambda_{\min}$ and $\lambda_{\max}$. Then the spectrum of the iteration matrix M_ϑ is real for any $\vartheta \in \mathbb{R}$, i.e., $\sigma(M_\vartheta) \subset \mathbb{R}$. The spectral radius is characterised by*

$$\rho(M_\vartheta) = \max\{|1 - \vartheta\lambda_{\min}|, |1 - \vartheta\lambda_{\max}|\} \qquad \textit{for all } \vartheta \in \mathbb{R}. \tag{6.4}$$

Exercise 6.5. Characterise $\rho(M_\vartheta)$ under the above assumptions for complex ϑ.

Theorem 6.6. *Assume that condition (6.1a) holds and let $\lambda_{\max}(NA)$ be the maximal eigenvalue of NA. Then, for real ϑ, the damped iteration Φ_ϑ converges if and only if*

$$0 < \vartheta < 2/\lambda_{\max}(NA). \tag{6.5}$$

The convergence rate is described by (6.4).

Theorem 6.7 (optimal ϑ). *Under the assumptions of Theorem 6.6, the optimal convergence rate of Φ_ϑ is attained for*

$$\vartheta_{\text{opt}} = \frac{2}{\lambda_{\max} + \lambda_{\min}} \quad \textit{with} \quad \rho(M_{\vartheta_{\text{opt}}}) = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} = \frac{\kappa(NA) - 1}{\kappa(NA) + 1}. \tag{6.6a}$$

$\kappa(NA) = \lambda_{\max}/\lambda_{\min}$ is the spectral condition number of NA (cf. (B.13)). For large $\kappa(NA) \gg 1,$ the asymptotic behaviour is

$$\frac{\kappa(NA) - 1}{\kappa(NA) + 1} = 1 - \frac{2}{\kappa(NA)} + \mathcal{O}\left(\kappa(NA)^{-2}\right). \tag{6.6b}$$

The expression $1-2/\kappa$ has to be compared with the rate $1 - 1/\kappa$ for iterations with $\sigma(M) \subset [0, 1)$.

Remark 6.8. Assume (6.1a) and $\sigma(M) \subset [0, \infty)$. The optimal scaling factor ϑ satisfying $\sigma(M_\vartheta) \subset [0, 1)$ is $\vartheta_+ = 1/\lambda_{\max}$. The corresponding rate is $\rho(M_{\vartheta_+}) = 1 - \frac{1}{\kappa(NA)}$.

For a complex spectrum of $NA,$ compare with Exercise 3.26 and Theorem 3.27.

6.2.2 *Case 2: Positive Definite NA*

Theorem 6.9. *Assume $NA > 0$ and $\vartheta \in \mathbb{R}$. Then iteration (5.8) converges if and only if*

$$0 < \vartheta < 2/\left\|NA\right\|_2.$$

The convergence is monotone with respect to the Euclidean norm $\|\cdot\|_2$ and the energy norm $\|\cdot\|_{NA}$. Furthermore, the convergence rate and the contraction number coincide:

$$\rho(M_\vartheta) = \|M_\vartheta\|_2 = \|M_\vartheta\|_{NA}.$$

The optimal convergence rate (6.6a) can be expressed as a function of the condition number $\kappa(NA) = \operatorname{cond}_2(NA)$:

$$\rho(M_{\vartheta_{\text{opt}}}) = \frac{\kappa(NA) - 1}{\kappa(NA) + 1} \quad \textit{for } \vartheta_{\text{opt}} = \frac{2}{\lambda_{\max}(NA) + \lambda_{\min}(NA)}. \tag{6.7}$$

Proof. Use the results in §6.2.1, $\lambda_{\max}(NA) = \left\|NA\right\|_2$, and Remark 6.2b. $\square$

6.2.3 Case 3: Positive Definite Iteration

Now we assume (6.1c). This case is already treated by Theorem 3.34. Since the proof is still missing, we repeat the statements in short. Note that (6.8a) describes a sufficient and necessary condition for convergence.

Theorem 6.10. *Let (6.1c) be valid. Then, for* $0 \le \lambda \le \Lambda$, *the following equivalence relations hold:*

$$\begin{aligned} 2W > A > 0 &\iff \rho(M) < 1, &\text{(6.8a)}\\ 0 < \lambda W \le A \le \Lambda W &\iff \sigma(M) \subset [1-\Lambda, 1-\lambda], &\text{(6.8b)}\\ 0 \le \lambda W < A < \Lambda W &\iff \sigma(M) \subset (1-\Lambda, 1-\lambda), &\text{(6.8c)}\\ W \ge A > 0 &\iff \sigma(M) \subset [0,1). &\text{(6.8d)} \end{aligned}$$

Proof. Using the matrix $\hat{M}$ in (6.2a), $\sigma(M) = \sigma(\hat{M})$ allows us to reformulate $\sigma(\hat{M}) \subset [1-\Lambda, 1-\lambda]$ as

$$(1-\Lambda)\, I \le \hat{M} = I - A^{1/2} N A^{1/2} \le (1-\lambda)\, I$$

(cf. (C.3e)). Applying (C.3b′) with $C := A^{-1/2}$, we get the equivalent inequalities

$$(1-\Lambda)\, A^{-1} \le A^{-1} - N \le (1-\lambda)\, A^{-1}.$$

The left inequality yields $-\Lambda A^{-1} \le -N \Leftrightarrow \Lambda A^{-1} \ge N$. Applying (C.3g), we arrive at $\frac{1}{\Lambda} A \le N^{-1} = W$, i.e., $A \le \Lambda W$. The proof of $\lambda W \le A$ is analogous. This proves (6.8b). Replacing '$\le$' with '$<$', we obtain (6.8c). The implications (6.8a,d) follow for special values of λ and Λ. □

Denote the iteration defined by the matrices (6.1c) by Φ. Below we discuss the damped iteration Φ_ϑ. Theorems 6.6 and 6.7 and (6.2c,d) yield the following result.

Theorem 6.11. *Assume (6.1c). The damped iteration* Φ_ϑ *defined by (5.8) converges if and only if* ϑ *satisfies*

$$\begin{aligned} &0 < \vartheta < 2/\lambda_{\max} \quad \textit{with}\\ &\lambda_{\max} := \|N^{1/2} A N^{1/2}\|_2 = \|A^{1/2} N A^{1/2}\|_2 = \rho(NA). \end{aligned} \tag{6.9}$$

An equivalent formulation of condition (6.9) using $W = N^{-1}$ *is*

$$0 < \vartheta A < 2W.$$

The convergence rate (even for general $\vartheta \in \mathbb{C}$*) is*

$$\rho(M_\vartheta) = \|M_\vartheta\|_A = \|M_\vartheta\|_W = \max\{|1-\vartheta\lambda_{\min}|, |1-\vartheta\lambda_{\max}|\},$$

where $\lambda_{\min}$ *is the minimal eigenvalue of* NA. *The optimal value of* ϑ *minimising* $\rho(M_\vartheta)$ *is* ϑ_{opt} *in (6.7).*

Corollary 6.12. (a) If $A > 0$ and $N < 0$ ($\Leftrightarrow W < 0$), then Φ_ϑ converges if and only if $0 > \vartheta > 2/\rho(NA)$.
(b) If $A > 0$, $N = N^{\mathsf{H}}$, and neither $N > 0$ nor $N < 0$, Φ_ϑ diverges for all $\vartheta \in \mathbb{C}$.

6.2.4 Case 4: Positive Definite $W+W^{\mathsf{H}}$ or $N+N^{\mathsf{H}}$

First, we assume

$$W + W^{\mathsf{H}} > A > 0. \tag{6.10}$$

The first part of the next theorem coincides with Theorem 3.35.

Theorem 6.13. *Under condition (6.10), the iteration converges monotonically with respect to the energy norm:*

$$\rho(M) \le \|M\|_A < 1 \qquad \text{for } M = I - W^{-1}A.$$

$W+W^{\mathsf{H}} > A$ is also necessary for $\|M\|_A < 1$ (but even without condition (6.10), $\rho(M) < 1$ is possible).

Proof. Assume that $W + W^{\mathsf{H}} - A$ has a nonpositive eigenvalue. Then, by (3.37), $\hat{M}^{\mathsf{H}}\hat{M} = I - A^{1/2}W^{\mathsf{H}}(W + W^{\mathsf{H}} - A)WA^{1/2}$ has an eigenvalue ≥ 1 implying $\|M\|_A \ge 1$. □

Next, we assume

$$W + W^{\mathsf{H}} > 0 \qquad \text{and} \qquad A > 0.$$

To regain inequality (6.10), we have to apply a suitable damping, since Φ_ϑ is associated with $W_\vartheta = \frac{1}{\vartheta}W$. For instance, the choice

$$\vartheta < \frac{\lambda_{\min}(W + W^{\mathsf{H}})}{\lambda_{\max}(A)} \tag{6.11}$$

ensures that $W_\vartheta + W_\vartheta^{\mathsf{H}} > A > 0$.

Exercise 6.14. The sharper estimate $\vartheta < \lambda_{\min}\left(A^{-1/2}\left(W + W^{\mathsf{H}}\right)A^{-1/2}\right)$ also implies $W_\vartheta + W_\vartheta^{\mathsf{H}} > A > 0$.

Remark 6.15. Theorem 6.13 proves that Φ_ϑ with ϑ in (6.11) is convergent. The convergence is monotone with respect to the energy norm: $\|M_\vartheta\|_A = \|\hat{M}_\vartheta\|_2 = \rho(\hat{M}_\vartheta^{\mathsf{H}}\hat{M}_\vartheta)^{1/2}$ ($\hat{M}_\vartheta$ as in (6.2a)).

Optimising the damping factor ϑ leads us to the quadratic inequality

$$\vartheta\left(W + W^{\mathsf{H}}\right) \ge \vartheta^2 A + \alpha W^{\mathsf{H}}AW, \qquad \alpha = \alpha(\vartheta) > 0. \tag{6.12}$$

For each sufficiently small $\vartheta > 0$, there is a maximal $\alpha(\vartheta)$ satisfying (6.12). ϑ_{opt} is the maximiser of $\alpha(\vartheta)$.

Theorem 6.16. *Let Φ_ϑ satisfy (6.12). Then Φ_ϑ converges with the contraction number*

$$\|M_\vartheta\|_A = \sqrt{1-\alpha}\,.$$

Proof. Repeat the estimate of $\hat{M}_\vartheta^{\mathsf{H}}\hat{M}_\vartheta$ in (3.37) and use (6.12). □

The assumption $N + N^{\mathsf{H}} > 0$ does not yield new results because of the next lemma, but in concrete cases the matrix $N + N^{\mathsf{H}}$ may be easier to analyse than $W + W^{\mathsf{H}}$.

Lemma 6.17. $N + N^{\mathsf{H}} > 0$ *and* $W + W^{\mathsf{H}} > 0$ *are equivalent.*

Proof. $N + N^{\mathsf{H}} > 0 \Leftrightarrow W^{\mathsf{H}}\left(N + N^{\mathsf{H}}\right)W = W + W^{\mathsf{H}} > 0$ by (C.3a). □

Remark 6.18. Assume $N + N^{\mathsf{H}} > 0$. With a suitable scaling, N_ϑ satisfies

$$N_\vartheta + N_\vartheta^{\mathsf{H}} > N_\vartheta^{\mathsf{H}} A\, N_\vartheta$$

which is equivalent to $W_\vartheta + W_\vartheta^{\mathsf{H}} > A > 0$ and allows applying Theorem 6.13. The estimate

$$N_\vartheta + N_\vartheta^{\mathsf{H}} - N_\vartheta^{\mathsf{H}} A\, N_\vartheta \geq \alpha A$$

is equivalent to (6.12).

6.2.5 *Case 5: Symmetrised Iteration Φ^{sym}*

Below we use the notation defined in (6.3). In particular, M^{sym} and M are the respective iteration matrices of $\Phi^{\mathrm{sym}} = \Phi^* \circ \Phi$ and Φ.

Remark 6.19. Assume $A > 0$. Then

$$\sigma(M^{\mathrm{sym}}) = \|\hat{M}^{\mathrm{sym}}\|_2 = \|M^{\mathrm{sym}}\|_A \subset [0, \infty)$$

holds, where $\hat{M}^{\mathrm{sym}} = A^{1/2} M^{\mathrm{sym}} A^{-1/2} > 0$. The connection to the iteration Φ is given by

$$\sigma(M^{\mathrm{sym}}) = \|M\|_A^2 = \|\hat{M}\|_2^2 \qquad (\hat{M} := A^{1/2} M A^{-1/2}). \tag{6.13}$$

If Φ^{sym} converges, the convergence is monotone with respect to the energy norm $\|\cdot\|_A$, and $\sigma(M^{\mathrm{sym}}) \subset [0, 1)$ holds.

Proof. Use $\hat{M}^{\mathrm{sym}} = \hat{M}^{\mathsf{H}}\hat{M} \geq 0$ with $\hat{M} = A^{1/2} M A^{-1/2}$ and the similarity of M^{sym} and $\hat{M}^{\mathrm{sym}}$. □

Equation (6.13) yields the following important conclusion. In general, the condition $\|M\|_A < 1$ (monotone convergence with respect to the energy norm) is only sufficient for convergence. Because of the next statement this is even a necessary condition for Φ^{sym}. Therefore estimates of $\|M\|_A$ become important.

Conclusion 6.20. $\Phi^{\mathrm{sym}} = \Phi^* \circ \Phi$ *converges if and only if* Φ *is monotonically converging with respect to the energy norm, i.e.,* $\|M_\Phi\|_A < 1$.

The construction of $\Phi^{\mathrm{sym}} = \Phi^* \circ \Phi$ in §5.4.2 ensures that $A > 0$ implies $N^{\mathrm{sym}} = (N^{\mathrm{sym}})^{\mathsf{H}}$. N^{sym} is Hermitian, but not necessarily positive definite. By Corollary 6.12b, convergence of the damped version of Φ^{sym} requires either $N^{\mathrm{sym}} > 0$ or $N^{\mathrm{sym}} < 0$. The second case is completely nonstandard. Since $N^{\mathrm{sym}} = N + N^{\mathsf{H}} - N^{\mathsf{H}}AN$ (cf. (6.3)), the condition $N^{\mathrm{sym}} > 0$ is equivalent to the identical conditions $N + N^{\mathsf{H}} > N^{\mathsf{H}}A\,N$ and $W + W^{\mathsf{H}} > A > 0$ in Remark 6.18. As stated in Remark 6.18, these inequalities can be guaranteed by a suitable scaling if $N + N^{\mathsf{H}} > 0$ or equivalently $W + W^{\mathsf{H}} > 0$.

Next, we investigate the properties of $(\Phi_\vartheta)^{\mathrm{sym}} = \Phi_\vartheta^* \circ \Phi_\vartheta$. For a proof, use Remark 6.19.

Proposition 6.21. *Assume $A > 0$. Let M, N, W and M_ϑ, N_ϑ, W_ϑ be the matrices associated with Φ and the damped iteration Φ_ϑ, while $M^{\vartheta,\mathrm{sym}}$, $N^{\vartheta,\mathrm{sym}}$, $W^{\vartheta,\mathrm{sym}}$ are those of $(\Phi_\vartheta)^{\mathrm{sym}} = \Phi_\vartheta^* \circ \Phi_\vartheta$.*

(a) Positive definite case $N + N^{\mathsf{H}} > 0$: For a suitable scaling factor $\vartheta > 0$, $W_\vartheta + W_\vartheta^{\mathsf{H}} > A$ holds and $(\Phi_\vartheta)^{\mathrm{sym}}$ converges. Since $N^{\vartheta,\mathrm{sym}} = N_\vartheta + N_\vartheta^{\mathsf{H}} - N_\vartheta^{\mathsf{H}} A N_\vartheta$, the statements of Remark 6.18 apply. In the convergent case, the transformed iteration matrix $\hat{M}^{\vartheta,\mathrm{sym}} := A^{1/2} M^{\vartheta,\mathrm{sym}} A^{-1/2}$ satisfies

$$0 \leq \hat{M}^{\vartheta,\mathrm{sym}} < I\,,$$

and $(\Phi_\vartheta)^{\mathrm{sym}}$ is a positive definite iteration.

(b) Negative definite case $N + N^{\mathsf{H}} < 0$: A negative ϑ leads us back to case (a).

(c) Otherwise, $(\Phi_\vartheta)^{\mathrm{sym}}$ diverges for any choice of ϑ.

Let ϑ be a suitable scaling of Φ so that $W_\vartheta + W_\vartheta^{\mathsf{H}} > A$ holds. Rename Φ_ϑ, M_ϑ, N_ϑ, W_ϑ, $(\Phi_\vartheta)^{\mathrm{sym}}$ by Φ, M_Φ, N_Φ, W_Φ, Φ^{sym}. The statements of Remark 6.19, together with the convergence criterion $W_\Phi + W_\Phi^{\mathsf{H}} > A > 0$, yield the next result.

Theorem 6.22. *Assume that*

$$W_\Phi + W_\Phi^{\mathsf{H}} > A > 0.$$

Then the symmetrised iteration $\Phi^{\mathrm{sym}} := \Phi^* \circ \Phi$ converges monotonically:

$$\rho(M_{\Phi^{\mathrm{sym}}}) = \|M_\Phi\|_A^2 < 1. \tag{6.14}$$

Moreover the spectrum is nonnegative:

$$\sigma(M_{\Phi^{\mathrm{sym}}}) \subset [0, \rho(M_{\Phi^{\mathrm{sym}}})] \subset [0, 1).$$

Proof. For the last equality combine (6.13) in the form $\sigma(M_{\Phi^{\mathrm{sym}}}) \subset [0, \rho(M_{\Phi^{\mathrm{sym}}})]$ with (6.14). ☐

Since $\sigma(M_{\Phi^{\mathrm{sym}}}) \subset [0, 1)$ holds, Remark 6.8 shows that the convergence rate can be improved by damping (extrapolation).

Concerning the contraction number with respect to the energy norm, Φ and Φ^{sym} behave the same: Φ^{sym} consists of two iteration steps and yields the same bound $\|M_\Phi\|_A^2$ as two steps of Φ. However, concerning the convergence rate, the symmetric iteration performs worse. While $\rho(M_\Phi)$ may be strictly smaller than $\|M_\Phi\|_A$ (cf. Remark 6.15), $\rho(M_{\Phi^{\mathrm{sym}}})$ is equal to $\|M_\Phi\|_A^2$; i.e., the inequality

$$\rho(M_{\Phi^{\mathrm{sym}}}) \geq \rho(M_\Phi)^2$$

holds and may possibly be a strict inequality.

When assessing Φ^{sym} and Φ only with regard to convergence speed, Φ should be preferred. The advantage of Φ^{sym} will be seen in connection with Krylov methods. Another advantage is the possibility to perform Φ^{sym} with less cost than two steps of Φ (cf. Remark 6.27).

6.2.6 Case 6: Perturbed Positive Definite Case

The next generalisation splits A into $A_0 + \mathrm{i}A_1$ according to (3.27). The condition $A > 0$ is weakened by $A_0 > 0$.

Theorem 6.23. *Assume that $A = A_0 + \mathrm{i}A_1$ according to (3.27) satisfies $A_0 > 0$. Let $W = N[A]^{-1} > 0$ hold for the matrix of the third normal form of $\Phi(\cdot,\cdot,A)$. The optimal constants $0 < \lambda \leq \Lambda$ and $\tau \geq 0$ in*

$$\lambda W \leq A_0 \leq \Lambda W, \qquad -\tau W \leq A_1 \leq \tau W \tag{6.15}$$

are $\lambda = \lambda_{\min}(NA_0)$, $\Lambda = \lambda_{\max}(NA_0)$, and $\tau := \rho(NA_1)$. Then the damped iteration (5.8) converges for

$$0 < \vartheta < \frac{2\lambda}{\lambda\Lambda + \tau^2}$$

monotonically with respect to the norm $\|\cdot\|_W$:

$$\rho(M_\vartheta) \leq \|M_\vartheta\|_W \;\leq\; \frac{1}{2}\vartheta(\Lambda - \lambda) + \sqrt{\left[1 - \tfrac{1}{2}\Theta\,(\Lambda + \lambda)\right]^2 + \Theta^2\tau^2} \;<\; 1.$$

The optimal ϑ can be determined as in (3.31c).

Proof. M_ϑ is similar to $M := N^{-1/2}M_\vartheta N^{1/2} = I - \vartheta N^{1/2}AN^{1/2}$. M can be regarded as the iteration matrix of the Richardson method for $\Theta := \vartheta$ and $A' := N^{1/2}AN^{1/2}$ instead of A. The splitting $A = A_0 + \mathrm{i}A_1$ induces the splitting $A' = A'_0 + \mathrm{i}A'_1$ with the Hermitian matrices

$$A'_0 = N^{1/2}A_0\,N^{1/2}, \qquad A'_1 = N^{1/2}A_1\,N^{1/2}.$$

The inequalities (3.30a,b) applied to A' are equivalent to (6.15). The estimate (3.31b) following from Theorem 3.30 refers to the iteration matrix M and reads as $\|M\|_2 = \|W^{1/2}M_\vartheta W^{-1/2}\|_2 = \|M_\vartheta\|_W$. $\square$

The counterpart of Theorem 3.31 reads as follows.

Theorem 6.24. *Under the assumption (3.30a,b), the estimate*

$$\rho(M_\vartheta) \le r_\vartheta := \sqrt{\vartheta^2\tau^2 + \max\{|1-\vartheta\lambda|, |1-\vartheta\Lambda|\}}$$

holds for the damped iteration (5.8) with λ *and* Λ *as in Theorem 6.23. The convergence is ensured in the form* $r_\vartheta < 1$ *if*

$$0 < \vartheta < \overline{\vartheta} \quad \text{with } \overline{\vartheta} := \begin{cases} 2\Lambda \,/\, \left(\Lambda^2 + \tau^2\right) & \text{if } \tau^2 < \lambda\Lambda, \\ 2\lambda \,/\, \left(\Lambda^2 + \tau^2\right) & \text{if } \tau^2 \ge \lambda\Lambda. \end{cases}$$

r_ϑ *is minimal for* $\vartheta' := \min\{\frac{\lambda}{\lambda^2+\tau^2}, \frac{2}{\lambda+\Lambda}\}$. *Moreover, the norm estimate (6.16) holds:*

$$\|(M_\vartheta)^m\|_W \le 2r_\vartheta^m \qquad (m \ge 0). \tag{6.16}$$

Exercise 6.25. Reformulate Corollary 3.32 for the damped iteration (5.8).

In the case of a matrix $NA = C_0 + \mathrm{i}C_1$ decomposed into a Hermitian part $C_0 := (NA + A^{\mathsf{H}}N^{\mathsf{H}})/2$ and a skew-Hermitian part $C_1 := (NA - A^{\mathsf{H}}N^{\mathsf{H}})/(2\mathrm{i})$, we can apply the counterparts of Theorems 3.28, 3.30, 3.31 and Corollaries 3.32, 3.33 to get similar results as above.

6.3 Symmetric Gauss–Seidel Iteration and SSOR

The symmetric Gauss–Seidel method $\Phi^{\mathrm{symGS}} = \Phi^{\mathrm{GS}}_{\mathrm{backw}} \circ \Phi^{\mathrm{GS}} \in \mathcal{L}_{\mathrm{sym}}$ and the symmetric SOR method (SSOR) $\Phi^{\mathrm{SSOR}}_\omega = \Phi^{\mathrm{backwSOR}}_\omega \circ \Phi^{\mathrm{SOR}}_\omega \in \mathcal{L}_{\mathrm{sym}}$ are defined in §5.4.3. In 1955, the SSOR method is first described by Sheldon [339].

Since $\Phi^{\mathrm{SOR}}_1 = \Phi^{\mathrm{GS}}$ (cf. Proposition 3.13c), the symmetric Gauss–Seidel iteration also satisfies $\Phi^{\mathrm{symGS}} = \Phi^{\mathrm{SSOR}}_1$. Therefore the symmetric Gauss–Seidel method does not require a separate analysis.

6.3.1 The Case $A > 0$

Theorem 6.26. *Let* A *be positive definite. The symmetric SOR method* $\Phi^{\mathrm{SSOR}}_\omega$ *converges for* $0 < \omega < 2$ *with*

$$\rho(M^{\mathrm{SSOR}}_\omega) = \|M^{\mathrm{SOR}}_\omega\|_A^2 < 1, \quad \text{where} \quad M^{\mathrm{SSOR}}_\omega = M^{\mathrm{backwSOR}}_\omega M^{\mathrm{SOR}}_\omega$$

(cf. Remark 5.2 and (3.15b)). The spectrum $\sigma(M^{\mathrm{SSOR}}_\omega)$ *is contained in* $[0,1)$. $\Phi^{\mathrm{SSOR}}_\omega$ *diverges for all real* $\omega \notin (0,2)$. *The same statements hold for the block-SSOR version.*

Proof. Combine the result of Theorem 3.41 (Ostrowski) with Theorem 6.22. Concerning $\omega \notin (0,2)$ use $\rho(M^{\mathrm{SSOR}}_\omega) = \|M^{\mathrm{SOR}}_\omega\|_A^2 \ge \rho(M^{\mathrm{SOR}}_\omega)^2$ and (3.41). □

The amount of work required by the symmetric SOR iteration seems to be twice as large as that for the original SOR method, since one SSOR step consists of two SOR steps (cf. (5.14)). However, this disadvantage can be overcome.

Remark 6.27 (Niethammer [292, 293]). The SSOR iteration requires essentially the same amount of work as the SOR method if one tolerates additional storage needed for an auxiliary vector. The cost factor (cf. §2.3) amounts to

$$C_\Phi^{\mathrm{SSOR}} = C_\Phi^{\mathrm{SOR}} + 5/C_A = 2 + 6/C_A$$

for an optimal implementation instead of $2\,C_\Phi^{\mathrm{SOR}} = 4 + 2/C_A$ for the naive implementation (5.14).

Proof. The first SSOR half-step $x^m \longmapsto x^{m+1/2}$ can be rewritten as

$$x^{m+1/2} = x^m + \omega\left(Lx^{m+1/2} - x^m + Ux^m + D^{-1}b\right) \tag{6.17a}$$

(cf. (3.15f)). The second backward SOR step

$$x^{m+1} = x^{m+1/2} + \omega\left(Ux^{m+1} - x^{m+1/2} + Lx^{m+1/2} + D^{-1}b\right) \tag{6.17b}$$

contains the term $Lx^{m+1/2}$ which is already evaluated in (6.17a). Analogously, the term Ux^{m+1} computed in (6.17b) can be used in the following half-step:

$$x^{m+3/2} = x^{m+1} + \omega\left(Lx^{m+3/2} - x^{m+1} + Ux^{m+1} + D^{-1}b\right).$$

On the average, one SSOR step requires one evaluation of Lx and Ux. $\square$

The statements of Theorem 3.44 can be translated into the following statement about the SSOR method.

Theorem 6.28. *Let $A = D - E - E^{\mathsf{H}} > 0$ and $0 < \omega < 2$. Furthermore, assume that there are constants $\gamma > 0$ and $\varGamma$ with (6.18a,b) (cf. (3.46a,b)):*

$$0 < \gamma D \le A\,, \tag{6.18a}$$

$$\left(\frac{1}{2}D - E\right) D^{-1} \left(\frac{1}{2}D - E^{\mathsf{H}}\right) \le \frac{1}{4}\varGamma A\,. \tag{6.18b}$$

Then the following estimate holds:

$$\rho(M_\omega^{\mathrm{SSOR}}) = \|M_\omega^{\mathrm{SSOR}}\|_A \le 1 - \frac{2\varOmega}{\frac{\varOmega^2}{\gamma} + \varOmega + \frac{\varGamma}{4}} \quad \textit{with } \varOmega := \frac{2-\omega}{2\omega}\,. \tag{6.18c}$$

For $\omega' = 2/(1+\sqrt{\gamma\varGamma}\,)$, the bound in (6.18c) becomes a minimum:

$$\rho(M_\omega^{\mathrm{SSOR}}) \le \frac{\sqrt{\varGamma} - \sqrt{\gamma}}{\sqrt{\varGamma} + \sqrt{\gamma}} = \frac{1 - \sqrt{\gamma/\varGamma}}{1 + \sqrt{\gamma/\varGamma}}\,.$$

Proof. Combine (6.14) with Theorem 3.44. $\square$

The following statement is analogous to Conclusion 3.46.

Conclusion 6.29 (order improvement). *In the case of* $\rho(D^{-1}ED^{-1}E^{\mathsf{H}}) \leq 1/4$ *(or* $\leq 1/4 + \mathcal{O}(1-\rho(M^{\mathrm{Jac}}))$*), the choice* $\omega = \omega'$ *enables an order improvement. If* τ *is the order of the Jacobi (and of the symmetric Gauss–Seidel) method, then* $\tau/2$ *is the order of the SSOR method with* $\omega = \omega'$.

The condition $\rho(D^{-1}ED^{-1}E^{\mathsf{H}}) \leq 1/4$ is essential. This inequality does not hold for the model problem with chequer-board ordering. Then, as we shall see in §6.3.4, no order improvement is possible.

For completeness, we repeat the properties of the symmetric Gauss–Seidel iteration.

Proposition 6.30. *(a) The iteration matrix of the symmetric Gauss–Seidel iteration and the matrices of the second and third normal forms are*

$$\begin{aligned} M^{\mathrm{symGS}} &= (D-F)^{-1}E(D-E)^{-1}F, \\ N^{\mathrm{symGS}} &= (D-F)^{-1}D(D-E)^{-1}, \\ W^{\mathrm{symGS}} &= (D-E)D^{-1}(D-F) \;=\; A + ED^{-1}F. \end{aligned}$$

(b) The symmetric Gauss–Seidel iteration is a symmetric iteration in the sense of Definition 5.3, provided that $D \in \mathbb{R}^{I\times I}$.

(c) If $A > 0$, *the matrix* W^{symGS} *of the third normal form is also positive definite, so that the symmetric Gauss–Seidel iteration is a positive definite iteration.*

(d) The symmetric Gauss–Seidel iteration converges and the spectrum of the iteration matrix is nonnegative:

$$\sigma(M^{\mathrm{symGS}}) \subset [0,1)\,.$$

6.3.2 SSOR in the 2-Cyclic Case

In the 2-cyclic case, we can rewrite the backward SOR iteration as $\Phi_\omega^{\mathrm{backwSOR}} = \Phi_\omega^{(1)} \circ \Phi_\omega^{(2)}$ with the partial steps defined in (6.19a,b). Therefore, the symmetric SOR iteration takes the form

$$\Phi_\omega^{\mathrm{SSOR}} = \Phi_\omega^{(1)} \circ \Phi_\omega^{(2)} \circ \Phi_\omega^{(2)} \circ \Phi_\omega^{(1)} \in \mathcal{L}_{\mathrm{sym}}.$$

Exercise 6.31. Prove: (a) The SSOR iteration matrix $M_\omega^{\mathrm{SSOR}} = M_\omega^{(1)} M_\omega^{(2)} M_\omega^{(2)} M_\omega^{(1)}$ leads to the rate

$$\rho(M_\omega^{(1)} M_\omega^{(2)} M_\omega^{(2)} M_\omega^{(1)}) = \rho(M_\omega^{(2)} M_\omega^{(2)} M_\omega^{(1)} M_\omega^{(1)}).$$

(b) $M_\omega^{(1)} M_\omega^{(1)} = M_{\omega'}^{(1)}$ and $M_\omega^{(2)} M_\omega^{(2)} = M_{\omega'}^{(2)}$ hold with $\omega' := \omega(2-\omega)$.
(c) $0 < \omega < 2$ implies $0 < \omega' \leq 1$. $\omega' = 1$ is only achieved for $\omega = 1$.

Exercise 6.31 entails the following negative conclusion.

Conclusion 6.32. *In the 2-cyclic case, $\rho(M_\omega^{\mathrm{SSOR}}) = \rho(M_{\omega'}^{\mathrm{SOR}})$ holds with $\omega' := \omega(2-\omega) \leq 1$ for all $0 < \omega < 2$. According to Theorem 4.27, underrelaxation ($\omega' < 1$) is always slower than the Gauss–Seidel iteration ($\omega' = 1$). Hence, $\omega = 1$ is the optimal parameter and SSOR simplifies to the symmetric Gauss–Seidel iteration (cf. Alefeld [2]).*

The reason for the missing order improvement is that, differently from the situation discussed in Remark 6.29, the condition $\rho(D^{-1}ED^{-1}E^{\mathsf{H}}) \leq \frac{1}{4}$ is not satisfied. In the 2-cyclic case, we have $\rho(D^{-1}ED^{-1}E^{\mathsf{H}}) = \rho(D_1^{-1}A_1D_2^{-1}A_2) = \rho(M^{\mathrm{GS}}) \approx 1$ (cf. Theorem 4.20).

Exercise 6.33. Let (A, D) be 2-cyclic. Prove that

$$M^{\mathrm{symGS}} = \begin{bmatrix} 0 & -D_1^{-1}A_1D_2^{-1}A_2D_1^{-1}A_1 \\ 0 & D_2^{-1}A_2D_1^{-1}A_1 \end{bmatrix}$$

is the iteration matrix of the symmetric Gauss–Seidel method and that

$$\rho(M^{\mathrm{symGS}}) = \rho(M^{\mathrm{GS}}).$$

6.3.3 *Modified SOR*

In the 2-cyclic case, we can regard the SOR method as a product iteration $\Phi_\omega^{\mathrm{SOR}} = \Phi_\omega^{(2)} \circ \Phi_\omega^{(1)}$ (cf. §5.4), where $\Phi_\omega^{(1)}$ involves only the first block of the vector and $\Phi_\omega^{(2)}$ only the second one:

$$\Phi_\omega^{(1)}(x,b) = \begin{pmatrix} x^1 - \omega\left[x^1 - D_1^{-1}\left(A_1x^2 - b^1\right)\right] \\ x^2 \end{pmatrix} \tag{6.19a}$$

$$\Phi_\omega^{(2)}(x,b) = \begin{pmatrix} x^1 \\ x^2 - \omega\left[x^2 - D_2^{-1}\left(A_2x^1 - b^2\right)\right] \end{pmatrix} \tag{6.19b}$$

where $x = \left[\begin{smallmatrix} x^1 \\ x^2 \end{smallmatrix}\right]$, $b = \left[\begin{smallmatrix} b^1 \\ b^2 \end{smallmatrix}\right]$, and A is split as in (4.3). The corresponding iteration matrices are

$$M_\omega^{(1)} = \begin{bmatrix} (1-\omega)\,I & \omega D_1^{-1}A_1 \\ 0 & I \end{bmatrix}, \quad M_\omega^{(2)} = \begin{bmatrix} I & 0 \\ \omega D_2^{-1}A_2 & (1-\omega)\,I \end{bmatrix}.$$

Thus we have $M_\omega^{\mathrm{SOR}} = M_\omega^{(2)}M_\omega^{(1)}$. The modified SOR iteration (MSOR) makes use of different relaxation parameters ω and ω' in both of the half-steps:

$$\Phi_{\omega,\omega'}^{\mathrm{mod\,SOR}} = \Phi_{\omega'}^{(2)} \circ \Phi_\omega^{(1)}.$$

Again the comment in Remark 3.6 about multiple parameters applies. Concerning convergence analysis and optimal parameters, we refer to Young [412, §8] and Hadjidimos [211, §3].

6.3.4 Unsymmetric SOR Method

The only reason for mentioning the unsymmetric SOR method is that it is constructed in analogy to the modified Gauss–Seidel method in §6.3.3. The SSOR method $\Phi_\omega^{\mathrm{SSOR}} = \Phi_\omega^{\mathrm{backwSOR}} \circ \Phi_\omega^{\mathrm{SOR}}$ (cf. (5.15)) can be modified by choosing different parameters ω, ω' in both factors. Accordingly, the unsymmetric SOR iteration reads

$$\Phi_{\omega,\omega'}^{\mathrm{unsymSOR}} := \Phi_{\omega'}^{\mathrm{backwSOR}} \circ \Phi_\omega^{\mathrm{SOR}}.$$

Again, this method is not notably better than the SSOR method. For more details and further references, see Hadjidimos [211, §4.1].

6.3.5 Numerical Results for the SSOR Iteration

For methods with an iteration matrix satisfying $0 < A^{1/2}MA^{-1/2} < \rho(M)I$, Remark 2.22d is applicable: the quotients $\|e^m\|_A/\|e^{m-1}\|_A$ converge monotonically to $\rho(M)$. Since $M = M_\omega^{\mathrm{SSOR}}$ satisfies this assumption, we observe this monotone behaviour for the SSOR iteration and the symmetric Gauss–Seidel method ($\omega = 1$). Table 6.1 (left) contains the results of the SSOR method with lexicographical ordering. For the Poisson model problem with step size $h = 1/32$, we obtain the convergence rate 0.98092. According to Table 3.2, $\omega = \omega' = 1.8213$ is the optimal value for the bound (3.55c), which becomes $\|M_\omega^{\mathrm{SSOR}}\|_A \le 0.9065$. Table 6.1 shows the convergence rates for different ω. Obviously, $\rho(M_\omega^{\mathrm{SSOR}})$ attains its minimum not at $\omega = \omega'$ but for $\omega_{\mathrm{opt}} \in [1.845, 1.846]$. The values of Table 6.1 demonstrate that, differently from the SOR method (cf. Fig. 4.1), the convergence rate has a flat minimum. Small errors in the choice of $\omega = \omega_{\mathrm{opt}}$ deteriorate the convergence rate only insignificantly. In this respect, the choice $\omega = \omega'$ is sufficiently good.

symmetric Gauss–Seidel iteration					SSOR with $\omega = 1.8213$	
m	$\|e^m\|_\infty$	$\|e^m\|_A$	$\frac{\|e^m\|_\infty}{\|e^{m-1}\|_\infty}$	$\frac{\|e^m\|_A}{\|e^{m-1}\|_A}$	$\|e^m\|_A$	$\frac{\|e^m\|_A}{\|e^{m-1}\|_A}$
1	1.48	202	0.79011	0.579572	2.3_{10}+02	0.67588
2	1.35	159	0.91627	0.790646	1.6_{10}+02	0.71534
3	1.27	137	0.94025	0.858495	1.2_{10}+02	0.72622
4	1.20	122	0.94528	0.891046	9.0_{10}+01	0.73679
5	1.14	111	0.94734	0.910237	6.7_{10}+01	0.74876
94	0.158	11.2	0.98074	0.980884	3.4_{10}–04	0.87961
95	0.155	11.0	0.98075	0.980891	2.8_{10}–04	0.87961
96	0.152	10.8	0.98075	0.980897	2.5_{10}–04	0.87961
97	0.149	10.6	0.98076	0.980903	2.2_{10}–04	0.87961
98	0.146	10.4	0.98076	0.980909	1.9_{10}–04	0.87961
99	0.144	10.2	0.98077	0.980914	1.7_{10}–04	0.87961
100	0.141	10.0	0.98077	0.980919	1.5_{10}–04	0.87961

ω	$\rho(M_\omega^{SSOR})$
1	0.98092
1.8	0.88376
1.81	0.88163
1.8213	0.87962
1.83	0.87845
1.84	0.87765
1.8450	0.877529
1.8455	0.877528
1.8460	0.877528
1.847	0.877528
1.85	0.87762
1.86	0.87855
1.87	0.88066

Table 6.1 *Left:* Symmetric Gauss–Seidel iteration and SSOR for $h = 1/32$. *Right:* Convergence rates of the SSOR method for $h = 1/32$ and different ω.