Numer. Math. 36, 405–429 (1981)

Numerische Mathematik

Solving Theodorsen's Integral Equation for Conformal Maps with the Fast Fourier Transform and Various Nonlinear Iterative Methods

Martin H. Gutknecht

Seminar für angewandte Mathematik, ETH-Zentrum HG, CH-8092 Zürich, Switzerland

Summary. We investigate several iterative methods for the numerical solution of Theodorsen's integral equation, the discretization of which is either based on trigonometric polynomials or function families with known attenuation factors. All our methods require simultaneous evaluations of a conjugate periodic function at each step and allow us to apply the fast Fourier transform for this. In particular, we discuss the nonlinear JOR iteration, the nonlinear SOR iteration, a nonlinear second order Euler iteration, the nonlinear Chebyshev semi-iterative method, and its cyclic variant. Under special symmetry conditions for the region to be mapped onto we establish local convergence in the case of discretization by trigonometric interpolation and give simple formulas for the optimal parameters (e.g., the underrelaxation factor) and the asymptotic convergence factor. Weaker related results for the general non-symmetric case are presented too. Practically, our methods extend the range of application of Theodorsen's method and improve its effectiveness strikingly.

Subject Classifications: AMS (1980): 30C30; 42A50, 65H10, 65R20. CR: 5.19; 5.15, 5.18.

1. Introduction

Solving Theodorsen's integral equation for the numerical computation of conformal mappings is one of the well-known methods [4] that take advantage of the fact that the dimension of the problem can be reduced from two to one if at first the boundary correspondence function θ is determined. The method yields this function for a map of the unit disk onto a starlike region given by the polar coordinates $\tau, \rho(\tau)$ of its boundary Γ.

This work was partially supported by NRC (Canada) grant No. A8240 while, in 1976, the author was at the Dept. of Computer Science, University of British Columbia, Vancouver, B.C., Canada. It is part of a thesis submitted for Habilitation at ETH Zurich

0029-599X/81/0036/0405/$05.00

The usual way to discretize Theodorsen's integral equation is based on interpolating $t \mapsto \log \rho(\theta(t))$ on an equidistant grid by a trigonometric polynomial T of degree N and evaluating the conjugate polynomial KT [which approximates $t \mapsto \theta(t) - t$] on the same grid. This process used to be replaced by the equivalent multiplication with Wittich's matrix $\mathbf{K}$ [4]. However, while this multiplication requires $O(N^2)$ multiplications, the fast Fourier transform (FFT) made it possible to implement the original idea of interpolation, conjugation, and evaluation with only $O(N \log N)$ multiplications [10, 13]. On proposal of Prof. P. Henrici, both Mrs. C. Lundwall-Skaar [16] and the author started in 1975 with numerical experiments to explore the progress made possible by the FFT in this field.

The discretization leads to a nonlinear system of $2N$ equations in $2N$ unknowns. In accordance with its structure the nonlinear Jacobi method is the simplest of several iterative numerical methods that have been proposed for its solution [4]. Unfortunately, the global and local convergence of this method are only ensured if $\varepsilon_\rho := \sup |\rho'(\tau)/\rho(\tau)| < 1$. Niethammer [18] proposed to permute the unknowns and the equations before applying the nonlinear SOR method or the Newton-SOR method. He pointed out that a nearly optimal underrelaxation factor can be determined a priori subject to the assumption that the eigenvalues of the Fréchet derivative $\mathbf{J}$ of the Jacobi iteration function are exactly or nearly imaginary. For boundaries Γ satisfying a special symmetry condition Gekeler [6] found rather large rectangular regions containing these eigenvalues. However, we state in Theorem 3.4 that these eigenvalues are in fact purely imaginary under weaker conditions. Essentially, we only require that Γ should be symmetric about one or several axes passing through the origin and that ρ' should not change sign except at the points on these axes. On the basis of this theorem it is possible to establish for arbitrary ε_ρ the local convergence of the nonlinear SOR iteration, of the Newton-SOR method, and of several other numerical methods if the exact solution of the discrete equation satisfies one additional condition, which is fulfilled by every useful solution.

In order that the SOR method be competitive, it is important that the FFT can be applied too. The corresponding algorithm is described in detail in [10]. (Independently, this has been discovered by Hübner [14] too, who, using our Theorem 3.4, has also established the local convergence of the SOR method.)

In another related paper [12] we show that the FFT can also be applied if the discretization of Theodorsen's integral equation is based on spline functions (or on another one of a large number of families of approximants). Wittich's matrix $\mathbf{K}$ is then replaced by a product $\mathbf{K\Sigma}$, where $\mathbf{\Sigma}$ is a symmetric circulant Toeplitz matrix. Unfortunately, this modification prevents the direct application of the SOR iteration, but we present other methods that work. Also, Theorem 3.4 seems to hold no more in general.

All the numerical methods we discuss require the simultaneous evaluation of a conjugate periodic function at each step and allow one to apply the FFT for this. All of them are linear-nonlinear methods [20] for which no secondary iteration is necessary due to the special structure of the discrete Theodorsen equation. (But some of our results are also useful if the linear versions of these methods are used as secondary iterations in Newton's method, cf. [20].) In particular, we consider the Jacobi overrelaxation method (JOR), the SOR

iteration, a second order generalized Euler method (which turns out to be a second order Richardson iteration too), and, for completeness, the related Chebyshev semi-iterative method and its cyclic variant. Under the mentioned assumptions of Theorem 3.4 we are able in each case to determine a priori nearly optimal parameters (e.g., underrelaxation factors) and the corresponding asymptotic (root-)convergence factor. If these assumptions are violated, it is still possible to find the optimal parameters with respect to an elliptic region (or, for the JOR method, the intersection of two disks) that is supposed to contain the eigenvalues of $\mathbf{J}$.

The SOR and the cyclic Chebyshev iterations have the best convergence rate, but they are only applicable for trigonometric interpolation, i.e. $\Sigma = \mathbf{I}$. The second order Euler method and the Chebyshev iteration, both allowing $\Sigma \neq \mathbf{I}$, are approximately half as efficient. Moreover, as partly known [7, 8, 15, 19], all of these methods are closely related to each other in our case.

If $\varepsilon_\rho < 1$, global convergence is established in Theorems 3.1 and 7.3. If $\varepsilon_\rho > 1$, however, our results only guarantee local convergence. In fact, our numerical experiments have shown that it is often necessary to combine those methods with the so-called continuation method [20]: One after the other we solve the conformal mapping problems with radius function $\rho^{l/L}(\tau)$, $l = 1, \ldots, L$, using the result of each problem as starting point for the next one.

Some other questions arising in practice, including the one of how to evaluate the mapping function in the interior of the unit disk, will be discussed in another closely related paper [11]. There we will also present numerical experiments. They indicate that in practice our choice of parameters based on Theorem 3.4 is still very good if the assumptions of this theorem do not hold.

2. Theodorsen's Integral Equation and its Discretization

Let Γ be a Jordan curve that is starlike with respect to the origin and given by its polar coordinates $\tau, \rho(\tau)$:

$$\Gamma := \{\zeta \in \mathbb{C}: \ \zeta = \rho(\tau)\, e^{i\tau}, \ \tau \in \mathbb{R}\}.$$

(ρ is a 2π-periodic positive function.) Assume Γ satisfies an ε-condition, i.e. ρ is absolutely continuous and

$$\varepsilon_\rho := \operatorname*{ess\,sup}_{0 \leq \tau < 2\pi} \left| \frac{\rho'(\tau)}{\rho(\tau)} \right| < \infty.$$

Moreover, suppose D is the open unit disk in the w-plane and Δ is the interior of Γ in the ζ-plane. Let $\bar{D}$ and $\bar{\Delta}$ denote the closure of D and Δ, respectively. There is a unique topological mapping $g: \bar{D} \to \bar{\Delta}$ that is the extension of a conformal mapping of D onto Δ and fulfills $g(0) = 0$, $g'(0) > 0$. The restriction of g to the unit circle ∂D may be written in terms of the *boundary correspondence function* θ, a real continuous function that is implicitly defined by

$$g|_{\partial D}: \ e^{it} \mapsto g(e^{it}) = \rho(\theta(t))\, e^{i\theta(t)},$$

and may be normalized by

$$\int_0^{2\pi} \theta(t)\,dt = 2\pi^2.$$

This boundary correspondence function satisfies Theodorsen's equation, a singular and nonlinear integral equation [4, p. 65]:

$$\theta(t) - t = K[\log \rho(\theta(.))](t), \qquad \forall t \in \mathbb{R}. \tag{2.1}$$

Here K denotes the *conjugation operator* [28]: For $X \in L[0, 2\pi]$ the *conjugate periodic function* KX is defined a.e. by the principal value integral

$$KX(t) := \frac{1}{2\pi} \text{P.V.} \int_0^{2\pi} X(s) \cot \frac{t-s}{2}\,ds.$$

The restriction of K to $L_2[0, 2\pi]$ is a skew-symmetric endomorphism with operator norm 1. In fact, if $X \in L_2[0, 2\pi]$ has the real Fourier coefficients $a_0, a_1, a_2, \ldots, b_1, b_2, \ldots$, then $KX \in L_2[0, 2\pi]$ has the coefficients $0, -b_1, -b_2, \ldots, a_1, a_2, \ldots$ [28, p. 128].

For numerical computation one has to discretize the conjugation operator. The most common way to do this is based on replacing X by the trigonometric polynomial T of degree N normalized by a vanishing $\sin(Nt)$ coefficient and interpolating X at the $2N$ equidistant points $t_k := k\pi/N$, $k = 0, \ldots, 2N-1$. The coefficients of T as well as all the values $y_k := (KT)(t_k)$ can be computed very efficiently with the fast Fourier transform (FFT) [10, 13].

On the other hand, since y_k depends linearly on the given values $x_j := X(t_j)$, $j = 0, \ldots, 2N-1$, there exists a matrix $\mathbf{K}$ of order $2N$ such that $\mathbf{y} = \mathbf{K}\mathbf{x}$ if $\mathbf{y} := (y_0, \ldots, y_{2N-1})^T$, $\mathbf{x} := (x_0, \ldots, x_{2N-1})^T$ (where T denotes transposition). For our theoretical investigations we will use this classical matrix representation of the discrete conjugation operator, although it hides the connection with trigonometric interpolation and the related efficient implementation with the FFT.

$\mathbf{K}$ is called Wittich's matrix and has the elements [4, p. 76]

$$(\mathbf{K})_{lj} = \begin{cases} 0, & \text{if } j-l \text{ is even,} \\ \dfrac{1}{N} \cot \dfrac{(l-j)\pi}{2N}, & \text{if } j-l \text{ is odd} \end{cases}$$

$(l, j = 0, \ldots, 2N-1)$. It is a circulant skew-symmetric Toeplitz matrix, which contains zeros in a checkerboard layout including the diagonal. Due to this last property $\mathbf{K}$ is weakly cyclic of index 2 [24], cf. Sect. 5. Owing to the first property it is easy to show that all non-vanishing eigenvalues are $\pm i$, namely, $N-1$ of each [27, § 17.6]. Hence $\|\mathbf{K}\| = 1$ for the spectral norm.

For any $\mathbf{s} \in \mathbb{R}^{2N}$ and any scalar function σ let us define

$$\sigma(\mathbf{s}) := (\sigma(s_0), \sigma(s_1), \ldots, \sigma(s_{2N-1}))^T.$$

Then, the *discrete Theodorsen equation* replacing Eq. (2.1) is

$$\mathbf{y} = \mathbf{K} \log \rho(\mathbf{t} + \mathbf{y}), \tag{2.2}$$

where $\mathbf{y}$ is supposed to approximate $\theta(\mathbf{t}) - \mathbf{t}$.

It has been proved by the author [9], and in a private communication by O. Hübner, that Eq. (2.2) always has at least one solution. However, as we will see in computed examples [11], some solutions may yield useless approximations of the exact boundary correspondence function θ, which, of course, is strictly monotone in $[0, 2\pi]$.

We may equally well approximate X by an interpolating spline function with knots t_k instead of using the interpolating trigonometric polynomial T. In fact, this just means that the coefficients of the polynomial KT have to be modified before its evaluation with the FFT [12]. We can even implement a number of other approximation functions with the same modification but using different weights. In the sequel we will only use the fact that $\mathbf{K}$ is then replaced by

$$\mathbf{K}_\Sigma := \mathbf{K}\boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is a circulant positive semi-definite symmetric Toeplitz matrix commuting with $\mathbf{K}$ and satisfying

$$\|\boldsymbol{\Sigma}\| \leqq 1.$$

Thus, $\mathbf{K}_\Sigma$ is still a circulant skew-symmetric Toeplitz matrix, but, in general, it is no longer weakly cyclic of index 2. We will call

$$\mathbf{y} = \mathbf{K}_\Sigma \log \rho(\mathbf{t} + \mathbf{y}) \tag{2.3}$$

the *modified discrete Theodorsen equation*. Equation (2.2) corresponding to $\boldsymbol{\Sigma} = \mathbf{I}$ (the identity matrix) will be considered as a special case of (2.3). The existence proof for (2.2) carries over to (2.3). As we will see, the constant ε_ρ, which is very important in the discussion of numerical methods for (2.2), is replaced by

$$\varepsilon := \varepsilon_\rho \|\mathbf{K}_\Sigma\|$$

when we are dealing with (2.3).

3. On the Jacobi Iteration Function and the Eigenvalues of its Fréchet Derivative

The classical method to solve Eq. (2.2), or now (2.3), is the *nonlinear Jacobi iteration*

$$\mathbf{y}_{m+1} := \boldsymbol{\Phi}(\mathbf{y}_m) := \mathbf{K}_\Sigma \log \rho(\mathbf{t} + \mathbf{y}_m), \qquad m = 0, 1, \ldots. \tag{3.1}$$

It may be started with $\mathbf{y}_0 := \mathbf{0}$ or with another approximation of the solution.

Since ρ is absolutely continuous, ρ' exists a.e. and

$$|\log \rho(\tau_1) - \log \rho(\tau_2)| = \left| \int_{\tau_1}^{\tau_2} \frac{\rho'(\tau)}{\rho(\tau)} d\tau \right| \leqq \varepsilon_\rho |\tau_1 - \tau_2| \tag{3.2}$$

for every τ_1 and τ_2. Using the Euclidean norm we get

$$\|\boldsymbol{\Phi}(\mathbf{y}) - \boldsymbol{\Phi}(\tilde{\mathbf{y}})\| \leqq \varepsilon_\rho \|\mathbf{K}_\Sigma\| \, \|\mathbf{y} - \tilde{\mathbf{y}}\| = \varepsilon \|\mathbf{y} - \tilde{\mathbf{y}}\|, \tag{3.3}$$

which means that ε is a Lipschitz constant for the iteration function $\boldsymbol{\Phi}$. Thus, the contraction theorem yields the following simple generalization of the classical convergence theorem [4, p. 87].

Theorem 3.1. *If* $\varepsilon<1$, *Eq.* (2.3) *has exactly one solution* $\mathbf{y}^*$, *and iteration* (3.1) *started with an arbitrary* $\mathbf{y}_0$ *converges to* $\mathbf{y}^*$.

But what can we say if $\varepsilon\geqq 1$? For any iteration $\mathbf{y}_{m+1}=\boldsymbol{\Psi}(\mathbf{y}_m)$ with Fréchet differentiable (F-differentiable) function $\boldsymbol{\Psi}$ the local convergence to a fixed point $\mathbf{y}^*$ of $\boldsymbol{\Psi}$ hinges on the spectral radius Λ of the Fréchet derivative (F-derivative) $\boldsymbol{\Psi}'(\mathbf{y}^*)$ of $\boldsymbol{\Psi}$ at $\mathbf{y}^*$ [21, Chap. 22]. Moreover, if $\Lambda\in(0,1)$, the convergence is linear and the asymptotic convergence factor (called root-convergence factor R_1 in [20]) equals Λ [20, 10.1.4]. Since most of our convergence results will be based on these facts, we need

Assumption (D). $\mathbf{y}^*$ is a solution of (2.3), and ρ is differentiable at each component of $\mathbf{y}^*$.

First, it follows that the F-derivative at $\mathbf{y}^*$ of the Jacobi iteration function $\boldsymbol{\Phi}$ [defined in (3.1)] exists:

$$\mathbf{J}:=\boldsymbol{\Phi}'(\mathbf{y}^*)=\mathbf{K}_\Sigma\mathbf{D},$$

where (3.4)

$$\mathbf{D}:=\operatorname{diag}\left(\frac{\rho'}{\rho}(t_k+y_k^*)\right)_{k=0,\ldots,2N-1}.$$

[Note that even if ρ were not differentiable at $\mathbf{y}^*$, there would exist according to (3.2) a diagonal matrix $\tilde{\mathbf{D}}_m$ with $\|\tilde{\mathbf{D}}_m\|\leqq\varepsilon_\rho$ such that

$$\log\rho(\mathbf{t}+\mathbf{y}_m)-\log\rho(\mathbf{t}+\mathbf{y}^*)=\tilde{\mathbf{D}}_m(\mathbf{y}_m-\mathbf{y}^*),$$

and hence

$$\mathbf{y}_{m+1}-\mathbf{y}^*=\mathbf{K}_\Sigma\tilde{\mathbf{D}}_m(\mathbf{y}_m-\mathbf{y}^*). \tag{3.5}$$

Our following investigations on the eigenvalues of $\mathbf{J}$ carry over to the equally structured matrices $\mathbf{K}_\Sigma\tilde{\mathbf{D}}_m$. This motivates our opinion that, practically, the differentiability of ρ is not crucial for the local convergence of the nonlinear methods to be discussed.]

Of course, the spectral radius $\Lambda(\mathbf{J})$ of $\mathbf{J}$ satisfies $\Lambda(\mathbf{J})\leqq\|\mathbf{J}\|\leqq\varepsilon$. In the case $\boldsymbol{\Sigma}=\mathbf{I}$ numerical experiments have shown that, in general, $\Lambda(\mathbf{J})$ is almost equal to ε if N is large. Consequently, if $\varepsilon<1$, the convergence is indeed linear with an asymptotic factor close to ε. However, if $\varepsilon>1$ and N is large, $\mathbf{y}^*$ is usually a point of repulsion; in its neighborhood iteration (3.1) can be expected to diverge, although convergence could occur [21, Chap. 22]. In Appendix 1 we will prove the related

Lemma 3.2. *Under assumption* (D), *if* $N\to\infty$ *and the third largest absolute value of an element of* $\mathbf{D}$ *tends to* ε_ρ, *then* $\lim\|\mathbf{K}\mathbf{D}\|=\varepsilon_\rho$.

In particular, this lemma applies if ρ is continuously differentiable and the solution $\mathbf{y}^*$ of (2.2) "converges" to the function $\theta(t)-t$. (A correct formulation of this fact would require additional notation.)

The question, whether the spectral radius $\Lambda(\mathbf{KD})$ tends to ε_ρ too, is still open.

To choose a numerical method that is more appropriate than Jacobi's iteration (3.1) requires additional information on the spectrum of $\mathbf{J}$.

Lemma 3.3. *Under assumption* (D) *the spectrum of* $\mathbf{J}$ *is symmetric about* 0.

Proof. If $\mathbf{D}$ is regular, $\mathbf{W}:=\mathbf{D}^{1/2}$ can be defined and is regular. So, $\mathbf{WJW}^{-1}=\mathbf{WK}_\Sigma\mathbf{W}$ is a complex skew-symmetric matrix similar to $\mathbf{J}$, and its spectrum is symmetric about 0 [5, p. 12]. Since the eigenvalues of $\mathbf{J}$ depend continuously on the elements of $\mathbf{D}$, the statement remains true for singular $\mathbf{D}$. □

(In case $\Sigma=\mathbf{I}$, there is another simple proof based on the fact that $\mathbf{J}$ is weakly cyclic of index 2, cf. [27, § 17.4].)

Additional information on the spectrum of $\mathbf{J}$ can be deduced for some boundaries Γ symmetric about an axis. More precisely, we have to replace assumption (D) by

Assumption (SD). Γ is symmetric about the real axis and, in addition, v-fold rotationally symmetric about 0, where $v>1$ is a divisor of N, $M:=N/v$. $\mathcal{S}$ is the $(M-1)$-dimensional subspace of vectors $\mathbf{y}\in\mathbb{R}^{2N}$ satisfying

$$y_0=y_N=0, \qquad y_k=-y_{2N-k}, \qquad k=1,\ldots,N-1; \tag{3.6a}$$

$$y_k=y_{k+2M}, \qquad k=0,\ldots,2N-2M-1 \quad \text{if } v>1. \tag{3.6b}$$

$\mathbf{y}^*\in\mathcal{S}$ is a solution of (2.3), for which

$$t_k+y_k^*\in(0,\pi/v), \qquad k=1,\ldots,M-1. \tag{3.7}$$

Moreover, ρ is differentiable at $t_k+y_k^*$, $k=1,\ldots,M-1$, and $\rho'(t_k+y_k^*)\geqq 0$ (or $\leqq 0$) for these values of k. Finally, $\Sigma=\mathbf{I}$.

Of course, the last condition is fulfilled if ρ is differentiable and weakly monotone in $(0,\pi/v)$. Simple examples of curves Γ satisfying this and the symmetry condition are: an ellipse ($v=2$) or a square ($v=4$) symmetric about real and imaginary axis, and the curve defined by $\rho(\tau)=\exp|\tau|$, $|\tau|\leqq\pi$, for which $v=1$.

As already noticed by Niethammer [18], there are good reasons to believe that the eigenvalues of $\mathbf{J}$ are often close to the imaginary axis. Gekeler [6], assuming (SD) with $v=2$, established a rather weak related result. In Appendix 2 we will prove the much stronger

Theorem 3.4. *Under assumption* (SD) $\mathcal{S}$ *is an invariant subspace of* $\mathbf{J}$ [*defined by* (3.4)], *and the restriction* $\mathbf{J}|_{\mathcal{S}}$ *of* $\mathbf{J}$ *to* $\mathcal{S}$ *has purely imaginary eigenvalues.*

In the following sections we will discuss the impact of this theorem on several numerical methods for the solution of (2.3) or (2.2). We will also consider weaker assumptions on the eigenvalues of $\mathbf{J}$. An important but simple tool in applying Theorem 3.4 will be

Lemma 3.5. *Assume* $\mathcal{S}$ *is any subspace of* $\mathbb{R}^n$, *and* $\Psi\colon\mathbb{R}^n\to\mathbb{R}^n$ *is F-differentiable at* $\mathbf{y}\in\mathcal{S}$ *and satisfies* $\Psi(\mathcal{S})\subset S$. *Denote the restriction of* Ψ *and* $\Psi'(\mathbf{y})$ *to* $\mathcal{S}$ *by* $\Psi|_{\mathcal{S}}$ *and* $\Psi'(\mathbf{y})|_{\mathcal{S}}$, *respectively. Then* $\Psi'(\mathbf{y})\mathcal{S}\subset\mathcal{S}$ *and* $\Psi'(\mathbf{y})|_{\mathcal{S}}=(\Psi|_{\mathcal{S}})'(\mathbf{y})$.

Proof. By definition of the F-derivative $\boldsymbol{\Psi}'(\mathbf{y})$ of $\boldsymbol{\Psi}$ at $\mathbf{y}$,

$$\|\boldsymbol{\Psi}(\mathbf{y}+\mathbf{h})-\boldsymbol{\Psi}(\mathbf{y})-\boldsymbol{\Psi}'(\mathbf{y})\,\mathbf{h}\|=o(\|\mathbf{h}\|) \quad \text{as } \mathbf{h}\to 0.$$

Assuming $\mathbf{y}\in\mathscr{S}$, $\mathbf{h}\in\mathscr{S}$ we see that $\boldsymbol{\Psi}'(\mathbf{y})\,\mathbf{h}\in\mathscr{S}$ too since the left side would be $O(\|\mathbf{h}\|)$ otherwise. Hence, $\boldsymbol{\Psi}'(\mathbf{y})\,\mathscr{S}\subset\mathscr{S}$ and

$$\|\boldsymbol{\Psi}|_{\mathscr{S}}(\mathbf{y}+\mathbf{h})-\boldsymbol{\Psi}|_{\mathscr{S}}(\mathbf{y})-\boldsymbol{\Psi}'(\mathbf{y})|_{\mathscr{S}}\,\mathbf{h}\|=o(\|\mathbf{h}\|).$$

So, $\boldsymbol{\Psi}'(\mathbf{y})|_{\mathscr{S}}=(\boldsymbol{\Psi}|_{\mathscr{S}})'(\mathbf{y})$ by definition of the latter. □

In particular, Lemma 3.5 applies to the iteration function $\boldsymbol{\Phi}$ of (3.1) and the subspace $\mathscr{S}$ defined by (3.6) for it will be proved in Appendix 2 that $\boldsymbol{\Phi}(\mathscr{S})\subset\mathscr{S}$.

4. The Nonlinear JOR Method

The *nonlinear Jacobi iteration with overrelaxation (nonlinear* JOR *iteration)* to solve (2.3) is defined by

$$\mathbf{y}_{m+1}:=\boldsymbol{\Phi}_\omega^J(\mathbf{y}_m):=\omega\,\mathbf{K}_\Sigma\log\rho(\mathbf{t}+\mathbf{y}_m)+(1-\omega)\,\mathbf{y}_m,\ m=0,1,2,\ldots, \tag{4.1}$$

where $\mathbf{y}_0$ is a given starting approximation and ω is the overrelaxation factor. We will choose $\omega\in(0,1]$, i.e. apply *underrelaxation* rather than overrelaxation, but nevertheless use the abbreviation JOR.

The eigenvalues μ of $\mathbf{J}$ [defined by (3.4)] and the eigenvalues λ of the F-derivative

$$\mathbf{J}_\omega^J:=\omega\,\mathbf{J}+(1-\omega)\,\mathbf{I} \tag{4.2}$$

of $\boldsymbol{\Phi}_\omega^J$ at $\mathbf{y}^*$ are related by

$$\lambda=\omega\,\mu+1-\omega. \tag{4.3}$$

Again, we use $\Lambda(\mathbf{J}_\omega^J)$ to denote the spectral radius of $\mathbf{J}_\omega^J$.

Theorem 4.1. (i) *Let* $\omega\in(0,1]$, $\sigma\in(1-\omega,1)$, *and let assumption* (D) *be satisfied. Then* $\Lambda(\mathbf{J}_\omega^J)\leqq\sigma$ *iff all eigenvalues* μ *of* $\mathbf{J}$ *lie in the point set*

$$L_{\sigma,\omega}^J:=\left\{\zeta:\ \left|\zeta+\frac{1}{\omega}-1\right|\leqq\frac{\sigma}{\omega},\ \left|\zeta-\frac{1}{\omega}+1\right|\leqq\frac{\sigma}{\omega}\right\},$$

which is an intersection of two disks with radius σ/ω *and center* $\pm(1-\omega)/\omega$.

(ii) *Under assumption* (SD),

$$\Lambda(\mathbf{J}_\omega^J|_{\mathscr{S}})\leqq\sigma:=\sigma_J(\varepsilon,\omega):=|1-\omega+i\,\varepsilon\,\omega|,$$

and

$$\sigma_J(\varepsilon,\omega)<1 \quad \textit{iff} \quad 0<\omega<\frac{2}{1+\varepsilon^2}.$$

Moreover,

$$\min_\omega\sigma_J(\varepsilon,\omega)=\sigma_J^*(\varepsilon):=\frac{\varepsilon}{\sqrt{1+\varepsilon^2}}$$

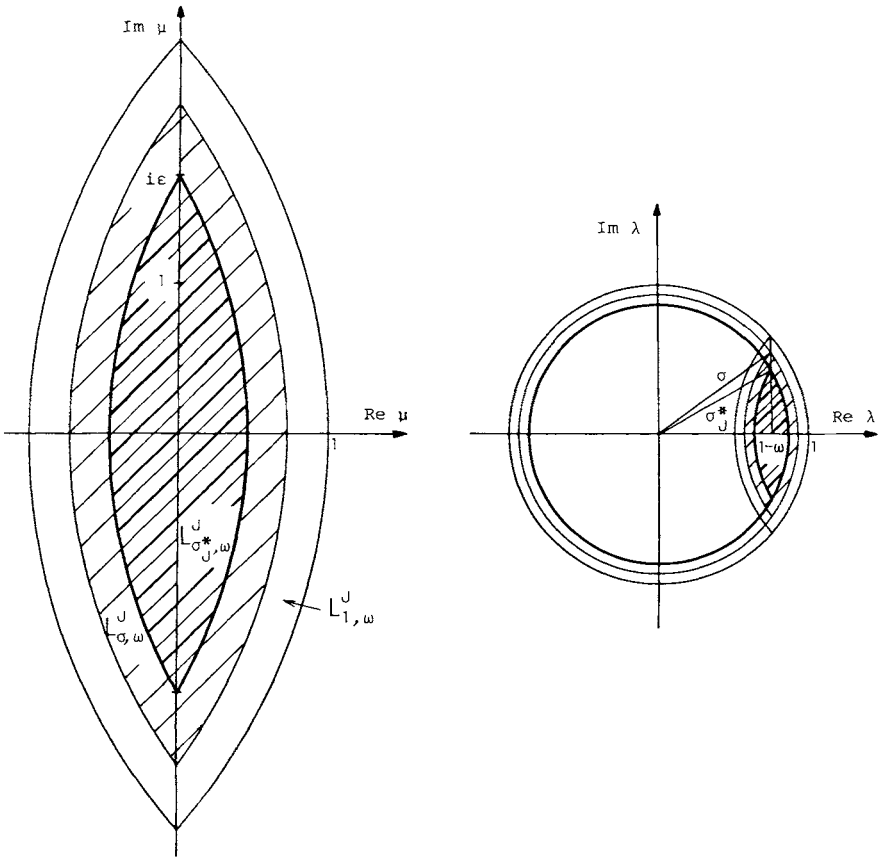

Fig. 1. The mapping of eigenvalues by the JOR method [$\varepsilon = 1/3$, $\omega = \omega_J^*(\varepsilon)$]

is attained at

$$\omega = \omega_J^*(\varepsilon) := \frac{1}{1+\varepsilon^2},$$

and, for $\varepsilon \to \infty$,

$$\sigma_J^*(\varepsilon) = 1 - \frac{1}{2\varepsilon^2} + O(\varepsilon^{-4}), \qquad \omega_J^*(\varepsilon) = \frac{1}{\varepsilon^2} + O(\varepsilon^{-4}).$$

(iii) *In both cases, iteration* (4.1) *converges locally and (except if* $\varepsilon = \omega - 1 = 0$*) linearly to* $\mathbf{y}^*$*, and the asymptotic convergence factor is not worse than* σ.

Proof. Relation (4.3) and elementary geometrical considerations easily yield part (i); cf. Fig. 1. To obtain (ii) we note that $\boldsymbol{\Phi}_\omega^J(\mathscr{S}) \subset \mathscr{S}$ (since $\boldsymbol{\Phi}(\mathscr{S}) \subset \mathscr{S}$) and apply Lemma 3.5 and Theorem 3.4 in addition. Finally, (iii) follows from Ostrowski's theorem and the linear convergence theorem [21, Chap. 22; 20, 10.1.4]. □

$\omega_J^*(\varepsilon)$ is the optimal underrelaxation factor with respect to our knowledge (exhibited in Theorem 3.4) about the eigenvalues of $\mathbf{J}$.

By geometric considerations and by a compactness argument we easily get

Corollary 4.2. (i) *If* $\omega \in (0, 1]$, *if assumption* (D) *holds, and if the interior of* $L^J_{1,\omega}$ *contains the spectrum of* $\mathbf{J}$, *then iteration* (4.1) *converges locally to* $\mathbf{y}^*$.

(ii) *If assumption* (D) *holds and the spectrum of* $\mathbf{J}$ *is contained in the strip* $\{\mu: |\mathrm{Re}\,\mu| < 1\}$, *then there exists* $\omega \in (0, 1]$ *such that* (4.1) *converges locally to* $\mathbf{y}^*$.

Finally, note that $\varepsilon < 1$, $\omega \leqq 1$, and (3.3) imply

$$\|\boldsymbol{\Phi}^J_\omega(\mathbf{y}) - \boldsymbol{\Phi}^J_\omega(\tilde{\mathbf{y}})\| \leqq \gamma \|\mathbf{y} - \tilde{\mathbf{y}}\|, \quad \gamma := 1 - \omega + \omega\varepsilon < 1. \tag{4.4}$$

Hence, $\gamma < 1$ is a Lipschitz constant for $\boldsymbol{\Phi}^J_\omega$, and we get

Theorem 4.3. *For* $\varepsilon < 1$ *iteration* (4.1) *converges globally.*

5. The Nonlinear SOR Method

Niethammer [18] showed that in the traditional case with $\boldsymbol{\Sigma} = \mathbf{I}$ two steps of the nonlinear Jacobi iteration (3.1) are essentially equivalent to one step of the nonlinear Gauss-Seidel method if for the latter the equations and unknowns in system (2.2) are suitably permuted. Moreover, on the basis of examples and his conjecture (mentioned in Sect. 3) concerning the eigenvalues of $\mathbf{J}$ he pointed out that underrelaxation may improve the convergence of the Gauss-Seidel method considerably. Owing to our Theorem 3.4 we are now able to prove his observation. We restrict ourselves to the case $\boldsymbol{\Sigma} = \mathbf{I}$ (as he did) since this is the only practical approximation that can be treated efficiently with the SOR method.

For any $\mathbf{y} = (y_0, y_1, \ldots, y_{2N-1})^T \in \mathbb{R}^{2N}$ we let

$$\mathbf{y}'' := (y_0, y_2, \ldots, y_{2N-2})^T, \qquad \mathbf{y}' := (y_1, y_3, \ldots, y_{2N-1})^T,$$

and define the permutation matrix $\mathbf{P}$ by

$$\mathbf{P}\,\mathbf{y} = \begin{pmatrix} \mathbf{y}'' \\ \mathbf{y}' \end{pmatrix}, \qquad \forall\, \mathbf{y} \in \mathbb{R}^{2N}. \tag{5.1}$$

Then, since $\mathbf{K}$ features zeros in a checkerboard layout,

$$\mathbf{P}\mathbf{K}\mathbf{P}^T = \begin{pmatrix} \mathbf{0} & -\mathbf{L}^T \\ \mathbf{L} & \mathbf{0} \end{pmatrix}, \tag{5.2}$$

where $\mathbf{L}$ is an $N \times N$ matrix. So, the discrete Theodorsen equation (2.2) is equivalent to the system

$$\mathbf{y}'' = -\mathbf{L}^T \log \rho(\mathbf{t}' + \mathbf{y}'), \tag{5.3a}$$

$$\mathbf{y}' = \mathbf{L} \log \rho(\mathbf{t}'' + \mathbf{y}''). \tag{5.3b}$$

(5.2) exhibits that $\mathbf{K}$ and $\mathbf{J} = \mathbf{K}\mathbf{D}$ are weakly cyclic of index 2 [24, p. 39], which implies that $\mathbf{I} - \mathbf{J} = \mathbf{I} - \mathbf{K}\mathbf{D}$ is 2-cyclic and has property A [24, p. 99]. However, opposite to $\mathbf{J}$, the matrix $\mathbf{P}\mathbf{J}\mathbf{P}^T$ is consistently ordered with the σ_1 ordering [25]. Since $\mathbf{P}\mathbf{J}\mathbf{P}^T$ is the F-derivative of the right side of (5.3) (considered

as function of $\mathbf{y}''$ and $\mathbf{y}'$), Young's SOR theory [25] can be applied for the computation of a nearly optimal relaxation factor and the correspondent asymptotic convergence factor of either the nonlinear SOR method or the Newton-SOR method.

To apply the *nonlinear SOR method* to (5.3) means to compute

$$\mathbf{y}''_{m+1} := -\omega \mathbf{L}^T \log \rho(\mathbf{t}' + \mathbf{y}'_m) + (1-\omega)\mathbf{y}''_m, \tag{5.4a}$$

$$\mathbf{y}'_{m+1} := \omega \mathbf{L} \log \rho(\mathbf{t}'' + \mathbf{y}''_{m+1}) + (1-\omega)\mathbf{y}'_m, \tag{5.4b}$$

$$m = 0, 1, \ldots,$$

where $\mathbf{y}'_0$, and, if $\omega \neq 1$, $\mathbf{y}''_0$ are given starting approximations. In practice it is most important that there exist a fast algorithm based on the FFT for the multiplication of a vector in $\mathbb{R}^N$ by $\mathbf{L}$ or $\mathbf{L}^T$; for large N each of these multiplications costs only half as much as the multiplication of a vector in $\mathbb{R}^{2N}$ by $\mathbf{K}$. This algorithm and an ALGOL 60 program are listed in [10]. (Independently, this possibility has also been recognized by Hübner [14].)

If $\omega = 1$, $\mathbf{y}''_m$ is now in fact seen to consist of the even indexed components of $\mathbf{y}_{2m-1}$, constructed by Jacobi's method (3.1), while $\mathbf{y}'_m$ contains exactly the odd indexed components of $\mathbf{y}_{2m}$ [18]. However, the effect of underrelaxation on the Gauss-Seidel method is much greater than on Jacobi's method.

The iteration function $\boldsymbol{\Phi}^S_\omega$ of the nonlinear SOR method becomes evident, if we replace in (5.4) the second definition (5.4b) by the equivalent assignment

$$\mathbf{y}'_{m+1} := \omega \mathbf{L} \log \rho[\mathbf{t}'' - \omega \mathbf{L}^T \log \rho(\mathbf{t}' + \mathbf{y}'_m) + (1-\omega)\mathbf{y}''_m] + (1-\omega)\mathbf{y}'_m. \tag{5.4b'}$$

In particular, the F-derivative of $\boldsymbol{\Phi}^S_\omega$ at $\mathbf{y}^*$ is

$$\mathbf{J}^S_\omega := \begin{pmatrix} (1-\omega)\mathbf{I} & -\omega \mathbf{L}^T \mathbf{D}' \\ \omega(1-\omega)\mathbf{L}\mathbf{D}'' & (1-\omega)\mathbf{I} - \omega^2 \mathbf{L}\mathbf{D}''\mathbf{L}^T\mathbf{D}' \end{pmatrix}, \tag{5.5}$$

where $\mathbf{D}''$ and $\mathbf{D}'$ are in an obvious way defined as submatrices of $\mathbf{P}\mathbf{D}\mathbf{P}^T$. Formally a more common formula for $\mathbf{J}^S_\omega$ is obtained if we linearize the right side of (5.3) at $y = y^*$ and delete higher order terms:

$$\begin{aligned} \mathbf{y}'' &= -\mathbf{L}^T\mathbf{c}' - \mathbf{L}^T\mathbf{D}'\mathbf{y}', \quad \text{where } \mathbf{c}' := \log \rho(\mathbf{t}' + \mathbf{y}^{*\prime}) - \mathbf{D}'\mathbf{y}^{*\prime}, \\ \mathbf{y}' &= \mathbf{L}\mathbf{c}'' + \mathbf{L}\mathbf{D}''\mathbf{y}'', \quad \text{where } \mathbf{c}'' := \log \rho(\mathbf{t}'' + \mathbf{y}^{*\prime\prime}) - \mathbf{D}''\mathbf{y}^{*\prime\prime}. \end{aligned} \tag{5.6}$$

The SOR iteration for this linear system would be

$$\begin{aligned} \mathbf{y}''_{m+1} &:= -\omega \mathbf{L}^T\mathbf{c}' - \omega \mathbf{L}^T\mathbf{D}'\mathbf{y}'_m + (1-\omega)\mathbf{y}''_m, \\ \mathbf{y}'_{m+1} &:= \omega \mathbf{L}\mathbf{c}'' + \omega \mathbf{L}\mathbf{D}''\mathbf{y}''_{m+1} + (1-\omega)\mathbf{y}'_m, \quad m = 0, 1, \ldots \end{aligned} \tag{5.7}$$

By noting that the inverse of a triangular matrix with diagonal elements equal to 1 always exists, we get now (cf. [20], Th. 10.3.5)

$$\mathbf{J}^S_\omega = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\omega \mathbf{L}\mathbf{D}'' & \mathbf{I} \end{pmatrix}^{-1} \begin{pmatrix} (1-\omega)\mathbf{I} & -\omega \mathbf{L}^T\mathbf{D}' \\ \mathbf{0} & (1-\omega)\mathbf{I} \end{pmatrix},$$

which is easily verified to be in accordance with (5.5).

Young [25, Theorem 2.3] showed: If $\omega \neq 0$, if λ is a non-zero eigenvalue of $\mathbf{J}_\omega^S$, and if μ satisfies

$$(\lambda+\omega-1)^2=\omega^2\mu^2\lambda, \tag{5.8}$$

then μ is an eigenvalue of $\mathbf{J}$. Conversely, if μ is an eigenvalue of $\mathbf{J}$ and λ satisfies (5.8), then λ is an eigenvalue of $\mathbf{J}_\omega^S$.

A further discussion of this relationship (for details see [19], Sect. 6, [25], Sect. 2, or [26], or Sect. 6.4) leads to part (i) illustrated by Fig. 2 of

Theorem 5.1. (i) *Let* $\omega\in(0,1]$, $\sigma\in[1-\omega, 1)$, *but not* $\omega-1=\sigma=0$. *Under assumption* (D), *iteration* (5.4) *converges linearly with an asymptotic convergence factor not greater than* σ *locally to* $\mathbf{y}^*$ *iff all eigenvalues* μ *of* $\mathbf{J}$ *lie in the closed point set* $L^S_{\sigma,\omega}$ *bounded by the ellipse* $E_{\gamma,\delta}$ *with foci* $\pm i\gamma$ *and semi-axes*

$$\alpha:=\frac{\gamma}{2}\left(\frac{1}{\delta}+\delta\right), \qquad \beta:=\frac{\gamma}{2}\left(\frac{1}{\delta}-\delta\right), \tag{5.9}$$

where γ, δ, σ, *and* ω *are related by*

$$\begin{aligned} &\gamma=\frac{2}{\omega}\sqrt{1-\omega}, \quad \text{i.e.} \quad \omega=\frac{2}{1+\sqrt{1+\gamma^2}},\\ &\delta=\sqrt{\frac{1-\omega}{\sigma}}\,(\leqq 1), \quad \text{i.e.} \quad \sigma=\frac{1-\omega}{\delta^2}. \end{aligned} \tag{5.10}$$

(*In particular,* $\alpha=\beta=\sqrt{\sigma}$ *if* $\omega=1$, *while for* $\delta=1$ *the ellipse reduces to the segment with endpoints* $-i\gamma$ *and* $i\gamma$.)

(ii) *Under assumption* (SD), *we may choose* $\gamma=\varepsilon$ *and* $\delta=1$ *in* (i): *If*

$$\omega=\omega_S^*(\varepsilon):=\frac{2}{1+\sqrt{1+\varepsilon^2}},$$

iteration (5.4) *converges locally and, if* $\varepsilon>0$, *linearly to* $\mathbf{y}^*$; *the asymptotic convergence factor is*

$$\sigma_S^*(\varepsilon):=1-\omega_S^*(\varepsilon)=\frac{\varepsilon^2}{(1+\sqrt{1+\varepsilon^2})^2}<1.$$

For $\varepsilon\to\infty$,

$$\sigma_S^*(\varepsilon)=1-\frac{2}{\varepsilon}+O(\varepsilon^{-2}), \qquad \omega_S^*(\varepsilon)=\frac{2}{\varepsilon}+O(\varepsilon^{-2}).$$

Proof of part (ii). If we let

$$\mathscr{S}':=\{\mathbf{y}\in\mathscr{S}: y_{2k}=0, k=0,\ldots,N-1\}, \qquad \mathscr{S}'':=(\mathscr{S}')^\perp\subset\mathscr{S},$$

then $\dim\mathscr{S}'=[M/2]$, $\dim\mathscr{S}''=[(M-1)/2]$, and

$$\mathbf{J}:\mathscr{S}'\to\mathscr{S}'', \qquad \mathbf{J}:\mathscr{S}''\to\mathscr{S}'.$$

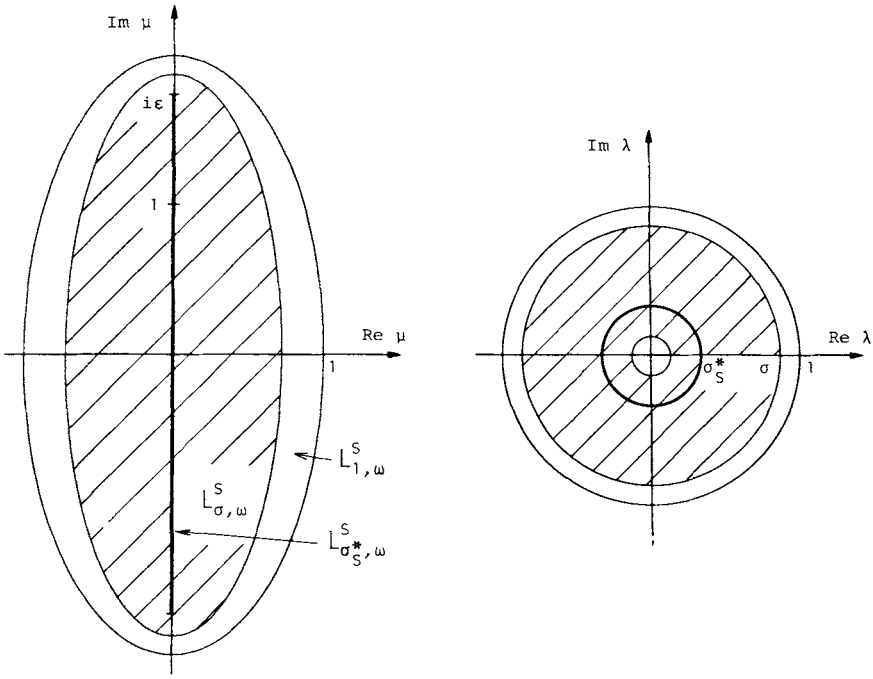

Fig. 2. The mapping of eigenvalues by the SOR method [$\gamma=\varepsilon=\sqrt{3}$, $\omega=\omega^*_S(\varepsilon)$]

The same holds for $\boldsymbol{\Phi}^S_\omega$ [defined by (5.4a) and (5.4b')] and for $\mathbf{J}^S_\omega$. If we use in $\mathscr{S}$ an orthogonal basis the first $[(M-1)/2]$ vectors of which are a basis of $\mathscr{S}''$, then the matrix corresponding to the restriction of the operator $\mathbf{J}$ to $\mathscr{S}$ is weakly cyclic of index 2 and consistently ordered. This fact is automatically capitalized upon if we use iteration (5.4) with a starting vector $\mathbf{y}_0\in\mathscr{S}$ (implying $\mathbf{y}_m\in\mathscr{S}$, $\forall m>0$); actually, if thought of as an iteration in $\mathscr{S}$, (5.4) just yields some additional redundant components of $\mathbf{y}_{m+1}$. But, as proposed by Lemma 3.5, the local convergence only depends on the eigenvalues of the restriction of $\mathbf{J}$ to $\mathscr{S}$, which are purely imaginary according to Theorem 3.4. Part (ii) then results from the linear theory [19, Theorem 6.3]. □

Using Niethammer's results [18, 19] Gekeler [6] has stated some facts related to our Theorem 5.1; however, part (ii) is an essential improvement of his result for doubly symmetric regions. On the basis of our Theorem 3.4, which was already presented in a previous version of this paper, Hübner [14] has independently proved part (ii) as well.

The underrelaxation factor ω used in Theorem 5.1 is optimal for every $\delta\in(0,1]$ with respect to the ellipse $E_{\gamma,\delta}$ [19, Theorem 6.3]. Iteration (5.4) would converge locally whenever the relaxation factor were chosen in the interval $(0, 2/(1+\alpha))$ [25, Corollary 3.1].

For $\sigma\to 1$ the parameters in (5.9) and (5.10) become

$$\gamma=\frac{2}{\omega}\sqrt{1-\omega},\quad \delta=\sqrt{1-\omega},\quad \alpha=\frac{2-\omega}{\omega},\quad \beta=1.$$

Using a compactness argument, we get in analogy to Corollary 4.2:

Corollary 5.2. (i) *If* $\omega \in (0, 1]$, *if assumption* (D) *holds, and if the spectrum of* $\mathbf{J}$ *lies inside the ellipse with foci* $\pm 2i(1-\omega)^{1/2}/\omega$ *and semi-axes* $\alpha = (2-\omega)/\omega$ *and* $\beta = 1$, *the iteration* (5.4) *converges locally to* $\mathbf{y}^*$.

(ii) *If assumption* (D) *holds and the spectrum of* $\mathbf{J}$ *is contained in the strip* $\{\mu : |\mathrm{Re}\,\mu| < 1\}$, *then there exists* $\omega \in (0, 1]$ *such that* (4.1) *converges locally to* $\mathbf{y}^*$.

If $\boldsymbol{\Sigma} \neq \mathbf{I}$ one might consider the nonlinear SOR iteration

$$\begin{aligned} \mathbf{y}_{m+1}^{(1)} &:= \omega \mathbf{K}_{\Sigma} \log \rho(\mathbf{t} + \mathbf{y}_m^{(2)}) + (1-\omega)\,\mathbf{y}_m^{(1)}, \\ \mathbf{y}_{m+1}^{(2)} &:= \omega \mathbf{K}_{\Sigma} \log \rho(\mathbf{t} + \mathbf{y}_{m+1}^{(1)}) + (1-\omega)\,\mathbf{y}_m^{(2)}, \qquad m = 0, 1, \ldots, \end{aligned} \tag{5.11}$$

where $\mathbf{y}_0^{(1)}$ and $\mathbf{y}_0^{(2)}$ are given (cf. [7] for the linear case). Every solution $\mathbf{y}$ of (2.3) defines a fixed point $\mathbf{y}^{(1)} := \mathbf{y}^{(2)} := \mathbf{y}$ of (5.11). But the converse need not be true since the coupled system

$$\begin{aligned} \mathbf{y}^{(1)} &= \mathbf{K}_{\Sigma} \log \rho(\mathbf{t} + \mathbf{y}^{(2)}) \\ \mathbf{y}^{(2)} &= \mathbf{K}_{\Sigma} \log \rho(\mathbf{t} + \mathbf{y}^{(1)}) \end{aligned} \tag{5.12}$$

might have additional solutions. The F-derivative of the right side of (5.12) is

$$\tilde{\mathbf{J}} = \begin{pmatrix} \mathbf{0} & \mathbf{K}_{\Sigma}\mathbf{D} \\ \mathbf{K}_{\Sigma}\mathbf{D} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{J} \\ \mathbf{J} & \mathbf{0} \end{pmatrix}. \tag{5.13}$$

By Lemma 3.3 and a standard argument based on forming $\tilde{\mathbf{J}}^2$ [27, § 17.4] it is easy to conclude that $\tilde{\mathbf{J}}$ has the same eigenvalues as $\mathbf{J}$ but with double geometric multiplicity. Hence, the F-derivative $\tilde{\mathbf{J}}_{\omega}^{S}$ of the iteration function corresponding to (5.11), which in analogy to (5.5) is

$$\tilde{\mathbf{J}}_{\omega}^{S} = \begin{pmatrix} (1-\omega)\mathbf{I} & \omega\mathbf{J} \\ \omega(1-\omega)\mathbf{J} & (1-\omega)\mathbf{I} + \omega^2\mathbf{J}^2 \end{pmatrix}, \tag{5.14}$$

has the same eigenvalues as $\mathbf{J}_{\omega}^{S}$ but with double multiplicity. We conclude

Theorem 5.3. *For iteration* (5.11), *which does not require* $\boldsymbol{\Sigma} = \mathbf{I}$, *Theorem* 5.1 *and Corollary* 5.2 *still hold.*

However, each step of iteration (5.11) costs approximately twice as much as a step of iteration (5.4). This is the price we have to pay for working with the generalized discrete Theodorsen equation (2.3), where $\boldsymbol{\Sigma} \neq \mathbf{I}$, and $\mathbf{I} - \mathbf{J}$ has lost property A.

The global convergence of the iterations (5.4) and (5.11) will be established in Sect. 7 under the assumption $\varepsilon < 1$.

6. Nonlinear Second Order Euler Methods

Generalized Euler methods for linear systems have been proposed by Kublanovskaya (see [3], pp. 532–548) and Niethammer [19]. In the case of systems with 2-cyclic matrices Niethammer has also investigated the relationship with the

SOR method. Here we modify some of these second order methods to apply them to the nonlinear equations (2.2) and (2.3) It turns out that the method most appropriate to (2.3) is equivalent to the SOR iteration (5.11). Moreover, the corresponding method for linear Hermitian systems would be identical with a particular optimal second order Richardson iteration [7].

For the moment, let us assume $\Sigma = \mathbf{I}$ again. The coupled nonlinear system (5.3) is equivalent to the uncoupled system

$$\mathbf{y}'' = -\mathbf{L}^T \log \rho(\mathbf{t}' + \mathbf{L} \log \rho(\mathbf{t}'' + \mathbf{y}'')), \tag{6.1a}$$

$$\mathbf{y}' = \mathbf{L} \log \rho(\mathbf{t}'' - \mathbf{L}^T \log \rho(\mathbf{t}' + \mathbf{y}')). \tag{6.1b}$$

Similarly, the linear system (5.6) may be written as an uncoupled system. As shown in Sect. 6 of [19], solving the first part of it with a particular second order Euler method yields the same iterates $\mathbf{y}_k''$ as the SOR method (5.7) with special starting values $\mathbf{y}_0''$, $\mathbf{y}_0'$.

In analogy, the *nonlinear second order Euler iteration* appropriate for (6.1a) has the form

$$\begin{aligned} \mathbf{y}_{m+1}'' := & -\omega^2 \mathbf{L}^T \log \rho(\mathbf{t}' + \mathbf{L} \log \rho(\mathbf{t}'' + \mathbf{y}_m'')) \\ & + 2(1-\omega)\, \mathbf{y}_m'' - (1-\omega)^2 \mathbf{y}_{m-1}'', \qquad m = 0, 1, \ldots, \end{aligned} \tag{6.2}$$

where $\omega \in (0, 1)$. Every fixed point $\mathbf{y}''$ of this iteration is a solution of (6.1a). Moreover, if we then define $\mathbf{y}'$ by (5.3b), both equations (5.3) are satisfied. We start with given $\mathbf{y}_{-1}'' = \mathbf{y}_0''$ possibly using a different $\omega = \omega_0 \in [0, 1]$ for $m = 0$. In any case, $\mathbf{y}_m'' = \mathbf{y}_0''$, $\forall m \geqq 0$, if $\mathbf{y}_0''$ happens to be a fixed point already.

Now, note that $\mathbf{y}_{m+1}''$ in (6.2) just contains the even indexed components of

$$\begin{aligned} \mathbf{y}_{m+1} := & \omega^2 \mathbf{K}_\Sigma \log \rho(\mathbf{t} + \mathbf{K}_\Sigma \log \rho(\mathbf{t} + \mathbf{y}_m)) \\ & + 2(1-\omega)\, \mathbf{y}_m - (1-\omega)^2 \mathbf{y}_{m-1}, \qquad m = 0, 1, \ldots, \end{aligned} \tag{6.3}$$

if this iteration is started with $\mathbf{y}_{-1} = \mathbf{y}_0$ and the same ω_0. Of course, this method is also well defined if $\Sigma \neq \mathbf{I}$. It relates to the system (5.12) and the SOR method (5.11) as iteration (6.2) relates to (5.3) and the SOR method (5.4). Again, it is twice as expensive as (6.2), and it might have fixed points that are not solutions of the given problem.

However, both iterations (6.2) and (6.3) have a serious disadvantage: while in the linear case the SOR iteration (5.7) (with special choice of $\mathbf{y}_0'$, $\mathbf{y}_0''$) is equivalent to an Euler method, this is not true in the nonlinear case. As we know from Jacobi's method, even if $\mathbf{y}_m$ is not far from $\mathbf{y}^*$, the vector $\mathbf{K}_\Sigma \log \rho(\mathbf{t} + \mathbf{y}_m)$ may be relatively far away (if $\varepsilon \gg 1$) since it is not modified by underrelaxation; therefore, in (6.2) and (6.3) ρ is sometimes evaluated at unfavorable points. Hence, although these iterations feature the same excellent asymptotic rate of convergence as the SOR methods (5.4) and (5.11), they are less recommended.

Fortunately, there is another *nonlinear second order Euler iteration that is adapted to* (2.3) *directly*. Its linear version is based on Niethammer's Theorem 5.1 in [19] (cf. [3], p. 542, and [15], too), which we state in slightly generalized form. (Also, the original formulas in [19] contain two misprints.)

Theorem 6.1. *Suppose all eigenvalues of the matrix* $\mathbf{B}$ *lie in a closed domain* L *bounded by the ellipse with foci* $a, b \in \mathbb{C}$ *and semiaxes of length*

$$\alpha = \frac{\gamma}{2}\left(\frac{1}{\delta} + \delta\right), \qquad \beta = \frac{\gamma}{2}\left(\frac{1}{\delta} - \delta\right),$$

where $\gamma = \frac{1}{2}|a-b|$, $\delta \in (0, 1]$. *Assume that* $1 \notin L$, *and that* ζ *is the solution of*

$$(b-a)\zeta^2 + 2(b+a-2)\zeta + (b-a) = 0$$

that lies in the unit disk. Then, the following second order Euler method for the linear system $\mathbf{x} - \mathbf{B}\mathbf{x} = \mathbf{c}$ *is optimal with respect to* L*:*

$$\begin{aligned} &\mathbf{x}_{-1} := \mathbf{x}_0 := \mathbf{c}, \\ &\mathbf{x}_{m+1} := \omega\hat{\omega}(\mathbf{B}\mathbf{x}_m + \mathbf{c}) + \omega(1-\hat{\omega})\,\mathbf{x}_m + (1-\omega)\,\mathbf{x}_{m-1}, \qquad m = 0, 1, \ldots, \end{aligned} \tag{6.4}$$

where

$$\omega := 1 + \zeta^2, \qquad \hat{\omega} := \frac{2}{2-a-b}.$$

Its asymptotic convergence factor is not worse than $|\zeta|/\delta$.

In our application the foci are at $\pm i\gamma$, and hence

$$\zeta = \frac{-i\gamma}{1+\sqrt{1+\gamma^2}}, \qquad \omega = \frac{2}{1+\sqrt{1+\gamma^2}}, \qquad \hat{\omega} = 1.$$

We have to replace $\mathbf{x}_m$ by $\mathbf{y}_m$ and $\mathbf{B}\mathbf{x}_m + \mathbf{c}$ by $\mathbf{K}_\Sigma \log \rho(\mathbf{t} + \mathbf{y}_m)$. Then (6.4) becomes

$$\mathbf{y}_{m+1} := \omega \mathbf{K}_\Sigma \log \rho(\mathbf{t} + \mathbf{y}_m) + (1-\omega)\,\mathbf{y}_{m-1}, \qquad m = 0, 1, \ldots. \tag{6.5}$$

Again, we let $\mathbf{y}_{-1} := \mathbf{y}_0$ and may use a special $\omega = \omega_0 \in [0, 1]$ for $m = 0$.

One might also call (6.5) a *nonlinear second order Richardson iteration* since in the linear Hermitian case the second order Euler method reduces to a particular Richardson iteration [7]. Young [26] calls it second-degree Jacobi method. Surprisingly, (6.5) is nearly identical with the JOR iteration (4.1), where the last term contains $\mathbf{y}_m$ instead of $\mathbf{y}_{m-1}$. However, the iterates $\mathbf{y}_m$ of (6.5) actually are identical with those constructed by the SOR method (5.11):

$$\mathbf{y}_{2m} = \mathbf{y}_m^{(1)}, \qquad \mathbf{y}_{2m+1} = \mathbf{y}_m^{(2)}, \qquad \forall m \geqq 0,$$

if the starting values are chosen appropriately. The optimal relaxation factor ω is the same, too. [Note that in general, if $\mathbf{J}$ were an arbitrary real or complex matrix, e.g., $\mathbf{J} = (i)$ or $\mathbf{J} = (0.5)$, the matrix (5.13) would not have the same eigenvalues as $\mathbf{J}$, and the optimal second order Euler method would be different from the SOR iteration (5.11). It is essential that Lemma 3.3 holds for our $\mathbf{J}$.]

(6.5) is a nonlinear system of second order difference equations. As usual there is an equivalent first order system of double size. The corresponding

iteration function Φ_ω^E and its F-derivative

$$\mathbf{J}_\omega^E = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ (1-\omega)\mathbf{I} & \omega\mathbf{J} \end{pmatrix} \tag{6.6}$$

are easily written down. Note that $(\mathbf{J}_\omega^E)^2 = \tilde{\mathbf{J}}_\omega^S$. Thus, both the equivalence with the SOR method (5.11) and the linear theory (Theorem 6.1) lead to

Theorem 6.2. *Theorem* 5.1 *and Corollary* 5.2 *hold for the nonlinear second order Euler iteration* (6.5) *as well, except that the asymptotic convergence factor is not worse than* $\sigma' = \sigma^{1/2}$ *in part* (i) *and equal to*

$$\sigma_E^*(\varepsilon) := [\sigma_S^*(\varepsilon)]^{1/2} = \frac{\varepsilon}{1+\sqrt{1+\varepsilon^2}} < 1$$

in part (ii), *respectively. In particular,*

$$\sigma_E^*(\varepsilon) = 1 - \frac{1}{\varepsilon} + O(\varepsilon^{-2}) \quad \text{as} \quad \varepsilon \to \infty.$$

Although the SOR methods (5.4) and (5.11) converge twice as fast, we must be aware that the second one is exactly as efficient as (6.5) since each step of (5.11) consists of two steps of (6.5). However, iteration (5.4), which takes advantage of the cyclic structure in the case $\Sigma = \mathbf{I}$, is roughly twice as efficient.

But if $\Sigma = \mathbf{I}$, iteration (6.5) can also be simplified by a standard trick due to Riley [23]. At first, by permuting the equations we get

$$\mathbf{y}''_{m+1} := -\omega \mathbf{L}^T \log \rho(\mathbf{t}' + \mathbf{y}'_m) + (1-\omega)\,\mathbf{y}''_{m-1}, \tag{6.7a}$$

$$\mathbf{y}'_{m+1} := \omega \mathbf{L} \log \rho(\mathbf{t}'' + \mathbf{y}''_m) + (1-\omega)\,\mathbf{y}'_{m-1}. \tag{6.7b}$$

It is now possible to compute the subsequences $\{\mathbf{y}''_{2m-1}\}$ and $\{\mathbf{y}'_{2m}\}$ by using (6.7a) for even m and (6.7b) for odd m only. Consequently, the iterates

$$\hat{\mathbf{y}}''_m := \mathbf{y}''_{2m-1}, \qquad \hat{\mathbf{y}}'_m := \mathbf{y}'_{2m}, \qquad m = 0, 1, \ldots \tag{6.8}$$

are obtained with half the effort. However, this method is easily seen to be identical with the SOR method (5.4).

For a global convergence theorem, see Sect. 7.

7. Nonlinear Chebyshev Iteration

The generalized Euler methods discussed in the last section are but one type of semi-iterative methods. Another type, used very often for Hermitian systems of linear equations, is the Chebyshev iteration, which was developed mainly by Golub and Varga [7, 8, 24]. Its generalization to nonsymmetric systems has been investigated by Manteuffel [17]. Unfortunately, in general the method does not feature the optimal average rate of convergence as in the Hermitian case,

where two facts are exploited: first, the spectral norm is equal to the spectral radius, and, second, the Chebyshev polynomials, translated to a real interval containing the eigenvalues, are the unique solution of a related optimization problem. Moreover, the optimality mentioned refers to linear systems, while we are considering a nonlinear iteration in this section.

The appropriate *nonlinear Chebyshev iteration* for the solution of the modified discrete Theodorsen equation (2.3) with $\Sigma \neq \mathbf{I}$ is

$$\mathbf{y}_{m+1} := \omega_{m+1} \mathbf{K}_\Sigma \log \rho(\mathbf{t}+\mathbf{y}_m) + (1-\omega_{m+1})\,\mathbf{y}_{m-1}, \qquad m=0,1,2,\ldots, \tag{7.1a}$$

where $\mathbf{y}_0$ has to be given, and where for any $\omega \in (0,1]$

$$\begin{aligned} &\omega_1 := 1, \\ &\omega_{m+1} := \frac{2}{i\gamma}\frac{T_m\left(\frac{1}{i\gamma}\right)}{T_{m+1}\left(\frac{1}{i\gamma}\right)} = \omega\frac{1+(-1)^m(1-\omega)^m}{1-(-1)^m(1-\omega)^{m+1}}, \qquad m=1,2,\ldots, \\ &\gamma := \frac{2}{\omega}\sqrt{1-\omega}, \qquad \text{i.e. } \omega = \frac{2}{1+\sqrt{1+\gamma^2}}. \end{aligned} \tag{7.1b}$$

Here T_m denotes the Chebyshev polynomial of degree m:

$$T_m(z) := \cosh(m \cosh^{-1}(z)), \qquad z \in \mathbb{C}.$$

(Any branch of $\cosh^{-1}$ may be used in this definition.) The equality in (7.1b) is easily verified; if Log denotes the principal branch of the logarithm, we get:

$$\begin{aligned} T_m\left(\frac{1}{i\gamma}\right) &= \cosh\left\{m \operatorname{Log}\left[\frac{1}{i\gamma}(1+\sqrt{1+\gamma^2})\right]\right\} \\ &= \cosh\left\{-m \operatorname{Log}\frac{\omega\gamma}{2} - im\frac{\pi}{2}\right\} \\ &= \tfrac{1}{2}(-i)^m(1-\omega)^{m/2}\{1+(-1)^m(1-\omega)^m\}. \end{aligned} \tag{7.2}$$

and the asserted equality follows.

Note that ω and γ are related as in Euler's method and the SOR iteration, cf. (5.10), and that

$$\sqrt{1-\omega} = \frac{\gamma}{1+\sqrt{1+\gamma^2}} = \sigma_E^*(\gamma) < 1. \tag{7.3}$$

Hence

$$\omega_{m+1} \to \omega \quad \text{as} \quad m \to \infty, \tag{7.4}$$

which means that the nonlinear Chebyshev method (7.1) approaches asymptotically the nonlinear Euler method (6.5). (For linear Hermitian systems, this fact is well known [7, 19].)

Since iteration (7.1) is non-stationary, we cannot apply Ostrowski's theorem to establish convergence. However, since $\mathbf{y}^*$ is a fixed point of $\boldsymbol{\Phi}$ [defined by (3.1)], (7.1a) is equivalent to

$$\begin{aligned}\mathbf{x}_{m+1}-\mathbf{y}^* &= \mathbf{y}_m-\mathbf{y}^*,\\ \mathbf{y}_{m+1}-\mathbf{y}^* &= \omega_{m+1}[\boldsymbol{\Phi}(\mathbf{y}_m)-\boldsymbol{\Phi}(\mathbf{y}^*)]+(1-\omega_{m+1})(\mathbf{x}_m-\mathbf{y}^*).\end{aligned} \tag{7.5}$$

The second equation may be written in the form

$$\begin{aligned}\mathbf{y}_{m+1}-\mathbf{y}^* = \omega\mathbf{J}(\mathbf{y}_m-\mathbf{y}^*)&+(1-\omega)(\mathbf{x}_m-\mathbf{y}^*)\\ &+(\omega_{m+1}-\omega)[\mathbf{J}(\mathbf{y}_m-\mathbf{y}^*)-\mathbf{x}_m+\mathbf{y}^*]\\ &+\omega_{m+1}[\boldsymbol{\Phi}(\mathbf{y}_m)-\boldsymbol{\Phi}(\mathbf{y}^*)-\mathbf{J}(\mathbf{y}_m-\mathbf{y}^*)].\end{aligned}$$

Here the last bracket is $o(\|\mathbf{y}_m-\mathbf{y}^*\|)$ as $\mathbf{y}_m\to\mathbf{y}^*$. Hence,

$$\begin{pmatrix}\mathbf{x}_{m+1}-\mathbf{y}^*\\ \mathbf{y}_{m+1}-\mathbf{y}^*\end{pmatrix}=\mathbf{J}_\omega^E\begin{pmatrix}\mathbf{x}_m-\mathbf{y}^*\\ \mathbf{y}_m-\mathbf{y}^*\end{pmatrix}+\boldsymbol{\Xi}(m,x_m,y_m),$$

where $\mathbf{J}_\omega^E$ defined by (6.6) is regular if $\omega\in(0,1)$,

$$\|\boldsymbol{\Xi}(m,\mathbf{x}_m,\mathbf{y}_m)\|\leqq\text{const}(\|\mathbf{x}_m-\mathbf{y}^*\|+\|\mathbf{y}_m-\mathbf{y}^*\|)$$

for every m and every $\mathbf{x}_m$, $\mathbf{y}_m$ in a neighborhood of $\mathbf{y}^*$, and, cf. (7.4),

$$\frac{\|\boldsymbol{\Xi}(m,\mathbf{x}_m,\mathbf{y}_m)\|}{\|\mathbf{x}_m-\mathbf{y}^*\|+\|\mathbf{y}_m-\mathbf{y}^*\|}\to 0\quad\text{as}\quad m\to\infty,\ \mathbf{x}_m\to\mathbf{y}^*,\ \mathbf{y}_m\to\mathbf{y}^*.$$

Thus, Perron's [22] conditions A and B are fulfilled and his Theorems 5 and 11 can be applied: If $\Lambda(\mathbf{J}_\omega^E)<1$, if $\mathbf{x}_0$ and $\mathbf{y}_0$ are sufficiently close to $\mathbf{y}^*$, and if $\mathbf{y}_m\neq\mathbf{y}^*$ for all m, then $\mathbf{x}_m\to\mathbf{y}^*$, $\mathbf{y}_m\to\mathbf{y}^*$ linearly (as $m\to\infty$), and the asymptotic convergence factor only depends on the dominant eigenvalues of the matrix $\mathbf{J}_\omega^E$, which originally has been defined as F-derivative of the Euler iteration (6.5). We conclude

Theorem 7.1. *Theorem* 6.2 *holds for the nonlinear Chebyshev iteration* (7.1) *also.*

On the other hand, if $\boldsymbol{\Sigma}=\mathbf{I}$, then the discrete Theodorsen equation (2.2) is equivalent to the system (5.3), for which an appropriate *nonlinear cyclic Chebyshev iteration* exists:

$$\begin{aligned}\mathbf{y}''_{m+1} &:= -\omega_{2m+1}\mathbf{L}^T\log\rho(\mathbf{t}'+\mathbf{y}'_m)+(1-\omega_{2m+1})\,\mathbf{y}''_m,\\ \mathbf{y}'_{m+1} &:= \omega_{2m+2}\mathbf{L}\log\rho(\mathbf{t}''+\mathbf{y}''_{m+1})+(1-\omega_{2m+2})\,\mathbf{y}'_m,\\ &\qquad m=0,1,2,\ldots\end{aligned} \tag{7.6}$$

Here, the factors ω_m are again defined by (7.1b), and only $\mathbf{y}'_0$ has to be given. In view of (7.4) this method, which is a particular nonlinear modified SOR (i.e., MSOR) iteration, approaches the nonlinear SOR method as $m\to\infty$. Applying Perron's theory again, we get

Theorem 7.2. *Theorem* 5.1 *and Corollary* 5.2 *hold for the nonlinear cyclic Chebyshev iteration* (7.6) *as well.*

Theorems 7.1 and 7.2 would hold as well if $\{\omega_m\}$ were any real sequence converging to ω.

Finally, supposing $\varepsilon<1$ we are going to establish the gloval convergence. As we have seen, the nonlinear Chebyshev iteration (7.1a) satisfies Eqs. (7.5), which also describe the nonlinear Euler iteration (6.5) if we set $\omega_m := \omega$, $\forall m>0$. In view of (3.3), with any $\kappa>0$,

$$\delta_m'' := \kappa \|\mathbf{x}_m - \mathbf{y}^*\|, \qquad \delta_m' := \|\mathbf{y}_m - \mathbf{y}^*\|$$

fulfill the inequalities

$$\begin{aligned} &0 \leqq \delta_{m+1}'' = \kappa \delta_m', \\ &0 \leqq \delta_{m+1}' \leqq \omega_{m+1}\varepsilon\delta_m' + (1-\omega_{m+1})\kappa^{-1}\delta_m''. \end{aligned} \tag{7.7}$$

The right side can be written in the form

$$\mathbf{H}(\omega_{m+1}) \begin{pmatrix} \delta_m'' \\ \delta_m' \end{pmatrix}, \qquad \text{where} \quad \mathbf{H}(\omega) := \begin{pmatrix} 0 & \kappa \\ (1-\omega)\kappa^{-1} & \omega\varepsilon \end{pmatrix}$$

If we denote by $\delta_m := [(\delta_m'')^2 + (\delta_m')^2]^{1/2}$ the Euclidean norm of $(\delta_m'', \delta_m')^T$ and use the (compatible) spectral norm of $\mathbf{H}(\omega_{m+1})$, it follows from (7.7) that

$$0 \leqq \delta_{m+1} \leqq \|\mathbf{H}(\omega_{m+1})\|\, \delta_m, \qquad \forall m \geqq 0. \tag{7.8}$$

Now, suppose $\omega\in(0,1)$, $\omega_m \to \omega$ (or $\omega_m = \omega$, $\forall m>0$), $\kappa := (1-\omega)^{1/2}$, and $\Lambda_0 := (2+\varepsilon\omega-\omega)/2$. Then, the eigenvalues λ of $\mathbf{H}(\omega)$ satisfy

$$|\lambda| = \tfrac{1}{2}|\varepsilon\omega \pm (\varepsilon^2\omega^2 - 4\omega + 4)^{1/2}| < \Lambda_0 < 1.$$

Since $\mathbf{H}(\omega)$ is symmetric, $\|\mathbf{H}(\omega)\| = \Lambda(\mathbf{H}(\omega)) < \Lambda_0$. Consequently, $\|\mathbf{H}(\omega_{m+1})\| < \Lambda_0$ for, say, $m \geqq m_0$, and, by (7.8), $\delta_m \to 0$ as $m \to \infty$.

In view of the fact that the SOR method (5.4) yields exactly part of the components of the Euler iterates (6.5) [cf. (6.8)], and since the same relation holds between the nonlinear MSOR iteration (7.6) and the iteration (7.1a), we conclude

Theorem 7.3. *Assume* $\varepsilon<1$, $\omega\in(0,1)$ *arbitrarily, and* $\{\omega_m\}$ *is any real sequence converging to* ω. *Then the nonlinear second order iteration* (7.1a), *the nonlinear Euler iteration* (6.5), *the equivalent nonlinear SOR method* (5.11), *as well as the nonlinear MSOR method* (7.6), *and the nonlinear SOR method* (5.4) *(the last two requiring* $\mathbf{\Sigma}=\mathbf{I}$*) converge globally to the unique solution* $\mathbf{y}^*$ *of* (2.3), *to* $(\mathbf{y}^{*T}, \mathbf{y}^{*T})^T$, *or to* $\mathbf{P}\mathbf{y}^*$, *respectively.*

Consequently, the system (5.12) also has the unique solution $\mathbf{y}^{(1)} = \mathbf{y}^{(2)} = \mathbf{y}^*$ if $\varepsilon<1$.

If $\omega=1$, the Euler iteration reduces to the Jacobi iteration (3.1), and the SOR iteration reduces to the Gauss-Seidel method, both of which are known to converge globally, cf. Theorem 3.1 and [18], respectively.

Appendix 1. Proof of Lemma 3.2

The proof is mainly based on the well-known minimax theorem for the eigenvalues of a symmetric matrix [2, Chap. 1, § 4] and on the fact that the restriction of $-\mathbf{K}^2$ to a distinct $2N-2$ dimensional subspace is the identity.

We let $t_k := k\pi/N$, $k=0, \ldots, 2N-1$, and

$$\mathscr{P}_N := \{\mathbf{x}\in\mathbb{R}^{2N}: \; x_k = \sum_{m=1}^{N-1} (a_m \cos m t_k + b_m \sin m t_k),$$
$$k=0, \ldots, 2N-1; a_m\in\mathbb{R}, b_m\in\mathbb{R}, m=1, \ldots, N-1\}.$$

From the relation sketched in Sect. 2 (cf. [10] for details) between $\mathbf{K}$ and the discrete Fourier transform it follows that $\mathbb{R}^{2N} = \mathscr{P}_N \oplus \ker \mathbf{K}$ (the kernel of $\mathbf{K}$) and that $\mathbf{K}^2\mathbf{x} = -\mathbf{x}$ for every $\mathbf{x}\in\mathscr{P}_N$, i.e. the restriction of $-\mathbf{K}^2$ to $\mathscr{P}_N$ is the identity. The discrete Fourier transform also yields

$$\mathscr{P}_N = \left\{\mathbf{x}\in\mathbb{R}^{2N}: \sum_{k=0}^{2N-1} x_k = \sum_{k=0}^{2N-1} (-1)^k x_k = 0\right\}.$$

If we let

$$\mathscr{Q}_N := \{\mathbf{y}\in\mathbb{R}^{2N}: \mathbf{D}\mathbf{y}\in\mathscr{P}_N\},$$
$$\mathbf{D} := \operatorname{diag}(d_0, \ldots, d_{2N-1}),$$

then $\mathbf{K}^T\mathbf{K}\mathbf{D}\mathbf{y} = -\mathbf{K}^2\mathbf{D}\mathbf{y} = \mathbf{D}\mathbf{y}$ for every $\mathbf{y}\in\mathscr{Q}_N$, and, on the other hand,

$$\mathscr{Q}_N = \left\{\mathbf{y}\in\mathbb{R}^{2N}: \sum_{k=0}^{2N-1} d_k y_k = \sum_{k=0}^{2N-1} (-1)^k d_k y_k = 0\right\}.$$

So, $\mathscr{Q}_N$ is a subspace of dimension at least $2N-2$. Now,

$$\|\mathbf{K}\mathbf{D}\| = \max_{\substack{\mathbf{y}\in\mathbb{R}^{2N}\\ \mathbf{y}\neq\mathbf{0}}} \left[\frac{(\mathbf{K}\mathbf{D}\mathbf{y}, \mathbf{K}\mathbf{D}\mathbf{y})}{(\mathbf{y}, \mathbf{y})}\right]^{1/2} \geqq \max_{\substack{\mathbf{y}\in\mathscr{Q}_N\\ \mathbf{y}\neq\mathbf{0}}} \left[\frac{(\mathbf{K}\mathbf{D}\mathbf{y}, \mathbf{K}\mathbf{D}\mathbf{y})}{(\mathbf{y}, \mathbf{y})}\right]^{1/2}$$
$$= \max_{\substack{\mathbf{y}\in\mathscr{Q}_N\\ \mathbf{y}\neq\mathbf{0}}} \left[\frac{(\mathbf{y}, \mathbf{D}^2\mathbf{y})}{(\mathbf{y}, \mathbf{y})}\right]^{1/2}.$$

According to the minimax theorem cited, the last maximum is not smaller than the square root of the third largest eigenvalue of the diagonal matrix $\mathbf{D}^2$. □

Appendix 2. Proof of Theorem 3.4

Suppose that assumption (SD) holds, and let $\mathscr{R}_0 := \{\mathbf{y}\in\mathbb{R}^{2N}: \mathbf{y} \text{ satisfies (3.6b)}\}$. If $\mathbf{y}\in\mathscr{R}_0$, then $\mathbf{D}\mathbf{y}\in\mathscr{R}_0$ and $\log\rho(\mathbf{t}+\mathbf{y})\in\mathscr{R}_0$. We might call these vectors $2M$-periodic. If we denote the complex discrete Fourier transform in $\mathbb{C}^{2N}$ and $\mathbb{C}^{2M}$ by $\mathscr{F}$ and $\mathscr{F}_0$, respectively, then, cf. [10, 12],

$$\hat{\mathbf{K}} := \mathscr{F}\mathbf{K}\mathscr{F}^{-1} = \operatorname{diag}(0, -i, \ldots, -i, 0, i, \ldots, i),$$
$$\hat{\boldsymbol{\Sigma}} := \mathscr{F}\boldsymbol{\Sigma}\mathscr{F}^{-1} = \operatorname{diag}(\sigma_0, \ldots, \sigma_N, \sigma_{N-1}, \ldots, \sigma_1) \tag{1}$$

with real nonnegative weights $\sigma_0, \ldots, \sigma_N$. We let $\mathbf{K}_0$ be the $2M \times 2M$ Wittich matrix, and set

$$\mathbf{\Sigma}_0 := \mathscr{F}_0^{-1} \operatorname{diag}(\sigma_0, \sigma_v, \sigma_{2v}, \ldots, \sigma_N, \sigma_{N-v}, \ldots, \sigma_v) \mathscr{F}_0,$$

$$\mathbf{D}_0 := \operatorname{diag}\left(\frac{\rho'}{\rho}(t_k + y_k^*)\right)_{k=0,\ldots,2M-1}.$$

It can be proved that $\mathbf{\Sigma}_0$ is again a real symmetric circulant Toeplitz matrix. Now, if $\mathbf{y} \in \mathscr{R}_0$, those components of $\mathscr{F}\mathbf{D}\mathbf{y}$ (i.e. those Fourier coefficients), where the index is not a multiple of v, vanish. Thus, we see from

$$\mathbf{u} := \mathbf{J}\mathbf{y} = \mathbf{K}\mathbf{\Sigma}\mathbf{D}\mathbf{y} = \mathscr{F}^{-1}\hat{\mathbf{K}}\hat{\mathbf{\Sigma}}\mathscr{F}\mathbf{D}\mathbf{y},$$

that $\mathbf{u} \in \mathscr{R}_0$ too, and

$$(u_0, \ldots, u_{2M-1})^T = \mathbf{K}_0\mathbf{\Sigma}_0\mathbf{D}_0(y_0, \ldots, y_{2M-1})^T. \tag{2}$$

By the same argument, $\mathbf{\Phi}(\mathbf{y}) \in \mathscr{R}_0$.

This means that $\mathscr{R}_0$ is an invariant subspace of $\mathbf{J}$ and that $\mathbf{J}_0 := \mathbf{K}_0\mathbf{\Sigma}_0\mathbf{D}_0$ is a matrix representation (with respect to a particular basis) of the restriction $\mathbf{J}|_{\mathscr{R}_0}$ of $\mathbf{J}$ to $\mathscr{R}_0$. Since $\mathbf{J}_0$ is exactly the F-derivative of an iteration function $\mathbf{\Phi}_0$ belonging to another conformal mapping problem satisfying assumption (SD) with $v=1$ (and $N:=M$), we may assume $v=1$ for the rest of the proof.

We call an $\mathbf{y}$ satisfying (3.6a) an odd vector. If $\mathbf{y}$ is odd, then $\mathbf{x} := \log\rho(\mathbf{t}+\mathbf{y})$ is called even since $x_k = x_{2N-k}$, $k=1,\ldots,N$. This implies that $\mathscr{F}\mathbf{x} \in \mathbb{R}^{2N}$, while the components of $\hat{\mathbf{K}}\hat{\mathbf{\Sigma}}\mathscr{F}\mathbf{x}$ are purely imaginary. Hence $\mathbf{\Phi}(\mathbf{y}) = \mathscr{F}^{-1}\hat{\mathbf{K}}\hat{\mathbf{\Sigma}}\mathscr{F}\mathbf{x}$ fulfills (3.6a) too. We conclude that $\mathbf{\Phi}(\mathscr{S}) \subset \mathscr{S}$, and that $\mathscr{S}$ is an invariant subspace of $\mathbf{J}$, cf. Lemma 3.5. In fact, for $\mathbf{y} \in \mathscr{S}$ we get

$$\mathbf{\Phi}(\mathbf{y}) = (0, z_1, z_2, \ldots, z_{N-1}, 0, -z_{N-1}, \ldots, -z_1)^T,$$

where

$$\hat{\mathbf{z}} := (z_1, \ldots, z_{N-1})^T = \mathbf{\Phi}_S(\hat{\mathbf{y}}) := \mathbf{K}_S\mathbf{L}_S(\hat{\mathbf{y}}),$$

$\mathbf{K}_S$ being an $(N-1)\times(N+1)$-matrix and $\mathbf{L}_S\colon \mathbb{R}^{N-1} \to \mathbb{R}^{N+1}$ a nonlinear operator:

$$(\mathbf{K}_S)_{kj} := \begin{cases} (\mathbf{K}_\Sigma)_{kj}, & j=0,N;\ k=1,\ldots,N-1; \\ (\mathbf{K}_\Sigma)_{kj} + (\mathbf{K}_\Sigma)_{k,2N-j}, & j,k=1,\ldots,N-1; \end{cases}$$

$$\mathbf{L}_S(\hat{\mathbf{y}}) := (\log\rho(0), \log\rho(t_1+y_1), \ldots, \log\rho(t_{N-1}+y_{N-1}), \log\rho(\pi))^T.$$

The F-derivative of $\mathbf{\Phi}_S$ at $\hat{\mathbf{y}}^*$, which is a matrix representation of $\mathbf{J}|_{\mathscr{S}}$, is

$$\mathbf{J}_S = \mathbf{K}_T\mathbf{D}_T, \tag{3}$$

where $\mathbf{K}_T$ and $\mathbf{D}_T$ are now square matrices of order $N-1$ defined by

$$\begin{aligned} (\mathbf{K}_T)_{kj} &:= (\mathbf{K}_\Sigma)_{kj} + (\mathbf{K}_\Sigma)_{k,2N-j}, \quad j,k=1,\ldots,N-1, \\ \mathbf{D}_T &:= \operatorname{diag}\left(\frac{\rho'}{\rho}(t_k+y_k^*)\right)_{k=1,\ldots,N-1}. \end{aligned} \tag{4}$$

The basic tool for the proof of Theorem 3.4 is the following

Lemma 1. *Assume that* $\sigma_k > 0$, $k = 1, \ldots, N-1$, *in* (1), *and define the square matrices* $\mathbf{S}$, $\mathbf{D}_{\sin}$, $\tilde{\mathbf{D}}_\Sigma$, $\mathbf{D}_\Sigma$, $\tilde{\mathbf{H}}_\Sigma$, *and* $\mathbf{H}_\Sigma$, *all of order* $N-1$, *by*

$$(\mathbf{S})_{kj} := \sin(j t_k) = \sin(j k \pi/N), \quad j, k = 1, \ldots, N-1,$$

$$\mathbf{D}_{\sin} := \operatorname{diag}(\sin t_1, \ldots, \sin t_{N-1}),$$

$$\tilde{\mathbf{D}}_\Sigma := \operatorname{diag}(\sqrt{\sigma_k/\sigma_1})_{k=1,\ldots,N-1}, \quad \mathbf{D}_\Sigma := \mathbf{S}\tilde{\mathbf{D}}_\Sigma\mathbf{S}^{-1},$$

$$\tilde{\mathbf{H}}_\Sigma := \frac{1}{2}\begin{pmatrix} 0 & \sigma_1 & & & 0 \\ -\sigma_2 & 0 & \sigma_2 & & \\ & \ddots & \ddots & \ddots & \\ & 0 & -\sigma_{N-2} & 0 & \sigma_{N-2} \\ & & & -\sigma_{N-1} & 0 \end{pmatrix}, \quad \mathbf{H}_\Sigma := \tilde{\mathbf{D}}_\Sigma^{-1}\tilde{\mathbf{H}}_\Sigma\tilde{\mathbf{D}}_\Sigma.$$

Then

$$\mathbf{D}_\Sigma^{-1}\mathbf{K}_T\mathbf{D}_{\sin}\mathbf{D}_\Sigma = \mathbf{S}\mathbf{H}_\Sigma\mathbf{S}^{-1} = \frac{2}{N}\mathbf{S}\mathbf{H}_\Sigma\mathbf{S}, \tag{5}$$

and this matrix is skew-symmetric.

Proof. The matrix $\mathbf{S}$ is symmetric, and as is known from the finite sine transform [1], $\mathbf{S}^{-1} = (2/N)\mathbf{S}$. This justifies the equality sign at right in (5). It is easy to verify that

$$\mathbf{H}_\Sigma = \begin{pmatrix} 0 & \sqrt{\sigma_1\sigma_2} & & 0 \\ -\sqrt{\sigma_1\sigma_2} & 0 & \sqrt{\sigma_2\sigma_3} & \\ & \ddots & \ddots & \\ 0 & & -\sqrt{\sigma_{N-2}\sigma_{N-1}} & 0 \end{pmatrix}.$$

So, $\mathbf{H}_\Sigma$ and $\mathbf{S}\mathbf{H}_\Sigma\mathbf{S}$ are skew-symmetric.

Due to (1), $\mathbf{K}_\Sigma \cos j\mathbf{t} = \mathscr{F}^{-1}\hat{\mathbf{K}}\hat{\Sigma}\mathscr{F}\cos j\mathbf{t} = \sigma_j \sin j\mathbf{t}$,

$$\mathbf{K}_\Sigma[\cos(j-1)\mathbf{t} - \cos(j+1)\mathbf{t}] = \sigma_{j-1}\sin(j-1)\mathbf{t} - \sigma_{j+1}\sin(j+1)\mathbf{t}, \quad j = 1, \ldots, N-1.$$

Here, the vector in brackets is even and both its first and its N-th component vanish. Hence, we get as well

$$\mathbf{K}_T[\cos(j-1)\hat{\mathbf{t}} - \cos(j+1)\hat{\mathbf{t}}] = \sigma_{j-1}\sin(j-1)\hat{\mathbf{t}} - \sigma_{j+1}\sin(j+1)\hat{\mathbf{t}}, \quad j = 1, \ldots, N-1,$$

where $\hat{\mathbf{t}} := (t_1, \ldots, t_{N-1})^T$. Finally, since

$$(\mathbf{D}_{\sin}\mathbf{S})_{kj} = \sin t_k \sin j t_k = \tfrac{1}{2}[\cos(j-1)t_k - \cos(j+1)t_k]$$

(for $k, j = 1, \ldots, N-1$), it follows that

$$(\mathbf{K}_T\mathbf{D}_{\sin}\mathbf{S})_{kj} = \tfrac{1}{2}[\sigma_{j-1}\sin(j-1)t_k - \sigma_{j+1}\sin(j+1)t_k] = (\mathbf{S}\tilde{\mathbf{H}}_\Sigma)_{kj},$$

$$\mathbf{K}_T\mathbf{D}_{\sin} = \mathbf{S}\tilde{\mathbf{H}}_\Sigma\mathbf{S}^{-1} = \mathbf{S}\tilde{\mathbf{D}}_\Sigma\mathbf{H}_\Sigma\tilde{\mathbf{D}}_\Sigma^{-1}\mathbf{S}^{-1} = \mathbf{D}_\Sigma\mathbf{S}\mathbf{H}_\Sigma\mathbf{S}^{-1}\mathbf{D}_\Sigma^{-1},$$

which shows that the equality at left in (5) holds. □

In view of (3) and our previous arguments Theorem 3.4 is just a corollary of

Lemma 2. *If either* $\mathbf{D}_T = \mathbf{D}_{\sin}$ *or* $\mathbf{D}_\Sigma = \mathbf{I}$ *and* $\mathbf{D}_T$ *is any diagonal matrix [not necessarily the one defined by* (4)*] with either no negative or no positive element, then* $\mathbf{K}_T\mathbf{D}_T$ *has purely imaginary eigenvalues.*

Proof. If $\mathbf{D}_T$ has strictly positive diagonal elements, it can be written in the form

$$\mathbf{D}_T = \mathbf{D}_{\sin}\mathbf{D}_R^2,$$

where $\mathbf{D}_R$ is a positive definite diagonal matrix, too. We assume $\sigma_k \neq 0$, $k = 1, \ldots, N-1$ and use (5) and $\mathbf{D}_R\mathbf{D}_\Sigma^{-1} = \mathbf{D}_\Sigma^{-1}\mathbf{D}_R$ (where one of the matrices is equal to $\mathbf{I}$) to conclude that

$$\mathbf{G} := \frac{2}{N}\mathbf{D}_R\mathbf{S}\mathbf{H}_\Sigma\mathbf{S}\mathbf{D}_R = \mathbf{D}_R\mathbf{D}_\Sigma^{-1}\mathbf{K}_T\mathbf{D}_{\sin}\mathbf{D}_\Sigma\mathbf{D}_R = \mathbf{D}_\Sigma^{-1}\mathbf{D}_R\mathbf{K}_T\mathbf{D}_T\mathbf{D}_R^{-1}\mathbf{D}_\Sigma \tag{6}$$

is a skew-symmetric matrix, which is similar to $\mathbf{K}_T\mathbf{D}_T$. Thus, $\mathbf{K}_T\mathbf{D}_T$ has purely imaginary eigenvalues. Since these eigenvalues depend continuously on the elements of $\mathbf{D}_T$ and $\hat{\boldsymbol{\Sigma}} = \mathscr{F}\boldsymbol{\Sigma}\mathscr{F}^{-1}$, they are still purely imaginary if some diagonal elements vanish. (Note that we may still define $\mathbf{D}_R$ and $\mathbf{G}$ in this case, but that $\mathbf{G}$ is no more similar to $\mathbf{K}_T\mathbf{D}_T$.) □

Finally, we would like to state the related

Lemma 3. $\|\mathbf{K}_T\| \leqq \|\mathbf{K}_\Sigma\|$.

Proof. Let $\mathbf{D}_T := \mathbf{I}$, $\mathbf{D} := \operatorname{diag}(0, 1, \ldots, 1, 0, -1, \ldots, -1)$. Assuming $\mathbf{y} \in \mathscr{S}$ we get $\|\mathbf{y}\| = \sqrt{2}\,\|\hat{\mathbf{y}}\|$ and

$$\mathbf{u} = \mathbf{K}_\Sigma\mathbf{D}\,\mathbf{y} \Leftrightarrow \hat{\mathbf{u}} = \mathbf{K}_T\mathbf{D}_T\hat{\mathbf{y}}.$$

Hence,

$$\|\mathbf{K}_T\| = \max_{\hat{\mathbf{y}} \neq \mathbf{0}} \frac{\|\mathbf{K}_T\mathbf{D}_T\hat{\mathbf{y}}\|}{\|\hat{\mathbf{y}}\|} = \max_{\mathbf{y} \in \mathscr{S} \setminus \{\mathbf{0}\}} \frac{\|\mathbf{K}_\Sigma\mathbf{D}\,\mathbf{y}\|}{\|\mathbf{y}\|}$$
$$\leqq \|\mathbf{K}_\Sigma\mathbf{D}\| \leqq \|\mathbf{K}_\Sigma\|\,\|\mathbf{D}\| = \|\mathbf{K}_\Sigma\|. \quad \square$$

Acknowledgment. The author is indebted to a number of persons including Dr. R. Bloch, Prof. D. Gaier, Prof. P. Henrici, Dr. O. Hübner, and Prof. J.M. Varah for valuable discussions and suggestions.

References

1. Cooley, J.W., Lewis, P.A.W., Welch, P.D.: The fast Fourier transform algorithm: Programming considerations in the calculation of sine, cosine and Laplace transforms. J. Sound Vib. **12**, 315–337 (1970)
2. Courant, R., Hilbert, D.: Methoden der mathematischen Physik, Band 1, 3. Auflage. Berlin Heidelberg New York: Springer 1968
3. Faddeev, D.K., Faddeeva, V.N.: Computational methods of linear algebra. San Francisco: W.H. Freemann, 1963
4. Gaier, D.: Konstruktive Methoden der konformen Abbildung. Berlin Heidelberg New York: Springer 1964
5. Gantmacher, F.R.: The theory of matrices, Vol. II. New York: Chelsea Publishing 1959
6. Gekeler, E.: Über Iterationsverfahren bei einer Klasse nichtlinearer Gleichungssysteme. Dissertation, Ruhr-Universität Bochum, 1969

7. Golub, G.H., Varga, R.S.: Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second order Richardson iterative methods, Part I. Numer. Math. **3**, 147-156 (1961)
8. Golub, G.H., Varga, R.S.: Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second order Richardson iterative methods, Part II. Numer. Math. **3**, 157-168 (1961)
9. Gutknecht, M.H.: Existence of a solution to the discrete Theodorsen equation for conformal mappings. Math. Comput. **31**, 478-480 (1977)
10. Gutknecht, M.H.: Fast algorithms for the conjugate periodic function. Computing **22**, 79-91 (1979)
11. Gutknecht, M.H.: Numerical experiments on solving Theodorsen's integral equation for conformal maps with the fast Fourier transform and various nonlinear iterative methods. SIAM J. Sci. Stat. Comput. (in press, 1981)
12. Gutknecht, M.H.: The evaluation of the conjugate function of a periodic spline on a uniform mesh. Report 79-05, Sem. ang. Math., ETH Zürich, 1979
13. Henrici, P.: Fast Fourier methods in computational complex analysis. SIAM Rev. **21**, 481-527 (1979)
14. Hübner, O.: Zur Numerik der Theodorsenschen Integralgleichung in der konformen Abbildung. Mitt. Math. Sem. Gießen **140**, 1-32 (1979)
15. Kincaid, D.R.: On complex second-degree iterative methods. SIAM J. Numer. Anal. **11**, 211-218 (1974)
16. Lundwall-Skaar, C.: Konforme Abbildung mit Fast Fourier Transformationen. Diplomarbeit, ETH Zürich, 1975
17. Manteuffel, T.A.: The Tchebychev iteration for nonsymmetric linear systems. Numer. Math. **28**, 307-327 (1977)
18. Niethammer, W.: Iterationsverfahren bei der konformen Abbildung. Computing **1**, 146-153 (1966)
19. Niethammer, W.: Iterationsverfahren und allgemeine Euler-Verfahren. Math. Z. **102**, 288-317 (1967)
20. Ortega, J.M., Rheinbolt, W.C.: Iterative solution of nonlinear equations in several variables. New York: Academic Press 1970
21. Ostrowski, A.M.: Solution of equations and systems of equations. New York: Academic Press 1966
22. Perron, O.: Über Stabilität und asymptotisches Verhalten der Lösungen eines Systems endlicher Differenzengleichungen. J. Reine Angew. Math. **161**, 41-64 (1929)
23. Riley, J.D.: Iteration procedures for the Dirichlet difference problem. Math. Tables Aids Comput. **8**, 125-131 (1954)
24. Varga, R.S.: Matrix iterative analysis. Englewood Cliffs, N.J.: Prentice-Hall 1962
25. Young, D.: Iterative methods for solving partial difference equations of elliptic type. Trans. Amer. Math. Soc. **76**, 92-111 (1954)
26. Young, D.: Iterative solution of large linear systems. New York: Academic Press, 1971
27. Zurmühl, H.: Matrizen und ihre technischen Anwendungen, 4. Aufl.. Berlin Heidelberg New York: Springer 1964
28. Zygmund, A.: Trigonometric series, Vol. I. Cambridge: University Press 1959

Received February 26, 1980