Applied Numerical Mathematics 6 (1989/90) 487–500
North-Holland

APPLICATION OF VECTOR EXTRAPOLATION METHODS TO CONSISTENT SINGULAR LINEAR SYSTEMS

Avram SIDI
Computer Science Department, Technion — Israel Institute of Technology, Haifa 32000, Israel

Consider the linear system of equations $Bx = f$, where B is an $N \times N$ singular matrix, but the system is consistent. In this work we show that iterative techniques coupled with vector extrapolation methods can be used to obtain (approximations to) a solution of $Bx = f$. We do this by extending the results of some previous work on vector extrapolation methods as they apply to nonsingular matrices B. In particular, we show that the minimal polynomial, reduced rank, and modified minimal polynomial extrapolation methods, and the scalar, topological, and vector epsilon algorithms all produce a solution of $Bx = f$ in at most $\operatorname{rank}(B) \leqslant N-1$ steps, and that this solution depends on the initial approximation in a simple way. Asymptotic error estimates and error bounds are given for two different limiting procedures that have been considered in previous work. Although we demonstrate all our results for Richardson's iterative method, they are equally valid for any other iterative method.

1. Introduction

Consider the system of linear equations

$$Bx = f, \tag{1.1}$$

where B is a complex $N \times N$ (not necessarily Hermitian) matrix and f is an N-dimensional vector. We are interested in obtaining a solution of (1.1) when B is singular, but the system (1.1) is consistent. We recall that in this case (1.1) has an infinity of solutions.

The purpose of the present work is to show that iterative techniques coupled with vector extrapolation methods can be used to obtain (approximations to) a solution of (1.1) when B is singular and (1.1) is consistent, under certain conditions.

The iterative techniques for (1.1) that we consider are all of the form

$$x_{j+1} = Ax_j + b, \quad j = 0, 1, \ldots, x_0 \text{ arbitrary}, \tag{1.2}$$

in which $A = M^{-1}Q$ and $B = M - Q$ for some matrices M (nonsingular) and Q, and $b = M^{-1}f$. The Jacobi, Gauss–Seidel, and SOR methods and Richardson's iterative method are but a few examples of such techniques.

Acceleration methods for the problem considered in this work with B symmetric positive-semi-definite have been considered before. For example, Hageman and Young [10, pp. 134–136] show that the Chebyshev acceleration method can be applied in a slightly modified form and give also a convergence analysis. In his paper, Axelsson [1] proposes a conjugate gradient type method and shows that this method is applicable to the problem considered in the present work. Recently, Notay [16] has given a treatment of polynomial acceleration methods, including steepest descent, conjugate gradients, and Chebyshev acceleration, for the same problem.

The extrapolation methods that we consider are the minimal polynomial extrapolation (MPE) of Cabay and Jackson [6], the reduced rank extrapolation (RRE) of Eddy [7] and Mešina [15], the modified minimal polynomial extrapolation (MMPE) of Sidi, Ford and Smith [21], the scalar epsilon algorithm (SEA) of Wynn [24], the vector epsilon algorithm (VEA) of Wynn [25], and the topological epsilon algorithm (TEA) of Brezinski [3]. We note that Sidi, Ford and Smith [21] provide a general framework within which most of these methods and many new ones can be obtained. An extensive survey of vector extrapolation methods has been carried out by Smith, Ford and Sidi [22], and some of the convergence and stability properties for MPE, RRE, MMPE, and TEA have been given in [18,21], and more recently in [20]. We note that, while in [18,21] the iteration matrix A in (1.2) is assumed to be diagonalizable, [20] analyzes the case in which A is not necessarily diagonalizable. Furthermore, it is mentioned in [20] that the analysis of TEA covers also SEA. In a recent work by Sidi [19] the connections between extrapolation and various Krylov subspace methods have been explored, and some new convergence results for them have been proved. For the appropriate references pertaining to Krylov subspace methods and more details on them, see [19]. Some of the results of [19] pertaining to the equivalence of vector extrapolation and Krylov subspace methods have also been given by Beuneu [2], and for TEA by Brezinski [4, pp. 186–189]. Different recursive techniques for the implementation of vector extrapolation methods such as MPE, RRE, MMPE, TEA, and other related methods have been derived by Ford and Sidi in [8], in which a very interesting four-term (lozenge) recursion relation for MPE and RRE is also given. Similar recursive techniques, by which methods such as MMPE and TEA can be implemented, can also be found in [5].

We now give a very brief description of the above-mentioned extrapolation methods, which is based on the developments in [18–21].

Let $x_0, x_1, x_2, \ldots$, be a given sequence of vectors whose limit or antilimit is s. For MPE, RRE, MMPE, and TEA, $s_{n,k}$, the approximation to s, is defined through

$$s_{n,k} = \sum_{j=0}^{k} \gamma_j^{(n,k)} x_{n+j}, \tag{1.3}$$

where the $\gamma_j^{(n,k)}$ are scalars that satisfy the linear equations,

$$\sum_{j=0}^{k} \gamma_j^{(n,k)} = 1, \qquad \sum_{j=0}^{k} u_{ij} \gamma_j^{(n,k)} = 0, \quad 0 \leqslant i \leqslant k-1, \tag{1.4}$$

with

$$u_{ij} = \begin{cases} (u_{n+i}, u_{n+j}) & \text{for MPE,} \\ (w_{n+i}, u_{n+j}) & \text{for RRE,} \\ (q_i, u_{n+j}) & \text{for MMPE,} \\ (q, u_{n+i+j}) & \text{for TEA.} \end{cases} \tag{1.5}$$

Here

$$u_i = x_{i+1} - x_i, \quad w_i = u_{i+1} - u_i, \qquad i = 0, 1, \ldots, \tag{1.6}$$

$q_0, q_1, \ldots$, are fixed and linearly independent vectors, q is a fixed nonzero vector, and $(\cdot, \cdot)$ is any inner product with homogeneity property $(\alpha y, \beta z) = \bar{\alpha}\beta(y, z)$, α, β complex numbers and

y, z vectors. For the N-dimensional vectors considered in this work we take $(x, y) = x^*y$, where x^* denotes the Hermitian conjugate of x.

An acceleration method almost identical to RRE has been proposed by Kaniel and Stein [12]. In this method, x_{n+j} in (1.3) is replaced by x_{n+j+1}, the rest of the method being the same as that for RRE.

If we denote $s_{n,k} = \varepsilon_{2k}^n$ in TEA, the following recursion relations may be shown to hold for ε_k^n, when the vectors q and $x_0, x_1, \ldots$, are real:

$$\begin{aligned}
&\varepsilon_{-1}^n = 0, \quad \varepsilon_0^n = x_n, \qquad n = 0, 1, \ldots, \\
&\varepsilon_{2k+1}^n = \varepsilon_{2k-1}^{n+1} + \frac{q}{\left(q, \varepsilon_{2k}^{n+1} - \varepsilon_{2k}^n\right)}, \qquad n,k = 0, 1, \ldots, \\
&\varepsilon_{2k+2}^n = \varepsilon_{2k}^{n+1} + \frac{\varepsilon_{2k}^{n+1} - \varepsilon_{2k}^n}{\left(\varepsilon_{2k+1}^{n+1} - \varepsilon_{2k+1}^n, \varepsilon_{2k}^{n+1} - \varepsilon_{2k}^n\right)}, \quad n,k = 0, 1, \ldots.
\end{aligned} \tag{1.7}$$

VEA is strictly defined through the recursion relations

$$\begin{aligned}
&\varepsilon_{-1}^n = 0, \quad \varepsilon_0^n = x_n, \qquad n = 0, 1, \ldots, \\
&\varepsilon_{k+1}^n = \varepsilon_{k-1}^{n+1} + \frac{1}{\varepsilon_k^{n+1} - \varepsilon_k^n}, \quad n,k = 0, 1, \ldots,
\end{aligned} \tag{1.8}$$

where $1/z$ is the Samelson inverse of the vector z, defined through $\bar{z}/(z, z)$, where $\bar{z} = z^{*\mathrm{T}}$.

Finally SEA is the result of componentwise application of the Shanks [17] transformation $e_k(S_n)$ (or its equivalent epsilon algorithm of Wynn [24]) to the vector sequence $x_0, x_1, x_2, \ldots$.

Numerical results and theoretical considerations suggest that $s_{n,k}$ for MPE, RRE, and MMPE, and ε_{2k}^n for all three epsilon algorithms, have similar convergence properties. In this respect we note that $s_{n,k}$ for MPE, RRE, and MMPE are obtained from the vectors $x_n, x_{n+1}, \ldots, x_{n+k+1}$, whereas ε_{2k}^n in the epsilon algorithms are obtained from $x_n, x_{n+1}, \ldots, x_{n+2k}$.

In Section 2 we first discuss the consistency of (1.1) in terms of the eigenvectors and principal vectors of B. Next we give a sufficient condition for the convergence of Richardson's iteration method, which generalizes a known result.

In Section 3 we show in the light of the results of Section 2 that extrapolation methods can be applied to the singular but consistent system (1.1) with no modifications to produce approximations to a solution. In particular, we show that, for some $k \leqslant \operatorname{rank}(B) \leqslant N - 1$, $s_{n,k}$ is a solution of (1.1). We also give convergence results for $s_{n,k}$ as $n \to \infty$ and k is held fixed, and we provide error bounds on $s_{n,k}$ for fixed n and increasing k. The results of this section are extensions of those that were obtained in [18–22].

2. Theoretical preliminaries

Consider the linear system in (1.1) with B and f as described in the first paragraph of Section 1. There exists a nonsingular matrix V such that

$$V^{-1}BV = J = \begin{bmatrix} J_1 & & & \\ & J_2 & & 0 \\ & & \ddots & \\ & 0 & & J_\nu \end{bmatrix}, \tag{2.1}$$

where J_i are Jordan blocks of dimension r_i and have the form

$$J_i = \begin{bmatrix} \mu_i & 1 & & & \\ & \mu_i & 1 & 0 & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & 0 & & & \mu_i \end{bmatrix}_{r_i \times r_i}, \quad \mu_i \text{ eigenvalue.} \tag{2.2}$$

If V has the columnwise partition

$$V = \left[v_{11} \mid v_{12} \mid \cdots \mid v_{1r_1} \mid v_{21} \mid v_{22} \mid \cdots \mid v_{2r_2} \mid \cdots \mid v_{\nu 1} \mid v_{\nu 2} \mid \cdots \mid v_{\nu r_\nu} \right], \tag{2.3}$$

then v_{i1} is the eigenvector corresponding to the eigenvalue μ_i, and in case $r_i > 1$, $v_{i2}, \ldots, v_{ir_i}$ are the principal vectors (or the generalized eigenvectors) corresponding to μ_i. We actually have

$$\begin{aligned} Bv_{i1} &= \mu_i v_{i1}, \\ Bv_{ij} &= \mu_i v_{ij} + v_{ij-1}, \qquad j = 2, \ldots, r_i \quad \text{for } r_i > 1. \end{aligned} \tag{2.4}$$

Let us denote

$$\begin{aligned} &\mathcal{S}(B) = \operatorname{span}\{ v_{ij},\ 1 \leqslant j \leqslant r_i \colon\ \mu_i \neq 0 \}, \\ &\mathcal{N}(B) = \operatorname{span}\{ v_{i1} \colon\ \mu_i = 0 \}, \quad \text{null space of } B, \\ &\mathcal{M}(B) = \operatorname{span}\{ v_{ij},\ 2 \leqslant j \leqslant r_i \colon\ \mu_i = 0,\ r_i > 1 \}. \end{aligned} \tag{2.5}$$

Obviously, the intersection of any two of these subspaces consists of the zero vector only, and

$$\mathbb{C}^N = \mathcal{S}(B) \oplus \mathcal{N}(B) \oplus \mathcal{M}(B).$$

If $r_i = 1$ for all the eigenvalues $\mu_i = 0$, $\mathcal{M}(B)$ is defined to be the empty set, and $y \in \mathcal{M}(B)$ is interpreted as $y = 0$.

2.1. Consistency of (1.1) in terms of eigenvectors and principal vectors

It is known that the system in (1.1) is consistent if and only if f is in the column space of B. We now give another necessary and sufficient condition for consistency of (1.1), which we state in terms of the eigenvectors and principal vectors of B.

Theorem 2.1. *When B is singular, the system in* (1.1) *has a solution if and only if f can be expanded in terms of the columns of the matrix V, excluding the vectors v_{ir_i} corresponding to zero eigenvalues. If a solution s exists, then it is of the form $s = s' + s'' + s'''$, where $s' \in \mathcal{S}(B)$ and $s'' \in \mathcal{M}(B)$ are uniquely determined, and $s''' \in \mathcal{N}(B)$ is nonunique. (Recall that $s'' = 0$ if $\mathcal{M}(B)$ is the empty set.)*

Proof. Let us expand f in terms of the columns of V. Then

$$f = \sum_{i=1}^{\nu} \sum_{j=1}^{r_i} \beta_{ij} v_{ij}. \tag{2.6}$$

A solution s of (1.1) can also be expressed in a similar way, namely,

$$s = \sum_{i=1}^{\nu} \sum_{j=1}^{r_i} \alpha_{ij} v_{ij}. \tag{2.7}$$

Substituting (2.6) and (2.7) in (1.1), and invoking (2.4), we obtain the following equations for the α_{ij}:

$$\mu_i \alpha_{ij} + \alpha_{ij+1} = \beta_{ij}, \quad 1 \leqslant j \leqslant r_i - 1, \qquad \mu_i \alpha_{ir_i} = \beta_{ir_i}. \tag{2.8}$$

For $\mu_i \neq 0$ these equations can be solved uniquely for α_{ij}, $1 \leqslant j \leqslant r_i$. For $\mu_i = 0$, however, a solution exists if and only if $\beta_{ir_i} = 0$, and in this solution $\alpha_{ij} = \beta_{ij-1}$, $2 \leqslant j \leqslant r_i$, while α_{i1} is arbitrary. This completes the proof. □

Corollary. *Let the matrix B in Theorem 2.1 be normal. Then any solution s of (1.1) is of the form $s = s' + s'''$, where $s' \in \mathcal{S}(B)$ and $s''' \in \mathcal{N}(B)$. Here $s' = B^+ f$ is unique. (B^+ is the Moore–Penrose generalized inverse of B.)*

The proof of this corollary makes use of the explicit form of B^+ for the case in which B is normal. We do not go into the details of the proof.

2.2. Richardson's iterative method for (1.1)

Let $\omega \neq 0$ be a complex number, and consider Richardson's iterative method for (1.1), namely,

$$x_{j+1} = x_j + \omega(f - Bx_j), \quad j = 0, 1, \ldots, \tag{2.9}$$

x_0 being an arbitrary initial vector, see [23, p. 141]. We note that this method is an iterative method of the form given in (1.2) with

$$A = I - \omega B, \qquad b = \omega f. \tag{2.10}$$

The following is an extension of a known result concerning Richardson's iterative method for the case in which B is Hermitian positive-semidefinite.

Theorem 2.2. *Assume that*

$$\begin{aligned} &\arg \mu_i \in (\theta - \tfrac{1}{2}\pi, \theta + \tfrac{1}{2}\pi) \qquad \text{for } \mu_i \neq 0, \quad \text{some } \theta, \\ &0 < |\omega| < (2 \cos \alpha)/\rho(B), \quad \alpha = \max\{|\arg \mu_i - \theta| : \mu_i \neq 0\}, \\ &\arg \omega = -\theta, \end{aligned} \tag{2.11}$$

where $\rho(B) = \max_i |\mu_i|$ is the spectral radius of B. Then the sequence $\{x_m\}_{m=0}^{\infty}$ obtained from (2.9) converges if and only if (1.1) has a solution s, and x_0 is such that $x_0 - s \in \mathcal{S}(B) \oplus \mathcal{N}(B)$. (Thus if $\mathcal{M}(B)$ is the empty set, no conditions need to be imposed on x_0.) When $\{x_m\}_{m=0}^{\infty}$ converges,

$$\lim_{m \to \infty} x_m = s_* + z_0,$$

where s_* is the only solution of (1.1) *of the form*

$$s_* = s' + s'', \quad s' \in \mathcal{S}(B), \quad s'' \in \mathcal{M}(B),$$

and z_0 depends solely on x_0 through

$$x_0 = x_0' + s'' + z_0, \qquad x_0' \in \mathcal{S}(B), \quad z_0 \in \mathcal{N}(B).$$

Proof. Assume that a solution s to (1.1) exists. Then it is easy to show that

$$x_m - s = A^m(x_0 - s), \quad m = 0, 1, \ldots, \tag{2.12}$$

with A as in (2.10). Let us expand $x_0 - s$ in terms of the columns of the matrix V. We have

$$\begin{aligned} &x_0 - s = r + y + z, \\ &r \in \mathcal{S}(B), \quad y \in \mathcal{M}(B), \quad z \in \mathcal{N}(B). \end{aligned} \tag{2.13}$$

With x_0 fixed and s varying, r and y remain fixed, while z varies. Now the eigenvalues of the matrix A are $\lambda_i = 1 - \omega\mu_i$. The eigenvector and principal vectors of A corresponding to the eigenvalue λ_i are $\hat{v}_{i1} = v_{i1}$ and $\hat{v}_{ij} = (-\omega)^{-j+1} v_{ij}$, $2 \leqslant j \leqslant r_i$, respectively. Also corresponding to $\mu_i = 0$ we have $\lambda_i = 1$, hence $Av_{i1} = v_{i1}$ for this case. Consequently,

$$x_m - s = A^m r + A^m y + z. \tag{2.14}$$

We recall now that (see, for example, [20, Section 2])

$$A^m \hat{v}_{ij} = \sum_{l=1}^{j} \binom{m}{j-l} \lambda_i^{m-j+l} \hat{v}_{il}, \tag{2.15}$$

thus $A^m v_{ij} = O(\lambda_i^m m^{j-1})$ as $m \to \infty$. As a result $\lim_{m\to\infty} A^m v_{ij} = 0$ if and only if $|\lambda_i| < 1$. With the condition (2.11) on the μ_i and ω, $|\lambda_i| < 1$ is satisfied for $\mu_i \neq 0$, as will be shown at the end of this proof. Consequently, $\lim_{m\to\infty} A^m r = 0$. Since $y \in \mathcal{M}(B)$, and $\mathcal{M}(B)$ is the subspace of the principal vectors corresponding to the zero eigenvalues of B, $\mu_i = 0$, for which $\lambda_i = 1$, we have from (2.15) that, when $y \neq 0$, $A^m y = O(m^\sigma)$ as $m \to \infty$, for some integer σ, $1 \leqslant \sigma \leqslant \max\{r_i - 1: \mu_i = 0\}$. Combining all the above, we see that $\lim_{m\to\infty} x_m$ exists provided $y = 0$, and we have

$$\lim_{m\to\infty} x_m = s + z = s_* + z_0$$

in this case. If $y \neq 0$, then $x_m = O(m^\sigma)$ as $m \to \infty$, i.e., $\{x_m\}_{m=0}^{\infty}$ diverges.

Assume now that $\lim_{m\to\infty} x_m$ exists. From (2.9) it follows automatically that (1.1) must have a solution, hence we are back to the case treated in the previous paragraph.

The only thing that remains to be shown is that when $\mu_k \neq 0$ and ω are as in (2.11) $|\lambda_k| = |1 - \omega\mu_k| < 1$. Now let

$$\omega = |\omega| \exp(i\phi), \qquad \mu_k = |\mu_k| \exp(i\psi_k).$$

Therefore, we should have

$$|\lambda_k| = |(1 - |\omega||\mu_k| \cos\alpha_k) - |\omega||\mu_k| \sin\alpha_k| < 1, \quad \alpha_k \equiv \phi + \psi_k, \tag{2.16}$$

or equivalently

$$|\omega|^2 |\mu_k|^2 - 2|\omega||\mu_k| \cos\alpha_k < 0. \tag{2.17}$$

This forces

$$0 < |\omega| < \frac{2\cos\alpha_k}{|\mu_k|} \quad \text{and} \quad |\alpha_k| < \tfrac{1}{2}\pi. \tag{2.18}$$

Picking $\phi = -\theta$, the result follows. □

Remarks.

(1) If there is no θ for which $\arg \mu_i \in (\theta - \frac{1}{2}\pi, \theta + \frac{1}{2}\pi)$, all $\mu_i \neq 0$, then there exists no complex ω for which $\{x_m\}_{m=0}^{\infty}$ converges. Such a situation arises, for example, when B has both positive and negative eigenvalues.

(2) When $\operatorname{Re} \mu_i > 0$ for $\mu_i \neq 0$, we can take $\theta = 0$; thus we can choose ω to be real, as can be seen from (2.11).

(3) When $\mathscr{M}(B)$ is empty, by picking $x_0 = f$ or $x_0 = B\hat{x}_0$ for arbitrary $\hat{x}_0$, we can cause $z_0 = 0$ everywhere, and have $\lim_{m\to\infty} x_m = s_*$. If the matrix B is normal, then, from the corollary to Theorem 2.1, $\lim_{m\to\infty} x_m = B^+ f$ for this choice of x_0.

(4) From Theorem 2.2 we see that when $\mathscr{M}(B)$ is not empty, the sequence $\{x_m\}_{m=0}^{\infty}$ does not converge if $y \neq 0$. This implies that when the zero eigenvalues of B have corresponding principal vectors, the sequence $\{x_m\}_{m=0}^{\infty}$ will not converge, in general, for arbitrary x_0.

Before closing this section we mention that any iterative technique of the form (1.2) with $A = M^{-1}Q$ and $B = M - Q$, is a Richardson iterative method with $\omega = 1$ for the "preconditioned" linear system $M^{-1}Bx = M^{-1}f$ that is equivalent to $Bx = f$. Therefore, there is no loss of generality in considering only Richardson's iterative method. We continue to do so in the next section.

3. Application of extrapolation methods

In this section we show that the vector extrapolation methods mentioned in Section 1 can be applied to the sequence $\{x_m\}_{m=0}^{\infty}$ obtained from Richardson's iterative method for (1.1). We do this by extending the results of [18–22].

A concept that will be of use in the sequel is that of the minimal polynomial of a matrix with respect to a vector. We will call the polynomial

$$P(\lambda) = \sum_{i=0}^{k} c_i \lambda^i, \qquad c_k = 1,$$

the minimal polynomial of the matrix A with respect to the vector u if

$$P(A)u \equiv \left(\sum_{i=0}^{k} c_i A^i\right) u = 0$$

and

$$k = \min\left\{ p: \left(\sum_{i=0}^{p} \beta_i A^i\right) u = 0,\ \beta_p = 1 \right\}.$$

It is known that $P(\lambda)$ exists and is unique. It is also known that if $R(\lambda)$ is another polynomial for which $R(A)u = 0$, then $P(\lambda)$ divides $R(\lambda)$. Consequently, $P(\lambda)$ divides the minimal polynomial of A, which in turn divides the characteristic polynomial of A. For details see [11, pp. 18–19; 19, 22].

Theorem 3.1. *Let $\{x_m\}_{m=0}^{\infty}$ be the sequence generated by Richardson's iterative method* (2.9). *Assume that a solution s for* (1.1) *exists, and let s_* be the (unique) solution of* (1.1) *of the form*

$$s_* = s' + s'', \qquad s' \in \mathcal{S}(B), \quad s'' \in \mathcal{M}(B).$$

Write $x_0 = x_0' + x_0'' + z_0$, where $x_0' \in \mathcal{S}(B)$, $x_0'' \in \mathcal{M}(B)$, and $z_0 \in \mathcal{N}(B)$. Then for any fixed integer $n \geq 0$ a solution s to (1.1) *can be constructed from the vectors $x_n, x_{n+1}, \ldots, x_{n+k_0+1}$, where k_0 is the degree of $P(\lambda)$, the minimal polynomial of A with respect to the vector $x_n - s_* - z_0$, provided $x_0'' = s''$. Under the same condition, $P(\lambda)$ is also the minimal polynomial of A with respect to $u_n = x_{n+1} - x_n$. (As mentioned in Theorem* 2.2, *this condition is automatically satisfied when $\mathcal{M}(B)$ is the empty set.) In fact, if we let*

$$P(\lambda) = \sum_{i=0}^{k_0} c_i \lambda^i, \qquad c_{k_0} = 1, \tag{3.1}$$

be this (unique) polynomial, then

$$s = s_* + z_0 = \sum_{i=0}^{k_0} c_i x_{n+i} \Big/ \sum_{i=0}^{k_0} c_i, \quad \sum_{i=0}^{k_0} c_i \neq 0. \tag{3.2}$$

Furthermore,

$$k_0 \leq \sum_{\substack{i=1 \\ \mu_i \neq 0}}^{\nu} r_i = N - \sum_{\substack{i=1 \\ \mu_i = 0}}^{\nu} r_i \equiv \bar{N} \leq \operatorname{rank}(B) \leq N - 1. \tag{3.3}$$

(*Note that $\bar{N} = \operatorname{rank}(B)$ when $\mathcal{M}(B)$ is empty.*) *The problem of determining the c_i for the construction of s is treated in the proof below.*

Proof. We have that $x_0 - s_* - z_0 \in \mathcal{S}(B)$. This with (2.12) implies that $x_m - x_* - z_0 \in \mathcal{S}(B)$ for any $m \geq 0$. Recall that $\mathcal{S}(B)$ is the subspace spanned by the eigenvectors and principal vectors of B (or A) that correspond to the nonzero eigenvalues of B (or to the eigenvalues of A that are not unity), hence its dimension is $\bar{N}$. Therefore, the minimal polynomial of A with respect to any vector in $\mathcal{S}(B)$ has degree $\leq \bar{N}$, and $\lambda - 1$ is not a divisor of this polynomial. Consequently, $\sum_{i=0}^{k_0} c_i = P(1) \neq 0$. Now

$$\left(\sum_{i=0}^{k_0} c_i A^i \right)(x_n - s_* - z_0) = 0. \tag{3.4}$$

By (2.12) this is equivalent to

$$\sum_{i=0}^{k_0} c_i (x_{n+i} - s_* - z_0) = 0. \tag{3.5}$$

Equality (3.2) now follows from (3.5). The problem that remains is that of determining the c_i. Let us multiply (3.4) on the left by A, and invoke (2.12) again. This results in

$$\sum_{i=0}^{k_0} c_i(x_{n+i+1} - s_* - z_0) = 0. \tag{3.6}$$

Subtracting (3.5) from (3.6), we obtain the linear system of equations

$$\sum_{i=0}^{k_0} c_i u_{n+i} = 0. \tag{3.7}$$

With $c_{k_0} = 1$, this system consists of N equations in k_0 unknowns, hence is overdetermined. However, since the c_i are known to exist uniquely, this system is consistent. Thus the c_i can be uniquely determined from this system. Note that (3.2) and (3.7) involve only the vectors x_{n+i}, $i = 0, 1, \ldots, k_0 + 1$. Furthermore, by the fact that $u_{j+1} = Au_j$, $j \geq 0$, (3.7) can be rewritten as $P(A)u_n = 0$.

Finally, we show that $P(\lambda)$ is the minimal polynomial of A with respect to u_n. Suppose that this is not true. Then there exists another polynomial $\hat{P}(\lambda)$ of degree less than k_0 such that $\hat{P}(A)u_n = 0$. But $u_n = (A - I)(x_n - s_* - z_0)$. Consequently, $(A - I)\hat{P}(A)(x_n - s_* - z_0) = 0$. Observe that $A - I = -\omega B$, so that the last equality is possible if either $\tilde{x} = \hat{P}(A)(x_n - s_* - z_0) = 0$ or $\tilde{x} = \hat{P}(A)(x_n - s_* - z_0) \in \mathcal{N}(B)$. Now $\tilde{x} = 0$ is impossible, since the degree of $\hat{P}(\lambda)$ is less than k_0. Also $\tilde{x} \in \mathcal{N}(B)$ is impossible since $\tilde{x} \in \mathcal{S}(B)$ by the assumption that $x_0 - s_* - z_0 \in \mathcal{S}(B)$. We thus have a contradiction. The result now follows from this, from $P(A)u_n = 0$, and from the uniqueness of $P(\lambda)$. This completes the proof. □

Remarks.

(1) In Theorem 3.1 we have assumed all the conditions of Theorem 2.2, with the exception of those in (2.11). This means that convergence of Richardson's iterative method is not necessary for Theorem 3.1 to be true.

(2) It seems that the condition $x_0 - s_* - z_0 \in \mathcal{S}(B)$ is absolutely necessary. We have not been able to give a construction of s from the vector sequence $\{x_m\}_{m=0}^{\infty}$ in the absence of this condition. We also recall that in the absence of this condition the sequence $\{x_m\}_{m=0}^{\infty}$ is unbounded for any value of ω and any spectrum for B.

(3) In case $\mathcal{M}(B)$ is the empty set, we can choose $x_0 = f$ or $x_0 = B\hat{x}_0$ for some arbitrary vector $\hat{x}_0$, and cause $z_0 = 0$ everywhere in Theorem 3.1. If the matrix B is normal, then the left-hand side of (3.2) becomes $s = s_* = B^+ f$ with this choice of x_0. This remark also applies to Theorem 3.2 and subsequent developments below.

Another important result follows from (2.12) and can be stated as follows:

Theorem 3.2. *With the sequence* $\{x_m\}_{m=0}^{\infty}$ *exactly as described in Theorem* 3.1 *with the notation therein, for all m sufficiently large we have*

$$x_m = s_* + z_0 + \sum_{i=1}^{\bar{\nu}} P_i(m)\lambda_i^m \quad \textit{for some } \bar{\nu} \leq \nu. \tag{3.8}$$

Here λ_i *are complex numbers satisfying*

$$\lambda_i \neq 0, \qquad \lambda_i \neq 1, \qquad \lambda_i \neq \lambda_j, \quad \textit{if } i \neq j, \tag{3.9}$$

and they are ordered such that

$$|\lambda_1| \geqslant |\lambda_2| \geqslant |\lambda_3| \geqslant \cdots. \tag{3.10}$$

The $P_i(m)$ are polynomials in m with vector coefficients, given in the form

$$P_i(m) = \sum_{j=0}^{p_i} y_{ij}\binom{m}{j}, \quad \text{some integer } p_i \geqslant 0, \tag{3.11}$$

where $\binom{m}{j}$ is the binomial coefficient, and the vectors y_{ij}, $j = 0, 1, \ldots, p_j$, $i = 1, 2, \ldots, \bar{\nu}$, form a linearly independent set. (Specifically, the λ_i are distinct nonzero eigenvalues of A corresponding to the nonzero eigenvalues μ_q of B, and y_{ij} are some linearly independent combinations of the eigenvectors and principal vectors corresponding to nonzero equal eigenvalues μ_q. As a result, the integer $p_i + 1$ is smaller than or equal to the maximum of the r_q for which $1 - \omega\mu_q = \lambda_i$. Also m sufficiently large means $m \geqslant \max\{r_i\colon 1 - \omega\mu_i = 0\}$.)

We shall not prove Theorem 3.2. We shall only mention that (3.8)–(3.11) follow from (2.12) with the help of (2.15). For the exact details see [20, Section 2].

The implication of Theorems 3.1 and 3.2 is that all vector extrapolation methods mentioned in Section 1 of the present work and their equivalent Krylov subspace methods can be applied, with no changes, to the sequence $\{x_m\}_{m=0}^{\infty}$ of this section to obtain approximations to $s_* + z_0$, provided $x_0 - s_* - z_0 \in \mathcal{S}(B)$, which we assume to be satisfied in the sequel. In Section 1, we used $s_{n,k}$ to denote the approximations obtained employing MPE, RRE, MMPE, and TEA, and ε_{2k}^n to denote those obtained employing the epsilon algorithms. For convenience we will let $s_{n,k}$ stand also for ε_{2k}^n.

3.1. Implications of Theorem 3.1

3.1.1. s_{n,k_0} is a solution

Employing Theorem 3.1 of the present work in [19, Theorem 2.1] we have

$$s_{n,k_0} = s_* + z_0 \tag{3.12}$$

for MPE and RRE unconditionally, and for MMPE and TEA under certain mild conditions. Equality (3.12) for SEA is very simple to show. Its truth for VEA follows by combining Theorem 3.1 with McLeod's theorem. This theorem was originally proved by McLeod [14] for the case in which the c_i in Theorem 3.1 are real, while the case of complex c_i was proved recently by Graves-Morris [9].

The implication of (3.12) is that Krylov subspace methods will produce $s_* + z_0$, a solution to (1.1), in exactly k_0, hence in at most $\bar{N} \leqslant \operatorname{rank}(B) \leqslant N - 1$, steps. (All conjugate gradient type methods are Krylov subspace methods.)

3.1.2. Conditions for existence of $s_{n,k}$

Conditions for the existence of $s_{n,k}$ with $k \leqslant k_0$ for MPE, RRE, MMPE, and TEA are given in [19, Theorem 2.1] and they apply to the problem treated in the present work with no modification.

According to this theorem, $s_{n,k}$ for RRE always exists uniquely for $k \leqslant k_0$. The proof of this is achieved by showing that $\det(\tilde{U}^*\tilde{U}) \neq 0$, with $\tilde{U} = (A - I)U$, where U is the $N \times k$ matrix defined by

$$U = [u_n \mid u_{n+1} \mid \cdots \mid u_{n+k-1}]. \tag{3.13}$$

Since the argument given in [19] in the proof of $\det(\tilde{U}^*\tilde{U}) \neq 0$ is not valid when B is singular, we need to provide another argument, to which we now turn. First, since k_0 is the degree of the minimal polynomial of A with respect to u_n, and $k \leqslant k_0$, the k columns $u_n, u_{n+1}, \ldots, u_{n+k-1}$ of U are linearly independent. Consequently, $U\xi \neq 0$ for any $\xi \neq 0$, $\xi \in \mathbb{C}^k$. Consider now $\tilde{U}\xi = (A - I)U\xi$, $\xi \in \mathbb{C}^k$. It is easy to see that $\det(\tilde{U}^*\tilde{U}) \neq 0$ if and only if $\xi^*\tilde{U}^*U\xi = (\tilde{U}\xi)^*(\tilde{U}\xi) > 0$, which is true if and only if $\tilde{U}\xi \neq 0$. But $\tilde{U}\xi = 0$ if and only if either $U\xi = 0$ or $U\xi \in \mathcal{N}(B)$ since $A - I = -\omega B$ (cf. the last part of the proof of Theorem 3.1), both of which are impossible. Therefore, $\tilde{U}\xi \neq 0$ for $\xi \neq 0$. This completes the proof of $\det(\tilde{U}^*\tilde{U}) \neq 0$.

Following [19, Theorem 2.1] it is mentioned that $s_{n,k}$ with $k < k_0$ does not always exist uniquely for MPE although it does for RRE. In [19, Theorem 2.2] a sufficient condition is given for the unique existence of $s_{n,k}$ with $k < k_0$ for MPE when B is not singular, this condition being that the matrix $I - A$ have a positive-definite Hermitian part. When B is singular, this condition needs to be modified, and we have that, if the restriction of $I - A = \omega B$ to the subspace $\mathcal{S}(B)$ has a positive-definite Hermitian part, then $s_{n,k}$ with $k < k_0$ exists uniquely for MPE. The proof of this is as that of [19, Theorem 2.2] once we recall that $U\xi \in \mathcal{S}(B)$, $\xi \in \mathbb{C}^k$.

Before proceeding further, we mention some useful facts about the restriction $\hat{B}$ of B to the subspace $\mathcal{S}(B)$, as we will need them below. The eigenvalues of $\hat{B}$ are the nonzero eigenvalues μ_i of B, and its corresponding eigenvectors and principal vectors are precisely v_{ij}, $1 \leqslant j \leqslant r_i$. If B is normal, then so is $\hat{B}$. If B is Hermitian, then so is $\hat{B}$.

3.1.3. Error bounds for increasing k

Bounds on different norms of the errors $e_k = s_{0,k} - s_* - z_0$ can be obtained exactly as in [19, Section 4] if the operator $C = I - A = \omega B$ there is replaced by its restriction $\hat{C}$ to the subspace $\mathcal{S}(B)$, and all operator norms there are defined in $\mathcal{S}(B)$. The Hermitian part $\hat{C}_h = \frac{1}{2}(\hat{C} + \hat{C}^*)$ of $\hat{C}$ is assumed to be positive-definite. With these changes all the results of [19, Section 4] remain intact. We elaborate on this below.

Let us denote the vector l_2-norm in $\mathbb{C}^N$ by $\| x \|$, $x \in \mathbb{C}^N$. We will use $\| D \|_{\mathcal{S}(B)}$ to denote the operator norm of D in $\mathcal{S}(B)$, induced by the vector l_2-norm in $\mathbb{C}^N$, i.e.,

$$\| D \|_{\mathcal{S}(B)} = \max_{\substack{x \in \mathcal{S}(B) \\ x \neq 0}} \frac{\| Dx \|}{\| x \|}. \tag{3.14}$$

Obviously, if $\hat{D}$ is the restriction of D to $\mathcal{S}(B)$, then

$$\| \hat{D} \|_{\mathcal{S}(B)} = \| D \|_{\mathcal{S}(B)}. \tag{3.15}$$

Denote also

$$\| x \|'_{\mathcal{S}(B)} = \sqrt{x^*\hat{C}_h x}, \quad x \in \mathcal{S}(B). \tag{3.16}$$

This is a true vector norm in $\mathcal{S}(B)$, since $\hat{C}_h$ is assumed to be positive-definite on $\mathcal{S}(B)$.

Let $s_0 \equiv x_0$ and $s_k \equiv s_{0,k}$, $k = 1, 2, \ldots$. The residual vector for arbitrary x is defined as $r(x) = f - Bx$. It is easy to see that $r(x) = -B(x - s_* - z_0)$. Thus $r(x_0) \in \mathscr{S}(B)$, and consequently $r(s_k) \in \mathscr{S}(B)$ for $k = 0, 1, \ldots$.

Let us denote by π_k the set of all polynomials $Q_k(\lambda)$ of degree at most k satisfying $Q_k(0) = 1$.

In view of the above, [19, Theorem 4.2], for the case treated in the present work, reads as follows:

Theorem 3.3. *For RRE*

$$\| r(s_k) \| \leqslant \hat{\Gamma}_k \| r(s_0) \|, \tag{3.17}$$

where

$$\hat{\Gamma}_k = \min_{Q_k \in \pi_k} \| Q_k(\hat{C}) \|_{\mathscr{S}(B)}. \tag{3.18}$$

Similarly, [19, Theorem 4.5] takes the following form:

Theorem 3.4. *For MPE*

$$\| e_k \|'_{\mathscr{S}(B)} \leqslant \hat{L}\sqrt{\operatorname{cond}(\hat{C}_h)}\, \hat{\Gamma}_k \| e_0 \|'_{\mathscr{S}(B)}, \tag{3.19}$$

where $\hat{\Gamma}_k$ *is exactly as in* (3.18),

$$\hat{L} = \sqrt{1 + \hat{\Lambda}^2}, \qquad \hat{\Lambda} = \rho\big(\hat{C}_h^{-1}\hat{C}_a\big), \quad \hat{C}_a = \tfrac{1}{2}(\hat{C} - \hat{C}^*), \tag{3.20}$$

and $\operatorname{cond}(\hat{C}_h)$ *is defined to be the ratio of the largest eigenvalue of* $\hat{C}_h$ *to its smallest eigenvalue. If* B *is a normal operator, then*

$$\| e_k \|'_{\mathscr{S}(B)} \leqslant \hat{L}\hat{\Gamma}_k \| e_0 \|'_{\mathscr{S}(B)}. \tag{3.21}$$

We note that the quantity $\hat{\Gamma}_k$ that appears both in (3.17) and (3.19) and (3.21) is used in obtaining different bounds for $\| r(s_k) \|$ in RRE and for $\| e_k \|'_{\mathscr{S}(B)}$ in MPE. For details see [19]. A special case is considered in the next paragraph.

When B is a Hermitian positive-semidefinite operator, then $C = \omega B$ implies that C is Hermitian positive-semidefinite with $\omega = 1$. Therefore, $C_h = C = B$. If we now denote the largest eigenvalue of B by $\mu_{\max}$ and its smallest nonzero eigenvalue by $\mu_{\min}$, and define $\kappa = \mu_{\max}/\mu_{\min}$ and $\eta = (\sqrt{\kappa} - 1)/(\sqrt{\kappa} + 1)$, then

$$\| r(s_k) \| \leqslant 2\eta^k \| r(s_0) \| \quad \text{for RRE} \tag{3.22}$$

and

$$\| e_k \|'_{\mathscr{S}(B)} \leqslant 2\eta^k \| e_0 \|'_{\mathscr{S}(B)} \quad \text{for MPE.} \tag{3.23}$$

Cf. [19, eqs. (4.36) and (4.37)]. We also note that for this case

$$\| e_k \|'_{\mathscr{S}(B)} = \sqrt{(s_k - s_* - z_0)^* B (s_k - s_* - z_0)}\,. \tag{3.24}$$

As is shown in [19, Theorem 2.4], MPE and RRE for this case become equivalent to the method of conjugate gradients and the method of conjugate residuals respectively.

Additional results can be obtained as in [19, Section 4].

We also mention that the Chebyshev acceleration method as generalized to the complex plane by Manteuffel [13], can be used to obtain approximations to a solution of (1.1), provided the ellipse F chosen in this method is taken to contain only the nonzero eigenvalues μ_i of B in its interior. If we start with the vector $x_0 = x_0' + s'' + z_0$, and denote by s_k^{Ch} the approximation to $s_* + z_0$ obtained at the kth step, then we have

$$\| s_k^{\mathrm{Ch}} - s_* - z_0 \| \leq \frac{\| T_k((dI - B)/c) \|_{\mathscr{S}(B)}}{| T_k(d/c) |} \| x_0 - s_* - z_0 \|, \tag{3.25}$$

where d is the center of the ellipse F and $d \pm c$ are its foci. Here d and c are complex numbers. $T_k(\lambda)$ is the Chebyshev polynomial of order k. The parameters that are needed for the Chebyshev acceleration are c and d.

3.2. *Implications of Theorem 3.2: An acceleration result for $n \to \infty$*

From Theorem 3.2 we see that the sequence $\{x_m\}_{m=0}^{\infty}$ is exactly of the form assumed in [20, Section 1]. Consequently, [20, Theorem 3.1] applies to $s_{n,k}$ of the present work too, providing a powerful convergence acceleration result for k fixed and $n \to \infty$ for MPE, RRE, MMPE, TEA, and SEA. Roughly speaking, provided

$$|\lambda_t| > |\lambda_{t+1}| \tag{3.26}$$

in (3.8), and

$$k = \sum_{i=1}^{t} (p_i + 1), \tag{3.27}$$

we have

$$s_{n,k} - (s_* + z_0) = \mathrm{O}(n^{\alpha} |\lambda_{t+1}|^n) \quad \text{as } n \to \infty, \tag{3.28}$$

for some nonnegative integer α, and also

$$\lim_{n \to \infty} \sum_{j=0}^{k} \gamma_j^{(n,k)} \lambda^j = \prod_{i=1}^{t} \left(\frac{\lambda - \lambda_i}{1 - \lambda_i} \right)^{p_i + 1}, \tag{3.29}$$

unconditionally for MPE and RRE, and under some mild conditions for MMPE and TEA. From the note at the end of [20, Section 3] it also follows that the above results hold also for SEA unconditionally. For the precise statement of these and other related results see [20].

Remark. The results concerning the application of VEA to singular systems that are given in the present work can be considered a generalization of those given in the paper by C. Brezinski, "Some results in the theory of the vector ε-algorithm" (*Linear Algebra Appl.* 8 (1974) 77–86). The method MMPE was also defined in [3] as a generalization of the Shanks transformation. The paper by B.P. Pugachev, "Acceleration of the convergence of iterative processes and a method of solving systems of nonlinear equations" (*U.S.S.R. Comput. Math. Math. Phys.* 17 (1978) 199–207), provides a convergence analysis for MMPE in case the iteration matrix is diagonalizable. This analysis is similar to the one given in [21]. The author is grateful to the referee for pointing these out to him.

References

[1] O. Axelsson, Conjugate gradient type methods for unsymmetric and inconsistent systems of linear equations, *Linear Algebra Appl.* 29 (1980) 1–16.

[2] J. Beuneu, Minimal polynomial projection methods, Preprint (1984).

[3] C. Brezinski, Généralisations de la transformation de Shanks, de la table de Padé et de l'ε-algorithme, *Calcolo* 12 (1975) 317–360.

[4] C. Brezinski, *Padé-Type Approximation and Generalized Orthogonal Polynomials* (Birkhäuser, Basel, 1980).

[5] C. Brezinski, Recursive interpolation, extrapolation and projection, *J. Comput. Appl. Math.* 9 (1983) 369–376.

[6] S. Cabay and L.W. Jackson, A polynomial extrapolation method for finding limits and antilimits of vector sequences, *SIAM J. Numer. Anal.* 13 (1976) 734–752.

[7] R.P. Eddy, Extrapolating to the limit of a vector sequence, in: P.C.C. Wang, ed., *Information Linkage between Applied Mathematics and Industry* (Academic Press, New York, 1979) 387–396.

[8] W.F. Ford and A. Sidi, Recursive algorithms for vector extrapolation methods, *Appl. Numer. Math.* 4 (1988) 477–489.

[9] P.R. Graves-Morris, Vector valued rational interpolants I, *Numer. Math.* 42 (1983) 331–348.

[10] L.A. Hageman and D.M. Young, *Applied Iterative Methods* (Academic Press, New York, 1981).

[11] A.S. Householder, *The Theory of Matrices in Numerical Analysis* (Blaisdell, New York, 1964).

[12] S. Kaniel and J. Stein, Least-square acceleration of iterative methods for linear equations, *J. Optim. Theory Appl.* 14 (1974) 431–437.

[13] T.A. Manteuffel, The Chebyshev iteration for nonsymmetric linear systems, *Numer. Math.* 28 (1977) 307–327.

[14] J.B. McLeod, A note on the ε-algorithm, *Computing* 7 (1971) 17–24.

[15] M. Mešina, Convergence acceleration for the iterative solution of the equations $X = AX + f$, *Comput. Methods Appl. Mech. Engrg.* 10 (1977) 165–173.

[16] Y. Notay, Polynomial acceleration of iterative schemes associated with subproper splittings, *J. Comput. Appl. Math.* 24 (1988) 153–167.

[17] D. Shanks, Non-linear transformations of divergent and slowly convergent sequences, *J. Math. Phys.* 34 (1955) 1–42.

[18] A. Sidi, Convergence and stability properties of minimal polynomial and reduced rank extrapolation algorithms, *SIAM J. Numer. Anal.* 23 (1986) 197–209; originally: NASA TM-83443 (1983).

[19] A. Sidi, Extrapolation vs. projection methods for linear systems of equations, *J. Comput. Appl. Math.* 22 (1988) 71–88.

[20] A. Sidi and J. Bridger, Convergence and stability analyses for some vector extrapolation methods in the presence of defective iteration matrices, *J. Comput. Appl. Math.* 22 (1988) 35–61.

[21] A. Sidi, W.F. Ford and D.A. Smith, Acceleration of convergence of vector sequences, *SIAM J. Numer. Anal.* 23 (1986) 178–196; originally: NASA TP-2193 (1983).

[22] D.A. Smith, W.F. Ford and A. Sidi, Extrapolation methods for vector sequences, *SIAM Rev.* 29 (1987) 199–233. Erratum: Correction to "Extrapolation methods for vector sequences", *SIAM Rev.* 30 (1988) 623–624.

[23] R.S. Varga, *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1962).

[24] P. Wynn, On a device for computing the $e_m(S_n)$ transformation, *MTAC* 10 (1956) 91–96.

[25] P. Wynn, Acceleration techniques for iterated vector and matrix problems, *Math. Comp.* 16 (1962) 301–322.