Journal of Computational and Applied Mathematics 22 (1988) 71–88
North-Holland

Extrapolation vs. projection methods for linear systems of equations

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

Received 24 April 1987
Revised 12 October 1987

Abstract: It is shown that the four vector extrapolation methods, minimal polynomial extrapolation, reduced rank extrapolation, modified minimal polynomial extrapolation, and topological epsilon algorithm, when applied to linearly generated vector sequences, are Krylov subspace methods, and are equivalent to some well known conjugate gradient type methods. A unified recursive method that includes the conjugate gradient, conjugate residual, and generalized conjugate gradient methods is developed. Finally, the error analyses for these methods are unified, and some known and some new error bounds for them are given.

Keywords: Iterative methods, vector sequences, extrapolation, projection, Krylov subspace methods, conjugate gradient type methods, Chebyshev acceleration.

1. Introduction

The purpose of the present work is to investigate the connection between extrapolation (or convergence acceleration) methods for sequences of vectors and projection methods for solving systems of linear equations. The extrapolation methods that we wish to consider are the minimal polynomial extrapolation (MPE) of Cabay and Jackson [7], the reduced rank extrapolation (RRE) of Eddy [9] and Mešina [20], the modified minimal polynomial extrapolation (MMPE) of Sidi, Ford and Smith [26], and the topological epsilon algorithm (TEA) of Brezinski [4]. The projection methods of interest, on the other hand, are Krylov subspace methods for linear equations; in particular, the conjugate gradient method (CG) of Hestenes and Stiefel [15], an extension due to Saad [21] of the method of Arnoldi [1] for eigenvalue problems, to which we shall refer as the Arnoldi method for short, the conjugate residual method (CR) of Stiefel [28], the generalized conjugate gradient method (GCG) of Concus and Golub [8] and Widlund [30], the generalized conjugate residual method (GCR) of Eisenstat, Elman, and Schultz [11], and the method of Lanczos [17]. Other related Krylov subspace methods will be mentioned later in this work.

A survey of vector extrapolation methods has been carried out by Smith, Ford, and Sidi [27], and some of the convergence and stability properties for MPE, RRE, MMPE, and TEA are given by Sidi [24,26], and more recently by Sidi and Bridger [25]. Recursive algorithms that can be used for implementing all of these methods have recently been given by Ford and Sidi [13]. Of these methods MMPE and TEA can also be implemented by the algorithms of Brezinski [6].

In Section 2 we show that the four extrapolation methods, when applied to linearly generated vector sequences, are bona fide Krylov subspace methods and conjugate gradient type methods as well. In particular, we show that MPE, RRE, and TEA are equivalent to the Arnoldi method, GCR, and the Lanczos method respectively. Some of the results of this section have been obtained also by Beuneu [3]. In Section 3 we give a unified recursive algorithm from which complex versions of CG, CR, and GCG can be obtained as special cases. Finally, in Section 4, we give a unified approach for error bounds in MPE and RRE (equivalently the Arnoldi method and GCR respectively) based on the approaches of [30] for GCG and of Manteuffel [19] for Chebyshev acceleration, from which there follow some known and some new results.

Before closing, we mention that the four extrapolation methods, unlike conjugate gradient type algorithms, are designed to work directly with the vector sequence, whose limit or antilimit is being sought, and do not depend on how this sequence is generated. Thus they are quite different than the extensions of conjugate gradient type methods when applied to sequences that are generated nonlinearly.

2. Extrapolation vs. projection methods

Let $\boldsymbol{B}$ be an inner product space over $\mathbb{C}$, the field of complex numbers, and let (x, y) and $\|x\| = \sqrt{(x, x)}$ be respectively the inner product and norm associated with $\boldsymbol{B}$. The homogeneity property of the inner product is such that, for α and β complex numbers and x and y vectors, $(\alpha x, \beta y) = \bar{\alpha}\beta(x, y)$.

2.1. *Summary of algebraic properties of vector extrapolation methods*

Let $x_0, x_1, x_2, \ldots,$be a sequence of vectors in $\boldsymbol{B}$, and define the first and second order forward differences of the x_i by

$$u_i = \Delta x_i = x_{i+1} - x_i, \qquad w_i = \Delta^2 x_i = \Delta u_i, \quad i = 0, 1, \ldots . \tag{2.1}$$

As is shown in [24] and [26], all four vector extrapolation methods, MPE, RRE, MMPE, and TEA, when applied to the sequence $x_0, x_1, \ldots,$ produce approximations $s_{n,k}$ to the limit or antilimit of this sequence, which are of the form

$$s_{n,k} = \sum_{j=0}^{k} \gamma_j^{(n,k)} x_{n+j}, \tag{2.2}$$

with

$$\sum_{j=0}^{k} \gamma_j^{(n,k)} = 1. \tag{2.3}$$

The $\gamma_j = \gamma_j^{(n,k)}$, in addition to (2.3), satisfy and are determined by the linear equations

$$\sum_{j=0}^{k} u_{i,j}\gamma_j = 0, \quad 0 \leqslant i \leqslant k - 1, \tag{2.4}$$

where the scalars $u_{i,j}$ are

$$u_{i,j}=(u_{n+i},\, u_{n+j}) \quad \text{for MPE,} \tag{2.5a}$$

$$u_{i,j}=(w_{n+i},\, u_{n+j}) \quad \text{for RRE,} \tag{2.5b}$$

$$u_{i,j}=(q_i,\, u_{n+j}) \quad \text{for MMPE,} \tag{2.5c}$$

$$u_{i,j}=(q,\, u_{n+i+j}) \quad \text{for TEA,} \tag{2.5d}$$

provided the matrix of the equations (2.3) and (2.4) is nonsingular. In (2.5c) $\{q_0,\dots,q_{k-1}\}$ is a linearly independent set of fixed vectors in $\boldsymbol{B}$, and in (2.5d) q is a fixed vector in $\boldsymbol{B}$. As such, $s_{n,k}$, for MPE, RRE, and MMPE, is based solely on the vectors x_i, $n \leqslant i \leqslant n+k+1$, and for TEA it is based on x_i, $n \leqslant i \leqslant n+2k$.

By inspection it can be shown that $s_{n,k}$ for all four methods can be written as the quotient of two determinants in the form

$$s_{n,k}=\frac{D(x_n,\, x_{n+1},\dots,x_{n+k})}{D(1,1,\dots,1)}, \tag{2.6}$$

where

$$D(\sigma_0,\, \sigma_1,\dots,\sigma_k)=\begin{vmatrix} \sigma_0 & \sigma_1 & \cdots & \sigma_k \\ u_{0,0} & u_{0,1} & \cdots & u_{0,k} \\ \vdots & \vdots & & \vdots \\ u_{k-1,0} & u_{k-1,1} & \cdots & u_{k-1,k} \end{vmatrix}, \tag{2.7}$$

provided $D(1,1,\dots,1) \neq 0$. Here the determinant $D(\sigma_0,\dots,\sigma_k)$ is to be taken as its expansion with respect to its first row in case σ_i are vectors. See [24] and [26] for details.

We note that the convergence and stability of $s_{n,k}$, for all four methods, have been analyzed in [24], [25] and [26] for the limiting case in which k is held fixed and $n \to \infty$, and for sequences of vectors such as those obtained by iterative solution of linear systems of equations. In the remainder of this work we turn our attention exactly to these sequences and analyze the behavior of $s_{n,k}$ for fixed n and increasing k.

Note: The original definitions given for $u_{i,j}$ in [26] regarding the methods MMPE and TEA are more general than those given in (2.5c) and (2.5d), and are good also for the case in which $\boldsymbol{B}$ is a normed linear space only, i.e., $\boldsymbol{B}$ is not required to be an inner product space. In this case (2.5c) and (2.5d) are replaced by

$$u_{i,j}=Q_i(u_{n+j}) \quad \text{for MMPE} \tag*{(2.5c)$'$}$$

and

$$u_{i,j}=Q(u_{n+i+j}) \quad \text{for TEA,} \tag*{(2.5d)$'$}$$

respectively. Here Q_i and Q are fixed bounded linear functionals on $\boldsymbol{B}$, and Q_i, $i=0,1,\dots$, are linearly independent. Further generalizations are obtained by allowing these functionals to depend on n, i.e., by letting

$$u_{i,j}=Q_i^{(n)}(u_{n+j}) \quad \text{for MMPE} \tag*{(2.5c)$''$}$$

and

$$u_{i,j}=Q^{(n)}(u_{n+i+j}) \quad \text{for TEA.} \tag*{(2.5d)$''$}$$

This, of course, results in the vectors q_i and q in (2.5c) and (2.5d) being replaced by some new vectors $q_i^{(n)}$ and $q^{(n)}$ respectively, in case $\boldsymbol{B}$ is an inner product space. In fact, by choosing $q_i^{(n)} = u_{n+i}$ or $q_i^{(n)} = w_{n+i}$, (2.5c)″ reduces to (2.5a) or (2.5b) respectively. Obviously, when $q_i^{(n)}$ are independent of n (2.5c)″ is the same as (2.5c). Thus this new general version of MMPE, which we shall designate generalized MPE (GMPE), provides us with a very comprehensive class of vector extrapolation methods that includes MPE, RRE, and MMPE and other new methods as well. We add that the recursive algorithms developed in [13] for methods like MPE and RRE are suitable for implementing GMPE too. An interesting result derived in [13] states that there is a four-term (lozange) recursion relation involving the $s_{n,k}$.

2.2. *Conditions for existence of $s_{n,k}$ for linear systems*

Let now A be a linear operator mapping $\boldsymbol{B}$ to itself, and consider the operator equation

$$x = Ax + b. \tag{2.8}$$

We assume that $I - A$ is nonsingular so that (2.8) has a unique solution, which we denote by s. We now pick the vector x_0 arbitrarily and generate the sequence $x_0, x_1, x_2, \ldots,$ by

$$x_{j+1} = Ax_j + b, \quad j = 0, 1, \ldots . \tag{2.9}$$

Define the residual vector $r(x)$ for a vector x by

$$r(x) = Ax + b - x. \tag{2.10}$$

Obviously

$$r(s) = 0 \quad \text{and} \quad r(x_i) = u_i, \quad i = 0, 1, \ldots . \tag{2.11}$$

Furthermore, it can easily be shown that

$$x_j - s = A^j(x_0 - s), \qquad u_j = A^j u_0, \qquad w_j = A^j w_0, \quad j = 0, 1, \ldots . \tag{2.12}$$

The case in which $\boldsymbol{B}$ is the Euclidean space $\mathbb{C}^N$ for some positive integer N, and A is an $N \times N$ (complex) matrix is of special interest both mathematically and practically. For this case we take $(x, y) = x^*y$, where x^* denotes the hermitian conjugate of x. It is this case that we consider below and in Theorems 2.1 and 2.2, although most of our development is valid in the setting of the general vector space $\boldsymbol{B}$.

Let $P(\lambda) = \sum_{i=0}^{k_0} c_i \lambda^i$, $c_{k_0} = 1$, be the minimal polynomial of the matrix A with respect to the vector $x_n - s$, i.e.,

$$P(A)(x_n - s) \equiv \left(\sum_{i=0}^{k_0} c_i A^i \right)(x_n - s) = 0 \tag{2.13}$$

and

$$k_0 = \min\left\{ p \,\middle|\, \left(\sum_{i=0}^{p} \beta_i A^i \right)(x_n - s) = 0,\ \beta_p = 1 \right\}. \tag{2.14}$$

It is known that $P(\lambda)$ exists and is unique. It is also known that if $R(\lambda)$ is another polynomial for which $R(A)(x_n - s) = 0$, then $P(\lambda)$ divides $R(\lambda)$. Consequently, if $Q(\lambda)$ is the minimal polynomial of A, and has degree m, then $P(\lambda)$ divides $Q(\lambda)$, $Q(\lambda)$ divides the characteristic polynomial of A, and hence $k_0 \leqslant m \leqslant N$. For details see [16, pp. 18–19]. When $I-A$ is nonsingular it can be shown that $P(\lambda)$ is also the minimal polynomial of A with respect to u_n.

With $P(\lambda)$ and k_0 as above it is known that

$$s = \sum_{i=0}^{k_0} c_i x_{n+i} / \sum_{i=0}^{k_0} c_i, \tag{2.15}$$

and that $\sum_{i=0}^{k_0} c_i \neq 0$ since $I-A$ is nonsigular. It is also known that s_{n,k_0} always exists uniquely for MPE and RRE, and that

$$s_{n,k_0} = s \tag{2.16}$$

for these two methods. For more details on this see [27]. If $D(1, 1, \ldots, 1) \neq 0$ for $k = k_0$ in (2.6), then s_{n,k_0} exists uniquely and satisfies (2.16) also for MMPE and TEA; otherwise, s_{n,k_0} does not exist uniquely for these methods, hence (2.16) does not necessarily hold.

We now wish to investigate the conditions under which $D(1, 1, \ldots, 1) \neq 0$ for each of the four methods above.

Theorem 2.1. *With A, b, x_i, $P(\lambda)$, and k_0 as above, $k \leqslant k_0$ is necessary for $D(1, 1, \ldots, 1) \neq 0$ and hence for the existence of a unique $s_{n,k}$ for all four extrapolation methods. Let U be the $N \times k$ matrix whose columns are $u_n, u_{n+1}, \ldots, u_{n+k-1}$, i.e.,*

$$U = (u_n \,|\, u_{n+1} \,|\, \cdots \,|\, u_{n+k-1}). \tag{2.17}$$

Then

(a) *for MPE $s_{n,k}$ always exists uniquely when $k = k_0$ (as mentioned following* (2.15)), *and it exists uniquely when $k < k_0$ if U^* $(I-A)U$ is nonsingular.*

(b) *for RRE $s_{n,k}$ always exists uniquely when $k \leqslant k_0$.*

(c) *for MMPE $s_{n,k}$ exists uniquely when $k \leqslant k_0$ if $T^*(I-A)U$ is nonsingular, where*

$$T = (q_0 \,|\, q_1 \,|\, \cdots \,|\, q_{k-1}). \tag{2.18}$$

(d) *for TEA $s_{n,k}$ exists uniquely when $k \leqslant k_0$ if $\hat{T}^*(I-A)U$ is nonsingular, where*

$$\hat{T} = (q \,|\, A^*q \,|\, \cdots \,|\, A^{*k-1}q). \tag{2.19}$$

Proof. Without loss of generality we take $n = 0$. By appropriate column transformations we can show that

$$D(1, 1, \ldots, 1) = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ (u_0, u_0) & (u_0, w_0) & \cdots & (u_0, w_{k-1}) \\ \vdots & \vdots & & \vdots \\ (u_{k-1}, u_0) & (u_{k-1}, w_0) & \cdots & (u_{k-1}, w_{k-1}) \end{vmatrix} \quad \text{for MPE.} \tag{2.20}$$

By the fact that

$$w_j = (A-I)u_j, \quad j = 0, 1, \ldots, \tag{2.21}$$

which follows from (2.1) and (2.12), (2.20) becomes

$$D(1, 1, \ldots, 1) = \det\{U^*(A-I)U\} \quad \text{for MPE.} \tag{2.22a}$$

By similar means it can be shown that

$$D(1,1,\ldots,1)=\det\{[(A-I)U]^*[(A-I)U]\} \quad \text{for RRE,} \tag{2.22b}$$

$$D(1,1,\ldots,1)=\det\{T^*(A-I)U\} \quad \text{for MMPE,} \tag{2.22c}$$

and

$$D(1,1,\ldots,1)=\det\{\hat{T}^*(A-I)U\} \quad \text{for TEA.} \tag{2.22d}$$

Now, in order for $D(1,1,\ldots,1)\neq 0$ to hold the $k\times k$ matrices on the right hand sides of (2.22a)–(2.22d) need to have full rank. For this it is necessary that U as well as T and $\hat{T}$ be of rank k. We now show that a necessary and sufficient condition for U to have rank k is that $k\leqslant k_0$. For if $k\leqslant k_0$ and the rank of U is less than k, then the vectors $u_0, u_1,\ldots,u_{k-1}$ are linearly dependent. Thus, there exist scalars d_i, $i=0,\ldots,k-1$, not all zero, for which $\sum_{i=0}^{k-1}d_iu_i = 0$. By (2.12) this is equivalent to $(\sum_{i=0}^{k-1}d_iA^i)u_0=0$. By the assumption that k_0 is the degree of the minimal polynomial of A with respect to u_0, this implies that $k-1\geqslant k_0$, which contradicts $k\leqslant k_0$. By a similar argument we can show that if the rank of U is k, then $k\leqslant k_0$.

All this is sufficient for proving (a), (c), and (d).

Finally, in order to prove (b), we note that the $N\times N$ matrix $I-A$ is nonsingular. Thus the matrix $\tilde{U}=(A-I)U$ has rank k. Consequently, so does $\tilde{U}^*\tilde{U}$. □

Theorem 2.1 suggests that when $k<k_0$ a unique solution $s_{n,k}$ is guaranteed for RRE, but may not exist for MPE. This is surprising since both methods have very similar performance when applied to the same sequence. However, Theorem 2.1 is optimal in the sense that not always does $s_{n,k}$ exist for MPE for $k<k_0$. The following simple example demonstrates this.

Example. Let A be a hermitian $N\times N$ matrix with real eigenvalues λ_i and corresponding orthogonal eigenvectors v_i, $i=1,\ldots,N$. Normalize the v_i such that $(v_i, v_j)=\delta_{ij}$, where δ_{ij} is the Kronecker delta. Assume that $\lambda_1\neq\lambda_2$, $\lambda_1\neq 0$, $\lambda_2\neq 0$, and consider an initial vector x_0 for which $u_0=v_1+v_2$. This implies that $k_0=2$. Let us now investigate the determinant $D(1,1)$ for MPE when $k=1$ $(<k_0)$. We have

$$D(1,1)=\begin{vmatrix}1 & 1\\ (u_0,u_0) & (u_0,u_1)\end{vmatrix}=(u_0,w_0).$$

Since $w_0=(A-I)u_0=(\lambda_1-1)v_1+(\lambda_2-1)v_2$, we see that $D(1,1)=\lambda_1+\lambda_2-2$. Now it is possible for λ_1 and λ_2 to be such that $I-A$ is nonsingular and $\lambda_1+\lambda_2-2=0$, in which case $D(1,1)=0$, hence $s_{0,1}$ does not exist uniquely for MPE.

The following theorem, however, gives a sufficient condition for existence and uniqueness of $s_{n,k}$ for MPE when $k<k_0$.

Theorem 2.2. *If the matrix $C=I-A$ has positive definite hermitian part, then $s_{n,k}$ for MPE exists and is unique for $k<k_0$.*

Proof. It is enough to show that for any nonzero vector $\xi\in\mathbb{C}^k$ $F(\xi)=\xi^*U^*CU\xi\neq 0$. In the proof of Theorem 2.1 we showed that the columns of U are linearly independent if $k\leqslant k_0$. Consequently $\eta=U\xi\neq 0$. Let us write $C=C_\text{h}+C_\text{a}$, where $C_\text{h}=\frac{1}{2}(C+C^*)$ and $C_a=\frac{1}{2}(C-C^*)$ are the hermitian and antihermitian parts of C. Then $F(\xi)$ can be reexpressed as $F(\xi)=\alpha+\text{i}\beta$, where $\alpha = \eta^* C_h \eta$ and $i\beta = \eta^* C_a \eta$ and α and β are real. Now since C_h is positive definite and $\eta \neq 0$, $\alpha > 0$. Thus $|F(\xi)| = (\alpha^2 + \beta^2)^{1/2} \geqslant \alpha > 0$. □

2.3. Equivalence of vector extrapolation and Krylov subspace methods for linear systems

We now state the first main result of this section. For this we go back to our general inner product space setting. In regard to (2.26d) below we recall that the concept of the adjoint D^* of a linear operator D mapping $\boldsymbol{B}$ to itself makes sense only with respect to the inner product associated with $\boldsymbol{B}$. Specifically, D^* is that operator satisfying $(D^*x, y) = (x, Dy)$ for all $x, y \in \boldsymbol{B}$.

Theorem 2.3. *Let A, b, and x_i be as above and let $s_{n,k}$ be well defined for all four methods of extrapolation. Then these methods are Krylov subspace methods. If we let*

$$W_j = \operatorname{span}\{u_n, u_{n+1}, \ldots, u_{n+j}\} = \operatorname{span}\{u_n, Au_n, \ldots, A^j u_n\}, \tag{2.23}$$

then

$$r(s_{n,k}) = \sum_{i=0}^{k} \gamma_i^{(n,k)} u_{n+i} \in W_k. \tag{2.24}$$

Furthermore,

$$(t, r(s_{n,k})) = 0 \quad \textit{for all } t \in V_{k-1}, \tag{2.25}$$

where V_j are subspaces defined by

$$V_j = W_j \quad \textit{for MPE}, \tag{2.26a}$$

$$V_j = \operatorname{span}\{w_n, w_{n+1}, \ldots, w_{n+j}\} = \operatorname{span}\{w_n, Aw_n, \ldots, A^j w_n\} \quad \textit{for RRE}, \tag{2.26b}$$

$$V_j = \operatorname{span}\{q_0, q_1, \ldots, q_j\} = \operatorname{span}\{q_0, Gq_0, \ldots, G^j q_0\} \quad \textit{for MMPE}, \tag{2.26c}$$

for some fixed matrix G, and

$$V_j = \operatorname{span}\{q, A^*q, \ldots, A^{*j}q\} \quad \textit{for TEA}. \tag{2.26d}$$

Proof. (2.24) is seen to hold by (2.3), (2.10)–(2.12) and (2.23). That (2.25) holds for MPE, RRE, and MMPE is seen from (2.24), (2.4), and (2.5a)–(2.5c). The second equality in (2.26c) follows from the fact that it is possible to construct a linear operator G for which $q_{j+1} = Gq_j$, $j = 0, \ldots, k-1$, for $k \leqslant N-1$. That (2.25) holds for TEA with (2.26d) follows from (2.4) once we rewrite (2.5d) in the form $u_{i,j} = (A^{*i}q, u_{n+j})$, which in turn follows from (2.12). This completes the proof. □

From Theorem 2.3, we see that MPE is an orthogonal projection method, while the remaining three methods are oblique projection methods, in general. We also observe from the details of the proof of Theorem 2.3 that MPE, RRE, and MMPE are projection methods (but not necessarily Krylov subspace methods), even when the sequence $x_0, x_1, x_2, \ldots,$ is generated nonlinearly. That MPE, RRE, and TEA are Krylov subspace methods has also bee shown in [3].

When we let $q_i = A^{*i}q$, $i = 0, 1, \ldots,$ in MMPE, with q as in TEA, V_j for MMPE becomes identical to V_j for TEA. Thus MMPE reduces to TEA for this case, in the sense that $s_{n,k}$ for

MMPE is the same as $s_{n,k}$ for TEA. In case A is hermitian and $q = u_n$, we see that V_j for TEA is identical to V_j for MPE, thus $s_{n,k}$ for TEA is the same as $s_{n,k}$ for MPE. In both of these cases we should remember that the $s_{n,k}$ in question are being obtained from the $2k+1$ vectors $x_n, x_{n+1}, \ldots, x_{n+2k}$ for TEA, and from the $k+2$ vectors $x_n, x_{n+1}, \ldots, x_{n+k+1}$ for MMPE and MPE.

We also note that the inner product $(\cdot,\cdot)$ can be replaced by any other inner product $(\cdot,\cdot)_M$, where $(x, y)_M = (x, My)$, M being a hermitian positive definite operator. This becomes useful when dealing with GCG. The remark prior to the statement of Theorem 2.3 concerning the adjoint of an operator should be kept in mind, however.

We now state the second main result of this section showing the connection between extrapolation methods and some Krylov subspace methods that were mentioned in the introduction to this work. (For a discussion of Krylov subspace methods and also the method of Lanczos see Saad [22].) For simplicity we shall set $n = 0$ and denote $s_{0,k} \equiv s_k$. Also we shall define $I - A \equiv C$. Thus s is the solution to $Cx = b$. (Conversely, given C we define A through $A \equiv I - C$.) If we let $\tilde{r}(x)$ be the residual for the vector x, i.e., $\tilde{r}(x) = b - Cx$, then by $I - A = C$ we have $\tilde{r}(x) = r(x)$. We shall denote by $z_0, z_1, z_2, \ldots,$ the sequence of approximations to s obtained by applying the Arnoldi method or GCR or the Lanczos method to the linear operator equation $Cx = b$ starting with $z_0 = x_0$.

Theorem 2.4. *With s_k and z_k as above,*

$$s_k = z_k, \quad k = 0, 1, \ldots, \tag{2.27}$$

(a) *for MPE and the Arnoldi method, or*

(b) *for RRE and GCR, or*

(c) *for TEA and the Lanczos method.*

Remark. With $V_k = \operatorname{span}\{q_0, Gq_0, \ldots, G^k q_0\}$ for MMPE with some linear operator G, this method seems to be equivalent to a generalized version of the Lanczos method.

Proof. For the Arnoldi method, GCR, and the Lanczos method z_i for the system $Cx = b$ are of the form

$$z_0 = x_0, \qquad z_k = x_0 + \sum_{i=0}^{k-1} \delta_i C^i r(x_0), \quad k = 0, 1, \ldots . \tag{2.28}$$

Thus, by the facts that $r(x) = b - Cx$ and $C = I - A$ it follows that

$$r(z_k) = r(x_0) - \sum_{i=0}^{k-1} \delta_i C^{i+1} r(x_0) = u_0 - \sum_{i=0}^{k-1} \delta_i C^{i+1} u_0 \in W_k, \tag{2.29}$$

with W_k as defined in (2.23) with $n = 0$. At this point recall that $r(s_k) \in W_k$ for all methods of extrapolation considered in this work.

For the Arnoldi method the δ_i are determined by the requirement that $r(z_k)$ be orthogonal to $r(z_0), \ldots, r(z_{k-1})$, i.e., $(t, r(z_k)) = 0$, for every $t \in W_{k-1}$. This proves (a).

For GCR the δ_i are determined by the requirement that $(Cr(z_i), r(z_k)) = 0$, $i = 0, \ldots, k-1$, i.e., $(Ct, r(z_k)) = 0$ for every $t \in W_{k-1}$. But the subspace $\{Ct \mid t \in W_j\}$ is the same as V_j in (2.26b). This proves (b).

For the Lanczos method the δ_i are determined by the requirement that $(t, r(z_k)) = 0$ for every $t \in V_{k-1}$ with V_j as given in (2.26d). This proves (c). □

All three projection methods above, namely, the Arnoldi method, GCR, and the Lanczos method, were devised to solve the system of linear equations $Cx = b$ for an arbitrary nonsingular matrix C. Note that the conjugate gradient type method of Axelsson [2], the method of Young and Jea [31] that has been designated ORTHODIR, and the recent generalized minimum residual method (GMRES) of Saad and Schultz [23], for solving $Cx = b$ with an arbitrary nonsingular matrix C, are all mathematically equivalent to GCR. Similarly, CG and CR were devised for solving the same system for a hermitian matrix C.

Let us now consider the case in which C is hermitian. From the theory of CG it follows immediately that if the Arnoldi method and CG are implemented beginning with the same vector $z_0 = x_0$, then they are equivalent. But in this case $A = I - C$ is hermitian too. Therefore, as mentioned following Theorem 2.3, if we pick $q = u_0$ for TEA, then $s_{0,k}$ for TEA is identical to $s_{0,k}$ for MPE. Combining both of these observations with parts (a) and (c) of Theorem 2.4, we conclude that, in the case under consideration and with the assumptions above, MPE, TEA, the Arnoldi method, the Lanczos method, and CG are equivalent. It is interesting to note that when C is hermitian and positive definite, s_k (or equivalently z_k), as obtained by these methods, are such that

$$E(z_k) = \min_{\Delta \in W_{k-1}} E(x_0 + \Delta), \tag{2.30}$$

where

$$E(z) = ((z - s), C(z - s)) \tag{2.31}$$

is a positive definite quadratic form.

When the matrix A in $C = I - A$ is antihermitian (a complex version of) GCG can be used to implement the method of Arnoldi recursively. Actually, GCG is designed to solve a system of linear equations $\hat{C}x = d$ for an arbitrary matrix $\hat{C}$. If we let $\hat{C}_\text{h}$ and $\hat{C}_\text{a}$ be respectively the hermitian and antihermitian parts of $\hat{C}$, and assume that $\hat{C}_\text{h}$ is positive definite, then GCG is actually a Krylov subspace method equivalent to MPE, for which the vectors x_j are generated by $x_{j+1} = Ax_j + b$, with $A = -\hat{C}_\text{h}^{-1}\hat{C}_\text{a}$ and $b = \hat{C}_\text{h}^{-1}d$, and the inner product $(\cdot,\cdot)$ is replaced by the inner product $(\cdot,\cdot)_{\hat{C}_\text{h}}$, where $(x, y)_{\hat{C}_\text{h}} = (x, \hat{C}_\text{h}y)$. With respect to the new inner product the operator A is antihermitian, i.e., $(x, Ay)_{\hat{C}_\text{h}} = -(Ax, y)_{\hat{C}_\text{h}}$. Since $\hat{C}x = d$ is also equivalent to $Cx = b$, with $C = \hat{C}_\text{h}^{-1}\hat{C} = I - A$, we are back at the case discussed above, only with a different inner product.

Again when C is hermitian from the theory of CR it follows immediately that if GCR and CR are implemented beginning with the same vector $z_0 = x_0$, then they are equivalent. Combining this with part (b) of Theorem 2.4, we conclude that in this case RRE, GCR, and CR are equivalent. For arbitrary nonsingular (not necessarily hermitian positive definite) C, s_k (or equivalently z_k), as obtained by these methods, are such that

$$F(z_k) = \min_{\Delta \in W_{k-1}} F(x_0 + \Delta), \tag{2.32}$$

where

$$F(z) = (r(z), r(z)) = \| r(z) \|^2 \tag{2.33}$$

is a positive definite quadratic form.

We wish to mention one more method that has been proposed for solving $Cx = b$ with arbitrary nonsingular C, namely the biconjugate gradient method of Fletcher [12]. This method is equivalent to the Lanczos method. Consequently, by Theorem 2.4, the biconjugate gradient method is equivalent to TEA too. The connection of TEA with CG and the biconjugate gradient method has also been studied in Brezinski [5, pp. 186–189].

Finally, we recall that when $\boldsymbol{B}$ is the Euclidean space $\mathbb{C}^N$ and C is a (complex) $N \times N$ matrix all of the projection methods mentioned above terminate in a finite number of steps. Actually $z_k = s$ for some $k \leqslant N$. From the discussion prior to Theorem 2.1 and from the equivalence of the vector extrapolation methods and the various projection methods as they are applied to linearly generated sequences beginning with the same initial vector x_0, it is now obvious that for all Krylov subspace methods mentioned above $z_k = s$ for $k = k_0$, where k_0 is the degree of the minimal polynomial of C (or $A = I - C$) with respect to the vector $x_0 - s$ (or u_0).

3. A unified treatment for CG, GCG, and CR

In this section we propose a unified recursive algorithm from which CG, GCG, and CR can be obtained as special cases. We shall deal with the linear operator equation $Cx = b$, whose solution we denote by s. In case C is a matrix, we shall allow it to be complex. Furthermore, we shall assume that C satisfies

$$C^* = \sigma I + \tau C, \quad \sigma,\ \tau \text{ scalars.} \tag{3.1}$$

Let M be an operator that commutes with C, i.e., $MC = CM$, and define

$$(x,\ y)_M = (x,\ My). \tag{3.2}$$

Of course, when M is hermitian positive definite $(x,\ y)_M$ is a true inner product; otherwise it is not.

The unified algorithm now reads as follows:

$$\left\{\begin{array}{l}\text{Pick } x_0 \text{ arbitrarily, set } z_0 = x_0,\ r_0 = b - Cz_0,\ p_0 = r_0.\\ \text{Do for } j = 0, 1, \ldots,\\ z_{j+1} = z_j + \alpha_j p_j, \text{ where } \alpha_j = (p_j,\ r_j)_M/(p_j,\ Cp_j)_M;\\ r_{j+1} = b - Cz_{j+1}; \text{ if } r_{j+1} = 0, \text{ then set } s = z_{j+1} \text{ and stop, else}\\ p_{j+1} = r_{j+1} + \beta_j p_j, \text{ where } \beta_j = -(p_j,\ Cr_{j+1})_M/(p_j,\ Cp_j)_M.\end{array}\right. \tag{3.3}$$

Needless to say, we assume that $(p_j,\ Cp_j)_M \neq 0$ and $\alpha_j \neq 0$ when $r_j \neq 0$. Under these circumstances the algorithm does not break down.

Theorem 3.1. *Provided each step in the algorithm above is well defined, we have*

$$(p_i,\ r_k)_M = 0, \quad i \leqslant k-1, \tag{3.4}$$

$$(p_i,\ Cp_k)_M = 0, \quad i \leqslant k-1, \tag{3.5}$$

$$\operatorname{span}\{p_0, \ldots, p_k\} = \operatorname{span}\{r_0, \ldots, r_k\} = \operatorname{span}\{r_0,\ Cr_0, \ldots, C^k r_0\}, \tag{3.6}$$

$$\alpha_k = (r_k,\ r_k)_M/(p_k,\ Cp_k)_M, \tag{3.7}$$

and

$$\beta_k = \bar{\tau}\, \mathrm{e}^{-2\mathrm{i}\theta_k}(r_{k+1},\ r_{k+1})_M/\overline{(r_k,\ r_k)_M}, \qquad \theta_k = \arg(p_k,\ Cp_k)_M. \tag{3.8}$$

Proof. Following Luenberger [18, Ch. 8], we shall prove (3.4)–(3.6) by induction on k.

For $k=1$ these assertions are true as can be verified directly. Suppose they are true for all $k \leqslant j$. We need to show that they are true for $k=j+1$ too.

First, we note that by $z_{j+1}=z_j+\alpha_j p_j$

$$r_{j+1}=r_j-\alpha_j C p_j. \tag{3.9}$$

Thus $(p_j, r_{j+1})_M=0$ is satisfied by the definition of α_j. Similarly, $(p_j, Cp_{j+1})_M=0$ is satisfied by $p_{j+1}=r_{j+1}+\beta_j p_j$ and the definition of β_j. Now for any $i \leqslant j-1$

$$(p_i, r_{j+1})_M=(p_i, r_j)_M-\alpha_j(p_i, Cp_j)_M \tag{3.10}$$

and

$$(p_i, Cp_{j+1})_M=(p_i, Cr_{j+1})_M+\beta_j(p_i, Cp_j)_M. \tag{3.11}$$

By applying the induction hypothesis to each of the terms on the right hand side of (3.10), $(p_i, r_{j+1})_M=0$ for $i \leqslant j-1$ follows. Again by the induction hypothesis $(p_i, Cp_j)_M=0$ for $i \leqslant j-1$, so that (3.11) becomes

$$(p_i, Cp_{j+1})_M=(p_i, Cr_{j+1})_M. \tag{3.12}$$

Since $CM=MC$, (3.12) can be expressed as

$$(p_i, Cp_{j+1})_M=(C^*p_i, r_{j+1})_M. \tag{3.13}$$

Invoking (3.1), (3.13) becomes

$$(p_i, Cp_{j+1})_M=([\sigma I+\tau C]p_i, r_{j+1})_M. \tag{3.14}$$

Since $(p_i, r_{j+1})_M=0$ for $i \leqslant j$ has already been shown, (3.14) now becomes

$$(p_i, Cp_{j+1})_M=\bar{\tau}(Cp_i, r_{j+1})_M. \tag{3.15}$$

Now since $Cp_i=(r_i-r_{i+1})/\alpha_i$ by (3.9), $Cp_i \in \operatorname{span}\{p_0, p_1, \ldots, p_j\}$ by the induction hypothesis and (3.6). This, combined with $(p_i, r_{j+1})_M=0$ for $i \leqslant j$, which has already been proved, results in $(p_i, Cp_{j+1})_M=0$ for $i \leqslant j-1$.

The relations $p_{j+1}=r_{j+1}+\beta_j p_j$ and (3.9), the induction hypothesis, and the assumption that $\alpha_j \neq 0$, together with the fact that the set $\{r_0, Cr_0, \ldots, C^k r_0\}$ is linearly independent provided $k \leqslant k_0-1$, can be used to prove (3.6) for $k=j+1$.

The proof of (3.7) can be achieved by substituting $p_j=r_j+\beta_{j-1}p_{j-1}$ in the expression for α_j in (3.3), and invoking (3.4).

For the proof of (3.8) we proceed as follows: By $MC=CM$ we have $(p_j, Cr_{j+1})_M=(C^*p_j, r_{j+1})_M$. Invoking now (3.1), and using (3.4), this becomes $(p_j, Cr_{j+1})_M=\bar{\tau}(Cp_j, r_{j+1})_M$. Substituting now $Cp_j=(r_j-r_{j+1})/\alpha_j$, and using $(r_j, r_{j+1})_M=0$, which follows from (3.4) and (3.6), we obtain $(p_j, Cr_{j+1})_M=-(\bar{\tau}/\bar{\alpha}_j)(r_{j+1}, r_{j+1})_M$. (3.8) follows by using this and (3.7) in the expression for β_j given in (3.3). □

As can be seen from Theorem 3.1, a sufficient condition for the algorithm not to break down is that MC and M have positive definite hermitian parts.

Note: If C is not a constant multiple of I and satisfies (3.1), then $|\tau|=1$ and $\sigma+\bar{\sigma}\tau=0$. It can be shown that this is possible if and only if $C=\lambda I+D$, where λ is some appropriate scalar, and

μD, for some $\mu \neq 0$, is hermitian. In fact, it turns out that $\mu = \tau^{1/2}$. Thus the algorithm in (3.6) can be specialized to produce all the known recursive algorithms, extending them to complex matrices at the same time:

(1) When $\sigma = 0$ and $\tau = 1$, $C^* = C$. Letting $M = I$, we obtain CG. Letting $M = C$, we obtain CR.

(2) When $\sigma = 2$ and $\tau = -1$, $C = I + D$ with $D^* = -D$. Letting $M =$ I, we obtain (another form of) GCG. The method obtained by letting $M = C^*$ is equivalent to ORTHOMIN(1) by Vinsome [29]. When $\boldsymbol{B}$ is a real space, the new form of GCG (with $M =$ I) can be simplified considerably by noting that $(x, Dx)_M = 0$ hence $(x, Cx)_M = (x, x)$ for all $x \in \boldsymbol{B}$. This results in $\alpha_j = (r_j, r_j)/(p_j, p_j)$ from (3.7). (3.8), on the other hand, can be simplified to read $\beta_j = (r_{j+1}, r_{j+1})/(r_j, r_j)$. Other simplifications can be made for the case $M = C^*$. We omit the details.

Finally, we note that when C and M are hermitian positive definite, z_k in the algorithm given in (3.3) satisfies

$$\tilde{F}(z_k) = \min_{\Delta \in Y} \tilde{F}(x_0 + \Delta), \tag{3.16}$$

where $\tilde{F}(x)$ is the positive definite quadratic form

$$\tilde{F}(x) = ((x - s), C(x - s))_M \tag{3.17}$$

and

$$Y = \text{span}\{r_0, Cr_0, \ldots, C^{k-1}r_0\}. \tag{3.18}$$

4. Error analysis for MPE and RRE

In this section we wish to give error bounds related to $e_k = s_k - s$ for MPE and RRE, where $s_k \equiv s_{0,k}$. We recall that when $\boldsymbol{B}$ is the Euclidean space $\mathbb{C}^N$ and C is an $N \times N$ (complex) matrix s_{k_0} exists and $s_{k_0} = s$ for both methods. When $k < k_0$, s_k exists and is unique for RRE always. For MPE, however, s_k for $k < k_0$ exists and is unique if C has a positive definite hermitian part. Our analysis is similar to and generalizes that of GCG that was given in [30], in conjunction with that of [19], to cover all the methods mentioned above. (For further developments concerning GCG and its convergence properties see Hageman, Luk, and Young [14] and Eisenstat [10].) Furthermore, it also reproduces the results known for CG, GCG, and CR. We shall state most of our results within the framework of the general vector space $\boldsymbol{B}$ with its inner product $(\cdot,\cdot)$ and we shall treat A and C as bounded linear operators on $\boldsymbol{B}$.

Let M be a bounded linear operator on $\boldsymbol{B}$, and define

$$(x, y)_M = (x, My). \tag{4.1}$$

Here M is not necessarily hermitian positive definite, consequently $(\cdot,\cdot)_M$ is not necessarily an inner product. We set

$$M = \begin{cases} C^* & \text{for RRE,} \\ I & \text{for MPE.} \end{cases} \tag{4.2}$$

Lemma 4.1. *The errors $e_k = s_k - s$ for MPE and RRE satisfy*

$$(e_k, Ce_k)_M = -(Q_k(C)(x_0 - s), Ce_k)_M, \tag{4.3}$$

where $Q_k(\lambda)$ is an arbitrary polynomial of degree $\leqslant k$ in λ, normalized such that $Q_k(0) = 1$.

Proof. By (2.2), (2.3), and (2.12)

$$e_k = \sum_{i=0}^{k} \gamma_i^{(n,k)}(x_i - s) = \left(\sum_{i=0}^{k} \gamma_i^{(n,k)} A^i \right)(x_0 - s). \tag{4.4}$$

By (4.4) and (2.10)

$$Ce_k = Cs_k - Cs = -(b - Cs_k) = -r(s_k). \tag{4.5}$$

By Theorem 2.3 $(t, r(s_k)) = 0$ for all $t \in W_{k-1}$ for MPE, and $(Ct, r(s_k)) = 0$ for all $t \in W_{k-1}$ for RRE. This can be expressed in a unified way as

$$(t, r(s_k))_M = (t, Ce_k)_M = 0. \tag{4.6}$$

Thus

$$(e_k, Ce_k)_M = -(e_k, r(s_k))_M = -(e_k + t, r(s_k))_M \quad \text{for all } t \in W_{k-1}. \tag{4.7}$$

Now if $t \in W_{k-1}$, then it can be expressed in the form

$$t = \sum_{i=0}^{k-1} \delta_i u_i = \left(\sum_{i=0}^{k-1} \delta_i A^i \right) u_0, \tag{4.8}$$

which, by $u_0 = -C(x_0 - s)$, becomes

$$t = -\left(\sum_{i=0}^{k-1} \delta_i A^i \right) C(x_0 - s). \tag{4.9}$$

By (4.4), (4.9), and $C = I - A$

$$\begin{aligned} e_k + t &= \left\{ \sum_{i=0}^{k} \gamma_i^{(n,k)} (I - C)^i - C \sum_{i=0}^{k-1} \delta_i (I - C)^i \right\} (x_0 - s) \\ &= Q_k(C)(x_0 - s). \end{aligned} \tag{4.10}$$

Obviously, since $\delta_0, \ldots, \delta_{k-1}$ are arbitrary, $Q_k(\lambda)$ is an arbitrary polynomial of degree $\leqslant k$ satisfying

$$Q_k(0) = \sum_{i=0}^{k} \gamma_i^{(n,k)} = 1. \tag{4.11}$$

Substituting (4.5) and (4.10) in (4.6), the result follows. □

From Lemma 4.1 we can now obtain an error bound for $\| r(s_k) \|$ for RRE. For any linear operator D on $\boldsymbol{B}$ denote by $\| D \|$ the operator norm induced by the vector norm $\| x \|$.

Theorem 4.2. *For RRE*

$$\| r(s_k) \| \leqslant \| Q_k(C) \| \, \| r(s_0) \|. \tag{4.12}$$

Proof. Letting $M = C^*$ in (4.3), we obtain

$$(Ce_k, Ce_k) = (Q_k(C)C(x_0 - s), Ce_k). \tag{4.13}$$

(4.12) follows from (4.13) by applying the Cauchy-Schwarz inequality on the right hand side, using $Ce_k = -r(s_k)$ and $C(x_0 - s) = -r(s_0)$, and cancelling a $\| r(s_k) \|$ factor from both sides. □

Note that since C is nonsingular and $r(s_k) = -Ce_k$, $\| r(s_k) \|$ is a true norm for e_k. The result in (4.12) is identical to that obtained for GCR in [11].

We now turn our attention to the analysis of MPE. As before, let us denote $C_\mathrm{h} = \frac{1}{2}(C + C^*)$ and $C_\mathrm{a} = \frac{1}{2}(C - C^*)$, and assume in the sequel that C_h is positive definite. Then

$$(x, y)' = (x, C_\mathrm{h} y) \tag{4.14}$$

is a true inner product in $\boldsymbol{B}$. Consequently,

$$\| x \|' = \sqrt{(x, x)'} = \| C_\mathrm{h}^{1/2} x \| \tag{4.15}$$

is a true norm in $\boldsymbol{B}$, and it satisfies

$$\| x \|' \leqslant | (x, Cx) |^{1/2}. \tag{4.16}$$

The proof of (4.16) can be accomplished by the technique used in the proof of Theorem 2.2. For any linear operator D on $\boldsymbol{B}$, let us denote by $\| D \|$ and $\| D \|'$ the operator norms induced by the vector norms $\| x \|$ and $\| x \|'$ respectively.

Lemma 4.3. *In general*

$$\| D \|' = \| C_\mathrm{h}^{1/2} D C_\mathrm{h}^{-1/2} \| \leqslant \sqrt{\operatorname{cond}(C_\mathrm{h})}\, \| D \|, \tag{4.17}$$

where $\operatorname{cond}(G) = \| G \| \, \| G^{-1} \|$, *and*

$$\| D \|' = \| D \| \quad \text{if } DC_\mathrm{h} = C_\mathrm{h} D. \tag{4.18}$$

Proof. The first part of (4.17) follows from

$$\| D \|' = \sup_{x \neq 0} \frac{(Dx, C_\mathrm{h} Dx)^{1/2}}{(x, C_\mathrm{h} x)^{1/2}} = \sup_{x \neq 0} \frac{(\tilde{D} C_\mathrm{h}^{1/2} x, \tilde{D} C_\mathrm{h}^{1/2} x)^{1/2}}{(C_\mathrm{h}^{1/2} x, C_\mathrm{h}^{1/2} x)^{1/2}}, \tag{4.19}$$

where $\tilde{D} = C_\mathrm{h}^{1/2} D C_\mathrm{h}^{-1/2}$. The second part of (4.17) follows from the fact that $\| C_\mathrm{h}^\alpha \| = \rho(C_\mathrm{h}^\alpha)$ for any real α. Finally, (4.18) is a corollary of (4.17). □

Theorem 4.4. *The error e_k for MPE satisfies*

$$\begin{aligned} \| e_k \|' \leqslant \| C_\mathrm{h}^{-1} C^* Q_k(C)(x_0 - s) \|' &\leqslant L \| Q_k(C)(x_0 - s) \|' \\ &\leqslant L \| Q_k(C) \|' \, \| e_0 \|', \end{aligned} \tag{4.20}$$

where $e_0 = s_0 - s = x_0 - s$ *and*

$$L = \| C_\mathrm{h}^{-1} C^* \|' = \| C_\mathrm{h}^{-1} C \|' = \sqrt{1 + \Lambda^2}, \tag{4.21}$$

with $\Lambda = \rho(C_h^{-1} C_a)$, *the spectral radius of the operator* $C_h^{-1} C_a$.

Proof. Invoking (4.14)–(4.16) in (4.3), we have

$$(\| e_k \|')^2 \leqslant | (C_\mathrm{h}^{-1} C^* Q_k(C)(x_0 - s), e_k)' |, \tag{4.22}$$

which, upon using the Cauchy-Schwarz inequality, becomes

$$\left(\| e_k \|' \right)^2 \leqslant \| C_{\mathrm{h}}^{-1} C^* Q_k(C)(x_0 - s) \|' \, \| e_k \|'. \tag{4.23}$$

This proves (4.20). To prove (4.21), we note that from Lemma 4.3

$$\| C_{\mathrm{h}}^{-1} C \|' = \| C_{\mathrm{h}}^{-1/2} C C_{\mathrm{h}}^{-1/2} \| = \| I + \tilde{C}_{\mathrm{a}} \|, \qquad \tilde{C}_{\mathrm{a}} = C_{\mathrm{h}}^{-1/2} C_{\mathrm{a}} C_{\mathrm{h}}^{-1/2}. \tag{4.24}$$

Now since C_{a} is antihermitian, $\tilde{C}_{\mathrm{a}}$ is antihermitian too, consequently $I + \tilde{C}_{\mathrm{a}}$ is normal. Therefore,

$$\| I + \tilde{C}_{\mathrm{a}} \| = \rho\left(I + \tilde{C}_{\mathrm{a}} \right). \tag{4.25}$$

By $\tilde{C}_{\mathrm{a}} = C_{\mathrm{h}}^{1/2}(C_h^{-1} C_{\mathrm{a}}) C_{\mathrm{h}}^{-1/2}$, $\tilde{C}_{\mathrm{a}}$ and $C_{\mathrm{h}}^{-1} C_{\mathrm{a}}$ have the same spectrum. Since $\tilde{C}_{\mathrm{a}}$ is antihermitian, all of its eigenvalues are zero or pure imaginary, and so are those of $C_{\mathrm{h}}^{-1} C_{\mathrm{a}}$. Combining all this in (4.24), we obtain $\| C_{\mathrm{h}}^{-1} C \| = \sqrt{1 + \Lambda^2}$. That $\| C_{\mathrm{h}}^{-1} C^* \| = \sqrt{1 + \Lambda^2}$ follows by exactly the same argument. This proves (4.21). □

We now aim at obtaining bounds on $\| e_k \|'$ that involve $\| Q_k(C) \|$ and not $\| Q_k(C) \|'$. This enables us to unify the treatments for CR, CG, GCG, and MPE and RRE in general. To this effect we have the following result:

Theorem 4.5. *In general, for MPE,*

$$\| e_k \|' \leqslant L \sqrt{\operatorname{cond}(C_{\mathrm{h}})} \, \| Q_k(C) \| \, \| e_0 \|', \tag{4.26}$$

and if C is a normal operator, i.e., $CC^* = C^*C$,

$$\| e_k \|' \leqslant L \| Q_k(C) \| \, \| e_0 \|'. \tag{4.27}$$

Proof. (4.26) follows by invoking (4.17) in (4.20). Now if C is normal, then C_{h} and C_{a} commute. Consequently C_{h} commutes with C. This, combined with (4.18), implies $\| Q_k(C) \|' = \| Q_k(C) \|$, from which (4.27) follows. □

Theorem 4.2 for RRE and Theorem 4.5 for MPE provide us with upper bounds on different norms of $e_k(\| r(s_k) \| = \| Ce_k \|$ for RRE and $\| e_k \|' = \| C_{\mathrm{h}}^{1/2} e_k \|$ for MPE), and these upper bounds have all been expressed in terms of $\| Q_k(C) \|$, where $Q_k(\lambda)$ is an arbitrary polynomial of degree at most k, normalized such that $Q_k(0) = 1$. Therefore, in (4.12), (4.26), and (4.27) $\| Q_k(C) \|$ can be replaced by $\Gamma_k \equiv \min_{Q_k \in \pi_k} \| Q_k(C) \|$, where we have denoted by π_k the set of all polynomials $Q_k(\lambda)$ of degree at most k satisfying $\mathrm{Q}_k(0) = 1$.

We now try to bound Γ_k in terms of the eigenvalues λ_i, $i = 1, 2, \ldots,$of C. Let us denote by $\sigma(C)$ the spectrum of C, namely the set $\{ \lambda_1, \lambda_2, \ldots, \}$. Since C_{h} is hermitian positive definite, $\sigma(C)$ is contained in the open right half of the complex λ-plane. That this is so can be seen by observing that if (λ, v) is an eigenvalue-eigenvector pair of C, then $\operatorname{Re} \lambda = (v, C_{\mathrm{h}} v)/(v, v) > 0$. Following [19], $\sigma(C)$ is contained in an ellipse $F(d, c, a)$ of smallest size, which in turn is contained in the open right half of the λ-plane. Here d is the center of the ellipse, $d \pm c$ are its foci, and a is its semimajor axis length, such that $|c| \leqslant a$ and $|\operatorname{Re} c| < \operatorname{Re} d$. Let us now set

$$Q_k(\lambda) = T_k\left(\frac{d - \lambda}{c} \right) \Big/ T_k\left(\frac{d}{c} \right) \equiv \tilde{T}_k(\lambda). \tag{4.28}$$

Obviously $\tilde{T}_k \in \pi_k$ and hence it is true that

$$\Gamma_k \leqslant \| \tilde{T}_k(C) \|. \tag{4.29}$$

Let us define η_k by

$$\rho(\tilde{T}_k(C)) = \max_{\lambda_i \in \sigma(C)} |\tilde{T}_k(\lambda_i)| \leqslant \max_{\lambda \in F(d,\,c,\,a)} |\tilde{T}_k(\lambda)| = \eta_k. \tag{4.30}$$

From known properties of Chebyshev polynomials it follows that

$$\eta_k = \frac{\cosh(k\phi)}{|\cosh(k\omega)|}, \quad \phi = \cosh^{-1}\left(\frac{a}{|c|}\right) \quad \text{and} \quad \omega = \cosh^{-1}\left(\frac{d}{c}\right). \tag{4.31}$$

Thus

$$\lim_{k\to\infty} \eta_k^{1/k} \leqslant \mathrm{e}^{\phi - \operatorname{Re}\omega} \equiv \eta, \quad \eta < 1. \tag{4.32}$$

We now bound $\|\tilde{T}_k(C)\|$ for different operators C in terms of η_k and η.

(1) If C is normal, $\|\tilde{T}_k(C)\| = \rho(\tilde{T}_k(C))$, thus (4.12) and (4.27) become

$$\|r(s_k)\| \leqslant \eta_k \|r(s_0)\| \leqslant \alpha\eta^k \|r(s_0)\| \quad \text{for RRE} \tag{4.33}$$

and

$$\|e_k\|' \leqslant L\eta_k \|e_0\|' \leqslant \alpha L\eta^k \|e_0\|' \quad \text{for MPE,} \tag{4.34}$$

for some $\alpha > 0$. In particular, if $C_{\mathrm{h}} = I$, then all eigenvalues of C are of the form $1 + \mathrm{i}\mu$ with $|\mu| \leqslant \Lambda$. Thus $F(d, c, a)$ degenerates to the line segment $[1 - \mathrm{i}\Lambda, 1 + \mathrm{i}\Lambda]$, i.e., $F(d, c, a) = F(1, \mathrm{i}\Lambda, \Lambda)$. Consequently, $\eta_k = |T_k(\mathrm{i}/\Lambda)|^{-1}$ in (4.33) and (4.34). The result for MPE in this case becomes, by $\|\cdot\|' = \|\cdot\|$ for this case,

$$\|e_k\| \leqslant \sqrt{1 + \Lambda^2} / |T_k(\mathrm{i}/\Lambda)| \, \|e_0\|, \tag{4.35}$$

and this result is identical to the one given in [30] for GCG. When C is hermitian positive definite $C = C_{\mathrm{h}}$ and all eigenvalues of C are real and positive. If we let $\lambda_{\min}$ and $\lambda_{\max}$ be respectively the smallest and the largest eigenvalues of C, then $F(d, c, a)$ degenerates to the line segment $[\lambda_{\min}, \lambda_{\max}]$, and its parameters become $d = \frac{1}{2}(\lambda_{\max} + \lambda_{\min})$, $c = \frac{1}{2}(\lambda_{\max} - \lambda_{\min}) = a$. Consequently $\eta_k = |T_k((\lambda_{\max} + \lambda_{\min})/(\lambda_{\max} - \lambda_{\min}))|^{-1}$ both in (4.33) and (4.34). After some manipulation, it can also be shown that $\eta_k < 2\eta^k$ for all k, with $\eta = (\sqrt{\kappa} - 1)/(\sqrt{\kappa} + 1)$, where $\kappa = \lambda_{\max}/\lambda_{\min} = \operatorname{cond}(C)$. Thus (4.33) and (4.34) become

$$\|r(s_k)\| \leqslant 2\eta^k \|r(s_0)\| \quad \text{for RRE} \tag{4.36}$$

and

$$\|e_k\|' \leqslant 2\eta^k \|e_0\|' \quad \text{for MPE,} \tag{4.37}$$

the latter being the well-known result for CG.

(2) If $\boldsymbol{B}$ is the Euclidean space $\mathbb{C}^N$, and C is an arbitrary $N \times N$ matrix, then as mentioned in [19],

$$\|\tilde{T}_k(C)\| \leqslant \alpha k^m \eta^k, \tag{4.38}$$

m being a positive integer and α a positive constant. In fact, $m + 1$ is the size of the largest Jordan block of C. (This implies that $m = 0$ if C is diagonalizable.) In this case (4.33) and (4.34) become

$$\|r(s_k)\| \leqslant \alpha k^m \eta^k \|r(s_0)\| \quad \text{for RRE} \tag{4.39}$$

and

$$\|e_k\|' \leqslant \alpha L \sqrt{\operatorname{cond}(C_h)}\, k^m \eta^k \|e_0\|' \quad \text{for MPE.} \tag{4.40}$$

In all our results above, we have obtained upper bounds for $\| r(s_k) \|$ or $\| e_k \|'$ in terms of the quantity $\Gamma_k = \min_{Q_k \in \pi_k} \| Q_k(C) \|$, which after a certain value of k are decreasing monotonically towards zero. Needless to say, these results can also be used to give convergence rates for restarted forms of the methods under consideration. The application of the Krylov subspace methods in their restarted forms is termed 'cycling' in the context of vector extrapolation methods.

Note that $\Gamma_k = \min_{Q_k \in \pi_k} \| Q_k(C) \|$ can also be bounded by using an approach suggested in [11]. In this approach we make use of

$$\Gamma_k \leqslant (\Gamma_1)^k = \left(\min_{\alpha \in \mathbb{C}} \| I + \alpha C \| \right)^k \leqslant \left(\min_{\alpha \in \mathbb{R}} \| I + \alpha C \| \right)^k, \tag{4.41}$$

where $\mathbb{R}$ denotes the field of real numbers. Now

$$\| I + \alpha C \|^2 = \max_{\| x \| = 1} \left[1 + 2 \operatorname{Re} \alpha(x, Cx) + | \alpha |^2 (x, C^*Cx) \right]. \tag{4.42}$$

When C_h is positive definite

$$\operatorname{Re}(x, Cx) = (x, C_\mathrm{h} x) \geqslant \mu \| x \|^2, \tag{4.43}$$

where μ is the smallest eigenvalue of C_h. Let ω be the largest singular value of C. Then for α real and negative

$$\| I + \alpha C \|^2 \leqslant 1 + 2\mu\alpha + \omega^2\alpha^2. \tag{4.44}$$

The minimum of $1 + 2\mu\alpha + \omega^2\alpha^2$ is obtained for $\alpha = -\mu/\omega^2 < 0$, and is $1 - \mu^2/\omega^2$. Combining this with (4.42) and (4.41), we finally obtain

$$\Gamma_k \leqslant \left(1 - \mu^2/\omega^2\right)^{k/2}, \tag{4.45}$$

which is the same as that given in [11] for a real operator C.

Finally, when $\boldsymbol{B}$ is a Hilbert space and C is a compact operator, superlinear convergence results similar to the ones given in [30] for GCG can also be proved. We omit the details.

Acknowledgement

This research was supported in part by the Fund for the Promotion of Research at the Technion. The author would like to thank the two anonymous referees for drawing his attention to some missing references and for their critical comments and suggestions that lead to the present version of the original manuscript.

Note

The author has been informed by one of the referees that the unified algorithm of (3.3) follows also from the recent work of V. Faber and T.A. Manteuffel, Orthogonal error methods, *SIAM J. Numer. Anal.* **24** (1987) 170–187.

References

[1] W.E. Arnoldi, The principle of minimized iterations in the solution of the matrix eigenvalue problem, *Quart. Appl. Math.* **9** (1951) 17–29.

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

[3] J. Beuneu, Minimal polynomial projection methods, Preprint, 1984.

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

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

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

[7] 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.

[8] P. Concus and G.H. Golub, A generalized conjugate gradient method for nonsymmetric systems of linear equations, in: R. Glowinski and J.L. Lions, Eds., *Proc. Second Internat. Symp. on Computing Methods in Applied Sciences and Engineering*, IRIA, Paris, Dec. 1975. Lecture Notes in Economics and Mathematical Systems **134** (Springer, Berlin, 1976) 56–65.

[9] 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.

[10] S.C. Eisenstat, A note on the generalized conjugate gradient method, *SIAM J. Numer. Anal.* **20** (1983) 358–361.

[11] S.C. Eisenstat, H.C. Elman and M.H. Schultz, Variational iterative methods for nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.* **20** (1983) 345–357.

[12] R. Fletcher, Conjugate gradient methods for indefinite systems, in: G.A. Watson, Ed., *Proc. Dundee Bienial Conference on Numerical Analysis* (1975), Lecture Notes in Mathematics **506** (Springer, Heidelberg, 1976) 73–89.

[13] W.F. Ford and A. Sidi, Recursive algorithms for vector extrapolation methods, *Appl. Numer. Math.* **4** (1988) to appear.

[14] L.A. Hageman, F.T. Luk, and D.M. Young, On the equivalence of certain iterative acceleration methods, *SIAM J. Numer. Anal.* **17** (1980) 852–873.

[15] M. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. N.B.S.* **49** (1952) 409–436.

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

[17] C. Lanczos, Solution of systems of linear equations by minimized iteration, *J. Res. N.B.S.* **49** (1952) 33–53.

[18] D.G. Luenberger, *Linear and Nonlinear Programming* (Addison-Wesley, Reading, MA, 2nd ed., 1983).

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

[20] M. Mešina, Convergence acceleration for the iterative solution of the equations $X = AX + f$, *Comp. Meth. Appl. Mech. Eng.* **10** (1977) 165–173.

[21] Y. Saad, Krylov subspace methods for solving large unsymmetric linear systems, *Math. Comp.* **37** (1981) 105–126.

[22] Y. Saad, The Lanczos biorthogonalization algorithm and other oblique projection methods for solving large unsymmetric linear systems, *SIAM J. Numer. Anal.* **19** (1982) 485–506.

[23] Y. Saad and M.H. Schultz, A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Stat. Computing* **7** (1986) 856–869.

[24] A. Sidi, Convergence and stability properties of minimal polynomial and reduced rank extrapolation algorithms, *SIAM J. Numer. Anal.* **23** (1986) 197–209.

[25] 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.* **23** (1988) (this issue) 35–61.

[26] A. Sidi, W.F. Ford and D.A. Smith, Acceleration of convergence of vector sequences, *SIAM J. Numer. Anal.* **23** (1986) 178–196.

[27] D.A. Smith, W.F. Ford and A. Sidi, Extrapolation methods for vector sequences, *SIAM Rev.* **29** (1987) 199–233; see also Correction to "Extrapolation methods for vector sequences", submitted.

[28] E.L. Stiefel, Relaxationsmethoden bester Strategie zur losung linearer Gleichungssystems, *Comment. Math. Helv.* **29** (1955) 157–179.

[29] P.K.W. Vinsome, Orthomin, an iterative method for solving sparse sets of simultaneous linear equations, in: Proc. Fourth Symposium on Reservoir Simulation, Society of Petroleum Engineers of AIME, 1976, pp. 149–159.

[30] O. Widlund, A Lanczos method for a class of nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.* **15** (1978) 801–812.

[31] D.M. Young and K.C. Jea, Generalized conjugate gradient acceleration of nonsymmetrizable iterative methods, *Lin. Alg. Appl.* **34** (1980) 159–194.