Generalized Conjugate-Gradient Acceleration of Nonsymmetrizable Iterative Methods*

David M. Young and Kang C. Jea
Center for Numerical Analysis
University of Texas at Austin
Austin, Texas 78712

Submitted by Robert J. Plemmons

ABSTRACT

Conjugate-gradient acceleration provides a powerful tool for speeding up the convergence of a symmetrizable basic iterative method for solving a large system of linear algebraic equations with a sparse matrix. The object of this paper is to describe three generalizations of conjugate-gradient acceleration which are designed to speed up the convergence of basic iterative methods which are not necessarily symmetrizable. The application of the procedures to some commonly used basic iterative methods is described.

1. INTRODUCTION

In this paper we are concerned with certain iterative methods for solving the linear system

$$Au = b \tag{1.1}$$

where A is a real nonsingular $N \times N$ matrix and b is a given real $N \times 1$ column matrix. We are primarily interested in cases where the matrix A is very large and very sparse. We consider the acceleration of "basic" iterative methods of the form

$$u^{(n+1)} = Gu^{(n)} + k, \qquad n = 0, 1, 2, \ldots, \tag{1.2}$$

*The work was supported in part by the National Science Foundation under Grants MCS76-03141 and MCS-7919829 with The University of Texas at Austin.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/80/060159+36$01.75

where $u^{(0)}$ is arbitrary and for some nonsingular "splitting" matrix Q we have

$$G = I - Q^{-1}A, \qquad k = Q^{-1}b. \tag{1.3}$$

We say the iterative method is *symmetrizable* if for some matrix H, which is symmetric and positive definite (SPD), the matrix $HQ^{-1}A$ is SPD. Otherwise, the method is said to be *nonsymmetrizable*.

In the symmetrizable case, the eigenvalues of the iteration matrix G are real and less than unity. With Chebyshev acceleration (see, e.g., [16]) the rate of convergence of the method (1.2) can be substantially increased. However, to apply Chebyshev acceleration it is necessary to either estimate upper and lower bounds on the eigenvalues of G or else determine them adaptively as done by Hageman and Young [19]. As an alternative to Chebyshev acceleration, conjugate-gradient acceleration (CG acceleration) provides a powerful tool for the acceleration of (1.2) in the symmetrizable case; see, for instance, [20], [21], [9], [10], [25], [26], [1], [2], and [8]. The amount of work required per iteration by CG acceleration is slightly greater than for Chebyshev acceleration. However, no information concerning the eigenvalues of G is required. Also, based on a certain norm of the error vector, the convergence of CG acceleration is at least as fast as that of Chebyshev acceleration.

If the basic iterative method (1.2) is not symmetrizable, the eigenvalues of G may be complex. Even in this case, however, one can often accelerate the convergence using Chebyshev acceleration provided one has sufficient information on the location of the eigenvalues of G; see, e.g., Manteuffel [24]. Manteuffel also showed how regions containing the eigenvalues of G can often be determined adaptively.

Based on the situation in the symmetrizable case, one might hope for the existence of a generalized CG acceleration procedure which could be applied to a nonsymmetrizable iterative method, would converge at least as fast as Chebyshev acceleration, and would not require any knowledge of the eigenvalues of G. The object of this paper is to contribute to the search for such a method.

In Sec. 2 we introduce an acceleration procedure which we refer to as the "idealized generalized CG acceleration" procedure (IGCG procedure). The IGCG procedure is derived as a polynomial acceleration procedure where a Galerkin condition is assumed. In some cases, including the symmetrizable case, this condition implies that the error vector is minimized with respect to a certain norm. In the symmetrizable case one obtains standard CG acceleration. Three equivalent forms of the IGCG method are considered in Secs. 3 through 5. We refer to these forms as ORTHODIR,

ORTHOMIN, and ORTHORES. In the symmetrizable case ORTHOMIN closely corresponds to the usual two-term form of CG acceleration (see [20]), while ORTHORES corresponds to the usual three-term form (see [13]).

The three equivalent versions of the IGCG method are in most cases impractical for numerical computation, since to obtain $u^{(n+1)}$ one would have to have available in storage an unacceptably large number of previously determined vectors. (This is not true in the symmetrizable case, where $u^{(n+1)}$ can be obtained in terms of a few previously computed vectors.) For each of the three forms of the IGCG method we consider "truncated" versions where only a few previously determined vectors are used. The idea of truncating seems to have been first used by Vinsome [28]; see also Axelsson [3–5]. The truncated versions of the three procedures are not in general equivalent even though the idealized versions are equivalent. We consider some of the properties of the truncated schemes in Sec. 3 through 6. One cannot, of course, expect that the truncated procedures will have all of the properties enjoyed by the nontruncated, or "idealized," processes. However, it is desired that, at least in nearly symmetrizable cases, the truncated procedures will in many cases behave much like the idealized procedures and, indeed, much like the CG method.

Each of the procedures which we consider requires the choice of an auxiliary matrix Z such that, at the least, $ZQ^{-1}A$ is positive real, i.e., $ZQ^{-1}A + (ZQ^{-1}A)^T$ is SPD. In Secs. 7 and 8 we consider various choices of Z. In some cases our choice of Z is based on the desire for the behavior of the procedure involved to be as close as possible to that of CG acceleration in the "almost symmetrizable" case. The choice of Z in the application of the acceleration procedures to specific basic iteration methods is considered in Sec. 9.

Numerical experiments based on some of the methods have been carried out by Eisenstat, Elman, Schultz, and Sherman [11], with very promising results. We have also carried out some preliminary experiments which indicate that the methods are effective in a number of cases. These experiments are continuing and the results will be given in a later report [34]. A great deal of additional theoretical analysis and numerical experimentation is clearly needed.

Further theoretical and numerical studies of methods for accelerating nonsymmetrizable iterative methods should include the Lanczos method; see, e.g., [22], [23], [15], [30], [32], and [34]. The Lanczos method has the advantage that at any stage only information concerning at most two previous iterations is required. The method converges in at most N iterations in the absence of rounding error provided that it does not break down. Unfortunately, there do not seem to be any general theorems concerning when such failures may and may not occur.

2. THE IDEALIZED GENERALIZED CONJUGATE GRADIENT ACCELERATION PROCEDURE

We now derive an acceleration procedure based on the basic iterative method (1.2) which is a form of polynomial acceleration; see [6] and [16]. Using the notation of Young [31, Chapter 11], we can write the general polynomial acceleration procedure based on (1.2) in the form

$$u^{(n)} = \sum_{i=0}^{n} \alpha_{n,i}\phi^{(i)} \tag{2.1}$$

where $u^{(0)}$ is arbitrary, $\phi^{(0)} = u^{(0)}$, and

$$\phi^{(i)} = G\phi^{(i-1)} + k, \qquad i = 1,2,\ldots . \tag{2.2}$$

For each n, the numbers $\alpha_{n,0}, \alpha_{n,1}, \ldots, \alpha_{n,n}$ may depend on $u^{(0)}, u^{(1)}, \ldots, u^{(n-1)}$ and must satisfy the condition

$$\sum_{i=0}^{n} \alpha_{n,i} = 1. \tag{2.3}$$

It can easily be shown that the error vector $\epsilon^{(n)} = u^{(n)} - \bar{u}$, where $\bar{u} = A^{-1}b$ is the true solution of (1.1), satisfies the condition

$$\epsilon^{(n)} = P_n(G)\epsilon^{(0)}. \tag{2.4}$$

Here $P_n(G)$ is the polynomial

$$P_n(G) = \sum_{i=0}^{n} \alpha_{n,i}G^i. \tag{2.5}$$

Corresponding to the initial vector $u^{(0)}$, let the *pseudoresidual vector* $\delta^{(0)}$ associated with the basic iterative method (1.2) be defined by

$$\delta^{(0)} = k - (I-G)u^{(0)}. \tag{2.6}$$

We assume that $\delta^{(0)} \neq 0$; otherwise, $u^{(0)} = \bar{u}$. Let t be the largest integer such that the vectors $\delta^{(0)}, (I-G)\delta^{(0)}, \ldots, (I-G)^t\delta^{(0)}$ are linearly independent. Evidently, $0 \leqslant t \leqslant N-1$. For $n = 1,2,\ldots,t+1$, we define the *Krylov space*

$K_n(\delta^{(0)})$ as the vector space spanned by $\delta^{(0)}$, $(I-G)\delta^{(0)},\ldots,(I-G)^{n-1}\delta^{(0)}$. Thus we write

$$K_n(\delta^{(0)})=\mathrm{Sp}\{\delta^{(0)},(I-G)\delta^{(0)},\ldots,(I-G)^{n-1}\delta^{(0)}\}. \tag{2.7}$$

We now give a well-known alternative characterization of polynomial acceleration. For details of the proof see, for instance, [33].

THEOREM 2.1. *A procedure for generating vectors* $u^{(1)}$, $u^{(2)},\ldots$ *corresponding to a given* $u^{(0)}$ *such that* $\delta^{(0)}=k-(I-G)u^{(0)}\neq 0$ *is a polynomial acceleration procedure based on the basic iterative method* (1.2) *if and only if for* $n=1, 2,\ldots,$ *we have*

$$u^{(n)}-u^{(0)}\in K_n(\delta^{(0)}). \tag{2.8}$$

For the balance of this section and for several subsequent sections we will assume that the basic iterative method which we wish to accelerate is the *RF method*[1] defined by

$$u^{(n+1)}=(I-A)u^{(n)}+b. \tag{2.9}$$

For each n we define the *residual vector* $r^{(n)}$ by

$$r^{(n)}=b-Au^{(n)}. \tag{2.10}$$

Evidently, in this case $\delta^{(0)}=r^{(0)}$. The Krylov spaces $K_1(r^{(0)})$, $K_2(r^{(0)}),\ldots,$ are defined by

$$K_n(r^{(0)})=\mathrm{Sp}\{r^{(0)},Ar^{(0)},\ldots,A^{n-1}r^{(0)}\}. \tag{2.11}$$

Moreover, by Theorem 2.1, for any polynomial acceleration procedure based on the RF method we have

$$u^{(n)}-u^{(0)}\in K_n(r^{(0)}). \tag{2.12}$$

[1] In the terminology of Young [31], the RF method is a special case of the method of Richardson [27] where the iteration parameter, which for the general method varies from iteration to iteration, is fixed.

We now consider an acceleration procedure which involves the use of an *auxiliary matrix* Z such that ZA is positive real (PR). One choice of Z would be $Z=A^T$. In that case ZA would be SPD. Another choice would be $Z=A^TY$ where Y is PR. In that case ZA would be PR. Other choices are discussed in Sec. 9.

For the symmetrizable case where Z and ZA are SPD, we could derive the standard CG acceleration procedure by requiring that $u^{(n)}-u^{(0)}\in K_n(r^{(0)})$ and requiring that

$$\|u^{(n)}-\bar{u}\|_{(ZA)^{1/2}}\leq\|w-\bar{u}\|_{(ZA)^{1/2}} \tag{2.13}$$

for all w such that $w-u^{(0)}\in K_n(r^{(0)})$. For details, see, e.g., [33]. We could also derive a generalized CG acceleration procedure using (2.13) if ZA is SPD. However, in order to treat the more general case where ZA may be PR, we replace (2.13) by the *Galerkin condition*

$$\big(ZA(u^{(n)}-\bar{u}),v\big)=(Zr^{(n)},v)=0 \tag{2.14}$$

for all $v\in K_n(r^{(0)})$. It is easy to show (see, e.g., [9], [5], or [33]) that if ZA is SPD, then (2.14) holds if and only if (2.13) holds.

As shown in Sec. 3, the conditions (2.12) and (2.14) uniquely determine a set of vectors $u^{(1)}, u^{(2)},\ldots,u^{(t+1)}$ corresponding to a given initial vector $u^{(0)}$. Moreover, we have "finite termination" in the sense that

$$u^{(t+1)}=\bar{u}. \tag{2.15}$$

We refer to the method thus defined as the *idealized generalized conjugate-gradient acceleration procedure* (IGCG procedure). One implementation of the IGCG procedure, called ORTHODIR, is given in Sec. 3. Other implementations, which can be used if Z as well as ZA is PR, are given in Secs. 4 and 5.

Axelsson [5] derived several generalized CG acceleration procedures based on the use of (2.14). However, he assumed, in effect, that Z is SPD as well as that ZA is PR. For the case where Z is SPD and ZA is PR, his procedures are equivalent to the IGCG procedure.

3. ORTHODIR

We now describe a procedure for implementing the IGCG procedure described in Sec. 2. Given the auxiliary matrix Z such that ZA is PR, we construct an ordered set of vectors $q^{(0)}, q^{(1)},\ldots,q^{(t)}$ which are pairwise

"semiorthogonal" with respect to ZA in the sense that

$$(ZAq^{(i)}, q^{(j)}) = 0, \qquad j<i, \quad i,j=0,1,\ldots,t. \tag{3.1}$$

Here, as before, t is the largest integer such that $r^{(0)}, Ar^{(0)},\ldots, A^t r^{(0)}$ are linearly independent. If ZA is SPD, then the vectors $\{q^{(i)}\}$ are orthogonal with respect to ZA in the sense that

$$(ZAq^{(i)}, q^{(j)}) = 0, \qquad j\neq i, \quad i,j=0,1,\ldots,t. \tag{3.2}$$

The existence of the vectors $\{q^{(i)}\}$ and a procedure for their construction is given by the following theorem, which we state without proof. Details are given in [34].

THEOREM 3.1. *Let v be any nonzero vector in R^N, and let A be any nonsingular matrix in $R^{N,N}$. Let H be any PR matrix in $R^{N,N}$, and let t be any nonnegative integer such that the vectors $v, Av,\ldots, A^t v$ are linearly independent. Then the set of vectors $w^{(0)}, w^{(1)},\ldots, w^{(t)}$ defined by*

$$w^{(0)} = v$$

$$w^{(n)} = Aw^{(n-1)} + \beta_{n,n-1}w^{(n-1)} + \ldots + \beta_{n,0}w^{(0)}, \qquad n=1,2,\ldots,t, \tag{3.3}$$

where

$$\beta_{n,i} = -\frac{(HAw^{(n-1)}, w^{(i)}) + \sum_{j=0}^{i-1} \beta_{n,j}(Hw^{(j)}, w^{(i)})}{(Hw^{(i)}, w^{(i)})},$$

$$i=0,1,\ldots,n-1, \quad n=1,2,\ldots,t, \tag{3.4}$$

are linearly independent and satisfy

$$(Hw^{(i)}, w^{(j)}) = 0, \qquad j<i, \quad i,j=0,1,\ldots,t. \tag{3.5}$$

Moreover, for each $n=1,2,\ldots,t$ there exists coefficients $c_{n,0}, c_{n,1},\ldots, c_{n,n-1}$ such that

$$w^{(n)} = c_{n,0}v + c_{n,1}Av + \ldots + c_{n,n-1}A^{n-1}v + A^n v. \tag{3.6}$$

Also, for $n=0,1,\ldots,t$, *there exists coefficients* $e_{n,0}, e_{n,1},\ldots,e_{n,n-1}$ *such that*

$$A^n v = e_{n,0}w^{(0)} + e_{n,1}w^{(1)} + \ldots + e_{n,n-1}w^{(n-1)} + w^{(n)}. \tag{3.7}$$

COROLLARY 3.2. *If* H *is* SPD, *then Theorem* 3.1 *remains true if we replace* (3.4) *by*

$$\beta_{n,i} = -\frac{(HAw^{(n-1)}, w^{(i)})}{(Hw^{(i)}, w^{(i)})}, \qquad i=0,1,2,\ldots,n-1, \quad n=1,2,\ldots,t, \tag{3.8}$$

and if we replace (3.5) *by*

$$(Hw^{(i)}, w^{(j)}) = 0, \qquad j \neq i, \quad i,j=0,1,\ldots,t. \tag{3.9}$$

If we apply Theorem 3.1 and Corollary 3.2 with $q^{(n)}=w^{(n)}$, $v=r^{(0)}$, and $H=ZA$, we get the "direction vectors" $q^{(0)}, q^{(1)},\ldots$ by

$$\begin{aligned} &q^{(0)} = r^{(0)}, \\ &q^{(n)} = Aq^{(n-1)} + \beta_{n,n-1}q^{(n-1)} + \ldots + \beta_{n,0}q^{(0)}, \qquad n=1,2,\ldots,t, \end{aligned} \tag{3.10}$$

where $\beta_{n,0}, \beta_{n,1},\ldots,\beta_{n,n-1}$ are determined by

$$\beta_{n,i} = -\frac{(ZA^2q^{(n-1)}, q^{(i)}) + \sum_{j=0}^{i-1} \beta_{n,j}(ZAq^{(j)}, q^{(i)})}{(ZAq^{(i)}, q^{(i)})},$$

$$i=0,1,\ldots,n-1, \quad n=1,2,\ldots,t. \tag{3.11}$$

Evidently, the vectors $q^{(0)}, q^{(1)},\ldots,q^{(t)}$ are linearly independent and satisfy (3.1). Moreover, for each i we have

$$q^{(i)} = c_{i,0}r^{(0)} + c_{i,1}Ar^{(0)} + \ldots + c_{i,i-1}A^{i-1}r^{(0)} + A^i r^{(0)} \tag{3.12}$$

for some coefficients $c_{i,0}, c_{i,1},\ldots,c_{i,i-1}$. Also for each i we have

$$A^i r^{(0)} = e_{i,0}q^{(0)} + \ldots + e_{i,i-1}q^{(i-1)} + q^{(i)} \tag{3.13}$$

for some coefficients $e_{i,0}, e_{i,1}, \ldots, e_{i,i-1}$. Thus for each i the set of vectors $\{q^{(0)}, q^{(1)}, \ldots, q^{(i)}\}$ spans $K_{i+1}(r^{(0)})$.

The procedure used to construct the direction vectors $q^{(0)}, q^{(1)}, \ldots$ is a slight generalization of that described by Faddeev and Faddeeva [14, pp. 277–279], and which is referred to as the "method of orthogonalized iterations."

We remark that another set of vectors could be constructed by using an extension of the Gram-Schmidt process, similar to that used in Theorem 3.1, for the vectors $r^{(0)}, Ar^{(0)}, \ldots, A^t r^{(0)}$. The vectors thus obtained would be the same, except for length, as those given by (3.10) and (3.11).

We now seek to find $u^{(n)}$, for each n, such that $u^{(n)} - u^{(0)} \in K_n(r^{(0)})$ and such that the conditions

$$(Zr^{(n)}, q^{(i)}) = 0, \qquad i = 0, 1, \ldots, n-1, \tag{3.14}$$

hold. Evidently, by (3.12) and (3.13), (3.14) is a necessary and sufficient condition for (2.14) to hold. Since $u^{(n)} - u^{(0)} \in K_n(r^{(0)})$, we have, by (3.13),

$$u^{(n)} = u^{(0)} + \lambda_{n,0} q^{(0)} + \lambda_{n,1} q^{(1)} + \ldots + \lambda_{n,n-1} q^{(n-1)} \tag{3.15}$$

for some $\lambda_{n,0}, \lambda_{n,1}, \ldots, \lambda_{n,n-1}$, and since $r^{(n)} = b - Au^{(n)}$, it follows that

$$r^{(n)} = r^{(0)} - \lambda_{n,0} Aq^{(0)} - \lambda_{n,1} Aq^{(1)} - \ldots - \lambda_{n,n-1} Aq^{(n-1)}. \tag{3.16}$$

If $n = 1$, we have

$$r^{(1)} = r^{(0)} - \lambda_{1,0} Aq^{(0)}$$

and

$$(Zr^{(1)}, q^{(0)}) = (Zr^{(0)}, q^{(0)}) - \lambda_{1,0}(ZAq^{(0)}, q^{(0)}),$$

so that

$$\lambda_{1,0} = \frac{(Zr^{(0)}, q^{(0)})}{(ZAq^{(0)}, q^{(0)})}. \tag{3.17}$$

We seek to show by induction that

$$r^{(n)} = r^{(n-1)} - \hat{\lambda}_{n-1} A q^{(n-1)},$$

$$\lambda_{n,i} = \hat{\lambda}_i, \qquad i = 0, 1, \ldots, n-1, \tag{3.18}$$

where we define $\hat{\lambda}_i$ by

$$\hat{\lambda}_i = \frac{(Zr^{(i)}, q^{(i)})}{(ZAq^{(i)}, q^{(i)})}, \qquad i = 0, 1, \ldots, t. \tag{3.19}$$

By (3.16) we have

$$r^{(n+1)} = r^{(0)} - \lambda_{n+1,0} A q^{(0)} - \lambda_{n+1,1} A q^{(1)} - \ldots - \lambda_{n+1,n} A q^{(n)}. \tag{3.20}$$

Since $(Zr^{(n+1)}, q^{(0)}) = 0$ we get, by (3.1),

$$\lambda_{n+1,0} = \frac{(Zr^{(0)}, q^{(0)})}{(ZAq^{(0)}, q^{(0)})} = \hat{\lambda}_0. \tag{3.21}$$

Thus $r^{(0)} - \lambda_{n+1,0} A q^{(0)} = r^{(0)} - \hat{\lambda}_0 A q^{(0)} = r^{(1)}$ and

$$r^{(n+1)} = r^{(1)} - \lambda_{n+1,1} A q^{(1)} - \ldots - \lambda_{n+1,n} A q^{(n)}. \tag{3.22}$$

Since $(Zr^{(n+1)}, q^{(1)}) = 0$, we get, by (3.1),

$$\lambda_{n+1,1} = \frac{(Zr^{(1)}, q^{(1)})}{(ZAq^{(1)}, q^{(1)})} = \hat{\lambda}_1. \tag{3.23}$$

Continuing, we get

$$r^{(n+1)} = r^{(n)} - \lambda_{n+1,n} A q^{(n)}. \tag{3.24}$$

Since $(Zr^{(n+1)}, q^{(n)}) = 0$, we get

$$\lambda_{n+1,n} = \frac{(Zr^{(n)}, q^{(n)})}{(ZAq^{(n)}, q^{(n)})} = \hat{\lambda}_n. \tag{3.25}$$

Moreover,

$$r^{(n+1)}=r^{(n)}-\hat{\lambda}_n Aq^{(n)}, \tag{3.26}$$

and (3.18) is proved.

From the above discussion it follows that, for each $n \leqslant t+1$, $u^{(n)}$ is given by

$$u^{(n)}=u^{(0)}+\hat{\lambda}_0 q^{(0)}+\hat{\lambda}_1 q^{(1)}+\ldots+\hat{\lambda}_{n-1}q^{(n-1)}, \tag{3.27}$$

where

$$\hat{\lambda}_i=\frac{(Zr^{(i)}, q^{(i)})}{(ZAq^{(i)}, q^{(i)})}, \qquad i=0,1,2,\ldots,t. \tag{3.28}$$

We now show that the true solution $\bar{u}$ can be written in the form

$$\bar{u}=u^{(0)}+\hat{\lambda}_0 q^{(0)}+\hat{\lambda}_1 q^{(1)}+\ldots+\hat{\lambda}_t q^{(t)}. \tag{3.29}$$

To do this we state without proof the following lemma. Details of the proof are given in [34].

LEMMA 3.3. *Let $r^{(0)} \neq 0$, and let t be the largest integer such that $r^{(0)}, Ar^{(0)}, \ldots, A^t r^{(0)}$ are linearly independent. Then*

$$\bar{u}-u^{(0)} \in K_{t+1}(r^{(0)}). \tag{3.30}$$

Moreover, if $\bar{u}-u^{(0)} \in K_m(r^{(0)})$ for some integer m, then $m \geqslant t+1$.

From Lemma 3.3 and (3.13) it follows that $\bar{u}$ can be written in the form

$$\bar{u}=u^{(0)}+\hat{\lambda}'_0 q^{(0)}+\hat{\lambda}'_1 q^{(1)}+\ldots+\hat{\lambda}'_t q^{(t)}. \tag{3.31}$$

It can then be shown (see [34] for details) that

$$\hat{\lambda}'_i=\hat{\lambda}_i, \qquad i=0,1,\ldots,t. \tag{3.32}$$

Using (3.27), (3.10), (3.28), and (3.11), we now define the following procedure:

$$u^{(n+1)}=u^{(n)}+\hat{\lambda}_n q^{(n)},$$

$$q^{(n)}=\begin{cases} r^{(0)}, & n=0, \\ Aq^{(n-1)}+\beta_{n,n-1}q^{(n-1)}+\ldots+\beta_{n,0}q^{(0)}, & n\geqslant 1, \end{cases}$$

$$\hat{\lambda}_n=\frac{(Zr^{(n)},q^{(n)})}{(ZAq^{(n)},q^{(n)})}, \tag{3.33}$$

$$\beta_{n,i}=-\frac{(ZA^2q^{(n-1)},q^{(i)})+\sum_{j=0}^{i-1}\beta_{n,j}(ZAq^{(j)},q^{(i)})}{(ZAq^{(i)},q^{(i)})},$$

$$i=0,1,2,\ldots,n-1, \quad n=1,2,,\ldots,t$$

We refer to this procedure as ORTHODIR.

It is evident from (3.27) and (3.29) that in the absence of rounding errors ORTHODIR converges in $t+1$ steps, where $t+1\leqslant N$. We now derive some additional properties of the direction vectors $\{q^{(n)}\}$ and the residual vectors $\{r^{(n)}\}$. First by (3.14) we have

$$(Zr^{(i)},q^{(j)})=0, \qquad j<i, \quad i,j=0,1,\ldots,t. \tag{3.34}$$

Next we show that

$$(Zr^{(i)},r^{(j)})=0, \qquad j<i, \quad i,j=0,1,\ldots,t. \tag{3.35}$$

Again we use induction. Clearly, $(Zr^{(1)},r^{(0)})=(Zr^{(1)},q^{(0)})=0$. If the result holds for $j<i$, $i,j=0,1,\ldots,n$, where $n\leqslant t-1$, then for $j\leqslant n$ we have

$$(Zr^{(n+1)},r^{(j)})=\left(Zr^{(n+1)},r^{(j-1)}-\hat{\lambda}_{j-1}Aq^{(j-1)}\right).$$

Continuing, we get

$$(Zr^{(n+1)},r^{(j)})=\left(Zr^{(n+1)},r^{(0)}-\hat{\lambda}_{j-1}Aq^{(j-1)}-\hat{\lambda}_{j-2}Aq^{(j-2)}-\ldots-\hat{\lambda}_0Aq^{(0)}\right). \tag{3.36}$$

But by (3.10) we have

$$Aq^{(k)}=q^{(k+1)}-\beta_{k+1,k}q^{(k)}-\ldots-\beta_{k+1,0}q^{(0)}. \tag{3.37}$$

Thus, since $r^{(0)}=q^{(0)}$, the result (3.35) follows from (3.34).

It follows from (3.35) and (3.34) that

$$\begin{aligned}(Zr^{(n)}, r^{(n)}) &= \left(Zr^{(n)}, r^{(n-1)}-\hat{\lambda}_{n-1}Aq^{(n-1)}\right)\\ &= \left(Zr^{(n)}, r^{(n-1)}-\hat{\lambda}_{n-1}\left\{q^{(n)}-\beta_{n,n-1}q^{(n-1)}-\ldots-\beta_{n,0}q^{(0)}\right\}\right)\\ &= -\hat{\lambda}_{n-1}(Zr^{(n)}, q^{(n)}).\end{aligned} \tag{3.38}$$

We remark that it is easy to show (see, e.g., [34]) that if $u^{(n)}=\bar{u}$ for some $n<t+1$, then $r^{(0)}, Ar^{(0)},\ldots, A^n r^{(0)}$ are linearly dependent. This contradiction shows that if $u^{(n)}=\bar{u}$, then $n\geqslant t+1$.

We will sometimes refer to the method (3.33) as the "idealized" ORTHODIR method. We now consider a "truncated" version of ORTHODIR obtained from (3.33) as follows. We choose an integer s, and we set $\beta_{n,i}=0$ for $i+s<n$. We thus obtain the method defined by (3.33) and

$$\beta_{n,i}=0 \qquad \text{if} \quad i+s<n. \tag{3.39}$$

We refer to the above process as "ORTHODIR(s)." The idealized process is sometimes referred to as "ORTHODIR(∞)."

The truncated version of ORTHODIR has many of the properties of the idealized version. Thus, for example, one can show that

$$(ZAq^{(i)}, q^{(j)})=0, \qquad j=i-1, i-2,\ldots, i-s, \tag{3.40}$$

and

$$(Zr^{(i)}, q^{(j)})=0, \qquad j=i-1, i-2,\ldots, i-s-1 \tag{3.41}$$

(see Young and Jea [34]). Also, as shown in Sec. 6, ORTHODIR(s) has an error-minimization property. On the other hand, it is *not* true in general that

$$(Zr^{(i)}, r^{(j)})=0, \qquad j=i-1, i-2,\ldots, i-s. \tag{3.42}$$

If Z and ZA are SPD, we have the symmetrizable case. It is easy to show that, for $s \geqslant 2$, ORTHODIR(s) is equivalent to ORTHODIR(∞). Details are given in [34]. In Sec. 4 we will show that in this case ORTHODIR(∞) is equivalent to the CG method.

The process ORTHODIR(s) can be formally carried out unless, for some n, $q^{(n)}=0$ but $r^{(n)} \neq 0$. If $q^{(n)}=0$ and $r^{(n)} \neq 0$, we say that the process has "broken down." For the idealized case we are assured that the process will not break down. However, we have no proof that the truncated procedure will not break down. On the contrary, Schultz and Elman (private communication) have found cases where the process does indeed break down.

We can, however, show that $q^{(n)} \neq 0$ for $n < t+1$ where t is defined above. Thus, it can be shown, using (3.33) and induction, that $q^{(n)}$ can be written in the form

$$q^{(n)} = A^n r^{(0)} + c_{n,n-1} A^{n-1} r^{(0)} + \ldots + c_{n,0} r^{(0)} \tag{3.43}$$

for some coefficients $c_{n,n-1}, c_{n,n-2}, \ldots, c_{n,0}$. Hence, if $q^{(n)}=0$, then $r^{(0)}, Ar^{(0)}, \ldots, A^n r^{(0)}$ are linearly dependent. It thus follows that ORTHODIR(s) cannot break down for at least t iterations.

For the idealized version of ORTHODIR we require only that ZA be PR. On the other hand, for the truncated version ORTHODIR(s) we do not have necessary or sufficient conditions to guarantee against a breakdown of the process. One might conjecture that the condition that Z is PR would be sufficient. However, to our knowledge this question remains open. These considerations make the idea of "restarting," as considered by Eisenstat, Elman, Schultz, and Sherman [11], seem very attractive. With restarting, one iterates for a fixed number, say k, of iterations using the idealized procedure, obtaining an approximate solution vector u. One then starts over with $u^{(0)}=u^{(k)}$ and performs the same number of iterations. The process is repeated. It is by no means clear that the process will necessarily converge, but at least it will not break down.

4. ORTHOMIN

We now develop an alternative form of the IGCG method which, in the symmetrizable case, reduces to the standard two-term form of the CG method. We refer to this modified procedure as ORTHOMIN. In addition to the assumption which was made in Secs. 2 and 3, that ZA is PR, we now assume that Z is PR.

The ORTHOMIN procedure is based on the use of the modified direction vectors $p^{(0)}, p^{(1)}, \ldots, p^{(t)}$ defined by

$$p^{(n)} = \begin{cases} q^{(0)}, & n=0, \\ -\hat{\lambda}_{n-1} q^{(n)}, & n=1,2,\ldots,t. \end{cases} \tag{4.1}$$

Here the $\{q^{(n)}\}$ are the direction vectors used for ORTHODIR and are given by (3.10). The $\{\hat{\lambda}_i\}$ are given by (3.19). We note that since Z and ZA are PR, then by (3.38), none of the $\hat{\lambda}_i$ vanishes. Therefore, none of the $p^{(i)}$ vanishes for $i=0,1,\ldots,t$.

We now show that the $\{p^{(i)}\}$ satisfy

$$p^{(n)} = \begin{cases} r^{(0)}, & n=0, \\ r^{(n)} + \alpha_{n,n-1} p^{(n-1)} + \ldots + \alpha_{n,0} p^{(0)}, & n=1,2,\ldots,t, \end{cases} \tag{4.2}$$

where

$$\alpha_{n,i} = -\frac{(ZAr^{(n)}, p^{(i)}) + \sum_{j=0}^{i-1} \alpha_{n,j} (ZAp^{(j)}, p^{(i)})}{(ZAp^{(i)}, p^{(i)})},$$

$$i=0,1,2,\ldots,n-1, \quad n=1,2,\ldots,t. \tag{4.3}$$

We use induction. By (3.10) we have

$$q^{(1)} = Aq^{(0)} + \beta_{1,0} q^{(0)}, \tag{4.4}$$

and by (3.18),

$$r^{(1)} = r^{(0)} - \hat{\lambda}_0 Aq^{(0)}. \tag{4.5}$$

Therefore,

$$q^{(1)} = -\frac{1}{\hat{\lambda}_0}(r^{(1)} - r^{(0)}) + \beta_{1.0} q^{(0)} \tag{4.6}$$

and

$$p^{(1)} = r^{(1)} + \left(-1 - \hat{\lambda}_0 \beta_{1,0}\right) p^{(0)}$$

$$= r^{(1)} + \alpha_{1,0} p^{(0)}, \tag{4.7}$$

where we let $\alpha_{1,0} = -1 - \hat{\lambda}_0 \beta_{1,0}$. Since $(ZAq^{(1)}, q^{(0)}) = 0$, it follows that $(ZAp^{(1)}, p^{(0)}) = 0$, and hence

$$\alpha_{1,0} = -\frac{(ZAr^{(1)}, p^{(0)})}{(ZAp^{(0)}, p^{(0)})}. \tag{4.8}$$

Suppose now that (4.2) and (4.3) hold for $k = 1, 2, \ldots, n$. Then

$$q^{(n+1)} = Aq^{(n)} + \beta_{n+1,n} q^{(n)} + \ldots + \beta_{n+1,0} q^{(0)}. \tag{4.9}$$

By (3.18) we have

$$Aq^{(n)} = -\frac{1}{\hat{\lambda}_n}\left(r^{(n+1)} - r^{(n)}\right) \tag{4.10}$$

and

$$p^{(n+1)} = -\hat{\lambda}_n q^{(n+1)}$$

$$= r^{(n+1)} - r^{(n)} - \hat{\lambda}_n \beta_{n+1,n} q^{(n)} - \ldots - \hat{\lambda}_n \beta_{n+1,0} q^{(0)}. \tag{4.11}$$

But

$$r^{(n)} = p^{(n)} - \alpha_{n,n-1} p^{(n-1)} - \ldots - \alpha_{n,0} p^{(0)} \tag{4.12}$$

Thus, for some $\alpha_{n+1,0}, \alpha_{n+1,1}, \ldots, \alpha_{n+1,n}$ we have

$$p^{(n+1)} = r^{(n+1)} + \alpha_{n+1,n} p^{(n)} + \ldots + \alpha_{n+1,0} p^{(0)}. \tag{4.13}$$

By requiring that $(ZAp^{(n+1)}, p^{(i)}) = 0$, $i = 0, 1, \ldots, n$, we get (4.2) and (4.3).

From (3.33) and (4.1) we have

$$u^{(n+1)}=u^{(n)}+\hat{\lambda}_n q^{(n)}=u^{(n)}-\frac{\hat{\lambda}_n}{\hat{\lambda}_{n-1}}p^{(n)} \tag{4.14}$$

and

$$\hat{\lambda}_n=-\hat{\lambda}_{n-1}\frac{(Zr^{(n)},p^{(n)})}{(ZAp^{(n)},p^{(n)})}. \tag{4.15}$$

From this it follows that (3.33) can be written in the form

$$\begin{aligned}
&u^{(n+1)}=u^{(n)}+\lambda_n p^{(n)}\\
&p^{(n)}=\begin{cases} r^{(0)}, & n=0\\ r^{(n)}+\alpha_{n,n-1}p^{(n-1)}+\ldots+\alpha_{n,0}p^{(0)}, & n\geqslant 1\end{cases}\\
&\lambda_n=\frac{(Zr^{(n)},p^{(n)})}{(ZAp^{(n)},p^{(n)})}\\
&\alpha_{n,i}=-\frac{(ZAr^{(n)},p^{(i)})+\sum_{j=0}^{i-1}\alpha_{n,j}(ZAp^{(j)},p^{(i)})}{(ZAp^{(i)},p^{(i)})},\\
&\times\qquad\qquad i=0,1,\ldots,n-1,\quad n=1,2,\ldots .
\end{aligned} \tag{4.16}$$

The method defined by (4.16) is essentially the method given by Vinsome [28] and which he called "ORTHOMIN." We will use the same terminology here.

As we did for ORTHODIR, we now consider a "truncated" version of ORTHOMIN. We choose an integer s and let $\alpha_{n,i}=0$ for $i+s<n$. We thus obtain the method defined by (4.16) and

$$\alpha_{n,i}=0 \qquad \text{if} \quad i+s<n. \tag{4.17}$$

We refer to the above process as "ORTHOMIN(s)." Sometimes we refer to the "idealized" procedure (4.16) as "ORTHOMIN(∞)."

It can be shown that (3.40) and (3.41) hold for the $\{p^{(i)}\}$. It is also shown in Sec. 6 that ORTHOMIN(s) has a certain error-minimization property. Eisenstat, Elman, and Schultz [12] proved that the process cannot break down. It is sufficient to show that if $r^{(0)}, r^{(1)}, \ldots, r^{(n)}$ do not vanish, then $p^{(0)}, p^{(1)}, \ldots, p^{(n)}$ and $\lambda_0, \lambda_1, \ldots, \lambda_n$ do not vanish. This can be done by induction. If $r^{(0)} \neq 0$, then $p^{(0)} \neq 0$ and $\lambda_0 = (Zr^{(0)}, p^{(0)})/(ZAp^{(0)}, p^{(0)}) \neq 0$, since ZA and Z are PR. Suppose that $r^{(0)}, \ldots, r^{(n+1)}$ do not vanish and that $p^{(0)}, \ldots, p^{(n)}$ and $\lambda_0, \ldots, \lambda_n$ do not vanish. We seek to show that $p^{(n+1)}$ and λ_{n+1} do not vanish. But

$$\begin{aligned}(Zr^{(n+1)}, p^{(n+1)}) &= (Zr^{(n+1)}, r^{(n+1)} + \alpha_{n+1,n}p^{(n)} + \ldots + \alpha_{n+1,n-s+1}p^{(n-s+1)}) \\ &= (Zr^{(n+1)}, r^{(n+1)}) \end{aligned} \tag{4.18}$$

by (3.41). Hence $(Zr^{(n+1)}, p^{(n+1)}) \neq 0$, and λ_{n+1} and $p^{(n+1)}$ do not vanish by (4.16).

We now show that

$$\lambda_n = \frac{(Zr^{(n)}, r^{(n)})}{(ZAp^{(n)}, r^{(n)})}. \tag{4.19}$$

Thus, we have

$$\begin{aligned}(ZAp^{(n)}, r^{(n)}) &= (ZAp^{(n)}, p^{(n)} - \alpha_{n,n-1}p^{(n-1)} - \ldots - \alpha_{n,n-s}p^{(n-s)}) \\ &= (ZAp^{(n)}, p^{(n)}) \end{aligned} \tag{4.20}$$

by (3.40). Thus (4.19) holds by (4.16) and the fact, as shown above, that $(Zr^{(n)}, p^{(n)}) = (Zr^{(n)}, r^{(n)})$.

Axelsson [3–5] presented a number of generalized CG acceleration schemes. One of these schemes is similar to ORTHOMIN(s). It is conjectured that it is actually equivalent to ORTHOMIN(s). Numerical experiments were carried out using the method with very good results.

In the symmetrizable case, where Z and ZA are SPD, ORTHOMIN(1) is equivalent to ORTHOMIN(∞) and both procedures reduce to the standard

two-term form of the CG method; see, e.g., [34]. Thus we have

$$u^{(n+1)}=u^{(n)}+\lambda_n p^{(n)},$$

$$p^{(n)}=\begin{cases} r^{(0)} & n=0, \\ r^{(n)}+\alpha_{n,n-1}p^{(n-1)}, & n\geqslant 1, \end{cases}$$

$$\lambda_n=\frac{(Zr^{(n)},p^{(n)})}{(ZAp^{(n)},p^{(n)})}, \tag{4.21}$$

$$\alpha_{n,n-1}=-\frac{(ZAr^{(n)},p^{(n-1)})}{(ZAp^{(n-1)},p^{(n-1)})}.$$

It also follows from the discussion of the previous section that in the symmetrizable case ORTHODIR(2) is equivalent to (4.21).

Let us illustrate a difference between the idealized versions of ORTHODIR and ORTHOMIN by considering the following example:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}=\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \tag{4.22}$$

where $\bar{u}^T=(1\ 3)$ and

$$Z=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{4.23}$$

$$u^{(0)}=\begin{pmatrix} 1 \\ 2 \end{pmatrix}. \tag{4.24}$$

While ZA is SPD, Z is not positive real. For ORTHOMIN we get $\lambda_0=0$ and $u^{(0)}=u^{(1)}=\cdots$. Thus the process breaks down. For ORTHODIR, on the other hand, we get

$$\begin{aligned} &\hat{\lambda}_0=0, \qquad q^{(1)}=\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ &u^{(1)}=u^{(0)}, \qquad \hat{\lambda}_1=1, \qquad u^{(2)}=\bar{u}, \end{aligned} \tag{4.25}$$

and the process converges in two iterations.

5. ORTHORES

Engeli, et al. [13] used a three-term form of the CG method (4.21) for the symmetrizable case. The scheme which they used can be written in the form

$$
\begin{aligned}
u^{(n+1)} &= \rho_{n+1}\left(u^{(n)} + \gamma_{n+1} r^{(n)}\right) + (1-\rho_{n+1}) u^{(n-1)}, \\
\gamma_{n+1} &= \frac{(r^{(n)}, Zr^{(n)})}{(r^{(n)}, ZAr^{(n)})}, \\
\rho_{n+1} &= \begin{cases} 1 & n=0, \\ \left[1 - \dfrac{\gamma_{n+1}}{\gamma_n} \dfrac{(r^{(n)}, Zr^{(n)})}{(r^{(n-1)}, Zr^{(n-1)})} \dfrac{1}{\rho_n}\right]^{-1} & n \geqslant 1. \end{cases}
\end{aligned}
\tag{5.1}
$$

We now seek to transform the IGCG method into a form which resembles (5.1). We continue to assume that Z and ZA are PR. From (4.16) we can write

$$
\begin{aligned}
u^{(n+1)} &= u^{(n)} + \lambda_n r^{(n)} + \lambda_n \alpha_{n,n-1} p^{(n-1)} + \cdots + \lambda_n \alpha_{n,0} p^{(0)} = u^{(n)} + \lambda_n r^{(n)} \\
&\quad + \frac{\lambda_n \alpha_{n,n-1}}{\lambda_{n-1}} (u^{(n)} - u^{(n-1)}) + \cdots + \frac{\lambda_n \alpha_{n,0}}{\lambda_0} (u^{(1)} - u^{(0)}).
\end{aligned}
\tag{5.2}
$$

(We note that, as shown in Sec. 4, none of the λ_i vanishes.) From this and from the fact that the residual vectors $r^{(0)}, r^{(1)}, \ldots, r^{(t)}$ are semiorthogonal with respect to Z [see (3.35)], we can derive the following procedure:

$$
\begin{aligned}
u^{(n+1)} &= \gamma_{n+1} f_{n+1,n} r^{(n)} + f_{n+1,n} u^{(n)} + \cdots + f_{n+1,0} u^{(0)}, \\
\gamma_{n+1} &= \frac{1}{\sigma_{n+1,n}}, \\
f_{n+1,n} &= \left(1 + \gamma_{n+1} \sum_{i=0}^{n-1} \sigma_{n+1,i}\right)^{-1}, \\
\sigma_{n+1,i} &= \frac{(ZAr^{(n)}, r^{(i)}) - \sum_{j=0}^{i-1} \sigma_{n+1,j} (Zr^{(j)}, r^{(i)})}{(Zr^{(i)}, r^{(i)})}, \\
&\qquad\qquad i = 0,1,2,\ldots,n, \quad 0,1,\ldots \\
f_{n+1,i} &= \gamma_{n+1} f_{n+1,n} \sigma_{n+1,i}, \qquad i = 0,1,\ldots,n-1.
\end{aligned}
\tag{5.3}
$$

We refer to procedure thus defined as "ORTHORES."

While the idealized forms of ORTHODIR, ORTHOMIN, and ORTHORES are equivalent, nevertheless, the computational effort required may differ. Thus, for instance, if Z is SPD while ZA is PR but not SPD, it would be preferable to use ORTHORES instead of ORTHODIR or ORTHOMIN, since for ORTHORES we would have full orthogonality of the residual vectors with respect to Z, while for ORTHODIR and ORTHOMIN the direction vectors would be semiorthogonal with respect to ZA. On the other hand, if Z is PR but not SPD and if ZA is SPD, then ORTHODIR and ORTHOMIN would be preferable, since for them the direction vectors would be fully orthogonal with respect to ZA, while for ORTHORES the residual vectors would be semiorthogonal with respect to Z.

One might ask whether it is necessary to assume for ORTHORES that ZA is PR as well as that Z is PR. To show that the condition that Z is PR is not sufficient, we consider the example (4.22) given at the end of Sec. 4 where we let $Z=I$. If we let $u^{(0)}=(1\ \ 2)^T$ and attempt to apply (5.3), we find that $(r^{(0)}, ZAr^{(0)})=0$; hence γ_1 cannot be computed and the process breaks down.

Analogous to ORTHODIR(s) and ORTHOMIN(s), we define the truncated ORTHORES process, which we call "ORTHORES(s)," by (5.3) and

$$\sigma_{n+1,i}=0, \qquad i+s<n. \tag{5.4}$$

It can be shown that with ORTHORES(s) we have

$$(Zr^{(i)}, r^{(j)})=0, \qquad j=i-1, i-2, \ldots, i-s. \tag{5.5}$$

However, in general (3.40) and (3.41) do not hold. Consequently, it is not true, in general, that ORTHORES(s) is equivalent to ORTHODIR(s) or to ORTHOMIN(s) except in the idealized case. We also note that we have at this time no assurance that the truncated process will not break down.

It is easy to show (see, e.g., [34]) that in the symmetrizable case ORTHORES(s) for $s \geqslant 1$ is equivalent to ORTHORES(∞) and hence to the CG method. In fact, one can derive (5.1) from (5.3) after some algebraic manipulation; see, e.g., [8].

6. ERROR MINIMIZATION

Throughout this section we assume that ZA is SPD. We have already seen in Sec. 2 that if $u^{(1)}, u^{(2)}, \ldots$ are generated by the IGCG method starting with $u^{(0)}$, then

$$\|u^{(n)}-\bar{u}\|_{(ZA)^{1/2}} \leqslant \|w-\bar{u}\|_{(ZA)^{1/2}} \tag{6.1}$$

for any w such that $w-u^{(0)}\in K_n(r^{(0)})$. We now show that a similar error-minimization property holds for the truncated versions of ORTHODIR and ORTHOMIN.

Let us consider the truncated version, ORTHODIR(s), of ORTHODIR. We make the assumption that $q^{(0)}, q^{(1)},\dots, q^{(n)}$ do not vanish. We show that ORTHODIR(s) minimizes $\|u^{(n+1)}-\bar{u}\|_{(ZA)^{1/2}}$ where $u^{(n+1)}$ has the form

$$u^{(1)}=u^{(0)}+\hat{\lambda}_0^* r^{(0)}$$

$$u^{(n+1)}=u^{(n)}+\hat{\lambda}_n^* Aq^{(n-1)}+d_{n,n-1}q^{(n-1)}+\cdots+d_{n,n-s}q^{(n-s)}, \qquad n\geqslant 1. \tag{6.2}$$

In other words, the minimization condition leads to the same values of $\hat{\lambda}_n^*$ and $d_{n,n-1},\dots,d_{n,n-s}$ as are used for ORTHODIR(s). To do this, we consider the (in general) inconsistent system

$$\text{Ⓐ}X=\text{Ⓑ}, \tag{6.3}$$

where

$$\begin{aligned}\text{Ⓐ}&=\left(Aq^{(n-1)}, q^{(n-1)},\dots, q^{(n-s)}\right),\\ X^T&=\left(\hat{\lambda}_n^*, d_{n,n-1},\dots, d_{n,n-s}\right),\\ \text{Ⓑ}&=-\epsilon^{(n)}=-\left(u^{(n)}-\bar{u}\right).\end{aligned} \tag{6.4}$$

To minimize the $(ZA)^{1/2}$-norm of $\epsilon^{(n+1)}=u^{(n+1)}-\bar{u}=\epsilon^{(n)}+\text{Ⓐ}X$, we consider the normal equation

$$\text{Ⓐ}^T ZA\,\text{Ⓐ}X=\text{Ⓐ}^T ZA\,\text{Ⓑ}, \tag{6.5}$$

or

$$\left[\begin{array}{c|ccc} (Aq^{(n-1)}, ZA^2q^{(n-1)}) & (Aq^{(n-1)}, ZAq^{(n-1)}) & \cdots & (Aq^{(n-1)}, ZAq^{(n-s)}) \\ \hline (Aq^{(n-1)}, ZAq^{(n-1)}) & (q^{(n-1)}, ZAq^{(n-1)}) & & \\ \vdots & & \ddots & \\ (Aq^{(n-1)}, ZAq^{(n-s)}) & & & (q^{(n-s)}, ZAq^{(n-s)}) \end{array}\right]\left[\begin{array}{c}\hat{\lambda}_n^* \\ d_{n,n-1} \\ \vdots \\ d_{n,n-s}\end{array}\right]$$

$$=-\left[\begin{array}{c}(Aq^{(n-1)}, ZA\epsilon^{(n)}) \\ (q^{(n-1)}, ZA\epsilon^{(n)}) \\ \vdots \\ (q^{(n-s)}, ZA\epsilon^{(n)})\end{array}\right]=\left[\begin{array}{c}(Aq^{(n-1)}, Zr^{(n)}) \\ (q^{(n-1)}, Zr^{(n)}) \\ \vdots \\ (q^{(n-s)}, Zr^{(n)})\end{array}\right]=\left[\begin{array}{c}(Aq^{(n-1)}, Zr^{(n)}) \\ 0 \\ \vdots \\ 0\end{array}\right]. \tag{6.6}$$

We note that $(q^{(i)}, ZAq^{(j)})=0$ for $i\neq j$ by (3.40) and that $(Zr^{(n)}, q^{(j)})=0$ for $j=n-1, n-2, \ldots, n-s$ by (3.41).

The last s equations of (6.6) give

$$d_{n,i}=-\frac{(Aq^{(n-1)}, ZAq^{(i)})}{(q^{(i)}, ZAq^{(i)})}\hat{\lambda}_n^*, \qquad i=n-s, \ldots, n-1,$$

$$=\beta_{n,i}\hat{\lambda}_n^* \tag{6.7}$$

by (3.33). The first equation gives

$$\begin{aligned}(Aq^{(n-1)}, Zr^{(n)}) &= \big(Aq^{(n-1)}, ZA\big(\hat{\lambda}_n^* Aq^{(n-1)} + d_{n,n-1}q^{(n-1)} \\ &\qquad + \cdots + d_{n,n-s}q^{(n-s)}\big)\big) \\ &= \hat{\lambda}_n^*\big(Aq^{(n-1)}, ZA\big(Aq^{(n-1)} + \beta_{n,n-1}q^{(n-1)} + \cdots + \beta_{n,n-s}q^{(n-s)}\big)\big) \\ &= \hat{\lambda}_n^*(Aq^{(n-1)}, ZAq^{(n)}). \end{aligned} \tag{6.8}$$

Therefore, since $Aq^{(n-1)}=q^{(n)}-\beta_{n,n-1}q^{(n-1)}-\cdots-\beta_{n,n-s}q^{(n-s)}$ and since (3.40) and (3.41) hold, it follows by (3.33) that

$$\hat{\lambda}_n^*=\frac{(q^{(n)}, Zr^{(n)})}{(q^{(n)}, ZAq^{(n)})}=\hat{\lambda}_n. \tag{6.9}$$

Thus, by (6.7), (6.9), and (6.2) we get ORTHODIR(s). Hence ORTHODIR(s) has the desired minimization property.

Let us consider ORTHOMIN(s). In addition to assuming that ZA is SPD we also assume, as in the case of ORTHOMIN (idealized), that Z is PR. As shown in Sec. 4, this guarantees us that if $r^{(0)}, r^{(1)}, \ldots, r^{(n)}$ do not vanish, then $p^{(0)}, p^{(1)}, \ldots, p^{(n)}$ and $\lambda_0, \lambda_1, \ldots, \lambda_n$ do not vanish either. We show that ORTHOMIN(s) minimizes $\|u^{(n+1)}-\bar{u}\|_{(ZA)^{1/2}}$, where $u^{(n+1)}$ has the form

$$\begin{aligned} u^{(1)} &= u^{(0)}+\lambda_0^* r^{(0)}, \\ u^{(n+1)} &= u^{(n)}+\lambda_n^* r^{(n)}+c_{n,n-1}p^{(n-1)}+\cdots+c_{n,n-s}p^{(n-s)}, \qquad n\geqslant 1. \end{aligned} \tag{6.10}$$

Thus we show that the minimization condition leads to the same values of λ_n^* and $c_{n,n-1},\dots,c_{n,n-s}$ as are used for ORTHOMIN(s). The proof is very similar to that for ORTHODIR. The normal equations lead to the condition

$$c_{n,i}=\alpha_{n,i}\lambda_n^* \tag{6.11}$$

$$\lambda_n^*=\frac{(r^{(n)},Zr^{(n)})}{(r^{(n)},ZAp^{(n)})}. \tag{6.12}$$

But since $r^{(n)}=p^{(n)}-\alpha_{n,n-1}p^{(n-1)}-\cdots-\alpha_{n,n-s}p^{(n-s)}$, it follows from (4.21) that

$$\lambda_n^*=\frac{(p^{(n)},Zr^{(n)})}{(p^{(n)},ZAp^{(n)})}=\lambda_n. \tag{6.13}$$

Thus ORTHOMIN(s) has the desired property.

It can be shown that ORTHORES(s) does not have a minimization property analogous to that enjoyed by ORTHODIR(s) and ORTHOMIN(s); see [34].

7. GENERALIZED CG ACCELERATION OF BASIC ITERATIVE METHODS

We now consider the application of ORTHODIR, ORTHOMIN, and ORTHORES to accelerate the convergence of the basic iterative method (1.2). We are interested in the nonsymmetrizable case where no SPD matrix Z is conveniently available so that $ZQ^{-1}A$ is SPD.

By (2.9), (1.2), and (1.3) one can regard the basic iterative method (1.2) as equivalent to the RF method for solving the system

$$Q^{-1}Au=Q^{-1}b. \tag{7.1}$$

All of the discussion of Secs. 2 through 6 concerning the RF method applies to (7.1) if we replace A by $Q^{-1}A$ and if we replace the residual vectors $r^{(n)}$ by the pseudoresidual vectors $\delta^{(n)}=Q^{-1}r^{(n)}$.

For the RF method we can apply ORTHODIR provided that ZA is PR. Hence, for the method (1.2) we can apply ORTHODIR provided that $ZQ^{-1}A$ is PR. Similarly, for the RF method we can apply ORTHOMIN or ORTHORES if ZA and Z are PR. Hence for the method (1.2) we can apply ORTHOMIN and ORTHORES if $ZQ^{-1}A$ and Z are PR. The word "apply" in reference to a given procedure means that the idealized version of the procedure will converge in

at most N iterations. In the case of ORTHOMIN the truncated version will not break down. However, as described in Secs. 3 and 5, we have no guarantee, except in certain cases, that the truncated versions of ORTHODIR and ORTHORES will not break down.

Before proceeding further, we state the following result without proof.

THEOREM 7.1. *If E is a nonsingular matrix, then* ZE *is* SPD (*or* PR) *if and only if*

$$Z = E^T Y \tag{7.2}$$

for some matrix Y which is SPD (*or* PR).

From the above result it follows that to apply ORTHODIR we can let $Z = (Q^{-1}A)^T Y$ for some PR matrix Y. We can apply ORTHOMIN or ORTHORES if $YQ^{-1}A$ is PR for some PR matrix Y. For ORTHORMIN we choose $Z = (Q^{-1}A)^T Y$, and for ORTHORES we choose $Z = Y$. If Z is SPD and $ZQ^{-1}A$ is PR but not SPD, then the formulas for ORTHORES are simpler than those for ORTHODIR or ORTHOMIN, since we have full orthogonality for the $\{\delta^{(n)}\}$, with respect to Z, but not full orthogonality for the $\{q^{(n)}\}$ or the $\{p^{(n)}\}$ with respect to $ZQ^{-1}A$. On the other hand, if $ZQ^{-1}A$ is SPD and Z is PR but not SPD, then the formulas for ORTHODIR and ORTHOMIN are simpler than those for ORTHORES, since we have full orthogonality, with respect to $ZQ^{-1}A$ for the $\{q^{(n)}\}$ and the $\{p^{(n)}\}$ but not full orthogonality for the $\{\delta^{(n)}\}$ with respect to Z.

From now on we will restrict ourselves to the case where $ZQ^{-1}A$ is SPD when considering ORTHODIR and ORTHOMIN, and to the case where Z is SPD when considering ORTHORES. This means that the direction vectors $\{q^{(n)}\}$ and $\{p^{(n)}\}$ will be fully orthogonal with respect to $ZQ^{-1}A$ for ORTHODIR and ORTHOMIN, respectively, and the pseudoresidual vectors $\{\delta^{(n)}\}$ will be fully orthogonal with respect to Z for ORTHORES. From the above discussion this means that $Z = (Q^{-1}A)^T Y$ for some SPD matrix Y (for ORTHODIR), that $Z = (Q^{-1}A)^T Y$ for some SPD matrix Y such that $YQ^{-1}A$ is PR (for ORTHOMIN), and that $Z = Y$ for some SPD matrix Y such that $YQ^{-1}A$ is PR (for ORTHORES). Formulas for the three procedures are given in Table 1.

8. CHOICE OF AUXILIARY MATRIX FOR THE ALMOST SYMMETRIZABLE CASE

In this section we discuss the choice of the auxiliary matrix Z so that for the "almost symmetrizable case" the generalized CG acceleration procedures will, in some sense, behave much like standard CG acceleration for the

TABLE 1

FORMULAS FOR ORTHODIR(s), ORTHOMIN(s), AND ORTHORES(s)

Let

$$\delta^{(n)}=k-(I-G)u^{(n)}=Q^{-1}r^{(n)} \qquad \text{(pseudoresidual)},$$

$$r^{(n)}=b-Au^{(n)}. \qquad \text{(residual)}.$$

Y is an SPD matrix.

For ORTHOMIN(s) and ORTHORES(s), the matrix $YQ^{-1}A$ should be PR.

ORTHODIR(s):

$$u^{(n+1)}=u^{(n)}+\hat{\lambda}_n q^{(n)},$$

$$\hat{\lambda}_n=\frac{(Y\delta^{(n)},Q^{-1}Aq^{(n)})}{(YQ^{-1}Aq^{(n)},Q^{-1}Aq^{(n)})},$$

$$q^{(0)}=\delta^{(0)},$$

$$q^{(n)}=Q^{-1}Aq^{(n-1)}+\beta_{n,n-1}q^{(n-1)}+\cdots+\beta_{n,n-s}q^{(n-s)}, \qquad n\geq 1,$$

$$\beta_{n,i}=-\frac{\left(Y(Q^{-1}A)^2q^{(n-1)},Q^{-1}Aq^{(i)}\right)}{(YQ^{-1}Aq^{(i)},Q^{-1}Aq^{(i)})}, \qquad i=n-s,\ldots,n-1, \quad n=1,2,\ldots .$$

ORTHOMIN(s):

$$u^{(n+1)}=u^{(n)}+\lambda_n p^{(n)},$$

$$\lambda_n=\frac{(Y\delta^{(n)},Q^{-1}Ap^{(n)})}{(YQ^{-1}Ap^{(n)},Q^{-1}Ap^{(n)})},$$

$$p^{(0)}=\delta^{(0)},$$

$$p^{(n)}=\delta^{(n)}+\alpha_{n,n-1}p^{(n-1)}+\cdots+\alpha_{n,n-s}p^{(n-s)}, \qquad n\geq 1,$$

$$\alpha_{n,i}=-\frac{(YQ^{-1}A\delta^{(n)},Q^{-1}Ap^{(i)})}{(YQ^{-1}Ap^{(i)},Q^{-1}Ap^{(i)})}, \qquad i=n-s,\ldots,n-1, \qquad n=1,2,\ldots .$$

ORTHORES(s):

$$u^{(n+1)}=\gamma_{n+1}f_{n+1,n}\delta^{(n)}+f_{n+1,n}u^{(n)}+\cdots+f_{n+1,n-s}u^{(n-s)},$$

$$\sigma_{n+1,i}=\frac{(YQ^{-1}A\delta^{(n)},\delta^{(i)})}{(Y\delta^{(i)},\delta^{(i)})}, \qquad i=n-s,\ldots,n,$$

$$\gamma_{n+1}=\frac{1}{\sigma_{n+1,n}},$$

$$f_{n+1,n}=\left[1+\gamma_{n+1}\sum_{i=n-s}^{n-1}\sigma_{n+1,i}\right]^{-1},$$

$$f_{n+1,i}=\sigma_{n+1,i}\gamma_{n+1}f_{n+1,n}, \qquad i=n-s,\ldots,n-1.$$

symmetrizable case. Let us assume that the matrix A belongs to a family $\mathcal{F}$ of matrices and that for each $A \in \mathcal{F}$ there exists a unique nonsingular splitting matrix $Q = Q(A)$ corresponding to the basic iterative method (1.2). We make the following additional assumption.

ASSUMPTION I. There exists $\hat{A} \in \mathcal{F}$ such that $Q(A)$ is continuous at $\hat{A}$ and such that $HQ(\hat{A})^{-1}\hat{A}$ is SPD for some SPD matrix H.

EXAMPLE. Let us consider the usual five-point central-difference discretization of the convection-diffusion equation

$$u_{xx} + u_{yy} + \beta u_x = 0, \tag{8.1}$$

where β is a constant and where values of u are prescribed on the boundary of the square $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$. Let the mesh size h be such that h^{-1} is an integer. The difference equation obtained after multiplying by $-h^2$ is

$$4u(x,y) - \left(1 + \frac{h\beta}{2}\right)u(x+h,y) - \left(1 - \frac{h\beta}{2}\right)u(x-h,y)$$
$$- u(x,y+h) - u(x,y-h) = 0. \tag{8.2}$$

The coefficient matrix $A = A(\beta)$ corresponding to the linear system associated with (8.2) is a function of β, and $\hat{A} = A(0)$ is an SPD matrix. If the basic method is the RF method, then $Q(A) \equiv I$, which is continuous everywhere. Assumption I is satisfied with $H = I$. We remark that each $A \in \mathcal{F}$ is PR, since the matrix $\frac{1}{2}(A + A^T)$ corresponds to the differential equation $u_{xx} + u_{yy} = 0$.

For convenience, let us define the following conditions.

Condition (a). The matrix-valued function $K(A)$ satisfies Condition (a) if $K(A)$ is continuous at $\hat{A}$, if $K(A)$ is SPD for all $A \in \mathcal{F}$, and if $K(\hat{A})Q(\hat{A})^{-1}\hat{A}$ is SPD.

Condition (b). The matrix-vlaued function $K(A)$ satisfies Condition (b) if $K(A)$ is SPD and $K(A)Q(A)^{-1}A$ is PR for all $A \in \mathcal{F}$.

Because of Assumption I, the function $Y(A) \equiv H$ satisfies Condition (a). The same is true of the function $Y(A) = \frac{1}{2}(A + A^T)$ if A is PR for all $A \in \mathcal{F}$. For any choice of $Y(A)$ satisfying Condition (a) we can apply ORTHODIR with $Z = [Q(A)^{-1}A]^T Y(A)$. We can also apply ORTHOMIN with $Z = [Q(A)^{-1}A]^T Y(A)$ and ORTHORES with $Z = Y(A)$ provided that A is sufficiently close to $\hat{A}$. To see this we show that $\Gamma(A) = Y(A)Q(A)^{-1}A$ is PR for all A sufficiently close to $\hat{A}$. But since $\Gamma(\hat{A})$ is SPD, it follows that $\Gamma(\hat{A}) +$

$\Gamma(\hat{A})^T$ is SPD. For A close enough to $\hat{A}$ it follows by the continuity of the eigenvalues of $\Delta(A)=\Gamma(A)+\Gamma(A)^T$ as a function of its elements and from the symmetry of $\Delta(A)$ that the eigenvalues of $\Delta(A)$ are real and positive. Hence $\Gamma(A)$ is PR for A sufficiently close to $\hat{A}$.

If $Y(A)$ satisfies Condition (a) and if $A=\hat{A}$, then we have the symmetrizable case, since $Y(\hat{A})Q(\hat{A})^{-1}\hat{A}$ and $Y(\hat{A})$ are SPD. We refer to the case where A is close to $\hat{A}$ as the "almost symmetrizable case." The above discussion shows that all three procedures can be applied in the almost symmetrizable case. For the case $A=\hat{A}$ each of the methods ORTHODIR(2), ORTHOMIN(1), and ORTHORES(1) is equivalent to CG acceleration. For A close to $\hat{A}$ one might expect that each of these three schemes would converge almost as rapidly as for the case where $A=\hat{A}$. If not, one might consider trying such methods as ORTHODIR(3), ORTHOMIN(2), and ORTHORES(2).

Suppose now that A is not close to $\hat{A}$. By Assumption I, the choice $Y(A)\equiv H$ satisfies Condition (a). We can therefore apply ORTHODIR. In order to apply ORTHOMIN and ORTHORES we need to find a function $Y(A)$ satisfying Condition (b). This is not always possible. Ideally, we would like to find $Y(A)$ satisfying both Conditions (a) and (b).

Let us consider the following special cases.

(1) *A is* PR *and $Q(A)$ is* SPD *for all $A\in\mathcal{F}$; $\hat{A}$ is* SPD. *If we choose $Y(A)\equiv Q(A)$, then Conditions (a) and (b) hold.*

(2) *A is* PR *and $Q(A)$ is* SPD *for all $A\in\mathcal{F}$; $HQ(\hat{A})^{-1}\hat{A}$ is* SPD *for some* SPD *matrix H, but $\hat{A}$ is not* SPD. *If we choose $Y(A)\equiv Q(A)$, then Condition (b) holds but not Condition (a) in general. If we choose $Y(A)\equiv H$, then Condition (a) holds but Condition (b) need not hold. However, $Y(A)Q(A)^{-1}A$ will be PR for A sufficiently close to $\hat{A}$.*

(3) *A is* SPD *and $Q(A)$ is* PR *for all A; $Q(\hat{A})$ is* SPD. *If we choose $Y(A)\equiv A$, then Conditions (a) and (b) hold.*

(4) *A is* SPD *and $Q(A)$ is* PR *for all A; $HQ(\hat{A})^{-1}\hat{A}$ is* SPD *for some* SPD *matrix H, but $Q(\hat{A})$ is not* SPD. *If we choose $Y(A)\equiv A$, then Condition (b) holds. However, Condition (a) need not hold. If we choose $Y(A)\equiv H$, then Condition (a) holds but Condition (b) need not hold. However, $Y(A)Q(A)^{-1}A$ will be PR for A sufficiently close to $\hat{A}$.*

9. APPLICATION OF GENERALIZED CG ACCELERATION TO SPECIFIC ITERATIVE METHODS

In this section we discuss the application of the generalized CG acceleration procedures described above to specific basic iterative methods. We assume that A is nonsingular and, in nearly every case, we assume that Assumption I holds with $H=I$.

For convenience we introduce the following notation. Given a matrix A, let $D=D(A)$ be the diagonal matrix with the same diagonal elements as A, and let

$$A=D-C_L-C_U, \tag{9.1}$$

where $C_L=C_L(A)$ and $C_U=C_U(A)$ are strictly lower triangular and strictly upper triangular matrices, respectively.

In each case we seek to determine a matrix function $Y(A)$ of A which is SPD for all A. We can then use ORTHODIR. However, in order that for A close to $\hat{A}$ the acceleration procedures will behave like standard CG acceleration, we desire that $Y(A)$ satisfy Condition (a) and if possible Condition (b) as well. If Condition (a) holds, then we can use ORTHODIR or, if A is sufficiently close to $\hat{A}$, we can use ORTHOMIN or ORTHORES. For A close enough to $\hat{A}$ we might try ORTHODIR(2), ORTHOMIN (1), or ORTHORES(1), since if $A=\hat{A}$ these procedures would reduce to standard CG acceleration. If Condition (b) holds, we can use any of the three methods; however, if Condition (a) does not hold, then the procedure will not behave like CG acceleration in general, even if A is very close to $\hat{A}$.

The RF Method

Let us assume that Assumption I holds but that $\hat{A}$ is not necessarily SPD. In the general case Condition (a) holds for $Y(A)\equiv H$. If A is PR, we can satisfy Condition (b), but not Condition (a) in general, by letting $Y(A)\equiv I$. If A is PR for all $A\in\mathcal{F}$ and if $\hat{A}$ is SPD, then we can satisfy both Conditions (a) and (b) by choosing $Y(A)\equiv I$.

The Jacobi Method

We assume that for all matrices A of the family $\mathcal{F}$ all diagonal elements of A are positive and that $\hat{A}$ is SPD. This is true for the example (8.2). Hence the splitting matrix $Q(A)=D(A)$ for the Jacobi method is SPD. Evidently, Condition (a) holds if $Y(A)\equiv\hat{A}$, $Y(A)\equiv D(A)$, $Y(A)\equiv D(\hat{A})$, or—given that A is PR for all $A\in\mathcal{F}$—if $Y(A)=\frac{1}{2}(A+A^T)$. If A is PR for all $A\in\mathcal{F}$, then the choice $Y(A)\equiv D(A)$ will satisfy both Conditions (a) and (b).

The RF Method for the Reduced System

Let us now consider the case where the matrix A can be written in the partitioned form

$$A=\begin{bmatrix} D_R & H \\ K & D_B \end{bmatrix}, \tag{9.2}$$

where D_R and D_B are diagonal matrices. Such a matrix is often referred to as a "redblack" matrix. The example given in Sec. 8 yields this form of A if we use the "redblack ordering"; see [31]. We also assume that the diagonal elements of A are positive for all $A \in \mathcal{F}$ and that $\hat{A}$ is SPD. If we partition the vectors u and b in accordance with the partitioning of A, then we have

$$\begin{pmatrix} D_R & H \\ K & D_B \end{pmatrix} \begin{pmatrix} u_R \\ u_B \end{pmatrix} = \begin{pmatrix} b_R \\ b_B \end{pmatrix}, \tag{9.3}$$

which is equivalent to

$$\begin{aligned} u_R &= F_R u_B + c_R, \\ u_B &= F_B u_R + c_B, \end{aligned} \tag{9.4}$$

where

$$\begin{aligned} F_R &= -D_R^{-1} H, \qquad c_R = D_R^{-1} b_R, \\ F_B &= -D_B^{-1} K, \qquad c_B = D_B^{-1} b_B. \end{aligned} \tag{9.5}$$

Eliminating u_R, we get the *reduced system*

$$(I_B - F_B F_R) u_B = F_B c_R + c_B, \tag{9.6}$$

or

$$\tilde{A} u_B = \tilde{b}, \tag{9.7}$$

where I_B is an identity matrix and

$$\begin{aligned} \tilde{A} &= I_B - F_B F_R, \\ \tilde{b} &= F_B c_R + c_B. \end{aligned} \tag{9.8}$$

Evidently, the matrix $\tilde{A} = \tilde{A}(A)$ belongs to a family $\tilde{\mathcal{F}}$ of matrices which contains the matrix $\tilde{A}(\hat{A}) = \hat{\tilde{A}}$. Moreover, $\tilde{A}(A) \to \hat{\tilde{A}}$ as $A \to \hat{A}$. We let $Y(\tilde{A}) = D_B(A)$. Since $Y(\hat{\tilde{A}})\hat{\tilde{A}} = D_B(\hat{A})\hat{\tilde{A}}$ is SPD, it follows that $Y(\tilde{A})$ satisfies Condition (a).

The SSOR Method

We again assume that for all matrices A in the family $\mathcal{F}$, all diagonal elements of A are positive and that $\hat{A}$ is SPD.

The splitting matrix for the SSOR method is

$$Q=\frac{\omega}{2-\omega}\left(\frac{1}{\omega}D-C_L\right)D^{-1}\left(\frac{1}{\omega}D-C_U\right). \tag{9.9}$$

(See, e.g., [31].) Here it is assumed that the relaxation factor ω lies in the interval $0<\omega<2$. One possible choice of $Y(A)$ is

$$Y(A)=\frac{\omega}{2-\omega}\left(\frac{1}{\omega}D-C_U\right)^T D^{-1}\left(\frac{1}{\omega}D-C_U\right). \tag{9.10}$$

This has the property that $Y(A)$ is SPD for all $A\in\mathcal{F}$ and that $Y(\hat{A})=Q(\hat{A})$. Moreover, $Q(\hat{A})$ is SPD. Thus $Y(\hat{A})Q(\hat{A})^{-1}\hat{A}$ is SPD. Therefore, $Y(A)$ satisfies Condition (a).

We remark that if A is PR, one might consider using the splitting matrix

$$Q'=Q'(A)=\frac{\omega}{2-\omega}\left(\frac{1}{\omega}D-C_L(A_S)\right)D^{-1}\left(\frac{1}{\omega}D-C_U(A_S)\right), \tag{9.11}$$

where

$$A_S=\tfrac{1}{2}(A+A^T). \tag{9.12}$$

Evidently, Q' is SPD and we can apply generalized CG acceleration with $Y(A)=Q'(A)$. Since $Y(A)Q'(A)^{-1}A$ is PR for all A, it follows that Conditions (a) and (b) hold. Research and numerical experimentation are needed to determine how the convergence rate of the SSOR method would be affected by this modification.

Incomplete Facotrization Methods

Incomplete factorization methods involve the use of a splitting matrix $Q=Q(A)$ of the form

$$Q=Q(A)=L(A)U(A) \tag{9.13}$$

where $L(A)$ is lower triangular and $U(A)$ is upper triangular. The matrices $L(A)$ and $U(A)$ are chosen so that $L(A)U(A)$ is approximately equal to A.

Let us assume that for all A in $\mathcal{F}$ no diagonal element of $L(A)$ or $U(A)$ vanishes and that $\hat{A}$ is SPD. We can thus write

$$Q=L_1(A)D_1(A)U_1(A), \tag{9.14}$$

where $L_1(A)$ is unit lower triangular, $D_1(A)$ is diagonal, and $U_1(A)$ is unit upper triangular.

We assume that our procedure for choosing $L(A)$ and $U(A)$ is continuous at $\hat{A}$ and that $L(\hat{A})=U(\hat{A})^T$, so that $\hat{Q}=Q(\hat{A})$ is SPD. We can then apply generalized CG acceleration with $Y(A)=\hat{A}$, with $Y(A)=\hat{Q}$, or with

$$Y(A)=U_1(A)^TD_1(A)U_1(A). \tag{9.15}$$

In each case Condition (a) holds but Condition (b) may not hold.

As in the case of the SSOR method, if A is PR, one might consider using the splitting matrix

$$Q'=Q'(A)=L(A_S)U(A_S), \tag{9.16}$$

where $L(A_S)=U(A_S)^T$ and where $L(A_S)U(A_S)$ is an approximate factorization of the SPD matrix $A_S=\frac{1}{2}(A+A^T)$. We can apply generalized CG acceleration with $Y(A)=Q'(A)$; Conditions (a) and (b) will hold.

The GCW Method

Here we assume that A is PR for all $A\in\mathcal{F}$ and that $\hat{A}$ is SPD. Let the splitting matrix Q be given by

$$Q=Q(A)=\tfrac{1}{2}(A+A^T). \tag{9.17}$$

Evidently, $Q(A)$ is a continuous function of A, $Q(A)$ is SPD for all A, and $\hat{Q}=Q(\hat{A})=\hat{A}$. This splitting was considered by Concus and Golub [7] and by Widlund [29].

We can apply generalized CG acceleration to the GCW method with $Y(A)\equiv Q(A)$. Evidently, Conditions (a) and (b) hold.

Let us consider the application of the idealized version of ORTHORES to the GCW method. It is easy to show that for all n we have $\gamma_{n+1}=1$ and $f_{n+1,i}=0$ for $i=n-2,n-3,\ldots,0$. Hence, the idealized procedure is the

same as ORTHORES(1). Moreover, we can derive the following formulas:

$$u^{(n+1)}=\rho_{n+1}(\delta^{(n)}+u^{(n)})+(1-\rho_{n+1})u^{(n-1)},$$

$$\rho_{n+1}=\begin{cases}1, & n=0,\\ \left[1-\dfrac{(\delta^{(n)},Q\delta^{(n)})}{(\delta^{(n-1)},Q\delta^{(n-1)})}\dfrac{1}{\rho_n}\right]^{-1}, & n\geqslant 1.\end{cases} \tag{9.18}$$

(Here we let $f_{n+1,n}=\rho_{n+1}$, $f_{n+1,n-1}=1-\rho_{n+1}$.) This method is thus equivalent to the procedure described by Concus and Golub [7] and by Widlund [29].

We remark that Widlund [29] derived the procedure by requiring the pseudoresiduals to be pairwise orthogonal with respect to the matrix Q. Hence, the above result is by no means unexpected.

An alternative derivation of (9.18) which is given by Hageman, Luk, and Young [17, 18] is based on the "double" GCW method, which has G^2 as its iteration matrix, where $G=I-Q^{-1}A$, and which is symmetrizable. Convergence estimates can be made by considering Chebyshev acceleration of the double method.

The Peaceman-Rachford Alternating-direction Implicit Method

In the Peaceman-Rachford method we represent the matrix A in the form

$$A=H+V, \tag{9.19}$$

where H and V are symmetric and positive semidefinite matrices. The Peaceman-Rachford method involves the choice of a positive parameter ρ and the use of the splitting matrix

$$Q=\frac{1}{2\rho}(H+\rho I)(V+\rho I). \tag{9.20}$$

If A is SPD and if $HV=VH$ (the commutative case), then Q is SPD and the method is symmetrizable with $Z=A$ or $Z=Q$. However, in general $HV\neq VH$ even when A is SPD.

We now consider the application of generalized CG acceleration to the Peaceman-Rachford method. We assume that we have a family $\mathcal{F}$ of SPD matrices A such that the SPD matrices $H(A)$ and $V(A)$ are defined for all

$A \in \mathcal{F}$. We also assume that for some $\hat{A} \in \mathcal{F}$, $H(\hat{A})V(\hat{A}) = V(\hat{A})H(\hat{A})$ and that $H(A)$ and $V(A)$ are continuous at $\hat{A}$. Evidently $Q(A)$ is continuous at $\hat{A}$ and $Q(\hat{A})$ is SPD.

If $Y(A) \equiv \hat{A}$, $Y(A) \equiv Q(\hat{A})$ or if $Y(A) \equiv A$, then Condition (a) holds but not, in general, Condition (b).

As an example, consider the solution of the Dirichlet problem for the differential equation

$$(1+\beta^2 x^2)u_{xx} + u_{yy} = 0 \tag{9.21}$$

in the square $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$. Here β is a real number. We use the customary five-point discretization based on the use of central differences with mesh size h where h^{-1} is an integer. We let $H = H(A)$ and $V = V(A)$ be matrices corresponding to the discretizations of $(1+\beta^2 x^2)u_{xx}$ and u_{yy}, respectively (after multiplying by $-h^2$). Evidently, $H(A)$ and $V(A)$ are SPD for all β and vary continuously with β. Moreover, $\hat{A} = A(0)$ and $H(\hat{A})V(\hat{A}) = V(\hat{A})H(\hat{A})$.

The preparation of this paper was greatly aided by a number of discussions held over the past few years with Owe Axelsson, Paul Concus, Stanley Eisenstat, H. Elman, Gene Golub, Louis Hageman, Tom Manteuffel, Martin Schultz, and Olof Widlund. The help received from these discussions is gratefully acknowledged. Extensive use has been made of papers by Axelsson [2–5], Concus, Golub, and O'Leary [7, 8], Eisenstat, Elman, and Schultz [12], Vinsome [28], and many others. The helpful and constructive suggestions and criticisms of the referee are also gratefully acknowledged.

REFERENCES

1 O. Axelsson, On preconditioning and convergence acceleration in sparse matrix problems, CERN 74-10, European Organization for Nuclear Research (CERN), Data Handling Division, 1974.

2 O. Axelsson, Solution of linear systems of equations: Iterative methods, in *Sparse Matrix Techniques* (V. A. Barker, Ed.), Lecture Notes in Mathematics #572, Springer, Copenhagen, 1976.

3 O. Axelsson, On preconditioned conjugate gradient methods, CNA-143, Center for Numerical Analysis, Univ. of Texas, Austin, 1978.

4 O. Axelsson, Conjugate gradient type methods for unsymmetric and inconsistent systems of linear equations, Report 78.03R, Dept. of Computer Sciences, Chalmers Univ. of Technology, Göteborg, Sweden; *Linear Algebra and Appl.*, to appear.

5 O. Axelsson, A generalized conjugate gradient direction method and its application on a singular perturbation problem, *Proceedinngs, 8th Biennial Numerical Analysis Conference*, Dundee, Scotland, June 26–29, 1979 (G. A. Watson, Ed.), Lecture Notes in Mathematics #773, Springer, Berlin, 1980.

6 A. Blair, N. Metropolis, J. von Neumann, A. H. Taub, and M. Tsingori, A study of a numerical solution of a two-dimensional hydrodynamical problem, *Math. Tables Aids Comput.* 13:145–184 (1979).

7 Paul Concus and Gene H. Golub, A generalized conjugate gradient method for nonsymmetric systems of linear equations, Report Stan-CS-76-535, Computer Science Dept., Stanford Univ., Calif., 1976.

8 P. Concus, G. H. Golub, and D. P. O'Leary, A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations, in *Sparse Matrix Computation* (J. R. Bunch and D. J. Rose, Eds.), Academic, 1976, pp. 309–332.

9 J. W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, Ph.D. thesis, Stanford Univ., Calif., 1965.

10 J. W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, *SIAM J. Numer. Anal.* 4:10–26 (1967).

11 S. C. Eisenstat, H. Elman, M. H. Schultz, and A. H. Sherman, Solving approximations to the convection diffusion equation, in *Proceedings of the Society of Petroleum Engineers of the* AIME, Fifth Symposium on Reservoir Simulation, Denver, Colo., 1979.

12 S. C. Eisenstat, H. Elman, and M. H. Schultz, Variational iterative methods for nonsymmetric systems of linear equations, unpublished.

13 M. Engeli, M. Ginsburg, H. Rutishauser, and E. Stiefel, Refined iterative methods for the computation of the solution and the eigenvalues of self-adjoint boundary value problems, *Mitt. Inst. Angew. Math. ETH, Zürich*, No. 8, Basel-Stuttgart, 1959.

14 D. K. Faddeev and V. N. Faddeeva, *Computational Methods of Linear Algebra*, Freeman, San Francisco, 1963.

15 R. Fletcher, Conjugate gradient methods for indefinite systems, Lecture Notes in Mathematics 506, Springer, 1976.

16 G. H. Golub and R. S. Varga, Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second-order Richardson iterative methods, Parts I, II, *Numer. Math.*, 3:147–168 (1961).

17 L. A. Hageman, F. Luk, and D. M. Young, On the acceleration of iterative methods: Preliminary report, CNA-129, Center for Numerical Analysis, Univ. of Texas, Austin, 1977.

18 L. A. Hageman, F. Luk, and D. M. Young, On the equivalence of certain iterative acceleration methods, *SIAM J. Numer. Anal.*, to appear.

19 L. A. Hageman and David M. Young, *Applied Iterative Methods*, Academic, 1981.

20 M. R. Hestenes and E. L. Stiefel, Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bur. Standards* 49:409–436 (1952).

21 Magnus R. Hestenes, The conjugate-gradient method for solving linear systems, in *Numberical Analysis*, Vol. VI (John Curtiss, Ed.), McGraw-Hill, New York, 1956.

22 C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bur. Standards* 45:255–282 (1950).

23 C. Lanczos, Solution of systems of linear equations by minimized iterations, *J. Res. Nat. Bur. Standards* 49:33–53 (1952).

24 T. A. Manteuffel, The Tchebyshev iteration for nonsymmetric linear systems, *Numer. Math.* 28:307–327 (1977).

25 J. K. Reid, On the method of conjugate gradients for the solution of large sparse systems of linear equations, in *Proceedings of the Conference on Large Sparse Sets of Linear Equations*, Academic, New York, 1971, pp. 231–254.

26 J. K. Reid, The use of conjugate gradients for systems of linear equations possessing "Property A", *SIAM J. Numer. Anal.*9:325–332 (1972).

27 L. F. Richardson, The approximate arithmetical solution by finite differences of physical problems involving differential equations with an application to the stresses in a masonry dam, *Philos. Trans. Roy. Soc. London Ser. A* 210:307–357 (1910).

28 P. K. W. Vinsome, ORTHOMIN, an iterative method for solving sparse sets of simultaneous linear equations, Paper SPE 5729, 4th Symposium of Numerical Simulation of Reservoir Performance of the Society of Petroleum Engineers of the AIME, Los Angeles, 19–20 February 1976.

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

30 Y. S. Wong, Conjugate gradient type methods for unsymmetric matrix problems, Technical Report No. 79-36, The Institute of Applied Mathematics and Statistics, Univ. of British Columbia, Canada, 1979.

31 David M. Young, *Iterative Solution of Large Linear Systems*, Academic, New York, 1971.

32 David M. Young, and Kang C. Jea, Some generalizations of conjugate gradient acceleration for nonsymmetrizable iterative methods (preliminary report), CNA-149, Center for Numerical Analysis, Univ. of Texas, Austin, 1979.

33 David M. Young, Linda Hayes, and Kang C. Jea, Conjugate gradient acceleration of iterative methods: Part I, the symmetrizable case, CNA-162, Center for Numerical Analysis, Univ. of Texas, Austin, 1980.

34 David M. Young and Kang C. Jea, Conjugate gradient acceleration of iterative methods: Part II, the nonsymmetrizable case, CNA-163, Center for Numerical Analysis, Univ. of Texas, Austin, 1980.

Received January 1980; revised 2 May 1980