MATHEMATICS OF COMPUTATION
Volume 69, Number 232, Pages 1559–1575
S 0025-5718(99)01171-0
Article electronically published on August 19, 1999

ANALYSIS OF THE FINITE PRECISION BI-CONJUGATE GRADIENT ALGORITHM FOR NONSYMMETRIC LINEAR SYSTEMS

CHARLES H. TONG AND QIANG YE

ABSTRACT. In this paper we analyze the bi-conjugate gradient algorithm in finite precision arithmetic, and suggest reasons for its often observed robustness. By using a tridiagonal structure, which is preserved by the finite precision bi-conjugate gradient iteration, we are able to bound its residual norm by a minimum polynomial of a perturbed matrix (i.e. the residual norm of the exact GMRES applied to a perturbed matrix) multiplied by an amplification factor. This shows that occurrence of near-breakdowns or loss of biorthogonality does not necessarily deter convergence of the residuals provided that the amplification factor remains bounded. Numerical examples are given to gain insights into these bounds.

1. INTRODUCTION

Since its introduction by Lanczos [16] and later re-discovery by Fletcher [7] in its present form, the bi-conjugate gradient (BiCG) algorithm has evolved many variations (e.g. CGS, BiCGSTAB, QMR, CSBCG [22, 25, 8, 2]), each of which was specially designed to overcome some of its inherent difficulties (the need for adjoint matrix vector product, potential breakdowns, erratic convergence behavior, etc.). However, it has been observed by Bank and Chan [2] and Tong [23] that, in many cases, BiCG may still be competitive (in terms of convergence and convergence rates), especially when coupled with no or relatively poor preconditioners.

One major concern in using BiCG is two types of potential breakdown problems, which can cause numerical instability. In addition, in finite precision arithmetic the biorthogonality is lost, as is experienced by other Lanczos-type algorithms. As a result, the finite precision BiCG iteration can deviate significantly from the exact one. However, in many cases where such difficulties arise, BiCG exhibits exceptional numerical robustness in practice as far as the convergence of the residual norm is concerned [2, 23]. Specifically, very often, occurrence of near-breakdowns or loss of biorthogonality does not deter convergence of the residuals. On the other

Received by the editor October 6, 1998.
1991 *Mathematics Subject Classification.* Primary 65F10, 65N20.
Key words and phrases. Bi-conjugate gradient algorithm, error analysis, convergence analysis, nonsymmetric linear systems.
The first author's research was supported by Research Grant Council of Hong Kong.
The second author's research was supported by Natural Sciences and Engineering Research Council of Canada. Part of this work was completed while this author visited Stanford University during the summer of 1995. He would like to thank Professor Gene Golub for providing this opportunity and for his great hospitality.

hand, it has been observed by Golub and Overton [9, 10] that the preconditioned conjugate gradient method with inexact preconditioner, which amounts to relatively large perturbations to the CG recurrence, could still converge. These phenomena suggest that the residual convergence property of BiCG (or CG) may be relatively insensitive to perturbations in the recurrence, even though many other properties such as biorthogonality are sensitive to them. However, there is no theoretical result to explain such robustness. It is the purpose of the present paper to study the sometimes surprising convergence behavior of BiCG in finite precision arithmetic. We remark that our intention is not to suggest that BiCG will be robust in all cases, as there are plenty of divergent examples of BiCG, but rather to understand when and why BiCG iterations show more stability than expected.

In exact arithmetic, bounds on approximation errors (or residuals) of the BiCG iteration have been obtained by Bank and Chan [2], showing the fast decrease of the errors under certain conditions. Furthermore, it was also shown recently by Barth and Manteufel [3] that the BiCG residual indeed gives optimal approximation from Krylov subspaces, considered in a certain metric. Since proving these results [2] relies on the Galerkin condition (i.e., bi-orthogonality) of BiCG, which is usually lost in finite precision arithmetic, it is difficult to apply or generalize these analyses to the finite precision case. Furthermore, occurrence of near-breakdown could cause BiCG to generate completely different sequences for exact arithmetic and for a finite precision arithmetic. Nevertheless, we will show in this paper that the convergence property of residuals may still be preserved.

We shall prove *a posteriori* residual bounds similar to those in [2] for the exact case, using an approach that is based on a tridiagonal structure implicit in the algorithm. This approach was also used by Ye [27] to analyze convergence of the Lanczos algorithms for eigenvalue problems. An advantage of analyzing BiCG using its tridiagonal structure is that our results include the finite precision case and the near-breakdown case, and explain its sometimes observed convergence under quite general conditions.

Finite precision analyses of conjugate gradient-type and Lanczos-type algorithms have played an important role in understanding these algorithms. The pioneering work is due to C. Paige [19, 20] and A. Greenbaum [12]. Paige showed in [19, 20] that the loss of orthogonality comes with but does not prevent convergence of the Ritz values, i.e., useful results can still be obtained from the algorithm even when the iterates deviate significantly from what would have been produced in exact arithmetic. A generalization to the nonsymmetric case was given by Bai [1], and the near-breakdowns and loss of biorthogonality were discussed by Day [5]. Greenbaum established backward stability results in a generalized sense [12], showing that the iterative residuals produced by the finite precision conjugate gradient algorithm are equivalent to what would have been produced by applying the exact CG to a larger matrix. Some estimates on the larger matrix were given, which may vary from step to step. It would be interesting to see if Greenbaum's backward stability results can be generalized to BiCG; we are not aware of any such generalization. One analysis is given by Cullum and Greenbaum [4], which relates BiCG type methods to QMR. We note that a recent work by Greenbaum, Druskin and Knizhnerman [15] on the finite precision CG also uses the approach of bounding the residuals. Other recent works on the finite precision CG include [13, 14, 18, 25].

The paper is organized as follows. In section 2, we review the BiCG algorithm and discuss its theoretical properties. We then present our results in section 3, with

a roundoff error analysis in section 3.1, then various approximation bounds on the BiCG residuals in sections 3.2 and 3.3. We present our numerical experiments in section 4 and concluding remarks in section 5.

Notation. We shall use the standard notation in numerical analysis. In particular, for the roundoff error analysis, absolute values and inequalities of matrices are componentwise, i.e., $|(a_{ij})| = (|a_{ij}|)$, and $(a_{ij}) < (b_{ij})$ means $a_{ij} < b_{ij}$. $A \otimes B$ denotes the Kronecker product of A and B, and $\lambda(A)$ denotes the spectrum of A. A^+ denotes the Moore-Penrose generalized inverse of A. I_n is the $n \times n$ identity matrix $[e_1, \ldots, e_n] = I_n$. $\mathcal{P}_n$ denotes the set of polynomials of degree not exceeding n.

2. The BiCG algorithm

The BiCG algorithm for solving the linear system $Ax = b$ is implemented as a pair of coupled two-term recurrences as follows.

Input initial approximation x_1;
Initialize $r_1 = p_1 = \tilde{r}_1 = \tilde{p}_1 = b - Ax_1$;
$\rho_1 = \tilde{r}_1^T r_1$;
For $n = 1, 2, \cdots\cdots$
$\quad \sigma_n = \tilde{p}_n^T A p_n$
$\quad \alpha_n = \frac{\rho_n}{\sigma_n}$
$\quad r_{n+1} = r_n - \alpha_n A p_n$
$\quad x_{n+1} = x_n + \alpha_n p_n$
$\quad \tilde{r}_{n+1} = \tilde{r}_n - \alpha_n A^T \tilde{p}_n$
$\quad \rho_{n+1} = \tilde{r}_{n+1}^T r_{n+1}$;
$\quad \beta_{n+1} = \frac{\rho_{n+1}}{\rho_n}$
$\quad p_{n+1} = r_{n+1} + \beta_{n+1} p_n$
$\quad \tilde{p}_{n+1} = \tilde{r}_{n+1} + \beta_{n+1} \tilde{p}_n$
end for

The algorithm breaks down if either $\sigma_n = 0$ (called *pivotal breakdown*) or $\rho_n = 0$ (*breakdown in the underlying Lanczos process*). Instability may be expected when the iteration is close to breakdown (called *near-breakdown*). We shall assume in this paper that no exact breakdown occurs.

The sequence generated by the algorithm satisfies the following biorthogonality and biconjugacy conditions:

$$\tilde{r}_n^T r_m = 0, \quad \tilde{p}_n^T A p_m = 0, \quad \text{for} \quad m \neq n.$$

r_{n+1} is called the computed residual, and in exact arithmetic, it is equal to the true residual $b - Ax_{n+1}$. The biorthogonality of the residuals implies that r_{n+1} is orthogonal to the dual Krylov subspace

$$K_n(A^T, r_1) = \text{span}\{r_1, A^T r_1, \cdots, (A^T)^{n-1} r_1\}.$$

This is also called the Galerkian condition.

Writing $R_n = [r_1, \cdots, r_n]$ and $P_n = [p_1, \cdots, p_n]$, we can derive from the recurrence the following matrix relations:

$$AP_n = R_n L_n \Lambda_N^{-1} - \frac{1}{\alpha_n} r_{n+1} e_n^T \quad \text{and} \quad R_n = P_n U_n, \tag{1}$$

where $\Lambda_n = \mathrm{diag}[\alpha_1, \cdots, \alpha_n], e_n^T = [0, \ldots, 0, 1]$ and

(2)
$$L_n = \begin{pmatrix} 1 & & & & & \\ -1 & 1 & & & & \\ & -1 & \cdot & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & 1 & \\ & & & & -1 & 1 \end{pmatrix}, \qquad U_n = \begin{pmatrix} 1 & -\beta_2 & & & & \\ & 1 & -\beta_3 & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & 1 & -\beta_n \\ & & & & & 1 \end{pmatrix}.$$

Combining the two equations in (1), we obtain the governing equation

$$AR_n = R_n \hat{T}_n - \frac{1}{\alpha_n} r_{n+1} e_n^T, \tag{3}$$

where $\hat{T}_n = L_n \Lambda_n^{-1} U_n$ is an invertible tridiagonal matrix such that

$$e_n^T \hat{T}_n^{-1} e_1 = e_n^T U_n^{-1} \Lambda_n L_n^{-1} e_1 = e_n^T \Lambda_n v = \alpha_n,$$

where $v = [1 \quad 1 \quad \cdots \quad 1]^T$. Similar results hold for the dual sequences $\tilde{p}_n$ and $\tilde{r}_n$.

3. Analysis of the finite precision BiCG

In this section we present the main results in several subsections. We begin with an outline of the main ideas of this paper.

It is well known that the biorthogonality (or orthogonality in the symmetric case) in a Lanczos-type algorithm is usually lost in finite precision arithmetic, and thus the iterates computed in a finite precision arithmetic may differ significantly from the corresponding exact quantities. In addition to this difficulty, BiCG may encounter near-breakdowns, and as a result, large roundoff errors may occur at two possible places. One is in the computation of α's and β's, and the other is in the computation of r's and p's. However, the errors in the computation of α's and β's do not translate into errors in the governing equation (3), which depends on computations in the local vector recurrence only, not on how accurate the α's and β's are. This is demonstrated in Sec. 3.1 by showing that (3) is still valid to within a small perturbation. So roundoff errors may cause the iterates to deviate from the exact ones and loss of biorthogonality; but they may not destroy the underlying tridiagonal structure in the BiCG algorithm. The importance of (3) was originally pointed out by Paige for the Lanczos algorithm and by Greenbaum for the CG algorithm.

We claim that the existence of this tridiagonal structure also plays a major role in the convergence of the BiCG residuals in a finite precision arithmetic. We shall prove in Sec. 3.2 some bounds on the computed residual r_n based on a perturbed (3) (see equation (11) below). The bounds are of *a posteriori* type and are in terms of the residual norm of the exact GMRES applied to a perturbed matrix multiplied by an amplification factor.

We mention that in finite precision arithmetic, the computed residuals r_n differ from the true residual $b - Ax_n$; however, they have the same magnitudes before roundoff error accumulation dominates (cf. [12, Theorem 2]). In investigating the robustness of BiCG, we are only interested in the convergence of r_n because it is the convergence of r_n that drives the convergence of the true residual $b - Ax_n$. This is also consistent with the analysis of the CG case [12].

3.1. **Roundoff error analysis of BiCG.** We provide in this section a roundoff error analysis of the BiCG algorithm in finite precision arithmetic. A similar analysis for the Lanczos algorithm has been given in [1]. Our analysis is based on the following simplified model of roundoff errors in basic matrix computations [11, p. 66] (recall that inequalities are componentwise)

$$fl(\alpha x + y) = \alpha x + y + \mathbf{u}e \quad \text{with} \quad |e| \le 2|\alpha x| + |y| + O(\mathbf{u}), \tag{4}$$

$$fl(Ax) = Ax + \mathbf{u}g \quad \text{with} \quad |g| \le N|A||x| + O(\mathbf{u}), \tag{5}$$

where $\mathbf{u}$ is the machine precision unit, $x, y \in R^N$, and $\alpha \in R$. Note that $O(\mathbf{u})$ denotes a term containing $\mathbf{u}$ and can be bounded rigorously.

For ease of notation, we shall use $r_n, x_n, p_n, \alpha_n, \beta_n$ etc. directly to denote the computed quantities in finite precision arithmetic in the rest of the paper.

Theorem 3.1. *Let* $\mathbf{u}$ *be the machine precision unit and let* $r_n, x_n, p_n, \alpha_n, \beta_n$ *be the computed quantities in the finite precision BiCG algorithm. Then*

$$AR_n = R_n\hat{T}_n - \frac{1}{\alpha_n}r_{n+1}e_n^T + \mathbf{u}\hat{\Delta}_n, \tag{6}$$

where $R_n = [r_1, \cdots, r_n]$, $\hat{T}_n = L_n\Lambda_n^{-1}U_n$ *(given in (2)) and* $\hat{\Delta}_n = [\delta_1, \cdots, \delta_n]$ *with*

$$|\delta_i| \le ((N+6)|A| + \frac{1}{|\alpha_i|} + \frac{|\beta_i|}{|\alpha_{i-1}|})|r_i| + (2N+7)|A||p_i| + O(\mathbf{u}). \tag{7}$$

Proof. At the nth iteration, to compute r_{n+1}, we first compute Ap_n and have $fl(Ap_n) = Ap_n + g$ with $|g| \le \mathbf{u}N|A||p_n| + O(\mathbf{u}^2)$ by (5). Then

$$\begin{aligned} r_{n+1} &= fl(r_n - \alpha_n fl(Ap_n)) = r_n - \alpha_n fl(Ap_n) + g' \\ &= r_n - \alpha_n Ap_n + \alpha_n g + g', \end{aligned}$$

where, by (4),

$$|g'| \le \mathbf{u}(|r_n| + 2|\alpha_n||fl(Ap_n)| + O(\mathbf{u})).$$

Letting $\delta_{r_n} = (\alpha_n g + g')/(\mathbf{u}|\alpha_n|)$, we obtain

$$\frac{1}{\alpha_n}(r_{n+1} - r_n) = -Ap_n + \mathbf{u}\delta_{r_n} \tag{8}$$

with $|\delta_{r_n}| \le N|A||p_n| + |r_n|/|\alpha_n| + 2|Ap_n| + O(\mathbf{u}) \le |r_n|/|\alpha_n| + (N+2)|A||p_n| + O(\mathbf{u})$. Similarly

$$p_{n+1} = r_{n+1} + \beta_{n+1}p_n + \mathbf{u}\delta_{p_{n+1}} \tag{9}$$

with $|\delta_{p_{n+1}}| \le |r_{n+1}| + 2|\beta_{n+1}||p_n| + O(\mathbf{u})$. Writing $\Delta_{R_n} = [\delta_{r_1}, \cdots, \delta_{r_n}]$ and $\Delta_{P_n} = [0, \delta_{p_2}, \cdots, \delta_{p_n}]$, we obtain from the above

$$AP_n = R_nL_n\Lambda_N^{-1} - \frac{1}{\alpha_n}r_{n+1}e_n^T + \mathbf{u}\Delta_{R_n} \quad \text{and} \quad R_n = P_nU_n + \mathbf{u}\Delta_{P_n}, \tag{10}$$

where R_n, etc., are defined in the same way as in section 2. Combining the two equations, we obtain

$$AR_n = R_n\hat{T}_n - \frac{1}{\alpha_n}r_{n+1}e_n^T + \mathbf{u}\hat{\Delta}_n$$

with $\hat{\Delta}_n = A\Delta_{P_n} + \Delta_{R_n}U_n$. Write $\hat{\Delta}_n = [\delta_1, \cdots, \delta_n]$. Then we have $\delta_i = A\delta_{p_i} + \delta_{r_i} - \delta_{r_{i-1}}\beta_i$. So for $i \geq 2$, we have (with $N' = N + 2$)

$$\begin{aligned}
|\delta_i| &\leq |A||r_i| + 2|A||\beta_i||p_{i-1}| + \frac{|r_i|}{|\alpha_i|} + N'|A||p_i| \\
&\quad + |\beta_i|(\frac{|r_{i-1}|}{|\alpha_{i-1}|} + N'|A||p_{i-1}|) + O(\mathbf{u}) \\
&\leq (|A| + \frac{1}{|\alpha_i|} + \frac{|\beta_i|}{|\alpha_{i-1}|})|r_i| + N'|A||p_i| + (N'+3)|A|(|p_i| + |r_i|) + O(\mathbf{u}) \\
&\leq ((N'+4)|A| + \frac{1}{|\alpha_i|} + \frac{|\beta_i|}{|\alpha_{i-1}|})|r_i| + (2N'+3)|A||p_i| + O(\mathbf{u}),
\end{aligned}$$

where we have used $|r_{i-1}| \leq |r_i| + |\alpha_{i-1}||Ap_{i-1}| + O(\mathbf{u})$ and $|\beta_i p_{i-1}| \leq |p_i| + |r_i| + O(\mathbf{u})$. The same bound holds for the case $i = 1$, and hence the theorem is proved. □

Remark. We can conclude from bound (7) that the accuracy in the computations of the coefficients α and β does not directly affect the perturbation $\mathbf{u}\hat{\Delta}_n$. In particular, occurrence of near-breakdowns may cause the computed α or β and hence the iterates to be inaccurate, but it does not necessarily cause a large error δ_i (see Example 2 in Sec. 4), and thus the fundamental equation (6) is still nearly satisfied. Also note that $|p_i| \leq |A^{-1}|(|r_i| + |r_{i+1}|)/|\alpha_i| + O(\mathbf{u})$, and therefore the relative magnitude of the perturbation $\delta_i/\|r_i\|$ does depend on the magnitude of $1/\alpha_i$ and β_i.

In our later analysis, it is more convenient to work with an equivalent form of (6) with r_i scaled to the unit norm. It will also become clear that only the relative perturbation $\delta_i/\|r_i\|$ is of importance. Let

$$D_n = \mathrm{diag}\{\|r_1\|, \cdots, \|r_n\|\} \text{ and } Z_n = [z_1, \cdots, z_n] = R_n D_n^{-1}.$$

Then we obtain scaled (6) with $\|z_i\| = 1$ as

$$AZ_n = Z_n T_n - \frac{1}{\alpha'_n}\frac{r_{n+1}}{\|r_1\|}e_n^T + \mathbf{u}\Delta_n, \tag{11}$$

where $T_n = D_n\hat{T}_n D_n^{-1}$ is an invertible tridiagonal matrix, $\alpha'_n = \|r_n\|\alpha_n/\|r_1\| = e_n^T T_n^{-1} e_1$ and

$$\Delta_n = \hat{\Delta}_n D_n^{-1} = [\delta_1/\|r_1\|, \cdots, \delta_n/\|r_n\|].$$

This will be the only equation that we assume for BiCG in the rest of this section.

3.2. **Bounds on $\|r_{n+1}\|$ for the finite precision BiCG.** In this section, we present the main result on bounding the computed residual (Theorem 3.6). We first give a few lemmas. Recall that $\mathcal{P}_n$ denote the set of polynomials of degree not exceeding n.

Lemma 3.2. *Assume $AZ_n = Z_nT_n - \frac{1}{\alpha'_n}\frac{r_{n+1}}{\|r_1\|}e_n^T$ with $e_n^T T_n^{-1} e_1 = \alpha'_n$ and $r_1 = \|r_1\|z_1$. Then, for any polynomial $p(x) = \sum_{k=0}^n \psi_k x^k$ of degree not exceeding n,*

$$p(A)z_1 = Z_n p(T_n)e_1 + c_n r_{n+1}, \tag{12}$$

where $c_n = -\psi_n(\alpha_1 \cdots \alpha_n\|r_1\|)^{-1}$.

Proof. We first prove by induction that for any k with $1 \le k \le n-1$,

$$A^k Z_n e_1 = Z_n T_n^k e_1. \tag{13}$$

Clearly it is true for $k = 1$. Assume that (13) is true for some $k \le n-2$. First, we have $e_n^T T_n^k e_1 = 0$ (see [27, Lemma 3.1]). Then

$$\begin{aligned} A^{k+1} Z_n e_1 &= AZ_n T_n^k e_1 = (Z_n T_n - \frac{1}{\alpha_n' \|r_1\|} r_{n+1} e_n^T) T_n^k e_1 \\ &= Z_n T_n^{k+1} e_1. \end{aligned}$$

Therefore (13) holds. Consider now

$$\begin{aligned} A^n Z_n e_1 &= AA^{n-1} Z_n e_1 = AZ_n T_n^{n-1} e_1 \\ &= (Z_n T_n - \frac{1}{\alpha_n' \|r_1\|} r_{n+1} e_n^T) T_n^{n-1} e_1 \\ &= Z_n T_n^n e_1 - \frac{1}{\alpha_1 \cdots \alpha_n \|r_1\|} r_{n+1}, \end{aligned}$$

where $e_n^T T_n^{n-1} e_1 = \|r_n\| (\alpha_1 \cdots \alpha_{n-1} \|r_1\|)^{-1}$ (see [27, Theorem 3.2]). Combining the above with (13), we obtain (12). □

Note that the above lemma also holds if T_n is Hessenberg. In particular, if p is a polynomial of degree $n-1$ (i.e. $\psi_n = 0$), then (12) becomes $p(A)z_1 = Z_n p(T_n) e_1$, a case that has been proved in [17, 21].

From this lemma, we have the following identity concerning r_{n+1}.

Lemma 3.3. *Assume*

$$AZ_n = Z_n T_n - \frac{1}{\alpha_n'} \frac{r_{n+1}}{\|r_1\|} e_n^T \tag{14}$$

with $e_n^T T_n^{-1} e_1 = \alpha_n'$ and $r_1 = \|r_1\| z_1$, and assume that $V^T \in R^{n \times N}$ is a matrix such that $V^T Z_n = I$ and $V^T r_{n+1} = 0$. Then for any polynomial $p(x)$ of degree not exceeding n with $p(0) = 1$, we have

$$r_{n+1} = (I - AZ_n T_n^{-1} V^T) p(A) r_1. \tag{15}$$

Proof. First, multiplying (14) by $T_n^{-1} e_1$, we obtain

$$r_{n+1}/\|r_1\| = z_1 - AZ_n T_n^{-1} e_1. \tag{16}$$

Write $p(x) = 1 + xq(x)$, with $q(x) = \sum_{k=0}^{n-1} \psi_{k+1} x^k$ a polynomial of degree not exceeding $n-1$. Then

$$\begin{aligned} r_{n+1}/\|r_1\| &= z_1 - AZ_n T_n^{-1} e_1 - p(A) z_1 + p(A) z_1 \\ &= -Aq(A) z_1 - AZ_n T_n^{-1} e_1 + p(A) z_1 \\ &= -AZ_n q(T_n) e_1 - AZ_n T_n^{-1} e_1 + p(A) z_1 \\ &= -AZ_n (q(T_n) + T_n^{-1}) e_1 + p(A) z_1 \\ &= -AZ_n T_n^{-1} p(T_n) e_1 + p(A) z_1, \end{aligned} \tag{17}$$

where we note that by Lemma 3.2, $Aq(A)z_1 = Aq(A)Z_n e_1 = AZ_n q(T_n) e_1$ with $\deg(q) < n$.

Now, for the polynomial p, equation (12) holds. Multiplying it by V^T, we obtain

$$V^T p(A) z_1 = p(T_n) e_1. \tag{18}$$

Finally, substituting $p(T_n)e_1$ into the equation above, we have

$$\frac{r_{n+1}}{\|r_1\|} = (I - AZ_nT_n^{-1}V^T)p(A)z_1.$$

Now, the lemma follows from using $r_1 = \|r_1\|z_1$. □

Note that the matrix V above exists if and only if $z_1, \cdots, z_n, z_{n+1}$ are linearly independent. In the case of exact BiCG, this is the case because of biorthogonality. Indeed, a natural choice for V is $V^T = (\hat{R}_n^T R_n)^{-1}\hat{R}_n^T$, since by the biorthogonality condition $\hat{R}_n^T R_n = D$ is a diagonal matrix, and $\hat{R}_n^T r_{n+1} = 0$. Using this V, the lemma leads to the known results for the exact BiCG [2].

When an biorthogonal basis is not explicitly available, the next lemma gives a construction of V.

Lemma 3.4. *Assume that $z_1, z_2, \cdots, z_n, z_{n+1} \in R^N$ are linearly independent, and write $Z_k = [z_1\ z_2\ \cdots\ z_k]$. Then $V_0^T = [I_n\ \ 0]Z_{n+1}^+$ (i.e. the matrix consisting of the first n rows of Z_{n+1}^+) has the property*

$$V_0^T Z_n = I \quad \text{and} \quad V_0^T z_{n+1} = 0. \tag{19}$$

Furthermore, its spectral norm is minimal among all the matrices having this property.

Proof. From the definition of V_0, $Z_{n+1}^+ = [V_0, v]^T$ for some v. Since $z_1, \cdots, z_n, z_{n+1}$ are linearly independent, $[V_0, v]^T[Z_n\ \ z_{n+1}] = Z_{n+1}^+ Z_{n+1} = I$. Then $V_0^T Z_n = I_n$ and $V_0^T z_{n+1} = 0$.

Now, if V is another matrix having property (19), then $V^T[Z_n\ \ z_{n+1}] = [I\ \ 0]$. Thus $V^T Z_{n+1}Z_{n+1}^+ = [I\ \ 0]Z_{n+1}^+ = V_0^T$. Hence $\|V_0\| \le \|V\| \cdot \|Z_{n+1}Z_{n+1}^+\| \le \|V\|$. □

We now present our main result on bounding the residuals.

Theorem 3.5. *Assume equation (11) and let $V_0^T = [I_n\ \ 0]Z_{n+1}^+ \in \mathbf{R}^{n\times N}$ (the matrix consisting of the first n rows of Z_{n+1}^+). If $z_1, z_2, \cdots, z_n, z_{n+1}$ are linearly independent, then*

$$\|r_{n+1}\| \le (1 + K_n) \min_{p\in\mathcal{P}_n, p(0)=1} \|p(A + \delta A_n)r_1\|, \tag{20}$$

where $K_n = \|(AZ_n - \mathbf{u}\Delta_n)T_n^{-1}V_0^T\|$ and $\delta A_n = -\mathbf{u}\Delta_n Z_n^+$.

Furthermore, as n increases, $\epsilon_n = \min_{p\in\mathcal{P}_n, p(0)=1} \|p(A + \delta A_n)r_1\|$ decreases monotonically.

Proof. Since $z_1, \cdots, z_n, z_{n+1}$ are linearly independent, $Z_n^+ Z_n = I$. Then $\delta A_n = -\mathbf{u}\Delta_n Z_n^+ \in R^{N\times N}$ satisfies $\delta A_n Z_n = -\mathbf{u}\Delta_n$. Thus (11) can be rewritten as

$$(A + \delta A_n)Z_n = Z_n T_n - \frac{1}{\alpha_n'}\frac{r_{n+1}}{\|r_1\|}e_n^T. \tag{21}$$

Now for any $p \in \mathcal{P}_n$ with $p(0) = 1$, we use Lemmas 3.3 and 3.4 to obtain

$$\begin{aligned} r_{n+1} &= (I - (A + \delta A_n)Z_n T_n^{-1}V_0^T) \cdot p(A + \delta A_n)r_1 \\ &= (I - (AZ_n - \mathbf{u}\Delta_n)T_n^{-1}V_0^T) \cdot p(A + \delta A_n)r_1. \end{aligned}$$

Thus

$$\|r_{n+1}\| \le (1 + \|(AZ_n - \mathbf{u}\Delta_n)T_n^{-1}V_0^T\|)\|p(A + \delta A_n)r_1\|.$$

Since this is true for any $p(x)$ with $p(0) = 1$, the inequality is true for the minimizing polynomial, which leads to the bound.

For the second part, we show that for any polynomial p with $\deg(p) = k < m$,

$$p(A + \delta A_m)z_1 = p(A + \delta A_k)z_1,$$

which leads to the monotonicity result

$$\min_{p\in\mathcal{P}_m, p(0)=1} \|p(A + \delta A_m)z_1\| \le \min_{p\in\mathcal{P}_k, p(0)=1} \|p(A + \delta A_k)z_1\|.$$

For any i with $0 \le i \le k-1$, $T_m^i e_1 = \begin{pmatrix} T_k^i e_1 \\ 0 \end{pmatrix}$ (see [27, Lemma 3.1]). Then using Lemma 3.2, $(A + \delta A_m)^i z_1 = Z_m T_m^i e_1 = Z_k T_k^i e_1 = (A + \delta A_k)^i z_1$. Furthermore,

$$\begin{aligned}
(A + \delta A_m)^k z_1 &= (A + \delta A_m) Z_m T_m^{k-1} e_1 \\
&= (Z_m T_m - \frac{1}{\alpha'_m \|r_1\|} r_{m+1} e_m^T) T_m^{k-1} e_1 \\
&= Z_m T_m \begin{pmatrix} T_k^{k-1} e_1 \\ 0 \end{pmatrix} \\
&= Z_k T_k^k e_1 - \frac{\|r_{k+1}\|}{\alpha_k \|r_k\|} z_{k+1} e_k^T T_k^{k-1} e_1 \\
&= (Z_k T_k - \frac{1}{\alpha'_k \|r_1\|} r_{k+1} e_k^T) T_k^{k-1} e_1 \\
&= (A + \delta A_k) Z_k T_k^{k-1} e_1 \\
&= (A + \delta A_k)^k z_1.
\end{aligned}$$

Thus, $p(A + \delta A_m)z_1 = p(A + \delta A_k)z_1$, and the proof is complete. □

Remark. For each n, $\epsilon_n = \min_{p\in\mathcal{P}_n, p(0)=1} \|p(A + \delta A_n)r_1\|$ is the nth residual norm of exact GMRES applied to the perturbed matrix $A + \delta A_n$. Explicit bounds on ϵ_n have been discussed extensively in the literature; see [24] for example. Our numerical experiments show that the perturbation term δA_n has little effect, i.e. usually $\epsilon_n \sim \min_{p\in\mathcal{P}_n, p(0)=1} \|p(A)r_1\|$.

Remark. If T_n is not too close to being singular, and the elements of the basis $z_1, \cdots, z_n, z_{n+1}$ are not too close to being linearly dependent, i.e. $\|T_n^{-1}\|$ and $\|V_0\| \le \|Z_{n+1}^+\|$ are bounded, then it follows from

$$K_n \le (\sqrt{n}\|A\| + \mathbf{u}\|\Delta_n\|)\|T_n^{-1}\| \cdot \|V_0\| \tag{22}$$

that the residual norm $\|r_{n+1}\|$ is within a moderate factor of ϵ_n. Hence, convergence of the BiCG residual can be achieved provided K_n increases at a rate slower than the rate at which ϵ_n decreases (see the examples in Sec. 4). In particular, this is still possible even when $\|\Delta_n\|$ is large (Example 1 in Sec. 4). We note that K_n could grow out of bound; but this simply reflects the situations when BiCG diverges. Unfortunately K_n cannot be determined *a priori*.

Remark. If T_n is close to being singular at some step n, then r_{n+1} becomes (relatively) large in that step; but if at the next step T_{n+1} becomes well conditioned, r_{n+2} will be small. So occurrence of a nearly singular T_n would cause local spikes in the convergence curve but may not affect the global convergence trend even in finite precision arithmetic, assuming that not all T_n are nearly singular.

Remark. The above bound assumes linear independence among the r_n, and $\|V_0\|$ is a measure of the linear independence. It suggests the possible effect of the linear dependence on the convergence, which is known, for example, for the exact QMR (see [8]). However, this dependence on $\|V_0\|$ is not essential, and in the next subsection we give some results that avoid such an assumption at the expense of a more complicated bound.

3.3. **Bounds for the linearly dependent case.** While it is well-known that loss of linear independence among the r_n could cause deterioration of convergence in BiCG, there are cases where the convergence has been observed even when the linear independence of the r_n is completely lost (the case $n > N$ being one such example). We next present two bounds that avoid the linear independence condition and may, therefore, partially explain the convergence in such situations.

First, recall that for $A \in \mathbf{R}^{N\times N}$ and $B \in \mathbf{R}^{n\times n}$, the matrix equation $AE - EB = Z$ corresponds to a linear system with $A \otimes I_n - I_N \otimes B$ as the coefficient matrix. It has a unique solution if and only if $\lambda(A) \cap \lambda(B) = \emptyset$ or $\mathrm{sep}(A, B) = \|(A \otimes I_n - I_N \otimes B)^{-1}\|^{-1} > 0$ [11, p. 389]. Note that $\mathrm{sep}(A, B)$ depends on the spectral gap of A and B.

Theorem 3.6. *Assume equation* (11) *and let* μ *be a complex number such that* $\mathrm{sep}(A - \mu I, T_n) >> 0$. *Then*

$$\|r_{n+1}\| \le K_n \min_{p\in\mathcal{P}_n, p(0)=1} (\|p(T_n)\| + \|p(A - \mu I)\|)\, \|r_1\|, \tag{23}$$

where

$$K_n = \frac{(|\mu| + \mathrm{sep}(A - \mu I, T_n))\sqrt{n} + \mathbf{u}\|\Delta_n\|_F}{\mathrm{sep}(A - \mu I, T_n)} \max\{1,\ \|A - \mu I\| \cdot \|T_n^{-1}\|\}$$

Proof. First, we rewrite (11) as

$$(A - \mu I)Z_n = Z_n T_n - \frac{1}{\alpha_n' \|r_1\|} r_{n+1} e_n^T + \mathbf{u}\Delta_n - \mu Z_n.$$

Since $\mathrm{sep}(A - \mu I, T_n) > 0$, the matrix equation

$$(A - \mu I)E_n = E_n T_n - \mathbf{u}\Delta_n + \mu Z_n$$

has a unique solution E_n [11, p.389] with

$$\|E_n\|_F \le \frac{\|-\mathbf{u}\Delta_n + \mu Z_n\|_F}{\mathrm{sep}(A - \mu I, T_n)} \le \frac{\mathbf{u}\|\Delta_n\|_F + |\mu|\sqrt{n}}{\mathrm{sep}(A - \mu I, T_n)}.$$

Then

$$(A - \mu I)(Z_n + E_n) = (Z_n + E_n)T_n - \frac{1}{\alpha_n'} \frac{r_{n+1}}{\|r_1\|} e_n^T.$$

Thus, for any $p \in \mathcal{P}_n$ with $p(0) = 1$, we have by (17)

$$\frac{r_{n+1}}{\|r_1\|} = p(A - \mu I)(Z_n + E_n)e_1 - (A - \mu I)(Z_n + E_n)T_n^{-1} p(T_n)e_1,$$

and hence

$$\begin{aligned}\frac{\|r_{n+1}\|}{\|r_1\|} &\le (\|Z_n\| + \|E_n\|)\, \|p(A - \mu I)\| \\ &\quad + \|A - \mu I\|\, (\|Z_n\| + \|E_n\|)\|T_n^{-1}\|\, \|p(T_n)e_1\|,\end{aligned}$$

which leads to the theorem, since $\|Z_n\| + \|E_n\| \le \sqrt{n} + \|E_n\|_F$. ∎

Again the bound contains two parts. K_n depends mainly on $\|T_n^{-1}\|$ if the parameter μ has been chosen to create a large $\mathrm{sep}(A-\mu I, T_n)$. The second part now depends on minimization of $p(T_n)$ in addition to $p(A-\mu I)$. Note that if μ is simply chosen to be 0, then we have a simpler bound depending on $\mathrm{sep}(A, T_n)$; but this gap is usually small because of convergence of the eigenvalues of T_n to those of A.

We further consider a simplified special case where r_i becomes nearly linearly dependent because of the appearance of spurious Ritz values and Ritz vectors (in the underlying Lanczos process). It is known, in the symmetric case, that gradual appearance of spurious Ritz vectors leads to linear dependence [19, 20]. Our next result will be formulated in terms of the Ritz basis, rather than in terms of the basis Z_n.

An eigenvalue θ of T_n is called a Ritz value and $Z_n u$ is called a corresponding Ritz vector if $T_n u = \theta u$. By considering the basis of Ritz vectors rather than Z_n, the pattern of loss of linear independence becomes clear, i.e., the Ritz vectors corresponding to the spurious Ritz values become nearly linearly dependent on the others. For this reason, we divide Ritz vectors into two groups, one consisting of $n-l$ Ritz vectors ($Z_n U_1$ in the theorem below) that are well linearly independent, and the other consisting of the rest, which are nearly linearly dependent on the first group.

Theorem 3.7. *Assume equation* (11) *and that* $z_1, \cdots, z_n$ *are linearly independent. Let* $T_n = USU^*$ *be a Schur decomposition ordered such that the columns of* $[Z_n U_1, z_{n+1}]$ *are linearly independent, where*

$$S \overset{\cdot}{=} \begin{pmatrix} \overset{n-l}{S_{11}} & \overset{l}{S_{12}} \\ 0 & S_{22} \end{pmatrix} \begin{matrix} n-l \\ l \end{matrix}, \qquad U = \overset{n-l}{(U_1,} \ \overset{l}{U_2}).$$

Let $V_0^T \in \mathbf{R}^{(n-l)\times N}$ *be the first* $n-l$ *rows of* $[Z_n U_1, z_{n+1}]^+$. *Then*

$$\|r_{n+1}\| \le (1+K_n) \min_{p\in\mathcal{P}_{n-l}, p(0)=1} \|p(A+\delta A_n)\hat{r}_1\|, \tag{24}$$

where $\hat{r}_1 = (I-(A+\delta A_n)/\theta_1)\cdots(I-(A+\delta A_n)/\theta_l)r_1$, $\lambda(S_{22}) = \{\theta_1, \ldots, \theta_l\}$,

$$K_n = \|(AZ_n - \mathbf{u}\Delta_n)U_1 S_{11}^{-1} V_0^T\| \le (\sqrt{n}\|A\| + \mathbf{u}\|\Delta_n\|)\|T_n^{-1}\| \cdot \|V_0\|$$

and $\delta A_n = -\mathbf{u}\Delta_n Z_n^+$.

Proof. Let $q(x) = (1-x/\theta_1)\cdots(1-x/\theta_l)$. Then for any polynomial $p \in \mathcal{P}_{n-l}$ with $p(0)=1$, $\hat{p}(x) = p(x)q(x)$ is of degree n and $\hat{p}(0)=1$. Then $\hat{p}(S_{22}) = 0$, and thus

$$\hat{p}(S) = \begin{pmatrix} \hat{p}(S_{11}) & M \\ 0 & \hat{p}(S_{22}) \end{pmatrix} \begin{matrix} n-l \\ l \end{matrix} = \begin{pmatrix} \hat{p}(S_{11}) & M \\ 0 & 0 \end{pmatrix} \begin{matrix} n-l \\ l \end{matrix},$$

where M is some matrix. Now, as in the proof of Theorem 3.5, we rewrite (11) as

$$(A+\delta A_n)Z_n = Z_n T_n - \frac{1}{\alpha_n'}\frac{r_{n+1}}{\|r_1\|}e_n^T.$$

Applying Lemma 3.2 (12) to the above with $\hat{p}$, we obtain

$$\begin{aligned} \hat{p}(A+\delta A_n)z_1 &= Z_n\hat{p}(T_n)e_1 + cr_{n+1} \\ &= [Z_nU_1,\ Z_nU_2]\hat{p}(S)U^*e_1 + cr_{n+1} \\ &= [Z_nU_1,\ 0]\hat{p}(S)U^*e_1 + cr_{n+1}. \end{aligned}$$

Multiplying the above by V_0^T, we obtain

$$V_0^T \hat{p}(A+\delta A_n) z_1 = [I_{n-l},\ 0]\hat{p}(S)U^* e_1,$$

where $V_0^T Z_n U_1 = I_{n-l}$ and $V_0^T r_{n+1} = 0$ by Lemma 3.4. Also, noting that S^{-1} is block upper triangular of the same structure as S, we have

$$\begin{aligned}
(A+\delta A_n) Z_n T_n^{-1} \hat{p}(T_n) e_1 &= (A+\delta A_n)[Z_n U_1,\ Z_n U_2] S^{-1} \hat{p}(S) U^* e_1 \\
&= (A+\delta A_n)[Z_n U_1,\ Z_n U_2]\mathrm{diag}[S_{11}^{-1},\ 0]\hat{p}(S)U^* e_1 \\
&= (A+\delta A_n)[Z_n U_1 S_{11}^{-1},\ 0]\hat{p}(S)U^* e_1 \\
&= (A+\delta A_n) Z_n U_1 S_{11}^{-1}[I_{n-l},\ 0]\hat{p}(S)U^* e_1 \\
&= (AZ_n - \mathbf{u}\Delta_n) U_1 S_{11}^{-1} V_0^T \hat{p}(A+\delta A_n) z_1.
\end{aligned}$$

Now, by (17), we have

$$\begin{aligned}
\frac{r_{n+1}}{\|r_1\|} &= \hat{p}(A+\delta A_n) z_1 - (A+\delta A_n) Z_n T_n^{-1} \hat{p}(T_n) e_1 \\
&= (I - (AZ_n - \mathbf{u}\Delta_n) U_1 S_{11}^{-1} V_0) \hat{p}(A+\delta A_n) z_1
\end{aligned}$$

Hence

$$\|r_{n+1}\| \le (1+K_n)\|\hat{p}(A+\delta A_n) z_1\| \|r_1\| = (1+K_n)\|p(A+\delta A_n)\hat{r}_1\|.$$

which leads to the theorem by taking p to be the minimizing polynomial. □

Essentially, the above bound replaces Z_n (containing n vectors) in Theorem 3.5, which may become nearly linearly dependent, by $Z_n U_1$ (containing $n-l$ vectors). This eases the dependence of the bound on loss of linear independence among the z_i. In doing so, however, the degree of the minimizing polynomial is reduced to $n-l$ and thus the convergence is expected to slow down.

4. Numerical experiments

We present some numerical examples in this section. The purposes of these experiments are to verify the bounds, and to observe how the different parameters in the bounds affect the convergence behavior. All quantities in the bounds are computed with quadruple precision.

We examine the following residual bound (Theorem 3.5) in our numerical experiment:

$$\|r_{n+1}\| \le (1+K_n)\epsilon_n, \tag{25}$$

where $\epsilon_n = \min_{p\in\mathcal{P}_n, p(0)=1} \|p(A+\delta A_n) r_1\|$ is the residual norm of GMRES applied to $A+\delta A_n$, $K_n = \|(AZ_n - \mathbf{u}\Delta_n) T_n^{-1} V_0^T\|$ and $\delta A_n = -\mathbf{u}\Delta_n Z_n^+$.

Example 1. Our first example is a diagonal matrix A ($N = 32$) with $A(i,i) = 1.01^i$. We consider injecting respectively 1% and 10% random perturbations in the computation of r_{n+1} (but not in $\tilde{r}_{n+1}$) and compare the residual norms with the unperturbed case. Note that most known properties of BiCG except the perturbed equation (11) have been lost under the perturbations of this magnitude. In Figure 1, we give the convergence curve for the perturbed (solid) and the unperturbed (dashed) cases and the bound (the + sign) for the perturbed case. To further analyze the various quantities in the bound, we give the magnitudes of K_n and the minimum polynomials ϵ_n in Figure 2, and the values $\|T_n^{-1}\|$ and $\|V_0\|$, which bound K_n, in Figure 3.

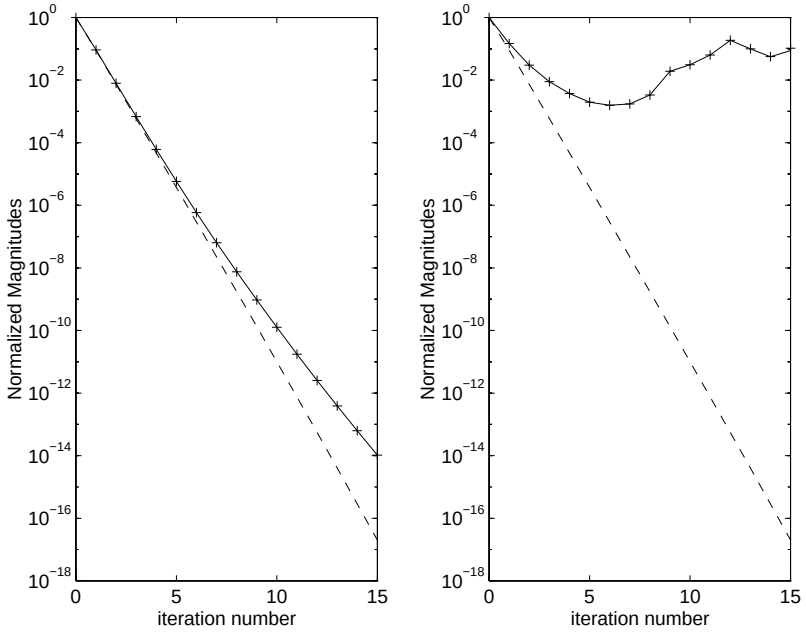

FIGURE 1. Perturbed BiCG convergence history and bounds in Example 1: (left) 0.01, (right) 0.1 perturbation (solid - with perturbation, dashed - no perturbation, '+' - residual bound)

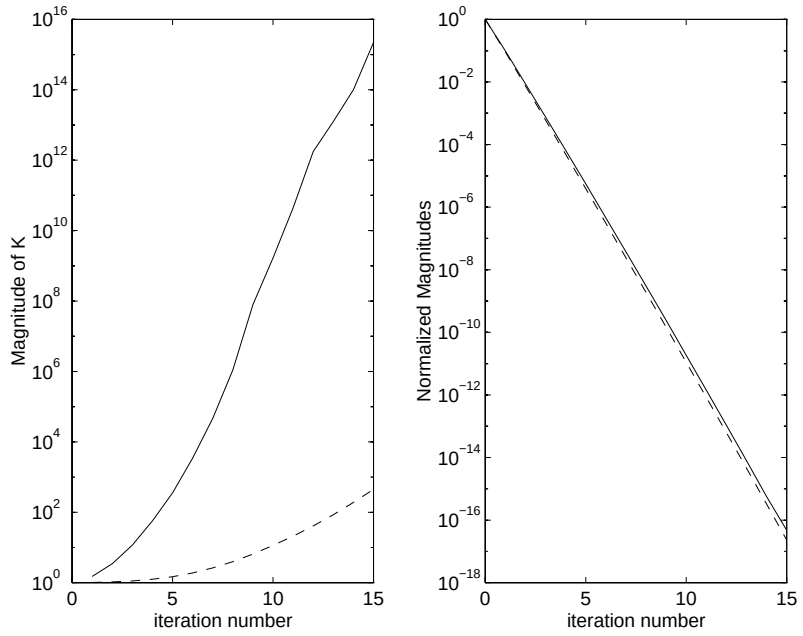

FIGURE 2. BiCG bounds in Example 1: (left) values of K_n, (right) GMRES residual norm ϵ_n (solid - 0.1, dashed - 0.01 perturbation)

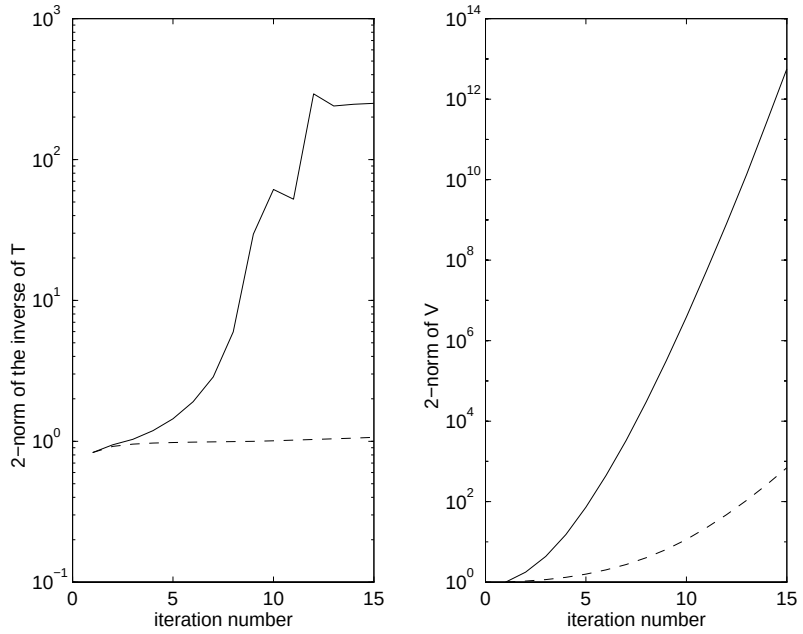

FIGURE 3. Components of K_n in Example 1: (left) $\|T_n^{-1}\|$ (right) $\|V_0\|$ (solid - 0.1, dashed - 0.01 perturbation)

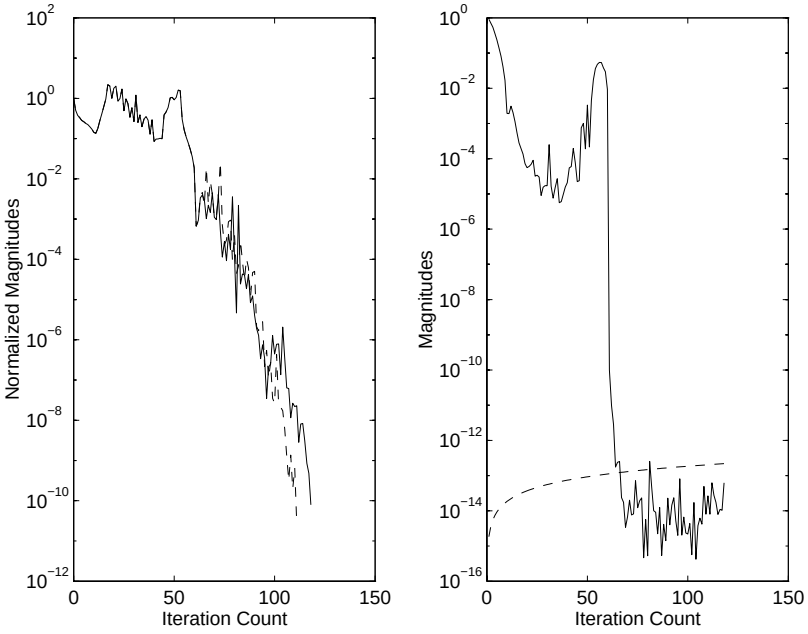

FIGURE 4. BiCG for the convection diffusion equation in Example 2: (left) convergence history (solid - double precision, dashed - quadruple precision), (right) Lanczos pivots $\cos\langle r_n, \tilde{r}_n\rangle$ (solid) and roundoff error term $\mathbf{u}\|\Delta_n\|_F$ (dashed)

We first observe that the given bound is satisfied (see Figure 1). We also observe that with 1% injected error, the BiCG residual still converges to the desired precision, while it diverges for 10% injected error. Since ϵ_n converges monotonically to 0 at essentially the same rate for both cases, the reason for the divergence in the latter case is that the K_n term grows faster than the GMRES contraction ϵ_n. A further study on K_n (Fig. 3) reveals that $\|V_0\|$, which is a measure of loss of linear independence among the r_i, plays a major role in determining convergence or divergence. Therefore, it appears that with the smaller injected error (1%) and relatively short iterations, the loss of linear independence is modest and $\|V_0\|$ grows slowly, so that the GMRES reduction ϵ_n can still drive the residual to convergence. However, with the larger injected error (10%), the loss of linear independence is much more severe and fast ($\|V_0\|$ increases to 10^{12}), which results in divergence.

In a separate experiment, we inject artificial perturbations in the computation of α_i instead of r_i. Similar behavior was observed. We omit the details.

Example 2. The second numerical example investigates the effect of near-breakdowns on the convergence of BiCG. The matrix used here is the following convection diffusion equation discretized on a 31×31 uniform rectangular grid:

$$-\triangle u + 50(xu_x + yu_y) - 25u = f(x,y) \quad \text{on} \quad \Omega = [0,1]^2.$$

Near-breakdowns occur in this example, and we run BiCG in double and quadruple precisions to show the impact of the roundoff errors. Figure 4 (left) shows the convergence curves of BiCG in double and quadruple precisions, and Figure 4 (right) shows the magnitude of the Lanczos pivots

$$\cos\langle r_n, \tilde{r}_n\rangle = \tilde{r}_n^T r_n/(\|\tilde{r}_n\|_2\|r_n\|_2)$$

and the rounding error term $\mathbf{u}\|\Delta_n\|_F$ for the double precision case.

We observe that the two convergence curves start to deviate at around iteration 60, after a near-breakdown occurs (the Lanczos pivot drops to about 10^{-14}; see Fig. 4). Indeed, the residual vectors as generated by the double and quadruple precisions are completely different after this point. Despite this drop in the Lanczos pivot, however, β_n remains small throughout, and so is the Frobenius norm of the error term $\mathbf{u}\Delta_n$. This confirms the analysis of Theorem 3.1, namely, occurrence of near breakdown does not necessarily cause large error $\mathbf{u}\Delta_n$ in the fundamental equation (11). Thus, it does not come as a surprise that the iteration counts required to achieve convergence are very close for the double and quadruple precision cases. In other words, the residual vectors in double and quadruple precision may be completely different, but their norms, bounded by similar quantities, may still be comparable.

5. Conclusion

The analysis presented in this paper attempts to shed some light on the important computational issue of how roundoff errors affect convergence of the BiCG algorithm. In particular, we show that even in the presence of roundoff errors, the BiCG residual bound can still be formulated in a way similar to the exact case in terms of a GMRES residual norm. There is, however, an amplification factor K_n, which cannot be bounded analytically without any *a posteriori* information. We remark that this is also the case even in exact arithmetic (see [2]), and it seems to be the nature of BiCG type algorithms that convergence cannot be determined *a*

priori. The BiCG residual bound given here also suggests and explains that it is still possible to achieve convergence even with relatively large errors in the recurrences, as confirmed by some numerical results.

Acknowledgments

We are grateful to Professors Tony Chan, Gene H. Golub, Anne Greenbaum and Beresford Parlett and Dr. David Day for discussions and comments. In particular, we thank Gene Golub for pointing out a connection to inexact CG, and Anne Greenbaum for showing us her related work [15].

References

1. Z. Bai, *Error Analysis of the Lanczos Algorithm for the Nonsymmetric Eigenvalue Problem,* Math. Comp., 62:209-226 (1994). MR **94c:**65045
2. R. E. Bank and T. F. Chan, *An Analysis of the Composite Step Biconjugate Gradient Algorithm for Solving nonsymmetric Systems,* Numer. Math., 66:295-319 (1993). MR **94i:**65043
3. T. Barth and T. A. Manteuffel, *Variable Metric Conjugate Gradient Methods*, Proceedings of the 10th International Symposium on Matrix Analysis and Parallel Computing, Keio University, Yokohama, Japan, March 14-16, 1994.
4. J. Cullum and A. Greenbaum, *Relation between Galerkian and norm-minimizing iterative methods for solving linear systems* SIAM J. Matrix Anal. Appl. 17:223–247 (1996). MR **97b:**65035
5. D. Day, *Semi-duality in the two-sided Lanczos algorithm.* Ph.D. Thesis, University of Californial, Berkeley, 1993.
6. I. S. Duff, R. G. Grimes, and J. G. Lewis, *Sparse Matrix Test Problems,* ACM Trans. Math. Softw., 15:1-14 (1989).
7. R. Fletcher, *Conjugate Gradient Methods for Indefinite Systems,* in Proc. Dundee Conference on Numerical Analysis, 1975, Lecture Notes in Mathematics 506, G. A. Watson, ed., Springer-Verlag, Berlin, pp. 73-89 (1976). MR **57:**1841
8. R. W. Freund and N. M. Nachtigal, *QMR : a Quasi-minimal Residual Method for non-Hermitian Linear Systems,* Numer. Math., 60:315-339 (1991). MR **92g:**65034
9. G. H. Golub and M. L. Overton, *Convergence of a two-stage Richardson iterative procedure for solving systems of linear equations.* in Numerical Analysis, Lect. Notes Math 912, (ed. G.A. Watson), Springer, New York Heidelberg Berlin pp.128-139. MR **83f:**65045
10. G. H. Golub and M. L. Overton, *The convergence of inexact Chebyshev and Richardson iterative methods for solving linear systems.* Numer. Math. 53:571-593 (1988). MR **90b:**65054
11. G. H. Golub, C. F. Van Loan, *Matrix Computations,* The Johns Hopkins University Press, Baltimore, 1983. MR **85h:**65063
12. A. Greenbaum, *Behavior of Slightly Perturbed Lanczos and Conjugate-Gradient Recurrences,* Lin. Alg. and its Appl., 113:7-63 (1989). MR **90e:**65044
13. A. Greenbaum, *Accuracy of Computed Solutions from Conjugate-Gradient-Like Methods,* PCG '94, Matrix Analysis and Parallel Computing, Keio University, March 14-16, 1994.
14. A. Greenbaum and Z. Strakos, *Predicting the behavior of finite precision Lanczos and Conjugate Gradient computations,* SIAM J. Matrix Anal. Appl. 13:121-137 (1992). MR **92j:**65043
15. A. Greenbaum, V. Druskin, and L. Knizhnerman, *Private communication.*
16. C. Lanczos, *Solution of Systems of Linear Equations by Minimized Iterations,* J. Res. Natl. Bur. Stand. 49:33-53 (1952). MR **14:**501g
17. R. Lehoucq, *Analysis and implementation of an implicitly restarted iteration* Ph.D. Thesis, Rice University, Houston, Texas, May 1995.
18. Y. Notay, *On the convergence rate of the conjugate gradients in presence of rounding errors*, Numer. Math. 65:301-317 (1993). MR **94j:**65050
19. C. Paige, *Error Analysis of the Lanczos Algorithm for Tridiagonalizing a Symmetric Matrix,* J. Inst. Math. Appl., 18:341-349 (1976). MR **58:**19082
20. C. Paige, *Accuracy and Effectiveness of the Lanczos Algorithm for the Symmetric Eigenproblem,* Linear Alg. Appl. 34(1980):235-258. MR **82b:**65025
21. Y. Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operators*, SIAM J. Numer. Anal. 29:209-228 (1992). MR **92m:**65050

22. P. Sonneveld, *CGS, a Fast Lanczos-type Solver for Nonsymmetric Linear Systems,* SIAM J. Sci. Stat. Comput., 10:36-52 (1989). MR **89k:**65052
23. C. H. Tong, *A Comparative Study of Preconditioned Lanczos Methods for Nonsymmetric Linear Systems,* Tech. Report SAND91-9240B, Sandia National Lab., Livermore, 1992.
24. L. N. Trefethen, *Approximation theory and numerical linear algebra,* in Algorithms for Approximation II, J.C. Mason and M.G. Cox eds., Chapman and Hall, London, 1990, pp. 336–360. MR **91j:**65063
25. H. A. Van der Vorst, *Bi-CGSTAB : A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems,* SIAM J. Sci. Stat. Comput., 13:631-644 (1992). MR **92j:**65048
26. H. A. Van der Vorst, *The convergence behaviour of preconditioned CG and CGS* in Lect. Notes in Math. 1457, ed. O. Axelsson and L. Kolotilina, pp. 121-136, Springer, Berlin Heidelberg New York (1990), pp. 126–136. MR **92a:**65141
27. Q. Ye, *A convergence analysis of nonsymmetric Lanczos algorithms*, Math. Comp. 56:677-691 (1991). MR **91m:**65115

Sandia National Laboratories, Livermore, CA 94551
E-mail address: chtong@california.sandia.gov

Department of Mathematics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2
E-mail address: ye@gauss.amath.umanitoba.ca