SIAM J. MATRIX ANAL. APPL.
Vol. 13, No. 3, pp. 778–795, July 1992
© 1992 Society for Industrial and Applied Mathematics
007

HOW FAST ARE NONSYMMETRIC MATRIX ITERATIONS?*

NOËL M. NACHTIGAL†, SATISH C. REDDY‡, AND LLOYD N. TREFETHEN§

Abstract. Three leading iterative methods for the solution of nonsymmetric systems of linear equations are CGN (the conjugate gradient iteration applied to the normal equations), GMRES (residual minimization in a Krylov space), and CGS (a biorthogonalization algorithm adapted from the biconjugate gradient iteration). Do these methods differ fundamentally in capabilities? If so, which is best under which circumstances? The existing literature, in relying mainly on empirical studies, has failed to confront these questions systematically. In this paper it is shown that the convergence of CGN is governed by singular values and that of GMRES and CGS by eigenvalues or pseudo-eigenvalues. The three methods are found to be fundamentally different, and to substantiate this conclusion, examples of matrices are presented for which each iteration outperforms the others by a factor of size $O(\sqrt{N})$ or $O(N)$ where N is the matrix dimension. Finally, it is shown that the performance of iterative methods for a particular matrix cannot be predicted from the properties of its symmetric part.

Key words. iterative method, conjugate gradient iteration, normal equations, Krylov space, pseudospectrum, CGN, GMRES, BCG, CGS

AMS(MOS) subject classification. 65F10

1. Introduction.

More than a dozen parameter-free iterative methods have been proposed for solving nonsymmetric systems of linear equations

$$Ax = b, \qquad A \in \mathbb{C}^{N \times N}. \tag{1.1}$$

A rough list is given in Table 1, and for a more detailed classification we recommend [1], [6], and [12]. In this paper we concentrate on the three methods that we believe are the most important: CGN, GMRES, and CGS. Quickly summarized, CGN is a name for the conjugate gradient iteration applied to the normal equations; this idea can be implemented in various ways, of which the most robust in the presence of rounding errors may be the program LSQR [18]. GMRES is the most robust of the Krylov space orthogonalization and residual minimization methods. CGS is a modification of BCG, the biconjugate gradient iteration, that appears to outperform BCG consistently. To the best of our knowledge none of the other iterations proposed to date significantly outperform CGN, GMRES, and CGS.

This leaves us with the questions: do CGN, GMRES, and CGS themselves differ significantly in capabilities? If so, which of them is best for which matrices? In the literature, these questions have for the most part been approached empirically by case studies of "real world" matrices and preconditioners. However, although such case studies are indispensable as proofs of feasibility, the answers they provide are not very sharp or general. We believe that the experimental approach is an inefficient route to the understanding of fundamental properties of algorithms and a poor basis for predicting the results of future computations.

In this paper we attempt a more systematic assessment of the convergence of nonsymmetric matrix iterations. The first half of the paper deals with generalities, presenting

* Received by the editors April 5, 1990; accepted for publication (in revised form) April 17, 1991.

† Research Institute for Advanced Computer Science, Mail Stop T041-5, NASA Ames Research Center, Moffett Field, California 94035. This research was supported by National Science Foundation grant 8814314-DMS.

‡ Courant Institute of Mathematical Sciences, 251 Mercer St., New York, New York 10012.

§ Department of Computer Science, Cornell University, Ithaca, New York 14853 (lnt@cs.cornell.edu). The research of this author was supported by a National Science Foundation Presidential Young Investigator Award and Air Force grant AFOSR-87-0102.

TABLE 1

Iterative methods for nonsymmetric systems $Ax = b$. (The differences among these algorithms are slight in some cases.) The terminology approximately follows Elman [6] *and Gutknecht* [12]. *References not listed can be found in those two papers and in* [21].

I. Methods based on the normal equations	
CGN = CGNR	Hestenes and Stiefel '52 [14]
CGNE	Craig '55
LSQR	Paige and Saunders '82 [18]
II. Orthogonalization methods	
GCG	Concus and Golub '76, Widlund '78
—	Axelsson '79, '80
ORTHOMIN	Vinsome '76
ORTHORES	Young and Jea '80
ORTHODIR	Young and Jea '80
FOM	Saad '81
GCR	Elman '82 [6], Eisenstat et al. '83 [5]
GMRES	Saad and Schultz '86 [23]
III. Biorthogonalization methods	
BIOMIN = BCG	Lanczos '52 [16], Fletcher '76 [8]
BIORES = BO	Lanczos '50, Jea and Young '83
BIODIR	Jea and Young '83
BIOMIN^2 = CGS	Sonneveld '89 [25]
BIORES^2	Gutknecht '90 [12]
BIODIR^2	Gutknecht '90 [12]
BiCGSTAB	Van der Vorst '90 [30]
QMR	Freund '90 [9], [10]
IV. Other methods	
USYMLQ	Saunders, Simon, and Yip '88 [24]
USYMQR	Saunders, Simon, and Yip '88 [24]

various results concerning the matrix properties that control the convergence of CGN (§ 2), GMRES (§ 3), and CGS (§ 4). In particular, we show that the convergence of CGN depends on the singular values of A, whereas the convergence of GMRES and CGS depends on its eigenvalues (if A is close to normal) or pseudo-eigenvalues (if A is far from normal). Many of the results we present are already known, especially those connected with CGN, but the fundamental distinction between the roles of eigenvalues and singular values seems to be often overlooked.

These general considerations lead to the conclusion that CGN, GMRES, and CGS indeed differ fundamentally in capabilities. In § 5 we substantiate this claim by constructing simple, artificial examples which show that in certain circumstances each of these three iterations outperforms the others by a factor on the order of $\sqrt{N}$ or N or more. We emphasize that these examples are in no way intended to be representative of realistic computations. They are offered entirely for the insight they provide.

Section 6 discusses the relationship between convergence rates and the properties of the symmetric part of a matrix, or as we prefer to think of it, the field of values. Using the examples of § 5 for illustration, we argue that a well-behaved symmetric part is neither

necessary nor sufficient for rapid convergence, and that therefore, considering the symmetric part is not a reliable way to analyze iterative methods.

Our discussion of all of these iterations is intentionally simplified. We largely ignore many important issues such as sparsity and other structure, machine architecture, rounding errors, storage limitations, the effect of truncation or restarts, and the possibility of hybrid Krylov space iterations, which in some cases may be the fastest of all [17]. Most important, we ignore the issue of preconditioning, without which all of these methods are often useless (see Example R, below). For a broader view of matrix iterations the reader should consult references such as [13], [21], and [22]. For an empirical comparison of CGN, GMRES, and CGS, see [19].

Throughout the paper we use the following standard notation:

$\|\ \|$ = 2-norm,
N = dimension of A,
Λ = spectrum of A,
Σ = set of singular values of A,
A^* = conjugate transpose of A,
x_0 = initial guess,
x_n = nth iterate,
$e_n = A^{-1}b - x_n$ = nth error,
$r_n = b - Ax_n = Ae_n$ = nth residual.

For any set $S \subseteq \mathbb{C}$ and function $f(z)$ defined on S we shall also find it convenient to write

$$\|f\|_S = \sup_{z \in S} |f(z)|.$$

CGN, GMRES, and CGS can each be described in a few lines of pseudocode—or programmed in a few lines of Matlab. The formulas are given in Fig. 1.

2. CGN. Perhaps the most obvious nonsymmetric iterative method is the application of the conjugate gradient iteration to the normal equations

$$A^*Ax = A^*b, \tag{2.1}$$

an idea that dates to the original CG paper by Hestenes and Stiefel [14]. (Of course, A^*A is never formed explicitly.) This algorithm, which we shall call CGN,[1] constructs the unique sequence of vectors

$$x_n \in x_0 + \langle A^*r_0, (A^*A)A^*r_0, \cdots, (A^*A)^{n-1}A^*r_0 \rangle \tag{2.2}$$

with minimal residual at each step:

$$\|r_n\| = \text{minimum}. \tag{2.3}$$

A statement equivalent to (2.3) is the orthogonality condition

$$r_n \perp \langle AA^*r_0, (AA^*)^2r_0, \cdots, (AA^*)^nr_0 \rangle. \tag{2.4}$$

The beauty of this algorithm is that thanks to the CG connection, x_n can be found by a three-term recurrence relation. For details, see [6].

[1] Though algorithms of the normal equations type are usually based on (2.1), an alternative (often called Craig's method) is the sequence $AA^*y = b$, $x = A^*y$. There is no universal agreement on names for these algorithms, but the most common choices are CGNR and CGNE, respectively. In this paper we use the neutral term CGN in place of CGNR, since except for a few details, most of what we say applies to CGNE as well.

CGN

$\beta_0 := 0;\ p_0 := 0$
For $n := 1, 2, \cdots$
$p_n := A^* r_{n-1} + \beta_{n-1} p_{n-1}$
$\alpha_n := \|A^* r_{n-1}\|^2 / \|A p_n\|^2$
$x_n := x_{n-1} + \alpha_n p_n$
$r_n := r_{n-1} - \alpha_n A p_n$
$\beta_n := \|A^* r_n\|^2 / \|A^* r_{n-1}\|^2$

GMRES

$v_1 := r_0 / \|r_0\|;\ e_1 = (1, 0, 0, \cdots)^T$
For $n := 1, 2, \cdots$
For $j := 1, \cdots, n$
$h_{jn} := v_j^* A v_n$
$\hat{v}_{n+1} := A v_n - \sum_{j=1}^{n} h_{jn} v_j$
$h_{n+1,n} := \|\hat{v}_{n+1}\|$
$v_{n+1} := \hat{v}_{n+1} / h_{n+1,n}$.
$y_n :=$ least-squares solution to $H_n y_n \approx e_1 \|r_0\|$
$x_n := x_0 + \sum_{j=1}^{n} (y_n)_j v_j$

CGS

$q_0 := p_0 := 0;\ \rho_0 := 1,\ \tilde{r}_0 = r_0$ or some other choice
For $n := 1, 2, \cdots$
$\rho_n := \tilde{r}_0^* r_{n-1}$
$\beta_n := \rho_n / \rho_{n-1}$
$u_n := r_{n-1} + \beta_n q_{n-1}$
$p_n := u_n + \beta_n (q_{n-1} + \beta_n p_{n-1})$
$v_n := A p_n$
$\sigma_n := \tilde{r}_0^* v_n$
$\alpha_n := \rho_n / \sigma_n$
$q_n := u_n - \alpha_n v_n$
$r_n := r_{n-1} - \alpha_n A (u_n + q_n)$
$x_n := x_{n-1} + \alpha_n (u_n + q_n)$

FIG. 1. CGN, GMRES, *and* CGS. *Each iteration begins with an initial guess* x_0 *and initial residual* $r_0 = b - Ax_0$. *See* [23] *for details of more efficient implementations of* GMRES *and* [18] *for the* LSQR *implementation of* CGN.

To investigate convergence rates we note that at each step we have

$$x_n = x_0 + q_{n-1}(A^*A) A^* r_0 \tag{2.5}$$

for some polynomial q_{n-1} of degree $n - 1$. Subtracting this equation from the exact solution $A^{-1}b$ gives

$$e_n = p_n(A^*A) e_0 \tag{2.6}$$

for some polynomial $p_n(z) = 1 - z q_{n-1}(z)$ of degree n with $p_n(0) = 1$. Since $r_n = A e_n$ and $A p_n(A^*A) = p_n(AA^*)A$, multiplying (2.6) by A gives

$$r_n = p_n(AA^*) r_0 \tag{2.7}$$

for the same polynomial $p_n(z)$. We conclude that convergence will be rapid if and only if polynomials p_n exist for which $\|p_n(AA^*) r_0\|$ decreases rapidly, and a sufficient condition for this is that $\|p_n(AA^*)\|$ should decrease rapidly. Exact convergence in exact arithmetic occurs in at most n steps if the degree of the minimal polynomial of AA^* is n. Convergence to a tolerance ε occurs if AA^* has a "pseudominimal polynomial" p_n with $p_n(0) = 1$ and $\|p_n(AA^*)\| \leqq \varepsilon$.

At this point singular values enter into the picture. Since AA^* is a normal matrix with spectrum Σ^2, we have

$$\|p_n(AA^*)\| = \|p_n\|_{\Sigma^2} \tag{2.8}$$

for any polynomial p_n, where we have defined $\|p_n\|_{\Sigma^2} = \sup_{z\in\Sigma^2} |p_n(z)|$ as mentioned in the Introduction. In other words, the rate of convergence of CGN is determined by the real approximation problem of minimizing $\|p_n\|_{\Sigma^2}$ subject to $p_n(0) = 1$. We have proved the following theorem.

THEOREM 1. *For the* CGN *iteration applied to an arbitrary matrix* A,

$$\frac{\|r_n\|}{\|r_0\|} \leqq \inf_{\substack{p_n\in P_n \\ p_n(0)=1}} \|p_n\|_{\Sigma^2}. \tag{2.9}$$

Greenbaum has shown that for each n, there exists an initial residual r_0 such that equality in (2.9) is attained [11]. Thus this theorem describes the upper envelope of the convergence curves corresponding to all possible initial guesses for the CGN iteration applied to a fixed matrix A and right-hand side b. Particular initial guesses make observed convergence curves lie below the envelope, but the improvement is rarely dramatic.

We emphasize that *the convergence of* CGN *is determined solely by the singular values of* A. Any two matrices with the same singular values have identical worst-case convergence rates.[2] If A is normal, the moduli of the eigenvalues are equal to the singular values, but the arguments of the eigenvalues are irrelevant to convergence. If A is not normal, convergence rates cannot be determined from eigenvalues alone.

One choice of a polynomial p_n in (2.9) is the Chebyshev polynomial T_n transplanted to the interval $[\sigma_{\min}^2, \sigma_{\max}^2]$ and normalized by $p_n(0) = 1$, where $\sigma_{\min}$ and $\sigma_{\max}$ denote the extreme singular values of A. Elementary estimates lead from here to the familiar corollary

$$\frac{\|r_n\|}{\|r_0\|} \leqq 2\left(\frac{\kappa-1}{\kappa+1}\right)^n, \tag{2.10}$$

where $\kappa = \sigma_{\max}/\sigma_{\min}$ is the condition number of A. Thus, loosely speaking, CGN converges in at most $O(\kappa)$ iterations. Unlike (2.9), however, this inequality is far from sharp in general, unless the singular values of A are smoothly distributed.

Another choice of p_n is a product of two polynomials p_k and p_{n-k} of lower degree. Together with Greenbaum's sharp form of Theorem 1, this yields another corollary of Theorem 1:

$$\frac{\|r_n\|}{\|r_0\|} \leqq \frac{\|r_k\|}{\|r_0\|}\left(\sup_{\tilde{r}_0\in\mathbb{C}^N} \frac{\|\tilde{r}_{n-k}\|}{\|\tilde{r}_0\|}\right) \tag{2.11}$$

for any $k \leqq n$. To put it in words: the envelope described by (2.9) is concave downwards, so the convergence of CGN tends to accelerate in the course of the iteration.

The convergence of CGN is strictly monotonic:

$$\|r_{n+1}\| < \|r_n\| \quad \text{if } \|r_n\| > 0. \tag{2.12}$$

---

[2] In fact the determining effect of the singular values applies to all initial vectors, not just to the worst case; see the penultimate paragraph of § 3.

One of the many ways to prove this is to note that for sufficiently small ε, $\|I - \varepsilon AA^*\| < 1$ (see § 6). Equation (2.12) follows since $p_{n+1}(z)$ must be at least as good as the product of $p_n(z)$ and the monomial $1 - \varepsilon z$.

The results of this section are essentially all known. In particular, theorems related to (2.11) can be found in [28].

3. GMRES. Residual minimization methods minimize the residual in a simpler Krylov space at the price of more arithmetic. They construct the unique sequence $\{x_n\}$ with

$$x_n \in x_0 + \langle r_0, Ar_0, \cdots, A^{n-1} r_0 \rangle \tag{3.1}$$

satisfying

$$\|r_n\| = \text{minimum}. \tag{3.2}$$

An equivalent statement is the orthogonality condition

$$r_n \perp \langle Ar_0, A^2 r_0, \cdots, A^n r_0 \rangle. \tag{3.3}$$

This condition is implemented by "brute force" in the sense that at the nth step, linear combinations of n vectors are manipulated. The GMRES iteration is a robust implementation of (3.1)–(3.3) by means of an Arnoldi construction of an orthonormal basis for the Krylov space, which leads to an $(n + 1) \times n$ Hessenberg least-squares problem [23]. At each step we have

$$e_n = p_n(A)e_0, \qquad r_n = p_n(A)r_0, \tag{3.4}$$

where $p_n(z)$ is a polynomial of degree n with $p_n(0) = 1$. Convergence will be rapid if and only if polynomials p_n exist for which $\|p_n(A)r_0\|$ decreases rapidly, and a sufficient condition for this is that $\|p_n(A)\|$ should decrease rapidly. Convergence to a tolerance ε occurs in n steps if there exists a polynomial p_n with $p_n(0) = 1$ and $\|p_n(A)\| \leqq \varepsilon$.

These formulas lead us to look at eigenvalues rather than singular values. If A is a normal matrix with spectrum Λ, then for any polynomial p_n,

$$\|p_n(A)\| = \|p_n\|_\Lambda. \tag{3.5}$$

From this we obtain the following analogue of Theorem 1.

THEOREM 2. *For the* GMRES *iteration applied to a normal matrix* A,

$$\frac{\|r_n\|}{\|r_0\|} \leqq \inf_{\substack{p_n \in P_n \\ p_n(0)=1}} \|p_n\|_\Lambda. \tag{3.6}$$

As in Theorem 1, we expect that this bound will be reasonably sharp in practice, though it is not known that equality need be attained for any r_0. Thus if A is normal, the convergence of GMRES is determined by the eigenvalues of A via the complex approximation problem of minimizing $\|p_n\|_\Lambda$ subject to $p_n(0) = 1$. Complex approximation problems are harder than real ones, and no convergence bound as memorable as (2.10) results. Equation (2.11), on the other hand, carries over to this case without modification.

Unfortunately, nonsymmetric matrices are rarely normal. Two methods of analysis of the convergence of GMRES for general matrices have been proposed. The first, the standard approach in the literature, is based on the assumption that A is not too far from normal. For any matrix A that can be diagonalized as $A = V\Lambda V^{-1}$, the natural generalization of (2.8) is

$$\|p_n\|_\Lambda \leqq \|p_n(A)\| \leqq \kappa(V)\|p_n\|_\Lambda. \tag{3.7}$$

Combining (3.4) and (3.7) gives the following theorem.

THEOREM 3. *For the* GMRES *iteration applied to a diagonalizable matrix* A,

$$\frac{\|r_n\|}{\|r_0\|} \leqq \kappa(V) \inf_{\substack{p_n \in P_n \\ p_n(0)=1}} \|p_n\|_\Lambda, \tag{3.8}$$

where $\kappa(V)$ *is the condition number of any matrix of eigenvectors of* A.

This theorem indicates that if $\kappa(V)$ is not too large, it is still a reasonable approximation to say that the convergence of GMRES is determined by the eigenvalues of A.

The second approach is motivated by matrices for which $\kappa(V)$ is huge or infinite, that is, matrices whose eigenvalues are highly sensitive to small perturbations in the matrix entries. Let $\Lambda_\varepsilon \supseteq \Lambda$ denote the ε-pseudospectrum of A, i.e., its set of ε-pseudo-eigenvalues: those points $z \in \mathbb{C}$ that are eigenvalues of some matrix $A + E$ with $\|E\| \leqq \varepsilon$ or, equivalently, those points $z \in \mathbb{C}$ with $\|(zI - A)^{-1}\| \geqq \varepsilon^{-1}$. Let L be the arc length of the boundary $\partial\Lambda_\varepsilon$. By a contour integral we can readily show that

$$\|p_n\|_\Lambda \leqq \|p_n(A)\| \leqq \frac{L}{2\pi\varepsilon} \|p_n\|_{\Lambda_\varepsilon} \tag{3.9}$$

for any $\varepsilon > 0$ [26]. This inequality leads to the following theorem.

THEOREM 4. *For the* GMRES *iteration applied to an arbitrary matrix* A,

$$\frac{\|r_n\|}{\|r_0\|} \leqq \frac{L}{2\pi\varepsilon} \inf_{\substack{p_n \in P_n \\ p_n(0)=1}} \|p_n\|_{\Lambda_\varepsilon} \tag{3.10}$$

for any $\varepsilon > 0$.

Loosely speaking, if A is far from normal, then the convergence of GMRES depends on polynomial approximation problems defined on the pseudospectra, not just the spectrum. See [17], [26], and [27] for examples and further discussion of this phenomenon.

The convergence of GMRES, unlike CGN, is not always strictly monotonic; we can have $\|r_{n+1}\| = \|r_n\|$. A necessary and sufficient condition for strict monotonicity at every step n (and for all r_0) is that the field of values of A should lie in an open half-plane with respect to the origin. This half-plane condition is discussed further in § 6.

Neither Theorem 3 nor Theorem 4 is sharp, nor necessarily close to sharp even for worst-case initial residuals r_0. To the best of our knowledge the convergence of GMRES, unlike that of CGN, cannot be reduced completely to a problem in approximation theory.

It is readily shown that if A and $\tilde{A}$ are unitarily similar, then their behaviors under GMRES are identical in the sense that there exists a bijection $r_0 \mapsto \tilde{r}_0$ on $\mathbb{C}^N$ such that the convergence curve for A with initial vector r_0 is the same as the convergence curve for $\tilde{A}$ with initial vector $\tilde{r}_0$. The analogous statement for CGN would be that the behaviors of A and $\tilde{A}$ under CGN are identical in the same sense if AA^* and $\tilde{A}\tilde{A}^*$ are unitarily similar, which is equivalent to A and $\tilde{A}$ having the same singular values. See the remarks following Theorem 1 in § 2.

We cannot complete a discussion of GMRES without mentioning the important point that in practice, residual minimization methods are usually not applied in the "pure" form described above. To keep storage requirements under control, GMRES is often restarted after each k steps for some integer k (e.g., 5 or 10 or 20), and ORTHOMIN is generally truncated in a different but analogous way so that the algorithm works always with a k-dimensional Krylov substance. Besides the desire to keep the discussion simple, we have avoided mentioning this issue because we believe that restarting or truncating

these iterations is not an entirely satisfactory idea, since the resulting algorithms tend to spend a great deal of time relearning information obtained in previous cycles. For a discussion of this point, see [17], where we advocate the use of hybrid methods instead.

4. BCG and CGS. The BCG, or biconjugate gradient iteration, constructs non-optimal approximations in the same Krylov subspace as GMRES, but with less work per step [8], [16]. Thus, like GMRES, BCG constructs a sequence of vectors

$$x_n \in x_0 + \langle r_0, Ar_0, \cdots, A^{n-1} r_0 \rangle, \tag{4.1}$$

which implies

$$e_n = p_n(A)e_0, \qquad r_n = p_n(A)r_0 \tag{4.2}$$

for some polynomial p_n of degree n. The difference is that instead of (3.3), p_n is now determined by the orthogonality condition

$$r_n \perp \langle \tilde{r}_0, A^*\tilde{r}_0, \cdots, (A^*)^{n-1}\tilde{r}_0 \rangle, \tag{4.3}$$

where $\tilde{r}_0 \in \mathbb{C}^N$ is a vector often taken equal to r_0. Since GMRES is optimal in the sense of (3.2), BCG can never outperform it if one measures performance by the number of iterations required to reduce $\|r_n\|$ by a certain amount. However, BCG computes its choice of x_n by three-term recurrence relations. Consequently the nth step of BCG requires $O(1)$ vector operations rather than the $O(n)$ vector operations required by GMRES, making it potentially much faster in total work. Equally important, the amount of storage required does not grow with n.

CGS, which stands for "CG squared," is a modification of BCG due to Sonneveld [25]. Sonneveld's observation is that by reorganizing the BCG algorithm in a certain way one can replace (4.2) by

$$e_n = p_n^2(A)e_0, \qquad r_n = p_n^2(A)r_0 \tag{4.4}$$

for the *same* polynomial p_n, with no increase in the amount of work per step. Furthermore, whereas BCG (like CGN) requires vector multiplications by both A and A^*, which may be awkward for certain sparse data structures or parallel machines, or may be impossible when matrix-free algorithms are in use, CGS only requires multiplications by A.

We will not give further details of these algorithms or much information about their convergence properties, which are less well understood than for CGN and GMRES. For discussion of these matters, including remarkable connections with orthogonal polynomials, continued fractions, Padé approximation, and the *qd* algorithm, see [2], [12], [20], and [29]. The following remarks, most of which can be derived from the description above, will suffice.

First, thanks to (4.4), CGS typically converges (or diverges) faster than BCG by a factor of between 1 and 2.

Second, except for that factor of 2, CGS can outperform GMRES in total work but not in number of iterations. In fact, at each step we obviously have that

$$\|r_n^{\mathrm{GMRES}}\| \leqq \|r_n^{\mathrm{BCG}}\|, \qquad \|r_{2n}^{\mathrm{GMRES}}\| \leqq \|r_n^{\mathrm{CGS}}\| \tag{4.5}$$

if all three methods begin with the same r_0, regardless of the choice of $\tilde{r}_0$.

Third, for a symmetric matrix and $\tilde{r}_0 = r_0$, BCG reduces to the CG iteration [8].

Finally, far from converging monotonically, BCG and CGS are susceptible to the possibility of breakdown—division by zero—if $\rho_{n-1} = 0$ or $\sigma_n = 0$ at some step (see Fig. 1). Breakdown will not occur in the generic case, but numerical analysts are well trained to expect that where infinities may arise with probability zero, numbers large enough to

be troublesome in floating-point arithmetic are likely to appear more often than that. Moreover, as our example S below will show, the mere requirement that r_0 and $\tilde{r}_0$ be real is enough to guarantee breakdown in certain cases. In the face of such reasonable grounds for suspicion, it is remarkable how frequently BCG and CGS turn out to be effective.

Various results are known about conditions under which BCG and CGS break down or converge exactly, assuming exact arithmetic [20], [12]. For example, it can be shown that if GMRES obtains the exact solution at a certain step n, then BCG and CGS do the same if they do not break down [20]. Unfortunately, much less is known about what matters in practice: approximate breakdown and approximate convergence.

5. Eight examples. So much for the generalities. Now back to the original questions: how different are CGN, GMRES, and CGS, and when? What convergence curves—$\log \|r_n\|$ as a function of n—are possible?

To show that none of these algorithms is dispensable, three examples would suffice. As our goal has been to learn as much as possible in the process, however, we have actually constructed $2^3 = 8$ examples in an attempt to nail down the space of matrices at every corner. Table 2 summarizes these examples by listing numbers of iterations—not work estimates. For CGN and CGS the two are proportional, but for GMRES the work per step increases linearly with the number of iterations if the matrix is sparse, and so does the storage. Thus if a sparse matrix requires $O(\sqrt{N})$ iterations for both GMRES and CGS, CGS is the winner in both work and storage by a factor $O(\sqrt{N})$.

GMRES and CGS construct iterates in essentially the same Krylov space and are relatively hard to distinguish. Therefore, we begin the discussion with the first four examples in the table, for which these two behave comparably. With each example we present a computed convergence curve corresponding to dimension $N = 40$, except in two cases with $N = 400$, and a random real initial vector x_0 and right-hand side b with independent normally distributed elements of mean 0 and variance 1. Bigger dimensions do not change the curves significantly. For CGS we take $\tilde{r}_0 = r_0$, except in Example $B_{\pm 1}$.

To fully explain these experiments we mention that the curves plotted below represent actual residuals, not residual estimates computed by the iterative algorithm; as it happens, in these examples it makes little difference. Plots of errors rather than residuals also look qualitatively similar for these examples.

Example I: all methods good (*Fig.* 2). By Theorem 1, CGN converges in one step (for all initial data) if and only if all the singular values of A are equal, that is, if and

TABLE 2

*Numbers of iterations required for convergence to a fixed precision for our eight example matrices for worst-case initial residuals. * denotes divergence.*

Name of matrix	CGN	GMRES	CGS	
I	1	1	1	all methods good
R	N	N	N	all methods bad
C	1	N	N	CGN wins
B_1	N	2	2	CGN loses
D	N	$2\sqrt{N}$	$\sqrt{N}$	CGS wins
S	1	2	*	CGS loses
$B_{\pm 1}$	N	2	*	GMRES wins
B_κ	2	$2\sqrt{N}$	$\sqrt{N}$	GMRES loses

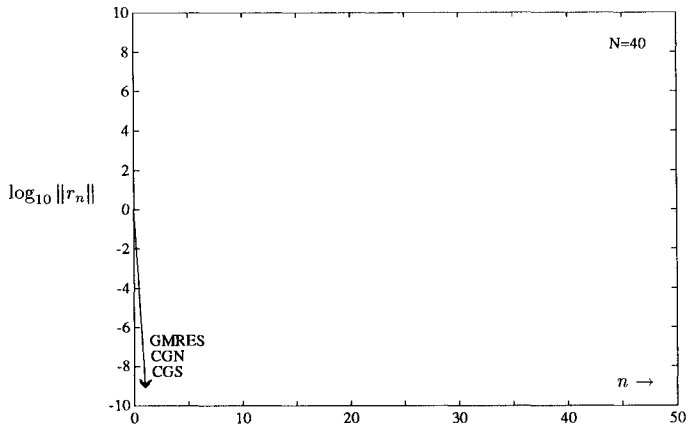

FIG. 2. *Example I (identity). All three iterations converge in one step.*

only if A is a multiple of an orthogonal matrix. By a slight extension of Theorem 3, GMRES converges in one step (and CGS also, by the remark at the end of § 4) if and only if A is diagonalizable and all its eigenvalues are equal, that is, if and only if A is a multiple of the identity. Since the identity is orthogonal, the latter condition implies the former, and these conditions are simultaneously satisfied if and only if A is a scalar multiple of the identity. Thus up to a scale factor there is a unique matrix that is handled perfectly by CGN, GMRES, and CGS: $A = I$.

Example R: all methods bad (Fig. 3). The opposite extreme would be a matrix for which all three iterations made no progress whatever until step N. By (2.12) no such example exists, but we can easily find a matrix for which all three algorithms make negligible progress until step N. By Theorems 1 and 2 any normal matrix with suitably troublesome eigenvalues and singular values will suffice, such as $A = \text{diag}(1, 4, 9, \cdots, N^2)$. For a more interesting example, consider a random matrix R of dimension N. To be precise (although the details are not very important), let the elements of R be independent normally distributed random numbers with mean 0 and variance 1. Such a matrix has condition number $O(N)$ on average and smoothly distributed singular values [4], so by

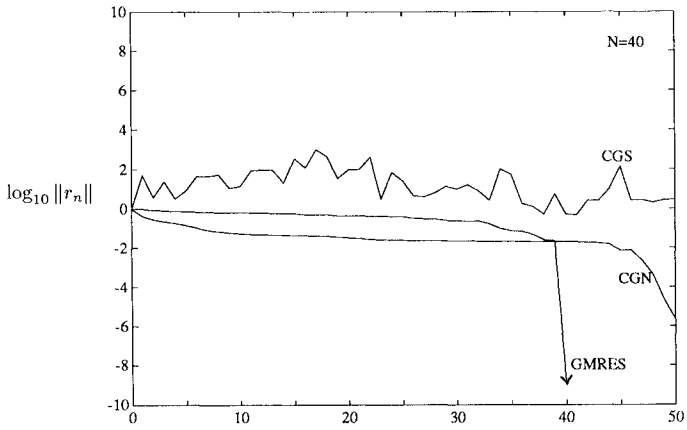

FIG. 3. *Example R (random). All three iterations require N steps.*

Theorem 1, CGN will require N steps for convergence. The eigenvalues are approximately uniformly distributed in a disk of radius $\sqrt{N}$ about the origin, suggesting that GMRES and CGS will also require N steps. In other words, *no known iterative method solves random matrix problems in better than $O(N)$ iterations*. (It would certainly be startling if this were not true, since such an iteration would beat Gaussian elimination on average even in the absence of preconditioning.) These predictions are confirmed by the experiment presented in Fig. 3. Note that the CGS convergence curve is wiggly, while the other two are monotonic, and that only GMRES exhibits the convergence in N steps that would be achieved by all three methods in exact arithmetic.

Example C: CGN *wins* (*Fig.* 4). Suppose we want a matrix for which CGN converges in one step but GMRES and CGS make no progress at all (for worst-case initial data) until step N. As mentioned above, the first requirement will be met if and only if A is a multiple of an orthogonal matrix. For the second, we must have $r_0 = r_1 = \cdots = r_{N-1}$, or by (3.3) and (4.3), $r_0 \perp \langle Ar_0, A^2r_0, \cdots, A^{N-1}r_0 \rangle$ and $r_0 \perp \langle \tilde{r}_0, A^*\tilde{r}_0, \cdots, (A^*)^{N-2}\tilde{r}_0 \rangle$. These conditions are simultaneously satisfied for suitable r_0 if A is a multiple of an orthogonal matrix with minimal polynomial $1 - z^N$, such as the circulant matrix

$$C = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & 1 & \\ & & & 0 & 1 \\ 1 & & & & 0 \end{pmatrix} \quad (N \times N). \tag{5.1}$$

It is obvious why this matrix is indigestible by GMRES and CGS: C represents a circulant shift upwards by one position, while C^{-1} is a circulant shift downwards. It takes $N - 1$ shifts in one direction to approximate a single shift in the other direction, and thus Krylov spaces provide very poor approximations. This example has been mentioned before by Brown [3], van der Vorst [29], and undoubtedly others.

Example B_1: CGN *loses* (*Fig.* 5). Now we want to reverse the pattern of the last example. As mentioned above, convergence in one step of GMRES and CGS implies that the matrix has just a single nondefective eigenvalue, hence is a multiple of the identity, entailing convergence in one step of CGN also. Thus a perfect example in this category cannot exist. However, a nearly perfect example can be found if we settle for

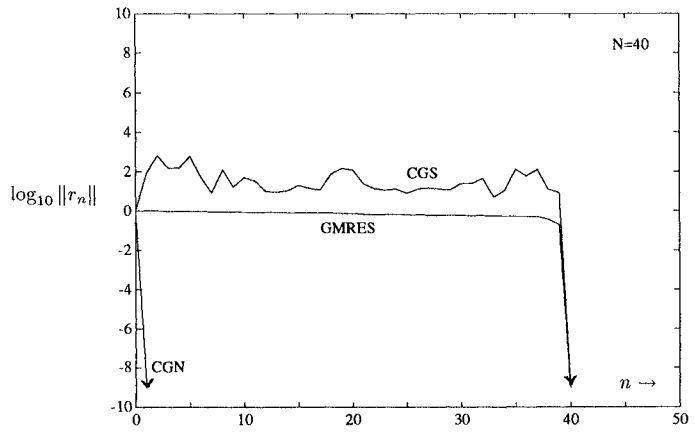

FIG. 4. *Example C (circulant shift)*. CGN *converges in one step, but* GMRES *and* CGS *require N steps.*

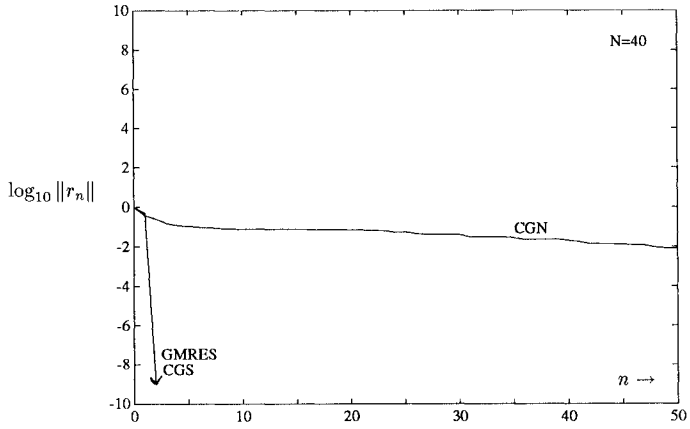

FIG. 5. *Example B_1 (block-diagonal matrix with eigenvalue 1).* CGN *requires N steps for convergence, but* GMRES *and* CGS *converge in two steps.*

convergence in *two* steps of GMRES and CGS. Thus we need a matrix whose minimal polynomial has degree 2 but which is ill-conditioned, with singular values spread over a wide range. Such an example is the block-diagonal matrix

$$B_1 = \begin{pmatrix} M_1 & & & & \\ & M_2 & & & \\ & & M_3 & & \\ & & & \ddots & \\ & & & & M_{N/2} \end{pmatrix} \quad (N \times N), \tag{5.2}$$

with

$$M_j = \begin{pmatrix} 1 & j-1 \\ 0 & 1 \end{pmatrix} \qquad 1 \le j \le N/2. \tag{5.3}$$

Obviously the minimal polynomial has degree 2, while the varying values of j ensure a troublesome distribution of singular values in the range approximately $[2/N, N/2]$. Incidentally, the diagonal elements of M_j might just as well have been taken to be any two numbers α and β of the same sign, so long as they remain the same in every block.

The four examples above show that CGN is sometimes better than GMRES and CGS by a factor $O(N)$ and sometimes worse by the same factor. This leaves us with the problem of distinguishing GMRES and CGS, which calls for examples of a different style. To make CGS look worse than GMRES, we construct examples in which CGS breaks down, at least for worst-case initial data. To make CGS look better than GMRES, we construct sparse examples in which both iterations take $O(\sqrt{N})$ steps, implying that the work and storage estimates for GMRES are $O(\sqrt{N})$ times larger. Alternatively, $O(\sqrt{N})$ may be replaced by a constant and these examples may be interpreted as showing that CGS may outperform GMRES by an arbitrary factor.

Example D: CGS *wins* (*Fig.* 6). For an example in this category it suffices to pick any diagonal matrix with condition number $\kappa = O(N)$ and smoothly distributed positive entries. BCG then behaves exactly like CG, requiring $O(\sqrt{N})$ iterations, since the condition number is $O(N)$, and GMRES behaves almost the same but not identically since it is

minimizing a different norm. CGS does better by at most a factor of 2. CGN, however, squares the condition number and requires $O(N)$ steps.

For a particularly clean version of this idea, define

$$D = \text{diag}\,(x_1, x_2, \cdots, x_N), \tag{5.4}$$

where $\{x_j\}$ denotes the set of Chebyshev extreme points scaled to the interval $[1, \kappa]$ for some $\kappa > 1$,

$$y_j = \cos\frac{(j-1)\pi}{N-1}, \quad x_j = 1 + \frac{1}{2}(y_j + 1)(\kappa - 1), \quad 1 \leqq j \leqq N. \tag{5.5}$$

Then we expect steady convergence of GMRES at the rate indicated by (2.10) with κ replaced by $\sqrt{\kappa}$, and convergence of CGS at about twice this rate. If we set

$$\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2\sqrt{N}} = \varepsilon, \quad \text{i.e., } \kappa = \left(\frac{1+\varepsilon^{1/2\sqrt{N}}}{1-\varepsilon^{1/2\sqrt{N}}}\right)^2, \tag{5.6}$$

then GMRES and CGS will converge to accuracy ε in about $2\sqrt{N}$ and $\sqrt{N}$ steps, respectively. Confirming this prediction, Fig. 6 shows the results of an experiment with $\varepsilon = 10^{-10}$ and dimension $N = 400$ rather than the usual $N = 40$.

Example S: CGS *loses* (*Fig.* 7). Let S be the skew-symmetric matrix

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \otimes I_{N/2}, \tag{5.7}$$

that is, an $N \times N$ block-diagonal matrix with 2×2 blocks. This matrix is normal and has eigenvalues $\pm i$ and singular value 1. Therefore, by Theorems 1 and 2, CGN converges in one step and GMRES in two steps, as shown in Fig. 7. On the other hand, CGS encounters a division by zero at the first step for any real initial vector r_0, assuming $\tilde{r}_0 = r_0$. If $\tilde{r}_0$ is chosen at random, the zero denominator is avoided generically and convergence is achieved in practice, but the expected result of that division remains infinite.

An analogous example, though essentially of dimension 4 rather than 2, has been discussed by Joubert [15].

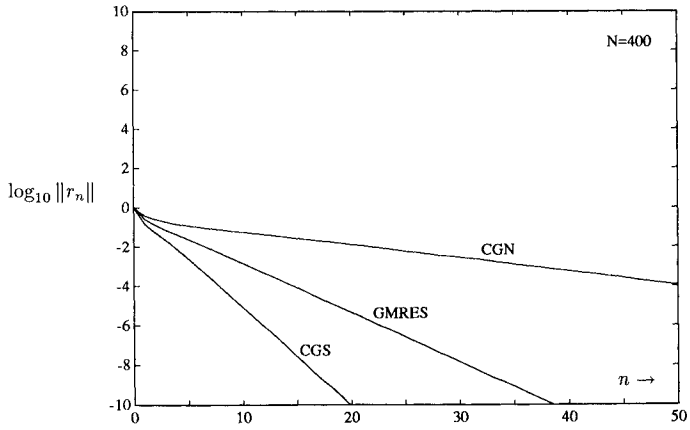

FIG. 6. *Example D (diagonal matrix with condition number N). The dimension is* $N = 400$. CGS *requires* $\sqrt{N}$ *steps for convergence, while* CGN *and* GMRES *require* $O(N)$ and $2\sqrt{N}$ *steps, respectively—hence a total work estimate in both cases comparable to* $O(N)$ *steps of* CGS.

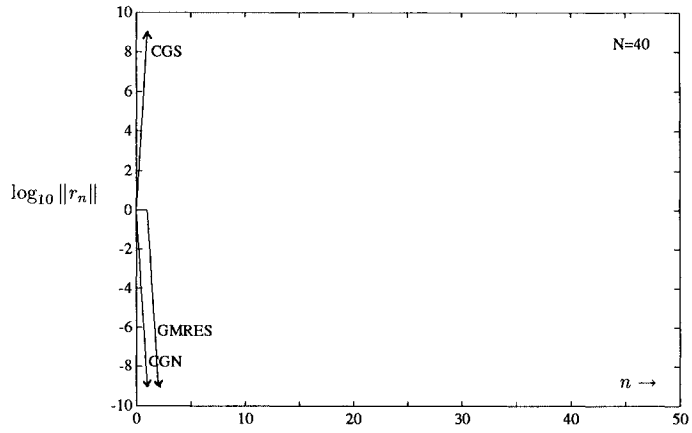

FIG. 7. *Example S (skew-symmetric).* CGS *breaks down at the first step, while* CGN *and* GMRES *converge in one and two steps, respectively.*

Example $B_{\pm 1}$: GMRES *wins* (*Fig.* 8). For this example we want a matrix like that of Fig. 5, except for which CGS breaks down. This is easily accomplished by defining a matrix $B_{\pm 1}$ by (5.2) but with (5.3) replaced by

$$M_j = \begin{pmatrix} 1 & j-1 \\ 0 & -1 \end{pmatrix}, \qquad 1 \leqq j \leqq N/2. \tag{5.8}$$

As with the matrix S above, CGS will encounter a division by zero at the first step if r_0 and $\tilde{r}_0$ are chosen appropriately, and this is what we have done in Fig. 8. Generically, however, this example does not break down.

Example B_κ: GMRES *loses* (*Fig.* 9). For this final example it is natural to modify the idea of matrices B_1 and $B_{\pm 1}$ again so that instead of fixed eigenvalues and varying singular values, we have fixed singular values and varying eigenvalues. In particular, let B_κ be defined as in (5.2) but with (5.3) replaced by

$$M_j = \begin{pmatrix} x_j & \gamma_j \\ 0 & \kappa/x_j \end{pmatrix}, \quad \gamma_j = (\kappa^2 + 1 - x_j^2 - \kappa^2/x_j^2)^{1/2}, \quad 1 \leqq j \leqq N/2, \tag{5.9}$$

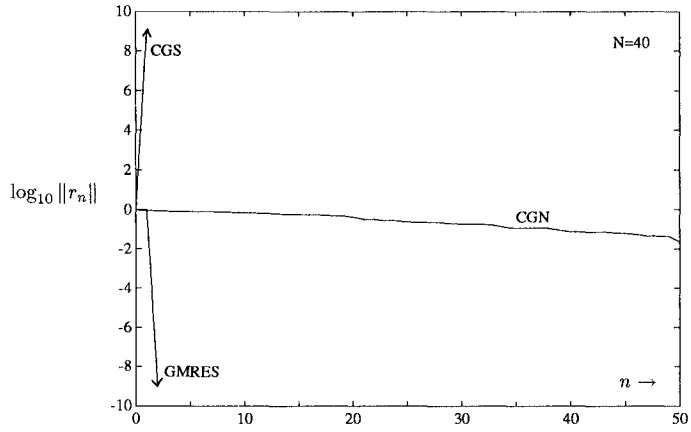

FIG. 8. *Example* $B_{\pm 1}$ *(block diagonal matrix with eigenvalues* ± 1*). To make* CGS *break down,* r_0 *and* $\tilde{r}_0$ *have been chosen diabolically.*

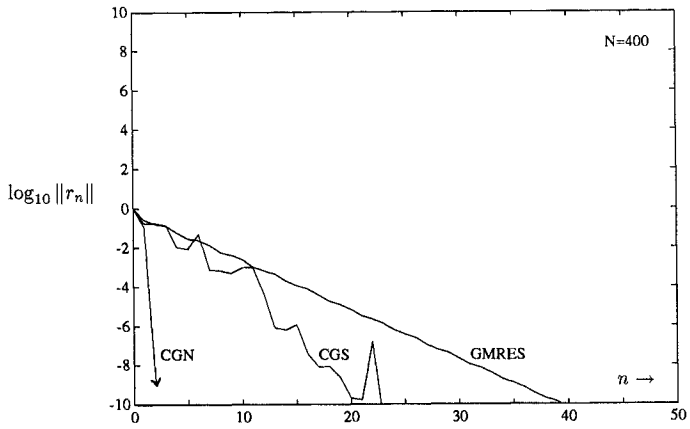

FIG. 9. *Example B_κ (block-diagonal matrix with singular values 1, κ). As in Fig. 6, the dimension is N = 400.* CGN *takes two steps for convergence,* CGS *takes $\sqrt{N}$ steps, and* GMRES *takes $2\sqrt{N}$ steps, for a total* GMRES *work estimate comparable to $O(N)$ steps of* CGS.

where $\{x_j\}$ are again Chebyshev points scaled to the interval $[1, \kappa]$ as in (5.5), but with N replaced by $N/2$. It is readily verified that each block M_j has the same singular values 1 and κ, whereas the eigenvalues lie throughout the interval $[1, \kappa]$. Taking again N = 400, $\varepsilon = 10^{-10}$, and κ defined by (5.6) gives the results shown in Fig. 9.

6. Symmetric parts and half-plane conditions. In the literature on nonsymmetric matrix iterations, much attention has been given to the behavior of the *symmetric* or more properly *Hermitian part* of a matrix, defined by $M = \frac{1}{2}(A + A^*)$. In particular, Eisenstat, Elman, and Schultz [5] and Elman [6] show that if M is positive definite, then various truncated and restarted Krylov space iterations are guaranteed to converge with a convergence rate bounded according to

$$\frac{\|r_n\|}{\|r_0\|} \leqq \left[1 - \frac{\lambda_{\min}(M)^2}{\sigma_{\max}(A)^2}\right]^{n/2}, \tag{6.1}$$

where $\sigma_{\max}(A) = \lambda_{\max}(A^*A)^{1/2}$ is the largest singular value of A. Among other algorithms, these results apply to GMRES (k) for any $k \geqq 1$, that is, GMRES restarted every k steps.

Theorems of this kind can be made rotationally invariant by restating them in terms of the *field of values* of a matrix, defined by $W = \{x^*Ax/x^*x,\ x \in \mathbb{C}^N\}$. The real part of W is equal to the interval $[\lambda_{\min}(M), \lambda_{\max}(M)]$, and therefore the statement that M is positive definite is equivalent to the statement that W lies in the open right half-plane. More generally, it is enough to assume that W lies in any open half-plane $\{z : \mathrm{Re}\,(e^{-i\theta}z) > 0\}$. We call this assumption the *half-plane condition*; it is also sometimes said that A is *definite*. The basis of these convergence theorems is the observation that the half-plane condition implies $\|1 - \varepsilon e^{-i\theta}A\| < 1$ for all sufficiently small $\varepsilon > 0$. The mathematics involved is the same as in standard results in numerical analysis on logarithmic norms, or in functional analysis, the Hille–Yosida theorem [27].

These theorems are important, but we believe they are of limited utility for choosing between iterative methods. The reason is that they are based on the relatively trivial case in which $k = 1$, analogous to a steepest-descent iteration; for $k > 2$ the half-plane condition is sufficient but not necessary for convergence. This fact is well known in principle, but nevertheless the opinion seems to have become widespread that the half-plane condition is what matters in practice. See, for example, [7] and [24].

To show that a well-behaved symmetric part is not necessary for rapid convergence of GMRES, it is enough to look at the matrices S, B_1, or $B_{\pm 1}$. For example, consider B_1. The field of values is the disk about $z = 1$ of radius $N/4$, which implies $\lambda_{\min}(M) = 1 - N/4$, $\lambda_{\max}(M) = 1 + N/4$. We could hardly be further from satisfying the half-plane condition, but GMRES converges in two steps.

Conversely, (6.1) shows that a sufficiently well-behaved symmetric part guarantees rapid convergence of GMRES. To show that mere positive definiteness of M is not enough, however, consider a normal matrix along the lines of the matrix C of (5.1), but with eigenvalues only at the roots of unity in the right half-plane. Since the condition number is 1, CGN converges in one step, whereas GMRES still requires many steps.

7. Conclusions and exhortation. Of the many parameter-free nonsymmetric matrix iterations proposed to date, we believe that CGN, GMRES, and CGS are the best. So far as we know, for calculations in exact arithmetic with performance measured by the residual norm $\|r_n\|$, no other iteration ever outperforms these three by more than a constant factor, except in certain examples involving special initial residuals r_0.

The convergence of CGN is determined by the singular values of A; the eigenvalues have nothing to do with it except insofar as they determine the singular values. If A is normal or close to normal, the convergence of GMRES is determined by the eigenvalues of A; the singular values, and in particular the condition number, have nothing to do with it. More precisely, by Theorems 1 and 2, the convergence of GMRES and CGN for a normal matrix depends on how well 0 can be approximated on the spectrum Λ by polynomials $p_n(z)$ and $p_n(r^2)$, respectively, with $p_n(0) = 1$ and $r = |z|$. It follows that we can expect CGN to be the winner if the singular values are clustered but the eigenvalues tend to surround the origin, whereas GMRES will be the winner if the eigenvalues are as tightly clustered as the singular values.

If A is far from normal, on the other hand, the convergence of GMRES becomes slower by a potentially unbounded factor than eigenvalues alone would suggest. In some such cases, the convergence is approximately determined by the pseudospectra of A instead.

The above statements about GMRES apply also, approximately, to CGS, but the convergence of CGS is affected additionally by instabilities that are not yet fully understood. When matrix-vector multiplications are much more expensive than vector operations and storage, CGS can outperform GMRES by at most a factor of 2. When the cost of vector operations and storage is significant, however, as is typical in sparse matrix calculations, Examples D and B_κ have established that CGS may outperform GMRES by a factor of order $\sqrt{N}$. Taken together, our examples show that CGN, GMRES, and CGS each outperform the others in some cases by factors of order $\sqrt{N}$ or N.

In summary, these three algorithms are genuinely distinct in their behavior. Until something better comes along, there is a place for all of them in scientific computing.

Having confined the discussion to generalities and contrived examples throughout the paper, we close with two editorial remarks of the more usual kind. First, we believe CGN is underrated. Despite the squaring of the condition number, this algorithm sometimes outperforms the competition; too many authors dismiss it with a flurry of rhetoric.[3] Second, CGS is a remarkable algorithm that deserves attention. It outperforms the more

[3] The "squaring of the condition number" we refer to is the fact that Σ^2 rather than Σ or Λ is what governs the convergence of CGN in exact arithmetic (Theorem 1). Whether rounding errors are amplified by a factor on the order of the square of the condition number is quite a different matter and is not discussed here. With the LSQR implementation, they need not be [18].

familiar BCG frequently by a factor of 1 to 2, and it converges in a number of iterations as low as GMRES far more often than the available theory might suggest. Yet, despite these impressive results, the convergence curves generated by CGS are frequently so erratic that it is hard to imagine that this algorithm can be completely right. We suspect an even better algorithm may be waiting to be discovered.[4]

CGN, GMRES, and CGS are so easy to program that there is little excuse for not taking the trouble to do so. We propose that until a fundamentally superior matrix iteration is invented, researchers in this field adopt the policy that no plot of convergence rates is complete unless it includes curves for CGN, GMRES, and CGS.

Acknowledgments. We have benefited from the generous advice of many people, including Peter Brown, Howard Elman, Roland Freund, Joe Grcar, Anne Greenbaum, Martin Gutknecht, David Keyes, Tom Manteuffel, Kapil Mathur, Olavi Nevanlinna, Chris Paige, Lothar Reichel, Youcef Saad, Michael Saunders, Robert Schreiber, Henk van der Vorst, and Jacob White. We especially thank Anne Greenbaum for pointing out an error in an earlier version of Theorems 1 and 2 and for several other valuable suggestions.

And none of this would have been any fun without Matlab.

REFERENCES

[1] S. F. ASHBY, T. A. MANTEUFFEL, AND P. E. SAYLOR, *A taxonomy for conjugate gradient methods*, SIAM J. Numer. Anal., 27 (1990), pp. 1542–1568.

[2] C. BREZINSKI, *Padé-Type Approximation and General Orthogonal Polynomials*, Birkhäuser-Verlag, Boston, 1980.

[3] P. BROWN, *A theoretical comparison of the Arnoldi and* GMRES *algorithms*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 58–78.

[4] A. EDELMAN, *Eigenvalues and condition numbers of random matrices*, Ph.D. thesis and Numer. Anal. Report 89-7, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[5] S. C. EISENSTAT, H. C. ELMAN, AND M. H. SCHULTZ, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 20 (1983), pp. 345–357.

[6] H. C. ELMAN, *Iterative methods for large, sparse, nonsymmetric systems of linear equations*, Ph.D. thesis and Res. Report #229, Department of Computer Science, Yale University, New Haven, CT, 1982.

[7] H. ELMAN AND R. STREIT, *Polynomial iteration for nonsymmetric indefinite linear systems*, in Numerical Analysis, J. D. Hennert, ed., Lecture Notes in Mathematics 1230, Springer-Verlag, Berlin, New York, 1986.

[8] R. FLETCHER, *Conjugate gradient methods for indefinite systems*, in Proc. of the Dundee Biennial Conference on Numerical Analysis, G. A. Watson, ed., Springer-Verlag, New York, 1975.

[9] R. W. FREUND, *Conjugate gradient-type methods for linear systems with complex symmetric coefficient matrices*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 425–448.

[10] R. W. FREUND AND N. M. NACHTIGAL, *An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices*, Part II, Tech. Report 90.46, Research Institute for Advanced Computer Science, NASA Ames Research Center, Moffett Field, CA, 1990.

[11] A. GREENBAUM, *Comparison of splittings used with the conjugate gradient algorithm*, Numer. Math., 33 (1979), pp. 181–194.

[12] M. H. GUTKNECHT, *The unsymmetric Lanczos algorithms and their relations to Padé approximation, continued fractions, and the qd algorithm*, Preliminary Proceedings, Copper Mountain Conference on Iterative Methods, 1990.

[4] Footnote added in revision. Since this paper was first submitted for publication, two notable additions to the field of CGS-type iterations have been introduced: the Bi-CGSTAB algorithm of van der Vorst [30] and the QMR algorithm of Freund [9], [10], both of which we inserted into Fig. 1 at the revision stage. It appears that these algorithms represent progress towards improving the erratic convergence of CGS. For a survey of these and related developments, see [31].

[13] L. A. HAGEMAN AND D. M. YOUNG, *Applied Iterative Methods*, Academic Press, New York, 1981.
[14] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 409–436.
[15] W. D. JOUBERT, *Generalized conjugate gradient and Lanczos methods for the solution of nonsymmetric systems of linear equations*, Ph.D. thesis and Report CNA-238, Center for Numerical Analysis, University of Texas, Austin, TX, 1990.
[16] C. LANCZOS, *Solution of systems of linear equations by minimized iteration*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 33–53.
[17] N. M. NACHTIGAL, L. REICHEL, AND L. N. TREFETHEN, *A hybrid* GMRES *algorithm for nonsymmetric linear systems*, SIAM J. Matrix Anal. Appl., this issue, pp. 796–825.
[18] C. C. PAIGE AND M. A. SAUNDERS, LSQR: *An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71.
[19] G. RADICATI, Y. ROBERT, AND S. SUCCI, *Iterative algorithms for the solution of nonsymmetric systems in the modelling of weak plasma turbulence*, J. Comp. Phys., 80 (1989), pp. 489–497.
[20] Y. SAAD, *The Lanczos biorthogonalization algorithm and other oblique projection methods for solving large unsymmetric systems*, SIAM J. Numer. Anal., 19 (1982), pp. 485–506.
[21] Y. SAAD, *Krylov subspace methods on supercomputers*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 1200–1232.
[22] Y. SAAD AND M. H. SCHULTZ, *Conjugate gradient-like algorithms for solving nonsymmetric linear systems*, Math. Comp., 44 (1985), pp. 417–424.
[23] Y. SAAD AND M. H. SCHULTZ, GMRES: *A generalized minimum residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comp., 7 (1986), pp. 856–869.
[24] M. A. SAUNDERS, H. D. SIMON, AND E. L. YIP, *Two conjugate-gradient-type methods for unsymmetric linear equations*, SIAM J. Numer. Anal., 25 (1988), pp. 927–940.
[25] P. SONNEVELD, CGS, *a fast Lanczos-type solver for nonsymmetric linear systems*, SIAM J. Sci. Statist. Comp., 10 (1989), pp. 36–52.
[26] 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.
[27] L. N. TREFETHEN, *Pseudospectra of matrices*, in Proc. 14th Dundee Biennial Conf. on Numer. Anal., D. F. Griffiths and G. A. Watson, eds., to appear.
[28] A. VAN DER SLUIS AND H. A. VAN DER VORST, *The rate of convergence of conjugate gradients*, Numer. Math., 48 (1986), pp. 543–560.
[29] H. A. VAN DER VORST, *The convergence behavior of some iterative solution methods*, in Proc. Fifth Internat. Symposium Numer. Meth. Engrg., Vol. 1, R. Gruber, J. Periaux, and R. P. Shaw, eds., Springer-Verlag, 1989.
[30] ———, *Bi*-CGSTAB: *A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 631–644.
[31] R. W. FREUND, G. H. GOLUB, AND N. M. NACHTIGAL, *Iterative solution of linear systems*, Acta Numerica, to appear.