SIAM J. SCI. STAT. COMPUT.
Vol. 12, No. 1, pp. 58–78, January 1991

A THEORETICAL COMPARISON OF THE ARNOLDI AND GMRES ALGORITHMS*

PETER N. BROWN†

Abstract. Two recently developed Krylov methods for solving linear systems are *Arnoldi's method* and the *Generalized Minimum Residual* (GMRES) *method.* The GMRES method has been considered superior to Arnoldi's method due in part to the fact that GMRES never breaks down in the way Arnoldi's algorithm can. However, it is shown that there is a relationship between breakdowns in the two methods. Specifically, it is shown that GMRES does exhibit breakdowns very similar to that of Arnoldi, often referred to as the "stagnation" of GMRES. A relationship between the norms of the residuals for Arnoldi and GMRES is also given which shows exactly how much larger the residual norm for Arnoldi is than that for GMRES. In general, the results in the paper suggest that if one of the methods performs poorly on a particular problem, then so will the other.

Key words. linear systems, iterative methods, stagnation, breakdown

AMS(MOS) subject classification. 65F10

1. Introduction.

Two recently developed Krylov methods for solving linear systems are *Arnoldi's method* [9] and the *Generalized Minimum Residual* (GMRES) *method* [12]. Both methods have been used successfully in applications, but no theoretical comparison has been done. We give here a theoretical analysis of the related performance aspects of the two methods. We note that Arnoldi and GMRES are equivalent to other methods for various special cases of the coefficient matrix A. For example, when A is symmetric and positive definite, Arnoldi is equivalent to the conjugate gradient (CG) method and GMRES is equivalent to the conjugate residual (CR) method. This then implies that the results presented here for Arnoldi and GMRES will also apply to the CG and CR algorithms, to the extent that they are relevant. Additionally, when the matrix A is symmetric and indefinite, Arnoldi is equivalent to the SYMMLQ algorithm and GMRES is equivalent to the MINRES algorithm. The results presented here then apply to these related methods, and are similar to those presented by Paige and Saunders in [8] for the SYMMLQ and MINRES algorithms.

One of the disadvantages of the Arnoldi algorithm is that it can break down and be unable to calculate an approximate solution. Partly for this reason, the GMRES method has been considered superior since it never breaks down in the way Arnoldi's algorithm can [12]. However, we show that GMRES does exhibit "breakdowns" very similar to that of Arnoldi, and that there is a relationship between breakdowns in the two methods. This type of breakdown for GMRES is often referred to as "stagnation." A second perceived disadvantage of Arnoldi is that it does not produce a sequence of approximations whose residual norms are nonincreasing, as GMRES does. We give a result which shows exactly how much larger the residual norm is for the Arnoldi algorithm than that for GMRES, when the approximations come from the same Krylov subspace. The results in the paper indicate that the performance of the two methods are related, and suggest that if one method performs poorly on a particular problem, then so will the other. This is consistent with the numerical results reported in Brown

* Received by the editors May 16, 1988; accepted for publication (in revised form) December 1, 1989. This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-Eng-48, and was supported by the Department of Energy Office of Energy Research, Applied Mathematical Sciences Research Program.

† Computing and Mathematics Research Division, L-316, Lawrence Livermore National Laboratory, Livermore, California 94550 (brown@lll-crg.llnl.gov).

and Hindmarsh [3].

In §2, we review the description of the Arnoldi and GMRES algorithms and discuss some of the implementation details of the methods. In §3, we discuss how the Arnoldi algorithm can break down, and we then show the existence of a similar "breakdown" in the GMRES algorithm. We also present several results which give a relationship between breakdowns in the two methods, and then in §4 we give the proofs of these results. In §5 we prove a relationship between the size of the norms of the residuals for the Arnoldi and GMRES methods, and then in §6 we consider some examples illustrating the results in the paper. Finally, in §7 we make some concluding remarks.

2. The Arnoldi and GMRES algorithms. In this section we will review the Krylov subspace methods under consideration, and discuss their main properties. We start with a brief description of the Arnoldi and GMRES algorithms. Next, we comment on some implementation details of Arnoldi and GMRES, and then briefly discuss restarted versions of the two methods.

We are interested in solving linear systems of the form

$$Ax = b, \tag{2.1}$$

where A is an $n \times n$ real matrix which is nonsingular, and x and b are vectors of length n. Both Arnoldi and GMRES are iterative in nature, and so generate a sequence of iterates converging to the solution of (2.1). If x_0 is an initial guess for the true solution of (2.1), then letting $x = x_0 + z$, we have the equivalent system

$$Az = r_0, \tag{2.2}$$

where $r_0 = b - Ax_0$ is the initial *residual.* Let K_m be the *Krylov subspace*

$$K_m \equiv \text{span}\{r_0, Ar_0, \cdots, A^{m-1}r_0\}.$$

Arnoldi's method and GMRES both find an approximate solution

$$x_m = x_0 + z_m \text{ with } z_m \in K_m, \tag{2.3}$$

such that

$$(b - Ax_m) \perp K_m \text{ (equivalently } (r_0 - Az_m) \perp K_m) \tag{2.4}$$

for Arnoldi's method, and

$$\|b - Ax_m\|_2 = \min_{x \in x_0 + K_m} \|b - Ax\|_2 \ \ (= \min_{z \in K_m} \|r_0 - Az\|_2) \tag{2.5}$$

for GMRES. (Note that (2.5) is equivalent to requiring $(b - Ax_m) \perp AK_m$.) Here, $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbf{R}^n$ and orthogonality is meant in the usual Euclidean sense.

The following algorithm is a version of the Arnoldi (GMRES) algorithm which generates orthonormal vectors $v_i, i = 1, 2, \cdots, m$, and then builds the vector x_m that satisfies (2.4) (or (2.5) for GMRES). In both algorithms, the v_i's are computed such that they form an orthonormal basis of the Krylov subspace K_m, where the first vector v_1 is obtained by normalizing r_0. The parameter ϵ below is a user-supplied tolerance on the norm of the residual which is used in deciding when to stop the iteration, and $(\cdot, \cdot)$ denotes the Euclidean inner product.

Algorithm: Arnoldi (GMRES)

(1) *Arnoldi process:*
- For an initial guess x_0, form $r_0 = b - Ax_0$.
- Compute $\beta = \|r_0\|_2$ and $v_1 = r_0/\beta$.
- For $j = 1, 2, \cdots$, do:

(a) Form Av_j and orthogonalize it against the previous $v_1, \cdots, v_j$ via

$$
\begin{aligned}
&h_{i,j} = (Av_j, v_i), \qquad i = 1, 2, \cdots, j, \\
&\hat{v}_{j+1} = Av_j - \sum_{i=1}^{j} h_{i,j} v_i, \\
&h_{j+1,j} = \|\hat{v}_{j+1}\|_2, \quad \text{and} \\
&v_{j+1} = \hat{v}_{j+1}/h_{j+1,j}.
\end{aligned}
\tag{2.6}
$$

(b) Compute the residual norm $\rho_j = \|b - Ax_j\|_2$ of the solution x_j that would be obtained if we stopped at this step.

(c) If $\rho_j \leq \epsilon$ set $m = j$ and go to (2).

(2) *Form the approximate solution:*

Arnoldi: Define H_m to be the $m \times m$ (Hessenberg) matrix whose nonzero entries are the coefficients $h_{i,j}$, $1 \leq i \leq \min\{j+1, m\}$, $1 \leq j \leq m$ and define $V_m \equiv [v_1, v_2, \cdots, v_m]$.
- Compute $x_m^A = x_0 + z_m^A$, where $z_m^A = \beta V_m H_m^{-1} e_1$.

GMRES: Define $\bar{H}_m$ to be the $(m+1) \times m$ (Hessenberg) matrix whose nonzero entries are the coefficients $h_{i,j}$, $1 \leq i \leq j+1$, $1 \leq j \leq m$ and define $V_m \equiv [v_1, v_2, \cdots, v_m]$.
- Find the vector y_m^{GM} which minimizes $\|\beta e_1 - \bar{H}_m y\|_2$ over all vectors y in $\mathbf{R}^m$, where $e_1 = [1, 0, \cdots, 0]^T$.
- Compute $x_m^{GM} = x_0 + V_m y_m^{GM}$.

The first stage of the algorithms above is an Arnoldi loop which builds an orthonormal basis $V_m = [v_1, v_2, \cdots, v_m]$ of the Krylov subspace K_m. If we denote by V_j the $n \times j$ matrix with column vectors $v_1, v_2, \cdots, v_j$, then it follows immediately from (2.6) that (see also [9] and [12])

$$
AV_m = V_m H_m + \hat{v}_{m+1} e_m^T, \tag{2.7}
$$

where $e_m = [0, \cdots, 0, 1]^T \in \mathbf{R}^m$. This relation can be rewritten as

$$
AV_m = V_{m+1} \bar{H}_m, \tag{2.8}
$$

and is crucial in the development of the Arnoldi and GMRES methods.

The details on the practical implementation of the methods above are discussed at length in [9], [3], and [12]. We discuss here only those details relevant to the current situation. First, step (2) above computes the approximate solution x_m^A or x_m^{GM} in $x_0 + K_m$ that solves either (2.4) or (2.5). For Arnoldi, first note that condition (2.4) is equivalent to

$$
V_m^T (b - Ax_m^A) = 0.
$$

Since x_m^A is in the affine space $x_0 + K_m$, we have $x_m^A = x_0 + V_m y$ for some $y \in \mathbf{R}^m$. The above system is then equivalent to

$$
V_m^T A V_m y = V_m^T r_0. \tag{2.9}
$$

Next, from (2.7) it follows that $H_m = V_m^T A V_m$, using the fact that $\hat{v}_{m+1}$ is orthogonal to K_m. Also, since $r_0 = \beta v_1$, the right-hand side reduces to $V_m^T r_0 = \beta e_1$. Hence, we must solve the m-dimensional linear system

$$H_m y = \beta e_1. \tag{2.10}$$

This is accomplished by finding an LU decomposition of the Hessenberg matrix H_m, and we denote the solution of (2.10) by y_m^A. The solution x_m^A is then given by $x_m^A = x_0 + V_m y_m^A$.

For the GMRES algorithm, first let $z = V_m y$ for $y \in \mathbf{R}^m$. Then $\|r_0 - Az\|_2 = \|\beta v_1 - AV_m y\|_2$. Using (2.8), we have

$$\begin{aligned} \|\beta v_1 - AV_m y\|_2 &= \|V_{m+1}(\beta e_1 - \bar{H}_m y)\|_2 \\ &= \|\beta e_1 - \bar{H}_m y\|_2, \end{aligned}$$

since V_{m+1} has orthonormal columns. We denote by y_m^{GM} the solution of the minimization problem

$$\min_{y \in \mathbf{R}^m} \|\beta e_1 - \bar{H}_m y\|_2. \tag{2.11}$$

Then the optimal x is given by $x_m^{GM} = x_0 + V_m y_m^{GM}$. Note that (2.11) is a least squares problem of size $m+1$, and its coefficient matrix is upper Hessenberg. Problem (2.11) is solved using a QR factorization of the matrix $\bar{H}_m$, and in §4 we give a detailed discussion of its solution. See [3], [9], [10], and [12] for more specifics about Arnoldi and GMRES.

One important property of both algorithms is that the residual norm ρ_j referred to in step (1) above does not require the computation of the approximate solution x_j at every step. Instead an inexpensive formula is updated at each step while the factorization of the Hessenberg matrix H_m or $\bar{H}_m$ is updated. We derive this result for GMRES in §4 while discussing the QR factorization of $\bar{H}_m$. For the Arnoldi algorithm, see [2], [9], and [10]. In §5 we also give an alternative formula for Arnoldi which uses a QR factorization of H_m.

Another aspect of the above algorithms we have not yet considered is the ability to restart the methods. In a typical implementation of the above Krylov methods, a maximum value of m is dictated by storage considerations. If we let $m_{\max}$ be this value, then it is possible that $m = m_{\max}$ in the Arnoldi process, and yet ρ_m is still greater than ϵ, the tolerance on the residual r_m. In this case, we can set x_0 equal to x_m^A (or x_m^{GM}) and restart the Arnoldi process, effectively restarting the Krylov method. The convergence of such a procedure is not always guaranteed, but the idea seems to work well in practice.

Finally, we note that the Arnoldi algorithm is theoretically equivalent to the conjugate gradient method when A is symmetric and positive definite, and to the Lanczos method for solving linear systems when A is symmetric [9]. The GMRES algorithm is theoretically equivalent to the conjugate residual method when A is symmetric and positive definite, to GCR [4], [5], and to ORTHODIR [7] when A is only positive definite. Both Arnoldi and GMRES are guaranteed to obtain the exact solution of (2.1) in at most n steps using exact arithmetic, for any nonsingular $n \times n$ matrix A and right-hand side vector b.

3. Breakdowns. In this section we discuss how the Arnoldi and GMRES algorithms can break down, and then present several relationships between breakdowns in the two methods. The proofs of these results are then given in the following section.

We first discuss the two ways in which the Arnoldi algorithm can break down. This can happen if either $\hat{v}_{m+1} = 0$ so that v_{m+1} cannot be formed, or $H_{m_{\max}}$ is singular, which means that the maximum number of Arnoldi steps has been taken, but the final iterate cannot be formed. The first situation has often been referred to as a *happy* breakdown, since $\hat{v}_{m+1} = 0$ implies H_m is nonsingular and x_m^A is the exact solution of (2.1) (cf. Brown [1] or Saad [9]). The second case is more serious in that it causes a convergence failure of the algorithm. One possible recourse is to hope that H_j is nonsingular for some j among $1, \cdots, m_{\max} - 1$. If such a j exists, we can compute x_j^A and then restart the algorithm using x_j^A as the new initial guess x_0. It may not always be possible to do this, however, as the following example shows.

Example 3.1. Let A be the permutation matrix sending $e_1 \to e_2 \to \cdots \to e_n \to e_1$, where e_i is the ith standard basis vector in $\mathbf{R}^n$. Then

$$A = \begin{pmatrix} 0 & & & & 1 \\ 1 & \ddots & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & 1 & 0 \end{pmatrix}.$$

Consider solving (2.1), where $b = e_1$ and $x_0 = 0$. Then the true solution is $x_* = A^{-1}b = e_n$. We have $V_m = [e_1, \cdots, e_m]$, and H_m is given by

$$H_m = \begin{pmatrix} 0 & & & & 0 \\ 1 & \ddots & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & 1 & 0 \end{pmatrix} \quad \text{for } m < n,$$

with $H_n = A$. Hence, H_m is singular for $m < n$, and the Arnoldi iterates x_m^A do not exist for $m < n$, but $x_n^A = x_*$. Next, for GMRES

$$\bar{H}_m = \begin{pmatrix} H_m \\ 0 \cdots 0 \; 1 \end{pmatrix},$$

and the solution of (2.11) is given by

$$y_m^{GM} = (\bar{H}_m^T \bar{H}_m)^{-1} \bar{H}_m^T V_{m+1}^T b.$$

Since $V_{m+1}^T b = e_1$, and since $\bar{H}_m^T e_1 = 0$ $(m = 1, \cdots, n-1)$ because the first row of $\bar{H}_m$ is all zeros, the GMRES iterates satisfy

$$x_m^{GM} = V_m y_m^{GM} = 0 \; (m = 1, \cdots, n-1), \qquad x_n^{GM} = x_*.$$

Thus, for this example neither Arnoldi nor GMRES make any progress until $m = n$. Note that restarting would be of no use here. This example was also reported on in [3].

A second possibility when H_m is singular would be to try to find the minimum norm solution of

$$\min_{x \in x_0 + K_m} \|V_m^T (b - Ax)\|_2 = \min_{y \in \mathbf{R}^m} \|\beta e_1 - H_m y\|_2. \tag{3.1}$$

For the H_m in the example, it is trivial to form its singular value decomposition. We have $H_m = U_m \Sigma_m V_m^T$, where $U_m = [e_2, \cdots, e_m, e_1]$, $V_m = I_m$ (the $m \times m$ identity matrix), and Σ_m is given by

$$\Sigma_m = \begin{pmatrix} & & 0 \\ & I_{m-1} & \vdots \\ & & 0 \\ 0 \cdots 0 & & 0 \end{pmatrix}.$$

The solution of (3.1) is then given by $y_m^{LS} = \beta H_m^+ e_1$, where $H_m^+ = \Sigma_m U_m^T$ is the pseudoinverse of H_m. This then gives $y_m^{LS} = \beta \Sigma_m U_m^T e_1 = \beta \Sigma_m e_m = 0$ for all $m < n$. Thus, this second possibility produces no better results than GMRES. We show below that this will always happen. That is, when H_m is singular, solving (3.1) gives exactly the same solution as solving (2.11).

The example above illustrates how Arnoldi and GMRES can have catastrophic breakdowns. Even though the GMRES solution can always be calculated, the approximations are of no use since they are all zero until the very last one, which gives the exact solution. This example is also indicative of several relationships between the singularity of the Hessenberg matrix H_m and the solution of the least squares problem (2.11). We first state the results and discuss their theoretical importance, and then give their proofs in the next section.

THEOREM 3.1. *Suppose m steps of the Arnoldi process have been taken, and assume that H_m is singular. Then*

$$\min_{y \in \mathbf{R}^m} \|\beta e_1 - \bar{H}_m y\|_2 = \min_{y \in \mathbf{R}^{m-1}} \|\beta e_1 - \bar{H}_{m-1} y\|_2. \tag{3.2}$$

If y_j^{GM} is the solution of (2.11) *with m replaced by j, $j = m$ or $m-1$, then $y_m^{GM} = ((y_{m-1}^{GM})^T, 0)^T$, and it follows that $x_m^{GM} = x_{m-1}^{GM}$. Conversely, suppose m steps of the Arnoldi process have been taken and that* (3.2) *holds. Then H_m is singular.*

This theorem gives a clear indication of how the performance of the Arnoldi and GMRES algorithms are interrelated. If H_m is singular, so that x_m^A does not exist, then GMRES was unsuccessful in reducing the norm of the residual, and $x_m^{GM} = x_{m-1}^{GM}$. The converse is also true. Thus, either both methods make progress, or both fail. However, the result says nothing about the errors in the two approximations. If one uses the size of the residual norm as a measure of accuracy, then clearly x_m^{GM} is to be preferred over x_m^A in view of (2.5).

Next, define the function $g(x)$ by

$$g(x) \equiv \frac{1}{2} \|Ax - b\|_2^2.$$

We say that the vector v is an *ascent* direction for g at x if

$$g(x + \lambda v) > g(x)$$

for all nonzero λ. We also have the following result which gives an equivalent condition for the matrix H_m to be singular.

COROLLARY 3.2. *Suppose that m steps of the Arnoldi process have been taken. Then H_m is singular if and only if v_m is an ascent direction for g at x_{m-1}^{GM}.*

These two results indicate what happens when the Arnoldi and GMRES algorithms begin performing poorly. If H_m is nearly singular, or equivalently if $\|r_m^{GM}\|_2$ is

only slightly smaller than $\|r_{m-1}^{GM}\|_2$, then the new Krylov vector v_m is close to being an ascent direction for g at x_{m-1}^{GM}.

THEOREM 3.3. *Suppose m steps of the Arnoldi process have been taken, and assume that H_m is singular. Then*

$$\min_{x \in x_0 + K_m} \|V_m^T(b - Ax)\|_2 = \min_{y \in \mathbf{R}^m} \|\beta e_1 - H_m y\|_2 = \min_{y \in \mathbf{R}^m} \|\beta e_1 - \bar{H}_m y\|_2. \tag{3.3}$$

Furthermore, $y_m^{LS} = y_m^{GM}$, where y_m^{LS} and y_m^{GM} are the solutions to the last two problems (respectively) in (3.3).

This last result also suggests that the GMRES algorithm may be the better of the two methods when using the residual norm as the measure of accuracy in the approximate solutions.

4. Proofs. Before we begin the proofs of the theorems stated in the last section, it will be necessary to review some of the details involved in the QR factorization of the Hessenberg matrix $\bar{H}_m$. This is accomplished using Givens rotations. As Saad and Schultz have noted [12], it is desirable to be able to update the factorization of $\bar{H}_m$ progressively as each column of it appears. This allows the residual norm $\|b - Ax_m^{GM}\|_2$ to be calculated without computing x_m^{GM} at each step.

From Brown and Hindmarsh [3], let $F_{j,k} \in \mathbf{R}^{(k+1)\times(k+1)}$ be the rotation matrix which rotates e_j and e_{j+1} by the angle θ_j, namely,

$$F_{j,k} = \begin{pmatrix} I_{j-1} & & & \\ & c_j & -s_j & \\ & s_j & c_j & \\ & & & I_{k-j} \end{pmatrix} \leftarrow \text{row } j+1,$$

where $c_j = \cos(\theta_j)$ and $s_j = \sin(\theta_j)$ for $j \le k$. From the notation, θ_j is assumed independent of k, and so the rotations $F_{j,k}$ and $F_{j,k+1}$ differ by only the addition of a column and row of the identity matrix. Now, suppose that the rotations $F_{1,j}, \cdots, F_{j,j}$ have been applied to $\bar{H}_j$, giving

$$F_{j,j}F_{j-1,j}\cdots F_{1,j}\bar{H}_j = R_j \in \mathbf{R}^{(j+1)\times j},$$

where R_j is upper triangular with its last row containing all zeros. The next step of the Arnoldi process produces the last row and column of the matrix $\bar{H}_{j+1}$. Let $d = (d', h)^T$, where $d' \in \mathbf{R}^{j+1}$ and $h = h_{j+2,j+1}$, be the new column. R_{j+1} is obtained by first forming $\bar{d} = F_{j,j+1}F_{j-1,j+1}\cdots F_{1,j+1}d$ (note that the last component h is unaffected by this multiplication), and then choosing the rotation $F_{j+1,j+1}$ to eliminate h in $\bar{d}$. This gives

$$\begin{aligned} c_{j+1} &= r/\sqrt{r^2 + h^2}, \\ s_{j+1} &= -h/\sqrt{r^2 + h^2}, \end{aligned}$$

letting r denote the next-to-last component of $\bar{d}$. After m steps, the resulting decomposition is

$$Q_m^T \bar{H}_m = R_m, \tag{4.1}$$

where $Q_m^T = F_{m,m}F_{m-1,m}\cdots F_{1,m} \in \mathbf{R}^{(m+1)\times(m+1)}$, and $R_m \in \mathbf{R}^{(m+1)\times m}$ is upper triangular with zeros in its last row. We note that if m steps of the Arnoldi loop

have been possible, then $h_{j+1,j} \neq 0$ for $j = 1, \cdots, m-1$. If $h_{m+1,m} = 0$, then H_m is nonsingular from the remarks at the beginning of §3, and so $\bar{H}_m$ would have full column rank. If $h_{m+1,m} \neq 0$, then $\bar{H}_m$ still has full column rank due to its upper Hessenberg structure. Thus, the solution of the least squares problem (2.11) is unique, and it is found using (4.1). We have

$$\begin{aligned} \|\beta e_1 - \bar{H}_m y\|_2 &= \|\beta e_1 - Q_m R_m y\|_2 \\ &= \|\beta Q_m^T e_1 - R_m y\|_2, \end{aligned}$$

since Q_m is orthogonal. Letting

$$R_m = \begin{pmatrix} \bar{R}_m \\ 0 \cdots 0 \end{pmatrix},$$

and $g_m = \beta Q_m^T e_1 = (\bar{g}_m, g)^T$ with $\bar{g}_m \in \mathbf{R}^m$ and $g \in \mathbf{R}$, we have that the value of y which solves (2.11) is then

$$y_m^{GM} = (\bar{R}_m)^{-1} \bar{g}_m, \tag{4.2}$$

and it follows that

$$\|b - Ax_m^{GM}\|_2 = \|\beta Q_m^T e_1 - R_m y_m^{GM}\|_2 = |g|.$$

We have $g = \beta e_1^T Q_m e_{m+1}$, and letting $Q_m = [q_1^m, \cdots, q_{m+1}^m]$ gives

$$\|b - Ax_m^{GM}\|_2 = |g| = \beta |(q_{m+1}^m)^T e_1|.$$

In [3] it is shown (cf. [3, eq. (2.26)]) that

$$q_{m+1}^m = \left(\prod_{i=1}^{m} s_i, c_1 \prod_{i=2}^{m} s_i, c_2 \prod_{i=3}^{m} s_i, \cdots, c_{m-1} s_m, c_m \right)^T. \tag{4.3}$$

It now follows that $|g| = \beta |s_1 s_2 \cdots s_m|$, and thus we have

$$\|b - Ax_m^{GM}\|_2 = \beta |s_1 s_2 \cdots s_m|. \tag{4.4}$$

This last equation gives an inexpensive way to calculate the norm of the residual associated with x_m^{GM}. See [3] and [12] for more details.

Proof of Theorem 3.1. First, note that if m steps of the Arnoldi process have been possible, then it must be the case that $h_{j+1,j} \neq 0$ for $j = 1, \cdots, m-1$. Otherwise, the Arnoldi process would have broken down at an earlier stage. From above, if $h_{m+1,m} = 0$ (i.e., if $\hat{v}_{m+1} = 0$), then H_m is nonsingular and $x_m^A = x_m^{GM}$ is the exact solution to (2.1). Thus, H_m singular implies that $h_{m+1,m} \neq 0$. Also, when H_m is singular the dimension of its null space is one due to its upper Hessenberg structure and $h_{j+1,j} \neq 0$ for $j = 1, \cdots, m-1$.

Next, consider the QR factorizations of $\bar{H}_m$ and $\bar{H}_{m-1}$. Let

$$\bar{H}_m = Q_m R_m \quad \text{and} \quad \bar{H}_{m-1} = Q_{m-1} R_{m-1},$$

where

$$R_m = \begin{pmatrix} \bar{R}_m \\ 0 \cdots 0 \end{pmatrix} \quad \text{and} \quad R_{m-1} = \begin{pmatrix} \bar{R}_{m-1} \\ 0 \cdots 0 \end{pmatrix}.$$

Note that

$$\bar{H}_m = \begin{pmatrix} & & h_{1,m} \\ & \bar{H}_{m-1} & \vdots \\ & & h_{m,m} \\ 0 \cdots 0 & & h_{m+1,m} \end{pmatrix} = \begin{pmatrix} \bar{H}_{m-1} & h_m \\ 0 \cdots 0 & h_{m+1,m} \end{pmatrix}$$

and $H_m = [\bar{H}_{m-1}, h_m]$, if we let $h_m = (h_{1,m}, \cdots, h_{m,m})^T$. Since H_m is singular, and from the fact that $\bar{H}_{m-1}$ has full column rank, we must have $h_m = \bar{H}_{m-1}\tilde{y}$ for some nonzero $\tilde{y} \in \mathbf{R}^{m-1}$. Thus,

$$h_m = Q_{m-1}R_{m-1}\tilde{y} = Q_{m-1}\begin{pmatrix} \bar{R}_{m-1}\tilde{y} \\ 0 \end{pmatrix}. \tag{4.5}$$

As noted above, the QR factorizations of $\bar{H}_m$ and $\bar{H}_{m-1}$ are related, and we can write $Q_m = \tilde{Q}_{m-1}F_{m,m}^T$, where $F_{m,m}$ is the last Givens rotation defined by

$$F_{m,m} = \begin{pmatrix} I_{m-1} & & \\ & c_m & -s_m \\ & s_m & c_m \end{pmatrix} \quad \text{and} \quad \tilde{Q}_{m-1} = \begin{pmatrix} & & 0 \\ & Q_{m-1} & \vdots \\ & & 0 \\ 0 \cdots 0 & & 1 \end{pmatrix}.$$

Hence, we have

$$F_{m,m}\tilde{Q}_{m-1}^T\bar{H}_m = R_m$$

or

$$F_{m,m}\begin{pmatrix} Q_{m-1}^T\bar{H}_{m-1} & Q_{m-1}^T h_m \\ 0 \cdots 0 & h_{m+1,m} \end{pmatrix} = R_m. \tag{4.6}$$

From (4.5), we have $Q_{m-1}^T h_m = (\tilde{y}^T\bar{R}_{m-1}^T, 0)^T$. Using this last equation, and the fact that $Q_{m-1}^T\bar{H}_{m-1} = R_{m-1}$, we have

$$F_{m,m}\begin{pmatrix} R_{m-1} & \begin{pmatrix} \bar{R}_{m-1}\tilde{y} \\ 0 \end{pmatrix} \\ 0 \cdots 0 & h_{m+1,m} \end{pmatrix} = R_m.$$

Thus, the c_m and s_m in $F_{m,m}$ are chosen so that

$$\begin{pmatrix} c_m & -s_m \\ s_m & c_m \end{pmatrix}\begin{pmatrix} 0 \\ h_{m+1,m} \end{pmatrix} = \begin{pmatrix} \pm h_{m+1,m} \\ 0 \end{pmatrix},$$

with $c_m^2 + s_m^2 = 1$. The values of c_m and s_m must then satisfy $c_m = 0$ and $|s_m| = 1$. If y_m^{GM} is the solution to (2.11), then from (4.4) we have

$$\begin{aligned} \|\beta e_1 - \bar{H}_m y_m^{GM}\|_2 &= \beta \cdot |s_1 s_2 \cdots s_m| = \beta \cdot |s_1 s_2 \cdots s_{m-1}| \\ &= \|\beta e_1 - \bar{H}_{m-1} y_{m-1}^{GM}\|_2. \end{aligned} \tag{4.7}$$

It follows from the uniqueness of y_m^{GM} that we must have $y_m^{GM} = ((y_{m-1}^{GM})^T, 0)$. This then also gives $x_m^{GM} = x_{m-1}^{GM}$.

For the converse, suppose m steps of the Arnoldi process have been taken and assume that (3.2) holds. It follows from (4.4) that $c_m = 0$ and $s_m = \pm 1$. Using (4.6), we can write

$$F_{m,m} \begin{pmatrix} R_{m-1} & Q_{m-1}^T h_m \\ 0 \cdots 0 & h_{m+1,m} \end{pmatrix} = R_m, \tag{4.8}$$

since $R_{m-1} = Q_{m-1}^T \bar{H}_{m-1}$. (Note that (4.6) holds whether or not H_m is singular.) Now, the last Givens rotation $F_{m,m}$ is chosen so that R_m is upper triangular with its last row containing all zeros. But we know that $F_{m,m}$ has the form

$$F_{m,m} = \begin{pmatrix} I_{m-1} & & \\ & 0 & \pm 1 \\ & \mp 1 & 0 \end{pmatrix}.$$

Hence, it must be the case that the last component of the m-vector $Q_{m-1}^T h_m$ is zero. Now, $Q_{m-1}^T h_m = (q_{1,m-1}^T h_m, \cdots, q_{m-1,m-1}^T h_m, 0)^T \equiv (w, 0)^T$. If $w = 0$, then $h_m = 0$ because Q_{m-1}^T is orthogonal, and so $H_m = [\bar{H}_{m-1}, h_m]$ is singular. So assume that $w \neq 0$ in the remainder of the proof.

Let $s = (\bar{s}^T, t)$, with $\bar{s} \in \mathbf{R}^{m-1}$ and $t \in \mathbf{R}$. We then have

$$\begin{aligned} H_m s &= [\bar{H}_{m-1}, h_m]s \\ &= [Q_{m-1}R_{m-1}, h_m]s \\ &= Q_{m-1}[R_{m-1}, Q_{m-1}^T h_m]s \\ &= Q_{m-1}(R_{m-1}\bar{s} + tQ_{m-1}^T h_m). \end{aligned}$$

Using $R_{m-1} = (\bar{R}_{m-1}^T, 0)^T$ and the definition of w, we have

$$H_m s = Q_{m-1} \begin{pmatrix} \bar{R}_{m-1}\bar{s} + tw \\ 0 \end{pmatrix}. \tag{4.9}$$

Since $\bar{R}_{m-1}$ is nonsingular, given any t there exists an $\bar{s}$ such that the right-hand side of (4.9) is zero. Thus, H_m is singular. □

Before we begin the proof of Corollary 3.2, recall that the function $g(x)$ is defined by

$$g(x) \equiv \frac{1}{2}\|Ax - b\|_2^2,$$

and that the new Krylov basis vector v_m is an ascent direction for g at x_{m-1}^{GM} whenever

$$g(x_{m-1}^{GM} + \lambda v_m) > g(x_{m-1}^{GM})$$

for all nonzero λ.

Proof of Corollary 3.2. Note that x_m^{GM} is the solution to the minimization problem

$$\min_{x \in x_0 + K_m} g(x).$$

So, if H_m is singular we also have $x_m^{GM} = x_{m-1}^{GM}$ and hence

$$g(x_{m-1}^{GM} + \lambda v_m) \geq g(x_{m-1}^{GM}) \tag{4.10}$$

for all λ. Defining $h(\lambda)$ to be the function

$$h(\lambda) = g(x_{m-1}^{GM} + \lambda v_m),$$

we have

$$\begin{aligned} h'(\lambda) = \frac{dh(\lambda)}{d\lambda} &= \nabla g(x_{m-1}^{GM} + \lambda v_m)^T v_m \\ &= (A^T(A(x_{m-1}^{GM} + \lambda v_m) - b), v_m) \\ &= (r_{m-1}^{GM} + \lambda A v_m, A v_m) \\ &= (r_{m-1}^{GM})^T A v_m + \lambda \|A v_m\|_2^2, \end{aligned}$$

where $(\cdot,\cdot)$ is the Euclidean inner product and $\nabla = (\partial/\partial x_1, \cdots, \partial/\partial x_n)^T$. It follows from (4.10) that $h'(0) = 0$, and so we must have $(r_{m-1}^{GM})^T A v_m = 0$, or that r_{m-1}^{GM} is orthogonal to Av_m. It then follows from the nonsingularity of A that when λ is nonzero there is strict inequality in (4.10).

Next, suppose that (4.10) holds with strict inequality when λ is nonzero. Then it again follows that $(r_{m-1}^{GM})^T A v_m = 0$. Since it is already known that r_{m-1}^{GM} is orthogonal to AK_{m-1} (see [11]), r_{m-1}^{GM} is also orthogonal to AK_m. But r_m^{GM} is the unique residual which is orthogonal to AK_m. Hence, $r_{m-1}^{GM} = r_m^{GM}$, which gives H_m singular by Theorem 3.1. □

One interesting fact in the proof above is that whenever H_m is singular the residual r_{m-1}^{GM} is already orthogonal to AK_m. Since requiring the residual r_m^{GM} to be orthogonal to AK_m uniquely determines the GMRES solution, we must then have $x_m^{GM} = x_{m-1}^{GM}$.

Proof of Theorem 3.3. From the proof of Theorem 3.1, we have $H_m = [\bar{H}_{m-1}, h_m]$, where $h_m = (h_{1,m}, \cdots, h_{m,m})^T$. Using the QR factorization of $\bar{H}_{m-1}$ it is trivial to obtain the corresponding factorization of H_m. Indeed, we have

$$\begin{aligned} H_m = [\bar{H}_{m-1}, h_m] &= [Q_{m-1}R_{m-1}, h_m] \\ &= Q_{m-1}[R_{m-1}, Q_{m-1}^T h_m] \\ &= Q_{m-1}\tilde{R}_m, \end{aligned} \tag{4.11}$$

letting $\tilde{R}_m = [R_{m-1}, Q_{m-1}^T h_m] \in \mathbf{R}^{m\times m}$. When H_m is singular, $\tilde{R}_m$ will also be singular. Thus, to find the minimum norm solution of

$$\min_{y\in\mathbf{R}^m} \|\beta e_1 - H_m y\|_2,$$

we must further factor $\tilde{R}_m$. Following Golub and Van Loan [6], we can use a sequence of Householder transformations $E_1, \cdots, E_m$ to perform the final reduction, giving

$$E_m E_{m-1} \cdots E_1 \tilde{R}_m^T = S_m^T = \begin{pmatrix} & & 0 \\ & \bar{S}_m^T & \vdots \\ & & 0 \\ 0 \cdots 0 & & 0 \end{pmatrix},$$

where $\bar{S}_m \in \mathbf{R}^{(m-1)\times(m-1)}$ is nonsingular and upper triangular. Then letting $Z_m = E_m \cdots E_1$, we have

$$\begin{aligned} \beta e_1 - H_m y &= \beta e_1 - Q_{m-1}\tilde{R}_m y \\ &= Q_{m-1}(\beta Q_{m-1}^T e_1 - S_m Z_m y) \\ &= Q_{m-1}(\beta Q_{m-1}^T e_1 - S_m w), \end{aligned}$$

where $w = Z_m y$. Thus,

$$\begin{aligned}\|\beta e_1 - H_m y\|_2^2 &= \|\beta Q_{m-1}^T e_1 - S_m w\|_2^2 \\ &= \left\| \beta Q_{m-1}^T e_1 - \begin{pmatrix} & & 0 \\ & \bar{S}_m & \vdots \\ & & 0 \\ 0 \cdots 0 & & 0 \end{pmatrix} w \right\|_2^2 \\ &= |g|^2 + \|\bar{g} - \bar{S}_m \bar{w}\|_2^2,\end{aligned}$$

letting $Q_{m-1}^T e_1 = (\bar{g}, g)^T$, and $w = (\bar{w}, t)^T$. Within w, it is clear that $\bar{w}$ is uniquely determined. The last component t contributes nothing to the residual, and hence the minimum norm solution would have $w = (\bar{w}, 0)^T$. Hence, the minimum norm solution of (3.1) is $y_m^{LS} = Z_m^T(\bar{w}, 0)^T$. Since $g = \beta s_1 \cdots s_{m-1}$ from (4.3), we have

$$\min_{y \in \mathbf{R}^m} \|\beta e_1 - H_m y\|_2 = |g| = \beta |s_1 \cdots s_{m-1}|.$$

Thus, relation (3.3) follows immediately from (4.7). To see that $y_m^{LS} = y_m^{GM}$, note first that by (3.2) and (4.7)

$$\min_{y \in \mathbf{R}^m} \|\beta e_1 - H_m y\|_2 = \min_{y \in \mathbf{R}^{m-1}} \|\beta e_1 - \bar{H}_{m-1} y\|_2.$$

Since the second minimum is taken over a subset of the first minimum, and since the two are equal, it is clear (by uniqueness) that $y_m^{LS} = ((y_{m-1}^{GM})^T, 0)^T = y_m^{GM}$. □

5. A relationship between the residual norms for Arnoldi and GMRES. In this section we first discuss a variant of the Arnoldi algorithm which uses a QR factorization of the upper Hessenberg matrix H_m to solve (2.10) instead of the usual LU factorization. We next derive an expression for the norm of the residual r_m^A analogous to that for GMRES. This expression is then used in deriving the main result of this section which gives an equation for $\|r_m^A\|_2$ in terms of $\|r_m^{GM}\|_2$.

As noted above, there is a relationship between H_m and $\bar{H}_{m-1}$, namely,

$$H_m = [\bar{H}_{m-1}, h_m],$$

where $h_m = (h_{1,m}, \cdots, h_{m,m})^T$. Using the QR factorization of $\bar{H}_{m-1}$ it is trivial to obtain the corresponding factorization of H_m. We have

$$\begin{aligned}H_m = [\bar{H}_{m-1}, h_m] &= [Q_{m-1} R_{m-1}, h_m] \\ &= Q_{m-1}[R_{m-1}, Q_{m-1}^T h_m] \\ &= Q_{m-1} \tilde{R}_m,\end{aligned}$$

letting $\tilde{R}_m = [R_{m-1}, Q_{m-1}^T h_m] \in \mathbf{R}^{m \times m}$. Assuming H_m is nonsingular, the solution of (2.10) is found by solving the upper triangular system

$$\tilde{R}_m y = \beta Q_{m-1}^T e_1 = \beta \begin{pmatrix} q_{1,m-1}^T e_1 \\ \vdots \\ q_{m,m-1}^T e_1 \end{pmatrix}. \tag{5.1}$$

Let y_m^A denote the solution of this system. Next, using (2.7) we have

$$\begin{aligned} b - Ax_m^A &= r_0 - AV_m y_m^A \\ &= r_0 - V_m H_m y_m^A - \hat{v}_{m+1} e_m^T y_m^A \\ &= V_m(\beta e_1 - H_m y_m^A) - \hat{v}_{m+1} e_m^T y_m^A \\ &= -\hat{v}_{m+1} e_m^T y_m^A. \end{aligned}$$

Thus,

$$\|b - Ax_m^A\|_2 = h_{m+1,m} |e_m^T y_m^A|,$$

since $h_{m+1,m} = \|\hat{v}_{m+1}\|_2$. From (4.3) and (5.1), it follows that

$$e_m^T y_m^A = \beta s_1 \cdots s_{m-1} / \tilde{r}_{m,m},$$

where $\tilde{r}_{m,m}$ denotes the (m,m) element of $\tilde{R}_m$. Therefore,

$$\|b - Ax_m^A\|_2 = \beta |s_1 \cdots s_{m-1} / \tilde{r}_{m,m}| \cdot h_{m+1,m}. \tag{5.2}$$

This last equation gives an economical way to update the residual norm $\|r_m^A\|_2$ as the QR factorization of H_m is updated in a manner very similar to that done for the GMRES algorithm. This version of the Arnoldi algorithm could then easily handle the case where $H_{m_{\max}}$ is singular by using Householder transformations as in the proof of Theorem 3.3. Although this implementation would be more expensive, it would nonetheless never break down in the way the LU factorization version of the algorithm can.

Our main result in this section makes use of (5.2), and gives a relationship between the norms of the residuals for Arnoldi and GMRES.

THEOREM 5.1. *Suppose m steps of the Arnoldi process have been taken, and H_m is nonsingular. Then*

$$\|r_m^A\|_2 = \|r_m^{GM}\|_2 \cdot \sqrt{1 + (h_{m+1,m}/\tilde{r}_{m,m})^2}. \tag{5.3}$$

Proof. It is clear from (4.4) and (5.2) that (5.3) is equivalent to

$$\beta |s_1 \cdots s_{m-1} / \tilde{r}_{m,m}| \cdot h_{m+1,m} = \beta |s_1 \cdots s_m| \sqrt{1 + (h_{m+1,m}/\tilde{r}_{m,m})^2},$$

or that

$$|s_m| = h_{m+1,m} / \sqrt{\tilde{r}_{m,m}^2 + h_{m+1,m}^2}.$$

But this last equation follows easily from (4.6) and the definition of the Givens rotation $F_{m,m}$. □

From (5.2) and (4.4), we also have

$$\|r_m^A\|_2 = \frac{h_{m+1,m}}{|\tilde{r}_{m,m}|} \cdot \|r_{m-1}^{GM}\|_2.$$

Next, from (5.3)

$$\frac{h_{m+1,m}}{|\tilde{r}_{m,m}|} = \sqrt{\frac{\|r_m^A\|_2^2}{\|r_m^{GM}\|_2^2} - 1}.$$

Thus,

$$\frac{\|r_m^A\|_2^2}{\|r_{m-1}^{GM}\|_2^2} = \frac{\|r_m^A\|_2^2}{\|r_m^{GM}\|_2^2} - 1,$$

from which it follows that

$$\|r_m^A\|_2 = \|r_m^{GM}\|_2(1 - s_m^2)^{-1/2}. \tag{5.4}$$

This last result then shows that the residual norm for Arnoldi grows without bound as $s_m \to 1$, which corresponds to the stagnation of GMRES.

6. Numerical results. In this final section we first present a somewhat more realistic example which illustrates the main results of the paper. This example also exhibits poor performance of the Arnoldi and GMRES algorithms, and we present it partly because of the lack of tractable examples when studying the Arnoldi and GMRES algorithms. As we show below, for this problem it is possible to write down explicit expressions for the Krylov basis vectors, and for the residual norm calculation. There are very few examples of this type, and so we present it in an effort to facilitate the investigation of the peformance of Krylov subspace projection methods. We then consider Arnoldi and GMRES applied to the solution of a related two-point boundary value problem. We note that these examples are indicative of the fact claimed above that either both methods (Arnoldi and GMRES) will perform well or both will perform poorly. However, we make no claim as to why the methods perform as they do. This will be the subject of future work.

Example 6.1. Let A be a skew-symmetric matrix of the form

$$A = \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 0 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ -\alpha \end{pmatrix},$$

where $\alpha = 1/\sqrt{2}$. Consider solving (2.1), where n is even and $x_0 = 0$. Then the true solution is $x_* = A^{-1}b = (\alpha, \cdots, \alpha)^T$. An easy calculation gives

$$H_m = \begin{pmatrix} 0 & -1 & & & \\ 1 & 0 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & -1 \\ & & & 1 & 0 \end{pmatrix} \in \mathbf{R}^{m \times m}$$

for $m \leq n$, and $h_{m+1,m} = 1$ for $m < n$, with the Krylov basis vectors given by

$$v_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ (-1)^{i+1}\alpha \\ 0 \\ \vdots \\ 0 \\ -\alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow \text{element } i \\ \\ \\ \\ \leftarrow \text{element } n+1-i \\ \\ \\ \\ \end{matrix} \qquad \text{for } i = 1, \cdots, n.$$

Since A is skew symmetric, so is $H_m = V_m^T A V_m$ for all m. Hence, H_m is singular for odd m, and the Arnoldi iterates x_m^A do not exist for m odd. For GMRES, we then have $\|r_{2k}^{GM}\|_2 = \|r_{2k+1}^{GM}\|_2$ for $k = 1, \cdots, n/2 - 1$. For $k = 1, \cdots, n/2 - 1$, the Givens sine coefficients are $s_{2k} = -(k/(k+1))^{1/2}$ and $s_{2k-1} = -1$. The element $\tilde{r}_{2k,2k}$ of $\tilde{R}_{2k}$ in the QR factorization of H_{2k} is given by $\tilde{r}_{2k,2k} = 1/\sqrt{k}$. Thus, we have

$$\|r_{2k}^A\|_2 = \beta |s_1 \cdots s_{2k-1}/\tilde{r}_{2k,2k}| h_{2k+1,2k} = 1, \quad \text{and}$$
$$\|r_{2k}^{GM}\|_2 = \|r_{2k+1}^{GM}\|_2 = \beta |s_1 \cdots s_{2k}| = 1/\sqrt{k+1}$$

for all $k = 1, \cdots, n/2 - 1$, since $\beta = 1$. Thus, Arnoldi is ineffective in reducing the norm of the residual until the very last iterate, and GMRES makes very slow progress until the last iterate. The restarted algorithms also perform poorly on this example, and we note that the poor performance does not seem to be linked to the condition number of A, as Fig. 6.1 shows a plot of the 2-norm condition number of A versus n for various even values of n from 2 to 40. (For this example, the eigenvalues of A are $\lambda_k = i\cos(k\pi/(n+1))$, with $k = 1, \cdots, n$, and so the condition number is $\cos(\pi/(n+1))/\cos(n\pi/(2n+2))$.)

We also consider A above perturbed by a small scalar times the identity matrix to give $A(\epsilon) = A + \epsilon I$, for $\epsilon > 0$ and small. Arnoldi applied to the linear system $A(\epsilon)x = b$ has a sawtooth graph for iteration count versus the residual norm, while GMRES will have a nonincreasing staircase graph. While both methods perform poorly for this perturbed problem, GMRES would be preferred due to the fact that the residual norm will never increase. Figs. 6.2 and 6.3 show the behavior of GMRES and Arnoldi for various values of ϵ when $n = 40$. Notice the sawtooth behavior of Arnoldi for $\epsilon = 0.05$.

We should also note that the poor performance here is not necessarily linked to skew-symmetric matrices, since a similar situation occurs for the linear system $|A|x = |b|$ with A and b defined as above, and where the absolute value sign simply means to take the absolute value of all the elements in A and b. In this case, the spectrum of $|A|$ is that of A rotated by 90 degrees, and it is possible to derive simple relationships for H_m and the v_i's similar to those obtained above for $Ax = b$. It happens that with a zero initial guess the initial residual $r_0 = |b|$ is orthogonal to half of the eigenspace of $|A|$, and so Arnoldi and GMRES will take at most $n/2$ iterations to obtain the solution in exact arithmetic. Figs. 6.4 and 6.5 show the behavior of

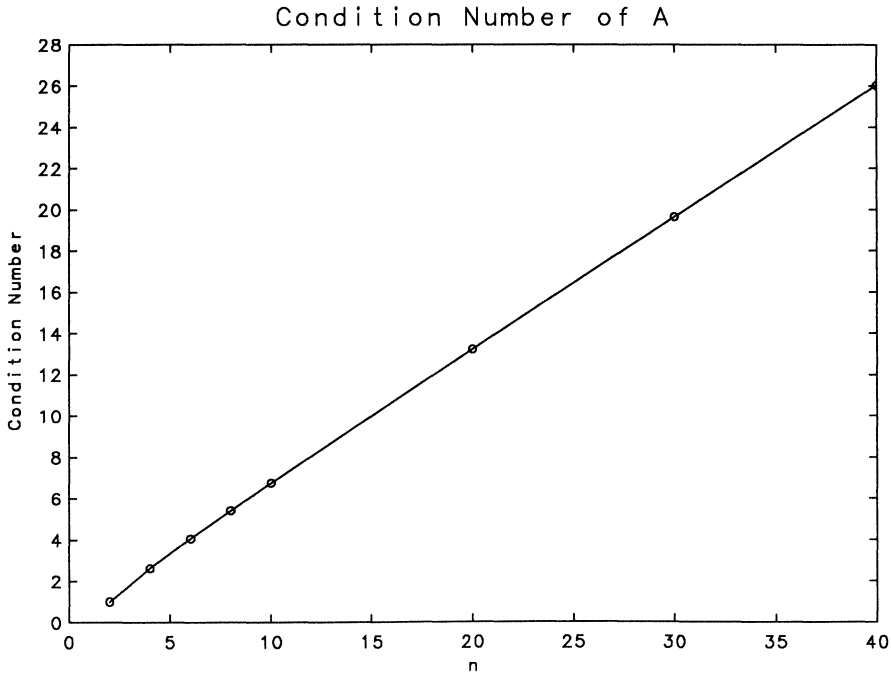

FIG. 6.1. *Condition number of A versus n.*

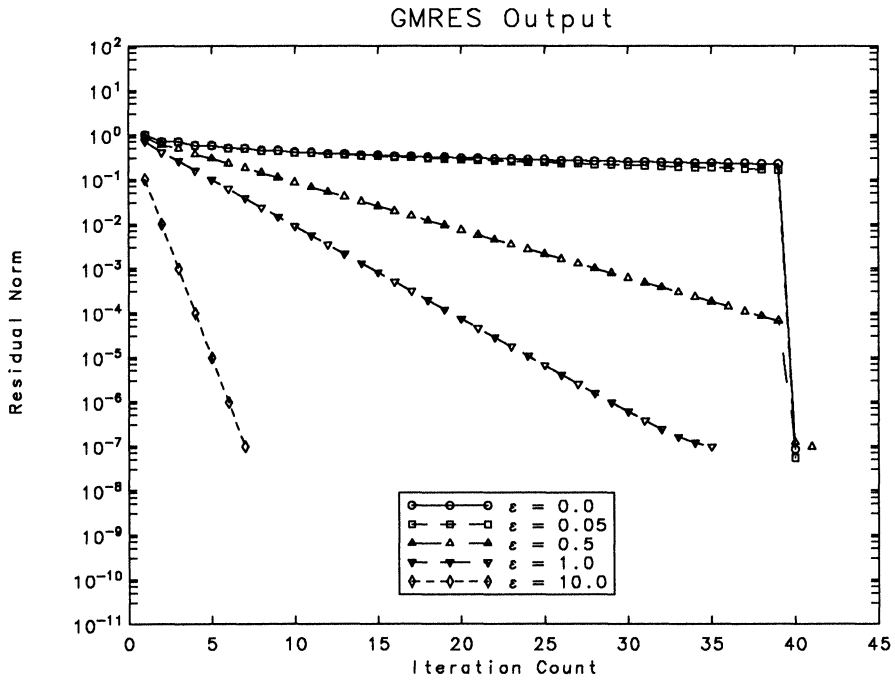

FIG. 6.2. *GMRES applied to* $(A + \epsilon I)x = b$.

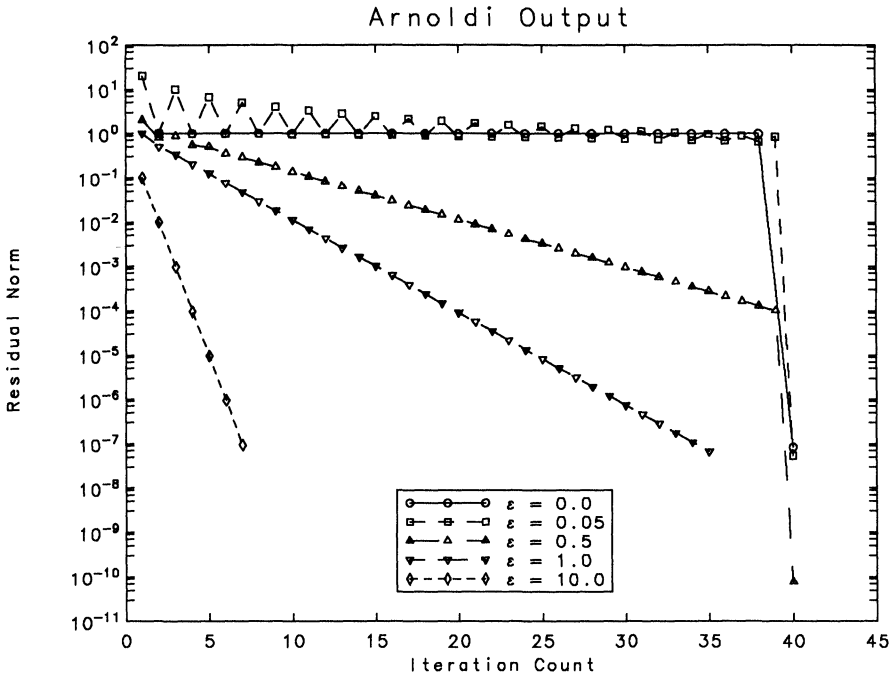

FIG. 6.3. *Arnoldi applied to* $(A + \epsilon I)x = b$.

GMRES and Arnoldi applied to the problem $(|A| + \epsilon I)x = |b|$ for various values of ϵ when $n = 40$. Again note the sawtooth behavior of the Arnoldi iteration.

Finally, the above matrix A (up to a scalar factor) results when letting $a \to 0$ in the two-point boundary value problem

$$au_{xx} + u_x = f \quad \text{on } 0 < x < 1,$$
$$u(0) = u(1) = 0,$$

where $f(x) \equiv 1$ and $a > 0$ is a constant, and if centered finite differences are used to approximate the first- and second-order derivatives. Figs. 6.6 and 6.7 show the behavior of GMRES and Arnoldi on this problem for various values of the scalar a when the interval $0 < x < 1$ is discretized into 40 subintervals (i.e., $n = 40$). Figs. 6.8 and 6.9 are magnified portions of Figs. 6.6 and 6.7, respectively. Again, note the sawtooth behavior of Arnoldi for small a. We also remark that, in general, centered differences for the first derivative terms are not recommended. However, it is worth pointing out that both GMRES and Arnoldi perform poorly for the two largest values of a, and for these two values the approximate solution has at least four significant digits.

7. Concluding remarks. In this paper we have given several relationships between the residual norms for Arnoldi and GMRES which suggest that either both methods will perform well on a particular problem, or both will perform poorly. We have also given a relationship which shows exactly how much larger the residual norm for the Arnoldi algorithm is than that for the GMRES algorithm. Unfortunately, these results give no indication of how well either method will perform on a particular

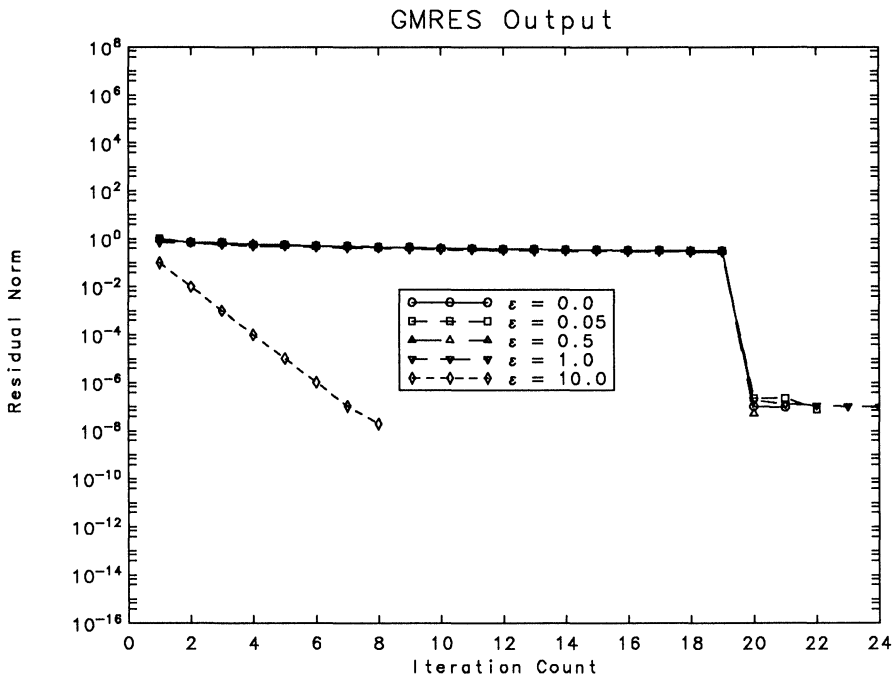

FIG. 6.4. *GMRES applied to* $(|A| + \epsilon I)x = |b|$.

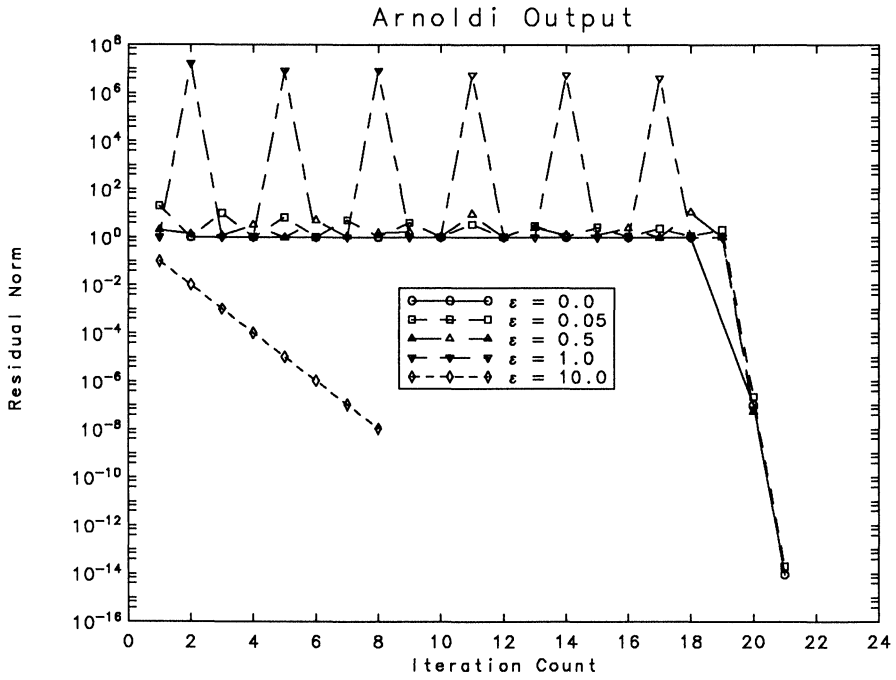

FIG. 6.5. *Arnoldi applied to* $(|A| + \epsilon I)x = |b|$.

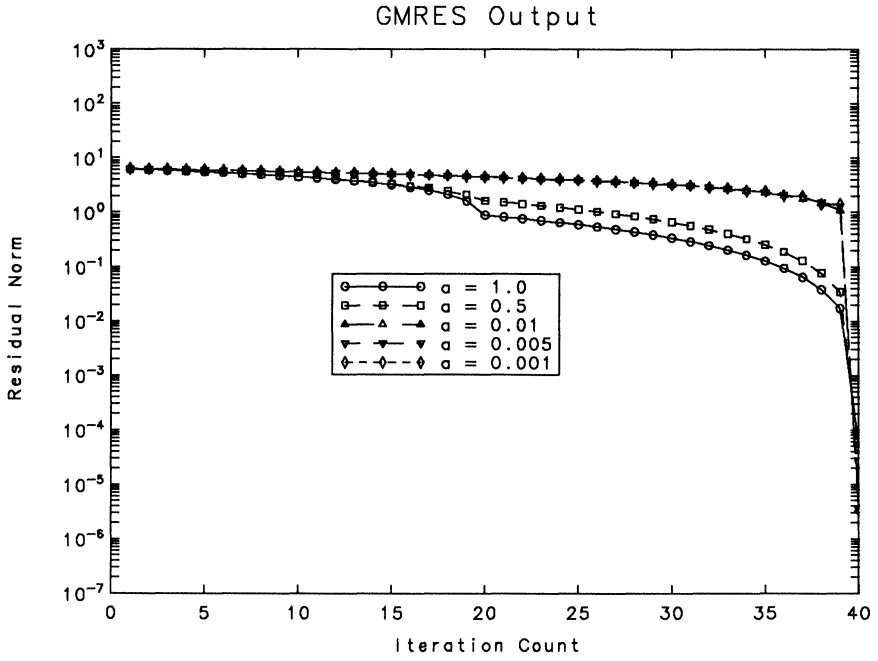

FIG. 6.6. *GMRES applied to* $au_{xx} + u_x = f$.

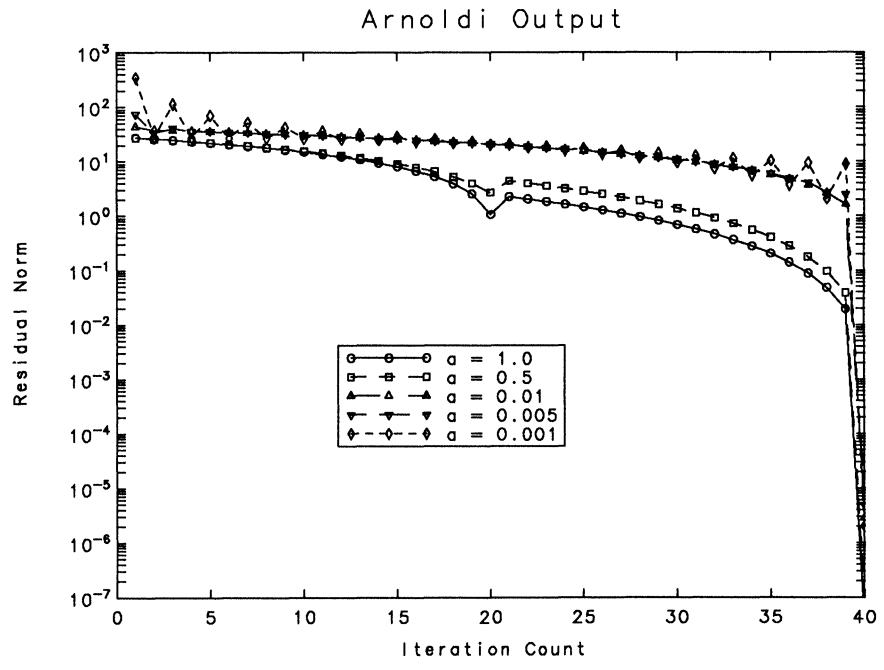

FIG. 6.7. *Arnoldi applied to* $au_{xx} + u_x = f$.

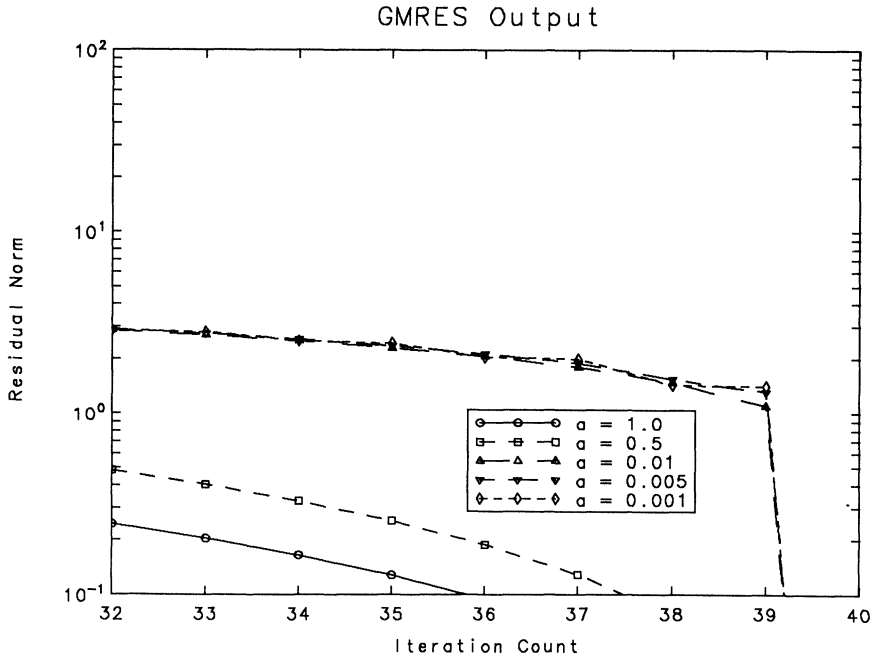

FIG. 6.8. *GMRES applied to* $au_{xx} + u_x = f$.

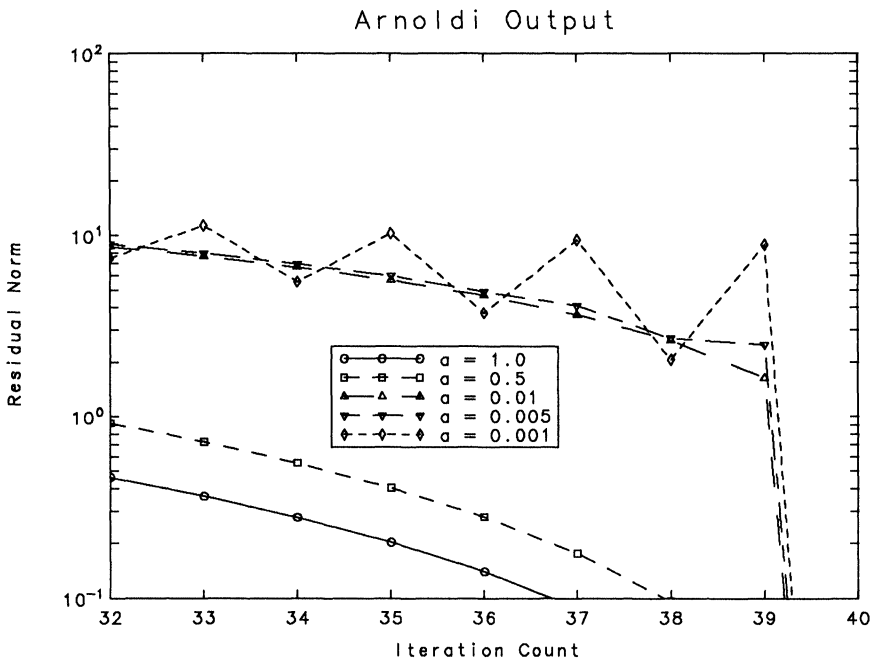

FIG. 6.9. *Arnoldi applied to* $au_{xx} + u_x = f$.

problem. They do, however, show that when Arnoldi or GMRES begins breaking down or stagnating, the latest Krylov basis vectors are poor descent directions for reducing the residual norm. In general, we prefer the GMRES algorithm due to its minimal norm property (and we would recommend GMRES over Arnoldi based upon this fact), but much more work is needed in order to better understand the performance characteristics of these methods. This will also be important in understanding what determines a good preconditioner for a particular problem.

Acknowledgment. The author would like to thank Alan Hindmarsh for his careful reading of a preliminary version of this manuscript.

REFERENCES

[1] P. N. BROWN, *A local convergence theory for combined inexact-Newton/finite-difference projection methods,* SIAM J. Numer. Anal., 24 (1987), pp. 407-434.

[2] P. N. BROWN AND A. C. HINDMARSH, *Matrix-free methods for stiff systems of* ODE*'s,* SIAM J. Numer. Anal., 23 (1986), pp. 610-638.

[3] _____, *Reduced storage methods in stiff* ODE *systems,* J. Appl. Math. Comput., 31 (1989), pp. 40-91.

[4] 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.

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

[6] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations,* The Johns Hopkins University Press, Baltimore, MD, 1983.

[7] A. L. HAGEMAN AND D. M. YOUNG, *Applied Iterative Methods,* Academic Press, New York, 1981.

[8] C. C. PAIGE AND M. A. SAUNDERS, *Solution of sparse indefinite systems of linear equations,* SIAM J. Numer. Anal., 12 (1975), pp. 617-629.

[9] Y. SAAD, *Krylov subspace methods for solving unsymmetric linear systems,* Math. Comp., 37(1981), pp. 105-126.

[10] _____, *Practical use of some Krylov subspace methods for solving indefinite and nonsymmetric linear systems,* SIAM J. Sci. Statist. Comput., 5 (1984), pp. 203-228.

[11] Y. SAAD AND M. H. SCHULTZ, *Conjugate gradient-like algorithms for solving nonsymmetric linear systems,* Math. Comp., 44/170(1985), pp. 417-424.

[12] _____, GMRES: *A generalized minimal residual algorithm for solving nonsymmetric linear systems,* SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856-869.