SIAM J. Matrix Anal. Appl.
Vol. 32, No. 2, pp. 394–411

ON THE RESIDUAL NORM IN FOM AND GMRES*

GÉRARD MEURANT†

Abstract. We provide expressions for the residual norms when using the full orthogonalization method and the generalized minimum residual method for solving linear systems. They involve a triangular submatrix of the Hessenberg matrix generated by the Arnoldi process. This allows one to obtain bounds showing that the norm of the residual decreases to zero when the smallest singular value of this triangular matrix goes to zero. Numerical examples show that even though these bounds are not sharp they describe quite well the rate of decrease of the residual norm.

Key words. full orthogonalization method, generalized minimum residual method, residual norm

AMS subject classifications. 15A06, 65F10

DOI. 10.1137/100807831

1. Introduction.

We consider solving a linear system

$$Ax = b,$$

where A is a (real or complex) nonsingular matrix of order n with the full orthogonalization method (FOM) and the generalized minimum residual method (GMRES), which are Krylov methods based on the Arnoldi orthogonalization process; see Saad [16], [17] and Saad and Schultz [18]. The initial residual is denoted as $r^0 = b - Ax^0$, where x^0 is the starting vector. The Krylov subspace of order k based on A and r^0, which is denoted as $\mathcal{K}_k(r^0, A)$, is $\text{span}\{r^0, Ar^0, \ldots, A^{k-1}r^0\}$. The approximate solution x^k at iteration k is sought as $x^k \in x^0 + \mathcal{K}_k(r^0, A)$ such that the residual vector $r^k = b - Ax^k$ satisfies an orthogonality condition with the Krylov subspace.

There are many papers and books giving bounds for the norms of the residual r^k; see, for instance, [18], [3], [4], [15], [9], [21]. In most of these papers the authors write the residual as a polynomial p_k in A applied to the initial residual and obtain bounds for the norm of the residual by bounding the norm $\|p_k(A)\|$ on a suitable subspace. The bounds on $\|p_k(A)\|$ do not depend on the right-hand side of the linear system. For some problems they can be far from being sharp. A study of the residual norms in an abstract setting was published in [5]. Residual bounds were given in [14]. They involve singular values of the matrix AV_kD_k, where V_k will be defined below and D_k is a diagonal scaling matrix. Expressions for the norm of the residual involving the Krylov matrices (whose columns are the vectors of the natural basis of $\mathcal{K}_k(r^0, A)$) were given in [12] and [19].

In this paper we derive expressions for the l_2 norm of the residual which involve the Hessenberg matrix H_k constructed by the Arnoldi process during the FOM or GMRES iterations. From these expressions we obtain bounds which show that for FOM or GMRES, the residual norm convergence depends on the smallest singular value of an upper triangular submatrix of H_k (denoted by $\tilde{H}_k$). The norm of the residual goes

*Received by the editors September 7, 2010; accepted for publication (in revised form) by A. Frommer March 18, 2011; published electronically June 24, 2011. This paper was partly written in 2009 while the author was visiting the Institute of Computer Science of the Czech Academy of Sciences in Prague with the support of a GAAS grant IAA100300802 and with the support of the Institutional Research Plan AV0Z10300504.
http://www.siam.org/journals/simax/32-2/80783.html
†CEA, 30, rue du sergent, Bauchat, 75012 Paris, France (gerard.meurant@gmail.com).

to zero when this smallest singular value goes to zero. Even though many papers have been written on this topic, FOM and GMRES convergence is still not fully understood. The goal of this paper is not to provide new expressions to compute the norm of the residual during the iterations in practical computations, since this can be done easily using rotations (see [18]), but to relate the behavior of this norm to some submatrices of H_k. It is hoped that these properties could eventually lead to a better understanding of FOM and GMRES convergence. However, it is an open (and interesting) research topic to study which properties of the matrix A and the right-hand side b could imply that the smallest singular value of $\tilde{H}_k$ would decrease fast with k.

The contents of the paper are as follows. Section 2 briefly recalls the definitions of FOM and GMRES proposed by Saad [17] and Saad and Schultz [18]. Section 3 derives expressions for the l_2 norm of the residual in FOM. We use these expressions in section 4 to obtain bounds that involve the smallest singular value of the triangular submatrix $\tilde{H}_k$. Section 5 gives a simple proof of a result proven in [2] concerning the GMRES residual norm. We also briefly discuss the relationships between FOM and GMRES. From this, we derive bounds for the GMRES residual norm in section 6. Section 7 provides numerical experiments comparing the bounds to the actual values of the residual norms. Our bounds are not sharp, but they describe quite accurately the rate of decrease of the norm of the residual. Finally, section 8 provides some conclusions and perspectives.

Throughout this paper, e^k denotes the kth column of the identity matrix of different orders, and an upper index $*$ denotes the conjugate transpose of a vector or a matrix. The inner product (x, y) is defined as $\sum_{i=1}^{n} x_i \bar{y}_i = y^* x$, where the bar denotes the conjugate of a complex number. In this paper we assume exact arithmetic. For Krylov methods with (rounding error) perturbations, see [23].

2. FOM and GMRES. The first step in the FOM or GMRES algorithms is to compute an orthogonal basis of the Krylov subspace $\mathcal{K}_k(r^0, A)$. Let v^j, $j = 1, \ldots, n$, be the Arnoldi basis vectors of unit norm (see [1]) and V_k be the matrix whose columns are $v^1, \ldots, v^k$. The orthonormal basis vectors v^j are usually computed recursively with the modified Gram–Schmidt Arnoldi algorithm. We write the iterates as $x^k = x^0 + V_k z^k$. The matrix relation for V_k given by the Arnoldi process is

$$A V_k = V_k H_k + h_{k+1,k} v^{k+1} (e^k)^T, \tag{2.1}$$

where H_k is an upper Hessenberg matrix of order k with elements $h_{i,j}$. Therefore,

$$H_k = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & \cdots & h_{1,k} \\ h_{2,1} & h_{2,2} & & & \vdots \\ & h_{3,2} & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & h_{k,k-1} & h_{k,k} \end{pmatrix}. \tag{2.2}$$

Note that in the Arnoldi process $h_{k+1,k}$ is defined as a norm of a vector (see [18]) and is therefore real and positive.

In FOM, which is an orthogonal residual method, one asks for the residual vector $r^k = b - Ax^k$ to be orthogonal to the Krylov subspace. It gives $V_k^* r^k = 0$. It is well known that the vector of coordinates z^k is obtained by writing

$$H_k z^k = V_k^* r^0 = \|r^0\| V_k^* v^1 = \|r^0\| e^1, \tag{2.3}$$

since in the Arnoldi process one takes $v^1 = r^0/\|r^0\|$. When H_k is nonsingular, the linear system (2.3) is usually solved using a QR factorization of the Hessenberg matrix H_k. The matrix H_k is transformed to triangular form by Givens rotations. The vector of coordinates z^k is a scalar multiple of the first column of the inverse of H_k.

In GMRES the l_2 norm of the residual is minimized at each iteration. The matrix relation in (2.1) can also be written as

$$AV_k = V_{k+1}H_k^{(e)}, \tag{2.4}$$

where $H_k^{(e)}$ is the $(k+1) \times k$ extended matrix

$$H_k^{(e)} = \begin{pmatrix} H_k \\ h_{k+1,k}(e^k)^T \end{pmatrix}.$$

Since the norm of the residual can be written as

$$\|b - Ax^k\| = \|\|r^0\|e^1 - H_k^{(e)}z^k\|,$$

the vector of coordinates z^k is computed by solving the (small) least squares problem

$$\min_z \|\|r^0\|e^1 - H_k^{(e)}z\|. \tag{2.5}$$

The theoretical solution is given by the pseudoinverse of $H_k^{(e)}$,

$$z^k = \|r^0\|([H_k^{(e)}]^* H_k^{(e)})^{-1}[H_k^{(e)}]^* e^1.$$

In practical computations, the solution of the least squares problem (2.5) is obtained by using a QR factorization of $H_k^{(e)}$. Contrary to FOM, GMRES cannot break down as long as $h_{k+1,k} \neq 0$, since the least squares problem can always be solved; see [18].

3. The residual norm in FOM. Even though the FOM residual norm is computable as $\|b - Ax^k\|$ or preferably by using the rotations computed for the QR factorization of H_k (see [18]), it is interesting to study other expressions of this norm to try to understand how and why the residual goes to zero. We first have the following result which was proved in [16]; see also [17], [18]. The proof is so simple that we give it for the reader's convenience.

LEMMA 3.1. *Assuming that H_k is nonsingular, the norm of the residual in FOM is given by*

$$\|r^k\| = \|r^0\|h_{k+1,k}|(H_k^{-1}e^1, e^k)|. \tag{3.1}$$

Proof. We have

$$r^k = b - Ax^k = b - A(x^0 + V_k z^k) = r^0 - (V_k H_k z^k + h_{k+1,k}v^{k+1}(e^k)^T z^k).$$

The first two terms in this last expression cancel because of the definition of z^k, and we have $r^k = -h_{k+1,k}z_k^k v^{k+1}$ which shows, as it is well known, that the residual is proportional to the basis vector v^{k+1}. Hence, $\|r^k\| = h_{k+1,k}|z_k^k|$. Since $z^k = \|r^0\|H_k^{-1}e^1$ we obtain the result. ∎

Considering the result of Lemma 3.1, it is first interesting to derive expressions for $(H_k^{-1}e^1, e^k)$ which is the $(k,1)$ element of the inverse of H_k (when it exists), and more generally for the norm of the residual in FOM. This can be done in several ways.

First, as in [18], we may consider reducing the Hessenberg matrix H_k to upper triangular form by unitary transformations. One can obtain expressions for the norm of the residual using the sines of the Givens rotations; see [18], [21].

Another way to proceed is to directly look at the first column of the inverse, $H_k^{-1}e^1$ (when H_k is nonsingular). Properties of inverses of Hessenberg matrices were studied in [11] and [6]; see also [22]. In these papers, it is proved that the lower triangular part of the inverse is the lower triangular part of a rank-one matrix. Since we are also interested in $h_{k+1,k}(H_k^{-1}e^1, e^k)$, we are going to proceed in a different way. A simple method to obtain the first column of the inverse of H_k is to consider an LU factorization of a permutation of the matrix.

THEOREM 3.2. *Let H_k be the Hessenberg matrix defined in* (2.2)*. Let us assume that H_k is nonsingular and $h_{i+1,i} \neq 0$, $i = 1, \ldots, k-1$. Let*

$$\tilde{H}_{k-1} = \begin{pmatrix} h_{2,1} & h_{2,2} & \cdots & \cdots & h_{2,k-1} \\ & h_{3,2} & \ddots & \vdots & \vdots \\ & & \ddots & \vdots & \vdots \\ & & & h_{k-1,k-2} & h_{k-1,k-1} \\ & & & & h_{k,k-1} \end{pmatrix},$$

which is an upper triangular matrix. Let h^{k-1} be the conjugate transpose of the $k-1$ first elements of the first row of H_k and w^{k-1} the last $k-1$ elements of the last column of H_k. Then, in blockwise form, H_k is written as

$$H_k = \begin{pmatrix} (h^{k-1})^* & h_{1,k} \\ \tilde{H}_{k-1} & w^{k-1} \end{pmatrix},$$

and

$$(H_k^{-1}e^1, e^k) = \frac{1}{h_{1,k} - (\tilde{H}_{k-1}^{-1}w^{k-1}, h^{k-1})}, \tag{3.2}$$

$$h_{k+1,k}(H_k^{-1}e^1, e^k) = \frac{1}{(e^k, \tilde{H}_k^{-*}h^k)}. \tag{3.3}$$

Consequently, the FOM residual norm is given by

$$\|r^k\| = \frac{\|r^0\|}{|(e^k, \tilde{H}_k^{-*}h^k)|}. \tag{3.4}$$

Proof. Let us compute an LU factorization of a permutation of H_k. Let P be the matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{3.5}$$

We multiply H_k from the left with the matrix P. The permuted matrix is

$$PH_k = \begin{pmatrix} h_{2,1} & h_{2,2} & \cdots & & h_{2,k} \\ & h_{3,2} & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & h_{k,k-1} & h_{k,k} \\ h_{1,1} & h_{1,2} & \cdots & h_{1,k-1} & h_{1,k} \end{pmatrix} = \begin{pmatrix} \tilde{H}_{k-1} & w^{k-1} \\ (h^{k-1})^* & h_{1,k} \end{pmatrix}.$$

From the structure of $PH_k = L_k U_k$, we see that there is no fill-in in the lower triangular factor L_k. So, we look for

$$L_k = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ l_1 & l_2 & \cdots & l_{k-1} & 1 \end{pmatrix}.$$

The matrix U_k is upper triangular with elements denoted by $u_{i,j}$. We write L_k and U_k in block form as

$$L_k = \begin{pmatrix} I & 0 \\ (l^{k-1})^* & 1 \end{pmatrix}, \qquad U_k = \begin{pmatrix} \tilde{U}_{k-1} & u^{k-1} \\ 0 & \alpha_k \end{pmatrix},$$

where $\tilde{U}_{k-1}$ of order $k-1$ is upper triangular. Then, the product $L_k U_k$ is

$$PH_k = \begin{pmatrix} \tilde{U}_{k-1} & u^{k-1} \\ (l^{k-1})^* \tilde{U}_{k-1} & (l^{k-1})^* u^{k-1} + \alpha_k \end{pmatrix}.$$

This shows that $\tilde{U}_{k-1} = \tilde{H}_{k-1}$ and that the vector u^{k-1} is equal to the last $k-1$ elements of the last column of H_k, that is, w^{k-1}. Hence, $u_{i,j} = h_{i+1,j}$, $i = 1, \ldots, k-1$, $j = 1, \ldots, k$. It remains to compute l^{k-1} and α_k. By identification

$$(l^{k-1})^* \tilde{U}_{k-1} = (l^{k-1})^* \tilde{H}_{k-1} = (h^{k-1})^*.$$

The vector l^{k-1} is found by solving a lower triangular linear system

$$\tilde{H}_{k-1}^* l^{k-1} = h^{k-1}.$$

This gives $l^{k-1} = \tilde{H}_{k-1}^{-*} h^{k-1}$. Finally, we have $(l^{k-1})^* u^{k-1} + \alpha_k = h_{1,k}$, and using the expression of l^{k-1} we obtain

$$\alpha_k = h_{1,k} - (l^{k-1})^* u^{k-1} = h_{1,k} - (\tilde{H}_{k-1}^{-1} w^{k-1}, h^{k-1}).$$

From this LU factorization, we are interested in solving the linear system $H_k z = e^1$. Multiplying by the permutation matrix P, we have to solve $PH_k z = Pe^1 = e^k$. We first solve $L_k y = e^k$, which yields $y = e^k$. Then, $U_k z = y = e^k$ has the solution

$$z = H_k^{-1} e^1 = U_k^{-1} e^k.$$

Hence, the first column of the inverse of H_k is given by the last column of the inverse of U_k. From the results above, we obtain

$$\begin{pmatrix} z_1 \\ \vdots \\ z_{k-1} \end{pmatrix} = -\frac{1}{\alpha_k} \tilde{U}_{k-1}^{-1} u^{k-1}, \qquad z_k = \frac{1}{\alpha_k}.$$

The first column of the inverse of H_k is

$$H_k^{-1} e^1 = \frac{1}{\alpha_k} \begin{pmatrix} -\tilde{U}_{k-1}^{-1} u^{k-1} \\ 1 \end{pmatrix} = \frac{1}{h_{1,k} - (\tilde{H}_{k-1}^{-1} w^{k-1}, h^{k-1})} \begin{pmatrix} -\tilde{H}_{k-1}^{-1} w^{k-1} \\ 1 \end{pmatrix}.$$

The element which is of interest for us is

$$(H_k^{-1} e^1, e^k) = \frac{1}{h_{1,k} - (\tilde{H}_{k-1}^{-1} w^{k-1}, h^{k-1})}.$$

This proves (3.2). Concerning (3.3), we have

$$l^k = \begin{pmatrix} \frac{\bar{\alpha}_1}{h_{2,1}} \\ \vdots \\ \frac{\bar{\alpha}_{k-1}}{h_{k,k-1}} \\ \frac{\bar{\alpha}_k}{h_{k+1,k}} \end{pmatrix}.$$

To prove this, we use the expression obtained for l^k,

$$l^k = \tilde{H}_k^{-*} h^k = \begin{pmatrix} \tilde{H}_{k-1}^{-*} & 0 \\ -\frac{(w^{k-1})^* \tilde{H}_{k-1}^{-*}}{h_{k+1,k}} & \frac{1}{h_{k+1,k}} \end{pmatrix} \begin{pmatrix} h^{k-1} \\ \bar{h}_{1,k} \end{pmatrix} = \begin{pmatrix} \tilde{H}_{k-1}^{-*} h^{k-1} \\ -\frac{(w^{k-1})^* \tilde{H}_{k-1}^{-*}}{h_{k+1,k}} h^{k-1} + \frac{\bar{h}_{1,k}}{h_{k+1,k}} \end{pmatrix}.$$

Therefore,

$$l^k = \begin{pmatrix} l^{k-1} \\ \frac{\bar{\alpha}_k}{h_{k+1,k}} \end{pmatrix}.$$

This proves the result by induction since $l^1 = \bar{h}_{1,1}/h_{2,1}$. Looking at the last entry of l^k, we have

$$\frac{\alpha_k}{h_{k+1,k}} = (e^k, \tilde{H}_k^{-*} h^k),$$

which proves (3.3). Equation (3.4) is a consequence of the previous results and Lemma 3.1. ∎

Equation (3.4) is interesting since it gives the norm of the relative residual as a function of a single quantity $|(e^k, \tilde{H}_k^{-*} h^k)|$ rather than two in $h_{k+1,k}|(H_k^{-1} e^1, e^k)|$. The norm $\|r^k\|$ is small if and only if $|(e^k, \tilde{H}_k^{-*} h^k)|$ is large.

Another way to obtain the last element of the first column of the inverse of H_k is to use Cramer's rule as in [19].

THEOREM 3.3. *Using the notation of Theorem* 3.2*, we have*

$$(H_k^{-1} e^1, e^k) = \frac{\det(\tilde{H}_{k-1})}{\det(H_k)}. \tag{3.6}$$

Proof. To obtain the last component of the first column of the inverse, we have to compute the determinant of the matrix obtained from H_k by replacing the last column by e^1. This determinant is obviously equal to $\det(\tilde{H}_{k-1})$. ∎

When looking at $\|r^k\|$ we have to consider $|(H_k^{-1} e^1, e^k)|$. This gives the following result.

THEOREM 3.4. *With the notation of Theorem* 3.2*, the norm of the FOM residual is*

$$\|r^k\| = \|r^0\| h_{k+1,k} \left| \frac{\det(\tilde{H}_{k-1})}{\det(H_k)} \right| = \|r^0\| \left| \frac{\det(\tilde{H}_k)}{\det(H_k)} \right|. \tag{3.7}$$

Let $\sigma_i(M)$ be the singular values of M. Then,

$$\|r^k\| = \|r^0\| h_{k+1,k} \frac{\prod_{i=1}^{k-1} \sigma_i(\tilde{H}_{k-1})}{\prod_{i=1}^{k} \sigma_i(H_k)} = \|r^0\| \frac{\prod_{i=1}^{k} \sigma_i(\tilde{H}_k)}{\prod_{i=1}^{k} \sigma_i(H_k)}. \tag{3.8}$$

Proof. From Theorem 3.3 we have

$$|(H_k^{-1} e^1, e^k)| = \frac{|\det(\tilde{H}_{k-1})|}{|\det(H_k)|} = \sqrt{\frac{\det(\tilde{H}_{k-1}^* \tilde{H}_{k-1})}{\det(H_k^* H_k)}}.$$

Then,

$$|(H_k^{-1} e^1, e^k)| = \frac{\prod_{i=1}^{k-1} \sigma_i(\tilde{H}_{k-1})}{\prod_{i=1}^{k} \sigma_i(H_k)}.$$

For the second formula we note that $\det(\tilde{H}_{k-1}) = \prod_{i=1}^{k-1} h_{i+1,i}$. Therefore, $h_{k+1,k} \det(\tilde{H}_{k-1}) = \det(\tilde{H}_k)$. ∎

4. Bounds for the FOM residual norm. The easiest expression to obtain a lower bound for the norm of the residual in FOM is probably (3.4) from which we obtain the following result.

THEOREM 4.1. *Using the notation of Theorem* 3.2 *we have the following lower bounds for the FOM residual norm:*

$$\|r^k\| \geq \|r^0\| \frac{\sigma_{\min}(\tilde{H}_k)}{\|h^k\|}, \tag{4.1}$$

$$\|r^k\| \geq \|r^0\| \frac{1}{\|\tilde{H}_k^{-*} h^k\|}, \tag{4.2}$$

$$\|r^k\| \geq \|r^0\| \frac{1}{\|\tilde{H}_k^{-1} e^k\| \|h^k\|}. \tag{4.3}$$

Proof. We have

$$|(e^k, \tilde{H}_k^{-*} h^k)| \leq \|\tilde{H}_k^{-*} h^k\| \leq \|\tilde{H}_k^{-1}\| \|h^k\|,$$

and this gives (4.1) and (4.2). The lower bound (4.2) is almost trivial since we have

$$\tilde{H}_k^{-*} h^k = \begin{pmatrix} \frac{\bar{\alpha}_1}{h_{2,1}} \\ \vdots \\ \frac{\bar{\alpha}_k}{h_{k+1,k}} \end{pmatrix}.$$

Therefore,

$$\|\tilde{H}_k^{-*} h^k\|^2 = \sum_{i=1}^{k} \left| \frac{\alpha_i}{h_{i+1,i}} \right|^2.$$

But we have

$$\left| \frac{\alpha_k}{h_{k+1,k}} \right| = \frac{\|r^0\|}{\|r^k\|}$$

and

$$\|\tilde{H}_k^{-*} h^k\|^2 = \|r^0\|^2 \sum_{i=1}^{k} \frac{1}{\|r^i\|^2}.$$

The bound (4.3) is obtained by writing

$$|(e^k, \tilde{H}_k^{-*} h^k)| = |(\tilde{H}_k^{-1} e^k, h^k)| \leq \|\tilde{H}_k^{-1} e^k\| \|h^k\|. \qquad \square$$

The lower bounds (4.1) and (4.3) involve the norm of h^k that can be written and bounded in different ways. We have

$$|h_{1,j}| = |(v^1, Av^j)| \leq \|A\|, \qquad j = 1, \ldots, k.$$

Therefore, $\|h^k\| \leq \sqrt{k}\|A\|$, and

$$\|r^k\| \geq \|r^0\| \frac{\sigma_{\min}(\tilde{H}_k)}{\sqrt{k}[\sigma_{\max}(A)]}. \tag{4.4}$$

Theorem 4.1 essentially shows that there is no convergence of FOM as long as $\sigma_{\min}(\tilde{H}_k)$ is large. The norm of h^k is related to the singular values of $H_k^{(e)}$ and $\tilde{H}_k$ since we have

$$\tilde{H}_k^* \tilde{H}_k + h^k (h^k)^* = [H_k^{(e)}]^* H_k^{(e)}.$$

Taking traces as in [19], we obtain

$$\|h^k\|^2 = \sum_{i=1}^{k} ([\sigma_i(H_k^{(e)})]^2 - [\sigma_i(\tilde{H}_k)]^2).$$

The norm of h^k is small when the singular values of $\tilde{H}_k$ are close to those of $H_k^{(e)}$.

It does not seem to be easy to find a lower bound for $|(e^k, \tilde{H}_k^{-*} h^k)|$ to obtain an upper bound of the norm of the residual. Therefore, let us consider the other expression for the residual norm. Following the expression in Lemma 3.1 we are looking for upper bounds for $|(H_k^{-1} e^1, e^k)|$.

THEOREM 4.2.

$$\|r^k\| \leq \|r^0\| \frac{h_{k+1,k}}{\sigma_{\min}(H_k)}. \tag{4.5}$$

Proof. Straightforwardly we have

$$|(H_k^{-1} e^1, e^k)| \leq \|H_k^{-1} e^1\| \leq \|H_k^{-1}\|. \qquad \square$$

In many cases the smallest singular value $\sigma_{\min}(H_k)$ is far from zero for $k < n$. Hence, Theorem 4.2 says that $\|r^k\|$ goes to zero with $h_{k+1,k}$. However, this bound is too crude, since it is seen in many numerical experiments that $|(H_k^{-1} e^1, e^k)|$ goes to zero before $h_{k+1,k}$ becomes small.

Other lower and upper bounds can be obtained using the characterization of the residual norm in Theorem 3.4, particularly in (3.8). We need the following interlacing result for the singular values that are ordered as usual in descending order, σ_1 being the largest one.

LEMMA 4.3. *Let C be a square matrix of order n and A a matrix of order $n-1$ obtained from C by deleting one row and one column. Then, we have*

$$\sigma_i(C) \geq \sigma_i(A) \geq \sigma_{i+2}(C), \qquad i = 1, \ldots, n-2,$$
$$\sigma_{n-1}(C) \geq \sigma_{n-1}(A).$$

Proof. See [20] or [10]. $\square$

THEOREM 4.4. *Using the notation of Theorem* 3.2*, we have*

$$\|r^k\| \leq \|r^0\| h_{k+1,k} \frac{\sigma_{\min}(\tilde{H}_{k-1})}{\sigma_{\min}(H_k)\sigma_{k-1}(H_k)}. \tag{4.6}$$

Proof. From (3.8) we have

$$\|r^k\| = \|r^0\| h_{k+1,k} \frac{\prod_{i=1}^{k-1} \sigma_i(\tilde{H}_{k-1})}{\prod_{i=1}^{k} \sigma_i(H_k)}.$$

Lemma 4.3 gives that

$$\frac{\sigma_i(\tilde{H}_{k-1})}{\sigma_i(H_k)} \leq 1, \qquad i = 1, \ldots, k-2.$$

Bounding these ratios by 1 gives the upper bound for the residual norm. $\square$

Theorem 4.4 clearly shows that the convergence to zero of the norm of the residual in FOM depends on $\sigma_{\min}(\tilde{H}_k)$ since we have

$$\|r^0\| \frac{\sigma_{\min}(\tilde{H}_k)}{\|h^k\|} \leq \|r^k\| \leq \|r^0\| h_{k+1,k} \frac{\sigma_{\min}(\tilde{H}_{k-1})}{\sigma_{\min}(H_k)\sigma_{k-1}(H_k)}.$$

Results relating the norm of the residual to singular values were also proven in [19].

5. Expression for the GMRES residual norm. To distinguish between the two methods of interest, we denote the variables in FOM by an index O and in GMRES by an index M.

To our knowledge, the following result was first proven in [2]. Unfortunately, it seems that this paper is not well known. Here we give a different and simpler proof of its main result.

THEOREM 5.1. *Using the notation of Theorem* 3.2*, we have*

$$\|r_M^k\|^2 = \frac{\|r^0\|^2}{1 + \|\tilde{H}_k^{-*} h^k\|^2}. \tag{5.1}$$

Proof. The solution z^k of the GMRES least squares problem is theoretically given by solving $(H_k^{(e)})^* H_k^{(e)} z^k = \|r^0\| (H_k^{(e)})^* e^1$. We have

$$(H_k^{(e)})^* H_k^{(e)} = \tilde{H}_k^* \tilde{H}_k + h^k (h^k)^*.$$

Moreover, $(H_k^{(e)})^* e^1 = h^k$. Hence,

$$z^k = \|r^0\| [\tilde{H}_k^* \tilde{H}_k + h^k (h^k)^*]^{-1} h^k.$$

Let us denote $\tilde{z}^k = z^k / \|r^0\|$ and $\xi_k = 1/(1 + \|\tilde{H}_k^{-*} h^k\|^2)$. Using the Sherman–Morrison formula (see [8]), and after some manipulations, we obtain

$$\tilde{z}^k = [\tilde{H}_k^* \tilde{H}_k + h^k (h^k)^*]^{-1} h^k = \tilde{H}_k^{-1} \tilde{H}_k^{-*} h^k - \frac{\|\tilde{H}_k^{-*} h^k\|^2}{1 + \|\tilde{H}_k^{-*} h^k\|^2} \tilde{H}_k^{-1} \tilde{H}_k^{-*} h^k.$$

Then,

$$\tilde{z}^k = \xi_k \tilde{H}_k^{-1} \tilde{H}_k^{-*} h^k.$$

The norm of the GMRES residual is

$$\begin{aligned} \|r_M^k\| &= \|r^0\| \left\| e^1 - \begin{pmatrix} (h^k)^* \\ \tilde{H}_k \end{pmatrix} \tilde{z}^k \right\| \\ &= \|r^0\| \left\| \begin{pmatrix} 1 - (h^k)^* \tilde{z}^k \\ -\tilde{H}_k \tilde{z}^k \end{pmatrix} \right\|. \end{aligned}$$

Therefore,

$$\frac{\|r_M^k\|^2}{\|r^0\|^2} = |1 - (h^k)^* \tilde{z}^k|^2 + \|\tilde{H}_k \tilde{z}^k\|^2.$$

But we have

$$\begin{aligned} 1 - (h^k)^* \tilde{z}^k &= 1 - \xi_k (h^k)^* \tilde{H}_k^{-1} \tilde{H}_k^{-*} h^k \\ &= 1 - \xi_k \| \tilde{H}_k^{-*} h^k \|^2 \\ &= \xi_k. \end{aligned}$$

It implies that

$$\frac{\| r_M^k \|^2}{\| r^0 \|^2} = \xi_k^2 + \xi_k^2 \| \tilde{H}_k^{-*} h^k \|^2 = \xi_k.$$

This proves the result. ∎

From Theorem 5.1 we clearly see the difference between FOM and GMRES. In FOM the denominator of the square of the norm of the relative residual is $|(e^k, \tilde{H}_k^{-*} h^k)|^2$. Therefore, only the last element of $\tilde{H}_k^{-*} h^k$ is involved, and its modulus may eventually be small, giving a large residual norm. In GMRES the denominator is $1 + \| \tilde{H}_k^{-*} h^k \|^2$. All the elements of the vector are involved, and, moreover, the 1 which is added prevents the residual norm from going to infinity. There is quasi stagnation as long as $\| \tilde{H}_k^{-*} h^k \|^2$ is small relative to 1. Note that because $\tilde{H}_k$ is triangular we have

$$\| \tilde{H}_k^{-*} h^k \|^2 = |(e^k, \tilde{H}_k^{-*} h^k)|^2 + \| \tilde{H}_{k-1}^{-*} h^{k-1} \|^2,$$

and, as it is well known, the decrease of the residual norm is monotone.

The result of Theorem 5.1 shows that as long as $\| \tilde{H}_k^{-*} h^k \|$ is small, there is no GMRES convergence. The discussions about the FOM and GMRES residual norms are summarized in the following result.

THEOREM 5.2. *Using the notation of Theorem* 3.2*, the FOM residual norm* $\| r_O^k \|$ *is small if and only if* $|(e^k, \tilde{H}_k^{-*} h^k)|$ *is large. The GMRES residual norm* $\| r_M^k \|$ *is small if and only if* $\| \tilde{H}_k^{-*} h^k \|$ *is large.*

Convergence of FOM implies the convergence of GMRES. From Theorem 5.1 we also obtain easily that

$$\| r_M^k \|^2 = \frac{\| r^0 \|^2}{|(e^k, \tilde{H}_k^{-*} h^k)|^2 + \frac{\| r^0 \|^2}{\| r_M^{k-1} \|^2}} = \frac{1}{\frac{1}{\| r_O^k \|^2} + \frac{1}{\| r_M^{k-1} \|^2}}.$$

This is a well-known result; see [3], [4]. We see that if $|(e^k, \tilde{H}_k^{-*} h^k)|$ is small, $\| r_O^k \|$ is large and we have

$$\| r_M^k \| \approx \| r_M^{k-1} \|.$$

Hence GMRES is almost stagnating. This is the well-known peak-plateau phenomenon. On the contrary, if $|(e^k, \tilde{H}_k^{-*} h^k)|$ is large, the FOM residual norm is small and the GMRES norm is even smaller.

6. Bounds for the GMRES residual norm. From the expression of the residual norm derived in the previous section we can obtain bounds.

THEOREM 6.1. *Using the notation of Theorem* 3.2*, we have*

$$\|r^0\|^2 \frac{[\sigma_{\min}(\tilde{H}_k)]^2}{[\sigma_{\min}(\tilde{H}_k)]^2 + \|h^k\|^2} \le \|r_M^k\|^2 \le \|r_O^k\|^2. \tag{6.1}$$

Proof. We have

$$1 + \|\tilde{H}_k^{-*} h^k\|^2 \le 1 + \|\tilde{H}_k^{-*}\|^2 \|h^k\|^2 = 1 + \frac{\|h^k\|^2}{[\sigma_{\min}(\tilde{H}_k)]^2}. \qquad \square$$

7. Numerical experiments. In this section we compare the bounds we have obtained in the previous sections with the actual values of the residual norm on two examples. We use real matrices and right-hand sides. We first briefly describe the examples and summarize the bounds to set up the notation used in the numerical experiments.

7.1. Examples. The first example (E1) is the matrix Steam2 from the Matrix Market arising from a three-dimensional steam model of oil reservoir. The order is $n = 600$, the condition number is $\kappa(A) = 3.78\ 10^6$, and the extreme singular values are $\min(\sigma_i) = 1238.55$, $\max(\sigma_i) = 4.68\ 10^9$. The eigenvalues of the matrix are real and negative. We use a random right-hand side.

The second example (E2) comes from Liesen and Strakoš in [13]; see also [7]. They discretized

$$-\nu \Delta u + w \cdot \nabla u = 0$$

with $w = [0, 1]^T$ in $\Omega = (0, 1)^2$ with Dirichlet boundary conditions $u = g$ on $\partial\Omega$, using a streamlined upwind Petrov–Galerkin method with bilinear finite elements on a regular Cartesian mesh. The matrix is

$$A = \nu K \otimes M + M \otimes ((\nu + \delta h)K + C),$$

where δ is the stabilization parameter, h is the mesh size, and

$$M = \frac{h}{6}\text{tridiag}(1, 4, 1), \qquad K = \frac{1}{h}\text{tridiag}(-1, 2, -1),$$

$$C = \frac{1}{2}\text{tridiag}(-1, 0, -1)$$

are tridiagonal matrices with constant diagonals. The right-hand side is Example 2.1, page 1995, of Liesen and Strakoš [13]. We use $h = 1/16$, $\nu = 0.01$, and $\delta = 0.34$. This gives a linear system of order 225.

7.2. Bounds. Let us summarize the bounds we would like to compare with the residual norm computed as $\|b - Ax^k\|$. We first have several lower bounds for FOM:

$$\text{flb1: } \|r^0\| \frac{\sigma_{\min}(\tilde{H}_k)}{\|h^k\|} \le \|r_O^k\|,$$

$$\text{flb2: } \|r^0\| \frac{1}{\|\tilde{H}_k^{-*} h^k\|} \le \|r_O^k\|,$$

$$\text{flb3: } \|r^0\| \frac{1}{\|\tilde{H}_k^{-1} e^k\| \|h^k\|} \le \|r_O^k\|.$$

The upper bound for the FOM residual norm is

$$\text{fub: } \|r_O^k\| \leq \|r^0\| h_{k+1,k} \frac{\sigma_{\min}(\tilde{H}_{k-1})}{\sigma_{\min}(H_k)\sigma_{k-1}(H_k)}.$$

For GMRES we have

$$\text{glb: } \|r^0\|^2 \frac{[\sigma_{\min}(\tilde{H}_k)]^2}{[\sigma_{\min}(\tilde{H}_k)]^2 + \|h^k\|^2} \leq \|r_M^k\|^2.$$

7.3. First example. The left part of Figure 7.1 shows the $\log_{10}$ of the FOM error and residual norms for the matrix Steam2. The residual norm is widely oscillating with large peaks but decreases in average from the start. There is stagnation of the norm after 180 iterations. It is interesting to note that the error norm is much smoother (which means that A^{-1} is a smoothing operator) and decreases at almost the same rate as the residual norm. However, it is smaller by several orders of magnitude.

The right part of Figure 7.1 displays the $\log_{10}$ of $\sigma_{\min}(H_k)$, $\sigma_{\min}(H_k^{(e)})$, and $\sigma_{\min}(\tilde{H}_k)$. The smallest singular values of H_k and $H_k^{(e)}$ are clearly different for approximately 30 iterations, and then they are closer and stagnating. The smallest singular value of $\tilde{H}_k$ is relatively close to $\sigma_{\min}(H_k^{(e)})$ for about 30 iterations, but it continues to decrease, whence $\sigma_{\min}(H_k^{(e)})$ stagnates.

The left part of Figure 7.2 shows the norm of the residual and the lower bounds we have established. We see that flb2 (the trivial bound) is very close to the norm of the residual. The other bounds flb1 and flb3 are much smaller, but they exhibit the same rate of decrease as the norm of the residual. The bound flb3 oscillates like the residual norm. The norm of h^k is almost constant during the 200 iterations. The right part of Figure 7.2 displays the upper bound fub. Even though it is not close to the residual norm, it shows its large oscillations and decreases as the same rate.

The left part of Figure 7.3 shows the $\log_{10}$ of the GMRES error and residual norms for the matrix Steam2. Of course, the residual norm is monotonely decreasing but with some intervals of almost stagnation which correspond to the large peaks of the FOM

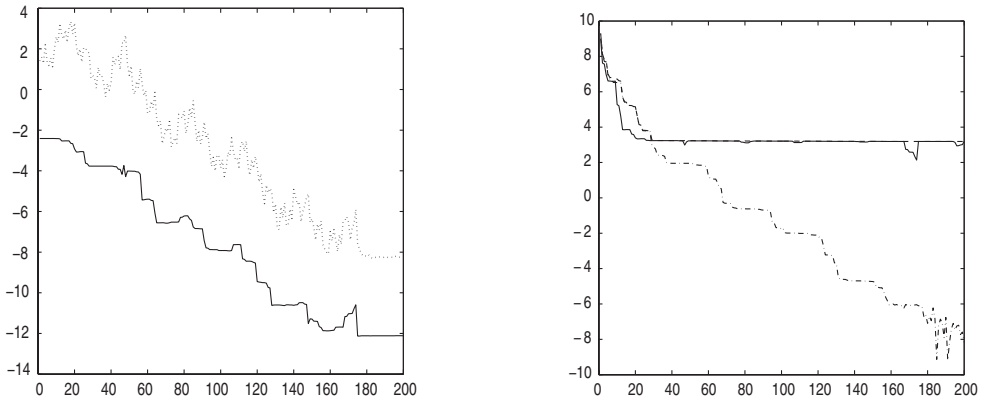

FIG. 7.1. *FOM E*1*: (left)* $\log_{10}$ *of the error norm (solid) and residual norm (dotted); (right)* $\log_{10}$ *of* $\sigma_{\min}(H_k)$ *(solid),* $\sigma_{\min}(H_k^{(e)})$ *(dashed), and* $\sigma_{\min}(\tilde{H}_k)$ *(dot-dashed).*

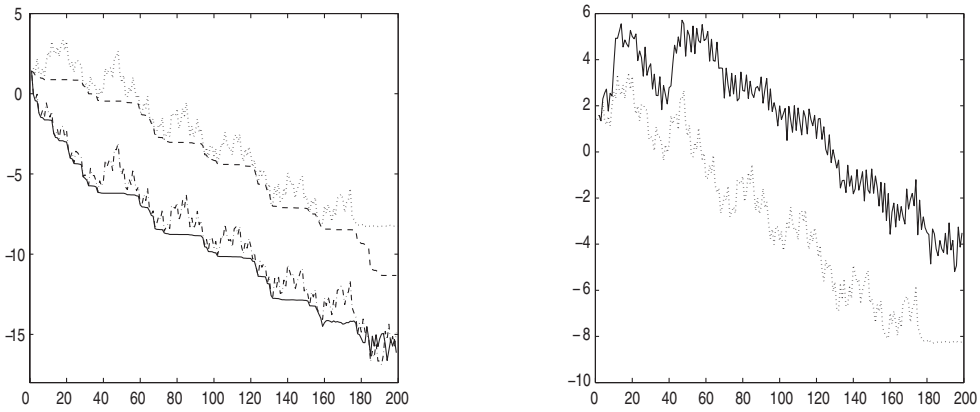

FIG. 7.2. *FOM E*1*: (left)* $\log_{10}$ *of* $\|r^k\|$ *(dotted),* $flb1$ *(solid),* $flb2$ *(dashed), and* $flb3$ *(dot-dashed); (right)* $\log_{10}$ *of* $\|r^k\|$ *(dotted) and fub (solid).*

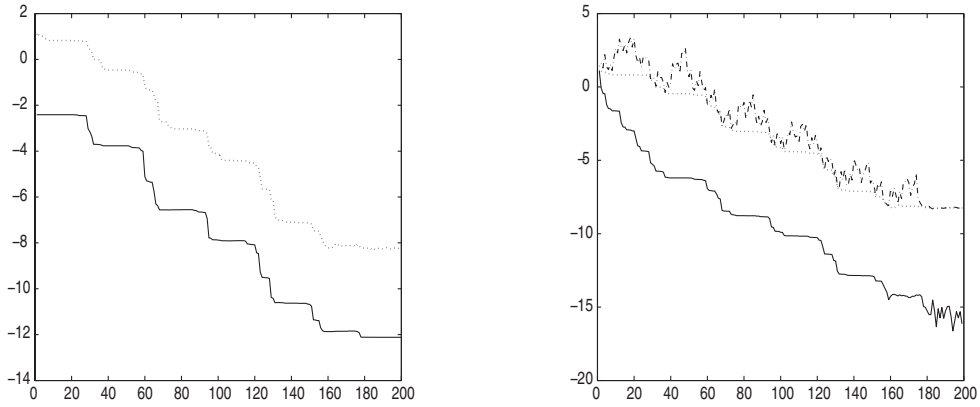

FIG. 7.3. *GMRES E*1*: (left)* $\log_{10}$ *of the norm of the error (solid), residual (dotted); (right)* $\log_{10}$ *of* $\|r_M^k\|$ *(dotted), glb (solid), and* $\|r_O^k\|$ *(dot-dashed).*

residual norm. Nevertheless, the error norms are not much different in both algorithms. As we can see in the right part of the figure, the norm of the FOM residual is from time to time a very good upper bound for the GMRES residual norm. The lower bound has the right rate of decrease and shows some of the plateaus except at the beginning of the iterations where it is decreasing too fast.

Figure 7.4 displays the norm of the last column of $\tilde{H}_k^{-1}$ as a function of k. It increases from the beginning. This explains the decrease of $\sigma_{\min}(\tilde{H}_k)$ since after a few iterations we have

$$\sigma_{\min}(\tilde{H}_k) \approx \frac{1}{\|\tilde{H}_k^{-1}\|_F}.$$

7.4. Second example. The left part of Figure 7.5 shows the $\log_{10}$ of the FOM error and residual norms. Both norms decrease after a stagnation period of 15 iterations.

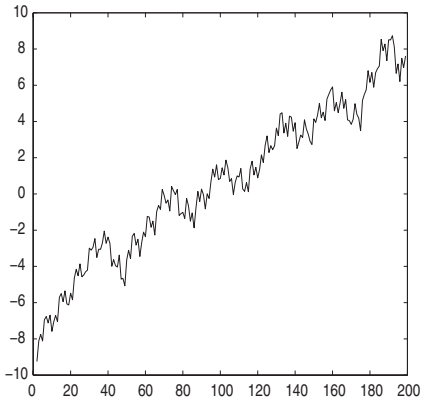

FIG. 7.4. *E1:* $\log_{10}$ *of the norm of the last column of the inverse of* $\tilde{H}_k$.

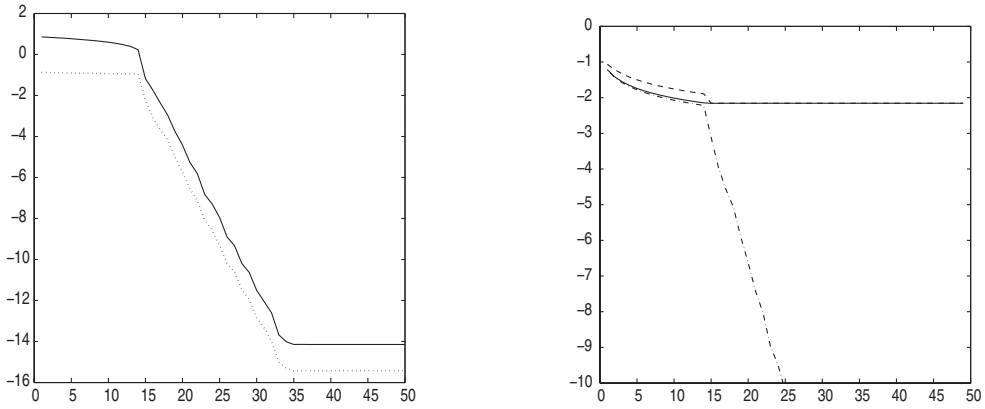

FIG. 7.5. *FOM E2: (left)* $\log_{10}$ *of the norm of the error (solid), residual (dotted); (right)* $\log_{10}$ *of* $\sigma_{\min}(H_k)$ *(solid),* $\sigma_{\min}(H_k^{(e)})$ *(dashed), and* $\sigma_{\min}(\tilde{H}_k)$ *(dot-dashed).*

They reach their smallest values after 35 iterations. The decrease starting at iteration 15 arises because of a small value of the subdiagonal entry of H_k which leads to much larger elements in the inverse of $\tilde{H}_k$. The right part of the figure displays the $\log_{10}$ of $\sigma_{\min}(H_k)$, $\sigma_{\min}(H_k^{(e)})$, and $\sigma_{\min}(\tilde{H}_k)$. The smallest singular values of H_k and $H_k^{(e)}$ are first decreasing and then stagnating after iteration 15. The smallest singular value of $\tilde{H}_k$ follows the same path at the beginning and then decreases very fast.

The lower bounds of $\|r_O^k\|$ are given in the left part of Figure 7.6. They are all quite close to the norm of the residual. The upper bound is displayed in the right part of the figure. The upper bound fub gives good results describing quite well the decrease of the norm.

Figure 7.7 shows the $\log_{10}$ of the GMRES error and residual norms in the left part. The lower bound displayed in the right part of the figure well describes the behavior of the norm of the residual.

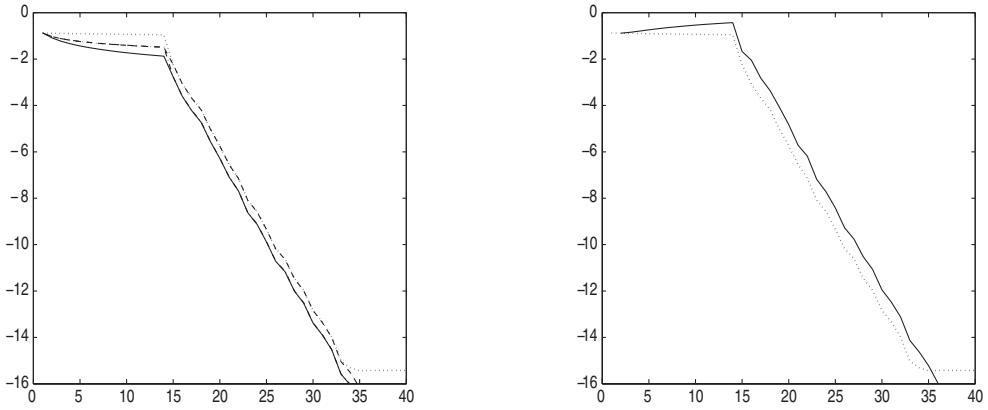

FIG. 7.6. *FOM E2: (left)* $\log_{10}$ *of* $\|r^k\|$ *(dotted), flb*1 *(solid), flb*2 *(dashed), and flb*3 *(dot-dashed); (right)* $\log_{10}$ *of* $\|r_M^k\|$ *(dotted) and fub (solid).*

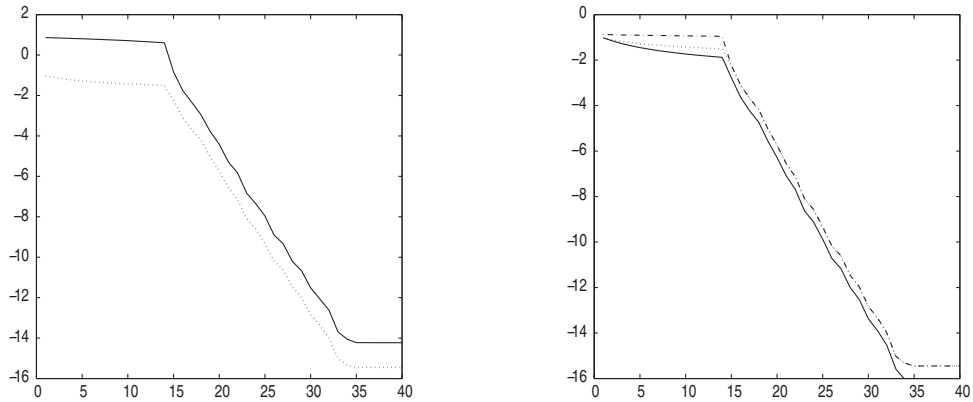

FIG. 7.7. *GMRES E2: (left)* $\log_{10}$ *of the norm of the error (solid), residual (dotted); (right)* $\log_{10}$ *of* $\|r_M^k\|$ *(dotted), glb (solid), and* $\|r_O^k\|$ *(dot-dashed).*

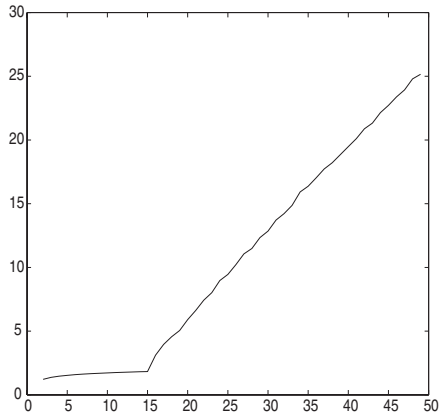

FIG. 7.8. *E2:* $\log_{10}$ *of the norm of the last column of the inverse of* $\tilde{H}_k$.

The norm of the last column of the inverse of $\tilde{H}_k$ is given in Figure 7.8. There is an initial stagnation phase, but after 15 iterations it is linearly increasing in this logarithmic scale. For this example our bounds describe the FOM and GMRES behavior quite accurately.

8. Conclusions. In this paper we have given expressions for the norm of the residual in FOM and GMRES involving a triangular submatrix $\tilde{H}_k$ of the Hessenberg matrix computed by the Arnoldi process during the iterations. We derived lower and upper bounds for the norm of the residual, showing that its decrease depends on the smallest singular value of $\tilde{H}_k$.

The main questions to be addressed in the future are, of course, to find how and why this smallest singular value decreases to zero. Knowing which properties of the matrix A and the right-hand side b could imply this decrease is an interesting and challenging topic of research.

Acknowledgments. The author thanks Miroslav Rozložník and Zdeněk Strakoš for their hospitality at the Institute of Computer Science of the Czech Academy of Sciences. The author also thanks the reviewers for helpful corrections and suggestions which improved the exposition of the paper.

REFERENCES

[1] W. E. Arnoldi, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.
[2] E. H. Ayachour, *A fast implementation for GMRES method*, J. Comput. Appl. Math., 159 (2003), pp. 269–283.
[3] P. N. Brown, *A theoretical comparison of the Arnoldi and GMRES algorithms*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 58–78.
[4] J. Cullum and A. Greenbaum, *Relations between Galerkin and norm-minimizing iterative methods for solving linear systems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 223–247.
[5] M. Eiermann and O. Ernst, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.
[6] D. K. Fadeev, *Properties of the inverse of a Hessenberg matrix*, in Numerical Methods and Computational Issues, Vol. 5, V. P. Ilin and V. N. Kublanovskaya, eds., 1981, pp. 177–179 (in Russian).
[7] B. Fischer, A. Ramage, D. J. Silvester, and A. J. Wathen, *On parameter choice and iterative convergence for stabilised discretisations of advection-diffusion problems*, Comput. Methods Appl. Mech. Engrg., 179 (1999), pp. 179–195.
[8] G. H. Golub and C. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[9] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.
[10] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.
[11] Y. Ikebe, *On inverses of Hessenberg matrices*, Linear Algebra Appl., 24 (1979), pp. 93–97.
[12] I. C. F. Ipsen, *Expressions and bounds for the residual in GMRES*, BIT, 40 (2000), pp. 524–33.
[13] J. Liesen and Z. Strakoš, *GMRES convergence analysis for a convection-diffusion model problem*, SIAM J. Sci. Comput., 26 (2005), pp. 1989–2009.
[14] C. C. Paige and Z. Strakoš, *Residual and backward error bounds in minimum residual Krylov subspace methods*, SIAM J. Sci. Comput., 23 (2002), pp. 1898–1923.
[15] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
[16] Y. Saad, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comp., 37 (1981), pp. 105–126.
[17] Y. Saad, *Practical use of some Krylov subspace methods for solving indefinite and unsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 203–228.
[18] Y. Saad and M. H. Schultz, *GMRES: A generalized minimum residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

[19] H. Sadok, *Analysis of the convergence of the minimal and orthogonal residual methods*, Numer. Algorithms, 40 (2005), pp. 201–216.
[20] R. C. Thompson, *Principal submatrices* IX. *Interlacing inequalities for singular values of submatrices*, Linear Algebra Appl., 5 (1972), pp. 1–12.
[21] H. A. van der Vorst, *Iterative Krylov Methods for Large Linear Systems*, Cambridge University Press, Cambridge, UK, 2003.
[22] J.-P. M. Zemke, *Hessenberg eigenvalue-eigenmatrix relations*, Linear Algebra Appl., 414 (2006), pp. 589–606.
[23] J.-P. M. Zemke, *Abstract perturbed Krylov methods*, Linear Algebra Appl., 424 (2007), pp. 405–434.