SIAM J. SCI. COMPUT.
Vol. 23, No. 5, pp. 1503–1525

LEAST SQUARES RESIDUALS AND MINIMAL RESIDUAL METHODS*

J. LIESEN†, M. ROZLOŽNÍK‡, AND Z. STRAKOŠ§

Abstract. We study Krylov subspace methods for solving unsymmetric linear algebraic systems that minimize the norm of the residual at each step (minimal residual (MR) methods). MR methods are often formulated in terms of a sequence of least squares (LS) problems of increasing dimension. We present several basic identities and bounds for the LS residual. These results are interesting in the general context of solving LS problems. When applied to MR methods, they show that the size of the MR residual is strongly related to the conditioning of different bases of the same Krylov subspace. Using different bases is useful in theory because relating convergence to the characteristics of different bases offers new insight into the behavior of MR methods.

Different bases also lead to different implementations which are mathematically equivalent but can differ numerically. Our theoretical results are used for a finite precision analysis of implementations of the GMRES method [Y. Saad and M. H. Schultz, *SIAM J. Sci. Statist. Comput.*, 7 (1986), pp. 856–869]. We explain that the choice of the basis is fundamental for the numerical stability of the implementation. As demonstrated in the case of Simpler GMRES [H. F. Walker and L. Zhou, *Numer. Linear Algebra Appl.*, 1 (1994), pp. 571–581], the best orthogonalization technique used for computing the basis does not compensate for the loss of accuracy due to an inappropriate choice of the basis. In particular, we prove that Simpler GMRES is inherently less numerically stable than the Classical GMRES implementation due to Saad and Schultz [*SIAM J. Sci. Statist. Comput.*, 7 (1986), pp. 856–869].

Key words. linear systems, least squares problems, Krylov subspace methods, minimal residual methods, GMRES, convergence, rounding errors

AMS subject classifications. 65F10, 65F20, 65G50, 15A42

PII. S1064827500377988

1. Introduction.

Consider a linear algebraic system $Ax = b$, where $A \in \mathcal{R}^{N,N}$ is a nonsingular matrix (generally unsymmetric) and $b \in \mathcal{R}^N$. Krylov subspace methods for solving such systems start with an initial approximation x_0, compute the initial residual $r_0 = b - Ax_0$, and then determine a sequence of approximate solutions $x_1, \ldots, x_n, \ldots$ such that x_n belongs to the linear manifold determined by x_0 and the nth Krylov subspace

$$x_n \in x_0 + \mathcal{K}_n(A, r_0), \quad \mathcal{K}_n(A, r_0) \equiv \mathrm{span}\{r_0, Ar_0, \ldots, A^{n-1}r_0\}\,. \tag{1.1}$$

The nth residual then belongs to the manifold given by r_0 and the shifted Krylov subspace (also called the Krylov residual subspace)

$$r_n \equiv b - Ax_n \in r_0 + A\mathcal{K}_n(A, r_0)\,. \tag{1.2}$$

*Received by the editors September 12, 2000; accepted for publication (in revised form) July 27, 2001; published electronically January 4, 2002.
http://www.siam.org/journals/sisc/23-5/37798.html

†Center for Simulation of Advanced Rockets, University of Illinois at Urbana-Champaign, Urbana, IL 61801 (liesen@uiuc.edu). This author's work was partly supported by the SFB 343, Fakultät für Mathematik, Universität Bielefeld, Germany.

‡Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague (miro@cs.cas.cz). This author's work was supported by the GA CR under grant 101/00/1035. Part of the work was performed while this author visited ETH Zürich from 1998–2000 and Emory University in May, 1999.

§Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322 and Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague (strakos@cs.cas.cz). This author's work was supported by the GA AS CR under grant A1030103.

The choice of x_n is based on some particular additional condition. In this paper we focus on the *minimal residual (MR) principle*

$$\|r_n\| = \min_{u \in x_0 + \mathcal{K}_n(A,r_0)} \|b - Au\|\,, \tag{1.3}$$

which can be equivalently formulated as the *orthogonal projection principle*

$$r_n \perp A\mathcal{K}_n(A, r_0)\,. \tag{1.4}$$

Since A is assumed to be nonsingular, both (1.3) and (1.4) determine the unique sequence of approximate solutions $x_1, \ldots, x_n$; see [26]. Mathematically (in exact arithmetic), there are several different methods and many of their algorithmic variants for generating this sequence. Computationally (in finite precision arithmetic), however, different algorithms for computing the same mathematical sequence may produce different results.

We will call the Krylov subspace methods (1.1) generating mathematically the approximate solutions $x_1, \ldots, x_n$ uniquely determined by the MR principle (1.3) (or by the equivalent orthogonal projection principle (1.4)) *MR Krylov subspace methods* (MR methods).

The MR principle (1.3) represents a least squares (LS) problem, and thus MR methods are often described as a sequence of LS least problems of increasing dimension [26]. In this paper we use general results about LS residuals to analyze the properties of different implementations of MR methods in exact as well as finite precision arithmetic. Our approach is as follows.

In section 2 we present several basic identities and bounds for the norm of the residual $r = c - By$ of the overdetermined LS problem $Bu \approx c$. Specifically, our results relate $\|r\|$ to the singular values of the matrix $[c\gamma, B]$, where $\gamma > 0$ is a scaling parameter, and occasionally some other data. Results of this type have been presented in the literature before (see, e.g., [29]), and they are of importance in studying LS problems in general. While our main focus is on MR methods, only a part of our general LS results are used later in the paper. We believe, however, that the presented LS results which are not directly applied here might be found useful in the context of MR methods in the future.

In section 3 we apply results from section 2 to MR methods for the problem $Ax = b$. In particular, we relate the norm of the MR residual to the conditioning of different bases of $\mathcal{K}_n(A, r_0)$. We derive several necessary and sufficient conditions for fast convergence as well as for stagnation of MR methods. Our results are significantly stronger and more complete than the corresponding results published previously [16, 17]. We point out that our results should not be interpreted as bounds for measuring convergence. As demonstrated in the further sections, results relating residual norm to the conditioning of different bases lead to a new understanding of MR methods.

Section 4 describes the main examples of the MR methods, in particular various forms of the GMRES method [26]. We then apply our theoretical results about the MR residual to finite precision analysis of the important implementations in section 5. Our results explain why the choice of the basis is fundamental for the numerical stability of the implementation. As demonstrated on the example of Simpler GMRES [34], which constructs in exact arithmetic an orthonormal basis of $A\mathcal{K}_n(A, r_0)$, the best orthogonalization technique (Householder reflections) in computing the basis does not compensate for the loss of accuracy due to the inappropriate choice of the basis.

Simpler GMRES is proved inherently less stable than the Classical GMRES implementation [26], which constructs in exact arithmetic an orthonormal basis of $\mathcal{K}_n(A, r_0)$. Our findings are illustrated by numerical experiments.

We denote by $\sigma_i(\cdot)$ the ith largest singular value and by $\sigma_{min}(\cdot)$ the smallest singular value of a given matrix. By $\kappa(\cdot)$ we denote the ratio of the largest to the smallest singular value (condition number). We use $\|\cdot\|$ to denote the 2-norm, e_i to denote the ith vector of the standard Euclidean basis, and I to denote the identity matrix.

2. Basic relations for the LS residual. As the MR methods can be expressed as sequences of LS problems, it will prove useful to collect some basic relations for the LS residual. We will recall some known results, prove several new results, and put the ones known previously in a new context. Most of the results of this section will be used in our analysis of MR methods later in the paper. We believe that they are also of interest in the LS context in general.

Consider an overdetermined linear approximation problem

$$Bu \approx c, \quad B \in \mathcal{R}^{N,n},\ c \in \mathcal{R}^N,\ n < N,\ \mathrm{rank}(B) = n. \tag{2.1}$$

We denote by y the LS solution of (2.1) and by $r = c - By$ the corresponding LS residual,

$$\|r\| \ = \ \|c - By\| \ = \ \min_u \|c - Bu\|\,. \tag{2.2}$$

We introduce a real scaling parameter $\gamma > 0$ and consider a scaled version of (2.1),

$$Bz \approx c\gamma, \quad B \in \mathcal{R}^{N,n},\ c \in \mathcal{R}^N,\ n < N,\ \mathrm{rank}(B) = n. \tag{2.3}$$

Note that if the right-hand side c is replaced in (2.1) and (2.2) by the scaled vector $c\gamma$, the LS solution and the LS residual scale trivially to $z = y\gamma$ and $r\gamma$. We start with general identities relating r to the matrix $[c\gamma, B]$.

THEOREM 2.1. *Suppose that $[c, B] \in \mathcal{R}^{N,n+1}$ has full column rank, and $r \neq 0$ is the residual of the LS problem* (2.1)–(2.2). *Let $\gamma > 0$ be a real parameter. Then*

$$e_1^T\,[c\gamma, B]^\dagger = \frac{r^T}{\gamma\|r\|^2} \qquad \textit{and} \qquad \gamma\|r\| \ = \ \frac{1}{\{e_1^T([c\gamma, B]^T[c\gamma, B])^{-1}e_1\}^{\frac{1}{2}}}, \tag{2.4}$$

where $X^\dagger$ denotes the Moore–Penrose generalized inverse of a matrix X.

Proof. For any matrix X the Moore–Penrose pseudoinverse $X^\dagger$ satisfies $XX^\dagger X = X$ (see, e.g., [5]), which using the symmetry of $XX^\dagger$ gives $X^T = X^T XX^\dagger$. Substituting $X = [c\gamma, B]$, we receive the following simple identities:

$$\begin{aligned}\gamma r^T \ &= [1, -\gamma y^T]\,[c\gamma, B]^T = [1, -\gamma y^T]\,[c\gamma, B]^T\,[c\gamma, B][c\gamma, B]^\dagger \\ &= \gamma r^T\,[c\gamma, B][c\gamma, B]^\dagger\,.\end{aligned}$$

Since r is orthogonal to the columns of B, $\gamma r^T\,[c\gamma, B] \ = \ \gamma^2(r^T c)\,e_1^T \ = \ \gamma^2\|r\|^2\,e_1^T$, which proves the first part of the theorem. The second part follows from the identity $\|e_1^T\,[c\gamma, B]^\dagger\|^2 = e_1^T([c\gamma, B]^T[c\gamma, B])^{-1}e_1$, which can be verified by a straightforward calculation. ∎

The first equality in (2.4) was essentially proven (though neither the statement nor the proof were formulated explicitly in the form presented here) in [28, relations

(2.5), (2.6), (3.7), and (3.8)]. Later it was presented (with $\gamma = 1$) in [16, Lemma 7.1] (see also other references therein).

It is important to notice that for an arbitrary nonsingular matrix $M \in \mathcal{R}^{n,n}$,

$$\|r\| \;=\; \|c - By\| \;=\; \|c - (BM)(M^{-1}y)\| \;=\; \min_u \|c - (BM)u\| \,.$$

As a consequence of this simple observation, (2.4) will hold when B is replaced by BM. A particularly useful choice is $M = R^{-1}$, where R is the upper triangular factor of a QR-factorization of B.

COROLLARY 2.2. *Using the assumptions and the notation of Theorem* 2.1*, and a QR-factorization $B = QR$ of the matrix B,*

$$e_1^T\,[c\gamma, Q]^\dagger = \frac{r^T}{\gamma\|r\|^2} \qquad \textit{and} \qquad \gamma\|r\| \;=\; \frac{1}{\{e_1^T([c\gamma, Q]^T[c\gamma, Q])^{-1}e_1\}^{\frac{1}{2}}}\,. \tag{2.5}$$

It may look a bit surprising that the first rows of the matrices $[c\gamma, B]^\dagger$ and $[c\gamma, Q]^\dagger$ are identical. A second look reveals that this fact is simple and natural.

Consider a full column rank matrix $X = [c\gamma, B] \in \mathcal{R}^{N,n+1}$. Then the rows of $X^\dagger$ are linear combinations of the rows of X^T (the transposed columns of X), and $X^\dagger X = I$. The last relation can be interpreted geometrically as an orthogonal relation between the rows of $X^\dagger$ and the columns of X. Denote by $s = e_1^T X^\dagger$ the first row of $X^\dagger$. Then s is orthogonal to all but the first column of X; i.e., it is orthogonal to the columns of the matrix B. Because s represents a linear combination of c^T and the transposed columns of B, it must be equal to a scalar multiple of the transposed residual $r^T = (c - By)^T$ for the LS problem (2.1)–(2.2). The identity $(\zeta r^T)\,(c\gamma) = 1$ then immediately gives $\zeta = \gamma^{-1}\|r\|^{-2}$.

The orthogonality idea clearly applies with no change when B is replaced by any matrix BM, where $M \in \mathcal{R}^{n,n}$ is nonsingular. The geometrical interpretation of the generalized inverse is simple but revealing.

The following theorem relates the norm of the LS residual (2.2) to the singular values of the matrices B, $[c\gamma, B]$, and $[c\gamma, Q]$. This theorem plays a substantial role in our further analysis.

THEOREM 2.3. *Suppose that $[c, B] \in \mathcal{R}^{N,n+1}$ has full column rank, and $r \neq 0$ is the residual of the LS problem* (2.1)–(2.2)*. Let $B = QR$ be a QR-factorization of the matrix B and $\gamma > 0$ be a real parameter. Then*

$$\|r\| \;=\; \frac{\sigma_{min}([c\gamma, B])}{\gamma} \prod_{j=1}^{n} \frac{\sigma_j([c\gamma, B])}{\sigma_j(B)} \tag{2.6}$$

$$\;=\; \frac{1}{\gamma}\,\sigma_{min}([c\gamma, Q])\,\sigma_1([c\gamma, Q]). \tag{2.7}$$

Furthermore,

$$\kappa([c\gamma, Q]) \;=\; \frac{\alpha + \left(\alpha^2 - 4\gamma^2\|r\|^2\right)^{1/2}}{2\gamma\,\|r\|}\,, \quad \|r\| \;=\; \frac{\alpha}{\gamma}\,\frac{\kappa([c\gamma,\, Q])}{\kappa([c\gamma,\, Q])^2 + 1}, \tag{2.8}$$

where $\alpha \equiv 1 + \gamma^2\|c\|^2$.

Proof. Using the orthogonality of the columns of the matrix Q, the right-hand side c and the residual r are related by the identity

$$c = Qh + r, \quad h \equiv Q^T c, \quad \|c\|^2 = \|h\|^2 + \|r\|^2\,.$$

Now consider an orthogonal matrix $U \in \mathcal{R}^{n,n}, U^T U = I$, such that $Uh = \|h\|e_1$. Then

$$[c\gamma, Q]^T[c\gamma, Q] = \begin{bmatrix} 1 & 0 \\ 0 & U^T \end{bmatrix} \begin{bmatrix} \gamma^2\|c\|^2 & \gamma\|h\|e_1^T \\ \gamma\|h\|e_1 & I \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U \end{bmatrix}, \tag{2.9}$$

$$[c\gamma, B]^T[c\gamma, B] = \begin{bmatrix} 1 & 0 \\ 0 & R^T \end{bmatrix} [c\gamma, Q]^T[c\gamma, Q] \begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}. \tag{2.10}$$

Identity (2.9) shows that all but two of the eigenvalues of $[c\gamma, Q]^T[c\gamma, Q]$ are equal to one. The two remaining eigenvalues are easily determined as the eigenvalues of the left principal two-by-two block,

$$\sigma_1^2([c\gamma, Q]) = \frac{\alpha + (\alpha^2 - 4\gamma^2\|r\|^2)^{1/2}}{2}, \tag{2.11}$$

$$\sigma_{min}^2([c\gamma, Q]) = \frac{\alpha - (\alpha^2 - 4\gamma^2\|r\|^2)^{1/2}}{2}, \tag{2.12}$$

where $\alpha \equiv 1 + \gamma^2\|c\|^2$. (Notice that $\alpha^2 - 4\gamma^2\|r\|^2 \geq (1 - \gamma^2\|c\|)^2 \geq 0$.) Using

$$\kappa([c\gamma, Q]) = \sigma_1([c\gamma, Q])/\sigma_{min}([c\gamma, Q]),$$

(2.8) is obtained by a simple algebraic manipulation.

Evaluating the determinants on both sides of (2.9) yields

$$\det([c\gamma, Q]^T[c\gamma, Q]) = \sigma_1^2([c\gamma, Q])\, \sigma_{min}^2([c\gamma, Q]) = \gamma^2\|r\|^2,$$

which shows (2.7). Similarly, transformation (2.10) yields

$$\begin{aligned} \det([c\gamma, B]^T[c\gamma, B]) &= \prod_{j=1}^{n+1} \sigma_j^2([c\gamma, B]) \\ &= \det([c\gamma, Q]^T[c\gamma, Q])\, \det(R^T R) = \gamma^2\|r\|^2 \prod_{j=1}^{n} \sigma_j^2(B), \end{aligned}$$

which proves (2.6). ∎

The relations (2.8) generalize results presented in [19, section 5.5.2]. The identity (2.6) (with $\gamma = 1$) was first shown by Van Huffel and Vandewalle [29, Theorem 6.9], and it also appeared (with a different proof) in [20].

Van Huffel and Vandewalle [29, Theorem 6.10] gave the following lower and upper bounds for $\|r\|$ (with $\gamma = 1$) in terms of $\sigma_{min}([c\gamma, B])$ (see also [20]). Let

$$\delta(\gamma) \equiv \sigma_{min}([c\gamma, B])/\sigma_{min}(B). \tag{2.13}$$

Then

$$\frac{\sigma_{min}([c\gamma, B])}{\gamma} \leq \|r\| \leq \frac{\sigma_{min}([c\gamma, B])}{\gamma} \left\{1 - \delta(\gamma)^2 + \frac{\gamma^2\|c\|^2}{\sigma_{min}^2(B)}\right\}^{\frac{1}{2}}. \tag{2.14}$$

Bounds for $\|r\|$ in terms of the minimal singular values of B and $[c\gamma, B]$, and as little additional information as possible, were studied in detail in [20]. In particular, when B has full column rank and

$$c \not\perp \{\text{left singular vector subspace of } B \text{ corresponding to } \sigma_{min}(B)\}, \tag{2.15}$$

then the following bounds were given in [20]:

$$\begin{aligned} \sigma_{min}([c\gamma, B])\,\{\gamma^{-2} + \|y\|^2\}^{\frac{1}{2}} \;\le\; \|r\| \\ \le\; \sigma_{min}([c\gamma, B])\,\left\{\gamma^{-2} + \frac{\|y\|^2}{1-\delta(\gamma)^2}\right\}^{\frac{1}{2}}. \end{aligned} \tag{2.16}$$

Though (2.14) can be derived from (2.16) (and not vice versa; see [20]), the upper bound in (2.16) is not always tighter than the upper bound in (2.14). When $\delta(\gamma) \approx 1$ and $\|r\| \approx \|c\|$, it is possible for the upper bound in (2.14) to be smaller than that in (2.16). However, in this case the upper bound in (2.14) becomes trivial. For details, see [20].

For $\delta(\gamma) = 1$ the upper bound in (2.16) does not exist. It was shown in [22] that if (2.15) holds, then $\delta(\gamma) < 1$ for all $\gamma > 0$. As explained in [22], the role of the assumption (2.15) is truly fundamental. If it does not hold, both theory and computation in errors-in-variables modeling are enormously complicated by the possible case $\delta(\gamma) = 1$. Fortunately, nearly all practical problems will satisfy (2.15). Nevertheless, it is instructive to consider possible cases where (2.15) does not hold, so that $\delta(\gamma) = 1$ is possible.

The lower bound in (2.14) shows that we can make $\sigma_{min}([c\gamma, B])$ arbitrarily small by taking γ small and thus ensure $\delta(\gamma) < 1$ in (2.13). How small must γ be to ensure this? The next theorem answers a variant of this question. Given $\sigma_{min}(B)$ and $\|c\|$, it shows that there is a γ_0 such that $\gamma < \gamma_0$ ensures $\delta(\gamma) < 1$, but $\gamma = \gamma_0$ does not.

THEOREM 2.4. *Suppose that $[c, B] \in \mathcal{R}^{N,n+1}$ has full column rank, y is the solution, and $r \neq 0$ is the residual of the LS problem (2.1)–(2.2). Let $\gamma > 0$ be a real parameter, and $\delta(\gamma) \equiv \sigma_{min}([c\gamma, B])/\sigma_{min}(B)$. Define $\gamma_0 \equiv \sigma_{min}(B)/\|c\|$. Then*

$$\delta(\gamma) < 1 \;\; \textit{for all} \;\; \gamma < \gamma_0. \tag{2.17}$$

Moreover,

$$y = 0 \;\; (r = c) \quad \textit{if and only if} \quad \delta(\gamma_0) = 1. \tag{2.18}$$

Proof. Note that when $\gamma < \gamma_0$, then $\|c\gamma\| < \sigma_{min}(B)$. Therefore $\sigma_{min}([c\gamma, B]) < \sigma_{min}(B)$, i.e., $\delta(\gamma) < 1$.

Now assume that the LS problem (2.1)–(2.2) has the trivial solution $y = 0$ $(r = c)$. Then $B^T c = 0$, which yields

$$[c\gamma, B]^T [c\gamma, B] = \begin{bmatrix} \|c\|^2\gamma^2 & 0 \\ 0 & B^T B \end{bmatrix}.$$

Thus, $\sigma_{min}([c\gamma, B]) = \min\{\|c\|\gamma,\ \sigma_{min}(B)\}$, $\delta(\gamma) = \min\{\|c\|\gamma/\sigma_{min}(B),\ 1\}$, and $\delta(\gamma_0) = 1$. Conversely, (2.14) gives with $\gamma = \gamma_0$,

$$\delta(\gamma_0)\,\|c\| \;\le\; \|r\| \;\le\; \delta(\gamma_0)\,\|c\|\,\{2 - \delta(\gamma_0)^2\}^{1/2}, \tag{2.19}$$

which for $\delta(\gamma_0) = 1$ reduces to $\|c\| \le \|r\| \le \|c\|$, i.e., $\|r\| = \|c\|$, which completes the proof. ∎

We see that $\gamma_0 \equiv \sigma_{min}(B)/\|c\|$ represents an important number. For $\gamma < \gamma_0$ the value of $\delta(\gamma)$ is always strictly less than one, and $\delta(\gamma_0) = 1$ if and only if the LS problem (2.1)–(2.2) has the trivial solution $y = 0$. Moreover, (2.19) shows that $\|r\|$ is significantly smaller than $\|c\|$ if and only if $\delta(\gamma_0)$ is significantly smaller than one. As an application, we will show in section 3 how these results characterize stagnation or near stagnation of MR methods.

One consequence of Theorem 2.4 can be stated as follows: Consider a rectangular matrix (here B) having full column rank and an additional column (here $c\gamma$). If the norm of the additional column is smaller than the smallest singular value of the matrix (here if $\gamma < \gamma_0$), then appending the column necessarily decreases the smallest singular value. If the norm of the appended column is equal to the smallest singular value of the matrix (here if $\gamma = \gamma_0$), then appending the column to the matrix does not change the smallest singular value if and only if the appended column is orthogonal to all the columns (all the left singular vectors) of the original matrix. This is a somewhat specialized result because of the norm of the added column fixed to $\sigma_{min}(B)$. Note that the condition is linear. The general necessary and sufficient condition under which adding a column (with a norm larger or equal to $\sigma_{min}(B)$) to a matrix B does not alter the smallest singular value was given in [22]. The added column must be orthogonal to the left singular vector subspace of B corresponding to $\sigma_{min}(B)$, and the left-hand side of the (deflated) secular equation [22, relation (3.4)] must be nonnegative at $\sigma_{min}(B)$. Theorem 2.4 can also be derived from this. The second part of the condition from [22] is nonlinear.

Theorem 2.4 and the consequence stated above must be understood in their proper context. It was pointed out in [22] that nearly all practical problems will satisfy (2.15), that any problem $Bu \approx c$ can be reduced to a core problem satisfying (2.15), and that for many formulations it makes sense only to consider problems satisfying (2.15). Also, if the problem satisfies (2.15), then $\delta(\gamma) < 1$ for all $\gamma > 0$, and in this case (2.17) and (2.18) are irrelevant. On the other hand, (2.19) seems to be a generally useful result. Thus $\gamma_0 \equiv \sigma_{min}(B)/\|c\|$ is a significant quantity, as can be seen from the interesting but rarely practical properties (2.17) and (2.18), and the interesting and compact bounds (2.19). Note also that in many practical problems of interest, $\gamma_0 \equiv \sigma_{min}(B)/\|c\|$ will be a *very* small number. In particular, this suggests that for a general LS problem the above "column addition" result will be of minor practical use. It is, however, important theoretically because it offers a new insight into the stagnation or near stagnation of the MR methods.

Finally, for completeness, consider a QR-factorization $B = QR$. Replacing B by $BR^{-1} = Q$ and y by Ry (notice that $\|Ry\| = \|By\|$) gives the analogies of (2.14) and (2.16),

$$\frac{\sigma_{min}([c\gamma, Q])}{\gamma} \;\le\; \|r\| \;\le\; \frac{\sigma_{min}([c\gamma, Q])}{\gamma}\,\{1 - \sigma^2_{min}([c\gamma, Q]) + \gamma^2\|c\|^2\}^{\frac{1}{2}}, \tag{2.20}$$

$$\begin{aligned}\sigma_{min}([c\gamma, Q])\,\{\gamma^{-2} + \|By\|^2\}^{\frac{1}{2}} \;&\le\; \|r\| \\ &\le\; \sigma_{min}([c\gamma, Q])\,\left\{\gamma^{-2} + \frac{\|By\|^2}{1 - \sigma_{min}([c\gamma, Q])^2}\right\}^{\frac{1}{2}}.\end{aligned} \tag{2.21}$$

Theorem 2.4 can be reformulated in a similar way. It is interesting to note that the bounds (2.20) do not give additional information. Indeed, since $\sigma_1([c\gamma, Q]) \ge 1$, the lower bound in (2.20) follows immediately from (2.7). And since $\{1 - \sigma^2_{min}([c\gamma, Q]) + \gamma^2\|c\|^2\}^{1/2} = \sigma_1([c\gamma, Q])$, the upper bound is a weak reformulation of (2.7) only.

In the following section we apply results of this section to the MR Krylov subspace methods.

3. Characteristics of the basis and the size of the MR residual. Let $\rho_0 \equiv \|r_0\|$, $v_1 \equiv r_0/\rho_0$, $w_1 \equiv Av_1/\|Av_1\|$. Consider two sequences of orthonormal vectors, $v_1, v_2, \ldots$ and $w_1, w_2, \ldots$, such that for each iterative step n,

$$\mathcal{K}_n(A, r_0) \;=\; \mathrm{span}\{v_1, \ldots, v_n\}, \;\; V_n \equiv [v_1, \ldots, v_n], \;\; V_n^T V_n = I, \tag{3.1}$$

$$A\mathcal{K}_n(A, r_0) \;=\; \mathrm{span}\{w_1, \ldots, w_n\}, \;\; W_n \equiv [w_1, \ldots, w_n], \;\; W_n^T W_n = I. \tag{3.2}$$

Then the MR principle (1.3) can be formulated as

$$\|r_n\| \;=\; \min_{u \in \mathcal{R}^n} \|r_0 - AV_n u\| \tag{3.3}$$

$$\;=\; \min_{u \in \mathcal{R}^n} \|r_0 - W_n u\| \,. \tag{3.4}$$

The MR residual at step n is therefore the LS residual for the LS problems $AV_n u \approx v_1\rho_0$ and $W_n u \approx v_1\rho_0$.

The application of the results presented in section 2 to (3.3) and (3.4) is straightforward: For the nth step of an MR method we consider $c \equiv r_0 = v_1\rho_0$, $B \equiv AV_n$, $Q \equiv W_n$, and $r \equiv r_n$. The scaling parameter $\gamma > 0$ offers some flexibility. While it seems natural to use $\gamma \equiv \|r_0\|^{-1} = \rho_0^{-1}$, other values of γ also prove useful; cf. [21] and our discussion below.

With $\gamma \equiv \|r_0\|^{-1} = \rho_0^{-1}$, Theorem 2.3 and relations (2.7) and (2.8) give the following identities for the relative residual norm $\|r_n\|/\rho_0$:

$$\|r_n\|/\rho_0 \;=\; \sigma_{min}([v_1,\, W_n])\; \sigma_1([v_1,\, W_n]) \tag{3.5}$$

$$\;=\; \frac{2\,\kappa([v_1,\, W_n])}{\kappa([v_1,\, W_n])^2 + 1}. \tag{3.6}$$

Identities (3.5) and (3.6) show that the conditioning of the basis $[v_1,\, W_n]$ of the Krylov subspace $\mathcal{K}_{n+1}(A, r_0)$ is fully determined (except for an unimportant multiplicative factor) by the convergence of the MR methods, and vice versa. In other words,

$$\|r_n\| = \rho_0 \quad \text{if and only if} \quad \kappa([v_1,\, W_n]) = 1, \tag{3.7}$$

and the relative residual norm $\|r_n\|/\rho_0$ is small if and only if $[v_1,\, W_n]$ is ill-conditioned.

The previous statement can also be mathematically expressed by inequalities. Dividing both the numerator and the denominator in (3.6) by $\kappa([v_1, W_n])$ gives in a simple way the bounds

$$\kappa([v_1, W_n])^{-1} \;\le\; \|r_n\|/\rho_0 \;\le\; 2\,\kappa([v_1, W_n])^{-1}. \tag{3.8}$$

The upper bound in (3.8) was published by Walker and Zhou [34, Lemma 3.1]. It is interesting to note that, because of (2.11),

$$1 \;\le\; \sigma_1([v_1, W_n]) \;\le\; \sqrt{2}, \tag{3.9}$$

which shows that the size of $\kappa([v_1, W_n])$ is in fact determined by the smallest singular value $\sigma_{min}([v_1, W_n])$.

Relations between the size of the residuals of the MR methods and the condition number of matrices $[v_1, W_n]$ and $[r_0, W_n]$ were studied in [19, section 5.5.2]. We will generalize the result [19, relation (5.48)] and develop an elegant tool for quantification of the influence of the scaling parameter γ.

THEOREM 3.1. *Let r_0, r_n, and W_n be as in (3.4), $\rho_0 \equiv \|r_0\|, v_1 \equiv r_0/\rho_0$, and $\gamma > 0$. Then*

$$\kappa([r_0\gamma, W_n]) \;\geq\; \kappa([v_1, W_n]) \;+\; \frac{\gamma\,(\rho_0 - \gamma^{-1})^2}{2\|r_n\|}. \tag{3.10}$$

Proof. Using (2.8) with the particular choices $c \equiv r_0,\ Q \equiv W_n,\ \gamma > 0$, and $c \equiv r_0,\ Q \equiv W_n,\ \gamma \equiv \rho_0^{-1} = \|r_0\|^{-1}$ gives

$$\begin{aligned}
&\kappa([r_0\gamma, W_n]) - \kappa([v_1, W_n]) \\
&\quad= \frac{1 + \gamma^2\rho_0^2 + [(1+\gamma^2\rho_0^2)^2 - 4\gamma^2\|r_n\|^2]^{1/2}}{2\gamma\|r_n\|} - \frac{2 + [4 - 4\rho_0^{-2}\|r_n\|^2]^{1/2}}{2\rho_0^{-1}\|r_n\|} \\
&\quad= \frac{\gamma^{-1} + \gamma\rho_0^2 + [(\gamma^{-1}+\gamma\rho_0^2)^2 - 4\|r_n\|^2]^{1/2} - 2\rho_0 - [4\rho_0^2 - 4\|r_n\|^2]^{1/2}}{2\|r_n\|} \\
&\quad= \frac{\gamma\,(\rho_0 - \gamma^{-1})^2}{2\|r_n\|} \;+\; \frac{[(\gamma^{-1}+\gamma\rho_0^2)^2 - 4\|r_n\|^2]^{1/2} - [4\rho_0^2 - 4\|r_n\|^2]^{1/2}}{2\|r_n\|} \\
&\quad\geq \frac{\gamma\,(\rho_0 - \gamma^{-1})^2}{2\|r_n\|}. \qquad \square
\end{aligned}$$

Clearly, $\kappa([r_0\gamma, W_n])$ is minimal for $\gamma = \rho_0^{-1}$, and the minimum is equal to $\kappa([v_1, W_n])$ (see also [8]). If $\gamma \neq \rho_0^{-1}$, then with the residual norm $\|r_n\|$ decreasing towards zero the condition number $\kappa([r_0\gamma, W_n])$ grows much faster than $\kappa([v_1, W_n])$. The results considering the matrix $[r_0\gamma, W_n]$ will be particularly useful for our discussion of MR implementations based on the orthogonal projection principle (1.4) in section 5.

With $c \equiv r_0$, $r \equiv r_n$, $y \equiv y_n$, and $B \equiv AV_n$, (2.16) gives the following bounds for the residual norm in terms of $\sigma_{min}([r_0\gamma, AV_n])$:

$$\begin{aligned}
\sigma_{min}([r_0\gamma, AV_n])\,\{\gamma^{-2} + \|y_n\|^2\}^{\frac{1}{2}} \;&\leq\; \|r_n\| \\
&\leq\; \sigma_{min}([r_0\gamma, AV_n])\,\left\{\gamma^{-2} + \frac{\|y_n\|^2}{1 - \delta_n(\gamma)^2}\right\}^{\frac{1}{2}},
\end{aligned} \tag{3.11}$$

where $\delta_n(\gamma) \equiv \sigma_{min}([r_0\gamma, AV_n])/\sigma_{min}(AV_n)$. As mentioned in section 2, the upper bound in (3.11) becomes intriguing for $\delta_n(\gamma) \approx 1$, and for $\delta_n(\gamma) = 1$ it is not defined.

The convergence of the MR methods and the situation $\delta_n(\gamma) = 1$ or $\delta_n(\gamma) \approx 1$ are related by Theorem 2.4. Define $\gamma_0^{(n)} \equiv \sigma_{min}(AV_n)/\rho_0$. Then $\delta_n(\gamma) < 1$ for all $\gamma < \gamma_0^{(n)}$ and

$$\begin{aligned}
\|r_n\| \;&=\; \rho_0 \quad \Leftrightarrow \\
\delta_n(\gamma_0^{(n)}) \equiv \frac{\sigma_{min}([v_1\sigma_{min}(AV_n), AV_n])}{\sigma_{min}(AV_n)} \;&=\; \sigma_{min}([v_1, AV_n/\sigma_{min}(AV_n)]) = 1.
\end{aligned} \tag{3.12}$$

Moreover, (2.19) gives

$$\delta_n(\gamma_0^{(n)}) \;\leq\; \|r_n\|/\rho_0 \;\leq\; \sqrt{2}\,\delta_n(\gamma_0^{(n)}), \tag{3.13}$$

which shows that the rate of convergence of the MR methods is determined by the size of $\delta_n(\gamma_0^{(n)})$. Summarizing, the MR methods stagnate in steps 1 through n if and only if $\delta_n(\gamma_0^{(n)}) = 1$, and they nearly stagnate in steps 1 through n if and only if $\delta_n(\gamma_0^{(n)}) \approx 1$. However, this specific link between convergence of the MR methods and the value of $\delta_n(\gamma)$ can be made for $\gamma = \gamma_0^{(n)}$ only. In particular, when $\delta_n(\gamma_1) = 1$ for some $\gamma_1 > \gamma_0^{(n)}$, the MR methods do not necessarily stagnate or nearly stagnate. They may exhibit very fast convergence while $\delta_n(\gamma_1) \approx 1$ and very slow convergence while $\delta_n(\gamma_1) \ll 1$. (For more details, see [21].)

For $\gamma = \gamma_0^{(n)}$ there is an interesting relationship between the smallest singular values of the matrices $[v_1, AV_n/\sigma_{min}(AV_n)]$ and $[v_1, W_n]$: (3.5), (3.9), and (3.13) yield

$$\sigma_{min}([v_1, AV_n/\sigma_{min}(AV_n)]) \leq \sqrt{2}\,\sigma_{min}([v_1, W_n]) \leq 2\,\sigma_{min}([v_1, AV_n/\sigma_{min}(AV_n)]),$$

which shows that these smallest singular values are very close to each other.

Using the matrix $[r_0\gamma, AV_n]$ instead of $[v_1, W_n]$ may seem unwise because it necessarily brings into play the potentially ill-conditioned matrix AV_n (in comparison to W_n having orthonormal columns). However, as shown in [22, 21], bounds using the matrix $[r_0\gamma, AV_n]$ are very useful for the analysis of the modified Gram–Schmidt implementation of Classical GMRES. Notice that the bounds (3.11) are not based on singular values only. Using $\|y_n\|$, the norm of the MR approximate solution, makes (3.11) amazingly tight [22]. The parameter γ offers flexibility required for the analysis of the GMRES method [21].

It is also possible to consider other bases of the Krylov subspaces or Krylov residual subspaces which lead to other matrices, identities, and bounds. For example, Ipsen [16, 17] used the matrix $K_{n+1} = [r_0, Ar_0, \ldots, A^n r_0]$, got the identity

$$\|r_n\| = 1/\|e_1^T\, K_{n+1}^{\dagger}\| \tag{3.14}$$

(cf. Theorem 2.1), and developed the bound $\|r_n\|/\rho_0 \geq 1/(\|K_{n+1}\|\,\|K_{n+1}^{\dagger}\|)$. However, any bound based directly on the matrix K_{n+1} necessarily suffers from the potential ill-conditioning of the matrix $[Ar_0, \ldots, A^n r_0]$. Consider the QR-decomposition $[Ar_0, \ldots, A^n r_0] = W_n R_n$. In light of the results presented above (see, in particular, (2.5), (3.5), and (3.6)), the upper triangular factor R_n containing *all* the potential ill-conditioning of the matrix $[Ar_0, \ldots, A^n r_0]$ is mathematically in no relation whatsoever to the residual r_n and to the convergence of any MR method measured by the residual norm. Except for some (rather special) examples, bounds based on the matrix K_{n+1} are therefore necessarily much weaker than the bounds based on the matrices $[r_0\gamma, W_n]$ and $[r_0\gamma, AV_n]$.

In the following we use our theoretical results to obtain new insight into the numerical behavior of MR methods.

4. Implementations of the MR methods. Numerous residual norm minimizing Krylov subspace methods have been proposed in the last decades [18, 30, 35, 1, 11, 26]. Resulting from different approaches, they generate (under different assumptions) approximate solutions satisfying (1.3) and (1.4). Though they are, under some particular assumptions, mathematically equivalent, they differ in various algorithmic aspects, and, consequently, in their numerical behavior.

We will concentrate on two main approaches which explicitly compute the basis vectors $v_1, v_2, \ldots, v_n$ (respectively, $v_1, w_1, \ldots, w_{n-1}$) defined in (3.1) and (3.2). In the

first approach, the approximate solution x_n is expressed as

$$x_n \;=\; x_0 \;+\; V_n\, y_n$$

for some y_n, and the residual norm is bounded in terms of $\sigma_{min}([v_1\rho_0\gamma, AV_n])$ via (3.11). In the second approach the approximate solution is expressed as

$$x_n \;=\; x_0 \;+\; [v_1, W_{n-1}]\, t_n$$

for some t_n, and for the residual norm we have the identities (3.5)–(3.6). At first sight the second approach seems more attractive because it gives a cleaner relation between the residual norm (which is minimized at every step) and the conditioning of the computed basis. Its implementation is also simpler. On the other hand, the fact that the approximate solution is in this approach determined via the basis vectors $v_1, w_1, \ldots, w_{n-1}$ which *are not mutually orthogonal* raises some suspicions about potential numerical problems. In this section we recall implementations of both approaches resulting in different variants of the GMRES algorithm. In section 5 we will discuss their numerical properties.

A variety of MR methods that do not explicitly compute the vectors $v_1, v_2, \ldots, v_n$ or $v_1, w_1, \ldots, w_{n-1}$ have been proposed. For example, the method by Khabaza [18] uses the vectors $r_0, Ar_0, \ldots, A^{n-1}r_0$; Orthomin [30], Orthodir [35], Generalized Conjugate Gradient (GCG) [1, 2] and Generalized Conjugate Residual (GCR) [10, 11] compute an A^TA-orthogonal basis of $\mathcal{K}_n(A, r_0)$. These methods played an important role in the development of the field. In comparison to the approaches discussed in this paper they are, however, less numerically stable. Therefore we will not consider them below.

4.1. Minimal residual principle: Classical GMRES. Consider an initial approximation x_0 and the initial residual $r_0 = b - Ax_0,\ \rho_0 \equiv \|r_0\|$. In their classical paper [26], Saad and Schultz used the orthonormal basis (3.1) (Arnoldi basis). As noted in [33], this basis can be mathematically expressed as the Q-factor of a recursive columnwise QR-factorization

$$[r_0, AV_n] = V_{n+1}\,[e_1\rho_0, H_{n+1,n}], \qquad V_{n+1} \equiv [v_1, \ldots, v_{n+1}], \qquad V_{n+1}^T V_{n+1} = I. \tag{4.1}$$

Here $H_{n+1,n}$ is an $(n+1)$-by-n upper Hessenberg matrix with elements $h_{i,j}$, $h_{j+1,j} \neq 0, j = 1, 2, \ldots, n-1$. If at any stage $h_{n+1,n} = 0$, the algorithm would stop with $[r_0, AV_n] = V_n\,[e_1\rho_0, H_{n,n}]$. Using the substitution

$$x_n \;=\; x_0 \;+\; V_n\, y_n \tag{4.2}$$

and (4.1), the MR principle (1.3) gives the LS problem for the vector of coefficients y_n:

$$\|r_n\| \equiv \|b - Ax_n\| = \min_{y\in\mathcal{R}^n} \|r_0 - AV_n\, y\| \;=\; \min_{y\in\mathcal{R}^n} \|V_{n+1}\,(e_1\rho_0 - H_{n+1,n}\, y)\| \tag{4.3}$$

$$= \min_{y\in\mathcal{R}^n} \|e_1\rho_0 - H_{n+1,n}\, y\|. \tag{4.4}$$

To solve (4.3) we apply orthogonal rotations $J_1, J_2, \ldots, J_n$ sequentially to $H_{n+1,n}$ to bring it to the upper triangular form T_n:

$$J_n \cdots J_2 J_1\, H_{n+1,n} \;=\; \begin{bmatrix} T_n \\ 0 \end{bmatrix}.$$

The vectors y_n and $\|r_n\|$ then satisfy

$$\begin{bmatrix} T_n\, y_n \\ \|r_n\| \end{bmatrix} \;=\; J_1^T J_2^T \cdots J_n^T\, e_1 \rho_0. \tag{4.5}$$

The value of the (nonincreasing) residual norm is available without determining y_n, and it can be easily updated at each iteration, while y_{n+1} and x_{n+1} will usually differ in every element from y_n and x_n, respectively. We refer to this algorithm as Classical GMRES.

Several variants for computing the basis vectors $v_1, \ldots, v_n$ were proposed. Saad and Schultz [26] considered the modified Gram–Schmidt process. Walker [32, 33] presented Classical GMRES based on Householder transformations. Iterated classical and iterated modified Gram–Schmidt versions were studied in [9].

A variety of parallel implementations [6, 3, 12, 23, 7, 27] use various techniques to increase the parallel efficiency of the basically sequential basis-generating process. Parallel aspects are out of the scope of this paper.

4.2. Orthogonal projection principle: Simpler GMRES. We now consider an implementation of an MR method derived from the orthogonal projection principle (1.4). The approach proposed by Walker and Zhou [34], called Simpler GMRES, uses the orthonormal basis (3.2).

This basis is computed by a recursive columnwise QR-factorization of the matrix $[Ar_0\gamma, AW_{n-1}]$. Based on Theorem 3.1 we set $\gamma = \rho_0^{-1}$, and we will use this value of the scaling parameter γ throughout the rest of this paper. Then

$$A[v_1, W_{n-1}] = [Av_1, AW_{n-1}] = W_n S_n, \;\; W_n \equiv [w_1, \ldots, w_n], \;\; W_n^T W_n = I, \tag{4.6}$$

where S_n is an n-by-n upper triangular matrix with elements $s_{i,j}$, $s_{j,j} \neq 0$. If at any stage $s_{n,n} = 0$, the algorithm would stop with $[Av_1, AW_{n-1}] = W_{n-1}[S_{n-1}, \hat{s}_n]$. Using the substitution

$$x_n \;=\; x_0 \,+\, [v_1, W_{n-1}]\, t_n\,, \tag{4.7}$$

the vector $t_n \in \mathcal{R}^n$ solves the LS problem

$$\|r_n\| \equiv \|b - Ax_n\| = \min_{t \in \mathcal{R}^n} \|r_0 - A[v_1, W_{n-1}]\, t\| \tag{4.8}$$

$$= \min_{t \in \mathcal{R}^n} \|r_0 - W_n S_n\, t\|\,. \tag{4.9}$$

Solving the LS problem (4.8)–(4.9) in a numerically stable way represents a more subtle task then solving (4.3)–(4.4). The main difference is in handling the right-hand side vector r_0. In (4.3)–(4.4), r_0 is expressed in terms of the vectors $v_1, v_2, \ldots, v_n$ simply as $r_0 = v_1\rho_0$. In finite precision arithmetic, until the linear independence of the vectors $v_1, v_2, \ldots, v_n$ is lost, this expression is maximally accurate. On the other hand, application of the orthogonal projection principle (1.4) directly to (4.8)–(4.9) gives the upper triangular system

$$S_n\, t_n \;=\; W_n^T r_0. \tag{4.10}$$

As demonstrated in [25], computing the vector of coefficients t_n from (4.10) leads to numerical difficulties. Numerically more stable implementations are described next.

First consider the implementation of Simpler GMRES using the modified Gram–Schmidt process for generating the basis vectors $w_1, \ldots, w_n$. A properly implemented

algorithm for solving the LS problem (4.8)–(4.9) applies the orthogonalization process also to the right-hand side r_0 (see [4, pp. 64–65]). Then, using the recursive columnwise modified Gram–Schmidt QR-factorization of the extended matrix $[Av_1, AW_{n-1}, r_0]$,

$$[Av_1, AW_{n-1}, r_0] \;=\; W_n\,[S_n, z_n] \;+\; \left[0, \frac{r_n}{\|r_n\|}\right] \begin{bmatrix} 0 \\ \|r_n\| \end{bmatrix}, \tag{4.11}$$

the vector t_n solves the upper triangular system

$$S_n\, t_n \;=\; z_n. \tag{4.12}$$

The jth component of $z_n \equiv (\zeta_1, \ldots, \zeta_n)^T$ is determined by

$$\zeta_j \;=\; w_j^T\,(I - w_{j-1}w_{j-1}^T)\cdots(I - w_1 w_1^T)\,r_0 \;=\; w_j^T\,r_{j-1}\,, \tag{4.13}$$

where we use

$$r_j \;=\; (I - w_j w_j^T)\cdots(I - w_1 w_1^T)\,r_0 \;=\; r_{j-1} \;-\; (w_j^T\,r_{j-1})\,w_{j-1}\,. \tag{4.14}$$

A complete algorithm of the modified Gram–Schmidt implementation of Simpler GMRES is given in the appendix.

Now we consider the implementation of Simpler GMRES based on Householder reflections. It transforms the matrix $[Av_1, AW_{n-1}]$ into upper triangular form,

$$P_n \cdots P_2 P_1\,[Av_1, AW_{n-1}] \;=\; \begin{bmatrix} S_n \\ 0 \end{bmatrix}, \tag{4.15}$$

where P_j, $j = 1, \ldots, n$, are elementary Householder matrices. (For details, see [9, p. 312].) Then the transformed right-hand side is determined as

$$z_n \;=\; [P_n \cdots P_1\, r_0]_{\{1:n\}},$$

where $[\cdot]_{\{1:n\}}$ denotes the first n elements of a vector. The vector of coefficients t_n is determined from (4.12). A complete algorithm of the Householder implementation of Simpler GMRES is given in the appendix.

Related to Simpler GMRES are stabilized Orthodir [31] and the recent A^TA-variant of GMRES [25]. Both compute an A^TA-orthogonal basis of $\mathcal{K}_n(A, r_0)$, and thus each step of these methods requires about twice as much storage and also slightly more arithmetic operations than Simpler GMRES. They are also numerically less stable than Simpler GMRES. On the other hand, they allow easier parallel implementations because they feature step by step updates of both the approximate solution and the residual vectors.

5. Numerical stability. In this section we analyze and compare the numerical stability of Classical and Simpler GMRES. As mentioned in section 4, different orthogonalization techniques for computing the columns of V_n or W_n can be applied. Here we focus on implementations based on Householder transformations [32, 33] and on the modified Gram–Schmidt process [26].

For distinction, we denote quantities computed in finite precision arithmetic (with the machine precision ε) by bars. We assume the standard model of floating point arithmetic (see, e.g., [15, equation (2.4)]). In our bounds we present only those terms which are linear in ε and do not account for the terms proportional to higher powers

of ε. By $p_k(n,m,N)$, $k=1,2,\ldots$, we denote low degree polynomials in the number of iteration steps n, the maximum number of nonzeros per row in the system matrix m, and the dimension of the system N. They are introduced in a number of places in the text; some of them depend only on one or two variables. In all cases, $p_k(n,m,N) \leq c_k N^{5/2}$, where $c_k > 0$ is a constant independent of n, m, and N. This bound is, in general, very pessimistic; it accounts for the worst possible case. For details, see [9, 14, 24].

5.1. Classical GMRES. In the Classical GMRES implementation the computed approximate solution $\bar{x}_n$ satisfies

$$\begin{aligned} \bar{x}_n &= x_0 + \bar{V}_n \bar{y}_n + g_n, \\ \|g_n\| &\leq \varepsilon \|x_0\| + p_1(n)\, \varepsilon \|\bar{V}_n\| \|\bar{y}_n\|. \end{aligned} \tag{5.1}$$

It was shown in [9, 14] that the computed matrix $\bar{V}_n = [\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n]$ satisfies the recurrence

$$\begin{aligned} [\bar{r}_0, A\bar{V}_n] &= \hat{V}_{n+1} \left[\bar{\rho}_0 e_1, \bar{H}_{n+1,n}\right] + [f_n, F_n], \\ \|f_n\| &\leq p_2(m,N)\, \varepsilon \|A\| \|x_0\| + p_3(N)\, \varepsilon \|b\|, \\ \|F_n\| &\leq p_4(n,m,N)\, \varepsilon \|A\| \|\bar{V}_n\|, \end{aligned} \tag{5.2}$$

where the matrix $\hat{V}_{n+1}$ has exactly orthogonal columns ($\hat{V}_{n+1}^T \hat{V}_{n+1} = I_{n+1}$). The vector $\bar{y}_n$ is a computed solution of the finite precision analogue of the transformed LS problem (4.4), and $\bar{r}_0$ satisfies

$$\|\bar{r}_0 - (b - Ax_0)\| \leq p_5(m,N)\, \varepsilon \|A\| \|x_0\| + p_6(N)\, \varepsilon \|b\|\,. \tag{5.3}$$

For details, we refer to [9] and also to [24, pp. 25–26].

Our goal is not to give a complete rounding error analysis of GMRES. (For the Householder implementation of Classical GMRES this was published in [9], and the modified Gram–Schmidt implementation of Classical GMRES has been analyzed in [14, 24, 21].) We wish to explain that there is a potential weakness of Simpler GMRES which may negatively affect its computational behavior in comparison with Classical GMRES. For this purpose we can simplify our description of the GMRES convergence. This allows us to avoid tedious details which would make reading of this section difficult. We will describe the convergence of Classical GMRES by the norm of the LS residual associated with the matrix $A\bar{V}_n$ and the computed initial residual $\bar{r}_0$:

$$\|\hat{r}_n\| \equiv \|\bar{r}_0 - A\bar{V}_n \hat{y}_n\| = \min_y \|\bar{r}_0 - A\bar{V}_n y\|. \tag{5.4}$$

The analysis in [14, section 3] as well as numerical experiments confirm that for Classical GMRES $\|\hat{r}_n\|$ is close to the norm of the actually computed GMRES residual $\|b - A\bar{x}_n\|$.

It follows immediately from (2.14) that the residual norm (5.4) can be bounded in terms of the minimal singular values of matrices $[\bar{r}_0, A\bar{V}_n]$ and $A\bar{V}_n$ as

$$\sigma_{min}([\bar{r}_0, A\bar{V}_n]) \leq \|\hat{r}_n\| \leq \sigma_{min}([\bar{r}_0, A\bar{V}_n]) \left\{ 1 - \frac{\sigma_{min}^2([\bar{r}_0, A\bar{V}_n])}{\sigma_{min}^2(A\bar{V}_n)} + \frac{\|\bar{r}_0\|^2}{\sigma_{min}^2(A\bar{V}_n)} \right\}^{1/2}. \tag{5.5}$$

We see that convergence of the residual $\hat{r}_n$ is closely related to ill-conditioning of the matrix $[\bar{r}_0, A\bar{V}_n]$; i.e., decreasing $\|\hat{r}_n\|$ leads to ill-conditioning of $[\bar{r}_0, A\bar{V}_n]$. Moreover, it follows from (5.2) and from classical perturbation theory (see, e.g., [13, p. 449]), that the minimum singular values of the matrices $[\bar{r}_0, A\bar{V}_n]$ and $[\bar{\rho}_0 e_1, \bar{H}_{n+1,n}]$ are close to each other:

$$\left|\sigma_{min}([\bar{\rho}_0 e_1, \bar{H}_{n+1,n}]) - \sigma_{min}([\bar{r}_0, A\bar{V}_n])\right| \leq \|[f_n, F_n]\|\,. \tag{5.6}$$

Consequently, decreasing $\|\hat{r}_n\|$ leads to ill-conditioning of the matrix $[\bar{\rho}_0 e_1, \bar{H}_{n+1,n}]$. The vector $\bar{y}_n$ from (5.1) is a computed solution of the LS problem

$$\min_y \|e_1 \bar{\rho}_0 - \bar{H}_{n+1,n} y\|. \tag{5.7}$$

Using (5.2), the extremal singular values of $\bar{H}_{n+1,n}$ can be bounded by

$$\|\bar{H}_{n+1,n}\| \leq \|A\bar{V}_n\| + \|F_n\| \leq \|A\|\|\bar{V}_n\| + \|F_n\|, \tag{5.8}$$

$$\sigma_{min}(\bar{H}_{n+1,n}) \geq \sigma_{min}(A\bar{V}_n) - \|F_n\| \geq \sigma_{min}(A)\,\sigma_{min}(\bar{V}_n) - \|F_n\|. \tag{5.9}$$

When $\|\hat{r}_n\|$ (and $\|b - A\bar{x}_n\|$) decreases, $\sigma_{min}([\bar{r}_0, A\bar{V}_n])$ and $\sigma_{min}([\bar{\rho}_0 e_1, \bar{H}_{n+1,n}])$ also decrease. However, while the columns of $\bar{V}_n$ (the Arnoldi vectors) keep their linear independence (while $\sigma_{min}(\bar{V}_n) \approx 1$), *the condition number of the computed upper Hessenberg matrix $\bar{H}_{n+1,n}$ is essentially bounded by the condition number of A.* Consequently, until the linear independence of the Arnoldi vectors begins to deteriorate, the solution $\bar{y}_n$ of the transformed LS problem and the GMRES solution $\bar{x}_n$ are affected by rounding errors in a minimal possible way. This distinguishes Classical GMRES from the other MR methods, in particular from Simpler GMRES. Finite precision analysis of the QR-factorization of the matrix $\bar{H}_{n+1,n}$ via Givens rotations and of forming the GMRES solution can be found in [9] or [24, equations (4.6)–(4.12)].

It is important to note that not the orthogonality but the linear independence of the columns of $\bar{V}_n$ (measured by its extremal singular values) plays a decisive role in the relations (5.8) and (5.9). If we use Householder reflections in the Arnoldi process, the loss of orthogonality among the computed columns of $\bar{V}_n$ and the extremal singular values of $\bar{V}_n$ are bounded independent of the system parameters

$$1 - p_7(n, N)\,\varepsilon \leq \sigma_n(\bar{V}_n) \leq \|\bar{V}_n\| \leq 1 + p_7(n, N)\,\varepsilon. \tag{5.10}$$

Moreover, it was shown in [9] that the Householder implementation of Classical GMRES is backward stable. Assuming that a conjecture similar to (5.10) holds, the same result can also be shown for the iterated classical or modified Gram–Schmidt implementations; see [9].

In practical computations, cheaper orthogonalization techniques like the modified Gram–Schmidt algorithm are used. It is well known that the orthogonality among the columns of $\bar{V}_n$ computed via the modified Gram–Schmidt process will gradually deteriorate. However, from [14, equation (1.7) and Corollary 2.4] it follows that

$$\|\hat{V}_n - \bar{V}_n\| \leq p_8(n, m, N)\,\varepsilon\kappa([\bar{v}_1, A\bar{V}_{n-1}]), \tag{5.11}$$

and the minimal singular value and the norm of $\bar{V}_n$ are bounded by

$$1 - \frac{p_9(n, m, N)\,\varepsilon\kappa(A)}{\|\hat{r}_{n-1}\|/\bar{\rho}_0} \leq \sigma_n(\bar{V}_n) \leq \|\bar{V}_n\| \leq 1 + \frac{p_9(n, m, N)\,\varepsilon\kappa(A)}{\|\hat{r}_{n-1}\|/\bar{\rho}_0}. \tag{5.12}$$

The columns of $\bar{V}_n$ will thus begin to lose their linear independence *only after* the relative residual norm is reduced close to the level $\varepsilon\kappa(A)$. Up to that point the modified Gram–Schmidt implementation of Classical GMRES behaves about as well as the Householder implementation.

It was shown in [20, 21] that there is a tight relation between the normwise backward error $\|b - Ax_n\|/(\|A\|\|x_n\| + \|b\|)$ associated with the approximate solution x_n and the condition number of the matrix $[r_0, AV_n]$. A finite precision analogy of this statement will prove normwise backward stability of the modified Gram–Schmidt implementation of Classical GMRES. A formal proof will be given elsewhere.

The results in [20, 21] are based on (2.16). We could have also used (2.16) instead of (2.14) in our derivation here, which would lead to tighter estimates. However, using (2.14) makes our derivation much simpler, and the results are fully sufficient for our purpose.

5.2. Simpler GMRES. In Simpler GMRES the approximate solution $\bar{x}_n$ computed in finite precision arithmetic satisfies

$$\begin{aligned} \bar{x}_n &= x_0 + [\bar{v}_1, \bar{W}_{n-1}]\,\bar{t}_n + g_n, \\ \|g_n\| &\le \varepsilon\|x_0\| + p_1(n)\,\varepsilon\|[\bar{v}_1, \bar{W}_{n-1}]\|\,\|\bar{t}_n\|. \end{aligned} \tag{5.13}$$

Analogously to (5.2), for every iteration step n there exists a matrix $\hat{W}_n$ with exactly orthonormal columns ($\hat{W}_n^T\hat{W}_n = I$) such that

$$\begin{aligned} A[\bar{v}_1, \bar{W}_{n-1}] &= \hat{W}_n\bar{S}_n + F_n, \\ \|F_n\| &\le p_4(n, m, N)\,\varepsilon\|A\|\|[\bar{v}_1, \bar{W}_{n-1}]\|. \end{aligned} \tag{5.14}$$

The vector of coefficients $\bar{t}_n$ is computed by solving the upper triangular system with the matrix $\bar{S}_n$. From (5.14) the extremal singular values of the matrix $\bar{S}_n$ are bounded by

$$\|\bar{S}_n\| \le \|A[\bar{v}_1, \bar{W}_{n-1}]\| + \|F_n\| \le \|A\|\|[\bar{v}_1, \bar{W}_{n-1}]\| + \|F_n\|, \tag{5.15}$$

$$\begin{aligned} \sigma_{min}(\bar{S}_n) &\ge \sigma_{min}(A[\bar{v}_1, \bar{W}_{n-1}]) - \|F_n\| \\ &\ge \sigma_{min}(A)\,\sigma_{min}([\bar{v}_1, \bar{W}_{n-1}]) - \|F_n\|. \end{aligned} \tag{5.16}$$

The minimal singular value of the matrix $[\bar{v}_1, \bar{W}_{n-1}]$ can further be related to the minimal singular value of the matrix $[\bar{r}_0/\|\bar{r}_0\|, \hat{W}_{n-1}]$, where $\hat{W}_{n-1}$ comes from the recurrence (5.14),

$$\sigma_n([\bar{v}_1, \bar{W}_{n-1}]) \ge \sigma_n([\bar{r}_0/\|\bar{r}_0\|, \hat{W}_{n-1}]) - \|[\bar{v}_1 - \bar{r}_0/\|\bar{r}_0\|, \bar{W}_{n-1} - \hat{W}_{n-1}]\|. \tag{5.17}$$

For the condition number $\kappa([\bar{r}_0/\|\bar{r}_0\|, \hat{W}_{n-1}])$ it follows from (3.6) that

$$\kappa([\bar{r}_0/\|\bar{r}_0\|,\, \hat{W}_{n-1}]) = \frac{\|\bar{r}_0\| + \left(\|\bar{r}_0\|^2 - \|\hat{r}_{n-1}\|^2\right)^{1/2}}{\|\hat{r}_{n-1}\|}, \tag{5.18}$$

where $\hat{r}_{n-1} \equiv (I - \hat{W}_{n-1}\hat{W}_{n-1}^T)\,\bar{r}_0$ is the LS residual associated with the matrix $\hat{W}_{n-1}$, $\|\hat{r}_{n-1}\| = \min_y \|\bar{r}_0 - \hat{W}_{n-1}y\|$. The identity (5.18) proves that *convergence of the residual norm* $\|\hat{r}_{n-1}\|$ *and ill-conditioning of the matrix* $[\bar{r}_0/\|\bar{r}_0\|, \hat{W}_{n-1}]$ *are closely related.*

Summarizing, small $\|\bar{W}_{n-1} - \hat{W}_{n-1}\|$ means $\kappa([\bar{v}_1, \bar{W}_{n-1}]) \approx \kappa([\bar{r}_0/\|\bar{r}_0\|, \hat{W}_{n-1}])$. (It can be shown that $\|\bar{v}_1 - \bar{r}_0/\|\bar{r}_0\|\| \le (N+4)\varepsilon$; see [9].) Using (5.15) and (5.16), we

conclude that *decreasing* $\|\hat{r}_{n-1}\|$ *may lead to ill-conditioning of the upper triangular matrix* $\bar{S}_n$*, and thus to a potentially large error in computing the vector* $\bar{t}_n$*, independent of the (well-) conditioning of the matrix* A. This important fact may negatively affect the numerical accuracy of the approximate solution $\bar{x}_n$ in Simpler GMRES in comparison to Classical GMRES.

Until $\bar{S}_n$ becomes pathologically ill-conditioned, $\|\hat{r}_n\|$ is (similarly to subsection 5.1) close to $\|b - A\bar{x}_n\|$. After that the behavior of $\|\hat{r}_n\|$ and $\|b - A\bar{x}_n\|$ may be significantly different.

We have seen that the relation between the condition number of the matrix $\bar{S}_n$ and the condition number of the matrix $[\bar{r}_0/\|\bar{r}_0\|, \hat{W}_{n-1}]$ (the decrease of $\|\hat{r}_n\|$) is strongly affected by the size of the term $\|\bar{W}_{n-1} - \hat{W}_{n-1}\|$. In the Householder implementation the computed matrix $\bar{W}_{n-1}$ is, up to a small multiple of the machine precision, close to the matrix $\hat{W}_{n-1}$ with exactly orthogonal columns,

$$\|\bar{W}_{n-1} - \hat{W}_{n-1}\| \le p_7(n, N)\,\varepsilon. \tag{5.19}$$

It follows from (5.19) that the condition number $\kappa([\bar{v}_1, \bar{W}_{n-1}])$ is, up to terms proportional to the machine precision, equal to $\kappa([\bar{r}_0/\|\bar{r}_0\|, \hat{W}_{n-1}])$. In practice one frequently observes that after $\|b - A\bar{x}_n\|/\|\bar{r}_0\|$ reaches some particular point the norm of the computed vector $\bar{t}_n$ starts to increase dramatically (the computed results become irrelevant due to rounding errors), and the residual norm $\|b - A\bar{x}_n\|$ diverges.

For the modified Gram–Schmidt implementation we have

$$\|\bar{W}_{n-1} - \hat{W}_{n-1}\| \le p_8(n, m, N)\,\varepsilon\kappa(A[\bar{v}_1, \bar{W}_{n-1}]). \tag{5.20}$$

Because $\kappa(A[\bar{v}_1, \hat{W}_{n-1}])$ is potentially much worse than $\kappa([\bar{v}_1, A\hat{V}_{n-1}])$, the linear independence of the columns of $\bar{W}_n$ often begins to deteriorate much sooner than the linear independence of the columns of $\bar{V}_n$ in Classical GMRES. Until that point the modified Gram–Schmidt and Householder implementations of Simpler GMRES behave similarly. In subsequent iterations, surprisingly, the behavior of the modified Gram–Schmidt implementation of Simpler GMRES may be better than the behavior of the Householder implementation. For the Householder implementation of Simpler GMRES the true residual $b - A\bar{x}_n$ often diverges. This has been linked to the tight relation between $\kappa([\bar{r}_0/\|\bar{r}_0\|, \hat{W}_{n-1}])$ and $\kappa([\bar{v}_1, \bar{W}_{n-1}])$, and, consequently, to the relation between the decrease of $\|\hat{r}_n\|$ and the simultaneous increase of $\kappa(\bar{S}_n)$. For the modified Gram–Schmidt implementation, after reaching a certain point no such relations hold. The norm of $\bar{t}_n$ does not diverge, and the norm of the true residual remains (and often slightly oscillates) on or below the level corresponding to the turning point for the Householder implementation.

5.3. Numerical experiments. The different behavior of Classical and Simpler GMRES implementations is demonstrated by numerical examples with the matrix FS1836 from the Harwell–Boeing collection, $N = 183$, $\kappa(A) = 1.5 \times 10^{11}$, $\|A\| = 1.2 \times 10^9$. Experiments were performed using MATLAB 5.2, $\varepsilon = 1.1 \times 10^{-16}$. Householder and modified Gram–Schmidt orthogonalizations have been considered for both Classical and Simpler GMRES. In all experiments we used $x = (1, \ldots, 1)^T$, $b = Ax$, and $x_0 = 0$ ($\|\bar{r}_0\| = \|b\|$).

Figures 1 and 2 illustrate characteristics of the transformed LS problem (5.7) for the Householder and the modified Gram–Schmidt implementations of Classical GMRES. In both figures horizontal dotted lines represent $\|A\|$ and the minimal singular value $\sigma_{min}(A)$. The dashed lines show $\|\bar{H}_{n+1,n}\|$, the norm of the computed

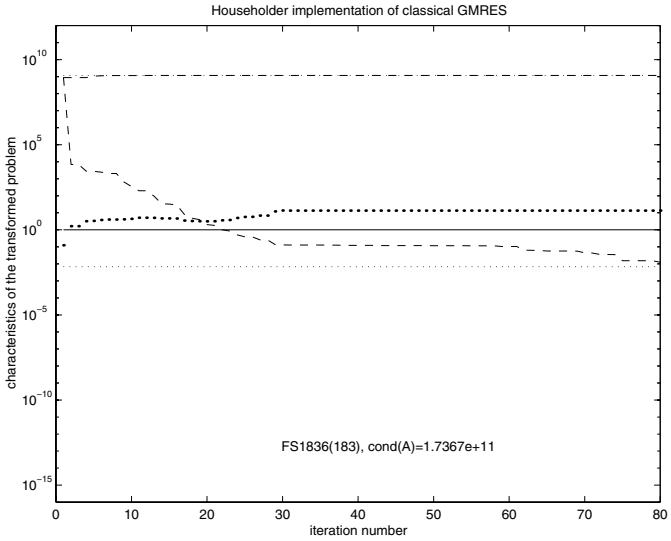

FIG. 1. *Householder implementation of Classical GMRES:* $\|A\|$ *and* $\sigma_{min}(A)$ *(dotted lines),* $\|\bar{H}_{n+1,n}\|$ *and* $\sigma_{min}(\bar{H}_{n+1,n})$ *(dashed lines),* $\sigma_{min}(\bar{V}_n)$ *(solid line), and* $\|\bar{y}_n\|$ *(dots).*

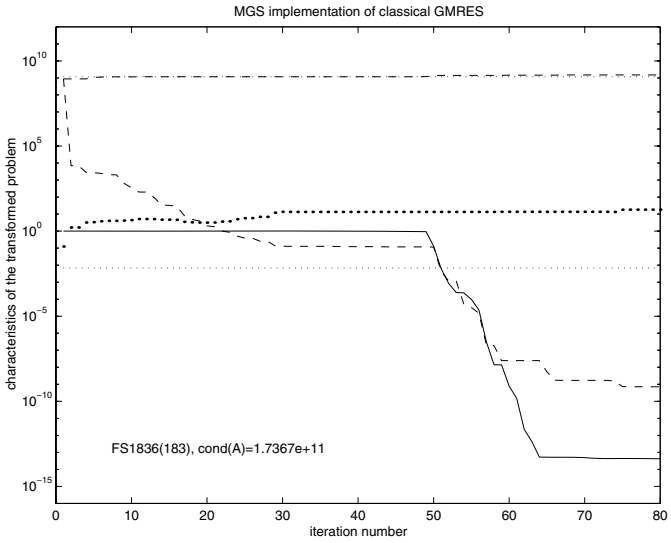

FIG. 2. *Modified Gram–Schmidt implementation of Classical GMRES:* $\|A\|$ *and* $\sigma_{min}(A)$ *(dotted lines),* $\|\bar{H}_{n+1,n}\|$ *and* $\sigma_{min}(\bar{H}_{n+1,n})$ *(dashed lines),* $\sigma_{min}(\bar{V}_n)$ *(solid line), and* $\|\bar{y}_n\|$ *(dots).*

upper Hessenberg matrix (it almost coincides with $\|A\|$), and the minimal singular value $\sigma_{min}(\bar{H}_{n+1,n})$. The solid line stands for $\sigma_{min}(\bar{V}_n)$, the minimal singular value of the matrix of computed Arnoldi vectors, and the dots depict $\|\bar{y}_n\|$, the norm of the computed solution vector of (5.7). We see that until the linear independence of the columns of $\bar{V}_n$ in the modified Gram–Schmidt implementation begins to deteriorate, Figures 1 and 2 are almost identical. There is no substantial growth in $\|\bar{y}_n\|$ even after the linear independence of the computed Arnoldi vectors is completely lost (cf. Figure 2).

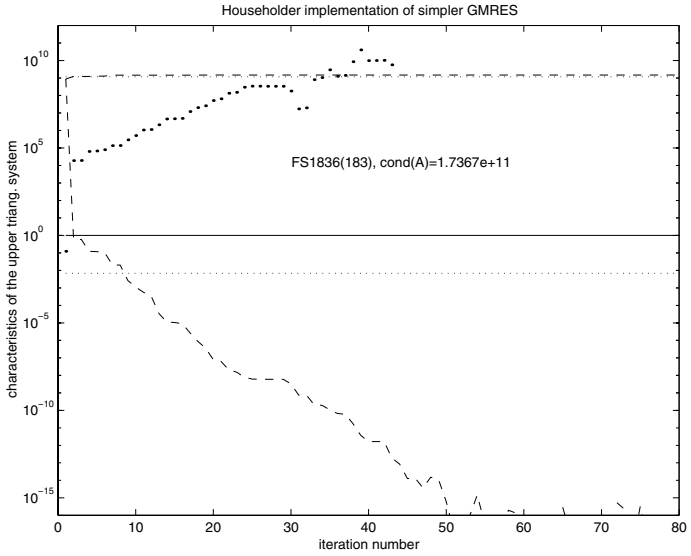

FIG. 3. *Householder implementation of Simpler GMRES:* $\|A\|$ *and* $\sigma_{min}(A)$ *(dotted lines),* $\|\bar{S}_n\|$ *and* $\sigma_{min}(\bar{S}_n)$ *(dashed lines),* $\sigma_{min}(\bar{W}_n)$ *(solid line), and* $\|\bar{t}_n\|$ *(dots).*

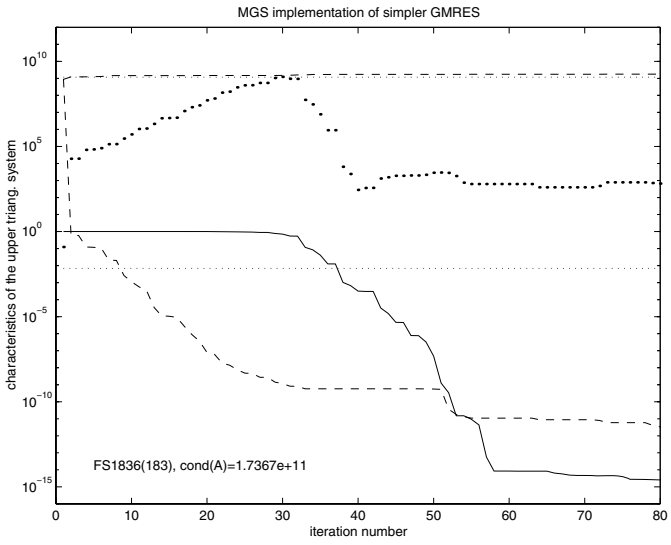

FIG. 4. *Modified Gram–Schmidt implementation of Simpler GMRES:* $\|A\|$ *and* $\sigma_{min}(A)$ *(dotted lines),* $\|\bar{S}_n\|$ *and* $\sigma_{min}(\bar{S}_n)$ *(dashed lines),* $\sigma_{min}(\bar{W}_n)$ *(solid line), and* $\|\bar{t}_n\|$ *(dots).*

Similar quantities are illustrated in Figures 3 and 4 for the Householder and the modified Gram–Schmidt implementations of Simpler GMRES. The dashed lines here represent $\|\bar{S}_n\|$, the norm of the computed upper triangular matrix, and its minimal singular value $\sigma_{min}(\bar{S}_n)$. The dots denote $\|\bar{t}_n\|$, the norm of the computed solution of the upper triangular system with the matrix $\bar{S}_n$.

We see that the condition number of the matrix $\bar{H}_{n+1,n}$ is in Figure 1 (the House-

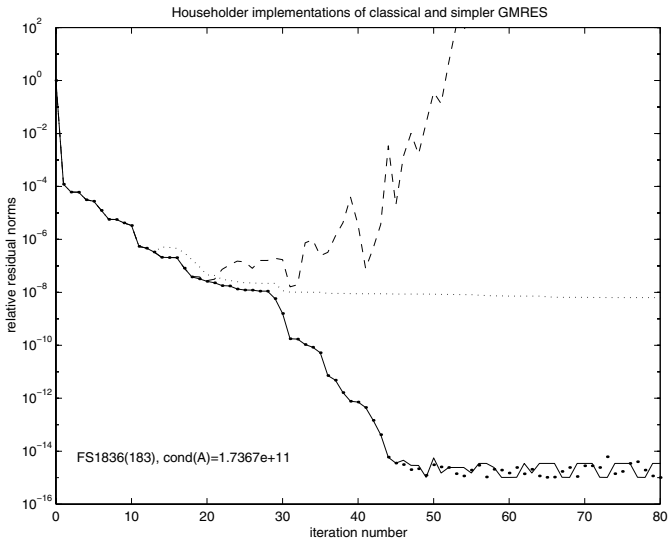

FIG. 5. *Householder implementation of Classical and Simpler GMRES: Normalized true residual norm* $\|b - A\bar{x}_n\|/\|b\|$ *(solid line—Classical GMRES, dashed line—Simpler GMRES), and* $\|\hat{r}_n\|/\|b\|$ *(dots—Classical GMRES, dotted line—Simpler GMRES).*

holder implementation of Classical GMRES) approximately bounded by the condition number of A, and for Figure 2 (the modified Gram–Schmidt implementation of Classical GMRES) the same is true until $\sigma_{min}(\bar{V}_n)$ begins to deteriorate. In contrast, in both implementations of Simpler GMRES, the minimal singular value of $\bar{S}_n$ decreases very soon far below $\sigma_{min}(A)$. Consequently, the accuracy of the computed vector $\bar{t}_n$ deteriorates, and for the Householder implementation $\|\bar{t}_n\|$ diverges. Also note the difference between $\sigma_{min}(\bar{V}_n)$ and $\sigma_{min}(\bar{W}_n)$ in Figures 2 and 4.

In Figure 5 we compare the convergence characteristics for the Householder implementations of both Classical GMRES ($\|b - A\bar{x}_n\|/\|b\|$ is represented by the solid line, $\|\hat{r}_n\|/\|b\|$ by dots) and Simpler GMRES ($\|b - A\bar{x}_n\|/\|b\|$ is represented by the dashed line, $\|\hat{r}_n\|/\|b\|$ by the dotted line). Figure 5 illustrates our theoretical considerations and shows that the true residual norm of the Householder implementation of Simpler GMRES may after some initial reduction diverge. Figure 6 uses similar notation for the illustration of the modified Gram–Schmidt implementations. The residual norm of Simpler GMRES again exhibits worse behavior than the residual norm corresponding to Classical GMRES.

6. Conclusions. MR methods can be formulated and implemented using different bases and different orthogonalization processes. Using general theoretical results about the LS residual, this paper shows that the choice of the basis is fundamental for getting revealing theoretical results about convergence of MR methods. It is also important for getting a numerically stable implementation. The choice of the computed basis may strongly affect the numerical behavior of the implementation. It is explained that using the best orthogonalization technique in building the basis does not compensate for the possible loss of accuracy in a given method which is related to the choice of the basis. In particular, it is shown that the classical implementation of GMRES, which is based on the Arnoldi process starting from the normalized initial residual (as proposed by Saad and Schultz), has numerical advantages over Simpler

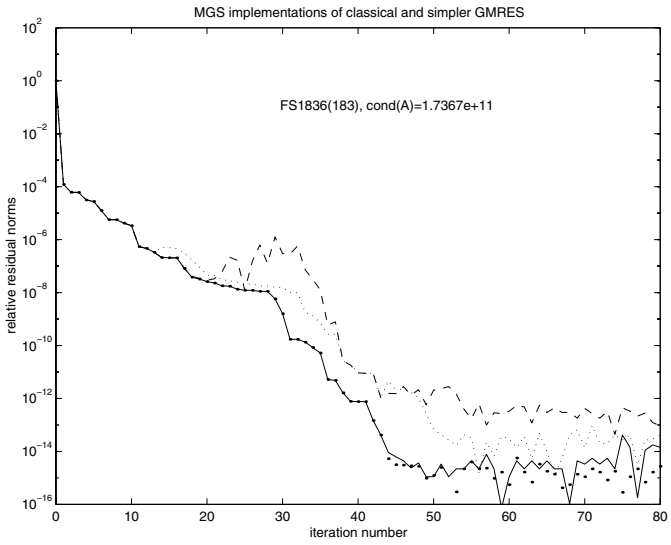

FIG. 6. *Modified Gram–Schmidt implementation of Classical and Simpler GMRES: Normalized true residual norm* $\|b - A\bar{x}_n\|/\|b\|$ *(solid line—Classical GMRES, dashed line—Simpler GMRES), and* $\|\hat{r}_n\|/\|b\|$ *(dots—Classical GMRES, dotted line—Simpler GMRES).*

GMRES, which is based on the shifted Arnoldi process.

7. Appendix. Here we present the implementations of Simpler GMRES used throughout the paper.

Modified Gram–Schmidt implementation of Simpler GMRES:

x_0, $r_0 = b - Ax_0$, $v_1 = r_0/\|r_0\|$, $w_0 = v_1$

$n = 1, 2, \ldots$

$w_n = Aw_{n-1}$

$j = 1, 2, \ldots, n-1$

$w_n \leftarrow w_n - \rho_{j,n} w_j, \;\; \rho_{j,n} = (w_n, w_j)$

$w_n \leftarrow w_n/\rho_{n,n}, \;\; \rho_{n,n} = \|w_n\|$

$$S_n = \begin{pmatrix} S_{n-1} & \rho_{1,n} \\ & \vdots \\ 0 & \rho_{n,n} \end{pmatrix}, \;\; S_1 = (\rho_{1,1})$$

$r_n = r_{n-1} - \zeta_n w_n, \;\; \zeta_n = (r_{n-1}, w_n)$

Solve $S_n\, t_n = (\zeta_1, \ldots, \zeta_n)^T$

$x_n = x_0 + [v_1, w_1, \ldots, w_{n-1}]\, t_n$

Householder implementation of Simpler GMRES:

$x_0,\ r_0 = b - Ax_0,\ v_1 = r_0/\|r_0\|,\ (\zeta_1, \ldots, \zeta_N)^T = r_0,\ w_0 = v_1$

$n = 1, 2, \ldots$

Compute P_n such that $P_n\, Aw_{n-1} = (\rho_{1,n}, \ldots, \rho_{n,n}, 0, \ldots, 0)^T$

$$S_n = \begin{pmatrix} S_{n-1} & \rho_{1,n} \\ & \vdots \\ 0 & \rho_{n,n} \end{pmatrix},\ S_1 = (\rho_{1,1})$$

$(\zeta_1, \ldots, \zeta_N)^T \leftarrow P_n\, (\zeta_1, \ldots, \zeta_N)$

Solve $\quad S_n\, t_n = (\zeta_1, \ldots, \zeta_n)^T$

$r_n = r_{n-1} - \zeta_n w_n$

$w_n = P_1 \ldots P_n\, e_n$

$x_n = x_0 + [v_1, w_1, \ldots, w_{n-1}]\, t_n$

Acknowledgments. The authors are indebted to Chris Paige for his comments which improved the formulation and interpretation of several results. They also wish to thank the referees for their comments which improved presentation of the results.

REFERENCES

[1] O. Axelsson, *Conjugate gradient type methods for unsymmetric and inconsistent systems of linear equations*, Linear Algebra Appl., 29 (1980), pp. 1–16.
[2] O. Axelsson, *A generalized conjugate gradient, least square method*, Numer. Math., 51 (1987), pp. 209–227.
[3] Z. Bai, D. Hu, and L. Reichel, *A Newton basis GMRES implementation*, IMA J. Numer. Anal., 14 (1994), pp. 563–581.
[4] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, 1996.
[5] S. L. Campbell and C. D. Meyer, Jr., *Generalized Inverses of Linear Transformations*, Pitman, Boston, MA, 1979.
[6] E. de Sturler, *A parallel variant of GMRES(m)*, in Proceedings of the 13th World Congress on Computation and Applied Mathematics, Dublin, Ireland, 1991, pp. 682–683.
[7] E. de Sturler and H. van der Vorst, *Reducing the effect of global communication in GMRES(m) and CG on parallel distributed memory computers*, Appl. Numer. Math., 18 (1995), pp. 441–459.
[8] J. Demmel, *The condition number of equivalence transformations that block diagonalize matrix pencils*, SIAM J. Numer. Anal., 20 (1983), pp. 599–610.
[9] J. Drkošová, A. Greenbaum, M. Rozložník, and Z. Strakoš, *Numerical stability of GMRES*, BIT, 35 (1995), pp. 309–330.
[10] 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.
[11] H. Elman, *Iterative Methods for Large Sparse Nonsymmetric Systems of Linear Equations*, Ph.D. thesis, Yale University, New Haven, CT, 1982.
[12] J. Erhel, *A parallel GMRES version for general sparse matrices*, Electron. Trans. Numer. Anal., 3 (1995), pp. 160–176.
[13] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[14] A. Greenbaum, M. Rozložník, and Z. Strakoš, *Numerical behaviour of the modified Gram-Schmidt GMRES implementation*, BIT, 37 (1997), pp. 706–719.
[15] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, 1996.

[16] I. Ipsen, *A Different Approach to Bounding the Minimal Residual Norm in Krylov Methods*, Technical report, CRSC, Department of Mathematics, North Carolina State University, Raleigh, NC, 1998.
[17] I. C. F. Ipsen, *Expressions and bounds for the GMRES residual*, BIT, 40 (2000), pp. 524–535.
[18] I. M. Khabaza, *An iterative least-square method suitable for solving large sparse matrices*, Comput. J., 6 (1963/1964), pp. 202–206.
[19] J. Liesen, *Construction and Analysis of Polynomial Iterative Methods for Non-Hermitian Systems of Linear Equations*, Ph.D. thesis, Fakultät für Mathematik, Universität Bielefeld, Bielefeld, Germany, 1998; also available online from http://archiv.ub.uni-bielefeld.de/disshabi/mathe.htm.
[20] C. Paige and Z. Strakoš, *Bounds for the least squares distance via scaled total least squares problems*, Numer. Math., to appear.
[21] C. Paige and Z. Strakoš, *Residual and backward error bounds in minimum residual Krylov subspace methods*, SIAM J. Sci. Comput., to appear.
[22] C. Paige and Z. Strakoš, *Scaled total least squares fundamentals*, Numer. Math., to appear.
[23] B. Philippe and R. Sidje, *Parallel algorithms for the Arnoldi procedure*, in Iterative Methods in Linear Algebra, II, IMACS Ser. Comput. Appl. Math. 3, IMACS, New Brunswick, NJ, 1995, pp. 156–165.
[24] M. Rozložník, *Numerical Stability of the GMRES Method*, Ph.D. thesis, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, 1997; also available online from http://www.cs.cas.cz/~miro.
[25] M. Rozložník and Z. Strakoš, *Variants of the residual minimizing Krylov space methods*, in Proceedings of the 11th Summer School on Software and Algorithms of Numerical Mathematics, I. Marek, ed., 1995, pp. 208–225.
[26] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
[27] R. B. Sidje, *Alternatives for parallel Krylov subspace basis computation*, Numer. Linear Algebra Appl., 4 (1997), pp. 305–331.
[28] G. W. Stewart, *Collinearity and least squares regression*, Statist. Sci., 2 (1987), pp. 68–100.
[29] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia, PA, 1991.
[30] P. Vinsome, *Orthomin, an iterative method for solving sparse sets of simultaneous linear equations*, in Proceedings of the Fourth Symposium on Reservoir Simulation, Society of Petroleum Engineers of AIME, 1976, pp. 149–159.
[31] C. Vuik and H. A. van der Vorst, *A comparison of some GMRES-like methods*, Linear Algebra Appl., 160 (1992), pp. 131–162.
[32] H. F. Walker, *Implementation of the GMRES method using Householder transformations*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 152–163.
[33] H. F. Walker, *Implementations of the GMRES method*, Comput. Phys. Comm., 53 (1989), pp. 311–320.
[34] H. F. Walker and L. Zhou, *A simpler GMRES*, Numer. Linear Algebra Appl., 1 (1994), pp. 571–581.
[35] D. M. Young and K. C. Jea, *Generalized conjugate-gradient acceleration of nonsymmetrizable iterative methods*, Linear Algebra Appl., 34 (1980), pp. 159–194.