BY HOW MUCH CAN RESIDUAL MINIMIZATION ACCELERATE THE CONVERGENCE OF ORTHOGONAL RESIDUAL METHODS?

MARTIN H. GUTKNECHT* AND MIROSLAV ROZLOŽNÍK* †

Version of May 16, 2001.

Abstract. We capitalize upon the known relationship between pairs of orthogonal and minimal residual methods (or, biorthogonal and quasi-minimal residual methods) in order to estimate how much smaller the residuals or quasi-residuals of the minimizing methods can be compared to the those of the corresponding Galerkin or Petrov-Galerkin method. Examples of such pairs are the conjugate gradient (CG) and the conjugate residual (CR) methods, the full orthogonalization method (FOM) and the generalized minimal residual (GMRes) method, the CGNE and CGNR versions of applying CG to the normal equations, as well as the biconjugate gradient (BiCG) and the quasi-minimal residual (QMR) methods. Also the pairs consisting of the (bi)conjugate gradient squared (CGS) and the transpose-free QMR (TFQMR) methods can be added to this list if the residuals at half-steps are included, and further examples can be created easily.

The analysis is more generally applicable to the minimal residual (MR) and quasi-minimal residual (QMR) smoothing processes, which are known to provide the transition from the results of the first method of such a pair to those of the second one. By an interpretation of these smoothing processes in coordinate space we deepen the understanding of some of the underlying relationships and introduce a unifying framework for minimal residual and quasi-minimal residual smoothing. This framework includes the general notion of QMR-type methods.

Key words. system of linear algebraic equations, iterative method, Krylov space method, conjugate gradient method, biconjugate gradient method, CG, CGNE, CGNR, CGS, FOM, GMRes, QMR, TFQMR, residual smoothing, MR smoothing, QMR smoothing

AMS subject classifications. 65F10

1. Introduction. The convergence of an iterative method for solving linear systems $\mathbf{Ax} = \mathbf{b}$ is normally monitored by checking the 2-norm of the *residual vectors* $\mathbf{r}_n :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_n$, because the error vectors $\mathbf{x}_n - \mathbf{x}$ (where $\mathbf{x}$ denotes the exact solution) are not available. Therefore, a method sells well, if the *residual norm history* shows a quick decline, although this does not necessarily mean that the error norms also decline as quickly. We will restrict ourselves here to Krylov space solvers, where, by definition, $\mathbf{x}_n$ is of the form

$$\mathbf{x}_n - \mathbf{x}_0 \in \mathcal{K}_n :\equiv \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) :\equiv \mathrm{span}\,(\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{n-1}\mathbf{r}_0) \tag{1.1}$$

and, hence,

$$\mathbf{r}_n - \mathbf{r}_0 \in \mathbf{A}\mathcal{K}_n = \mathrm{span}\,(\mathbf{A}\mathbf{r}_0, \mathbf{A}^2\mathbf{r}_0, \ldots, \mathbf{A}^n\mathbf{r}_0). \tag{1.2}$$

Methods of this class that are optimal in the sense of producing residuals of minimal norm are the conjugate residual (CR) method for Hermitian positive definite (Hpd) matrices [23], the minimal residual (MinRes) algorithm of Paige and Saunders [17], which applies also to indefinite Hermitian matrices, and the generalized minimal residual (GMRes) algorithm of Saad and Schultz [19] for arbitrary nonsingular square matrices. Another, mathematically nearly equivalent form of the latter method is the generalized conjugate residual (GCR) algorithm; in exact arithmetic it produces the same iterates unless it breaks down.

*Seminar for Applied Mathematics, Swiss Federal Institute of Technology (ETH) Zurich, ETH-Zentrum, CH-8092 Zurich, Switzerland. (mhg@sam.math.ethz.ch, miro@sam.math.ethz.ch)

†Part of this author's work was supported by the Grant Agency of the Czech Republic under grant No. 201/98/P108.

These residual minimizing methods are in competition with closely related Krylov space solvers that satisfy a Galerkin condition and feature orthogonal residuals, namely the conjugate gradient (CG) method of Hestenes and Stiefel [15] for Hpd matrices and the full orthogonalization method (FOM) (also called Arnoldi method) for arbitrary nonsingular square matrices, whose iterates $\mathbf{x}_n$ may not be defined for certain n, however. The SymmLQ algorithm [17] for symmetric indefinite matrices also generates orthogonal residuals, but the search space is different from the Krylov space in (1.1); hence, SymmLQ does not fit into our framework. However, SymmLQ allows to easily construct a second sequence of approximate solutions, which define an orthogonal residual method of the type considered here and is mathematically equivalent to FOM when the system is symmetric.

A further option for non-Hermitian systems is the biconjugate gradient (BiCG) method of Lanczos [16] and Fletcher [6]. The related quasi-minimal residual (QMR) method of Freund and Nachtigal [10] modifies BiCG in the same way as MinRes is obtained from CG: minimization is enforced on the coordinate vector of the residuals with respect to the normalized Lanczos basis. (Additionally, in the original version of QMR, the look-ahead Lanczos process is applied for generating this basis in order to overcome the breakdowns of the Lanczos algorithm [12, 14, 9].) Since the Lanczos basis is not orthogonal, the norm of the QMR residuals is, in general, not minimal. However, the norm of the coordinate vectors still is. These are often referred to as *quasi-residuals*.

Most of the aforementioned methods can be implemented in various ways. But since finite precision effects are not considered here, the details of the implementation do not matter. That is why we refer here to 'methods' and not to 'algorithms'. For example, our results on CR residuals also hold for MinRes residuals, and those for GMRes residuals are valid for GCR residuals.

The transition from CG to CR, and the one from FOM to GMRes, can be simulated by a smoothing process proposed by Schönauer [21], which is called *minimal residual (MR) smoothing* [29]. For FOM and GMRes (and thus, *a fortiori* for CG and CR) this was shown by Weiss [26, 27]. In the Appendix we establish such a result also for the transition from CGNE to CGNR, the two well-known versions of applying CG to the normal equations. This result is mentioned, but not explicitly proven in Weiss [28]. The pair consisting of the aforementioned second set of approximations from SymmLQ and the results of MinRes could also be added to the list.

For BiCG and QMR the process has to be adapted suitably, as shown by Zhou and Walker [29, 25]. They call this variation *quasi-minimal residual (QMR) smoothing*. It can be understood as applying the minimal residual smoothing process to the coordinates of the BiCG residuals with respect to the Lanczos basis, that is, to the quasi-residuals. As pointed out by Zhou and Walker, QMR smoothing also mimics the transition from the (bi)conjugate gradient squared (CGS) method of Sonneveld [22] to the transpose-free QMR (TFQMR) method of Freund [7] if iterates and residuals at half-steps are included. And similarly, one obtains the QMRCGStab method of Chan et al. [2] from Van der Vorst's BiCGStab [24]. TFQMR and QMRCGStab are readily seen to be examples for the class of QMR-type methods described in Section 3. These have generally the property that their iterates and residuals can be found by QMR smoothing. The MR and QMR smoothing processes can be applied to any Krylov space solver, and they can therefore serve as a general framework for the above mentioned transitions. We will review the main results about these smoothing process in Sections 2 and 3 and give a simple account of the basic relationships.

The interest in these smoothing processes arose because the BiCG and CGS methods often show an erratic convergence behavior: the residual norm fluctuates heavily. In contrast, after MR smoothing the residuals decline monotonically, and QMR smoothing normally produces a nearly monotonic residual norm history. The QMR method generates approximations $\widetilde{\mathbf{x}}_n$ that coincide (in exact arithmetic) with those obtained by QMR smoothing from the BiCG iterates $\mathbf{x}_n$. Therefore, its residuals also decline nearly monotonously.

The fundamental question to be posed is whether this transition from an orthogonal residual method (or, more generally, from any *primary* iterative method) to the corresponding minimal residual method (or, in general, to the corresponding *smoothed* method) enables us to find the solution of $\mathbf{Ax} = \mathbf{b}$ faster. If we could measure convergence in terms of the 2-norms of the errors $\mathbf{x}_n - \mathbf{x}$, the answer would be hard; if we measure it as usual in terms of the 2-norm of the residuals, we will clearly gain something, but it is not so clear how much.

A qualitative answer was given by Cullum and Greenbaum [3, 4] by capitalizing upon a known relationship between the residual norms of the primary method and the residual or quasi-residual norms of the smoothed method. In particular, this relationship enabled them to explain the so-called peak-plateau connection: at places where the residual norm history of FOM or BiCG has a peak, the smoothed methods, GMRes and QMR, exhibit a plateau where the residual norm remains more or less constant. In Section 4 we will use essentially the same relation to derive upper and lower bounds for the norms of the smoothed residuals or quasi-residuals.

The aforementioned relationship between the residual norms of the primary and the smoothed methods was established in the thesis of Weiss [26, 27]. Independently, Brown [1] proved it around the same time for FOM and GMRes. It also appeared in Zhou and Walker's common framework for residual smoothing methods [29, 25], and, as mentioned above, in Cullum and Greenbaum [3, 4]. It must be pointed out, however, that another relationship from which the above mentioned follows easily, appeared already in Paige and Saunder's derivation of MinRes [17] and, in a more general setting, in Freund's analysis of TFQMR [7]. A different interpretation of these relations in terms of singular values of upper Hessenberg matrices was given by Sadok [20]. Recently, Eiermann and Ernst [5] presented for the relationship between orthogonal and minimal residual methods a new, geometric framework , in which angles between subspaces play a crucial role. This abstract framework includes projection methods that are not Krylov space solvers, a generality we do not aim at here. Eiermann and Ernst make also use of the fact that methods based on oblique projections can be viewed as orthogonal or minimal residual methods with respect to a problem-dependent, only *a posteriori* available inner product — another aspect we do not touch. We will present in Section 2 a different, less general, but simpler geometric interpretation; its main point is that in each smoothing step, all the action occurs in a two-dimensional plane.

2. Orthogonal and minimal residual methods; MR smoothing. Given a real or complex nonsingular square linear system $\mathbf{Ax} = \mathbf{b}$, a *Krylov space solver* is an iterative method generating approximate solutions $\mathbf{x}_n$ and corresponding residuals $\mathbf{r}_n :\equiv \mathbf{b} - \mathbf{Ax}_n$ (or, at least, as in GMRes, the norm of the latter) so that (1.1) holds and $\mathbf{x}_n \to \mathbf{x}$ in a finite or infinite number of steps under certain assumptions on $\mathbf{A}$. A basis or, more generally, a generating set $\{\mathbf{y}_n\}_{n=0}^{m}$ for $\mathcal{K}_{m+1}$ can be built up by a recursion of the form

$$\mathbf{y}_{n+1} := (\mathbf{Ay}_n - \mathbf{y}_n\,\eta_{n,n} - \mathbf{y}_{n-1}\,\eta_{n-1,n} - \cdots - \mathbf{y}_0\,\eta_{0,n})\,/\eta_{n+1,n} \tag{2.1}$$

$(0 \le n < m)$, where $\mathbf{y}_0 :\equiv \mathbf{r}_0/\rho_0$ with suitable ρ_0, for example, $\rho_0 := \|\mathbf{r}_0\|$. If we let

$$\mathbf{Y}_m :\equiv \left[\ \mathbf{y}_0 \quad \mathbf{y}_1 \quad \cdots \quad \mathbf{y}_{m-1}\ \right],$$

then, for $n < m$, these recursions can be summarized in

$$\mathbf{A}\mathbf{Y}_m = \mathbf{Y}_{m+1}\underline{\mathbf{H}}_m\,, \tag{2.2}$$

where $\underline{\mathbf{H}}_m :\equiv \{\eta_{k,l}\}$ is an $(m+1) \times m$ extended, irreducible Hessenberg matrix. In particular, one can choose the coefficients in (2.1) so that an orthogonal or even orthonormal basis results, a version referred to as the *Arnoldi process*. In this section we assume that this choice has been made. Alternatives to the recursion (2.1) exist, but this is not of importance here since we are not concerned with roundoff errors.

We can always express $\mathbf{x}_n - \mathbf{x}_0$ and $\mathbf{r}_n - \mathbf{r}_0$ in terms of the generating set and a coefficient vector $\mathbf{k}_n$ according to

$$\mathbf{x}_n = \mathbf{x}_0 + \mathbf{Y}_n\mathbf{k}_n\,, \qquad \mathbf{r}_n = \mathbf{r}_0 - \mathbf{A}\mathbf{Y}_n\mathbf{k}_n\,. \tag{2.3}$$

Then global minimization of the 2-norm of the residual (under the condition (1.2)) means to minimize the following functional $\widetilde{\Phi}_n$ of $\mathbf{k}_n$:

$$\widetilde{\Phi}_n(\mathbf{k}_n) :\equiv \tfrac{1}{2}\|\mathbf{r}_n\|^2 = \tfrac{1}{2}\|\mathbf{r}_0 - \mathbf{A}\mathbf{Y}_n\mathbf{k}_n\|^2\,. \tag{2.4}$$

We will denote the resulting optimal nth iterate and residual by $\widetilde{\mathbf{x}}_n$ and $\widetilde{\mathbf{r}}_n$, respectively. In view of

$$\nabla\widetilde{\Phi}_n(\mathbf{k}_n) = \mathbf{Y}_n^\star\mathbf{A}^\star(\mathbf{r}_0 - \mathbf{A}\mathbf{Y}\mathbf{k}_n) = -\,\mathbf{Y}_n^\star\mathbf{A}^\star\mathbf{r}_n\,, \tag{2.5}$$

minimization of $\widetilde{\Phi}_n$ yields the Galerkin condition

$$\widetilde{\mathbf{r}}_n^\star\mathbf{A}\mathbf{Y}_n = \mathbf{o}^\star \qquad \text{or} \qquad \widetilde{\mathbf{r}}_n \perp \mathbf{A}\mathcal{K}_n\,, \tag{2.6}$$

which implies that the residuals must be $\mathbf{A}$-orthogonal or so-called *conjugate*. (The star denotes the Hermitian transpose, and $\mathbf{o}$ is the zero vector of appropriate size.) This situation applies to the CR and GMRES methods, though, of course, the algorithmic realizations of these methods are quite different for each one and from what might be indicated by the above formulas.

When $\mathbf{A}$ is Hermitian positive definite (Hpd), minimization of the $\mathbf{A}$-norm of the error according to

$$\Phi_n(\mathbf{k}_n) :\equiv \tfrac{1}{2}\|\mathbf{x}_n - \mathbf{x}\|_{\mathbf{A}}^2 = \tfrac{1}{2}\|\mathbf{x}_0 - \mathbf{x} + \mathbf{Y}_n\mathbf{k}_n\|_{\mathbf{A}}^2 \tag{2.7}$$

(where $\mathbf{x} :\equiv \mathbf{A}^{-1}\mathbf{b}$) yields in view of

$$\nabla\Phi_n(\mathbf{k}_n) = \mathbf{Y}_n^\star\mathbf{A}(\mathbf{x}_0 - \mathbf{x} + \mathbf{Y}_n\mathbf{k}_n) = \mathbf{Y}_n^\star\mathbf{A}(\mathbf{x}_n - \mathbf{x}) = -\,\mathbf{Y}_n^\star\mathbf{r}_n \tag{2.8}$$

likewise the Galerkin condition

$$\mathbf{r}_n^\star\mathbf{Y}_n = \mathbf{o}^\star \qquad \text{or} \qquad \mathbf{r}_n \perp \mathcal{K}_n\,, \tag{2.9}$$

which implies that the residuals need to be orthogonal. This is the setting of the classical CG method. Here, clearly, the minimization property does not generalize to matrices that are not Hpd, but for such matrices one may nevertheless use the

Galerkin condition (2.9), which implies the mutual orthogonality of the residuals, as the basis of a more generally applicable algorithm. For example, FOM is an algorithm that fits into this pattern, but like others it may break down.

We will refer to the methods based on (2.9) as *orthogonal residual methods* and to those based on (2.4) or (2.6) as *minimal residual methods*, although there is a danger of mixing up the latter general class with MinRes, which is a particular algorithm. While for the former $\widetilde{\mathbf{x}}_n$ is guaranteed to exist for all n, the approximation $\mathbf{x}_n$ (and thus also the residual $\mathbf{r}_n$) of an orthogonal residual method may not exist for some n.

According to (1.1), (1.2), (2.6), and (2.9) the orthogonal residual $\mathbf{r}_n$ (if it exists) and the minimizing residuals $\widetilde{\mathbf{r}}_n$ and $\widetilde{\mathbf{r}}_{n-1}$ satisfy the conditions

$$\mathbf{r}_n\,,\ \widetilde{\mathbf{r}}_n \in \mathbf{r}_0 + \mathbf{A}\mathcal{K}_n \subseteq \mathcal{K}_{n+1}\,, \qquad \widetilde{\mathbf{r}}_{n-1} \in \mathbf{r}_0 + \mathbf{A}\mathcal{K}_{n-1} \subseteq \mathcal{K}_n \subseteq \mathcal{K}_{n+1} \tag{2.10}$$

and

$$\mathbf{r}_n \perp \mathcal{K}_n \supseteq \mathbf{A}\mathcal{K}_{n-1}\,, \qquad \widetilde{\mathbf{r}}_n \perp \mathbf{A}\mathcal{K}_n\,, \qquad \widetilde{\mathbf{r}}_{n-1} \perp \mathbf{A}\mathcal{K}_{n-1}\,. \tag{2.11}$$

Consequently,

$$\mathbf{r}_n\,,\ \widetilde{\mathbf{r}}_n\,,\ \widetilde{\mathbf{r}}_{n-1} \in \mathcal{K}_{n+1} \ominus \mathbf{A}\mathcal{K}_{n-1} \tag{2.12}$$

lie in an at most two-dimensional subspace and are therefore linearly dependent. Since $\mathbf{r}_n - \mathbf{r}_0\,,\ \widetilde{\mathbf{r}}_n - \mathbf{r}_0\,,\ \widetilde{\mathbf{r}}_{n-1} - \mathbf{r}_0 \in \mathbf{A}\mathcal{K}_n$ the following more precise statement holds.

Lemma 1. *The orthogonal residual* $\mathbf{r}_n$ *(if it exists) and the minimizing residuals* $\widetilde{\mathbf{r}}_n$ *and* $\widetilde{\mathbf{r}}_{n-1}$*, which satisfy* (2.10)–(2.12)*, lie in a one-dimensional linear manifold: their differences*

$$\mathbf{r}_n - \widetilde{\mathbf{r}}_{n-1}\,, \quad \widetilde{\mathbf{r}}_n - \widetilde{\mathbf{r}}_{n-1}\,, \quad \mathbf{r}_n - \widetilde{\mathbf{r}}_n \ \in\ \mathbf{A}\mathcal{K}_n \ominus \mathbf{A}\mathcal{K}_{n-1} \tag{2.13}$$

lie in a one-dimensional subspace.

Now, Krylov spaces have the fundamental property that $\dim \mathcal{K}_{n+1} = \dim \mathcal{K}_n + 1$ unless one of the following equivalent properties hold:

(i) $\dim \mathcal{K}_m = \dim \mathcal{K}_n$ for all $m > n$,
(ii) $\mathbf{r}_n = \mathbf{o}$,
(iii) $\widetilde{\mathbf{r}}_n = \mathbf{o}$,
(iv) $\mathbf{r}_m = \widetilde{\mathbf{r}}_m = \mathbf{o}$ for all $m \geq n$

In particular, if $\widetilde{\mathbf{r}}_{n-1} = \mathbf{o}$, all three vectors in (2.12) are zero; this case is of no interest and is therefore excluded from now on. Moreover, $\widetilde{\mathbf{r}}_n = \mathbf{o}$ if and only if $\mathbf{r}_n = \mathbf{o}$.

Since $\mathbf{r}_n \perp \widetilde{\mathbf{r}}_{n-1} \in \mathcal{K}_n$, the difference $\mathbf{r}_n - \widetilde{\mathbf{r}}_{n-1}$ is nonzero, and, hence, by (2.13), $\widetilde{\mathbf{r}}_n - \widetilde{\mathbf{r}}_{n-1}$ must be a multiple of $\mathbf{r}_n - \widetilde{\mathbf{r}}_{n-1}$. In case $\mathbf{r}_n \neq \mathbf{o}$ the situation of Figure 1 applies and it is clear that $\widetilde{\mathbf{r}}_n \neq \widetilde{\mathbf{r}}_{n-1}$ too, and thus $\mathbf{r}_n - \widetilde{\mathbf{r}}_{n-1}$ is also a multiple of $\widetilde{\mathbf{r}}_n - \widetilde{\mathbf{r}}_{n-1}$. (This will follow formally too, in a moment.) If $\mathbf{r}_n = \mathbf{o}$, then $\widetilde{\mathbf{r}}_n = \mathbf{o}$, and thus the two first differences in (2.13) both equal $-\widetilde{\mathbf{r}}_{n-1}$.

Consequently, whenever $\mathbf{r}_n$ exists, there are recursions of the form

$$\widetilde{\mathbf{r}}_n := \widetilde{\mathbf{r}}_{n-1}(1-\theta_n) + \mathbf{r}_n\theta_n\,, \qquad \widetilde{\mathbf{x}}_n := \widetilde{\mathbf{x}}_{n-1}(1-\theta_n) + \mathbf{x}_n\theta_n\,, \tag{2.14}$$

and

$$\mathbf{r}_n := \widetilde{\mathbf{r}}_n(1-\widetilde{\theta}_n) + \widetilde{\mathbf{r}}_{n-1}\widetilde{\theta}_n\,, \qquad \mathbf{x}_n := \widetilde{\mathbf{x}}_n(1-\widetilde{\theta}_n) + \widetilde{\mathbf{x}}_{n-1}\widetilde{\theta}_n\,, \tag{2.15}$$

Note that the recurrences for the iterates are consistent with those for the residual: $\mathbf{r}_n = \mathbf{b} - \mathbf{A}\mathbf{x}_n$ and $\widetilde{\mathbf{r}}_n = \mathbf{b} - \mathbf{A}\widetilde{\mathbf{x}}_n$. It is easily verified that the two coefficients θ_n and $\widetilde{\theta}_n$ are related by

$$\theta_n = \frac{1}{1-\widetilde{\theta}_n}\,, \qquad \widetilde{\theta}_n = 1 - \frac{1}{\theta_n}\,. \tag{2.16}$$

We show next, that the coefficients can be determined by enforcing the "missing" orthogonality condition. To make $\|\widetilde{\mathbf{r}}_n\|$ minimal in (2.14) we need

$$\widetilde{\mathbf{r}}_n \perp (\widetilde{\mathbf{r}}_{n-1} - \mathbf{r}_n)\ , \tag{2.17}$$

which means that necessarily

$$\theta_n := \frac{\langle \widetilde{\mathbf{r}}_{n-1} - \mathbf{r}_n, \widetilde{\mathbf{r}}_{n-1}\rangle}{\|\widetilde{\mathbf{r}}_{n-1} - \mathbf{r}_n\|^2}\,. \tag{2.18}$$

Conversely, if we assume that $\mathbf{r}_n$, $\widetilde{\mathbf{r}}_{n-1}$, and $\mathbf{r}_n - \widetilde{\mathbf{r}}_{n-1}$ satisfy the respective relations in (2.11) and (2.13), then it follows from the recursion (2.14) that $\widetilde{\mathbf{r}}_n \perp \mathbf{A}\mathcal{K}_{n-1}$ for any θ_n. Choosing θ_n according to (2.18), so that (2.17) holds, then implies that even $\widetilde{\mathbf{r}}_n \perp \mathbf{A}\mathcal{K}_n$ is valid, which is the characteristic property of minimal residual methods given in (2.6). Here we have used the fact that $\widetilde{\mathbf{r}}_{n-1} - \mathbf{r}_n \in \mathbf{A}\mathcal{K}_n$ has a nonzero component in the direction of $\mathbf{A}^n\mathbf{y}_0$ if $\mathbf{r}_n \neq \mathbf{o}$. If $\mathbf{r}_n = \mathbf{o}$, then (2.14) and (2.17) imply $\widetilde{\mathbf{r}}_n = \mathbf{o}$, so (2.6) holds too. Hence, we can conclude by induction that the smoothing process (2.14) with (2.18) turns a minimal residual sequence into an orthogonal residual sequence.

Likewise, to make $\mathbf{r}_n$ from (2.15) orthogonal to $\mathcal{K}_n$, we only have to enforce

$$\left(\widetilde{\mathbf{r}}_n(1-\widetilde{\theta}_n) + \widetilde{\mathbf{r}}_{n-1}\widetilde{\theta}_n\right) \perp \mathbf{r}_0\,, \tag{2.19}$$

which means to choose in (2.15)

$$\widetilde{\theta}_n := -\frac{\langle \mathbf{r}_0, \widetilde{\mathbf{r}}_n\rangle}{\langle \mathbf{r}_0, \widetilde{\mathbf{r}}_{n-1} - \widetilde{\mathbf{r}}_n\rangle}\,. \tag{2.20}$$

Recall that we assumed that $\mathbf{x}_n$ and, hence, $\mathbf{r}_n$ exist. If this is not the case, they can be considered to have infinite length: the orthogonal residual $\mathbf{r}_n$ can move in the one-dimensional manifold $\widetilde{\mathbf{r}}_{n-1} + (\mathbf{A}\mathcal{K}_n \ominus \mathbf{A}\mathcal{K}_{n-1})$ to infinity, in which case $\widetilde{\mathbf{r}}_n \to \widetilde{\mathbf{r}}_{n-1}$, $\theta_n \to +0$, and $\widetilde{\theta}_n \to -\infty$; see Figure 1. Hence, as is well known, the residual minimizing methods stagnate if and only if the orthogonal residual methods break down due to a "non-existing", or rather infinite, $\mathbf{x}_n$. Note that (2.14) can still be used in this case if we set $\theta_n := 0$, $\mathbf{r}_n\theta_n := \mathbf{o}$.

While (2.15) is an explicit formula showing that the basis transformation from $\{\mathbf{r}_n\}$ to $\{\widetilde{\mathbf{r}}_n\}$ is given by an upper bidiagonal matrix with elements $\widetilde{\theta}_n$ and $1-\widetilde{\theta}_n$ in column n, (2.15) is a recursive formula for the inverse transformation, whose matrix is of course the inverse of this bidiagonal matrix, and is therefore upper triangular, as is also seen by resolving the recursion: if we set $\theta_0 :\equiv 1$,

$$\widetilde{\mathbf{r}}_n = \sum_{k=0}^{n} \mathbf{r}_k\theta_k \prod_{l=k+1}^{n} (1-\theta_l)\,. \tag{2.21}$$

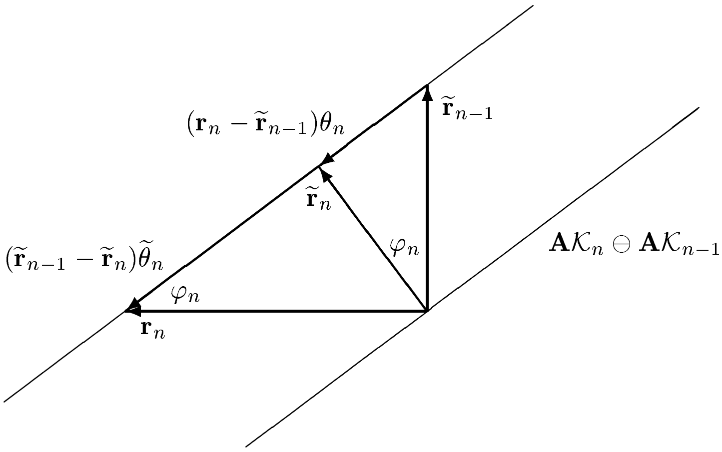

FIG. 1. *The relationship between an orthogonal residual method and the corresponding minimal residual method is best seen by displaying the 2-dimensional space $\mathcal{K}_{n+1} \ominus \mathbf{A}\mathcal{K}_{n-1}$, which contains $\mathbf{r}_n$, $\widetilde{\mathbf{r}}_n$, and $\widetilde{\mathbf{r}}_{n-1}$.*

Let us draw some further conclusions from the above formulas. First, from (2.14) and (2.17) we conclude by Pythagoras' theorem that

$$\|\widetilde{\mathbf{r}}_n\|^2 = \|\widetilde{\mathbf{r}}_{n-1}\|^2 - \|\widetilde{\mathbf{r}}_{n-1} - \mathbf{r}_n\|^2 |\theta_n|^2 \,, \tag{2.22}$$

see Figure 1, and likewise, (2.15) and (2.19) yield

$$\|\widetilde{\mathbf{r}}_n\|^2 = \|\mathbf{r}_n\|^2 - \|\widetilde{\mathbf{r}}_{n-1} - \widetilde{\mathbf{r}}_n\|^2 |\widetilde{\theta}_n|^2 . \tag{2.23}$$

Next, we further capitalize upon the fact that $\widetilde{\mathbf{r}}_{n-1} \perp \mathbf{r}_n$ by (2.10) and (2.11). Firstly, Pythagoras' theorem yields

$$\|\widetilde{\mathbf{r}}_{n-1} - \mathbf{r}_n\|^2 = \|\widetilde{\mathbf{r}}_{n-1}\|^2 + \|\mathbf{r}_n\|^2 \,. \tag{2.24}$$

Secondly, (2.18) simplifies to

$$\theta_n := \frac{\|\widetilde{\mathbf{r}}_{n-1}\|^2}{\|\widetilde{\mathbf{r}}_{n-1}\|^2 + \|\mathbf{r}_n\|^2} \qquad \in \ (0, 1] \,, \tag{2.25}$$

and inserting into (2.22) and taking the reciprocal yields

$$\frac{1}{\|\widetilde{\mathbf{r}}_n\|^2} = \frac{1}{\|\widetilde{\mathbf{r}}_{n-1}\|^2} + \frac{1}{\|\mathbf{r}_n\|^2} = \sum_{k=0}^{n} \frac{1}{\|\mathbf{r}_k\|^2} \tag{2.26}$$

and

$$\|\mathbf{r}_n\|^2 = \frac{\|\widetilde{\mathbf{r}}_n\|^2}{1 - \|\widetilde{\mathbf{r}}_n\|^2 / \|\widetilde{\mathbf{r}}_{n-1}\|^2} \qquad (n \geq 1). \tag{2.27}$$

In particular, if $\mathbf{r}_n$ is finite, then $\theta_n \neq 0$ and $\widetilde{\mathbf{r}}_n \neq \widetilde{\mathbf{r}}_{n-1}$ as claimed.

For the angle φ_n shown in Figure 1 holds on the one hand, in view of (2.25) and (2.24),

$$\sin^2 \varphi_n = \theta_n \,, \qquad \cos^2 \varphi_n = 1 - \theta_n \tag{2.28}$$

and on the other hand,

$$\cos\varphi_n = \frac{\|\widetilde{\mathbf{r}}_n\|}{\|\widetilde{\mathbf{r}}_{n-1}\|}\,, \qquad \sin\varphi_n = \frac{\|\widetilde{\mathbf{r}}_n\|}{\|\mathbf{r}_n\|}\,, \qquad \tan\varphi_n = \frac{\|\widetilde{\mathbf{r}}_{n-1}\|}{\|\mathbf{r}_n\|}\,. \tag{2.29}$$

In particular, if $\rho_0 :\equiv \|\mathbf{r}_0\|$,

$$\|\widetilde{\mathbf{r}}_n\| = \rho_0 \prod_{k=1}^{n} \cos\varphi_k\,, \qquad \|\mathbf{r}_n\| = \frac{\rho_0}{\sin\varphi_n} \prod_{k=1}^{n} \cos\varphi_k\,. \tag{2.30}$$

These well-known formulas are, in the context of the CG to CR transition, due to Paige and Saunders [17], where φ_n is the angle of a Givens rotation required for updating the LQ decomposition of a triangular matrix. The same formulas have since come up in a variety of related situations.

Independent of the relation between orthogonal and minimal residual methods discussed here, Schönauer [21] suggested the recursions (2.14) with the choice (2.18) for θ_n as a process now called *minimal residual (MR) smoothing* that can be applied to the iterates $\mathbf{x}_n$ and residuals $\mathbf{r}_n$ of any Krylov space solver and produces iterates $\widetilde{\mathbf{x}}_n$ and residuals $\widetilde{\mathbf{r}}_n$ of a new solver for the same Krylov space. Let us call the former the *primary* iterates and residuals and the latter the *smoothed* ones. Note that (2.18) does not rely on any orthogonality of the given residuals, but is chosen to minimize the new residual in the one-dimensional linear manifold spanned by $\widetilde{\mathbf{r}}_{n-1}$ and $\mathbf{r}_n$. In contrast, (2.25)–(2.27) assume orthogonality of the primary residuals.

The fact that MR smoothing of the FOM iterates yields the GMRES iterates was established by Weiss [26, 27] as a side-result of his investigation of Schönauer's MR smoothing procedure. In the Hermitian positive definite case, this means that MR smoothing produces CR iterates from CG iterates. The derivation given above for (2.14) and (2.18) is a "turned over" version of a proof in [13]. Weiss [26] also derived (2.25)–(2.27), along with many other relationships. Brown [1] had a formula equivalent to (2.27) too, as well as many related results.

The relations (2.26) and (2.27) are the basis of the analysis of the *peak–plateau connection*; see Brown [1], Cullum [3], Walker [25], Zhou and Walker [29] and, in particular, Cullum and Greenbaum [4]. We will return to it in Section 4 and see that it is even easier to understand from our Figure 1.

If the primary residuals are not orthogonal, it is difficult to relate the norms of the primary and the smoothed residuals. However, we will discuss next a variation of the above, in which we can at least relate the norms of their coordinate vectors with respect to the Krylov space basis.

3. QMR smoothing and a related framework for QMR-type methods.

Zhou and Walker [29] introduced *quasi-minimal residual (QMR) smoothing* as a variation of MR smoothing with the property that it generates the iterates and residuals of the QMR method of Freund and Nachtigal [10] if applied to the BiCG iterates and residuals, but which, like MR smoothing, is also generally applicable to the output of an iterative method. If applied to an orthogonal residual method, MR and QMR smoothing are equivalent. We introduce here the QMR smoothing as a process that is equivalent to applying MR smoothing in the coordinate space, or, more exactly, to the coordinate vectors of the primary residuals with respect to the Krylov space basis generated by these residuals. We will see that this implies that we can relate the norms of the coordinate vectors of the primary and smoothed residuals with respect to this basis.

We also define a general framework for *QMR-type methods*, and will conclude that QMR smoothing generates the same iterates and residuals as these methods.

We generalize the situation considered in the previous section by allowing the usage of different bases $\{\mathbf{v}_n\}$ and $\{\mathbf{y}_n\}$ for the iterates and the residuals. This will allow us to cover also the transitions from CGS to TFQMR and from BiCGStab to QMRCGStab, or, more generally, to produce the results of any QMR-type methods (defined later) by smoothing.

We let $\mathbf{V}_m :\equiv \begin{bmatrix} \mathbf{v}_0 & \mathbf{v}_1 & \cdots & \mathbf{v}_{m-1} \end{bmatrix}$ and replace (2.3) by

$$\mathbf{x}_n = \mathbf{x}_0 + \mathbf{V}_n\mathbf{k}_n\,, \qquad \mathbf{r}_n = \mathbf{r}_0 - \mathbf{A}\mathbf{V}_n\mathbf{k}_n \tag{3.1}$$

and (2.2) by

$$\mathbf{A}\mathbf{V}_m = \mathbf{Y}_{m+1}\underline{\mathbf{L}}_m\,, \tag{3.2}$$

where $\underline{\mathbf{L}}_m$ is in general still an $(m+1)\times m$ extended, irreducible Hessenberg matrix, though in well-known examples it is just lower bidiagonal. To describe the generation of both bases we need then at least another relation, such as $\mathbf{Y}_m = \mathbf{V}_m\mathbf{U}_m$ with an upper triangular matrix $\mathbf{U}_m$, but this relation does not play a role here. Inserting (3.1) and its analog for the smoothed quantities,

$$\widetilde{\mathbf{x}}_n = \mathbf{x}_0 + \mathbf{V}_n\widetilde{\mathbf{k}}_n\,, \qquad \widetilde{\mathbf{r}}_n = \mathbf{r}_0 - \mathbf{A}\mathbf{V}_n\widetilde{\mathbf{k}}_n\,, \tag{3.3}$$

into the smoothing formula (2.14) leads to an analogue formula for the coefficient vectors:

$$\widetilde{\mathbf{k}}_n := \begin{bmatrix} \widetilde{\mathbf{k}}_{n-1} \\ 0 \end{bmatrix}(1-\theta_n) + \mathbf{k}_n\theta_n\,. \tag{3.4}$$

(In the case where the set $\{\mathbf{y}_j\}_{j=0}^{n-1}$ is not linearly independent, (3.4) still implies that (2.14) is valid; so, we may suppose that (3.4) holds.)

On the other hand, by a standard argument, inserting (3.2) into the right-hand side formulas in (3.1) and (3.3) leads in view of $\mathbf{r}_0 = \mathbf{y}_0\rho_0$ to

$$\begin{aligned} \mathbf{r}_n &= \mathbf{Y}_{n+1}(\underline{\mathbf{e}}_1\rho_0 - \underline{\mathbf{L}}_n\mathbf{k}_n) = \mathbf{Y}_{n+1}\mathbf{q}_n\,, \quad \text{where} \quad \mathbf{q}_n :\equiv \underline{\mathbf{e}}_1\rho_0 - \underline{\mathbf{L}}_n\mathbf{k}_n\,, \\ \widetilde{\mathbf{r}}_n &= \mathbf{Y}_{n+1}(\underline{\mathbf{e}}_1\rho_0 - \underline{\mathbf{L}}_n\widetilde{\mathbf{k}}_n) = \mathbf{Y}_{n+1}\widetilde{\mathbf{q}}_n\,, \quad \text{where} \quad \widetilde{\mathbf{q}}_n :\equiv \underline{\mathbf{e}}_1\rho_0 - \underline{\mathbf{L}}_n\widetilde{\mathbf{k}}_n\,, \end{aligned} \tag{3.5}$$

with $\underline{\mathbf{e}}_1 :\equiv \begin{bmatrix} 1 & 0 & 0 & \cdots \end{bmatrix}^T \in \mathbb{R}^{n+1}$ and $\widetilde{\mathbf{q}}_0 :\equiv \mathbf{q}_0 :\equiv [\rho_0] \in \mathbb{R}^1$. We call $\mathbf{q}_n$ the *primary quasi-residual* and $\widetilde{\mathbf{q}}_n$ the *smoothed quasi-residual*. They are the coordinate vectors of the residuals with respect to our Krylov space basis $\{\mathbf{v}_n\}$. They can be related by multiplying (3.4) by $\underline{\mathbf{L}}_n$ and subtracting the result from $\underline{\mathbf{e}}_1\rho_0$:

$$\widetilde{\mathbf{q}}_n := \begin{bmatrix} \widetilde{\mathbf{q}}_{n-1} \\ 0 \end{bmatrix}(1-\theta_n) + \mathbf{q}_n\theta_n\,. \tag{3.6}$$

Assume now that a total of m steps are executed with the primary method, and let us extend the quasi-residual vectors $\mathbf{q}_n$ and $\widetilde{\mathbf{q}}_n$ by $m-n$ zero components and call the extended vectors $\widetilde{\mathbf{q}}_n^\circ$ and $\mathbf{q}_n^\circ$, respectively. They all lie in $\mathbb{C}^{m+1}$, and, clearly, in the 2-norm holds $\|\mathbf{q}_n^\circ\| = \|\mathbf{q}_n\|$ and $\|\widetilde{\mathbf{q}}_n^\circ\| = \|\widetilde{\mathbf{q}}_n\|$. In particular, $\mathbf{q}_0^\circ = \widetilde{\mathbf{q}}_0^\circ = \mathbf{e}_1^\circ\rho_0$. Moreover, by (3.6),

$$\widetilde{\mathbf{q}}_n^\circ := \widetilde{\mathbf{q}}_{n-1}^\circ(1-\theta_n) + \mathbf{q}_n^\circ\theta_n\,, \tag{3.7}$$

which is just a special case of the smoothing process (2.14). In other words, *in the coordinate space, the smoothed quasi-residuals $\widetilde{\mathbf{q}}_n^\circ$ can be generated by smoothing the primary quasi-residuals $\mathbf{q}_n^\circ$.*

We further define $\mathbf{k}_n^\circ$, $\widetilde{\mathbf{k}}_n^\circ \in \mathbb{C}^m$ by extending $\mathbf{k}_n$, $\widetilde{\mathbf{k}}_n \in \mathbb{C}^n$ with $m-n-1$ zeros, and we let $\mathbf{e}_n^\circ$ be the nth standard basis vector in m-space and set

$$\mathcal{E}_n :\equiv \operatorname{span}\{\mathbf{e}_1^\circ, \ldots, \mathbf{e}_n^\circ\}.$$

Then we have according to the definition of $\mathbf{q}_n^\circ$ and $\widetilde{\mathbf{q}}_n^\circ$, and since $\underline{\mathbf{L}}_m$ is an extended Hessenberg matrix,

$$\begin{aligned} \mathbf{q}_n^\circ &= \mathbf{e}_1^\circ \rho_0 - \underline{\mathbf{L}}_m \mathbf{k}_n^\circ \in \mathbf{q}_0^\circ + \underline{\mathbf{L}}_m \mathcal{E}_n \subseteq \mathcal{E}_{n+1}, \\ \widetilde{\mathbf{q}}_n^\circ &= \mathbf{e}_1^\circ \rho_0 - \underline{\mathbf{L}}_m \widetilde{\mathbf{k}}_n^\circ \in \mathbf{q}_0^\circ + \underline{\mathbf{L}}_m \mathcal{E}_n \subseteq \mathcal{E}_{n+1}. \end{aligned} \tag{3.8}$$

So far we have made no assumption on the primary residuals and the smoothing coefficients θ_n. From now on we assume that for all $n \le m$ we have $\|\mathbf{y}_n\| = 1$ and

$$\mathbf{r}_n = \mathbf{y}_n \rho_n, \qquad \text{that is,} \quad \mathbf{q}_n = \mathbf{e}_{n+1}\rho_n, \quad \mathbf{q}_n^\circ = \mathbf{e}_{n+1}^\circ \rho_n \tag{3.9}$$

with $\rho_n :\equiv \|\mathbf{r}_n\| \ge 0$, which implies that $\mathbf{k}_n = \mathbf{L}_n^{-1}\mathbf{e}_1\rho_0$, where $\mathbf{L}_n$ is the upper $n \times n$ submatrix of $\underline{\mathbf{L}}_n$. Moreover, we assume that θ_n is chosen so that in (3.6) $\widetilde{\mathbf{q}}_n$ is as short as possible or, equivalently, in (3.7) $\widetilde{\mathbf{q}}_n^\circ$ has minimal length.

As before, we let the basis (or, more generally, the generating set) of $\mathcal{K}_{m+1}$ be given by the normalized primary residuals $\mathbf{r}_n/\rho_n$ and choose θ_n so that analogously to (2.17)

$$\widetilde{\mathbf{q}}_n^\circ \perp (\widetilde{\mathbf{q}}_{n-1}^\circ - \mathbf{q}_n^\circ) = \widetilde{\mathbf{q}}_{n-1}^\circ - \mathbf{e}_{n+1}^\circ \rho_n. \tag{3.10}$$

By induction we can show that $\widetilde{\mathbf{q}}_n^\circ \perp \underline{\mathbf{L}}_m \mathcal{E}_n$, so that in analogy to (2.11)

$$\mathbf{q}_n^\circ \perp \mathcal{E}_n \supseteq \underline{\mathbf{L}}_m \mathcal{E}_{n-1}, \qquad \widetilde{\mathbf{q}}_n^\circ \perp \underline{\mathbf{L}}_m \mathcal{E}_n, \qquad \widetilde{\mathbf{q}}_{n-1}^\circ \perp \underline{\mathbf{L}}_m \mathcal{E}_{n-1}. \tag{3.11}$$

In fact, the first statement is immediate from the choice (3.9), and the last is the induction assumption. Hence, both terms on the right side of (3.7) are orthogonal to $\underline{\mathbf{L}}_m \mathcal{E}_{n-1}$, and their sum lies in view of (3.8) in $\mathbf{q}_0^\circ + \underline{\mathbf{L}}_m \mathcal{E}_n \subseteq \mathcal{E}_{n+1}$. Making the sum orthogonal to $\widetilde{\mathbf{q}}_{n-1}^\circ - \mathbf{q}_n^\circ \in \underline{\mathbf{L}}_m \mathcal{E}_n \ominus \underline{\mathbf{L}}_m \mathcal{E}_{n-1}$, as required in (3.10), suffices to attain $\widetilde{\mathbf{q}}_n^\circ \perp \underline{\mathbf{L}}_m \mathcal{E}_n$.

From (3.8) and (3.11) we conclude that, in summary,

$$\mathbf{q}_n^\circ, \ \widetilde{\mathbf{q}}_n^\circ, \ \widetilde{\mathbf{q}}_{n-1}^\circ \in \mathcal{E}_{n+1} \ominus \underline{\mathbf{L}}_m \mathcal{E}_{n-1}, \tag{3.12}$$

and that, in view of $\mathbf{q}_n^\circ - \mathbf{q}_0^\circ$, $\widetilde{\mathbf{q}}_n^\circ - \mathbf{q}_0^\circ$, $\widetilde{\mathbf{q}}_{n-1}^\circ - \mathbf{q}_0^\circ \in \underline{\mathbf{L}}_m \mathcal{E}_n$,

$$\mathbf{q}_n^\circ - \widetilde{\mathbf{q}}_{n-1}^\circ, \ \widetilde{\mathbf{q}}_n^\circ - \widetilde{\mathbf{q}}_{n-1}^\circ, \ \mathbf{q}_n^\circ - \widetilde{\mathbf{q}}_n^\circ \in \underline{\mathbf{L}}_m \mathcal{E}_n \ominus \underline{\mathbf{L}}_m \mathcal{E}_{n-1}. \tag{3.13}$$

Hence, the following analog of Lemma 1 is valid:

LEMMA 2. *If the assumptions* (3.9) *and* (3.10) *hold, the extended primary quasi-residual $\mathbf{q}_n^\circ$ (if it exists) and the extended smoothed quasi-residuals $\widetilde{\mathbf{q}}_n^\circ$ and $\widetilde{\mathbf{q}}_{n-1}^\circ$, which according to* (3.7) *all lie in a one-dimensional linear manifold, satisfy* (3.8) *and* (3.11)–(3.13). *In particular, this manifold is given by $\mathbf{q}_n^\circ + (\underline{\mathbf{L}}_m \mathcal{E}_n \ominus \underline{\mathbf{L}}_m \mathcal{E}_{n-1})$.*

Consequently, Figure 1 essentially still holds if we replace $\mathbf{r}_n$, $\widetilde{\mathbf{r}}_n$, and $\widetilde{\mathbf{r}}_{n-1}$ by $\mathbf{q}_n$, $\widetilde{\mathbf{q}}_n$, and $\widetilde{\mathbf{q}}_{n-1}$, respectively, and the space $\mathbf{A}\mathcal{K}_n \ominus \mathbf{A}\mathcal{K}_{n-1}$ by $\underline{\mathbf{L}}_m \mathcal{E}_n \ominus \underline{\mathbf{L}}_m \mathcal{E}_{n-1}$; see Figure 2.

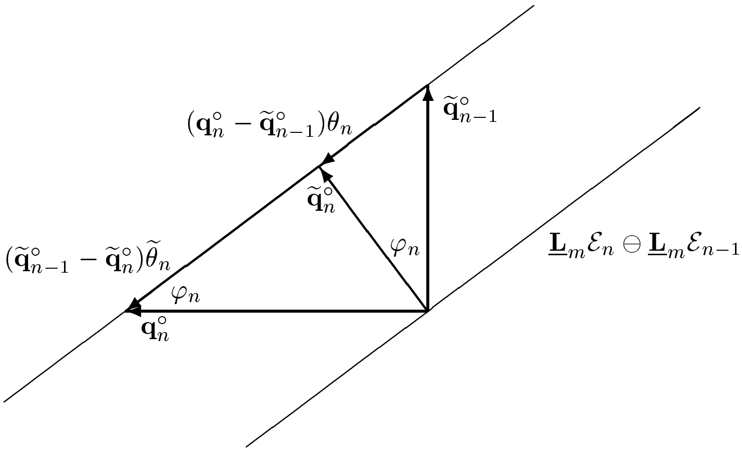

FIG. 2. *The relationship between the coordinates of the residuals of a primary method and those of the corresponding coordinates after QMR smoothing is best seen by displaying the 2-dimensional space $\mathcal{E}_{n+1} \ominus \underline{\mathbf{L}}_m \mathcal{E}_{n-1}$, which contains $\mathbf{q}_n^\circ$, $\widetilde{\mathbf{q}}_n^\circ$, and $\widetilde{\mathbf{q}}_{n-1}^\circ$.*

By (3.6) and (3.9) we have

$$\widetilde{\mathbf{q}}_n := \begin{bmatrix} \widetilde{\mathbf{q}}_{n-1} \\ 0 \end{bmatrix} (1-\theta_n) + \mathbf{e}_{n+1}\rho_n\theta_n \tag{3.14}$$

[or, $\widetilde{\mathbf{q}}_n^\circ := \widetilde{\mathbf{q}}_{n-1}^\circ(1-\theta_n) + \mathbf{e}_{n+1}^\circ\rho_n\theta_n$], where clearly the two vectors on the right-hand side are orthogonal, so that by Pythagoras' theorem

$$\|\widetilde{\mathbf{q}}_n\|^2 = \|\widetilde{\mathbf{q}}_{n-1}\|^2 |1-\theta_n|^2 + \rho_n^2\theta_n^2 \qquad \text{with} \quad \rho_n :\equiv \|\mathbf{r}_n\| \,. \tag{3.15}$$

This orthogonality is a consequence of the primary quasi-residuals being orthogonal to each other and implies that the analogs of (2.25)–(2.27) hold:

$$\theta_n := \frac{\|\widetilde{\mathbf{q}}_{n-1}\|^2}{\|\widetilde{\mathbf{q}}_{n-1}\|^2 + \|\mathbf{q}_n\|^2} \qquad \in \ (0,1]\,, \tag{3.16}$$

$$\frac{1}{\|\widetilde{\mathbf{q}}_n\|^2} = \frac{1}{\|\widetilde{\mathbf{q}}_{n-1}\|^2} + \frac{1}{\|\mathbf{q}_n\|^2} = \sum_{k=0}^{n} \frac{1}{\|\mathbf{q}_k\|^2}\,, \tag{3.17}$$

and

$$\|\mathbf{q}_n\|^2 = \frac{\|\widetilde{\mathbf{q}}_n\|^2}{1 - \|\widetilde{\mathbf{q}}_n\|^2 / \|\widetilde{\mathbf{q}}_{n-1}\|^2} \qquad (n \geq 1). \tag{3.18}$$

Everywhere we could insert $\|\mathbf{q}_n\|^2 = \rho_n^2$. Note that once the norm ρ_n of the primary residual has been computed, θ_n can be obtained from (3.16) without spending an inner product. This θ_n can then be inserted into the smoothing formulas (2.14) and (3.11) in order to generate $\widetilde{\mathbf{x}}_n$, $\widetilde{\mathbf{r}}_n$, and $\widetilde{\mathbf{q}}_n$.

Actually, there is no need to update $\widetilde{\mathbf{q}}_n$. It suffices to compute

$$\xi_n :\equiv \frac{1}{\|\widetilde{\mathbf{q}}_n\|^2}\,, \tag{3.19}$$

for which we get from (3.17) and (3.16):

$$\xi_n := \xi_{n-1} + \frac{1}{\rho_n^2}, \qquad \theta_n := \frac{1}{1 + \xi_{n-1}\,\rho_n^2} = \frac{1}{\xi_n\,\rho_n^2}. \tag{3.20}$$

The smoothing process based on (3.20) and (2.14) is due to Zhou and Walker [29] and is called *quasi-minimal residual smoothing* or, briefly, *QMR smoothing*.

There is an alternative way to construct the same smoothed iterates and residuals, which only requires knowing the basis $\{\mathbf{y}_n\}$ and the matrix $\underline{\mathbf{L}}_m$, but not the lengths ρ_n of the primary residuals. In fact, the condition $\widetilde{\mathbf{q}}_n^\circ \perp \underline{\mathbf{L}}_m \mathcal{E}_n$ means that $\widetilde{\mathbf{k}}_n$ solves the least squares problem

$$\|\widetilde{\mathbf{q}}_n\|^2 = \|\underline{\mathbf{e}}_1\rho_0 - \underline{\mathbf{L}}_n\widetilde{\mathbf{k}}_n\|^2 = \min! \tag{3.21}$$

This is no surprise since we applied MR smoothing to the orthogonal primary coordinates $\mathbf{q}_n$ to find $\widetilde{\mathbf{q}}_n$, and in this situation local minimization implies global minimization. The least squares problem can be solved in every step by updating the QR decomposition of $\underline{\mathbf{L}}_n$, exactly as done in the GMRes, QMR, and TFQMR methods. Inserting $\widetilde{\mathbf{k}}_n$ into (3.3) allow us then to determine the corresponding iterates $\widetilde{\mathbf{x}}_n$ (and, optionally, the residuals $\widetilde{\mathbf{r}}_n$). It may also be possible to find efficient recursions for updating $\widetilde{\mathbf{x}}_n$, as they exist for MinRes, QMR, and TFQMR.

In view of the analogy to the QMR method, we call such a method that minimizes $\|\widetilde{\mathbf{q}}_n\|$ a *QMR-type method.* TFQMR and QMRCGStab belong to this class because they make use of the representation (3.2), where the even numbered $\mathbf{y}_{2k}$ are normalized residual vectors of CGS and BiCGStab, respectively, while the odd numbered $\mathbf{y}_{2k+1}$ are intermediate quantities that can be considered as residual vectors at half steps. The choice of the vectors $\mathbf{v}_n$ is not important here. Consequently, the transitions from CGS to TFQMR and from BiCGStab to QMRCGStab can be understood as QMR smoothing. Given the generality of (3.2) it would be easy to introduce further methods that fit into this pattern.

If the columns of $\mathbf{Y}_m$ are orthogonal, then, for $n < m$, QMR smoothing is equivalent to MR smoothing since the coordinate map is isometric in the 2-norm:

$$\|\mathbf{q}_n\| = \|\mathbf{r}_n\| = \rho_n\,, \qquad \|\widetilde{\mathbf{q}}_n\| = \|\widetilde{\mathbf{r}}_n\|\,. \tag{3.22}$$

However, if the columns of $\mathbf{Y}_m$ are not necessarily orthogonal to each other, as, *e.g.*, in the biconjugate gradient (BiCG) method, then the two smoothing processes differ also in exact arithmetic. In the case of BiCG, QMR smoothing transforms the (primary) BiCG iterates and residuals into the iterates and residuals of the quasi-minimal residual (QMR) method since in its version based on three-term recurrences, where $\mathbf{V}_m = \mathbf{Y}_m$, the latter solves in coordinate space exactly the same least-square problem (3.21) as MinRes and GMRes do.

4. Bounds for the norms of the smoothed residuals or quasi-residuals. The two equivalent relations (2.26) and (2.27) are the basis of the so-called *peak-plateau connection* between the residual norms of primary and smoothed methods, as has been clarified by Cullum and Greenbaum [4], following earlier work of Brown [1], Cullum [3], Walker [25], and others. For generality, we will start here from the corresponding relations (3.17) and (3.18) for the quasi-residual norms, from which we can in fact immediately conclude that

$$\|\mathbf{q}_n\| \ll \|\widetilde{\mathbf{q}}_{n-1}\| \quad \Longleftrightarrow \quad \|\widetilde{\mathbf{q}}_n\| \approx \|\mathbf{q}_n\| \quad \Longleftrightarrow \quad \|\widetilde{\mathbf{q}}_n\| \ll \|\widetilde{\mathbf{q}}_{n-1}\| \tag{4.1}$$

and

$$(4.2) \qquad \|\mathbf{q}_n\| \gg \|\widetilde{\mathbf{q}}_{n-1}\| \quad \Longleftrightarrow \quad \|\widetilde{\mathbf{q}}_n\| \approx \|\widetilde{\mathbf{q}}_{n-1}\| \quad \Longleftrightarrow \quad \|\mathbf{q}_n\| \gg \|\widetilde{\mathbf{q}}_n\| \,.$$

In words: if $\|\mathbf{q}_n\|$ decreases fast, then $\|\widetilde{\mathbf{q}}_n\|$ is not much smaller than $\|\mathbf{q}_n\|$, while if $\|\mathbf{q}_n\|$ is much larger than $\|\widetilde{\mathbf{q}}_{n-1}\|$ (as, *e.g.*, when the primary residual norm history has a peak), then $\|\widetilde{\mathbf{q}}_n\|$ essentially stagnates, that is, the smoothed residual norm history has a plateau.

Since $\|\mathbf{q}_n^\circ\| = \|\mathbf{q}_n\|$ and $\|\widetilde{\mathbf{q}}_n^\circ\| = \|\widetilde{\mathbf{q}}_n\|$, the connections (4.1) and (4.2) also hold for the extended vectors $\mathbf{q}_n^\circ$, $\widetilde{\mathbf{q}}_n^\circ$, and $\widetilde{\mathbf{q}}_{n-1}^\circ$ shown in Figure 2. In fact, these connections are seen to be an immediate consequences of this figure, in particular of the fact that the angle between $\mathbf{q}_n^\circ$ and $\widetilde{\mathbf{q}}_{n-1}^\circ$ is a right one.

This relation between peaks of CG or FOM and plateaux of CR or GMRes is by now well-known. This effect appears even more often and more pronounced when we compare the residual norms of BiCG and those of QMR. For the first four methods, the underlying equations hold both for the residual norms and the quasi-residual norms (since they are equal). In contrast, for BiCG and QMR they only apply to the quasi-residual norms, but in a vaguer sense they remain valid for the residual norms. Of course, the non-orthogonal basis of the Krylov space introduces some distortion, which could be expressed in terms of its condition number, but the closer to orthogonal the primary residuals are, the better the relations hold approximately also for the residuals.

Using only the two equivalent relations (3.17) and (3.18) we will now derive bounds that give a quantitative answer to the vague statement of the peak-plateaux behavior and also answer the question of how much we can gain by applying a smoothed method (like CR, GMRes, or QMR) instead of a primary method (like CG, FOM, or BiCG) if convergence is measured in terms of the residual norm or the quasi-residual norm. We formulate all estimates in terms of quasi-residual norms, but keep in mind that they hold for the residual norms if the primary residuals are orthogonal.

First, (3.17) implies

$$\frac{1}{\|\widetilde{\mathbf{q}}_n\|^2} = \sum_{k=0}^{n} \frac{1}{\|\mathbf{q}_k\|^2} \le \sum_{k=0}^{n} \frac{1}{\min_{0\le j\le n} \|\mathbf{q}_j\|^2} = \frac{n+1}{\min_{0\le j\le n} \|\mathbf{q}_j\|^2}$$

and

$$\frac{1}{\|\widetilde{\mathbf{q}}_n\|^2} = \sum_{k=0}^{n} \frac{1}{\|\mathbf{q}_k\|^2} \ge \frac{1}{\min_{0\le j\le n} \|\mathbf{q}_j\|^2} \,.$$

This simple estimate yields a first lower bounds for the quasi-residual norm, which we state together with a trivial upper bound reflecting the minimality property of $\widetilde{\mathbf{q}}_n$.

Theorem 3. *The norms of the primary and the smoothed quasi-residuals* ($\mathbf{q}_n$ *and* $\widetilde{\mathbf{q}}_n$*, respectively*) *satisfy*

$$(4.3) \qquad \frac{1}{\sqrt{n+1}} \min_{0\le j\le n} \|\mathbf{q}_j\| \le \|\widetilde{\mathbf{q}}_n\| \le \min_{0\le j\le n} \|\mathbf{q}_j\| \,.$$

Equality holds at left if and only if $\|\mathbf{q}_k\| = \min_{0\le j\le n} \|\mathbf{q}_j\|$*,* $k = 0, \ldots, n$*, that is, if and only if* $\|\mathbf{q}_k\| = \|\mathbf{q}_0\|$*,* $k = 0, \ldots, n$*.*

Consequently, if CG or FOM stagnates initially for n steps, the quasi-residual norms of CR and GMRes decrease roughly by a factor $\sqrt{n}$.

Theorem 3 can be generalized to provide a lower bound for the smoothed quasi-residual in a later phase:

THEOREM 4. *Let $\chi \geq 1$ be such that*

$$\|\widetilde{\mathbf{q}}_m\|^2 = \frac{1}{\chi}\,\|\mathbf{q}_m\|^2. \tag{4.4}$$

Then, for $n > m$,

$$\frac{1}{\sqrt{n-m+\chi}}\,\min_{m\leq j\leq n}\|\mathbf{q}_j\| \leq \|\widetilde{\mathbf{q}}_n\| \leq \frac{1}{\sqrt{\chi}}\,\|\mathbf{q}_m\|. \tag{4.5}$$

Proof. By (3.17) and by definition of χ,

$$\begin{aligned}\frac{1}{\|\widetilde{\mathbf{q}}_n\|^2} &= \frac{1}{\|\widetilde{\mathbf{q}}_m\|^2} + \sum_{k=m+1}^{n}\frac{1}{\|\mathbf{q}_k\|^2} = \frac{\chi}{\|\mathbf{q}_m\|^2} + \sum_{k=m+1}^{n}\frac{1}{\|\mathbf{q}_k\|^2}\\ &\leq \frac{\chi}{\min_{m\leq j\leq n}\|\mathbf{q}_j\|^2} + \sum_{k=m+1}^{n}\frac{1}{\min_{m\leq j\leq n}\|\mathbf{q}_j\|^2}\,,\end{aligned}$$

so that in view of $\|\widetilde{\mathbf{q}}_n\| \leq \|\widetilde{\mathbf{q}}_m\|$ and (4.4)

$$\frac{\chi}{\|\mathbf{q}_m\|^2} \leq \frac{1}{\|\widetilde{\mathbf{q}}_n\|^2} \leq \frac{n-m+\chi}{\min_{m\leq j\leq n}\|\mathbf{q}_j\|^2}.$$

Taking the square root of the reciprocal yields the claimed result. □

Another useful variation of this result is obtained when we assume that the primary method converges with at least a certain geometric rate.

THEOREM 5. *Let $\chi \geq 1$ be such that* (4.4) *holds and assume that for some $\gamma > 0$ and some $n > m$*

$$\|\mathbf{q}_k\| \leq \|\mathbf{q}_m\|\,\gamma^{k-m} \qquad (m < k \leq n)\,. \tag{4.6}$$

Then

$$\|\widetilde{\mathbf{q}}_n\| \leq \|\mathbf{q}_m\|\frac{\gamma^{n-m}}{\sqrt{\chi\gamma^{2(n-m)} + \frac{1-\gamma^{2(n-m)}}{1-\gamma^2}}}\,. \tag{4.7}$$

If $\gamma < 1$, the reciprocal of the square root in (4.7) *approaches $\sqrt{1-\gamma^2}$ as $n\to\infty$:*

$$\|\widetilde{\mathbf{q}}_n\| \leq \|\mathbf{q}_m\|\gamma^{n-m}\sqrt{1-\gamma^2}\left(1+\mathcal{O}(\gamma^{2(n-m)})\right) \qquad \textit{as} \quad n\to\infty\,. \tag{4.8}$$

The results also hold when in the inequalities (4.6)–(4.8) *$\leq$ is replaced by $\geq$.*

Proof. By (3.17) and by definition of χ,

$$\frac{1}{\|\widetilde{\mathbf{q}}_n\|^2} = \frac{1}{\|\widetilde{\mathbf{q}}_m\|^2} + \sum_{k=m+1}^{n}\frac{1}{\|\mathbf{q}_k\|^2} \geq \frac{\chi}{\|\mathbf{q}_m\|^2} + \sum_{k=m+1}^{n}\frac{\gamma^{-2(k-m)}}{\|\mathbf{q}_m\|^2}\,.$$

Consequently,

$$\frac{1}{\|\widetilde{\mathbf{q}}_n\|^2} \geq \frac{1}{\|\mathbf{q}_m\|^2}\left(\chi + \gamma^{-2(n-m)}\frac{1-\gamma^{2(n-m)}}{1-\gamma^2}\right).$$

Taking the square root of the reciprocal yields the claimed result. If the inequality signs are changed, the derivation persists. $\square$

Since the previous theorem holds with inequalities of both types, the following corollary on a primary sequence that converges exactly geometrically is immediate.

COROLLARY 6. *Let $\chi \geq 1$ be given by (4.4), and assume that for some $\gamma > 0$ and some $n > m$*

$$\|\mathbf{q}_k\| = \|\mathbf{q}_m\|\,\gamma^{k-m} \qquad (m < k \leq n)\,. \tag{4.9}$$

Then

$$\|\widetilde{\mathbf{q}}_n\| = \|\mathbf{q}_n\|\,\frac{1}{\sqrt{\chi\gamma^{2(n-m)} + \frac{1-\gamma^{2(n-m)}}{1-\gamma^2}}}\,, \tag{4.10}$$

and thus, if $\gamma < 1$,

$$\|\widetilde{\mathbf{q}}_n\| = \|\mathbf{q}_n\|\sqrt{1-\gamma^2}\,\Big(1 + \mathcal{O}(\gamma^{2(n-m)})\Big) \qquad as \quad n \to \infty. \tag{4.11}$$

This corollary describes the interesting but somewhat unlikely case of exactly geometrically decreasing residuals. Other special cases could also be treated analytically. Such results may be used for providing upper and lower bounds for more general cases, as can be seen from the following simple inclusion result.

THEOREM 7. *Let three sequences of primary (quasi-)residuals, $\{\mathbf{q}_n\}$, $\{\underline{\mathbf{q}}_n\}$, and $\{\overline{\mathbf{q}}_n\}$ satisfying*

$$\|\underline{\mathbf{q}}_n\| \leq \|\mathbf{q}_n\| \leq \|\overline{\mathbf{q}}_n\| \qquad (0 \leq n \leq m) \tag{4.12}$$

be given. Then the corresponding sequences of smoothed (quasi-)residuals $\{\widetilde{\mathbf{q}}_n\}$, $\{\widetilde{\underline{\mathbf{q}}}_n\}$, and $\{\widetilde{\overline{\mathbf{q}}}_n\}$ satisfy analogously

$$\|\widetilde{\underline{\mathbf{q}}}_n\| \leq \|\widetilde{\mathbf{q}}_n\| \leq \|\widetilde{\overline{\mathbf{q}}}_n\| \qquad (0 \leq n \leq m). \tag{4.13}$$

Proof. This result is an immediate consequence of (3.17). $\square$

5. Illustrations and numerical results.

Let us first illustrate the results of Theorems 3–7 and Corollary 6 by some figures reflecting the behavior of artificial, but archetypical examples. Consider first the primary quasi-residuals (or residuals if orthogonal)

$$\|\mathbf{q}_k\| = \begin{cases} 2^k\,, & k = 1,\ldots,10\,,\\ 2^{20-k}\,, & k = 10,\ldots,30\,,\\ 2^{-10}\,, & k = 30,\ldots,40\,,\\ 2^{30-k}\,, & k = 40,\ldots,50\,. \end{cases} \tag{5.1}$$

Their quasi-residual norm history (solid line) and that of the corresponding smoothed quasi-residuals (dot-dashed) are given in Figure 3, along with the lower bound from (4.3) (dotted line) and the upper and lower bounds from (4.5) with $m = 25$ (dashed lines). It is obvious that the smoothed quasi-residuals are considerably better when the primary quasi-residuals do not improve or have not yet caught up with previous losses, but that there is very little gain where the primary quasi-residual norms decrease fast.

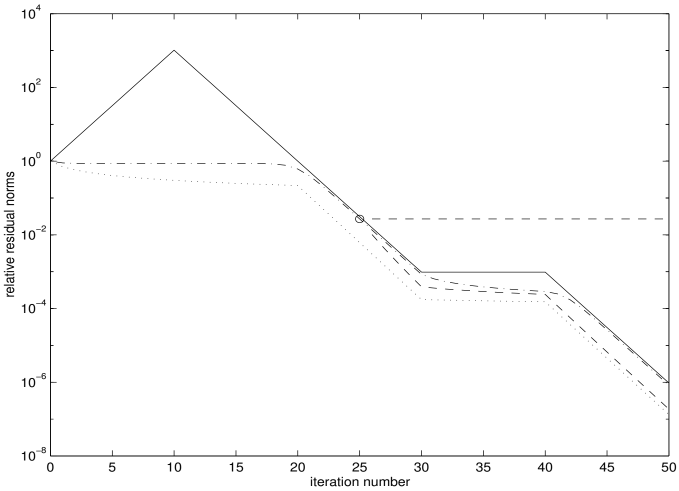

FIG. 3. *The (quasi-)residual norm history of example* (5.1) *(solid line), the corresponding smoothed (quasi-)residual norm history (dot-dashed), the lower bound from* (4.3) *(dotted line) and the upper and lower bounds from* (4.5) *with* $m = 25$ *(dashed lines).*

In Figure 4 we consider two slowly, namely logarithmically converging quasi-residual norm histories,

$$\|\overline{\mathbf{q}}_n\| :\equiv \frac{1}{n+1}\,, \qquad \|\underline{\mathbf{q}}_n\| :\equiv \frac{1}{2(n+1)} = \tfrac{1}{2}\,\|\overline{\mathbf{q}}_n\| \tag{5.2}$$

shown as solid lines. According to formula (3.17) the corresponding smoothed quasi-residuals satisfy, respectively,

$$\|\widetilde{\overline{\mathbf{q}}}_n\| = \sqrt{\frac{6}{(n+1)(n+2)(2n+3)}}\,, \qquad \|\widetilde{\underline{\mathbf{q}}}_n\| = \tfrac{1}{2}\,\|\widetilde{\overline{\mathbf{q}}}_n\|\,. \tag{5.3}$$

They are shown as dot-dashed lines. Clearly, both pairs of curves have a constant vertical distance of $\log 2$. By Theorem 7, whenever the norms of quasi-residual or orthogonal residuals are limited to the band between the pair of solid lines, the corresponding smoothed (quasi-)residuals are limited to the band between the dot-dashed lines.

In Figure 5 we illustrate additionally Theorems 3 and 4 for the first example, $\|\mathbf{q}_n\| :\equiv \|\overline{\mathbf{q}}_n\| = \frac{1}{n+1}$, shown again as solid line. The dotted line is the corresponding lower bound in (4.3), while the upper bound is given by the upper border of the figure. The smoothed quasi-residual norms $\|\widetilde{\mathbf{q}}_n\| :\equiv \|\widetilde{\overline{\mathbf{q}}}_n\|$ of (5.3) are shown as dot-dashed line. If we apply Theorem 4 with $m = 50$, the upper bound of (4.5) yields the horizontal dashed line, and the lower bound leads to the lower dashed curve.

Note that in Figures 4 and 5 the vertical scale is rather large for residual norms. Since the convergence is so slow, smoothing does speed up the quasi-residual norm convergence here. However, logarithmic convergence is very unusual for Krylov space methods. Typically, they converge linearly or superlinearly. Therefore, we next consider in Figure 6 the two geometric sequences

$$\|\mathbf{q}_n\| :\equiv \gamma_i^n \quad (i = 1, 2)\,, \qquad \gamma_1 :\equiv \tfrac{3}{4}\,, \qquad \gamma_2 :\equiv \tfrac{3}{5}\,, \tag{5.4}$$

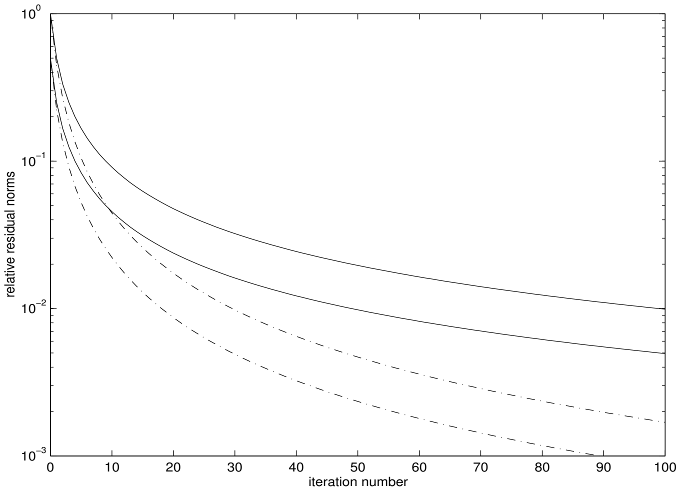

FIG. 4. *The (quasi-)residual norm histories of the untypical, logarithmically converging examples* (5.2) *(solid lines) and the corresponding smoothed (quasi-)residual norm histories (dot-dashed lines) given by* (5.3).

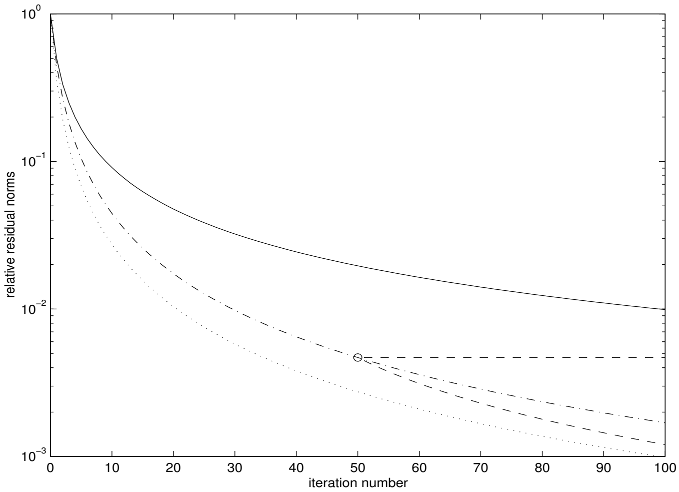

FIG. 5. *The (quasi-)residual norm history of* $\|\mathbf{q}_n\| :\equiv \frac{1}{n+1}$ *(solid line) and the corresponding smoothed (quasi-)residual norm history (dot-dashed line). The dotted line shows the lower bound from* (4.3), *and the upper and lower dashed lines illustrate the upper and lower bounds from* (4.5) *with* $m = 50$.

and the corresponding two smoothed sequences given by (4.10) with $m = 0$ and $\chi = 1$:

$$\|\widetilde{\mathbf{q}}_n\| = \|\mathbf{q}_n\| \, \frac{1}{\sqrt{\gamma_i^{2n} + \frac{1-\gamma_i^{2n}}{1-\gamma_i^2}}} = \|\mathbf{q}_n\| \sqrt{\frac{1-\gamma_i^2}{1-\gamma_i^{2n+2}}}\,. \tag{5.5}$$

Now, the improvement due to smoothing is very limited and does not merit any additional cost.

In Figure 7 we illustrate again the bounds from the Theorems 3 and 4 for the

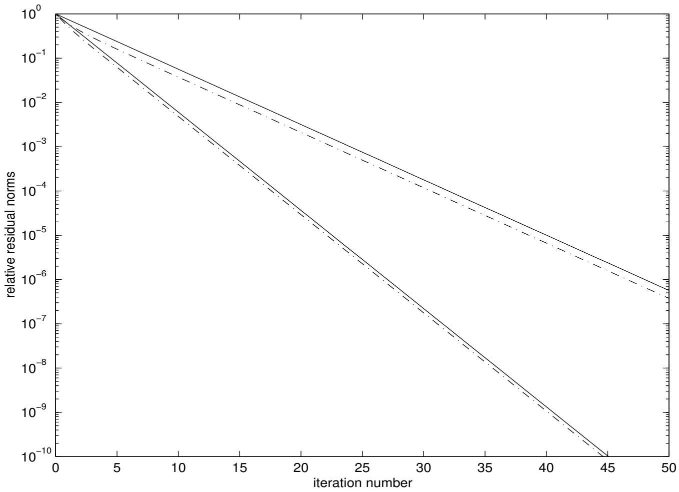

FIG. 6. *The (quasi-)residual norm histories of the geometrically converging examples* (5.4) *(solid line) and the corresponding smoothed (quasi-)residual norm histories (dot-dashed) given by* (5.5).

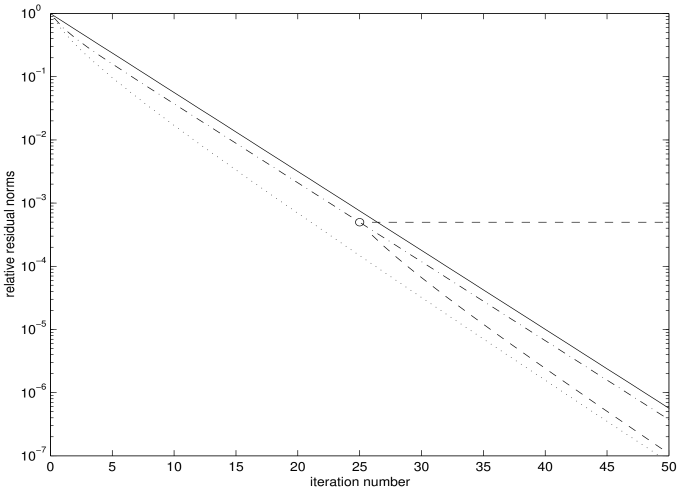

FIG. 7. *The (quasi-)residual norm history of* $\|\mathbf{q}_n\| :\equiv \gamma_1^n :\equiv (\frac{3}{4})^n$ *(solid line) and the corresponding smoothed (quasi-)residual norm history (dot-dashed line). The dotted line shows the lower bound from* (4.3)*, and the upper and lower dashed lines illustrate the upper and lower bounds in* (4.5) *with* $m = 25$.

slower converging of these two examples, in a manner analogous to that of Figure 5. Here, the value of the trivial lower bound (4.3) (dotted line) for the smoothed quasi-residuals turns out to be too optimistic and even the improved lower bound from (4.5) with $m = 25$ (lower dashed line) is only for a small number of steps much closer to the true behavior.

Finally we present the numerical results of a realistic example, $\mathbf{Ax} = \mathbf{b}$ with the matrix NOS6 from the MatrixMarket (`http://math.nist.gov/MatrixMarket/`) for $\mathbf{A}$ and a right-hand side $\mathbf{b}$ of all ones. NOS6 is a symmetric positive definite

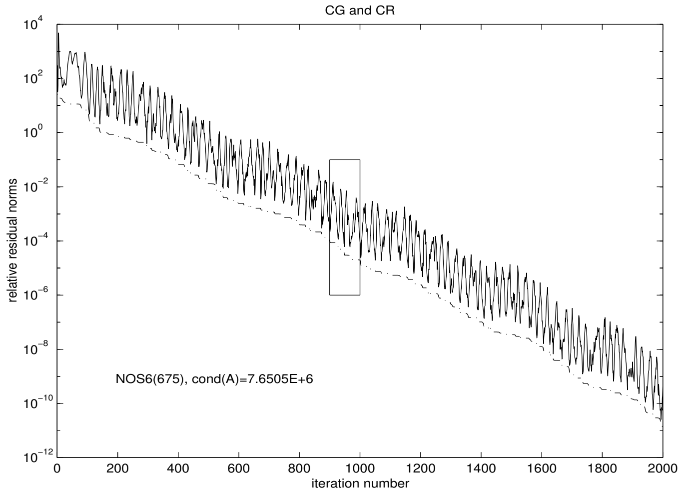

FIG. 8. *The residual norm history of CG applied to NOS6 with all ones as right-hand side (solid line) and the corresponding smoothed residuals, which can be viewed as CR residuals (dot-dashed line).*

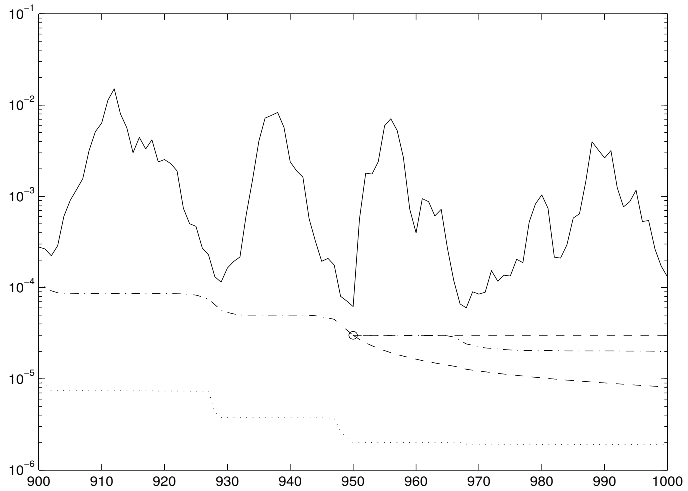

FIG. 9. *The window in Figure 8 enlarged. Shown are in addition to the residual norm histories of CG (solid line) and smoothed CG (dot-dashed line) the lower bound from* (4.3) *(dotted line) and the upper and lower bounds from* (4.5) *with* $m = 950$ *(upper and lower dashed lines).*

matrix of order 675 and condition number 7.6505×10^6. Figure 8 shows the residual norm histories of CG (solid line) and smoothed CG (equivalent to CR; dashed line) and specifies the window between the iterations 900 and 1000, which is enlarged in Figure 9. There, like in Figures 3, 5, and 7, we also show the lower bound given in (4.3) (dotted line) and the bounds from (4.5) with $m = 950$ (upper and lower dashed lines).

6. Conclusions. We have first shown by a simple proof that the iterates and residuals of a minimal residual method can be obtained from those of the corresponding orthogonal residual method by a three-term smoothing process, called MR smoothing, because in each step the action is restricted to a two-dimensional subspace. Pairs of such methods include CG and CR (or MinRes), FOM and GMRes, as well as CGNE and CGNR. For the first two pairs, the result was given by Weiss [26, 27]. The last pair is treated in the Appendix. The same process can also be applied in coordinate space, the coordinates being those of the residuals with respect to the basis consisting of the normalized residual vectors of a first or, primary, Krylov space solver. It yields then in general a QMR smoothing process, which generates the iterates and residuals of what we call a QMR-type method. The large class of these methods includes QMR, TFQMR, and QMRCGStab; their iterates and residuals can be found by QMR smoothing applied to BiCG, CGS, and BiCGStab, respectively, three cases that were treated by Zhou and Walker [29].

Then, from known relations between the norms of the residuals or their coordinate vectors (the quasi-residuals) of primary and smoothed methods, we have derived several estimates for what can be gained in the 2-norm of the residual or quasi-residual by the transition from the original to the smoothed method. The results have been illustrated by artificial, but archetypical examples, which allow us to specify bounds for more general cases, as well as by a realistic application of the CG method.

Appendix. The transition from CGNE to CGNR.

There are two well-known standard approaches for solving the normal equations by the conjugate gradient method: CGNR and CGNE, which correspond to applying CG (suitably adopted to the special situations) to

$$\mathbf{A}^\star\mathbf{A}\mathbf{x} = \mathbf{A}^\star\mathbf{b} \qquad \text{or} \qquad \mathbf{A}\mathbf{A}^\star\mathbf{z} = \mathbf{b} \quad \text{with} \quad \mathbf{x} \equiv: \mathbf{A}^\star\mathbf{z}\,, \tag{A.1}$$

respectively; see, for example, [8, 11, 18]. In both methods,

$$\mathbf{x}_n - \mathbf{x}_0 \in \mathcal{K}_n(\mathbf{A}^\star\mathbf{A}, \mathbf{A}^\star\mathbf{r}_0) = \text{span}\,(\mathbf{A}^\star\mathbf{r}_0, (\mathbf{A}^\star\mathbf{A})\mathbf{A}^\star\mathbf{r}_0, \dots, (\mathbf{A}^\star\mathbf{A})^{n-1}\mathbf{A}^\star\mathbf{r}_0) \tag{A.2}$$

and, hence, with $\mathbf{B} :\equiv \mathbf{A}\mathbf{A}^\star$,

$$\mathbf{r}_n - \mathbf{r}_0 \in \mathbf{A}\,\mathcal{K}_n(\mathbf{A}^\star\mathbf{A}, \mathbf{A}^\star\mathbf{r}_0) = \mathbf{A}\mathbf{A}^\star\,\mathcal{K}_n(\mathbf{A}\mathbf{A}^\star, \mathbf{r}_0) = \mathbf{B}\mathcal{K}_n(\mathbf{B}, \mathbf{r}_0)\,. \tag{A.3}$$

Recall that CG applied to a Hermitian positive definite (hpd) system $\mathbf{A}\mathbf{x} = \mathbf{b}$ yields iterates $\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0)$ that minimize the $\mathbf{A}$-norm of the error. Therefore, in CGNR, $\mathbf{x}_n$ minimizes the $\mathbf{A}^\star\mathbf{A}$-norm of the error $\mathbf{x}_n - (\mathbf{A}^\star\mathbf{A})^{-1}\mathbf{A}^\star\mathbf{b}$ of the first system in (A.1) subject to the condition (A.2). As is well-known, if $\mathbf{A}$ is nonsingular, this is the same as minimizing the 2-norm of the residual of the original system, $\|\mathbf{b} - \mathbf{A}\mathbf{x}_n\|$, subject to the condition (A.2). This, on the other hand, is equivalent to minimizing

$$\|\mathbf{b} - \mathbf{A}\mathbf{A}^\star\mathbf{z}_n\| = \|\mathbf{b} - \mathbf{B}\mathbf{z}_n\|\,, \tag{A.4}$$

that is, the 2-norm of the residual of the second system in (A.1), subject to

$$\mathbf{z}_n - \mathbf{z}_0 \in (\mathbf{A}^\star)^{-1}\mathcal{K}_n(\mathbf{A}^\star\mathbf{A}, \mathbf{A}^\star\mathbf{r}_0) = \mathcal{K}_n(\mathbf{A}\mathbf{A}^\star, \mathbf{r}_0) = \mathcal{K}_n(\mathbf{B}, \mathbf{r}_0)\,. \tag{A.5}$$

In CGNE, $\mathbf{x}_n$ minimizes the $\mathbf{A}\mathbf{A}^\star$-norm or $\mathbf{B}$-norm of the error $\mathbf{z}_n - (\mathbf{A}\mathbf{A}^\star)^{-1}\mathbf{b}$ of the second system in (A.1) subject to the condition (A.5). As is well-known, this is the same as minimizing the 2-norm of the error of the original system, $\|\mathbf{x}_n - \mathbf{A}^{-1}\mathbf{b}\|$, subject to the condition (A.2). But this fact will not be needed here.

We summarize these partly well-known results:

THEOREM 8. *The two algorithms CGNR and CGNE for applying the CG method to the two normal equations* (A.1) *that can be associated with a nonsingular system* $\mathbf{Ax} = \mathbf{b}$ *have, in exact arithmetic, the following properties:*

In CGNR, $\mathbf{x}_n = \mathbf{A}^\star \mathbf{z}_n$ *minimizes the 2-norm of the residual of the original system,* $\|\mathbf{b} - \mathbf{Ax}_n\|$, *subject to the condition* (A.2), *and* $\mathbf{z}_n$ *minimizes the 2-norm of the residual of*

$$\mathbf{Bz} = \mathbf{b}\,, \qquad \textit{where} \quad \mathbf{B} = \mathbf{AA}^\star\,, \tag{A.6}$$

that is, $\|\mathbf{b} - \mathbf{Bz}_n\|$, *subject to* (A.5). *In other words, it is equivalent to applying the CR method to* (A.6).

In CGNE, $\mathbf{x}_n = \mathbf{A}^\star \mathbf{z}_n$ *minimizes the 2-norm of the error of the original system,* $\|\mathbf{x}_n - \mathbf{A}^{-1}\mathbf{b}\|$, *subject to the condition* (A.2), *and* $\mathbf{z}_n$ *minimizes the B-norm of the error of* (A.6), *that is,* $\|\mathbf{z}_n - \mathbf{B}^{-1}\mathbf{b}\|_\mathbf{B}$, *subject to* (A.5).

COROLLARY 9. *The iterates* $\widetilde{\mathbf{x}}_n$ *of CGNR can be computed from the iterates* $\mathbf{x}_n = \mathbf{A}^\star \mathbf{z}_n$ *of CGNE by applying MR or QMR smoothing.*

Proof. If we consider CGNE and CGNR as methods for solving the hpd system (A.6), then CGNE is CG and CGNR is CR. The relevant residuals $\mathbf{r}_n$ and $\widetilde{\mathbf{r}}_n$ that are needed in (2.18) or (2.25) and in (3.19) (recall that (3.22) holds here), are just the ordinary residuals of the original system since $\mathbf{b} - \mathbf{Bz}_n = \mathbf{b} - \mathbf{Ax}_n$. □

REFERENCES

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

[2] T. F. CHAN, E. GALLOPOULOS, V. SIMONCINI, T. SZETO, AND C. H. TONG, *A quasi-minimal residual variant of the Bi-CGSTAB algorithm for nonsymmetric systems*, SIAM J. Sci. Comput., 15 (1994), pp. 338–347.

[3] J. CULLUM, *Peaks, plateaus, numerical instabilities in a Galerkin/minimal residual pair of methods for solving $Ax = b$*, Appl. Numer. Math., 19 (1995), pp. 255–278.

[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 in the theory of Krylov space methods*, Acta Numerica, 10 (2001). To appear.

[6] R. FLETCHER, *Conjugate gradient methods for indefinite systems*, in Numerical Analysis, Dundee, 1975, G. A. Watson, ed., vol. 506 of Lecture Notes in Mathematics, Springer, Berlin, 1976, pp. 73–89.

[7] R. W. FREUND, *A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems*, SIAM J. Sci. Comput., 14 (1993), pp. 470–482.

[8] R. W. FREUND, G. H. GOLUB, AND N. M. NACHTIGAL, *Iterative solution of linear systems*, Acta Numerica, (1992), pp. 57–100.

[9] R. W. FREUND, M. H. GUTKNECHT, AND N. M. NACHTIGAL, *An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices*, SIAM J. Sci. Comput., 14 (1993), pp. 137–158.

[10] R. W. FREUND AND N. M. NACHTIGAL, *QMR: a quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–339.

[11] A. GREENBAUM, *Estimating the attainable accuracy of recursively computed residual methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 535–551.

[12] M. H. GUTKNECHT, *A completed theory of the unsymmetric Lanczos process and related algorithms, Part I*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 594–639.

[13] ———, *Changing the norm in conjugate gradient type algorithms*, SIAM J. Numer. Anal., 30 (1993), pp. 40–56.

[14] ———, *A completed theory of the unsymmetric Lanczos process and related algorithms, Part II*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 15–58.

[15] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bureau Standards, 49 (1952), pp. 409–435.

[16] C. LANCZOS, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bureau Standards, 49 (1952), pp. 33–53.

[17] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
[18] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing, Boston, 1996.
[19] Y. Saad and M. H. Schultz, *Conjugate gradient-like algorithms for solving nonsymmetric linear systems*, Math. Comp., 44 (1985), pp. 417–424.
[20] H. Sadok, *Analysis of the convergence of the minimal and the orthogonal residual methods.* Preprint, 1997.
[21] W. Schönauer, *Scientific Computing on Vector Computers*, Elsevier, Amsterdam, 1987.
[22] P. Sonneveld, *CGS, a fast Lanczos-type solver for nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 36–52.
[23] E. Stiefel, *Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme*, Comm. Math. Helv., 29 (1955), pp. 157–179.
[24] H. A. van der Vorst, *Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 631–644.
[25] H. F. Walker, *Residual smoothing and peak/plateau behavior in Krylov subspace methods*, Appl. Numer. Math., 19 (1995), pp. 279–286.
[26] R. Weiss, *Convergence behavior of generalized conjugate gradient methods*, PhD thesis, University of Karlsruhe, 1990.
[27] ———, *Properties of generalized conjugate gradient methods*, J. Numer. Linear Algebra Appl., 1 (1994), pp. 45–63.
[28] ———, *Parameter-Free Iterative Linear Solvers*, vol. 97 of Mathematical Research, Akademie Verlag, Berlin, 1996.
[29] L. Zhou and H. F. Walker, *Residual smoothing techniques for iterative methods*, SIAM J. Sci. Comput., 15 (1994), pp. 297–312.