Universität Karlsruhe
Rechenzentrum
Postfach 6980
76128 Karlsruhe
Germany

Interner Bericht Nr. 53/94

Relations between Smoothing and QMR

Abteilung "Numerikforschung für Supercomputer"
Rechenzentrum der Universität Karlsruhe
January 1994
Author: Rüdiger Weiss

RELATIONS BETWEEN SMOOTHING AND QMR

RÜDIGER WEISS
NUMERIKFORSCHUNG FÜR SUPERCOMPUTER
RECHENZENTRUM UNIVERSITÄT KARLSRUHE
POSTFACH 6980
D-76128 KARLSRUHE
GERMANY
E-MAIL: WEISS@RZ.UNI-KARLSRUHE.DE

Summary. It is known from practice that the norms of the residuals of the biconjugate gradients smoothed by Schönauer's residual-minimizing algorithm and of Freund's and Nachtigal's QMR method are very close. However, a proof for this behavior has been still missing. Here the same bounds for the residuals of Schönauer's smoothing applied to biconjugate gradients as for QMR are given. Therefrom result estimates for the relation between the residuals of both techniques, confirming theoretically close interconnections.

Mathematics Subject Classification (1991): 65F10

1. Background. The norm of the residuals of generalized conjugate gradient methods may heavily oscillate. This fact is known from practice. Therefore, a smoothing algorithm was introduced by Schönauer [7, 8] in order to get a norm non-increasing function of the index for the residuals. In each iteration step a one-dimensional minimization is performed. However, the concept of smoothing is quite more general. Quasi-residual smoothing was introduced initially by Freund and Nachtigal [3] without showing the connection to smoothing. It was analyzed very elaborately by Zhou and Walker [12] who made the connection to smoothing and thus created the name of this kind of smoothing in analogy to Freund's and Nachtigal's QMR algorithm.

Definition 1. *Let x_k be an approximation resulting from an iterative method for the solution of the linear system $Ax = b$ and let $r_k = Ax_k - b$ be the residual. The sequence*

$$s_0 = r_0 \ , z_0 = x_0 \ , \tag{1}$$

$$s_k = s_{k-1} + \gamma_k(r_k - s_{k-1}) \ , \tag{2}$$

$$z_k = z_{k-1} + \gamma_k(x_k - z_{k-1}) \tag{3}$$

is a corresponding smoothed method delivering approximations z_k and residuals s_k.

The smoothed sequence is a mean of the actual element of the original sequence and the previous element of the smoothed sequence. In general the smoothed sequence does not react on the original sequence. Different smoothing techniques vary by the determination of the coefficients γ_k. We do not address in the following the problem of stable implementations for smoothing because these implementations are mathematically equivalent and we refer to [12] for this topic.

1985	Schönauer [8]:	oscillating residual	$\rightarrow$	monot. decr. res.
1990	Weiss [10]:	pseudo-res.-min. cg	$\rightarrow$	res.-min. cg
1991	Gutknecht [4]:	res.-min. cg	$\rightarrow$	pseudo-res.-min. cg
1992	Zhou & Walker [12]:	Biconjugate gradients	$\rightarrow$	QMR
	Zhou & Walker [12]:	CGS	$\rightarrow$	TFQMR
1993	Weiss [11]:	oscillating error	$\rightarrow$	monot. decr. error (GMERR)

TABLE 1
Transformation of methods by smoothing

Since Schönauer's development [8] a lot of theoretical investigations have been performed. In general smoothing transforms iterative methods. In table 1 different transformations are given.

In [10] smoothing was investigated for basic methods with orthogonal residuals. For example ORTHORES is transformed to GMRES and CGNE to CGNR by smoothing. The further main results in [10] are briefly summarized. Let Z be a positive definite matrix and $||y||_Z = \sqrt{y^T Z y}$ for any vector y. If r_k is Z-orthogonal to r_i for $i = 0, \ldots, k-1$, i. e.

$$r_k^T Z r_i = 0 \ ,$$

then the following equations are equivalent

$$||s_k||_Z = min_{\gamma_k} ||s_{k-1} + \gamma_k (r_k - s_{k-1})||_Z \ , \tag{4}$$

$$||s_k||_Z = min_{\gamma_1,\ldots,\gamma_k} ||s_{k-1} + \sum_{i=1}^{k} \gamma_i (r_i - s_{i-1})||_Z \ , \tag{5}$$

$$s_k = \frac{1}{\sum_{i=0}^{k} \frac{1}{||r_i||_Z^2}} \sum_{i=0}^{k} \frac{1}{||r_i||_Z^2} r_i \ , \tag{6}$$

$$\frac{1}{||s_k||_Z^2} = \sum_{i=0}^{k} \frac{1}{||r_i||_Z^2} \ . \tag{7}$$

The meaning of (5) is that the smoothing technique is equivalent to a k-dimensional minimization, though only a 1-dimensional minimization is performed in each step resulting from (4). We call (7) Kirchhoff's rule because the norm of the smoothed sequence squared behaves like the global resistance of parallel-connected resistances in an electrical network. By (7) we also get the norm of the unique minimizer:

$$||s_k||_Z \ = \ \frac{1}{\sqrt{\sum_{i=0}^{k} \frac{1}{||r_i||_Z^2}}} \ . \tag{8}$$

If the residuals of the original sequence are in a Krylov space, i. e.

$$r_k = \sum_{i=1}^{k} \alpha_i A^i r_0 + r_0 \ ,$$

then

$$(9) \qquad ||s_k||_Z = \min_{\eta_1,\ldots,\eta_k} ||\sum_{i=1}^{k} \eta_i A^i r_0 + r_0||_Z \ .$$

2. Smoothing applied to non-orthogonal residuals. In the following we will consider only the case $Z = I$ because it is the most important from a practical viewpoint and the extension to arbitrary matrices Z is quite simple. If the r_i are non-orthogonal, then the equations (4) – (7) are not equivalent and smoothing can be applied by using (4) or (6) for determining γ_k.

If we choose (4), then the technique is called minimal residual (MR) smoothing. This is the original algorithm of Schönauer [7]. We get

$$(10) \qquad \gamma_k^{MR} = -\frac{\left(s_{k-1}^{MR}\right)^T (r_k - s_{k-1}^{MR})}{||r_k - s_{k-1}^{MR}||^2} \ .$$

Let us denote by s^{MR} the residual of such a method.

$$(11) \qquad s_k^{MR} = s_{k-1}^{MR} + \gamma_k^{MR}(r_k - s_{k-1}^{MR}) \ .$$

The residual is a function of the iteration index, with a non-increasing norm, i. e.

$$(12) \qquad ||s_k^{MR}|| \leq ||s_{k-1}^{MR}|| \ ,$$

and

$$(13) \qquad ||s_k^{MR}|| \leq ||r_k|| \ .$$

This sort of smoothing is locally optimal, because the residual is minimized in each step. Practical experiences show that the convergence behavior is quite good. However, very little has been known about the global behavior. We can prove the following estimate for the convergence that is the same as for QMR.

THEOREM 2.

$$(14) \qquad ||s_k^{MR}|| \leq \sqrt{k+1} \, \frac{1}{\sqrt{\sum_{i=0}^{k} \frac{1}{||r_i||^2}}} \ .$$

If the original sequence is generated by biconjugate gradients, i. e.

$$(15) \qquad x_k = x_0 + V_k \bar{z} \ ,$$

$$(16) \qquad r_k = r_0 + A V_k \bar{z} \ ,$$

with $\bar{z} \in \mathbb{R}^k$ *and* $V_k = (\frac{r_0}{||r_0||}, \ldots, \frac{r_{k-1}}{||r_{k-1}||})$ *and*

$$(17) \qquad A V_k = V_{k+1} H_k \ ,$$

where $H_k \in I\!R^{(k+1)\times k}$ is the tridiagonal matrix resulting from the non-symmetric Lanczos process, or a block tridiagonal matrix resulting from a look-ahead Lanczos process to avoid breakdowns, then

$$||s_k^{MR}|| \;\leq\; \sqrt{k+1}\;\min_{z_1,\ldots,z_k} ||\; ||r_0||e_1 - H_k z|| \;, \tag{18}$$

where e_1 is the first unit vector and $z = (z_1, \ldots, z_k)^T \in I\!R^k$.

Proof. By (12) and (13)

$$\begin{aligned} ||s_k^{MR}|| &\leq \min_{i=0,\ldots,k} ||r_i|| \\ &= \frac{1}{\sum_{i=0}^k \frac{1}{||r_i||^2}} \sum_{i=0}^k \frac{1}{||r_i||^2} \min_{i=0,\ldots,k} ||r_i|| \\ &\leq \frac{1}{\sum_{i=0}^k \frac{1}{||r_i||^2}} \sum_{i=0}^k \frac{1}{||r_i||^2} ||r_i|| \\ &= \frac{1}{\sum_{i=0}^k \frac{1}{||r_i||^2}} \sum_{i=0}^k \frac{1}{||r_i||} \\ &\leq \frac{1}{\sum_{i=0}^k \frac{1}{||r_i||^2}} \sqrt{k+1} \sqrt{\sum_{i=0}^k \frac{1}{||r_i||^2}} \\ &= \sqrt{k+1}\; \frac{1}{\sqrt{\sum_{i=0}^k \frac{1}{||r_i||^2}}} \;. \end{aligned}$$

The equivalence of the right-hand sides of (14) and (18) for the biconjugate gradients is proven in [12]. □

Note that also the minimum of the original residuals fulfils (14) and (18) resulting from the proof of Theorem 2. Thus this minimum would be a serious smoother, but worse than minimal residual smoothing.

The idea of quasi-minimal residual smoothing [12] is to satisfy (6). This can be realized as follows.

$$\tau_0^2 = ||r_0||^2 \;.$$

Calculate for $k \geq 1$:

$$\begin{aligned} \tau_k^2 &= \frac{1}{\frac{1}{\tau_{k-1}^2} + \frac{1}{||r_k||^2}} \;, \\ \gamma_k^{QMR} &= \frac{\tau_k^2}{||r_k||^2} \;. \end{aligned}$$

For the residual we get

$$s_k^{QMR} = s_{k-1}^{QMR} + \gamma_k^{QMR}(r_k - s_{k-1}^{QMR}) \;. \tag{19}$$

It is obvious that

$$0 \leq \gamma_k^{QMR} \leq 1 \ . \tag{20}$$

In [12] the convergence was investigated and it was proven for the biconjugate gradients and CGS as basic methods:

$$||s_k^{QMR}|| \ \leq \ \sqrt{k+1} \ \frac{1}{\sqrt{\sum_{i=0}^{k} \frac{1}{||r_i||^2}}} \ . \tag{21}$$

(21) also holds in general by an analoguous proof as in Theorem 2. As in [12] it is shown that QMR is a quasi-minimal smoothed biconjugate gradient algorithm, it follows from the original paper on QMR [3]: If the original sequence is generated by biconjugate gradients, i. e.

$$x_k = x_0 + V_k \bar{z} \tag{22}$$

$$r_k = r_0 + AV_k \bar{z} \tag{23}$$

with $\bar{z} \in \mathbb{R}^k$ and $V_k = (\frac{r_0}{||r_0||}, \ldots, \frac{r_{k-1}}{||r_{k-1}||})$ and

$$AV_k = V_{k+1} H_k \ , \tag{24}$$

where $H_k \in \mathbb{R}^{(k+1)\times k}$ is the tridiagonal matrix resulting from the non-symmetric Lanczos process, or a block tridiagonal matrix resulting from a look-ahead Lanczos process to avoid breakdowns, then

$$||s_k^{QMR}|| \ \leq \ \sqrt{k+1} \ \min_{z_1,\ldots,z_k} || \ ||r_0||e_1 - H_k z|| \ , \tag{25}$$

where e_1 is the first unit vector and $z = (z_1, \ldots, z_k)^T \in \mathbb{R}^k$. Note the equivalence of (25) and (21) with Theorem 2. As a result the global behavior of quasi-minimal residual smoothing is known, but still very little is known theoretically about the local behavior, except heuristics.

Zhou and Walker [12] show that quasi-minimal residual smoothing transforms distinct cg methods to QMR methods. It transforms biconjugate gradients [1, 5] to QMR [3]. Moreover quasi-minimal residual smoothing transforms an auxiliary sequence based on CGS [9] to TFQMR [2].

3. Comparison of minimal residual smoothing with QMR. Numerical tests of Zhou and Walker [12] show that quasi-minimal residual smoothing and minimal residual smoothing applied to the biconjugate gradients or CGS gives quite similar results.

Indeed the estimates (18) and (25) are the same and of course the residuals of both techniques are greater than the residuals of the optimal minimizer in the Krylov space, GMRES [6], i. e.

$$||r_k^{GMRES}|| \leq \quad ||s_k^{MR}|| \quad \leq \sqrt{k+1} \min_{z_1,\ldots,z_k} || \ ||r_0||e_1 - H_k z|| \ , \tag{26}$$

$$||r_k^{GMRES}|| \leq \quad ||s_k^{QMR}|| \quad \leq \sqrt{k+1} \min_{z_1,\ldots,z_k} || \ ||r_0||e_1 - H_k z|| \ . \tag{27}$$

Thus we have got the same lower and upper bounds for both techniques. From these inequalities we get a relation for the two techniques:

THEOREM 3. *If* $V_n = (\frac{r_0}{||r_0||}, \ldots, \frac{r_n}{||r_n||})$ *is invertible, then for biconjugate gradients as basic method*

$$||s_k^{MR}|| \leq \sqrt{k+1}\ ||V_n^{-1}||\ ||s_k^{QMR}|| \ , \tag{28}$$

$$||s_k^{QMR}|| \leq \sqrt{k+1}\ ||V_n^{-1}||\ ||s_k^{MR}|| \tag{29}$$

is satisfied.

Proof. Let $\tilde{H}_n \in I\!R^{n\times n}$ be the tridiagonal matrix resulting from the non-symmetric Lanczos process in the step n with

$$AV_n = V_n \tilde{H}_n \ ,$$

$\tilde{z} = (z_1, \ldots, z_k, 0, \ldots, 0)^T \in I\!R^n$ and $\tilde{e}_1$ the first unit vector in $I\!R^n$, then

$$\begin{aligned} \min_{z_1,\ldots,z_k} ||\ ||r_0||e_1 - H_k z|| &= \min_{z_1,\ldots,z_k} ||\ ||r_0||\tilde{e}_1 - \tilde{H}_n \tilde{z}|| \\ &= \min_{z_1,\ldots,z_k} ||V_n^{-1} V_n (||r_0||\tilde{e}_1 - \tilde{H}_n \tilde{z})|| \\ &= \min_{z_1,\ldots,z_k} ||V_n^{-1}(r_0 - AV_n \tilde{z})|| \\ &\leq ||V_n^{-1}(r_0 - AV_n \hat{z})|| \ , \end{aligned}$$

where $\hat{z} = (\hat{z}_1, \ldots, \hat{z}_k, 0, \ldots, 0)^T$ is chosen so that

$$||r_0 - AV_n \hat{z}|| \ = \ \min_{z_1,\ldots,z_k} ||r_0 - AV_k z|| = ||r_k^{GMRES}|| \ .$$

Thus

$$\begin{aligned} \min_{z_1,\ldots,z_k} ||\ ||r_0||e_1 - H_k z|| &\leq ||V_n^{-1}(r_0 - AV_n \hat{z})|| \\ &\leq ||V_n^{-1}||\ ||r_0 - AV_n \hat{z}|| \\ &= ||V_n^{-1}||\ ||r_k^{GMRES}|| \ . \end{aligned}$$

By comparing (26) with (27) and by inserting the last inequality we obtain the statements. □.

Thus the norm of the residuals of minimal residual and quasi-minimal residual smoothing are closely connected, though $||V_n^{-1}||$ cannot be estimated. $\sqrt{k+1}$ is slowly growing and it seems to overestimate the true bound in general as can be seen from tests.

Both techniques were tested for a rough model of the 3-dimensional Navier-Stokes equations:

$$\Delta v + v + R(v^T \nabla) v = h \ , \tag{30}$$

with $v = (v_1, v_2, v_3)^T$, $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})^T$, $\Delta = \nabla^T \nabla$. The calculations have been performed on a $20 \times 20 \times 20$ grid with Dirichlet boundary conditions on the unit cube.

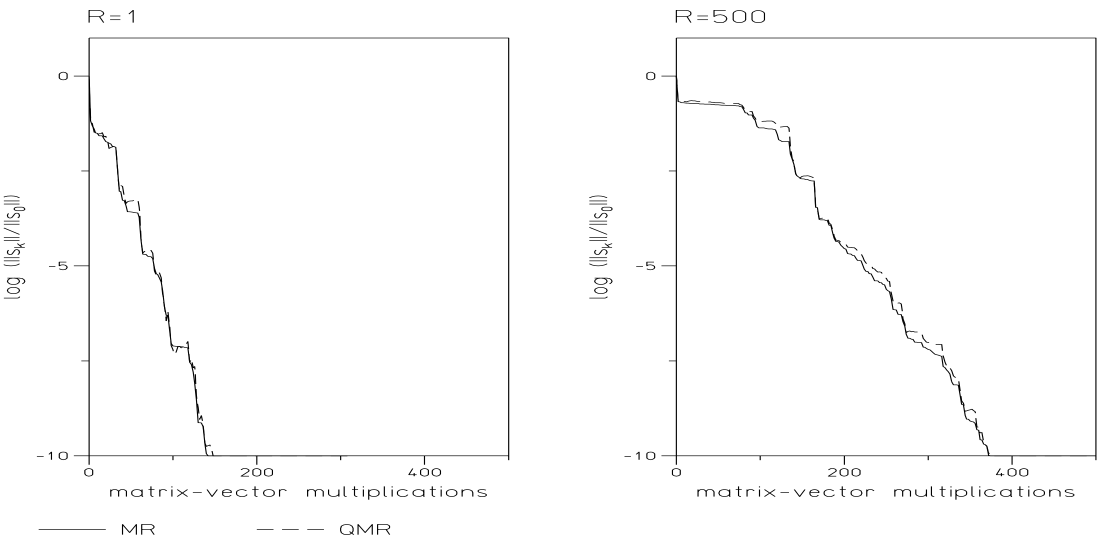

FIG. 1. *Residuals of MR-smoothing for biconjugate gradients and QMR for R=1 and R=2000*

The right-hand side is determined so that equation (30) has trigonometric functions as exact solution. The linear system arises from a finite difference discretization with consistency order 4 and from the linearization in the first Newton step. The matrix is normalized, i.e. every row is divided by the sum of the absolute entries in that row and all diagonal entries have a positive sign. The parameter R simulates a Reynolds number. For increasing R the imaginary part of the eigenvalues of the matrix increases and for $R > 500$ the matrix becomes indefinite; see [10] for a detailed investigation on a coarser grid. We tested $R = 1$, $R = 500$ (see figure 1), $R = 2000$ and $R = 10^5$ (see figure 2). The difference of the two smoothing techniques is far less than the theoretical bound. Mostly minimal residual smoothing is slightly better than quasi-minimal residual smoothing, but not always. Minimal residual smoothing seems to be more robust because it converges better for high values of R as for example $R = 2000$ where the techniques come close to the bound of non-convergence. For $R = 10^5$ both techniques cannot compensate the bad behavior of the underlying biconjugate gradients and do not converge.

4. Conclusion. For orthogonal residuals of the original sequence minimal and quasi-minimal residual smoothing coincide. For arbitrary relations between the residuals of the basic technique the same bounds for minimal residual smoothing were proven as for QMR. These estimates establish the possibility to compare the techniques theoretically. It turns out that both techniques are closely connected. Practical tests show that they perform indeed quite similar. It does not matter which one of the both techniques is applied, but it seems to be important to use smoothing at all in order to get a smooth decrease of the residuals. We think that all kinds of smoothing are closely related and that no further progress in convergence acceleration for the residuals will be achieved by other smoothing approaches. The situation may be different for the decrease of the errors. A minimization of the energy norm or the Euclidean norm of

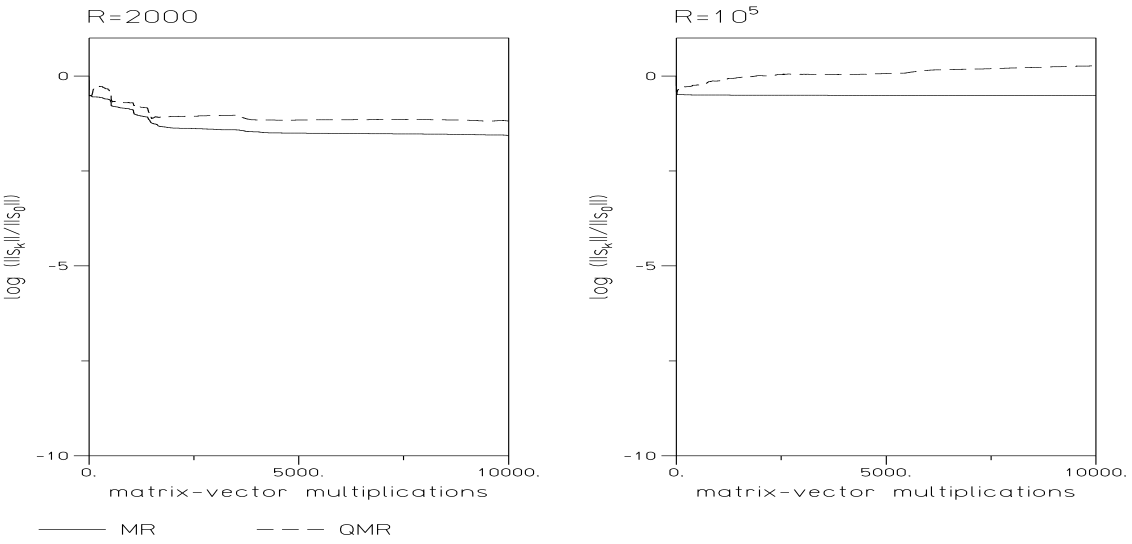

FIG. 2. *Residuals of MR-smoothing for biconjugate gradients and QMR for R=2000 and R=*10^5

the error by smoothing as proposed in [11] may be advantageous.

REFERENCES

[1] R. Fletcher. Conjugate gradient methods for indefinite systems. In G.A. Watson, editor, *Proc. Dundee Biennial Conf. on Num. Anal.*, pages 73–89. Springer-Verlag, 1975.

[2] R.W. Freund. A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems. *SIAM J. Sci. Comput.*, 14(2):470–482, 1993.

[3] R.W. Freund and N.M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numer. Math.*, 60:315–339, 1991.

[4] M.H. Gutknecht. Changing the norm in conjugate gradient type algorithms. *SIAM J. Numer. Anal.*, 30(1):40–56, 1993.

[5] C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bur. Standards*, 49:33–53, 1952.

[6] Y. Saad and M.H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 7(3):856–869, 1986.

[7] W. Schönauer. *Scientific Computing on Vector Computers.* North-Holland, Amsterdam, New York, Oxford, Tokyo, 1987.

[8] W. Schönauer, H. Müller, and E. Schnepf. Pseudo-residual type methods for the iterative solution of large linear systems on vector computers. In M. Feilmeier, J. Joubert, and U. Schendel, editors, *Parallel Computing 85*, pages 193–198. Elsevier Science Publishers B.V., North-Holland, 1986.

[9] P. Sonneveld. CGS, a fast Lanczos-type solver for nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 10(1):36–52, 1989.

[10] R. Weiss. Convergence behavior of generalized conjugate gradient methods. Internal report 43/90, University of Karlsruhe, Computing Center, 1990.

[11] R. Weiss. Error-minimizing Krylov subspace methods. *SIAM J. Sci. Comput.*, 15(3):511–527, 1994.

[12] L. Zhou and H.F. Walker. Residual smoothing techniques for iterative methods. *SIAM J. Sci. Comput.*, 15(2):297–312, 1994.