A NEW KRYLOV-SUBSPACE METHOD FOR SYMMETRIC INDEFINITE LINEAR SYSTEMS

Roland W. Freund*
AT&T Bell Laboratories
Room 2C–420
600 Mountain Avenue
Murray Hill, NJ 07974–0636
freund@research.att.com

Noël M. Nachtigal†
Mathematical Sciences Section
Oak Ridge National Laboratory
P.O. Box 2008, Building 6012
Oak Ridge, TN 37831–6367
santa@msr.epm.ornl.gov

Abstract

Many important applications involve the solution of large linear systems with symmetric, but indefinite coefficient matrices. For example, such systems arise in incompressible flow computations and as subproblems in optimization algorithms for linear and nonlinear programs. Existing Krylov-subspace iterations for symmetric indefinite systems, such as SYMMLQ and MINRES, require the use of symmetric positive definite preconditioners, which is a rather unnatural restriction when the matrix itself is highly indefinite with both many positive and many negative eigenvalues. In this note, we describe a new Krylov-subspace iteration for solving symmetric indefinite linear systems that can be combined with arbitrary symmetric preconditioners. The algorithm can be interpreted as a special case of the quasi-minimal residual method for general non-Hermitian linear systems, and like the latter, it produces iterates defined by a quasi-minimal residual property. The proposed method has the same work and storage requirements per iteration as SYMMLQ or MINRES, however, it usually converges in considerably fewer iterations. Results of numerical experiments are reported.

1 Introduction

The solution of large linear systems

$$Ax = b, \tag{1}$$

where $A \in \mathbb{R}^{N\times N}$ is symmetric, but indefinite, is required in many important applications. Often the matrix A is *highly indefinite*, i.e., it has both many positive and many negative eigenvalues. For example,

*Part of this research was performed while this author was in residence at the Research Institute for Advanced Computer Science (RIACS), NASA Ames Research Center, Moffett Field, CA 94035, and it was supported in part by Cooperative Agreement NCC 2-387 between the National Aeronautics and Space Administration and the Universities Space Research Association.

†The research of this author was supported in part by the Applied Mathematical Sciences Research Program, Office of Energy Research, U.S. Department of Energy, under contract DE–AC05–840R21400 with the Martin Marietta Energy Systems, Inc.

consider the Stokes problem of determining the velocity field $\mathbf{u}$ and pressure p of an incompressible fluid in a flow region $\Omega \subset \mathbb{R}^d$, $d = 2, 3$. The flow is described by the coupled partial differential equations

$$\begin{aligned} -\nu \nabla^2 \mathbf{u} + \operatorname{grad} p &= \mathbf{f} \quad \text{in } \Omega, \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega, \\ &+ \quad \text{appropriate boundary conditions,} \end{aligned} \tag{2}$$

where $\nu > 0$ denotes the viscosity of the fluid. Finite-difference or finite-element approximations of (2), possibly using stabilized discretizations, lead to linear systems (1) with symmetric matrices A of the form

$$A = \begin{bmatrix} B & C^T \\ C & -D \end{bmatrix}, \quad B = B^T \in \mathbb{R}^m, \quad D = D^T \in \mathbb{R}^{N-m}; \tag{3}$$

see, e.g., [10, 14]. The matrix B in (3) represents the discretization of the term $-\nu \nabla^2 \mathbf{u}$ in (2), and thus B is positive definite. Furthermore, the matrix C in (3) usually has full row rank, and D is a positive semi-definite matrix. These properties imply that A is an indefinite matrix with m positive and $N - m$ negative eigenvalues. Note that $N - m \approx m/d$, and therefore, the matrix A is highly indefinite.

Linear systems (1) with coefficient matrices of the form (3) (often with C the zero matrix) also arise as subproblems in optimization algorithms, such as interior-point methods for linear and nonlinear programs, or sequential programming methods for constrained nonlinear programs; see, e.g., [4, 9]. Again, the matrix A is typically highly indefinite.

The standard conjugate gradient-type Krylov-subspace algorithms for the iterative solution of symmetric indefinite linear systems are SYMMLQ and MINRES due to Paige and Saunders [13]. Both SYMMLQ and MINRES are based on the Lanczos process for symmetric matrices. Consequently, when preconditioning is used, then the coefficient matrix of the preconditioned systems needs to be symmetric. This condition implies that the preconditioner itself needs to be a symmetric positive definite matrix. This restriction for the choice of possible preconditioners for SYMMLQ and MINRES is rather unnatural when the coefficient matrix itself is highly indefinite. In this note, we describe a new iterative method for solving symmetric indefinite linear systems that can be combined with arbitrary symmetric preconditioners. The algorithm can be interpreted as a special case of the quasi-minimal residual (QMR) method [5, 6] for general non-Hermitian linear systems, and like the latter, it generates iterates defined by a quasi-minimal residual property. The proposed method has the same work and storage requirements per iteration as SYMMLQ or MINRES, however, it usually converges in considerably fewer iterations.

2 The QMR Method for Non-Hermitian Linear Systems

In this section, we briefly recall the QMR method [5, 6] for the solution of linear systems (1), where $A \in \mathbb{C}^{N \times N}$ is now allowed to be any nonsingular, in general non-Hermitian, matrix.

Let $x_0 \in \mathbb{C}^N$ be an arbitrary initial guess for (1), let $r_0 := b - Ax_0$ be the associated residual vector, and set $v_1 = r_0 / \|r_0\|_2$. Moreover, let $w_1 \in \mathbb{C}^N$ be an arbitrary vector with $\|w_1\|_2 = 1$. The original QMR algorithm proposed in [5] uses the look-ahead variant [3] of the classical Lanczos process [11] to generate two sequences of so-called Lanczos vectors

$$\{v_j\}_{j=1}^n \quad \text{and} \quad \{w_j\}_{j=1}^n, \quad n = 1, 2, \ldots,$$

that span the Krylov subspaces

$$K_n(A, r_0) := \operatorname{span}\{r_0, Ar_0, \ldots, A^{n-1} r_0\} \quad \text{and} \quad K_n(A^T, w_1) := \operatorname{span}\{w_1, A^T w_1, \ldots, (A^T)^{n-1} w_1\},$$

respectively. These vectors are constructed to be block-biorthogonal; for example, if no look-ahead steps occur, then they are even biorthogonal, i.e.,

$$w_j^T v_k = 0 \quad \text{for all} \quad j \neq k = 1, 2, \ldots, n.$$

We mention that, in our implementation of the look-ahead Lanczos algorithm, the Lanczos vectors are always scaled to unit length:

$$\|v_n\|_2 = \|w_n\|_2 = 1 \quad \text{for all} \quad n.$$

The Lanczos vectors can be generated by means of short vector recurrences. For the *right* Lanczos vectors $\{v_j\}_{j=1}^{n+1}$, these recursions can be summarized compactly in the relation

$$AV_n = V_{n+1} T_n. \tag{4}$$

Here, $V_n := [\, v_1 \quad v_2 \quad \cdots \quad v_n \,]$, and T_n is an $(n+1) \times n$ matrix whose entries are the coefficients of the Lanczos recursions. The matrix T_n is both block tridiagonal and upper Hessenberg, and it always has full rank n. Setting $W_n := [\, w_1 \quad w_2 \quad \cdots \quad w_n \,]$, he recurrences for the *left* Lanczos vectors $\{w_j\}_{j=1}^{n+1}$ can be summarized in the relation

$$A^T W_n = W_{n+1} \tilde{T}_n.$$

Here, $\tilde{T}_n$ is again a block tridiagonal $(n+1) \times n$ matrix that is also upper Hessenberg.

The QMR method is a Krylov-subspace iteration, i.e., it produces a sequence of approximations to the solution of (1) of the form

$$x_n \in x_0 + K_n(A, r_0), \quad n = 1, 2, \ldots .$$

Since the right Lanczos vectors $\{v_j\}_{j=1}^{n}$ span $K_n(A, r_0)$, any possible iterate can be written as

$$x_n = x_0 + V_n z_n, \quad \text{where} \quad z_n \in \mathbb{C}^n,$$

and z_n is a free parameter vector. In QMR, z_n is chosen as the (unique) solution of the least-squares problem

$$\|f_n - T_n z_n\|_2 = \min_{z \in \mathbb{C}^n} \|f_n - T_n z\|_2, \quad \text{where} \quad f_n := [\, \|r_0\|_2 \quad 0 \quad \cdots \quad 0 \,]^T \in \mathbb{R}^{n+1}. \tag{5}$$

We remark that, by (4), the residual vector of the nth QMR iterate x_n is given by

$$b - Ax_n = V_{n+1} \left(f_n - T_n z_n \right). \tag{6}$$

Hence, in view of (5), the QMR iterate x_n is characterized by a minimization of the second factor in the representation (6) of its residual vector. This is called a *quasi-minimal residual* (QMR) property, and the resulting iterative scheme for solving linear systems (1) is the QMR method. The solution of the least-squares problem (5) and the QMR iterates can be easily updated from step to step, using only short vector updates; see [5]. It is also possible to replace the three-term look-ahead Lanczos algorithm by a coupled two-term look-ahead Lanczos algorithm; the resulting implementation of QMR is discussed in [6]. Finally, a very simple no-look-ahead version of the coupled two-term QMR algorithm was derived by Freund and Szeto in [7]. They showed that the no-look-ahead QMR iterates can be obtained from the classical biconjugate gradient (BCG) algorithm [12] by performing one additional vector update and a few scalar updates at each BCG iteration. The resulting algorithm is called “QMR-from-BCG”.

3 Simplification of the Lanczos Process

The Lanczos process requires matrix-vector products $A \cdot v$ for the construction of the right Lanczos vectors, as well as products $A^T \cdot w$ with the transpose A^T for the construction of the left Lanczos vectors. Freund and Zha [2, 8] observed that the Lanczos process (and thus the QMR method) simplifies if one knows a nonsingular matrix P with

$$A^T P = PA, \tag{7}$$

and if the initial vector w_1 is chosen as $w_1 = Pv_1 / \|Pv_1\|_2$. In fact, it can be shown that the Lanczos vectors satisfy

$$w_n = \frac{Pv_n}{\|Pv_n\|_2} \quad \text{for all} \quad n.$$

Therefore, each matrix-vector product with A^T can be replaced by a matrix-vector multiplication with P. In particular, if the computation of products with P is cheap, then the resulting simplified Lanczos algorithm, as well as the QMR method, requires only half the work and storage of the general Lanczos process. For example, if P is a sparse matrix, then computing products with P is cheap. There are also important cases of dense matrices for which matrix-vector products $P \cdot v$ can be computed efficiently; for instance, this is the case if P is a Toeplitz matrix, see [8].

4 The Symmetric QMR Algorithm

We now return to linear systems (1) with symmetric indefinite coefficient matrices $A \in \mathbb{R}^{N \times N}$. Let $M = M^T \in \mathbb{R}^{N \times N}$ be an arbitrary nonsingular matrix that we want to use as a preconditioner for (1). To allow for left, right, or two-sided preconditioning, we write the preconditioner M in the form $M = M_1 \cdot M_2$. We stress that M_1 and M_2 are arbitrary matrices such that

$$M = M_1 M_2 = M_2^T M_1^T = M^T,$$

and M_1 and M_2 need not be the transpose of each other. Then, instead of (1), we solve the equivalent preconditioned system

$$A'x' = b', \quad \text{where} \quad A' = M_1^{-1} A M_2^{-1}, \quad x' = M_2 x, \quad \text{and} \quad b' = M_1^{-1} b. \tag{8}$$

Note that $M_1 = I$ for right preconditioning, and $M_2 = I$ for left preconditioning. We now set $P := M_1^T M_2^{-1}$. Using $A = A^T$ and $M_1 M_2 = M_2^T M_1^T$, one readily verifies that

$$(A')^T P = M_2^{-T} A M_2^{-1} = M_1^T M_2^{-1} M_1^{-1} A M_2^{-1} = PA',$$

i.e., the matrix A' satisfies the relation (7). Consequently, the QMR algorithm applied to (8) simplifies. Furthermore, by rewriting the resulting QMR algorithm in terms of the original system (1), it can be seen that the multiplications with P need not be done explicitly. In fact, the matrices M_1 and M_2 appear in the algorithm only in the usual preconditioning solves of systems of the form $M_1 r = t$ and $M_2 u = t$. We refer to the resulting QMR method for symmetric indefinite systems (1) with symmetric preconditioners M as the *symmetric* QMR algorithm.

A simple no-look-ahead version of the symmetric QMR algorithm derived from Freund and Szeto's QMR-from-BCG can be stated as follows.

Algorithm 1 (Symmetric QMR without look-ahead for $Ax = b$ with preconditioner $M = M_1M_2$)

0) *Choose* $x_0 \in \mathbb{R}^N$.

Set $r_0 = b - Ax_0$, $t = M_1^{-1}r_0$, $\tau_0 = \|t\|_2$, $q_0 = M_2^{-1}t$, $\vartheta_0 = 0$, *and* $\rho_0 = r_0^T q_0$.

For $n = 1, 2, \ldots,$ *do* :

1) *Compute*

$$t = Aq_{n-1} \quad \textit{and} \quad \sigma_{n-1} = q_{n-1}^T t.$$

If $\sigma_{n-1} = 0$, *then stop.*

Otherwise, set

$$\alpha_{n-1} = \frac{\rho_{n-1}}{\sigma_{n-1}} \quad \textit{and} \quad r_n = r_{n-1} - \alpha_{n-1}t.$$

2) *Set*

$$t = M_1^{-1}r_n, \quad \vartheta_n = \frac{\|t\|_2}{\tau_{n-1}}, \quad c_n = \frac{1}{\sqrt{1+\vartheta_n^2}}, \quad \tau_n = \tau_{n-1}\vartheta_n c_n,$$

$$d_n = c_n^2\vartheta_{n-1}^2 d_{n-1} + c_n^2\alpha_{n-1}q_{n-1}, \quad \textit{and} \quad x_n = x_{n-1} + d_n.$$

If x_n *has converged, then stop.*

3) *If* $\rho_{n-1} = 0$, *then stop.*

Otherwise, set

$$u_n = M_2^{-1}t, \quad \rho_n = r_n^T u_n, \quad \beta_n = \frac{\rho_n}{\rho_{n-1}}, \quad \textit{and} \quad q_n = u_n + \beta_n q_{n-1}.$$

We remark that, in Algorithm 1, r_n denotes the BCG residual, and not the residual vector associated with the nth QMR iterates x_n. Furthermore, $u_n = M^{-1}r_n = M_2^{-1}\left(M_1^{-1}r_n\right)$ is the preconditioned BCG residual.

5 Numerical Examples

We performed extensive numerical tests with Algorithm 1 for symmetric, highly indefinite linear systems arising in fluid flow computations and as subproblems within optimization algorithms. We now present two typical examples.

The first example is a linear system (1) with A of the form (3) that results from a finite-element approximation of a Stokes problem of the type (2), using the stabilized discretization from [14]. The matrix A is of the order 1634 with 1122 positive and 512 negative eigenvalues. The preconditioner M was chosen as the diagonal of A, and it was applied from the right, i.e., $M_1 = I$ and $M_2 = M$ in (8). In Figure 1, we show the relative residual norms

$$\frac{\|b - Ax_n\|_2}{\|b - Ax_0\|_2} \tag{9}$$

for the symmetric QMR Algorithm 1 (solid line) and MINRES [13] (dotted line). Recall that MINRES requires a positive definite preconditioner, and we used the diagonal matrix $|M|$ consisting of the absolute values of the diagonal elements of A as the preconditioner. Notice that QMR converges almost twice as fast as MINRES. We stress that both algorithms have essentially the same work and storage requirements.

The second example is a linear system (1) with a nonsingular symmetric indefinite coefficient matrix A of the form (3) (with the lower right block D being the zero matrix). This system arose in connection with

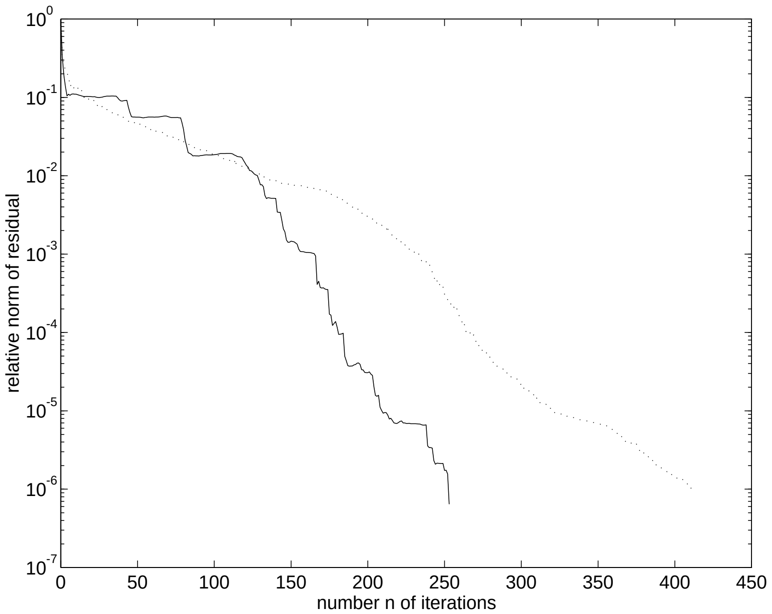

Figure 1: *Convergence curves for Stokes example*

the Karush-Kuhn-Tucker optimality conditions for an optimization problem, and we will refer to it as a KKT system. The order of A is $N = 1000$, and the upper left block B in (3) is a positive definite 600×600 matrix. The matrix A is highly indefinite with 600 positive and 400 negative eigenvalues. For this system we obtained a symmetric indefinite preconditioner by computing a Bunch-Parlett factorization [1] of the form

$$M = LDL^T \tag{10}$$

of an approximation M to A. Here L is a unit lower triangular matrix and D is a nonsingular block diagonal matrix with 1×1 and 2×2 blocks. We ran the symmetric QMR Algorithm 1 with right preconditioner M, and we compared it to MINRES with a positive definite right preconditioner $\hat{M}$. Following a proposal made in [9], we used the positive definite matrix $\hat{M} = L\hat{D}L^T$ that is obtained from (10) by constructing a block diagonal matrix $\hat{D}$ whose eigenvalues are just the absolute values of those of the indefinite matrix D. The resulting relative residual norms (9) for symmetric QMR (solid line) and MINRES (dotted line) are shown in Figure 2. In this example, symmetric QMR converges more than twice as fast as MINRES.

If no preconditioning is used, then the symmetric QMR algorithm 1 and MINRES are mathematically equivalent. In Figure 2, we also plotted the residual norm (9) for a run of Algorithm 1 without preconditioning (dashed line). Note that the method does not converge in a reasonable number of iterations. We remark that all Krylov-subspace iterations typically converge very slowly for highly indefinite coefficient matrices.

Finally, we refer the reader to [4] for further results of numerical experiments with linear systems from optimization problems. In [4], Freund and Jarre use the symmetric QMR Algorithm 1 to solve the symmetric indefinite linear systems that arise in each iteration of a primal-dual interior-point method for linear programs.

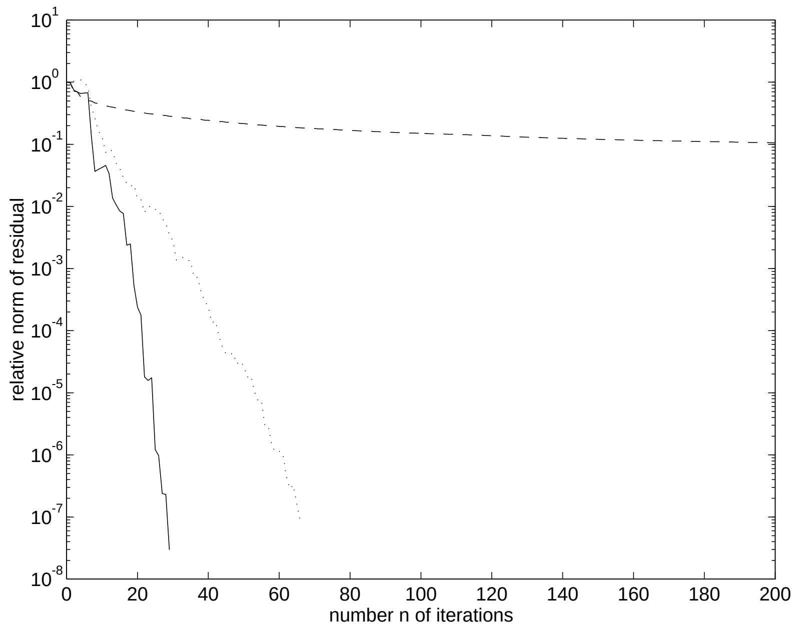

Figure 2: *Convergence curves for KKT system*

References

[1] J.R. Bunch and B.N. Parlett, "Direct methods for solving symmetric indefinite systems of linear equations", SIAM J. Numer. Anal. 8, pp. 639–655, 1971.

[2] R.W. Freund, "Transpose-free quasi-minimal residual methods for non-Hermitian linear systems", Recent Advances in Iterative Methods (G. Golub, A. Greenbaum, and M. Luskin, eds.), The IMA Volumes in Mathematics and its Applications, Vol. 60, pp. 69–94, Springer-Verlag, New York, 1994.

[3] 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, pp. 137–158, 1993.

[4] R.W. Freund and F. Jarre, "An interior-point method for large linear programs based on the symmetric QMR algorithm", in preparation.

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

[6] R.W. Freund and N.M. Nachtigal, "An implementation of the QMR method based on coupled two-term recurrences", SIAM J. Sci. Comput. 15, pp. 313–337, 1994.

[7] R.W. Freund and T. Szeto, "A transpose-free quasi-minimal residual squared algorithm for non-Hermitian linear systems", Advances in Computer Methods for Partial Differential Equations – VII (R. Vichnevetsky, D. Knight, and G. Richter, eds.), pp. 258–264, IMACS, 1992.

[8] R.W. Freund and H. Zha, "Simplifications of the nonsymmetric Lanczos process and a new algorithm for Hermitian indefinite linear systems", AT&T Numerical Analysis Manuscript, Bell Labs, Murray Hill, NJ, 1994.

[9] P.E. Gill, W. Murray, D.B. Ponceleón, and M.A. Saunders, "Preconditioners for indefinite systems arising in optimization", SIAM J. Matrix Anal. Appl. 13, pp. 292–311, 1992.

[10] M.D. Gunzburger, Finite Element Methods for Viscous Incompressible Flows, Academic Press, San Diego, 1989.

[11] C. Lanczos, "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators", J. Res. Nat. Bur. Standards 45, pp. 255–282, 1950.

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

[13] C.C. Paige and M.A. Saunders, "Solution of sparse indefinite systems of linear equations", SIAM J. Numer. Anal. 12, pp. 617–629, 1975.

[14] D.J. Silvester and K. Kechkar, "Stabilised bilinear-constant velocity-pressure finite elements for the conjugate gradient solution of the Stokes problem", Comput. Methods Appl. Mech. Engrg. 79, pp. 71–86, 1990.