NASA-CR-197948

IN-64-CR
0043879
398232 P-17

RIACS

Research Institute for Advanced Computer Science
NASA Ames Research Center

(NASA-CR-197948) QUASI-KERNEL POLYNOMIALS AND CONVERGENCE RESULTS FOR QUASI-MINIMAL RESIDUAL ITERATIONS (Research Inst. for Advanced Computer Science) 17 p

N95-23593

Unclas

G3/64 0043879

Quasi-Kernel Polynomials and Convergence Results for Quasi-Minimal Residual Iterations

Roland W. Freund

RIACS Technical Report 92.07

March 1992

Submitted to Numerical Methods of Approximation Theory

Quasi-Kernel Polynomials and Convergence Results for Quasi-Minimal Residual Iterations

Roland W. Freund

The Research Institute for Advanced Computer Science is operated by
Universities Space Research Association (USRA),
The American City Building, Suite 311, Columbia, MD 21044, (301)730–2656.

Work reported herein was supported by Cooperative Agreement NCC 2–387
between NASA and USRA.

Quasi-Kernel Polynomials and Convergence Results for Quasi-Minimal Residual Iterations

Roland W. Freund

Abstract. Recently, Freund and Nachtigal [9] have proposed a novel polynomial-based iteration, the quasi-minimal residual algorithm (QMR), for solving general nonsingular non-Hermitian linear systems. Motivated by the QMR method, in [6] we have introduced the general concept of quasi-kernel polynomials, and we have shown that the QMR algorithm is based on a particular instance of quasi-kernel polynomials. In this paper, we continue our study of quasi-kernel polynomials. In particular, we derive bounds for the norms of quasi-kernel polynomials. These results are then applied to obtain convergence theorems both for the QMR method and for a transpose-free variant of QMR, the TFQMR algorithm.

§1. Introduction

Many iterative algorithms for solving large nonsingular systems of linear equations

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

are based on polynomials. For example, the classical conjugate gradient algorithm [13] for Hermitian positive definite matrices A and its various extensions to non-Hermitian linear systems all fall into the category of polynomial-based methods. For a survey of iterative methods, we refer the reader to [7] and the references given there.

Typically, polynomial-based algorithms with optimal convergence properties correspond to polynomials that are characterized as solutions of constrained least squares approximation problems of the form

$$\min_{\psi \in \mathcal{P}_n:\, \psi(0)=1} \langle \psi, \psi \rangle. \tag{2}$$

Here and in the sequel, we use the notation

$$\mathcal{P}_n := \{\psi(\lambda) \equiv \gamma_0 + \gamma_1 \lambda + \cdots + \gamma_n \lambda^n \mid \gamma_0, \gamma_1, \ldots, \gamma_n \in \mathbf{C}\} \tag{3}$$

for the set of all complex polynomials of degree at most n. Moreover, $\langle \cdot, \cdot \rangle$ in (2) is an inner product that is positive definite on $\mathcal{P}_n$. The exact solutions of problems of the form (2) are normalized kernel polynomials, and they can be expressed in terms of the orthogonal polynomials associated with the inner product $\langle \cdot, \cdot \rangle$.

For the special case of Hermitian matrices A, only inner products on the real line arise, and then the corresponding kernel and orthogonal polynomials satisfy three-term recurrences. As a result, optimal polynomial-based

algorithms for Hermitian linear systems can be implemented using only three-term vector recursions. The situation is less satisfactory for the case of general non-Hermitian matrices A. Here, inner products in the complex plane arise, and the corresponding complex kernel and orthogonal polynomials no longer satisfy short recurrences. Consequently, optimal polynomial-based methods for non-Hermitian linear systems in general involve long vector recurrences. Usually, it is too expensive to run the full versions of these optimal algorithms. Instead, restarted or truncated variants are used, which often leads to rather slow convergence.

Recently, Freund and Nachtigal [9] proposed the quasi-minimal residual algorithm (QMR) for solving general nonsingular non-Hermitian linear systems (1). The QMR method produces approximate solutions of (1) that are characterized by a quasi-optimality condition, namely a quasi-minimization of the residual norm. The point is that the QMR algorithm—unlike optimal polynomial-based iteration schemes—can be implemented using only short vector recurrences.

Motivated by the QMR method, in [6] we have introduced the general concept of quasi-kernel polynomials, and we have shown that the QMR algorithm is based on a particular instance of quasi-kernel polynomials. Moreover, in [6] some general theory for quasi-kernel polynomials was given, such as recurrence relations and a characterization of roots of quasi-kernel polynomials as generalized eigenvalues.

In this paper, we continue our study of quasi-kernel polynomials. In particular, we derive bounds for the norms of quasi-kernel polynomials. These results are then applied to obtain convergence theorems both for the original QMR method and for a transpose-free variant of QMR, the TFQMR algorithm [5].

The remainder of the paper is organized as follows. In §2 we briefly recall the definition of quasi-kernel polynomials. In §3 we present bounds for the norms of quasi-kernel polynomials. We then turn to non-Hermitian linear systems (1), and in §4 we describe the general setting of polynomial-based methods for the iterative solution of such systems. Finally, in §5 and §6 we consider the QMR and TFQMR algorithms, respectively, and we give convergence results for both methods.

Throughout the paper, all vectors and matrices are allowed to have real or complex entries. As usual, $M^T = [\, m_{kj} \,]$ and $M^H = [\, \overline{m_{kj}} \,]$ denote the transpose and the conjugate transpose, respectively, of the matrix $M = [\, m_{jk} \,]$. We use the notation $\sigma_{\max}(M)$ and $\sigma_{\min}(M)$ for the largest and smallest singular value, respectively, of the matrix M. The vector norm $\|x\|_2 = \sqrt{x^H x}$ is always the Euclidean norm, $\|M\|_2 = \sigma_{\max}(M)$ is the corresponding matrix norm, and for nonsingular square matrices M, $\kappa_2(M) = \sigma_{\max}(M)/\sigma_{\min}(M)$ is the Euclidean condition number. Moreover, for square matrices M, we denote by $\lambda(M)$ the set of all eigenvalues of M. In addition to (3), we use the notation $\mathcal{P}_\infty$ for the set of all complex polynomials. The symbol 0 will be used for the number zero, the zero matrix, and the polynomial $\psi(\lambda) \equiv 0$; its actual meaning and, in the case of the zero matrix, its dimension will be apparent

from the context. Finally, I_n denotes the $n \times n$ identity matrix.

§2. Quasi-Kernel Polynomials

We consider constrained approximation problems of the form (2), where $\langle \cdot, \cdot \rangle$ is assumed to be a given positive semidefinite inner product on $\mathcal{P}_\infty$, i.e., for all φ_1, φ_2, φ, $\psi \in \mathcal{P}_\infty$ and γ_1, $\gamma_2 \in \mathbf{C}$:

$$\begin{aligned} \langle \gamma_1 \varphi_1 + \gamma_2 \varphi_2, \psi \rangle &= \gamma_1 \langle \varphi_1, \psi \rangle + \gamma_2 \langle \varphi_2, \psi \rangle, \\ \overline{\langle \varphi, \psi \rangle} &= \langle \psi, \varphi \rangle, \\ \langle \varphi, \varphi \rangle &\geq 0. \end{aligned}$$

Notice that we do not require $\langle \varphi, \varphi \rangle > 0$ for $\varphi \neq 0$, i.e., $\langle \cdot, \cdot \rangle$ is not assumed to be positive definite. It will be convenient to rewrite (2) in the form

$$\min_{\psi \in \mathcal{P}_n : \psi(0)=1} \|\psi\|. \tag{4}$$

Here and in the sequel,

$$\|\varphi\| := \langle \varphi, \varphi \rangle^{1/2}, \quad \varphi \in \mathcal{P}_\infty,$$

always denotes the seminorm induced by $\langle \cdot, \cdot \rangle$.

If $\langle \cdot, \cdot \rangle$ is positive definite on $\mathcal{P}_n$, then there always exists a unique optimal solution ψ_n^{K} of (4). It is well known (see, e.g., [21, Chapter XVI], [1, Chapter I], [20], or [6, Corollary 3.3]) that ψ_n^{K} is just the suitably normalized nth *kernel polynomial* associated with $\langle \cdot, \cdot \rangle$. Furthermore, ψ_n^{K} is given by

$$\psi_n^{\mathrm{K}}(\lambda) \equiv \frac{\sum_{j=0}^{n} \overline{\varphi_j(0)} \varphi_j(\lambda)}{\sum_{j=0}^{n} |\varphi_j(0)|^2}, \tag{5}$$

where $\varphi_j \in \mathcal{P}_j$, $j = 0, 1, \ldots, n$, is any set of *orthonormal polynomials* with respect to $\langle \cdot, \cdot \rangle$, i.e.,

$$\langle \varphi_j, \varphi_k \rangle = \begin{cases} 0, & \text{if } j \neq k, \\ 1, & \text{if } j = k. \end{cases}$$

In the following, we will refer to ψ_n^{K} as a *true* nth kernel polynomial.

Roughly speaking, quasi-kernel polynomials are approximations to true kernel polynomials, which are obtained from a set of arbitrary basis polynomials, rather than a set of orthonormal polynomials. More precisely, the setting is as follows. Let $\Pi_J := \{\varphi_j\}_{j=0}^{J}$, where J can either be a finite integer or equal to ∞, be a given set of polynomials

$$\varphi_j \in \mathcal{P}_j \quad \text{with} \quad \deg \varphi_j = j, \tag{6}$$

which span $\mathcal{P}_J$. We note that, by (6), each polynomial $\lambda \varphi_{j-1}$, $1 \leq j \leq J$, has a unique representation of the form

$$\lambda \varphi_{j-1}(\lambda) \equiv \sum_{k=0}^{j} h_{k+1,j} \varphi_k(\lambda), \quad \text{where} \quad h_{j+1,j} \neq 0. \tag{7}$$

In the sequel, we always assume that $n \in \{1,2,\ldots,J\}$, if J is finite, respectively $n \in \{1,2,\ldots\}$, if $J = \infty$. Moreover, we denote by

$$H_n = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & h_{n,n-1} & h_{nn} \\ 0 & \cdots & 0 & h_{n+1,n} \end{bmatrix} \tag{8}$$

the $(n+1) \times n$ *upper Hessenberg* matrix, which contains the recurrence coefficients $h_{k+1,j}$ from (7) as entries. We remark that, in view of (7), all subdiagonal elements $h_{j+1,j}$ in (8) are different from zero, i.e., H_n is an *unreduced* upper Hessenberg matrix.

It follows from (6) that any polynomial $\psi_n \in \mathcal{P}_n$ with $\psi_n(0) = 1$ can be parametrized in the form

$$\psi_n(\lambda) \equiv 1 - \lambda \sum_{j=0}^{n-1} \zeta_{j+1} \varphi_j(\lambda), \quad \text{where} \quad z := \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_n \end{bmatrix} \in \mathbf{C}^n. \tag{9}$$

In [6] we have shown that each polynomial (9) satisfies

$$\begin{aligned} \|\psi_n\| &= \left((d_n - H_n z)^H G_n (d_n - H_n z)\right)^{1/2} \\ &\leq \left(\|G_n\|_2\right)^{1/2} \|d_n - H_n z\|_2. \end{aligned} \tag{10}$$

Here

$$G_n := \left[\, \langle \varphi_j, \varphi_k \rangle \,\right]_{j,k=0,1,\ldots,n} \tag{11}$$

is the Gram matrix of $\varphi_0, \varphi_1, \ldots, \varphi_n$, and

$$d_n := \left[\, 1/\varphi_0 \quad 0 \quad \cdots \quad 0 \,\right]^T \in \mathbf{C}^{n+1}. \tag{12}$$

Note that, by (6), φ_0 is a nonzero constant.

The estimate (10) is the basis for the definition of quasi-kernel polynomials. We choose the parameter vector z in (9) such that, instead of $\|\psi_n\|$, only the second factor in the upper bound in (10) is minimal. This leads to the following definition.

Definition 1. [6, Definition 3.1]. *Let $z_n \in \mathbf{C}^n$ be the solution of the least squares problem*

$$\|d_n - H_n z_n\|_2 = \min_{z \in \mathbf{C}^n} \|d_n - H_n z\|_2. \tag{13}$$

The polynomial ψ_n given by (9) (*with $z := z_n$*) *is called the nth quasi-kernel polynomial* (*corresponding to the inner product $\langle \cdot, \cdot \rangle$ and derived from Π_J*).

Recall that H_n is an unreduced $(n+1) \times n$ Hessenberg matrix, and thus it has full rank n. This guarantees that (13) always has a unique solution z_n.

In [6] we have established some general theory for quasi-kernel polynomials, such as recurrence relations and a characterization of roots of quasi-kernel polynomials as generalized eigenvalues. For example, in [6, Theorem 3.2], it is shown that ψ_n satisfies the same relation (5) as true kernel polynomials, where the φ_j's are now the given basis polynomials (6). In particular, if the polynomials in Π_J are orthonormal, then the quasi-kernel polynomial ψ_n coincides with the true kernel polynomial.

§3. Bounds for the Norms of Quasi-Kernel Polynomials

In this section, we present some bounds for the norm $\|\psi_n\|$ of the nth quasi-kernel polynomial ψ_n.

First, note that, from (10) and (13), one obtains the estimates

$$\begin{aligned} \min_{\psi\in\mathcal{P}_n:\,\psi(0)=1} \|\psi\| &\le \|\psi_n\| \\ &\le (\|G_n\|_2)^{1/2} \min_{z\in\mathbf{C}^n} \|d_n - H_n z\|_2. \end{aligned} \tag{14}$$

For later use, we remark that the first factor of the upper bound in (14) satisfies the inequality

$$(\|G_n\|_2)^{1/2} \le \sqrt{n+1}\, \max_{j=0,\ldots,n} \|\varphi_j\|, \tag{15}$$

which follows from

$$\|G_n\|_2 = \max_{y\in\mathbf{C}^{n+1}:\,\|y\|_2=1} y^H G_n y$$

and the Cauchy-Schwarz inequality.

Clearly, the Gram matrix G_n defined in (11) is positive definite if the inner product $\langle\cdot,\cdot\rangle$ is positive definite on $\mathcal{P}_n$. In this case, the condition number $\kappa_2(G_n)$ of G_n is a measure for how far $\|\psi_n\|$ can be from the minimal value $\|\psi_n^{\mathrm{K}}\|$ of (4).

Theorem 2. *Assume that the inner product $\langle\cdot,\cdot\rangle$ is positive definite on $\mathcal{P}_n$, and let ψ_n^{K} be the true nth kernel polynomial. Then:*

$$\|\psi_n^{\mathrm{K}}\| \le \|\psi_n\| \le (\kappa_2(G_n))^{1/2}\, \|\psi_n^{\mathrm{K}}\|. \tag{16}$$

In particular, $\psi_n = \psi_n^{\mathrm{K}}$ if $G_n = I_n$.

Proof: The polynomial ψ_n^{K} is the optimal solution of the approximation problem on the left-hand side of (14). This implies the first inequality in (16).

Let ψ_n^{K} be represented in the form (9), and let $z_n^{\mathrm{K}} \in \mathbf{C}^n$ be the corresponding parameter vector. Using the first relation in (10), it follows that

$$\begin{aligned} \|\psi_n^{\mathrm{K}}\| &= \left((d_n - H_n z_n^{\mathrm{K}})^H G_n (d_n - H_n z_n^{\mathrm{K}})\right)^{1/2} \\ &\ge (\sigma_{\min}(G_n))^{1/2}\, \|d_n - H_n z_n^{\mathrm{K}}\|_2. \end{aligned} \tag{17}$$

On the other hand, from (14), we have

$$\begin{aligned} \|\psi_n\| &\leq (\sigma_{\max}(G_n))^{1/2} \min_{z \in \mathbf{C}^n} \|d_n - H_n z\|_2 \\ &\leq (\sigma_{\max}(G_n))^{1/2} \|d_n - H_n z_n^{\mathrm{K}}\|_2. \end{aligned} \tag{18}$$

Finally, by combining the estimates (17) and (18), we obtain the second inequality in (16). ■

Next, we show that the upper bound in (14) can be related to certain constrained approximation problems. This result will be used in §5 and §6 to derive convergence theorems for QMR and TFQMR, respectively.

We use the following setting. Let $m \leq J$ be an arbitrary, but fixed integer. We denote by $\lambda_1, \lambda_2, \ldots, \lambda_l \in \mathbf{C}$ the distinct zeros of the mth basis polynomial φ_m, and μ_j is the multiplicity of λ_j. Note that

$$\varphi_m(\lambda) \equiv \gamma \prod_{j=1}^{l} (\lambda - \lambda_j)^{\mu_j} \quad \text{for some} \quad \gamma \in \mathbf{C},\ \gamma \neq 0.$$

Furthermore, we denote by

$$H := [\, I_m \quad 0 \,]\, H_m \tag{19}$$

the $m \times m$ matrix obtained by deleting the last row of the mth Hessenberg matrix H_m in (8). Using (7) and (19), one easily verifies the identity

$$\lambda \Phi(\lambda) \equiv \Phi(\lambda) H + [\, 0 \quad \cdots \quad 0 \quad h_{m+1,m} \varphi_m(\lambda) \,], \tag{20}$$

where $\Phi(\lambda)$ is a row vector defined by

$$\Phi(\lambda) := [\, \varphi_0(\lambda) \quad \cdots \quad \varphi_{m-1}(\lambda) \,].$$

By setting $\lambda = \lambda_j$ in (20), we deduce that the zeros of φ_m are just the eigenvalues of H, i.e.,

$$\{\lambda_1, \lambda_2, \ldots, \lambda_l\} = \lambda(H), \tag{21}$$

and each λ_j is an eigenvalue of algebraic multiplicity μ_j. Next, we recall that H_m (and thus H) is an unreduced upper Hessenberg matrix. This implies that, for each λ_j, the matrix $H - \lambda_j I_m$ has rank $m - 1$, and hence the eigenspace corresponding to λ_j is of dimension 1. As a result, there is exactly one Jordan block for each eigenvalue of H. Therefore, H has the Jordan canonical form

$$J = SHS^{-1}, \tag{22}$$

where S is a nonsingular $m \times m$ matrix and

$$J = \operatorname{diag}(J(\lambda_1), J(\lambda_2), \ldots, J(\lambda_l)) \tag{23}$$

is a block diagonal matrix with Jordan blocks

$$J(\lambda_j) = \lambda_j I_{\mu_j} + N_j, \quad N_j := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \in \mathbf{R}^{\mu_j \times \mu_j}. \tag{24}$$

After these preliminaries, our result can now be stated as follows.

Theorem 3. *Let $m \leq J$ be a fixed integer, and let H be the $m \times m$ matrix defined by (19) with spectrum (21). Let $J = SHS^{-1}$ be the Jordan normal form (22), (23), with Jordan blocks (24). Then, for $n = 1, 2, \ldots, m-1$, the nth quasi-kernel polynomial ψ_n satisfies*

$$\|\psi_n\| \leq \frac{\kappa_2(S)}{|\varphi_0|} \left(\|G_n\|_2\right)^{1/2} \varepsilon_n, \tag{25}$$

where

$$\varepsilon_n := \min_{\psi \in \mathcal{P}_n:\, \psi(0)=1} \max_{j=1,\ldots,l} \|\psi(J(\lambda_j))\|_2. \tag{26}$$

Moreover, if H is diagonalizable, then:

$$\varepsilon_n = \min_{\psi \in \mathcal{P}_n:\, \psi(0)=1} \max_{\lambda \in \lambda(H)} |\psi(\lambda)|. \tag{27}$$

Proof: Let $n \in \{1, 2, \ldots, m-1\}$. By using (12) and by setting $u := \varphi_0 z$, we obtain from (14) the estimate

$$\|\psi_n\| \leq \frac{1}{|\varphi_0|} \left(\|G_n\|_2\right)^{1/2} \vartheta_n, \tag{28}$$

where

$$\vartheta_n := \min_{u \in \mathbf{C}^n} \|e_1^{(n+1)} - H_n u\|_2, \quad e_1^{(n+1)} := [1 \quad 0 \quad \cdots \quad 0]^T \in \mathbb{R}^{n+1}. \tag{29}$$

In view of (19), H_n is the $(n+1) \times n$ leading submatrix of H, i.e.,

$$H = \begin{bmatrix} H_n & * \\ 0 & * \end{bmatrix},$$

and we have the relation

$$H \begin{bmatrix} u \\ 0 \end{bmatrix} = \begin{bmatrix} H_n u \\ 0 \end{bmatrix} \quad \text{for all} \quad u \in \mathbf{C}^n. \tag{30}$$

Next, using the fact that H is an unreduced upper Hessenberg matrix, we deduce that

$$\left\{ \begin{bmatrix} u \\ 0 \end{bmatrix} \,\middle|\, u \in \mathbf{C}^n \right\} = \{\phi(H) e_1^{(m)} \mid \phi \in \mathcal{P}_{n-1}\}. \tag{31}$$

By means of (30) and (31), the quantity ϑ_n in (29) can be rewritten in the form

$$\vartheta_n = \min_{\phi \in \mathcal{P}_{n-1}} \|e_1^{(m)} - \phi(H) e_1^{(m)}\|_2 = \min_{\psi \in \mathcal{P}_n:\, \psi(0)=1} \|\psi(H) e_1^{(m)}\|_2. \tag{32}$$

With (22)–(24) and since $\|e_1^{(m)}\|_2 = 1$, it follows that, for all $\psi \in \mathcal{P}_\infty$,

$$\|\psi(H) e_1^{(m)}\|_2 \leq \kappa_2(S) \|\psi(J)\|_2 = \kappa_2(S) \max_{j=1,\ldots,l} \|\psi(J(\lambda_j))\|_2. \tag{33}$$

By combining (28), (32), and (33), we obtain the desired estimate (25), (26).

Finally, if H is diagonalizable, then all Jordan blocks $J(\lambda_j) = [\lambda_j]$ are of size 1, and (26) reduces to (27). ■

We remark that the norms in (26) can be estimated by means of the inequality

$$\|\psi(J(\lambda_j))\|_2 \leq \sum_{i=0}^{\mu_j - 1} \frac{1}{i!} \left|\psi^{(i)}(\lambda_j)\right|, \quad \psi \in \mathcal{P}_\infty, \tag{34}$$

which is an immediate consequence of the standard relation

$$\psi(J(\lambda_j)) = \sum_{i=0}^{\mu_j - 1} \frac{1}{i!} \psi^{(i)}(\lambda_j) N_j^i.$$

Here N_j is the nilpotent matrix defined in (24), and $\psi^{(i)}$ denotes the ith derivative of ψ.

§4. Polynomial-Based Matrix Iterations

We now return to systems of linear equations (1). From now on, it is always assumed that A in (1) is a nonsingular non-Hermitian $N \times N$ matrix.

Many iterative algorithms for solving linear systems (1) are *polynomial-based methods*: they produce approximations x_n to $A^{-1}b$ of the form

$$x_n = x_0 + \phi_n(A)r_0 \quad \text{with} \quad \phi_n \in \mathcal{P}_{n-1}, \quad n = 1, 2, \ldots . \tag{35}$$

Here $x_0 \in \mathbf{C}^N$ is any initial guess for the solution of (1), and $r_0 := b - Ax_0$. We remark that the residual vector $r_n := b - Ax_n$ corresponding to the nth iterate (35) of any polynomial-based algorithm is given by

$$r_n = \psi_n(A)r_0, \quad \text{where} \quad \psi_n(\lambda) \equiv 1 - \lambda\phi_n(\lambda). \tag{36}$$

Note that

$$\psi_n \in \mathcal{P}_n \quad \text{and} \quad \psi_n(0) = 1. \tag{37}$$

Generally, any ψ_n that satisfies (37) is called an nth *residual polynomial*.

The goal in designing a polynomial-based method is to choose at each step the residual polynomial ψ_n in (36) such that some norm of r_n is as small as possible. A standard approach [18] is to require that the Euclidean norm of r_n is minimal, i.e.,

$$\|r_n\|_2 = \min_{\psi \in \mathcal{P}_n : \psi(0)=1} \|\psi(A)r_0\|_2. \tag{38}$$

By setting

$$\langle \varphi, \psi \rangle := r_0^H (\psi(A))^H \varphi(A) r_0 \quad \text{and} \quad \|\psi\| := \langle \psi, \psi \rangle^{1/2}, \tag{39}$$

we see that (38) is an instance of a constrained polynomial approximation problem of the form (4). We remark that other strategies (see [17,19]) for choosing the residual polynomial ψ_n in (36) also lead to problems of the form (4), e.g., with inner products given by

$$\langle \varphi, \psi \rangle := \iint_G \varphi(\lambda)\overline{\psi(\lambda)}\omega(\lambda)\, d\sigma \quad \text{or} \quad \langle \varphi, \psi \rangle := \int_C \varphi(\lambda)\overline{\psi(\lambda)}\omega(\lambda)\, |d\lambda|\,.$$

Typically, $G \subset \mathbf{C}$ is a compact set containing the spectrum $\lambda(A)$ of A or some approximation to $\lambda(A)$, and C is a curve bounding such a set G.

In the remainder of this paper, we always assume that $\langle \cdot, \cdot \rangle$ is the inner product defined by (39). Note that the corresponding polynomial norm $\|\cdot\|$ is equivalent to the Euclidean vector norm in $\mathbf{C}^N$, in the sense that

$$\|\psi\| = \|v\|_2 \quad \text{for all} \quad v = \psi(A)r_0 \quad \text{with} \quad \psi \in \mathcal{P}_\infty. \tag{40}$$

Finally, we call a polynomial-based method *optimal* if its residual vectors satisfy (38). We recall (see, e.g., [6]) that an optimal iteration scheme terminates after a finite number $J_{\text{opt}} (\leq N)$ of steps with the exact solution $x_{J_{\text{opt}}} = A^{-1}b$ of (1). Furthermore, the residual vectors r_n^{opt} of an optimal polynomial-based method are unique, and they are given by

$$r_n^{\text{opt}} = \psi_n^{\text{K}}(A)r_0, \quad n = 1, 2, \ldots, J_{\text{opt}}, \tag{41}$$

where ψ_n^{K} is the nth true kernel polynomial associated with $\langle \cdot, \cdot \rangle$.

§5. Convergence Results for the QMR Algorithm

Faber and Manteuffel [2] have shown that—except for a very special class of non-Hermitian matrices—optimal polynomial-based methods for general non-Hermitian linear systems can be implemented only with long recurrences, which involve vectors from all previous iterations. Their result motivated the development of a quasi-optimal polynomial-based iteration scheme, the QMR algorithm [9], which can be implemented using only short vector recurrences. In this section, we apply the results from §3 to obtain two convergence theorems for the QMR algorithm.

The QMR iteration uses the Lanczos process [14] with look-ahead [16,8] to generate basis vectors of the form

$$v_j = \phi_j(A)r_0, \quad j = 0, 1, \ldots, J_{\text{L}}. \tag{42}$$

Here

$$\phi_j \in \mathcal{P}_j \quad \text{with} \quad \deg \phi_j = j, \quad j = 0, 1, \ldots, J_{\text{L}}, \tag{43}$$

are the so-called Lanczos polynomials, and J_{L} is the termination index of the look-ahead Lanczos algorithm. We note that $J_{\text{L}} \leq J_{\text{opt}}$. For a detailed description of the Lanczos process, we refer the reader to [8] and the references given there.

In [6] we have shown that the QMR algorithm is based on quasi-kernel polynomials. More precisely, the QMR residual vectors are given by

$$r_n^{\mathrm{QMR}} = \psi_n(A)r_0, \quad n = 1, 2, \ldots, J, \tag{44}$$

where $J := J_{\mathrm{L}}$, and ψ_n is the nth quasi-kernel polynomial corresponding to the inner product (39) and derived from the normalized Lanczos polynomials

$$\varphi_j = \begin{cases} \dfrac{\phi_j}{\|\phi_j\|}, & j = 0, 1, \ldots, J_{\mathrm{L}} - 1, \\ \phi_{J_{\mathrm{L}}}, & j = J_{\mathrm{L}}. \end{cases} \tag{45}$$

For later use, we remark that the condition number of the Gram matrix G_n associated with the basis polynomials (45) can be expressed in terms of the extreme singular values of the matrix of normalized Lanczos vectors

$$V_n := \left[\frac{v_0}{\|v_0\|_2} \quad \frac{v_1}{\|v_1\|_2} \quad \cdots \quad \frac{v_n}{\|v_n\|_2} \right].$$

More precisely, for $n = 0, 1, \ldots, J_{\mathrm{L}} - 1$, we have

$$(\kappa_2(G_n))^{1/2} = \frac{\sigma_{\max}(V_n)}{\sigma_{\min}(V_n)}. \tag{46}$$

The relation (46) is an immediate consequence of the identity

$$y^H G_n y = \|V_n y\|^2, \quad y \in \mathbf{C}^{n+1},$$

which follows with (11), (39), (42), (45), and (40).

Next, we apply Theorem 2, and we rewrite, by means of (40) and (46), the bounds in (16) in terms of the residual vectors (41) and (44). This gives the following result.

Theorem 4. *For $n = 1, 2, \ldots, J_{\mathrm{L}} - 1$, the* QMR *residual vectors satisfy*

$$\|r_n^{\mathrm{opt}}\|_2 \leq \|r_n^{\mathrm{QMR}}\|_2 \leq \frac{\sigma_{\max}(V_n)}{\sigma_{\min}(V_n)} \|r_n^{\mathrm{opt}}\|_2. \tag{47}$$

In particular, $r_n^{\mathrm{QMR}} = r_n^{\mathrm{opt}}$ if the columns of V_n are orthonormal.

We remark that a different proof for (47), without using quasi-kernel polynomials, was given by Nachtigal [15, Section 4.3].

Next, we derive upper bounds for $\|r_n^{\mathrm{QMR}}\|_2$ by applying Theorem 3 (with $m = J_{\mathrm{L}}$) for the special case of the QMR quasi-kernel polynomials.

In the sequel, the same notation as in §3 is used. We set $m := J_{\mathrm{L}}$, and let

$$H := [\, I_{J_{\mathrm{L}}} \quad 0 \,]\, H_{J_{\mathrm{L}}} \tag{48}$$

be the Hessenberg matrix of recurrence coefficients for the basis polynomials (45). As before, we denote by $\lambda_1, \lambda_2, \ldots, \lambda_l$ the distinct zeros of φ_m. We remark that, by (45), the polynomial $\varphi_m = \phi_{J_L}$ is just the last Lanczos polynomial. It is well known (see [16,12]) that each zero of ϕ_{J_L} is also an eigenvalue of A. Therefore, together with (21), we have

$$\{\lambda_1, \lambda_2, \ldots, \lambda_l\} = \lambda(H) \subseteq \lambda(A). \tag{49}$$

Furthermore, we will need the relation

$$|\varphi_0| = 1/\|r_0\|_2, \tag{50}$$

which follows from (45) (with j=0) and (40), and the inequality

$$(\|G_n\|_2)^{1/2} \leq \sqrt{n+1}, \quad n = 1, 2, \ldots, m-1, \tag{51}$$

which follows from (15) and (45).

Finally, by applying Theorem 3, we obtain the following convergence result for the QMR algorithm.

Theorem 5. *Let H be the $J_L \times J_L$ matrix* (48) *corresponding to the normalized Lanczos polynomials* (45), *and let* (49) *be the spectrum of H. Let $J = SHS^{-1}$ be the Jordan normal form* (22), (23) *of H, with Jordan blocks* (24). *Then, for $n = 1, 2, \ldots, J_L - 1$, the* QMR *residual vectors satisfy*

$$\|r_n^{\mathrm{QMR}}\|_2 \;\leq\; \|r_0\|_2 \, \kappa_2(S) \, \sqrt{n+1} \, \varepsilon_n, \tag{52}$$

where

$$\varepsilon_n := \min_{\psi \in \mathcal{P}_n : \psi(0)=1} \; \max_{j=1,\ldots,l} \|\psi(J(\lambda_j))\|_2.$$

Moreover, if H is diagonalizable, then:

$$\varepsilon_n \leq \min_{\psi \in \mathcal{P}_n : \psi(0)=1} \; \max_{\lambda \in \lambda(A)} |\psi(\lambda)|. \tag{53}$$

Proof: The inequality (52) is obtained by rewriting (25) in terms of the QMR residuals (44) instead of ψ_n, and by using the relations (50) and (51). The estimate (53) follows from (27) and (49). ■

We remark that, for the special case of diagonalizable H, the bound (52), (53) was first given by Freund and Nachtigal [9, Theorem 6.1].

Of course, the bound in (53) is not very practical since it would require the knowledge of all eigenvalues of A. However, often one knows some compact set $G \subset \mathbf{C}$ such that

$$\lambda(A) \subset G \quad \text{and} \quad 0 \notin G.$$

In this case, we can replace (53) by the estimate

$$\varepsilon_n \leq \min_{\psi \in \mathcal{P}_n : \psi(0)=1} \; \max_{\lambda \in G} |\psi(\lambda)|. \tag{54}$$

The solutions of constrained polynomial approximation problems of the form (54) are known explicitly for some "simple" sets, such as disks [22], certain ellipses [3,4], and line segments [11].

§6. A Convergence Result for the TFQMR Algorithm

The QMR algorithm involves matrix-vector multiplications with the coefficient matrix A, as well as its transpose A^T. This is a disadvantage for certain applications, where matrix-vector products $A \cdot v$, but not $A^T \cdot w$, can be computed efficiently without ever explicitly generating the matrix A. This was the motivation for developing a transpose-free variant of the standard QMR method, the TFQMR algorithm [5,10], which involves matrix-vector products with A only. We remark that the QMR and TFQMR methods are not equivalent, and the two algorithms produce different sets of iterates.

In this section, we present a convergence result for the TFQMR algorithm. It appears that this is the first convergence theorem for a transpose-free Lanczos-based iterative method.

In [6] we have shown that—like the QMR algorithm—the TFQMR method is also based on quasi-kernel polynomials. Its residual vectors are given by

$$r_n^{\text{TFQMR}} = \psi_n(A) r_0, \quad n = 1, 2, \ldots, J,$$

where $J := 2J_{\text{L}}$ and ψ_n is the nth quasi-kernel polynomial corresponding to the inner product (39) and derived from basis polynomials $\varphi_0, \varphi_1, \ldots, \varphi_J$ that are now defined as squares or products of Lanczos polynomials (43). More precisely, the basis polynomials are defined as follows:

$$\varphi_j = \begin{cases} \dfrac{\phi_k^2}{\|\phi_k^2\|}, & \text{if } j = 2k,\ k = 0, 1, \ldots, J_{\text{L}} - 1, \\ \dfrac{\phi_{k-1}\phi_k}{\|\phi_{k-1}\phi_k\|}, & \text{if } j = 2k-1,\ k = 1, 2, \ldots, J_{\text{L}} - 1, \\ \dfrac{\phi_{J_{\text{L}}-1}\phi_{J_{\text{L}}}}{\|\phi_{J_{\text{L}}-1}\phi_{J_{\text{L}}}\|}, & \text{if } j = 2J_{\text{L}} - 1 \text{ and } \|\phi_{J_{\text{L}}-1}\phi_{J_{\text{L}}}\| \neq 0, \\ \phi_{J_{\text{L}}-1}\phi_{J_{\text{L}}}, & \text{if } j = 2J_{\text{L}} - 1 \text{ and } \|\phi_{J_{\text{L}}-1}\phi_{J_{\text{L}}}\| = 0, \\ \phi_{J_{\text{L}}}^2, & \text{if } j = 2J_{\text{L}}. \end{cases} \tag{55}$$

We remark that, as in the previous section, J_{L} denotes the termination index of the look-ahead Lanczos algorithm.

Again, we apply Theorem 3, but this time with $m := 2J_{\text{L}}$. Let

$$H' := [\, I_{J_{2\text{L}}} \quad 0 \,]\, H_{J_{2\text{L}}}, \tag{56}$$

where $H_{J_{2\text{L}}}$ now denotes the Hessenberg matrix (8) containing the recurrence coefficients of the basis polynomials (55). Here, the notation H' is used to distinguish the matrix (56) from the Hessenberg matrix (48) associated with the QMR algorithm.

Recall from §3 that the eigenvalues of H' are just the distinct zeros of the polynomial φ_m. Now, from (55) (with $j = 2J_{\text{L}}$), we have that

$$\varphi_m = \phi_{J_{\text{L}}}^2 \tag{57}$$

is the square of the last Lanczos polynomial. It follows that

$$\lambda(H') = \{\lambda_1, \lambda_2, \ldots, \lambda_l\},$$

where, as in §5, the λ_j's denote the distinct zeros of ϕ_{J_L}. Hence, in view of (49), the eigenvalues of the TFQMR and QMR Hessenberg matrices H' and H coincide, and we have

$$\lambda(H') = \lambda(H) \subseteq \lambda(A). \tag{58}$$

However, their algebraic multiplicities are different. Indeed, by (57), each eigenvalue λ_j of H' has even algebraic multiplicity $2\mu_j$, while λ_j is an eigenvalue of H with algebraic multiplicity $\mu_j (\geq 1)$. In particular, the blocks in the Jordan normal form (22), (23) of H' are now of the form

$$J(\lambda_j) = \lambda_j I_{2\mu_j} + N_j, \quad N_j := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{2\mu_j \times 2\mu_j}. \tag{59}$$

Finally, we note that, for the basis polynomials (55), the relations (50) and (51) hold too.

Therefore, we can proceed as in the previous section, and by applying Theorem 3, we obtain the following convergence result for the TFQMR algorithm.

Theorem 6. *Let H' be the $2J_L \times 2J_L$ matrix (56) corresponding to the basis polynomials (55). Then, the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_l$ of H' are the zeros of the last Lanczos polynomial ϕ_{J_L}, and each λ_j is an eigenvalue of even algebraic multiplicity $2\mu_j$.*

Let $J = SH'S^{-1}$ be the Jordan normal form (22), (23), with Jordan blocks (59). Then, for $n = 1, 2, \ldots, 2J_L - 1$, the TFQMR *residual vectors satisfy*

$$\|r_n^{\mathrm{TFQMR}}\|_2 \leq \|r_0\|_2 \, \kappa_2(S) \sqrt{n+1} \, \varepsilon_n,$$

where

$$\varepsilon_n := \min_{\psi \in \mathcal{P}_n : \psi(0)=1} \max_{j=1,\ldots,l} \|\psi(J(\lambda_j))\|_2. \tag{60}$$

Moreover, if all zeros of the last Lanczos polynomial ϕ_{J_L} are simple, then:

$$\varepsilon_n \leq \min_{\psi \in \mathcal{P}_n : \psi(0)=1} \max_{\lambda \in \lambda(A)} \left(|\psi(\lambda)| + |\psi'(\lambda)| \right). \tag{61}$$

We remark that the estimate (61) follows from (60) by using (58) and inequality (34) (with μ_j replaced by 2).

References

1. Chihara, T. S., *An Introduction to Orthogonal Polynomials*, Gordon and Breach, New York, 1978.
2. Faber, V. and T. Manteuffel, Necessary and sufficient conditions for the existence of a conjugate gradient method, SIAM J. Numer. Anal. **21** (1984), 352–362.
3. Fischer, B. and R. W. Freund, On the constrained Chebyshev approximation problem on ellipses, J. Approx. Th. **62** (1990), 297–315.
4. Fischer, B. and R. W. Freund, Chebyshev polynomials are not always optimal, J. Approx. Th. **65** (1991), 261–272.
5. Freund, R. W., A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems, RIACS Technical Report 91.18, NASA Ames Research Center, September 1991.
6. Freund, R. W., Quasi-kernel polynomials and their use in non-Hermitian matrix iterations, J. Comput. Appl. Math., to appear.
7. Freund, R. W., G. H. Golub, and N. M. Nachtigal, Iterative solution of linear systems, Acta Numerica **1** (1992), to appear.
8. Freund, R. W., M. H. Gutknecht, and N. M. Nachtigal, An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices, SIAM J. Sci. Stat. Comput., to appear.
9. Freund R. W. and N. M. Nachtigal, QMR: a quasi-minimal residual method for non-Hermitian linear systems, Numer. Math. **60** (1991), 315–339.
10. Freund R. W. and N. M. Nachtigal, TFQMR: a maximal subspace look-ahead QMR algorithm for non-Hermitian linear systems, in preparation.
11. Freund R. W. and St. Ruscheweyh, On a class of Chebyshev approximation problems which arise in connection with a conjugate gradient type method, Numer. Math. **48** (1986), 525–542.
12. Gutknecht, M. H., A completed theory of the unsymmetric Lanczos process and related algorithms, Part I, SIAM J. Matrix Anal. Appl. **13** (1992), to appear.
13. Hestenes, M. R. and E. L. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Stand. **49** (1952), 409–436.
14. Lanczos, C., An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Nat. Bur. Stand. **45** (1950), 255–282.
15. Nachtigal, N. M., A look-ahead variant of the Lanczos algorithm and its application to the quasi-minimal residual method for non-Hermitian linear systems, Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, 1991.
16. Parlett, B. N., D. R. Taylor, and Z. A. Liu, A look-ahead Lanczos algorithm for unsymmetric matrices, Math. Comp. **44** (1985), 105–124.
17. Saad, Y., Least squares polynomials in the complex plane and their use for solving nonsymmetric linear systems, SIAM J. Numer. Anal. **24** (1987), 155–169

18. Saad, Y. and M. H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput. **7** (1986), 856–869.
19. Smolarski, D. C. and P. E. Saylor, An optimum iterative method for solving any linear system with a square matrix, BIT **28** (1988), 163–178.
20. Stiefel, E. L., Kernel polynomials in linear algebra and their numerical applications, U.S. National Bureau of Standards, Applied Mathematics Series **49** (1958), 1–22.
21. Szegö, G., *Orthogonal Polynomials*, Fourth Edition, Amer. Math. Soc., Providence, 1975.
22. Varga, R. S., A comparison of the successive overrelaxation method and semi-iterative methods using Chebyshev polynomials, J. Soc. Indust. Appl. Math. **5** (1957), 39–46.

Supported by Cooperative Agreement NCC 2-387 between the National Aeronautics and Space Administration (NASA) and the Universities Space Research Association (USRA).

RIACS
Mail Stop 230-5
NASA Ames Research Center
Moffett Field, CA 94035