SIAM J. SCI. COMPUT.
Vol. 15, No. 2, pp. 427–439, March 1994


012



# A ROBUST GMRES-BASED ADAPTIVE POLYNOMIAL PRECONDITIONING ALGORITHM FOR NONSYMMETRIC LINEAR SYSTEMS*

WAYNE JOUBERT†

**Abstract.** In this study a hybrid generalized minimal residual (GMRES) / polynomial preconditioning algorithm for solving nonsymmetric systems of linear equations is defined. The algorithm uses the results from cycles of restarted GMRES to form an effective polynomial preconditioner, typically resulting in decreased work requirements. The algorithm has the advantage over other hybrid algorithms in that its convergence behavior is well understood: the new algorithm converges for all starting vectors if and only if restarted GMRES converges. The results of numerical experiments with the algorithm are presented.

**Key words.** iterative methods, nonsymmetric linear systems, GMRES, polynomial preconditioning

**AMS subject classification.** 65F10

## 1. Introduction.

A desirable method for solving the linear system

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

where $A \in \mathbb{C}^{N\times N}$ is nonsingular, is the minimal residual method

$$u^{(n)} = u^{(0)} + q_{n-1}(A)r^{(0)}, \quad \deg q_{n-1} \le n-1, \quad q_{n-1} \text{ such that } ||r^{(n)}|| \text{ minimized}, \tag{2a}$$

or, alternatively,

$$u^{(n)} \in u^{(0)} + \mathbf{K}_n(r^{(0)}, A), \qquad Ar^{(n)} \perp \mathbf{K}_n(r^{(0)}, A), \tag{2b}$$

where $\{u^{(i)}\}_{i\ge 0}$ denote the iterates, $r^{(i)} = b - Au^{(i)}$ denotes the associated residuals, $\mathbf{K}_n(v, A) = \text{span}\{A^i v\}_{i=0}^{n-1}$ is the associated Krylov space, and $||\cdot||$ is the standard 2-norm of vectors. We also let $P_n(z) = 1 - zq_{n-1}(z)$, so that $r^{(n)} = P_n(A)r^{(0)}$, which condition identifies (2a), (2b) as a polynomial method.

When $A$ is Hermitian positive definite (HPD) or possesses a similar property, the polynomial $q_{n-1}$ may be determined by only $n$ matrix-vector products $A \cdot v$ along with $2n$ inner products. This fact leads to effective iterative methods such as the conjugate residual method. On the other hand, to form $q_{n-1}$ in the general case requires at least $n^2/2$ inner products, a requirement that appears insurmountable and that renders standard iterative methods such as full GMRES for performing (2) prohibitively costly due to long recurrence relations [2], [3], [12], and [9, §2.5].

A typical remedy to the order-$n^2$ work problem is to restart (2) periodically every $s$ iterations. This is the basis of the GMRES($s$) algorithm, for example, [18]; see [11] for a survey of restarted methods. Choosing $s$ large improves the convergence of the method; however, an increased average work per iteration results (see [10]).

A further remedy to the long recurrence problem comes in the form of hybrid methods. These methods typically apply GMRES until sufficient information is extracted from $A$, typically in the form of eigenvalue information, so that an effective polynomial method such as polynomial preconditioning or Chebyshev acceleration may be applied based on this information (see, e.g., [15], [1], [16]). This approach has the advantage of using short recurrences

*Received by the editors May 18, 1992; accepted for publication (in revised form) April 12, 1993. This work was supported in part by National Science Foundation grant DCR-8518722, and Department of Energy grant DE-FG05-87ER25048, with the University of Texas at Austin. This work was also supported in part by Department of Energy grant W-7405-ENG-36 with Los Alamos National Laboratory.

†Los Alamos National Laboratory, Los Alamos, New Mexico 87545 (wdj@lanl.gov).

for the sections of the algorithm not using GMRES. On the other hand, the composite nature of these algorithms often makes it difficult to prove rigorous convergence results for the algorithms or to predict their behavior compared to other iterative methods.

In this study, an alternative hybrid algorithm is proposed. For this algorithm, the basic framework of restarted GMRES is maintained, but for each restart cycle the $(s-1)$ degree polynomial that would normally be calculated by (2) is optionally replaced by a polynomial preconditioner of the same degree that is based on information from previous GMRES cycles. The new algorithm has the advantage of converging for exactly the same class of matrices as restarted GMRES, and often at significantly less cost. The savings are even greater on parallel machines and other architectures for which inner product computations are particularly expensive.

The remainder of this paper is organized as follows. In §2 we consider the theoretical feasibility of replacing GMRES cycles with polynomial preconditioning. Then in §3 the new algorithm is defined. In §4 the results of numerical experiments are presented.

**2. The feasibility of polynomial preconditioning.** In this section we consider the relative convergence behavior of (2) compared to the best fixed polynomial preconditioning that can be applied to solve (1). We assume throughout that $A$ is nonsingular.

We begin with definitions of basic convergence bounds for polynomial methods. Following [10], for $\mathbb{K}$ denoting the reals $\mathbb{R}$ or the complex numbers $\mathbb{C}$, and $A \in \mathbb{K}^{N\times N}$ and $n \geq 0$, let

$$\psi_{n,\mathbb{K}}(A) = \sup_{\substack{r\in\mathbb{K}^N \\ r\neq 0}} \inf_{\substack{\deg P_n \leq n \\ P_n(0)=1}} \frac{||P_n(A)r||}{||r||} = \sup_{\substack{r^{(0)}\in\mathbb{K}^N \\ r^{(0)}\neq 0}} \frac{||r^{(n)}||}{||r^{(0)}||},$$

$$\varphi_{n,\mathbb{K}}(A) = \inf_{\substack{\deg P_n \leq n \\ P_n(0)=1}} \sup_{\substack{r\in\mathbb{K}^N \\ r\neq 0}} \frac{||P_n(A)r||}{||r||},$$

where $r^{(n)}$ refers to the $n$th residual generated by the minimal residual method (2), and $P_n(z) = 1 - zq_{n-1}(z)$ is a polynomial over $\mathbb{C}$. When the field $\mathbb{K}$ is omitted in the notation, it is assumed to be $\mathbb{C}$. Here and throughout, $||v|| = \sqrt{v^*v}$, where $v^*$ denotes the conjugate transpose of $v$, and $||M||$ denotes the induced matrix norm.

The quantities $\psi_{n,\mathbb{K}}(A)$ and $\varphi_{n,\mathbb{K}}(A)$ are bounds that denote the possible residual norm decrease for two particular polynomial methods. They differ in the following important respect. The first quantity gives a sharp bound for the convergence of GMRES or any other algorithm implementing (2). In this case the polynomial $P_n$ is chosen as a function of $r^{(0)}$. On the other hand, the second quantity denotes the performance of the best polynomial preconditioner (defined here as $q_{n-1}$) that is chosen independent of $r^{(0)}$ and that must be effective for all such vectors in $\mathbb{K}^N$. The main concern of this section is the relationship between these two quantities; in particular, it is of interest to determine when they are equal.

Basic results on these quantities are set forth in [10]. Of interest to us here are the following properties for $A \in \mathbb{K}^{N\times N}$:

(i) $0 \leq \psi_{n,\mathbb{K}}(A) \leq \varphi_{n,\mathbb{K}}(A) \leq 1$ for all $n$.

(ii) $\psi_{0,\mathbb{K}}(A) = \varphi_{0,\mathbb{K}}(A) = 1$, while $\psi_{n,\mathbb{K}}(A) = \varphi_{n,\mathbb{K}}(A) = 0$ for any $n$ satisfying $n \geq d(A) = \min\{\deg P : P \text{ monic}, P(A) = 0\}$.

(iii) When $A \in \mathbb{R}^{N\times N}$, $\psi_{n,\mathbb{R}}(A) \leq \psi_{n,\mathbb{C}}(A)$, while $\varphi_{n,\mathbb{R}}(A) = \varphi_{n,\mathbb{C}}(A)$.

(iv) For fixed $n$, $\psi_{n,\mathbb{K}}(A)$ and $\varphi_{n,\mathbb{K}}(A)$ are each continuous in $A$.

(v) $\psi_{n,\mathbb{K}}(A)^2 = \sup_{r\in\mathbb{K}^N} F_n(r)$ holds for $A \in \mathbb{K}^{N\times N}$, where $F_n(r)$ is defined to be zero for $r$ satisfying $d(r, A) < n$ and elsewhere $F_n(r)$ is defined to be

$$1 - r^* A K_n(r, A)[K_n(r, A)^* A^* A K_n(r, A)]^{-1} K_n(r, A)^* A^* r / r^* r. \tag{3}$$

Here, $d(v, A) = \min\{\deg P : P \text{ monic}, P(A)v = 0\}$ denotes the degree of a vector, and $K_n(v, A) = \begin{bmatrix} v & Av & \dots & A^{n-1}v \end{bmatrix}$ is the standard Krylov basis matrix. Note that $F_n(r^{(0)}) = ||r^{(n)}||^2/||r^{(0)}||^2$, where $r^{(n)}$ is defined by (2), which gives the square of the residual decrease from $n$ steps of GMRES applied to a vector.

(vi) The map $F_n$ of (3) is the zero map when $d(A) \le n$. Otherwise it is continuous, and in fact it is $C^\infty$ on the complement of the closed measure-zero set $\{r : d(r, A) < n\}$, where $F_n$ is considered as a map on $\mathbb{R}^{2N}$. The latter fact follows directly from the fact that $F_n(r)$ is a rational function of the entries of $\operatorname{Re} r$ and $\operatorname{Im} r$.

From the standpoint of iterative methods, it is desirable that one or both of the following conditions be satisfied.

*Condition* 1. For given $A \in \mathbb{K}^{N\times N}$, $\psi_{s,\mathbb{K}}(A) < 1 \Leftrightarrow \varphi_{s,\mathbb{K}}(A) < 1$. If this property holds, then a convergent polynomial preconditioner of degree $s - 1$ necessarily exists (in the sense that $\varphi_{s,\mathbb{K}}(A) < 1$) if and only if GMRES($s$) is guaranteed to converge for all $r^{(0)} \in \mathbb{K}^N$.

*Condition* 2. For given $A \in \mathbb{K}^{N\times N}$, $\psi_{s,\mathbb{K}}(A) = \varphi_{s,\mathbb{K}}(A)$. If this stronger condition holds, then not only does a convergent polynomial preconditioner exist if and only if GMRES($s$) must converge, but one exists that has the same convergence bound $||r^{(ms)}||/||r^{(0)}|| \le \psi_s(A)^m$ as GMRES($s$).

It is not clear whether generally $\psi_{n,\mathbb{K}}(A) = \varphi_{n,\mathbb{K}}(A)$. However, the following results do hold.

THEOREM 1. *For* $\mathbb{K} = \mathbb{R}$ *or* $\mathbb{C}$, *and* $A \in \mathbb{K}^{N\times N}$, $\psi_{1,\mathbb{K}}(A) = \varphi_{1,\mathbb{K}}(A)$.

*Proof.* Let $P$ be a minimizing polynomial for $\varphi_{1,\mathbb{K}}(A)$. As shown in [10], it is possible to choose $P$ with coefficients in $\mathbb{K}$.

We first show that for any nonzero $\alpha \in \mathbb{K}$, there exists nonzero $v$ within $V$, $V$ here defined to be the eigenspace (over $\mathbb{K}$) associated with the maximal eigenvalue of $P(A)^* P(A)$, such that $\operatorname{Re} \alpha^* v^* A^* P(A) v \ge 0$. Note that for any $v$ and $\alpha$, and any $\epsilon > 0$, $||[P(A) + \epsilon\alpha A]v||^2 = ||P(A)v||^2 + 2\epsilon \operatorname{Re}[\alpha^* v^* A^* P(A)v] + \epsilon^2|\alpha|^2||Av||^2$. If the result is false, then for some $\alpha$, $\operatorname{Re} \alpha^* v^* A^* P(A) v < 0$ for all nonzero $v \in V$, in which case there exists $\tilde{c}_1 > 0$ with $2 \operatorname{Re} \alpha^* v^* A^* P(A) v \le -\tilde{c}_1 ||P(A)v||^2$ for all $v \in V$.

Let $\tilde{P}(z) = P(z) + \epsilon\alpha z$ and $w = v \cos\theta + v^\perp \sin\theta$, $v \in V$, $v^\perp \in V^\perp$, $||v|| = ||v^\perp|| = 1$. We seek to show for some $\epsilon$, $||\tilde{P}(A)w|| < ||P(A)||$ for all $w$, a contradiction to $P$ being a minimizer. Note that

$$\begin{aligned} ||\tilde{P}(A)w||^2 &= \cos^2\theta ||\tilde{P}(A)v||^2 + 2\cos\theta \sin\theta \operatorname{Re} v^* \tilde{P}(A)^* \tilde{P}(A) v^\perp + \sin^2\theta ||\tilde{P}(A)v^\perp||^2 \\ &= \cos^2\theta ||P(A)v||^2 + 2\epsilon \cos^2\theta \operatorname{Re} \alpha^* v^* A^* P(A) v \\ &\quad + 2\epsilon \cos\theta \sin\theta \operatorname{Re}[\alpha v^* P(A)^* A v^\perp + \alpha^* v^* A^* P(A) v^\perp] \\ &\quad + \sin^2\theta ||P(A)v^\perp||^2 + 2\epsilon \sin^2\theta \operatorname{Re} \alpha^* v^{\perp *} A^* P(A) v^\perp + \mathcal{O}(\epsilon^2). \end{aligned}$$

Thus there exist constants $c_2$, $c_3$, $c_4$, and $c_5$ (with $c_5 > 0$, such that the condition $||P(A)v^\perp||^2 \le ||P(A)v||^2 - c_5$ is satisfied, by definition of $V$), such that for any $v$, $v^\perp$ as defined above, $||\tilde{P}(A)w||^2$ is bounded by

$$\begin{aligned} &\cos^2\theta(1 - \epsilon\tilde{c}_1)||P(A)v||^2 + \epsilon c_2 \cos\theta \sin\theta + \epsilon c_3 \sin^2\theta + \sin^2\theta ||P(A)v^\perp||^2 + \epsilon^2 c_4 \\ &\quad \le ||P(A)v||^2 - \epsilon c_1 \cos^2\theta + \epsilon c_2 \cos\theta \sin\theta + \epsilon c_3 \sin^2\theta - c_5 \sin^2\theta + \epsilon^2 c_4, \end{aligned}$$

where $c_1 = \tilde{c}_1 ||P(A)||^2$. It is sufficient to find $\epsilon > 0$ such that for all $\theta$,

$$\epsilon(-c_1 \cos^2\theta + c_2 \cos\theta \sin\theta + c_3 \sin^2\theta) - c_5 \sin^2\theta + \epsilon^2 c_4 < 0.$$

But the left-hand side of this inequality equals

$$-c_5 \sin^2\theta + \epsilon^2 c_4 - \epsilon c_1 + \epsilon \sin\theta(c_1 \sin\theta + c_2 \cos\theta + c_3 \sin\theta)$$
$$\leq -c_5 \sin^2\theta + \epsilon^2 c_4 - \epsilon c_1 + \epsilon c_6 \sin\theta$$

for some $c_6$, for all $\theta$. Letting $z = \sin\theta$ yields the quadratic $-c_5 z^2 + \epsilon c_6 z + (\epsilon^2 c_4 - \epsilon c_1)$, which can be made negative for all real $z$ by choosing $\epsilon$ sufficiently small.

This establishes that for any $\alpha \in \mathbb{K}$ there exists a nonzero $v \in V$ such that $\mathrm{Re}\,\alpha^* v^* A^* \cdot P(A)v \geq 0$. We now show that for any $c \in \mathbb{K}$ there is nonzero $v \in V$ such that $\mathrm{Re}[c^* v^* A^* P(A)v] = 0$. Otherwise, letting $\alpha = \pm c$ yields $v_1$, $v_2 \in V$ such that $\mathrm{Re}[c^* v_1^* A^* P(A)v_1] > 0$ and $\mathrm{Re}[c^* v_2^* A^* P(A)v_2] < 0$. Note $v_1$ and $v_2$ must be linearly independent; otherwise, $v_1 = \xi v_2$, $\xi \in \mathbb{K}$, yields $v_1^* A^* P(A)v_1 = |\xi|^2 v_2^* A^* P(A)v_2$, a contradiction. Setting $v_t = t v_1 + (1-t)v_2$, $0 \leq t \leq 1$, yields by a continuity argument some $t$ such that $\mathrm{Re}[c^* v_t^* A^* P(A)v_t] = 0$, giving the result.

Let $\underline{V}$ be a matrix whose columns form a basis over $\mathbb{K}$ of $V$. We have shown for any $c \in \mathbb{K}$ there exists nonzero $u \in \mathbb{K}^{\mathrm{Dim}V}$ such that $\mathrm{Re}[c^* u^* \underline{V}^* A^* P(A)\underline{V}u] = 0$. When $\mathbb{K} = \mathbb{R}$, take $c = 1$ to give $r = \underline{V}u \neq 0$ with $r^* A^* P(A)r = 0$. When $\mathbb{K} = \mathbb{C}$, we have shown $0 \in F(H(c^* \underline{V} A^* P(A)\underline{V}))$ for any $c \in \mathbb{C}$, where $H(M) = (M + M^*)/2$ denotes the Hermitian part of a matrix $M$ and $F(M) = \{x^* M x : x \in \mathbb{C}^N, \ ||x|| = 1\}$ the field of values [8]. This implies that $0 \in F(\underline{V} A^* P(A)\underline{V})$: otherwise, by the convexity of the field of values, there exists $c = e^{i\theta}$ such that $H(c^* \underline{V} A^* P(A)\underline{V})$ is positive definite, i.e., $0 \notin F(H(c^* \underline{V} A^* P(A)\underline{V}))$. Thus there exists nonzero $r = \underline{V}u \in V$ such that $r^* A^* P(A)r = 0$.

Thus, for either the real or complex case, $Ar \perp P(A)r$ for nonzero $r \in V$. It is easily seen that this $P(z) = 1 - z q_0(z)$ satisfies (2b) for this $r$, and in fact $\psi_{1,\mathbb{K}}(A)^2 \geq ||P(A)r||^2/||r||^2 = ||P(A)||^2 \geq \varphi_{1,\mathbb{K}}(A)^2$. □

THEOREM 2. *For $A \in \mathbb{R}^{N\times N}$, $\psi_{2,\mathbb{R}}(A) < 1 \Leftrightarrow \varphi_{2,\mathbb{R}}(A) < 1$.*

*Proof.* Let $n = 2$. It is clear from (3) that $\psi_{n,\mathbb{R}}(A) < 1$ implies that for all real $v$ satisfying $d(v, A) \geq n$ and $||v|| = 1$, $v^* A K_n(v, A) \neq 0$. The same is true when $d = d(v, A) < n$ and $||v|| = 1$: in this case, we have

$$0 = v^* v - v^* A K_d(v, A)[K_d(v, A)^* A^* A K_d(v, A)]^{-1} K_d(v, A)^* A^* v,$$

so that $0 \neq v^* A K_d(v, A)$, a subvector of $v^* A K_n(v, A)$. Defining

$$\mathcal{F}_{n,\mathbb{K}}(A) = \{v^* A K_n(v, A) : v \in \mathbb{K}^N, \ ||v|| = 1\} \subseteq \mathbb{K}^n$$

for $A \in \mathbb{K}^{N\times N}$, we thus have that $0 \notin \mathcal{F}_{n,\mathbb{R}}(A)$. According to [7, p. 86], $\mathcal{F}_{2,\mathbb{R}}(A)$ is convex in $\mathbb{R}^2$ when $N \geq 3$.

When $N \leq 2$, $\psi_{2,\mathbb{R}}(A) = \varphi_{2,\mathbb{R}}(A) = 0$. When $N \geq 3$, since $0 \notin \mathcal{F}_{n,\mathbb{R}}(A)$, by the Hahn–Banach theorem there is a hyperplane through zero that does not intersect $\mathcal{F}_{2,\mathbb{R}}(A)$. Letting $c = [c_1\ c_2]^T$ denote the appropriate normal vector to this hyperplane, we have that $c^* \mathcal{F}_{2,\mathbb{R}}(A)$ consists only of positive numbers. Thus $c_1 v^* A v + c_2 v^* A^2 v$ is bounded above zero for all real $v$ satisfying $||v|| = 1$. Now let $P(z) = 1 - \alpha(c_1 z + c_2 z^2)$. Then $||P(A)v||^2/||v||^2 = 1 - 2\alpha v^*(c_1 A v + c_2 A^2)v/v^* v + \alpha^2 ||(c_1 A v + c_2 A^2)v||^2/v^* v$ which, for $\alpha$ sufficiently small, is bounded beneath one for all nonzero $v$. □

THEOREM 3. *For $A \in \mathbb{R}^{N\times N}$ HPD, $\psi_{n,\mathbb{R}}(A) = \varphi_{n,\mathbb{R}}(A)$.*

*Proof.* See [4]. □

THEOREM 4. *For $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and $A \in \mathbb{K}^{N\times N}$ normal, $\psi_{n,\mathbb{K}}(A) = \varphi_{n,\mathbb{K}}(A)$.*

*Proof.* See [5] for an alternate proof for the $\mathbb{K} = \mathbb{R}$ case. If $n \geq d(A)$ then we are done. Otherwise for $r$, $s \in \mathbb{K}^N$ and $\epsilon > 0$ small, let

$$r' = (1 - \epsilon^2)^{1/2} r + \epsilon s = (1 - \epsilon^2/2)r + \epsilon s + \mathcal{O}(\epsilon^4)$$

with $d(r, A) \geq n$. Also let $||r|| = ||s|| = 1$ and $r \perp s$, so $||r'|| = 1$. Let $P_r$ be the least squares polynomial (2a), (2b) of degree not greater than $n$ associated with $r$, and let $K_r = K_n(r, A)$, $K_s = K_n(s, A)$. Note that $P_r(A)r = r - AK_r[K_r^*A^*AK_r]^{-1}K_r^*A^*r$, and $P_r$ is a polynomial over $\mathbb{K}$. Note also that:

(i) $K_{r'} \equiv K_n(r', A) = (1 - \epsilon^2/2)K_r + \epsilon K_s + \mathcal{O}(\epsilon^3)$;

(ii) $K_{r'}^*A^*AK_{r'} = (1-\epsilon^2)K_r^*A^*AK_r + \epsilon K_r^*A^*AK_s + \epsilon K_s^*A^*AK_r + \epsilon^2 K_s^*A^*AK_s + \mathcal{O}(\epsilon^3)$;

(iii) $[K_{r'}^*A^*AK_{r'}]^{-1} = (1 + \epsilon^2)K_r^*A^*AK_r^{-1}$
$-\epsilon K_r^*A^*AK_r^{-1}[K_r^*A^*AK_s + K_s^*A^*AK_r]K_r^*A^*AK_r^{-1}$
$+\epsilon^2 K_r^*A^*AK_r^{-1}[K_r^*A^*AK_s + K_s^*A^*AK_r]$
$\cdot K_r^*A^*AK_r^{-1}[K_r^*A^*AK_s + K_s^*A^*AK_r]K_r^*A^*AK_r^{-1}$
$-\epsilon^2 K_r^*A^*AK_r^{-1}[K_s^*A^*AK_s]K_r^*A^*AK_r^{-1} + \mathcal{O}(\epsilon^3)$.

Then, since $s - AK_s[K_r^*A^*AK_r]^{-1}K_r^*A^*r = P_r(A)s$, we have after some manipulation

$$F_n(r') = [r^*P_r(A)^*P_r(A)r] + 2\epsilon \operatorname{Re}[s^*P_r(A)^*P_r(A)r]$$
$$+\epsilon^2\Big[s^*P_r(A)^*P_r(A)s - r^*P_r(A)^*P_r(A)r - [K_r^*A^*P_r(A)s + K_s^*A^*P_r(A)r]^*$$
$$\cdot[K_r^*A^*AK_r]^{-1}[K_r^*A^*P_r(A)s + K_s^*A^*P_r(A)r]\Big] + \mathcal{O}(\epsilon^3).$$

If $r$ is a (global) maximizer for $F_n$ (and thus a local maximizer), then $F_n$ is smooth near $r$, so we must have

$$\operatorname{Re} s^*P_r(A)^*P_r(A)r = 0, \tag{4a}$$

$$\begin{aligned} s^*P_r(A)^*P_r(A)s &- r^*P_r(A)^*P_r(A)r \\ &- [K_r^*A^*P_r(A)s + K_s^*A^*P_r(A)r]^*[K_r^*A^*AK_r]^{-1} \\ &\cdot [K_r^*A^*P_r(A)s + K_s^*A^*P_r(A)r] \leq 0 \end{aligned} \tag{4b}$$

for any $s \perp r$.

Now $\operatorname{Re} s^*P_r(A)^*P_r(A)r = 0$ for any $s \perp r$. If $\mathbb{K} = \mathbb{R}$, then $s^*P_r(A)^*P_r(A)r = 0$; otherwise, by replacement of $s$ with the quantity $(is)$, we obtain $\operatorname{Im} s^*P_r(A)^*P_r(A)r = 0$, and thus $s^*P_r(A)^*P_r(A)r = 0$ for all $s \perp r$. Thus $P_r(A)^*P_r(A)r$ is a scalar multiple of $r$, namely,

$$P_r(A)^*P_r(A)r = \frac{r^*P_r(A)^*P_r(A)r}{r^*r}r.$$

Thus $r$ is a (right) singular vector of $P_r(A)$.

If $r$ is a maximal singular vector (in the sense that its associated singular value is maximal among the singular values), then for all $v \in \mathbb{K}^N$,

$$\varphi_{n,\mathbb{K}}(A) \leq \frac{||P_r(A)v||}{||v||} \leq \frac{||P_r(A)r||}{||r||} = \psi_{n,\mathbb{K}}(A) \leq \varphi_{n,\mathbb{K}}(A),$$

and thus $P_r$ is a minimizer for $\varphi_{n,\mathbb{K}}(A)$ and we are done. Otherwise, there exists $s \perp r$ a maximal singular vector for $P_r(A)$, and $r$ is not a maximal singular vector.

Since $A$ is normal, then $A = U\Lambda U^*$, with $U$ unitary and $\Lambda$ diagonal. Then $U^*r$ is an eigenvector of $\Lambda^*\Lambda$, and is thus composed of a linear combination of standard unit basis vectors $e_{i_j}$, and similarly $U^*s$ is composed of a set of vectors $e_{i_k}$ all distinct from $e_{i_j}$. Since then $r$ and $s$ are contained in the linear spans of distinct (orthogonal) eigenspaces of $A$, it follows that $A^l r \perp A^m s$ for any $l, m$. Then (4b) reduces to

$$s^*P_r(A)^*P_r(A)s - r^*P_r(A)^*P_r(A)r \leq 0$$

or

$$r^* P_r(A)^* P_r(A) r \geq s^* P_r(A)^* P_r(A) s$$

which contradicts the maximality of $s$ over $r$. □

The results of this section have the following significance. In certain cases such as when $A$ is normal, a fixed polynomial preconditioning of degree $(s-1)$ exists independent of $r^{(0)}$, which has the same convergence bound as a cycle of GMRES($s$). In such cases it is reasonable to seek a polynomial preconditioner to replace (2) so that inner products need not be computed. Of course, since (2) finds the best polynomial for each $r^{(0)}$, the *average* behavior of (2) may be better than the average behavior of the best polynomial preconditioner, though the worst case behavior for a cycle applied to a given vector is the same.

It is not clear whether $\psi_{n,\mathbb{K}}(A) = \varphi_{n,\mathbb{K}}(A)$ holds for general $A$, though numerical experiments suggest that it does hold for cases other than $A$ normal and $n = 1$. The exact behavior of these functions is an open area of research.

The algorithm presented in §3 takes advantage of the cases when a fixed polynomial preconditioning of degree $s$ can be found that yields a convergence rate as good as that of the restarted GMRES algorithm with restart frequency $s$.

**3. Definition of the hybrid algorithm.** We wish to define a modification of the restarted GMRES algorithm that allows for alternate means of calculating the polynomial $P_n$ of (2) besides the least squares minimization of (2a), (2b).

To make use of the relevant inner product information from the Krylov spaces obtained from GMRES cycles in the history of a run, we define a fixed polynomial basis to be used to represent the Krylov space at each cycle. The use of a fixed polynomial basis simplifies the recording of Krylov space inner products for each cycle. In particular, we define the polynomials $\{p_i\}_{i=0}^{s}$ that are fixed independent of the cycle number, and use these to represent the Krylov space of each cycle by the matrix $\mathcal{P}_{s+1}(r) = [p_0(A)r \ldots p_s(A)r]$ and the associated inner product matrix $\mathcal{P}_{s+1}(r)^* \mathcal{P}_{s+1}(r)$. Means of choosing the polynomials so that the basis is well conditioned are described in [11]. Here we employ Chebyshev polynomials which possess a three-term recurrence.

The idea of the algorithm is as follows: given $r^{(ms)}$, the Krylov basis vectors $\mathcal{P}_{s+1}(r^{(ms)})$ may be generated, and then a decision can be made as to whether $r^{(ms+s)}$ is calculated by (2) or alternatively by some polynomial preconditioner represented in the basis $\mathcal{P}_{s+1}(r^{(ms)})$. The polynomial preconditioner may be based on information from the history of the run, in particular, the previous cycles that were performed in GMRES mode (2).

The structure of the algorithm is shown below. Here $m$ denotes the cycle number, $\mathcal{S}$ denotes the set of cycles performed in GMRES mode, and the parameter $t$ is a user-supplied tolerance, $0 \leq t \leq 1$.

1. Initialize. $m \leftarrow 0$.
2. Perform the stopping test.
3. Compute $r^{(s)}$ via a cycle of GMRES. $m \leftarrow 1$.
4. Determine well-conditioned basis polynomials $\{p_i\}$.
5. Determine $\mathcal{P}_{s+1}(r^{(0)})^* \mathcal{P}_{s+1}(r^{(0)})$ from the GMRES Arnoldi vector information from the first cycle (see below). Initialize $\mathcal{S} \leftarrow \{0\}$.
6. Perform the stopping test.
7. Select a polynomial preconditioner $q$, based on the information in $\{\mathcal{P}_{s+1}(r^{(is)})^* \cdot \mathcal{P}_{s+1}(r^{(is)})\}_{i \in \mathcal{S}}$, and represented in the basis $\{p_i\}$.
8. Compute $\tilde{r}^{(ms+s)} = [I - Aq(A)]r^{(ms)}$ and its norm.
9. Test: If $||\tilde{r}^{(ms+s)}||/||r^{(ms)}|| \leq [\max_{i \in \mathcal{S}} ||r^{(is+s)}||/||r^{(is)}||] \cdot (1-t) + [1] \cdot (t)$, then set $r^{(ms+s)} = \tilde{r}^{(ms+s)}$; else, add $m$ to $\mathcal{S}$, calculate the inner products

$\mathcal{P}_{s+1}(r^{(ms)})^*\mathcal{P}_{s+1}(r^{(ms)})$, and then compute $r^{(ms+s)}$ via the minimal residual method (2) using the existing basis vectors $\mathcal{P}_{s+1}(r^{(ms)})$.

10. Form $u^{(ms+s)}$ corresponding to $r^{(ms+s)}$.
11. Increment $m$; go to 6.

The details of executing the minimal residual component of step 9 are found in [11]. In particular,

$$u^{(ms+s)} = u^{(ms)} + \mathcal{P}_s(r^{(ms)})[\mathcal{P}_s(r^{(ms)})^* A^* A \mathcal{P}_s(r^{(ms)})]^{-1}\mathcal{P}_s(r^{(ms)})^* A r^{(ms)},$$

$$r^{(ms+s)} = r^{(ms)} - A\mathcal{P}_s(r^{(ms)})[\mathcal{P}_s(r^{(ms)})^* A^* A \mathcal{P}_s(r^{(ms)})]^{-1}\mathcal{P}_s(r^{(ms)})^* A r^{(ms)}.$$

The parameter $t$ controls how good the polynomial preconditioner is required to be in order to be used. If $t = 0$, then the polynomial preconditioner applied to the current cycle must do at least as well as the worst GMRES cycle in the history of the run; $t$ set closer to 1 allows a worse preconditioner to be tolerated, while any $t < 1$ insures convergence unless GMRES($s$) can stagnate for this matrix (i.e., $\psi_{s,\mathbb{K}}(A) = 1$). This equivalence of the convergence behavior of the new method to that of GMRES($s$) assumes, of course, that the Krylov basis for each cycle is well conditioned; for a full discussion of such considerations, see [11].

The calculation of the inner product information of step 5 from the GMRES information of that first cycle requires a more detailed description. Let $AQ_s = Q_{s+1}H_{s+1}$, where $H_{s+1}$ is the upper Hessenberg matrix from the Arnoldi sequence of the first GMRES cycle, with $Q_{s+1}$ the associated matrix of Arnoldi vectors, $Q_{s+1}^* Q_{s+1} = I$. We seek a new basis $\tilde{Q}_s \equiv \mathcal{P}_s(r^{(0)})$ such that $A\tilde{Q}_s = \tilde{Q}_{s+1}T_{s+1}$, where $T_{s+1}$ is upper Hessenberg (for the Chebyshev basis case, tridiagonal) and defines the set of recurrences associated with $\{p_i\}$ (see [11] for complete details). Let $\tilde{Q}_{s+1} = Q_{s+1}\tilde{T}$, so that $\tilde{T}$ is the change of basis matrix. We seek $\tilde{Q}_{s+1}^*\tilde{Q}_{s+1} = \tilde{T}^*\tilde{T}$. Note $\tilde{T}$ is upper triangular. After some manipulation we have

$$\tilde{T}T_{s+1} = H_{s+1}\begin{bmatrix} I_s \\ 0 \end{bmatrix}^* \tilde{T}\begin{bmatrix} I_s \\ 0 \end{bmatrix},$$

where $I_s$ is the identity. By applying the standard unit basis vector $e_i$ to the right side of each side of the equation, we obtain a recursion for the columns of $\tilde{T}$. Specifically, we let $\tilde{t}_i = \tilde{T}e_i$ and $T_{s+1} = \{t_{i,j}\}$, and then

$$t_{i+1,i}\tilde{t}_{i+1} = H_{s+1}\begin{bmatrix} I_s \\ 0 \end{bmatrix}^* \tilde{t}_i - \sum_{j=1}^{i} t_{j,i}\tilde{t}_j.$$

Finally, $\tilde{t}_1$ is a multiple of $e_1$ based on the scaling relationship between the initial basis vectors $Q_{s+1}e_1$ and $\tilde{Q}_{s+1}e_1$.

It should be emphasized that the algorithm presented thus far makes no restriction on the type of preconditioner used. Any type of preconditioner may be applied, and the mechanism of the algorithm insures that if the residual decrease is not adequate, then the cycle can be completed using GMRES. Furthermore, the cost of the algorithm when in GMRES mode is only slightly greater (about one SAXPY operation per step) than a standard GMRES-type algorithm for computing (2), due to the computation of $\tilde{r}^{(ms+s)}$.

We now define a particular preconditioner to be used in this study. This preconditioner is based on the inner product information from past GMRES cycles of the run and approximates the minimizing polynomial of $\varphi_{s,\mathbb{K}}(A)$.

Given a set of vectors $\{r_i\} \subseteq \mathbb{K}^N$, for $A \in \mathbb{K}^{N\times N}$, it is assumed that the inner product information contained in the matrices $K_{s+1}(r_i, A)^* K_{s+1}(r_i, A)$ is known. Then we note

$$\inf_{e_1^*c=1} \sup_{r_i} \frac{||K_{s+1}(r_i, A)c||}{||r_i||} \leq \inf_{e_1^*c=1} \sup_{r\in\mathbb{K}^N} \frac{||K_{s+1}(r, A)c||}{||r||}$$
$$= \inf_{P_s(0)=1} \sup_{r\in\mathbb{K}^N} \frac{||P_s(A)r||}{||r||} = \varphi_{s,\mathbb{K}}(A),$$

where $P_s(z) = \sum_{i=0}^{s}(e_{i+1}^*c)z^i$. Since each $F_i(c) = ||K_{s+1}(r_i, A)c||/||r_i||$ is convex (and in fact $F_i^2$ is quadratic), the function $F(c) = \sup_i F_i(c)$ is convex. Optimization routines may be used to find the minimizer of $F$ numerically. Furthermore, when the set $\{r_i\}$ is sufficiently expanded, $\inf_{e_1^*c=1} F(c)$ becomes closer to $\varphi_{s,\mathbb{K}}(A)$.

The algorithm considered here, then, uses a polynomial based on the minimizer of $F = \sup_i F_i$, which yields a polynomial that in some sense is optimal for all previous GMRES cycles.

Let us consider for a moment the computational cost of this algorithm. Following the analysis techniques of [10], let $w_{mv}$, $w_{\text{sax}}$, and $w_{\text{dot}}$ represent the computational cost on a given computer of the matrix-vector product $A \cdot v$, SAXPY $y \leftarrow y + \alpha x$, and dot product $u^*v$, respectively. The cost for an $s$-step cycle in polynomial preconditioning mode using Chebyshev basis polynomials is $[sw_{mv} + w_{\text{dot}} + (4s + 2)w_{\text{sax}}]$, while a cycle in GMRES mode adds $[(s + 1)(s + 2)/2]w_{\text{dot}} + (s + 1)w_{\text{sax}}$ to this. The number of cycles to converge to $||r^{(ms)}||/||r^{(0)}|| \leq \zeta$ is approximately $\log(\zeta)/\log((1 - t)\varphi_{s,\mathbb{K}}(A) + t)$, assuming $0 \leq t \leq 1$ for this analysis, since $||r^{(ms+s)}|| \leq ((1 - t)\varphi_{s,\mathbb{K}}(A) + t)||r^{(ms)}||$ is assured. If $l$ cycles in GMRES mode are required, then the total work for solution is approximately

$$l\left[\frac{(s + 1)(s + 2)}{2}w_{\text{dot}} + (s + 1)w_{\text{sax}}\right]$$
$$+[sw_{mv} + w_{\text{dot}} + (4s + 2)w_{\text{sax}}]\frac{\log(\zeta)}{\log[(1 - t)\varphi_{s,\mathbb{K}}(A) + t]}. \tag{5}$$

For comparison purposes we consider the case when $A$ is HPD. Then $\varphi_{s,\mathbb{K}}(A)$ may be approximated using Chebyshev polynomials to yield

$$\varphi_{s,\mathbb{K}}(A) \doteq \frac{1}{\cosh\left(s \log\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)\right)}, \tag{6}$$

where $\kappa = ||A|| \cdot ||A^{-1}||$ is the condition number of $A$. This approximation is most valid when the spectrum $\sigma(A)$ does not have large gaps.

For the purpose of this analysis we assume $l$ is fixed independently of $\kappa$; this may not hold in practice, depending in part on the strategy used to determine the preconditioner. Let $s = c\kappa^{\xi}$ and assume $\kappa$ is large. Now, if $\xi \geq \frac{1}{2}$, then (5) grows at least as fast as a multiple of $\kappa$. Otherwise, when $0 < \xi < \frac{1}{2}$, we expand by Taylor series for $\kappa$ large and after some manipulation and dropping of low order terms, we obtain from (5) the estimated asymptotic total cost,

$$w_\infty \equiv ls^2\frac{w_{\text{dot}}}{2} + \frac{\kappa}{s}[w_{mv} + 4w_{\text{sax}}]\frac{\log\zeta^{-1}}{2(1-t)}$$
$$= lc^2\kappa^{2\xi}\frac{w_{\text{dot}}}{2} + \frac{\kappa^{1-\xi}}{c}[w_{mv} + 4w_{\text{sax}}]\frac{\log\zeta^{-1}}{2(1-t)}. \tag{7}$$

It is clear that setting $\xi = \frac{1}{3}$ minimizes the maximal exponent of $\kappa$ and thus the growth rate of this quantity. In this case, $s$ grows at a rate proportional to $\kappa^{1/3}$, and the total work estimate

$w_\infty$ grows at a rate proportional to $\kappa^{2/3}$. Under these assumptions, minimizing the asymptotic total work estimate (7) over choices of the constant $c$ yields after some manipulation

$$s = \left[ \frac{w_{mv} + 4w_{\text{sax}}}{2w_{\text{dot}}} \frac{\log \zeta^{-1}}{l(1-t)} \kappa \right]^{1/3},$$

$$w_\infty = \frac{3}{2}(l w_{\text{dot}})^{1/3} \left[ (w_{mv} + 4w_{\text{sax}}) \frac{\log \zeta^{-1}}{2(1-t)} \right]^{2/3} \kappa^{2/3}.$$

These results show that for the HPD case under these assumptions, it is desirable to let the restart frequency $s$ for this algorithm grow in proportion to $\kappa^{1/3}$ to get best performance, and the resulting total work required (measured in vector operations) grows in proportion to $\kappa^{2/3}$.

This compares favorably with a total work growth rate proportional to $\kappa$ for standard restarted GMRES with the best choice of $s$ (see [10]), while it is not as good as the conjugate residual method whose growth rate is proportional to $\kappa^{1/2}$ due to its short recurrence. In other words, the new method under the given assumptions is able to recover part of the effective performance of short recurrence methods.

The algorithm presented here is similar to the hybrid GMRES algorithm of [16] but has significant differences. Both algorithms apply a cycle of GMRES to the linear problem and use the resulting polynomial as a preconditioner. However, if the polynomial is not adequate, the algorithm of [16] increases the value of $s$, in an attempt to form a better polynomial. On the other hand, the algorithm presented here keeps $s$ fixed but uses more GMRES cycles to improve the preconditioner.

**4. Numerical experiments.** We now present the results of numerical experiments with the new algorithm. For these experiments we consider the model problem

$$-u_{xx}(x, y) - u_{yy}(x, y) + Du_x(x, y) = g(x, y) \quad \text{on } \Omega = [0, 1]^2,$$

$$u(x, y) = 1 + xy \quad \text{on } \partial\Omega.$$

Here $D$ is constant, and $g$ is chosen so that the true solution is $u(x, y) = 1 + xy$ on $\Omega$. Five-point central differencing is applied, with uniform mesh spacing $h = 1/n_h$ in each direction. No preconditioning is used.

The mesh size for these experiments is $h = \frac{1}{128}$. Initial guess $u^{(0)} = 0$ is used, and a vector of random entries is used for $b$. In practice, a random vector may be used for $u^{(0)}$, as described in [16]. The stopping test $||r^{(n)}||/||r^{(0)}|| < \zeta = 10^{-8}$ is used.

The experiments are run in double precision on a Sun 4/330 Sparcstation with 25 Mhz clock speed. The optimization to find the polynomial preconditioner is performed by the GRG2 package of [13], [14], which uses a generalized reduced gradient technique for convex optimization. For each run, the approximation to the best polynomial for the previous cycle is used as the initial guess to the next call of the optimizer package to find an improved value.

The timings in Tables 1–6 are in seconds and the optimizer time is not counted. The case of $t = -\infty$ denotes the case when GMRES mode forced for every cycle. In these instances, the least squares polynomial (2) is used for each cycle, and the method reduces to the GMRES/Chebyshev basis algorithm described in [11], which typically gives the same iterates as standard GMRES at reduced cost. The performance values for these instances are indicative of the performance for the standard restarted GMRES algorithm for these cases.

Several conclusions may be drawn from these results (cf. [16], [17]):

1. In most cases only a couple of GMRES cycles are required to form an adequate preconditioner.

TABLE 1
*Timings; case of* $s = 20$.

| Meth \ Dh: | 0 | $2^{-3}$ | $2^{-2}$ | $2^{-1}$ | $2^{0}$ | $2^{1}$ | $2^{2}$ | $2^{3}$ | $2^{4}$ | $2^{5}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $t = -\infty$ | 1703 | 405 | 255 | 196 | 200 | 199 | 213 | 199 | 174 | 273 |
| $t = .2$ | 724 | 446 | 262 | 192 | 163 | 143 | 122 | 122 | 150 | 236 |
| $t = .5$ | 559 | 446 | 255 | 186 | 159 | 139 | 118 | 118 | 146 | 228 |

TABLE 2
*Total cycles/GMRES cycles; case of* $s = 20$.

| Meth \ Dh: | 0 | $2^{-3}$ | $2^{-2}$ | $2^{-1}$ | $2^{0}$ | $2^{1}$ | $2^{2}$ | $2^{3}$ | $2^{4}$ | $2^{5}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $t = -\infty$ | 143/143 | 33/33 | 21/21 | 16/16 | 16/16 | 16/16 | 17/17 | 16/16 | 14/14 | 22/22 |
| $t = .2$ | 86/25 | 61/3 | 35/2 | 25/2 | 21/2 | 18/2 | 15/2 | 15/2 | 19/2 | 31/2 |
| $t = .5$ | 62/24 | 63/2 | 35/2 | 25/2 | 21/2 | 18/2 | 15/2 | 15/2 | 19/2 | 31/2 |

TABLE 3
*Timings; case of* $s = 40$.

| Meth \ Dh: | 0 | $2^{-3}$ | $2^{-2}$ | $2^{-1}$ | $2^{0}$ | $2^{1}$ | $2^{2}$ | $2^{3}$ | $2^{4}$ | $2^{5}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $t = -\infty$ | 1354 | 498 | 406 | 477 | 412 | 412 | 421 | 360 | 354 | 385 |
| $t = .2$ | 748 | 404 | 294 | 253 | 229 | 201 | 188 | 187 | 201 | 244 |
| $t = .5$ | 681 | 418 | 302 | 262 | 238 | 207 | 195 | 191 | 201 | 246 |

TABLE 4
*Total cycles/GMRES cycles; case of* $s = 40$.

| Meth \ Dh: | 0 | $2^{-3}$ | $2^{-2}$ | $2^{-1}$ | $2^{0}$ | $2^{1}$ | $2^{2}$ | $2^{3}$ | $2^{4}$ | $2^{5}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $t = -\infty$ | 40/40 | 15/15 | 12/12 | 14/14 | 12/12 | 12/12 | 12/12 | 10/10 | 10/10 | 11/11 |
| $t = .2$ | 37/12 | 25/2 | 17/2 | 14/2 | 12/2 | 10/2 | 9/2 | 9/2 | 10/2 | 13/2 |
| $t = .5$ | 33/10 | 25/2 | 17/2 | 14/2 | 12/2 | 10/2 | 9/2 | 9/2 | 10/2 | 13/2 |

TABLE 5
*Timings; case of* $s = 60$.

| Meth\ Dh: | 0 | $2^{-3}$ | $2^{-2}$ | $2^{-1}$ | $2^{0}$ | $2^{1}$ | $2^{2}$ | $2^{3}$ | $2^{4}$ | $2^{5}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $t = -\infty$ | 1343 | 603 | 665 | 740 | 798 | 758 | 763 | 617 | 604 | 671 |
| $t = .2$ | 798 | 443 | 423 | 403 | 423 | 423 | 402 | 402 | 261 | 323 |
| $t = .5$ | 733 | 424 | 383 | 320 | 301 | 281 | 240 | 241 | 261 | 323 |

TABLE 6
*Total cycles/GMRES cycles; case of* $s = 60$.

| Meth \ Dh: | 0 | $2^{-3}$ | $2^{-2}$ | $2^{-1}$ | $2^{0}$ | $2^{1}$ | $2^{2}$ | $2^{3}$ | $2^{4}$ | $2^{5}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $t = -\infty$ | 21/21 | 9/9 | 10/10 | 11/11 | 12/12 | 11/11 | 11/11 | 9/9 | 9/9 | 10/10 |
| $t = .2$ | 21/8 | 15/2 | 14/2 | 13/2 | 14/2 | 14/2 | 13/2 | 13/2 | 8/1 | 11/1 |
| $t = .5$ | 31/1 | 16/1 | 14/1 | 11/1 | 10/1 | 9/1 | 7/1 | 7/1 | 8/1 | 11/1 |

2. The run times for the new method are typically better than those for the pure GMRES method (denoted by $t = -\infty$), sometimes by as much as a factor of three, which is a substantial savings. This is coupled with the fact that the GMRES/Chebyshev basis algorithm [11] used here for the $t = -\infty$ case is up to twice as fast as the standard Arnoldi-based GMRES implementation, leading to solution time reductions of up to a factor of six.

3. It is easier to determine a good polynomial when $s$ is chosen larger. This is consistent with the results of [16]. Of course, if $s = d(A) = d(r^{(0)}, A)$, then the polynomial from the first cycle matches the spectrum of $A$ exactly (in exact arithmetic).

4. For the cases for which only one GMRES cycle is performed, the result of the algorithm applied here is identical to that of the hybrid algorithm of [16], assuming the value of $s$ is forced to be fixed for that algorithm. For the cases for which more than one GMRES cycle is performed, the algorithms are different. In these cases, the polynomial $P_s$ from the first cycle of GMRES is found by this algorithm to be inadequate, and if used would lead to slower convergence or to divergence (cf. [17]). This is because a single GMRES cycle may yield insufficient information to insure that $||P_s(A)r||/||r|| \leq 1$ for all $r$, even though $||P_s(A)r^{(0)}||/||r^{(0)}|| \leq 1$ is imposed. Such problems may occur when $r^{(0)}$ lacks a representative distribution of eigencomponents. In such cases, the algorithm of [16] with fixed $s$ may diverge.

5. In the HPD cases, many GMRES cycles are required to obtain an effective polynomial. This is due in part to the fact that for the HPD case it is especially critical to estimate the smallest eigenvalue accurately, which is difficult. Furthermore, the polynomial preconditioner used here does not take advantage of the fact that for the HPD case, effective polynomials may be found based solely on the endpoints of the spectrum. Therefore, the algorithm spends a large amount of time making the polynomial small throughout the interior of the spectrum. On the other hand, when the matrix is nonnormal, the behavior of polynomials on the spectrum is less determinative of the convergence behavior of the polynomial preconditioner, and the approach used here has greater merit.

6. For these problems, $s = 20$ generally gave the best timings. This indicates a tradeoff between choosing $s$ large to keep small the required number of GMRES cycles for the preconditioner calculation, and choosing $s$ small to avoid the need to do even a single very large GMRES cycle. Furthermore, the aspect of GMRES which makes its average performance better than its worst case performance may be more pronounced when $s$ is large.

7. These experiments do not shed light on the growth rate of the best value of $s$ as the problem size grows, since only one problem size is used. The earlier asymptotic analysis suggests that at least for the HPD case, the best value of $s$ grows for larger problem sizes. However, this may not always occur in practice due to such factors as superlinear convergence of GMRES due to gaps in the spectrum of $A$, or the need for differing numbers of GMRES cycles as the problem size grows, factors not accounted for in the model given earlier.

8. Though the timings for the new algorithm are generally less than those for the $t = -\infty$ (pure GMRES($s$)) case, the difference is not as dramatic as might be expected. This is because more overall cycles are often required by the hybrid algorithm: the true minimal residual algorithm for each cycle has an *average* performance which may be better than the average performance of a good polynomial preconditioner, though the worst case bounds may be the same. The reasons for this are a topic of further study.

Next, in Table 7 the execution times for the convex optimizer are compared to the total run times, for the $s = 20$ case shown in Table 2.

It should be emphasized that this study does not fully investigate the issue of fast convex minimization algorithms for this problem. Alternate algorithms might give faster results; this is a topic for further research. On the other hand, it can be expected that for sufficiently large problems, the time for the optimization relative to the time for the rest of the algorithm will be small, since the optimization does not require the use of long vectors of size $N$. Thus, for many practical problems, an extremely fast minimizer may not be necessary.

The optimization times shown in Table 2 are fairly small compared to the total time, except for the HPD ($Dh = 0$) case. Even for the $s = 60$ case, the optimizer time did not exceed approximately 20% of the total time, except for the HPD case. On the other hand, when $A$ is HPD, as noted earlier, many cycles are executed in GMRES mode, requiring a large number of calls to the optimizer to revise the preconditioner. One possible remedy to avoid

TABLE 7
*Optimizer times / total times (seconds); case of $s = 20$.*

| Meth \ Dh: | 0 | $2^{-3}$ | $2^{-2}$ | $2^{-1}$ | $2^{0}$ | $2^{1}$ | $2^{2}$ | $2^{3}$ | $2^{4}$ | $2^{5}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $t = .2$ | 407/1131 | 10/456 | 5/267 | 8/200 | 5/168 | 5/148 | 7/129 | 8/130 | 10/160 | 7/243 |
| $t = .5$ | 2055/2614 | 6/452 | 5/260 | 7/193 | 5/164 | 5/144 | 7/125 | 8/126 | 10/156 | 7/235 |

such cases is to set an upper limit on the number of optimizations done, and when this number is exceeded the algorithm could revert to using standard restarted GMRES for every cycle. In any case, it should be emphasized that performing even a few optimizations yields a better polynomial than the simple first-cycle polynomial, which is employed by the fixed-$s$ version of the algorithm of [16].

In short, the algorithm shows benefit in terms of CPU savings over restarted GMRES, particularly for large values of $s$. Even greater savings may be expected on computer architectures such as distributed memory machines for which inner product operations are particularly costly.

**5. Conclusions.** In this paper we have shown results on the effectiveness of replacing GMRES with a polynomial preconditioning, defined an algorithm for safely adapting between GMRES and polynomial preconditioning, and defined a new preconditioning which approximates the minimax polynomial preconditioning for a given matrix.

The algorithm is shown to be effective compared to restarted GMRES for a number of cases. Further research is needed to shed more light on the behavior of polynomial methods for nonsymmetric problems.

REFERENCES

[1] H. C. ELMAN, Y. SAAD, AND P. SAYLOR, *A hybrid Chebyshev Krylov subspace algorithm for solving nonsymmetric systems of linear equations*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 840–855.
[2] V. FABER AND T. MANTEUFFEL, *Necessary and sufficient conditions for the existence of a conjugate gradient method*, SIAM J. Numer. Anal., 21 (1984), pp. 352–362.
[3] ———, *Orthogonal error methods*, SIAM J. Numer. Anal., 24 (1987), pp. 170–187.
[4] A. GREENBAUM, *Comparison of splittings used with the conjugate gradient algorithm*, Numer. Math. 33 (1979), pp. 181–194.
[5] A. GREENBAUM AND L. GURVITS, *Max-Min properties of matrix factor norms*, SIAM J. Sci. Comput., 15 (1994), pp. 348–358.
[6] A. GREENBAUM AND L. N. TREFETHEN, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–368.
[7] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
[8] A. S. HOUSEHOLDER, *The Theory of Matrices in Numerical Analysis*, Dover, New York, 1964.
[9] W. JOUBERT, *Iterative methods for the solution of nonsymmetric systems of linear equations*, Report CNA-242, Center for Numerical Analysis, The Univ. Texas at Austin, Feb. 1990.
[10] ———, *On the convergence behavior of the restarted GMRES algorithm for solving nonsymmetric linear systems*, J. Numer. Linear Algebra Appl., to appear.
[11] W. D. JOUBERT AND G. F. CAREY, *Parallelizable restarted iterative methods for nonsymmetric linear systems. Part* I: *Theory*, *Part* II: *Parallel implementation*, Internat. J. Computer Math., 44 (1992), pp. 243–267.
[12] W. D. JOUBERT AND D. M. YOUNG, *Necesssary and sufficient conditions for the simplification of generalized conjugate gradient algorithms*, Linear Algebra Appl., 88/89 (1987), pp. 449–485.
[13] L. S. LASDON AND A. D. WARREN, GRG2 *User's Guide*, unpublished report, Department of Management Science and Information Systems, The University of Texas at Austin, Dec. 1989.
[14] L. S. LASDON, A. D. WARREN, A. JAIN, AND M. RATNER, *Design and testing of a generalized reduced gradient code for nonlinear programming*, ACM Trans. Math. Software, 4 (1978), pp. 34–50.
[15] T. A. MANTEUFFEL, *Adaptive procedure for estimation of parameters for the nonsymmetric Tchebychev iteration*, Numer. Math., 31 (1978), pp. 187–208.

[16] N. M. NACHTIGAL, L. REICHEL, AND L. N. TREFETHEN, *A hybrid GMRES algorithm for nonsymmetric linear systems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 796–825.

[17] D. P. O'LEARY, *Yet another polynomial preconditioner for the conjugate gradient algorithm*, Linear Algebra Appl., 154–158 (1991), pp. 377–388.

[18] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.