SIAM J. MATRIX ANAL. APPL.
Vol. 34, No. 3, pp. 1066–1088

GMRES CONVERGENCE FOR PERTURBED COEFFICIENT MATRICES, WITH APPLICATION TO APPROXIMATE DEFLATION PRECONDITIONING*

JOSEF A. SIFUENTES†, MARK EMBREE‡, AND RONALD B. MORGAN§

Abstract. How does GMRES convergence change when the coefficient matrix is perturbed? Using spectral perturbation theory and resolvent estimates, we develop simple, general bounds that quantify the lag in convergence such a perturbation can induce. This analysis is particularly relevant to preconditioned systems, where an ideal preconditioner is only approximately applied in practical computations. To illustrate the utility of this approach, we combine our analysis with Stewart's invariant subspace perturbation theory to develop rigorous bounds on the performance of approximate deflation preconditioning using Ritz vectors.

Key words. GMRES, deflation preconditioning, resolvent, perturbation theory

AMS subject classifications. 65F10, 65F08, 47A10, 47A55

DOI. 10.1137/120884328

1. Introduction.

Large-scale systems of linear algebraic equations with nonsymmetric coefficient matrices are often solved using some variant of the generalized mininum residual (GMRES) algorithm [46]. In a variety of circumstances, GMRES can be analyzed for an idealized case that differs in some boundable manner from the actual system that is solved in practice. For example, certain preconditioners $\mathbf{M}^{-1}$ (e.g., involving exact inverses of certain constituents) give matrices $\mathbf{A}\mathbf{M}^{-1}$ with appealing spectral properties, yet practical considerations lead to approximate implementations that spoil the precise spectral structure. (For example, the saddle-point preconditioner in [41] yields a matrix $\mathbf{A}\mathbf{M}^{-1}$ with three distinct eigenvalues, which are rendered into three clusters by inexact implementations.) In another context, one might have a sequence of nearby linear systems that arise from sweeping through values of a physical parameter; see, e.g., [35].

How do such deviations affect GMRES convergence? We address this question by developing rigorous upper bounds on the maximum amount by which GMRES applied to $\mathbf{A}+\mathbf{E}$ can lag behind GMRES applied to $\mathbf{A}$, imposing no assumptions on $\mathbf{A}$ beyond nonsingularlity. Our main results (Theorems 2.1 and 2.3) rely on perturbation theory for the resolvent and thus account in a natural way for the potential sensitivity of the eigenvalues of the original matrix. Though the perturbation to $\mathbf{A}$ can move its eigenvalues significantly (as much as $O(\|\mathbf{E}\|^{1/n})$ as $\|\mathbf{E}\| \to 0$), Theorem 2.1 implies that the GMRES residual increases in norm by at most $O(\|\mathbf{E}\|)$. However, these bounds are not just asymptotic: they hold for a finite range of $\|\mathbf{E}\|$ values. To study

*Received by the editors July 12, 2012; accepted for publication (in revised form) by V. Simoncini May 13, 2013; published electronically July 18, 2013. The first and second authors were supported in part by National Science Foundation grant DMS-CAREER-0449973.

http://www.siam.org/journals/simax/34-3/88432.html

†Courant Institute of Mathematical Sciences, New York University, New York, NY 10012–1185 (sifuentes@courant.nyu.edu). This author was supported in part by a National Science Foundation graduate research fellowship, National Science Foundation grant DMS06-02235, and Air Force NSSEFF program award FA9550-10-1-018.

‡Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005–1892 (embree@rice.edu).

§Department of Mathematics, Baylor University, Waco, TX 76798–7328 (Ronald_Morgan@baylor.edu).

their merits and limitations, we illustrate these results for a matrix with a significant departure from normality.

We believe this approach to be widely applicable. To demonstrate its potential we analyze deflation preconditioning [1, 6, 16, 17, 21, 32], where an ideal preconditioner $\mathbf{M}^{-1}$ is constructed from some exact eigenvectors of $\mathbf{A}$. In practice one readily draws approximations to these eigenvectors from the Arnoldi process that underlies the GMRES algorithm. Combining our GMRES analysis with Stewart's invariant subspace perturbation theory [53, 54], we can bound convergence behavior for this practical preconditioner. While approximate deflation is often recommended as a practical alternative to the easily analyzed exact deflation (see, e.g., [1]), to our best knowledge this is the first result to rigorously bound how much small perturbations can slow the convergence of exact deflation preconditioners. Other approximately applied preconditioners could be similarly analyzed.

Simoncini and Szyld offer a complementary analysis for the general case that models inexact matrix-vector products at each GMRES step as a single aggregate perturbation to the coefficient matrix [50]. In that setting, early matrix-vector products must satisfy a strict accuracy requirement, which can be relaxed as convergence proceeds. This perspective is well-suited to cases where the perturbation changes at each iteration, e.g., as the result of a secondary iterative method whose convergence tolerance can be varied, and the resulting bounds involve quantities generated during the GMRES iteration. In contrast, our analysis models any fixed perturbation smaller than the distance of $\mathbf{A}$ to singularity, i.e., $\|\mathbf{E}\| < 1/\|\mathbf{A}^{-1}\|$. This is the best one can accomplish without imposing special structure on the perturbation $\mathbf{E}$, since for singular matrix GMRES stagnates with a residual bounded away from zero (unless $\mathbf{b}$ is special) [5]. Our resulting bounds rely primarily on spectral properties of the coefficient matrix, with little or no reliance on information generated by GMRES. Simoncini and Szyld have also analyzed how perturbations affect the superlinear convergence of GMRES [51, sect. 6], giving bounds that involve spectral projectors of the perturbed matrix. The resolvent approach taken below in section 2.2 could be used to relate these spectral projectors to those of the unperturbed problem.

Throughout, $\|\cdot\|$ denotes the vector 2-norm and the matrix norm it induces, $\sigma(\mathbf{A})$ denotes the set of eigenvalues of the matrix $\mathbf{A}$, $\mathbf{e}_k$ denotes the kth column of the identity matrix, and $\mathrm{Ran}(\cdot)$ denotes the range (column space).

2. GMRES on perturbed coefficient matrices. Consider a linear system $\mathbf{Ax} = \mathbf{b}$ with nonsingular $\mathbf{A} \in \mathbb{C}^{n\times n}$, $\mathbf{b} \in \mathbb{C}^n$, and unknown $\mathbf{x} \in \mathbb{C}^n$. Given the initial solution estimate $\mathbf{x}_0 = \mathbf{0} \in \mathbb{C}^n$, step k of the GMRES algorithm [47] finds the approximate solution $\mathbf{x}_k$ from the Krylov subspace

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \mathrm{span}\{\mathbf{b}, \mathbf{Ab}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\}$$

that minimizes the 2-norm of the residual $\mathbf{r}_k := \mathbf{b} - \mathbf{Ax}_k$:

$$\|\mathbf{r}_k\| = \min_{\widehat{\mathbf{x}}\in\mathcal{K}_k(\mathbf{A},\mathbf{b})} \|\mathbf{b} - \mathbf{A}\widehat{\mathbf{x}}\| \tag{2.1}$$

$$= \min_{q\in\mathcal{P}_{k-1}} \|\mathbf{b} - \mathbf{A}q(\mathbf{A})\mathbf{b}\|$$

$$= \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\mathbf{b}\|, \tag{2.2}$$

where $\mathcal{P}_d$ denotes the set of all polynomials of degree d or less.

How does a small modification to $\mathbf{A}$ affect GMRES convergence? Let $p_k \in \mathcal{P}_k$ denote the optimal polynomial in (2.2), and suppose $\mathbf{E} \in \mathbb{C}^{n\times n}$ is a perturbation that we presume to be small relative to $\mathbf{A}$, in a flexible way that the analysis will make precise. This perturbation gives a new system, $(\mathbf{A}+\mathbf{E})\boldsymbol{\xi} = \mathbf{b}$, and GMRES applied to that system (with initial guess $\boldsymbol{\xi}_0 = \mathbf{0}$) produces the iterate $\boldsymbol{\xi}_k$ with residual $\boldsymbol{\rho}_k := \mathbf{b} - (\mathbf{A}+\mathbf{E})\boldsymbol{\xi}_k$ satisfying

$$\|\boldsymbol{\rho}_k\| = \min_{\substack{\phi\in\mathcal{P}_k \\ \phi(0)=1}} \|\phi(\mathbf{A}+\mathbf{E})\mathbf{b}\|.$$

Let $\phi_k \in \mathcal{P}_k$ denote the polynomial that optimizes this last expression, so that

$$\|\boldsymbol{\rho}_k\| = \|\phi_k(\mathbf{A}+\mathbf{E})\mathbf{b}\| \le \|p_k(\mathbf{A}+\mathbf{E})\mathbf{b}\|.$$

Our ultimate concern is to understand how much the perturbation $\mathbf{E}$ impedes the convergence of GMRES. (Of course it is possible for this perturbation to accelerate convergence; in that case, one might switch the roles of $\mathbf{A}$ and $\mathbf{A}+\mathbf{E}$ in this analysis.) Toward this end, we can replace the optimal polynomial ϕ_k for the perturbed problem with p_k:

$$\begin{aligned}
\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\| &= \|\phi_k(\mathbf{A}+\mathbf{E})\mathbf{b}\| - \|p_k(\mathbf{A})\mathbf{b}\| \\
&\le \|p_k(\mathbf{A}+\mathbf{E})\mathbf{b}\| - \|p_k(\mathbf{A})\mathbf{b}\| \\
&\le \big\|\big(p_k(\mathbf{A}+\mathbf{E}) - p_k(\mathbf{A})\big)\mathbf{b}\big\| \\
&\le \|p_k(\mathbf{A}+\mathbf{E}) - p_k(\mathbf{A})\|\|\mathbf{b}\|. \qquad (2.3)
\end{aligned}$$

How much does $p_k(\mathbf{A}+\mathbf{E})$ differ from $p_k(\mathbf{A})$? Such an analysis can be approached by applying perturbation theory for functions of matrices. Expanding the polynomial term by term can yield crude results: if $p_k(z) = 1 + c_1 z + \cdots + c_k z^k$, then

$$p_k(\mathbf{A}+\mathbf{E}) = p_k(\mathbf{A}) + c_1\mathbf{E} + c_2(\mathbf{A}\mathbf{E} + \mathbf{E}\mathbf{A} + \mathbf{E}^2) + \cdots,$$

suggesting the coarse bound

$$\|p_k(\mathbf{A}+\mathbf{E}) - p_k(\mathbf{A})\| \le \sum_{j=1}^{k}\sum_{\ell=1}^{j} |c_j| \binom{j}{\ell} \|\mathbf{A}\|^{j-\ell}\|\mathbf{E}\|^{\ell}. \qquad (2.4)$$

Involving large powers of $\|\mathbf{A}\|$ and coefficients c_j (about which little is typically known), this bound is unlikely to be descriptive. Much more can be learned from analysis based on the spectrum and resolvent.

2.1. Spectral analysis. If $\mathbf{A}$ and $\mathbf{A}+\mathbf{E}$ are Hermitian, there exist unitary matrices $\mathbf{U}$ and $\widehat{\mathbf{U}}$ such that

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*, \qquad \mathbf{A}+\mathbf{E} = \widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}\widehat{\mathbf{U}}^*$$

for diagonal matrices $\boldsymbol{\Lambda}$ and $\widehat{\boldsymbol{\Lambda}}$. Using the unitary invariance of the norm,

$$\begin{aligned}
\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\| &\le \|\widehat{\mathbf{U}}p_k(\widehat{\boldsymbol{\Lambda}})\widehat{\mathbf{U}}^*\mathbf{b}\| - \|\mathbf{U}p_k(\boldsymbol{\Lambda})\mathbf{U}^*\mathbf{b}\| \\
&= \|p_k(\widehat{\boldsymbol{\Lambda}})\widehat{\mathbf{U}}^*\mathbf{b}\| - \|p_k(\boldsymbol{\Lambda})\mathbf{U}^*\mathbf{b}\| \\
&\le \|\big(p_k(\widehat{\boldsymbol{\Lambda}}) - p_k(\boldsymbol{\Lambda})\mathbf{U}^*\widehat{\mathbf{U}}\big)\widehat{\mathbf{U}}^*\mathbf{b}\| \\
&\le \|\big(p_k(\widehat{\boldsymbol{\Lambda}}) - p_k(\boldsymbol{\Lambda})\mathbf{U}^*\widehat{\mathbf{U}}\big)\|\|\mathbf{b}\|. \qquad (2.5)
\end{aligned}$$

Label the eigenvalues of $\mathbf{A}$ and $\mathbf{A}+\mathbf{E}$ as $\lambda_1 \le \cdots \le \lambda_n$ and $\widehat{\lambda}_1 \le \cdots \le \widehat{\lambda}_n$, so that $|\widehat{\lambda}_j - \lambda_j| \le \|\mathbf{E}\|$ by Weyl's theorem [4, Thm. VI.2.1]. Then

$$\begin{aligned}\frac{\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\|}{\|\mathbf{b}\|} &\le \max_{1\le j\le n} |p_k(\widehat{\lambda}_j) - p_k(\lambda_j)| + \|p_k(\boldsymbol{\Lambda})\|\|\mathbf{I} - \mathbf{U}^*\widehat{\mathbf{U}}\| \\ &= \max_{1\le j\le n} |p_k'(\lambda_j)|\|\mathbf{E}\| + \|p_k(\boldsymbol{\Lambda})\|\|\mathbf{I} - \mathbf{U}^*\widehat{\mathbf{U}}\| + \mathcal{O}(\|\mathbf{E}\|^2) \qquad (2.6)\end{aligned}$$

as $\|\mathbf{E}\| \to 0$, using first-order perturbation theory [31, Chap. 2]. The term $\|\mathbf{I} - \mathbf{U}^*\widehat{\mathbf{U}}\|$ measures the departure of the respective eigenvectors from biorthogonality: when $\mathbf{A}$ has several eigenvalues close together, this term can be considerably larger than $\|\mathbf{E}\|$.

For nonnormal $\mathbf{A}$, the style of analysis just described is significantly complicated by the nonorthogonality of the eigenvectors and the potential sensitivity of the eigenvalues (which can split like $O(\|\mathbf{E}\|^{1/d})$ for $d \times d$ Jordan blocks). Even then, as in the Hermitian case, the fruits of this labor only yield results that hold asymptotically as $\|\mathbf{E}\| \to 0$; one cannot readily establish a finite range of $\|\mathbf{E}\|$ values for which these bounds apply. Rather than pursuing this course in detail, we prefer analysis based on resolvent integrals that results in more descriptive quantitative bounds that are valid for finite values of $\|\mathbf{E}\|$.

2.2. Resolvent integrals and pseudospectra. Begin by writing the matrix function $p_k(\mathbf{A})$ as a Cauchy integral (see, e.g., [27, Chap. 1], [28, Chap. 5]):

$$p_k(\mathbf{A}) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma p_k(z)(z\mathbf{I} - \mathbf{A})^{-1}\,\mathrm{d}z, \qquad (2.7)$$

$$p_k(\mathbf{A}+\mathbf{E}) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma p_k(z)(z\mathbf{I} - \mathbf{A} - \mathbf{E})^{-1}\,\mathrm{d}z, \qquad (2.8)$$

where Γ is a finite union of Jordan curves in the complex plane whose interior contains the spectra of both $\mathbf{A}$ and $\mathbf{A}+\mathbf{E}$. While the eigenvalues of $\mathbf{A}$ and $\mathbf{A}+\mathbf{E}$ can differ considerably, the *resolvent* $(z\mathbf{I} - \mathbf{A})^{-1}$ is robust to perturbations. Indeed,

$$(z\mathbf{I} - \mathbf{A} - \mathbf{E})^{-1} = (\mathbf{I} - (z\mathbf{I} - \mathbf{A})^{-1}\mathbf{E})^{-1}(z\mathbf{I} - \mathbf{A})^{-1},$$

which is known as the *second resolvent identity*. If $\varepsilon := \|\mathbf{E}\| < 1/\|(z\mathbf{I} - \mathbf{A})^{-1}\| =: \delta$ (i.e., $\|\mathbf{E}\|$ is smaller than the distance of $z\mathbf{I} - \mathbf{A}$ to singularity), we can expand in a Neumann series to obtain

$$(z\mathbf{I} - \mathbf{A} - \mathbf{E})^{-1} = \sum_{j=0}^{\infty} \big((z\mathbf{I} - \mathbf{A})^{-1}\mathbf{E}\big)^j (z\mathbf{I} - \mathbf{A})^{-1}, \qquad (2.9)$$

from which it follows that

$$(z\mathbf{I} - \mathbf{A} - \mathbf{E})^{-1} - (z\mathbf{I} - \mathbf{A})^{-1} = \sum_{j=1}^{\infty} \big((z\mathbf{I} - \mathbf{A})^{-1}\mathbf{E}\big)^j (z\mathbf{I} - \mathbf{A})^{-1}.$$

The submultiplicativity of the matrix 2-norm yields the bound

$$\|(z\mathbf{I} - \mathbf{A} - \mathbf{E})^{-1} - (z\mathbf{I} - \mathbf{A})^{-1}\| \le \frac{\varepsilon/\delta}{\delta - \varepsilon}, \qquad (2.10)$$

which is small when $\varepsilon \ll \delta$. This analysis, for generic functions of $\mathbf{A}+\mathbf{E}$, was applied by Rinehart in 1956 [44] (see Higham [27, Prob. 3.4]) and more recently by Davies [10].

More concrete bounds on the difference (2.3) follow from strategic choices for the contour of integration Γ, e.g., as level sets of the resolvent norm: $\|(z\mathbf{I}-\mathbf{A})^{-1}\| = 1/\delta$ for all $z \in \Gamma$. This curve forms the boundary of the δ-pseudospectrum of $\mathbf{A}$,

$$\sigma_\delta(\mathbf{A}) = \{z \in \mathbb{C} : \|(z\mathbf{I}-\mathbf{A})^{-1}\| > 1/\delta\},$$

a set that can be defined equivalently in terms of perturbed eigenvalues:

$$\sigma_\delta(\mathbf{A}) = \{z \in \mathbb{C} : z \in \sigma(\mathbf{A}+\mathbf{D}) \text{ for some } \mathbf{D} \text{ with } \|\mathbf{D}\| < \delta\};$$

see, e.g., [57] for further details. From this latter definition it is evident that $\sigma(\mathbf{A}) \subset \sigma_\delta(\mathbf{A})$ for all $\delta > 0$. Denote the boundary of the δ-pseudospectrum by $\partial\sigma_\delta(\mathbf{A})$. Taking the contour Γ in (2.7)–(2.8) to be this boundary ensures that Γ will encircle all eigenvalues of $\mathbf{A}$ and, provided $\delta > \epsilon = \|\mathbf{E}\|$, all eigenvalues of $\mathbf{A}+\mathbf{E}$ as well. Conveniently, the requirement $\delta > \epsilon$ is precisely the condition $1/\|(z\mathbf{I}-\mathbf{A})^{-1}\| > \|\mathbf{E}\|$ that ensured convergence of the Neumann expansion of the perturbed resolvent (2.9).

Returning to the bound (2.3), we see

$$\begin{aligned}
\frac{\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\|}{\|\mathbf{b}\|} &\le \|p_k(\mathbf{A}+\mathbf{E}) - p_k(\mathbf{A})\| \\
&= \frac{1}{2\pi}\Big\| \int_\Gamma p_k(z)\big((z\mathbf{I}-\mathbf{A}-\mathbf{E})^{-1} - (z\mathbf{I}-\mathbf{A})^{-1}\big)\,\mathrm{d}z\Big\| \\
&\le \frac{L_\delta}{2\pi}\Big(\max_{z\in\Gamma}|p_k(z)|\Big)\Big(\max_{z\in\Gamma}\big\|(z\mathbf{I}-\mathbf{A}-\mathbf{E})^{-1} - (z\mathbf{I}-\mathbf{A})^{-1}\big\|\Big) \\
&\le \Big(\frac{\varepsilon}{\delta-\varepsilon}\Big)\Big(\frac{L_\delta}{2\pi\delta}\Big)\max_{z\in\Gamma}|p_k(z)|,
\end{aligned}$$

where L_δ denotes the arc length of $\Gamma = \partial\sigma_\delta(\mathbf{A})$. Applying the maximum modulus principle yields the following bound.

THEOREM 2.1. *Let $\mathbf{r}_k = p_k(\mathbf{A})\mathbf{b}$ denote the kth residual vector produced by GMRES applied to $\mathbf{A}\mathbf{x} = \mathbf{b}$ with the residual polynomial $p_k \in \mathcal{P}_k$ satisfying $p_k(0) = 1$. Then for all $\delta > 0$, the residual $\boldsymbol{\rho}_k$ produced by GMRES applied to $(\mathbf{A}+\mathbf{E})\mathbf{x} = \mathbf{b}$ satisfies*

$$\frac{\|\boldsymbol{\rho}_k\|}{\|\mathbf{b}\|} \le \frac{\|\mathbf{r}_k\|}{\|\mathbf{b}\|} + \left(\frac{\varepsilon}{\delta-\varepsilon}\right)\left(\frac{L_\delta}{2\pi\delta}\right)\sup_{z\in\sigma_\delta(\mathbf{A})}|p_k(z)|, \tag{2.11}$$

where $\|\mathbf{E}\| =: \varepsilon < \delta$ and L_δ denotes the arc length of the boundary of $\sigma_\delta(\mathbf{A})$.

As noted earlier, the perturbation $\mathbf{E}$ can change the eigenvalues of $\mathbf{A}$ significantly, as much as $\mathcal{O}(\varepsilon^{1/n})$ if $\mathbf{A}$ has an $n \times n$ Jordan block. While eigenvalues alone do not determine GMRES convergence [24], they are the first objects one typically examines when analyzing GMRES. Theorem 2.1 implies that, regardless of how much the eigenvalues change, the perturbation $\mathbf{E}$ increases the norm of the GMRES residual by at most $\mathcal{O}(\varepsilon)$ as $\varepsilon \to 0$, as we now make precise.

COROLLARY 2.2. *For $\varepsilon \ge 0$ and some $\mathbf{B} \in \mathbb{C}^{n\times n}$ with $\|\mathbf{B}\| = 1$, consider the family of linear systems $(\mathbf{A}+\varepsilon\mathbf{B})\mathbf{x}(\varepsilon) = \mathbf{b}$. Let $\mathbf{r}_k(\varepsilon)$ denote the residual produced by k steps of GMRES applied to this system (with $\mathbf{r}_k(\varepsilon) = \mathbf{0}$ if GMRES has converged*

exactly at or before the kth iteration). There exist constants C_k and ε_k (depending on $\mathbf{A}$, $\mathbf{b}$, *and* k*) such that for all* $\varepsilon \in [0, \varepsilon_k]$,

$$\|\mathbf{r}_k(\varepsilon)\| \leq \|\mathbf{r}_k(0)\| + C_k \varepsilon.$$

Proof. Pick any $\delta > 0$. Theorem 2.1 implies that

$$\|\mathbf{r}_k(\varepsilon)\| \leq \|\mathbf{r}_k(0)\| + \left(\frac{\varepsilon}{\delta - \varepsilon}\right)\left(\frac{L_\delta \|\mathbf{b}\|}{2\pi\delta}\right) \sup_{z \in \sigma_\delta(\mathbf{A})} |p_k(z)|.$$

Set $\varepsilon_k := \delta/2$ so that if $\varepsilon \in [0, \varepsilon_k]$, then $1/(\delta - \varepsilon) \leq 2/\delta$. The corollary follows with

$$C_k := \left(\frac{L_\delta \|\mathbf{b}\|}{\pi\delta^2}\right) \sup_{z \in \sigma_\delta(\mathbf{A})} |p_k(z)|. \qquad \square$$

We have just investigated Theorem 2.1 as the size of the perturbation, ε, goes to zero with the iteration number k fixed. Here the choice of $\delta > 0$ is arbitrary, only influencing the constant in the asymptotic term; one could have $0 \in \sigma_\delta(\mathbf{A})$. This corollary complements the more abstract discussion of the continuity of GMRES provided by Faber et al. [19, sect. 2].

In practice, we are often more interested in how the bound in Theorem 2.1 performs for fixed $\varepsilon > 0$ as k increases. Suppose $\|\mathbf{r}_k\|$ from the unperturbed problem converges steadily. Does the right-hand side of (2.11) ensure that the same will be true of $\|\boldsymbol{\rho}_k\|$? In this case the choice of δ is critical, since $\sup_{z \in \sigma_\delta(\mathbf{A})} |p_k(z)|$ is the only means by which the perturbation term in (2.11) can decrease to zero as k increases (ε now being fixed). How should one choose δ? If δ exceeds the distance of $\mathbf{A}$ to singularity, $\delta > 1/\|\mathbf{A}^{-1}\|$, then $0 \in \sigma_\delta(\mathbf{A})$, and since $p_k(0) = 1$ for all k, the term $\sup_{z \in \sigma_\delta(\mathbf{A})} |p_k(z)|$ in (2.11) cannot be smaller than one: hence the right-hand side of (2.11) does not go to zero as $\|\mathbf{r}_k\|$ does. Thus beyond a certain threshold, Theorem 2.1 would not give any insight into the convergence of $\boldsymbol{\rho}_k$. To avoid this problem, for a fixed perturbation size $\varepsilon = \|\mathbf{E}\|$, we restrict attention to $\delta \in (\varepsilon, 1/\|\mathbf{A}^{-1}\|)$; Figure 2.1 illustrates this setting in the complex plane. As δ decreases from $1/\|\mathbf{A}^{-1}\|$ down to ε, the term $L_\delta/(2\pi\delta)$ in (2.11) typically *increases*, while the polynomial approximation term *decreases* (as the maximization is posed on smaller and smaller nested sets). As evident from the numerical examples in section 3, larger values of δ usually give bounds that are sharper at early iterations, while smaller values of δ are often better at later iterations. To get the tightest overall bound, one should take the envelope of the upper bounds in (2.11) over a range of $\delta \in (\varepsilon, 1/\|\mathbf{A}\|)$, as illustrated in Figure 3.1.

Some additional hints about the choice of δ come from the GMRES algorithm itself. Recall that k steps of the Arnoldi process produce a factorization

$$\begin{aligned} \mathbf{A}\mathbf{V}_k &= \mathbf{V}_{k+1}\widetilde{\mathbf{H}}_k \\ &= \mathbf{V}_k\mathbf{H}_k + h_{k+1,k}\mathbf{v}_{k+1}\mathbf{e}_k^*, \end{aligned} \tag{2.12}$$

where the first j columns of $\mathbf{V}_{k+1} = [\mathbf{V}_k \ \mathbf{v}_{k+1}]$ form an orthonormal basis for the Krylov subspace $\mathcal{K}_j(\mathbf{A}, \mathbf{b})$ ($j = 1, \ldots, k+1$), the matrix $\widetilde{\mathbf{H}}_k \in \mathbb{C}^{(k+1)\times k}$ is upper Hessenberg, and $\mathbf{H}_k$ is the top $k \times k$ portion of $\widetilde{\mathbf{H}}_k$. The $(k+1) \times k$ entry of $\widetilde{\mathbf{H}}_k$,

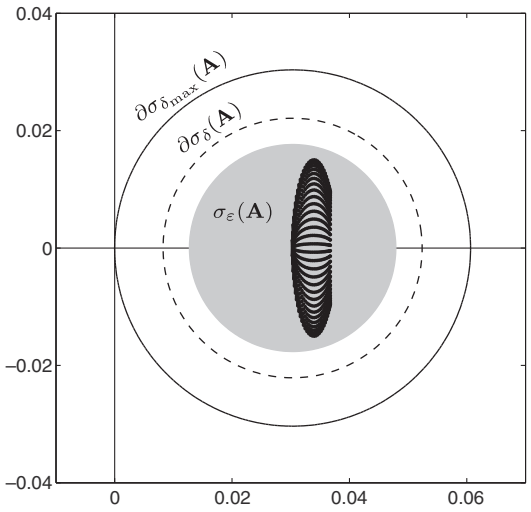

FIG. 2.1. *The proposed bounds involve several sets in the complex plane, illustrated here for a model problem described in section* 3. *Black dots denote the eigenvalues; the gray region shows* $\sigma_\varepsilon(\mathbf{A})$*: eigenvalues of* $\mathbf{A}+\mathbf{E}$ *can fall anywhere in this set. The outermost curve shows the boundary of* $\sigma_{\delta_{\max}}(\mathbf{A})$ *for* $\delta_{\max} = 1/\|\mathbf{A}^{-1}\|$*, which passes through the origin. Theorems* 2.1 *and* 2.3 *pertain to any* $\delta \in (\varepsilon, \delta_{\max})$*, such as the one for which the boundary* $\partial\sigma_\delta(\mathbf{A})$ *is shown by the dashed line.*

$h_{k+1,k}$, is a nonnegative real number. See, e.g., [46, Chap. 6] for further details. The eigenvalues $\theta_1, \ldots, \theta_k$ of $\mathbf{H}_k$, the *Ritz values*, satisfy

$$\theta_j \in \sigma_\gamma(\mathbf{A}) \quad \text{for all } \gamma > h_{k+1,k}. \tag{2.13}$$

The roots $\nu_1, \ldots, \nu_k$ of the residual polynomial p_k, called *harmonic Ritz values* [22, 36, 42], similarly satisfy

$$\nu_j \in \sigma_\gamma(\mathbf{A}) \quad \text{for all } \gamma > h_{k+1,k} + s_{\min}(\mathbf{H}_k)h_{k+1,k}^2, \tag{2.14}$$

where $s_{\min}(\mathbf{H}_k)$ is the smallest singular value of $\mathbf{H}_k$; see, e.g., [33, sect. 4.6], [49], [57, sect. 26]. Thus, one might take $\delta = h_{k+1,k}$ or $\delta = h_{k+1,k} + s_{\min}(\mathbf{H}_k)h_{k+1,k}^2$, though this only gives a rough starting point for larger values of δ. Our computational results illustrate that a range of δ values often gives more insight than can be derived from a single δ.

Theorem 2.1 requires knowledge of the GMRES residual polynomial p_k. There are settings where p_k is known explicitly.[1] In some such cases, Theorem 2.1 can give good insight [59]; in others, p_k can cause the bound (2.11) to be too large, e.g., when $\delta \gg 0$ and $p_k(z)$ is large for some $z \in \sigma_\delta(\mathbf{A})$, or when $\mathbf{b}$ has small components in certain eigenvectors of $\mathbf{A}$.

[1] One knows p_k in important special cases. Some ideal preconditioners yield a coefficient matrix whose minimal polynomial has low degree, implying that GMRES must converge in a few steps, e.g., three or four steps for the preconditioners in [29, 41]. Suppose $\mathbf{r}_k = p_k(\mathbf{A})\mathbf{b} = \mathbf{0}$. Then for any $m > 0$, in the above analysis one can take $\mathbf{r}_{k+m} = p_{k+m}(\mathbf{A})\mathbf{b} = p_k(\mathbf{A})q_m(\mathbf{A})\mathbf{b}$ for *any* degree m polynomial q_m with $q_m(0) = 1$. This allows one to compare cases where $\mathbf{r}_k = \mathbf{0}$ but $\boldsymbol{\rho}_k \neq \mathbf{0}$, e.g., when $\mathbf{b}$ is special for $\mathbf{A}$ but not $\mathbf{A}+\mathbf{E}$, or when $\mathbf{E}$ destroys special spectral structure in $\mathbf{A}$.

One can avoid p_k by using the series (2.9) to bound the perturbed residual

$$\|\boldsymbol{\rho}_k\| = \min_{\substack{\phi\in\mathcal{P}_k\\ \phi(0)=1}} \|\phi(\mathbf{A}+\mathbf{E})\mathbf{b}\| = \min_{\substack{\phi\in\mathcal{P}_k\\ \phi(0)=1}} \left\| \frac{1}{2\pi \mathrm{i}} \int_\Gamma \phi(z)(z\mathbf{I}-\mathbf{A}-\mathbf{E})^{-1}\,\mathrm{d}z\,\mathbf{b} \right\|$$

directly. Picking $\Gamma = \sigma_\delta(\mathbf{A})$ and bounding the norm of the integral as above, we obtain the following bound. Note that this new bound does not generalize Theorem 2.1, which bounds the lag of $\|\boldsymbol{\rho}_k\|$ behind $\|\mathbf{r}_k\|$.

THEOREM 2.3. *The residual $\boldsymbol{\rho}_k$ produced by GMRES applied to $(\mathbf{A}+\mathbf{E})\mathbf{x} = \mathbf{b}$ satisfies*

$$\frac{\|\boldsymbol{\rho}_k\|}{\|\mathbf{b}\|} \le \left(1 + \frac{\varepsilon}{\delta-\varepsilon}\right)\left(\frac{L_\delta}{2\pi\delta}\right) \min_{\substack{\phi\in\mathcal{P}_k\\ \phi(0)=1}} \sup_{z\in\sigma_\delta(\mathbf{A})} |\phi(z)|, \tag{2.15}$$

where $\|\mathbf{E}\| =: \varepsilon < \delta$ and L_δ denotes the arc length of the boundary of $\sigma_\delta(\mathbf{A})$.

Contrast the estimate (2.15) with Trefethen's bound on the GMRES residual norm for the unperturbed problem [56], [57, sect. 26]:

$$\frac{\|\mathbf{r}_k\|}{\|\mathbf{b}\|} \le \left(\frac{L_\delta}{2\pi\delta}\right) \min_{\substack{p\in\mathcal{P}_k\\ p(0)=1}} \sup_{z\in\sigma_\delta(\mathbf{A})} |p(z)|. \tag{2.16}$$

The bound (2.15) for the perturbed problem differs from the bound for the unperturbed problem (2.16) only by the k-independent scaling factor $\delta/(\delta-\varepsilon) > 1$. In fact, (2.15) can be manipulated to resemble (2.11). Denoting the right-hand side of (2.16) by $\eta_k/\|\mathbf{b}\|$, we have

$$\frac{\|\boldsymbol{\rho}_k\|}{\|\mathbf{b}\|} \le \frac{\eta_k}{\|\mathbf{b}\|} + \left(\frac{\varepsilon}{\delta-\varepsilon}\right)\left(\frac{L_\delta}{2\pi\delta}\right) \min_{\substack{p\in\mathcal{P}_k\\ p(0)=1}} \sup_{z\in\sigma_\delta(\mathbf{A})} |p(z)|, \tag{2.17}$$

which replaces the GMRES polynomial p_k in (2.11) with the best polynomial for the pseudospectral bound, at the expense of replacing the true residual norm $\|\mathbf{r}_k\|$ on the right-hand side with the upper bound $\eta_k \ge \|\mathbf{r}_k\|$. (This amounts to saying that Trefethen's bound (2.16) is robust to perturbations. Trefethen's bound may provide a good description of GMRES convergence [57, sect. 26], though an example of Greenbaum and Strakoš [25] illustrates that it is sometimes a significant overestimate.)

3. An illustrative example. To demonstrate the efficacy of this analysis, we apply our bounds to a well-studied model problem in the GMRES literature [18, 20, 34], the streamline-upwind/Petrov–Galerkin (SUPG) discretization of the two-dimensional convection-diffusion problem

$$-\nu\Delta u(x,y) + u_x(x,y) = 0$$

on $(x,y) \in [0,1]\times[0,1]$ with inhomogeneous Dirichlet boundary conditions that induce a boundary layer and an interior layer in the solution (u is zero on the boundary except for $u(x,0) = 1$ for $x \in [1/2,1]$ and $u(1,y) = 1$ for $y \in [0,1]$), a common test problem [14, 20, 34]. See Fischer et al. [20] for a full description of the problem, its discretization on an $N \times N$ grid of square elements that gives rise to a matrix $\mathbf{A}$ of dimension $n = N^2$, and the eigenvalues and eigenvectors of the discretization. (Having explicit formulas for the spectral information is especially helpful, given the

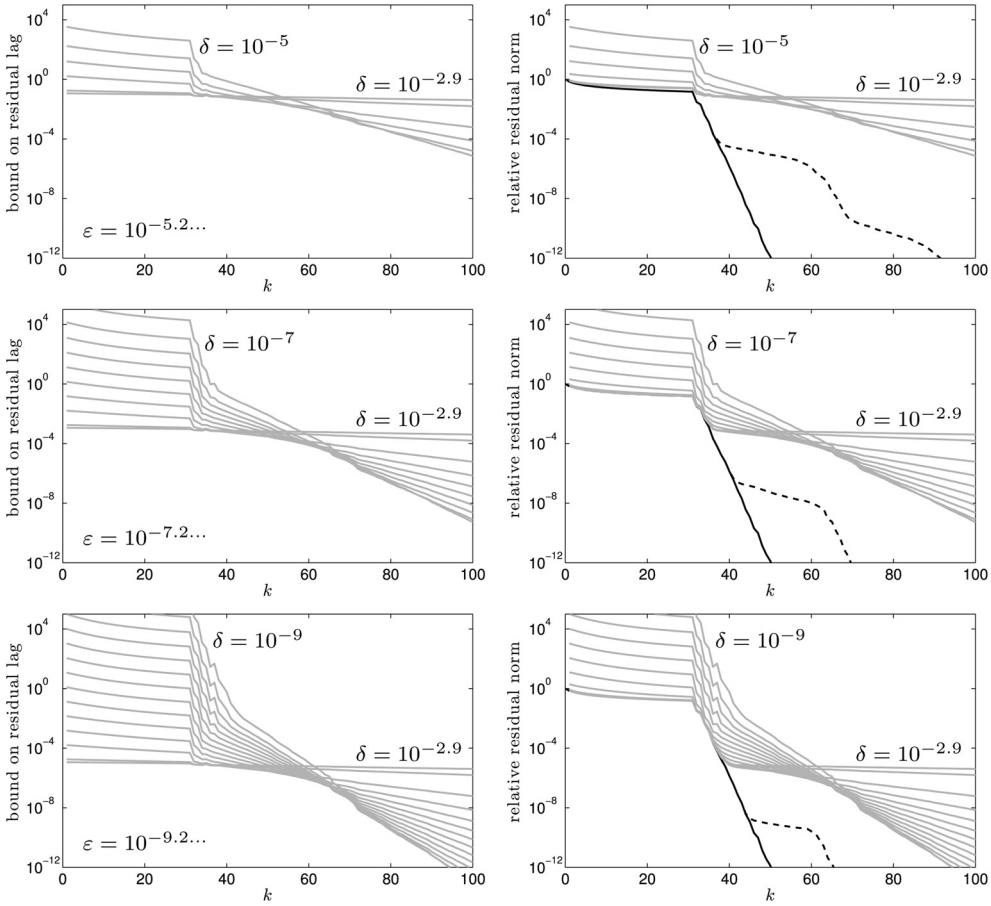

FIG. 3.1. *Example of Theorem* 2.1 *for the SUPG problem with* $N = 32$ *and* $\nu = .01$*, with a perturbation of size* $10^{-4}\|\mathbf{A}\|$ *(top),* $10^{-6}\|\mathbf{A}\|$ *(middle), and* $10^{-8}\|\mathbf{A}\|$ *(bottom), with* $\|\mathbf{A}\| = 10^{-1.2\cdots}$*. The plots on the left show the bound on* $\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\|$*; the plots on the right compare* $\|\mathbf{r}_k\|$ *(black solid line),* $\|\boldsymbol{\rho}_k\|$ *(dashed line), and the upper bound* (2.11) *on* $\|\boldsymbol{\rho}_k\|$ *for various* δ *(gray lines). Top:* $\delta = 10^{-2.9}, 10^{-3}, 10^{-3.5}, \ldots, 10^{-5} > \varepsilon = 10^{-5.2\cdots}$*; middle:* $\delta = 10^{-2.9}, 10^{-3}, 10^{-3.5}, \ldots, 10^{-7} > \varepsilon = 10^{-7.2\cdots}$*; bottom:* $\delta = 10^{-2.9}, 10^{-3}, 10^{-3.5}, \ldots, 10^{-9} > \varepsilon = 10^{-9.2\cdots}$.

strong departure from normality exhibited for small values of the diffusion parameter, ν.) We use this model problem simply to illustrate the efficacy of our analysis, even on a problem that has posed considerable challenge to GMRES analysts. For the large convection-diffusion problems encountered in practice, one should certainly use a more sophisticated approach than unpreconditioned full GMRES; see [15, Chap. 4] for good alternatives.

We begin with $\nu = .01$ and $N = 32$ ($n = 1024$), using the upwinding parameter advocated in [20]. This gives the matrix whose spectrum and pseudospectra were illustrated in Figure 2.1. (For $N = 8$, MATLAB computes the condition number of the matrix of eigenvectors of $\mathbf{A}$ to be roughly 4×10^9, a number that gets worse as N increases. Conventional analysis of GMRES based on the matrix of eigenvectors is not at all effective for this challenging problem.) Figure 3.1 shows approximations to the bound in Theorem 2.1 for perturbations $\mathbf{E}$ of size $10^{-4}\|\mathbf{A}\|$, $10^{-6}\|\mathbf{A}\|$, and $10^{-8}\|\mathbf{A}\|$, along with the convergence curve for the unperturbed problem and one

random choice of $\mathbf{E}$ (with entries drawn from the normal distribution), based on bounding the $\sigma_\delta(\mathbf{A})$ with a circle. Especially for the two smaller perturbations, our bound accurately describes the residual convergence to a level of $\|\mathbf{E}\|/\|\mathbf{A}\|$, before predicting a departure of $\|\boldsymbol{\rho}_k\|$ from $\|\mathbf{r}_k\|$ several orders of magnitude earlier than observed for these particular random perturbations.

Figure 3.2 reveals a subtlety that can emerge when applying Theorem 2.1. In this case, we take $\nu = .001$ and $N = 48$ ($n = 2304$) with a perturbation of size $\varepsilon = 10^{-4}\|\mathbf{A}\|$ and focus on iteration $k = 70$. The polynomial p_{70} (computed via harmonic Ritz values) is larger than one in magnitude at points $z \in \sigma_\varepsilon(\mathbf{A})$, so the polynomial term in (2.11) does not help make the bound small, as one might expect. While one often observes a correspondence between pseudospectra and lemniscates of the GMRES residual polynomial [55], this agreement is not always sufficiently close to yield useful bounds in Theorem 2.1 for a particular $\|\mathbf{E}\|$. In such circumstances, the flexibility afforded by Theorem 2.3 can be helpful. Figure 3.3 shows approximations to the bound from Theorem 2.3 for this example. Small values of δ appear to accurately capture the asymptotic convergence rate for the perturbed problem. (For careful computations of the pseudospectral bound (2.16) as in the bottom-left figure here, see [57, p. 252].)

To calculate the bounds in Figures 3.1 and 3.3, we replace $\sigma_\delta(\mathbf{A})$ (computed using EigTool [58]) with a slightly larger region Ω having a circular boundary, $\partial\Omega$. Since $\|(z\mathbf{I} - \mathbf{A})^{-1}\| \leq 1/\delta$ for all $z \in \partial\Omega$, the analysis in section 2 follows with Ω replacing $\sigma_\delta(\mathbf{A})$ and the length of $\partial\Omega$ replacing L_δ. We obtain the polynomial p_k in Theorem 2.1 from its roots, the computed harmonic Ritz values for the unperturbed problem. For Theorem 2.3, the optimal polynomial for the disk Ω centered at $c \in \mathbb{C}$ is given by $p_k(z) = (1 - z/c)^k$ [12, p. 553].

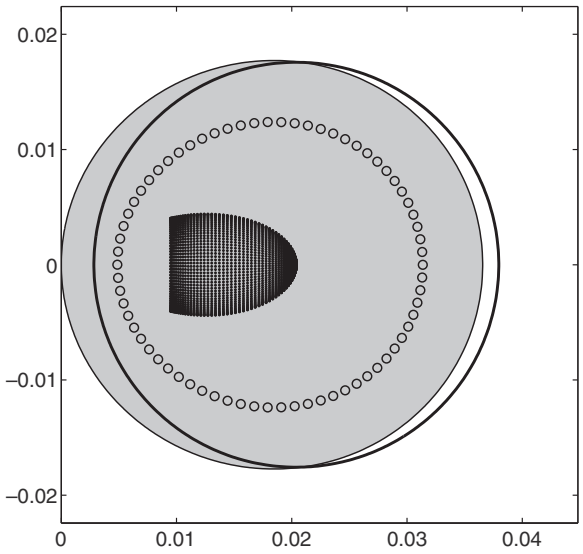

FIG. 3.2. *SUPG example for $N = 48$ and $\nu = .001$ with a perturbation of size $\varepsilon = 10^{-4}\|\mathbf{A}\|$. The solid dots denote the eigenvalues of $\mathbf{A}$; the gray region shows those $z \in \mathbb{C}$ for which $|p_{70}(z)| \leq 1$; the roots of p_{70} (the harmonic Ritz values) are denoted by $\circ$; the solid black line denotes the boundary of $\sigma_\varepsilon(\mathbf{A})$. Since $|p_{70}(z)| > 1$ for some $z \in \sigma_\varepsilon(\mathbf{A})$, the polynomial term in (2.11) does not contribute to convergence at iteration $k = 70$ for this value of ε. (It will contribute for smaller values of ε, for which $|p_{70}(z)| < 1$ for all $z \in \sigma_\varepsilon(\mathbf{A})$.)*

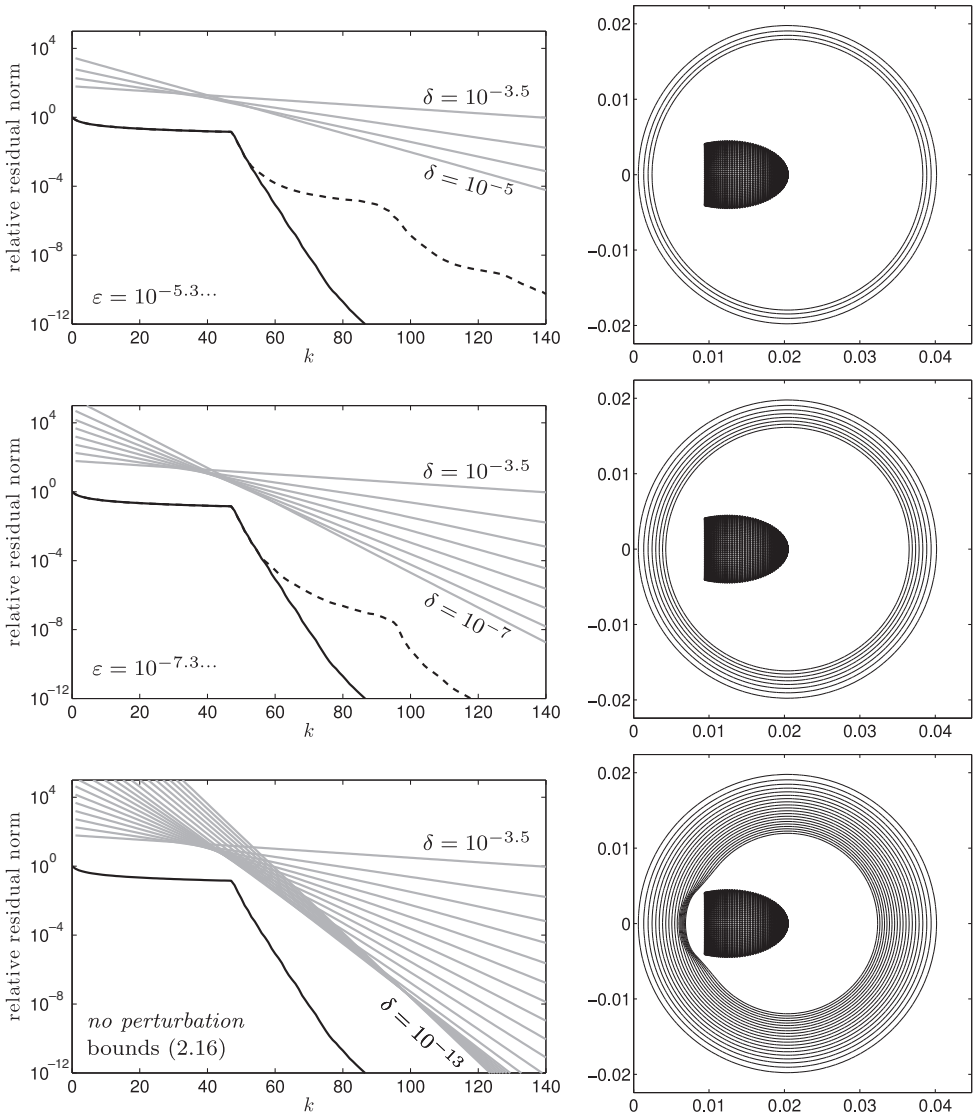

FIG. 3.3. *Example of Theorem* 2.3 *for the SUPG problem with* $N = 48$ *and* $\nu = .001$. *The left plots compare GMRES convergence for the unperturbed problem (solid black line) to that of perturbed problems (dashed black line) for perturbations of size* $10^{-4}\|\mathbf{A}\|$ *(top) and* $10^{-6}\|\mathbf{A}\|$ *(middle) with* $\|\mathbf{A}\| = 10^{-1.3\cdots}$. *The gray lines in the top two plots show the bound from Theorem* 2.3*; the gray lines in the bottom plot give the standard pseudospectra bound* (2.16) *for the unperturbed problem. The plots on the right show the boundaries of* $\sigma_\delta(\mathbf{A})$ *used for the bounds in the left with the* 2304 *eigenvalues of* $\mathbf{A}$ *appearing as a dark region inside these boundaries. On the top,* $\delta = 10^{-3.5}, 10^{-4}, \ldots, 10^{-5} > \varepsilon = 10^{-5.3\cdots}$*; in the middle,* $\delta = 10^{-3.5}, 10^{-4}, \ldots, 10^{-7} > \varepsilon = 10^{-7.3\cdots}$*; on the bottom,* $\delta = 10^{-3.5}, 10^{-4}, 10^{-4.5}, \ldots, 10^{-13}$.

4. Approximate deflation preconditioning. To illustrate how the preceding analysis can be applied in the context of preconditioning, we study the behavior of GMRES with approximate deflation. Deflation preconditioners are in a class of algorithms that apply knowledge (or estimates) of eigenvectors of $\mathbf{A}$ to remove certain components from the initial residual $\mathbf{r}_0 = \mathbf{b}$, thus accelerating convergence.

In practice, eigenvector approximations are built up from scratch during the iterations of GMRES. When the GMRES process is periodically restarted to allay growing computation and storage requirements, this approximate spectral information can be incorporated in various ways. In augmented subspace methods, the usual Krylov subspace from which approximate solutions are constructed is enlarged to include approximations to certain eigenvectors (e.g., those associated with eigenvalues near the origin); see, e.g., [8, 13, 37, 38, 39]. Deflation preconditioners differ in that they use (approximate) eigenvector information to build a matrix $\mathbf{M}^{-1}$ that maps certain eigenvalues of $\mathbf{AM}^{-1}$ (approximately) to one, while leaving the other eigenvalues of $\mathbf{A}$ fixed. Such methods have been developed in [1, 6, 11, 16, 17, 21, 32, 40, 43]; recently Gutknecht and coauthors have developed a general theoretical framework for deflation preconditioners [23, 26].

Here we consider a deflation preconditioner built with eigenvector estimates drawn from the Arnoldi process within GMRES. First we address the ideal case proposed by Erhel, Burrage, and Pohl [16, Thm. 3.1]. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a nonsingular matrix with (possibly repeated) eigenvalues $\lambda_1, \ldots, \lambda_n$, and suppose the columns of $\mathbf{X} \in \mathbb{C}^{n\times r}$ form an orthonormal basis for an r-dimensional invariant subspace of $\mathbf{A}$ associated with eigenvalues $\lambda_1, \ldots, \lambda_r$. Then the matrix

$$\mathbf{M}_D := \mathbf{I} - \mathbf{XX}^* + \mathbf{X}(\mathbf{X}^*\mathbf{AX})\mathbf{X}^*$$

is invertible, with

$$\mathbf{M}_D^{-1} = \mathbf{I} - \mathbf{XX}^* + \mathbf{X}(\mathbf{X}^*\mathbf{AX})^{-1}\mathbf{X}^*, \tag{4.1}$$

and $\sigma(\mathbf{AM}_D^{-1}) = \{1, \lambda_{r+1}, \lambda_{r+2}, \ldots, \lambda_n\}$. To see this, construct $\mathbf{Y} \in \mathbb{C}^{n\times(n-r)}$ such that $\mathbf{Z} = [\mathbf{X}\ \mathbf{Y}]$ is unitary, and notice that

$$\mathbf{Z}^*\mathbf{AM}_D^{-1}\mathbf{Z} = \begin{bmatrix} \mathbf{I} & \mathbf{X}^*\mathbf{AY} \\ \mathbf{0} & \mathbf{Y}^*\mathbf{AY} \end{bmatrix}. \tag{4.2}$$

In particular, the eigenvalue 1 has (algebraic and geometric) multiplicity of at least r. (One can readily modify $\mathbf{M}_D$ with any real $\alpha \neq 0$ such that if $\mathbf{M}_D := \mathbf{I} - \mathbf{XX}^* + \alpha^{-1}\mathbf{X}(\mathbf{X}^*\mathbf{AX})^{-1}\mathbf{X}^*$, then $\mathbf{M}_D^{-1} = \mathbf{I} - \mathbf{XX}^* + \alpha\mathbf{X}(\mathbf{X}^*\mathbf{AX})^{-1}\mathbf{X}^*$ and $\sigma(\mathbf{AM}_D^{-1}) = \{\alpha, \lambda_{r+1}, \lambda_{r+2}, \ldots, \lambda_n\}$; e.g., Erhel, Burrage, and Pohl use $\alpha = |\lambda_n|$.)

Since an exact invariant subspace is generally unavailable (too expensive to compute independently), we shall build the preconditioning matrix $\mathbf{M}^{-1}$ using *approximations* to an invariant subspace. In [6, 16, 40] harmonic Ritz vectors are used to approximate an invariant subspace, though other options are possible (e.g., using coarse-grid eigenvector estimates for a differential or integral operator [48]).

In this paper we consider preconditioners built with Arnoldi vectors as a basis for an approximate invariant subspace. This could be done by approximating $\mathbf{X}$ with the matrix $\mathbf{V}_m$ from (2.12) upon restart. From (2.12) one can see that $\mathbf{H}_m = \mathbf{V}_m^*\mathbf{AV}_m$, and so $\mathbf{H}_m$ is an orthogonal compression of $\mathbf{A}$ onto an m-dimensional subspace. The eigenvalues of $\mathbf{H}_m$ (the Ritz values of $\mathbf{A}$ with respect to the subspace $\mathcal{K}_m(\mathbf{A}, \mathbf{v}_1)$, discussed in section 2.2), can approximate eigenvalues of $\mathbf{A}$, while subspaces of $\mathcal{K}_m(\mathbf{A}, \mathbf{v}_1)$ approximate corresponding invariant subspaces of $\mathbf{A}$ [2, 3, 30, 45].

Typically few of these Ritz values will be accurate approximations to eigenvalues; this, together with memory constraints, suggests that one build an approximate deflation preconditioner using some subspace of $\mathrm{Ran}(\mathbf{V}_m) = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$. To do so, the Arnoldi factorization (2.12) is reordered using implicit QR steps, as in Sorensen's

implicitly restarted Arnoldi algorithm [52]. With this approach, *any* desired subset of r eigenvalues of $\mathbf{H}_m$ can be placed as the eigenvalues of the $r \times r$ upper-left portion of $\mathbf{H}_m$. This is precisely the restarted preconditioning strategy developed by Baglama et al. [1, Alg. 3.6], who suggest that their approach mimics exact deflation provided the approximate invariant subspace is sufficiently accurate. Using our perturbed GMRES analysis, we can provide a rigorous convergence theory for sufficiently accurate subspaces.

Suppose the Arnoldi factorization (2.12) has been reordered so that the first $r \leq m$ columns $\mathbf{V}_r$ of the Arnoldi basis matrix $\mathbf{V}_m$ define the approximate invariant subspace we wish to deflate. Replace the exact deflation preconditioner (4.1) with

$$\mathbf{M}^{-1} := \mathbf{I} - \mathbf{V}_r\mathbf{V}_r^* + \mathbf{V}_r\mathbf{H}_r^{-1}\mathbf{V}_r^*, \tag{4.3}$$

where $\mathbf{H}_r = \mathbf{V}_r^*\mathbf{A}\mathbf{V}_r \in \mathbb{C}^{r\times r}$ is upper Hessenberg, $\mathbf{V}_r \in \mathbb{C}^{n\times r}$ has orthonormal columns, and together these matrices satisfy an Arnoldi relation,

$$\mathbf{A}\mathbf{V}_r = \mathbf{V}_r\mathbf{H}_r + h_{r+1,r}\mathbf{v}_{r+1}\mathbf{e}_r^*. \tag{4.4}$$

The preconditioner (4.3) requires $\mathbf{H}_r$ to be invertible, which is *not* guaranteed by the invertibility of $\mathbf{A}$.

How well does $\mathrm{Ran}(\mathbf{V}_r)$ approximate an invariant subspace of $\mathbf{A}$? Among all choices of $\mathbf{L} \in \mathbb{C}^{r\times r}$, the norm of the residual $\mathbf{R}_r = \mathbf{A}\mathbf{V}_r - \mathbf{V}_r\mathbf{L}$ is minimized by $\mathbf{L} = \mathbf{V}_r^*\mathbf{A}\mathbf{V}_r = \mathbf{H}_r$ [54, Thm. IV.1.15], so $\mathbf{R}_r = \mathbf{A}\mathbf{V}_r - \mathbf{V}_r\mathbf{H}_r = h_{r+1,r}\mathbf{v}_{r+1}\mathbf{e}_r^*$, giving

$$\|\mathbf{R}_r\| = h_{r+1,r}. \tag{4.5}$$

In the unusual case that $\mathbf{R}_r = \mathbf{0}$, the columns of $\mathbf{V}_r$ give an orthonormal basis for an invariant subspace, and we have the exact setting (4.1): the preconditioner moves r eigenvalues of $\mathbf{A}\mathbf{M}^{-1}$ to one. When $\mathbf{R}_r \neq \mathbf{0}$, one only has an approximation to an invariant subspace, so perhaps it is curious that the preconditioner still moves $r-1$ eigenvalues exactly to one. (Of course, unlike exact deflation, the remaining $n-r+1$ eigenvalues can be far from the eigenvalues of $\mathbf{A}$.)

THEOREM 4.1. *Given a nonsingular matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, suppose $\mathbf{H}_r$ in the Arnoldi factorization* (4.4) *is invertible. Then the matrix*

$$\mathbf{M} := \mathbf{I} - \mathbf{V}_r\mathbf{V}_r^* + \mathbf{V}_r\mathbf{H}_r\mathbf{V}_r^*$$

is invertible, with

$$\mathbf{M}^{-1} = \mathbf{I} - \mathbf{V}_r\mathbf{V}_r^* + \mathbf{V}_r\mathbf{H}_r^{-1}\mathbf{V}_r^*,$$

and 1 *is an eigenvalue of $\mathbf{A}\mathbf{M}^{-1}$ with multiplicity of at least $r-1$. Furthermore, let $\mathbf{V} \in \mathbb{C}^{n\times n}$ be a unitary matrix of the form $\mathbf{V} = [\mathbf{V}_r\ \widehat{\mathbf{V}}_r]$, such that $\mathbf{V}^*\mathbf{A}\mathbf{V}$ is upper Hessenberg. Then the eigenvalues of $\mathbf{A}\mathbf{M}^{-1}$ are contained in the δ-pseudospectrum of*

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{I}_r & \mathbf{H}_r^{-1}\mathbf{J} \\ \mathbf{0} & \widehat{\mathbf{H}}_r \end{bmatrix}, \tag{4.6}$$

where $\widehat{\mathbf{H}}_r := \widehat{\mathbf{V}}_r^*\mathbf{A}\widehat{\mathbf{V}}_r$, and $\mathbf{J} := \mathbf{V}_r^*\mathbf{A}\widehat{\mathbf{V}}_r$, for any $\delta > \min_{\mathbf{L}\in\mathbb{C}^{r\times r}}\|\mathbf{A}\mathbf{V}_r - \mathbf{V}_r\mathbf{L}\| = h_{r+1,r}$.

Proof. The formula for $\mathbf{M}^{-1}$ can be verified by direct computation. We are given the unitary transformation of $\mathbf{A}$ to upper Hessenberg form,

$$\mathbf{H} = \mathbf{V}^*\mathbf{A}\mathbf{V} = \begin{bmatrix} \mathbf{H}_r & \mathbf{J} \\ h_{r+1,r}\mathbf{e}_1\mathbf{e}_r^* & \widehat{\mathbf{H}}_r \end{bmatrix}.$$

The unitary matrix $\mathbf{V}$ transforms the preconditioner $\mathbf{M}^{-1}$ to block diagonal form,

$$\mathbf{V}^*\mathbf{M}^{-1}\mathbf{V} = \begin{bmatrix} \mathbf{H}_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-r} \end{bmatrix}.$$

The unitary transformation of the product $\mathbf{M}^{-1}\mathbf{A}$ gives

$$\mathbf{V}^*\mathbf{M}^{-1}\mathbf{A}\mathbf{V} = \begin{bmatrix} \mathbf{I}_r & \mathbf{H}_r^{-1}\mathbf{J} \\ h_{r+1,r}\mathbf{e}_1\mathbf{e}_r^* & \widehat{\mathbf{H}}_r \end{bmatrix}$$

$$= \left[\begin{array}{c|c|c} \mathbf{I}_{r-1} & 0 & [\mathbf{H}_r^{-1}\mathbf{J}]_{1:r-1,1:n-r} \\ \hline 0 & 1 & \mathbf{e}_r^*\mathbf{H}_r^{-1}\mathbf{J} \\ \hline \mathbf{0} & \begin{matrix} h_{r+1,r} \\ 0 \\ \vdots \\ 0 \end{matrix} & \widehat{\mathbf{H}}_r \end{array}\right]. \tag{4.7}$$

Thus the spectrum of $\mathbf{M}^{-1}\mathbf{A}$ (and hence also the spectrum of the similar matrix $\mathbf{A}\mathbf{M}^{-1}$) contains the eigenvalue one with multiplicity $r-1$. All that remains is to note that (4.7) implies

$$\mathbf{V}^*\mathbf{M}^{-1}\mathbf{A}\mathbf{V} = \begin{bmatrix} \mathbf{I}_r & \mathbf{H}_r^{-1}\mathbf{J} \\ \mathbf{0} & \widehat{\mathbf{V}}_r^*\mathbf{A}\widehat{\mathbf{V}}_r \end{bmatrix} + h_{r+1,r}\,\mathbf{e}_{r+1}\mathbf{e}_r^*. \tag{4.8}$$

The fact that $h_{r+1,r} = \min_{\mathbf{L}\in\mathbb{C}^{r\times r}} \|\mathbf{A}\mathbf{V}_r - \mathbf{V}_r\mathbf{L}\|$ completes the proof. ∎

The invariant subspace perturbation theory developed by Stewart [53, 54] provides an analytical framework in which we can extend our understanding of approximate deflation techniques and thus provide rigorous convergence bounds for such restarted GMRES methods. Stewart and Sun explicitly show how an approximate invariant subspace can be perturbed into a nearby true invariant subspace. We translate the main result [54, Thm. V.2.1] into our notation here, so as to identify an invariant subspace approximated by the Arnoldi vectors. A key element is the "sep" of two matrices, a measure of the proximity of the spectra of two matrices that accounts for the sensitivity of the eigenvalues to perturbations:

$$\operatorname{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r) := \inf_{\substack{\mathbf{S}\in\mathbb{C}^{(n-r)\times r} \\ \|\mathbf{S}\|=1}} \|\mathbf{S}\mathbf{H}_r - \widehat{\mathbf{H}}_r\mathbf{S}\|. \tag{4.9}$$

The spectra of $\mathbf{H}_r$ and $\widehat{\mathbf{H}}_r$ are disjoint if and only if $\operatorname{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r) > 0$. For further properties of sep, see [54, sect. V.2].

THEOREM 4.2 (see [53, 54]). *Let* $\mathbf{A}$, $\mathbf{V}_r$, *and* $\mathbf{H}_r$ *satisfy the hypotheses of Theorem* 4.1. *If* $\sigma(\mathbf{H}_r)$ *and* $\sigma(\widehat{\mathbf{H}}_r)$ *are disjoint and*

$$\frac{h_{r+1,r}\|\mathbf{J}\|}{\operatorname{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)^2} < \frac{1}{4}, \tag{4.10}$$

then there exists a unique $\mathbf{P} \in \mathbb{C}^{(n-r)\times r}$ *such that*

$$\mathbf{X} = (\mathbf{V}_r + \widehat{\mathbf{V}}_r\mathbf{P})(\mathbf{I} + \mathbf{P}^*\mathbf{P})^{-1/2}, \tag{4.11}$$

$$\mathbf{Y} = (\widehat{\mathbf{V}}_r + \mathbf{V}_r\mathbf{P}^*)(\mathbf{I} + \mathbf{P}\mathbf{P}^*)^{-1/2} \tag{4.12}$$

with the properties that $\mathbf{Z} = [\mathbf{X}\ \mathbf{Y}]$ *is unitary,* $\operatorname{Ran}(\mathbf{X})$ *is a right invariant subspace of* $\mathbf{A}$, *and* $\operatorname{Ran}(\mathbf{Y})$ *is a left invariant subspace of* $\mathbf{A}$.

Stewart and Sun provide a telling corollary to Theorem 4.2 [54, Cor. V.2.2]. The tangent of the largest canonical angle between $\mathcal{K}_r(\mathbf{A}, \mathbf{v}_1)$ and $\mathrm{Ran}(\mathbf{X})$ is bounded by $\|\mathbf{P}\|$, which itself is bounded by $2h_{r+1,r}/\mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)$. This result describes how well the Krylov subspace approximates an invariant subspace as a function of the residual norm $h_{r+1,r}$ and provides the means to express approximate deflation preconditioning as a perturbation of exact deflation preconditioning of size proportional to $\|\mathbf{R}_r\|$.

THEOREM 4.3. *Let* $\mathbf{A}$, $\mathbf{V}$, *and* $\mathbf{H}_r$ *satisfy the hypotheses of Theorem* 4.2 *and let* $\mathbf{M}_D^{-1}$ *be the exact deflation preconditioner defined in* (4.1). *Then*

$$\|\mathbf{A}\mathbf{M}_D^{-1} - \mathbf{Z}\mathbf{V}^*\mathbf{A}\mathbf{M}^{-1}\mathbf{V}\mathbf{Z}^*\| \leq \|\mathbf{R}_r\|\|\mathbf{H}_r^{-1}\| + 2\|\mathbf{R}_r\|\|\mathbf{A}\|\left(2\varphi(\mathbf{V}_r) + \varphi(\mathbf{V}_r)^2\right),$$

where

$$\varphi(\mathbf{V}_r) := \left(\frac{8}{\mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)^2 + 4h_{r+1,r}^2 + \mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)\sqrt{\mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)^2 + 4h_{r+1,r}^2}}\right)^{1/2}.$$

Proof. The representations (4.11) and (4.12) for $\mathbf{X}$ and $\mathbf{Y}$ given by Stewart allow for the expression of these matrices as measurable pertubations of $\mathbf{V}_r$ and $\widehat{\mathbf{V}}_r$. Let $\mathbf{X} =: \mathbf{V}_r + \mathbf{E}_\mathbf{X}$ and $\mathbf{Y} =: \widehat{\mathbf{V}}_r + \mathbf{E}_\mathbf{Y}$, and write the full singular value decomposition of the matrix $\mathbf{P}$ from Theorem 4.2 as $\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^*$ (i.e., $\mathbf{U} \in \mathbb{C}^{(n-r)\times(n-r)}$, $\mathbf{\Sigma} \in \mathbb{C}^{(n-r)\times r}$ with upper $r \times r$ block denoted $\mathbf{\Sigma}_0$, and $\mathbf{W} \in \mathbb{C}^{r\times r}$). Using the formula for $\mathbf{X}$ from (4.11),

$$\begin{aligned}
\|\mathbf{E}_\mathbf{X}\| &= \|\mathbf{V}_r - (\mathbf{V}_r + \widehat{\mathbf{V}}_r\mathbf{U}\mathbf{\Sigma}\mathbf{W}^*)(\mathbf{I} + \mathbf{W}\mathbf{\Sigma}_0^2\mathbf{W}^*)^{-1/2}\| \\
&= \|\mathbf{V}_r\mathbf{W}(\mathbf{I} - (\mathbf{I} + \mathbf{\Sigma}_0^2)^{-1/2}) + \widehat{\mathbf{V}}_r\mathbf{U}\mathbf{\Sigma}(\mathbf{I} + \mathbf{\Sigma}_0^2)^{-1/2}\| \\
&= \left\| \begin{bmatrix} \mathbf{V}_r\mathbf{W} & \widehat{\mathbf{V}}_r\mathbf{U} \end{bmatrix} \begin{bmatrix} \mathbf{I} - (\mathbf{I} + \mathbf{\Sigma}_0^2)^{-1/2} \\ \mathbf{\Sigma}(\mathbf{I} + \mathbf{\Sigma}_0^2)^{-1/2} \end{bmatrix} \right\| \\
&= \sqrt{2 - \frac{2}{\sqrt{1 + \|\mathbf{P}\|^2}}}, \qquad (4.13)
\end{aligned}$$

since $[\mathbf{V}_r\mathbf{W} \;\; \widehat{\mathbf{V}}_r\mathbf{U}] \in \mathbb{C}^{n\times n}$ is unitary and $\|\mathbf{\Sigma}\| = \|\mathbf{\Sigma}_0\| = \|\mathbf{P}\|$.[2] A similar argument applied to $\|\mathbf{E}_\mathbf{Y}\|$ gives the same expression, so $\|\mathbf{E}_\mathbf{Y}\| = \|\mathbf{E}_\mathbf{X}\|$. Since this formula (4.13) increases monotonically with $\|\mathbf{P}\|$, we can bound the subspace perturbation using the fact that $\|\mathbf{P}\| \leq 2h_{r+1,r}/\mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)$ (as shown by Stewart and Sun [54, p. 231]) to obtain

$$\begin{aligned}
\|\mathbf{E}_\mathbf{X}\| = \|\mathbf{E}_\mathbf{Y}\| &\leq \sqrt{2 - \frac{2\,\mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)}{\sqrt{\mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)^2 + 4h_{r+1,r}^2}}} \\
&= \sqrt{\frac{8h_{r+1,r}^2}{\mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)^2 + 4h_{r+1,r}^2 + \mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)\sqrt{\mathrm{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)^2 + 4h_{r+1,r}^2}}}. \qquad (4.14)
\end{aligned}$$

[2] We thank Valeria Simoncini for pointing out the formula (4.13) for $\|\mathbf{E}_\mathbf{X}\| = \|\mathbf{E}_\mathbf{Y}\|$, which improves our earlier bound for these quantities.

To compare deflation with exact and inexact invariant subspaces, first note that

$$\mathbf{Z}^*\mathbf{A}\mathbf{M}_D^{-1}\mathbf{Z} = \begin{bmatrix} \mathbf{X}^*\mathbf{A}\mathbf{M}_D^{-1}\mathbf{X} & \mathbf{X}^*\mathbf{A}\mathbf{M}_D\mathbf{Y} \\ \mathbf{Y}^*\mathbf{A}\mathbf{M}_D^{-1}\mathbf{X} & \mathbf{Y}^*\mathbf{A}\mathbf{M}_D\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{X}^*\mathbf{A}\mathbf{Y} \\ \mathbf{0} & \mathbf{Y}^*\mathbf{A}\mathbf{Y} \end{bmatrix},$$

$$\mathbf{V}^*\mathbf{A}\mathbf{M}^{-1}\mathbf{V} = \begin{bmatrix} \mathbf{V}_r^*\mathbf{A}\mathbf{M}^{-1}\mathbf{V}_r & \mathbf{V}_r^*\mathbf{A}\mathbf{M}\widehat{\mathbf{V}}_r \\ \widehat{\mathbf{V}}_r^*\mathbf{A}\mathbf{M}^{-1}\mathbf{V}_r & \widehat{\mathbf{V}}_r^*\mathbf{A}\mathbf{M}\widehat{\mathbf{V}}_r \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{V}_r^*\mathbf{A}\widehat{\mathbf{V}}_r \\ \mathbf{e}_1\mathbf{e}_r^* h_{r+1,r}\mathbf{H}_r^{-1} & \widehat{\mathbf{V}}_r^*\mathbf{A}\widehat{\mathbf{V}}_r \end{bmatrix}.$$

Using the expressions for $\mathbf{X}$ and $\mathbf{Y}$ derived above,

$$\mathbf{X}^*\mathbf{A}\mathbf{Y} = \mathbf{V}_r^*\mathbf{A}\widehat{\mathbf{V}}_r + \mathbf{E}_\mathbf{X}^*\mathbf{A}\widehat{\mathbf{V}}_r + \mathbf{V}_r^*\mathbf{A}\mathbf{E}_\mathbf{Y} + \mathbf{E}_\mathbf{X}^*\mathbf{A}\mathbf{E}_\mathbf{Y},$$
$$\mathbf{Y}^*\mathbf{A}\mathbf{Y} = \widehat{\mathbf{V}}_r^*\mathbf{A}\widehat{\mathbf{V}}_r + \mathbf{E}_\mathbf{Y}^*\mathbf{A}\widehat{\mathbf{V}}_r + \widehat{\mathbf{V}}_r\mathbf{A}\mathbf{E}_\mathbf{Y}^* + \mathbf{E}_\mathbf{Y}^*\mathbf{A}\mathbf{E}_\mathbf{Y},$$

and so

$$\mathbf{Z}^*\mathbf{A}\mathbf{M}_D^{-1}\mathbf{Z} = \mathbf{V}^*\mathbf{A}\mathbf{M}^{-1}\mathbf{V} + \begin{bmatrix} \mathbf{0} & \mathbf{E}_\mathbf{X}^*\mathbf{A}\widehat{\mathbf{V}}_r + \mathbf{V}_r^*\mathbf{A}\mathbf{E}_\mathbf{Y} + \mathbf{E}_\mathbf{X}^*\mathbf{A}\mathbf{E}_\mathbf{Y} \\ -\mathbf{e}_1\mathbf{e}_r^* h_{r+1,r}\mathbf{H}_r^{-1} & \mathbf{E}_\mathbf{Y}^*\mathbf{A}\widehat{\mathbf{V}}_r + \widehat{\mathbf{V}}_r\mathbf{A}\mathbf{E}_\mathbf{Y}^* + \mathbf{E}_\mathbf{Y}^*\mathbf{A}\mathbf{E}_\mathbf{Y} \end{bmatrix}.$$

Norm inequalities then yield the rough bound

$$\begin{aligned} &\|\mathbf{Z}^*\mathbf{A}\mathbf{M}_D^{-1}\mathbf{Z} - \mathbf{V}^*\mathbf{A}\mathbf{M}^{-1}\mathbf{V}\| \\ &\quad \le h_{r+1,r}\|\mathbf{H}_r^{-1}\| + \|\mathbf{A}\|(\|\mathbf{E}_\mathbf{X}\| + \|\mathbf{E}_\mathbf{X}\|\|\mathbf{E}_\mathbf{Y}\| + 3\|\mathbf{E}_\mathbf{Y}\| + \|\mathbf{E}_\mathbf{Y}\|^2). \end{aligned} \tag{4.15}$$

Combining (4.14) with (4.15) yields the result stated in the theorem. ∎

This result gives a clean interpretation of how Krylov subspaces, and thus the span of Ritz vectors, approximate invariant subspaces. Thus it provides a framework for expressing approximate deflation as a bounded perturbation of exact deflation, facilitating the analysis of approximate deflation using the results of section 2. We seek a convergence bound for GMRES applied to the system $\mathbf{A}\mathbf{M}^{-1}\mathbf{y} = \mathbf{b}$ with approximate deflation preconditioning, which is equivalent to solving $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{x} = \mathbf{M}^{-1}\mathbf{y}$. To begin, define

$$\varepsilon := h_{r+1,r}\left(\|\mathbf{H}_r^{-1}\| + 8\|\mathbf{A}\|\left(\varphi(\mathbf{V}_r) + \varphi(\mathbf{V}_r)^2 h_{r+1,r}\right)\right), \tag{4.16}$$

where $\varphi(\mathbf{V}_r)$ is defined in Theorem 4.3.

THEOREM 4.4. *Let $\|\boldsymbol{\rho}_k\|$ denote the norm of the kth residual produced by GMRES applied to $\mathbf{A}\mathbf{M}^{-1}\mathbf{y} = \mathbf{b}$. If the hypotheses of Theorem* 4.3 *are satisfied, then*

$$\frac{\|\boldsymbol{\rho}_k\|}{\|\mathbf{b}\|} \le \frac{L_\delta}{2\pi\delta}\left(1 + \frac{\varepsilon}{\delta - \varepsilon}\right) \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \max_{z\in\sigma_\delta(\mathbf{A}\mathbf{M}_D^{-1})} |p(z)| \tag{4.17}$$

where ε is defined in (4.16) *and $\delta > \varepsilon$.*

Proof. The system $\mathbf{A}\mathbf{M}^{-1}\mathbf{y} = \mathbf{b}$ is equivalent to

$$(\mathbf{Z}\mathbf{V}^*)\mathbf{A}\mathbf{M}^{-1}(\mathbf{V}\mathbf{Z}^*)\widehat{\mathbf{y}} = \widehat{\mathbf{b}}, \tag{4.18}$$

where $\widehat{\mathbf{y}} := \mathbf{Z}\mathbf{V}^*\mathbf{y}$ and $\widehat{\mathbf{b}} := \mathbf{Z}\mathbf{V}^*\mathbf{b}$, and since $\mathbf{Z}\mathbf{V}^*$ is unitary, the relative residual norm for GMRES applied to (4.18) is identical to that produced for the original problem. By Theorem 4.3, we can write

$$(\mathbf{Z}\mathbf{V}^*)\mathbf{A}\mathbf{M}^{-1}(\mathbf{V}\mathbf{Z}^*) = \mathbf{A}\mathbf{M}_D^{-1} + \mathbf{E},$$

where $\|\mathbf{E}\| \le \varepsilon$. The bound (4.17) follows from application of Theorem 2.3. ∎

Combining the same argument with Theorem 2.1 yields another bound.

THEOREM 4.5. *Let $\boldsymbol{\rho}_k$ denote the kth residual produced by GMRES applied to $\mathbf{A}\mathbf{M}^{-1}\mathbf{y} = \mathbf{b}$ and let $\mathbf{r}_k = p_k(\mathbf{A}\mathbf{M}_D^{-1})\mathbf{b}$ denote the kth residual produced by GMRES applied to the exactly deflated system $\mathbf{A}\mathbf{M}_D^{-1}\mathbf{y} = \mathbf{b}$. If the hypotheses of Theorem* 4.3 *are satisfied, then*

$$\frac{\|\boldsymbol{\rho}_k\|}{\|\mathbf{b}\|} \le \frac{\|\mathbf{r}_k\|}{\|\mathbf{b}\|} + \left(\frac{\varepsilon}{\delta - \varepsilon}\right)\left(\frac{L_\delta}{2\pi\delta}\right) \sup_{z \in \sigma_\delta(\mathbf{A}\mathbf{M}_D^{-1})} |p_k(z)|, \tag{4.19}$$

where ε is defined in (4.16) *and $\delta > \varepsilon$.*

5. Deflation preconditioning example. We illustrate the analysis of the last section with an example designed to show the merits of (approximate) deflation preconditioning. (For examples of effective approximate deflation for practical problems, see [1] and [48, Chap. 5].) Consider the block diagonal matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix}, \tag{5.1}$$

where

$$\mathbf{A}_1 = \begin{bmatrix} -0.01 & & & \\ & -0.02 & & \\ & & \ddots & \\ & & & -0.10 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 2 & 1.25 & & \\ & \ddots & \ddots & \\ & & \ddots & 1.25 \\ & & & 3 \end{bmatrix},$$

with $\mathbf{A}_1 \in \mathbb{C}^{10\times 10}$ and $\mathbf{A}_2 \in \mathbb{C}^{190\times 190}$ having its eigenvalues uniformly distributed in the interval $[2, 3]$. With eigenvalues close to the origin, the $\mathbf{A}_1$ block causes GMRES to initially converge slowly. After sufficiently many iterations GMRES can effectively eliminate $\mathbf{A}_1$, and convergence will proceed at a rate governed by the spectral properties of $\mathbf{A}_2$; see [7, 12]. (The 1.25 on the superdiagonal of $\mathbf{A}_2$ adds nonnormality, slowing convergence from what one would expect from a diagonal matrix whose spectrum is bounded in the interval $[2, 3]$.) If GMRES is restarted before $\mathbf{A}_1$ has been effectively eliminated, convergence will be slow indeed, as seen in our later computations.

Figure 5.1 shows convergence of full GMRES for the unpreconditioned problem and several flavors of deflation preconditioning. In all cases, we use $n = 200$ and each entry of the right-hand side vector $\mathbf{b} = \mathbf{r}_0$ is $1/\sqrt{n}$. (To ease the computation of our bounds, we use the shifted preconditioners $\mathbf{M}_D^{-1} = \mathbf{I} - \mathbf{X}\mathbf{X}^* + \alpha\mathbf{X}(\mathbf{X}^*\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^*$ and $\mathbf{M}^{-1} = \mathbf{I} - \mathbf{V}_r\mathbf{V}_r^* + \alpha\mathbf{V}_r\mathbf{H}_r^{-1}\mathbf{V}_r^*$ for $\alpha = 3/2$, which places r or $r - 1$ eigenvalues at α, rather than one.)

Exact deflation uses the eigenvectors associated with the $r = 10$ leftmost eigenvalues (from $\mathbf{A}_1$). For approximate deflation, we use the implicitly restarted Arnoldi algorithm [52] to develop increasingly accurate approximations to the eigenspace used in exact deflation while restricting the overall subspace size, as advocated for deflation preconditioning in [1]. At each restarted Arnoldi cycle, we expand the $r = 10$ dimensional approximate subspace into a larger subspace of dimension $r + p = 20$. The rightmost $p = 10$ Ritz values from this subspace are then used as shifts that drive

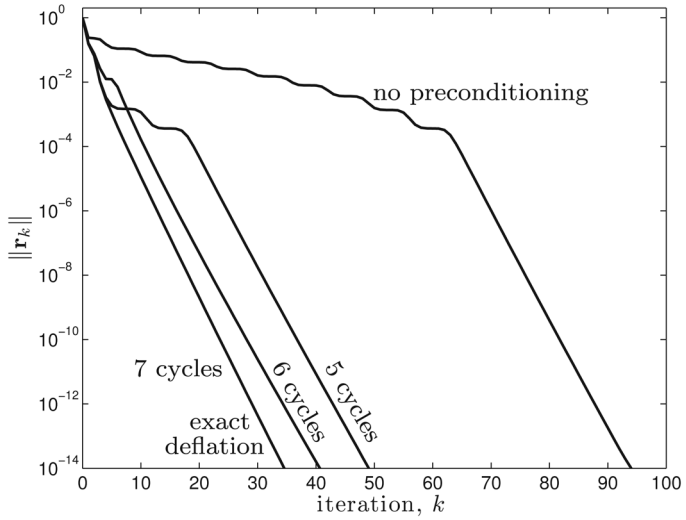

FIG. 5.1. *Full GMRES convergence for the unpreconditioned matrix* (5.1)*, compared to exact and approximate deflation. Exact deflation significantly accelerates convergence. The efficacy of approximate deflation depends on the accuracy of the approximated invariant subspace; here "cycle" refers to the number of restarted Arnoldi cycles used to develop that subspace. Seven cycles are required before the sufficient condition* 5.3 *for the theorems in section* 4 *holds; this approximation yields convergence that is essentially the same as exact deflation.*

the approximate subspace closer to the eigenspace for the leftmost $r = 10$ eigenvalues. How many cycles are required before the approximate invariant subspace produces a deflation preconditioner that satisfies the hypotheses of Theorems 4.4 and 4.5?

To check these hypotheses, we must evaluate the condition (4.10). The quantity $\text{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r)$ is not easy to compute directly, but it can be bounded [53, Thm. 4.10]:

$$\frac{\text{sep}_F(\mathbf{H}_r, \widehat{\mathbf{H}}_r)}{\sqrt{r}} \leq \text{sep}(\mathbf{H}_r, \widehat{\mathbf{H}}_r) \leq \text{sep}_F(\mathbf{H}_r, \widehat{\mathbf{H}}_r), \tag{5.2}$$

where sep_F uses the Frobenius norm instead of the 2-norm used for sep in (4.9). Since $\text{sep}_F(\mathbf{H}_r, \widehat{\mathbf{H}}_r)$ is the smallest singular value of $\mathbf{I}_r \otimes \widehat{\mathbf{H}}_r - \mathbf{H}_r^T \otimes \mathbf{I}_{n-r}$, we can verify

$$\frac{r h_{r+1,r} \|\mathbf{J}\|}{\text{sep}_F(\mathbf{H}_r, \widehat{\mathbf{H}}_r)^2} < \frac{1}{4}, \tag{5.3}$$

which is a *stricter* requirement than (4.10). (Thus to evaluate sep_F, we compute a singular value of a matrix of size $r(n-r) \times r(n-r)$. Indeed, to obtain the coefficient matrices $\mathbf{H}_r$ and $\widehat{\mathbf{H}}_r$, we compute a full n-dimensional Arnoldi decomposition (with DGKS reorthogonalization [9, 52]) for the experiments shown here. While such conditions cannot reasonably be verified in practice, we do so here to illuminate the hypothesis in question.)

To compute m cycles of the restarted Arnoldi method requires $r + pm$ matrix vector products with $\mathbf{A}$. For this problem, good approximations of the desired eigenspace begin to emerge after seven cycles. At cycle 6, the left-hand side of (5.3) is roughly $3.81 > 1/4$, so the condition is violated (though the preconditioner is effective); at cycle 7, (5.3) drops abruptly to $2.7 \times 10^{-7} < 1/4$, and our theorems hold. Figure 5.1 shows how the degree to which the approximate deflation preconditioners

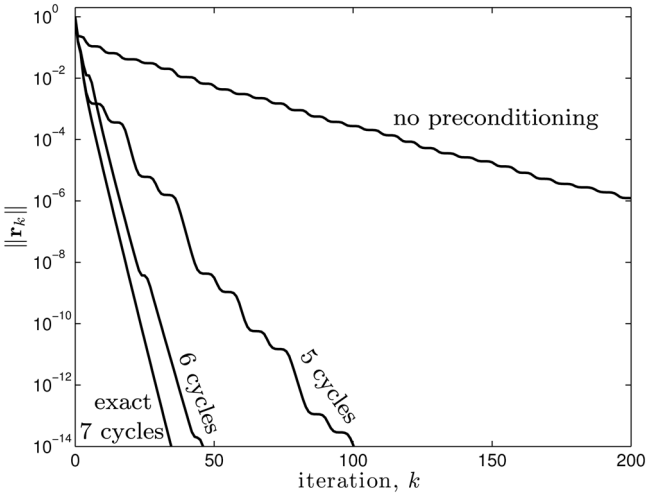

FIG. 5.2. *GMRES*(20) *convergence for the unpreconditioned matrix* (5.1)*, compared to exact and approximate deflation. For exact deflation and its accurate approximation (formed from seven cycles of restarted Arnoldi) convergence is similar to full GMRES; performance degrades for the less accurate approximations.*

mimic the performance of the exact depends on the accuracy of the subspace. This is seen more clearly if we replace full GMRES with a restarted version of the algorithm. Figure 5.2 repeats the above experiment, but now with GMRES(20) (since we use subspaces of up to dimension $r + p = 20$ in the restarted Arnoldi method for computing the deflation preconditioner). We first perform restarted Arnoldi cycles to construct the preconditioner (not shown), then invoke restarted GMRES. As for exact deflation, the accurate approximation derived from seven Arnoldi cycles yields convergence much like full GMRES. Performance degrades for the less accurate approximations. Insight into the behavior can be gleaned from Figure 5.3, which shows the eigenvalues of $\mathbf{AM}^{-1}$ for each of the preconditioners we have been considering. Exact deflation sends 10 eigenvalues to $\alpha = 3/2$, while all three approximate deflation preconditioners send $r - 1 = 9$ eigenvalues to α; cf. Theorem 4.1. The inaccurate approximation from five cycles leaves an eigenvalue quite close to the origin, explaining the slower convergence seen in Figure 5.2. Like the exact preconditioner, the accurate approximation derived from seven cycles deflates all the small magnitude eigenvalues, leaving a problem that restarted GMRES can effectively tackle. That said, the spectrum of this effectively preconditioned matrix bears little resemblance to the spectrum of the the exactly preconditioned matrix, though the pseudospectra shown in Figure 5.3 are quite similar. The pseudospectral analysis here predicts similar convergence for these problems, though their spectra differ so markedly.

Figure 5.4 shows approximations to the bound from Theorem 4.5 on full GMRES convergence for the approximately preconditioned matrices based on seven and eight cycles of the restarted Arnoldi process. (We use (5.2) to get an upper bound on ε in the theorem.) Both preconditioners provide convergence that is essentially the same as for the exact preconditioner. The bound captures this accurately up to the (rather coarse) bound ε (4.16) on the error in the preconditioner. One or two additional cycles of restarted Arnoldi (not shown) further refine the approximation and lead to bounds that are accurate at smaller residual norms.

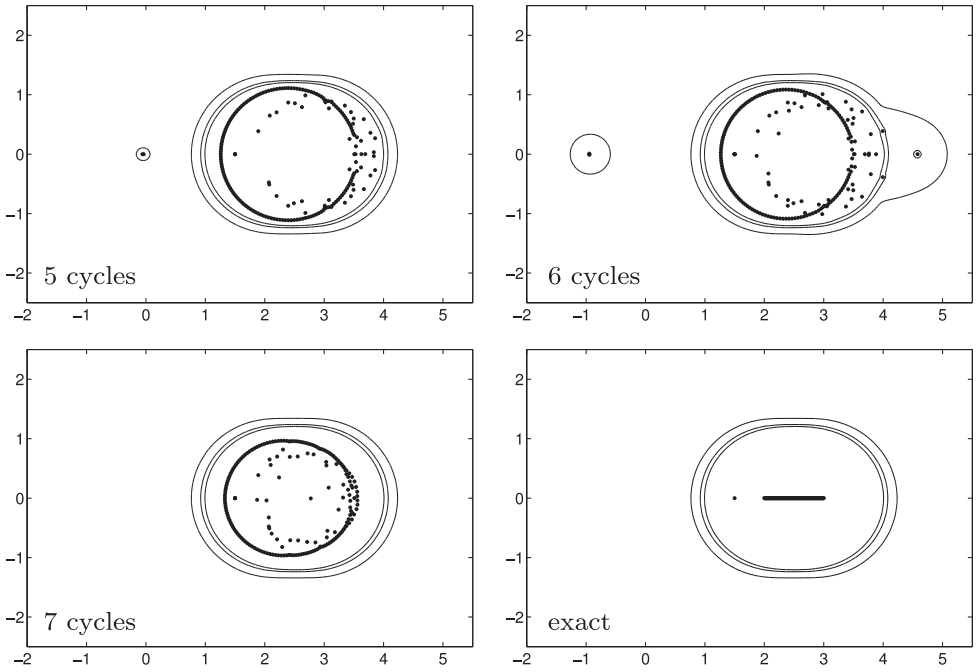

FIG. 5.3. *Eigenvalues and δ-pseudospectra ($\delta = 10^{-1}, 10^{-2}, 10^{-3}$) for $\mathbf{AM}^{-1}$ for the approximate deflation preconditioners for the matrix* (5.1) *after five, six, and seven cycles of the restarted Arnoldi method, and of $\mathbf{AM}_D^{-1}$ for exact deflation. After five cycles, one eigenvalue remains close to the origin; after six cycles, approximate deflation puts one eigenvalue near -1; after seven cycles, all small-magnitude eigenvalues are deflated, though the spectrum differs markedly from exact deflation because of the nonnormality of $\mathbf{A}$; the pseudospectra shown here are quite similar.*

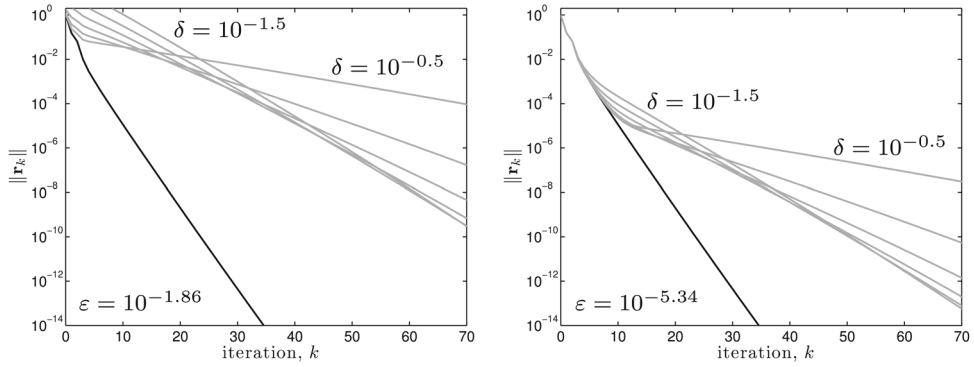

FIG. 5.4. *Bounds (gray) from Theorem* 4.5 *for full GMRES convergence (black) for the approximately preconditioned coefficient matrix* (5.1) *with $\mathbf{M}^{-1}$ constructed from seven (left) and eight (right) Arnoldi cycles. On the left, $\varepsilon = 10^{-1.86}$; on the right, $\varepsilon = 10^{-5.34}$. In both cases, we show bounds for $\varepsilon < \delta = 10^{-1.5}, 10^{-1.25}, \ldots, 10^{-0.5}$, which capture convergence nearly up to ε. Further cycles yield increasingly accurate preconditioners and bounds.*

6. Conclusions. We have introduced several approaches to bound the lag in GMRES convergence caused by a perturbation to the coefficient matrix. This style of analysis is well-suited for understanding the performance of practical preconditioners that can be related to idealized preconditioners that have favorable spectral

properties. We illustrated this approach by analyzing approximate deflation preconditioning, then showed the behavior of such a preconditioner applied to a linear system. Though applicable in many situations, our analysis does immediately pertain to preconditioners that vary at each iteration; the extension of our approach to this common situation is an important opportunity for further investigation. Structured perturbations provide another avenue for exploration: the resolvent integrals at the root of our analysis in section 2.2 presume no special structure in $\mathbf{E}$. For special classes of perturbation (e.g., $\mathbf{E} = \gamma\mathbf{I}$ for some constant γ, or $\mathbf{E} \in \mathbb{R}^{n\times n}$), a refined analysis could yield more precise bounds.

Acknowledgments. We thank Howard Elman, Sabaresan Mothi, Valeria Simoncini, Daniel Szyld, Gilbert Ymbert, and several anonymous referees for helpful comments and references, which have improved this work.

REFERENCES

[1] J. Baglama, D. Calvetti, G. H. Golub, and L. Reichel, *Adaptively preconditioned GMRES algorithms*, SIAM J. Sci. Comput., 20 (1998), pp. 243–269.
[2] C. Beattie, M. Embree, and J. Rossi, *Convergence of restarted Krylov subspaces to invariant subspaces*, SIAM J. Matrix Anal. Appl., 25 (2004), pp. 1074–1109.
[3] C. A. Beattie, M. Embree, and D. C. Sorensen, *Convergence of polynomial restart Krylov methods for eigenvalue computations*, SIAM Rev., 47 (2005), pp. 492–515.
[4] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1997.
[5] P. N. Brown and H. F. Walker, *GMRES on (nearly) singular systems*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 37–51.
[6] K. Burrage and J. Erhel, *On the performance of various adaptive preconditioned GMRES strategies*, Numer. Linear Algebra Appl., 5 (1998), pp. 101–121.
[7] S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, and C. D. Meyer, *GMRES and the minimal polynomial*, BIT, 36 (1996), pp. 664–675.
[8] A. Chapman and Y. Saad, *Deflated and augmented Krylov subspace techniques*, Numer. Linear Algebra Appl., 4 (1996), pp. 43–66.
[9] J. W. Daniel, W. B. Gragg, L. Kaufman, and G. W. Stewart, *Reorthogonalization and stable algorithms for updating the Gram–Schmidt QR factorization*, Math. Comp., 30 (1976), pp. 772–795.
[10] E. B. Davies, *Approximate diagonalization*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1051–1064.
[11] E. de Sturler, *Truncation strategies for optimal Krylov subspace methods*, SIAM J. Numer. Anal., 36 (1999), pp. 864–889.
[12] T. A. Driscoll, K.-C. Toh, and L. N. Trefethen, *From potential theory to matrix iterations in six steps*, SIAM Rev., 40 (1998), pp. 547–578.
[13] M. Eiermann, O. G. Ernst, and O. Schneider, *Analysis of acceleration strategies for restarted minimal residual methods*, J. Comput. Appl. Math., 123 (2000), pp. 261–292.
[14] H. C. Elman and A. Ramage, *An analysis of smoothing effects of upwinding strategies for the convection-diffusion equation*, SIAM J. Numer. Anal., 40 (2002), pp. 254–281.
[15] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite Elements and Fast Iterative Solvers with Applications in Incompressible Fluid Dynamics*, Oxford University Press, Oxford, UK, 2005.
[16] J. Erhel, K. Burrage, and B. Pohl, *Restarted GMRES preconditioned by deflation*, J. Comput. Appl. Math., 69 (1996), pp. 303–318.
[17] Y. A. Erlangga and R. Nabben, *Deflation and balancing preconditioners for Krylov subspace methods applied to nonsymmetric matrices*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 684–699.
[18] O. G. Ernst, *Residual-minimizing Krylov subspace methods for stabilized discretizations of convection-diffusion equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1079–1101.
[19] V. Faber, W. Joubert, E. Knill, and T. A. Manteuffel, *Minimal residual method stronger than polynomial preconditioning*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 707–729.
[20] B. Fischer, A. Ramage, D. J. Silvester, and A. J. Wathen, *On parameter choice and iterative convergence for stabilised discretizations of advection–diffusion problems*, Comput. Methods Appl. Mech. Engrg., 179 (1999), pp. 179–195.

[21] J. Frank and C. Vuik, *On the construction of deflation-based preconditioners*, SIAM J. Sci. Comput., 23 (2001), pp. 442–462.
[22] R. W. Freund, *Quasi-kernel polynomials and their use in non-Hermitian matrix iterations*, J. Comput. Appl. Math., 43 (1992), pp. 135–158.
[23] A. Gaul, M. H. Gutknecht, J. Liesen, and R. Nabben, *A Framework for Deflated and Augmented Krylov Subspace Methods*, http://arxiv.org/pdf/1206.1506v2.pdf (2012).
[24] A. Greenbaum, V. Pták, and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469.
[25] A. Greenbaum and Z. Strakoš, *Matrices that generate the same Krylov residual spaces*, in Recent Advances in Iterative Methods, G. Golub, A. Greenbaum, and M. Luskin, eds., Springer-Verlag, New York, 1994, pp. 95–118.
[26] M. H. Gutknecht, *Spectral deflation in Krylov solvers: A theory of coordinate space based methods*, Electron. Trans. Numer. Anal., 39 (2012), pp. 156–185.
[27] N. J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.
[28] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.
[29] I. C. F. Ipsen, *A note on preconditioning nonsymmetric matrices*, SIAM J. Sci. Comput., 23 (2001), pp. 1050–1051.
[30] Z. Jia, *The convergence of generalized Lanczos methods for large unsymmetric eigenproblems*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 843–862.
[31] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed., Springer-Verlag, Berlin, 1980.
[32] S. A. Kharchenko and A. Y. Yeremin, *Eigenvalue translation based preconditioners for the GMRES(k) method*, Numer. Linear Algebra Appl., 2 (1995), pp. 51–77.
[33] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, 1998.
[34] J. Liesen and Z. Strakoš, *GMRES convergence analysis for a convection-diffusion model problem*, SIAM J. Sci. Comput., 26 (2005), pp. 1989–2009.
[35] K. Meerbergen, *The solution of parameterized symmetric linear systems*, SIAM J. Matrix Anal. Appl., 24 (2003), pp. 1038–1059.
[36] R. B. Morgan, *Computing interior eigenvalues of large matrices*, Linear Algebra Appl., 154–156 (1992), pp. 289–309.
[37] R. B. Morgan, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 65 (1995), pp. 1154–1171.
[38] R. B. Morgan, *Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1112–1135.
[39] R. B. Morgan, *GMRES with deflated restarting*, SIAM J. Sci. Comput., 24 (2002), pp. 20–37.
[40] K. Moriya and T. Nodera, *The DEFLATED-GMRES(m,k) method with switching the restart frequency*, Numer. Linear Algebra Appl., 7 (2000), pp. 569–584.
[41] M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (2000), pp. 1969–1972.
[42] C. C. Paige, B. N. Parlett, and H. A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra Appl., 2 (1995), pp. 115–133.
[43] M. L. Parks, E. de Sturler, G. Mackey, D. D. Johnson, and S. Maiti, *Recycling Krylov subspaces for sequences of linear systems*, SIAM J. Sci. Comput., 28 (2006), pp. 1651–1674.
[44] R. F. Rinehart, *The derivative of a matric function*, Proc. Amer. Math. Soc., 7 (1956), pp. 2–5.
[45] Y. Saad, *Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices*, Linear Algebra Appl., 34 (1980), pp. 269–295.
[46] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
[47] 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.
[48] J. Sifuentes, *Preconditioned Iterative Methods for Inhomogeneous Acoustic Scattering Applications*, Ph.D. thesis, Rice University, Houston, TX, 2010.
[49] V. Simoncini and E. Gallopoulos, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra Appl., 247 (1996), pp. 97–119.
[50] V. Simoncini and D. B. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.
[51] V. Simoncini and D. B. Szyld, *On the occurrence of superlinear convergence of exact and inexact Krylov subspace methods*, SIAM Rev., 47 (2005), pp. 247–272.
[52] D. C. Sorensen, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.

[53] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Rev., 15 (1973), pp. 727–764.
[54] G. W. Stewart and J.-g. Sun, *Matrix Perturbation Theory*, Academic Press, San Diego, 1990.
[55] K.-C. Toh and L. N. Trefethen, *The Chebyshev polynomials of a matrix*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 400–419.
[56] L. N. Trefethen, *Approximation theory and numerical linear algebra*, in Algorithms for Approximation II, J. C. Mason and M. G. Cox, eds., Chapman and Hall, London, 1990.
[57] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.
[58] T. G. Wright, *EigTool*, http://www.cs.ox.ac.uk/pseudospectra/eigtool (2002).
[59] G. Ymbert III, M. Embree, and J. A. Sifuentes, *Approximate Murphy–Golub–Wathen preconditioning for saddle point problems*, in preparation.