THE STABILITY OF GMRES CONVERGENCE,
WITH APPLICATION TO APPROXIMATE DEFLATION
PRECONDITIONING

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

16 September 2011

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 in-
vARIANT 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

1. Introduction. Large scale systems of linear algebraic equations with non-
symmetric coefficient matrices are often solved using some variant of the Generalized
Minimum Residual (GMRES) algorithm [32]. 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
M⁻¹ (e.g., involving exact inverses of certain constituents) give matrices AM⁻¹ with
appealing spectral properties, yet practical considerations lead to approximate imple-
mentations that spoil the precise spectral structure. (For example, the saddle-point
preconditioner in [28] yields a matrix AM⁻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., [22].

How do such deviations affect GMRES convergence? We address this question by
bounding the maximum amount by which GMRES applied to A + E can lag behind
GMRES applied to A. Our main results 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.

We believe this approach to be widely applicable. To demonstrate its poten-
tial, we analyze approximate deflation preconditioning, where an ideal preconditioner
M⁻¹ is constructed from some exact eigenvectors of A. In practice one readily draws
approximations to these eigenvectors from the Arnoldi process that underlies the GM-
RES algorithm. Combining our GMRES analysis with Stewart’s invariant subspace
perturbation theory [39, 41], we can bound convergence behavior for this practical

†Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York,
New York 10012–1185 (sifuentes@courant.nyu.edu).
‡Department of Computational and Applied Mathematics, Rice University, 6100 Main Street—
MS 134, Houston, Texas 77005–1892 (embree@rice.edu).
§Department of Mathematics, Baylor University, One Bear Place #97328, Waco, Texas 76798–
7328 (Ronald_Morgan@baylor.edu).
¶Supported in part by National Science Foundation grant DMS-CAREER-0449973.
∥Supported in part by National Science Foundation grant DMS06-02235 and Air Force NSSEFF
Program Award FA9550-10-1-018.

1
preconditioner. These results are consistent with numerical computations for a discretized integral equation that arises in acoustic scattering, a problem for which deflation preconditioning is particularly well-suited. We also illustrate our basic bound for a perturbed convection–diffusion equation.

Simoncini and Szyld offer a complementary analysis that models inexact matrix-vector products at each GMRES step as a single aggregate perturbation to the coefficient matrix \[36\]. 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 \(A\) to singularity, i.e., \(\|E\| < 1/\|A^{-1}\|\). As GMRES stagnates for singular matrices (unless \(b\) is special), this is the best one can accomplish without imposing special structure on the perturbation \(E\). The 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 \([37, §6]\), giving bounds that involve spectral projectors of the perturbed matrix. The resolvent approach used 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; \(\sigma(A)\) denotes the set of eigenvalues of the matrix \(A\); \(e_k\) denotes the \(k\)th column of the identity matrix; \(\text{Ran}(\cdot)\) denotes the range (column space).

2. GMRES on perturbed coefficient matrices. Consider a linear system \(Ax = b\) with nonsingular \(A \in \mathbb{C}^{n \times n}\), \(b \in \mathbb{C}^n\), and unknown \(x \in \mathbb{C}^n\). Given an initial solution estimate \(x_0 = 0 \in \mathbb{C}^n\), step \(k\) of the GMRES algorithm \([33]\) finds the approximate solution \(x_k\) from the Krylov subspace

\[
K_k(A, b) = \text{span}\{b, Ab, \ldots, A^{k-1}b\}
\]

that minimizes the 2-norm of the residual \(r_k := b - Ax_k\):

\[
\|r_k\| = \min_{\bar{x} \in K_k(A, b)} \|b - A\bar{x}\| = \min_{q \in P_{k-1}} \|b - Aq(A)b\| = \min_{p \in P_k, p(0)=1} \|p(A)b\|,
\]

where \(P_d\) denotes the set of all polynomial of degree \(d\) or less.

How does a small modification to \(A\) affect GMRES convergence? Let \(p_k \in P_k\) denote the optimal polynomial in (2.2), and suppose \(E \in \mathbb{C}^{n \times n}\) is a perturbation that we presume to be small relative to \(A\), in a flexible way that the analysis will make precise. We shall denote perturbed quantities by Greek characters, so that GMRES applied to \((A + E)\xi = b\) produces an iterate \(\xi_k\) with residual \(\rho_k := b - (A + E)\xi_k\) satisfying

\[
\|\rho_k\| = \min_{\phi \in P_k, \phi(0)=1} \|\phi(A + E)b\|.
\]
Let $\phi_k \in \mathcal{P}_k$ denote the polynomial that optimizes this last expression, so that

$$\|\rho_k\| = \|\phi_k(A + E)b\| \leq \|p_k(A + E)b\|.$$  

Our ultimate concern is to understand how much the perturbation $E$ 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 $A$ and $A + E$ in this analysis.) Toward this end, we can replace the optimal polynomial $\phi_k$ for the perturbed problem with $p_k$:

$$\|\rho_k\| - \|r_k\| = \|\phi_k(A + E)b\| - \|p_k(A)b\|
\leq \|p_k(A + E)b\| - \|p_k(A)b\|
\leq \|(p_k(A + E) - p_k(A))b\|
\leq \|p_k(A + E) - p_k(A)\|\|b\|. $$

(2.3)

How much does $p_k(A + E)$ differ from $p_k(A)$? Such analysis amounts to an exercise in perturbation theory for functions of matrices. Expanding the polynomial term-by-term can yield crude results: if $p_k(z) = 1 + c_1z + \cdots + c_kz^k$, then

$$p_k(A + E) = p_k(A) + c_1E + c_2(AE + EA + E^2) + \cdots,$$  

suggesting the coarse bound

$$\|p_k(A + E) - p_k(A)\| \leq \sum_{j=1}^k \sum_{\ell=1}^j |c_j| \|A\|^j \|E\|^\ell.$$  

(2.4)

Involving large powers of $\|A\|$ 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 $A$ and $A + E$ are Hermitian, there exist unitary matrices $U$ and $\hat{U}$ such that

$$A = U\Lambda U^*, \quad A + E = \hat{U}\hat{\Lambda}\hat{U}^*$$

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

$$\|\rho_k\| - \|r_k\| = \|\hat{U}p_k(\hat{\Lambda})\hat{U}^*b\| - \|Up_k(A)U^*b\|
\leq \|p_k(\hat{\Lambda})\hat{U}^*b\| - \|p_k(\Lambda)U^*b\|
\leq \|(p_k(\hat{\Lambda}) - p_k(\Lambda)U^*\hat{U})U^*b\|
\leq \|(p_k(\hat{\Lambda}) - p_k(\Lambda)U^*\hat{U})\|\|b\|.$$  

(2.5)

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

$$\frac{\|\rho_k\| - \|r_k\|}{\|b\|} \leq \max_{1 \leq j \leq n} |p_k(\tilde{\lambda}_j) - p_k(\lambda_j)| + \|p_k(\Lambda)\|\|I - U^*\hat{U}\|
= \max_{1 \leq j \leq n} |p'_k(\lambda_j)|\|E\| + \|p_k(\Lambda)\||I - U^*\hat{U}| + O(\|E\|^2)$$

(2.6)
as \( \|E\| \to 0 \), using first-order perturbation theory [19, Ch. 2]. The term \( \|I - U^* \hat{U}\| \) measures the departure of the respective eigenvectors from biorthogonality: when \( A \) has several eigenvalues close together, this term can be considerably larger than \( \|E\| \).

For nonnormal \( A \), the style of analysis just described is complicated by the potential sensitivity of both the eigenvalues and eigenvectors; standard eigenvalue perturbation theory leads to asymptotic results that hold only in the limit of small \( \|E\| \). Rather than pursuing this course in detail, we prefer analysis based on resolvent integrals that leads to more descriptive results valid for finite \( \|E\| \).

2.2. Resolvent integrals and pseudospectra. Begin by writing the matrix function \( p_k(A) \) as a Cauchy integral (see, e.g., [16, Ch. 1], [17, Ch. 5]):

\[
p_k(A) = \frac{1}{2\pi i} \int_{\Gamma} p_k(z)(zI - A)^{-1} \, dz
\]

(2.7)

\[
p_k(A + E) = \frac{1}{2\pi i} \int_{\Gamma} p_k(z)(zI - A - E)^{-1} \, dz,
\]

(2.8)

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

\[
(zI - A - E)^{-1} = (I - (zI - A)^{-1}E)(zI - A)^{-1},
\]

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

\[
(zI - A - E)^{-1} = \sum_{j=0}^{\infty} ((zI - A)^{-1}E)^j(zI - A)^{-1},
\]

(2.9)

from which it follows that

\[
(zI - A - E)^{-1} - (zI - A)^{-1} = \sum_{j=1}^{\infty} ((zI - A)^{-1}E)^j(zI - A)^{-1}.
\]

The submultiplicativity of the matrix 2-norm yields the bound

\[
\|(zI - A - E)^{-1} - (zI - A)^{-1}\| \leq \frac{\varepsilon/\delta}{\delta - \varepsilon},
\]

(2.10)

which is small when \( \varepsilon \ll \delta \). This analysis, for generic functions of \( A + E \), was applied by Rinehart in 1956 [30] (see Higham [16, Prob. 3.4]), and more recently by Davies [8].

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

\[
\sigma_\delta(A) = \{ z \in \mathbb{C} : \|(zI - A)^{-1}\| > 1/\delta \},
\]

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

\[
\sigma_\delta(A) = \{ z \in \mathbb{C} : z \in \sigma(A + D) \text{ for some } D \text{ with } \|D\| < \delta \};
\]
see, e.g., [44] for further details. From this latter definition it is evident that \( \sigma(A) \subset \sigma_\delta(A) \) for all \( \delta > 0 \). Denote the boundary of the \( \delta \)-pseudospectrum by \( \partial \sigma_\delta(A) \).

Taking the contour \( \Gamma \) in (2.7)–(2.8) to be this boundary ensures that \( \Gamma \) will encircle all eigenvalues of \( A \) and, provided \( \delta > \epsilon = ||E|| \), all eigenvalues of \( A + E \) as well. Conveniently, the requirement \( \delta > \epsilon \) is precisely the condition \( 1/\| (zI - A)^{-1} \| > \| E \| \) that ensured convergence of the Neumann expansion of the perturbed resolvent (2.9).

Returning to the bound (2.3), we see

\[
\frac{\| \rho_k - r_k \|}{\| b \|} \leq \| p_k(A + E) - p_k(A) \|
\]

\[
= \frac{1}{2\pi} \left\| \int_\Gamma p_k(z)((zI - A - E)^{-1} - (zI - A)^{-1}) \, dz \right\|
\]

\[
\leq \frac{L_\delta}{2\pi} \left( \max_{z \in \Gamma} |p_k(z)| \right) \left( \max_{z \in \Gamma} \| (zI - A - E)^{-1} - (zI - A)^{-1} \| \right)
\]

\[
\leq \left( \frac{\epsilon}{\delta - \epsilon} \right) \left( \frac{L_\delta}{2\pi} \right) \max_{z \in \sigma_\delta(A)} |p_k(z)|,
\]

where \( L_\delta \) denotes the arc length of \( \Gamma = \partial \sigma_\delta(A) \). Applying the maximum modulus principle yields the following bound.

**Theorem 2.1.** Let \( r_k = p_k(A) b \) denote the \( k \)th residual vector produced by GMRES applied to \( Ax = b \), with the residual polynomial \( p_k \in P_k \) satisfying \( p_k(0) = 1 \). The residual \( p_k \) produced by GMRES applied to \( (A + E)x = b \) satisfies

\[
(2.11) \quad \frac{\| \rho_k \|}{\| b \|} \leq \frac{\| r_k \|}{\| b \|} + \left( \frac{\epsilon}{\delta - \epsilon} \right) \left( \frac{L_\delta}{2\pi} \right) \sup_{z \in \sigma_\delta(A)} |p_k(z)|,
\]

where \( ||E|| =: \epsilon < \delta \) and \( L_\delta \) denotes the arc length of the boundary of \( \sigma_\delta(A) \).

This bound substantiates the intuition that “a size \( \epsilon \) perturbation to \( A \) should only affect GMRES convergence by \( O(\epsilon) \)”. The relative increase in the GMRES residual norm induced by the perturbation \( E \) is bounded in terms of the GMRES polynomial \( p_k \) for the unperturbed problem on the \( \delta \)-pseudospectrum of \( A \). How should one choose \( \delta \)? If \( \delta \) exceeds the distance of \( A \) to singularity, \( \delta > 1/\| A^{-1} \| \), then \( 0 \in \sigma_\delta(A) \), and the polynomial optimization problem in Theorem 2.1 cannot be smaller than one, since \( p_k(0) = 1 \) for all \( k \): Theorem 2.1 cannot predict convergence as \( k \) increases. Hence, for a fixed perturbation size \( \epsilon = ||E|| \), we restrict attention to \( \delta \in (\epsilon, 1/\| A^{-1} \|) \); Figure 2.1 illustrates this setting in the complex plane. As \( \delta \) decreases from \( 1/\| A^{-1} \| \) down to \( \epsilon \), the leading constant 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 illustrated in Section 3, larger values of \( \delta \) usually give bounds that are sharper at early iterations, while smaller values of \( \delta \) are better at later iterations.

Some additional hint about the choice of \( \delta \) comes from the GMRES algorithm itself. Recall that \( k \) steps of the Arnoldi process produce a factorization

\[
AV_k = V_{k+1} \hat{H}_k
\]

\[
= V_k H_k + h_{k+1,k} v_{k+1} e_k^*,
\]

where the first \( j \) columns of \( V_{k+1} = [V_k \ v_{k+1}] \) form an orthonormal basis for the Krylov subspace \( K_j(A,b) \) \((j = 1, \ldots, k + 1)\), the matrix \( \hat{H}_k \in \mathbb{C}^{(k+1)\times k} \) is upper
Hessenberg, and $\tilde{H}_k$ is the top $k \times k$ portion of $\tilde{H}_k$. The $(k+1) \times k$ entry of $\tilde{H}_k$, $h_{k+1,k}$, is a nonnegative real number. See, e.g., [32, Ch. 6] for further details. The eigenvalues $\theta_1, \ldots, \theta_k$ of $H_k$, the Ritz values, satisfy

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

The roots $\nu_1, \ldots, \nu_k$ of the residual polynomial $p_k$, called harmonic Ritz values [15, 23, 29], similarly satisfy

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

where $s_{\min}(H_k)$ is the smallest singular value of $H_k$; see, e.g., [20, §4.6], [35], [44, §26]. Thus, one might take $\delta = h_{k+1,k}$ or $\delta = h_{k+1,k} + s_{\min}(H_k) h_{k+1,k}^2$, though this only gives a rough starting point for larger values of $\delta$. Our computational results illustrate that a range of $\delta$ values often give more insight than can be derived from a single $\delta$.

The utility of Theorem 2.1 is restricted by the presence of the GMRES residual polynomial $p_k$ in the bound (2.11). For example, $p_k$ may be large at some $z \in \sigma_{\delta}(A)$ when $\delta \gg 0$, or when $b$ has small components in certain eigenvectors of $A$. In some cases one might indeed know $p_k$ explicitly (e.g., when $A$ is an ideally preconditioned matrix that gives exact convergence at step $k$, so $r_k = 0$ and $p_k$ equals the minimal polynomial for $A$, as in [28] with $k = 3$).

To avoid use of $p_k$, one can instead use the series (2.9) to bound

$$\|\rho_k\| = \min_{\phi \in P_k} \|\phi(A + E)b\| = \min_{\phi \in P_k} \| \frac{1}{2\pi i} \int_{\Gamma} \phi(z)(zI - A - E)^{-1} dz b \|$$

directly. Picking $\Gamma = \sigma_{\delta}(A)$ and bounding the norm of the integral as above, we obtain a more flexible bound.

**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 $\sigma_{\varepsilon}(A)$: eigenvalues of $A + E$ can fall anywhere in this set. The outermost curve shows the boundary of $\sigma_{\delta_{\max}}(A)$ for $\delta_{\max} = 1/\|A^{-1}\|$, which passes through the origin. Theorems 2.1 and 2.2 pertain to any $\delta \in (\varepsilon, \delta_{\max})$, such as the one whose boundary is shown by the dashed line.
Theorem 2.2. The residual $\rho_k$ produced by GMRES applied to $(A + E)x = b$ satisfies

$$\|\rho_k\| \leq \left(1 + \frac{\varepsilon}{\delta - \varepsilon}\right) \left(\frac{L_\delta}{2\pi\delta}\right) \min_{\phi \in p_k, \phi(0) = 1} \sup_{z \in \sigma_\delta(A)} |\phi(z)|,$$

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

Contrast the estimate (2.15) with Trefethen’s bound on the GMRES residual norm for the unperturbed problem [43], [44, §26]:

$$\|r_k\| \leq \left(\frac{L_\delta}{2\pi\delta}\right) \min_{p \in P_k} \sup_{z \in \sigma(A)} |p(z)|.$$

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 $\eta_k/\|b\|$, we have

$$\|\rho_k\|/\|b\| \leq \eta_k/\|b\| + \left(\frac{\varepsilon}{\delta - \varepsilon}\right) \left(\frac{L_\delta}{2\pi\delta}\right) \min_{p \in P_k} \sup_{z \in \sigma(A)} |p(z)|,$$

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 $\|r_k\|$ on the right hand side with the upper bound $\eta_k \geq \|r_k\|$. (This amounts to saying that Trefethen’s bound (2.16) is robust to perturbations. This bound can provide a good description of GMRES convergence [44, §26], though an example of Greenbaum and Strakoš 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 [13, 14, 21], the SUPG discretization of the two-dimensional convection–diffusion problem

$$-\nu \Delta u(x, y) + a \cdot \nabla u(x, y) = f(x, y)$$

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. See Fischer et al. [14] for a full description of the problem and its discretization on an $N \times N$ grid of square elements that gives rise to a matrix $A$ of dimension $n = N^2$. The right hand side vector $b$ is also derived from the differential equation. In our examples we align the convection with the grid, $a = (1, 0)^T$, so that simple formulas for the eigenvalues follow from Fourier analysis [14]: this is especially helpful given the strong departure from normality exhibited for small values of the diffusion parameter, $\nu$. (We use this model problem simply to illustrate the efficacy of our bounds, even on a problem that has posed considerable challenge to GMRES analysts. For the large-scale convection–diffusion problems encountered in practice, one should use a more sophisticated approach than unpreconditioned full GMRES; see [10, Ch. 4] for alternatives.)

We begin with $\nu = .01$ and $N = 32$ ($n = 1024$). Figure 3.1 shows the bound in Theorem 2.1 for perturbations $E$ of size $10^{-4}\|A\|$, $10^{-6}\|A\|$, and $10^{-8}\|A\|$, along with the convergence curve for the unperturbed problem and one random choice of $E$ (with entries drawn from the normal distribution). Especially for the two smaller
Fig. 3.1. Example of Theorem 2.2 for the SUPG problem with $N = 32$ and $\nu = 0.01$, with a perturbation of size $10^{-4}\|A\|$ (top), $10^{-6}\|A\|$ (middle), and $10^{-8}\|A\|$ (bottom). The plots on the left show the bound on $\|\rho_k\| - \|r_k\|$; the plots on the right compare $\|r_k\|$ (black solid line), $\|\rho_k\|$ (dashed line), and the upper bound (2.11) on $\|\rho_k\|$ for various $\delta$ (gray lines). On the top, $\delta = 10^{-2.9}, 10^{-3}, 10^{-3.5}, \ldots, 10^{-9}, 10^{-9.2}$; in the middle, $\delta = 10^{-2.9}, 10^{-3}, 10^{-3.5}, \ldots, 10^{-7}, 10^{-7.2}$; on the bottom, $\delta = 10^{-2.9}, 10^{-3}, 10^{-3.5}, \ldots, 10^{-5}, 10^{-5.2}$.

Figure 3.2 reveals a subtlety that can emerge when applying Theorem 2.1. In this case, we take $\nu = 0.001$ and $N = 48$ ($n = 2304$) with a perturbation of size $\varepsilon = 10^{-4}\|A\|$, and focus on iteration $k = 70$. The polynomial $p_{70}$ (computed via harmonic Ritz values) is larger than one in magnitude at points $z \in \sigma_\varepsilon(A)$, 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 [42], this agreement need not always be close enough to yield helpful bounds in Theorem 2.1. In such circumstances, the flexibility afforded by Theorem 2.2 can be helpful. Figure 3.3 shows the bound from Theorem 2.2 for this example. Small values of $\delta$ appear to accurately capture the asymptotic convergence.
Fig. 3.2. SUPG example for \( N = 48 \) and \( \nu = 0.001 \) with a perturbation of size \( \varepsilon = 10^{-4} \|A\| \). The solid dots denote the eigenvalues of \( A \); 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 \( \sigma_\varepsilon(A) \). Since \( |p_{70}(z)| > 1 \) for some \( z \in \sigma_\varepsilon(A) \), the polynomial term in (2.15) does not contribute to convergence at iteration \( k = 70 \).

rate for the perturbed problem. (For careful computations of the pseudospectral bound (2.16) as in the bottom-left figure here, see [44, p. 252].)

To calculate the bounds in Figures 3.1 and 3.3, we replace \( \sigma_\delta(A) \) (computed using EigTool [45]) with a slightly larger region \( \Omega \) having a circular boundary, \( \partial \Omega \). Since \( \|(zI - A)^{-1}\| \leq 1/\delta \) for all \( z \in \partial \Omega \), the analysis in Section 2 follows with \( \Omega \) replacing \( \sigma_\delta(A) \) and the length of \( \partial \Omega \) replacing \( L_\delta \). We obtain the polynomial \( p_k \) in Theorem 2.1 from its roots, the computed harmonic Ritz values for the unperturbed problem. For Theorem 2.2, the optimal polynomial for the disk \( \Omega \) centered at \( c \in \mathbb{C} \) is given by \( p_k(z) = (1 - z/c)^k \).

4. Approximate deflation preconditioning. To illustrate the utility of the preceding analysis, we study the convergence of GMRES for a system with approximate deflation preconditioning. Deflation preconditioners are in a class of algorithms that apply knowledge (or estimates) of eigenvectors of \( A \) to remove certain components from the initial residual \( r_0 = b \), 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) [6, 9, 24, 25, 26]. Deflation preconditioners differ, in that they use approximate eigenvector information to build a matrix \( M^{-1} \) that sends certain eigenvalues of \( AM^{-1} \) (approximately) to one, while leaving the other eigenvalues of \( A \) fixed. Such methods have been developed in [1, 5, 11, 12, 27].

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 [11, Thm. 3.1]. Let \( A \in \mathbb{C}^{n \times n} \) be a nonsingular matrix
with (possibly repeated) eigenvalues $\lambda_1, \ldots, \lambda_n$, and suppose the columns of $X \in \mathbb{C}^{n \times r}$ form an orthonormal basis for an $r$-dimensional invariant subspace of $A$ associated with eigenvalues $\lambda_1, \ldots, \lambda_r$. Then the matrix

$$M_D := I - XX^* + X(X^*AX)X^*$$
is invertible, with

\begin{equation}
M_D^{-1} = I - XX^* + X(X^*AX)^{-1}X^*,
\end{equation}

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

\begin{equation}
Z^*AM_D^{-1}Z = \begin{bmatrix}
I & X^*AY \\
0 & Y^*AY
\end{bmatrix}.
\end{equation}

In particular, the eigenvalue 1 has (algebraic and geometric) multiplicity of at least \(r\). (One can readily modify \(M_D\) with any real \(\alpha \neq 0\) such that if \(M_D = I - XX^* + \alpha^{-1}X(X^*AX)^{-1}X^*\), then \(M_D^{-1} = I - XX^* + \alpha X(X^*AX)^{-1}X^*\) and \(\sigma(AM_D^{-1}) = \{\alpha, \lambda_{r+1}, \lambda_{r+2}, \ldots, \lambda_n\}\); 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 \(M^{-1}\) using approximations to an invariant subspace. In [5, 11, 27] 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 [34]).

In this paper we consider preconditioners built with Arnoldi vectors as a basis for an approximate invariant subspace. This could be done by approximating \(X\) with the matrix \(V_m\) from (2.12) upon restart. Keeping the entire subspace requires the same storage as GMRES, but accelerates computational time dramatically by limiting the computational complexity of the orthogonalization step in the Arnoldi iteration. (Of course, more practical implementations limit the dimension of the subspaces kept, as discussed in section 5.) From (2.12) one can see that \(H_m = V_m^*AV_m\), and so \(H_m\) is an orthogonal compression of \(A\) onto an \(m\)-dimensional subspace. The eigenvalues of \(H_m\), the Ritz values, estimate the eigenvalues of \(A\), while subspaces of \(V_m\) approximate the corresponding invariant subspaces of \(A\) [2, 3, 18, 31].

In general 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 \(\text{Ran}(V_m) = \mathcal{K}_m(A, b)\). To do so, the Arnoldi factorization (2.12) can be reordered using implicit QR steps, as is done in Sorensen’s implicitly restarted Arnoldi algorithm [38], in such a way that the eigenvalues of the \(r \times r\) upper-left portion of \(H_m\) are any desired subset of \(r\) eigenvalues of \(H_m\). This is precisely the restarted preconditioning strategy developed by Baglama, et al. [1, Alg. 3.6]. (Alternatively, in the Krylov–Schur setting, this step amounts to swapping diagonal blocks in a Schur factorization [40].) Suppose the Arnoldi factorization (2.12) has been reordered so that the first \(r \leq m\) columns \(V_r\) of the Arnoldi basis matrix \(V_m\) define the approximate invariant subspace we wish to deflate. Replace the exact deflation preconditioner (4.1) with

\begin{equation}
M^{-1} := I - V_rV_r^* + V_rH_r^{-1}V_r^*,
\end{equation}

where \(H_r = V_r^*AV_r \in \mathbb{C}^{r \times r}\) is upper Hessenberg, \(V_r \in \mathbb{C}^{n \times r}\) has orthonormal columns, and together these matrices satisfy an Arnoldi relation,

\begin{equation}
AV_r = V_rH_r + h_{r+1,r}v_{r+1}e_r^*.
\end{equation}

The preconditioner (4.3) requires \(H_r\) to be invertible, which is not guaranteed by the invertibility of \(A\).
How well does Ran($V_r$) approximate an invariant subspace of $A$? Among all choices of $L \in \mathbb{C}^{r \times r}$, the norm of the residual $R_r = AV_r - V_rL$ is minimized by $L = V_r^*AV_r = H_r$ [41, Thm. IV.1.15], so $R_r = AV_r - V_rH_r = h_{r+1,r}v_{r+1}^*$, giving

$$\|R_r\| = h_{r+1,r}. \quad (4.5)$$

In the unusual case that $R_r = 0$, the columns of $V_r$ give an orthonormal basis for an invariant subspace, and we have the exact setting (4.1): the preconditioner moves $r$ eigenvalues of $AM^{-1}$ to one. When $R_r \neq 0$, one only has an approximation to an invariant subspace, so perhaps it is surprising that the preconditioner still moves $r-1$ eigenvalues exactly to one.

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

$$M := I - V_rV_r^* + V_rH_rV_r^*$$

is invertible, with

$$M^{-1} = I - V_rV_r^* + V_rH_r^{-1}V_r^*,$$

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

$$\Phi = \begin{bmatrix} I_r & H_r^{-1}J \\ 0 & H_r \end{bmatrix}, \quad (4.6)$$

where $H_r = \tilde{V}_r^*A\tilde{V}_r$, and $J \equiv V_r^*A\tilde{V}_r$, for any $\delta > \min_{L \in \mathbb{C}^{r \times r}} \|AV_r - V_rL\| = h_{r+1,r}$.

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

$$H = V^*AV = \begin{bmatrix} H_r & J \\ h_{r+1,r}e_1e_1^* & H_r \end{bmatrix}. $$

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

$$V^*M^{-1}V = \begin{bmatrix} I_r & 0 \\ 0 & I_{n-r} \end{bmatrix}. $$

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

$$V^*M^{-1}AV = \begin{bmatrix} I_r & H_r^{-1}J \\ h_{r+1,r}e_1e_1^* & H_r \end{bmatrix}$$

$$= \begin{bmatrix} I_{r-1} & 0 & [H_r^{-1}J]_{1:r-1,1:n-r} \\ 0 & 1 & e_1^*H_r^{-1}J \\ 0 & h_{r+1,r} & \tilde{H}_r \end{bmatrix}. $$
Thus the spectrum of $M^{-1}A$ (and hence also the spectrum of the similar matrix $AM^{-1}$) contains the eigenvalue one with multiplicity $r - 1$. All that remains is to note that (4.7) implies

$$
V^*M^{-1}AV = \begin{bmatrix}
I_r & H_r^{-1}J \\
0 & \hat{V}_r^*A\hat{V}_r
\end{bmatrix} + h_{r+1,r}e_{r+1}e^*_r.
$$

The fact that $h_{r+1,r} = \min_{L \in \mathbb{C}^{n \times n}} \|AV_r - V_rL\|$ completes the proof. \( \Box \)

The invariant subspace perturbation theory developed by Stewart [39, 41] provides an analytical framework in which we can extend our understanding of approximate deflation techniques, and thus develop convergence bounds for such restarted GMRES methods. Stewart’s analysis shows that if the approximate invariant subspace is sufficiently close to a true invariant subspace, then the true subspace can be written as a perturbation of the approximate one. We translate the main result [41, Thm. V.2.1] into our notation here, so as to uniquely specify the 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:

$$
\text{sep}(H_r, \hat{H}_r) := \inf_{S \in \mathbb{C}^{r \times (n-r)}} \|H_rS - S\hat{H}_r\|.
$$

The spectra of $H_r$ and $\hat{H}_r$ are disjoint if and only if $\text{sep}(H_r, \hat{H}_r) > 0$. For further properties of sep, see [41, §V.2].

**Theorem 4.2.** (Stewart, [39, 41]) Let $A$, $V_r$, and $H_r$ satisfy the hypotheses of Theorem 4.1. If $\sigma(H_r) \cap \sigma(\hat{H}_r) = \emptyset$ and

$$
h_{r+1,r} \frac{\|J\|}{\text{sep}(H_r, \hat{H}_r)^2} < \frac{1}{4},
$$

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

$$
X = (V_r + \hat{V}_rP)(I + P^*P)^{-1/2},
$$

$$
Y = (\hat{V}_r + V_rP^*)(I + PP^*)^{-1/2},
$$

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

Stewart and Sun provide a telling corollary to Theorem 4.2 [41, Cor. V.2.2]. The tangent of the largest canonical angle between $K_r(A, v_1)$ and $\text{Ran}(X)$ is bounded by $\|P\|$, which itself is bounded by $2h_{r+1,r}/\text{sep}(H_r, \hat{H}_r)$. 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 $\|R_r\|$.

**Theorem 4.3.** Let $A$, $V$, $H_r$ satisfy the hypotheses of Theorem 4.2 and let $M_D^{-1}$ be the exact deflation preconditioner defined in (4.1). Assume further that the unique matrix $P$ defined in Theorem 4.2 is bounded in norm by one, i.e. $\rho := \|P\| < 1$. Then

$$
\|AM_D^{-1} - ZV^*AM^{-1}VZ^*\| \leq \|R_r\||H_r^{-1}| + \frac{8\|A\||\|R_r\|\|H_r\|}{(\text{sep}(H_r, \hat{H}_r) - 2\|R_r\|)^2}.
$$
Proof. Since $\|P\| < 1$, express $(I + P^*P)^{-1/2}$ as the convergent series

$$(I + P^*P)^{-1/2} = \sum_{n=0}^{\infty} \left( -\frac{1}{n} \right) (P^*P)^n.$$ 

This identity, along with the description of $X$ and $Y$ in (4.10) and (4.11), leads to an expression for $X$ as a perturbation of $V_r$:

$$X = (V_r + \hat{V}_r P) \sum_{n=0}^{\infty} \left( -\frac{1}{n} \right) (P^*P)^n = V_r + E_X,$$

for the perturbation $E_X := \hat{V}_r P + (V_r + \hat{V}_r P) \sum_{n=1}^{\infty} \left( -\frac{1}{n} \right) (P^*P)^n$. Continuing similarly with the left invariant subspace basis leads to the expression

$$Y = \hat{V}_r + E_Y,$$

for the perturbation $E_Y := V_r P^* + (\hat{V}_r + V_r P^*) \sum_{n=1}^{\infty} \left( -\frac{1}{n} \right) (PP^*)^n$. To compare deflation with exact and inexact invariant subspaces, first note that

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

and so

$$V^* \mathcal{AM}^{-1} V = \begin{bmatrix} V^*_r \mathcal{AM}^{-1} V_r & V^*_r \mathcal{AM} \hat{V}_r \\ \hat{V}_r^* \mathcal{AM}^{-1} \hat{V}_r & \hat{V}_r^* \mathcal{AM} \hat{V}_r \end{bmatrix} = \begin{bmatrix} 1 & V^*_r \hat{A} \hat{V}_r \\ e_1 e_1^* h_{r+1, r} H_r^{-1} & e_1 e_1^* \hat{A} \hat{V}_r \\ e_1 e_1^* h_{r+1, r} H_r^{-1} & e_1 e_1^* \hat{A} \hat{V}_r \end{bmatrix}.$$ 

Using the expressions for $X$ and $Y$ derived above,

$$X^* A Y = V^*_r \hat{A} \hat{V}_r + E_X \hat{A} \hat{V}_r + V^*_r A E_Y + E_X A E_Y$$

$$Y^* A Y = \hat{V}_r^* \hat{A} \hat{V}_r + E_Y \hat{A} \hat{V}_r + \hat{V}_r A E_Y + E_Y A E_Y,$$

and so

$$Z^* \mathcal{AM}_D^{-1} Z = V^* \mathcal{AM}^{-1} V + \begin{bmatrix} 0 & E_X \hat{A} \hat{V}_r + V^*_r A E_Y + E_X A E_Y \\ -e_1 e_1^* h_{r+1, r} H_r^{-1} & e_1 e_1^* \hat{A} \hat{V}_r + \hat{V}_r A E_Y + E_Y A E_Y \end{bmatrix}.$$ 

Norm inequalities then yield the rough bound

$$\|Z^* \mathcal{AM}_D^{-1} Z - V^* \mathcal{AM}^{-1} V\| \leq h_{r+1, r} \|H_r^{-1}\| + \|A\| (\|E_X\| + \|E_Y\| + 3\|E_Y\| + \|E_Y\|^2).$$

(4.12)

The perturbations norms $\|E_X\|$ and $\|E_Y\|$ both satisfy the same upper bound

$$\rho + (1 + \rho) \sum_{n=1}^{\infty} \left( -\frac{1}{n} \right) \rho^{2n} < \frac{\rho}{1 - \rho}$$

(4.13)

Combining this bound with (4.12) yields

$$\|A \mathcal{M}^{-1} - ZV^* \mathcal{AM}^{-1} VZ^*\| \leq h_{r+1, r} \|H_r^{-1}\| + \|A\| \frac{4\rho}{(1 - \rho)^2}$$

$$\leq h_{r+1, r} \left( \|H_r^{-1}\| + \frac{8\|A\| \operatorname{sep}(H_r, \hat{H}_r)}{(\operatorname{sep}(H_r, \hat{H}_r) - 2h_{r+1, r})^2} \right),$$

where $\operatorname{sep}(H_r, \hat{H}_r)$ denotes the minimum distance between $H_r$ and $\hat{H}_r$. This completes the proof.

---

[Note: The proof involves complex mathematical expressions and inequalities that are typical in operator theory or numerical linear algebra, focusing on the analysis of perturbed invariant subspaces and their bounds.]
where the last inequality follows from the bound $\|P\| \leq 2h_{r+1,r}/\text{sep}(H_r, \hat{H}_r)$ shown in the proof of Theorem 4.2 given in [41, p. 231].

This result gives a clean interpretation of how Krylov subspaces, and thus the span of Ritz vectors, approximate invariant subspaces—and identifies the invariant subspace that is approximated, provided the approximation is sufficiently close. This insight gives the necessary framework to express approximate deflation as a bounded perturbation of exact deflation, thus facilitating the analysis of approximate deflation using the the results of Section 2. Thus we now seek a convergence bound for GMRES applied to the system $AM^{-1}y = b$ with approximate deflation preconditioning, which is equivalent to solving $Ax = b$, where $x = M^{-1}y$. To begin, define

$$
\varepsilon := h_{r+1,r} \left( \left\| H_r^{-1} \right\| + \frac{8 \| A \|}{\text{sep}(H_r, \hat{H}_r)} \right).
$$

(4.14)

**THEOREM 4.4.** Let $\| \rho_k \|$ denote the norm of the $k$th residual produced by GMRES applied to $AM^{-1}y = b$. If the hypotheses of Theorem 4.3 are satisfied, then

$$
\frac{\| \rho_k \|}{\| b \|} \leq \frac{L_\delta}{2\pi\delta} \left( 1 + \frac{\varepsilon}{\delta - \varepsilon} \right) \min_{p \in P_k} \max_{z \in \sigma_\delta(AM_D^{-1})} |p(z)|
$$

(4.15)

where $\varepsilon$ is defined in (4.14) and $\delta > \varepsilon$.

**Proof.** The system $AM^{-1}y = b$ is equivalent to

$$
(ZV^*)AM^{-1}(VZ^*) \tilde{y} = \hat{b},
$$

(4.16)

where $\tilde{y} := ZV^*y$ and $\hat{b} = ZV^*b$, and since $ZV^*$ is unitary, the relative residual norm for GMRES applied to (4.16) is identical to that produced for the original problem. By Theorem 4.3, we can write

$$
(ZV^*)AM^{-1}(VZ^*) = AM_D^{-1} + E,
$$

where $\| E \| \leq \varepsilon$. The bound (4.15) follows from application of Theorem 2.2.

Combining the same argument with Theorem 2.1 yields another bound.

**THEOREM 4.5.** Let $\rho_k$ denote the $k$th residual produced by GMRES applied to $AM^{-1}y = b$ and $r_k = p_k(AM_D^{-1})b$ denote the $k$th residual produced by GMRES applied to the exactly deflated system $AM_D^{-1}y = b$. If the hypotheses of Theorem 4.3 are satisfied, then

$$
\frac{\| \rho_k \|}{\| b \|} \leq \frac{\| r_k \|}{\| b \|} + \left( \frac{\varepsilon}{\delta - \varepsilon} \right) \frac{L_\delta}{2\pi\delta} \sup_{z \in \sigma_\delta(AM_D^{-1})} |p_k(z)|
$$

(4.17)

where $\varepsilon$ is defined in (4.14) and $\delta > \varepsilon$.

5. Deflation preconditioning examples. We illustrate the approximate deflation scenario described in the last section on a model problem for which this preconditioner is effective, a Lippmann–Schwinger integral equation for acoustic scattering in two dimensions [7, §8.2], [34]. Matrices resulting from the discretization of the Lippmann–Schwinger operator (a compact perturbation of the identity) have many eigenvalues clustered around 1, with other eigenvalues spread out approximately on a circle tangent to the real line at 1 in the complex plane, whose diameter grows with increasing wave number, $\kappa$. Figure 5.1 shows the spectrum for three examples. In all
these cases, a subset of the spectrum falls quite close to the origin; these eigenvalues, as well as those scattered around the circle, cause GMRES to converge slowly at first. This phase becomes lengthier as the wave number increases: iteration counts increase quadratically with $\kappa$. The more complicated nature of this spectrum makes it difficult to precisely compute our bounds, as done for a cleaner example in Section 3. Instead, here we perform several qualitative experiments to compare exact and approximate deflation preconditioning, showing results consistent with the analysis in the last section.

For this model problem, deflation preconditioning using Arnoldi vectors to approximate eigenvectors significantly accelerates the convergence rate of restarted GMRES. As our analysis suggests, this approximate deflation method performs comparably to a restarted routine that utilizes exact eigenvectors. For details of the Lippmann–Schwinger problem and the Nyström quadrature discretization used here, see [34, Chap. 1]. The discretization gives the system $Ax = b$, with $A := I + K$,

where $K$ is a dense matrix that depends on the wave number, $\kappa$. Although $K$ is dense, the matrix vector product $Kv$ can be computed using a two dimensional FFT algorithm, and requires only $O(n \log n)$ floating point operations. All our GMRES experiments involve a matrix with $\kappa = 5$ of dimension $n = 900$.

To qualitatively illustrate the theoretical results of the last section, we present numerical comparisons of exact and approximate deflation. Theorems 4.4 and 4.5 suggest that the residual norm of an approximately deflated system lags behind that of an exactly deflated system by a factor proportional to $h_{r+1}^{r+1}$, where $r$ is the dimension of the Krylov subspace used for deflation. Figure 5.2 shows this lag as a function of $h_{r+1}^{r+1}$. First apply 80 steps of GMRES to the unpreconditioned system $Ax = b$; the iterate $x_{80}$ and residual $r_{80}$ will be used to start a preconditioned iteration. To construct the approximate deflation preconditioner, refine the 80 dimensional Krylov subspace to a 40 dimensional subspace via the implicitly restarted Arnoldi method. (The largest magnitude Ritz values are used as shifts, with the iteration proceeding until $h_{41,40}^{41,40} < 1$ or $h_{41,40}^{41,40} < .1$; see [38] for details of the method. The tolerance $h_{41,40}^{41,40} < 1$ was met by simply shifting out the largest Ritz values, without further Arnoldi steps; the tolerance $h_{41,40}^{41,40} < .1$ required 33 restarted Arnoldi cycles.) Now apply GMRES to the preconditioned problem $A M^{-1} y = b$ with initial guess $y_0 := M x_{80}$, where $M^{-1}$ is built according to (4.3) using the 40 refined Arnoldi vectors. To compare performance to exact deflation, we also run GMRES with the exact deflation preconditioner $M^{-1} D$ in (4.1) corresponding to the 40 dimensional subspace associated with the smallest magnitude eigenvalues, with $y_0 := M D x_{80}$.
For the example in Figure 5.2, a tolerance of $h_{r+1, r} = O(10^{-1})$ is sufficient for approximate deflation to closely track the convergence of GMRES with exact deflation. While this tolerance means that the approximating space is near some invariant subspace – and the implicit restarting procedure drives the approximating subspace toward the one associated with smallest magnitude eigenvalues – there is no guarantee that these approximate subspaces are good approximations to the invariant subspace used for exact deflation. A more apt example of the scenario outlined in Theorems 4.2 through 4.5 would pit approximate deflation against exact deflation using the invariant subspace closest to the approximate one.

In Figure 5.3, GMRES is applied to $Ax = b$ for 40 steps; $x_{40}$ and $r_{40}$ are then used to start GMRES on the preconditioned problem $AM^{-1}y = b$, where $M^{-1}$ is built from the 40 dimensional Krylov subspace $\mathcal{K}_{40}(A, b)$ (i.e., unlike the last example, no implicit restarting is used). GMRES is also applied to an exactly deflated system, using the 40 dimensional invariant subspace closest to $\mathcal{K}_{40}(A, r_0)$ (in terms of angles between subspaces) with starting vector $y_0 := Mx_{40}$. Although the residual is $h_{41, 40} = 2.3288$, the approximately deflated system matches the performance of the exactly deflated system, which both closely match the natural deflation of the full GMRES process. In this case, our intuition is guided by Theorem 4.1: while exact deflation moves 40 eigenvalues to 1, approximate deflation using Arnoldi vectors moves 39 eigenvalues to 1. This intuition is affirmed by the spectral plots in Figures 5.4.

Such results prompt a numerical experiment with a subtle variation of Algorithm 3.6 of [1]. We denote this algorithm GMRES-ADAP($m, r$), for GMRES with approximate deflation-based adaptive preconditioning, where $m$ is the restart frequency and $r$ is the dimension of the approximate invariant subspace kept at each restart. (We fix the algorithm parameter $\beta_0 = 1$, the number of implicitly restarted Arnoldi cycles used to create the Ritz vectors for deflation.) Let $x_{k,j}$ denote the $j$th iterate of the $k$th cycle of GMRES-ADAP($m, r$); similarly let $v_{k,j}$ and $H_{k,j}$ denote the $j$th Arnoldi vector and $j \times j$ Hessenberg matrix at step $j$ of cycle $k$. At cycle $k$, GMRES-ADAP($m, r$) applies $m$ steps of GMRES to the preconditioned problem

(5.1) \[ AM_k^{-1}y = b \]
with preconditioner $\overline{M}_k^{-1} := M_1^{-1} \cdots M_{k-1}^{-1} M_k^{-1}$, where

\begin{equation}
M_j^{-1} := I - U_r U_r^* + U_r G_r^{-1} U_r^*,
\end{equation}

for $j \geq 2$ and $M_1^{-1} := I$; the initial iterate is $y_{k,0} = \overline{M}_k x_{k,0} = \overline{M}_k x_{k-1,m}$. The matrix $U_r \in \mathbb{C}^{n \times r}$ ($U_r^* U_r = I$) and upper Hessenberg matrix $G_r \in \mathbb{C}^{r \times r}$ satisfy the Arnoldi relation $A U_r = U_r G_r + g_{r+1} u_r e_r$ and are computed using one cycle of the implicitly restarted Arnoldi iteration with exact shifts (see [20, 38]). A practical implementation would pre-bound, according to memory constraints, the number of approximate invariant subspaces kept. However, we leave the practical considerations and numerical efficiency to the persuasive results presented in [1]. The aim of this section is to examine the convergence of deflation algorithms that use Arnoldi vectors.

We compare GMRES-ADAP to a method that applies exact deflation preconditioning at each cycle, i.e., the preconditioner defined in (5.2) replaces $U_r$ with the matrix $X_r$ whose columns form an orthonormal basis for an $r$-dimensional invariant subspace, and $G_r$ with $X_r^* A X_r$. The invariant subspace is computed to minimize the angle between $\text{Ran}(X_r)$ and $\text{Ran}(U_r)$. We call this method GMRES-EDAP$(m,r)$, for GMRES with exact deflation adaptive preconditioning. Figure 5.5 shows that GMRES-EDAP performs only slightly better than GMRES-ADAP on this problem. Indeed this result is not unexpected, given Theorem 4.1: approximate deflation using Arnoldi vectors moves $r-1$ eigenvalues to 1, while exact deflation moves $r$ eigenvalues moved to 1. However, approximate deflation could indeed deflate $r-1$ eigenvalues, while shifting the others in an unfavorable manner (e.g., moving them close to the origin), especially if the matrix $\Phi$ defined in (4.6) is poorly conditioned, or the residual $\delta$ is too large. For such problems, GMRES-ADAP$(m,k)$ can be altered so that $U_r$ and $G_r$ are computed from several cycles of the implicitly restarted Arnoldi iteration until a predetermined tolerance is met, as described in [1, Alg. 3.6].

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 having favorable spectral properties. We illustrated this approach by analyzing approximate deflation preconditioning, then

![Fig. 5.3. GMRES convergence restarted once after 40 iterations, then preconditioned with a 40 dimensional approximate invariant subspace (dashed black line) and the nearby 40 dimensional exact invariant subspace given by Theorem 4.2 (solid black line). Both methods compare very favorably with full GMRES (no restart, no preconditioning: solid black line).](image-url)
Fig. 5.4. Spectra for approximate versus exact deflation. Top left: eigenvalues of $A$ (·) and Ritz values (+) from $X_{40}(A, r_0)$. Top right: eigenvalues of $A$ (·) with circles denoting the eigenvalues corresponding to the 40-dimensional invariant subspace closest to $X_{40}(A, r_0)$. Bottom left: eigenvalues of the approximately deflated matrix $AM^{-1}$. Bottom right: eigenvalues of the exactly deflated matrix $AM^{-1}D$ corresponding to the invariant subspace identified in the top-right plot. One would expect all circled eigenvalues in the top right to be absent from the plot on the bottom right; however, some of those circled eigenvalues are double eigenvalues (or nearly so), and only a single copy is deflated.

Illustrated the behavior of such a preconditioner applied to a linear system from acoustic scattering. 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 $E$. For special classes of perturbation (e.g., $E = \alpha I$ for some constant $\alpha$, or $E \in \mathbb{R}^{n \times n}$), a refined analysis may yield more precise bounds.

Acknowledgements. We thank Howard Elman, Sabaresan Mothi, Daniel Szyld, and Gilbert Ymbert for helpful comments on this work, and McKay Hyde for introducing us to the Lippmann–Schwinger application discussed in Section 5.

REFERENCES
Fig. 5.5. Approximately preconditioned GMRES: In both plots, the gray shows full GMRES convergence; the dashed line shows approximate deflation, GMRES-ADAP(40, r); and the solid line shows exact deflation, GMRES-EDAP(40, r). The left plot deflates with r = 20 dimensional subspaces; the right plot shows the advantage gained by using r = 40 dimensional subspaces.