HOW DESCRIPTIVE ARE GMRES CONVERGENCE BOUNDS?

MARK EMBREE*

Abstract. Eigenvalues with the eigenvector condition number, the field of values, and pseudospectra have all been suggested as the basis for convergence bounds for minimum residual Krylov subspace methods applied to non-normal coefficient matrices. This paper analyzes and compares these bounds, illustrating with six examples the success and failure of each one. Refined bounds based on eigenvalues and the field of values are suggested to handle low-dimensional non-normality. It is observed that pseudospectral bounds can capture multiple convergence stages. Unfortunately, computation of pseudospectra can be rather expensive. This motivates an adaptive technique for estimating GMRES convergence based on approximate pseudospectra taken from the Arnoldi process that is the basis for GMRES.

Key words. Krylov subspace methods, GMRES convergence, non-normal matrices, pseudospectra, field of values

AMS subject classifications. 15A06, 65F10, 15A18, 15A60, 31A15

1. Introduction. Popular algorithms for solving large, sparse systems of linear equations construct iterates that attempt to minimize the residual norm over all candidates in an affine Krylov subspace whose dimension grows at each step. For non-symmetric matrices, the GMRES algorithm of Saad and Schultz [33] generates such optimal iterates. This method, based on the Arnoldi process with its long vector recurrences, is intractable for problems that converge slowly. Practical algorithms, such as BiCGSTAB or QMR (see, e.g., [18, Ch. 5], [32, Ch. 7]), reduce this computational expense by only approximating the optimality property. It is tough to characterize the convergence that results from compromising optimality. The residual norms cannot be smaller than those produced by GMRES (as the algorithms choose iterates from the same Krylov subspace), but they can sometimes be related to the GMRES residual norm (e.g., for QMR [16]). Understanding GMRES convergence, facilitated by its optimality property, is thus an important step towards convergence analysis for general algorithms. It can also inform the construction and evaluation of preconditioners for non-symmetric problems.

Given a system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$, with $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\mathbf{x}, \mathbf{b} \in \mathbb{C}^n$, the GMRES algorithm [33] iteratively generates solution estimates $\mathbf{x}_k$ based on an initial guess $\mathbf{x}_0$. The residuals induced by these iterates, $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$, satisfy the minimum residual property,

$$\|\mathbf{r}_k\|_2 = \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\mathbf{r}_0\|_2, \tag{1.1}$$

where $\mathcal{P}_k$ is the set of polynomials of degree k or less.

What properties of the coefficient matrix $\mathbf{A}$ govern convergence? In this paper, we examine three prominently proposed answers to this question (see [18]): eigenvalues with eigenvector condition number; the field of values; and pseudospectra. When $\mathbf{A}$ is normal (i.e., it has an orthogonal basis of eigenvectors or, equivalently, it commutes with its adjoint), convergence can be accurately bounded using the eigenvalues alone. This is not the case for non-normal matrices, as the construction of Greenbaum, Pták, and Strakoš dramatically illustrates [19]. When $\mathbf{A}$ is significantly non-normal, the

*Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, UK (embree@comlab.ox.ac.uk).

sequence of residual norms $\{\|\mathbf{r}_k\|_2\}$ often exhibits a period of initial stagnation before converging at a quicker asymptotic rate. The bounds we study here essentially differ from the standard eigenvalue-only bounds for normal GMRES in the mechanisms they use to predict the duration of this transient period of convergence.

We describe three standard GMRES convergence bounds in Section 2. These characterizations can be misleading when non-normality is only associated with a few eigenvalues (e.g., several nearly aligned eigenvectors orthogonal to all other eigenvectors). To circumvent this, we apply spectral projectors to modify the traditional formulations. This strategy bounds GMRES convergence using the condition numbers of individual eigenvalues, and leads to a flexible generalization of the field of values bound. In Section 3, we present six examples to illustrate that the three standard bounds can each be tricked into dramatic overestimates, but each can also be rather descriptive. The bounds are also compared via the analytic relationships between the eigenvectors, field of values, and pseudospectra. The examples highlight a strength of pseudospectral bounds: convergence rates based on different pseudospectral sets can accurately describe different phases of convergence. Unfortunately, the cost of pseudospectral computation makes this bound rather expensive for large, practical problems. This motivates our Section 4, where we suggest an alternative that leads to GMRES convergence *estimates* at a lower computational expense based on approximate pseudospectra taken from data generated by the Arnoldi process within the standard GMRES implementation.

Though we are implicitly interested in a linear system with a specific initial residual, all the analysis described here first employs the inequality

$$\|\mathbf{r}_k\|_2 \le \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \, \|\mathbf{r}_0\|_2, \tag{1.2}$$

and then studies $\|p(\mathbf{A})\|_2$ independently of $\mathbf{r}_0$. This leads to upper bounds for worst case GMRES convergence. With a carefully crafted example, Toh proved that this inequality can be arbitrarily misleading for non-normal matrices [35]. There may be *no* vector $\mathbf{r}_0 \in \mathbf{C}^n$ for which $\|p(\mathbf{A})\|_2$ characterizes $\|\mathbf{r}_k\|_2/\|\mathbf{r}_0\|_2$ at iteration k. Examples of this extreme behavior are thought to be rare in practice [34, §3.6] and thus we are typically satisfied with the inequality (1.2) and the general analysis of $\|p(\mathbf{A})\|_2$ that follows from it.

2. Three Convergence Bounds and Variations. In this section, we derive three familiar convergence bounds for GMRES, based on eigenvalues with the eigenvector condition number, the field of values, and pseudospectra. These bounds all fail to accurately describe convergence when non-normality is primarily associated with just part of the spectrum. This motivates the use of spectral projectors to develop localized versions of these bounds that can be sharper than the traditional versions.

2.1. Eigenvalues with Eigenvector Conditioning. The first convergence bound suggested for GMRES predicts convergence at a rate determined by the set of eigenvalues of $\mathbf{A}$, denoted $\Lambda(\mathbf{A})$. If $\mathbf{A}$ is normal, $\Lambda(\mathbf{A})$ determines convergence. Non-normality may delay the onset of convergence at this spectral rate; to account for such stagnation, this bound scales the spectral convergence prediction by the condition number of the matrix having the eigenvectors of $\mathbf{A}$ as its columns [10],[33]. Provided that $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, we have

$$\|\mathbf{r}_k\|_2 = \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\mathbf{r}_0\|_2 \le \|\mathbf{V}p(\mathbf{\Lambda})\mathbf{V}^{-1}\|_2 \, \|\mathbf{r}_0\|_2,$$

implying the bound

$$\text{(EV)} \qquad \frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \leq \kappa(\mathbf{V}) \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \max_{\lambda\in\Lambda(\mathbf{A})} |p(\lambda)|.$$

Here, $\kappa(\mathbf{V}) \equiv \|\mathbf{V}\|_2 \, \|\mathbf{V}^{-1}\|_2$ is the 2-norm condition number of the eigenvector matrix $\mathbf{V}$. If $\mathbf{A}$ is normal, then $\kappa(\mathbf{V}) = 1$; if, in addition, the eigenvalues are real, then (EV) reduces to the standard convergence bound for MINRES [13]. If $\mathbf{A}$ is non-normal, then $\kappa(\mathbf{V}) > 1$ and determining the optimal value of $\kappa(\mathbf{V})$ can be a challenge [20]; this task is further complicated if $\mathbf{A}$ has repeated eigenvalues. Throughout this work, the columns of $\mathbf{V}$ have unit 2-norm; provided each eigenvalue of $\mathbf{A}$ is simple, $\kappa(\mathbf{V})$ with this scaling can be no larger than $\sqrt{n}$ times the optimal value, where n is the matrix dimension [43].

Typically, the polynomial minimization in (EV) predicts a linear asymptotic convergence rate, and $\kappa(\mathbf{V})$ reflects the non-normality of the matrix $\mathbf{A}$. Since GMRES residual norms necessarily form a non-increasing sequence, a large $\kappa(\mathbf{V})$ implies that the bound (EV) will only possibly be descriptive for latter iterations. Even then, it can be grossly inaccurate. For example, $\kappa(\mathbf{V})$ will be large if only two eigenvectors are nearly aligned, or if all eigenvectors are. In the former case, the bound usually fails to predict convergence, while it may be more appropriate in the latter case. This is illustrated in Examples B and E of Section 3.

The bound (EV) can be sharpened to get around this difficulty by considering the conditioning of individual eigenvalues. Suppose $\lambda \in \Lambda(\mathbf{A})$ is simple with left and right eigenvectors $\mathbf{u}$ and $\mathbf{v}$ respectively. Then the condition number of λ [44, §2.8] is

$$\kappa(\lambda) \equiv \frac{\|\mathbf{u}\|_2 \, \|\mathbf{v}\|_2}{|\mathbf{u}^*\mathbf{v}|}.$$

Using these condition numbers leads to a bound that can be much sharper than (EV).

THEOREM 2.1. *Suppose every eigenvalue λ_j of $\mathbf{A}$ is simple. Then for any $p \in \mathcal{P}_k$,*

$$\text{(2.1)} \qquad \|p(\mathbf{A})\|_2 \leq \sum_{j=1}^{n} \kappa(\lambda_j) \, |p(\lambda_j)|.$$

Proof. Since $\mathbf{A}$ has simple eigenvalues, it is diagonalizable, $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$. Let $\{\mathbf{u}_j\}_{j=1}^n$ be the left eigenvectors ($\mathbf{u}_j^*$ is the jth row of $\mathbf{V}^{-1}$) and $\{\mathbf{v}_j\}_{j=1}^n$ the corresponding right eigenvectors (columns of $\mathbf{V}$), with $\mathbf{\Lambda}_{jj} = \lambda_j$. Then

$$\|p(\mathbf{A})\|_2 = \|\mathbf{V}p(\mathbf{\Lambda})\mathbf{V}^{-1}\|_2 = \left\| \sum_{j=1}^{n} p(\lambda_j)\mathbf{v}_j\mathbf{u}_j^* \right\|_2 \leq \sum_{j=1}^{n} |p(\lambda_j)| \, \|\mathbf{v}_j\mathbf{u}_j^*\|_2.$$

The result follows from noting that since $\mathbf{u}_j^*\mathbf{v}_j = 1$ by construction, $\|\mathbf{v}_j\mathbf{u}_j^*\|_2 = \|\mathbf{u}_j\|_2 \, \|\mathbf{v}_j\|_2 = \|\mathbf{u}_j\|_2 \, \|\mathbf{v}_j\|_2 / |\mathbf{u}_j^*\mathbf{v}_j| = \kappa(\lambda_j)$. ∎

Notice that the quantity on the right of equation (2.1) is simply the 1-norm of $p(\mathbf{\Lambda})\mathbf{r}$ where $r_j = \kappa(\lambda_j)$. Norm equivalence reduces this problem to a conventional GMRES problem involving a normal matrix, but with a very special right hand side.

COROLLARY 2.2. *Define $\mathbf{r} \in \mathbb{C}^n$ by $r_j = \kappa(\lambda_j)$. Then*

$$\text{(EV}') \qquad \frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \leq \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \leq \sqrt{n} \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{\Lambda})\mathbf{r}\|_2.$$

The jth component of $\mathbf{r}$ will be large if λ_j is ill-conditioned. If some eigenvalues are much better conditioned, GMRES applied to $\mathbf{\Lambda}$ with this $\mathbf{r}$ will focus on the largest entries first, leading to more rapid convergence than one would expect for a typical initial residual of similar magnitude. Figure 4.3 shows the ability of (EV′) to describe convergence for a highly non-normal matrix from a convection–diffusion problem.

Non-normality is closely related to eigenvalue instability [40], which can complicate the computation of $\Lambda(\mathbf{A})$ and $\kappa(\mathbf{V})$. This motivates us to look for more stable convergence bound formulations.

2.2. Field of Values. Recently, the field of values has garnered interest as an alternative foundation for GMRES bounds. The *field of values*, or *numerical range*, is the set of all Rayleigh quotients,

$$W(\mathbf{A}) \equiv \left\{ \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \;\middle|\; \mathbf{x} \in \mathbb{C}^n,\, \mathbf{x} \neq 0 \right\},$$

and the largest magnitude of a point in $W(\mathbf{A})$ is called the *numerical radius*, $\mu(\mathbf{A}) \equiv \max_{z \in W(\mathbf{A})} |z|$. The field of values is always a convex set containing $\Lambda(\mathbf{A})$. It is computationally more attractive than the eigenvalues and eigenvectors because it is not sensitive to perturbations and can be effectively estimated even for large, sparse matrices using, for example, techniques introduced by Braconnier and Higham [3].

Eiermann has recently developed GMRES bounds based on the field of values, provided $0 \notin W(\mathbf{A})$ [7],[8]. Note that for any $\mathbf{A}$, the matrix 2-norm is bounded by twice the numerical radius, $\|\mathbf{A}\|_2 \leq 2\mu(\mathbf{A})$ (see, e.g., [24, Ch. 1]), implying that for any $p \in \mathcal{P}_k$, $\|p(\mathbf{A})\|_2$ can be bounded by $2\mu(p(\mathbf{A}))$. Kato's field of values mapping theorem [26] ensures that $\mu(p(\mathbf{A})) \leq \max_{z \in W(\mathbf{A})} |p(z)|$, thus giving the bound

$$\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \leq 2 \min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \max_{z \in W(\mathbf{A})} |p(z)|. \tag{FOV}$$

Eiermann and Ernst have recently proposed a different field of values bound involving $W(\mathbf{A}^{-1})$ [9], but we focus here on the more accessible statement (FOV). Several limitations of this bound are clear from the start. Suppose that $W(\mathbf{A})$ contains points far from the eigenvalues due to high non-normality associated with a low-degree invariant subspace of $\mathbf{A}$. Then a matrix polynomial $p(\mathbf{A})$ may well annihilate the non-normal portion of $\mathbf{A}$, leaving Kato's inclusion $W(p(\mathbf{A})) \subseteq p(W(\mathbf{A}))$ rather slack. As the constant term in (FOV) is small and the approximation problem on a convex region predicts asymptotic linear convergence (see equation (2.5)), the bound (FOV) cannot be accurate for iterations that initially stagnate before converging at a more rapid asymptotic rate. This behavior is observed by Higham and Trefethen in the context of matrix powers [22], and identified by Ernst in the context of GMRES applied to convection–diffusion problems [12], as we see in Section 4.1.

Another drawback of this approach is its requirement that $0 \notin W(\mathbf{A})$, which rules out indefinite problems. This approach also suffers from outlying eigenvalues, which can artificially stretch the critical convergence region. We address these problems by working with projectors onto invariant subspaces of $\mathbf{A}$. Partition the spectrum of $\mathbf{A}$ into disjoint sets Λ_j, such that $\Lambda(\mathbf{A}) = \cup\Lambda_j$. Define the spectral projector

$$\mathbf{P}_j \equiv \int_{\Gamma_j} \frac{1}{2\pi \mathrm{i}} (z\mathbf{I} - \mathbf{A})^{-1} \, \mathrm{d}z,$$

where Γ_j is the union of Jordan curves containing the eigenvalues Λ_j in their collective interior, but not enclosing any other eigenvalues. Then $\mathbf{P}_j$ is a projector onto the invariant subspace of $\mathbf{A}$ associated with the eigenvalues Λ_j (see, e.g., [27, §I.5.3]).

THEOREM 2.3. *Let $\{\Lambda_j\}_{j=1}^m$ be a partition of $\Lambda(\mathbf{A})$ into m disjoint sets. For each $1 \le j \le m$, let $\mathbf{P}_j$ be the spectral projector onto the invariant subspace associated with Λ_j, and let the columns of $\mathbf{U}_j^{n\times \operatorname{rank}(\mathbf{P}_j)}$ be an orthonormal basis for* Ran $\mathbf{P}_j$. *Then for any polynomial $p \in \mathcal{P}_k$,*

$$\|p(\mathbf{A})\|_2 \le \sum_{j=1}^{m} \|\mathbf{P}_j\|_2 \, \|p(\mathbf{U}_j^* \mathbf{A} \mathbf{U}_j)\|_2.$$

Proof. Note that $\mathbf{\Pi}_j \equiv \mathbf{U}_j \mathbf{U}_j^*$ is the orthogonal projector onto Ran $\mathbf{P}_j$, an invariant subspace of $\mathbf{A}$. We apply two important identities for spectral projectors: $\sum_{j=1}^m \mathbf{P}_j = \mathbf{I}$ and $\|\mathbf{A}\mathbf{P}_j\|_2 = \|\mathbf{A}\mathbf{\Pi}_j\mathbf{P}_j\|_2$ (see, e.g., [27, §I.5.3],[31, §III.1]). Substituting the first identity into $\|p(\mathbf{A})\|_2$ yields

$$\begin{aligned} \|p(\mathbf{A})\|_2 = \left\| p(\mathbf{A}) \sum_{j=1}^{m} \mathbf{P}_j \right\|_2 &\le \sum_{j=1}^{m} \|p(\mathbf{A})\mathbf{P}_j\|_2 = \sum_{j=1}^{m} \|\mathbf{\Pi}_j \, p(\mathbf{A}) \, \mathbf{\Pi}_j \mathbf{P}_j\|_2 \\ &\le \sum_{j=1}^{m} \|\mathbf{\Pi}_j \, p(\mathbf{A}) \, \mathbf{\Pi}_j\|_2 \, \|\mathbf{P}_j\|_2. \end{aligned} \tag{2.2}$$

Notice that for each j, $\|\mathbf{\Pi}_j p(\mathbf{A}) \mathbf{\Pi}_j\|_2 = \|p(\mathbf{\Pi}_j \mathbf{A} \mathbf{\Pi}_j)\|_2 = \|\mathbf{U}_j \, p(\mathbf{U}_j^* \mathbf{A} \mathbf{U}_j) \, \mathbf{U}_j^*\|_2 \le \|p(\mathbf{U}_j^* \mathbf{A} \mathbf{U}_j)\|_2$. This combines with (2.2) to complete the proof. ∎

This theorem provides a natural tool for transferring between global statements, such as (EV), and localized statements like (EV′). In the former case, Theorem 2.3 is vacuous since the spectrum is partitioned into a single set; in the latter case, Theorem 2.3 reduces to Theorem 2.1, as each set Λ_j is taken as a single eigenvalue and thus $\|\mathbf{P}_j\|_2 = \kappa(\lambda_j)$. In summary, the cost of localizing non-normality is the introduction of the spectral projector norm.

Theorem 2.3 can be combined with the analysis leading to the bound (FOV) to provide a field of values analog to (EV′).

COROLLARY 2.4. *Partition the eigenvalues $\Lambda(\mathbf{A})$ into disjoint sets $\{\Lambda_j\}_{j=1}^m$ as before, with the orthogonal columns of each $\mathbf{U}_j$ spanning the invariant subspace of $\mathbf{A}$ associated with Λ_j. Then*

$$\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \le \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \le \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} 2 \sum_{j=1}^{m} \max_{z\in W(\mathbf{U}_j^*\mathbf{A}\mathbf{U}_j)} \|\mathbf{P}_j\|_2 \, |p(z)|. \tag{FOV′}$$

Provided one is willing to localize sufficiently, (FOV′) is applicable to a wide range of matrices, including all matrices with only simple eigenvalues. Some matrices are still beyond the reach of this field of values analysis, though, including any sufficiently non-normal Jordan block.

2.3. Pseudospectra. Pseudospectra provide an alternative stable platform for GMRES bounds. The ε-pseudospectrum [40] is defined by

$$\Lambda_\varepsilon(\mathbf{A}) \equiv \{z \in \mathbb{C} \mid \|(z\mathbf{I} - \mathbf{A})^{-1}\|_2 \ge \varepsilon^{-1}\};$$

an equivalent formulation is $\Lambda_\varepsilon(\mathbf{A}) = \{z \in \mathbb{C} \mid z \in \Lambda(\mathbf{A}+\mathbf{E}), \|\mathbf{E}\|_2 \le \varepsilon\}$. In an early application of pseudospectral theory, Trefethen [39] developed GMRES bounds by working from the Dunford–Taylor integral [27, §I.5.6] for any polynomial $p \in \mathcal{P}_k$,

$$p(\mathbf{A}) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma p(z)(z\mathbf{I}-\mathbf{A})^{-1}\,\mathrm{d}z, \tag{2.3}$$

where Γ is any union of Jordan curves containing $\Lambda(\mathbf{A})$ in its interior. For a fixed $\varepsilon > 0$, we can choose the contour Γ_ε to be the boundary of $\Lambda_\varepsilon(\mathbf{A})$. (If this isn't the union of Jordan curves, we can take Γ_ε to be slightly exterior.) Coarsely approximating the resolvent norm about Γ_ε yields

$$\|p(\mathbf{A})\|_2 \le \frac{1}{2\pi} \int_{\Gamma_\varepsilon} |p(z)|\,\|(z\mathbf{I}-\mathbf{A})^{-1}\|_2\,\mathrm{d}|z| \le \frac{\mathcal{L}(\Gamma_\varepsilon)}{2\pi\varepsilon} \max_{z\in\Lambda_\varepsilon(\mathbf{A})} |p(z)|,$$

where $\mathcal{L}(\Gamma_\varepsilon)$ is the contour length of Γ_ε. When applied to the GMRES problem, this gives the bound

$$\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \le \frac{\mathcal{L}(\Gamma_\varepsilon)}{2\pi\varepsilon} \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \max_{z\in\Lambda_\varepsilon(\mathbf{A})} |p(z)|. \tag{PSA}$$

A striking feature of this bound, perhaps not widely appreciated, is that it applies for different values of ε. Moreover, these different values of ε may apply at different stages of the iteration. This is strikingly observed in Figure 3.6. As ε shrinks, the associated pseudospectral sets shrink too, and in that sense the pseudospectra form a bridge between the field of values and the spectrum [22]. A localized version of (PSA) is obtained by simultaneously applying different values of ε on discrete pseudospectral components in (2.3); unlike (EV′) and (FOV′), this does not introduce spectral projector norms. Pseudospectral calculation is expensive, but improvements to the naive approach greatly expedite the process; see Trefethen's recent survey [41]. We explore an alternative approach in Section 4.

2.4. Computing the Convergence Bounds. With each of the bounds (EV), (FOV), and (PSA) is associated a constant, defined as

$$C_{\mathrm{EV}} \equiv \kappa(\mathbf{V}), \qquad C_{\mathrm{FOV}} \equiv 2, \qquad \text{and} \quad C_{\mathrm{PSA}}(\varepsilon) \equiv \frac{\mathcal{L}(\Gamma_\varepsilon)}{2\pi\varepsilon}.$$

The asymptotic behavior of each bound is determined by the associated complex approximation problem over $\Lambda(\mathbf{A})$, $W(\mathbf{A})$, or $\Lambda_\varepsilon(\mathbf{A})$.

Let $\Omega \subset \mathbb{C}$ be a compact set without isolated points tightly bounding $W(\mathbf{A})$, $\Lambda_\varepsilon(\mathbf{A})$, or the clustered eigenvalues of $\Lambda(\mathbf{A})$[1]. The error of the approximation problem, $\min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \max_{z\in\Omega} |p(z)|$, decreases asymptotically linearly in the polynomial degree k (see, e.g., [23, Ch. 16]). With this linear rate is associated a constant, ρ, the asymptotic convergence rate of Ω. Driscoll, Toh, and Trefethen demonstrate how this constant can be computed via conformal mapping [5]. When the set is a line segment or a disk, the rate is simple to compute. For arbitrary polygons, the rate can be computed using Driscoll's Schwarz–Christoffel MATLAB toolbox for numerical conformal mapping [4].

[1] If $\mathbf{A}$ has finite dimension, $\Lambda(\mathbf{A})$ is a discrete point set with no finite asymptotic convergence rate. If the eigenvalues are clustered, the asymptotic convergence rate of the bounding set Ω typically describes convergence. Outlying eigenvalues don't affect this convergence rate; see [5] for details.

More general sets may arise, for example, if $\Lambda(\mathbf{A})$ has outliers or for $\Lambda_\varepsilon(\mathbf{A})$ with sufficiently small ε. If each connected component of a multiply connected set is a polygon on the real axis and is symmetric about the real axis, the rate can still be computed [11]. More general sets present greater difficulty, and it may in practice be necessary to bound $\Lambda(\mathbf{A})$ or $\Lambda_\varepsilon(\mathbf{A})$ with a single over-sized polygon.

To obtain computable convergence bounds, we must relate minimization over Ω with a degree k polynomial p satisfying $p(0) = 1$ to ρ^k. We are guaranteed that

$$\min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \max_{z \in \Omega} |p(z)| \geq \rho^k \tag{2.4}$$

(see, e.g., [5]), and if Ω is a disk, this is sharp. When Ω is a segment $[a, b]$ of a line passing through the origin, shifted and scaled Chebyshev polynomials are optimal. In this case, the minimax error is bounded above by $2\rho^k$ and known explicitly for each k (see, e.g., [32, §6.11]). If Ω is convex, Eiermann [6],[8] uses Faber polynomial analysis based on an approximation theorem of Kövari and Pommerenke [28] to show that

$$\min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \max_{z \in \Omega} |p(z)| \leq \frac{2\rho^k}{1 - \rho^k}. \tag{2.5}$$

In particular, this bound always applies to the field of values [8]. In other circumstances, one can compute the Faber polynomials associated with Ω from the conformal map that determines ρ, and these polynomials will provide an upper bound on $\min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \max_{z \in \Omega} |p(z)|$. To unify notation, we label the rate associated with each of the three sets $\Lambda(\mathbf{A})$, $W(\mathbf{A})$, and $\Lambda_\varepsilon(\mathbf{A})$ as ρ_{EV}, ρ_{FOV}, and $\rho_{\text{PSA}}(\varepsilon)$.

3. Which Bounds are Useful? In the previous section, we developed the three standard bounds (EV), (FOV), and (PSA), along with "localized" versions (EV′) and (FOV′). How do these bounds compare to one another? Are they redundant, or can each provide specific insight? We begin by exploring the analytic relationships between the sets $\Lambda(\mathbf{A})$, $W(\mathbf{A})$, and $\Lambda_\varepsilon(\mathbf{A})$, and then turn to concrete examples illustrating the relative merits of the three standard bounds.

If the pseudospectra are large, then the field of values and the eigenvalue condition number must also be large, in the sense defined in the following theorems. The first, a version of the Bauer–Fike theorem [2],[42], bounds the pseudospectra by $\kappa(\mathbf{V})$. The second relates the pseudospectra to the field of values (see Gustafson and Rao [21, §4.6]). The third relates the field of values to the eigenvector condition number as a consequence of the basic inequality $\mu(\mathbf{A}) \leq \|\mathbf{A}\|_2$. Let $\Delta_r \equiv \{z \in \mathbb{C} \mid |z| \leq r\}$ denote the disk of radius r.

THEOREM 3.1 (Bauer–Fike). *Let* $\mathbf{A}$ *be diagonalizable,* $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$. *Then for any* $\varepsilon > 0$, $\Lambda_\varepsilon(\mathbf{A}) \subseteq \Lambda(\mathbf{A}) + \Delta_{\varepsilon\kappa(\mathbf{V})}$.

THEOREM 3.2. *For any* $\varepsilon > 0$, $\Lambda_\varepsilon(\mathbf{A}) \subseteq W(\mathbf{A}) + \Delta_\varepsilon$.

THEOREM 3.3. *Let* $\mathbf{A}$ *be diagonalizable,* $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$. *Then* $\mu(\mathbf{A}) \leq \kappa(\mathbf{V})\,\rho(\mathbf{A})$, *where* $\rho(\mathbf{A}) = \max_{\lambda \in \Lambda(\mathbf{A})} |\lambda|$ *is the spectral radius of* $\mathbf{A}$.

Theorems 3.1 and 3.3 are sharp whenever $\mathbf{A}$ is normal. Theorem 3.2 is sharp if $\mathbf{A}$ is a multiple of the identity, or if $\mathbf{A}$ is a Jordan block in the limit $n = \infty$ (see Example D). For non-normal matrices, all three of these theorems may be weak.

Theorem 3.1 leads to a simple bound on the constant $C_{\text{PSA}}(\varepsilon)$. Note that $\Lambda_\varepsilon(\mathbf{A})$ is bounded by the union of n disks each with radius $\varepsilon\kappa(\mathbf{V})$. Choosing Γ_ε to be the boundary of this union, $\mathcal{L}(\Gamma_\varepsilon)$ can be no larger than $2\pi n\varepsilon\kappa(\mathbf{V})$, so $C_{\text{PSA}}(\varepsilon) \leq n\kappa(\mathbf{V})$.

TABLE 3.1
Predicted iterations for the six examples of Section 3.1.

example		(EV)	(FOV)	(PSA)	*true iterations*
A:	all descriptive	1	1	1	1
B:	none descriptive	∞	∞	∞	*see note 1*
C:	(EV) wins		*see note 2*		
D:	(EV) loses	∞	1	1	1
	(FOV) wins		*see note 3*		
E:	(FOV) loses	2	∞	2	2
F:	(PSA) wins	∞	∞	2	2
	(PSA) loses		*see note 3*		

1. $2 + \log(2/\mathrm{TOL})/\log(\rho)$ iterations, for parameters $\mathrm{TOL} < 1$ and $\rho \in (0,1)$.
2. "(EV) wins" is more involved; details are given in the text.
3. (FOV) can't dramatically beat (PSA); see the "Non-Example" sections below.

But since $\Lambda(\mathbf{A}) \subset \Lambda_\varepsilon(\mathbf{A})$, $\rho_{\mathrm{EV}} < \rho_{\mathrm{PSA}}(\varepsilon)$ for all $\varepsilon > 0$, and thus (PSA) is only possibly useful for those values of ε with $C_{\mathrm{PSA}}(\varepsilon) < \kappa(\mathbf{V})$.

Similarly, Theorem 3.2 implies that $C_{\mathrm{PSA}}(\varepsilon) \to 1$ as $\varepsilon \to \infty$. Note that $\Lambda_\varepsilon(\mathbf{A})$ is bounded by the disk centered at 0 with radius $\mu(\mathbf{A}) + \varepsilon$. This bounding set gives the constant $C_{\mathrm{PSA}}(\varepsilon) = 1 + \mu(\mathbf{A})/\varepsilon$, and the bound follows as $\varepsilon \to \infty$. When the containment $\Lambda_\varepsilon(\mathbf{A}) \subset W(\mathbf{A}) + \Delta_\varepsilon$ is nearly equality even for small values of ε, the bound (FOV) can be slightly sharper than (PSA), as is seen in Figures 3.1, 3.3, and 3.4. In cases where the bound in Theorem 3.2 is weak, one often finds that (FOV) predicts slow, consistent convergence, while (PSA) predicts convergence that eventually accelerates to a more rapid rate, as in Figures 3.5, 3.6, and 4.2.

3.1. The Examples. The bounds (EV), (FOV), and (PSA) are descriptive in different situations. We demonstrate with six examples where the bounds succeed together, fail together, and, in turn, fail and succeed alone. These examples are summarized in Table 3.1. We only discuss the standard bounds, though in some instances a localized version would fix the flaw that causes the corresponding standard bound to fail. It is difficult to show the failure of (PSA) with the simultaneous success of (EV) or (FOV). Example C, showing success of (EV) with pessimistic (PSA) bounds, is the least satisfying of our six examples. We also discuss why (FOV) cannot dramatically outperform (PSA).

In the illustrations that follow, $\min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2$ is drawn as a solid line with dots at each iteration k. The bound (EV) is drawn as a solid line, (FOV) with a broken line, and (PSA) with dotted lines for various values of ε.

• **Example A: All descriptive.** All bounds accurately describe GMRES convergence for a scalar multiple of the identity,

$$\mathbf{A} = \alpha \mathbf{I}, \quad \alpha \in \mathbb{C} \setminus \{0\}.$$

Since $\mathbf{A}$ is normal with a single eigenvalue, $\Lambda(\mathbf{A}) = W(\mathbf{A}) = \{\alpha\}$ and $\Lambda_\varepsilon(\mathbf{A}) = \{\alpha\} + \Delta_\varepsilon$. The approximation problems in (EV) and (FOV) are on singleton sets, and

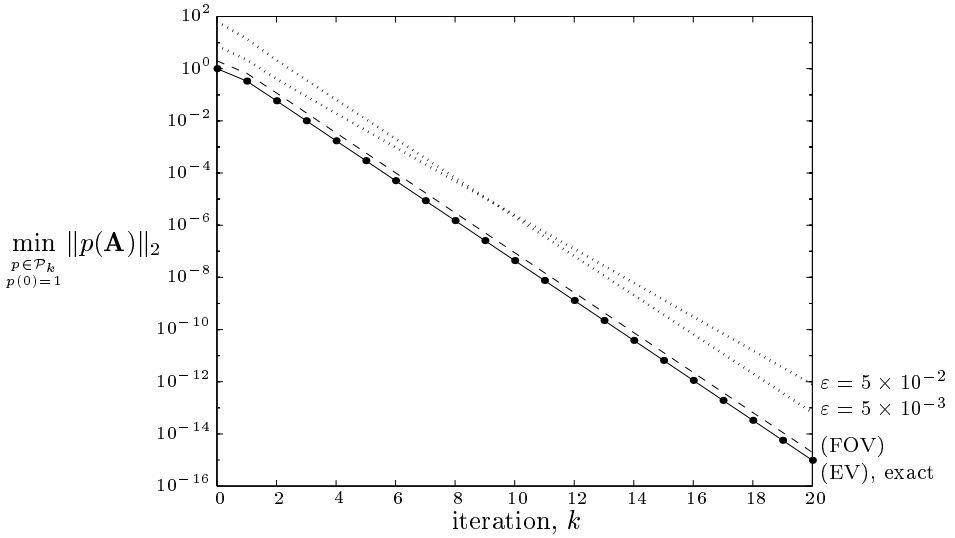

FIG. 3.1. *Convergence bounds for a normal matrix with* $\Lambda(\mathbf{A}) = [1, 2]$ *and dimension* $n = \infty$.

thus these bounds predict convergence in one iteration; the associated constants are $C_{\text{EV}} = 1$ and $C_{\text{FOV}} = 2$. The pseudospectral bound does as well. The constant term is $C_{\text{PSA}} = 1$ independent of ε and, with its approximation problem on a disk, (PSA) predicts $\min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \leq (\varepsilon/\alpha)^k$, implying convergence to arbitrary tolerance in a single iteration as $\varepsilon \to 0$.

What happens when $\mathbf{A}$ remains normal but $\Lambda(\mathbf{A})$ is the positive real interval $[a, b]$, rather than a single point? For dimension $n = \infty$, (FOV) and (EV) predict the same rate of convergence, $(\sqrt{b/a} - 1)/(\sqrt{b/a} + 1)$, with $C_{\text{EV}} = 1$ and $C_{\text{FOV}} = 2$ as before. But now (PSA) becomes slightly less accurate, as $\Lambda_\varepsilon(\mathbf{A})$ consists of $[a, b]$ surrounded by a border of radius ε. The constant C_{PSA} now involves a finite length scale, $C_{\text{PSA}}(\varepsilon) = (b-a)/(\pi\varepsilon)+1$. Descriptive pseudospectral bounds require balancing the more accurate convergence rates obtained for small ε against the growth of the constant $C_{\text{PSA}}(\varepsilon)$. Figure 3.1 illustrates this situation for $\Lambda(\mathbf{A}) = [1, 2]$.

• **Example B: None descriptive.** As described in Section 2, each bound can be deceived by low-dimensional non-normality. Example B exploits this shortcoming:

$$\mathbf{A} = \begin{pmatrix} 1 & \alpha \\ & 1 \end{pmatrix} \oplus \widehat{\mathbf{\Lambda}}, \qquad \alpha \gg 1,$$

where $\widehat{\mathbf{\Lambda}}$ is a diagonal matrix with uniformly distributed real positive entries in the positive real interval $[a, b]$. GMRES requires two iterations to eliminate the highly non-normal Jordan block, and then convergence depends only on the interval $[a, b]$. Thus, GMRES can be bounded from above independent of α.

Since $\mathbf{A}$ is non-diagonalizable, $C_{\text{EV}} = \infty$ and (EV) isn't helpful. The field of values grows ever larger with α, with $1 + \Delta_{\alpha/2} \subseteq W(\mathbf{A})$. Consequently, the field of values bound predicts no convergence when $\alpha \geq 2$. Analysis of the pseudospectral bound is slightly more involved. (PSA) accurately predicts convergence of the Jordan

block as $\varepsilon \to 0$ (see Example F), but such values of ε lead to large constant terms associated with $[1,2]$, the normal eigenvalues of $\mathbf{A}$ (as described in Example A). One can show that $C_{\text{PSA}}(\varepsilon) \geq \alpha/\sqrt{\varepsilon}$ for all $\varepsilon > 0$. For ε sufficiently small such that the component of the pseudospectrum generated by the Jordan block does not intersect the component associated with $\widehat{\mathbf{\Lambda}}$, we have $C_{\text{PSA}}(\varepsilon) \geq \alpha/\sqrt{\varepsilon} + (b-a)/(\pi\varepsilon) + 1$. This bound can be made arbitrarily poor by increasing α, and thus in Table 3.1 we say that (PSA) predicts infinitely many iterations.

This example is essentially an extreme version of the diagonalizable example constructed by Greenbaum and Strakoš to demonstrate the failure of the pseudospectral bound (PSA) [20].

• **Example C: Only (EV) descriptive.** The field of values and pseudospectral bounds are based on approximation problems on dense sets in the complex plane, and so the convergence rates predicted by these bounds depend on parts of the spectrum that may be eliminated at an early stage of iteration. When there are outlier eigenvalues, (EV) may be significantly more descriptive than (PSA) and (FOV) because its approximation problem is discrete and isolated outliers do not influence ρ_{EV}. Define

$$\mathbf{A} = \delta \oplus \widehat{\mathbf{\Lambda}}, \qquad 0 < -\delta \ll 1,$$

where $\widehat{\mathbf{\Lambda}}$ is a diagonal matrix with entries uniformly distributed in the positive real interval $[a,b]$.

Since $\mathbf{A}$ is normal, convergence is determined by the spectrum. The bound (EV), with $C_{\text{EV}} = 1$, is exact. This convergence can be bounded by the polynomial $p_k(z) = (1 - z/\delta)q_{k-1}(z)$, where q_k is the optimal degree k residual polynomial on the interval $[a,b]$. The outlier eigenvalue δ near the origin will lead to an initial stagnation [5]; the polynomial p_k suggests that this plateau will last no more than

$$1 + \frac{\log(\delta) - \log(2\delta + b)}{\log(\rho)} \text{ iterations},$$

where $\rho = (\sqrt{b/a} - 1)/(\sqrt{b/a} + 1)$. Note that this stagnation is not due to non-normality, as is the case in Example F.

By placing the outlier on the opposite side of the origin from the rest of the spectrum, we ensure that $0 \in W(\mathbf{A})$ and thus (FOV) predicts no convergence. (If δ were small and positive, (FOV) would predict non-trivial, but pessimistic, convergence.) The pseudospectral bound suffers from the fact that it can't identify δ as a single eigenvalue, eliminated at an early stage of convergence. For any finite ε, the pseudospectrum about $\lambda = \delta$ is a disk of radius ε. The approximation problem on $\Lambda_\varepsilon(\mathbf{A})$ must always incorporate it in the asymptotic bounds. This effect diminishes as ε decreases, but such small values of ε lead to large constants ($C_{\text{PSA}}(\varepsilon) = 2 + (b-a)/(\pi\varepsilon)$ for $n = \infty$) due to the interval $[a,b]$. In particular, it is necessary to take $\varepsilon < \delta$ to obtain non-trivial convergence rates.

This is the least analytically compelling of our six examples. In particular, it is difficult to cleanly describe the convergence rates associated with (PSA). In Figure 3.2, we illustrate (PSA) for $\delta = 0.01$, $[a,b] = [1,2]$, and $n = \infty$. This representation of (PSA) is obtained by *underestimating* the convergence rate of $\Lambda_\varepsilon(\mathbf{A})$, taking instead the rate associated with the union of two intervals, $[\delta - \varepsilon, \delta + \varepsilon] \cup [a - \varepsilon, b + \varepsilon]$. This rate can be expressed in terms of elliptic integrals, as described and implemented in MATLAB by Fischer [13]. The pseudospectral constants are $C_{\text{PSA}}(\varepsilon) = 2 + (b-a)/(\pi\varepsilon)$.

While a single outlier, as presented in this example, may appear to be an easily-overcome obstacle for the bounds (FOV) and (PSA), the same phenomenon can prove

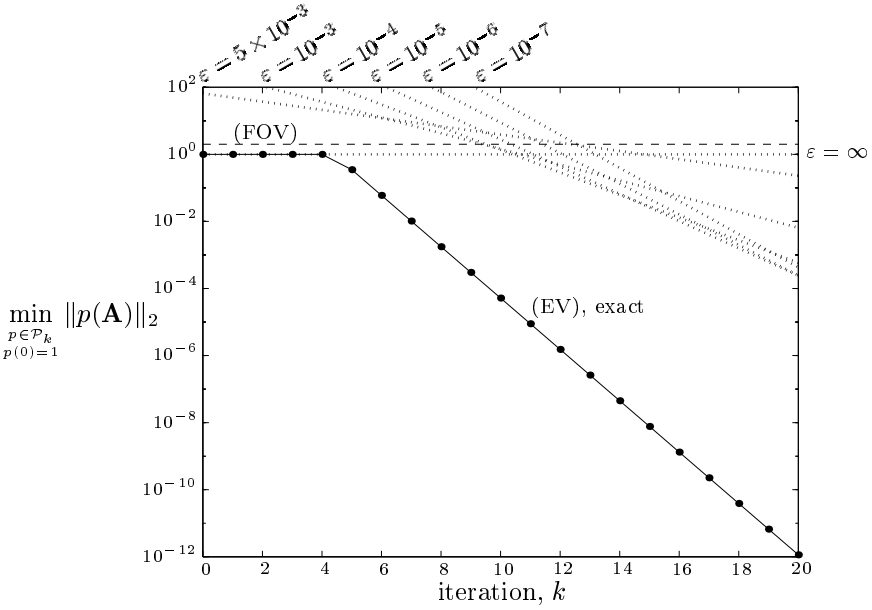

FIG. 3.2. *Convergence bounds for Example C with* $\delta = 0.01$ *and* $n = \infty$.

more difficult to identify when the outlier consists of a cluster of several eigenvalues, or if it is non-normal.

• **Example D: Only (EV) not descriptive.** The bound (EV) fails for any non-diagonalizable matrix. Yet defectiveness alone does not imply high non-normality. For the present example, consider small perturbations to the identity matrix,

$$\mathbf{A} = \begin{pmatrix} 1 & \delta & & \\ & 1 & \ddots & \\ & & \ddots & \delta \\ & & & 1 \end{pmatrix}, \qquad 0 < \delta \ll 1.$$

This matrix is completely defective for all $\delta \neq 0$, but small values of δ exert only the slightest impact on convergence. Unlike the case of $\delta = 0$, there will now be residuals that require n iterations to exactly converge. In fact, if $\mathbf{r}_0$ is the nth column of the identity matrix, Ipsen [25] shows that

$$\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} = \delta^k \sqrt{\frac{1-\delta^2}{1-\delta^{2(k+1)}}}. \tag{3.1}$$

This is nearly the worst case, as $\min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \leq \delta^k$.

Since $\mathbf{A}$ is non-diagonalizable, (EV) fails to predict convergence. The field of values is known explicitly for this example, $W(\mathbf{A}) = 1 + \Delta_{\delta \cos((n+1)^{-1})}$ [21, §1.3], leading to the exact formulation of (FOV): $\min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \leq 2(\delta \cos((1+n)^{-1}))^k$. The pseudospectra are also disks for this example [42], but the radius of these disks isn't generally known in closed form. In the limit $n \to \infty$, a theorem of Reichel and

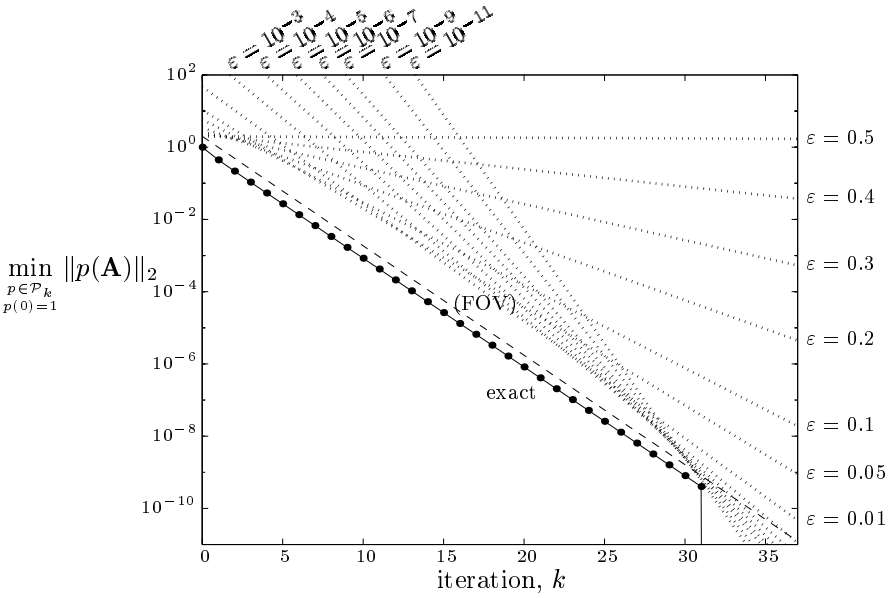

FIG. 3.3. *Convergence bounds for Example D with* $\delta = 1/2$ *and* $n = 32$.

Trefethen [30] shows that $\Lambda_\varepsilon(\mathbf{A}) = 1 + \Delta_{\delta+\varepsilon}$. In this limit, the bounds give:

$$\text{(FOV)}: \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \le 2\,\delta^k \qquad \text{(PSA)}: \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \le (1+\delta/\varepsilon)(\delta+\varepsilon)^k.$$

In the limit as $\delta \to 0$, both these bounds predict convergence to arbitrary desired tolerance in a single iteration, though (FOV) is asymptotically sharper for $n = \infty$. One must balance the size of the (PSA) constant against the accompanying convergence rate, just as for the normal matrix with $\Lambda(\mathbf{A}) = [a, b]$ discussed in Example A. Figures 3.3 and 3.4 illustrate this for $\delta = 1/2$ with $n = 32$ and $n = \infty$. For the "exact" curve, we plot the lower bound given by equation (3.1). The bound (PSA) is particularly interesting in the finite-dimensional case. In fact, for very small values of ε, (PSA) predicts convergence rates that are too quick, associated with large constants that ensure that bound doesn't intersect the convergence curve for $k < n$.

Taking $\mathbf{A}$ to be non-diagonalizable makes for a clean example, but it is not necessary. Moving the diagonal entries of $\mathbf{A}$ from λ to distinct nearby values, one obtains arbitrarily large values of $C_{\text{EV}} = \kappa(\mathbf{V})$ and corresponding convergence rates not necessarily descriptive of the observed rates.

• **Non-Example: Only (FOV) descriptive.** Theorem 3.2 indicates that examples where (FOV) dramatically outperforms (PSA) may be difficult to find. Since $\Lambda_\varepsilon(\mathbf{A}) \subset W(\mathbf{A}) + \Delta_\varepsilon$, the rate $\rho_{\text{PSA}}(\varepsilon)$ can only be significantly larger than ρ_{FOV} when this containment is sharp, ε is relatively large, and the sets $W(\mathbf{A})$ and $\Lambda_\varepsilon(\mathbf{A})$ are near the origin. In such cases, values of ε that yield $\rho_{\text{PSA}}(\varepsilon)$ similar to ρ_{FOV} will be associated with small ε and thus large constant terms $C_{\text{PSA}}(\varepsilon)$. But proximity to the origin implies that ρ_{FOV} will predict slow convergence, and the pseudospectral bounds, while less sharp, will still provide a decent indication of the number of iterations required to meet the convergence criteria. Taking the non-diagonalizable Example D with δ close to 1 and $n = \infty$ yields the most extreme behavior possible.

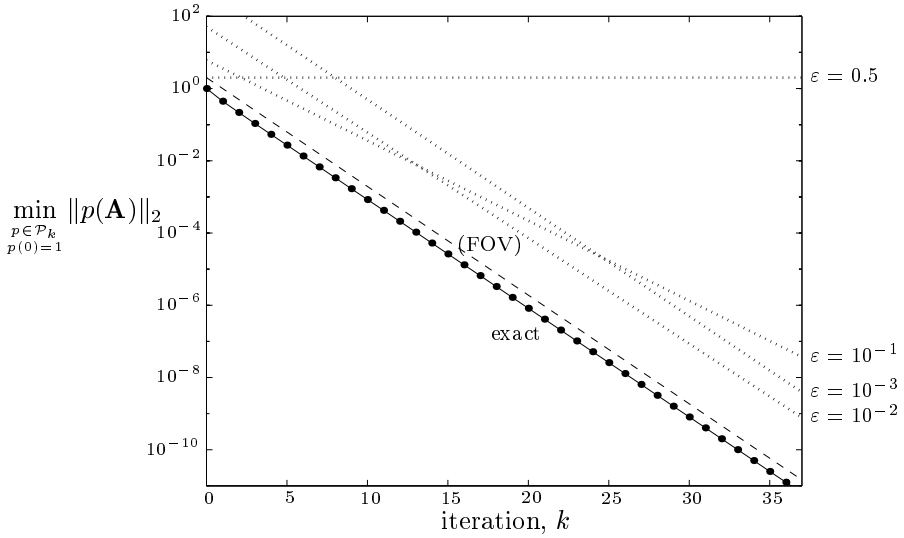

FIG. 3.4. *Convergence bounds for Example D with $\delta = 1/2$ and $n = \infty$.*

• **Example E: Only (FOV) not descriptive.** The field of values bound (FOV) is not descriptive when there is initial stagnation followed by more rapid convergence. The simplest example of GMRES stagnation at the first iteration occurs for the symmetric indefinite matrix

$$\mathbf{A} = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix},$$

as described by Saad and Schultz [33]. Since $\mathbf{A} \in \mathbb{C}^{2\times 2}$, the second iteration gives exact convergence.

Since $\mathbf{A}$ is normal, (EV) is exact, correctly predicting two iterations to solve the polynomial approximation problem on the discrete set of two eigenvalues. Since $W(\mathbf{A})$ is the convex hull of $\Lambda(\mathbf{A})$, the field of values contains the origin for this example, $W(\mathbf{A}) = [-1, 1]$, and thus (FOV) predicts no convergence. The ε-pseudospectrum consists of the union of two disks of radius ε, each centered at an eigenvalue. Thus, $C_{\text{PSA}}(\varepsilon) = 2$ independent of $0 < \varepsilon \leq 1$. Taking $\varepsilon \to 0$, (PSA) predicts arbitrarily good convergence in two iterations.

The bound (FOV) can also fail when $\mathbf{A}$ is a non-normal matrix with $\Lambda_\varepsilon(\mathbf{A})$ in the right half plane. The following example, with uniformly ill-conditioned eigenvectors, is the only case we discuss where $\kappa(\mathbf{V})$ is large and (EV) is descriptive. Let $\mathbf{\Lambda}$ be diagonal with eigenvalues uniformly distributed in the interval $[1, 2]$ and define

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}, \quad \text{where } \mathbf{V} = \begin{pmatrix} 1 & \sqrt{1-\delta} & \sqrt{1-\delta} & \cdots & \sqrt{1-\delta} \\ & \sqrt{\delta} & 0 & \cdots & 0 \\ & & \sqrt{\delta} & \ddots & \vdots \\ & & & \ddots & 0 \\ & & & & \sqrt{\delta} \end{pmatrix}, \tag{3.2}$$

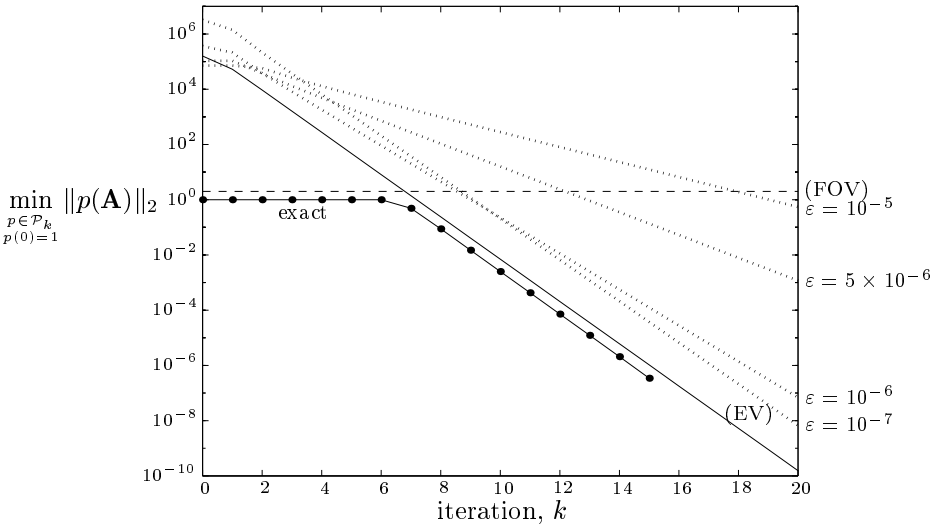

FIG. 3.5. *Convergence bounds for the matrix 3.2 with* $\delta = 10^{-8}$ *and* $n = 64$.

and $0 < \delta \ll 1$ is a small positive parameter.

Clearly, these eigenvectors are highly ill-conditioned; looking at the upper left 2×2 block of $\mathbf{V}$ alone shows that $\kappa(\mathbf{V}) \geq (1+\sqrt{\delta})/\sqrt{\delta}$. One can show $\kappa(\mathbf{V}) = \mathcal{O}(n/\sqrt{\delta})$ as $n \to \infty$ and $\delta \to 0$. If $0 < \delta < 1/9$, then $0 \in W(\mathbf{A})$ and (FOV) is vacuous; for particularly small δ, the field of values is extremely large. The pseudospectra are likewise large, but as ε is taken small enough such that $0 \notin \Lambda_\varepsilon(\mathbf{A})$, the bound becomes increasingly descriptive. This is another case where the inclusion in Theorem 3.2 is weak for small ε.

Figure 3.5 presents the bounds for $\delta = 10^{-8}$ and $n = 64$. The constant $C_{\text{EV}} = \kappa(\mathbf{V})$ is obtained by computing the condition number in MATLAB, with the first column of $\mathbf{V}$ multiplied by $\sqrt{n}$ to improve conditioning. The rate ρ_{EV} is taken from the optimal polynomial on the interval $[1, 2]$. The pseudospectral bound was determined for each ε by bounding $\Lambda_\varepsilon(\mathbf{A})$ by an exterior convex polygon, obtaining the convergence rate through numerical conformal mapping, and then applying the bound (2.5). The "exact" curve was computed using the semidefinite programming strategy of Toh and Trefethen [38], as implemented in the MATLAB SDPT3 toolbox [36]. For this example, random initial residuals do not exhibit the long plateau obtained in bound on worst case convergence computed from $\min_{\substack{p\in\mathcal{P}_k\\ p(0)=1}} \|p(\mathbf{A})\|_2$.

• **Example F: Only (PSA) descriptive.** For the Jordan block in Example D, the field of values bound (FOV) captured the single convergence rate perfectly. For the present example, we seek a non-diagonalizable matrix that initially stagnates but eventually converges more rapidly, making (FOV) misleading. The most extreme example of this behavior is the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & \alpha \\ & 1 \end{pmatrix}, \qquad \alpha \gg 1,$$

which featured as a submatrix in Example B. For $\alpha \geq 2$, there are right hand sides for which GMRES makes no progress at the first step. Yet at the second step, there is exact convergence since $\mathbf{A} \in \mathbb{C}^{2\times 2}$.

Since $\mathbf{A}$ is non-diagonalizable, $C_{\text{EV}} = \infty$ and (EV) does not apply. Since the field of values is a disk centered at 1 with radius $\alpha/2$, $W(\mathbf{A}) = 1 + \Delta_{\alpha/2}$ [21, §1.3], the field of values bound fails to predict convergence for $\alpha \geq 2$. The pseudospectral bound, however, captures the exact convergence in two iterations. As $\varepsilon \to 0$, $\Lambda_\varepsilon(\mathbf{A}) \to 1 + \Delta_{\alpha\sqrt{\varepsilon}}$. For such ε, $C_{\text{PSA}}(\varepsilon) \to \alpha/\sqrt{\varepsilon}$ and (PSA) predicts to leading order $\min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\|_2 \leq \alpha^{k+1}(\sqrt{\varepsilon})^{k-1}$. For any fixed finite α, we can make the right hand side of this bound arbitrarily small for $k \geq 2$, so we must have exact convergence in two iterations.

A more interesting example, without reliance on finite dimension, is obtained by varying the entries in the first upper diagonal of a Jordan block,

$$\mathbf{A} = \begin{pmatrix} 1 & \beta & & & \\ & 1 & \frac{\beta}{2} & & \\ & & 1 & \ddots & \\ & & & \ddots & \frac{\beta}{(n-1)} \\ & & & & 1 \end{pmatrix}, \tag{3.3}$$

for constant $\beta > 0$. Reichel and Trefethen called a closely related example the "integration matrix" [30].

Since $\mathbf{A}$ is non-diagonalizable, the bound (EV) can't be usefully applied. Both the field of values and the pseudospectra of $\mathbf{A}$ are circular disks [42]. If $n \geq 2$ and $\beta \geq 2$, then $0 \in W(\mathbf{A})$. For such values of β, (FOV) can only predict stagnation. Driscoll, Toh, and Trefethen discuss this particular example, arguing that pseudospectral bounds may be descriptive. As motivation, they note that the pseudospectra shrink rapidly as ε decreases. For small ε, $W(\mathbf{A})$ is much larger than $\Lambda_\varepsilon(\mathbf{A})$, and thus (PSA) predicts much faster convergence rates than (FOV).

Figure 3.6 shows the bounds for $\beta = 5/2$ and $n = 64$. The exact curve was computed using the technique of Toh and Trefethen [38]. The pseudospectral bounds were obtained by numerically determining the radius of each pseudospectral boundary. Increasing the dimension n doesn't significantly alter the pseudospectra for the values of ε shown here, as such an extension is a small perturbation of the direct sum of the original matrix with another integration matrix having smaller pseudospectra.

• **Non-Example: Only (PSA) not descriptive.** Success of (EV) and (FOV) together implies that $W(\mathbf{A})$ and $\Lambda(\mathbf{A})$ determine similar asymptotic convergence rates. There can be little initial stagnation if (FOV) is descriptive, and thus $\kappa(\mathbf{V})$ must be small if both bounds are to be descriptive. Thus, $\mathbf{A}$ must be normal (or nearly so). Theorems 3.1 and 3.2 then assure us that $\Lambda_\varepsilon(\mathbf{A})$ cannot be much larger than $\Lambda(\mathbf{A})$ and $W(\mathbf{A})$, so it is impossible to get an example where (PSA) performs dramatically worse than (EV) and (PSA). The closest one can come are normal matrices with $\Lambda(\mathbf{A}) = [a, b]$, where a is close enough to the origin to makes the pseudospectrum's rim of radius ε about the spectrum significant even for small ε, leading to an extreme version of Figure 3.1.

3.2. Summary of the Examples. Let us collect some of the points highlighted in these examples.

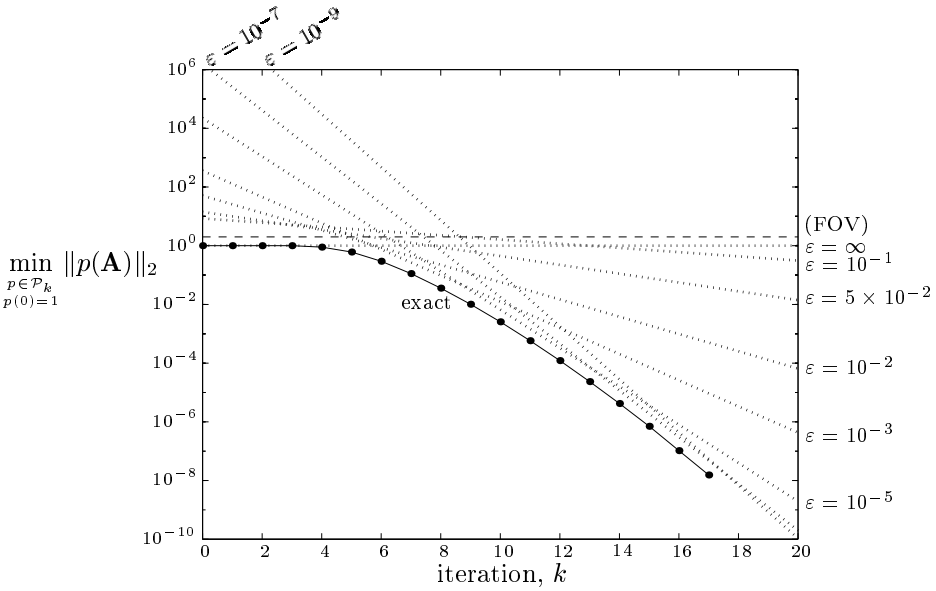

FIG. 3.6. *Convergence bounds for the integration matrix (3.3) with* $\alpha = 5/2$ *and* $n = 64$.

(EV) This bound worked perfectly for normal matrices and for (3.2), where all eigenvalues were uniformly non-normal, but failed when the matrix was non-diagonalizable or the non-normality was primarily associated with only part of the spectrum.

(FOV) The field of values performed well when only one convergence stage was observed, as in Examples A and D. Its primary advantages over (PSA) for these examples was sharpness (Figures 3.1, 3.3, and 3.4) and ease of computability. In the presence of transient stagnation, (FOV) failed, as in Examples B, C, E, and F.

(PSA) The pseudospectral bound inherits properties of both (EV) and (FOV), but can also capture interesting information between these extremes, as seen in Example F. The primary flaw in (PSA), exploited in Examples B and C, is its inability to recognize the finiteness of the spectrum, thus overestimating the influence of outliers eliminated at an early stage of convergence.

In the next section, we see how pseudospectra can yield convergence estimates during an iteration, and see how the bounds we have surveyed here apply to a matrix derived from a convection–diffusion problem.

4. Adaptive Pseudospectral Bounds. Often the pseudospectra bound provides a good indication of convergence behavior, especially when a range of values for ε are considered. The expense of full pseudospectral computation limits this bound's wide applicability, though. In this section, we see that information constructed within the GMRES iteration by the Arnoldi orthogonalization can give descriptive pseudospectral convergence estimates.

The first $k < n$ steps of the Arnoldi process [1] build an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ for the Krylov subspace $\mathcal{K}_{k+1}(\mathbf{A}, \mathbf{r}_0) \equiv \operatorname{span}\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^k\mathbf{r}_0\}$. This construction gives a partial upper Hessenberg decomposition of $\mathbf{A}$, such that if $\mathbf{V}_k =$

$(\mathbf{v}_1, \ldots, \mathbf{v}_k) \in \mathbb{C}^{n\times k}$, then

$$\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\overline{\mathbf{H}}_k \qquad \text{and} \qquad \mathbf{V}_k^*\mathbf{A}\mathbf{V}_k = \mathbf{H}_k, \tag{4.1}$$

where $\overline{\mathbf{H}}_k \in \mathbb{C}^{k+1\times k}$ is upper Hessenberg and $\mathbf{H}_k \in \mathbb{C}^{k\times k}$ consists of the first k rows of $\overline{\mathbf{H}}_k$ (see, e.g., [32, §6.3]). We take the subdiagonal entries of $\overline{\mathbf{H}}_k$ to be non-negative.

Toh and Trefethen [37] show that the pseudospectra of $\mathbf{H}_k$ and $\overline{\mathbf{H}}_k$ can well-approximate those of $\Lambda_\varepsilon(\mathbf{A})$ even when $k \ll n$. They define the pseudospectra of a rectangular matrix using the pseudoinverse, implying that $z \in \Lambda_\varepsilon(\overline{\mathbf{H}}_k)$ provided $\sigma_k(z\overline{\mathbf{I}} - \overline{\mathbf{H}}_k) \le \varepsilon$, where $\overline{\mathbf{I}}$ is the $k \times k$ identity matrix augmented by a row of zeros and σ_k is the kth largest singular value. With this definition, the relationship between $\Lambda_\varepsilon(\mathbf{H}_k)$, $\Lambda_\varepsilon(\overline{\mathbf{H}}_k)$, and $\Lambda_\varepsilon(\mathbf{A})$ can be quantified.

THEOREM 4.1. *Suppose* $\mathbf{V}_k^*\mathbf{A}\mathbf{V}_k = \mathbf{H}_k$ *and* $\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\overline{\mathbf{H}}_k$. *Then*

(1) $\Lambda_\varepsilon(\overline{\mathbf{H}}_1) \subseteq \Lambda_\varepsilon(\overline{\mathbf{H}}_2) \subseteq \cdots \subseteq \Lambda_\varepsilon(\overline{\mathbf{H}}_{n-1}) \subseteq \Lambda_\varepsilon(\mathbf{H}_n) = \Lambda_\varepsilon(\mathbf{A})$.

(2) $\Lambda_\varepsilon(\mathbf{H}_k) \subseteq \Lambda_{\widehat{\varepsilon}}(\overline{\mathbf{H}}_k) \subseteq \Lambda_{\widehat{\varepsilon}}(\mathbf{A})$, *where* $\widehat{\varepsilon} \equiv \varepsilon + h_{k+1,k}$.

Proof. Part (1) was proved by Toh and Trefethen; it follows immediately from noting that $\sigma_k(z\overline{\mathbf{I}} - \overline{\mathbf{H}}_k) \ge \sigma_{k+1}(z\overline{\mathbf{I}} - \overline{\mathbf{H}}_{k+1})$. For part (2), suppose that $z \in \Lambda_\varepsilon(\mathbf{H}_k)$. Observe that $\sigma_k(z\overline{\mathbf{I}} - \overline{\mathbf{H}}_k) \le \sigma_k(z\mathbf{I} - \mathbf{H}_k) + h_{k+1,k} \le \varepsilon + h_{k+1,k}$, where the first inequality follows Horn and Johnson's Theorem 3.3.16 [24]. Applying (1) to this bound completes the proof. ∎

Since the pseudospectra of $\mathbf{H}_k$ and $\overline{\mathbf{H}}_k$ approximate those of $\mathbf{A}$, it is natural to consider the bound (PSA) with $\mathbf{A}$ replaced by $\mathbf{H}_k$ or $\overline{\mathbf{H}}_k$. The resultant expression is no longer a convergence bound, but only an estimate. We have

$$\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \lesssim \frac{\mathcal{L}(\Gamma_\varepsilon)}{2\pi\varepsilon} \min_{\substack{p\in\mathcal{P}_k \\ p(0)=1}} \max_{z\in\Lambda_\varepsilon(\mathbf{A})} |p(z)|,$$

where Γ_ε is a Jordan curve enclosing $\Lambda_\varepsilon(\mathbf{H}_k)$ or $\Lambda_\varepsilon(\overline{\mathbf{H}}_k)$.

What value of ε is relevant at a specific iteration? The following bounds, while not necessarily sharp, indicate where GMRES is focusing its energy.

PROPOSITION 4.2. *For* $k < n$, *the eigenvalues of* $\mathbf{H}_k$ *(Ritz values) are contained in the* ε*-pseudospectrum of* $\mathbf{A}$ *for* $\varepsilon = h_{k+1,k}$. *If* $\mathbf{H}_k$ *is nonsingular, the roots of the GMRES residual polynomial (harmonic Ritz values [15], [17]) are contained in the* ε*-pseudospectrum of* $\mathbf{A}$ *for* $\varepsilon = h_{k+1,k} + h_{k+1,k}^2/\sigma_k(\mathbf{H}_k)$.

Proof. We prove the first part by constructing the specific perturbation matrix $\mathbf{E} = -h_{k+1,k}\mathbf{v}_{k+1}\mathbf{e}_k^*\mathbf{V}_k^*$. Then, from (4.1),

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

where $\mathbf{e}_k \in \mathbb{C}^k$ is the kth column of the $k \times k$ identity matrix. This implies that $\mathbf{V}_k$ is an exact invariant subspace of $\mathbf{A}+\mathbf{E}$, and thus $\Lambda(\mathbf{H}_k) \subseteq \Lambda(\mathbf{A}+\mathbf{E}) \subset \Lambda_\varepsilon(\mathbf{A})$, where $\varepsilon \equiv h_{k+1,k} \ge \|\mathbf{E}\|_2$.

The second part follows from the analysis of Goossens and Roose [17]. They characterize the harmonic Ritz values as the eigenvalues of $(\mathbf{H}_k + h_{k+1,k}^2\mathbf{f}_k\mathbf{e}_k^*)$, where $\mathbf{f}_k = \mathbf{H}_k^{-*}\mathbf{e}_k$. Defining the perturbation $\mathbf{E} = (\mathbf{V}_k h_{k+1,k}^2\mathbf{f}_k\mathbf{e}_k^* - h_{k+1,k}\mathbf{v}_{k+1}\mathbf{e}_k^*)\mathbf{V}_k^*$, it follows from (4.1) that $(\mathbf{A}+\mathbf{E})\mathbf{V}_k = \mathbf{V}_k(\mathbf{H}_k + h_{k+1,k}^2\mathbf{f}_k\mathbf{e}_k^*)$, and thus $\Lambda(\mathbf{H}_k + h_{k+1,k}^2\mathbf{f}_k\mathbf{e}_k^*) \subseteq \Lambda(\mathbf{A}+\mathbf{E}) \subset \Lambda_\varepsilon(\mathbf{A})$, where $\|\mathbf{E}\|_2 \le \varepsilon \equiv h_{k+1,k} + h_{k+1,k}^2/\sigma_{\min}(\mathbf{H}_k)$. ∎

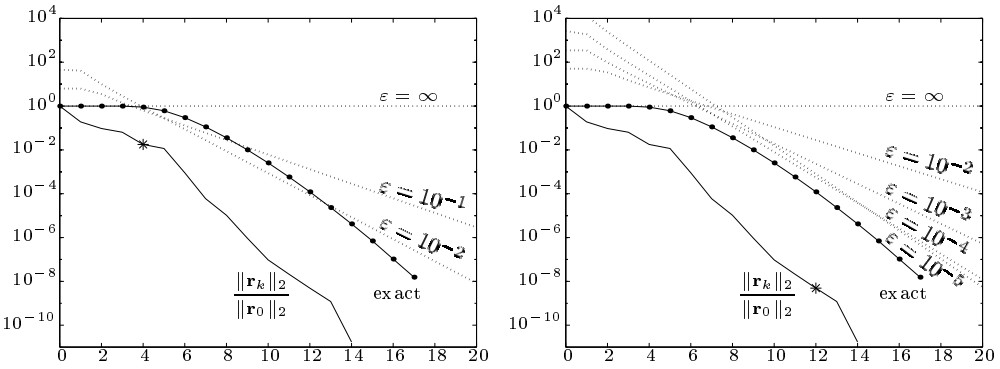

FIG. 4.1. *Adaptive convergence bounds for the integration matrix (3.3) at iteration $k = 4$ on the left and $k = 12$ on the right.*

GMRES convergence estimates derived from $\Lambda(\mathbf{H}_k)$ have several applications. If k is taken to be small, the estimate may give an indication of GMRES behavior at future iterations. If k is larger (e.g., the k that satisfies the GMRES convergence criterion), one may obtain a good description of worst case convergence over all initial residuals. This estimate depends on the particular initial residual $\mathbf{r}_0$ that determines the entries of $\mathbf{H}_k$. If $\mathbf{r}_0$ is deficient in all eigenvector directions associated with a particular eigenvalue, that eigenvalue cannot influence $\mathbf{H}_k$ nor the estimates derived from it. If $\mathbf{r}_0$ only has a small component in a certain eigenvector direction, that component may not exert much influence early iterations (and $\mathbf{H}_k$ for small k), but become significant at later iterations.

Toh and Trefethen observe qualitative links between the pseudospectra and the GMRES iteration polynomial [34], [38]. A deeper quantitative understanding of this relationship would be an important step toward appreciating the ability of pseudospectral bounds to describe GMRES convergence, the value of ε for which $\Lambda_\varepsilon(\mathbf{A})$ is relevant for each iteration, and the merits of adaptive strategies, like the one described here, to capture significant features of convergence behavior.

Figure 4.1 illustrates adaptive estimates drawn from the integration matrix with $n = 64$ and $\beta = 5/2$. In Figure 3.6, we saw that the bound (PSA) was ideal for this example, and thus hope these Arnoldi estimates would perform similarly well. The estimates shown were here are based on a random initial residual with entries drawn from the standard normal distribution. Taking estimates at iteration $k = 4$ gives some hint of the quick convergence that follows; when $k = 12$, the pseudospectra of $\mathbf{H}_k$ match those of $\mathbf{A}$ for the relevant values of ε and characterize the worst case convergence curve. The curve labeled $\|\mathbf{r}_k\|_2/\|\mathbf{r}_0\|_2$ is the convergence obtained for this particular initial residual, with an asterisk marking the iteration that yields the convergence estimate.

4.1. A Practical Example. We illustrate the use of this technique, and the bounds (EV), (EV′), (PSA), and (FOV), for a model problem from fluid dynamics. Let $\mathbf{A}$ be the matrix generated by a streamline upwinded Petrov–Galerkin finite element discretization of the two-dimensional convection–diffusion equation,

$$-\nu\Delta u + \mathbf{w}\cdot\nabla u = f \quad \text{on} \quad \Omega = [0,1]\times[0,1],$$

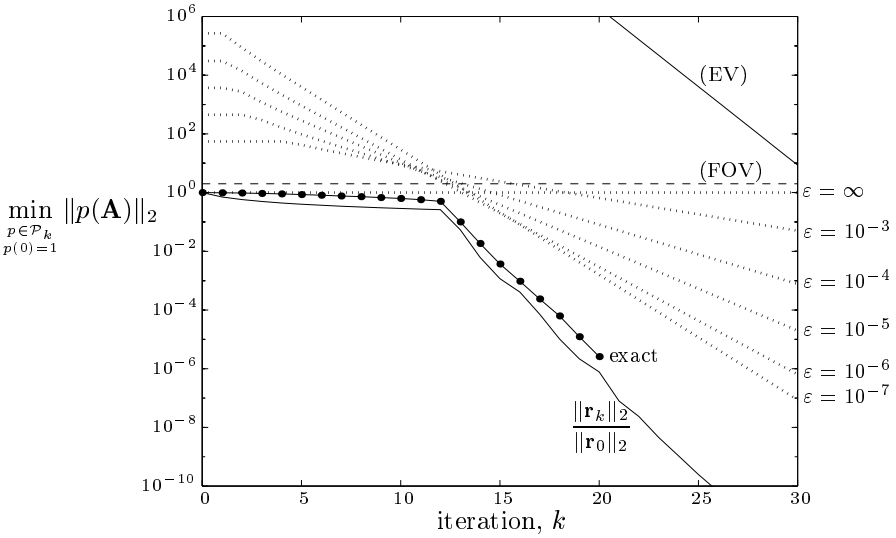

FIG. 4.2. *Convergence bounds for the convection–diffusion problem with* $N = 13$.

with diffusion $\nu = 0.01$, constant vertical wind $\mathbf{w} = (0, 1)$, and Dirichlet boundary conditions that induce an interior layer and a boundary layer. The solution is approximated with bilinear finite elements on a regular square grid with N unknowns in each coordinate direction, yielding a matrix $\mathbf{A}$ of dimension $n = N^2$. This problem is discussed by Fischer et al. [14]; we apply the upwinding parameter they suggest. This parameter value leads to good approximate solutions to the partial differential equation, but the corresponding matrix is highly non-normal. Its eigenvalues, though very sensitive to perturbations, are known analytically for this special wind direction [14]. These eigenvalues fall on N lines in the complex plane with constant real part, with N eigenvalues per line. The eigenvalues with largest real part are the most ill-conditioned. The coefficient matrix is non-diagonalizable in the limit $N \to \infty$, and the eigenvector matrix condition number is very large even for modest values of N.

The high non-normality of $\mathbf{A}$ leads early iterations of GMRES to stagnate for typical initial residuals. Ernst investigated the field of values bound (FOV) for this matrix [12]. Since $\mathbf{A}$ results from a coercive finite element discretization, $W(\mathbf{A})$ is contained in the open right half plane and (FOV) applies non-trivially. As expected from the initial stagnation, this bound is descriptive at early iterations but asymptotically inaccurate. As in Examples E and F, (PSA) does better. Though it is somewhat pessimistic asymptotically, it accurately captures the end of the initial stagnation, a feature that eludes the other two bounds.

Figure 4.2 illustrates (EV), (FOV), and (PSA) for $N = 13$. The eigenvector matrix $\mathbf{V}$ is known analytically [14]. When each column of $\mathbf{V}$ has unit 2-norm, we compute $\kappa(\mathbf{V}) \approx 4.6 \times 10^{16}$. We bound ρ_{EV} by the convergence rate associated with the convex hull of the eigenvalues of $\mathbf{A}$; this agrees with the observed asymptotic convergence rate, though C_{EV} is much too large. The bound (FOV) gives $\rho_{\text{FOV}} \approx 0.968$. The pseudospectral bound leads to better convergence rates; we compute these rates by calculating the convergence rate of an approximate convex hull of

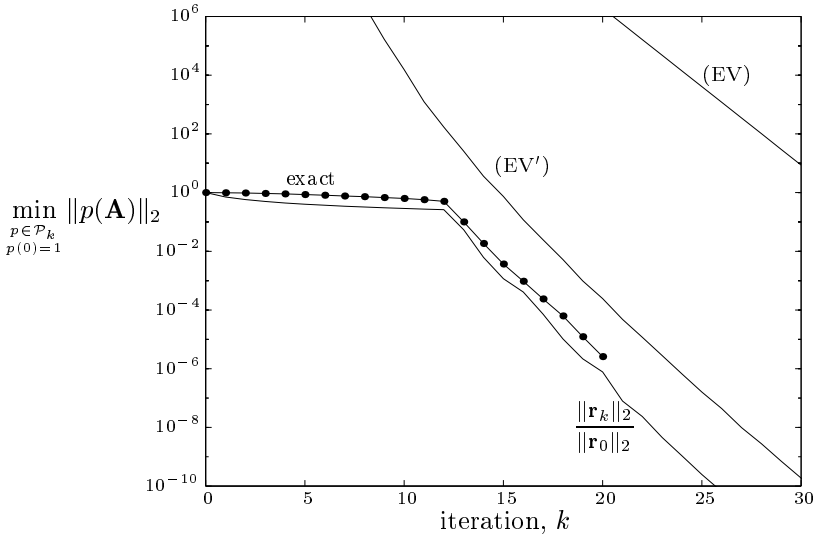

FIG. 4.3. *Convergence bound (EV′) for the convection–diffusion problem with* $N = 13$.

$\Lambda_\varepsilon(\mathbf{A})$ and applying the convergence bound for convex sets (2.5). The exact curve was computed using the techniques of Toh and Trefethen [38], and we compare it to GMRES convergence for an initial residual derived from boundary conditions that induce an internal layer and a boundary layer.

How does the bound (EV′) perform in this situation? Though there aren't any normal eigenvalues, the degree of eigenvalue ill-conditioning increases markedly from those with smallest real part to those with largest real part. Figure 4.3 illustrates that (EV′) handles this situation admirably; asymptotically, (EV′) is more accurate than the three standard bounds. For this example, the eigenvalue condition numbers were computed from analytic formulas for left and right eigenvectors, and (EV′) was calculated in quadruple precision arithmetic.

We test the pseudospectral estimates described earlier in this section for $N = 13$ with the same initial residual described above. The adaptive estimates taken during the initial stagnation give little hint of future convergence behavior; the approximate pseudospectra do not improve much from iteration to iteration during these iterations. For $k = 13$, at the onset of more rapid convergence, $\Lambda_\varepsilon(\mathbf{H}_k) \approx \Lambda_\varepsilon(\mathbf{A})$ for $\varepsilon \geq 10^{-4}$ and one gets an indication of the improved convergence to come. By the time the convergence criterion is satisfied at $k = 26$, $\Lambda_\varepsilon(\mathbf{H}_k) \approx \Lambda_\varepsilon(\mathbf{A})$ for those values of ε relevant to the bound (PSA). Figure 4.4 shows these convergence estimates for $k = 13$ and $k = 26$.

5. Summary. We have seen the relative merits of convergence bounds based on eigenvalues (with the eigenvector condition number), the field of values, and pseudospectra. In particular, these bounds have distinct weaknesses that indicate situations in which one bound may be preferred over the others. The standard bounds are global statements that can be refined; for (EV′) and (FOV′) this localization introduces spectral projector norms. Pseudospectra provide a convenient tool for transferring between the eigenvalues and the field of values, but they can be expen-

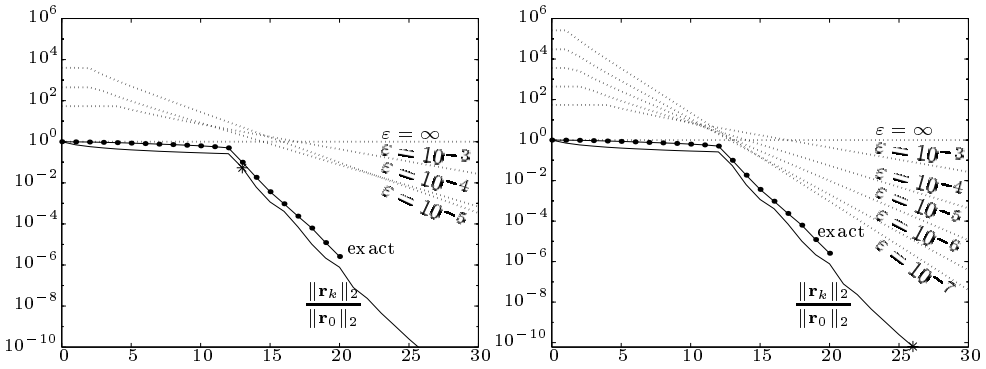

FIG. 4.4. *Adaptive convergence bounds for the convection–diffusion matrix for* $k = 13$ *on the left and* $k = 26$ *on the right.*

sive to compute. Approximate pseudospectra drawn from the Arnoldi process yield convergence estimates at a fraction of the cost of full pseudospectral computation.

Acknowledgements. I thank Andy Wathen for guiding this research and suggesting numerous improvements to this presentation. I am also grateful for Nick Trefethen's many helpful comments. The title and the format of Section 3 were inspired by a paper by Nachtigal, Reddy, and Trefethen [29]. I thank Anne Greenbaum for providing a copy of reference [8], and Chris Beattie, Bernd Fischer, and Henk van der Vorst for stimulating discussions related to this research.

REFERENCES

[1] W. E. ARNOLDI, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.
[2] F. L. BAUER AND C. T. FIKE, *Norms and exclusion theorems*, Numer. Math., (1960), pp. 137–141.
[3] T. BRACONNIER AND N. J. HIGHAM, *Computing the field of values and pseudospectra using the Lanczos method with continuation*, BIT, 36 (1996), pp. 422–440.
[4] T. A. DRISCOLL, *A* MATLAB *toolbox for Schwarz–Christoffel mapping*, ACM Trans. Math. Software, 22 (1996), pp. 168–186. For accompanying software, see `http://amath.colorado.edu/appm/faculty/tad/research/sc.html`.
[5] T. A. DRISCOLL, K.-C. TOH, AND L. N. TREFETHEN, *From potential theory to matrix iterations in six steps*, SIAM Review, 40 (1998), pp. 547–578.
[6] M. EIERMANN, *On semiiterative methods generated by Faber polynomials*, Numer. Math., 56 (1989), pp. 139–156.
[7] ———, *Fields of values and iterative methods*, Lin. Alg. Applics., 180 (1993), pp. 167–197.
[8] ———, *Field of values and iterative methods.* Slides from the Oberwolfach Conference on Iterative Methods, Oberwolfach, Germany, April 1997.
[9] M. EIERMANN AND O. ERNST, *Geometric aspects in the theory of Krylov subspace methods.* submitted, December 1998.
[10] S. C. EISENSTAT, H. C. ELMAN, AND M. H. SCHULTZ, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 20 (1983), pp. 345–357.
[11] M. EMBREE AND L. N. TREFETHEN, *Green's functions for multiply connected domains via Schwarz–Christoffel mapping.* To appear in *SIAM Review.*
[12] O. ERNST, *Residual-minimizing Krylov subspace methods for stabilized discretizations of convection-diffusion equations.* To appear in *SIAM J. Matrix Anal. Appl.*
[13] B. FISCHER, *Polynomial Based Iteration Methods for Symmetric Linear Systems*, Wiley–

Teubner, Chichester, 1996. For accompanying software, see
`http://www.math.mu-luebeck.de/workers/fischer/fischer-home.html`.

[14] B. Fischer, A. Ramage, D. Silvester, and A. J. Wathen, *Towards parameter–free streamline upwinding for advection–diffusion problems.* To appear in *Comp. Meths. Appl. Eng.* For related software, see `http://www.ma.umist.ac.uk/djs/software.html`.

[15] R. W. Freund, *Quasi-kernel polynomials and their use in non-Hermitian matrix iterations*, J. Comp. Appl. Math., 43 (1992), pp. 135–158.

[16] R. W. Freund and N. M. Nachtigal, *QMR: A quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–339.

[17] S. Goossens and D. Roose, *Ritz and harmonic Ritz values and the convergence of FOM and GMRES.* To appear in *Numer. Lin. Alg. Applics.*, 1999.

[18] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.

[19] 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.

[20] 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.

[21] K. E. Gustafson and D. K. M. Rao, *Numerical Range: The Field of Values of Linear Operators and Matrices*, Springer–Verlag, New York, 1997.

[22] D. J. Higham and L. N. Trefethen, *Stiffness of ODEs*, BIT, 33 (1993), pp. 285–303.

[23] E. Hille, *Analytic Function Theory*, vol. 2, Chelsea, New York, 1962.

[24] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[25] I. C. F. Ipsen, *Expressions and bounds for the GMRES residual.* Manuscript, 1999.

[26] T. Kato, *Some mapping theorems for the numerical range*, Proc. Japan Acad., 41 (1965), pp. 652–655.

[27] ———, *Perturbation Theory for Linear Operators*, Springer, New York, 2nd ed., 1980.

[28] T. Kövari and C. Pommerenke, *On Faber polynomials and Faber expansions*, Math. Zeit., 99 (1967), pp. 193–206.

[29] N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–795.

[30] L. Reichel and L. N. Trefethen, *Eigenvalues and pseudo-eigenvalues of Toeplitz matrices*, Lin. Alg. Applics., 162–164 (1992), pp. 153–185.

[31] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester Univerisity Press, Manchester, 1992.

[32] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS, Boston, 1996.

[33] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.

[34] K.-C. Toh, *Matrix Approximation Problems and Nonsymmetric Iterative Methods*, PhD thesis, Cornell University, August 1996.

[35] ———, *GMRES vs. ideal GMRES*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 30–36.

[36] K. C. Toh, M. J. Todd, and R. H. Tütüncü, *SDPT3—a* Matlab *software package for semidefinite programming.* To appear in *Optimization Methods and Software.* For accompanying software, see `http://www.math.nus.sg/~mattohkc/`.

[37] K.-C. Toh and L. N. Trefethen, *Calculation of pseudospectra by the Arnoldi iteration*, SIAM J. Sci. Comput., 17 (1996), pp. 1–15.

[38] ———, *The Chebyshev polynomials of a matrix*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 400–419.

[39] L. N. Trefethen, *Approximation theory and numerical linear algebra*, in Algorithms for Approximation II, J. C. Mason and M. G. Cox, eds., London, 1990, Chapman and Hall.

[40] ———, *Pseudospectra of matrices*, in Numerical Analysis 1991, D. F. Griffiths and G. A. Watson, eds., Harlow, Essex, UK, 1992, Longman Scientific and Technical, pp. 234–266.

[41] ———, *Computation of pseudospectra*, Acta Numerica, (1999), pp. 247–295.

[42] ———, *Spectra and Pseudospectra: The Behavior of Non-normal Matrices and Operators.* Book in preparation, 1999.

[43] A. van der Sluis, *Condition numbers and equilibration of matrices*, Numer. Math., 14 (1969), pp. 14–23.

[44] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford, 1965.