SIAM J. Matrix Anal. Appl.
Vol. 36, No. 1, pp. 289–301

ROOTS OF MATRICES IN THE STUDY OF GMRES CONVERGENCE AND CROUZEIX'S CONJECTURE*

DAESHIK CHOI† AND ANNE GREENBAUM‡

Abstract. For any nonsingular matrix A and any positive integer m, the mth root, $A^{1/m}$, can be defined using any branch cut of the mth root function that does not pass through an eigenvalue of A. Cleary, $A^{1/m}$ approaches the identity as $m \to \infty$, but we are interested in how it approaches the identity. It is shown that $\lim_{m\to\infty}[W(A^{1/m})]^m = \exp[W(\log A)]$, where W denotes the field of values and $\log A$ is defined using the same branch cut. It is also shown that $\|A^{1/m}\|^m$ approaches $\exp(\alpha(\log A))$, where $\|\cdot\|$ denotes the spectral norm and $\alpha(\cdot)$ is the numerical abscissa. Implications for the convergence rate of the GMRES algorithm are discussed, especially when $W(A)$ contains the origin. Additionally, it is shown that if A is a strict contraction and $\rho \in (\|A\|, 1)$, then the matrices $[(I + \rho A)^{-1}(A + \rho I)]^{1/m}$, $m = 1, 2, \ldots$, are all strict contractions. The significance of results of this sort in a possible approach to proving Crouzeix's conjecture is discussed.

Key words. roots of a matrix, field of values, GMRES

AMS subject classifications. 15A60, 65F10

DOI. 10.1137/140961742

1. Definitions and main results. Let A be an n by n nonsingular matrix. For any positive integer m (and, more generally, for any real number t), the matrix $A^{1/m}$ (A^t) can be defined using the Cauchy integral formula

$$A^{1/m} = \frac{1}{2\pi i}\int_\Gamma (zI - A)^{-1} z^{1/m}\, dz,$$

where Γ is any simple closed curve or union of simple closed curves enclosing a region about each eigenvalue of A and not crossing the chosen branch cut of the function $z^{1/m}$. An equivalent definition is

$$A^{1/m} = SJ^{1/m}S^{-1},$$

where $A = SJS^{-1}$ is a Jordan decomposition of A, and if J_k is an n_k by n_k Jordan block corresponding to eigenvalue λ_k and $f(z) = z^{1/m}$, then

$$J_k^{1/m} = \begin{pmatrix} f(\lambda_k) & f'(\lambda_k) & \dots & \frac{1}{(n_k-1)!}f^{(n_k-1)}(\lambda_k) \\ & \ddots & \ddots & \vdots \\ & & \ddots & f'(\lambda_k) \\ & & & f(\lambda_k) \end{pmatrix}.$$

Since $J^{1/m}$ approaches the identity as $m \to \infty$, it is clear that $\lim_{m\to\infty} A^{1/m} = I$. We are interested in how $A^{1/m}$ approaches the identity as $m \to \infty$.

*Received by the editors March 21, 2014; accepted for publication (in revised form) by J. Liesen December 18, 2014; published electronically March 19, 2015. The research of the authors was supported in part by NSF grant DMS-1210886.
http://www.siam.org/journals/simax/36-1/96174.html

†Department of Mathematics and Statistics, Southern Illinois University, Edwardsville, IL (dchoi@sine.edu).
‡Applied Mathematics Department, University of Washington, Seattle, WA (greenbau@uw.edu).

The *field of values* or *numerical range* of A is defined as

$$W(A) = \{q^* A q : q \in \mathbb{C}^n, \quad \|q\| = 1\},$$

where $\|\cdot\|$ denotes the 2-norm for vectors and will be used to denote the corresponding operator norm for matrices. The following theorem shows what happens to $[W(A^{1/m})]^m$ as $m \to \infty$.

THEOREM 1.1. *For any n by n nonsingular matrix A and any n-vector q with $\|q\| = 1$,*

$$\lim_{m\to\infty} (q^* A^{1/m} q)^m = e^{q^*(\log A)q}, \tag{1.1}$$

where $A^{1/m}$ and $\log A$ are defined using any branch cut of the functions $z^{1/m}$ and $\log z$ that does not pass through any eigenvalues of A. Hence,

$$\lim_{m\to\infty} [W(A^{1/m})]^m = \exp[W(\log A)]. \tag{1.2}$$

Proof. Taking logarithms on each side, the claim (1.1) is equivalent to

$$\lim_{m\to\infty} m \log(q^* A^{1/m} q) = q^*(\log A)q. \tag{1.3}$$

Write the left-hand side of (1.3) as

$$\lim_{m\to\infty} \frac{\log(q^* A^{1/m} q)}{1/m} = \lim_{t\to 0^+} \frac{\log(q^* A^t q)}{t} = \lim_{t\to 0^+} \frac{\log(q^* e^{t\log A} q)}{t}$$

and apply L'Hopital's rule to see that this is equal to

$$\frac{d}{dt}\left[\log(q^* e^{t\log A} q)\right]\Big|_{t=0} = \frac{q^*(\log A)e^{t\log A}q}{q^* e^{t\log A} q}\Big|_{t=0} = q^*(\log A)q. \qquad \square$$

The following theorem shows what happens to $\|A^{1/m}\|^m$ as $m \to \infty$.

THEOREM 1.2. *For any n by n nonsingular matrix A,*

$$\lim_{m\to\infty} \|A^{1/m}\|^m = \exp(\alpha(\log A)), \tag{1.4}$$

where $\|\cdot\|$ is the operator 2-norm and $\alpha(\cdot)$ denotes the numerical abscissa*:*

$$\alpha(\cdot) = \max_{z\in W(\cdot)} Re(z). \tag{1.5}$$

Proof. This is an immediate consequence of the fact that for any n by n matrix B,

$$\frac{d}{dt}\left(\|e^{tB}\|\right)\Big|_{t=0} = \alpha(B), \tag{1.6}$$

which is based on the Hille–Yosida theorem. See, for example, [26]. From (1.6) it follows that

$$\begin{aligned}\lim_{t\to 0^+} \|e^{tB}\|^{1/t} &= \lim_{t\to 0^+}\left[\exp\left(\frac{1}{t}\log(\|e^{tB}\|)\right)\right] \\ &= \exp\left[\lim_{t\to 0^+}\frac{\log(\|e^{tB}\|)}{t}\right] = \exp\left[\frac{d}{dt}(\log(\|e^{tB}\|))\Big|_{t=0}\right] = \exp(\alpha(B)).\end{aligned}$$

Replace B by $\log A$ and t by $1/m$ to obtain the desired result (1.4). $\square$

Another question of interest has to do with functions of the form

$$(I - \bar{\gamma} A)^{-1}(A - \gamma I), \quad |\gamma| < 1, \tag{1.7}$$

which are degree one *Blaschke products* [10], or Möbius transformations. If A is a *contraction*, that is, $\|A\| \le 1$, then von Neumann's inequality [27, 24] states that for any function f analytic in the open unit disk $\mathcal{D}$,

$$\|f(A)\| \le \sup_{z \in \mathcal{D}} |f(z)|. \tag{1.8}$$

Since a Blaschke product maps the unit disk onto itself, it follows that if A is a contraction, then for any $|\gamma| < 1$, the matrix in (1.7) is as well. It was further shown in [11] that A is a *strict* contraction, that is, $\|A\| < 1$, if and only if the matrix in (1.7) is also a strict contraction. In general, the square root or mth root of the matrix in (1.7) is *not* a contraction (e.g., if $\gamma = 0$ and $A^{1/m}$ is any matrix such that the norm of its mth power is less than or equal to 1 but the mth power of its norm is greater than 1), but as $|\gamma| \to 1$, this matrix approaches a unit multiple of the identity. We will show that for $|\gamma| < 1$ and sufficiently close to 1, if A is a strict contraction, then so are all *mth* roots of the matrix in (1.7), where we use the branch cut along the ray from 0 to γ.

To avoid using different branches of the mth root function, if $\gamma = \rho e^{i\theta}$, we can rotate through the angle $\pi - \theta$ and then take γ to be negative real:

$$(I - \rho e^{-i\theta} A)^{-1}(A - \rho e^{i\theta} I) = e^{i(\theta - \pi)}(I + \rho e^{i(\pi - \theta)} A)^{-1}(e^{i(\pi-\theta)} A + \rho I).$$

Thus A is replaced by $e^{i(\pi-\theta)}A$ and γ by $-\rho$.

THEOREM 1.3.[1] *Let C be a strict contraction ($\|C\| < 1$) and define*

$$f(C) = (I + \rho C)^{-1}(C + \rho I), \quad \rho \in (\|C\|, 1).$$

Then $f(C)$ is a strict contraction and its principal mth roots, $f(C)^{1/m}$, $m = 2, 3, \ldots$ are well-defined and are strict contractions as well. In fact,

$$\|f(C)^{1/m}\| \le \left(\frac{\|C\| + \rho}{1 + \rho \|C\|} \right)^{1/m} < 1. \tag{1.9}$$

Proof. First, note that real eigenvalues of $f(C)$ correspond to real eigenvalues of C and since real eigenvalues of C lie in $[-\|C\|, \|C\|]$ and $\|C\| < \rho < 1$, all real eigenvalues of $f(C)$ are positive. Thus, $f(C)$ has no eigenvalues in $(-\infty, 0]$ so the principal mth roots of $f(C)$ are well-defined. To establish the bound (1.9), we will use von Neumann's inequality (1.8), which implies that for any function p analytic on $\mathcal{D}_{\|C\|}$, the disk about the origin of radius $\|C\|$, we have $\|p(C)\| \le \|p\|_{\mathcal{D}_{\|C\|}}$. This is because if $q(z) = p(z\|C\|)$, then $\|p(C)\| = \|q(C/\|C\|)\| \le \|q\|_{\mathcal{D}} = \|p\|_{\mathcal{D}_{\|C\|}}$. Define

$$f(z) = \frac{z + \rho}{1 + \rho z}.$$

It can be checked that f maps $\mathcal{D}_{\|C\|}$ to a disk in the open right half-plane; this is because functions of this form map disks inside the unit disk to other disks inside the

[1] The authors thank an anonymous referee for an improvement to the statement and proof of this theorem, which originally was shown to hold for ρ "sufficiently close" to 1.

unit disk, f maps disks about the origin to disks that are symmetric about the real axis since $f(\bar{z}) = \overline{f(z)}$, and the image under f of the circle of radius $\|C\|$ crosses the real axis at

$$r_{\max} = f(\|C\|) = \frac{\|C\| + \rho}{1 + \rho\|C\|} > 0 \quad \text{and} \quad r_{\min} = f(-\|C\|) = \frac{-\|C\| + \rho}{1 - \rho\|C\|} > 0.$$

Thus $f(\mathcal{D}_{\|C\|})$ is the disk centered at $(r_{\min} + r_{\max})/2$ of radius $(r_{\max} - r_{\min})/2$, and the maximum modulus of f on $\mathcal{D}_{\|C\|}$ is $r_{\max}$. Since f is analytic on $\mathcal{D}_{\|C\|}$, it follows from von Neumann's inequality that (1.9) holds for $m = 1$, and since $h(z) = z^{1/m}$ is analytic on $f(\mathcal{D}_{\|C\|})$ (that is, $h \circ f$ is analytic on $\mathcal{D}_{\|C\|}$), it also follows from von Neumann's inequality that (1.9) holds for $m = 2, 3, \ldots$. ∎

While Theorem 1.3 does not have direct implications for the GMRES algorithm, which will be discussed in the next section, it may be of interest in relation to Crouzeix's conjecture [4], which will be used to bound the residual norm in GMRES.

2. Implications for the GMRES algorithm.

Let A be an n by n nonsingular matrix and b a given n-vector. The GMRES algorithm for solving the linear system $Ax = b$ starts with an initial guess $x^{(0)}$ for the solution and, at each step j, constructs the approximation $x^{(j)}$ for which the residual $r^{(j)} \equiv b - Ax^{(j)}$ satisfies

$$\|r^{(j)}\| = \min_{p_j \in \mathcal{P}_j(0)} \|p_j(A) r^{(0)}\|,$$

where $\mathcal{P}_j(0)$ denotes the set of polynomials of degree j or less with value 1 at the origin and $\|\cdot\|$ denotes the 2-norm. To eliminate dependence on the direction of the initial residual $r^{(0)}$, one sometimes studies the quantity

$$\min_{p_j \in \mathcal{P}_j(0)} \|p_j(A)\|, \tag{2.1}$$

where $\|\cdot\|$ now denotes the operator norm corresponding to the 2-norm for vectors, that is, the largest singular value. The quantity in (2.1) is sometimes referred to as the *ideal* GMRES residual norm [15], and it is an upper bound on $\|r^{(j)}\|/\|r^{(0)}\|$. Clearly, the quantity in (2.1) is always less than or equal to 1, since one could take $p_j(A) = I$, but it is useful to identify properties of the matrix A that will ensure that this quantity is strictly less than 1. This will ensure that j steps of the GMRES algorithm make at least some progress toward the solution of the linear system and hence that if the algorithm is restarted every j steps, then the restarted algorithm GMRES(j) will converge to the solution.

A number of bounds on $\|r^{(j)}\|$ or on the quantity in (2.1) have been derived based on the field of values of A, *assuming* that $W(A)$ does not contain the origin. Eisenstat, Elman, and Schultz [8, 9] showed that

$$\|r^{(j)}\| \le \left(\sqrt{1 - d^2/\|A\|^2}\right)^j \|r^{(0)}\|,$$

where d is the distance from the origin to $W(A)$. It was shown in [18] that this and another bound involving the distance from 0 to $W(A^{-1})$ as well [23, 7] hold also for the ideal GMRES residual norm in (2.1). More recently, Beckermann and others [1, 2, 3, 6] have improved this estimate by considering Faber polynomials corresponding to the field of values of A or various regions containing $W(A)$. Bounds of the form

$$\|r^{(j)}\| \le c\gamma^j \|r^{(0)}\|$$

have been established with an explicit constant $c \leq 3$ and a convergence factor γ depending only on $W(A)$ and its distance to the origin.

In [5], Crouzeix proved the following theorem relating the norm of any polynomial function of A to the maximum value of that polynomial on $W(A)$.

THEOREM 2.1 (Crouzeix). *For any n by n matrix A and any polynomial p,*

$$\|p(A)\| \leq 11.08\|p\|_{W(A)}, \tag{2.2}$$

where $\|\cdot\|_{W(A)}$ denotes the ∞-norm on $W(A)$.

Crouzeix conjectured that the constant 11.08 could be replaced by 2 [4]. We will use this theorem (or conjecture) to establish (or establish under the assumption that the conjecture is true) different bounds on the quantity in (2.1).

In order for the theorem or conjecture to provide useful information about the GMRES algorithm, it appears that the field of values still cannot contain the origin; for if $0 \in W(A)$, then any polynomial p_j with value 1 at 0 obviously satisfies $\|p_j\|_{W(A)} \geq 1$. Note, however, that if $A = \varphi(M)$ for some matrix M and some polynomial φ, then

$$\|p(A)\| = \|p(\varphi(M))\| \leq K\|p \circ \varphi\|_{W(M)} = K\|p\|_{\varphi(W(M))}, \tag{2.3}$$

where $K = 11.08$ (and possibly $K = 2$). Taking $M = A^{1/m}$ and $\varphi(z) = z^m$, we have

$$\min_{p_j \in \mathcal{P}_j(0)} \|p_j(A)\| \leq K \min_{p_j \in \mathcal{P}_j(0)} \|p_j \circ \varphi\|_{W(M)} = K \min_{p \in \mathcal{P}_j(0)} \|p_j\|_{[W(A^{1/m})]^m}. \tag{2.4}$$

For m sufficiently large, $W(A^{1/m})$ clearly excludes the origin and so there is a chance that the right-hand side of (2.4) could be less than 1, unless $[W(A^{1/m})]^m$ wraps entirely around the origin, in which case, by the maximum principle, $\min_{p_j \in \mathcal{P}_j(0)} \|p_j\|_{[W(A^{1/m})]^m} = 1$.

Figures 1, 3, and 5 illustrate the sets $W(A)$, $[W(A^{1/m})]^m$, and $\exp[W(\log A)]$ for some matrices A and different values of m. Here, we have used the principal mth root and principal logarithm since the matrices do not have eigenvalues on the negative real axis. The matrix in Figure 1 was used by Grcar [12] and subsequently many others as an illustration of a matrix with ill-conditioned eigenvalues. It has 1's on its

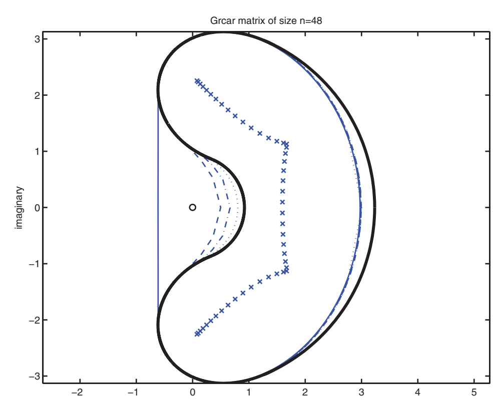

FIG. 1. *Field of values of A (thin solid curve enclosing the origin o) and the sets $[W(A^{1/m})]^m$, $m = 2, 3, 6$ (dashed, dash-dot, dotted) and $\exp[W(\log A)]$ (thick solid curve farthest from the origin). Eigenvalues are marked with x's.*

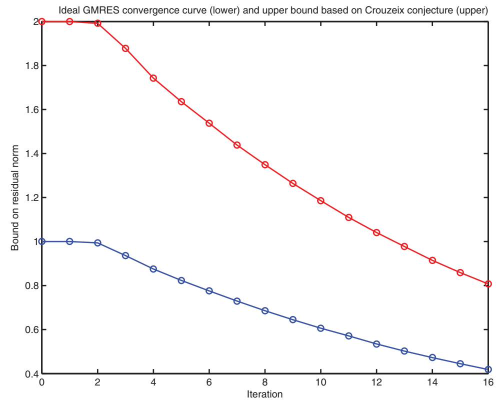

Fig. 2. *Convergence of ideal GMRES (lower curve) and upper bound* (2.5) *(upper curve).*

main diagonal and the first three upper diagonals, -1's on its first lower diagonal, and 0's elsewhere. The computed eigenvalues are marked with x's in the figure, and the boundary of the field of values is plotted with a thin solid line. The field of values of A contains the origin (denoted with an o), but the sets $[W(A^{1/m})]^m$ do not for $m = 2, 3, 6$, as shown by the dashed, dash-dot, and dotted curves in the figure. The thick solid curve is the boundary of the set $\exp[W(\log A)]$. Based on this, it is clear that for j sufficiently large, a polynomial of degree j with value 1 at the origin *can* be constructed so that its value is less than 1 or less than $1/K$, where $K = 11.08$ on any of the sets (referred to as K-spectral sets) pictured here, except for the field of values.

Figure 2 shows the quantity in (2.1) and the upper bound (assuming that Crouzeix's conjecture holds)

$$2 \min_{p_j \in \mathcal{P}_j(0)} \|p_j\|_{\exp[W(\log A)]} \tag{2.5}$$

for the GMRES residual norm, $\|r_j\|/\|r_0\|$. These curves were computed using the SDPT3 semidefinite programming package [25] to determine the ideal GMRES polynomial for A and for a diagonal matrix with diagonal elements around the boundary of $\exp[W(\log A)]$. It can be seen from the figure that this bound gives a fairly good estimate of the actual convergence rate of early steps of the GMRES algorithm for the Grcar matrix. Of course, the GMRES algorithm will find the exact solution after $n = 48$ steps and this will *not* be reflected in the upper bounds based on the field of values of A or related regions. In general, we have found no "monotonicity" in the bounds corresponding to $[W(A^{1/m})]^m$, $m \geq 1$, and $\exp[W(\log A)]$; often the bounds are fairly close once we reach a value of m for which $0 \notin [W(A^{1/m})]^m$, and one bound might be slightly better at one step while another is slightly better at another step.

As noted above, the bounds (2.2) and (2.5) do not always reflect the actual behavior of ideal GMRES. They do not account for the fact that GMRES finds the exact solution after at most n steps for an n by n matrix, and if significant parts of the matrix can be annihilated well before step n, these upper bounds will not indicate this either. As an example, consider a block diagonal matrix A with multiple scaled Jordan blocks. It was shown in [17] that matrices of approximately this form can arise from a streamline upwind Petrov–Galerkin discretization of the convection-diffusion

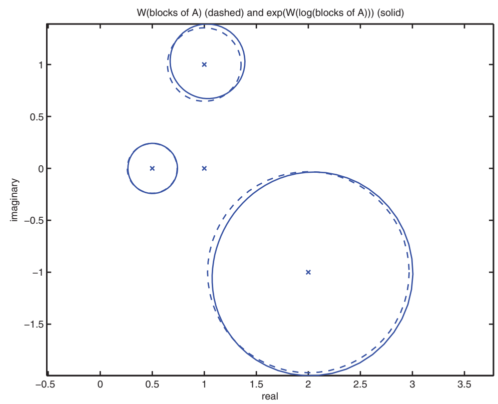

FIG. 3. *Fields of values of blocks A_ℓ (dashed) and the sets* $\exp[W(\log A_\ell)]$ *(solid). x's mark the eigenvalues.*

equation. For simplicity, we chose a matrix A with four scaled Jordan blocks,

$$A_\ell = \begin{pmatrix} \lambda_\ell & \gamma_\ell & & \\ & \ddots & \ddots & \\ & & \ddots & \gamma_\ell \\ & & & \lambda_\ell \end{pmatrix}, \quad \ell = 1, 2, 3, 4,$$

of dimensions 1, 2, 5, and 10, respectively, with $\lambda_1 = 1$, $\lambda_2 = 1 + i$, $\lambda_3 = 2 - i$, and $\lambda_4 = 0.5$. The superdiagonal entry γ_ℓ was taken to be $0.5\ |\lambda_\ell|$. The field of values of this matrix is the convex hull of the fields of values of the individual blocks, but clearly one can obtain a better bound on the GMRES residual by working with the union of the fields of values of the individual blocks:

$$\min_{p_j \in \mathcal{P}_j(0)} \|p_j(A)\| = \min_{p_j \in \mathcal{P}_j(0)} \max_{\ell=1,\ldots,4} \|p_j(A_\ell)\| \leq K \min_{p_j \in \mathcal{P}_j(0)} \|p_j\|_{\cup_{\ell=1}^4 W(A_\ell)}. \tag{2.6}$$

As before, each region $W(A_\ell)$ can be replaced by $[W(A_\ell^{1/m})]^m$, $m = 2, 3, \ldots$ or by $\exp[W(\log A_\ell)]$. Figure 3 shows the regions $W(A_\ell)$ and $\exp[W(\log A_\ell)]$, which are not very different in this case.

Figure 4 shows the ideal GMRES convergence curve and the upper bounds obtained from either $\cup_{\ell=1}^4 W(A_\ell)$ or $\cup_{\ell=1}^4 \exp[W(\log A_\ell)]$, assuming that the constant K in (2.6) is 2. The two bounds are indistinguishable in the figure. Initially, the bounds are too high by only about a factor of 2, but as the degree of the polynomial increases, the actual ideal GMRES polynomial is able to annihilate the Jordan blocks of smaller dimension and concentrate on the one of dimension 10, while the upper bounds require polynomials that are small throughout the fields of values of all blocks.

Just as K-spectral sets provide information about the convergence of the GMRES algorithm, knowing something about the convergence or lack of convergence of that algorithm for a class of matrices tells us something about the K-spectral sets for those

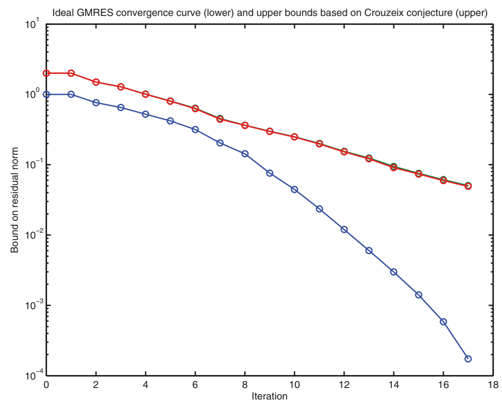

FIG. 4. *Convergence of ideal GMRES (lower curve) and upper bounds based on* $\cup_{\ell=1}^{4} W(A_\ell)$ *or* $\cup_{\ell=1}^{4} \exp[W(\log A_\ell)]$ *(upper curve).*

matrices. When a matrix has the sparsity pattern

$$\begin{pmatrix} 0 & * & 0 & \dots & 0 \\ 0 & * & * & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & * & * & \dots & * \\ * & * & * & \dots & * \end{pmatrix}, \tag{2.7}$$

it is easy to see that the GMRES algorithm makes no progress toward the solution until step n if the initial residual $r^{(0)}$ is a multiple of the first unit vector [13, p. 55]. In that case $Ar^{(0)}$ is a multiple of the nth unit vector, $A^2 r_0$ is a linear combination of the nth and $(n-1)$st unit vectors, etc. The residual at step j of the GMRES algorithm is $r^{(0)}$ minus its orthogonal projection onto $\text{span}\{Ar^{(0)}, A^2 r^{(0)}, \dots, A^j r^{(0)}\}$, but in this case that orthogonal projection is 0 for $j < n$. This means that any K-spectral set for the matrix must have the property that no polynomial of degree $n-1$ or less with value 1 at the origin can be less than $1/K$ in magnitude throughout the set. Hence, one might expect that a K-spectral set for such a matrix, if it does not contain the origin, might wrap completely around the origin. This is indeed the case for the matrix depicted in Figure 5. This is a random 6 by 6 matrix with the sparsity pattern in (2.7). Its field of values contains the origin, as shown in Figure 5(a), and the field of values of $\log(A)$ contains points with imaginary parts ranging over more than a 2π interval, as shown in Figure 5(b). This means that when those points are exponentiated, the values $e^{x+iy} = e^x \cdot e^{iy}$ will have arguments that range beyond 2π. The exponential of the boundary of $W[\log(A)]$ is shown in Figure 5(d), and Figure 5(c) shows a closeup of the region $\exp[W(\log A)]$ near the origin. It excludes the origin, but by the maximum principal, no polynomial with value 1 at the origin can have magnitude less than 1 throughout this region.

When A is an n by n matrix, the GMRES algorithm finds the exact solution to the linear system $Ax = b$ in at most n steps. When A is an infinite dimensional bounded linear operator on a Hilbert space, it can be shown that the GMRES algorithm converges to the solution of the operator equation $Ax = b$ if and only if the unbounded component of the resolvent set contains the origin [21, Theorem 3.3.4]; that is, the spectrum (which is the complement of the resolvent set) does not contain the origin or wind completely around the origin. In this case, the mth root of the

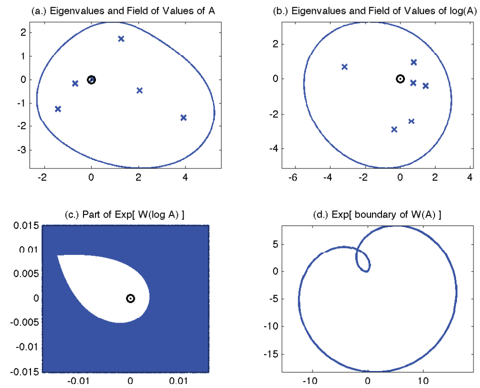

FIG. 5. *Sets associated with a random* 6 *by* 6 *matrix* A *with the sparsity pattern in* (2.7). (a) *Eigenvalues and field of values of* A. (b) *Eigenvalues and field of values of* $\log(A)$. (c) *Part of* $\exp[W(\log A)]$ *near the origin.* (d) $\exp[$*boundary of* $W(\log A)]$.

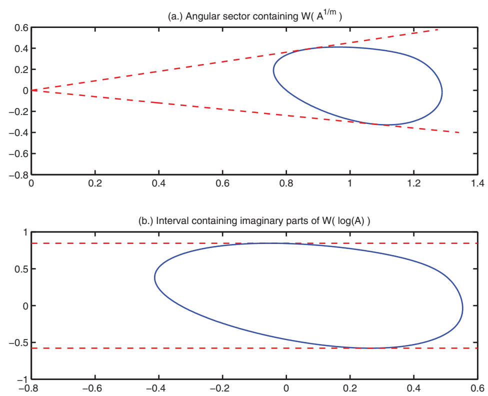

FIG. 6. (a) *Field of values of* $A^{1/m}$ *contained in an angular sector with angle less than* $2\pi/m$. (b) *Field of values of* $\log A$ *with imaginary parts in an interval of width less than* 2π.

operator can be defined using a branch cut that does not intersect the spectrum. We can now bound the rate of convergence of the GMRES algorithm in terms of an approximation problem on a K-spectral set related to the field of values. All of the previous results about matrices hold for infinite dimensional bounded linear operators on a Hilbert space if the field of values $W(A)$ is replaced by its closure $\mathrm{cl}(W(A))$; in finite dimensions, $W(A)$ is already closed.

THEOREM 2.2. *Let* A *be a bounded linear operator on a Hilbert space and suppose that either*

- *for some* $m = 1, 2, \ldots,$ *cl*$(W(A^{1/m}))$ *does not contain the origin and is a subset of an angular sector, anchored at the origin, of angle* θ/m, *where* $\theta < 2\pi$, *as pictured in Figure* 6(a), *or*
- *cl*$(W(\log A))$ *consists of points with imaginary parts in an interval of width less than* 2π, *as pictured in Figure* 6(b).

Then the GMRES algorithm for the operator equation $Ax = b$ *converges to the solution* x *and the residual at step* j *satisfies*

$$\|r_j\|/\|r_0\| \le K \min_{p_j \in \mathcal{P}_j(0)} \|p_j\|_\Omega, \tag{2.8}$$

where $\Omega = [cl(W(A^{1/m})]^m$ *or* $\Omega = \exp[cl(W(\log A))]$ *and* $K \leq 11.08$*. For* j *sufficiently large, the right-hand side of* (2.8) *is strictly less than* 1*.*

Proof. By Mergelyan's theorem [19, 20], any function that is holomorphic on the interior of a compact set S and continuous on its boundary, where $\mathbb{C} \setminus S$ is connected, can be arbitrarily well approximated on S by polynomials. Since, under the given assumptions, $[\text{cl}(W(A^{1/m}))]^m$ or $\exp[\text{cl}(W(\log A))]$ does not wind completely around the origin, one can draw a region S containing $[\text{cl}(W(A^{1/m}))]^m$ or $\exp[\text{cl}(W(\log A))]$ in its interior with the origin on its boundary and, say, let $g : S \to \mathcal{D}$ be a conformal mapping from this region to the unit disk with $g(0) = 1$. Then, since $\max\{|g(z)| : z \in [\text{cl}(W(A^{1/m}))]^m\}$ or $\max\{|g(z)| : z \in \exp[\text{cl}(W(\log A))]\}$ is strictly less than 1, by Mergelyan's theorem there is a polynomial p with $p(0) = 1$ that has this property as well. Appropriate powers of this polynomial would then be strictly less than $1/K$, where $K \leq 11.08$ for these regions. ∎

With more restrictive assumptions about the K-spectral sets $[\text{cl}(W(A^{1/m}))]^m$ and $\exp[\text{cl}(W(\log A))]$, one can give bounds on the rate of convergence for restarted GMRES. For example, using powers of a simple first degree polynomial, we can derive the following theorem.

THEOREM 2.3. *Suppose that either*

- *for some* $m = 1, 2, \ldots$, $cl(W(A^{1/m}))$ *does not contain the origin and is a subset of a piece of an angular sector of angle* θ/m*, where* $\theta < \pi$*, between radius* $r^{1/m}$ *and* $R^{1/m}$*, as pictured in Figure* 6(a)*, or*
- $cl(W(\log A))$ *consists of points with imaginary parts in an interval of width* $\theta < \pi$ *and real parts between* $\log r$ *and* $\log R$*, as pictured in Figure* 6(b)*.*

Let j *satisfy*

$$K \left(\sqrt{1 - 4\frac{rR}{(r+R)^2} \cos^2(\theta/2)} \right)^j < 1, \tag{2.9}$$

where $K = 11.08$ *(or perhaps* 2*). Then GMRES(*j*) converges and reduces the* 2*-norm of the residual at each cycle by at least the factor on the left-hand side of* (2.9)*.*

Proof. Assume without loss of generality that the angular sector is centered about the positive real axis so that arguments of the points in $[\text{cl}(W(A^{1/m}))]^m$ are all between $-\theta/2$ and $\theta/2$ and magnitudes are between r and R. Similarly, assume without loss of generality that the interval containing the imaginary parts of points in $\text{cl}(W(\log A))$ is $[-\theta/2, \theta/2]$, so that $\exp[\text{cl}(W(\log A))]$ consists of points with arguments between $-\theta/2$ and $\theta/2$ and magnitudes between r and R. Consider the first degree polynomial $p_1(z) = 1 - z/S$, where $S = (R + r)/(2\cos(\theta/2))$. The magnitude of this polynomial at a point $se^{i\varphi}$ in $[\text{cl}(W(A^{1/m}))]^m$ or $\exp[\text{cl}(W(\log A))]$ is

$$\left| 1 - \frac{se^{i\varphi}}{S} \right| = \sqrt{1 - 2\frac{s}{S}\cos\varphi + \left(\frac{s}{S}\right)^2}.$$

This is largest when $\varphi = \pm\theta/2$ and $s = r$ or $s = R$. At these points, the magnitude of the polynomial is

$$\sqrt{1 - 4\frac{rR}{(r+R)^2}\cos^2(\theta/2)}.$$

It follows that under assumption (2.9), $(p_1(z))^j$ is a jth degree polynomial with value 1 at the origin that is less than $1/K$ throughout the K-spectral set $[\text{cl}(W(A^{1/m}))]^m$

or $\exp[\mathrm{cl}(W(\log A))]$. Hence, $\|(p_1(A))^j\| < 1$ and the norm of the jth degree ideal GMRES polynomial in (2.1) is at least this small. ∎

3. Crouzeix's conjecture and the Koebe algorithm for conformal mapping. Crouzeix's conjecture—that for any square matrix A and any polynomial p, $\|p(A)\| \leq 2\|p\|_{W(A)}$—and its implications for the GMRES algorithm were discussed in the previous section. Let g be a bijective holomorphic mapping from $W(A)$ onto the unit disk $\mathcal{D}$. A conjecture mentioned in [14], which would imply Crouzeix's conjecture, is that $g(A)$ is 2-similar to a contraction; that is, there is an invertible matrix S with $\kappa(S) \equiv \|S\| \cdot \|S^{-1}\| \leq 2$ such that $C \equiv S^{-1}g(A)S$ is a contraction. If this (possibly stronger) conjecture holds, then, using von Neumann's inequality (1.8),

$$\|p(A)\| = \|(p \circ g^{-1})(g(A))\| \leq 2\|(p \circ g^{-1})(C)\| \leq 2\|(p \circ g^{-1})\|_{\mathcal{D}} = 2\|p\|_{W(A)}.$$

Let $\Omega \subset \mathcal{D}$ be a region containing the origin in its interior. The Koebe algorithm for the conformal mapping of Ω to $\mathcal{D}$, with 0 mapping to 0, is as follows [16]:

Repeat until convergence:

1. Find a point a_0 on $\partial\Omega$ that is closest to the origin. For convenience, rotate Ω so that this point lies on the negative real axis, and denote the rotated point as $-\rho$. Move this point to the origin through the transformation
$$z \to f_1(z) = \frac{z+\rho}{1+\rho z}.$$
2. Take the square root (or the mth root for $m = 3, 4, \ldots$):
$$z \to f_2(z) = z^{1/m}.$$
(The mth root function can be defined as a one to one analytic function in $f_1(\Omega)$ whose range is made unique by requiring that $\rho^{1/m} > 0$.)
3. Move the image of the origin, namely, $b_0 = \rho^{1/m}$, back to 0 via
$$z \to f_3(z) = \frac{z - b_0}{1 - b_0 z}.$$

By shifting and scaling a given matrix, one can move its field of values inside the unit disk with the origin in the interior. Let the shifted and scaled matrix be denoted as A. Then the Koebe algorithm can be used to map $W(A)$ to $\mathcal{D}$. Okubo and Ando showed that if the numerical radius of A, $r(A) = \max_{z \in W(A)} |z|$, is less than or equal to 1, then A is 2-similar to a contraction [22]. If the operations of the Koebe algorithm are applied to A, then $g(A)$ will be 2-similar to a contraction if and only if the matrices $f_1(A)$, $(f_2 \circ f_1)(A)$, $(f_3 \circ f_2 \circ f_1)(A)$, etc., arising at every step of the Koebe algorithm are 2-similar to contractions. The "if" part follows by taking the limit, and the "only if" part follows because the operations can all be inverted and the inverse operations can be shown to preserve contractions; that is, if C is a contraction, then

$$f_3^{-1}(C) = (I + b_0 C)^{-1}(C + b_0 I)$$

is a contraction since this function is analytic throughout the unit disk and maps the unit disk to itself,

$$f_2^{-1}(C) = C^m$$

is a contraction since $\|C^m\| \leq \|C\|^m \leq 1$, etc.

In the forward algorithm, the only operation that does *not* necessarily preserve contractions is the mth root operation, f_2. Assuming that the initial set $W(A) = \Omega$ is inside the open unit disk, the Okubo and Ando result [22] implies that A is 2-similar to a strict contraction C (i.e., to a matrix whose norm is less than or equal to $r(A) < 1$). It follows from Theorem 1.3 that $f_1(C)$ is a strict contraction and so $f_1(A)$ is 2-similar to a strict contraction via the same similarity transformation. Similarly, if $(f_2 \circ f_1)(A)$ is 2-similar to a strict contraction, then Theorem 1.3 implies that $(f_3 \circ f_2 \circ f_1)(A)$ is as well. This argument could then be repeated to show that 2-similarity to a strict contraction is maintained throughout the Koebe algorithm (assuming that it is maintained at step 2) and hence that $g(A)$ is 2-similar to a contraction.

Now, the points a_0 in step 1 do not have to be chosen as points on $\partial\Omega$ that are closest to the origin; other points on $\partial\Omega$ may work as well [16, exercise 2, p. 344]. Unfortunately, choosing points to satisfy the assumptions of Theorem 1.3 (namely, $\rho \in (\|f(C)\|, 1)$, where C is a strict contraction to which A is 2-similar and f is the result of all previous operations in the algorithm) may not lead to a convergent algorithm. This is not surprising since numerical experiments show that not every contraction to which A is 2-similar remains a contraction when operated on by the functions in the Koebe algorithm. Still, in all numerical experiments that we have performed, we have found some contraction C to which A is 2-similar that does remain a contraction throughout the Koebe algorithm and hence for which $g(A)$ is 2-similar to the contraction $g(C)$. This remains to be proved.

Acknowledgment. The authors thank two anonymous referees for helpful comments on a first version of this manuscript. In particular, one of the referees provided the current statement and proof of Theorem 1.3.

REFERENCES

[1] B. Beckermann, *Image numérique, GMRES et polynômes de Faber*, C.R. Acad. Sci. Paris Ser. I, 340 (2005), pp. 855–860.

[2] B. Beckermann and M. Crouzeix, *Faber polynomials of matrices for non-convex sets*, JAEN J. Approx., 6 (2014), pp. 219–231.

[3] B. Beckermann, S. Goreinov, and E. Tyrtyshnikov, *Some remarks on the Elman estimate for GMRES*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 772–778.

[4] M. Crouzeix, *Bounds for analytical functions of matrices*, Integral Equations Operator Theory, 48 (2004), pp. 461–477.

[5] M. Crouzeix, *Numerical range and functional calculus in Hilbert space*, J. Funct. Anal., 244 (2007), pp. 668–690.

[6] M. Eiermann, *Fields of values and iterative methods*, Linear Algebra Appl., 180 (1993), pp. 167–197.

[7] M. Eiermann and O. Ernst, *Geometric aspects of the theory of Krylov subspace methods*, Acta. Numer., 10 (2001), pp. 251–312.

[8] S. Eisenstat, H. Elman, and M. Schultz, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 20 (1983), pp. 345–357.

[9] H. C. Elman, *Iterative Methods for Sparse Nonsymmetric Systems of Linear Equations*, Ph.D. thesis, Department of of Computer Science, Yale University, New Haven, CT, 1982.

[10] J. B. Garnett, *Bounded Analytic Functions*, Springer, New York, 2007.

[11] H. Gau and P. Wu, *Finite Blaschke products of contractions*, Linear Algebra Appl., 368 (2003), pp. 359–370.

[12] J. Grcar, *Operator Coefficient Methods for Linear Equations*, Sandia National Laboratories report SAND89-8691, 1989.

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

[14] A. Greenbaum and D. Choi, *Crouzeix's Conjecture and Perturbed Jordan Blocks*, Linear Algebra Appl., 436 (2011), pp. 2342–2352.

[15] A. Greenbaum and L. N. Trefethen, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–368.

[16] P. Henrici, *Applied and Computational Complex Analysis*, Vol. III, Wiley & Sons, New York, 1986.
[17] J. Liesen and Z. Strakoš, *GMRES convergence analysis for a convection-diffusion problem*, SIAM J. Sci. Comput., 26 (2005), pp. 1989–2009.
[18] J. Liesen and P. Tichý, *The field of values bound on ideal GMRES*, arXiv:1211.5969, 2013.
[19] S. Mergelyan, *On the representation of functions by series of polynomials on closed sets*, Transl. Amer. Math. Soc., 3 (1962), pp. 287–293.
[20] S. Mergelyan, *Uniform approximations to functions of a complex variable*, Transl. Amer. Math. Soc., 3 (1962), pp. 294–391.
[21] O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.
[22] K. Okubo and T. Ando, *Constants related to operators of class* C_ρ, Manuscripta Math., 16 (1975) pp. 385–394.
[23] G. Starke, *Iterative Methods and Decomposition-Based Preconditioners for Nonsymmetric Elliptic Boundary Value Problems*, Habilitationsschrift, Universität Karlsruhe, Karlsruhe, 1994.
[24] B. Sz.-Nagy and C. Foias, *Harmonic Analysis of Operators on Hilbert Space*, North-Holland, New York, 1970.
[25] K. C. Toh, M. J. Todd, and R. H. Tütüncü, *SDPT3–A MATLAB software package for semidefinite programming*, Optim. Methods Softw., 11-12 (1999), pp. 545–581.
[26] L. N. Trefethen, *Pseudospectra of linear operators*, SIAM Rev., 39 (1997), pp. 383–406.
[27] J. von Neumann, *Eine spektraltheorie für allgemeine operatoren eines unitären raumes*, Math. Nachr., 4 (1951), pp. 258–281.