Some uses of the field of values in numerical analysis

Michele Benzi

Bollettino dell'Unione Matematica Italiana

ISSN 1972-6724
Volume 14
Number 1

Boll Unione Mat Ital (2021) 14:159-177
DOI 10.1007/s40574-020-00249-2

Springer

Bollettino dell'Unione Matematica Italiana (2021) 14:159–177
https://doi.org/10.1007/s40574-020-00249-2

Some uses of the field of values in numerical analysis

Michele Benzi[1]

Received: 13 March 2020 / Accepted: 14 July 2020 / Published online: 22 July 2020

Abstract
In this expository paper we illustrate the role that the field of values (or numerical range) of a matrix plays in connection with certain problems of numerical analysis. These include the approximation of matrix functions and the convergence of preconditioned Krylov subspace methods for solving large systems of equations arising from the discretization of partial differential equations.

Keywords Field of values · Numerical linear algebra · Functions of matrices · Krylov subspace methods

1 Introduction

The *field of values*, or *numerical range*, of a matrix (or operator in Hilbert space) is a well-studied object in linear algebra and functional analysis [28]. Some of its fundamental properties were identified by Hausdorff and Toeplitz over a century ago [30,45]. In recent decades, the field of values has become increasingly important in numerical analysis, in particular in certain problems of numerical linear algebra involving functions of matrices and iterative methods for solving large systems of linear equations. In such problems one has to deal with sequences of matrices of increasing (potentially unbounded) dimension.
For instance, the matrices may arise from the discretization of differential or integral operators, and their dimension tends to infinity as the discretization is refined; in other cases the discretization is fixed, but the size of the computational domain may increase without bounds. Analyzing the behavior of algorithms for the approximation of functions of such matrices (or, more typically, for the approximation of the action of matrix functions on vectors) as their size increases is of central importance in numerical linear algebra.

For sequences of *normal* matrices, the eigenvalues provide all the necessary information to establish the convergence rates of approximation algorithms; indeed, the spectral theorem for normal matrices (and bounded operators) allows one to translate the approximation problem for functions of matrices into one for functions of a real or complex variable, and to make use of classical results from approximation theory. For matrices that "stay close to normal", in a sense that can be made precise, the eigenvalues are still useful indicators of what's going

✉ Michele Benzi
michele.benzi@sns.it

[1] Scuola Normale Superiore, Piazza dei Cavalieri, 7, 56126 Pisa, Italy

on as the dimension grows. If, however, the matrices are far from normal, and particularly if the departure from normality grows as the dimension increases, then it has long been known that the eigenvalues alone are not sufficient to capture many phenomena of interest and may even paint a misleading picture [46].

Consider for example the following two (for simplicity, finite-dimensional) linear dynamical systems, the first one discrete, the second one continuous in time:

1. $x_{k+1} = Ax_k + b$ with $k = 0, 1, \ldots$
2. $\dot{x} = Ax + b$ with $x = x(t)$, $t \geq 0$, where $x(0) = x_0$.

Here $A \in \mathbb{C}^{n\times n}$ and $b \in \mathbb{C}^n$ are fixed, and $x_0 \in \mathbb{C}^n$ is prescribed and arbitrary. As is well known, the long-term behavior of both evolution processes is governed by the spectral properties of A. Specifically:

(i) In the discrete time case, as $k \to \infty$ the iterates x_k converge, for any choice of x_0, to the unique solution of $x = Ax + b$ if and only if the eigenvalues of A satisfy $|\lambda_i(A)| < 1$ for all $i = 1, \ldots, n$.
(ii) In the continuous time case, as $t \to \infty$, $x(t)$ converges, for any choice of x_0, to the steady state x_* (solution of $Ax+b = 0$) if and only if $\Re(\lambda_i(A)) < 0$ for all $i = 1, \ldots, n$.

In practice, we are interested in the rate of convergence. In the first case, the asymptotic rate of convergence is dictated by the *spectral radius* of A:

$$\rho(A) := \max_i \{\, |\lambda_i(A)| \,;\ \lambda_i(A) \text{ is an eigenvalue of } A \,\}.$$

In the second case, by the *spectral abscissa* of A:

$$\alpha(A) := \max_i \{\, \Re(\lambda_i(A)) \,;\ \lambda_i(A) \text{ is an eigenvalue of } A \,\}.$$

If A is normal (i.e., unitarily diagonalizable), then $\rho(A)$ and $\alpha(A)$ completely describe the evolution of x_k and $x(t)$, not just asymptotically, but for all $k = 0, 1, \ldots$ and $t \geq 0$, respectively. Indeed, if we denote by $\|\cdot\|$ the operator norm induced by the Euclidean norm on $\mathbb{C}^n$, we have, by unitary invariance of $\|\cdot\|$, $\|A\| = \rho(A)$, and therefore if A is normal and $\rho(A) < 1$ we have that

$$\|A^k\| = \|A\|^k = \rho(A)^k \to 0 \text{ monotonically as } k \to \infty.$$

Likewise, if A is normal and $\alpha(A) < 0$ we have that

$$\|\mathrm{e}^{tA}\| = \mathrm{e}^{t\alpha(A)} \to 0 \text{ monotonically as } t \to \infty.$$

Hence, in both cases the dynamics is governed at all times by the (extreme) eigenvalues, when A is normal. What happens, however, when A is non-normal? In particular, highly non-normal?

Suppose for the time being that A is diagonalizable: $A = XDX^{-1}$ with D diagonal, for some nonsingular $X \in \mathbb{C}^{n\times n}$. Then

$$\|A^k\| = \|XD^kX^{-1}\| \leq \kappa(X)\rho(A)^k, \tag{1}$$

where $\kappa(X) = \inf \|X\|\|X^{-1}\|$, where the infimum is taken over all nonsingular matrices X that diagonalize A; note that $\kappa(X) \geq 1$, and that $\kappa(X) = 1$ when A is normal. This quantity is known as the *spectral condition number of the eigenbasis* of A.

Similarly, in the continuous time case we have

$$\|\mathrm{e}^{tA}\| = \|X\mathrm{e}^{tD}X^{-1}\| \leq \kappa(X)\,\mathrm{e}^{t\alpha(A)}. \tag{2}$$

In numerical analysis one often deals not with a single problem of fixed size, but with sequences of problems of increasing size, usually due to some discretization parameter going to zero. It is easy to find examples of sequences of matrices of increasing size, of interest in applications, for which the condition number of the eigenbasis grows without bounds, even though their spectral radius or the spectral abscissa remain bounded away from their critical values, 1 and 0 (such an example is described in Sect. 5, see (11)). It is clear that in such cases the bounds (1)–(2) are virtually useless when trying to establish the actual rate of convergence: although the right-hand sides of both (1) and (2) eventually approach zero, if $\kappa(X)$ is very large then we cannot infer anything on the transient behavior of the quantities on the left-hand side. If A is not diagonalizable, the situation is even worse. Some asymptotic estimates involving the size of the largest Jordan block in the Jordan canonical form of A are known [47, Theorem 3.1], but they are of limited practical use; see also the discussion in [46, Chapter 16].

Informally, A is said to be *highly non-normal* if it is not diagonalizable or if the corresponding condition number of the eigenbasis, $\kappa(X)$, is very large. For matrices like these, the onset of the asymptotic convergence regime may manifest itself only after very large times; in other cases, the bounds (1)–(2) may be so loose as to be uninformative. Thus, the eigenvalues give at best a partial picture of the underlying behavior.

Another limitation of spectral analysis is that the eigenvalues of a non-normal matrix can be highly sensitive to perturbations. For instance, due to unavoidable rounding errors, finite precision approximations of the above linear dynamical systems 1–2 are governed not by the exact spectral radius or spectral abscissa of A but by those of a slightly perturbed matrix $\tilde{A} \approx A$. If A has highly sensitive eigenvalues, which is often the case when A is far from normal, it may happen that $\rho(\tilde{A}) > 1$ or $\alpha(\tilde{A}) > 0$, even though the unperturbed matrix A amply satisfies $\rho(A) < 1$ or $\alpha(A) < 0$, thus causing divergence or blow up of the computed quantities.

The limitations of eigenvalue analysis become even more apparent when we consider processes that are more complex than the convergence of simple linear (discrete or continuous) dynamical systems. In the rest of the paper we will focus on two such problems from the field of numerical linear algebra.

2 Two problems in numerical linear algebra

In this section, we briefly introduce two important problems in numerical linear algebra, one concerning functions of matrices, the other one the solution of large systems of linear equations; as we will see, while apparently rather different, the two problems are closely related.

2.1 Decay estimates for functions of large matrices

In several applications, given a complex-valued function f defined on the spectrum of A, we are interested in obtaining estimates, or bounds, for the entries of the matrix $f(A)$.

Typically, f is analytic and A is banded or sparse. We say that a matrix

$$A = [a_{ij}] \in \mathbb{C}^{n \times n}$$

is *k-banded* if $a_{ij} = 0$ for all i, j with $|i - j| > k$. For instance, a tridiagonal matrix is 1-banded. In the following discussion, one should think of k being fixed, while the dimension n of A grows unbounded ($n \to \infty$).

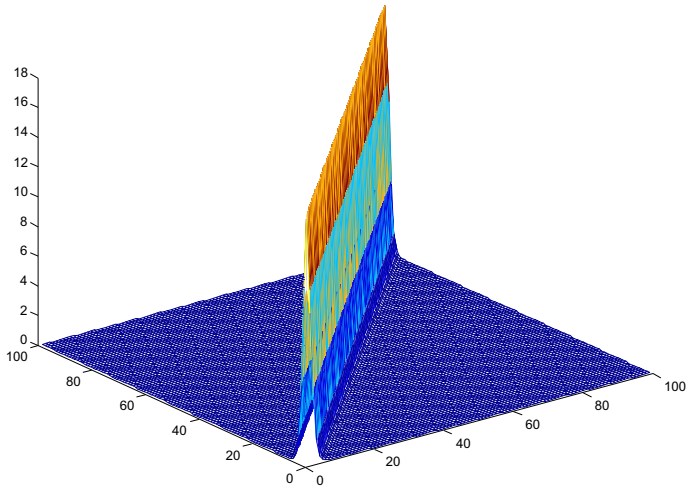

Fig. 1 Plot of $|[\mathrm{e}^A]_{ij}|$ for A tridiagonal (discrete 1D Laplacian)

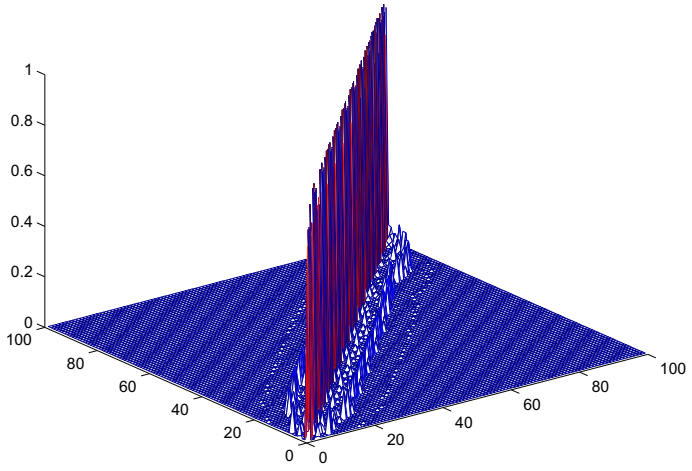

Fig. 2 Plot of $|[A^{1/2}]_{ij}|$ for matrix `nos4` from the SuiteSparse Collection [18] (scaled and reordered with reverse Cuthill–McKee)

Among the various equivalent definitions of a matrix function, we can use for instance the following one based on contour integration (and due to E. Cartan):

$$f(A) = \frac{1}{2\pi i} \int_\Gamma f(z)(zI - A)^{-1} \mathrm{d}z,$$

where Γ is a contour in $\mathbb{C}$, counterclockwise oriented, containing the eigenvalues of A in its interior, and such that f is analytic inside and on Γ. We refer to [31] for details and other, equivalent, definitions of matrix function.

It is frequently observed that while functions of banded or sparse matrices are fully populated, the magnitude of the entries in $f(A)$ that are in some sense far from the nonzero entries

in A are often small, and in fact they tend to decay with the distance; see Figs. 1 and 2 for two such examples. Based on this observation, several bounds or estimates for the entries of functions of banded or sparse matrices have been obtained. Typically, when f is analytic and A is banded these take the form of exponential off-diagonal decay bounds:

$$|[f(A)]_{ij}| \leq K\, \mathrm{e}^{-\alpha|i-j|}, \quad \forall i, j = 1, \ldots, n. \tag{3}$$

Note that for any fixed matrix A this inequality can always be trivially satisfied by taking K large enough, but here we are interested in non-trivial bounds where the constants K and $\alpha > 0$ are given explicitly in terms of properties of f and A, such as the location of the singularities of f, the spectral properties of A, and the bandwidth k. Special interest is placed in those cases where K and α are independent of the dimension n. In this case we speak of *localization* of the entries of $f(A)$. We refer to [6] for an extensive survey of matrix localization, and to the recent thesis [42] for further results and applications.

When A is *sparse*, but not necessarily banded, the bounds take the form

$$|[f(A)]_{ij}| \leq K\, \mathrm{e}^{-\alpha d(i,j)}, \quad \forall i, j = 1, \ldots, n, \tag{4}$$

where $d(i, j)$ is now the *geodesic distance* between nodes i and j, i.e., the length of the shortest path joining nodes i and j in the graph $G(A)$ associated with A, where there is an edge between node i and node j if and only if $a_{ij} \neq 0$. Note that this is a genuine distance only if A is structurally symmetric (i.e., $G(A)$ is undirected).

An example of such a bound is the following one from [11]: if $A = XDX^{-1}$ is diagonalizable and sparse, then

$$|[f(A)]_{ij}| \leq \underbrace{\kappa(X)K_0}_{=K}\, \mathrm{e}^{-\alpha d(i,j)}, \quad \forall i, j = 1, \ldots, n. \tag{5}$$

Here the positive constants K_0 and α depend only on the distance between the singularities of f (if any) and the spectrum of A and on the maximum of $|f|$ on the boundary of a region $\mathcal{F} \subset \mathbb{C}$ containing the eigenvalues of A and such that f is analytic on $\mathcal{F}$. Hence, (5) is not a single bound but a family of bounds, parameterized by the choice of $\mathcal{F}$. There is a trade-off involved: taking a larger set $\mathcal{F}$ may lead to faster exponential decay (larger α), but K_0 will also become larger. If f is entire, the set can be chosen arbitrarily, leading to *superexponential decay* estimates; that is, $\alpha > 0$ can be taken arbitrarily large, but of course K_0 will also grow without bounds (except for the trivial case of constant f), in view of Liouville's Theorem.

Clearly, the bound (5) suffers from the same limitations as the ones for $\|A^k\|$ or $\|\mathrm{e}^{tA}\|$ we discussed earlier: the presence of $\kappa(X)$ makes the bound virtually useless, unless A is normal ($\kappa(X) = 1$), or nearly normal ($\kappa(X)$ small). In particular, if $\kappa(X)$ depends on n, we don't obtain uniform bounds in n. Of course, if A is not diagonalizable then the bounds simply do not apply.

We shall come back to this problem in Sect. 5.

2.2 Convergence bounds for Krylov subspace methods

The second problem concerns the characterization of the convergence of minimal residual-type *Krylov subspace methods* to the solution of large-scale linear systems arising from the discretization of certain PDEs or systems of PDEs.

These methods construct polynomial approximations of the form $x_k = p_k(A)b$ to the solution $x_* = A^{-1}b$ of the system $Ax = b$. The polynomial p_k is chosen so as to satisfy an optimality condition [40]. When A is Hermitian there are two main approaches, both based

on the minimization of two different norms of the residual $r_k = b - Ax_k$ over a suitable subspace of dimension k at each step $k = 1, 2, \ldots$. Without loss of generality, here we assume that $x_0 = 0$. These two approaches lead to the *Minimal Residual (MINRES)* method and to the *Conjugate Gradient (CG)* method, respectively.

The Minimal Residual method determines the vector x_k which minimizes the ℓ^2-norm of the residual $\|r_k\| = \|b - Ax_k\|$ over the kth *Krylov subspace*

$$\mathcal{K}_k(A, b) := \text{span}\,\{b, Ab, A^2b, \ldots, A^{k-1}b\}.$$

Note that the vectors in this subspace are of the form $p_{k-1}(A)b$, where p_k is a polynomial of degree k, and that the Krylov subspaces form a nested sequence, $\mathcal{K}_k(A, b) \subseteq \mathcal{K}_{k+1}(A, b)$. Therefore, the sequence of residual norms $\|r_k\|$ is non-increasing.

On the other hand, if A is positive definite the Conjugate Gradient method minimizes

$$\|b - Ax_k\|_{A^{-1}} = \sqrt{(b - Ax_k)^* A^{-1}(b - Ax_k)} = \|A^{-1}b - x_k\|_A$$

over the same subspace. Again, the convergence is monotonic in this norm.

For both of these methods, the eigenvalues of A are descriptive of the convergence behavior. Indeed, for MINRES we have the following bound:

$$\|r_k\| \leq \min_{p \in \Pi_k} \max_{\lambda \in \Lambda(A)} |p(\lambda)|\, \|r_0\|\,, \tag{6}$$

where Π_k denotes the set of all polynomials of degree $\leq k$ that satisfy $p(0) = 1$.

For CG we have the analogous bound in the appropriate norm:

$$\|r_k\|_{A^{-1}} \leq \min_{p \in \Pi_k} \max_{\lambda \in \Lambda(A)} |p(\lambda)|\, \|r_0\|_{A^{-1}}\,. \tag{7}$$

We note that both bounds (6) and (7) are sharp; see, e.g., [26, Chapter 3], or [34, Theorems 5.6.6 and 5.7.4]. Hence, for both MINRES and CG the convergence will be fast if there exists a polynomial of low degree (having the value one at zero) that takes small values on the eigenvalues of A, and this depends only on the distribution of the eigenvalues of A.

For a general matrix A, GMRES (Generalized Minimum Residual method, [40]) minimizes the ℓ^2-norm of the residual over the Krylov subspace method $\mathcal{K}_k(A, b)$ at each step. If A is diagonalizable, $A = XDX^{-1}$, then the residual norm at step k satisfies

$$\|r_k\| = \min_{p \in \Pi_k} \|p(A)r_0\| \leq \min_{p \in \Pi_k} \|Xp(D)X^{-1}\|\|r_0\|,$$

leading again to a crude bound of the form

$$\frac{\|r_k\|}{\|r_0\|} \leq \kappa(X) \min_{p \in \Pi_k} \max_{\lambda \in \Lambda(A)} |p(\lambda)|. \tag{8}$$

If A is normal, $\kappa(X) = 1$ and we recover the bound for MINRES. If $\kappa(X)$ is large, however, the right-hand side of (8) may provide no information; in particular, if the right-hand side is > 1 the bound doesn't even capture the non-increasing behavior of the residual norms $\|r_k\|$.

Furthermore, it has been shown by Greenbaum et al. [27] that, given any set of n not necessarily distinct complex numbers $\lambda_1, \ldots, \lambda_n$ (for instance, all equal to 1) and any nonincreasing sequence of n nonnegative values $\rho_0, \ldots, \rho_{n-1}$, it is possible to construct a matrix $A \in \mathbb{C}^{n \times n}$ and a right-hand side $b \in \mathbb{C}^n$ such that A has the λ_i as its eigenvalues and GMRES with initial guess $x_0 = 0$ produces a sequence of residuals $\{r_k\}$ with $\|r_k\| = \rho_k$ for $k = 0, 1, \ldots, n - 1$.

In other words: any non-increasing convergence curve is possible for GMRES, and the eigenvalues of A, in general, do not contain enough information to describe the convergence

behavior. It follows that when A is far from normal, other tools must be sought. While it is unlikely that we will ever find a fully satisfactory answer to the problem of characterizing the convergence of GMRES in general (see [23]), in Sect. 6.1 we will see that in certain special cases it is possible to give reasonably satisfactory convergence bounds.

3 What else is there besides the spectrum?

As we have seen, when A is non-normal, eigenvalue information alone is not enough to analyze various fundamental problems in numerical linear algebra, and in some cases it can even be misleading. Moreover, when A is non-normal (for example, A is defective or close to a defective matrix) the spectrum lacks robustness in the presence of perturbations in the data, which are unavoidable in finite precision computations. It is also desirable to find approaches that do not assume the diagonalizability of A.

Among the sets associated to an operator A that have been proposed as substitutes for the spectrum $\Lambda(A)$, we mention the following:

1. The *pseudospectrum* $\Lambda_\epsilon(A)$;
2. The *field of values* $\mathcal{W}(A)$;
3. Various *spectral sets*, intermediate between $\Lambda(A)$ and $\mathcal{W}(A)$.

These sets allow us to do away with the diagonalizability assumption and lead to bounds that do not depend on $\kappa(X)$. After a brief discussion of the pseudospectrum, we will focus our attention on the field of values; other spectral sets are mentioned in passing in the conclusion section.

3.1 The pseudospectrum

Let $A \in \mathbb{C}^{n\times n}$ and let $\epsilon > 0$. The *ϵ-pseudospectrum* of A is the set

$$\Lambda_\epsilon(A) = \{z \in \mathbb{C};\ \|(zI - A)^{-1}\| > \epsilon^{-1}\}.$$

It can be equivalently defined as the set of all $z \in \mathbb{C}$ such that there exists a matrix $\Delta A \in \mathbb{C}^{n\times n}$ with $\|\Delta A\| < \epsilon$ and $z \in \Lambda(A + \Delta A)$. In other words, the ϵ-pseudospectrum of A is the set of all complex numbers that are eigenvalues of ϵ-perturbations of A [46].

When A is normal, the ϵ-pseudospectrum of A is just

$$\Lambda_\epsilon(A) = \Lambda(A) + \Delta_\epsilon, \quad \text{where} \quad \Delta_\epsilon = \{z \in \mathbb{C};\ |z| < \epsilon\},$$

where, as usual, the sum of sets is defined elementwise (Minkowski addition). However, when A is far from normal, $\Lambda_\epsilon(A)$ can be much larger than $\Lambda(A)$ even for very small values of ϵ.

Consider now the problem of bounding the approximation error

$$\|f(A) - q(A)\|$$

where $q(z)$ is a polynomial approximation of $f(z)$ on some region of $\mathbb{C}$ containing the eigenvalues of A. Note that both of our problems, obtaining bounds for the entries of $f(A)$ and bounding the error in the approximate solution of $Ax = b$ by Krylov subspace methods can be reduced to this one; in the latter case, we take $f(z) = z^{-1}$.

Recalling that

$$f(A) - q(A) = \frac{1}{2\pi i}\int_\Gamma (f(z) - q(z))(zI - A)^{-1}\mathrm{d}z,$$

and letting

$$\begin{aligned}\delta &= \sup_{z\in\Gamma} |f(z) - q(z)|,\\ L &= \frac{1}{2\pi} \times (\text{arclength of } \Gamma),\\ R &= \sup_{z\in\Gamma} \|(zI - A)^{-1}\|,\\ \sigma_{\min} &= R^{-1} = \inf_{z\in\Gamma} \sigma_n(zI - A),\end{aligned}$$

where $\sigma_n(zI - A)$ denotes the smallest singular value of $zI - A$, we obtain the bound

$$\|f(A) - q(A)\| \le L\, R\, \delta = \frac{L}{\sigma_{\min}}\, \delta.$$

In particular, if Γ is the boundary of the pseudospectrum $\Lambda_\epsilon(A)$, then $\sigma_{\min} = \epsilon$ and we get

$$\|f(A) - q(A)\| \le \frac{L}{\epsilon}\, \delta. \tag{9}$$

When A is normal, one can shrink the contours so that L/ϵ is arbitrarily close to 1, and thus the approximation error is given, in the limit as $\epsilon \to 0$, by

$$\delta = \max_{\lambda\in\Lambda(A)} |f(\lambda) - q(\lambda)|,$$

and we recover the fact that the eigenvalues suffice to fully describe the quality of the approximation.

If A is non-normal, however, we have to choose the contours (and thus ϵ) so as to balance the size of L with that of $R = \sigma_{\min}^{-1}$, which can be difficult. Nevertheless, there are cases where (9) can be used to obtain uniform error bounds, not containing the factor $\kappa(X)$, and thus applicable even if A is not diagonalizable.

Unfortunately, the need to choose a suitable value of ϵ and the fact that the geometry of the pseudospectra can be rather complicated make the use of this tool quite difficult in practice. Examples of successful uses of the pseudospectrum in a variety of problems in pure and applied mathematics, together with a discussion of its advantages and disadvantages, can be found in the (now classic) book [46]. We do not consider the pseudospectrum further, and move instead to the second alternative.

4 The field of values and some of its properties

If A is a bounded linear operator on a complex Hilbert space $\mathcal{H}$, the field of values (or *numerical range*) of A is the subset of $\mathbb{C}$ defined by

$$\mathcal{W}(A) = \{z = \langle Ax, x\rangle\,;\ \langle x, x\rangle = 1\}.$$

In other terms, $\mathcal{W}(A)$ is the range of the quadratic form $q(x) = \langle Ax, x\rangle$ as x varies over the unit sphere in $\mathcal{H}$. Depending on the problem, one may consider the field of values with respect to different inner products. When not explicitly indicated otherwise, we assune $\mathcal{H} = \mathbb{C}^n$ and the inner product will be the standard one.

Here are some properties of the field of values of a matrix $A \in \mathbb{C}^{n\times n}$:

1. Spectral containment: $\Lambda(A) \subset \mathcal{W}(A)$.

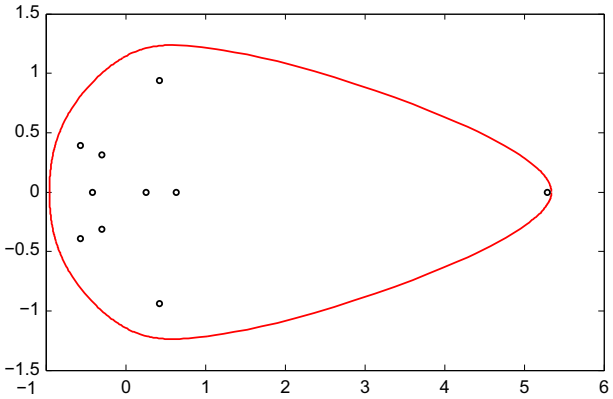

Fig. 3 The boundary of the field of values and the eigenvalues (small circles) of a random 10×10 matrix

2. $\|A\| \leq 2r(A)$, where $r(A) := \max\{|z| \,;\, z \in \mathcal{W}(A)\}$ is the *numerical radius* of A.
3. $\mathcal{W}(A) \subseteq D(0, \|A\|)$, the disk centered at 0 with radius $R = \|A\|$.
4. Subadditivity: $\mathcal{W}(A + B) \subseteq \mathcal{W}(A) + \mathcal{W}(B)$.
5. Translations: $\mathcal{W}(A + \alpha I) = \mathcal{W}(A) + \alpha$, for $\alpha \in \mathbb{C}$.
6. Scalings: $\mathcal{W}(\alpha A) = \alpha \mathcal{W}(A)$, for $\alpha \in \mathbb{C}$.
7. $\mathcal{W}(A)$ is compact.
8. Submatrix inclusion: $\mathcal{W}(A_k) \subseteq \mathcal{W}(A)$ for any principal submatrix A_k.
9. Unitary invariance: $\mathcal{W}(UAU^*) = \mathcal{W}(A)$, for any unitary matrix U.
10. Normal matrices: if A is normal, then $\mathcal{W}(A) = \text{co}(\Lambda(A))$ (the *convex hull* of $\Lambda(A)$).
11. Projection: $\Re(\mathcal{W}(A)) = \mathcal{W}(\frac{1}{2}(A + A^*))$ (a real interval).
12. Hausdorff–Toeplitz Theorem: $\mathcal{W}(A)$ is convex.

Several of this properties, but not all, retain meaning and remain true in infinite dimension. In particular, while the field of values of a bounded operator on an infinite-dimensional Hilbert space is bounded and convex, it may not be closed. We refer to [32] for detailed expositions of the properties of the field of values of $n \times n$ matrices, and to [28] for the operator case.

We also note that properties 3 and 4 together show that the field of values is stable under perturbations, in the sense that the field of values of a slightly perturbed matrix is a slight perturbation of the field of values of the original matrix.

In Fig. 3 we show the boundary of the field of values and the eigenvalues of a 10×10 matrix with randomly distributed entries in $(0, 1)$.

5 Functions of large, sparse matrices

Bounds on the entries of $f(A)$ for A banded, or sparse, can be obtained from bounds on the polynomial approximation error

$$\|f - p_N\|_{\infty,\mathcal{K}} = \max_{z \in \mathcal{K}} |f(z) - p_N(z)|, \quad N = 0, 1, , \ldots,$$

on a suitable compact set $\mathcal{K} \subset \mathbb{C}$. Here we assume that $\Lambda(A) \subset \mathcal{K}$ and that f is analytic on an open set containing $\mathcal{K}$.

Indeed, suppose A is k-banded and let p_N be the best approximation polynomial of degree N. Using the fact that $p_N(A)$ is kN-banded, it is possible to write

$$|[f(A)]_{ij}| = |[f(A)]_{ij} - [p_N(A)]_{ij}| \le \|f(A) - p_N(A)\|$$

for all i, j such that $|i - j| > kN$. Assume for a moment that there exist constants $C_0 > 0$, $\alpha > 0$ such that

$$\|f(A) - p_N(A)\| \le C_0 \, \mathrm{e}^{-\alpha(N+1)}, \quad N = 0, 1, \ldots \tag{10}$$

For $i \neq j$ we can write $|i - j| = kN + \ell$, $\ell = 1, 2, \ldots, k$. Observing that $|i - j| > kN$ implies $N + 1 < \frac{|i-j|}{k} + 1$, we can write for all $i \neq j$

$$|[f(A)]_{ij}| \le C_0 \, \mathrm{e}^{-\alpha\left(\frac{|i-j|}{k}+1\right)} = C \, \mathrm{e}^{-\alpha'|i-j|},$$

where $C = C_0 \mathrm{e}^{-\alpha}$, $\alpha' = \alpha/k$, i.e., an exponential off-diagonal decay bound.

Hence, we need to find a suitable set $\mathcal{K}$ such that (10) holds, and obtain explicit expressions for the constants C_0 and α. In particular, we seek bounds not containing the condition number of the eigenvector matrix; indeed, we do not want to assume that A is diagonalizable. Moreover, as already mentioned, we are especially interested in bounds that are independent of the dimension n, when possible.

For A Hermitian (more generally, normal), such bounds have been given in [8,11,12]. In these papers the solution is obtained through *Bernstein's Theorem* combined with the Spectral Theorem. Bernstein's Theorem states that if $\mathcal{K} \subset \mathbb{C}$ is a continuum (a nonempty, compact, connected set not reduced to a point) and f is analytic on an open subset Ω with $\mathcal{K} \subset \Omega$, then f can be approximated uniformly on $\mathcal{K}$ by a sequence of polynomials p_N such that the approximation error $\|f - p_N\|_{\infty,\mathcal{K}}$ decays at least exponentially in the degree, N (and viceversa). As it turns out, the p_N can be taken to be *Faber polynomials*.

The Spectral Theorem for normal matrices allows one to translate this result into the corresponding exponential decay bound for $[f(A)]_{ij}$ via the inequalities

$$|[f(A)]_{ij}| \le \|f(A) - p_N(A)\| \le \|f - p_N\|_{\infty,\mathcal{K}} \le C_0 \, \mathrm{e}^{-\alpha(N+1)} \le C \, \mathrm{e}^{-\alpha'|i-j|},$$

with C and α' as described above. Both C_0 and α (and thus C and α') depend on the choice of $\mathcal{K}$; taking a larger $\mathcal{K}$ makes both C and α' larger, as already mentioned; hence, there is a trade-off.

These results have been extended to the non-normal case by the author and Boito in [7] and, more recently, by Pozza and Simoncini in [39]. Specifically, if A is a banded normal matrix and f is analytic in the interior of $\mathcal{W}(A)$ and bounded on the boundary $\partial\mathcal{W}(A)$, then an exponential off-diagonal decay bound can be established for the entries of $f(A)$. A similar bound holds for sparse matrices with the geodesic distance on the graph of A replacing the distance from the main diagonal. Moreover, these results hold not just for functions of matrices over the complex field, but more generally for functions of matrices with entries in any complex C*-algebra.

The proof given in [7] was obtained combining Bernstein's Theorem and the following deep theorem of Crouzeix's:

Theorem 1 [15] *Let $A \in \mathbb{C}^{n\times n}$ and let f be analytic in the interior of $\mathcal{W}(A)$ and bounded on its boundary. There exists a universal constant $\mathcal{Q}$ such that*

$$\|f(A)\| \le \mathcal{Q} \sup_{z\in\mathcal{W}(A)} |f(z)|.$$

The constant $\mathcal{Q}$ satisfies $2 \le \mathcal{Q} \le 11.08$ and it is conjectured that $\mathcal{Q} = 2$. Moreover, the same result applies to analytic functions of bounded linear operators on a complex Hilbert space $\mathcal{H}$.

Recently, the upper bound on $\mathcal{Q}$ has been lowered to $1 + \sqrt{2}$ in [17]. Whether $\mathcal{Q} = 2$ remains an open question. The bounds in [7] contain the constant $\mathcal{Q}$, which can be taken to be equal to $1 + \sqrt{2}$.

The results of Pozza and Simoncini do not make use of Crouzeix's Theorem but instead rely on a result of Beckermann [4]. In both approaches, a key role in the analysis is played by Faber polynomials, which are briefly introduced next. For more details we refer to [21,36,44].

Recall that a *continuum* is any compact, connected set not reduced to a point. If $\mathcal{K}$ is a continuum with connected complement, the Riemann Mapping Theorem guarantees the existence of a function ϕ that maps the exterior of $\mathcal{K}$ conformally onto the set $\{z \in \mathbb{C}\,;\ |z| > 1\}$ and such that

$$\phi(\infty) = \infty, \quad \lim_{z\to\infty} \frac{\phi(z)}{z} = \rho > 0.$$

Such ϕ has the Laurent expansion

$$\phi(z) = \rho\, z + a_0 + \frac{a_1}{z} + \frac{a_2}{z^2} + \cdots$$

Furthermore, for every $N > 0$ we have

$$[\phi(z)]^N = \rho^N \left[z^N + \alpha_{N-1}^{(N)} z^{N-1} + \cdots + \alpha_0^{(N)} + \frac{\alpha_1^{(N)}}{z} + \cdots \right].$$

The polynomial parts,

$$F_N(z) = \rho^N \left[z^N + \alpha_{N-1}^{(N)} z^{N-1} + \cdots + \alpha_0^{(N)} \right],$$

are called the *Faber polynomials generated by the continuum* $\mathcal{K}$. The constant ρ is called the *logarithmic capacity* of $\mathcal{K}$.

Let $\mathcal{K} \subset \mathbb{C}$ be a continuum. As shown by Faber [24], every analytic function f defined on $\mathcal{K}$ can be expanded in the series

$$f(z) = \sum_{N=0}^{\infty} f_N F_N(z)$$

(uniformly convergent on $\mathcal{K}$), where the coefficients are given by

$$f_N = \frac{1}{2\pi i} \int_{|z|=\tau} \frac{f(\phi^{-1}(z))}{z^{N+1}}\, \mathrm{d}z.$$

Here $\tau > 1$ is chosen such that f is analytic on the complement of the set $\{\phi^{-1}(z)\,;\ |z| > \tau\}$ and ϕ maps the exterior of $\mathcal{K}$ conformally onto the set $\{z \in \mathbb{C}\,;\ |z| > 1\}$.

If $A \in \mathbb{C}^{n\times n}$ has spectrum contained in $\mathcal{K}$, then

$$f(A) = \sum_{N=0}^{\infty} f_N F_N(A).$$

Moreover, we have the following important result by Beckermann [4], the proof of which employs ideas from potential theory.

Theorem 2 [4] *Let $\mathcal{K} \subset \mathbb{C}$ be convex and compact. If $A \in \mathbb{C}^{n\times n}$ is such that*

$$\mathcal{W}(A) \subseteq \mathcal{K},$$

then the Faber polynomials generated by $\mathcal{K}$ satisfy $\|F_N(A)\| \leq 2$, for all N. The constant 2 is optimal.

Using this theorem, Pozza and Simoncini [39] obtained the following off-diagonal decay bound. We include the short and elegant proof for completeness.

Theorem 3 [39] *Let $A \in \mathbb{C}^{n\times n}$ be k-banded and such that $\mathcal{W}(A) \subseteq \mathcal{K}$, with $\mathcal{K}$ compact and convex. With ϕ and $\tau > 1$ defined as before, we have*

$$|[f(A)]_{ij}| \leq 2\,\frac{\tau}{\tau-1}\,\max_{|z|=\tau} |f(\phi^{-1}(z))| \left(\frac{1}{\tau}\right)^{\xi},$$

where

$$\xi = \lceil |i-j|/k \rceil.$$

Proof Since $[A^N]_{ij} = 0$ for $N < \xi$, we have

$$|[f(A)]_{ij}| = \left|\sum_{N=0}^{\infty} f_N [F_N(A)]_{ij}\right| = \left|\sum_{N=\xi}^{\infty} f_N [F_N(A)]_{ij}\right| \leq 2 \sum_{N=\xi}^{\infty} |f_N|$$

by Beckermann's Theorem. Using

$$f_N = \frac{1}{2\pi i} \int_{|z|=\tau} \frac{f(\phi^{-1}(z))}{z^{N+1}}\,\mathrm{d}z$$

we easily obtain

$$|f_N| \leq \left(\frac{1}{\tau}\right)^N \max_{|z|=\tau} |f(\phi^{-1}(z))|,$$

hence

$$|[f(A)]_{ij}| \leq 2 \max_{|z|=\tau} |f(\phi^{-1}(z))| \sum_{N=\xi}^{\infty} \left(\frac{1}{\tau}\right)^N = 2\frac{\tau}{\tau-1} \max_{|z|=\tau} |f(\phi^{-1}(z))| \left(\frac{1}{\tau}\right)^{\xi}.$$

□

A more precise statement is possible to account for matrices with lower bandwidth β and upper bandwidth γ with $\beta \neq \gamma$, see [39]. Moreover, the result can be extended to more general sparse matrices. Note, again, the trade-off involved in the choice of τ. If f is entire, τ can be arbitrarily large and the decay is superexponential.

When explicitly computing the bound, one can take $\mathcal{K} = \mathcal{W}(A)$, if the latter is known. For certain classes of matrices, $\mathcal{W}(A)$ itself is not known, but it is known to be bounded by some simple compact convex set, like an ellipse or a disk, which can be easily estimated. In some cases the corresponding bounds can be dramatically better than those containing the condition number of the eigenvector matrix. This is the case of families of $n \times n$ matrices such that $\kappa(X)$ grows unboundedly with the dimension n, while the field of values remains

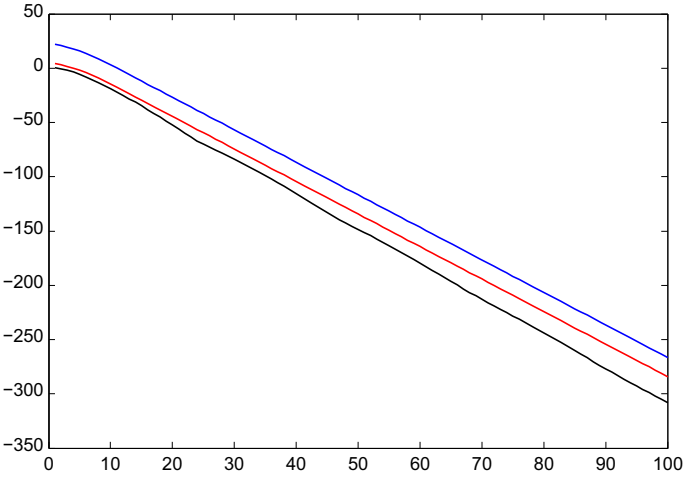

Fig. 4 Decay of the entries in the first row of the exponential of a non-normal tridiagonal matrix (black), together with the bounds depending on the eigenvectors (blue) and the field of values (red) (color figure online)

uniformly bounded. Consider for example the infinite tridiagonal Toeplitz matrix generated by the symbol $\varphi(z) = 2z^{-1} + 1 + 3z$, $|z| = 1$:

$$A = \begin{bmatrix} 1 & 3 & & & \\ -2 & 1 & 3 & & \\ & -2 & 1 & 3 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \tag{11}$$

The matrix represents a bounded linear operator on $\ell^2(\mathbb{N})$. Let A_n denote the finite section of A of dimension n, i.e., the $n \times n$ matrix formed by the first n rows and columns of A. Then all the fields of values $\mathcal{W}(A_n)$ are regions whose boundaries are ellipses, see [19, Corollary 4]. As $n \to \infty$, these ellipses converge to an ellipse which contains all the $\mathcal{W}(A_n)$ and is the boundary of $\mathcal{W}(A)$, therefore the fields of values of A_n are all uniformly bounded in n. In contrast, the condition number of the eigenbasis $\kappa(X_n)$ grows exponentially with n. Note that the infinite matrix (11) has no point spectrum, hence no eigenvectors in $\ell^2(\mathbb{N})$.

In Fig. 4, we illustrate the decay behavior of the order of magnitude of the entries in the first row of $f(A_n) = \mathrm{e}^{A_n}$ for $n = 100$ (black plot), together with the bounds obtained using the field of values (red) and the one containing $\kappa(X)$ (blue). Note the logarithmic scale on the vertical axis. This example shows that the eigenbasis-dependent bounds can overestimate the magnitude of the entries by many orders of magnitude, while the bounds based on the field of values can result in much more accurate estimates, especially at short distance from the main diagonal. For this matrix, the eigenbasis condition number is $\kappa(X) \approx 5.26 \cdot 10^8$. Taking larger values of n will make the eigenbasis-dependent bound much worse, while the field of values-dependent bound remains unchanged. This example can be easily generalized and extended.

Finally, we mention that while we have focused here on the derivation of bounds for the entries of $f(A)$, nearly identical considerations apply to the problem of polynomial (and also rational) approximations for computing the action of a function of a matrix on a vector, $v = f(A)b$; see, for instance, [5,48].

6 Convergence of Krylov methods for saddle point problems

In this section we review some convergence bounds for GMRES based on the field of values, and show how they lead to mesh-independent estimates of the rate of convergence of preconditioned GMRES applied to saddle point problems.

6.1 Field of values bounds for GMRES

The Generalized Minimal Residual (GMRES) method [40,41] is the most widely used algorithm for the solution of large, sparse, nonsymmetric systems of linear equations $Ax = b$. Starting from an initial guess x_0, GMRES constructs approximations x_k to the solution $x_* = A^{-1}b$ ($k = 0, 1, \ldots$) such that the kth residual vector $r_k = b - Ax_k$ satisfies

$$\frac{\|r_k\|}{\|r_0\|} = \min \left\{ \frac{\|p(A)r_0\|}{|p(0)|\|r_0\|} ;\ p \in \mathbb{C}[x],\ \deg(p) \le k \right\},$$

where $\deg(p)$ is the degree of the polynomial p. Using $\|p(A)r_0\| \le \|p(A)\|\|r_0\|$, we easily obtain the bound

$$\frac{\|r_k\|}{\|r_0\|} \le \min \left\{ \frac{\|p(A)\|}{|p(0)|} ;\ p \in \mathbb{C}[x],\ \deg(p) \le k \right\},$$

which no longer depends on b or r_0.

Over the years, there have been many attempts to derive descriptive error bounds for GMRES analogous to those available for MINRES or CG. This is a difficult task, see for example [23]. Results are known for matrices A such that $A + A^*$ is positive definite, see [20] (see also [40]). More generally, if $0 \notin \mathcal{W}(A)$, there are field of values-based bounds due to Eiermann [19] and to Beckermann [4], among others. The latter one is given next.

Theorem 4 [4] *Let $A \in \mathbb{C}^{n\times n}$ and let $\mathcal{K} \subset \mathbb{C}$ be convex, compact, and such that $\mathcal{W}(A) \subseteq \mathcal{K}$ and $0 \notin \mathcal{K}$. Let ϕ be the map in the statement of Theorem 3. Then the GMRES residuals satisfy*

$$\frac{\|r_k\|}{\|r_0\|} \le \left(\frac{2}{1 - \gamma_{\mathcal{K}}^{k+1}} \right) \gamma_{\mathcal{K}}^k, \quad k = 0, 1, \ldots,$$

where $\gamma_{\mathcal{K}} = \frac{1}{|\phi(0)|} < 1$.

Suppose now that we have a family of linear systems, $A_\nu x_\nu = b_\nu$, depending on a parameter ν. Here ν could be a physical parameter, such as the viscosity in a discretized convection-diffusion equation, or the dimension of the linear system, corresponding to finer and finer discretizations of some differential or integral operator. Of particular interest is the case where $\nu = O(h)$, where h is a discretization parameter. We have the following simple consequence of Beckermann's result:

Corollary 1 *Let $\mathcal{K} \subset \mathbb{C}$ be convex, compact, and such that*

$$\bigcup_\nu \mathcal{W}(A_\nu) \subseteq \mathcal{K}, \quad 0 \notin \mathcal{K}.$$

Then GMRES converges to the solution of each of the linear systems $A_\nu x_\nu = b_\nu$ in a number of steps that is bounded uniformly in ν.

In practice, this result will be applied not to the original linear system $A_\nu x_\nu = b_\nu$ but to a preconditioned version. Indeed, apart from very special situations, preconditioning is usually necessary to achieve ν-independent convergence. We turn to preconditioning next.

6.2 Field of values equivalence

We begin by reviewing the notion of spectral equivalence for families of Hermitian positive definite (HPD) matrices [3]. Recall that two families of HPD matrices $\{A_h\}$ and $\{B_h\}$ are said to be *spectrally equivalent* if there exist h-independent constants α and β with

$$0 < \alpha \leq \lambda_i(B_h^{-1}A_h) \leq \beta, \quad \forall i.$$

Equivalently, $\{A_h\}$ and $\{B_h\}$ are spectrally equivalent if the spectral condition number $\kappa(B_h^{-1}A_h)$ is uniformly bounded with respect to h.

Yet another equivalent condition is that the *generalized Rayleigh quotients* associated with A_h and B_h are uniformly bounded:

$$0 < \alpha \leq \frac{\langle A_h x, x\rangle}{\langle B_h x, x\rangle} \leq \beta, \quad \forall x \neq 0.$$

Note that this is an equivalence relation between families of matrices.

If the discretization of (say) an elliptic PDE leads to a sequence of linear systems $A_h u_h = b_h$, a family of spectrally equivalent preconditioners $\{B_h\}$ guarantees that the Preconditioned Conjugate Gradient (PCG) method will converge in a number of steps that is uniformly bounded with respect to the parameter h . If h denotes some measure of the mesh size (discretization parameter), the resulting PCG iteration exhibits *mesh-independent convergence*. If, in addition, the cost of applying the preconditioner B_h is linear in the number of degrees of freedom, we say that the preconditioner is *optimal* with respect to the mesh size h. In general, of course, the actual performance of the preconditioner can be affected by other factors, such as physical parameters. Good general references for the PCG method for the solution of discretized PDEs include [3,22,38].

When the preconditioned system is not symmetrizable with positive eigenvalues, for example because the preconditioner is indefinite or non-symmetric, then spectral equivalence is no longer the appropriate tool to analyze the convergence of preconditioned Krylov methods, and PCG cannot be applied. In this case, the more general concept of *field of values equivalence*, first proposed by G. Starke [43], can in some cases provide the theoretical framework needed to establish mesh-independent convergence for certain preconditioners for Krylov methods like GMRES. Examples include preconditioners for convection-diffusion equations [43], block preconditioners for the Stokes system and other problems of saddle point type [14,22,33,35], preconditioners for the incompressible linearized Navier–Stokes equations [10] and for Rayleigh–Bénard convection [2]. Field of values equivalence has also been applied to the analysis of preconditioned iterative solvers applied to discretizations of the Helmholtz equation; see, e.g., [25,29]. Finally, we refer to [37] for recent work on the use of the field of values to study the convergence of a class of two-grid iterative methods.

For reasons of space we can only give a very succinct overview of how field of values equivalence may be used to obtain h-independent convergence bounds for preconditioned GMRES applied to large linear systems in saddle point form, i.e.,

$$\mathcal{A}\mathbf{x} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} = \mathbf{f}.$$

Such systems arise frequently from the finite element discretization of boundary value problems for systems of PDEs. Examples include mixed formulations of the Poisson equation [13], the Stokes equations, the Oseen problem (obtained from the Navier–Stokes equations via Picard linearization) [22], or the coupled Stokes–Darcy system [14]. In most cases the matrix A is symmetric positive definite and B is rectangular and has full row rank [9].

We assume that the matrix $\mathcal{A} \in \mathbb{R}^{n\times n}$ satisfies the following (Babuška–Brezzi) boundedness and stability conditions:

$$\sup_{\mathbf{w}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \sup_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{w}^T\mathcal{A}\mathbf{v}}{\|\mathbf{w}\|_H\|\mathbf{v}\|_H} \le c_1, \tag{12a}$$

$$\inf_{\mathbf{w}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \sup_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{w}^T\mathcal{A}\mathbf{v}}{\|\mathbf{w}\|_H\|\mathbf{v}\|_H} \ge c_2, \tag{12b}$$

where c_1 and c_2 are positive constants independent of n, and the vector H-norm is defined by $\|x\|_H = (\langle Hx, x\rangle)^{\frac{1}{2}}$, where the matrix H is symmetric positive definite (SPD). A typical choice of H for finite element discretizations of incompressible flow problems is

$$H = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix}, \quad H_1 = \text{ discrete vector Laplacian}, \quad H_2 = M_p,$$

where M_p denotes the mass matrix for the pressure space.

Definition 1 Two nonsingular matrices $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n\times n}$ are said to be *H-field-of-values equivalent*, $\mathcal{A}\approx_H\mathcal{B}$, if there exist constants $\alpha_0 > 0$ and $\beta_0 > 0$ independent of n such that the following holds for all nonzero $\mathbf{x} \in \mathbb{R}^n$:

$$\alpha_0 \le \frac{\langle \mathcal{A}\mathcal{B}^{-1}\mathbf{x}, \mathbf{x}\rangle_H}{\langle \mathbf{x}, \mathbf{x}\rangle_H} \quad \text{and} \quad \left\|\mathcal{A}\mathcal{B}^{-1}\right\|_H \le \beta_0$$

We say in short that $\mathcal{A}$ and $\mathcal{B}$ are *FoV-equivalent*. Note that FoV-equivalence implies that the eigenvalues of $\mathcal{A}\mathcal{B}^{-1}$ are uniformly bounded: $\alpha_0 \le |\lambda_i(\mathcal{A}\mathcal{B}^{-1})| \le \beta_0$. The converse, however, is not true: FoV-equivalence is generally stronger than the condition that all the matrices $\mathcal{A}\mathcal{B}^{-1}$ have spectra that are uniformly bounded with respect to the dimension n. If, however, $\mathcal{A}$ and $\mathcal{B}$ are SPD and $H = I_n$, FoV-equivalence reduces to spectral equivalence. We also note that in the general case, FoV-equivalence is not an equivalence relation. We refer to [33,35,43] for details.

Introducing again the subscript h to denote dependence on the discretization parameter h (and therefore on the dimension n), we have the following: if a family of preconditioners $\{\mathcal{B}_h\}$ is H-FoV equivalent to a family of saddle point matrices $\{\mathcal{A}_h\}$, the H-FoVs of the preconditioned matrices $\mathcal{A}_h\mathcal{B}_h^{-1}$ lie in the right-half plane and are bounded independently of h. As a result, Krylov subspace methods like MINRES or GMRES converge at a rate that is h-independent. In the case of GMRES, this follows for instance from Theorem 4.

We mention that the use of the H-FoV implies that the GMRES residual convergence should be measured either in the H-norm for left preconditioning or in the H^{-1}-norm for right preconditioning. In finite element computations, the natural norm is the H^{-1}-norm, and it can be shown that h-independent convergence in this norm of the preconditioned Krylov method implies h-independent convergence in the standard Euclidean norm as well, see [1,22].

Generally speaking, showing FoV-equivalence for a given family of saddle point problems and a corresponding family of preconditioners is non-trivial. Nevertheless, it has been possible to establish it in the following important cases:

1. Block triangular preconditioners based on approximate Schur complements for the Stokes and Oseen problems [33];
2. Block diagonal preconditioning of Darcy's equations [35];
3. Augmented Lagrangian preconditioning of the Oseen problem [10];
4. Constraint preconditioning of the coupled Stokes–Darcy system [14];
5. Block triangular preconditioning of the Rayleigh–Bénard system [2].

We refer interested readers to the cited literature for details.

7 Conclusions

In this expository paper we have illustrated how the field of values has been used in the study of some important problems in numerical analysis, from the approximation of matrix functions to the convergence analysis of preconditioned GMRES for solving large-scale linear systems. While we have not discussed the actual numerical computation of the field of values of a matrix, which is a challenging task in the case of matrices of very large size, we have shown how a priori knowledge of certain properties of the field of values may be sufficient to prove certain useful bounds and even to obtain optimality results for a class of preconditioners for a given problem. Briefly stated, the fields of values must remain bounded and bounded away from any singularities of the underlying function, uniformly in the parameter of interest (which is often, but not always, the matrix dimension n).

Of course, the field of values is no panacea, and approaches based on it will fail if it contains any singularities of the underlying scalar function; for the convergence analysis of GMRES the function is $f(z) = z^{-1}$, and the field of values is useless if it contains the origin. Nevertheless, in this case it may still be possible to identify a *C-spectral set*, i.e., a subset $\mathcal{S}$ of the complex plane satisfying $\Lambda(A) \subset \mathcal{S} \subset \mathcal{W}(A)$, not containing 0 (or, more generally, any singularities of the function f), and such that

$$\|g(A)\| \leq C \sup_{z \in \mathcal{S}} |g(z)|$$

for all rational functions g bounded on $\mathcal{S}$, where C is a universal constant. We refer to [16] for some examples illustrating this technique. It is, however, too early to say if this approach can be successfully applied to prove convergence bounds for the preconditioned GMRES method in realistic applications.

Acknowledgements Open access funding provided by Scuola Normale Superiore within the CRUI-CARE Agreement. This paper faithfully represents the contents of a plenary lecture delivered by the author on the occasion of the XXI Congress of the Unione Matematica Italiana, held in Pavia on 2–7 September 2019. The author would like to express his gratitude to the Scientific Committee of UMI for the invitation. Thanks are due also to an anonymous referee for useful suggestions.

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

References

1. Arioli, M., Noulard, E., Russo, A.: Stopping criteria for iterative methods: applications to PDEs. Calcolo **38**, 97–112 (2001)
2. Aulisa, E., Bornia, G., Howle, V., Ke, G.: Field-of-values analysis of preconditioned linearized Rayleigh–Bénard convection problems. J. Comput. Appl. Math. **369**, Article 112582 (2020)
3. Axelsson, O., Barker, V.A.: Finite Element Solution of Boundary Value Problems: Theory and Computation, SIAM Classics in Applied Mathematics, vol. 35. Society for Industrial and Applied Mathematics, Philadelphia (2001)
4. Beckermann, B.: Image numérique, GMRES et polynomes de Faber. C. R. Acad. Sci. Paris Ser. **I**(340), 855–860 (2005)
5. Beckermann, B., Reichel, L.: Error estimates and evaluation of matrix functions via the Faber transform. SIAM J. Numer. Anal. **47**, 3849–3883 (2009)
6. Benzi, M.: Localization in matrix computations: theory and applications. In: Benzi, M., Simoncini, V. (eds.) Exploiting Hidden Structure in Matrix Computations: Algorithms and Applications (Cetraro, Italy, 2015), Lecture Notes in Mathematics 2173, pp. 211–317. Springer, Cham (2016)
7. Benzi, M., Boito, P.: Decay properties for functions of matrices over C*-algebras. Linear Algebra Appl. **456**, 174–198 (2014)
8. Benzi, M., Golub, G.H.: Bounds for the entries of matrix functions with applications to preconditioning. BIT **39**, 417–438 (1999)
9. Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. Acta Numer. **14**, 1–137 (2005)
10. Benzi, M., Olshanskii, M.A.: Field-of-values convergence analysis of augmented Lagrangian preconditioners for the linearized Navier-Stokes problem. SIAM J. Numer. Anal. **49**, 770–788 (2011)
11. Benzi, M., Razouk, N.: Decay rates and $O(n)$ algorithms for approximating functions of sparse matrices. Electr. Trans. Numer. Anal. **28**, 16–39 (2007)
12. Benzi, M., Simoncini, V.: Decay bounds for functions of Hermitian matrices with banded or Kronecker structure. SIAM J. Matrix Anal. Appl. **36**, 1263–1282 (2015)
13. Boffi, D., Brezzi, F., Fortin, M.: Mixed Finite Element Methods and Applications. Springer, Berlin (2013)
14. Chidyagwai, P., Ladenheim, S., Szyld, D.B.: Constraint preconditioning for the coupled Stokes–Darcy system. SIAM J. Sci. Comput. **38**, A668–A690 (2016)
15. Crouzeix, M.: Numerical range and functional calculus in Hilbert space. J. Funct. Anal. **244**, 668–690 (2007)
16. Crouzeix, M., Greenbaum, A.: Spectral sets: numerical range and beyond. SIAM J. Matrix Anal. Appl. **40**, 1087–1101 (2019)
17. Crouzeix, M., Palencia, C.: The numerical range is a $(1+\sqrt{2})$-spectral set. SIAM J. Matrix Anal. Appl. **38**, 649–655 (2017)
18. Davis, T.: SuiteSparse: A Suite of Sparse Matrix Software. http://faculty.cse.tamu.edu/davis/suitesparse.html. Accessed 7 June 2020
19. Eiermann, M.: Fields of values and iterative methods. Linear Algebra Appl. **180**, 167–197 (1993)
20. Eisenstat, S.C., Elman, H.C., Schultz, M.H.: Variational iterative methods for nonsymmetric systems of linear equations. SIAM J. Numer. Anal. **20**, 345–357 (1983)
21. Ellacott, S.W.: Computation of Faber series with application to numerical polynomial approximation in the complex plane. Math. Comput. **40**, 575–587 (1983)
22. Elman, H.C., Silvester, D., Wathen, A.J.: Finite Elements and Fast Iterative Solvers, 2nd edn. Oxford University Press, UK (2014)
23. Embree, M.: How Descriptive are GMRES Convergence Bounds? Tech. Rep. NA-99-08. University of Oxford (1999)
24. Faber, G.: Über polynomische Entwickelungen. Math. Annalen **57**, 389–408 (1903)
25. Gander, M.J., Graham, I.G., Spence, E.A.: Applying GMRES to the Helmholtz equation with shifted Laplacian preconditioning: what is the largest shift for which wavenumber-independent convergence is guaranteed? Numer. Math. **131**, 567–614 (2015)
26. Greenbaum, A.: Iterative Methods for Solving Linear Systems. Society for Industrial and Applied Mathematics, Philadelphia (1997)

27. Greenbaum, A., Pták, V., Strakoš, Z.: Any nonincreasing convergence curve is possible for GMRES. SIAM J. Matrix Anal. Appl. **17**, 465–469 (1996)
28. Gustafson, K.E., Rao, D.K.M.: Numerical Range. The Field of Values of Linear Operators and Matrices. Universitext Springer, Berlin (1997)
29. Hannukainen, A.: Field of values analysis of a two-level preconditioner for the Helmholtz equation. SIAM J. Numer. Anal. **51**, 1567–1584 (2013)
30. Hausdorff, F.: Der Wertvorrat einer bilinear Form. Math. Z. **3**, 314–316 (1919)
31. Higham, N.J.: Functions of Matrices: Theory and Computation. Society for Industrial and Applied Mathematics, Philadelphia (2008)
32. Horn, R.A., Johnson, C.A.: Topics in Matrix Analysis. Cambridge University Press, Cambridge (1991)
33. Klawonn, A., Starke, G.: Block triangular preconditioners for nonsymmetric saddle point problems: field-of-values analysis. Numer. Math. **81**, 577–594 (1999)
34. Liesen, J., Strakoš, Z.: Krylov Subspace Methods: Principles and Analysis. Oxford University Press, UK (2013)
35. Loghin, D., Wathen, A.J.: Analysis of preconditioners for saddle-point problems. SIAM J. Sci. Comput. **25**, 2029–2049 (2004)
36. Markushevich, A.I.: Theory of Functions of a Complex Variable, vol. III. Prentice-Hall, Englewood Cliffs (1967)
37. Notay, Y.: Analysis of two-grid methods: the nonnormal case. Math. Comput. **89**, 807–827 (2020)
38. Olshanskii, M.A., Tyrtyshnikov, E.E.: Iterative Methods for Linear Systems: Theory and Applications. Society for Industrial and Applied Mathematics, Philadelphia (2014)
39. Pozza, S., Simoncini, V.: Inexact Arnoldi residual estimates and decay properties for functions of non-Hermitian matrices. BIT **59**, 969–986 (2019)
40. Saad, Y.: Iterative Methods for Sparse Linear Systems, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia (2003)
41. Saad, Y., Schultz, M.H.: GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comput. **7**, 856–859 (1986)
42. Schimmel, C.: Bounds in the decay of matrix functions and its exploitation in matrix computations. PhD thesis, Bergische Universität Wuppertal, Fakultät für Mathematik und Naturwissenschaften (2019)
43. Starke, G.: Field of values analysis of preconditioned iterative methods for nonsymmetric elliptic problems. Numer. Math. **78**, 103–117 (1997)
44. Suetin, P.K.: Series of Faber Polynomials. Gordon and Breach Science Publishers, Amsterdam (1998). (Translated from the 1984 Russian original by E. V. Pankratiev)
45. Toeplitz, O.: Das algebraische Analogon zu einem Satz von Fejér. Math. Z. **2**, 187–197 (1918)
46. Trefethen, L.N., Embree, M.: Spectra and Pseudospectra. The Behavior of Nonnormal Matrices and Operators. Princeton University Press, Princeton (2005)
47. Varga, R.S.: Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs (1962)
48. Wang, H., Ye, Q.: Error bounds for the Krylov subspace methods for computation of matrix exponentials. SIAM J. Numer. Anal. **38**, 155–187 (2017)