Technical Report

Localization in Matrix Computations: Theory and Applications

by

Michele Benzi

MATHEMATICS AND COMPUTER SCIENCE

EMORY UNIVERSITY

Localization in Matrix Computations: Theory and Applications

Michele Benzi

Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA. Email: `benzi@mathcs.emory.edu`

Summary. Many important problems in mathematics and physics lead to (non-sparse) functions, vectors, or matrices in which the fraction of nonnegligible entries is vanishingly small compared the total number of entries as the size of the system tends to infinity. In other words, the nonnegligible entries tend to be *localized*, or concentrated, around a small region within the computational domain, with rapid decay away from this region (uniformly as the system size grows). When present, localization opens up the possibility of developing fast approximation algorithms, the complexity of which scales linearly in the size of the problem. While localization already plays an important role in various areas of quantum physics and chemistry, it has received until recently relatively little attention by researchers in numerical linear algebra. In this chapter we survey localization phenomena arising in various fields, and we provide unified theoretical explanations for such phenomena using general results on the decay behavior of matrix functions. We also discuss computational implications for a range of applications.

1 Introduction

In numerical linear algebra, it is common to distinguish between *sparse* and *dense* matrix computations. An $n \times n$ sparse matrix A is one in which the number of nonzero entries is much smaller than n^2 for n large. It is generally understood that a matrix is dense if it is not sparse.[1] These are not, of course, formal definitions. A more precise definition of a sparse $n \times n$ matrix, used by some authors, requires that the number of nonzeros in A is $O(n)$ as $n \to \infty$. That is, the average number of nonzeros per row must remain bounded by a constant for large n. Note that this definition does not apply to a single matrix, but to a family of matrices parameterized by the dimension, n. The definition can be easily adapted to the case of non-square matrices, in particular to vectors.

The latter definition, while useful, is rather arbitrary. For instance, suppose we have a family of $n \times n$ matrices in which the number of nonzero entries behaves like $O(n^{1+\varepsilon})$ as $n \to \infty$, for some $\varepsilon \in (0, 1)$. Clearly, for such matrix family the fraction of nonzero entries vanishes as $n \to \infty$, and yet such matrices would not be regarded as sparse according to this definition.[2]

---

[1] Note that we do not discuss here the case of *data-sparse matrices*, which are thoroughly treated elsewhere in this book.

[2] Perhaps a better definition is the one given in [73, page 1]: "A matrix is **sparse** if there is an advantage in exploiting its zeros."

Another limitation of the usual definition of sparsity is that it does not take into account the *size* of the nonzeros. All nonzeros are treated as equals: a matrix is either sparse or not sparse (dense). As we shall see, there are many situations in computational practice where one encounters vectors or matrices in which virtually every entry is nonzero, but only a very small fraction of the entries has nonnegligible magnitude. A matrix of this kind is close to being sparse: it would become truly sparse (according to most definitions) upon *thresholding*, or *truncation* (i.e., the setting to zero of matrix elements smaller than a prescribed, sufficiently small quantity in absolute value). However, this assumes that entries are first computed, then set to zero if small enough, which could be an expensive and wasteful task. Failing to recognize this may lead to algorithms with typical $O(n^2)$ or $O(n^3)$ scaling for most matrix computation tasks. In contrast, careful exploitation of this property can lead to *linear scaling algorithms*, i.e., approximation algorithms with $O(n)$ computational complexity (in some cases even sublinear complexity may be possible). One way to accomplish this is to derive *a priori* bounds on the size of the elements, so as to know in advance which ones *not* to compute.

Matrices with the above-mentioned property are often referred to as being *localized*, or to exhibit *decay*.[3] These terms are no more precise than the term "sparse" previously discussed, and one of the goals of these lectures is to provide precise formalizations of these notions. While the literature on sparse matrix computations is enormous, much less attention has been devoted by the numerical linear algebra community to the exploitation of localization in computational problems; it is our hope that these lectures will attract some interest in this interesting and important property, which is well known to computational physicists and chemists.

Just as sparse matrices are often structured, in the sense that the nonzeros in them are usually not distributed at random, so are localized matrices and vectors. The entries in them typically fit some type of decay behavior, such as exponential decay, away from certain clearly defined positions, for example the main diagonal. Many important computational problems admit localized solutions, and identifying this *hidden structure* (i.e., being able to predict the decay properties of the solution) can lead to efficient approximation algorithms. The aim of these lectures is to provide the reader with the mathematical background and tools needed to understand and exploit localization in matrix computations.

We now proceed to give a brief (and by no means complete) overview of localization in physics and in numerical mathematics. Some of these examples will be discussed in greater detail in later sections.

1.1 Localization in physics

Generally speaking, the term *locality* is used in physics to describe situations where the strength of interactions between the different parts of a system decay rapidly with the distance: in other words, correlations are *short-ranged*. Mathematically, this fact is expressed by saying that some function $\phi(\mathbf{r}, \mathbf{r}')$ decays rapidly to zero as the spatial separation $\|\mathbf{r} - \mathbf{r}'\|$ increases. The opposite of localization is *delocalization*: a function is delocalized if its values are nonnegligible on an extended region. In other words, if non-local (long-range) interactions are important, a system is delocalized. Locality (or lack of it) is of special importance in quantum chemistry and

---

[3] Occasionally, the term *pseudosparse* is used; see, e.g., [34].

solid state physics, since the properties of molecules and the behavior of materials are strongly dependent on the presence (or absence) of localization.

Recall that in quantum mechanics the stationary states of a system of N particles are described by wave functions, $\Psi_n \in L^2(\mathbb{R}^{3N})$, $n = 0, 1, \ldots$, normalized so that $\|\Psi_n\|_{L^2} = 1$. These states are stationary in the sense that a system initially in state Ψ_n will remain in it if left unperturbed. The probability that a system in the stationary state corresponding to Ψ_n is in a configuration $\mathbf{x}$ belonging to a given region $\Omega \subseteq \mathbb{R}^{3N}$ is given by

$$\mathbf{Pr}\,(\text{system configuration } \mathbf{x} \in \Omega) = \int_{\Omega} |\Psi_n(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x}.$$

As an example, consider the electron in a hydrogen atom. We let $\mathbf{r} = (x, y, z) \in \mathbb{R}^3$ be the position of the electron with respect to the nucleus (supposed to be at the origin) and $r = \sqrt{x^2 + y^2 + z^2}$. The radial part $\psi_0(r)$ of the first atomic orbital, the wave function $\Psi_0(\mathbf{r}) \in L^2(\mathbb{R}^3)$ corresponding to the lowest energy (ground state), is a decaying exponential:

$$\psi_0(r) = \frac{1}{\sqrt{\pi}\, a_0^{3/2}} \, \mathrm{e}^{-r/a_0}, \quad r \geq a_0,$$

where (using Gaussian units) $a_0 = \frac{\hbar^2}{me^2} = 0.0529$ nm is the Bohr radius. Thus, the wave function is strongly localized in space (see Fig. 1, left). Localization of the wave function Ψ_0 expresses the fact that in the hydrogen atom at ground state, the electron is bound to a small region around the nucleus, and the probability of finding the electron at a distance r decreases rapidly as r increases.

The wave function Ψ_0 satisfies the (stationary) Schrödinger equation:

$$H\,\Psi_0 = E_0\,\Psi_0$$

where the operator H (using now atomic units) is given by

$$H = -\frac{1}{2}\Delta - \frac{1}{r} \qquad (\Delta = \text{Laplacian})$$

is the *Hamiltonian*, or energy, operator, and E_0 is the *ground state energy*. That is, the ground state Ψ_0 is the eigenfunction of the Hamiltonian corresponding to the lowest eigenvalue E_0.

Note that the Hamiltonian is of the form $H = T + V$ where

$$T = -\frac{1}{2}\Delta = \text{kinetic energy}$$

and

$$V = -\frac{1}{r} = (\text{Coulomb}) \text{ potential.}$$

What happens if the Coulomb potential is absent? In this case there is no force binding the electron to the nucleus: the electron is "free." This implies *delocalization*: there are no eigenvalues (the spectrum is purely continuous) and therefore no eigenfunctions in $L^2(\mathbb{R}^3)$. Another example is the following. Consider a particle confined to the interval $[0, L]$, then the eigenfunction corresponding to the smallest eigenvalue of the Hamiltonian $H = -\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ (with zero Dirichlet boundary conditions)
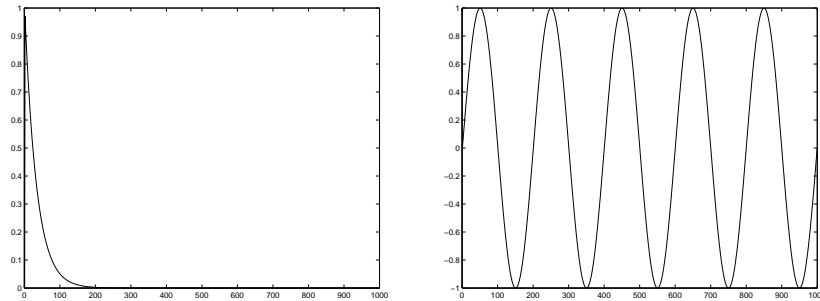
Fig. 1. Left: Localized eigenfunction. Right: Delocalized eigenfunction.

is given (up to a normalization factor) by $\Psi_0(x) = \sin\left(\frac{2\pi}{L}x\right)$, which is delocalized (see Fig. 1, right).

Consider now an extended system consisting of a large number of atoms, assumed to be in the ground state. Suppose the system is perturbed at one point space, for example by slightly changing the value of the potential V near some point $\mathbf{x}$. If the system is an insulator, then the effect of the perturbation will only be felt locally: it will not be felt outside of a small region. This "absence of diffusion" is also known as localization. W. Kohn [124, 170] called this behavior the "nearsightedness" of electronic matter. In insulators, and also in semi-conductors and in metallic systems under suitable conditions (such as room temperature), the electrons tend to stay put.

Localization is a phenomenon of major importance in quantum chemistry and in solid state physics. We will return on this in section 4.2, when we discuss applications to the electronic structure problem. Another important example is *Anderson localization*, which refers to the localization in systems described by Hamiltonians of the form $H = T + \gamma V$ where V is a random potential and $\gamma > 0$ a parameter that controls the "disorder strength" in the system [4]. Loosely speaking, once γ exceeds a certain threshold γ_0 the eigenfunctions of H abruptly undergo a transition from extended to localized with very high probability. Anderson localization is beyond the scope of the techniques discussed in these lectures. The interested reader is referred to [186] for a survey.

Locality (or lack thereof) is also of central importance in quantum information theory and quantum computing, in connection with the notion of entanglement of states [77].

1.2 Localization in numerical mathematics

In contrast to the situation in physics, the recognition of localization as an important property in numerical mathematics is relatively recent. It began to slowly emerge in the late 1970s and early 1980s as a results of various trends in numerical analysis, particularly in approximation theory (convergence properties of splines) and in numerical linear algebra. Researchers in these areas were the first to investigate the decay properties of inverses and eigenvectors of certain classes of banded

matrices; see [68, 69] and [60]. By the late 1990s, the decay behavior of the entries of fairly general functions of banded matrices had been analyzed [22, 117], and numerous papers on the subject have appeared since then. Figures 2-4 provide examples of localization for different matrix functions.

From a rather different direction, Banach algebras of infinite matrices with off-diagonal decay arising in computational harmonic analysis and other problems of a numerical nature were being investigated in the 1990s by Jaffard in France [119] and by Baskakov and others [12, 13, 34] in the former Soviet Union. In particular, much effort has been devoted to the study of classes of inverse-closed algebras of infinite matrices with off-diagonal decay.[4] This is now a well-developed area of mathematics; see, e.g., [97, 98, 190] as well as [141, 173]. We will return on this topic in section 3.8.
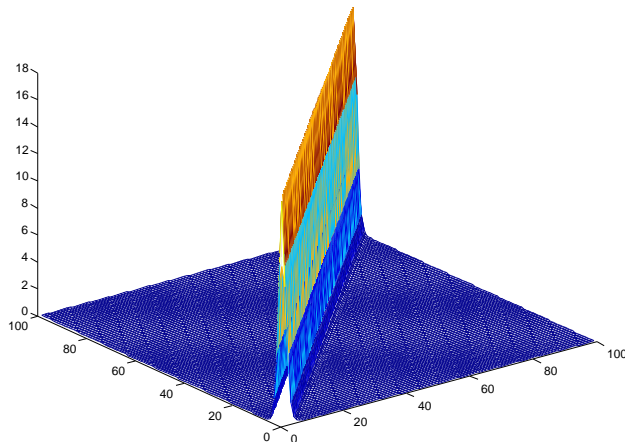

Fig. 2. Plot of $|[e^A]_{ij}|$ for A tridiagonal (discrete 1D Laplacian).

Locality in numerical linear algebra is related to, but should not be confused with, sparsity. A matrix can be localized even if it is a full matrix, although it will be close to a sparse matrix (in some norm).

Perhaps less obviously, a (discrete) system could well be described by a highly sparse matrix but be strongly delocalized. This happens when all the different parts comprising the system are "close together" in some sense. Network science provides striking examples of this: small diameter graphs, and particularly small-world networks, such as Facebook, and other online social networks, are highly sparse but delocalized, in the sense that there is no clear distinction between "short-range" and "long-range" interactions between the components of the system. Even if, on

---

[4] Let $\mathscr{A} \subseteq \mathscr{B}$ be two algebras with common identity. Then $\mathscr{A}$ is said to be *inverse-closed* in $\mathscr{B}$ if $A^{-1} \in \mathscr{A}$ for all $A \in \mathscr{A}$ that are invertible in $\mathscr{B}$ [97].
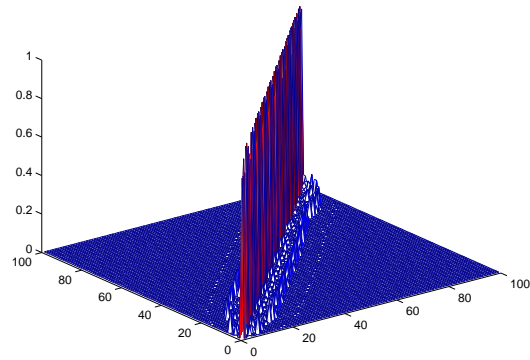
Fig. 3. Plot of $|[A^{1/2}]_{ij}|$ for matrix **nos4** from the University of Florida Sparse Matrix Collection [65] (scaled and reordered with reverse Cuthill–McKee).
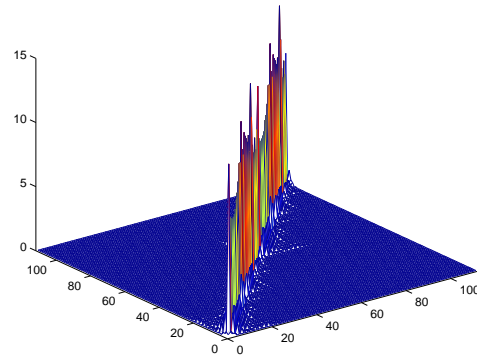

Fig. 4. Plot of $|[\log(A)]_{ij}|$ for matrix **bcsstk03** from the University of Florida Sparse Matrix Collection [65] (scaled and reordered with reverse Cuthill–McKee).

average, each component of such a system is directly connected to only a few other components, the system is strongly delocalized, since every node is only a few steps away from every other node. Hence, a "disturbance" at one node propagates quickly to the entire system. Every short range interaction is also long-range: locality is almost absent in such systems. We shall retrun to this topic in section 4.3.

Intuitively speaking, localization makes sense (for a system of N parts embedded in some n-dimensional space) when it is possible to let the system size N grow to infinity while keeping the density (number of parts per unit volume) constant. This situation is sometimes referred to as the *thermodynamic limit* (or *bulk limit* in solid state physics). We will provide a more formal discussion of this in a later section of the paper using notions from graph theory.

It is interesting to observe that both localization and delocalization can be advantageous from a computational perspective. Computing approximations to vectors or matrices that are strongly localized can be very efficient in terms of both storage and arithmetic complexity, but computations with systems that are both sparse and delocalized (in the sense just discussed) can also be very efficient, since information propagates very quickly in such systems. As a result, iterative methods based on matrix vector products for solving linear systems, computing eigenvalues and evaluating matrix functions tend to converge very quickly for sparse problems corresponding to small-diameter graphs; see, e.g., [5].

2 Notation and background in linear algebra and graph theory

In this chapter we provide the necessary background in linear algebra and graph theory. Excellent general references for linear algebra and matrix analysis are the two volumes by Horn and Johnson [113, 114]. For a thorough treatment of matrix functions, see the monograph by Higham [109]. A good general introduction to graph theory is Diestel [72].

We will be dealing primarily with matrices and vectors with entries in $\mathbb{R}$ or $\mathbb{C}$. The (i, j) entry of matrix A will be denoted either by a_{ij} or by $[A]_{ij}$. Throughout this chapter, I will denote the identity matrix (or operator); the dimension should be clear from the context.

Recall that a matrix $A \in \mathbb{C}^{n \times n}$ is *Hermitian* if $A^* = A$, *skew-Hermitian* if $A^* = -A$, *unitary* if $A^* = A^{-1}$, *symmetric* if $A^T = A$, *skew-symmetric* if $A^T = -A$, and *orthogonal* if $A^T = A^{-1}$. A matrix A is *diagonalizable* if it is similar to a diagonal matrix: there exist a diagonal matrix D and a nonsingular matrix X such that $A = XDX^{-1}$. The diagonal entries of D are the eigenvalues of A, denoted by λ_i, and they constitute the *spectrum* of A, denoted by $\sigma(A)$. The columns of X are the corresponding eigenvectors. A matrix A is *unitarily diagonalizable* if $A = UDU^*$ with D diagonal and U unitary. The *spectral theorem* states that a necessary and sufficient condition for a matrix A to be unitarily diagonalizable is that A is *normal*: $AA^* = A^*A$. Hermitian, skew-Hermitian and unitary matrices are examples of normal matrices.

Any matrix $A \in \mathbb{C}^{n \times n}$ can be reduced to *Jordan form*. Let $\lambda_1, \ldots, \lambda_s \in \mathbb{C}$ be the distinct eigenvalues of A. Then there exists a nonsingular $Z \in \mathbb{C}^{n \times n}$ such that $Z^{-1}AZ = J = \mathrm{diag}(J_1, J_2, \ldots, J_s)$, where each diagonal block $J_1, J_2, \ldots, J_s$

is block diagonal and has the form $J_i = \mathrm{diag}(J_i^{(1)}, J_i^{(2)}, \ldots, J_i^{(g_i)})$, where g_i is the geometric multiplicity of the λ_i,

$$
J_i^{(j)} = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix} \in \mathbb{C}^{\nu_i^{(j)} \times \nu_i^{(j)}},
$$

and $\sum_{i=1}^{s} \sum_{j=1}^{g_i} \nu_i^{(j)} = n$. The Jordan matrix J is unique up to the ordering of the blocks, but Z is not. The order n_i of the largest Jordan block in which the eigenvalue λ_i appears is called the *index* of λ_i. If the blocks J_i are ordered from largest to smallest, then $\mathrm{index}(\lambda_i) = \nu_i^{(1)}$. A matrix A is diagonalizable if and only if all the Jordan blocks in J are 1×1.

From the Jordan decomposition of a matrix $A \in \mathbb{C}^{n \times n}$ we obtain the following "coordinate-free" form of the Jordan decomposition of A:

$$
A = \sum_{i=1}^{s} [\lambda_i G_i + N_i] \tag{1}
$$

where $\lambda_1, \ldots, \lambda_s$ are the distinct eigenvalues of A, G_i is the projector onto the generalized eigenspace $\mathrm{Ker}((A - \lambda_i I)^{n_i})$ along $\mathrm{Ran}((A - \lambda_i I)^{n_i})$ with $n_i = \mathrm{index}(\lambda_i)$, and $N_i = (A - \lambda_i I)G_i = G_i(A - \lambda_i I)$ is nilpotent of index n_i. The G_i's are the *Frobenius covariants* of A.

If A is diagonalizable ($A = XDX^{-1}$) then $N_i = 0$ and the expression above can be written

$$
A = \sum_{i=1}^{n} \lambda_i \mathbf{x}_i \mathbf{y}_i^*
$$

where $\lambda_1, \ldots, \lambda_n$ are not necessarily distinct eigenvalues, and $\mathbf{x}_i$, $\mathbf{y}_i$ are right and left eigenvectors of A corresponding to λ_i. Hence, A is a weighted sum of at most n rank-one matrices (oblique projectors).

If A is normal then the spectral theorem yields

$$
A = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^*
$$

where $\mathbf{u}_i$ is eigenvector corresponding to λ_i. Hence, A is a weighted sum of at most n rank-one orthogonal projectors.

From these expressions one readily obtains for any matrix $A \in \mathbb{C}^{n \times n}$ that

$$
\mathrm{Tr}\,(A) := \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i
$$

and, more generally,

$$
\mathrm{Tr}\,(A^k) = \sum_{i=1}^{n} \lambda_i^k, \quad \forall k = 1, 2, \ldots
$$

Next, we recall the *singular value decomposition* (SVD) of a matrix. For any $A \in \mathbb{C}^{m \times n}$ there exist unitary matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ and a "diagonal" matrix $\Sigma \in \mathbb{R}^{m \times n}$ such that

$$U^*AV = \Sigma = \mathrm{diag}\,(\sigma_1, \ldots, \sigma_p)$$

where $p = \min\{m, n\}$. The σ_i are the *singular values* of A and satisfy (for $A \neq 0$)

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0\,,$$

where $r = \mathrm{rank}(A)$. The matrix Σ is uniquely determined by A, but U and V are not. The columns $\mathbf{u}_i$ and $\mathbf{v}_i$ of U and V are left and right singular vectors of A corresponding to the singular value σ_i. From $AA^* = U\Sigma\Sigma^T U^*$ and $A^*A = V\Sigma^T \Sigma V^*$ we deduce that the singular values of A are the (positive) square roots of the eigenvalues of the matrices AA^* and A^*A; the left singular vectors of A are eigenvectors of AA^*, and the right ones are eigenvectors of A^*A. Moreover,

$$A = \sum_{i=1}^{r} \sigma_i\, \mathbf{u}_i \mathbf{v}_i^*\,,$$

showing that any matrix A of rank r is the sum of exactly r rank-one matrices.

The notion of a *norm* on a vector space (over $\mathbb{R}$ or $\mathbb{C}$) is well known. A *matrix norm* on the matrix spaces $\mathbb{R}^{n \times n}$ or $\mathbb{C}^{n \times n}$ is just a vector norm $\|\cdot\|$ which satisfies the additional requirement of being *submultiplicative*:

$$\|AB\| \leq \|A\|\|B\|, \quad \forall A, B\,.$$

Important examples of matrix norms include the *Frobenius norm*

$$\|A\|_F := \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2}$$

as well as the norms

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|, \quad \|A\|_\infty = \|A^*\|_1 = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$$

and the *spectral norm* $\|A\|_2 = \sigma_1$. Note that $\|A\|_F = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$ and therefore $\|A\|_2 \leq \|A\|_F$ for all A. The inequality

$$\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty} \tag{2}$$

is often useful; note that for $A = A^*$ it implies that $\|A\|_2 \leq \|A\|_1 = \|A\|_\infty$.

The *spectral radius* $\varrho(A) := \max\{|\lambda| : \lambda \in \sigma(A)\}$ satisfies $\varrho(A) \leq \|A\|$ for all A and all matrix norms. For a normal matrix, $\varrho(A) = \|A\|_2$. But if A is nonnormal, $\|A\|_2 - \varrho(A)$ can be arbitrarily large. Also note that if A is diagonalizable with $A = XDX^{-1}$, then

$$\|A\|_2 = \|XDX^{-1}\|_2 \leq \|X\|_2 \|X^{-1}\|_2 \|D\|_2 = \kappa_2(X)\varrho(A)\,,$$

where $\kappa_2(X) = \|X\|_2 \|X^{-1}\|_2$ is defined as the infimum of the spectral condition numbers of X taken over the set of *all* matrices X which diagonalize A.

Clearly, the spectrum $\sigma(A)$ is entirely contained in the closed disk in the complex plane centered at the origin with radius $\varrho(A)$. Much effort has been devoted to

finding better "inclusion regions," i.e., subsets of $\mathbb{C}$ containing all the eigenvalues of a given matrix. We review some of these next.

Let $A \in \mathbb{C}^{n \times n}$. For all $i = 1, \ldots, n$, let

$$r_i := \sum_{j \neq i} |a_{ij}|, \quad D_i = D_i(a_{ii}, r_i) := \{z \in \mathbb{C} \, : \, |z - a_{ii}| \leq r_i\}.$$

The set D_i is called the *ith Geršgorin disk* of A. *Geršgorin's Theorem* (1931) states that $\sigma(A) \subset \cup_{i=1}^{n} D_i$. Moreover, each connected component of $\cup_{i=1}^{n} D_i$ consisting of p Geršgorin disks contains exactly p eigenvalues of A, counted with their multiplicities. Of course, the same result holds replacing the off-diagonal row-sums with off-diagonal column-sums. The spectrum is then contained in the intersection of the two resulting regions.

Also of great importance is the *field of values* (or *numerical range*) of $A \in \mathbb{C}^{n \times n}$, defined as the set

$$\mathcal{W}(A) := \{z = \langle A\mathbf{x}, \mathbf{x} \rangle \, : \, \mathbf{x}^* \mathbf{x} = 1\}. \tag{3}$$

This set is a compact subset of $\mathbb{C}$ containing the eigenvalues of A; it is also convex. This last statement is known as the *Hausdorff–Toeplitz Theorem*, and is highly nontrivial. If A is normal, the field of values is the convex hull of the eigenvalues; the converse is true if $n \leq 4$, but not in general. The eigenvalues and the field of values of a random 10×10 matrix are shown in Fig. 5.
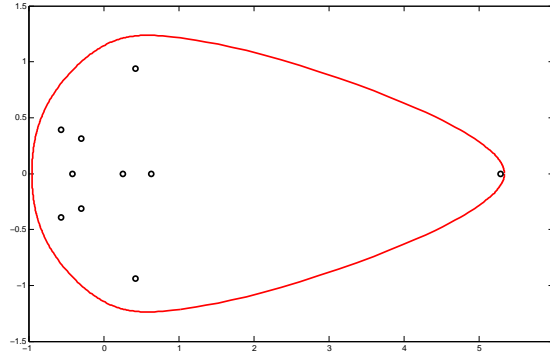

Fig. 5. Eigenvalues and field of values of a random 10×10 matrix.

For a matrix $A \in \mathbb{C}^{n \times n}$, let

$$H_1 = \frac{1}{2}(A + A^*), \qquad H_2 = \frac{1}{2i}(A - A^*). \tag{4}$$

Note that H_1, H_2 are both Hermitian. Let $a = \min \lambda(H_1)$, $b = \max \lambda(H_1)$, $c = \min \lambda(H_2)$, and $d = \max \lambda(H_2)$. Then for every eigenvalue $\lambda(A)$ of A we have that

$$a \leq \Re(\lambda(A)) \leq b, \qquad c \leq \Im(\lambda(A)) \leq d.$$

This is sometimes referred to as the *Bendixson–Hirsch Theorem*; see, e.g., [14, page 224]. Moreover, the field of values of A is entirely contained in the rectangle $[a, b] \times [c, d]$ in the complex plane [114, page 9]. Note that if $A \in \mathbb{R}^{n \times n}$, then $c = -d$.

The definition of field of values (3) also applies to *bounded linear operators* on a Hilbert space $\mathscr{H}$; however, $\mathcal{W}(A)$ may not be closed if $\dim(\mathscr{H}) = \infty$.

For a matrix $A \in \mathbb{C}^{n \times n}$ and a scalar polynomial

$$p(\lambda) = c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_k\lambda^k,$$

define

$$p(A) = c_0 I + c_1 A + c_2 A^2 + \cdots + c_k A^k.$$

Let $A = ZJZ^{-1}$ where J is the Jordan form of A. Then $p(A) = Zp(J)Z^{-1}$. Hence, the eigenvalues of $p(A)$ are given by $p(\lambda_i)$, for $i = 1, \ldots, n$. In particular, if A is diagonalizable with $A = XDX^{-1}$ then $p(A) = Xp(D)X^{-1}$. Hence, A and $p(A)$ have the same eigenvectors.

The *Cayley–Hamilton Theorem* states that for any matrix $A \in \mathbb{C}^{n \times n}$ it holds that $p_A(A) = 0$, where $p_A(\lambda) := \det(A - \lambda I)$ is the characteristic polynomial of A. Perhaps an even more important polynomial is the *minimum polynomial* of A, which is defined as the monic polynomial $q_A(\lambda)$ of least degree such that $q_A(A) = 0$. Note that $q_A | p_A$, hence $\deg(q_A) \leq \deg(p_A) = n$. It easily follows from this that for any nonsingular $A \in \mathbb{C}^{n \times n}$, the inverse A^{-1} can be expressed as a polynomial in A of degree at most $n - 1$:

$$A^{-1} = c_0 I + c_1 A + c_2 A^2 + \cdots + c_k A^k, \quad k \leq n - 1.$$

Note, however, that the coefficients c_i depend on A. It also follows that powers A^p with $p \geq n$ can be expressed as linear combinations of powers A^k with $0 \leq k \leq n - 1$. The same result holds more generally for matrix functions $f(A)$ that can be represented as power series in A (see below).

Indeed, let $\lambda_1, \ldots, \lambda_s$ be the distinct eigenvalues of $A \in \mathbb{C}^{n \times n}$ and let n_i be the index of λ_i. If f is a given function, we *define* the matrix function

$$f(A) := r(A),$$

where r is the unique Lagrange–Hermite interpolating polynomial of degree $< \sum_{i=1}^{s} n_i$ satisfying

$$r^{(j)}(\lambda_i) = f^{(j)}(\lambda_i) \qquad j = 0, \ldots, n_i - 1, \quad i = 1, \ldots, s.$$

Here $f^{(j)}$ denotes the jth derivative of f, with $f^{(0)} \equiv f$. Note that for the definition to make sense we must require that the values $f^{(j)}(\lambda_i)$ with $0 \leq j \leq n_i - 1$ and $1 \leq i \leq s$ exist. We say that f *is defined on the spectrum of* A. When all the eigenvalues are distinct, the interpolation polynomial has degree $n - 1$. In this case, the minimum polynomial and the characteristic polynomial of A coincide.

There are several other ways to define $f(A)$, all equivalent to the definition just given [109]. One such definition is through the Jordan canonical form. Let $A \in \mathbb{C}^{n \times n}$ have Jordan form $Z^{-1}AZ = J$ with $J = \mathrm{diag}(J_1, \ldots, J_s)$. We define

$$f(A) := Z\, f(J)\, Z^{-1} = Z\, \mathrm{diag}(f(J_1), f(J_2), \ldots, f(J_s))\, Z^{-1},$$

where $f(J_i) = \mathrm{diag}(f(J_i^{(1)}), f(J_i^{(2)}), \ldots, f(J_i^{(g_i)}))$ and

$$f(J_i^{(j)}) = \begin{bmatrix} f(\lambda_i) \ f'(\lambda_i) \ \dots \ \dfrac{f^{(\nu_i^{(j)}-1)}(\lambda_i)}{(\nu_i^{(j)}-1)!} \\ \quad f(\lambda_i) \ \ddots \ \vdots \\ \quad\quad \ddots \ f'(\lambda_i) \\ \quad\quad\quad f(\lambda_i) \end{bmatrix}.$$

An equivalent expression is the following:

$$f(A) = \sum_{i=1}^{s} \sum_{j=0}^{n_i-1} \frac{f^{(j)}(\lambda_i)}{j!} (A - \lambda_i I)^j G_i,$$

where $n_i = \text{index}(\lambda_i)$ and G_i is the Frobenius covariant associated with λ_i (see (1)). The usefulness of this definition is primarily theoretical, given the difficulty of determining the Jordan structure of a matrix numerically. If $A = XDX^{-1}$ with D diagonal, then $f(A) := Xf(D)X^{-1} = X\text{diag}(f(\lambda_i))X^{-1}$. Denoting with $\mathbf{x}_i$ the $i$the column of X and with $\mathbf{y}_i$ the ith column of X^{-1} we obtain the expression

$$f(A) = \sum_{i=1}^{n} f(\lambda_i) \, \mathbf{x}_i \mathbf{y}_i^*.$$

If in addition $A = UDU^*$ is normal then

$$f(A) = \sum_{i=1}^{n} f(\lambda_i) \, \mathbf{u}_i \mathbf{u}_i^*.$$

If f is analytic in a domain $\Omega \subseteq \mathbb{C}$ containing the spectrum of $A \in \mathbb{C}^{n \times n}$, then

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} \mathrm{d}z, \qquad (5)$$

where $i = \sqrt{-1}$ is the imaginary unit and Γ is any simple closed curve surrounding the eigenvalues of A and entirely contained in Ω, oriented counterclockwise. This definition has the advantage of being easily generalized to functions of bounded operators on Banach spaces, and it is also the basis for some of the currently most efficient computational methods for the evaluation of matrix functions.

Another widely used definition of $f(A)$ when f is analytic is through power series. Suppose f has a Taylor series expansion

$$f(z) = \sum_{k=0}^{\infty} a_k(z - z_0)^k \qquad \left(a_k = \frac{f^{(k)}(z_0)}{k!} \right)$$

with radius of convergence R. If $A \in \mathbb{C}^{n \times n}$ and each of the distinct eigenvalues $\lambda_1, \dots, \lambda_s$ of A satisfies

$$|\lambda_i - z_0| < R,$$

then

$$f(A) := \sum_{k=0}^{\infty} a_k(A - z_0 I)^k.$$

If $A \in \mathbb{C}^{n \times n}$ and f is defined on $\sigma(A)$, the following facts hold (see [109]):

(i) $f(A)A = Af(A)$;

(ii) $f(A^T) = f(A)^T$;

(iii) $f(XAX^{-1}) = Xf(A)X^{-1}$;

(iv) $\sigma(f(A)) = f(\sigma(A))$;

(v) (λ, x) eigenpair of $A \Rightarrow (f(\lambda), x)$ eigenpair of $f(A)$;

(vi) A is block triangular $\Rightarrow F = f(A)$ is block triangular with the same block structure as A, and $F_{ii} = f(A_{ii})$ where A_{ii} is the ith diagonal block of A;

(vii)   In particular, $f(\text{diag}\,(A_{11}, \ldots, A_{pp})) = \text{diag}\,(f(A_{11}), \ldots, f(A_{pp}))$;

(viii) $f(I_m \otimes A) = I_m \otimes f(A)$, where $\otimes$ is the Kronecker product;

(ix) $f(A \otimes I_m) = f(A) \otimes I_m$.

Another useful result is the following:

Theorem 1. ([110]) *Let f be analytic on an open set $\Omega \subseteq \mathbb{C}$ such that each connected component of Ω is closed under conjugation. Consider the corresponding matrix function f on the set $\mathcal{D} = \{A \in \mathbb{C}^{n \times n} : \sigma(A) \subseteq \Omega\}$. Then the following are equivalent:*

(a) $f(A^*) = f(A)^*$ *for all* $A \in \mathcal{D}$.

(b) $f(\overline{A}) = \overline{f(A)}$ *for all* $A \in \mathcal{D}$.

(c) $f(\mathbb{R}^{n \times n} \cap \mathcal{D}) \subseteq \mathbb{R}^{n \times n}$.

(d) $f(\mathbb{R} \cap \Omega) \subseteq \mathbb{R}$.

In particular, if $f(x) \in \mathbb{R}$ for $x \in \mathbb{R}$ and A is Hermitian, so is $f(A)$.

Important examples of matrix functions are the resolvent and the matrix exponential. Let $A \in \mathbb{C}^{n \times n}$, and let $z \notin \sigma(A)$. The *resolvent* of A at z is defined as

$$R(A; z) = (zI - A)^{-1}.$$

The resolvent is central to the definition of matrix functions via the contour integral approach (5). The resolvent also plays a fundamental role in spectral theory. For example, it can be used to define the spectral projector onto the eigenspace of a matrix or operator corresponding to an isolated eigenvalue $\lambda_0 \in \sigma(A)$:

$$P_{\lambda_0} := \frac{1}{2\pi i} \int_{|z - \lambda_0| = \varepsilon} (zI - A)^{-1} \mathrm{d}z\,,$$

where $\varepsilon > 0$ is small enough so that no other eigenvalue of A falls within ε of λ_0. It can be shown that $P_{\lambda_0}^2 = P_{\lambda_0}$ and that the range of P_{λ_0} is the one-dimensional subspace spanned by the eigenvector associated with λ_0. More generally, one can define the spectral projector onto the invariant subspace of A corresponding to a set of selected eigenvalues by integrating $R(A; z)$ along a countour surrounding those eigenvalues and excluding the others. It should be noted that the spectral projector is an orthogonal projector ($P = P^*$) if and only if A is normal. If A is diagonalizable, a spectral projector P is a simple function of A: if f is any function taking the value 1 at the eigenvalues of interest and 0 on the remaining ones, then $P = f(A)$.

The *matrix exponential* can be defined via the Maclaurin expansion

$$\mathrm{e}^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}A^k\,,$$

which converges for arbitrary $A \in \mathbb{C}^{n \times n}$. Just as the resolvent is central to spectral theory, the matrix exponential is fundamental to the solution of differential equations. For example, the solution to the inhomogeneous system

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathbf{y} + f(t, \mathbf{y}), \quad \mathbf{y}(0) = \mathbf{y}_0, \quad \mathbf{y} \in \mathbb{C}^n, \quad A \in \mathbb{C}^{n \times n}$$

is given (implicitly!) by

$$\mathbf{y}(t) = \mathrm{e}^{tA}\mathbf{y}_0 + \int_0^t \mathrm{e}^{A(t-s)} f(s, \mathbf{y}(s)) \mathrm{d}s.$$

In particular, $\mathbf{y}(t) = \mathrm{e}^{tA}\mathbf{y}_0$ when $f = 0$. It is worth recalling that $\lim_{t \to \infty} \mathrm{e}^{tA} = 0$ if and only if A is a *stable* matrix: $\Re(\lambda) < 0$ for all $\lambda \in \sigma(A)$.

When $f(t, \mathbf{y}) = \mathbf{b} \in \mathbb{C}^n$ (=const.), the solution can also be expressed as

$$\mathbf{y}(t) = t\psi_1(tA)(\mathbf{b} + A\mathbf{y}_0) + \mathbf{y}_0\,,$$

where

$$\psi_1(z) = \frac{\mathrm{e}^z - 1}{z} = 1 + \frac{z}{2!} + \frac{z^2}{3!} + \cdots$$

The matrix exponential plays an especially important role in quantum theory. Consider for instance the time-dependent Schrödinger equation:

$$i\frac{\partial \Psi}{\partial t} = H\Psi, \quad t \in \mathbb{R}, \quad \Psi(0) = \Psi_0, \tag{6}$$

where $\Psi_0 \in L^2$ is a prescribed initial state with $\|\Psi_0\|_2 = 1$. Here $H = H^*$ is the Hamiltonian, or energy operator (which we assume to be time-independent). The solution of (6) is given explicitly by $\Psi(t) = \mathrm{e}^{-itH}\Psi_0$, for all $t \in \mathbb{R}$; note that since itH is skew-Hermitian, the *propagator* $U(t) = \mathrm{e}^{-itH}$ is unitary, which guarantees that the solution has unit norm for all t:

$$\|\Psi(t)\|_2 = \|U(t)\Psi_0\|_2 = \|\Psi_0\|_2 = 1, \quad \forall t \in \mathbb{R}.$$

Also very important in many-body quantum mechanics is the *Fermi–Dirac operator*, defined as

$$f(H) := (I + \exp(\beta(H - \mu I)))^{-1}\,,$$

where $\beta = (\kappa_B T)^{-1}$ is the *inverse temperature*, κ_B the Boltzmann constant, and μ is the *Fermi level*, separating the eigenvalues of H corresponding to the first n_e eigenvectors from the rest, where n_e is the number of particles (electrons) comprising the system under study. This matrix function will be discussed in section 4.2.

Finally, in statistical quantum mechanics the state of a system is completely described (statistically) by the *density operator*:

$$\rho := \frac{\mathrm{e}^{-\beta H}}{Z}, \quad \text{where } Z = \mathrm{Tr}\,(\mathrm{e}^{-\beta H}).$$

The quantity $Z = Z(\beta)$ is known as the *partition function* of the system.

Trigonometric functions and square roots of matrices are also important in applications to differential equations. For example, the solution to the second-order system

$$\frac{\mathrm{d}^2\mathbf{y}}{\mathrm{d}t^2} + A\mathbf{y} = 0, \qquad \mathbf{y}(0) = \mathbf{y}_0, \qquad \mathbf{y}'(0) = \mathbf{y}_0'$$

(where A is SPD) can be expressed as

$$\mathbf{y}(t) = \cos(\sqrt{A}t)\,\mathbf{y}_0 + (\sqrt{A})^{-1}\sin(\sqrt{A}t)\,\mathbf{y}_0'\,.$$

Apart from the contour integration formula, the matrix exponential and the resolvent are also related through the *Laplace transform*: there exists an $\omega \in \mathbb{R}$ such that $z \notin \sigma(A)$ for $\Re(z) > \omega$ and

$$(zI - A)^{-1} = \int_0^\infty \mathrm{e}^{-zt}\mathrm{e}^{tA}\mathrm{d}t = \int_0^\infty \mathrm{e}^{-t(zI-A)}\mathrm{d}t\,.$$

Also note that if $|z| > \varrho(A)$, the following *Neumann series expansion* of the resolvent is valid:

$$(zI - A)^{-1} = z^{-1}(I + z^{-1}A + z^{-2}A^2 + \cdots) = z^{-1}\sum_{k=0}^\infty z^{-k}A^k.$$

Next, we recall a few definitions and notations associated with graphs. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with $n = |\mathcal{V}|$ nodes (or vertices) and $m = |\mathcal{E}|$ edges (or links). The elements of $\mathcal{V}$ will be denoted simply by $1, \ldots, n$. If for all $i, j \in \mathcal{V}$ such that $(i, j) \in \mathcal{E}$ then also $(j, i) \in \mathcal{E}$, the graph is said to be *undirected*. On the other hand, if this condition does not hold, namely if there exists $(i, j) \in \mathcal{E}$ such that $(j, i) \notin \mathcal{E}$, then the network is said to be *directed*. A directed graph is commonly referred to as a *digraph*. If $(i, j) \in \mathcal{E}$ in a digraph, we will write $i \to j$. A graph is *simple* if it is unweighted, contains no *loops* (edges of the form (i, i)) and there are no multiple edges with the same orientation between any two nodes. A simple graph can be represented by means of its *adjacency matrix* $A = [a_{ij}] \in \mathbb{R}^{n \times n}$, where

$$a_{ij} = \begin{cases} 1, & \text{if } (i, j) \in \mathcal{E}, \\ 0, & \text{else.} \end{cases}$$

Note that $A = A^T$ if, and only if, $\mathcal{G}$ is undirected. If the graph is weighted, then a_{ij} will be equal to the weight of the corresponding edge (i, j).

If $\mathcal{G}$ is undirected, the *degree* $\deg(i)$ of node i is the number of edges incident to i in $\mathcal{G}$. That is, $\deg(i)$ is the number of "immediate neighbors" of i in $\mathcal{G}$. Note that in terms of the adjacency matrix, $\deg(i) = \sum_{j=1}^n a_{ij}$. A *d-regular graph* is a graph where every node has the same degree d.

For an undirected graph we also define the *graph Laplacian* as the matrix

$$L := D - A, \qquad \text{where} \quad D := \mathrm{diag}(\deg(1), \deg(2), \ldots, \deg(n))$$

and, assuming $\deg(i) \neq 0$ for all i, the *normalized Laplacian*

$$\widehat{L} := I - D^{-1/2}AD^{-1/2}.$$

Both of these matrices play an important role in the structural analysis of networks and in the study of diffusion-type process on graphs, and matrix exponentials of the form e^{-tL} and $\mathrm{e}^{-t\widehat{L}}$, where $t > 0$ denotes time, are widely used in applications. Note that L and $\hat{L}$ are both symmetric positive semidefinite matrices. Moreover, if $\mathcal{G}$ is a d-regular graph, then the eigenvalues of L are given by $d - \lambda_i(A)$ (where

$\lambda_i(A)$ are the eigenvalues of the adjacency matrix A of $\mathcal{G}$) and L and A have the same eigenvectors. For more general graphs, however, there is no simple relationship between the spectra of L and A.

A *walk* of length k in $\mathcal{G}$ is a set of nodes $\{i_1, i_2, \ldots i_k, i_{k+1}\}$ such that for all $1 \leq j \leq k$, there is an edge between i_j and i_{j+1} (a directed edge $i_j \to i_{j+1}$ for a digraph). A *closed walk* is a walk where $i_1 = i_{k+1}$. A *path* is a walk with no repeated nodes.

There is a close connection between the walks in $\mathcal{G}$ and the entries of the powers of the adjacency matrix A. Indeed, let $k \geq 1$. For any simple graph $\mathcal{G}$, the following holds:

$[A^k]_{ii}$ = number of closed walks of length k starting and ending at node i;
$[A^k]_{ij}$ = number of walks of length k starting at node i and ending at node j.

Let now i and j be any two nodes in $\mathcal{G}$. In many situations in network science it is desirable to have a measure of how "well connected" nodes i and j are. Estrada and Hatano [80] have proposed to quantify the strength of connection between nodes in terms of the number of walks joining i and j, assigning more weight to shorter walks (i.e., penalizing longer ones). If walks of length k are downweighted by a factor $\frac{1}{k!}$, this leads [80] to the following definition of *communicability* between node i and node j:

$$C(i,j) := [\mathrm{e}^A]_{ij} = \sum_{k=0}^{\infty} \frac{[A^k]_{ij}}{k!}, \tag{7}$$

where by convention we assign the value 1 to the number of "walks of length 0." Of course, other matrix functions can also be used to define the communicability between nodes [82], but the matrix exponential has a natural physical interpretation (see [81]).

The *geodesic distance* $d(i,j)$ between two nodes i and j is the length of the shortest path connecting i and j. We let $d(i,j) = \infty$ if no such path exists. We note that $d(\cdot,\cdot)$ is a true distance function (i.e., a *metric* on $\mathcal{G}$) if the graph is undirected, but not in general, since only in an undirected graph the condition $d(i,j) = d(j,i)$ is satisfied for all $i \in \mathcal{V}$.

The *diameter* of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined as

$$\mathrm{diam}(\mathcal{G}) := \max_{i,j \in \mathcal{V}} d(i,j).$$

A digraph $\mathcal{G}$ is *strongly connected* (or, in short, *connected*) if for every pair of nodes i and j there is a path in $\mathcal{G}$ that starts at i and ends at j; i.e., $\mathrm{diam}(\mathcal{G}) < \infty$. We say that $\mathcal{G}$ is *weakly connected* if it is connected as an undirected graph (i.e., when the orientation of the edges is disregarded). Clearly, for an undirected graph the two notions coincide. It can be shown that for an undirected graph the number of connected components is equal to the dimension of $\mathrm{Ker}(L)$, the null space of the graph Laplacian.

Just as we have associated matrices to graphs, graphs can also be associated to matrices. In particular, to any matrix $A \in \mathbb{C}^{n \times n}$ we can associate a digraph $\mathcal{G}(A) = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{1, 2, \ldots, n\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, where $(i,j) \in \mathcal{E}$ if and only if $a_{ij} \neq 0$. Diagonal entries in A are usually ignored, so that there are no loops in $\mathcal{G}(A)$. We also note that for *structurally symmetric* matrices ($a_{ij} \neq 0 \Leftrightarrow a_{ji} \neq 0$) the associated graph $\mathcal{G}(A)$ with A is undirected.

Let $|A| := [\,|a_{ij}|\,]$, then the digraph $\mathcal{G}(|A|^2)$ is given by $(\mathcal{V}, \hat{\mathcal{E}})$ where $\hat{\mathcal{E}}$ is obtained by including all directed edges (i, k) such that there exists $j \in V$ with $(i, j) \in \mathcal{E}$ and $(j, k) \in \mathcal{E}$. (The reason for the absolute value is to disregard the effect of possible cancellations in A^2.) For higher powers ℓ, the digraph $\mathcal{G}(|A|^\ell)$ is defined similarly: its edge set consists of all pairs (i, k) such that there is a directed path of length at most ℓ joining node i with node k in $\mathcal{G}(A)$.
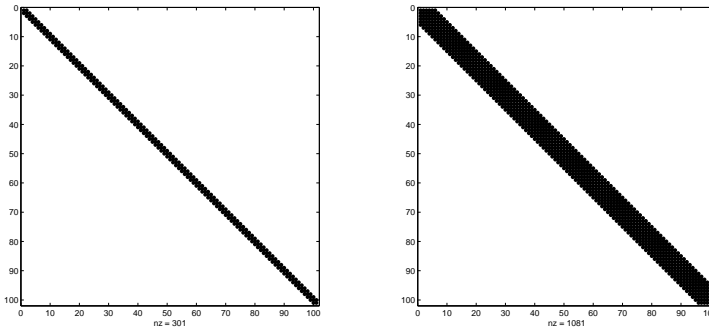

Fig. 6. Path graph. Left: nonzero pattern of Laplacian matrix L. Right: pattern of fifth power of L.
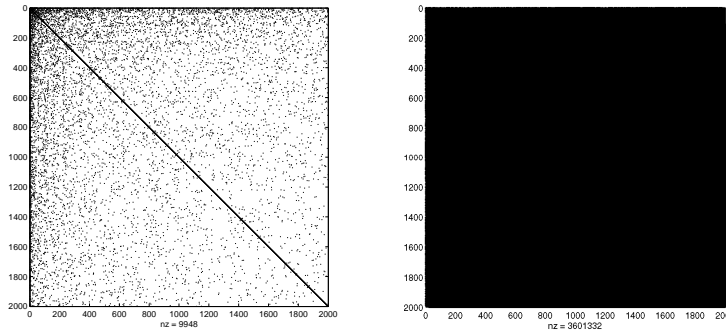

Fig. 7. Scale-free graph. Left: nonzero pattern of Laplacian matrix L. Right: pattern of fifth power of L.

Thus, for any square matrix A it is possible to predict the *structural nonzero pattern* of the powers A^ℓ for $\ell = 2, 3, \ldots$ from the connectivity of the graphs $\mathcal{G}(|A|^\ell)$. One of the first observations that can be made is that powers of narrow-banded matrices (corresponding to graphs with large diameter, for example paths) take large values of ℓ to fill, whereas the opposite happens with matrices that correspond to small-diameter graphs. Figures 6-7 illustrate this fact by displaying the graph

Laplacian L and the fifth power of L for two highly sparse undirected graphs, a path graph with $n = 100$ nodes and a scale-free graph on $n = 2000$ nodes built according to preferential attachment scheme (see, e.g., [79]). Graphs of this type are examples of *small-world graphs*, in particular they can be expected to have small diameter. It can be seen that in the case of the scale-free graph the fifth power of the Laplacian, L^5, is almost completely full (the number of nonzeros is $3,601,332$ out of a possible $4,000,000$), implying that in this graph most pairs of nodes are less than five degrees of separation away from one another.

The *transitive closure* of $\mathcal{G}$ is the graph $\bar{\mathcal{G}} = (\mathcal{V}, \bar{\mathcal{E}})$ where $(i,j) \in \bar{\mathcal{E}}$ if and only if there is a directed path from i to j in $\mathcal{G}(A)$. A matrix $A \in \mathbb{C}^{n \times n}$ is *reducible* if there exists a permutation matrix P such that

$$P^T A P = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

with A_{11} and A_{22} square submatrices. If no such P exists, A is said to be *irreducible*. Denote by K_n the *complete graph* on n nodes, i.e., the graph where every edge (i,j) is present (with $i \neq j$). The following statements are equivalent:

(i)   the matrix A is irreducible;
(ii)   the digraph $\mathcal{G}(A)$ is strongly connected;
(iii) the transitive closure $\bar{\mathcal{G}}(A)$ of $\mathcal{G}(A)$ is K_n.

Note that (iii) and the Cayley–Hamilton Theorem imply that the powers $(I + |A|)^k$ are completely full for $k \geq n - 1$. This has important implications for matrix functions, since it implies that for an irreducible matrix A a matrix function of the form

$$f(A) = \sum_{k=0}^{\infty} a_k (A - z_0 I)^k$$

is completely full, if no cancellation occurs and $a_k \neq 0$ for sufficiently many k. This is precisely formulated in the following result.

Theorem 2. ([24]) *Let f be an analytic function of the form*

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k \quad \left(a_k = \frac{f^{(k)}(z_0)}{k!} \right),$$

where $z_0 \in \mathbb{C}$ and the power series expansion has radius of convergence $R > 0$. Let A have an irreducible sparsity pattern and let l $(1 \leq l \leq n - 1)$ be the diameter of $\mathcal{G}(A)$. Assume further that there exists $k \geq l$ such that $f^{(k)}(z_0) \neq 0$. Then it is possible to assign values to the nonzero entries of A in such a way that $f(A)$ is defined and $[f(A)]_{ij} \neq 0$ for all $i \neq j$.

This result applies, in particular, to banded A and to such functions as the inverse (resolvent) and the matrix exponential.

3 Localization in matrix functions

We have just seen that if A is irreducible and f is a "generic" analytic function defined on the spectrum of A then we should expect $f(A)$ to be completely full

(barring fortuitous cancellation). For A large, this seems to make the explicit computation of $f(A)$ impossible, and this is certainly the case if all entries of $f(A)$ need to be accurately approximated.

As we have already mentioned in the Introduction, however, numerical experiments show that when A is a banded matrix and $f(z)$ is a smooth function for which $f(A)$ is defined, the entries of $f(A)$ often decay rapidly as one moves away from the diagonal. The same property is often (but not always!) satisfied by more general sparse matrices: in this case the decay is away from the support (nonzero pattern) of A. In other words, nonnegligible entries of $f(A)$ tend to be concentrated near the positions (i, j) for which $a_{ij} \neq 0$.

This observation opens up the possibility of approximating functions of sparse matrices, by neglecting "sufficiently small" matrix elements in $f(A)$. Depending on the rate of decay and on the accuracy requirements, it may be possible to develop approximation algorithms that exhibit optimal computational complexity, i.e., $O(n)$ (or *linear scaling*) methods.

In this section we review our current knowledge on localization in functions of large and sparse matrices. In particular, we consider the following questions:

1. Under which conditions can we expect decay in $f(A)$?
2. Can we obtain sharp bounds on the entries of $f(A)$?
3. Can we characterize the rate of decay in $f(A)$ in terms of
   - the bandwidth/sparsity of A?
   - the spectral properties of A?
   - the location of singularities of $f(z)$ in relation to the spectrum of A?
4. What if $f(z)$ is an entire[5] function?
5. When is the rate of decay independent of the matrix size n?

The last point is especially crucial if we want to develop $O(n)$ algorithms for approximating functions of sparse matrices.

3.1 Matrices with decay

A matrix $A \in \mathbb{C}^{n \times n}$ is said to have the *off-diagonal decay property* if its entries $[A]_{ij}$ satisfy a bound of the form

$$|[A]_{ij}| \leq K\phi(|i - j|), \qquad \forall\, i, j, \tag{8}$$

where $K > 0$ is a constant and ϕ is a function defined and positive for $x \geq 0$ and such that $\phi(x) \to 0$ as $x \to \infty$. Important examples of decay include exponential decay, corresponding to $\phi(x) = \mathrm{e}^{-\alpha x}$ for some $\alpha > 0$, and algebraic (or power-law) decay, corresponding to $\phi(x) = (1 + |i - j|^p)^{-1}$ for some $p \geq 1$.

As it stands, however, this definition is meaningless, since for any fixed matrix $A \in \mathbb{C}^{n \times n}$ the bound can always be achieved with an arbitrary choice of ϕ just by taking K sufficiently large. To give a meaningful definition we need to consider either infinite matrices (for example, bounded linear operators on some sequence space ℓ^p), or sequences of matrices of increasing dimension. The latter situation being the more familiar one in numerical analysis, we give the following definition.

---

[5] Recall that an *entire function* is a function of a complex variable that is analytic everywhere on the complex plane.

Definition 1. *Let* $\{A_n\}$ *be a sequence of* $n \times n$ *matrices with entries in* $\mathbb{C}$*, where* $n \to \infty$*. We say that the matrix sequence* $\{A_n\}$ *has the off-diagonal decay property if*

$$|[A_n]_{ij}| \leq K\phi(|i-j|), \qquad \forall\, i,j = 1, \ldots, n, \tag{9}$$

where the constant $K > 0$ *and the function* $\phi(x)$*, defined for* $x \geq 0$ *and such that* $\phi(x) \to 0$ *as* $x \to \infty$*, do not depend on* n*.*

Note that if A is an infinite matrix that satisfies (8) then its finite $n \times n$ sections (leading principal submatrices, see [141]) A_n form a matrix sequence that satisfies Def. 1. The definition can also be extended to block matrices in a natural way.

When dealing with non-Hermitian matrices, it is sometimes required to allow for different decay rates on either side of the main diagonal. For instance, one could have exponential decay on either side but with different rates:

$$|[A_n]_{ij}| \leq K_1\, e^{-\alpha(i-j)} \qquad \text{for} \quad i > j\,,$$

and

$$|[A_n]_{ij}| \leq K_2\, e^{-\beta(j-i)} \qquad \text{for} \quad j > i\,.$$

Here K_1, K_2 and α, β are all positive constants. It is also possible to have matrices where decay is present on only one side of the main diagonal (see [24, Theorem 3.5]). For simplicity, in the rest of the paper we will primarily focus on the case where the decay bound has the same form for $i > j$ and for $j > i$. However, most of the results can be extended easily to the more general case.

Also, in multidimensional problems it is important to be able to describe decay behavior not just away from the main diagonal but with a more complicated pattern. To this end, we can use any distance function (metric) d (with $d(i,j) = d(j,i)$ for simplicity) with the property that

$$\forall \varepsilon > 0 \,\, \exists\, c = c(\varepsilon) \quad \text{such that} \quad \sup_j \sum_i e^{-\varepsilon d(i,j)} \leq c(\varepsilon), \tag{10}$$

see [119]. Again, condition (10) is trivially satisfied for any distance function on a finite set $S = \{1, 2, \ldots, n\}$, but here we allow infinite ($S = \mathbb{N}$) or bi-infinite matrices ($S = \mathbb{Z}$). In practice, we will consider sequences of matrices of increasing size n and we will define for each n a distance d_n on the set $S = \{1, 2, \ldots, n\}$ and assume that each d_n satisfies condition (10) with respect to a constant $c = c(\varepsilon)$ independent of n.

We will be mostly concerned with decay away from a sparsity pattern. For banded sparsity patterns, this is just off-diagonal decay. For more general sparsity patterns, we assume that we are given a sequence of sparse graphs $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ with $|\mathcal{V}_n| = n$ and $|\mathcal{E}_n| = O(n)$ and a distance function d_n satisfying (10) uniformly with respect to n. In practice we will take d_n to be the geodesic distance on $\mathcal{G}_n$ and we will impose the following *bounded maximum degree* condition:

$$\sup_n \{\deg(i) \,|\, i \in \mathcal{G}_n\} < \infty\,. \tag{11}$$

This condition guarantees that the distance $d_n(i,j)$ grows unboundedly as $|i-j|$ does, at a rate independent of n for $n \to \infty$. In particular, we have that $\lim_{n\to\infty} \operatorname{diam}(\mathcal{G}_n) = \infty$. This is necessary if we want the entries of matrices with decay to actually go to zero with the distance as $n \to \infty$.

Let us now consider a sequence of $n \times n$ matrices A_n with associated graphs $\mathcal{G}_n$ and graph distances $d_n(i,j)$. We will say that A_n has the *exponential decay property relative to the graph* $\mathcal{G}_n$ if there are constants $K > 0$ and $\alpha > 0$ independent of n such that

$$|[A_n]_{ij}| \leq K\, \mathrm{e}^{-\alpha d_n(i,j)}, \quad \text{for all} \quad i, j = 1, \ldots, n, \ \forall\, n \in \mathbb{N}. \tag{12}$$

The following two results says that matrices with decay can be "uniformly well approximated" by sparse matrices.

Theorem 3. ([20]) *Let* $\{A_n\}$ *be a sequence of* $n \times n$ *matrices satisfying the exponential decay property (12) relative to a sequence of graphs* $\{\mathcal{G}_n\}$ *having uniformly bounded maximal degree. Then, for any given* $0 < \delta < K$, *each* A_n *contains at most* $O(n)$ *entries greater than* δ *in magnitude.*

Theorem 4. ([24]) *Let the matrix sequence* $\{A_n\}$ *satisfy the assumptions of Theorem 3. Then, for all* $\varepsilon > 0$ *and for all* n *there exists an* $n \times n$ *matrix* $\tilde{A}_n$ *containing only* $O(n)$ *nonzeros such that*

$$\|A_n - \tilde{A}_n\|_1 < \varepsilon. \tag{13}$$

For example, suppose the each matrix in the sequence $\{A_n\}$ satisfies the following exponential decay property: there exist $K, \alpha > 0$ independent of n such that

$$|[A_n]_{ij}| \leq K\mathrm{e}^{-\alpha|i-j|}, \quad \forall\, i, j = 1, \ldots, n, \ \forall\, n \in \mathbb{N}.$$

Then, for any $\varepsilon > 0$, there is a sequence of p-banded matrices $\tilde{A}_n$, with p independent of n, such that $\|A_n - \tilde{A}_n\|_1 < \varepsilon$. The matrices $\tilde{A}_n$ can be defined as follows:

$$[\tilde{A}_n]_{ij} = \begin{cases} [A_n]_{ij} & \text{if} \quad |i - j| \leq p; \\ 0 & \text{otherwise}, \end{cases}$$

where p satisfies

$$p \geq \left\lfloor \frac{1}{\alpha} \log \left(\frac{2K}{1 - \mathrm{e}^{-\alpha}} \varepsilon^{-1} \right) \right\rfloor. \tag{14}$$

Note that $\tilde{A}_n$ is the orthogonal projection of A_n, with respect to the inner product associated with the Frobenius norm, onto the linear subspace of $\mathbb{C}^{n \times n}$ of p-banded matrices.

Similar approximation results hold for other matrix norms. For instance, using the inequality (2) one can easily satisfy error bounds in the matrix 2-norm.

Remark 1. As mentioned in [20], similar results also hold for other types of decay; for instance, it suffices to have algebraic decay of the form

$$|[A_n]_{ij}| \leq K \left(|i - j|^p + 1 \right)^{-1} \quad \forall\, i, j, \ \forall\, n \in \mathbb{N},$$

with $p > 1$. However, this type of decay is often too slow to be useful in practice, in the sense that any sparse approximation $\tilde{A}_n$ to A_n would have to have $O(n)$ nonzeros with a huge prefactor in order to satisfy (13) for even moderately small values of ε.

3.2 Decay bounds for the inverse

It has long been known that the entries in the inverse of banded matrices are bounded in a decaying manner away from the main diagonal, with the decay being faster for more diagonally dominant matrices [68]. In 1984, Demko, Moss and Smith [69] proved that the entries of A^{-1}, where A is Hermitian positive definite and m-banded ($[A]_{ij} = 0$ if $|i - j| > m$), satisfy the following exponential off-diagonal decay bound:

$$|[A^{-1}]_{ij}| \leq K \rho^{|i-j|}, \quad \forall\, i, j. \tag{15}$$

Here we have set

$$K = \max\{a^{-1}, K_0\}, \quad K_0 = (1 + \sqrt{\kappa})/2b, \quad \kappa = \frac{b}{a}, \tag{16}$$

where $[a, b]$ is the smallest interval containing the spectrum $\sigma(A)$ of A, and

$$\rho = q^{1/m}, \qquad q = q(\kappa) = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}. \tag{17}$$

Hence, the decay bound deteriorates as the relative distance between the spectrum of A and the singularity at zero of the function $f(x) = x^{-1}$ tends to zero (i.e., as $\kappa \to \infty$) and/or if the bandwidth m increases. The bound is sharp (being attained for certain tridiagonal Toeplitz matrices). The result holds for $n \times n$ matrices as well as for bounded, infinite matrices acting on the Hilbert space ℓ^2. We also note that the bound (15) can be rewritten as

$$|[A^{-1}]_{ij}| \leq K e^{-\alpha|i-j|}, \quad \forall\, i, j, \tag{18}$$

where we have set $\alpha = -\log(\rho)$.

It should be emphasized that (15) is a just a bound: the off-diagonal decay in A^{-1} is in general not monotonic. Furthermore the bound, although sharp, may be pessimistic in practice.

The result of Demko et al. implies that if we are given a sequence of $n \times n$ matrices $\{A_n\}$ of increasing size, all Hermitian, positive definite, m-banded (with $m < n_0$) and such that

$$\sigma(A_n) \subset [a, b] \quad \forall n \geq n_0, \tag{19}$$

then the bound (15) holds for all matrices of the sequence; in other words, if the spectra $\sigma(A_n)$ are bounded away from zero and infinity uniformly in n, the entries of A_n^{-1} are uniformly bounded in an exponentially decaying manner (i.e., the decay rates are independent of n). Note that it is not necessary that all matrices have exactly the same bandwidth m, as long as they are banded with bandwidth less than or equal to a constant m.

The requirement that the matrices A_n have uniformly bounded condition number as $n \to \infty$ is restrictive. For example, it does not apply to banded or sparse matrices that arise from the discretization of differential operators, or in fact of any unbounded operator. Consider for example the sequence of tridiagonal matrices

$$A_n = (n + 1)^2 \operatorname{tridiag}(-1\,, 2\,, -1)$$

which arise from the three-point finite difference approximation with mesh spacing $h = \frac{1}{n+1}$ of the operator $T = -\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ with zero Dirichlet conditions at $x = 0$ and

$x = 1$. For $n \to \infty$ the condition number of A_n grows like $O(n^2)$, and although the entries of each inverse A_n^{-1} satisfy a bound of the type (15), the spectral condition number $\kappa_2(A_n)$ is unbounded and therefore the bound deteriorates since $K = K(n) \to \frac{1}{\pi^2}$ and $\rho = \rho(n) \to 1$ as $n \to \infty$. Moreover, in this particular example the actual decay in A_n^{-1} (and not just the bound) slows down as $h \to 0$. This is to be expected since A_n^{-1} is trying to approximate the Green's function of T, which does not fall off exponentially.

Nevertheless, this result is important for several reasons. First of all, families of banded or sparse matrices (parameterized by the dimension n) exhibiting bounded condition numbers do occur in applications. For example, under mild conditions, mass matrices in finite element analysis and overlap matrices in quantum chemistry satisfy such conditions (these matrices represent the identity operator with respect to some non-orthogonal basis set $\{\phi_i\}_{i=1}^n$, where the ϕ_i are strongly localized in space). Second, the result is important because it suggests a possible *sufficient* condition for the existence of a uniform exponential decay bound in more general situations: the relative distance of the spectra $\sigma(A_n)$ from the singularities of the function must remain strictly positive as $n \to \infty$. Third, it turns out that the method of proof used in [69] works with minor changes also for more general functions and matrix classes, as we shall see. The proof of (15) is based on a classical result of Chebyshev on the uniform approximation error

$$\min \max_{a \leq x \leq b} |p_k(x) - x^{-1}|$$

(where the minimum is taken over all polynomials p_k of degree $\leq k$), according to which the error decays exponentially in the degree k as $k \to \infty$. Combined with the spectral theorem (which allows to go from scalar functions to matrix functions, with the $\| \cdot \|_2$ matrix norm replacing the $\| \cdot \|_\infty$ norm), this result gives the exponential decay bound for $[A^{-1}]_{ij}$. A crucial ingredient of the proof is the fact that if A is m-banded, then A^k is km-banded, for all $k = 0, 1, 2, \ldots$.

The paper of Demko et al. also contains some extensions to the case of non-Hermitian matrices and to matrices with a general sparsity pattern. Invertible, non-Hermitian matrices are dealt with by observing that for any $A \in \mathbb{C}^{n \times n}$ one can write

$$A^{-1} = A^*(AA^*)^{-1} \tag{20}$$

and that if A is banded, then the Hermitian positive definite matrix AA^* is also banded (albeit with a larger bandwidth). It is not difficult to see that the product of two matrices, one of which is banded and the other has entries that satisfy an exponential decay bound, is also a matrix with entries that satisfy an exponential decay bound.

For a general sparse matrix, the authors of [69] observe that the entries of A^{-1} are bounded in an exponentially decaying manner away from the support (nonzero pattern) of A. This fact can be expressed in the form

$$|[A^{-1}]_{ij}| \leq K e^{-\alpha d(i,j)}, \quad \forall i, j, \tag{21}$$

where $d(i, j)$ is the geodesic distance between nodes i and j in the undirected graph $\mathcal{G}(A)$ associated with A.

Results similar to those in [69] where independently obtained by Jaffard [119], motivated by problems concerning wavelet expansions. In this paper Jaffard proves

exponential decay bounds for the entries of A^{-1} and mentions that similar bounds can be obtained for other matrix functions, such as $A^{-1/2}$ for A positive definite. Moreover, the bounds are formulated for (in general, infinite) matrices the entries of which are indexed by the elements of a suitable metric space, allowing the author to obtain decay results for the inverses of matrices with arbitrary nonzero pattern and even of dense matrices with decaying entries (we will return to this topic in section 3.8).

The exponential decay bound (15) together with Theorem 4 implies the following (asymptotic) uniform approximation result.

Theorem 5. *Let* $\{A_n\}$ *be a sequence of* $n \times n$ *matrices, all Hermitian positive definite and m-banded. Assume that there exists an interval* $[a, b]$, $0 < a < b < \infty$, *such that* $\sigma(A_n) \subset [a, b]$, *for all* n. *Then, for all* $\varepsilon > 0$ *and for all* n *there exist an integer* $p = p(\varepsilon, m, a, b)$ *(independent of* n*) and a matrix* $B_n = B_n^*$ *with bandwidth* p *such that* $\|A_n^{-1} - B_n\|_2 < \varepsilon$.

The smallest value of the bandwidth p needed to satisfy the prescribed accuracy can be easily computed via (14). As an example, for tridiagonal matrices A_n ($m = 1$), $K = 10$, $\alpha = 0.6$ (which corresponds to $\rho \approx 0.5488$) we find $\|A_n^{-1} - B_n\|_2 < 10^{-6}$ for all $p \geq 29$, regardless of n. In practice, of course, this result is of interest only for $n > p$ (in fact, for $n \gg p$).

We note that a similar result holds for sparse matrix sequences $\{A_n\}$ corresponding to a sequence of graphs $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ satisfying the assumption (11) of bounded maximum degree. In this case the matrices B_n will be sparse rather than banded, with a maximum number p of nonzeros per row which does not depend on n; in other words, the graph sequence $\mathcal{G}(B_n)$ associated with the matrix sequence $\{B_n\}$ will also satisfy a condition like (11).

The proof of the decay bound (15) shows that for any prescribed value of $\varepsilon > 0$, each inverse matrix A_n^{-1} can be approximated within ε (in the 2-norm) by a polynomial $p_k(A_n)$ of degree k in A_n, with k independent of n. To this end, it suffices to take the (unique) polynomial of best approximation of degree k of the function $f(x) = x^{-1}$, with k large enough that the error satisfies

$$\max_{a \leq x \leq b} |p_k(x) - x^{-1}| < \varepsilon.$$

In this very special case an exact, closed form expression for the approximation error is known ; see [150, pages 33–34]. This expression yields an upper bound for the error $\|p_k(A_n) - A_n^{-1}\|_2$, uniform in n. Provided that the assumptions of Theorem 5 are satisfied, the degree k of this polynomial does not depend on n, but only on ε. This shows that it is in principle possible to approximate A_n^{-1} using only $O(n)$ arithmetic operations and storage.

Remark 2. The polynomial of best approximation to the function $f(x) = x^{-1}$ found by Chebyshev does not yield a practically useful expression for the explicit approximation of A^{-1}. However, observing that for any invertible matrix A and any polynomial p

$$\frac{\|A^{-1} - p(A)\|_2}{\|A^{-1}\|_2} \leq \|I - p(A)A\|_2,$$

we can obtain an upper bound on the *relative* approximation error by finding the polynomial of smallest degree k for which

$$\max_{a \le x \le b} |1 - p_k(x)x| = \min . \tag{22}$$

Problem (22) admits an explicit solution in terms of shifted and scaled Chebyshev polynomials; see, e.g., [177, page 381]. Other procedures for approximating the inverse will be briefly mentioned in section 4.1.

A number of improvements, extensions, and refinements of the basic decay results by Demko et al. have been obtained by various authors, largely motivated by applications in numerical analysis, mathematical physics and signal processing, and the topic continues to be actively researched. Decay bounds for the inverses of M-matrices that are near to Toeplitz matrices (a structure that arises frequently in the numerical solution of partial differential equations) can be found in Eijkhout and Polman [76]. Freund [85] obtains an exponential decay bound for the entries of the inverse of a banded matrix A of the form

$$A = c\,I + d\,T, \quad T = T^*, \quad c, d \in \mathbb{C}.$$

Exponential decay bounds for resolvents ad eigenvectors of infinite banded matrices were obtained by Smith [185]. Decay bounds for the inverses of nonsymmetric band matrices can be found in a paper by Nabben [156]. The paper by Meurant [153] provides an extensive treatment of the tridiagonal and block tridiagonal cases. Inverses of triangular Toeplitz matrices arising from the solution of integral equations also exhibit interesting decay properties; see [84].

A recent development is the derivation of bounds that accurately capture the oscillatory decay behavior observed in the inverses of sparse matrices arising from the discretization of multidimensional partial differential equations. In [48], Canuto et al. obtain bounds for the inverse of matrices in *Kronecker sum* form, i.e., matrices of the type

$$A = T_1 \oplus T_2 := T_1 \otimes I + I \otimes T_2, \tag{23}$$

with T_1 and T_2 banded (for example, tridiagonal). For instance, the 5-point finite difference scheme for the discretization of the Laplacian on a rectangle produces matrices of this form. Generalization to higher-dimensional cases (where A is the Kronecker sum of three or more banded matrices) is also possible.

3.3 Decay bounds for the matrix exponential

As we have seen in the Introduction, the entries in the exponential of banded matrices can exhibit rapid off-diagonal decay (see Fig. 2). As it turns out, the actual decay rate is faster than exponential (the term *superexponential* is often used), a phenomenon common to all entire functions of a matrix. More precisely, we have the following definition.

Definition 2. *A matrix A has the* superexponential off-diagonal decay property *if for any $\alpha > 0$ there exists a $K > 0$ such that*

$$|[A]_{ij}| \le K \mathrm{e}^{-\alpha|i-j|} \qquad \forall i, j.$$

As usual, in this definition A is either infinite or a member of a sequence of matrices of increasing order, in which case K and α do not depend on the order.

The definition can be readily extended to decay with respect to a general nonzero pattern, in which case $|i - j|$ must be replaced by the geodesic distance on the corresponding graph.

A superexponential decay bound on the entries of the exponential of a *tridiagonal* matrix has been obtained by Iserles [117]. The bound takes the form

$$|[\mathrm{e}^A]_{ij}| \le \mathrm{e}^\rho I_{|i-j|}(2\rho), \quad i, j = 1, \dots, n \tag{24}$$

where $\rho = \max_{i,j} |[A_{ij}]|$ and $I_\nu(z)$ is the *modified Bessel function of the first kind*:

$$I_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^\infty \frac{\left(\frac{1}{4}z^2\right)^k}{k!\,\Gamma(\nu + k + 1)},$$

where $\nu \in \mathbb{R}$ and Γ is the gamma function; see [1]. For any fixed value of $z \in \mathbb{C}$, the values of $|I_\nu(z)|$ decay faster than exponentially for $\nu \to \infty$. The paper by Iserles also presents superexponential decay bounds for the exponential of more general banded matrices, but the bounds only apply at sufficiently large distances from the main diagonal. None of these bounds require A to be Hermitian.

In [25], new decay bounds for the entries of the exponential of a banded, Hermitian, positive semidefinite matrix A have been presented. The bounds are a consequence of fundamental error bounds for Krylov subspace approximations to the matrix exponential due to Hochbruch and Lubich [111]. The decay bounds are as follows.

Theorem 6. ([25]) *Let A be a Hermitian positive semidefinite matrix with eigenvalues in the interval $[0, 4\rho]$ and let $\tau > 0$. Assume in addition that A is m-banded. For $i \ne j$, let $\xi = \lceil |i - j|/m \rceil$. Then*
  i) For $\rho\tau \ge 1$ and $\sqrt{4\rho\tau} \le \xi \le 2\rho\tau$,

$$|[\exp(-\tau A)]_{ij}| \le 10 \exp\left(-\frac{1}{5\rho\tau}\xi^2\right);$$

ii) For $\xi \ge 2\rho\tau$,

$$|[\exp(-\tau A)]_{ij}| \le 10 \frac{\exp(-\rho\tau)}{\rho\tau} \left(\frac{\mathrm{e}\rho\tau}{\xi}\right)^\xi.$$

As shown in [25], these bounds are quite tight and capture the actual superexponential decay behavior very well. Similar bounds can be derived for the skew-Hermitian case ($A = -A^*$). See also [180], where decay bounds are derived for the exponential of a class of *unbounded* infinite skew-Hermitian tridiagonal matrices arising in quantum mechanical problems, and [203].

These bounds can also be adapted to describe the decay behavior of the exponential of matrices with a general sparsity pattern. See Fig. 8 for an example.

Bounds for the matrix exponential in the nonnormal case will be discussed in section 3.4 below, as special cases of bounds for general analytic functions of matrices.

We note that exploiting the well known identity

$$\exp(A \oplus B) = \exp(A) \otimes \exp(B) \tag{25}$$

(see [109, Theorem 10.9]), it is possible to use Theorem 6 to obtain bounds for the exponential of a matrix that is the Kronecker sum of two (or more) banded matrices; these bounds succeed in capturing the oscillatory decay behavior in the exponential of such matrices (see [25]).
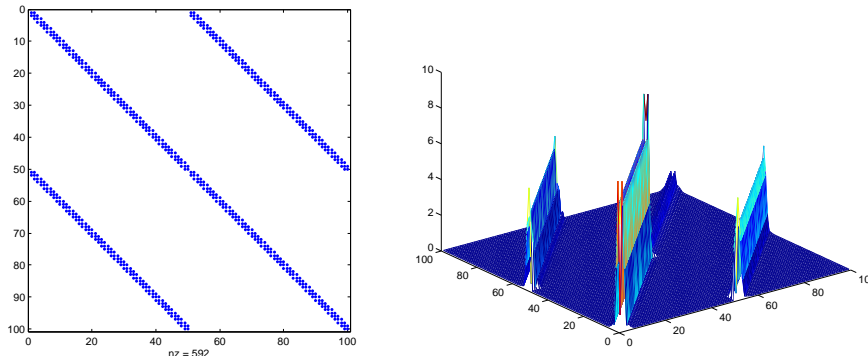
Fig. 8. Sparsity pattern of multi-banded matrix A and decay in e^A.

3.4 Decay bounds for general analytic functions

In this section we present decay bounds for the entries of matrix functions of the form $f(A)$ where f is analytic on an open connected set $\Omega \subseteq \mathbb{C}$ with $\sigma(A) \subset \Omega$ and A is banded or sparse. These bounds are obtained combining classical results on the approximation of analytic functions by polynomials with the spectral theorem, similar to the approach used by Demko et al. in [69] to prove exponential decay in the inverses of banded matrices. The classical Chebyshev expression for the error incurred by the polynomials of best approximation (in the infinity norm) of $f(x) = x^{-1}$ will be replaced by an equally classical bound (due to S. N. Bernstein) valid for arbitrary analytic functions. The greater generality of Bernstein's result comes at a price: instead of having an exact expression for the approximation error, it provides only an upper bound. This is sufficient, however, for our purposes.

We begin with the Hermitian case.[6] If $[a, b] \subset \mathbb{R}$ denotes any interval containing the spectrum of a (possibly infinite) matrix $A = A^*$, the shifted and scaled matrix

$$\hat{A} = \frac{2}{b - a}A - \frac{a + b}{a - b}I \tag{26}$$

has spectrum contained in $[-1, 1]$. Since decay bounds are simpler to express for functions of matrices with spectrum contained in $[-1, 1]$ than in a general interval $[a, b]$, we will make the assumption that A has already been scaled and shifted so that $\sigma(A) \subseteq [-1, 1]$. It is in general not difficult to translate the decay bounds in terms of the original matrix, if required. In practice it is desirable that $[-1, 1]$ is the smallest interval containing the spectrum of the scaled and shifted matrix.

Given a function f continuous on $[-1, 1]$ and a positive integer k, the kth best approximation error for f by polynomials is the quantity

$$E_k(f) = \inf\left\{\max_{-1 \leq x \leq 1} |f(x) - p(x)| : p \in \mathbb{P}_k\right\},$$

---

[6] The treatment is essentially the same for any normal matrix with eigenvalues lying on a line segment in the complex plane, in particular if A is skew-Hermitian.

where $\mathbb{P}_k$ is the set of all polynomials of degree less than or equal to k. Bernstein's Theorem describes the asymptotic behavior of the best polynomial approximation error for a function f analytic on a domain containing the interval $[-1, 1]$.

Consider now the family of ellipses in the complex plane with foci in -1 and 1. Any ellipse in this family is completely determined by the sum $\chi > 1$ of the lengths of its half-axes; if these are denoted by $\kappa_1 > 1$ and $\kappa_2 > 0$, it is well known that

$$\sqrt{\kappa_1^2 - \kappa_2^2} = 1, \quad \kappa_1 - \kappa_2 = 1/(\kappa_1 + \kappa_2) = 1/\chi\,.$$

We will denote the ellipse characterized by $\chi > 1$ by $\mathcal{E}_\chi$.

If f is analytic on a region (open simply connected subset) of $\mathbb{C}$ containing $[-1, 1]$, then there exists an infinite family of ellipses $\mathcal{E}_\chi$ with $1 < \chi < \bar{\chi}$ such that f is analytic in the interior of $\mathcal{E}_\chi$ and continuous on $\mathcal{E}_\chi$. Moreover, $\bar{\chi} = \infty$ if and only if f is entire.

The following fundamental result is known as *Bernstein's Theorem*.

Theorem 7. *Let the function f be analytic in the interior of the ellipse $\mathcal{E}_\chi$ and continuous on $\mathcal{E}_\chi$, for $\chi > 1$. Then*

$$E_k(f) \leq \frac{2M(\chi)}{\chi^k(\chi - 1)},$$

where $M(\chi) = \max_{z \in \mathcal{E}_\chi} |f(z)|$.

Proof. See, e.g., [145, Chapter 3.15].

Hence, if f is analytic, the error corresponding to polynomials of best approximation in the uniform convergence norm decays exponentially with the degree of the polynomial. As a consequence, we obtain the following exponential decay bounds on the entries of $f(A)$. We include the proof (modeled after the one in [69]) as it is instructive.

Theorem 8. ([22]) *Let $A = A^*$ be m-banded with spectrum $\sigma(A)$ contained in $[-1, 1]$ and let f be analytic in the interior of $\mathcal{E}_\chi$ and continuous on $\mathcal{E}_\chi$ for $1 < \chi < \bar{\chi}$. Let*

$$\rho := \chi^{-\frac{1}{m}}, \quad M(\chi) = \max_{z \in \mathcal{E}_\chi} |f(z)|, \quad \text{and} \quad K = \frac{2\chi M(\chi)}{\chi - 1}\,.$$

Then

$$|[f(A)]_{ij}| \leq K\,\rho^{|i-j|}, \quad \forall\, i, j. \tag{27}$$

Proof. Let p_k be the polynomial of degree k of best uniform approximation for f on $[-1, 1]$. First, observe that if A is m-banded then A^k (and therefore $p_k(A)$) is km-banded: $[p_k(A)]_{ij} = 0$ if $|i - j| > km$. For $i \neq j$ write $|i - j| = km + l$, $l = 1, 2, \ldots, m$, hence $k < |i - j|/m$ and $\chi^{-k} < \chi^{-\frac{|i-j|}{m}} = \rho^{|i-j|}$. Therefore, for all $i \neq j$ we have

$$|[f(A)]_{ij}| = |[f(A)]_{ij} - [p_k(A)]_{ij}| \leq \|f(A) - p_k(A)\|_2 \leq \|f - p_k\|_\infty \leq K\rho^{|i-j|}.$$

The last inequality follows from Theorem 7. For $i = j$ we have $|[f(A)]_{ii}| \leq \|f(A)\|_2 < K$ (since $2\chi/(\chi - 1) > 1$ and $\|f(A)\|_2 \leq M(\chi)$ for all $\chi > 1$ by the maximum principle). Therefore the bound (27) holds for all i, j.

Note that the bound can also be expressed as

$$|[f(A)]_{ij}| \leq K \, \mathrm{e}^{-\alpha|i-j|}, \quad \forall i, j,$$

by introducing $\alpha = -\log(\rho) > 0$.

Remark 3. An important difference between (27) and bound (15) is that (27) actually represents an infinite family of bounds, one for every $\chi \in (0, \bar{\chi})$. Hence, we cannot expect (27) to be sharp for any fixed value of χ. There is a clear trade-off involved in the choice of χ; larger values of χ result in faster exponential decay (smaller ρ) and smaller values of $2\chi/(\chi - 1) > 1$ (which is a monotonically decreasing function of χ for $\chi > 1$), but potentially much larger values of $M(\chi)$. In particular, as χ approaches $\bar{\chi}$ from below, we must have $M(\chi) \to \infty$. As noted in [20, pp. 27–28] and [180, p. 70], for any entry (i, j) of interest the bound (27) can be optimized by finding the value of $\chi \in (0, \bar{\chi})$ that minimizes the right-hand side of (27); for many functions of practical interest there is a unique minimizer which can be found numerically if necessary.

Remark 4. Theorem 8 can be applied to both finite matrices and bounded infinite matrices on ℓ^2. Note that infinite matrices may have continuous spectrum, and indeed it can be $\sigma(A) = [-1, 1]$. The result is most usefully applied to matrix sequences $\{A_n\}$ of increasing size, all m-banded (or with bandwidth $\leq m$ for all n) and such that

$$\bigcup_{n=1}^{\infty} \sigma(A_n) \subset [-1, 1],$$

assuming f is analytic on a region $\Omega \subseteq \mathbb{C}$ containing $[-1, 1]$ in its interior. For instance, each A_n could be a finite section of a bounded infinite matrix A on ℓ^2 with $\sigma(A) \subseteq [-1, 1]$. The bound (27) then becomes

$$|[f(A_n)]_{ij}| \leq K \, \rho^{|i-j|} \quad \forall i, j, \quad \forall n \in \mathbb{N}. \tag{28}$$

In other words, the bounds (28) are uniform in n. Analogous to Theorem 5, it follows that under the conditions of Theorem 8, for any prescribed $\varepsilon > 0$ there exists a positive integer p and a sequence of p-banded matrices $B_n = B_n^*$ such that

$$\|f(A_n) - B_n\|_2 < \varepsilon.$$

Moreover, the proof of Theorem 8 shows that each B_n can be taken to be a polynomial in A_n, which does not depend on n. Therefore, it is possible in principle to approximate $f(A_n)$ with arbitrary accuracy in $O(n)$ work and storage.

We emphasize again that the restriction to the interval $[-1, 1]$ is done for ease of exposition only; in practice, it suffices that there exists a bounded interval $\mathcal{I} = [a, b] \subset \mathbb{R}$ such that $\sigma(A_n) \subset [a, b]$ for all $n \in \mathbb{N}$. In this case we require f to be analytic on a region of $\mathbb{C}$ containing $[a, b]$ in its interior. The result can then be applied to the corresponding shifted and scaled matrices $\hat{A}_n$ with spectrum in $[-1, 1]$, see (26). The following example illustrates how to obtain the decay bounds expressed in terms of the original matrices in a special case.

Example 1. The following example is taken from [22]. Assume that $A = A^*$ is m-banded and has spectrum in $[a, b]$ where $b > a > 0$, and suppose we want to obtain decay bounds on the entries of $A^{-1/2}$. Note that there is an infinite family of ellipses $\{\mathscr{E}_\xi\}$ entirely contained in the open half plane with foci in a and b, such that the function $F(z) = z^{-1/2}$ is analytic on the interior of each $\mathscr{E}_\xi$ and continuous on it. If ψ denotes the linear affine mapping

$$\psi(z) = \frac{2z - (a + b)}{b - a}$$

which maps $[a, b]$ to $[-1, 1]$, we can apply Theorem 8 to the function $f = F \circ \psi^{-1}$, where

$$\psi^{-1}(w) = \frac{(b - a)w + a + b}{2} \ .$$

Obviously, f is analytic on the interior of a family $\mathcal{E}_\chi$ of ellipses (images via ψ of the $\mathscr{E}_\xi$) with foci in $[-1, 1]$ and continuous on each $\mathcal{E}_\chi$, with $1 < \chi < \bar{\chi}$. An easy calculation shows that

$$\bar{\chi} = \frac{b + a}{b - a} + \sqrt{\left(\frac{b + a}{b - a}\right)^2 - 1} = \frac{(\sqrt{\kappa + 1})^2}{\kappa - 1},$$

where $\kappa = \frac{b}{a}$. Finally, for any $\chi \in (1, \bar{\chi})$ we easily find (recalling that $\chi = \kappa_1 + \kappa_2$)

$$M(\chi) = \max_{z \in \mathcal{E}_\chi} |f(z)| = |f(-\kappa_1)| = \frac{\sqrt{2}}{\sqrt{(a - b)\kappa_1 + a + b}} = \frac{2}{\sqrt{\frac{(a-b)(\chi^2+1)}{2\chi} + a + b}}.$$

It is now possible to compute the bounds (27) for any $\chi \in (1, \bar{\chi})$ and for all i, j. Note that if b is fixed and $a \to 0+$, $M(\chi)$ grows without bound and $\rho \to 1-$, showing that the decay bound deteriorates as A becomes nearly singular. Conversely, for well-conditioned A decay can be very fast, since $\bar{\chi}$ will be large for small conditioned numbers κ. This is analogous to the situation for A^{-1}.

More generally, the decay rate in the bound (27) depends on the distance between the singularities of f (if any) and the interval $[a, b]$ (and, of course, on the bandwidth m). If f has any singularities near $[a, b]$ then $\bar{\chi}$ will be necessarily close to 1, and the bound (27) will decay very slowly. Conversely, if they are far from $[a, b]$ then χ can be taken large and decay will be fast.

In the case of an entire function, χ can be taken arbitrarily large, so that the exponential decay part of the bound decays arbitrarily fast; note, however, that this will cause K to increase. Thus, it is clear that for f entire and A banded, the entries of $f(A)$ are bounded in a superexponentially decay manner according to Definition 2; see [180]. As a special case, we have an alternative proof of the superexponential decay for the matrix exponential. Note, however, that in the case of the matrix exponential the specialized bounds given in Theorem 6 are generally tighter.

Remark 5. Let now $\{A_n\}$ be a sequence of m-banded matrices of increasing size. It is clear that if $\sigma(A_n)$ is not bounded away from the singularities of f for all n, then we cannot expect to have uniform decay bounds like (27) valid for all n. The same happens in the case of a (non-constant) entire function f if the smallest interval containing $\sigma(A_n)$ is unbounded as $n \to \infty$. Hence, the bounds (27) cannot

be expect to hold uniformly for matrices A_n arising from the discretization of unbounded differential operators if the size n is related to the mesh size h (in the sense that $n \to \infty$ if $h \to 0$). Nevertheless, we will see that there are important applications where the decay bounds (8) hold uniformly in n.

As in the case of the inverse, the bounds (8) can be extended, with some caution, from the banded case to the case of matrices with more general sparsity patterns. We formally state this result as follows.

Theorem 9. ([20]) *Let $\{A_n\}$ be a sequence of sparse Hermitian matrices of increasing size. Assume that there exists a bounded interval $[a,b] \subset \mathbb{R}$ such that $\sigma(A_n) \subset [a,b]$ for all $n \in \mathbb{N}$, and that the sequence of graphs $\mathcal{G}(A_n)$ has bounded maximum degree. If f is analytic on a region containing $[a,b]$ in its interior, there exist constants $K > 0$ and $\alpha > 0$ such that*

$$|[f(A_n)]_{ij}| \le K \, \mathrm{e}^{-\alpha d_n(i,j)}, \quad \forall \, i,j, \ \forall n \in \mathbb{N}, \tag{29}$$

where d_n denotes the geodesic distance on $\mathcal{G}(A_n)$. The constants K and α depend on $[a,b]$ and on the maximum degree of any node in the graph sequence $\{\mathcal{G}(A_n)\}$, but not on n.

As before, (29) is actually an infinite family of bounds parameterized by χ, the sum of the semi-axes of the infinitely many ellipses with foci in a and b entirely contained in the largest simply connected region Ω on which f is analytic. The expressions for K and α (equivalently, ρ) are exactly as in Theorem 8 when $a = -1$ and $b = -1$, otherwise they can be found as shown in Example 1.

Theorem 9 also holds if the sequence $\{A_n\}$ is replaced by a single bounded infinite matrix acting on ℓ^2 such that $\sigma(A) \subseteq [a,b]$ and such that the infinite graph $\mathcal{G}(A)$ has finite maximum degree.

Remark 6. The proof of Theorem 8 shows that the bounds (27) and (29) are in general pessimistic. Indeed, much can be lost when bounding $|[f(A)]_{ij} - [p_k(A)]_{ij}|$ with $\|f(A) - p_k(A)\|_2$ and the latter quantity with $\|f - p_k\|_\infty$. In particular, these bounds completely ignore the distribution of the eigenvalues of A in the interval $[-1, 1]$; in this sense, the situation is completely analogous to the well known error estimate for the A-norm of the error in the conjugate gradient method based on the condition number of A, see [96, page 636]. It is is well known that this bound, while sharp, can be overly conservative. The same holds for (29): the bound on the rate of decay depends on a single essential quantity, the distance between the spectrum of A and the singularities of f, and thus cannot be expected to be very accurate. More accurate bounds can likely be obtained given more information on the spectral distribution of A; for example, if most of the eigenvalues of A are tightly clustered, then the decay rate in $f(A)$ should be much faster than if the eigenvalues of A are distributed more or less evenly in the spectral interval. In the limiting case where A has only $k \ll n$ distinct eigenvalues (so that the minimum polynomial of A has degree k), then $f(A)$ can be represented exactly by a low degree polynomial, and many of the entries of A will be exactly zero as long as $\mathrm{diam}(\mathcal{G}(A)) \gg k$.

Bounds for the normal case. Owing to the fact that the spectral theorem applies not just to Hermitian matrices but more generally to normal matrices, it is not surprising that results completely analogous to Theorem 8 and Theorem 9 can be stated and proved assuming that $\{A_n\}$ is a sequence of normal matrices (banded or sparse) with eigenvalues lying on a line segment $[z_1, z_2] \subset \mathbb{C}$ entirely contained in a region Ω on which f is analytic. For instance, decay bounds for the entries of functions of banded skew-symmetric matrices have been given in [67, 143].

More generally, suppose A is normal and m-banded. Let $\mathcal{F} \subset \mathbb{C}$ be a compact, connected region containing $\sigma(A)$, and denote by $\mathbb{P}_k$ the set of complex polynomials of degree at most k. Then the argument in the proof of Theorem 8 still holds, except that now polynomial approximation is no longer applied on an interval, but on the complex region $\mathcal{F}$. Therefore, the following bound holds for all indices i, j such that $|i - j| > km$:

$$|[f(A)]_{ij}| \leq \max_{\lambda \in \sigma(A)} |f(\lambda) - p(\lambda)| \leq E_k(f, \mathcal{F}), \tag{30}$$

where

$$E_k(f, \mathcal{F}) := \min_{p \in \mathbb{P}_k} \max_{z \in \mathcal{F}} |f(z) - p(z)| =: \min_{p \in \mathbb{P}_k} \|f - p\|_{\infty, \mathcal{F}}.$$

Unless more accurate estimates for $\sigma(A)$ are available, a possible choice for $\mathcal{F}$ is the disk of center 0 and radius $\varrho(A)$.

If f is analytic on $\mathcal{F}$, bounds for $E_k(f, \mathcal{F})$ that decay exponentially with k are available through the use of Faber polynomials: see [24, Theorem 3.3] and the next subsection for more details. More precisely, there exist constants $\tilde{c} > 0$ and $0 < \tilde{\rho} < 1$ such that $E_k(f, \mathcal{F}) \leq \tilde{c} \, \tilde{\rho}^k$ for all $k \in \mathbb{N}$. This result, together with (30), yields for all i and j the bound

$$|[f(A)]_{ij}| \leq K \, \rho^{|i-j|} = K \, \mathrm{e}^{-\alpha|i-j|} \tag{31}$$

(where $\alpha = -\log(\rho)$) for suitable constants $K > 0$ and $0 < \rho < 1$, which do not depend on the size of the matrix n, although they generally depend on f and $\mathcal{F}$ (and of course on the bandwidth, m, in the case of ρ).

The extension of these bounds to sparse matrices with more general sparsity patterns is entirely straightforward; note, however, that unless A is structurally symmetric (in which case the graph $\mathcal{G}(A)$ is undirected), the distance $d(i, j)$, defined as the length of the shortest directed path starting at node i and ending at node j, will be different, in general, from $d(j, i)$. Hence, different rates of decay may be observed on either side of the main diagonal.

Bounds for the nonnormal case. As can be expected, the derivation of decay bounds for the entries of $f(A)$ when A is banded or sparse and nonnormal is more challenging than in the normal case, since in this case the spectral theorem can no longer be relied upon.

It is easy to give examples where decay fails to occur. The simplest one is probably the following. Let A_n be the $n \times n$ upper bidiagonal matrix

$$A_n = \begin{bmatrix} 1 & -\alpha & 0 & \dots & 0 \\ 0 & 1 & -\alpha & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & -\alpha \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \tag{32}$$

where $\alpha \in \mathbb{R}$. Then

$$
A_n^{-1} = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ 0 & 1 & \alpha & \dots & \alpha^{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \alpha \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} .
$$

Hence, for $\alpha \geq 1$ no decay is present[7] in the upper triangular part of A_n^{-1}.

Nevertheless, useful bounds can still be obtained in many cases. In order to proceed, we need some additional background in approximation theory, namely, some notions about *Faber polynomials* and their use in the approximation of analytic functions on certain regions of the complex plane. In a nutshell, Faber polynomials play for these regions the same role played by Taylor polynomials for disks. The following discussion is taken from [24] and follows closely the treatment in [145].

A *continuum* in $\mathbb{C}$ is a nonempty, compact and connected subset of $\mathbb{C}$. Let F be a continuum consisting of more than one point. Let G_∞ denote the component of the complement of F containing the point at infinity. Note that G_∞ is a simply connected domain in the extended complex plane $\bar{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. By the Riemann Mapping Theorem there exists a function $w = \Phi(z)$ which maps G_∞ conformally onto a domain of the form $|w| > \rho > 0$ satisfying the normalization conditions

$$
\Phi(\infty) = \infty, \quad \lim_{z \to \infty} \frac{\Phi(z)}{z} = 1; \tag{33}
$$

ρ is the *logarithmic capacity* of F. Given any integer $N > 0$, the function $[\Phi(z)]^N$ has a Laurent series expansion of the form

$$
[\Phi(z)]^N = z^N + \alpha_{N-1}^{(N)} z^{N-1} + \dots + \alpha_0^{(N)} + \frac{\alpha_{-1}^{(N)}}{z} + \dots \tag{34}
$$

at infinity. The polynomials

$$
\Phi_N(z) = z^N + \alpha_{N-1}^{(N)} z^{N-1} + \dots + \alpha_0^{(N)}
$$

consisting of the terms with nonnegative powers of z in the expansion (34) are called the *Faber polynomials* generated by the continuum F.

Let Ψ be the inverse of Φ. By C_R we denote the image under Ψ of a circle $|w| = R > \rho$. The (Jordan) region with boundary C_R is denoted by $I(C_R)$. By Theorem 3.17, p. 109 of [145], every function f analytic on $I(C_{R_0})$ with $R_0 > \rho$ can be expanded in a series of Faber polynomials:

$$
f(z) = \sum_{k=0}^{\infty} \alpha_k \Phi_k(z), \tag{35}
$$

where the series converges uniformly inside $I(C_{R_0})$. The coefficients are given by

---

[7] At first sight, this example seems to contradict the result by Demko et al. [69] based on the identity (20). However, the result of Demko et al. assumes that the condition number of AA^* (equivalently, of A itself) remains bounded as $n \to \infty$, which is not the case for this example.

$$\alpha_k = \frac{1}{2\pi i} \int_{|w|=R} \frac{f(\Psi(w))}{w^{k+1}} \mathrm{d}w$$

where $\rho < R < R_0$. We denote the partial sums of the series in (35) by

$$\Pi_N(z) := \sum_{k=0}^{N} \alpha_k \Phi_k(z). \tag{36}$$

Each $\Pi_N(z)$ is a polynomial of degree at most N, since each $\Phi_k(z)$ is of degree k. We are now ready to state the following generalization of Theorem 7, also due to Bernstein, which will be instrumental for what follows.

Theorem 10. *Let f be a function defined on F. Then given any $\varepsilon > 0$ and any integer $N \geq 0$, there exists a polynomial Π_N of degree at most N and a positive constant $c(\varepsilon)$ such that*

$$|f(z) - \Pi_N(z)| < c(\varepsilon)(q + \varepsilon)^N \quad (0 < q < 1) \tag{37}$$

for all $z \in F$ if and only if f is analytic on the domain $I(C_{R_0})$, where $R_0 = \rho/q$. In this case, the sequence $\{\Pi_N\}$ converges uniformly to f inside $I(C_{R_0})$ as $N \to \infty$.

In the special case where F is a disk of radius ρ centered at z_0, Theorem 10 states that for any function f analytic on the disk of radius ρ/q centered at z_0, where $0 < q < 1$, there exists a polynomial Π_N of degree at most N and a positive constant $c(\varepsilon)$ such that for any $\varepsilon > 0$

$$|f(z) - \Pi_N(z)| < c(\varepsilon)(q + \varepsilon)^N, \tag{38}$$

for all $z \in F$. We are primarily concerned with the sufficiency part of Theorem 10. Note that the choice of q (with $0 < q < 1$) depends on the region where the function f is analytic. If f is defined on a continuum F with logarithmic capacity ρ then we can pick q bounded away from 1 as long as the function is analytic on $I(C_{\rho/q})$. Therefore, the rate of convergence is directly related to the properties of the function f, such as the location of its poles (if there are any). Following [145], the constant $c(\varepsilon)$ can be estimated as follows. Let R_0, q and ε be given as in Theorem 10. Furthermore, let R' and R be chosen such that $\rho < R' < R < R_0$ and

$$\frac{R'}{R} = q + \varepsilon,$$

then we define

$$M(R) = \max_{z \in C_R} |f(z)|.$$

An estimate for the value of $c(\varepsilon)$ is asymptotically (i.e., for sufficiently large N) given by

$$c(\varepsilon) \approx \frac{3}{2} M(R) \frac{q + \varepsilon}{1 - (q + \varepsilon)}.$$

It may be necessary to replace the above expression for $c(\varepsilon)$ by a larger one to obtain validity of the bound (37) for all N. However, for certain regions (and in particular for convex F) it is possible to obtain an explicit constant valid for all $N \geq 0$; see [78] and [24, Section 3.7]. Based on this theorem, we can state the following result.

Theorem 11. *Let $\{A_n\}$ be a sequence of $n \times n$ diagonalizable matrices and assume that $\sigma(A_n)$ is contained in a continuum F, for all n. Assume further that the matrices A_n are sparse and that each graph $\mathcal{G}(A_n)$ satisfies the maximum bounded degree assumption. Let $\kappa_2(X_n)$ be the spectral condition number of the eigenvector matrix of A_n. Let f be a function defined on F. Furthermore, assume that f is analytic on $I(C_{R_0})$ $(\supset \sigma(A_n))$, where $R_0 = \frac{\rho}{q}$ with $0 < q < 1$ and ρ is the logarithmic capacity of F. Then there are positive constants K and α, independent of n, such that*

$$|[f(A_n)]_{ij}| < \kappa(X_n)\,K\mathrm{e}^{-\alpha\,d_n(i,j)}, \quad \forall\,i,j, \quad \forall\,n \in \mathbb{N}, \tag{39}$$

where d_n denotes the geodesic distance on $\mathcal{G}(A_n)$.

Proof. From Theorem 10 we know that for any $\varepsilon > 0$ there exists a sequence of polynomials Π_k of degree k which satisfies for all $z \in F$

$$|f(z) - \Pi_k(z)| < c(\varepsilon)(q + \varepsilon)^k, \quad \text{where} \quad 0 < q < 1.$$

Therefore, since $A_n = X_n D_n X_n^{-1}$ with D_n diagonal, we have

$$\|f(A_n) - \Pi_k(A_n)\|_2 \leq \kappa_2(X_n) \max_{z \in \sigma(A_n)} |f(z) - \Pi_k(z)| < \kappa_2(X_n)c(\varepsilon)(q + \varepsilon)^k,$$

where $0 < q < 1$. For $i \neq j$ we can write

$$d_n(i,j) = k + 1$$

and therefore, observing that $[\Pi_k(A_n)]_{ij} = 0$ for $d_n(i,j) > k$, we have

$$|[f(A_n)]_{ij}| = |[f(A_n)]_{ij} - [\Pi_k(A_n)]_{ij}| \leq \|f(A_n) - \Pi_k(A_n)\|_2 < \kappa_2(X_n)c(\varepsilon)(q+\varepsilon)^{d_n(i,j)-1}.$$

Hence, choosing $\varepsilon > 0$ such that $\rho_0 := q + \varepsilon < 1$ and letting $K_0 = c(\varepsilon)/(q + \varepsilon)$ we obtain

$$|[f(A_n)]_{ij}| < \kappa_2(X_n)\,K_0\,\rho_0^{d_n(i,j)}.$$

If $i = j$ then $|[f(A_n)]_{ii}| \leq \|f(A_n)\|_2$ and therefore letting $K = \max\{K_0, \|f(A)\|_2\}$ and $\alpha = -\log(\rho_0)$ we see that inequality (39) holds for all i, j.

It is worth noting that in the normal case we have $\kappa_2(X_n) = 1$ in (39), and therefore the bound (31) is proved. Bound (39) may also prove useful if the spectral condition numbers $\kappa_2(X_n)$ are uniformly bounded by a (moderate) constant. However, in the case of a highly nonnormal sequence (for which the $\kappa_2(X_n)$ are necessarily very large, see [96, P7.2.3]) the bound is virtually useless as it can can be a severe overestimate of the actual size of the elements of $f(A_n)$; see [24, page 25] for an example. The situation is entirely analogous to the standard residual bound for GMRES applied to diagonalizable matrices; see, e.g., [177, Proposition 6.32].

In this case a different approach, based on the field of values and not requiring diagonalizability, is often found to give better bounds. The following discussion is based on [19]. The field of values of a complex matrix appears in the context of bounds for functions of matrices thanks to a fundamental result by Crouzeix (see [58]):

Theorem 12. (Crouzeix) *There is a universal constant $2 \leq \mathcal{Q} \leq 11.08$ such that, given $A \in \mathbb{C}^{n \times n}$, $\mathcal{F}$ a convex compact set containing the field of values $\mathcal{W}(A)$, a function g continuous on $\mathcal{F}$ and analytic in its interior, then the following inequality holds:*

$$\|g(A)\|_2 \leq \mathcal{Q} \sup_{z \in \mathcal{F}} |g(z)|.$$

We mention that Crouzeix has conjectured that $\mathcal{Q}$ can be replaced by 2, but so far this has been proved only in some special cases. Combining Theorem 10 with Theorem 12 we obtain the following result from [19].

Theorem 13. ([19]) *Let $\{A_n\}$ be a sequence of banded $n \times n$ matrices, with bandwidths uniformly bounded by m. Let the complex function f be analytic on a neighborhood of a connected compact set $\mathcal{C} \subset \mathbb{C}$ containing $\mathcal{W}(A_n)$ for all n. Then there exist explicitly computable constants $K > 0$ and $\alpha > 0$, independent of n, such that*

$$|[f(A_n)]_{ij}| \leq K \, \mathrm{e}^{-\alpha|i-j|} \tag{40}$$

for all indices i, j, and for all $n \in \mathbb{N}$.

This result is similar to the one for the normal case, with the field of values $\mathcal{W}(A_n)$ now playing the roles of the spectra $\sigma(A_n)$. As long as the singularities of f (if any) stay bounded away from a fixed compact set $\mathcal{C}$ containing the union of all the fields of values $\mathcal{W}(A_n)$, and as long as the matrices A_n have bandwidths less than a fixed integer m, the entries of $f(A_n)$ are bounded in an exponentially decaying manner away from the main diagonal, at a rate bounded below by a fixed positive constant as $n \to \infty$. The larger the distance between the singularities of f and the compact $\mathcal{C}$, the larger this constant is (and the faster the bound decays).

As usual, the same result holds for sequences of sparse matrices A_n such that the graphs $\mathcal{G}(A_n)$ satisfy the bounded maximum degree assumption, in which case $|i - j|$ in (40) is replaced by the geodesic distance $d_n(i, j)$.

In Fig. 9 we show three plots which correspond to the first row of e^A where A is the 100×100 nonnormal tridiagonal Toeplitz matrix generated by the symbol $\phi(t) = 2t^{-1} + 1 + 3t$, see [42]. This matrix is diagonalizable with eigenvector matrix X such that $\kappa_2(X) \approx 5.26 \cdot 10^8$. The lowest curve is the actual magnitude of the entries $[\mathrm{e}^A]_{1j}$ for $j = 1, \ldots, 100$ (drawn as a continuous curve). The top curve is the bound (39), and the curve in the middle is the bound (40) obtained from Crouzeix's Theorem (with $\mathcal{C} = \mathcal{W}(A)$). Note the logarithmic scale on the vertical axis. Clearly, for this example the Crouzeix-type bound is a significant improvement over the earlier bound from [24].

A practical limitation of bound (40) is that it is in general difficult to find the constants K and α. This requires knowledge of the compact set $\mathcal{C}$ in the statement of Theorem 13. If such a set is known, a simple approach to the computation of constants K and α goes as follows [19, 149]. Suppose there is a disk of center 0 and radius $r > 0$ that contains $\mathcal{C}$, and such that f is analytic on an open neighborhood of some disk of center 0 and radius $R > r$. Define

$$E_k(f, \mathcal{C}) = \inf \max_{z \in \mathcal{C}} |f(z) - p_k(z)|,$$

where the infimum is taken over all polynomials with complex coefficients of degree $\leq k$. Then the standard theory of complex Taylor series gives the following estimate for the Taylor approximation error [78, Corollary 2.2]:
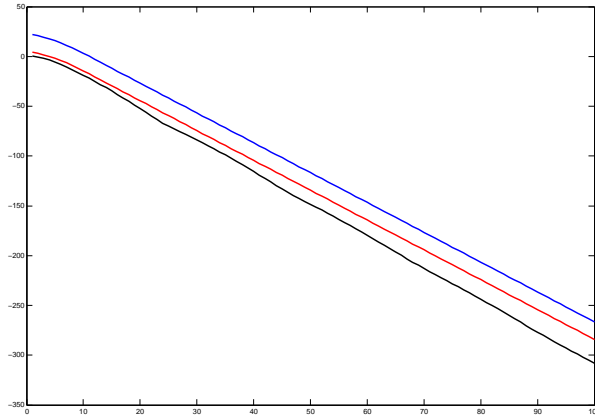
Fig. 9. Black: first row of e^A. Blue: bound (39). Red: bound (40).

$$E_k(f, \mathcal{C}) \leq \frac{M(R)}{1 - \frac{r}{R}} \left(\frac{r}{R}\right)^{k+1}, \tag{41}$$

where $M(R) = \max_{|z|=R} |f(z)|$. Therefore we can choose

$$K = \max \left\{ \|f(A)\|, \, \mathcal{Q} \, M(R) \frac{r}{R - r} \right\}, \qquad \hat{\rho} = \left(\frac{r}{R}\right)^{1/m}, \qquad \alpha = -\log(\hat{\rho}).$$

The choice of the parameter R in (41) is somewhat arbitrary: any value of R will do, as long as $r < R < \min |\zeta|$, where ζ varies over the poles of f (if f is entire, we let $\min |\zeta| = \infty$). As discussed earlier, there is a trade-off involved in the choice of R: choosing as large a value of R as possible gives a better asymptotic decay rate, but also a potentially large constant K. It is also clear that in the entire case the bound decays superexponentially.

We also mention that the choice of a disk can result in poor bounds, as it can give a crude estimate of the field of values. Better estimates can sometimes be obtained by replacing disks with rectangles: for instance, if good estimates for the smallest and largest eigenvalues of the real and imaginary parts H_1 and H_2 of A are known (see (4)), then one can approximate the rectangle of the Bendixson–Hirsch Theorem. This is a compact set containing $\mathcal{W}(A)$ and may be a much better estimate of $\mathcal{W}(A)$ than some disk containing the field of values. In [203] the authors show how these rectangles, combined with certain conformal mappings, may be useful in obtaining improved decay bounds in the case of the exponential of a matrix in upper Hessenberg form, which in turn provides accurate error estimates for Krylov subspace approximations of the action of the matrix exponential on a given vector in the nonnormal case. We shall return to this topic in section 4.1.

Further decay bounds for the entries of analytic functions of general nonnormal, nondiagonalizable band matrices based on Faber polynomials can be found in [167]; see also the comments at the end of the next section.

3.5 Bounds for matrix functions defined by integral transforms

In the Hermitian positive definite case, the available decay bounds (see (15) and Theorem 6) for the inverse and the exponential of a band matrix are generally better than the general bounds (27) based on Bernstein's Theorem. This is not surprising: the bound (15) of Demko et al. exploits the fact that the best approximation error of $f(x) = x^{-1}$ is known exactly, and similarly very good error bounds are known for the polynomial approximation of $f(x) = e^{-x}$. On the other hand, Bernstein's Theorem is much more general since it applies to any analytic function and thus the bounds on the entries of $f(A)$ obtained from it are likely to be less sharp.

This observation suggests that improved bounds should be obtainable for those matrix functions that are related in some manner to the exponential and the inverse function. As it turns out, many of the most important functions that arise in applications can be expressed as integral transforms involving either the exponential or the resolvent (and as we know, these two functions are related through the Laplace transform).

Here we focus on two (overlapping) classes of functions: the *strictly completely monotonic functions* (associated with the *Laplace–Stieltjes transform*) and the *Markov functions* (associated with the *Cauchy–Stieltjes transform*). We first review some basic properties of these functions and the relationship between the two classes, following closely the treatment in [25].

Definition 3. *Let f be defined in the interval (a, b) where $-\infty \leq a < b \leq +\infty$. Then, f is said to be* completely monotonic *in (a, b) if*

$$(-1)^k f^{(k)}(x) \geq 0 \quad \text{for all} \quad a < x < b \quad \text{and all} \quad k = 0, 1, 2, \ldots$$

Moreover, f is said to be strictly completely monotonic *in (a, b) if*

$$(-1)^k f^{(k)}(x) > 0 \quad \text{for all} \quad a < x < b \quad \text{and all} \quad k = 0, 1, 2, \ldots$$

Here $f^{(k)}$ denotes the kth derivative of f, with $f^{(0)} \equiv f$. A classical result of Bernstein (see [205]) states that a function f is completely monotonic in $(0, \infty)$ if and only if f is a Laplace–Stieltjes transform:

$$f(x) = \int_0^\infty e^{-\tau x} d\alpha(\tau), \tag{42}$$

where $\alpha(\tau)$ is nondecreasing and the integral in (42) converges for all $x > 0$. Furthermore, f is strictly completely monotonic in $(0, \infty)$ if it is completely monotonic there and the function $\alpha(\tau)$ has at least one positive point of increase, that is, there exists a $\tau_0 > 0$ such that $\alpha(\tau_0 + \delta) > \alpha(\tau_0)$ for any $\delta > 0$. Here we assume that $\alpha(\tau)$ is nonnegative and that the integral in (42) is a Riemann–Stieltjes integral.

Important examples of strictly completely monotonic functions include (see for instance [200]):

1. $f_1(x) = x^{-1} = \int_0^\infty e^{-x\tau} d\alpha(\tau)$ for $x > 0$, where $\alpha(\tau) = \tau$ for $\tau \geq 0$.
2. $f_2(x) = e^{-x} = \int_0^\infty e^{-x\tau} d\alpha(\tau)$ for $x > 0$, where $\alpha(\tau) = 0$ for $0 \leq \tau < 1$ and $\alpha(\tau) = 1$ for $\tau \geq 1$.
3. $f_3(x) = (1 - e^{-x})/x = \int_0^\infty e^{-x\tau} d\alpha(\tau)$ for $x > 0$, where $\alpha(\tau) = \tau$ for $0 \leq \tau \leq 1$, and $\alpha(\tau) = 1$ for $\tau \geq 1$.

Other examples include the functions $x^{-\sigma}$ (for any $\sigma > 0$), $\log(1 + 1/x)$ and $\exp(1/x)$, all strictly completely monotonic on $(0, \infty)$. Also, it is clear that products and positive linear combinations of strictly completely monotonic functions are strictly completely monotonic.

A closely related class of functions is given by the Cauchy–Stieltjes (or Markov-type) functions, which can be written as

$$f(z) = \int_\Gamma \frac{\mathrm{d}\gamma(\omega)}{z - \omega}, \quad z \in \mathbb{C} \setminus \Gamma, \tag{43}$$

where γ is a (complex) measure supported on a closed set $\Gamma \subset \mathbb{C}$ and the integral is absolutely convergent. In this paper we are especially interested in the special case $\Gamma = (-\infty, 0]$ so that

$$f(x) = \int_{-\infty}^0 \frac{\mathrm{d}\gamma(\omega)}{x - \omega}, \quad x \in \mathbb{C} \setminus (-\infty, 0], \tag{44}$$

where γ is now a (possibly signed) real measure. The following functions, which occur in various applications (see, e.g., [102] and references therein), fall into this class:

$$x^{-\frac{1}{2}} = \int_{-\infty}^0 \frac{1}{x - \omega} \frac{1}{\pi\sqrt{-\omega}} \mathrm{d}\omega,$$

$$\frac{\mathrm{e}^{-t\sqrt{x}} - 1}{x} = \int_{-\infty}^0 \frac{1}{x - \omega} \frac{\sin(t\sqrt{-\omega})}{-\pi\omega} \mathrm{d}\omega,$$

$$\frac{\log(1 + x)}{x} = \int_{-\infty}^{-1} \frac{1}{x - \omega} \frac{1}{(-\omega)} \mathrm{d}\omega.$$

The two classes of functions just introduced overlap. Indeed, it is easy to see (e.g., [152]) that functions of the form

$$f(x) = \int_0^\infty \frac{\mathrm{d}\mu(\omega)}{x + \omega},$$

with μ a positive measure, are strictly completely monotonic on $(0, \infty)$; but every such function can also be written in the form

$$f(x) = \int_{-\infty}^0 \frac{\mathrm{d}\gamma(\omega)}{x - \omega}, \quad \gamma(\omega) = -\mu(-\omega),$$

and therefore it is a Cauchy–Stieltjes function. We note, however, that the two classes do not coincide: e.g., $f(x) = \exp(-x)$ is strictly completely monotonic but is not a Cauchy–Stieltjes function. In the following, the term *Laplace–Stieltjes function* will be used to denote a function that is strictly completely monotonic on $(0, \infty)$.

If A is Hermitian and positive definite and f is a Laplace–Stieltjes function given by (42), we can write

$$f(A) = \int_0^\infty \mathrm{e}^{-\tau A} \mathrm{d}\alpha(\tau)$$

and therefore

$$|[f(A)]_{ij}| \leq \int_0^\infty |[\mathrm{e}^{-\tau A}]_{ij}|\,\mathrm{d}\alpha(\tau), \qquad \forall\, i,j = 1,\ldots,n.$$

We can now apply Theorem 6 on the off-diagonal decay behavior of $[\mathrm{e}^{-\tau A}]_{ij}$ to bound the entries of $f(A)$. We thus obtain the following result.

Theorem 14. ([25]) *Let* $A = A^*$ *be* m-*banded and positive definite, and let* $\widehat{A} = A - \lambda_{\min}(A)I$, *Let* $[0, 4\rho]$ *be the smallest interval containing* $\sigma(\widehat{A})$, *and assume* f *is a Laplace–Stieltjes function. Then, with the notation and assumptions of Theorem 6 and for* $\xi = \lceil |i - j|/m \rceil \geq 2$:

$$\begin{aligned}
|[f(A)]_{ij}| &\leq \int_0^\infty \exp(-\tau\lambda_{\min}(A))|[\exp(-\tau\widehat{A})]_{ij}|\mathrm{d}\alpha(\tau) \\
&\leq 10 \int_0^{\frac{\xi}{2\rho}} \exp(-\tau\lambda_{\min}(A))\frac{\exp(-\rho\tau)}{\rho\tau}\left(\frac{\mathrm{e}\rho\tau}{\xi}\right)^\xi \mathrm{d}\alpha(\tau) \qquad (45) \\
&\quad + 10 \int_{\frac{\xi}{2\rho}}^{\frac{\xi^2}{4\rho}} \exp(-\tau\lambda_{\min}(A))\exp\left(-\frac{\xi^2}{5\rho\tau}\right)\mathrm{d}\alpha(\tau) \\
&\quad + \int_{\frac{\xi^2}{4\rho}}^\infty \exp(-\tau\lambda_{\min}(A))[\exp(-\tau\widehat{A})]_{ij}\mathrm{d}\alpha(\tau) = I + II + III.
\end{aligned}$$

The nature of these bounds (45) is quite different from the ones previously seen, since they are given only implicitly by the integrals I, II and III. Note that the latter integral does not significantly contribute provided that $|i - j|$ is sufficiently large while ρ and m are not too large. In general, it will be necessary to evaluate these integrals numerically; in some special cases it may be possible to find explicit bounds for the various terms in (45). On the other hand, these bounds are much more accurate, in general, than those based on Bernstein's Theorem. We refer to [25] for numerical examples illustrating the tightness of these bounds.

Suppose now that f is a Cauchy–Stieltjes function of the form (44), then for any Hermitian positive definite matrix A we can write

$$f(A) = \int_{-\infty}^0 (A - \omega I)^{-1}\mathrm{d}\gamma(\omega),$$

Since $A - \omega I$ is Hermitian positive definite, if A is banded (or sparse) we can apply the bounds (15) of Demko et al. to the resolvent $(A - \omega I)^{-1}$ in order to derive bounds for the entries of $f(A)$.

For a given $\omega \in \Gamma = (-\infty, 0)$, let $\kappa = \kappa(\omega) = (\lambda_{\max}(A) - \omega)/(\lambda_{\min}(A) - \omega)$, $q = q(\omega) = (\sqrt{\kappa} - 1)/(\sqrt{\kappa} + 1)$, $C = C(-\omega) = \max\{1/(\lambda_{\min}(A) - \omega), C_0\}$, with $C_0 = C_0(-\omega) = (1 + \sqrt{\kappa})^2/(2(\lambda_{\max}(A) - \omega))$. We have the following result.

Theorem 15. ([25]) *Let* $A = A^*$ *be positive definite and let* f *be a Cauchy–Stieltjes function of the form (44), where* γ *is of bounded variation on* $(-\infty, 0]$. *Then for all* i *and* j *we have*

$$|[f(A)]_{ij}| \leq \int_{-\infty}^0 \left|[(A - \omega I)^{-1}]_{ij}\right||\mathrm{d}\gamma(\omega)| \leq \int_{-\infty}^0 C(\omega)q(\omega)^{\frac{|i-j|}{m}}|\mathrm{d}\gamma(\omega)|. \quad (46)$$

We remark that the first inequality in (46) is a consequence of the assumptions made on γ; see [205, Chapter I].

A natural question is the following: if f is both a Laplace–Stieltjes and a Cauchy–Stieltjes function, how do the bounds (45) and (46) compare? Is one better than the other? The answer is likely to depend on the particular function considered. However, (46) tends to lead to more explicit and more accurate bounds for some important functions, like $f(x) = x^{-1/2}$ (which is both a Laplace–Stieltjes and a Cauchy–Stieltjes function); see [25].

Remark 7. As always, all the decay bounds discussed in this section can be formulated for a sequence $\{A_n\}$ of Hermitian positive definite matrices of increasing size n as long as they are all banded (with bandwidth uniformly bounded by m) and such that the norms $\|A_n\|_2$ are uniformly bounded with respect to n. In this case, the decay bounds (45) (46) hold uniformly in n. Moreover, the bounds can be modified to accommodate more general sparsity patterns, and can be applied to sequences of sparse matrices of increasing size under the bounded maximum degree assumption (11).

Next, we briefly discuss the case of matrices that are Kronecker sums of banded or sparse matrices, again following the treatment in [25]. Recall that the Kronecker sum of two matrices $T_1 \in \mathbb{C}^{n \times n}$ and $T_2 \in T_2 \in \mathbb{C}^{m \times m}$ is defined as the $nm \times nm$ matrix

$$A = T_1 \oplus T_2 = T_1 \otimes I_m + I_n \otimes T_2.$$

A familiar example is given by the 5-point finite difference discretization of the Dirichlet Laplacian on a rectangular region. In this case T_1 and T_2 are tridiagonal, and A is block tridiagonal.

Suppose now that $f(A)$ is defined, then numerical experiments reveal that the entries of $f(A)$ decay in an oscillatory manner. In Fig. 10 we illustrate this behavior for two different matrix functions, the exponential e^{-tA} for $t = 5$ and the inverse square root $A^{-1/2}$. Here A is 100×100 and represents the 5-point finite difference discretization of the Dirichlet Laplacian on the unit square.

The question arises of whether such oscillatory behavior can be accurately captured in the form of non-monotonic decay bounds. One possibility is to treat A as a general sparse matrix, and to use the graph distance to measure decay. However, this approach does not exploit the relationship (25) between the exponential of A and the exponentials of T_1 and T_2; since we have very good bounds for the decay in the exponential of a banded matrix, it should be possible to derive similarly good bounds on the entries of the exponential of A by making use of (25). Moreover, suppose that f is a Laplace–Stieltjes functions and that $A = T_1 \oplus T_2$ (with T_1, T_2 Hermitian positive definite), such as $f(A) = A^{-1/2}$. Then we can write

$$f(A) = \int_0^\infty \exp(-\tau A) \, d\alpha(\tau) = \int_0^\infty \exp(-\tau T_1) \otimes \exp(-\tau T_2) \, d\alpha(\tau) \qquad (47)$$

for a suitable choice of $\alpha(\tau)$. Now (47) can be used to obtain bounds on the entries of $f(A)$ in terms of integrals involving the entries of $\exp(-\tau T_1)$ and $\exp(-\tau T_2)$, for which accurate bounds are available. It has been shown in [25] that these bounds are generally better than the ones obtained when A is treated as a general sparse matrix. A similar approach can also be used in the case of Cauchy–Stieltjes functions; see
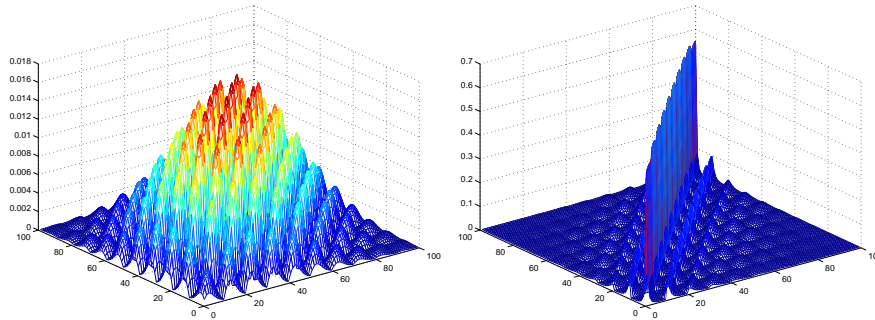
Fig. 10. Oscillatory behavior of $\exp(-5A)$ and $A^{-1/2}$ where A is a discrete Laplacian on a 10×10 grid.

[25] for details. These results can be extended to the case where A is the Kronecker sum of an arbitrary number of terms.

Finally, we mention that in the recent paper [167] new decay bounds for Cauchy–Stieltjes functions of banded non-Hermitian matrices have been obtained based on rational approximation. Specifically, the use of *Faber–Dzhrbashyan rational functions* (see [189]) allows the authors to derive exponential decay bounds which are found to be much less sensitive to the distance between the field of values of A and the singularities of f than the bounds based on polynomial approximation, in the sense that the quality of the bounds does not deteriorate nearly as much as this distance approaches zero. In [167] these bounds are also used to study the oscillatory decay behavior in functions of nonnormal matrices with Kronecker sum structure.

3.6 Functions of structured matrices

Up to now we have considered rather general classes of banded or sparse matrices, and (apart from the case where A is a Kronecker sum) we have not taken into account possible additional structure present in A. The question arises whether more can be said about the structural and decay properties of $f(A)$ when A is restricted to a specific class of structured matrices.

The simplest nontrivial example is perhaps that of circulant matrices [64]. Since any circulant $n \times n$ matrix with complex entries is of the form $A = F^* \Lambda F$, where Λ is diagonal and F is the (unitary) discrete Fourier transform matrix, we have that $f(A) = F^* f(\Lambda) F$ is also circulant. More generally, if A belongs to a subalgebra $\mathscr{A} \subseteq \mathbb{C}^{n \times n}$, then so does $f(A)$. Clearly, this poses strong constraints on the decay pattern of $f(A)$.

What if A is not circulant, but Toeplitz? Since Toeplitz matrices do not form an algebra, it is not generally true that $f(A)$ is Toeplitz if A is. Banded Toeplitz and block Toeplitz matrices arise in many important applications (see for example [41]), but relatively little has been done in the study of functions of banded Toeplitz

and block Toeplitz matrices. To our knowledge, most of the results in this area are found in [101], in which the authors study the structure of functions of banded Hermitian block Toeplitz matrices $A_n \in \mathbb{C}^{nN \times nN}$ in the limit as $n \to \infty$ (with N, the block size, being fixed). In this limit, $f(A_n)$ is asymptotically "close" to being a block Toeplitz matrix, in a precise sense. However, there is no explicit discussion of decay in [41].

Another very interesting example is that of functions of finite difference matrices (approximations of differential operators), which in [187] are shown to have a "Toeplitz-plus-Hankel" structure. Again, this fact imposes constraints on the decay behavior of the entries of $f(A)$.

Finally, it can happen that A and $f(A)$ belong to different, but related structures. For example, if A is skew-Hermitian, the exponential e^A is unitary. Hence, the exponential map takes elements of the *Lie algebra* of skew-Hermitian matrices to elements of the corresponding *Lie group*, the unitary matrices. Many more such examples are given in [109, Section 14.1.1]. Exploiting these structural properties may lead to improved bounds for the entries of $f(A)$; to our knowledge, however, this possibility has not been explored so far.

3.7 Some generalizations

So far we have only considered matrices over $\mathbb{R}$ or $\mathbb{C}$. In applications (especially in physics) it is sometimes necessary to consider functions of matrices over more general algebraic and topological structures. Depending on the problem, these could be non-commutative division algebras, algebras of operators on a Hilbert space (finite or infinite dimensional), algebras of continuous functions, etc.

The question arises then whether decay bounds such as those discussed up to now can be extended to these more general situations. The answer, as shown in [19], is largely affirmative. The most natural tool for carrying out the desired extension is the general theory of *complex C^*-algebras*. In particular, the *holomorphic functional calculus* allows one to define the notion of analytic function on such algebras, and to develop a theory of functions of matrices over such algebras. In this setting, almost all[8] the decay bounds described so far can be extended *verbatim*, the only difference being that the absolute value of $[f(A)]_{ij}$ must now be replaced by $\|[f(A)]_{ij}\|$, where $\|\cdot\|$ is the norm in the underlying C^*-algebra.

We proceed now to sketch the generalization of some of the decay bounds. The following discussion is based on [19]; for an excellent introduction to the basic theory of C^*-algebras, see [122].

Recall that a *Banach algebra* is a complex algebra $\mathbb{A}$ with a norm making $\mathbb{A}$ into a Banach space and satisfying

$$\|ab\| \leq \|a\|\|b\|$$

for all $a, b \in \mathbb{A}$. In this paper we consider only *unital* Banach algebras, i.e., algebras with a multiplicative unit I with $\|I\| = 1$.

An *involution* on a Banach algebra $\mathbb{A}$ is a map $a \mapsto a^*$ of $\mathbb{A}$ into itself satisfying

(i) $(a^*)^* = a$

---

[8] The exception is given by the bounds involving the condition number of the eigenvector matrix, see Theorem 11.

(ii) $(ab)^* = b^* a^*$
(iii) $(\lambda a + b)^* = \bar{\lambda} a^* + b^*$

for all $a, b \in \mathbb{A}$ and $\lambda \in \mathbb{C}$. A C^*-*algebra* is a Banach algebra with an involution such that the C^*-*identity*

$$\|a^* a\| = \|a\|^2$$

holds for all $a \in \mathbb{A}$. Note that we do not make any assumption on whether $\mathbb{A}$ is commutative or not. Basic examples of C^*-algebras are:

1. The commutative algebra $\mathcal{C}(\mathcal{X})$ of all continuous complex-valued functions on a compact Hausdorff space $\mathcal{X}$. Here the addition and multiplication operations are defined pointwise, and the norm is given by $\|f\|_\infty = \max_{t \in \mathcal{X}} |f(t)|$. The involution on $\mathcal{C}(\mathcal{X})$ maps each function f to its complex conjugate f^*, defined by $f^*(t) = \overline{f(t)}$ for all $t \in \mathcal{X}$.
2. The algebra $\mathcal{B}(\mathcal{H})$ of all bounded linear operators on a complex Hilbert space $\mathcal{H}$, with the operator norm $\|T\| = \sup \|T\mathbf{x}\|_{\mathcal{H}}/\|\mathbf{x}\|_{\mathcal{H}}$, where the supremum is taken over all nonzero $\mathbf{x} \in \mathcal{H}$. The involution on $\mathcal{B}(\mathcal{H})$ maps each bounded linear operator T on $\mathcal{H}$ to its adjoint, T^*.

Note that the second example contains as a special case the algebra $\mathcal{M}_n(\mathbb{C})$ $(= \mathbb{C}^{k \times k})$ of all $k \times k$ matrices with complex entries, with the norm being the usual spectral norm and the involution mapping each matrix $A = [a_{ij}]$ to its Hermitian conjugate $A^* = [\overline{a_{ji}}]$. This algebra is noncommutative for $k \geq 2$.

Examples 1 and 2 above provide, in a precise sense, the "only" examples of C^*-algebras. Indeed, every (unital) commutative C^*-algebra admits a faithful representation onto an algebra of the form $C(\mathcal{X})$ for a suitable (and essentially unique) compact Hausdorff space $\mathcal{X}$; and, similarly, every unital (possibly noncommutative) C^*-algebra can be faithfully represented as a norm-closed subalgebra of $\mathcal{B}(\mathcal{H})$ for a suitable complex Hilbert space $\mathcal{H}$.

More precisely, a map ϕ between two C^*-algebras is a $*$-*homomorphism* if ϕ is linear, multiplicative, and such that $\phi(a^*) = \phi(a)^*$; a $*$-*isomorphism* is a bijective $*$-homomorphism. Two C^*-algebras are said to be *isometrically* $*$-*isomorphic* if there is a norm-preserving $*$-isomorphism between them, in which case they are indistinguishable as C^*-algebras. A $*$-*subalgebra* $\mathbb{B}$ of a C^*-algebra $\mathbb{A}$ is a subalgebra that is $*$-closed, i.e., $a \in \mathbb{B}$ implies $a^* \in \mathbb{B}$. Finally, a C^*-subalgebra is a norm-closed $*$-subalgebra of a C^*-algebra. The following two results are classical [86, 87].

Theorem 16. (Gelfand) *Let $\mathbb{A}$ be a commutative C^*-algebra. Then there is a compact Hausdorff space $\mathcal{X}$ such that $\mathbb{A}$ is isometrically $*$-isomorphic to $\mathcal{C}(\mathcal{X})$. If $\mathcal{Y}$ is another compact Hausdorff space such that $\mathbb{A}$ is isometrically $*$-isomorphic to $\mathcal{C}(\mathcal{Y})$, then $\mathcal{X}$ and $\mathcal{Y}$ are necessarily homeomorphic.*

Theorem 17. (Gelfand–Naimark) *Let $\mathbb{A}$ be a C^*-algebra. Then there is a complex Hilbert space $\mathcal{H}$ such that $\mathbb{A}$ is isometrically $*$-isomorphic to a C^*-subalgebra of $\mathcal{B}(\mathcal{H})$.*

We will also need the following definitions and facts. An element a of a C^*-algebra is *unitary* if $aa^* = a^*a = I$, *Hermitian* (or *self-adjoint*) if $a^* = a$, *skew-Hermitian* if $a^* = -a$, *normal* if $aa^* = a^*a$. Clearly, unitary, Hermitian and skew-Hermitian elements are all normal. Any element a in a C^*-algebra can be written uniquely as $a = h_1 + i\,h_2$ with h_1, h_2 Hermitian and $i = \sqrt{-1}$.

For any (complex) Banach algebra $\mathbb{A}$, the *spectrum* of an element $a \in \mathbb{A}$ is the set of all $\lambda \in \mathbb{C}$ such that $\lambda I - a$ is not invertible in $\mathbb{A}$. We denote the spectrum of a by $\sigma(a)$. For any $a \in \mathbb{A}$, the spectrum $\sigma(a)$ is a non-empty compact subset of $\mathbb{C}$ contained in the closed disk of radius $r = \|a\|$ centered at 0. The *spectral radius* of a is defined as $\varrho(a) = \max\{|\lambda| : \lambda \in \sigma(A)\}$. *Gelfand's formula* for the spectral radius [86] states that

$$\varrho(a) = \lim_{m \to \infty} \|a^m\|^{\frac{1}{m}} . \tag{48}$$

Note that this identity contains the statement that the above limit exists.

If $a \in \mathbb{A}$ (a C*-algebra) is Hermitian, $\sigma(a)$ is a subset of $\mathbb{R}$. If $a \in \mathbb{A}$ is normal (in particular, Hermitian), then $\varrho(a) = \|a\|$. This implies that if a is Hermitian, then either $-\|a\| \in \sigma(a)$ or $\|a\| \in \sigma(a)$. The spectrum of a skew-Hermitian element is purely imaginary, and the spectrum of a unitary element is contained in the unit circle $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.

An element $a \in \mathbb{A}$ is *nonnegative* if $a = a^*$ and the spectrum of a is contained in $\mathbb{R}_+$, the nonnegative real axis; a is *positive* if $\sigma(a) \subset (0, \infty)$. Any linear combination with real nonnegative coefficients of nonnegative elements of a C*-algebra is nonnegative; in other words, the set of all nonnegative elements in a C*-algebra $\mathbb{A}$ form a (nonnegative) *cone* in $\mathbb{A}$. For any $a \in \mathbb{A}$, aa^* is nonnegative, and $I + aa^*$ is invertible in $\mathbb{A}$. Furthermore, $\|a\| = \sqrt{\varrho(a^*a)} = \sqrt{\varrho(aa^*)}$, for any $a \in \mathbb{A}$.

Note that if $\|\cdot\|_*$ and $\|\cdot\|_{**}$ are two norms with respect to which $\mathbb{A}$ is a C*-algebra, then $\|\cdot\|_* = \|\cdot\|_{**}$.

Let $\mathbb{A}$ be a C*-algebra. Given a positive integer n, let $\mathbb{A}^{n \times n} = \mathcal{M}_n(\mathbb{A})$ be the set of $n \times n$ matrices with entries in $\mathbb{A}$. Observe that $\mathbb{A}^{n \times n}$ has a natural C*-algebra structure, with matrix addition and multiplication defined in the usual way (in terms, of course, of the corresponding operations on $\mathbb{A}$). The involution is naturally defined as follows: given a matrix $A = [a_{ij}] \in \mathbb{A}^{n \times n}$, the adjoint of A is given by $A^* = [a_{ji}^*]$. The algebra $\mathbb{A}^{n \times n}$ is obviously unital, with unit

$$I_n = \begin{bmatrix} I & 0 & \ldots \ldots & 0 \\ 0 & I & \ldots \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots \ldots & I & 0 \\ 0 & \ldots \ldots & 0 & I \end{bmatrix}$$

where I is the unit of $\mathbb{A}$. The definition of unitary, Hermitian, skew-Hermitian and normal matrix are the obvious ones.

It follows from the Gelfand–Naimark Representation Theorem (Theorem 17 above) that each $A \in \mathbb{A}^{n \times n}$ can be represented as a matrix T_A of bounded linear operators, where T_A acts on the direct sum $\mathscr{H} = \mathcal{H} \oplus \cdots \oplus \mathcal{H}$ of n copies of a suitable complex Hilbert space $\mathcal{H}$. This fact allows us to introduce an operator norm on $\mathbb{A}^{n \times n}$, defined as follows:

$$\|A\| := \sup_{\|\mathbf{x}\|_{\mathscr{H}} = 1} \|T_A \mathbf{x}\|_{\mathscr{H}} , \tag{49}$$

where

$$\|\mathbf{x}\|_{\mathscr{H}} := \sqrt{\|\mathbf{x}_1\|_{\mathcal{H}}^2 + \cdots + \|\mathbf{x}_n\|_{\mathcal{H}}^2}$$

is the norm of an element $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathscr{H}$. Relative to this norm, $\mathbb{A}^{n \times n}$ is a C*-algebra. Note that $\mathbb{A}^{n \times n}$ can also be identified with the tensor product of C*-algebras $\mathbb{A} \otimes \mathcal{M}_n(\mathbb{C})$.

Similarly, Gelfand's Theorem (Theorem 16 above) implies that if $\mathbb{A}$ is commutative, there is a compact Hausdorff space $\mathcal{X}$ such that any $A \in \mathbb{A}^{n \times n}$ can be identified with a continuous matrix-valued function

$$A : \mathcal{X} \longrightarrow \mathcal{M}_n(\mathbb{C}).$$

In other words, A can be represented as an $n \times n$ matrix of continuous, complex-valued functions: $A = [a_{ij}(t)]$, with domain $\mathcal{X}$. The natural C*-algebra norm on $\mathbb{A}^{n \times n}$, which can be identified with $\mathcal{C}(\mathcal{X}) \otimes \mathcal{M}_n(\mathbb{C})$, is now the operator norm

$$\|A\| := \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|, \tag{50}$$

where $\mathbf{x} = (x_1, \ldots, x_n) \in [\mathcal{C}(\mathcal{X})]^n$ has norm $\|\mathbf{x}\| = \sqrt{\|x_1\|_\infty^2 + \cdots + \|x_n\|_\infty^2}$ with $\|x_i\|_\infty = \max_{t \in \mathcal{X}} |x_i(t)|$, for $1 \le i \le n$.

Since $\mathbb{A}^{n \times n}$ is a C*-algebra, all the definitions and basic facts about the spectrum remain valid for any matrix A with entries in $\mathbb{A}$. Thus, the spectrum $\sigma(A)$ of $A \in \mathbb{A}^{n \times n}$ is the set of all $\lambda \in \mathbb{C}$ such that $\lambda I_n - A$ is not invertible in $\mathbb{A}^{n \times n}$. The set $\sigma(A)$ is a nonempty compact subset of $\mathbb{C}$ completely contained in the disk of radius $\|A\|$ centered at 0. The definition of spectral radius and Gelfand's formula (48) remain valid. Hermitian matrices have real spectra, skew-Hermitian matrices have purely imaginary spectra, unitary matrices have spectra contained in $\mathbb{T}$, and so forth. Note, however, that it is not true in general that a normal matrix A over a C*-algebra can be unitarily diagonalized [121].

The standard way to define the notion of an analytic function $f(a)$ of an element a of a C*-algebra $\mathbb{A}$ is via contour integration. In particular, we can use this approach to define functions of a matrix A with elements in $\mathbb{A}$.

Let $f(z)$ be a complex function which is analytic in an open neighborhood U of $\sigma(a)$. Since $\sigma(a)$ is compact, we can always find a finite collection $\Gamma = \cup_{j=1}^{\ell} \gamma_j$ of smooth simple closed curves whose interior parts contain $\sigma(a)$ and entirely contained in U. The curves γ_j are assumed to be oriented counterclockwise.

Then $f(a) \in \mathbb{A}$ can be defined as

$$f(a) = \frac{1}{2\pi i} \int_\Gamma f(z)(zI - a)^{-1} \mathrm{d}z, \tag{51}$$

where the line integral of a Banach-space-valued function g defined on a smooth curve $\gamma : t \mapsto z(t)$ for $t \in [0, 1]$ is given by the norm limit of Riemann sums of the form

$$\sum_{j=1}^{\nu} g(z(\theta_j))[z(t_j) - z(t_{j-1})], \qquad t_{j-1} \le \theta_j \le t_j,$$

where $0 = t_0 < t_1 < \ldots < t_{\nu-1} < t_\nu = 1$.

Denote by $\mathcal{H}(a)$ the algebra of analytic functions whose domain contains an open neighborhood of $\sigma(a)$. The following well known result is the basis for the *holomorphic functional calculus*; see, e.g., [122, page 206].

Theorem 18. *The mapping $\mathcal{H}(a) \longrightarrow \mathbb{A}$ defined by $f \mapsto f(a)$ is an algebra homomorphism, which maps the constant function 1 to $I \in \mathbb{A}$ and maps the identity*

function to a. If $f(z) = \sum_{j=0}^{\infty} c_j z^j$ is the power series representation of $f \in \mathcal{H}(a)$ over an open neighborhood of $\sigma(a)$, then we have

$$f(a) = \sum_{j=0}^{\infty} c_j a^j.$$

Moreover, the following version of the *spectral theorem* holds:

$$\sigma(f(a)) = f(\sigma(a)). \tag{52}$$

If a is normal, the following properties also hold:

1. $\|f(a)\| = \|f\|_{\infty,\sigma(a)} := \max_{\lambda \in \sigma(a)} |f(\lambda)|$;
2. $\overline{f(a)} = [f(a)]^*$; in particular, if a is Hermitian then $f(a)$ is also Hermitian if and only if $f(\sigma(a)) \subset \mathbb{R}$;
3. $f(a)$ is normal;
4. $f(a)b = bf(a)$ whenever $b \in \mathbb{A}$ and $ab = ba$.

Obviously, these definitions and results apply in the case where a is a matrix A with entries in a C*-algebra $\mathbb{A}$. In particular, if $f(z)$ is analytic on a neighborhood of $\sigma(A)$, we define $f(A)$ via

$$f(A) = \frac{1}{2\pi i} \int_\Gamma f(z)(zI_n - A)^{-1} \mathrm{d}z, \tag{53}$$

with the obvious meaning of Γ.

The holomorphic functional calculus allows us to generalize most of the decay results that hold for analytic functions of matrices with entries in $\mathbb{C}$ to functions of matrices with entries in an arbitrary C*-algebra $\mathbb{A}$ almost without changes. The fact that finite matrices over $\mathbb{C}$ have finite spectra whereas matrices over a general C*-algebra $\mathbb{A}$ have in general continuous spectra makes no difference whatsoever; note that we have already encountered this situation when we discussed the case of functions of bounded infinite matrices. To see that the same arguments used for matrices over $\mathbb{C}$ carry over almost *verbatim* to this more general setting, consider for example an m-banded Hermitian matrix $A \in \mathbb{A}^{n \times n}$. Then $\sigma(A) \subseteq [a, b] \subset \mathbb{R}$ for some a, b with $-\infty < a < b < \infty$. Up to scaling and shift, we can assume that $[a, b] = \mathcal{I} = [-1, 1]$. Let f be analytic on a region $\Omega \subseteq \mathbb{C}$ containing $\mathcal{I}$ and let $\mathbb{P}_k$ denote the set of all complex polynomials of degree at most k on $\mathcal{I}$. Given $p \in \mathbb{P}_k$, the matrix $p(A) \in \mathbb{A}^{n \times n}$ is well defined and it is banded with bandwidth at most km. So for any polynomial $p \in \mathbb{P}_k$ and any pair of indices i, j such that $|i - j| > km$ we have

$$\|[f(A)]_{ij}\| = \|[f(A) - p(A)]_{ij}\| \tag{54}$$
$$\leq \|f(A) - p(A)\| \tag{55}$$
$$= \varrho(f(A) - p(A)) \tag{56}$$
$$= \max(\sigma(f(A) - p(A))) = \max(\sigma((f - p)(A))) \tag{57}$$
$$= \max((f - p)(\sigma(A))) \leq E_k(f, \mathcal{I}), \tag{58}$$

where $E_k(f, \mathcal{I})$ is the best uniform approximation error for the function f on the interval $\mathcal{I}$ using polynomials of degree at most k:

$$E_k(f, \mathcal{I}) := \min_{p \in \mathbb{P}_k} \max_{x \in \mathcal{I}} |f(x) - p(x)|.$$

In the above derivation we made use of the definition (49), of the fact that $A = A^*$, and of the spectral theorem (52), valid for normal elements of any C*-algebra. We can now apply Bernstein's Theorem to bound $E_k(f, \mathcal{I})$ in terms of ellipses contained in Ω having foci at -1, 1. From this we can deduce exponentially decaying bounds for $\|[f(A)]_{ij}\|$ with respect to $|i - j|$ in the usual manner:

$$\|[f(A)]_{ij}\| \le K \chi^{-\frac{|i-j|}{m}} = K \rho^{|i-j|}, \quad \rho = \chi^{-\frac{1}{m}}, \tag{59}$$

where $K = 2M(\chi)/(\chi - 1)$, $M(\chi) = \max_{z \in \mathcal{E}_\chi} |f(z)|$.

Analogous decay bounds can be derived in the normal case, without any changes to the proofs. The same is true for the general, nonnormal case, with one caveat: the notion of field of values is not well-defined, in general, for an element of a C*-algebra. One can attempt to define the field of values of an element a of a C*-algebra $\mathbb{A}$ by making use of the Gelfand–Naimark Representation Theorem (Theorem 17): since there is an isometric $*$-isomorphism ϕ from $\mathbb{A}$ into the algebra $\mathcal{B}(\mathcal{H})$ for some complex Hilbert space $\mathcal{H}$, we could define the field of values of a as the field of values of the bounded linear operator $T_a = \phi(a)$, i.e.,

$$\mathcal{W}(a) = \mathcal{W}(T_a) = \{\langle T_a \mathbf{x}, \mathbf{x} \rangle \mid \mathbf{x} \in \mathcal{H}, \|\mathbf{x}\| = 1\}.$$

Unfortunately, ϕ is not unique and it turns out that different choices of ϕ may give rise to different fields of values. Fortunately, however, the *closure* of the field of values is independent of the choice of representation [28]. Hence, if we replace the field of values $\mathcal{W}(T_a)$ with the closure $\overline{\mathcal{W}(T_a)}$, everything works as in the "classical" case.[9]

In order to achieve the desired generalization, we make use of the following theorem of Crouzeix, which is an extension of Theorem 12. Given a set $E \subset \mathbb{C}$, denote by $\mathcal{H}_b(E)$ the algebra of continuous and bounded functions in $\overline{E}$ which are analytic in the interior of E. Furthermore, for $T \in \mathcal{B}(\mathcal{H})$ let $\|p\|_{\infty, T} := \max_{z \in \overline{\mathcal{W}(T)}} |p(z)|$. Then we have ([58], Theorem 2):

Theorem 19. *For any bounded linear operator $T \in \mathcal{B}(\mathcal{H})$ the homomorphism $p \mapsto p(T)$ from the algebra $\mathbb{C}[z]$, with norm $\|\cdot\|_{\infty, T}$, into the algebra $\mathcal{B}(\mathcal{H})$, is bounded with constant $\mathcal{Q}$. It admits a unique bounded extension from $\mathcal{H}_b(\mathcal{W}(T))$ into $\mathcal{B}(\mathcal{H})$. This extension is also bounded with constant $\mathcal{Q}$.*

Making use of the notion of field of values for elements of a C*-algebra, we obtain the following corollary.

Corollary 1. *Given $A \in \mathbb{A}^{n \times n}$, the following bound holds for any complex function g analytic on a neighborhood of $\overline{\mathcal{W}(A)}$:*

$$\|g(A)\| \le \mathcal{Q}\|g\|_{\infty, A} = \mathcal{Q} \max_{z \in \overline{\mathcal{W}(A)}} |g(z)|.$$

---

[9] Recall that for $A \in \mathbb{C}^{n \times n}$ the field of values $\mathcal{W}(A)$ is compact and therefore always closed.

In order to obtain bounds on $\|[f(A)]_{ij}\|$, where the function $f(z)$ can be assumed to be analytic on an open set $S \supset \mathcal{W}(A)$, we can choose $g(z)$ in Corollary 1 as $f(z) - p_k(z)$, where $p_k(z)$ is any complex polynomial of degree bounded by k. The argument in (54)–(58) can then be adapted as follows:

$$\|[f(A)]_{ij}\| = \|[f(A) - p_k(A)]_{ij}\| \tag{60}$$

$$\leq \|f(A) - p_k(A)\| \tag{61}$$

$$\leq \mathcal{Q} \|f - p_k\|_{\infty, A} \tag{62}$$

$$= \mathcal{Q} \max_{z \in \overline{\mathcal{W}(A)}} |f(z) - p_k(z)| \tag{63}$$

$$\leq \mathcal{Q} \, E_k(f, \overline{\mathcal{W}(A)}), \tag{64}$$

where $E_k(f, \overline{\mathcal{W}(A)})$ is the degree k best approximation error for f on the compact set $\overline{\mathcal{W}(A)}$. In order to make explicit computations easier, we may of course replace $\overline{\mathcal{W}(A)}$ with a larger but more manageable set in the above argument.

Putting everything together, we obtain the following generalization of Theorem 13:

Theorem 20. ([19]) *Let $\{A_n\}$ be a sequence of matrices of increasing size over a complex C^*-algebra $\mathbb{A}$ with bandwidths uniformly bounded by m. Let the complex function f be analytic on a neighborhood of a connected compact set $\mathcal{C} \subset \mathbb{C}$ containing $\overline{\mathcal{W}(A_n)}$ for all n. Then there exist explicitly computable constants $K > 0$ and $\alpha > 0$, independent of n, such that*

$$\|[f(A_n)]_{ij}\| \leq K \, e^{-\alpha |i-j|}$$

for all indices i, j and for all $n \in \mathbb{N}$.

As always, analogous results hold for more general sparse matrices, with the geodesic distance on the matrix graphs $\mathcal{G}(A_n)$ replacing the distance from the main diagonal, as long as the bounded maximum degree condition (11) holds. Also, if f is entire, then the entries of $f(A_n)$ are bounded in a superexponentially decaying manner.

As a consequence of this extension, the decay bounds apply directly to functions of block-banded matrices (with blocks all of the same size), to functions of banded or sparse matrices of operators on a complex Hilbert space, and to functions of matrices the entries of which are complex-valued continuous functions. In [19] one can also find the results of numerical and symbolic computations illustrating the decay in $f(A)$ where A is a banded (in particular, tridiagonal) function over the function algebra $\mathcal{C}[0, 1]$ endowed with the infinity norm, showing the rapid decay of $\|[f(A)]_{ij}\|_\infty$ for increasing $|i - j|$.

In [19] it is further shown that the theory can be extended to cover analytic functions of matrices with entries in the *real* C^*-algebra $\mathbb{H}$ of quaternions, as long as these functions have power series expansions with real coefficients.

The time-ordered exponential. In mathematical physics, the *time-ordered exponential $OE[A]$* associated with a given time-dependent matrix $A = A(t) = [A(t)]_{ij}$, $t \in [a, b] \subset \mathbb{R}$, is defined as the unique solution to the system of ordinary differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t'}OE[A](t',t) = A(t')OE[A](t',t)$$

such that $OE[A](t,t) = I$ for all $t \in [a,b]$. Hence, the time-ordered exponential, also denoted by

$$OE[A](t',t) = \mathcal{T}\exp\left(\int_t^{t'} A(\tau)\mathrm{d}\tau\right),$$

with $\mathcal{T}$ the *time-ordering operator* (see, e.g., [133]), provides a way to express the solution of a linear first-order system of ordinary differential equations with variable coefficients. In the case of a constant A, the time-ordered exponential reduces to the usual matrix exponential: $OE[A](t',t) = \mathrm{e}^{(t'-t)A}$, where we assume $t' > t$.

When $A(t) \neq A(t')$ for $t \neq t'$, no simple, explicit expression is available for the time-ordered exponential. Techniques for evaluating $OE[A](t',t)$ (or its action) have been studied in, e.g., [89]. In that paper the authors also study the decay properties of the time-ordered exponential for the case of a sparse $A(t)$. Note that $OE[A](t',t)$ cannot be expressed in terms of countour integration, power series expansion, or other such device, therefore techniques different from those employed so far must be employed. The authors of [89] assume that $A = A(t)$ is a possibly infinite sparse matrix satisfying

$$M := \sup_{t\in[a,b]} \max_{i,j} |[A(t)]_{ij}| < \infty,$$

and that the nonzero pattern of $A(t)$ does not depend on t. The following result holds.

Theorem 21. ([89]) *If $d = d(i,j)$ is the geodesic distance between nodes i and j in the graph $\mathcal{G}(A)$, then the following bound holds for all i and j and for $t' > t$:*

$$\left|\left[OE[A](t',t)\right]_{ij}\right| \leq \sum_{k=d}^{\infty} \frac{M^k(t'-t)^k}{k!} W_{i,j;k}, \tag{65}$$

where $W_{i,j;k}$ is the number of walks of length k in $\mathcal{G}(A)$ between node i and node j. If Δ, the maximum degree of any node in $\mathcal{G}(A)$, is finite, then we also have the weaker bound

$$\left|\left[OE[A](t',t)\right]_{ij}\right| \leq \mathrm{e}^{\Delta M(t'-t)} \frac{(\Delta M(t'-t))^d}{d!}. \tag{66}$$

The bound (66) decays superexponentially with the distance $d(i,j)$. The same result can be restated for a sequence of sparse time-dependent matrices $\{A_n(t)\}$ of increasing size such that $\sup_n \sup_{t\in[a,b]} \max_{i,j} |[A_n(t)]_{ij}| < \infty$, as long as the corresponding graphs $\mathcal{G}(A_n)$ satisfy the bounded maximum degree assumption. In this case a bound of the type (66) holds uniformly in n. In [89] it is also shown by example that the superexponential decal fails, in general, if $\Delta = \infty$.

3.8 Decay algebras

Although so far our main emphasis has been on exponential decay, other types of decay occur frequently in applications. These different decay rates lead to the definition of various *decay algebras*, which are Banach algebras of infinite matrices, the entries of which satisfy different decay conditions.

Seminal works on decay algebras are the already cited paper by Jaffard [119] and papers by Baskakov [12, 13]. A key question addressed in these papers is that of inverse-closedness (see footnote 4). This question is highly relevant for us in view of the fact that the techniques covered so far all assume bandedness or sparsity of A in order to make statements about the decay behavior in A^{-1} or in more general matrix functions $f(A)$, but they are not applicable if A is a full matrix satisfying a decay condition. As noted in [98], if $\mathscr{A}$ and $\mathscr{B}$ are Banach algebras with $\mathscr{A} \subseteq \mathscr{B}$ and $\mathscr{A}$ is inverse-closed in $\mathscr{B}$, then using the contour integral definition of a matrix function

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} \mathrm{d}z \qquad (67)$$

(where the integral of a Banach-space-valued function has been defined in the previous section) we immediately obtain that $f(A) \in \mathscr{A}$ if $A \in \mathscr{A}$. Therefore, the entries of $f(A)$ must satisfy the same decay bound as those of A itself. Hence, inverse-closedness provides a powerful tool to establish the decay properties in the entries of $f(A)$ when A is not just a sparse or banded matrix, but more generally a matrix with certain types of decay. We emphasize that this approach is completely different from the techniques reviewed earlier, which are largely based on classical results on the approximation of analytic functions with polynomials.

Results on inverse-closedness of decay algebras can be regarded as noncommutative variants of *Wiener's Lemma*: if a periodic function f has an absolutely convergent Fourier series and is never zero, then $1/f$ has an absolutely convergent Fourier series.[10] There is a strong analogy between Wiener's Lemma and the inverse-closedness of matrix algebras. Just as a function with rapidly decaying Fourier coefficients can be well approximated by trigonometric polynomials, so a matrix with rapidly decaying off-diagonal entries can be well approximated by banded matrices. We refer the reader to [97] for details.

In this section we limit ourselves to a brief description of some of the most important decay algebras, and we refer to the original papers for details and applications. For simplicity we focus on matrices (bounded linear operators) of the form $A = [A_{ij}]_{i,j\in S}$ with $S = \mathbb{Z}$ or $S = \mathbb{N}$ and on off-diagonal decay measured in terms of the distance $d(i,j) = |i-j|$, although the same results hold more generally for matrices indexed by a set $T \times T$ where (T,d) is a metric space such that the distance function d on T satisfies condition (10).

The first two examples are due to Jaffard [119]:

Definition 4. *Let $\gamma > 0$. A matrix A belongs to the class $\mathscr{E}_\gamma$ if for all i,j:*

$$\forall \gamma' < \gamma, \quad |[A]_{ij}| \leq K(\gamma') \exp(-\gamma'|i - j|) \qquad (68)$$

for some constant $K = K(\gamma') > 0$.

Next, suppose that

---

[10] As is well known, Gelfand was able to give a short proof of Wiener's Lemma using his general theory of *commutative* Banach algebras; see, e.g., [88, page 33]. Wiener's Lemma is simply the statement that the *Wiener algebra* $\mathscr{A}(\mathbb{T})$ of all functions on the unit circle having an absolutely convergent Fourier expansion is inverse-closed in the algebra $\mathcal{C}(\mathbb{T})$ of continuous functions on $\mathbb{T}$.

$$\sup_{i \in S} \sum_{j \in S} (1 + |i - j|)^{-p} < \infty;$$

for example, $p > 1$ (for $S = \mathbb{N}$ or $\mathbb{Z}$). For such p we have the following definition.

Definition 5. *Let $\alpha > p$. A matrix A belongs to the class Q_α if for all i, j:*

$$|[A]_{ij}| \leq K(1 + |i - j|)^{-\alpha}, \tag{69}$$

for some constant $K > 0$.

Any matrix A in $\mathscr{E}_\gamma$ or in Q_α is a bounded linear operator on $\ell^2(S)$ (this is a consequence of *Schur's Lemma*, see [119]). Moreover, in [119] it is also shown that both $\mathscr{E}_\gamma$ and Q_α are algebras; Q_α is called the *Jaffard algebra*.

In [119], Jaffard proved the following theorems (the first straightforward, the second not).

Theorem 22. ([119]) *Let $A \in \mathscr{E}_\gamma$ and assume that A is invertible as an operator on $\ell^2(S)$. Then $A^{-1} \in \mathscr{E}_{\gamma'}$ for some $0 < \gamma' < \gamma$.*

Hence, if A has an exponential off-diagonal decay property and A is invertible in $\mathcal{B}(\ell^2)$, the entries of A^{-1} are also bounded in an exponentially decaying manner away from the main diagonal but with a different decay rate, the decay being generally slower. Note that this result generalizes many of the results known in the literature about the exponential decay in the inverses of band matrices.

A deeper, and *a priori* unexpected, result is the following.

Theorem 23. ([119]) *Let $A \in Q_\alpha$ and assumat that A is invertible as an operator on $\ell^2(S)$. Then $A^{-1} \in Q_\alpha$*

Thus, *the Jaffard algebra Q_α is inverse-closed in $\mathcal{B}(\ell^2)$*: if A satisfies the off-diagonal algebraic decay property (69) and is invertible in $\mathcal{B}(\ell^2)$, the inverse A^{-1} satisfies exactly the same decay property. Similar results were obtained by Baskakov in [12, 13].

Jaffard's and Baskakov's results have attracted considerable interest and have been generalized in various directions. Extensions to different types of decay can be found in [98] and [190]; the former paper, in particular, makes use of Banach algebra techniques (not mentioned in Jaffard's original paper) and points out the implications for the functional calculus.

Although concerned with infinite matrices, there is no lack of applications of the theory to concrete, finite-dimensional problems from numerical analysis. A connection is provided by the finite section method for the solution of operator equations of the form $A\mathbf{x} = \mathbf{b}$, where A is assumed to be boundedly invertible and $\mathbf{b} \in \ell^2$. In a nutshell, this method consists in considering the n-dimensional sections $A_n = P_n A P_n$ (where P_n is the orthogonal projector onto the subspace spanned by $\mathbf{e}_1, \ldots, \mathbf{e}_n$) of the infinite matrix A and the truncated vectors $\mathbf{b}_n = P_n \mathbf{b}$, solving the finite-dimensional problems $A_n \mathbf{x}_n = \mathbf{b}_n$, and letting $n \to \infty$. The component-wise convergence of the approximate solutions $\mathbf{x}_n$ to the solution $\mathbf{x} = A^{-1}\mathbf{b}$ of the original, infinite-dimensional problem requires that the sequence $\{A_n\}$ be *stable*, i.e., the inverses A_n^{-1} exist and have uniformly bounded norm with respect to n. These conditions are essentially those that guarantee off-diagonal decay in A^{-1}; hence,

decay algebras play a key role in the analysis, see for example [99], or [141] for a systematic treatment. Another approach, based on the notion of *nearest neighbor approximation*, is described in [61]; here again decay algebras play the main role. In the opposite direction, the authors of [174] develop an algorithm for solving large $n \times n$ Toeplitz systems by embedding the coefficient matrix A_n into a semi-infinite Toeplitz matrix A and making use of the (canonical) Wiener–Hopf factorization of the inverse of the symbol of A to obtain the solution, which is then truncated and corrected (via the solution of a much smaller Toeplitz system by conventional techniques) to yield the solution of the original problem. Ultimately, this approach works because of the exponential decay of the entries of the inverse of the infinite matrix A.

The finite section method, when applicable, can also be used to establish decay properties for functions of $n \times n$ matrices A_n with off-diagonal decay (with $n \to \infty$) by thinking of the A_n's as the finite sections of an infinite matrix $A \in \mathscr{A}$ for a suitable decay algebra $\mathscr{A}$, assumed to be inverse-closed in $\mathcal{B}(\ell^2)$. Suppose that the spectra of all the A_n are contained in a compact subset $\mathcal{C}$ of $\mathbb{C}$ and that the contour Γ in (67) surrounds $\mathcal{C}$. If the norms of the resolvents $(zI_n - A_n)^{-1}$ are bounded uniformly in n and in $z \in \Gamma$, then the entries of $(zI_n - A_n)^{-1}$ converge to those of $(zI - A)^{-1}$ as $n \to \infty$, and therefore the entries of $f(A_n)$ must converge to those of $f(A)$ as $n \to \infty$. This implies that, at least for n sufficiently large, the off-diagonal entries of $f(A_n)$ must decay like those of $f(A)$, therefore the decay is that of the algebra $\mathscr{A}$. Note that this approach does not work unless $\mathscr{A}$ is inverse-closed in $\mathcal{B}(\ell^2)$; thus, it cannot be used to prove exponential decay (or superexponential decay when f is entire), since the algebra $\mathscr{E}_\gamma$ is *not* inverse-closed in $\mathcal{B}(\ell^2)$.

3.9 Localization in matrix factorizations

So far we have focused on the decay properties of functions of matrices, including the inverse. In numerical linear algebra, however, matrix factorizations (LU, Cholesky, QR, and so forth) are even more fundamental. What can be said about the localization properties of the factors of a matrix which is itself localized? By *localized* here we mean banded, sparse, or satisfying an off-diagonal decay property.

We say that a matrix A is (m, p)-*banded* if $[A]_{ij} = 0$ for $i - j > m$ and for $j - i > p$. If A is (m, p)-banded and has the LU factorization $A = LU$ with L unit lower triangular and U upper triangular (without pivoting), it is clear that the triangular factors L and U have, respectively, lower bandwidth m and upper bandwidth p. A similar observation applies to the Cholesky and QR factors.

For more general sparse matrices the situation is more complicated, because of the fill-in that usually occurs in the factors of a sparse matrix and due to the fact that reorderings (row and column permutations) are usually applied in an attempt to preserve sparsity. Nevertheless, much is known about the nonzero structure of the triangular factors, especially in the case of Cholesky and QR factorizations.

The decay properties of the *inverse* factors of infinite banded matrices have been studied by a few authors. Existence and bounded invertibility results for triangular factorizations and block factorizations of have been obtained, e.g., in [198, 199], in particular for Toeplitz and block Toeplitz matrices, where the decay properties of the inverse factors were also considered.

For a banded $n \times n$ matrix A_n, example (32) shows that in general we cannot expect decay in the inverse triangular factors as $n \to \infty$ unless some uniform

boundedness condition is imposed on the condition number of A_n. One such result (from [27]) goes as follows. Recall the bound of Demko et al. for the entries of the inverse of an m-banded Hermitian and positive definite A:

$$|[A^{-1}]_{ij}| \leq K \, \rho^{|i-j|}, \quad \forall i, j$$

where $[a, b]$ is the smallest interval containing the spectrum $\sigma(A)$ of A, $K = \max\{a^{-1}, K_0\}$, $K_0 = (1 + \sqrt{\kappa_2(A)})/2b$, $\kappa = \frac{b}{a} = \|A\|_2 \|A^{-1}\|_2$, $\rho = q^{1/m}$, and $q = q(\kappa_2(A)) = \frac{\sqrt{\kappa_2(A)}-1}{\sqrt{\kappa_2(A)}+1}$. With these definitions of K and ρ, we have:

Theorem 24. ([27]) *Let $A = A^* \in \mathbb{C}^{n \times n}$ be positive definite and m-banded, and suppose A has been scaled so that $\max_{1 \leq i \leq n}[A]_{ii} = 1$. Let $A = LL^*$ denote the Cholesky factorization of A. Then*

$$|[L^{-1}]_{ij}| \leq K_1 \, \rho^{i-j}, \qquad \forall \, i > j, \tag{70}$$

where $K_1 = K \frac{1-\rho^m}{1-\rho}$.

In view of the identity $L^{-1} = L^T A^{-1}$, the decay in the inverse factor is a consequence of the fact that the product of a banded matrix times a matrix with exponential decay must necessarily decay as well. Since $K_1 > K$, the bound (70) indicates a potentially slower decay in L^{-1} than in the corresponding entries of A^{-1}, but this is not always the case. For instance, as noted in [27], the entries of L^{-1} must actually be smaller than the corresponding entries of A^{-1} when A is an M-matrix.

As usual, Theorem 24 can be applied to a sequence $\{A_n\}$ of m-banded, Hermitian positive definite matrices of increasing order, such that $\sigma(A_n) \subset [a, b]$ for all n, normalized so that their largest entry is equal to 1. The theorem then gives a uniform (in n) exponential decay bound on the entries of the inverse Cholesky factors L_n^{-1}, as $n \to \infty$. If, on the other hand, the condition numbers $\kappa_2(A_n)$ grow unboundedly for $n \to \infty$, the bounds (70) depend on n, and will deteriorate as $n \to \infty$. This is the case, for instance, of matrices arising from the discretization of partial differential equations. Nevertheless, sharp decay bounds on the elements of the inverse Cholesky factor of sparse matrices arising from the discretization of certain PDEs have been recently obtained in [49].

What about the case of matrices with decay, which may be full rather than banded or sparse? When are the factors of a localized matrix themselves localized? A wealth of results for matrices belonging to different decay algebras have been obtained by Blatov [33, 34] and more recently by Kryshtal et al. [129]. Roughly speaking, these papers show that for the most frequently encountered decay algebras $\mathscr{A}$, if a matrix $A \in \mathscr{A}$ admits the LU factorization in $\mathcal{B}(\ell^2)$, then the factors belong to $\mathscr{A}$, hence they satisfy the same decay condition as A; if, moreover, the algebra $\mathscr{A}$ is inverse-closed in $\mathcal{B}(\ell^2)$, then obviously the inverse factors L^{-1}, U^{-1} must satisfy the same decay bounds. Analogous results, under the appropriate technical conditions, apply to the Cholesky, QR, and polar factorizations. We refer to [33, 34] and [129] for details.

3.10 Localization in the unbounded case.

Throughout this paper, we have limited our discussion of localization to the following situations:

- finite matrices of fixed size;
- sequences of matrices of increasing size, with uniformly bounded spectra;
- bounded infinite matrices on ℓ^2.

A natural question is, to what extent can the decay results for matrix functions (especially the inverse, the exponential, and spectral projectors) be extended to unbounded operators[11] or to sequences of matrices of increasing size not having uniformly bounded spectra? We know from simple examples that in general we cannot hope to find straightforward extensions of (say) exponential decay bounds of the type (15) or (28) without the boundedness assumption on the spectra. On the other hand, classical exponential decay results for the eigenfunctions of certain elliptic operators (e.g., [2, 54, 75, 183]) and for the Green's function of parabolic operators (like the *heat kernel* for a fixed time t, see e.g. [204, page 328]) show that exponential and even superexponential decay with respect to space do occur for certain functions of unbounded operators. It is reasonable to expect that similar results should hold for discretizations of these operators that lead to sparse matrices; ideally, one would like to obtain decay rates that do not depend on discretization parameters, and unfortunately decay bounds like the ones in Theorem 8 fail to meet this requirement.

More in detail, suppose A is a self-adjoint, unbounded operator defined on a dense subspace of a Hilbert space $\mathscr{H}$, and that f is a function defined on the spectrum of $A = A^*$. If f is an essentially bounded function defined on $\sigma(A)$, the spectral theorem for self-adjoint operators (see, e.g., [176]) allows one to define the function $f(A)$ via the integral representation

$$f(A) = \int_{-\infty}^{\infty} f(\lambda) \, dE(\lambda) \, ,$$

where E is the spectral family (resolution of the identity) associated with A which maps Lebesgue-measurable subsets of $\sigma(A)$ to the algebra $\mathcal{B}(\mathscr{H})$, such that $E(\sigma(A)) = I$. Note that $\sigma(A) \subseteq \mathbb{R}$ is now unbounded. Since $f \in L^{\infty}(\sigma(A))$, clearly $f(A) \in \mathcal{B}(\mathscr{H})$. Thus, bounded functions of unbounded operators are bounded operators. Non-trivial examples include the Cayley transform,

$$\Psi(A) = (A - iI)(A + iI)^{-1},$$

a unitary (and therefore bounded) operator, and the exponential e^{itA}, also unitary. Furthermore, the exponential e^{-tA} (when A is positive definite) and the closely related resolvent $(A - zI)^{-1}$ (with $z \notin \sigma(A)$) are compact, and therefore bounded, for some important classes of unbounded operators. Another obvious example is given by the spectral projectors, since in this case the range of f is just the set $\{0, 1\}$. In such cases it is sometimes possible to obtain exponential decay results for certain (analytic) functions of banded, unbounded operators.

Remark 8. The case in which $f(A)$ is compact is especially favorable: since every compact operator on a separable Hilbert space is the norm-limit of finite rank operators, the entries of $f(A)$ (represented by an infinite matrix with respect to an arbitrary orthonormal basis on $\mathscr{H}$) must have rapid decay away from a *finite* set of positions, including down the main diagonal. If in addition $f(A)$ is *trace class*

---

[11] For the sake of simplicity, we only consider the self-adjoint case here.

$(\sum_{i=1}^{\infty}[f(A)]_{ii} < \infty)$ and in particular *Hilbert–Schmidt* $(\sum_{i,j=1}^{\infty}|[f(A)]_{ij}|^2 < \infty)$, decay must be quite fast, though not necessarily exponential.

To the best of our knowledge, only isolated results are available in the literature. An example, already mentioned, is that of e^{itA} for a specific class of unbounded tridiagonal matrices A on $\ell^2(\mathbb{Z})$. Additional examples can be found in [185] and [120]. In these papers one can find exponential localization results for the resolvent and eigenfunctions (and thus spectral projectors) of certain infinite banded matrices of physical interest. Very recently, sharp decay estimates of discretized Green's functions for a broad class of Schrödinger operators have been obtained in [138]. These decay results are established for finite difference and pseudo-spectral discretizations, using methods similar to those used to establish the decay properties of the continuous Green's function (see, e.g., [169]). The advantage of these bounds is that they do not deteriorate as the mesh parameter h tends to zero, and thus they are able to capture the exponential decay in the Green's function, when present.

It would be desirable to investigate to what extent one can derive general decay results for bounded analytic functions of unbounded banded (or sparse) infinite matrices, analogous to those available in the bounded case.

4 Applications

In this section we discuss a few selected applications of the theory developed so far. We focus on two broad areas: numerical linear algebra, and electronic structure computations. Algorithmic aspects are also briefly mentioned. The following is not intended as an exhaustive discussion, but rather as a sampling of current and potential applications with pointers to the literature for the interested reader.

4.1 Applications in numerical linear algebra

The decay properties of inverses and other matrix functions have been found useful in various problems of numerical linear algebra, from solving linear systems and eigenvalue problems to matrix function evaluation. Below we discuss a few of these problems.

Linear systems with localized solutions. A possible application of inverse decay occurs when solving linear systems $A\mathbf{x} = \mathbf{b}$ with a localized right-hand side $\mathbf{b}$. For example, if $\mathbf{b} = \alpha\mathbf{e}_i$ where $\mathbf{e}_i$ is the ith standard basis vector, the solution is given by $\mathbf{x} = \alpha A^{-1}\mathbf{e}_i$, a multiple of the ith column of A^{-1}. If it is known that A^{-1} decays rapidly away from certain positions, the solution vector will be localized around the corresponding positions in $\mathbf{x}$. The same holds if $\mathbf{b}$ contains not just one but $k \ll n$ nonzero entries, or if it is a dense but localized vector. Problems of this kind arise frequently in applications, where the right-hand side $\mathbf{b}$ often corresponds to a localized forcing term such as a point load or a source (of heat, of neutrons, etc.) located in a small subregion of the computational domain. In each of these cases, bounds on the entries of A^{-1} can be used to determine a priori an "envelope" containing those parts of the solution vector $\mathbf{x}$ in which the nonnegligible entries

are concentrated. Even if the bounds are pessimistic, this can lead to worthwhile computational savings.

Of course, this approach requires the use of algorithms for solving linear systems that are capable of computing only selected parts of the solution vector. Available methods include variants of Gaussian elimination [73, Chapter 7.9], Monte Carlo linear solvers [21], and quadrature rule-based methods for evaluating bilinear forms $\mathbf{u}^T A^{-1}\mathbf{v}$ (in our case $\mathbf{u} = \mathbf{e}_j$ and $\mathbf{v} = \mathbf{b}$, since $x_j = \mathbf{e}_j^T A^{-1}\mathbf{b}$), see [40, 95].

It is worth mentioning that this problem is somewhat different from that arising in compressed sensing, where one looks for solutions that have (near-)maximal sparsity in a non-standard basis, leading to the problem of finding the "sparsest" solution among the infinitely many solutions of an underdetermined system [46, 47].

Construction of preconditioners. The results on the exponential decay in the inverses of band and sparse matrices, originally motivated by the converge analysis of spline approximations [68], have been applied to the construction of preconditioners for large, sparse systems of linear equations. Specifically, such results have been used, either implicitly or explicitly, in the development of block incomplete factorizations and of sparse approximate inverse preconditioners.

A pioneering paper on block incomplete factorizations is [55], where various block incomplete Cholesky preconditioners are developed for solving large, symmetric positive definite block tridiagonal linear systems with the preconditioned conjugate gradient method. This paper has inspired many other authors to develop similar techniques, including preconditioners for nonsymmetric problems; see, e.g., [7, Chapter 7], and [8, 11, 33, 35, 201] among others.

Consider a large, sparse, block tridiagonal matrix (assumed to be symmetric positive definite for simplicity):

$$
A = \begin{bmatrix}
A_1 & B_2^T & & & & \\
B_2 & A_2 & B_3^T & & & \\
& B_3 & A_3 & B_4^T & & \\
& & \ddots & \ddots & \ddots & \\
& & & \ddots & \ddots & \ddots \\
& & & & B_{p-1} & A_{p-1} & B_p^T \\
& & & & & B_p & A_p
\end{bmatrix},
$$

where the blocks A_i and B_i are typically banded; for example, in [55] the diagonal blocks A_i are all tridiagonal, and the off-diagonal nonzero blocks B_i are diagonal. Then A has a block Cholesky factorization of the form

$$
A = (L + D)D^{-1}(L + D)^T
$$

where L is block strictly lower triangular and D is block diagonal with blocks

$$
\Delta_1 = A_1, \qquad \Delta_i := A_i - B_i \Delta_{i-1}^{-1} B_i^T, \quad i = 2, \ldots, p. \tag{71}
$$

The successive Schur complements Δ_i in (71) are the *pivot blocks* of the incomplete block Cholesky (more precisely, block LDL^T) factorization. They are dense matrices for $i = 2, \ldots, p$. An incomplete block factorization can be obtained by approximating them with sparse matrices:

$$\Delta_1^{-1} \approx \Sigma_1, \quad \Delta_i \approx A_i - B_i \Sigma_{i-1} B_i^T, \quad i = 2, \ldots, p,$$

where $\Sigma_i \approx \Delta_i^{-1}$ for $2 \le i \le p$ is typically a banded approximation. Estimates of the decay rates of the inverses of band matrices can then be used to determine the bandwidth of the successive approximations to the pivot blocks. We refer to the above-given references for details on how these banded approximations can be obtained.

Unfortunately, unless the pivot blocks are sufficiently diagonally dominant they cannot be well-approximated by banded matrices. For these reasons, more sophisticated (albeit generally more expensive) approximations have been developed based on hierarchical matrix techniques in recent years [11]. Nevertheless, cheap approximations to Schur complements using banded or sparse approximations to the inverses of the blocks may be sufficient in some applications; see, e.g., [144, 182].

Preconditioners for general sparse matrices based on sparse approximate inverses, the first examples of which date back to the 1970s, have been intensively developed beginning in the 1990s; see, e.g., [17] for a survey, and [177, Chapter 10.5] for a self-contained discussion. More recently, interest in these inherently parallel preconditioning methods has been revived, due in part to the widespread diffusion of Graphic Processing Units (GPUs). In these methods, the inverse of the coefficient matrix is approximated directly and explicitly by a sparse matrix $M \approx A^{-1}$; in some cases M is the product of two sparse triangular matrices which approximate the inverse triangular factors L, U of A. Applying the preconditioner only requires matrix-vector products, which are much easier to parallelize than triangular solves [100]. The main challenge in the construction of these preconditioners is the determination of a suitable sparsity pattern for M. Indeed, if a "good" sparsity pattern can be estimated in advance, the task of computing a sparse approximate inverse with a nonzero pattern that is a subset of the given one is greatly facilitated [53, 100, 115]. If A is banded, a banded approximate inverse may suffice. If A is not banded but strongly diagonally dominant, a sparse approximate inverse with the same nonzero pattern as A will usually do. In other cases, using the sparsity pattern of A^2 will give better results, although at a higher cost since A^2 may be considerably less sparse than A. The rationale for considering the patterns of successive powers of A is the following. Suppose A is diagonally dominant. Then, up to a diagonal scaling, it can be written as $A = I - B$ for some matrix $B \in \mathbb{C}^{n \times n}$ with $\varrho(B) < 1$. Therefore

$$A^{-1} = (I - B)^{-1} = I + B + B^2 + \cdots, \tag{72}$$

where the entries of B^k must decay rapidly to zero as $k \to \infty$ since A is diagonally dominant. Since B and A have the same pattern, (72) suggests that using the sparsity pattern of A for M may be sufficient, especially if A exhibits very strong diagonal dominance; if not, higher order terms in (72) may have to be considered. Note that considering higher powers of A means looking beyond the immediate neighbors of each node in the sparsity graph $\mathcal{G}(A)$ of A; in view of bounds like (21), according to which the entries in A^{-1} decay rapidly with the distance from the nonzeros in A, it shouldn't be necessary to look at high powers, and indeed in practice one rarely goes beyond A^2. For highly nonsymmetric matrices, Huckle [115] has shown that the nonzero pattern of A^T and its powers may have to be considered as well. We refer to [3, 53] for further details.
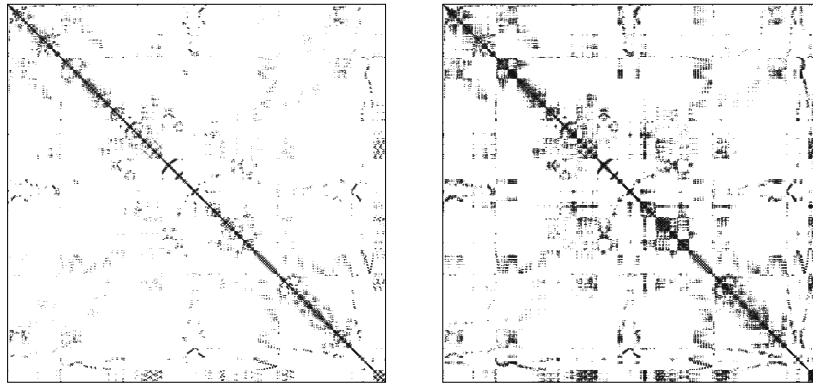
Fig. 11. Left: nonzero pattern of A. Right: pattern of approximate inverse of A.

As an example, in Fig. 11 we show on the left the nonzero pattern of a complex symmetric matrix A arising from an application to computational electromagnetics (see [3] for details), and on the right the sparsity pattern of an approximation M to A^{-1} corresponding to the nonzero pattern of A^2. The approximate inverse M was computed by minimizing the Frobenius norm $\|I - AM\|_F$ over all matrices with the same sparsity pattern of A^2. With this preconditioner, GMRES requires 80 iterations to reduce the relative residual norm below 10^{-8} (the method stagnates without preconditioning). It is worth noting that computing the exact inverse A^{-1} and dropping all entries with $|[A^{-1}]_{ij}| < \varepsilon\|A\|_F$ with $\varepsilon = 0.075$ produces a sparse matrix with a nonzero pattern virtually identical to the one in Fig. 11 (right).

In the construction of *factorized* approximate inverse preconditioners (like FSAI [127] and AINV [23, 26]), it is useful to know something about the decay in the inverse triangular factors of A. Hence, bounds like (70) provide some insight into the choice of a good approximate sparsity pattern and on the choice of ordering [27]. Similar remarks apply to more traditional incomplete LU and QR factorization using the results in section 3.9 on the localization in the factors of matrices with decay. See also [49] and [140] for other examples of situations where knowledge of decay bounds in the inverse Cholesky factor is useful for numerical purposes.

In recent years, much attention has been devoted to the numerical solution of fractional differential equations. Discretization of these non-local equations leads to dense matrices which are usually not formed explicitly. Matrix-vector multiplications (needed in the course of Krylov subspace iterations) can be efficiently computed in $O(n \log n)$ work using Fast Fourier Transforms (FFTs) and diagonal scalings. In [162], preconditioning techniques for linear systems arising from the discretization of certain initial boundary value problems for a fractional diffusion equation of order $\alpha \in (1, 2)$ are introduced and analyzed. At each time step, a nonsymmetric linear system $A\mathbf{x} = \mathbf{b}$ must be solved, where A is of the form

$$A = \eta I + DT + WT^T,$$

with $\eta > 0$, D, W diagonal and nonnegative, and T a lower Hessenberg Toeplitz matrix. As the matrices D and W change at each time step, A also changes, and

it is therefore important to develop preconditioners that are easy to construct and apply, while at the same time resulting in fast convergence rates of the preconditioned Krylov iteration. The preconditioners studied in [162] are based on circulant approximations, FFTs and interpolation and the authors show both theoretically and numerically that they are effective. The theoretical analysis in [162], which shows that the preconditioned matrices have spectra clustered around unity, makes crucial use of inverse-closed decay algebras. In particular, the authors show that T, and therefore A and the circulant approximations used in constructing the preconditioners, all belong to the Jaffard algebra $Q_{\alpha+1}$, where α is the fractional order of the spatial derivatives in the differential equation (see Def. 5). This fact is used in establishing the spectral properties of the preconditioned matrices.

Localization and eigenvalue problems. Parlett [163, 164] and Vömel and Parlett [202] have observed that the eigenvectors corresponding to isolated groups of eigenvalues of symmetric tridiagonal matrices are often localized—an observation already made by Cuppen [60].[12] In [202], Vömel and Parlett develop heuristics for estimating envelopes corresponding to nonnegligible entries of eigenvectors of tridiagonals, and show that knowledge of these envelopes can lead to substantial savings when the eigenvectors are localized and when solvers that are able to compute only prescribed components of eigenvectors are used, such as inverse iteration and the MRRR algorithm [71, 146, 165]. They also observe that eigenvectors corresponding to isolated eigenvalue clusters are not always localized, and that *a priori* detection of localization of the eigenvectors poses a challenge.

To see how the theory of decay in matrix functions can help address this challenge, we recast the problem in terms of spectral projectors instead of eigenvectors. Let $A = A^* \in \mathbb{C}^{n \times n}$ have eigenvalues

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

(in the tridiagonal case we can assume that A is irreducible so that its eigenvalues are all simple, see [96, page 467]). Suppose now that eigenvalue λ_{k+1} is well-separated from λ_k, and that eigenvalue λ_{k+p} is well-separated from λ_{k+p+1}. If $\mathbf{v}_i$ denotes an eigenvector corresponding to λ_i with $\|\mathbf{v}_i\|_2 = 1$, the spectral projector associated with the group of eigenvalues $\{\lambda_{k+1}, \ldots, \lambda_{k+p}\}$ can be written as

$$P = \mathbf{v}_{k+1}\mathbf{v}_{k+1}^* + \cdots + \mathbf{v}_{k+p}\mathbf{v}_{k+p}^* = VV^*, \tag{73}$$

where

$$V = [\mathbf{v}_{k+1}, \ldots, \mathbf{v}_{k+p}] \in \mathbb{C}^{n \times p} \tag{74}$$

is a matrix with orthonormal columns. Note that P is the orthogonal projector onto the A-invariant subspace $\mathcal{V} = \mathrm{span}\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_{k+p}\}$.

Let us first consider the case of a single, simple eigenvalue, $p = 1$. If $\mathbf{v}$ is the corresponding normalized eigenvector, the spectral projector is of the form

$$P = \mathbf{v}\mathbf{v}^* = \begin{bmatrix} |v_1|^2 & v_1\bar{v}_2 & \ldots & v_1\bar{v}_n \\ v_2\bar{v}_1 & |v_2|^2 & \ldots & v_2\bar{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ v_n\bar{v}_1 & v_n\bar{v}_2 & \ldots & |v_n|^2 \end{bmatrix}.$$

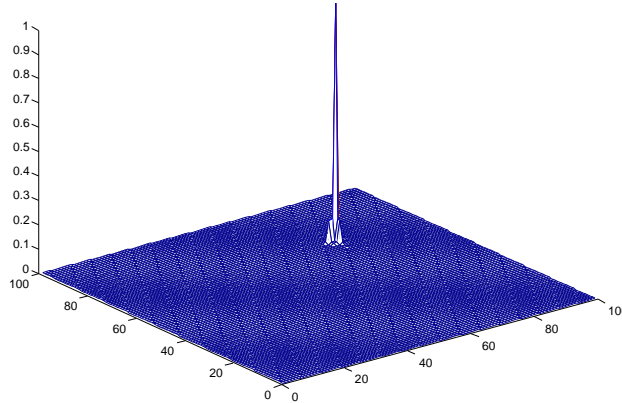[12] See also Exercise 30.7 in Trefethen and Bau [195].

Fig. 12. Plot of $|[P]_{ij}|$ where P is the spectral projector onto the eigenspace corresponding to an isolated eigenvalue of a tridiagonal matrix of order 100.

Conversely, given a rank-1 projector P, the eigenvector $\mathbf{v}$ is uniquely determined (up to a constant). It is also clear that if $\mathbf{v} = (v_i)$ is a vector such that $|v_i| \ll |v_j|$ for $i \neq j$, then the entries of P must decay rapidly away from $[P]_{ii} = |v_i|^2$ (not only away from the main diagonal, but also along the main diagonal). More generally, if most of the "mass" of $\mathbf{v}$ is concentrated in a few components, the entries of P must decay rapidly away from the corresponding diagonal entries. Conversely, it is evident that rapid decay in P implies that $\mathbf{v}$ must itself be localized. Hence, in the rank-1 case $P = \mathbf{v}\mathbf{v}^*$ is localized if and only if $\mathbf{v}$ is. An example is shown in Fig. 12.

On the other hand, in the case of spectral projectors of the form (73) with $p > 1$, localization in the eigenvectors $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_{k+p}$ is a *sufficient* condition for localization of P, but not a necessary one . This is due to the possible (near-)cancellation in the off-diagonal entries when adding up the rank-1 projectors $\mathbf{v}_j \mathbf{v}_j^*$. This fact becomes actually obvious if one observes that summing *all* the projectors $\mathbf{v}_j \mathbf{v}_j^*$ for $j = 1, \ldots, n$ must result in the identity matrix, which is maximally localized even though the eigenvectors may be strongly delocalized. Less trivial examples (with $1 < p \ll n$) can be easily constructed. Real-world instances of this phenomenon are actually well known in physics; see, e.g., [45] and section 4.2 below.

We also remark that if P is localized, there may well be another orthonormal basis $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_{k+p}\}$ of $\mathcal{V}$, different from the eigenvector basis, which is localized. When $p > 1$, P does not determine the basis vectors uniquely. Indeed, if $\Theta \in \mathbb{C}^{p \times n}$ is any matrix with orthonormal rows, we have that

$$P = VV^* = V\Theta\Theta^*V^* = UU^*, \qquad U = V\Theta,$$

which shows how $U = [\mathbf{u}_{k+1}, \ldots, \mathbf{u}_{k+p}]$ is related to V. Even if the columns of V are not localized, those of U may well be, for a suitable choice of Θ. We note, on the other hand, that if P is delocalized then there can be no strongly localized basis vectors $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_{k+p}\}$ for $\mathcal{V}$. Nevertheless, searching for an orthonormal

basis that is "as localized as possible" is an important problem in certain physics applications; see, e.g., [62, 63, 93].

Now that we have recast the problem in terms of spectral projectors, we can apply the theory of decay in matrix functions. Indeed, the spectral projector is a function of A: if $A = A^*$ has the spectral decomposition $A = Q\Lambda Q^*$, where Q is unitary and $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$, then

$$P = \phi(A) = Q\,\phi(\Lambda)\,Q^*, \tag{75}$$

where ϕ is any function such that

$$\phi(\lambda_i) = \begin{cases} 1, & \text{if } k + 1 \le i \le k + p, \\ 0, & \text{else.} \end{cases}$$

Hence, any analytic function that interpolates ϕ at the eigenvalues of A will do; in practice, it is sufficient to use an analytic function that approximates ϕ on the spectrum of A. For example, any function f such that $f(\lambda) \approx 1$ for $\lambda \in [\lambda_{k+1}, \lambda_{k+p}]$ which drops rapidly to zero outside this interval will be an excellent approximation of P. It is easy to see that the wider the gaps $(\lambda_k, \lambda_{k+1})$ and $(\lambda_{k+p}, \lambda_{k+p+1})$, the easier it is to construct such an analytic approximation of $\phi(\lambda)$, and the faster the off-diagonal decay is in $f(A)$ and therefore in P, assuming of course that A is banded or sparse.

As an illustration, consider the case of an isolated eigenvalue $\lambda \in \sigma(A)$. Since an eigenvector of A associated with λ is an eigenvector of $A - \lambda I$ associated with 0, we can assume without any loss of generality that $\lambda = 0$. Let $\mathbf{v}$ be this eigenvector (with $\|\mathbf{v}\|_2 = 1$) and let $P = \mathbf{v}\mathbf{v}^*$ be the corresponding spectral projector. The function ϕ such that $P = \phi(A)$ can be approximated to within arbitrary accuracy by a Gaussian

$$f(x) = \mathrm{e}^{-x^2/\xi}, \quad \text{where } \xi > 0. \tag{76}$$

The choice of ξ, which controls the rate of decay to zero of the Gaussian, will depend on the desired accuracy and thus on the distance between the eigenvalue $\lambda = 0$ and its nearest neighbor in the spectrum of A; we denote this distance by η. To determine ξ, suppose we wish to have $f(\pm\eta) \le \varepsilon$ for a prescribed $\varepsilon > 0$. Thus, we require that

$$\mathrm{e}^{-\eta^2/\xi} \le \varepsilon,$$

which yields

$$0 < \xi \le -\eta^2/\log(\varepsilon).$$

For instance, given $\varepsilon > 0$ we can approximate P by

$$P \approx f(A) = \exp(-A^2/\xi), \quad \xi = -\eta^2/\log(\varepsilon).$$

It is shown in [171] that this approximation works very well. For instance, for $\varepsilon = 10^{-8}$ and $\eta = 0.1$, choosing $\xi = 3 \cdot 10^{-4}$ yields $\|P - f(A)\|_F = 7 \cdot 10^{-14}$. Moreover, specializing Theorem 8 to the Gaussian (76) leads to the following off-diagonal decay result (Theorem 4.2 in [171]):

Theorem 25. ([171]) *Let $\varepsilon > 0$ and let f be given by (76) with $\xi = -c\eta^2$, $c = 1/\log(\varepsilon)$. Let $A = A^*$ be m-banded and assume that $[-1, 1]$ is the smallest interval containing $\sigma(A)$. Then, for $i \ne j$ we have*

$$|[\exp(-A^2/\xi)]_{ij}| \le K\,\rho^{|i-j|}, \tag{77}$$

where

$$K = \frac{2\chi e^{c\alpha^2/\eta^2}}{\chi-1}, \quad \alpha > 1, \quad \chi = \alpha + \sqrt{\alpha^2-1}, \quad \rho = \chi^{-1/m}.$$

Note that the assumption that $\sigma(A) \subset [-1,1]$ leads to no loss of generality, since spectral projectors are invariant under shifting and scaling of A. Recalling that for any projector $|[P]_{ij}| \le 1$ for all i,j, it is clear that the bound (77) is only informative if the quantity on the right-hand side is less than 1, which may require taking $|i-j|$ sufficiently large. This theorem provides an infinite family of bounds parameterized by $\alpha > 1$ (equivalently, by $\chi > 1$). Hence, the entries of $f(A)$, and thus of P, satisfy a superexponential off-diagonal decay (this is expected since f is entire). Note that there is fast decay also along the main diagonal: this is obvious since for an orthogonal projector,

$$\mathrm{Tr}(P) = \mathrm{rank}(P), \tag{78}$$

and since the diagonal entries of P are all positive, they must decrease rapidly away from the $(1,1)$ position for the trace of P to be equal to 1. With this, the localization of the spectral projector (and thus of a normalized eigenvector) corresponding to isolated eigenvalues of banded matrices (in particular, tridiagonal matrices) is rigorously established.

The above construction can be extended to approximate the spectral projector P corresponding to a group of k well-separated eigenvalues: in this case P can be well-approximated by a sum of rapidly decaying Gaussians centered at the eigenvalues in the given group, hence P will again exhibit superexponential off-diagonal decay, with k spikes appearing on the main diagonal.

The use of a single shifted Gaussian (centered at a prescribed value μ) with a suitable choice of the parameter ξ can also be used to approximate the spectral projector corresponding to a tight cluster of several eigenvalues falling in a small interval around μ. Combined with a divide-and-conquer approach, this observation is at the basis of the recently proposed *localized spectrum slicing* (LSS) technique for computing interior eigenpairs of large, sparse, Hermitian matrices, see [137]. Unlike most current methods, this technique does not require the solution of highly indefinite, shifted linear systems. The decay theory for analytic functions of sparse matrices plays a central role in the development and analysis of this algorithm, which is shown in [137] to have linear cost in n. It should also be noted that the LSS algorithm has controllable error.

On the other hand, a different function f must be used if the eigenvalues of interest form an isolated *band*, i.e., they densely fill an interval which is well separated from the rest of the spectrum. This situation, which is of importance in physical applications, will be discussed in section 4.2 below. As we shall see, in this case only exponential off-diagonal decay should be expected. Nevertheless, this gives a partial answer to the problem posed by Vömel and Parlett in [202]: even though the *eigenvectors* corresponding to an isolated cluster of eigenvalues of a tridiagonal A may fail to be localized, the corresponding spectral projector will be localized, the more so the larger the relative gap between the cluster and the rest of the spectrum.
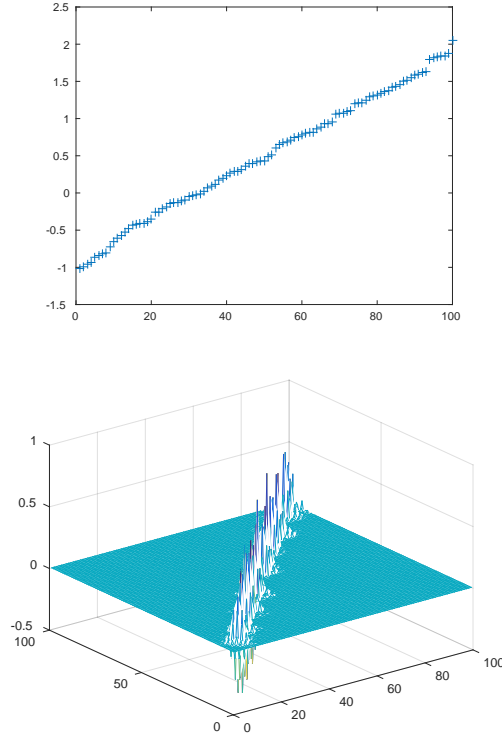
Fig. 13. Top: eigenvalues of a random tridiagonal. Bottom: spectral projector corresponding to 10 smallest eigenvalues.

We emphasize, however, that the gap assumption is only a *sufficient condition*, not a necessary one. If A is a large tridiagonal matrix without any discernible gap in the spectrum, its eigenvectors may or may not be localized. For instance, the $n \times n$ tridiagonal matrix

$$A_n = \text{tridiag}(-1, 2, -1)$$

presents no gaps in the spectrum as $n \to \infty$, and in fact the corresponding infinite tridiagonal matrix A, viewed as a bounded operator on ℓ^2, has purely continuous spectrum: $\sigma(A) = [0, 4]$. As is well known, the eigenvectors of A_n are delocalized, and the orthogonal projectors corresponding to individual eigenvalues or to groups of eigenvalues of A_n are also delocalized with very slow decay as $n \to \infty$ (see [20, Section 10] for a detailed analysis). On the other hand, with very high probability, the eigenvectors (and therefore the spectral projectors) of a randomly generated symmetric tridiagonal matrix will exhibit a high degree of localization, even in the absence of any clearly defined spectral gap between (groups of) eigenvalues. An instance of this behavior is shown in Fig. 13. At the top we show a plot of the eigenvalues of a random symmetric tridiagonal matrix A of order $n = 100$, and at the bottom we display the spectral projector P onto the invariant subspace spanned by the eigenvectors associated with the 10 smallest eigenvalues of A. Note

that there is no clear gap separating the eigenvalues $\lambda_1, \ldots, \lambda_{10}$ from the rest of the spectrum, and yet P exhibits rapid off-diagonal decay. The eigenvectors themselves are also strongly localized. See also the already referenced Exercise 30.7 in [195] and the examples discussed in [41, pages 374–375], where a connection with Anderson localization [4] is made.

Hence, the challenge posed by Vömel and Parlett in [202] remains in part open, both because localization can occur even in the absence of gaps in the spectrum, and because the presence of gaps may be difficult to determine *a priori*.

Another interesting application of off-diagonal decay is to eigenvalue perturbation theory. It turns out that for certain structured matrices, such as tridiagonal or block tridiagonal matrices, the effect of small perturbations in the matrix entries on some of the eigenvalues is much smaller than can be expected from the "standard" theory based on Weyl's Theorem.[13] It has been observed (see, e.g., [157]) that for tridiagonal matrices an eigenvalue is insensitive to perturbations in A if the corresponding eigenvector components are small. In [157], generalizations of this fact are established for block tridiagonal A under suitable assumptions. Hence, eigenvector localization plays an important role in proving much tighter perturbation results when A is block tridiagonal (including the special cases of tridiagonal and general m-banded matrices).

Here we show how localization results like Theorem 25 can shed some light on perturbation theory. Assume $A = A^* \in \mathbb{C}^{n \times n}$ is banded and consider the eigenvalue problem $A\mathbf{v} = \lambda\mathbf{v}$. For simplicity we assume that λ is a simple eigenvalue. If $\mathbf{v}$ is normalized ($\|\mathbf{v}\|_2 = 1$), then

$$\lambda = \mathbf{v}^* A \mathbf{v} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \bar{v}_i v_j.$$

The sensitivity of an eigenvalue λ to small changes in the entries of A can be estimated, to first order, by the partial derivative

$$\frac{\partial \lambda}{\partial a_{ij}} = \bar{v}_i v_j + v_i \bar{v}_j \quad \forall i, j.$$

Now, the (i, j) entry of the spectral projector $P = \mathbf{v}\mathbf{v}^*$ on the eigenspace associated with λ is $[P]_{ij} = v_i \bar{v}_j$. Therefore,

$$|[P]_{ij}| \approx 0 \Rightarrow \lambda \text{ is insensitive to small changes in } a_{ij}.$$

But we know from Theorem 25 that $|[P]_{ij}| \approx 0$ if $|i - j|$ is sufficiently large, since the entries of $|[P]_{ij}|$ satisfy a superexponential decay bound. Thus, perturbing entries of A at some distance from the main diagonal by a small amount $\delta \neq 0$ will cause a change in λ much smaller than δ. The change can be expected to be comparable to δ, on the other hand, if the perturbation occurs in a position (i, j) where $[P]_{ij}$ is not small.

---

[13] Weyl's Theorem implies that the eigenvalues of A and $A + E$ (both Hermitian) can differ by a quantity as large as $\|E\|_2$. See [113, Chapter 4.3] for precise statements.

Example 2. Consider the symmetric tridiagonal matrix

$$
A = \begin{bmatrix}
5 & 1 & & & & & & \\
1 & 0 & 1 & & & & & \\
& 1 & 0 & & 1 & & & \\
& & & \ddots & \ddots & \ddots & & \\
& & & & \ddots & \ddots & \ddots & \\
& & & & & 1 & 0 & 1 \\
& & & & & & 1 & 0
\end{bmatrix}
$$

of order 100. The spectrum of A consists of the eigenvalues $\lambda_1, \ldots, \lambda_{99}$, all falling in the interval $[-1.99902, 1.99902]$, plus the eigenvalue $\lambda = \lambda_{100} = 5.2$. As expected from Theorem 25, the (normalized) eigenvector $\mathbf{v}$ associated with λ is strongly localized:

$$
\mathbf{v} = \begin{bmatrix}
0.979795897113271 \\
0.195959179422654 \\
0.039191835884531 \\
0.007838367176906 \\
0.001567673435381 \\
0.000313534687076 \\
0.000062706937415 \\
0.000012541387483 \\
0.000002508277497 \\
0.000000501655499 \\
\vdots
\end{bmatrix}.
$$

The entries of $\mathbf{v}$ decay monotonically; they are all smaller than the double-precision machine epsilon from the 22nd one on.[14] Hence, $P = \mathbf{v}\mathbf{v}^T$ is strongly localized, and in fact its entries decay very fast away from the $(1,1)$ position. See also Fig. 12 for a similar case corresponding to a well-separated interior eigenvalue.

Let $\tilde{A}$ be the perturbed matrix obtained by replacing the 5 in position $(1,1)$ with the value 5.001. Clearly, $\|A - \tilde{A}\|_2 = 10^{-3}$. We find that the change in the largest eigenvalue is $|\lambda(\tilde{A}) - \lambda(A)| = 9.6 \cdot 10^{-4}$. Hence, the change in the isolated eigenvalue is essentially as large as the change in the matrix; note that the $(1,1)$ entry of the spectral projector, P, is equal to $0.96 \approx 1$.

On the other hand, suppose that the perturbed matrix $\tilde{A}$ is obtained from A by replacing the zero in positions $(10,1)$ and $(1,10)$ of A by $\delta = 10^{-3}$. Again, we have that $\|A - \tilde{A}\|_2 = 10^{-3}$, but the largest eigenvalue of the modified matrix is now $\lambda(\tilde{A}) = 5.2000002$. Hence, in this case a perturbation of size 10^{-3} in A only produces a change of $O(10^{-7})$ in the isolated eigenvalue; note that the $(10,1)$ entry of P is $\approx 4.9152 \cdot 10^{-7}$.

As we have mentioned, rapid decay in P is not limited to the off-diagonal entries: the diagonal entries $[P]_{ii}$ of P also decay superexponentially fast for increasing i. Perturbing the $(2,2)$ entry of A by 0.001 causes a change equal to $3.84 \cdot 10^{-5}$ in the largest eigenvalue, consistent with the fact that $[P]_{2,2} = 3.84 \cdot 10^{-2}$; perturbing the

---

[14] We mention in passing reference [142], where an alternative justification is given for the observed exponential decay in $\mathbf{v}$.

$(5,5)$ entry of A again by 0.001 causes a change equal to $2.458 \cdot 10^{-9}$, consistent with the fact that $[P]_{2,2} = 2.458 \cdot 10^{-6}$. After a perturbation by 0.001 in the (i,i) of A for $i \geq 12$, we find that the largest computed eigenvalue is numerically unchanged at 5.2.

Incidentally, we note that in this example the presence of an isolated eigenvalue can be determined *a priori* from Geršgorin's Theorem. More generally, this theorem can sometimes be used to determine the presence of groups or clusters of eigenvalues well-separated from the rest of the spectrum.

More generally, suppose we are interested in computing the quantity

$$\text{Tr}\,(PA) = \lambda_1 + \lambda_2 + \cdots + \lambda_k, \tag{79}$$

where P is the orthogonal projector onto the invariant subspace spanned by the eigenvectors corresponding to the k smallest eigenvalues of A, assumed to be banded or sparse. This is a problem that occurs frequently in applications, especially in physics.

If the relative gap $\gamma = (\lambda_{k+1} - \lambda_k)/(\lambda_n - \lambda_1)$ is "large", then the entries of P can be shown to decay exponentially away from the sparsity pattern of A, with larger γ leading to faster decay (see section 4.2). Differentiating (79) with respect to a_{ij} shows again that the quantity in (79) is insensitive to small perturbations in positions of A that are far from the nonzero pattern of A. This fact has important consequences in quantum chemistry and solid state physics.

Although we have limited our discussion to Hermitian eigenvalue problems, an identical treatment applies more generally to normal matrices. In the nonnormal case, it is an open question whether decay results (for oblique spectral projectors) can be used to gain insight into the stability of isolated components of the spectrum, or of the pseudo-spectra [197], of a matrix. For a study of localization in the case of *random* nonnormal matrices, see [196].

Approximation of matrix functions. In most applications involving functions of large, sparse matrices, it is required to compute the vector $\mathbf{x} = f(A)\mathbf{b}$ for given $A \in \mathbb{C}^{n \times n}$ and $\mathbf{b} \in \mathbb{C}^{n \times n}$. When $f(A) = A^{-1}$, this reduces to approximating the solution of a linear system. If $\mathbf{b}$ is localized, for example $\mathbf{b} = \mathbf{e}_i$ (or a linear combination of a few standard basis vectors), then the decay in $f(A)$ leads to a localized solution vector $\mathbf{x}$. In this case, similar observations to the ones we made earlier about localized linear system solutions apply.

Suppose now that we want to compute a sparse approximation to $f(A_n)$, where $\{A_n\}$ is a sequence of banded or sparse matrices of increasing size. If the conditions of Theorems 8, 11 or 13 are satisfied, the entries of $f(A_n)$ are bounded in an exponentially decaying manner, with decay rates independent of n; if f is entire, decay is superexponential. In all these cases Theorem 4 ensures that we can find a banded (or sparse) approximation to $f(A_n)$ to within an arbitrary accuracy $\varepsilon > 0$ in $O(n)$ work.

The question remains of *how* to compute these approximations. The above-mentioned theorems are based on the existence of best approximation polynomials $p_k(x)$, such that the error $\|p_k(A_n) - f(A_n)\|_2$ decays exponentially fast with the degree k. Under the assumptions of those theorems, for every $\varepsilon > 0$ one can determine a value of k, independent of n, such that $\|p_k(A_n) - f(A_n)\|_2 < \varepsilon$. Unfortunately, the form of the polynomial p_k is not known, except in very special cases. However,

it is not necessary to make use of the polynomial of best approximation: there may well be other polynomials, also exponentially convergent to f, which can be easily constructed explicitly.

From here on we drop the subscript n and we work with a fixed matrix A, but the question of (in-)dependence of n should always be kept in mind. Suppose first that $A = A^*$ is banded or sparse, with spectrum in $[-1, 1]$; shifting and scaling A so that $\sigma(A) \subset [-1, 1]$ requires bounds on the extreme eigenvalues of A, which can usually be obtained in $O(n)$ work, for instance by carrying out a few Lanczos iterations. Let f be a function defined on a region containing $[-1, 1]$. A popular approach is polynomial approximation of $f(A)$ based on Chebyshev polynomials; see, e.g., [10, 92]. For many analytic functions, Chebyshev polynomials are known to converge very fast; for example, convergence is superexponential for $f(x) = e^x$ and other entire functions.

The following discussion is based on [24] (see also [171]). We start by recalling the matrix version of the classical three-term recurrence relation for the Chebyshev polynomials:

$$T_{k+1}(A) = 2AT_k(A) - T_{k-1}(A), \qquad k = 1, 2, \ldots \tag{80}$$

(with $T_0(A) = I$, $T_1(A) = A$). These matrices can be used to obtain an approximation

$$f(A) = \sum_{k=1}^{\infty} c_k T_k(A) - \frac{c_1}{2} I \approx \sum_{k=1}^{N} c_k T_k(A) - \frac{c_1}{2} I =: p_N(A)$$

to $f(A)$ by truncating the Chebyshev series expansion after N terms. The coefficients c_k in the expansion only depend on f (not on A) and can be easily computed numerically at a cost independent of n using the approximation

$$c_k \approx \frac{2}{M} \sum_{j=1}^{M} f(\cos(\theta_j)) \cos((k-1)\theta_j),$$

where $\theta_j = \pi(j - \frac{1}{2})/M$ with a sufficiently large value of M. Thus, most of the computational work is performed in (80). The basic operation in (80) is the matrix–matrix multiply. If the initial matrix A is m-banded, then after k iterations the matrix $T_{k+1}(A)$ will be km-banded. The *Paterson–Stockmeyer algorithm* can be used to evaluate polynomials in a matrix A with minimal arithmetic complexity, see [166] and [109, pages 73–74]. We also mention [36], where sophisticated algorithms for matrix-matrix multiplication that take decay into account are developed.

In order to have a linear scaling algorithm, it is essential to fix a maximum bandwidth for the approximation $P_N(A)$, which must not depend on n. Then the cost is dominated by the matrix–matrix multiplies, and this is an $O(n)$ operation provided that the maximum bandwidth remains bounded as $n \to \infty$. Similar conclusions apply for more general sparsity patterns, which can be determined by using the structure of successive powers A^k of A. In alternative, dropping elements by size using a drop tolerance is often used, although rigorous justification of this procedure is more difficult.

Let us now consider the error incurred by the series truncation:

$$\|e_N(A)\|_2 = \|f(A) - P_N(A)\|_2, \tag{81}$$

where $P_N(A) = \sum_{k=1}^{N} c_k T_k(A) - \frac{c_1}{2} I$. We limit our discussion to the banded case, but the same arguments apply in the case of general sparsity patterns as well. Since $|T_k(x)| \leq 1$ for all $x \in [-1, 1]$ and $k = 1, 2, \ldots$, we have that $\|T_k(A)\|_2 \leq 1$ for all k, since $\sigma(A) \subset [-1, 1]$. Using this well known property to bound the error in (81), we obtain that

$$\|e_N(A)\|_2 = \left\| \sum_{k=N+1}^{\infty} c_k T_k(A) \right\|_2 \leq \sum_{k=N+1}^{\infty} |c_k|.$$

The last inequality shows that the error defined by (81) only depends on the sum of the absolute values of the coefficients c_k for $k = N + 1, N + 2, \ldots$, but these in turn do not depend on n, the dimension of the matrix we are approximating. Hence if we have a sequence of $n \times n$ matrices $\{A_n\}$ with $\sigma(A_n) \subset [-1, 1]$ for all n, we can use an estimate of the quantity $\sum_{k=N+1}^{\infty} |c_k|$ (see, for instance, [30, Equations (2.2)-(2.3)]) and use that to prescribe a sufficiently large bandwidth (sparsity pattern) to ensure a prescribed accuracy of the approximation. As long as the bandwidth of the approximation does not exceed the maximum prescribed bandwidth, the error is guaranteed to be n-independent. In practice, however, we found that this strategy is too conservative. Because of the rapid decay outside of the bandwidth of the original matrix, it is usually sufficient to prescribe a much smaller maximum bandwidth than the one predicted by the truncation error. This means that numerical dropping is necessary (see below for a brief discussion), since the bandwidth of $P_N(A)$ rapidly exceeds the maximum allowed bandwidth. Because of dropping, the simple error estimate given above is no longer rigorously valid. The numerical experiments reported in [24], however, suggest that n-independence (and therefore linearly scaling complexity and storage requirements) is maintained.

We now turn to the problem of approximating $f(A)$ for a general A with spectrum contained in an arbitrary continuum $\mathcal{F} \subset \mathbb{C}$; for a more detailed description of the technique we use, see [192]. In this case we can use a (Newton) interpolation polynomial of the form

$$P_N(A) = c_0 I + c_1(A - z_0 I) + c_2(A - z_0 I)(A - z_1 I) + \cdots + c_N(A - z_0 I)\ldots(A - z_{N-1} I)$$

where c_k is the divided difference of order k, i.e.,

$$c_k = f[z_0, \ldots, z_k], \qquad k = 0, \ldots, N.$$

For $k = 0, \ldots, N - 1$, the interpolation points are chosen as $z_k = \Psi(\omega_k)$, where ω_k are the $N - 1$ roots of the equation $\omega^{N-1} = \rho$ and $\Psi(z)$ is the inverse of the map $\Phi(z)$ that maps the complement of $\mathcal{F}$ to the outside of a disk with radius ρ and satisfies the normalization conditions (33). This method does not require the computation of Faber polynomials and their coefficients. However, it does require knowledge of the map $\Psi(z)$. For specific domains $\mathcal{F}$ this map can be determined analytically, see for example [29, 192]. In addition, $\Psi(z)$ may require information on the convex hull of the eigenvalues of A. For more general domains one may have to resort to numerical approximations to compute $\Psi(z)$; see [194]. Once again, the approximation algorithm requires mostly matrix–matrix multiplies with banded (or sparse) matrices and appropriate sparsification is generally required to keep the cost within $O(n)$ as the problem size n grows. A rigorous error analysis that takes dropping as well as truncation into account is however still lacking.

We briefly discuss now numerical dropping. The idea applies to more general sparsity patterns, but we restrict our discussion to the case where A is a banded matrix with bandwidth m. In this case we only keep elements inside a prescribed bandwidth $\hat{m}$ in every iteration. For given ρ and R (see (41)) we choose $\hat{m}$ a priori so as to guarantee that

$$(\rho/R)^{\hat{m}} \approx \varepsilon/K$$

where $K > 0$ is the constant for the bounds on $|[f(A)]_{ij}|$ (with $i \neq j$) appearing in Theorem 13 (for instance) and $\varepsilon > 0$ is a prescribed tolerance. As already noted, if A is banded with bandwidth m, then A^k has bandwidth km. This means that if we want the approximation to have a fixed bandwidth $\hat{m}$, where $\hat{m}$ is (say) an integer multiple of m corresponding to a prescribed approximation error ε, then we ought to truncate the expansion at the N^*th term, with $N^* = \hat{m}/m$. It may happen, however, that this value of N is actually too small to reduce the error below the prescribed threshold. In this case it is necessary to add extra terms to the Chebyshev expansion; but this would lead to an increase of the bandwidth beyond the prescribed limit. A solution that has been used by physicists is simply to continue the recurrence but ignoring all entries in positions outside the prescribed bandwidth; By restricting all the terms in the three-term recurrence (80) to have a fixed bandwidth (independent of n and N) we obtain an approximation scheme whose cost scales linearly in the size n of the problem. This, however, leaves open the problem of controlling the approximation error.

Approximation based on quadrature rules. Another approach that can sometimes be used to find a banded or sparse approximation of a rapidly decaying matrix $f(A)$ is based on Gaussian quadrature [95]. With this approach it is possible to compute or estimate individual entries in $f(A)$. There exist also block versions of these techniques which allow the computation of several entries of $f(A)$ at once [95, 172]. Thus, if we know that only the entries of $f(A)$ within a certain bandwidth or block structure are nonnegligible, one can use Gaussian quadrature rules to estimate the entries within this bandwidth or blocks. This approach has been used, e.g., in [22] to construct simple banded preconditioners for Toeplitz matrices with decay, using the function $f(A) = A^{-1}$.

Here we briefly sketch this technique. Suppose f is strictly completely monotonic on an interval (a, b) (see Definition 3). For instance, the function $f(x) = x^{-\sigma}$ is strictly completely monotonic on $(0, \infty)$ for any $\sigma > 0$, and $f(x) = e^{-x}$ is strictly completely monotonic on $\mathbb{R}$.

Now, let $A = A^T \in \mathbb{R}^{n \times n}$. Consider the eigendecompositions $A = Q\Lambda Q^T$ and $f(A) = Qf(\Lambda)Q^T$. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ we have

$$\mathbf{u}^T f(A)\mathbf{v} = \mathbf{u}^T Q f(\Lambda) Q^T \mathbf{v} = \mathbf{p}^T f(\Lambda)\mathbf{q} = \sum_{i=1}^{n} f(\lambda_i)p_i q_i, \qquad (82)$$

where $\mathbf{p} = Q^T \mathbf{u}$ and $\mathbf{q} = Q^T \mathbf{v}$. In particular, we have that $[f(A)]_{ij} = \mathbf{e}_i^T f(A)\,\mathbf{e}_j$.

Next, we rewrite the expression in (82) as a Riemann–Stieltjes integral with respect to the spectral measure:

$$\mathbf{u}^T f(A)\mathbf{v} = \int_a^b f(\lambda)\mathrm{d}\mu(\lambda), \quad \mu(\lambda) = \begin{cases} 0, & \text{if } \lambda < a = \lambda_1, \\ \sum_{j=1}^{i} p_j q_j, & \text{if } \lambda_i \leq \lambda < \lambda_{i+1}, \\ \sum_{j=1}^{n} p_j q_j, & \text{if } b = \lambda_n \leq \lambda. \end{cases}$$

The general Gauss-type quadrature rule gives in this case:

$$\int_a^b f(\lambda)\mathrm{d}\mu(\lambda) = \sum_{j=1}^{N} w_j f(t_j) + \sum_{k=1}^{M} v_k f(z_k) + R[f], \tag{83}$$

where the nodes $\{t_j\}_{j=1}^N$ and the weights $\{w_j\}_{j=1}^N$ are unknown, whereas the nodes $\{z_k\}_{k=1}^M$ are prescribed. We have

- $M = 0$ for the Gauss rule,
- $M = 1$, $z_1 = a$ or $z_1 = b$ for the Gauss–Radau rule,
- $M = 2$, $z_1 = a$ and $z_2 = b$ for the Gauss–Lobatto rule.

Also, for the case $\mathbf{u} = \mathbf{v}$, the remainder in (83) can be written as

$$R[f] = \frac{f^{(2N+M)}(\eta)}{(2N+M)!} \int_a^b \prod_{k=1}^{M}(\lambda - z_k)\Big[\prod_{j=1}^{N}(\lambda - t_j)\Big]^2 \mathrm{d}\mu(\lambda), \tag{84}$$

for some $a < \eta < b$. This expression shows that, if $f(x)$ is strictly completely monotonic on an interval containing the spectrum of A, then quadrature rules applied to (83) give bounds on $\mathbf{u}^T f(A)\mathbf{v}$. More precisely, the Gauss rule gives a lower bound, the Gauss–Lobatto rule gives an upper bound, whereas the Gauss–Radau rule can be used to obtain both a lower and an upper bound. In particular, they can be used to obtain bounds on $[f(A)]_{ij}$. The evaluation of these quadrature rules is reduced to the computation of orthogonal polynomials via three-term recurrence, or, equivalently, to the computation of entries and spectral information on a certain tridiagonal matrix via the Lanczos algorithm. We refer to [22, 95] for details. Here we limit ourselves to observe that the conditions under which one can expect rapid decay of the off-diagonal entries of $f(A)$ also guarantee fast convergence of the Lanczos process. In practice, this means that under such conditions a small number N of quadrature nodes (equivalently, Lanczos steps) are sufficient to obtain very good estimates of the entries of $f(A)$. In numerical experiments, this number is usually between 5 and 10, see [18].

Error bounds for Krylov subspace approximations. Another situation where the decay bounds for $f(A)$ have found application is in the derivation of error bounds for Krylov subspace approximations of $f(A)\mathbf{b}$, and in particular for the important case of the matrix exponential $f(A) = \mathrm{e}^{-tA}\mathbf{b}$ [203, 207]. Recall that Krylov subspace methods are examples of polynomial approximation methods, where $f(A)\mathbf{b}$ is approximated by $p(A)\mathbf{b}$ for some (low-degree) polynomial p. Since every matrix function $f(A)$ is a polynomial in A, this is appropriate. The kth Krylov subspace of $A \in \mathbb{C}^{n \times n}$ and a nonzero vector $\mathbf{b} \in \mathbb{C}^n$ is defined by

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

and it can be written as

$$\mathcal{K}_k(A, \mathbf{b}) = \{q(A)\mathbf{b} \,|\, q \text{ is a polynomial of degree } \le k - 1\}.$$

The successive Krylov subspaces form a nested sequence:

$$\mathcal{K}_1(A, \mathbf{b}) \subset \mathcal{K}_2(A, \mathbf{b}) \subset \cdots \subset \mathcal{K}_d(A, \mathbf{b}) = \cdots = \mathcal{K}_n(A, \mathbf{b}).$$

Here d is the degree of the minimum polynomial of A with respect to $\mathbf{b}$. This is just the monic polynomial p of least degree such that $p(A)\mathbf{b} = \mathbf{0}$.

The basic idea behind Krylov methods is to project the given problem onto the successive Krylov subspaces, solving the (low-dimensional) projected problems, and expand the solution back to n-dimensional space to yield the next approximation. An orthonormal basis for a Krylov subspace can be efficiently constructed using the *Arnoldi process*; in the Hermitian case, this reduces to the Lanczos process (see [96, 195, 177]). Both of these algorithms are efficient implementations of the classical Gram–Schmidt process. In Arnoldi's method, the projected matrix H_k has upper Hessenberg structure, which can be exploited in the computation. In the Hermitian case, H_k is tridiagonal.

Denoting by $Q_k = [\mathbf{q}_1, \dots, \mathbf{q}_k] \in \mathbb{C}^{n \times k}$, with $\mathbf{q}_1 = \mathbf{b}/\|\mathbf{b}\|_2$, the orthonormal basis for the kth Krylov subspace produced by the Arnoldi process, the kth approximation to the solution vector $f(A)\mathbf{b}$ is computed as

$$\mathbf{x}_k := \|\mathbf{v}\|_2 Q_k f(H_k)\mathbf{e}_1 = Q_k f(H_k)Q_k^* \mathbf{v}. \tag{85}$$

Typically, $k \ll n$ and computing $f(H_k)$ is inexpensive, and can be carried out in a number of ways. For instance, when $H_k = H_k^* = T_k$ (a tridiagonal matrix), it can be computed via explicit diagonalization of T_k. More generally, methods based on the Schur form of H_k can be used [96, Chapter 9.1.4].

The main remaining issue is to decide when to stop the iteration. Much effort has been devoted in recent years to obtained bounds for the error $\|\mathbf{x}_k - f(A)\mathbf{b}\|_2$. As it turns out, in the case of the matrix exponential $f(A) = \mathrm{e}^{-tA}$ the approximation error is mainly governed by the quantity

$$h(t) = \mathbf{e}_k^T \mathrm{e}^{-tH_k} \mathbf{e}_1, \tag{86}$$

i.e., by the last entry in the first column of e^{-tH_k}. Since H_k is upper Hessenberg (in particular, tridiagonal if $A = A^*$), the bottom entry in the first column of e^{-tH_k} should be expected to decay rapidly to zero as k increases. In [203] the authors show how the decay bounds in Theorem 13, combined with estimates of the field of values of A obtained from a clever use of the Bendixson–Hirsch Theorem, can lead to fairly tight bounds on $|h(t)|$, which in turn leads to an explanation of the superlinear convergence behavior of Krylov methods. These results can be seen as a generalization to the nonnormal case of the bounds obtained in [111] and [207] for the Hermitian case.

Exponentials of stochastic matrices. A real $n \times n$ matrix S is *row-stochastic* if it has nonnegative entries and its row sums are all equal to 1. It is *doubly stochastic* if its columns sums are also all equal to 1. Such matrices arise as transition matrices of discrete Markov chains, and play an important role in the analysis of large graphs and networks (see, e.g., [130] and [32]).

In the study of diffusion-type and other dynamical processes on graphs, it is often necessary to perform computations involving matrix exponentials of the form e^{tS}, where $t \in \mathbb{R}$; see, e.g., [90, page 357] and [91]. In some cases, one is interested in approximating selected columns of e^{tS}. Note that this is a special case of the problem of computing a matrix function times a given vector, where the given vector is now of the form $\mathbf{e}_i$. If the entries in the ith column of e^{tS} are strongly localized, it may be possible to compute reasonable approximations very cheaply. This is crucial given

the often huge size of graphs arising from real-world applications, for example in information retrieval.

For sparse matrices corresponding to graphs with maximum degree uniformly bounded in n, the decay theory for matrix functions guarantees superexponential decay in the entries of e^{tS}. This means that each column of e^{tS} only contains $O(1)$ nonnegligible entries (with a prefactor depending on the desired accuracy ε, of course). Localized computations such as those investigated in [91] may be able to achieve the desired linear, or even sublinear, scaling. Note, however, that the bounded maximum degree assumption is not always realistic. Whether strong localization can occur without this assumption is an open question. While it is easy to construct sparse graphs which violate the condition and lead to delocalized exponentials, the condition is only a sufficient one and it may well happen that e^{tS} remains localized even if the maximum degree grows as $n \to \infty$.

Localization in e^{tS} is also linked to localization in the PageRank vector $\mathbf{p}$, the unique stationary probability distribution vector (such that $\mathbf{p}^T = \mathbf{p}^T S$) associated with an irreducible row-stochastic matrix S [90, 130, 159]. When S is the "Google matrix" associated with the World Wide Web, however, the decay in the entries of $\mathbf{p}$ does not appear to be exponential, but rather to satisfy a power law of the form $p_k = O(k^{-\gamma})$ for $k \to \infty$, assuming the entries are sorted in nonincreasing order. The value of γ is estimated to be approximately 2.1; see [130, page 110]. This fact reflects the power law nature of the degree distribution of the Web. General conditions for strong localization in seeded PageRank vectors are discussed in [159].

A completely different type of stochastic process leading to exponentials of very large, structured matrices is the Markovian analysis of queuing networks, see [31, 32]. In [31] the authors study the exponential of huge block upper triangular, block Toeplitz matrices and show that this matrix function satisfies useful decay properties that can be exploited in the computations, leading to efficient algorithms.

Exponential integrators. Finally, we mention that the decay properties of the exponential of banded matrices have recently been used in [39] to develop and analyze a class of domain decomposition methods for the integration of time-dependent PDEs [39] and in the analysis of an infinite Arnoldi exponential integrator for systems of ODEs [128].

4.2 Linear scaling methods for electronic structure computations

In quantum chemistry and solid state physics, one is interested in determining the *electronic structure* of (possibly large) atomic and molecular systems [147]. The problem amounts to computing the ground state (smallest eigenvalue and corresponding eigenfunction) of the many-body quantum-mechanical Hamiltonian (Schrödinger operator), H. In variational terms, we want to minimize the Rayleigh quotient:

$$E_0 = \min_{\Psi \neq 0} \frac{\langle \mathsf{H}\Psi, \Psi \rangle}{\langle \Psi, \Psi \rangle} \quad \text{and} \quad \Psi_0 = \mathrm{argmin}_{\Psi \neq 0} \frac{\langle \mathsf{H}\Psi, \Psi \rangle}{\langle \Psi, \Psi \rangle} \tag{87}$$

where $\langle \cdot, \cdot \rangle$ denotes the L^2 inner product. In the Born–Oppenheimer approximation, the many-body Hamiltonian is given (in atomic units) by

$$\mathsf{H} = \sum_{i=1}^{n_e} \left(-\frac{1}{2}\Delta_i - \sum_{j=1}^{M} \frac{Z_j}{|\mathbf{x}_i - \mathbf{r}_j|} + \sum_{j \neq i}^{n_e} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|} \right)$$

where n_e = number of electrons and M = number of nuclei in the system. The electron positions are denoted by $\mathbf{x}_i$, those of the nuclei by $\mathbf{r}_j$; as usual, the charges are denoted by Z_j. The operator H acts on a suitable subspace of $H^1(\mathbb{R}^{3n_e})$ consisting of anti-symmetric functions (as a consequence of Pauli's Exclusion Principle for fermions). Here, the spin is neglected in order to simplify the presentation.

Unless n_e is very small, the "curse of dimensionality" makes this problem intractable; even storing the wave function Ψ becomes impossible already for moderately-sized systems [125]. In order to make the problem more tractable, various approximations have been devised, most notably:

− Wave function methods (e.g., Hartree–Fock);
− Density Functional Theory (e.g., Kohn-Sham);
− Hybrid methods (e.g., B3LYP).

In these approximations the original, linear eigenproblem $\mathsf{H}\Psi = E\Psi$ for the many-electrons Hamiltonian is replaced by a non-linear one-particle eigenproblem:

$$\mathcal{F}(\psi_i) = \lambda_i \psi_i, \quad \langle \psi_i, \psi_j \rangle = \delta_{ij}, \quad 1 \le i, j \le n_e, \tag{88}$$

where $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n_e}$ are the n_e smallest eigenvalues of (88). In the case of Density Functional Theory, (88) are known as the *Kohn–Sham equations*, and the nonlinear operator $\mathcal{F}$ in (88) has the form $\mathcal{F}(\psi_i) = \left(-\frac{1}{2}\Delta + V(\rho)\right)\psi_i$, where $\rho = \rho(\mathbf{x}) = \sum_{i=1}^{n_e} |\psi_i(\mathbf{x})|^2$ is a the *electronic density*, a function of only three variables that alone is sufficient to determine, in principle, all the properties of a system [112, 126]. The Kohn–Sham equations (88) are the Euler–Lagrange equations for the minimization of a functional $J = J[\rho]$ (the *density functional*) such that the ground state energy, E_0, is the minimum of the functional: $E_0 = \inf_\rho J$. While the exact form of this functional is not known explicitly, a number of increasingly accurate approximations have been developed since the original paper [126] appeared. The enormous success of Density Functional Theory (which led to the award of a share of the 1998 Nobel Prize for Chemistry to Kohn) is due to the fact that the high-dimensional, intractable minimization problem (87) with respect to $\Psi \in L^2(\mathbb{R}^{3n_e})$ is replaced by a minimization problem with respect to $\rho \in L^2(\mathbb{R}^3)$.

The nonlinear Kohn–Sham equations (88) can be solved by a "self-consistent field" (SCF) iteration, leading to a sequence of *linear* eigenproblems

$$\mathcal{F}^{(k)}\psi_i^{(k)} = \lambda_i^{(k)}\psi_i^{(k)}, \quad \langle \psi_i^{(k)}, \psi_j^{(k)} \rangle = \delta_{ij}, \quad k = 1, 2, \ldots \tag{89}$$

$(1 \le i, j \le n_e)$, where each $\mathcal{F}^{(k)} = -\frac{1}{2}\Delta + V^{(k)}$ is a one-electron Hamiltonian with potential

$$V^{(k)} = V^{(k)}(\rho^{(k-1)}), \quad \rho^{(k-1)} = \sum_{i=1}^{n_e} |\psi_i^{(k-1)}(\mathbf{x})|^2.$$

Solution of each of the (discretized) linear eigenproblems (89) leads to a typical $O(n_e^3)$ cost per SCF iteration. However, the actual eigenpairs $(\psi_i^{(k)}, \lambda_i^{(k)})$ are unnecessary; hence, diagonalization of the (discretized) one-particle Hamiltonians can be avoided. Indeed, all one really needs is the *density matrix*, i.e., the spectral projector P onto the invariant subspace

$$V_{occ} = \mathrm{span}\{\psi_1, \ldots, \psi_{n_e}\}$$

corresponding to the n_e lowest eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n_e}$ ("occupied states"). At the kth SCF cycle, an approximation $P^{(k)} \approx P$ to the orthogonal projector onto the occupied subspace V_{occ} needs to be computed. At convergence, all quantities of interest in electronic structure theory can be computed from P.

In practice, operators are replaced by matrices by Galerkin projection onto a finite-dimensional subspace spanned by a set of basis functions $\{\phi_i\}_{i=1}^n$, where n is a multiple of n_e. Typically, $n = n_b \cdot n_e$ where $n_b \geq 2$ is a moderate constant when linear combinations of Gaussian-type orbitals are used; often, $n_b \approx 10 - 25$ (see [134]). We assume that the basis functions are localized, so that the resulting discrete Hamiltonians (denoted by H) are, up to some small truncation tolerance, sparse: their pattern/bandwidth is determined by the form of the potential. Finite difference approximations can also be used, in which case the sparsity pattern is that of the discrete Laplacian, since the potential is represented by a diagonal matrix.

Non-orthogonal bases are easily accommodated into the theory but they may lead to algorithmic complications. They are often dealt with by a congruence transformation to an orthogonal basis, which can be accomplished via an inverse-Cholesky factorization; the transformed Hamiltonian is $\widehat{H} = Z^T H Z$ where $S^{-1} = ZZ^T$ is the inverse Cholesky factorization of the *overlap matrix*,

$$S = [S_{ij}], \quad S_{ij} = \int_\Omega \phi_i(\mathbf{r})\phi_j(\mathbf{r})\mathrm{d}\mathbf{r}\,.$$

The inverse factor Z can be efficiently approximated by the AINV algorithm [23, 50]. Often S is strongly localized and has condition number independent of n. As we have seen, under these conditions its inverse (and therefore Z) decays exponentially fast, with a rate independent of n; see Theorem 24. Hence, up to a small truncation tolerance Z is sparse and so is $\widehat{H}$, see [20, page 49]. In alternative, Z can be replaced by the inverse square root $S^{-1/2}$ of S; transformation from H to $S^{-1/2}HS^{-1/2}$ is known as *Löwdin orthogonalization*. Again, localization of $S^{-1/2}$ is guaranteed if S is banded, sparse, or localized and well-conditioned, so a sparse approximation to $S^{-1/2}$ is possible. It is important to stress that transformation of H into $\widehat{H}$ need not be carried out explicitly in most linear scaling algorithms. Rather, the transformed matrix is kept in factored form, similar to preconditioning. From here on, we assume that the transformation has already been performed and we denote the representation of the discrete Hamiltonian in the orthogonal basis by H instead of $\widehat{H}$.

Thus, the fundamental problem of (zero-temperature) electronic structure theory has been reduced to the approximation of the spectral projector P onto the subspace spanned by the n_e lowest eigenfunctions of H (occupied states):

$$P = \psi_1 \otimes \psi_1 + \cdots + \psi_{n_e} \otimes \psi_{n_e}\,, \tag{90}$$

where $H\psi_i = \lambda_i\,\psi_i,\ i = 1, \ldots, n_e$. Note that we can write $P = h(H)$, where f is the Heaviside (step) function

$$h(x) = \begin{cases} 1 & \text{if} \quad x < \mu \\ 0 & \text{if} \quad x > \mu \end{cases}$$

with $\lambda_{n_e} < \mu < \lambda_{n_e+1}$ (μ is the "Fermi level"). Alternatively, we can write $P = (I - \mathrm{sign}(H - \mu I))/2$, where sign denotes the sign function ($\mathrm{sign}(x) = 1$ if $x > 0$,

$\text{sign}(x) = -1$ if $x < 0$). We will come back to this representation at the end of this section.

As usual, we can assume that H has been scaled and shifted so that $\sigma(H) \subset [-1, 1]$. If the spectral gap $\gamma = \lambda_{n_e+1} - \lambda_{n_e}$ is not too small, h can be well approximated by a smooth function with rapid decreases from 1 to 0 within the gap $(\lambda_{n_e}, \lambda_{n_e+1})$. A common choice is to replace h by the Fermi–Dirac function

$$f(x) = \frac{1}{1 + e^{\beta(x-\mu)}}, \tag{91}$$

which tends to a step function as the parameter β increases.

Physicists have observed long ago that for "gapped systems" (like insulators and, under certain conditions, semiconductors) the entries of the density matrix P decay exponentially fast away from the main diagonal, reflecting the fact that interaction strengths decrease rapidly with the distance [9, 70, 107, 118, 123, 134, 148, 160, 168, 170]. We recall that, as already discussed, exponential decay in the ψ_i in (90) is sufficient for localization of the density matrix P, but not necessary; and indeed, situations can be found where P decays exponentially but the ψ_i do not, see [45] and the discussion in [20, section 4].

Localization of the density matrix is a manifestation of the "nearsightedness" of electronic matter discussed in section 1.1.[15] Localization is crucial as it provides the basis for so-called *linear scaling* (i.e., $O(n_e)$) methods for electronic structure calculations. These methods have been vigorously developed since the early 1990s, and they are currently able to handle very large systems, see, e.g., [9, 10, 37, 43, 50, 92, 124, 132, 135, 136, 139, 161, 175, 178, 206]. These methods including expansion of the Fermi–Dirac operator $f(H)$ in the Chebyshev basis, constrained optimization methods based on density matrix minimization (possibly with ℓ^1 regularization to enforce localized solutions), methods based on the sign matrix representation of P (such as "McWeeny purification"), multipole expansions, and many others. As we have seen, rapidly decaying matrices can be approximated by sparse matrices, uniformly in n. Hence, rigorously establishing the rate of decay in the density matrix provides a sound mathematical foundation for linear scaling methods in electronic structure calculations. A mathematical analysis of the asymptotic decay properties of spectral projectors associated with large, banded or sparse Hermitian matrices has been presented in [20]. The main result in [20] can be summarized in the following theorem.

Theorem 26. ([20]) *Let $n = n_b \cdot n_e$ where n_b is a fixed positive integer and the integers n_e form a monotonically increasing sequence. Let $\{H_n\}$ be a sequence of Hermitian $n \times n$ matrices with the following properties:*

1. Each H_n has bandwidth m independent of n;

---

[15] As we saw earlier (see (79)), rapid decay in P means that quantities like $\text{Tr}(PH)$, which in electronic structure theory has the interpretation of a single particle-energy [92, 161], are insensitive to small perturbations in the Hamiltonian in positions that correspond to small entries in P. Also, the fact that $\text{Tr}(P) = \text{rank}(P) = n_e \ll n$ implies that many of the diagonal entries of P will be tiny; hence, slight changes in the potential $V(\mathbf{x})$ at a point $\mathbf{x}$ are only felt locally, see Example 2.

2. *There exist two (fixed) intervals $I_1 = [a, b]$, $I_2 = [c, d] \subset \mathbb{R}$ with $\gamma = c - b >$
 0 such that for all $n = n_b \cdot n_e$, I_1 contains the smallest n_e eigenvalues of
 H_n (counted with their multiplicities) and I_2 contains the remaining $n - n_e$
 eigenvalues.*

Let P_n denote the $n \times n$ spectral projector onto the subspace spanned by the eigen-
vectors associated with the n_e smallest eigenvalues of H_n, for each n. Then there
exist constants $K > 0$, $\alpha > 0$ independent of n such that

$$|[P_n]_{ij}| \leq K \, e^{-\alpha|i-j|}, \quad \forall i \neq j. \tag{92}$$

Moreover, for any $\varepsilon > 0$ there is a matrix $\tilde{P}_n$ of bandwidth p independent of n such
that $\|P_n - \tilde{P}_n\| < \varepsilon$, for all n.

As usual, the bandedness assumption can be replaced with the assumption
that the Hamiltonians are sparse, with associated graphs that satisfy the bounded
maximum degree assumption. In this case the geodesic distance on the graphs
should be used to measure decay.

Different proofs of Theorem 26 can be found in [20]. One approach consists
in approximating P_n via the (analytic) Fermi–Dirac function (91), and exploiting
the fact that the rate of decay in $f(H_n)$ is independent of n thanks to the non-
vanishing gap assumption. This approach yields explicit, computable formulas for
the constants K and α in (92), see [20, pages 26–27]. Another approach is based on
results on the polynomial approximation of analytic functions on disjoint intervals
[59, 105]. Both proofs make crucial use of the general theory of exponential decay
in analytic functions of banded and sparse matrices discussed earlier, in particular
Theorem 9.

Some comments about the meaning of Theorem 26 are in order. The first thing
to observe is that the independence of the decay bounds on n follows from the
assumption that $n = n_b \cdot n_e$ where n_b, which controls the accuracy of the discretiza-
tion, is fixed, whereas n_e, which determines the system size (i.e., the number of
electrons in the system), is allowed to increase without bounds. This is sometimes
referred to as the *thermodynamic limit*, where the system size grows but the dis-
tance between atoms is kept constant. It should not be confused with the limit in
which n_e is fixed and $n_b \to \infty$, or with the case where both n_e and n_b are allowed
to grow without bounds. Keeping n_b constant ensures that the spectra of the dis-
crete Hamiltonians remain uniformly bounded, which guarantees (for insulators)
that the relative spectral gap γ does not vanish as $n \to \infty$. In practice a constant
n_b is a reasonable assumption since existing basis sets are not very large and they
are already highly optimized so as to achieve the desired accuracy.

Another issue that warrants comment is the choice of the parameter β (the
"inverse temperature") in the Fermi–Dirac function. Note that the Fermi–Dirac
function has two poles in the complex plane: if we assume, without loss of generality,
that the Fermi level is at $\mu = 0$, the poles are on the imaginary axis at $\pm i\pi/\beta$.
Since the rate of decay is governed by the distance between these two poles and
the smallest intervals I_1, I_2 containing all the spectra $\sigma(H_n)$, a large value of β
(corresponding to a small gap γ) means that the poles approach 0 and therefore b
and c. In this case, decay could be slow; indeed, the rate of decay α in the bounds
(92) tends to zero as $\gamma \to 0$ or, equivalently, as $\beta \to \infty$. On the other hand, a
relatively large value of γ means that β can be chosen moderate and this will imply

fast decay in the entries of the density matrix P_n, for all n. We refer to [20, pages 40–42] for details on the dependence of the decay rate in the density matrix as a function of the gap and of the temperature, in particular in the zero-temperature limit ($\beta \to \infty$), and to [191] for an example of how these results can be used in actual computations.

The case of metallic systems at zero temperature corresponds to $\gamma \to 0$. In this case the bound (92) becomes useless, since $\alpha \to 0$. The actual decay in the density matrix in this case can be as slow as $(1 + |i - j|)^{-1}$; see [20, section 10] for a detailed analysis of a model problem.

We conclude this section discussing alternative representations of the density matrix that could potentially lead to better decay bounds. Recall that the step function $h(x)$ can be expressed in terms of the sign function as $h(x) = (1 - \text{sign}(x))/2$. Hence, studying the decay behavior of the spectral projector $h(H)$ amounts to studying the decay behavior in $\text{sign}(H)$. Again, we assume for simplicity that the Fermi level is at $\mu = 0$. Now, the matrix sign function admits a well known integral representation, see [109]:

$$\text{sign}(H) = \frac{2}{\pi} H \int_0^\infty (t^2 I + H^2)^{-1} \mathrm{d}t. \tag{93}$$

One can now use available decay bounds on the inverse of the banded (or sparse) Hermitian positive definite matrix $t^2 I + H^2$ together with quadrature rules to obtain bounds on the entries of $\text{sign}(H)$ and therefore of P. Note the similarity of this approach with the one reviewed in section 3.5.

Another approach, which yields more explicit bounds, is based on the identity

$$\text{sign}(H) = H(H^2)^{-1/2}, \tag{94}$$

see again [109]. We can then use bounds for the inverse square root of the banded (or sparse) Hermitian positive definite matrix H^2 to obtain exponential decay bounds for $\text{sign}(H)$, using the fact that the product of a banded (or sparse) matrix times an exponentially decaying one retains the exponential decay property. Preliminary numerical experiments on simple model gapped Hamiltonians [38] suggest that the decay bounds obtained via the representations (93) and (94) can be more accurate than those obtained via the Fermi–Dirac representation.

4.3 Further applications

In this section we briefly mention a few other applications of localization and decay bounds in applied mathematics and physics.

Localized solutions to matrix equations. In the area of control of discrete-time, large-scale dynamical systems, a central role is played by the *Lyapunov equation* associated to a linear, time-varying dynamical system:

$$AX + XA^T = P, \tag{95}$$

with $A, P \in \mathbb{R}^{n \times n}$ given matrices and X unknown. If A is stable, equation (95) has a unique solution, which can be expressed as

$$X = -\int_0^\infty \mathrm{e}^{tA} P \mathrm{e}^{tA} \, \mathrm{d}t \, ; \tag{96}$$

and also as the solution of a linear system in Kronecker sum form:

$$(I \otimes A + A^T \otimes I)\mathrm{vec}(X) = \mathrm{vec}(P) \, , \tag{97}$$

where "vec" is the operator that takes an $n \times n$ matrix and forms the vector of length n^2 consisting of the columns of the matrix stacked one on top of one another; see, e. g. [131] or [184].

Several authors (e.g., [52, 104]) have observed that whenever A and P are banded, or sparse, the solution matrix X is localized, with rapid off-diagonal decay if A is well-conditioned. Moreover, the decay is oscillatory (see [104, Fig. 3]). This does not come as a surprise given the relation of X to the matrix exponential of A, see (96), and to the inverse of a matrix in Kronecker sum form, see (97). The decay in the solution has been exploited to develop efficient solution techniques for (95) with A and P banded. When $A = A^T$, an approximate solution to (95) can sometimes be obtained using polynomial expansion in the Chebyshev basis, as outlined in the previous section. We refer again to [104] for details.

Localized matrices (in the form of rapidly decaying inverse Gramians) also arise in another problem in control theory, namely, subspace identification of large-scale interconnected systems. Again, this fact can be exploited to develop fast approximation algorithms; see [103].

For a different application of exponential decay bounds for A^{-1} to the study of the behavior of dynamical systems, see [155].

Localization in graph and network analysis. Recently, several authors have proved that, with high probability, the eigenvectors of the adjacency matrix [66, 74, 193] and of the Laplacian [44] of large sparse undirected random graphs, in particular Erdős–Rényi graphs, are delocalized, a fact that was already known on the basis of empirical observation; see, e.g., [94]. On the other hand, localization in eigenvectors of scale-free networks, particularly those corresponding to the largest eigenvalues of the adjacency matrix or of the Laplacian, has been reported, for both synthetic [94, 151] and real-world [108, 158, 181] networks. For instance, in [94] the eigenvector corresponding to the largest eigenvalue of the adjacency matrix of a power-law graph was found to be localized at the *hub* (the node of maximum degree).

Power-law graphs are also studied in [108], where a class of graph substructures leading to locally supported eigenvectors is identified. In some cases these eigenvectors are actually *sparse*, not just localized: an extreme example is given by the so-called *Faria vectors* [83], i.e., eigenvectors that have only two nonzero components (one positive, the other necessarily negative). These eigenvectors are associated with eigenvalues of very high multiplicity (sometimes as large as $O(n)$), such as those associated with star graphs—graphs consisting of a central node connected to several peripheral nodes. Many star-like subgraphs, and thus Laplacian eigenvalues of very high multiplicity, are often found in real-world scale-free graphs, an observation that leads to the conjecture that such Laplacians may have a number of locally supported eigenvectors. We refer to [108] for a detailed study, including computational aspects. It should be noted that since power-law graphs do not satisfy the bounded maximum degree assumption, we cannot directly apply the decay

theory for matrix functions (and in particular for spectral projectors) to explain the eigenvector localization discussed in [94, 108].

Another class of graphs for which eigenvector localization has been observed is discussed in [158], motivated by the study of dendrites of retinal ganglion cells (RGCs). It turns out the Laplacian eigenvalues of a typical dendritic tree display a peculiar distribution: most of the λ_i are distributed according to a smooth curve in the interval $(0, 4)$, after which a jump occurs in the spectrum and the remaining eigenvalues are clustered around some value larger than 4. Moreover, the eigenvectors associated with eigenvalues less than 4 are delocalized, while those associated with eigenvalues greater than 4 are exponentially localized.

In order to find an explanation to this phenomenon, the authors of [158] consider simplified models of RGCs that are easier to analyze but at the same time capture the main properties of RGCs. The simplest among these models is a *star-like tree*, obtained by connecting one or more path graphs to the central hub of a star graph S_k (consisting of k peripheral nodes connected to a central node), where $k \geq 3$. In the case of a single path graph P_ℓ connected to the central hub of a star graph S_k, the resulting graph is called a *comet* of type (k, ℓ). The Laplacian of the resulting star-like tree is then obtained by *gluing* the Laplacians of S_k and P_ℓ For example, in the case $k = 4$, $\ell = 5$ the corresponding comet graph (with the hub numbered first) has 10 nodes and the associated Laplacian is given by

$$
L = \left[
\begin{array}{ccccc|ccccc}
5 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\hline
-1 & 0 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
\end{array}
\right],
$$

where the horizontal and vertical line have been added to more clearly show how the gluing of the Laplacians of S_4 and P_5 is carried out. The resulting L has nine eigenvalues < 4, all falling between 0 and 3.652007105, with the tenth eigenvalue $\lambda \approx 6.0550$. It is known that a star-like tree can have only one eigenvalue ≥ 4, with equality if and only if the graph is a *claw* (see [158] for details). In our example the (normalized) eigenvector associated with the largest eigenvalue is of the form

$$
\mathbf{v} = \left[\mathbf{v}_1, \, \mathbf{v}_2 \right],
$$

where

$$
\mathbf{v}_1 = \left[0.9012, \, -0.1783, \, -0.1783, \, -0.1783, \, -0.1783 \right],
$$

and

$$
\mathbf{v}_2 = \left[0.2377, \, -0.0627, \, 0.0165, \, -0.0043, \, 0.0008 \right]
$$

(rounding to four decimal places). We see that the dominant eigenvector is mostly concentrated at the hub, is constant on the peripheral nodes of the star S_4, and decays monotonically in magnitude along the path P_5 away from the hub. It can be shown that this phenomenon is common to all comet graphs and that the decay

of the dominant eigenvector along the path is exponential [158]. Moreover, similar behavior is also common to star-like trees in general; if there are multiple paths connected to the central node of a star, the dominant eigenvector will decay exponentially along the paths, away from the central hub, where the eigenvector is concentrated. The remaining eigenvectors, corresponding to eigenvalues less than 4, are delocalized (oscillatory).

The proofs in [158] are direct and are based on previous results on the eigenvalues of star-like trees, together with a careful analysis of certain recurrences that the entries of the dominant eigenvector of L must satisfy. Here we point out that the decay behavior of the dominant eigenvector of L for star-like trees (and also for more general, less structured graphs obtained by gluing one or more long paths to a graph having a hub of sufficiently high degree) is a byproduct of our general decay results for functions of banded or sparse matrices. To see this, we consider the case of a single path of length ℓ attached to the hub of a graph with order $k+1$ nodes. The resulting graph has $n = k+\ell+1$ nodes, and its Laplacian is of the form

$$L_n = \begin{bmatrix} L_{11} & L_{12} \\ L_{12}^T & L_{22} \end{bmatrix},$$

where L_{12} is a $(k+1) \times \ell$ matrix with all its entries equal to 0 except for a -1 in the upper left corner. For fixed k and increasing ℓ, the sequence $\{L_n\}$ satisfies the assumptions of Theorem 9, therefore for any analytic function f, the entries of $f(L_n)$ must decay at least exponentially fast away from the main diagonal (or nonzero pattern of L_n) with rate independent of n for ℓ sufficiently large. Moreover, assume that the dominant eigenvalue of L_n is well separated from the rest of the spectrum; this happens for example if one of the nodes, say the first, has a significantly higher degree than the remaining ones. Then the corresponding eigenvector is localized, and its entries decay exponentially along the path, away from the hub. To see this we can approximate the corresponding spectral projector, P, by a Gaussian in L_n and use the fact that $\mathrm{Tr}(P) = 1$, as discussed earlier.

Next, we consider the issue of localization in functions of matrices associated with graphs, limited to the undirected case. We are especially interested in the matrix exponential, which is widely used in the study of network structure and dynamics. We discuss first the communicability between nodes, see (7), which is measured by the entries of the exponential of the adjacency matrix. In many applications, see for example [81], it is desirable to identify pairs of nodes within a network having low communicability. For fairly regular sparse graphs with large diameter, communication between neighboring nodes is clearly much easier than communication between pairs of distant nodes, and this fact is well captured by the fact that $[\mathrm{e}^A]_{ij}$ decays superexponentially with the geodesic distance $d(i,j)$. The simplest example is that of a path graph P_ℓ, for which A is tridiagonal: as we have shown, in this case the entries of the exponential decay very fast with the distance $|i - j|$. On the other hand, for large connected graphs of small diameter, like many real world complex networks, the distance $d(i,j)$ is small for any i and j, and decay bounds like (29) cannot predict any small entries in functions of the adjacency matrix. This fact is related to the violation of the bounded maximum degree assumption. The simplest example is now that of a star graph S_n, which has diameter 2 and maximum degree n, and for which there is no decay whatsoever in e^A.

Analogous remarks apply to functions of the Laplacian of a connected graph. Consider for instance the heat kernel, e^{-tL}. As is well known, this matrix function occurs in the solution of initial value problems of the form

$$\dot{\mathbf{x}} = -L\mathbf{x} \quad \text{subject to} \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{98}$$

where the dot on a vector denotes differentiation with respect to time and $\mathbf{x}_0 \in \mathbb{R}^n$ is given. The solution of (98) is given, for all $t \geq 0$, by

$$\mathbf{x}(t) = \mathrm{e}^{-tL}\mathbf{x}_0.$$

Consider now the special case where $\mathbf{x}_0 = \mathbf{e}_i$, the ith vector of the standard basis. This means that a unit amount of some "substance" is placed at node i at time $t = 0$, the amount at all other nodes being zero. Alternatively, we could think of node i being at 1° temperature, with all other nodes having 0° temperature initially. Then the jth entry of the solution vector $\mathbf{x}(t)$ represents the fraction of the substance that has diffused to node j at time t or, alternatively, the temperature reached by node j at time t. This quantity is given by

$$x_j(t) = [\mathrm{e}^{-tL}]_{ij}\,.$$

Note that

$$\lim_{t \to \infty} x_j(t) = \frac{1}{n} \quad \forall\, j = 1, \ldots, n,$$

and moreover this limit is independent of i. This fact simply means that asymptotically, the system is at thermal equilibrium, with all the initial "substance" (e.g., heat) being equally distributed among the nodes of the network, regardless of the initial condition.

Another interpretation is possible: observing that e^{-tL} is a (doubly stochastic) matrix for all $t \geq 0$, we can interpret its entries $[\mathrm{e}^{-tL}]_{ij}$ as transition probabilities for a Markov chain, namely, for a (continuous-time) random walk on the graph. Then $[\mathrm{e}^{-tL}]_{ij}$ has the meaning of the probability of a "walker" being at node j at time t given that it was at node i at time $t = 0$.

No matter what the interpretation is, we see that the entries of e^{-tL} can serve as a measure of *communicability over time* between pairs of nodes in a network. We note that for $t = 1$ and regular graphs, this measure is identical (up to a constant factor) to the earlier notion of communicability (7). Note that for fairly regular, large diameter, sparse, "grid-like" graphs and for fixed $t > 0$ the entries $[\mathrm{e}^{-tL}]_{ij}$ decay superexponentially fast as the geodesic distance $d(i, j)$ increases, reflecting the fact that after a finite time only a relatively small fraction of the diffusing substance will have reached the furthest nodes in the network. Clearly, this amount increases with time. The rate of convergence to equilibrium is governed by the spectral gap (i.e., the smallest nonzero eigenvalue of L): the smaller it is, the longer it takes for the system to reach equilibrium. Since for grid-like graphs the smallest eigenvalue goes to zero rapidly as $n \to \infty$ ($\lambda_2(L) = O(n^{-2})$), convergence to equilibrium is slow. On the other hand, graphs with good expansion properties tend to have large spectral gap and equilibrium is reached much faster, even for large n. This is reflected in the fact that e^{-tL} is usually delocalized for such graphs. Note, however, that things are more complicated in the case of *weighted graphs*, or if a *normalized Laplacian*

$$\widehat{L} = I - D^{-1/2}AD^{-1/2} \quad \text{or} \quad \widetilde{L} = I - D^{-1}A$$

is used instad of L (see for instance [91]).

Finally, we consider a "quantum" form of network communicability, obtained by replacing the diffusion-type equation (98) with the Schrödinger-type equation:

$$i\,\dot{\mathbf{x}} = L\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{99}$$

where $i = \sqrt{-1}$, $\mathbf{x}_0 \in \mathbb{C}^n$ is given and such that $\|\mathbf{x}_0\|_2 = 1$. As we have seen, the solution of (99) is given by

$$\mathbf{x}(t) = \mathrm{e}^{-itL}\mathbf{x}_0, \quad \forall\, t \in \mathbb{R}.$$

Note that since $U(t) = \mathrm{e}^{-itL}$ is unitary, the solution satisfies $\|\mathbf{x}(t)\|_2 = 1$ for all $t \in \mathbb{R}$. Consider now the matrix family $\{S(t)\}_{t \in \mathbb{R}}$ defined as follows:

$$S(t) = [S(t)]_{ij}, \quad [S(t)]_{ij} = |[U(t)]_{ij}|^2 \quad \forall\, i,j. \tag{100}$$

Since $U(t)$ is unitary, $S(t)$ is doubly stochastic. Its entries are transition probabilities: they measure how likely a system (say, a particle) governed by equation (99), initially in state i, is to be in state j at time t. Here we have identified the nodes of the underlying graph with the states of the system. We see again that the magnitudes (squared) of the entries of e^{-itL} measure the communicability (which we could call "quantum communicability") between pairs of nodes, or states. Localization in e^{-itL} means low quantum communicability between far away pairs of nodes.

Log-determinant evaluation. In statistics, it is frequently necessary to evaluate the expression

$$\log(\det(A)) = \mathrm{Tr}\,(\log(A)), \tag{101}$$

where A is symmetric positive definite. When A is not too large, one can directly compute the determinant on the left-hand side of (101) via a Cholesky factorization of A. If A is too large to be factored, various alternative methods have been proposed, the most popular of which are randomized methods for trace estimation such as the one proposed by Hutchinson [116]:

$$\mathrm{Tr}\,(\log(A)) \approx \frac{1}{s}\sum_{j=1}^{s} \mathbf{v}_j^T \log(A)\mathbf{v}_j,$$

with $\mathbf{v}_1, \ldots, \mathbf{v}_s$ suitably chosen "sampling" vectors. This method requires the rapid evaluation of the matrix-vector products $\log(A)\mathbf{v}_j$ for many different vectors $\mathbf{v}_j$.

A number of methods for approximating $\log(A)\mathbf{v}_j$ are studied in [6]. Some of the techniques in [6] rely on the off-diagonal decay property in $\log(A)$ when A is sparse. The authors of [6] make the observation that the decay in $\log(A)$ will be different in different bases; using the fact that for any nonsingular matrix W the identities

$$\det(f(WAW^{-1})) = \det(Wf(A)W^{-1}) = \det(f(A))$$

hold, it may be possible to find a basis W in which $f(WAW^{-1})$ is highly localized, so that performing the computations in this basis might result in significant speed-ups. In particular, the use of an orthonormal ($W^{-1} = W^T$) wavelet basis is shown to result in considerable off-diagonal compression, as long as the entries in A vary smoothly (as is often the case). We refer to [6] for details. We note that a similar approach, with $f(A) = A^{-1}$, has been used (e.g., in [51]) to construct sparse approximate inverses for use as preconditioners.

Quantum information theory. Another area of research where decay bounds for matrix functions have proven useful is the study of many-body systems in quantum information theory; see, e.g., [56, 57, 77, 179]. For instance, relationships between spectral gaps and rates of decay for functions of local Hamiltonian operators have been derived in [56] based on Bernstein's Theorem, following [22].

As shown in [57], exponential decay bounds for matrix functions can be used to establish so-called *area laws* for the *entanglement entropy* of ground states associated with bosonic systems. In a nutshell, these area laws imply that the entanglement entropy associated with a 3D bosonic lattice is proportional to the surface area, rather than to the volume, of the lattice. It is noteworthy that such area laws are analogous to those governing the Beckenstein–Hawking black hole entropy. We refer the interested reader to the comprehensive survey paper [77], where implications for computer simulations of quantum states are also discussed.

5 Conclusions and future work

The traditional dichotomy between *sparse* and *dense* matrix computations is too restrictive and needs to be revised to allow for additional modes of computation in which other, less-obvious forms of (approximate) sparsity are present, either in the problem data or in the solution, or both.

In recent years there has been strong interest and many important developments in research areas like hierarchical matrices and data-sparse algorithms (discussed by Ballani and Kressner and by Bini in this same volume) and compressed sensing; a different direction, the exploitation of localization, or decay, has been the subject of this chapter. Localization has long played an important role (both conceptual and computational) in various areas of physics, but until recently it has received less attention from researchers in the field of numerical linear algebra. Here we have reviewed various notions of localization arising in different fields of mathematics and some of its applications in physics. We have attempted to provide a unified view of localization in numerical linear algebra using various types of decay bounds for the entries of matrix functions. Other useful tools include the use of decay algebras and C*-algebras, and integral representations of matrix functions. We have further indicated how exploitation of localization is being used for developing fast approximate solution algorithms, in some cases having linear complexity in the size of the problem.

There are numerous opportunities for further research in this area. At several points in the chapter we have pointed out a few open problems and challenges, which can be summarized briefly as follows.

1. Concerning functions of matrices, including the important special cases of inverses and spectral projectors, we have discussed several conditions for localization. We have seen that these conditions are sufficient, but not necessary in general. Finding necessary conditions would deepen our understanding of localization considerably. Deriving some lower bounds on the entries of $f(A)$ would be useful in this regard.
2. Many of the decay bounds we have seen are rather pessimistic in practice. Similar to the convergence theory for Krylov subspace methods, it should be

possible to obtain improved bounds by making use of more detailed spectral information on the matrix, at least in the Hermitian case.

3. As usual, the case of nonnormal matrices presents challenges and difficulties not present in the normal case. It would be useful to have a better understanding of decay properties in functions of highly nonnormal matrices, for example in oblique spectral projectors. This may have interesting applications in fields like non-Hermitian quantum mechanics [15, 16, 154, 196].

4. It is easy to see with examples that violating the bounded maximum degree assumption leads to failure of exponential decay in the limit $n \to \infty$; in practice, however, sufficiently rapid decay may persist for finite n to be useful in computation if the maximum degree increases slowly enough. This aspect seems to warrant further investigation, especially in view of applications in network analysis.

5. It would be interesting to develop general conditions under which bounded functions of unbounded, banded operators (or sequences of banded finite matrices without uniformly bounded spectra) exhibit decay behavior.

6. Outside of the broad area of linear scaling methods for electronic structure computations and in the solution of certain types of structured problems (e.g., [31, 188]), relatively little has been done so far to exploit advance knowledge of localization in designing efficient algorithms. It would be especially useful to develop approximation algorithms that can exploit localization in the solution of large linear systems and in the eigenvectors (or invariant subspaces) of large matrices, when present. The ideas and techniques set forth in [137] for Hermitian matrices are a good starting point.

7. Last but not least, error control techniques in algorithms based on neglecting small matrix or vector entries deserve careful study.

We hope hope that this chapter will stimulate progress on these and other problems related to localization.

Acknowledgments

References

1. M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, NY, 1965.

2. S. Agmon, *Lectures on Exponential Decay of Solutions of Second-Order Elliptic Equations: Bounds on Eigenfunctions of N-Body Schrödinger Operators*, Mathematical Notes, 29. Princeton University Press, Princeton, NJ, and University of Tokyo Press, Tokyo, 1982.

3. G. Alléon, M. Benzi, and L. Giraud, *Sparse approximate inverse preconditioning for dense linear systems arising in computational electromagnetics*, Numer. Algorithms, 16 (1997), pp. 1–15.

4. P. W. Anderson, *Absence of diffusion in certain random lattices*, Phys. Rev., 109 (1958), pp. 1492–1505.

5. M. Arioli and M. Benzi, *A finite element method for quantum graphs*, Math/CS Technical Report TR-2015-009, Emory University, October 2015.

6. E. Aune, D. P. Simpson, J. Eidsvik, *Parameter estimation in high dimensional Gaussian distributions*, Stat. Comput., 24 (2014), pp. 247–263.

7. O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, UK, 1994.

8. O. Axelsson and Polman, *On approximate factorization methods for block matrices suitable for vector and parallel processors*, Linear Algebra Appl., 77 (1986), pp. 3–26.

9. R. Baer and M. Head-Gordon, *Sparsity of the density matrix in Kohn–Sham density functional theory and an assessment of linear system-size scaling methods*, Phys. Rev. Lett., 79 (1997), pp. 3962–3965.

10. R. Baer and M. Head-Gordon, *Chebyshev expansion methods for electronic structure calculations on large molecular systems*, J. Chem. Phys., 107 (1997), pp. 10003–10013.

11. H. Bağci, J. E. Pasciak, and K. Y. Sirenko, *A convergence analysis for a sweeping preconditioner for block tridiagonal systems of linear equations*, Numer. Linear Algebra Appl., 22 (2015), pp. 371–392.

12. A. G. Baskakov, *Wiener's theorem and the asymptotic estimates of the elements of inverse matrices*, Funct. Anal. Appl., 24 (1990), pp. 222–224.

13. A. G. Baskakov, *Estimates for the entries of inverse matrices and the spectral analysis of linear operators*, Izvestiya: Mathematics, 61 (1997), pp. 1113–1135.

14. R. Bellman, *Introduction to Matrix Analysis. Second Edition*, McGraw–Hill, New York, NY, 1970.

15. C. M. Bender, S. Boettcher and P. N. Meisinger, *PT-symmetric quantum mechanics*, J. Math. Phys., 40 (1999), pp. 2201–2229.

16. C. M. Bender, D. C. Brody and H. F. Jones, *Must a Hamiltonian be Hermitian?*, Amer. J. Phys., 71 (2003), pp. 1095–1102.

17. M. Benzi, *Preconditioning techniques for large linear systems: a survey*, J. Comp. Phys., 182 (2002), pp. 418–477.

18. M. Benzi and P. Boito, *Quadrature rule-based bounds for functions of adjacency matrices*, Linear Algebra Appl., 433 (2010), pp. 637–652.

19. M. Benzi and P. Boito, *Decay properties for functions of matrices over C^*-algebras*, Linear Algebra Appl., 456 (2014), pp. 174–198.

20. M. Benzi, P. Boito and N. Razouk, *Decay properties of spectral projectors with applications to electronic structure*, SIAM Rev., 55 (2013), pp. 3–64.

21. M. Benzi, T. Evans, S. Hamilton, M. Lupo Pasini, and S. Slattery, *Analysis of Monte Carlo accelerated iterative methods for sparse linear systems*, Math/CS Technical Report TR-2016-002, Emory University, 2016.

22. M. Benzi and G. H. Golub, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT Numer. Math., 39 (1999), pp. 417–438.

23. M. Benzi, C. D. Meyer, and M. Tůma, *A sparse approximate inverse preconditioner for the conjugate gradient method*, SIAM J. Sci. Comput., 17 (1996), pp. 1135–1149.

24. M. Benzi and N. Razouk, *Decay bounds and $O(n)$ algorithms for approximating functions of sparse matrices*, Electr. Trans. Numer. Anal., 28 (2007), pp. 16–39.

25. M. Benzi and V. Simoncini, *Decay bounds for functions of Hermitian matrices with banded or Kronecker structure*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1263–1282.

26. M. Benzi and M. Tůma, *A sparse approximate inverse preconditioner for nonsymmetric linear systems*, SIAM J. Sci. Comput., 19 (1998), pp. 968–994.

27. M. Benzi and M. Tůma, *Orderings for factorized approximate inverse preconditioners*, SIAM J. Sci. Comput., 21 (2000), pp. 1851–1868.

28. S. K. Berberian and G. H. Orland, *On the closure of the numerical range of an operator*, Proc. Amer. Math. Soc., 18 (1967), pp. 499–503.

29. L. Bergamaschi, M. Caliari and M. Vianello, *Efficient approximation of the exponential operator for discrete 2D advection-diffusion problems*, Numer. Linear Algebra Appl., 10 (2003), pp. 271–289.

30. L. Bergamaschi and M. Vianello, *Efficient computation of the exponential operator for large, sparse, symmetric matrices*, Numer. Linear Algebra Appl., 7 (2000), pp. 27–45.

31. D. A. Bini, S. Dendievel, G. Latouche, and B. Meini, *Computing the exponential of large block-triangular block-Toeplitz matrices encountered in fluid queues*, Linear Algebra Appl. 502 (2016), pp. 387–419.

32. D. A. Bini, G. Latouche, and B. Meini, *Numerical Methods for Structured Markov Chains*, Oxford University Press, Oxford, UK, 2005.

33. I. A. Blatov, *Incomplete factorization methods for systems with sparse matrices*, Comput. Math. Math. Phys., 33 (1993), pp. 727–741.

34. I. A. Blatov, *On algebras and applications of operators with pseudosparse matrices*, Siber. Math. J., 37 (1996), pp. 32–52.

35. I. A. Blatov and A. A. Terteryan, *Estimates of the elements of the inverse matrices and pivotal condensation methods of incomplete block factorization*, Comput. Math. Math. Phys., 32 (1992), pp. 1509–1522.

36. N. Bock and M. Challacombe, *An optimized sparse approximate matrix multiply for matrices with decay*, SIAM J. Sci. Comput., 35 (2013), pp. C72–C98.

37. N. Bock, M. Challacombe, and L. V. Kalé, *Solvers for $\mathcal{O}(N)$ electronic structure in the strong scaling limit*, SIAM J. Sci. Comput., 38 (2016), pp. C1–C21.

38. P. Boito, *Private communication*, 2015.

39. L. Bonaventura, *Local exponential methods: a domain decomposition approach to exponential time integration of PDEs*, arXiv:1505.02248v1, May 2015.

40. F. Bonchi, P. Esfandiar, D. F. Gleich, C. Greif, and L. V. S. Lakshmanan, *Fast matrix computations for pair-wise and column-wise commute times and Katz scores*, Internet Math., 8 (2012), pp. 73–112.

41. A. Böttcher and S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2005.

42. A. Böttcher and B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*, Springer, New York, NY, 1998.

43. D. R. Bowler and T. Miyazaki, *$O(N)$ methods in electronic structure calculations*, Rep. Progr. Phys., 75 (2012), 036503.

44. S. Brooks, E. Lindenstrauss, *Non-localization of eigenfunctions on large regular graphs*, Israel J. Math., 193 (2013), pp. 1–14.
45. C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, *Exponential localization of Wannier functions in insulators*, Phys. Rev. Lett., 98 (2007), 046402.
46. S. Bruciapaglia, S. Micheletti, and S. Perotto, *Compressed solving: A numerical approximation technique for elliptic PDEs based on Compressed Sensing*, Comput. Math. Appl., 70 (2015), pp. 1306–1335.
47. K. Bryan and T. Lee, *Making do with less: an introduction to compressed sensing*, SIAM Rev., 55 (2013), pp. 547–566.
48. C. Canuto, V. Simoncini, and M. Verani, *On the decay of the inverse of matrices that are sum of Kronecker products*, Linear Algebra Appl., 452 (2014), pp. 21–39.
49. C. Canuto, V. Simoncini, and M. Verani, *Contraction and optimality properties of an adaptive Legendre–Galerkin method: the multi-dimensional case*, J. Sci. Comput., 63 (2015), pp. 769–798.
50. M. Challacombe, *A simplified density matrix minimization for linear scaling self-consistent field theory*, J. Chem. Phys., 110 (1999), pp. 2332–2342.
51. T. Chan, W.-P. Tang, and J. Wan, *Wavelet sparse approximate inverse preconditioners*, BIT Numer. Math., 37 (1997), pp. 644–660.
52. J. Chandrasekar and D. S. Bernstein, *Correlation bounds for discrete-time systems with banded dynamics*, Systems & Control Letters, 56 (2007), pp. 83–86.
53. E. Chow, *A priori sparsity patterns for parallel sparse approximate inverse preconditioners*, SIAM J. Sci. Comput., 21 (2000), pp. 1804–1822.
54. J.-M. Combes and L. Thomas, *Asymptotic behaviour of eigenfunctions for multiparticle Schrödinger operators*, Comm. Math. Phys., 34 (1973), pp. 251–270.
55. P. Concus, G. H. Golub, and G. Meurant, *Block preconditioning for the conjugate gradient method*, SIAM J. Sci. Stat. Comput., 6 (1985), pp. 220–252.
56. M. Cramer and J. Eisert, *Correlations, spectral gap and entanglement in harmonic quantum systems on generic lattices*, New J. Phys., 8 (2006), 71.
57. M. Cramer, J. Eisert, M. B. Plenio and J. Dreissig, *Entanglement-area law for general bosonic harmonic lattice systems*, Phys. Rev. A, 73 (2006), 012309.
58. M. Crouzeix, *Numerical range and functional calculus in Hilbert space*, J. Funct. Anal., 244 (2007), pp. 668–690.
59. C. K. Chui and M. Hasson, *Degree of uniform approximation on disjoint intervals*, Pacif. J. Math., 105 (1983), pp. 291–297.
60. J. J. M. Cuppen, *A divide and conquer method for the symmetric tridiagonal eigenproblem*, Numer. Math., 36 (1981), pp. 177–195.
61. S. Dahlke, M. Fornasier, and K. Gröchenig, *Optimal adaptive computations in the Jaffard algebra and localized frames*, J. Approx. Theory, 162 (2010), pp. 153–185.
62. A. Damle, L. Lin, and L. Ying, *Compressed representations of Kohn–Sham orbitals via selected columns of the density matrix*, J. Chem. Theory Comput., 11 (2015), pp. 1463–1469.
63. A. Damle, L. Lin, and L. Ying, *Accelerating selected columns of the density matrix computations via approximate column selection*, arXiv:1604.06830v1, April 2016.

64. P. J. DAVIS, *Circulant Matrices*, John Wiley and Sons, New York, NY, 1979.

65. T. A. DAVIS AND Y. HU, *The University of Florida Sparse Matrix Collection*, ACM Trans. Math. Software, 38 (2011), pp. 1–25.

66. Y. DEKEL, J. R. LEE, AND N. LINIAL, *Eigenvectors of random graphs: nodal domains*, Random Structures Algorithms, 39 (2011), pp. 39–58.

67. N. DEL BUONO, L. LOPEZ AND R. PELUSO, *Computation of the exponential of large sparse skew-symmetric matrices*, SIAM J. Sci. Comput., 27 (2005), pp. 278–293.

68. S. DEMKO, *Inverses of band matrices and local convergence of spline projections*, SIAM J. Numer. Anal., 14 (1977), pp. 616–619.

69. S. DEMKO, W. F. MOSS, AND P. W. SMITH, *Decay rates for inverses of band matrices*, Math. Comp., 43 (1984), pp. 491–499.

70. J. DES CLOIZEAUX, *Energy bands and projection operators in a crystal: analytic and asymptotic properties*, Phys. Rev., 135 (1964), pp. A685–A697.

71. I. S. DHILLON, B. S. PARLETT, AND C. VÖMEL, *The design and implementation of the MRRR algorithm*, ACM Trans. Math. Software, 32 (2006), pp. 533–560.

72. R. DIESTEL, *Graph Theory*, Springer-Verlag, Berlin, 2000.

73. I. S. DUFF, A. M. ERISMAN, AND J. K. REID, *Direct Methods for Sparse Matrices*, Oxford University Press, Oxford, UK, 1986.

74. I. DUMITRIU AND S. PAL, *Sparse regular random graphs: spectral density and eigenvectors*, Ann. Prob., 40 (2012), pp. 2197–2235.

75. W. E AND J. LU, *The electronic structure of smoothly deformed crystals: Wannier functions and the Cauchy–Born rule*, Arch. Ration. Mech. Anal., 199 (2011), pp. 407–433.

76. V. EIJKHOUT AND B. POLMAN, *Decay rates of inverses of banded M-matrices that are near to Toeplitz matrices*, Linear Algebra Appl., 109 (1988), pp. 247–277.

77. J. EISERT, M. CRAMER AND M. B. PLENIO, *Colloquium: Area laws for the entanglement entropy*, Rev. Modern Phys., 82 (2010), pp. 277–306.

78. S. W. ELLACOTT, *Computation of Faber series with application to numerical polynomial approximation in the complex plane*, Math. Comp., 40 (1983), pp. 575–587.

79. E. ESTRADA, *The Structure of Complex Networks: Theory and Applications*, Oxford University Press, Oxford, UK, 2012.

80. E. ESTRADA AND N. HATANO, *Communicability in complex networks*, Phys. Rev. E, 77 (2008), 036111.

81. E. ESTRADA, N. HATANO, AND M. BENZI, *The physics of communicability in complex networks*, Phys. Rep., 514 (2012), pp. 89–119.

82. E. ESTRADA AND D. J. HIGHAM, *Network properties revealed by matrix functions*, SIAM Rev., 52 (2010), pp. 696–714.

83. I. FARIA, *Permanental roots and the star degree of a graph*, Linear Algebra Appl., 64 (1985), pp. 255–265.

84. N. J. FORD, D. V. SAVOSTYANOV, AND N. L. ZAMARASHKIN, *On the decay of the elements of inverse triangular Toeplitz matrices*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1288–1302.

85. R. FREUND, *On polynomial approximations to $f_a(z) = (z - a)^{-1}$ with complex a and some applications to certain non-Hermitian matrices*, Approx. Theory Appl., 5 (1989), pp. 15–31.

86. I. M. Gelfand, *Normierte Ringe*, Mat. Sbornik, 9 (1941), pp. 3–23.

87. I. M. Gelfand and M. A. Neumark, *On the imbedding of normed rings in the ring of operators in Hilbert space*, Mat. Sbornik, 12 (1943), pp. 197–213.

88. I. M. Gelfand, D. A. Raikov, and G. E. Shilov, *Commutative Normed Rings*, Chelsea Publishing Co., Bronx, New York, NY, 1964.

89. P.-L. Giscard, K. Lui, S. J. Thwaite, and D. Jaksch, *An exact formulation of the time-ordered exponential using path-sums*, J. Math. Phys., 56 (2015), 053503.

90. D. F. Gleich, *PageRank beyond the Web*, SIAM Rev., 57 (2015), pp. 321–363.

91. D. F. Gleich and K. Kloster, *Sublinear column-wise actions of the matrix exponential on social networks*, Internet Math., 11 (2015), pp. 352–384.

92. S. Goedecker, *Linear scaling electronic structure methods*, Rev. Mod. Phys., 71 (1999), pp. 1085–1123.

93. S. Goedecker and O. V. Ivanov, *Frequency localization properties of the density matrix and its resulting hypersparsity in a wavelet representation*, Phys. Rev. B, 59 (1999), pp. 7270–7273.

94. K.-I. Goh, B. Khang, and D. Kim, *Spectra and eigenvectors of scale-free networks*, Phys. Rev. E, 64 (2001), 051903.

95. G. H. Golub and G. Meurant, *Matrices, Moments and Quadrature with Applications*, Princeton University Press, Princeton, NJ, 2010.

96. G. H. Golub and C. F. Van Loan, *Matrix Computations. Fourth Edition*, Johns Hopkins University Press, Baltimore and London, 2014.

97. K. Gröchenig and A. Klotz, *Noncommutative approximation: inverse-closed subalgebras and off-diagonal decay of matrices*, Constr. Approx., 32 (2010), pp. 429–466.

98. K. Gröchenig and M. Leinert, *Symmetry and inverse-closedness of matrix algebras and functional calculus for infinite matrices*, Trans. Amer. Math. Soc., 358 (2006), pp. 2695–2711.

99. K. Gröchenig, Z. Rzeszotnik, and T. Strohmer, *Convergence analysis of the finite section method and Banach algebras of matrices*, Integr. Equ. Oper. Theory, 67 (2010), pp. 183–202.

100. M. Grote and T. Huckle, *Parallel preconditioning with sparse approximate inverses*, SIAM J. Sci. Comput., 18 (1997), pp. 838–853.

101. J. Gutiérrez–Gutiérrez, P. M. Crespo, and A. Böttcher, *Functions of the banded Hermitian block Toeplitz matrices in signal processing*, Linear Algebra Appl., 422 (2007), pp. 788–807.

102. S. Güttel and L. Knizhnerman, *A black-box rational Arnoldi variant for Cauchy–Stieltjes matrix functions*, BIT Numer. Math., 53 (2013), pp. 595–616.

103. A. Haber and M. Verhaegen, *Subspace identification of large-scale interconnected systems*, IEEE Trans. Automat. Control, 59 (2014), pp. 2754–2759.

104. A. Haber and M. Verhaegen, *Sparse solution of the Lyapunov equation for large-scale interconnected systems*, Automatica, to appear (2016).

105. M. Hasson, *The degree of approximation by polynomials on some disjoint intervals in the complex plane*, J. Approx. Theory, 144 (2007), pp. 119–132.

106. M. B. Hastings and T. Koma, *Spectral gap and exponential decay of correlations*, Commun. Math. Phys., 265 (2006), pp. 781–804.

107. L. He and D. Vanderbilt, *Exponential decay properties of Wannier functions and related quantities*, Phys. Rev. Lett., 86 (2001), pp. 5341–5344.

108. V. E. Henson and G. Sanders, *Locally supported eigenvectors of matrices associated with connected and unweighted power-law graphs*, Electr. Trans. Numer. Anal., 39 (2012), pp. 353–378.

109. N. J. Higham, *Matrix Functions. Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

110. N. J. Higham, D. S. Mackey, N. Mackey, and F. Tisseur, *Functions preserving matrix groups and iterations for the matrix square root*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 1178–1192.

111. M. Hochbruck and Ch. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.

112. P. Hohenberg and W. Kohn, *Inhomogeneous electron gas*, Phys. Rev., 136 (1964), pp. B864–871.

113. R. A. Horn and C. R. Johnson, *Matrix Analysis. Second Edition*, Cambridge University Press, Cambridge, UK, 2013.

114. R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1994.

115. T. Huckle, *Approximate sparsity patterns for the inverse of a matrix and preconditioning*, Appl. Numer. Math., 30 (1999), pp. 291–303.

116. M. Hutchinson, *A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines*, Commun. Stat. Simul. Comput., 18 (1989), pp. 1059–1076.

117. A. Iserles, *How large is the exponential of a banded matrix?*, New Zealand J. Math., 29 (2000), pp. 177–192.

118. S. Ismail-Beigi and T. A. Arias, *Locality of the density matrix in metals, semiconductors, and insulators*, Phys. Rev. Lett., 82 (1999), pp. 2127–2130.

119. S. Jaffard, *Propriétés des matrices "bien localisées" près de leur diagonale et quelques applications*, Ann. Inst. Henri Poincarè, 7 (1990), pp. 461–476.

120. J. Janas, S. Naboko, and G. Stolz, *Decay bounds on eigenfunctions and the singular spectrum of unbounded Jacobi matrices*, Intern. Math. Res. Notices, 2009, No. 4, pp. 736–764.

121. R. Kadison, *Diagonalizing matrices*, Amer. J. Math., 106 (1984), pp. 1451–1468.

122. R. Kadison and J. Ringrose, *Fundamentals of the Theory of Operator Algebras. Vol. I, Elementary Theory*, Academic Press, Orlando, FL, 1983.

123. W. Kohn, *Analytic properties of Bloch waves and Wannier functions*, Phys. Rev., 115 (1959), pp. 809–821.

124. W. Kohn, *Density functional and density matrix method scaling linearly with the number of atoms*, Phys. Rev. Lett., 76 (1996), pp. 3168–3171.

125. W. Kohn, *Nobel Lecture: Electronic structure of matter—wave functions and density functionals*, Rev. Mod. Phys., 71 (1999), pp. 1253–1266.

126. W. Kohn and L. J. Sham, *Self-consistent equations including exchange and correlation effects*, Phys. Rev. Lett., 140 (1965), pp. A1133–1138.

127. L. Yu. Kolotilina and A. Yu. Yeremin, *Factorized sparse approximate inverse preconditioning I. Theory*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 45–58.

128. A. Koskela and E. Jarlebring, *The infinite Arnoldi exponential integrator for linear inhomogeneous ODEs*, arXiv:1502.01613v2, February 2015.

129. I. Kryshtal, T. Strohmer and T. Wertz, *Localization of matrix factorizations*, Found. Comput. Math., 15 (2015), pp. 931–951.

130. A. N. Langville and C. D. Meyer *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, NJ, 2006.

131. A. J. Laub, *Matrix Analysis for Scientists and Engineers*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2005.

132. R. Lai and J. Lu, *Localized density matrix minimization and linear-scaling algorithms*, J. Comp. Phys., 315 (2016), pp. 194–210.

133. C. S. Lam, *Decomposition of time-ordered products and path-ordered exponentials*, J. Math. Phys., 39 (1998), pp. 5543–5558.

134. C. Le Bris, *Computational chemistry from the perspective of numerical analysis*, Acta Numer., 14 (2005), pp. 363–444.

135. X.-P. Li, R. W. Nunes, and D. Vanderbilt, *Density-matrix electronic structure method with linear system-size scaling*, Phys. Rev. B, 47 (1993), pp. 10891–10894.

136. W. Liang, C. Saravanan, Y. Shao, R. Baer, A. T. Bell, and M. Head-Gordon, *Improved Fermi operator expansion methods for fast electronic structure calculations*, J. Chem. Phys., 119 (2003), pp. 4117–4124.

137. L. Lin, *Localized spectrum slicing*, Math. Comp., to appear (2016).

138. L. Lin and J. Lu, *Sharp decay estimates of discretized Green's functions for Schrödinger type operators*, arXiv:1511.07957v1, November 2015.

139. L. Lin, J. Lu, L. Ying, R. Car, and W. E, *Multipole representation of the Fermi operator with application to the electronic structure analysis of metallic systems*, Phys. Rev. B, 79 (2009), 115133.

140. F.-R. Lin, M. K. Ng, and W.-K. Ching, *Factorized banded inverse preconditioners for matrices with Toeplitz structure*, SIAM J. Sci. Comput., 26 (2005), pp. 1852–1870.

141. M. Lindner, *Infinite Matrices and their Finite Sections*, Birkhäuser, Basel, 2006.

142. X. Liu, G. Strang, and S. Ott, *Localized eigenvectors from widely spaced matrix modifications*, SIAM J. Discrete Math., 16 (2003), pp. 479–498.

143. L. Lopez and A. Pugliese, *Decay behaviour of functions of skew-symmetric matrices*, in Proceedings of HERCMA 2005, 7th Hellenic-European Conference on Computer Mathematics and Applications, September, 22-24, 2005, Athens, E. A. Lipitakis, ed., Electronic Editions LEA, Athens.

144. T. Malas and L. Gürel, *Schur complement preconditioners for surface integral-equation formulations of dielectric problems solved with the multilevel multipole algorithm*, SIAM J. Sci. Comput., 33 (2011), pp. 2440–2467.

145. A. I. Markushevich, *Theory of Functions of a Complex Variable, Vol. III*, Prentice-Hall, Englewood Cliffs, NJ, 1967.

146. O. A. Marques, B. N. Parlett, and C. Vömel, *Computation of eigenpair subsets with the MRRR algorithm*, Numer. Linear Algebra Appl., 13 (2006), pp. 643–653.

147. R. M. Martin, *Electronic Structure. Basic Theory and Practical Methods*. Cambridge University Press, Cambridge, UK, 2004.

148. P. E. Maslen, C. Ochsenfeld, C. A. White, M. S. Lee, and M. Head-Gordon, *Locality and sparsity of ab initio one-particle density matrices and localized orbitals*, J. Phys. Chem. A, 102 (1998), pp. 2215–2222.

149. N. Mastronardi, M. K. Ng, and E. E. Tyrtyshnikov, *Decay in functions of multi-band matrices*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2721–2737.

150. G. Meinardus, *Approximation of Functions: Theory and Numerical Methods*, Springer Tracts in Natural Philosophy, Vol. 13, Springer-Verlag, New York, NY, 1967.

151. P. N. McGraw and M. Menzinger, *Laplacian spectra as a diagnostic tool for network structure and dynamics*, Phys. Rev. E, 77 (2008), 031102.

152. N. Merkle, *Completely monotone functions—a digest*, arXiv:1211.0900v1, November 2012.

153. G. Meurant, *A review of the inverse of symmetric tridiagonal and block tridiagonal matrices*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 707–728.

154. N. Moiseyev, *Non-Hermitian Quantum Mechanics*, Cambridge University Press, Cambridge, UK, 2011.

155. L. Molinari, *Identities and exponential bounds for transfer matrices*, J. Phys. A: Math. and Theor., 46 (2013), 254004.

156. R. Nabben, *Decay rates of the inverse of nonsymmetric tridiagonal and band matrices*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 820–837,

157. Y. Nakatsukasa, *Eigenvalue perturbation bounds for Hermitian block tridiagonal matrices*, Appl. Numer. Math., 62 (2012), pp. 67–78.

158. Y. Nakatsukasa, N. Saito, E. Woei, *Mysteries around the graph Laplacian eigenvalue 4*, Linear Algebra Appl., 438 (2013), pp. 3231–3246.

159. H. Nassar, K. Kloster, and D. F. Gleich, *Strong localization in personalized PageRank vectors*, in D. F. Gleich et al. (Eds.), *Algorithms and Models for the Web Graph*, Lecture Notes in Computer Science, 9479 (2015), Springer, New York, NY, pp. 190–202.

160. G. Nenciu, *Existence of the exponentially localised Wannier functions*, Commun. Math. Phys., 91 (1983), pp. 81–85.

161. A. M. N. Niklasson, *Density matrix methods in linear scaling electronic structure theory*, in R. Zaleśny et al. (Eds.), *Linear-Scaling Techniques in Computational Chemistry and Physics*, Springer, New York, NY, 2011, pp. 439–473.

162. J. Pan, R. Ke, M. K. Ng, and H.-W. Sun, *Preconditioning techniques for diagonal-times-Toeplitz matrices in fractional diffusion equations*, SIAM. J. Sci. Comput., 36 (2014), pp. A2698–A2719.

163. B. N. Parlett, *Invariant subspaces for tightly clustered eigenvalues of tridiagonals*, BIT Numer. Math., 36 (1996), pp. 542–562.

164. B. N. Parlett, *A result complementary to Geršgorin's circle theorem*, Linear Algebra Appl., 432 (2009), pp. 20–27.

165. B. N. Parlett and I. S. Dhillon, *Relatively robust representations of symmetric tridiagonals*, Linear Algebra Appl., 309 (2000), pp. 121–151.

166. M. S. Paterson and L. J. Stockmeyer, *On the number of nonscalar multiplications necessary to evaluate polynomials*, SIAM J. Comput., 2 (1973), pp. 60–66.

167. S. Pozza and V. Simoncini, *Decay bounds for functions of banded non-Hermitian matrices*, arXiv.1605.01595v1, May 2016.

168. E. Prodan, *Nearsightedness of electronic matter in one dimension*, Phys. Rev. B, 73 (2006), 085108.

169. E. Prodan, S. R. Garcia, and M. Putinar, *Norm estimates of complex symmetric operators applied to quantum systems*, J. Phys. A: Math. Gen., 39 (2006), pp. 389–400.

170. E. Prodan and W. Kohn, *Nearsightedness of electronic matter*, Proc. Nat. Acad. Sci., 102 (2005), pp. 11635–11638.

171. N. Razouk, *Localization Phenomena in Matrix Functions: Theory and Algorithms*, PhD Thesis, Emory University, 2008.

172. L. Reichel, G. Rodriguez, and T. Tang, *New block quadrature rules for the approximation of matrix functions*, Linear Algebra Appl., 502 (2016), pp. 299–326.

173. S. Roch, *Finite Sections of Band-Dominated Operators*, Memoirs of the American Mathematical Society, 191(895), Providence, RI, 2008.

174. G. Rodriguez, S. Seatzu, and D. Theis, *An algorithm for solving Toeplitz systems by embedding in infinite systems*, Oper. Theory Adv. Appl., 160 (2005), pp. 383–401.

175. E. H. Rubensson, E. Rudberg, and P. Salek, *Methods for Hartree–Fock and density functional theory electronic structure calculations with linearly scaling processor time and memory usage*, in R. Zaleśny et al. (Eds.), *Linear-Scaling Techniques in Computational Chemistry and Physics*, Springer, New York, NY, 2011, pp. 269–300.

176. W. Rudin, *Functional Analysis*, McGraw–Hill, New York, NY, 1973.

177. Y. Saad, *Iterative Methods for Sparse Linear Systems. Second Edition*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.

178. Y. Saad, J. R. Chelikowsky, and S. M. Shontz, *Numerical methods for electronic structure calculations of materials*, SIAM Rev., 52 (2010), pp. 3–54.

179. N. Schuch, J. I. Cirac, M. M. Wolf, *Quantum states on harmonic lattices*, Comm. Math. Phys., 267 (2006), pp. 65–92.

180. M. Shao, *On the finite section method for computing exponentials of doubly-infinite skew-Hermitian matrices*, Linear Algebra Appl., 451 (2014), pp. 65–96.

181. D. I. Shuman, B. Ricaud, and P. Vandergheynst, *Vertex-frequency analysis on graphs*, Appl. Comput Harmon. Anal., 40 (2016), pp. 260–291.

182. C. Siefert and E. de Sturler, *Probing methods for saddle-point problems*, Electr. Trans. Numer. Anal., 22 (2006), pp. 163–183.

183. B. Simon, *Semiclassical analysis of low lying eigenvalues. I. Nondegenerate minima: asymptotic expansions*, Ann. Inst. H. Poincaré Sect. A, 38 (1983), pp. 295–308.

184. V. Simoncini, *Computational methods for linear matrix equations*, SIAM Review, to appear.

185. D. T. Smith, *Exponential decay of resolvents and discrete eigenfunctions of banded infinite matrices*, J. Approx. Theory, 66 (1991), pp. 83–97.

186. G. Stolz, *An introduction to the mathematics of Anderson localization*, in R. Sims and D. Ueltschi (Eds.), *Entropy and the Quantum II*, Contemp. Math. 552, pp. 71–108, American Mathematical Society, Providence, RI (2011).

187. G. Strang and S. MacNamara, *Functions of difference matrices are Toeplitz plus Hankel*, SIAM Rev., 56 (2014), pp. 525–546.

188. T. Strohmer, *Four short stories about Toeplitz matrix calculations*, Linear Algebra Appl., 343–344 (2002), pp. 321–344.

189. P. K. Suetin, *Series of Faber polynomials*, Gordon and Breach Science Publishers, Amsterdam, 1998.

190. Q. Sun, *Wiener's lemma for infinite matrices with polynomial off-diagonal decay*, C. R. Acad. Sci. Paris, Ser. I 340 (2005), pp. 567–570.

191. P. Suryanarayana, *On spectral quadrature for linear-scaling density functional theory*, Chem. Phys. Letters, 584 (2013), pp. 182–187.

192. H. Tal-Ezer, *Polynomial approximation of functions of matrices and applications*, J. Sci. Comput., 4 (1989), pp. 25–60.

193. L. V. Tran, V. H Vu, and K. Wang, *Sparse random graphs: eigenvalues and eigenvectors*, Random Structures Algorithms, 42 (2013), pp. 110–134.

194. L. N. Trefethen, *Numerical computation of the Schwarz–Christoffel transformation*, SIAM J. Sci. Stat. Comput., 1 (1980), pp. 82–102.

195. L. N. Trefethen and D. Bau, *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

196. L. N. Trefethen, M. Contedini and M. Embree, *Spectra, pseudospectra, and localization for random bidiagonal matrices*, Comm. Pure Appl. Math., 54 (2001), pp. 595–623.

197. L. N. Trefethen and M. Embree, *Spectra and Pseudospectra. The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.

198. C. V. M. van der Mee, G. Rodriguez, and S. Seatzu, *LDU factorization results for bi-infinite and semi-infinite scalar and block Toeplitz matrices*, Calcolo, 33 (1998), pp. 307–335.

199. C. V. M. van der Mee, G. Rodriguez, and S. Seatzu, *Block Cholesky factorization of infinite matrices and orthonormalization of vectors of functions*, in *Advances in Computational Mathematics (Guangzhou, 1997)*, pp. 423–455, Lecture Notes in Pure and Appl. Math., Dekker, New York, 1999.

200. R. S. Varga, *Nonnegatively posed problems and completely monotonic functions*, Lin. Algebra Appl., 1 (1968), pp. 329–347.

201. P. S. Vassilevski, *On some ways of approximating inverses of band matrices in connection with deriving preconditioners based on incomplete block factorizations*, Computing, 43 (1990), pp. 277–296.

202. C. Vömel and B. N. Parlett, *Detecting localization in an invariant subspace*, SIAM J. Sci. Comput., 33 (2011), pp. 3447–3467.

203. H. Wang and Q. Ye, *Error bounds for the Krylov subspace methods for computations of matrix exponentials*, Tech. Rep., Department of Mathematics, University of Kentucky, Lexington, KY, 2016.

204. H. F. Weinberger, *A First Course in Partial Differential Equations*, John Wiley and Sons, New York, NY, 1965.

205. D. V. Widder, *The Laplace Transform*, Princeton University Press, 1946.

206. W. Yang, *Direct calculation of electron density in Density-Functional Theory*, Phys. Rev. Lett., 66 (1991), pp. 1438–1441.

207. Q. Ye, *Error bounds for the Lanczos method for approximating matrix exponentials*, SIAM J. Numer. Anal., 51 (2013), pp. 68–87.