





Full Length Article

Approximating sparse matrices and their functions using matrix-vector products

Taejun Park ^{a,1,*}, Yuji Nakatsukasa ^{b,2}^a *Institute of Mathematics, EPF Lausanne, 1015, Lausanne, Switzerland*^b *Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK*

ARTICLE INFO

Editor: Thomas Strohmer

Keywords:

Matrix function
Banded matrix
Sparse matrix
Matrix-vector products
Compressed sensing

ABSTRACT

The computation of a matrix function $f(A)$ is an important task in scientific computing appearing in machine learning, network analysis and the solution of partial differential equations. In this work, we use only matrix-vector products $x \mapsto Ax$ to approximate functions of sparse matrices and matrices with similar structures such as sparse matrices A themselves or matrices that have a similar decay property as matrix functions. We show that when A is a sparse matrix with an unknown sparsity pattern, techniques from compressed sensing can be used under natural assumptions. Moreover, if A is a banded matrix then certain deterministic matrix-vector products can efficiently recover the large entries of $f(A)$. We describe an algorithm for each of the two cases and give error analysis based on the decay bound for the entries of $f(A)$. We finish with numerical experiments showing the accuracy of our algorithms.

1. Introduction

The evaluation of matrix functions is an important task arising frequently in many areas of scientific computing. They appear in the numerical solution of partial differential equations [1–3], network analysis [4,5], electronic structure calculations [6,7] and statistical learning [8,9], among others. In some problems, the matrix of interest has some underlying structure such as bandedness or sparsity. For example, in graphs with community structure one requires functions of sparse matrices [10] and for electronic structure calculations one computes functions of banded matrices [6,7]. In this paper, we consider the problem of approximating $f(A)$ through matrix-vector products $x \mapsto Ax$ only, assuming A is sparse or banded.

Let $A \in \mathbb{R}^{m \times n}$ and f be a function that is analytic on the closure of a domain $\Omega \subset \mathbb{C}$ that contains the eigenvalues of A . Then the matrix function $f(A)$ is defined as

$$f(A) = \frac{1}{2\pi i} \int_{\partial\Omega} f(z)(zI - A)^{-1} dz,$$

* Corresponding author.

E-mail addresses: taejun.park@epfl.ch (T. Park), nakatsukasa@maths.ox.ac.uk (Y. Nakatsukasa).¹ This work was done partially while the author was at the University of Oxford. TP is supported by the RandESC project funded by the Swiss Platform for Advanced Scientific Computing. TP was also supported by the Heilbronn Institute for Mathematical Research.² YN is supported by EPSRC grants EP/Y010086/1 and EP/Y030990/1.

where I denotes the identity matrix and $(zI - A)^{-1}$ is called the resolvent matrix of A . If $A \in \mathbb{R}^{n \times n}$ is diagonalizable with $A = XDX^{-1}$ where X is the eigenvector matrix and D is the diagonal matrix containing the eigenvalues $\{\lambda_i\}_{i=1}^n$ of A , then

$$f(A) = Xf(D)X^{-1}, \quad f(D) = \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix}.$$

There are many other equivalent ways of defining $f(A)$, see for example [11].

When the matrix dimension is large, explicitly computing $f(A)$ using standard algorithms [11,12] becomes prohibitive in practice, usually requiring $O(n^3)$ operations. Moreover, the matrix A may also be too large to store explicitly. In this scenario, A is often only available through matrix-vector products and $f(A)b^1$ is used to infer information about $f(A)$ [13,14]. In this paper, we show that the whole matrix $f(A)$ can still be approximated in a sparse format using $\ll n$ matrix-vector products with A if A is sparse and f is smooth. We exploit the decay bounds for the entries of $f(A)$ [15] to recover its large entries using matrix-vector products; see also [16]. We primarily study functions of sparse matrices with general sparsity pattern and matrices with few large entries in each row/column (see Section 2), but also consider functions of banded matrices and matrices with off-diagonal decay as a special case (see Section 3). Our work is particularly useful if $f(A)$ is sparse or has only a few entries that are large in each row/column or if A is banded with bandwidth $k \ll n$. Throughout this work, we assume that the function f is analytic on the closure of a domain that contains the eigenvalues of A .

Let A be a diagonalizable, sparse matrix with some sparsity pattern and f (either $f : \mathbb{R} \rightarrow \mathbb{R}$ or $f : \mathbb{C} \rightarrow \mathbb{C}$) a function that is analytic in a region containing the spectrum of A . Depending on the sparsity pattern of A and the smoothness of f , some entries of $f(A)$ can be shown to be exponentially smaller than others [15] and we can view them as noise. For a sparse matrix with a general sparsity pattern, it can be difficult to know the support (location of the large entries) of $f(A)$ in advance, for example, if the sparsity pattern of A is unknown. However, we can write $f(A) = \widehat{f(A)} + E$ with the matrix E having only small entries in magnitude and compute $\widehat{f(A)}$ by viewing $\widehat{f(A)}$ as signals and E as noise using *compressed sensing* row by row or column by column. Compressed sensing is a technique in signal processing for efficiently reconstructing a signal which is sparse in some domain by finding solutions to underdetermined linear systems [17–19]. As a special case (by taking $f(A) = A$), our algorithm can be used for recovering a sparse matrix A itself, where A is accessed only through matrix-vector multiplications $x \mapsto Ax$. The details will be discussed in Section 2.

Let A now be a banded matrix. Then $f(A)$ can usually be well approximated by a banded matrix [20]. There are theoretical results that confirm this property [21], which show that the entries of $f(A)$ decay exponentially away from the main diagonal. The observation is that $|[f(A)]_{ij}| \ll |[f(A)]_{ii}|, |[f(A)]_{jj}|$ for $|i - j| \gg 1$, that is, the large entries are typically located not too far away from the main diagonal. Therefore, we can dismiss the exponentially small entries and find a matrix that recovers the large entries of $f(A)$, i.e., find a matrix $\widehat{f(A)}$ such that the difference $f(A) - \widehat{f(A)}$ only contains small entries in magnitude. We can show that the large entries of $f(A)$ can be captured through matrix-vector products with the following simple deterministic matrix

$$I_n^{(s)} := [I_s, I_s, \dots, I_s, I_s(:, 1 : \{n/s\} \cdot s)]^T \in \mathbb{R}^{n \times s}$$

where $s < n$, I_s is the $s \times s$ identity matrix and $I_s(:, 1 : \{n/s\} \cdot s)$ are the first $\{n/s\} \cdot s$ columns of I_s , which ensures that $I_n^{(s)}$ has exactly n rows. Here $\{x\}$ is the fractional part of x . This is unlike the sparse case where we now have significant information on the support of $f(A)$. This set of vectors $I_n^{(s)}$ are identical to the probing vectors proposed in [22] for estimating functions of banded matrices and [23,24] for estimating sparse, banded Jacobian matrices. If we let s be proportional to the bandwidth of A , the matrix-vector products $f(A)I_n^{(s)}$ capture the large entries near the main diagonal, giving us a good approximation for $f(A)$. We discuss the details in Section 3.

Both cases require matrix-vector products with $f(A)$, i.e., $x \mapsto f(A)x$. Computing this matrix-vector product to high accuracy is crucial for our algorithms. Popular methods include the standard (polynomial) Krylov method [11,25], and contour integration [26]. In this paper, we discuss three functions: the exponential e^A , the square root function \sqrt{A} and the logarithm $\log(A)$. For the exponential we use the polynomial Krylov method, and for the square root function and the logarithm, we use contour integration as in [26, method 2] as both functions have a branch cut on the negative real axis $(-\infty, 0]$, making them difficult to approximate using polynomials if A has eigenvalues close to 0. For Krylov methods, block versions can also be used [27], which compute $f(A)B \in \mathbb{R}^{n \times s}$ at once instead of column by column. Although the computation of the matrix-vector product $f(A)b$ is important and crucial to the approximation of $f(A)$, we do not discuss the details further as they are not the focus of this paper; see for example [11, Ch.13].

1.1. Existing methods

First, Benzi and Razouk [15] describe an algorithm that exploits the decay property by first finding a good polynomial approximation p to f , for example using Chebyshev interpolation, and computing $p(A)$ while using a dropping strategy to keep the matrix as sparse as possible. However, this method can quickly become infeasible if f cannot be well approximated by a low-degree polynomial or if sparse data structures and arithmetic is not available. Frommer, Schimmel and Schweitzer [22] use probing methods to approximate matrix functions $f(A)$ and its related quantities, for instance the trace of $f(A)$. The method finds probing vectors to estimate the entries of $f(A)$ by partitioning the vertices of the directed graph $G(A)$ associated with the sparse matrix A using graph

¹ We note that the matrix-vector product $f(A)b$ is not computed by explicitly forming $f(A)$ and then multiplying it by b . Instead, it is approximated using matrix-vector products involving A alone, typically through Krylov subspace methods.

coloring and the sparsity pattern of \mathbf{A} . Our Section 3 on banded matrices can be seen as a special case of [22], but as banded matrices arise naturally in applications and come with strong theoretical results, we make the sensing matrix explicit and study them in detail. The strategy proposed in [22] costs $\mathcal{O}(\Delta^d n)$ to obtain the graph coloring where Δ is the maximal degree of $G(\mathbf{A})$ and d is a distance parameter. In the special case when \mathbf{A} is a banded matrix the graph coloring can be obtained with $\mathcal{O}(n)$ cost. Our work and [22] satisfies a similar theoretical accuracy bound up to a low-degree polynomial factor in n (cf. Section 3 of [22] and Theorem 1 in our work), but we use compressed sensing to estimate the large entries of each row/column. Cortinovis, Kressner and Massei [28] use a divide-and-conquer approach to approximate matrix functions where the matrix is either banded, hierarchically semiseparable or has a related structure. The divide-and-conquer method is based on (rational) Krylov subspace methods for performing low-rank updates of matrix functions. Under the assumption that the minimal block size in the divide-and-conquer algorithm is $\mathcal{O}(1)$ and the low-rank update in the Krylov method converges in $\mathcal{O}(1)$ iterations, the algorithm runs with complexity $\mathcal{O}(nk^2)$ for k -banded matrices.

There are other related works that use matrix-vector products to recover matrices with certain properties or using compressed sensing to recover matrices with only few large entries. In [29–31], the authors devise algorithms or strategies based on matrix-vector products for recovering structured matrices such as hierarchical low-rank matrices and Toeplitz matrices or for answering queries such as whether an unknown matrix is symmetric or not. In particular, Curtis, Powell and Reid [24] and later Coleman and Moré [23] use a certain graph coloring of the adjacency graph of a sparse or banded matrix to approximate sparse or banded Jacobian matrices using matrix-vector products. Bekas, Kokiopoulou and Saad [32] suggested using the same set of coloring as the banded case to recover matrices with rapid off-diagonal decay.

In [33,34], the authors formulate an algorithm based on compressed sensing techniques to recover sparse matrices with at most $\mathcal{O}(sn/\log^2 n)$ nonzero entries using s matrix-vector products (See [34, Theorem 6.3]). This method uses a vectorized form of the matrix in question after a certain transformation, which can become expensive. Lastly, Dasarathy, Shah and Bhaskar [35] use compressed sensing techniques by applying compression (sensing operators) on both sides, i.e. $\mathbf{Y}_1 \mathbf{X} \mathbf{Y}_2^T$, to recover an unknown sparse matrix \mathbf{X} . The authors use $\mathcal{O}(\sqrt{n} \log n)$ matrix-vector products with \mathbf{X} where the number of nonzero entries of \mathbf{X} is $\mathcal{O}(n)$. This can become prohibitive for large n as computing matrix-vector products is usually the dominant cost.

1.2. Contributions

Our first contribution is devising strategies for approximating (or sometimes recovering exactly, e.g. for sparse matrices) functions of sparse matrices and matrices with similar structures, e.g., matrices with only few large entries in each row/column or matrices with rapid off-diagonal decay, which is typically the case when the function f is analytic on the closure of a domain that contains the eigenvalues of \mathbf{A} . We provide two strategies, one for functions of sparse matrices or matrices with only few large entries in each row/column and another for functions of banded matrices or matrices with off-diagonal decay. The two strategies are based on approximating the large entries of the matrix in question using matrix-vector products only. This is advantageous if the matrix dimension is too large to store explicitly or the matrix is only available through matrix-vector products. For general sparse matrices and their functions, we use compressed sensing to estimate the large entries of each row/column without needing the sparsity pattern of \mathbf{A} . For the banded case, our work is analogous to the previous work done in [22–24] but we give a focused treatment of banded matrices.

Our second contribution is designing simple algorithms, Algorithm 1 (SpaMRAM) for sparse matrices with general sparsity pattern and Algorithm 2 (BaMRAM) for banded matrices, with the 2-norm error bounds based on the decay structure of the matrix. We show that SpaMRAM converges quickly as long as the matrix has only few large entries in each row/column and that BaMRAM converges exponentially with the number of matrix-vector products. Assuming that f is sufficiently smooth so that the Krylov method converges in $\mathcal{O}(1)$ iterations when computing $f(\mathbf{A})\mathbf{b}$, SpaMRAM runs with complexity $\mathcal{O}(n^2 \log n)$ for functions of general sparse matrices and BaMRAM runs linearly in n with complexity $\mathcal{O}(nk^2)$ for functions of k -banded matrices. For banded matrices, BaMRAM runs with complexity $\approx 4nk^2$ and for functions of banded matrices, BaMRAM has the same complexity, $\mathcal{O}(nk^2)$, as [28]. The output of the algorithm is also sparse and has $\mathcal{O}(nk)$ nonzero entries where k is the number of nonzero entries recovered per row/column, which can be advantageous in many situations as $f(\mathbf{A})$ need not be exactly sparse even if \mathbf{A} is sparse.

1.3. Notation

Throughout, we write $\lfloor x \rfloor$ and $\{x\}$ to denote the floor function of x and the fractional part of x respectively. We use $\|\cdot\|_p$ where $1 \leq p \leq \infty$ to denote the p -norm for vectors and matrices and define $\|\cdot\|_{\max}$ by $\|\mathbf{A}\|_{\max} = \max_{i,j} |A_{ij}|$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a matrix. We use $\|v\|_0$ to denote the number of nonzeros in the vector $v \in \mathbb{R}^n$. Lastly, we use MATLAB style notation for matrices and vectors. For example, for the k th to $(k + j)$ th columns of a matrix \mathbf{A} we write $\mathbf{A}(:, k : k + j)$.

2. Approximation for functions of general sparse matrices

Let \mathbf{A} be a diagonalizable sparse matrix with some sparsity pattern. Suppose that f is analytic in a region containing the spectrum of \mathbf{A} . Then depending on the sparsity pattern of \mathbf{A} , the entries of $f(\mathbf{A})$ decay exponentially away from certain small regions of the matrix. More formally, associate the unweighted directed graph $G(\mathbf{A}) = (V, E)$ to $\mathbf{A} = [A_{ij}]$ so that \mathbf{A} is the adjacency matrix of $G(\mathbf{A})$. V is the vertex set consisting of integers from 1 to n and E is the edge set containing all ordered pairs (i, j) with $A_{ij} \neq 0$. We define the

distance $d_G(i, j)$ between vertex i and vertex j to be the shortest directed path from i to j .² We then have that there exist constants $C_0 > 0$ and $0 < \lambda_0 < 1$ such that

$$|[f(\mathbf{A})]_{ij}| < C_0 \lambda_0^{d_G(i,j)} \tag{1}$$

for all $i, j = 1, 2, \dots, n$. For further details see [15]. This result tells us that if $d_G(i, j) \gg 1$ then $|[f(\mathbf{A})]_{ij}| \ll 1$, which in turn means that depending on the sparsity pattern of \mathbf{A} , there may only be few large entries in $f(\mathbf{A})$ with the rest being negligible. With this observation we can aim to recover the dominating entries of each row/column of $f(\mathbf{A})$. We solve this problem using *compressed sensing* by interpreting small entries as noise and the large entries as signals.³

Compressed sensing [17,19] is a technique in signal processing for efficiently reconstructing a sparse signal. Let $\mathbf{Y} \in \mathbb{R}^{n \times s}$ be a so-called *sensing operator* and $\mathbf{v} \in \mathbb{R}^n$ be an unknown sparse vector. Then the problem of recovering \mathbf{v} using the measurements $\mathbf{y} = \mathbf{Y}^T \mathbf{v}$ is a combinatorial optimization problem given by

$$\min_{\mathbf{z} \in \mathbb{R}^n} \|\mathbf{z}\|_0 \text{ subject to } \mathbf{y} = \mathbf{Y}^T \mathbf{z} \tag{2}$$

where $\|\mathbf{z}\|_0$ denotes the number of nonzeros in the vector \mathbf{z} . There are two established classes of algorithms, among others, to solve (2) in the compressed sensing literature, namely basis pursuit or ℓ_1 -minimization using convex optimization [36], and greedy methods [37]. For large n , convex optimization can become infeasible (especially given that we need to solve (2) n times to recover the whole matrix), so we use the cheaper option, greedy methods. Greedy methods are iterative and at each iteration attempt to find the correct location and the values of the dominant entries.

Let $\mathbf{v} \in \mathbb{R}^n$ be a k -sparse vector, that is, a vector with at most k nonzero entries. To recover \mathbf{v} with theoretical guarantee we typically require at least $s = \mathcal{O}(k \log n)$ measurements. In practice, $s = \mathcal{O}(k)$ measurements seem to work well [38]. Each measurement is taken through vector-vector multiply with \mathbf{v} . The three classes of frequently used sensing operators \mathbf{Y} are: a Gaussian matrix \mathcal{N} with entries drawn from i.i.d. $\mathcal{N}(0, s^{-1})$, s columns of the discrete cosine transform matrix \mathbf{C} sampled uniformly at random and a sparse matrix \mathcal{S}_ξ with ξ nonzero entries per row with the support set drawn uniformly at random and the nonzero values drawn uniformly at random from $\pm \xi^{-1/2}$. Here, ξ is the sparsity parameter for the sparse sensing operator.

There are a number of greedy methods available, including Normalized Iterative Hard Thresholding (NIHT) [39], Hard Thresholding Pursuit (HTP) [40] and Compressive Sampling Matching Pursuit (CoSaMP) [41]. Among the key differences and the trade-offs between the different methods is the cost per iteration in order to speed up support detection and the number of iterations for better asymptotic convergence rate. Many greedy methods have uniform recovery guarantees if the sensing operator \mathbf{Y} has sufficiently small restricted isometry constants [42].⁴ This means that if we want to recover a matrix then we can use the same sensing operator \mathbf{Y} for the recovery of all rows of a matrix, which makes it much more efficient than recovering n rows of a matrix individually using different sensing operators.

In this paper, we use NIHT as it is simple and has a low cost per iteration.⁵ NIHT is an iterative hard thresholding algorithm which performs gradient descent and then hard thresholding at each iteration to update the support and the value of the support set: the $(N + 1)$ th iterate $\mathbf{v}^{(N+1)}$ of NIHT is

$$\mathbf{v}^{(N+1)} = H_k(\mathbf{v}^{(N)} + \mu \mathbf{Y}(\mathbf{y} - \mathbf{Y}^T \mathbf{v}^{(N)})). \tag{3}$$

Here H_k is the hard thresholding operator that sends all but the largest k entries (in absolute value) of the input vector to 0. The stepsize μ in NIHT is chosen to be optimal when the support of the current iterate is the same as the sparsest solution. Assuming that \mathbf{Y} satisfies the restricted isometry property (RIP), then at iteration N , the output $\mathbf{v}^{(N)}$ of NIHT satisfies

$$\|\mathbf{v} - \mathbf{v}^{(N)}\|_2 \leq 2^{-N} \|\mathbf{v}^k\|_2 + 8\epsilon_k,$$

where \mathbf{v}^k is the best k -sparse vector approximation to \mathbf{v} and

$$\epsilon_k = \|\mathbf{v} - \mathbf{v}^k\|_2 + \frac{1}{\sqrt{k}} \|\mathbf{v} - \mathbf{v}^k\|_1,$$

and after at most $N^* = \lceil \log_2(\|\mathbf{v}^k\|_2/\epsilon_k) \rceil$ iterations, we have

$$\|\mathbf{v} - \mathbf{v}^{(N^*)}\|_2 \leq 9\epsilon_k. \tag{4}$$

Further details about NIHT can be found in [39].

The dominating cost for the greedy methods is the matrix-vector product between the sensing operator or the transpose of the sensing operator and the unknown sparse signal, i.e., the cost of $\mathbf{v} \mapsto \mathbf{Y}\mathbf{v}$ and $\mathbf{w} \mapsto \mathbf{Y}^T \mathbf{w}$ respectively. If the sensing operator is a

² Note that if \mathbf{A} is not symmetric then $d_G(i, j) \neq d_G(j, i)$ in general.

³ Another possibility would be to identify the nonzero structure using the graph structure (V, E) and f . This can quickly become intractable, and in some cases $f(\mathbf{A})$ can be sparse despite the graph structure suggesting otherwise, so here we opt for a generic approach.

⁴ Under mild assumptions on s , the three classes of sensing operators described in the previous paragraph have small restricted isometry constants with high probability.

⁵ Other greedy methods such as CoSaMP have similar (or slightly stronger) theoretical guarantees as NIHT, but is generally more expensive; hence making NIHT more preferable in our setting ($k \ll n$ and n problems to solve).

Gaussian \mathcal{N} then the cost is $\mathcal{O}(ns)$, if Y is a subsampled discrete cosine matrix C then the cost is $\mathcal{O}(n \log n)$ or $\mathcal{O}(n \log s)$ if we use the subsampled FFT algorithm [43] and lastly, if Y is a sparse matrix S_ξ then the cost is $\mathcal{O}(\xi n)$.

In our context, we are concerned with recovering the entire matrix. We now use compressed sensing row by row or column by column to approximate functions of sparse matrices.

2.1. Strategy for recovering general sparse matrices

In this section, we consider matrices $B \in \mathbb{R}^{n \times n}$ that satisfy

$$|B_{ij}| < C_B \lambda_B^{d(i,j)} \tag{5}$$

where $C_B > 0$ and $0 < \lambda_B < 1$ are constants and $d(i, j) \in \mathbb{R}_{\geq 0} \cup \{+\infty\}$ is some bivariate function that captures the magnitude of the entries of B . Functions of sparse matrices $B = f(A)$ with a smooth f is a special case of B that satisfies (1). Another special case is $B = A$, which recovers the sparse matrix A itself. The zero entries of B would correspond to $d(i, j) = \infty$ in (5).

We are concerned with recovering the whole matrix. We approximate the dominant entries of each row of B by taking measurements through matrix-vector products $B\mathbf{y}_i$ where \mathbf{y}_i is the i th column of the sensing operator $Y \in \mathbb{R}^{n \times s}$.⁶ Let $\text{CS}(Y, \mathbf{b}_i Y, k)$ be a function that solves (2) where $Y \in \mathbb{R}^{n \times s}$ is the sensing operator, $\mathbf{b}_i Y \in \mathbb{R}^{1 \times s}$ are the measurements where \mathbf{b}_i is the i th row of B (i.e., $\mathbf{b}_i Y$ is the i th row of BY) and k is the sparsity parameter. For the compressed sensing algorithm CS, we can take, for example CS = NIHT. The algorithm to recover the dominant entries of each row of B is given below in Algorithm 1, which we call SpaMRAM (Sparse Matrix Recovery Algorithm using Matvecs).

Algorithm 1 Sparse Matrix Recovery Algorithm using Matvecs (SpaMRAM).

Require: Matrix-vector product $\text{mvp} : \mathbf{x} \mapsto B\mathbf{x}$;

Y sensing operator;

k sparsity parameter;

CS function that solves (2) from input (Y, \mathbf{y}, k) using a compressed sensing algorithm, e.g. NIHT

Ensure: \hat{B} sparse matrix with at most k nonzero entries in each row.

1: Draw a sensing operator $Y \in \mathbb{R}^{n \times s}$, e.g. Gaussian.

▷ rec. $s = \lceil 2k \log(n/k) \rceil$

2: Compute matrix-vector products $F_s = \text{mvp}(Y)$

3: **for** $i = 1, 2, \dots, n$ **do**

4: $\hat{B}(i, :) = \text{CS}(Y, F_s(i, :), k)$

▷ Using compressed sensing to approximate the rows of B

5: **end for**

The algorithm SpaMRAM is a straightforward application of compressed sensing techniques used to recover a sparse matrix instead of a vector, particularly when the goal is to recover A itself. We are nonetheless unaware of existing work that spells out explicitly, and SpaMRAM can be significantly more efficient than alternative approaches [33,34] that vectorize the matrix. Since most sensing matrices satisfy the RIP with failure probability exponentially decaying in s [42], by the union bound we can conclude that the whole matrix can be recovered with high probability.

2.1.1. Complexity of SpaMRAM (Algorithm 1)

The complexity of SpaMRAM is s matrix-vector products with B and n times the complexity of solving the compressed sensing problem. The complexity for solving n compressed sensing problems using NIHT and the subsampled DCT matrix as Y is $\mathcal{O}(n^2 \log n)$, where we take the number of NIHT iterations to be $\mathcal{O}(1)$.

When $B = f(A)$ for sparse A , we can evaluate $f(A)Y$ using Krylov methods. Assuming that f is analytic in a region containing the eigenvalues of A ,⁷ the Krylov method converges (to sufficient accuracy) in $\mathcal{O}(1)$ iterations, then the complexity of SpaMRAM is $\mathcal{O}(s \cdot \text{nnz}(A) + n^2 \log n) = \mathcal{O}(k \log n \cdot \text{nnz}(A) + n^2 \log n)$ which consists of $\mathcal{O}(s \cdot \text{nnz}(A)) = \mathcal{O}(k \log n \cdot \text{nnz}(A))$ flops for computing $s = \mathcal{O}(k \log n)$ matrix-vector products with $f(A)$ and $\mathcal{O}(n^2 \log n)$ flops for solving n compressed sensing problems. Here $\text{nnz}(A)$ is the number of nonzero entries of A .

2.1.2. Dependency of the compressed sensing solver in SpaMRAM

The majority of the computational task in SpaMRAM (Algorithm 1) relies on compressed sensing. There are three main parameter choices in relation to compressed sensing for SpaMRAM: the sensing operator Y , the number of measurements s and the compressed sensing solver CS. For the solver, we suggest NIHT as it is the simplest and has a low cost per iteration. For better asymptotic guarantee, but with a higher cost per iteration, we suggest CoSaMP or its variants, e.g. CSMPSP discussed in [38]. For the sensing operator Y , all three choices mentioned previously are good choices. For robustness, we recommend the Gaussian matrix \mathcal{N} and for speed, we recommend the sparse matrix S_ξ with the sparsity parameter $\xi = \min\{8, s\}$ if a good sparse matrix library is available, otherwise we recommend the subsampled DCT matrix C . Lastly, for the number of measurements s we recommend $s = \lceil 2k \log \left(\frac{n}{k} \right) \rceil$ for theoretical

⁶ We can also recover the dominant entries of B column by column using $B^T Y$ if the matrix-vector product $\mathbf{x} \mapsto B^T \mathbf{x}$ is available.

⁷ It also suffices for f to be approximable to high accuracy with a low-degree polynomial.

guarantee [17,19] as suggested in SpaMRAM. However, a constant multiple of the sparsity parameter k , say $s = 10k$, is observed to work well in practice [38,42].

2.1.3. A posteriori error estimate for SpaMRAM

The behaviour of SpaMRAM can be difficult to grasp as it relies on various parameters that are sometimes implicit, e.g., (5). Therefore, we can incorporate a posteriori error testing for SpaMRAM to observe how well it might perform on a particular problem. Let us try to estimate the error using known quantities. We have the sensing operator Y , the measurements $F_s \in \mathbb{R}^{n \times s}$ and the output of SpaMRAM, \hat{B} . If the matrix-vector products (line 2 of SpaMRAM) have been computed exactly then we have $BY = F_s$, so we try to estimate the error by determining how close the measurements computed using the output of SpaMRAM, i.e. $\hat{B}Y$, are to the computed measurements F_s . We define

$$\delta_{RE}^{(S)} := \frac{\|\hat{B}Y - F_s\|_2}{\|F_s\|_2} \tag{6}$$

as the relative error of the computed solution of SpaMRAM quantified through the measurements. Note that the error estimate $\delta_{RE}^{(S)}$ can be computed a posteriori in $\mathcal{O}(nks)$ as \hat{B} has at most k nonzero entries in each row. The error estimate $\delta_{RE}^{(S)}$ can be used as a heuristic for how SpaMRAM performs. The main motivation for this approach is that a small number of Gaussian measurements estimate the norm well [44,45]. In our case, we essentially have $F_s = BY$ and $\hat{B}Y$ giving us $(B - \hat{B})Y$, which is a small number of Gaussian measurements of the absolute error $B - \hat{B}$ if Y is the Gaussian sensing operator.⁸ Therefore δ_{RE} provides a good guidance for how small the error is for SpaMRAM. For the other sensing operators, the behaviour is similar [45]. Later, we show in a numerical experiment (Fig. 2) in Section 4 that $\delta_{RE}^{(S)}$ can estimate the performance of SpaMRAM well.

2.1.4. Adaptive methods

There are adaptive methods that reduce the number of matrix-vector products required for sparse recovery, for example [46–49] by making adaptive measurements based on previous measurements. However, in our context we are concerned with recovering the whole matrix and it is difficult to make adaptive measurements that account for every row/column of B when we only have access to B using matrix-vector products. A naive strategy is to make adaptive measurements for each row/column, but this requires at least $\mathcal{O}(n)$ matrix-vector products, which is computationally infeasible and futile as we can recover B using n measurements using the $n \times n$ identity matrix.

2.1.5. Shortcomings

There are some limitations with SpaMRAM. First, the entries of $f(A)$ are only negligible when $d(i, j) \gg 1$. Indeed, there are matrices for which the sparsity pattern implies $d(i, j) = \mathcal{O}(1)$ for all or most i, j . An example is the arrow matrix which has nonzero entries along the first row, first column and the main diagonal. The sparsity pattern of the arrow matrix makes $d(i, j) \leq 2$, which can make SpaMRAM behave poorly as the a priori estimate from (1) tells us that $f(A)$ may have no entries that are significantly larger than others. A similar issue has been discussed in [35], where the authors assume that the matrix we want to approximate is d -distributed, that is, the matrix has at most d nonzero entries along any row or column. We will use a similar notion, which will be described in Subsection 2.2 to analyze SpaMRAM.

SpaMRAM requires k as an input. This may be straightforward from the sparsity pattern of A and the smoothness of f , but in some cases a good choice of k can be challenging. Moreover, when most rows of B are very sparse but a few are much denser, k would need to be large enough for the densest row to recover the whole matrix B . However, even with a small k , the sparse rows will be recovered.

Another downside is in the complexity of SpaMRAM for functions of sparse matrices. The cost of SpaMRAM is $\mathcal{O}(k \log n \cdot nnz(A) + n^2 \log n)$, which is dominated by the cost of solving n compressed sensing problems, $\mathcal{O}(n^2 \log n)$ but computing $f(A)I_n$ where I_n is the $n \times n$ identity matrix costs $\mathcal{O}(n \cdot nnz(A))$, which is cheaper if $nnz(A) = \mathcal{O}(n)$ and approximates $f(A)$ with higher accuracy. However, if matrix-vector products with $B = f(A)$ start dominating the cost of SpaMRAM, or when A is not so sparse with $nnz(A) > n \log n$, but $f(A)$ is approximately sparse, then there are merits in using SpaMRAM. It is also worth noting that storing the measurements $f(A)Y$ and only solving the compressed sensing problem to approximate specific row(s) when necessary would save computational time. This may be useful when we only need to approximate some subset of rows of $f(A)$, but we do not know the indices in advance.

2.2. Analysis of SpaMRAM (Algorithm 1)

The class of matrices B that satisfy the decay bound (5) can be difficult to analyze as the behaviour of $d(i, j)$ can give non-informative bounds in (5), i.e., there may be no entries that are small. For functions of matrices, the arrow matrix is an example, which has the sparsity pattern that makes $d_G(i, j) \leq 2$. When B has entries that all have similar magnitudes, SpaMRAM fails to compute a good approximation to B .

⁸ For theoretical guarantee, Y needs to be independent from B and \hat{B} [44], however the output of SpaMRAM, \hat{B} is dependent on Y . Nonetheless, the dependency is rather weak so $\delta_{RE}^{(S)}$ should be a good error estimate in practice. For theoretical guarantee, one can take, say 5 extra measurements for error estimation.

We now define a notion similar to d -distributed sparse matrices in [35] to analyze SpaMRAM. Let $d \in \mathbb{Z}^+$ be a distance parameter and define

$$k_d = \max_{1 \leq i \leq n} |\{j : d(i, j) \leq d\}|.$$

When $\mathbf{B} = f(\mathbf{A})$, k_d measures the maximum number of vertices that are at most distance d away from any given vertex. k_d also measures the combined sparsity of $\mathbf{I}_n, \mathbf{A}, \dots, \mathbf{A}^d$ because the cardinality of the union of the support locations in a fixed row in each of $\mathbf{I}_n, \mathbf{A}, \dots, \mathbf{A}^d$ will be at most k_d . If $k_d = \mathcal{O}(1)$ for d large enough such that f can be approximated by a polynomial of degree $d - 1$ on \mathbf{A} 's spectrum, then each row of $f(\mathbf{A})$ would only have $\mathcal{O}(1)$ dominating entries. Using the definition of k_d , we now analyze SpaMRAM in Theorem 1.

Theorem 1. (Algorithm 1) Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be a matrix satisfying

$$|B_{ij}| < C_B \lambda_B^{d(i,j)} \tag{7}$$

for constants $C_B > 0$, $1 > \lambda_B > 0$ and $d(i, j) \in \mathbb{R}_{\geq 0} \cup \{+\infty\}$ is some bivariate function that captures the magnitude of the entries of \mathbf{B} . Then the output of SpaMRAM, $\hat{\mathbf{B}}$, satisfies

$$\|\hat{\mathbf{B}} - \mathbf{B}\|_2 \leq 9C_B \left(n + \frac{n^{3/2}}{\sqrt{k_d}} \right) \lambda_B^{d+1} \tag{8}$$

where

$$k_d = \max_i |\{j : d(i, j) \leq d\}|$$

is the sparsity parameter in Algorithm 1 with the distance parameter d and the compressed sensing problem was solved row by row using NIHT until (4) was satisfied.

Proof. Define $\mathbf{E} = \hat{\mathbf{B}} - \mathbf{B}$. Let $E(i, \cdot)$ be the i th row of \mathbf{E} . Then from (4) we get

$$\|E(i, \cdot)\|_2 \leq 9 \left(\|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_2 + \frac{1}{\sqrt{k_d}} \|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_1 \right)$$

where \mathbf{b}_i is the i th row of \mathbf{B} and $\mathbf{b}_i^{k_d}$ is the best k_d -term approximation to \mathbf{b}_i . Since $\mathbf{b}_i^{k_d}$ is the best k_d -term approximation to \mathbf{b}_i , we have

$$\|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_{\max} \leq C_B \lambda_B^{d+1}$$

using (5). Therefore

$$\|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_2 \leq \sqrt{n - k_d} C_B \lambda_B^{d+1}, \quad \|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_1 \leq (n - k_d) C_B \lambda_B^{d+1}.$$

Therefore

$$\begin{aligned} \|\mathbf{E}\|_2 &\leq \sqrt{n} \max_{1 \leq i \leq n} \|E(i, \cdot)\|_2 \\ &\leq 9\sqrt{n} \max_{1 \leq i \leq n} \|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_2 + \frac{1}{\sqrt{k_d}} \|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_1 \\ &\leq 9\sqrt{n} \left(C_B \sqrt{n - k_d} \lambda_B^{d+1} + C_B \frac{n - k_d}{\sqrt{k_d}} \lambda_B^{d+1} \right) \\ &< 9C_B \left(n + \frac{n^{3/2}}{\sqrt{k_d}} \right) \lambda_B^{d+1}. \end{aligned}$$

□

Remark 1.

1. The bound in Theorem 1 can be pessimistic as we can replace the 2-norm bound by the Frobenius norm bound in (8). This can be shown by noting

$$\|\mathbf{E}\|_F \leq \sqrt{\sum_{1 \leq i \leq n} \|E(i, \cdot)\|_2^2} \leq \sqrt{n} \max_{1 \leq i \leq n} \|E(i, \cdot)\|_2 < 9C_B \left(n + \frac{n^{3/2}}{\sqrt{k_d}} \right) \lambda_B^{d+1}.$$

2. The bound in (8) can be improved if for each i , the nonzero entries of $\mathbf{b}_i - \mathbf{b}_i^{k_d}$ decay exponentially. This implies

$$\|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_1 \lesssim \lambda_B^{d+1}, \quad \|\mathbf{b}_i - \mathbf{b}_i^{k_d}\|_2 \lesssim \lambda_B^{d+1}$$

and the bound improves to

$$\|\mathbf{E}\|_2 \lesssim \sqrt{n} \left(1 + \frac{1}{\sqrt{k_d}} \right) \lambda_B^{d+1}.$$

3. In general, the constants C_B, λ_B and the bivariate function $d(i, j)$ are either unknown, implicit or too expensive to compute in applications. For matrix functions, [Theorem 1](#) relies on the decay bound by Benzi and Razouk [15], which contains constants that are implicit as they depend on the function f and the spectral property of the matrix A . Therefore, without a result that includes explicit constants, [Theorem 1](#) cannot serve as a priori estimate for [Algorithm 1](#). However, the analysis in [Theorem 1](#) can still serve as a heuristic for when [Algorithm 1](#) will be useful.

3. Approximation for functions of banded matrices

Let A be a diagonalizable banded matrix with upper bandwidth k_1 ($A_{ij} = 0$ if $j - i > k_1$) and lower bandwidth k_2 ($A_{ij} = 0$ if $i - j > k_2$). This is a more restrictive case than [Section 2](#), and so SpAMRAM can be used; however, in this case the sparsity pattern is predictable, and hence more efficient algorithms are possible.

Suppose that f is a function that is analytic in the region containing the spectrum of A . We then have that the entries of $f(A)$ decay exponentially away from the main diagonal. More specifically, we have that there exists constants $C_1, C_2 > 0$ and $0 < \lambda_1, \lambda_2 < 1$ dependent on A and f such that for $i < j$,

$$|[f(A)]_{ij}| < C_1 \lambda_1^{j-i} \tag{9}$$

and for $i \geq j$,

$$|[f(A)]_{ij}| < C_2 \lambda_2^{i-j}. \tag{10}$$

The proof of this result is based on polynomial approximation of f , together with the simple fact that powers of A are banded if A is. The exponents satisfy $\log \lambda_1^{-1} \propto k_1^{-1}$ and $\log \lambda_2^{-1} \propto k_2^{-1}$. For more details see [15,50,51]. Unlike the general sparse case, here the location of the large entries is known, i.e., near the main diagonal.

We will exploit the exponential off-diagonal decay given above in (9) and (10) to efficiently recover functions of banded matrices using matrix-vector products. We first design a way to recover banded matrices using matrix-vector products similarly to how banded Jacobian matrices were recovered in [23,24] and then use the strategy used for the banded matrices along with the exponential off-diagonal decay to approximate functions of banded matrices and matrices with rapid off-diagonal decay using matrix-vector products. The idea of using the same strategy for matrices with rapid off-diagonal decay was suggested in [32] and was later applied to functions of banded matrices in [22].

3.1. Strategy for recovering banded matrices

Let $A \in \mathbb{R}^{n \times n}$ be a banded matrix with upper bandwidth k_1 and lower bandwidth k_2 . A has at most $1 + k_1 + k_2$ nonzero entries in each row and column, which lie consecutively near the main diagonal. Therefore if we recover the $1 + k_1 + k_2$ consecutive nonzero entries near the main diagonal in each row or column then we can recover A . This motivates us to use a simple set of s vectors, which to the authors' knowledge, originates from [24] in the context of recovering matrices using matrix-vector products. The s vectors are

$$I_n^{(s)} = [I_s, I_s, \dots, I_s, I_s(:, 1 : \{n/s\} \cdot s)]^T \in \mathbb{R}^{n \times s} \tag{11}$$

where $s < n$ is a positive integer and $I_s(:, 1 : \{n/s\} \cdot s)$ is the first few columns of I_s to ensure $I_n^{(s)}$ has exactly n rows. This set of s vectors capture s consecutive nonzero entries of A in each row when right-multiplied to A if the other $(n - s)$ entries are all zeros. By choosing $s = 1 + k_1 + k_2$ we are able to recover every nonzero entry of A . We illustrate this idea with an example.

Example. Let $A_6 \in \mathbb{R}^{6 \times 6}$ be a banded matrix with upper bandwidth 2 and lower bandwidth 1. Then the matrix-vector products of A_6 with $I_6^{(1+2+1)}$ give

$$A_6 I_6^{(4)} = \begin{bmatrix} b_1 & c_1 & d_1 & 0 & 0 & 0 \\ a_1 & b_2 & c_2 & d_2 & 0 & 0 \\ 0 & a_2 & b_3 & c_3 & d_3 & 0 \\ 0 & 0 & a_3 & b_4 & c_4 & d_4 \\ 0 & 0 & 0 & a_4 & b_5 & c_5 \\ 0 & 0 & 0 & 0 & a_5 & b_6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} b_1 & c_1 & d_1 & 0 \\ a_1 & b_2 & c_2 & d_2 \\ d_3 & a_2 & b_3 & c_3 \\ c_4 & d_4 & a_3 & b_4 \\ b_5 & c_5 & 0 & a_4 \\ a_5 & b_6 & 0 & 0 \end{bmatrix}.$$

We observe in the example that every nonzero entry of A_6 in each row has been copied exactly once up to reordering onto the same row of $A_6 I_6^{(4)}$. Since we know A_6 has upper bandwidth 2 and lower bandwidth 1, we can reorganize the entries of $A_6 I_6^{(4)}$ to recover A_6 exactly. This idea can be generalized as follows [23,24].⁹

Proposition 1. *Let $A \in \mathbb{R}^{n \times n}$ be a banded matrix with upper bandwidth k_1 and lower bandwidth k_2 . Then A can be exactly recovered using $1 + k_1 + k_2$ matrix-vector products using $I_n^{(1+k_1+k_2)}$ as in (11).*

⁹ The work in [23,24] is based on partitioning/coloring of the adjacency graph of a banded matrix. In this work, we show the vectors more explicitly.

Proof. ¹⁰ Let $s = 1 + k_1 + k_2$ and for $1 \leq i \leq n$ and $1 \leq j \leq s$, define

$$\begin{aligned} t_{ij} &= \arg \min_{\substack{1 \leq j+st \leq n \\ t \in \mathbb{Z}}} \left| j + st - \left(i + (s-1) \left(\frac{1}{2} - \frac{k_1}{k_1 + k_2} \right) \right) \right| \\ &= \arg \min_{\substack{1 \leq j+st \leq n \\ t \in \mathbb{Z}}} \left| j + st - i - \frac{k_2 - k_1}{2} \right|. \end{aligned}$$

Equivalently, t_{ij} is the unique integer satisfying

$$k_1 \geq j + st_{ij} - i \geq -k_2, \text{ and } 1 \leq j + st_{ij} \leq n.$$

Then the (i, j) -entry of $\mathbf{A}\mathbf{I}_n^{(s)}$ is given by

$$[\mathbf{A}\mathbf{I}_n^{(s)}]_{ij} = \sum_{\substack{1 \leq j+st \leq n \\ t \in \mathbb{Z}}} A_{i,j+st} = A_{i,j+st_{ij}} \tag{12}$$

since $A_{k\ell} = 0$ for all $k_2 < k - \ell$ and $k_1 < \ell - k$.

Now we show that for all $1 \leq i, r \leq n$ with $k_1 \geq r - i \geq -k_2$, there is a unique entry of $\mathbf{A}\mathbf{I}_n^{(s)}$ equal to A_{ir} , i.e., we can use the entries of $\mathbf{A}\mathbf{I}_n^{(s)}$ to recover every nonzero entry of \mathbf{A} . Let $1 \leq i, r \leq n$ be integers such that $k_1 \geq r - i \geq -k_2$ and $1 \leq j \leq s$ be a unique integer such that $r = st^* + j$ for some integer t^* . Then

$$[\mathbf{A}\mathbf{I}_n^{(s)}]_{ij} = \sum_{\substack{1 \leq j+st \leq n \\ t \in \mathbb{Z}}} A_{i,j+st} = A_{i,j+st^*} = A_{ir}.$$

Therefore $t^* = t_{ij}$ and for each $k_1 \geq r - i \geq -k_2$, there is a unique entry $[\mathbf{A}\mathbf{I}_n^{(s)}]_{ij}$ equal to A_{ir} with $j = r - st_{ij}$. Therefore \mathbf{A} can be exactly recovered using $1 + k_1 + k_2$ matrix-vector products using $\mathbf{I}_n^{(1+k_1+k_2)}$. \square

We now extend this idea to approximate functions of banded matrices and matrices with off-diagonal decay.

3.2. Strategy for recovering functions of banded matrices

To accurately approximate $f(\mathbf{A})$ using matrix-vector products, we want to find a set of vectors when multiplied with $f(\mathbf{A})$, approximate the large entries of $f(\mathbf{A})$. We use the same set of vectors as for the banded case, namely $\mathbf{I}_n^{(s)}$. We assume that s is an odd integer $s = 2s_0 + 1$ throughout this section. We will see that this set of simple vectors can approximate functions of banded matrices. This is the same as the observation made in [32] and the work done later in [22], which are based on graph coloring of the adjacency graph of a matrix with off-diagonal decay.

In this section, we consider matrices $\mathbf{B} \in \mathbb{R}^{n \times n}$ with exponential¹¹ off-diagonal decay with $\mathbf{B} = f(\mathbf{A})$ for banded \mathbf{A} being the special case. For simplicity we consider matrices with the same decay rate on both sides of the diagonal, that is, \mathbf{B} satisfies

$$|\mathbf{B}_{ij}| < C\lambda^{|i-j|} \tag{13}$$

for some constant $C > 0$ and $0 < \lambda < 1$. The results in this section can easily be extended to the more general case, that is, the class of matrices with entries satisfying (9) and (10), i.e., the decay rate along the two sides of the diagonal are different. See Subsection 3.2.4 for a further discussion.

The key idea is to revisit (12) in the proof of Proposition 1. For $1 \leq i \leq n$ and $1 \leq j \leq s$, recall that t_{ij} is defined by

$$t_{ij} = \arg \min_{\substack{1 \leq j+st \leq n \\ t \in \mathbb{Z}}} |j + st - i|$$

when the decay rate is the same on both sides of the diagonal.¹² t_{ij} is an integer such that i and $j + st_{ij}$ are as close as possible, i.e., the index closest to the main diagonal. (12) gives us

$$[\mathbf{B}\mathbf{I}_n^{(s)}]_{ij} = \sum_{\substack{1 \leq j+st \leq n \\ t \in \mathbb{Z}}} B_{i,j+st}.$$

By (13), the sum is dominated by $B_{i,j+st_{ij}}$ and the remaining terms decay exponentially. Therefore we can view this as

$$[\mathbf{B}\mathbf{I}_n^{(s)}]_{ij} = \underbrace{B_{i,j+st_{ij}}}_{\text{Dominant term}} + \underbrace{\sum_{\substack{1 \leq j+st \leq n \\ t \neq t_{ij} \in \mathbb{Z}}} B_{i,j+st}}_{\text{Sum of exponentially decaying terms}}$$

¹⁰ The proof of Proposition 1 can be simplified by arguing that we need at most $1 + k_1 + k_2$ colors to color the columns of \mathbf{A} such that all the columns of the same color have disjoint support. Now, probing all the columns of the same color simultaneously yields the result. In this work, we provide a lengthier proof of Proposition 1 to make certain observations that lay the groundwork for our strategy to recover functions of banded matrices in Section 3.2.

¹¹ The analysis can be carried out similarly for any other type of decay, for example, algebraic.

¹² When the decay rate is different on each side of the diagonal we can carry out the analysis in a similar way by defining t_{ij} differently. See Section 3.2.4.

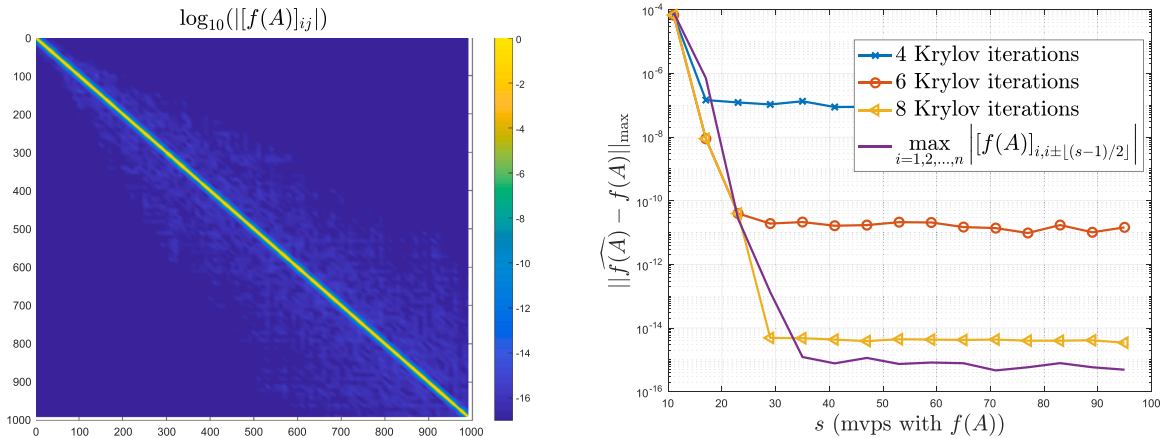


Fig. 1. f is the exponential function e^x and $A \in \mathbb{R}^{1000 \times 1000}$ is a symmetric 2-banded matrix. The left figure shows the exponential decay in the entries of $f(A)$ and the right figure shows that the error analysis in (16) captures the decay rate until the error is dominated by the error from evaluating matrix-vector products using the Krylov method. The stagnation of the $\max_{i=1,2,\dots,n} |[f(A)]_{i,i \pm (s-1)/2}|$ curve is caused by machine precision. The exponential decay in the max-norm error will be used to derive the 2-norm error bound in Section 3.2.3.

which motivates us to use $[B I_n^{(s)}]_{ij}$ as an approximation for $B_{i,j+st_{ij}}$.

Now create a matrix \hat{B} by

$$\hat{B}_{i,j+st_{ij}} = [B I_n^{(s)}]_{ij} = \sum_{\substack{1 \leq j+st \leq n \\ t \in \mathbb{Z}}} B_{i,j+st} \tag{14}$$

for $|j + st_{ij} - i| \leq s_0$ where $1 \leq i \leq n$ and $1 \leq j \leq s$ and zero everywhere else. Notice that by construction, \hat{B} is an s_0 -banded matrix. Roughly, the nonzero entries of the i th row of \hat{B} will be, with some small error, the largest entries of the i th row of B up to reordering. The entries of \hat{B} approximate those of B well because

$$|\hat{B}_{i,j} - B_{i,j}| = \left| \sum_{\substack{1 \leq j+st \leq n \\ t \neq 0 \in \mathbb{Z}}} B_{i,j+st} \right| \approx C \lambda^{s_0} \tag{15}$$

for $|j - i| \leq s_0$ and $1 \leq i, j \leq n$. The largest error incurred from (15) is at most $\mathcal{O}(|B_{i,i \pm s_0}|)$ for the i th row and the zero entries of \hat{B} , i.e. the entries not covered by (15), incur error of at most $\mathcal{O}(|B_{i,i \pm s_0}|)$ also. Therefore,

$$\|\hat{B} - B\|_{\max} = \mathcal{O}\left(\max_{1 \leq i \leq n} |B_{i,i \pm s_0}|\right). \tag{16}$$

This qualitative max-norm error bound of the approximant \hat{B} will be used to derive the 2-norm error bound in Section 3.2.3. The precise details will be laid out in Theorem 2.

Fig. 1 illustrates the error analysis given above for $B = f(A)$ where f is the exponential function e^x and $A \in \mathbb{R}^{1000 \times 1000}$ is a symmetric 2-banded matrix with entries drawn i.i.d. from $\mathcal{N}(0, n^{-1})$. This matrix has real eigenvalues with $\lambda(A) \in [-0.1919, 0.2011]$. The matrix-vector products $f(A)I_n^{(s)}$ were evaluated using the polynomial Krylov method with 4, 6 and 8 iterations (corresponding to polynomial approximation of degrees 3, 5, 7 respectively), which give approximations to $f(A)I_n^{(s)}$ of varying accuracy. In Fig. 1a, we see the exponential decay in the entries of $f(A)$ and in Fig. 1b, we see that (16) captures the decay rate until the error is dominated by the error from computing matrix-vector products using the Krylov method. The stagnation of the $\max_{i=1,2,\dots,n} |[f(A)]_{i,i \pm (s-1)/2}|$ curve is caused by machine precision.

We now devise an algorithm based on the above analysis to approximate matrices that satisfy the exponential off-diagonal decay. The algorithm is given in Algorithm 2, which we call BaMRAM (Banded Matrix Recovery Algorithm using Matvecs).

3.2.1. Complexity of BaMRAM (Algorithm 2)

The complexity of BaMRAM is s matrix-vector products with B . In the case $B = f(A)$ for a k -banded matrix A , we can evaluate $f(A)I_n^{(s)}$ using Krylov method. Assuming that f is analytic in a region containing the eigenvalues of A , the polynomial Krylov method converges (to sufficient accuracy) in $\mathcal{O}(1)$ iterations and $s = 2ck + 1$ for some constant c , then the complexity of BaMRAM is $\mathcal{O}(snk) = \mathcal{O}(nk^2)$.

If we only have access to the matrix using its transpose, i.e. $x \mapsto A^T x$ (or $x \mapsto B^T x$), then we can take $f(A^T)I_n^{(s)}$ (or $B^T I_n^{(s)}$) and use the rows of $f(A^T)I_n^{(s)}$ (or $B^T I_n^{(s)}$) to recover the columns of $f(A)$ (or B).

Algorithm 2 Banded Matrix Recovery Algorithm using Matvecs (BaMRRAM).

Require: Matrix-vector product function handle $\text{mvp} : \mathbf{x} \mapsto \mathbf{B}\mathbf{x}$;

$s = 2s_0 + 1$ the number of matrix-vector products; (rec. $s_0 = 10k$ for functions of k -banded matrices)

Ensure: $\hat{\mathbf{B}}$, s_0 -banded approximation to \mathbf{B} ;

- 1: Compute matrix-vector products $\mathbf{B}_s = \text{mvp}(\mathbf{I}_n^{(s)}) \in \mathbb{R}^{n \times s}$, where $\mathbf{I}_n^{(s)}$ is as in (11)
- 2: Set the entries of $\hat{\mathbf{B}}$ using \mathbf{B}_s and (14)

3.2.2. Error estimate for BaMRRAM

The qualitative behaviour of BaMRRAM is better understood than SpaMRRAM. Since \mathbf{B} is expected to have an off-diagonal decay, an increase in s would improve the quality of the estimate $\hat{\mathbf{B}}$ in BaMRRAM. In the context of matrix functions, the speed of the off-diagonal decay of $f(\mathbf{A})$ is positively correlated to the smoothness of f and negatively correlated to the size of the bandwidth of \mathbf{A} . However, unlike SpaMRRAM, BaMRRAM estimates the entries of \mathbf{B} using the entries of the measurements \mathbf{B}_s . Therefore, by construction, $\|\hat{\mathbf{B}}\mathbf{I}_n^{(s)} - \mathbf{B}_s\|_2 = 0$ and we cannot use the same strategy as SpaMRRAM in Section 2.1.3. However, we can estimate the error by making a small number of extra Gaussian measurements of \mathbf{B} . We can estimate the error using

$$\delta_{RE}^{(B)} := \frac{\|\hat{\mathbf{B}}\mathbf{X} - \text{mvp}(\mathbf{X})\|_2}{\|\text{mvp}(\mathbf{X})\|_2}$$

where $\mathbf{X} \in \mathbb{R}^{n \times s}$ is a Gaussian sensing operator. As previously mentioned, making a small constant number of Gaussian measurements, here we take 5, usually suffices in practice; see [44,45]. We recommend computing $\delta_{RE}^{(B)}$ when the decay behaviour of \mathbf{B} is unknown or in the context of matrix functions, when the smoothness of f or the bandwidth of the underlying matrix \mathbf{A} is unknown. Later in Section 4, we illustrate the error estimate $\delta_{RE}^{(B)}$ in a numerical experiment; see Fig. 2.

3.2.3. Analysis of BaMRRAM(Algorithm 2)

Here we give theoretical guarantees for BaMRRAM. The analysis for BaMRRAM is easier and it gives a stronger bound than SpaMRRAM as the precise location of the dominant entries and the qualitative behaviour of the decay is known. The proof is motivated by Proposition 3.2 in [15].

Theorem 2. (Algorithm 2) Let \mathbf{B} be a matrix satisfying

$$|\mathbf{B}_{ij}| < C\lambda^{|i-j|} \tag{17}$$

where $C > 0$ and $0 < \lambda < 1$ are constants. Then the output of BaMRRAM, $\hat{\mathbf{B}}$, satisfies

$$\|\hat{\mathbf{B}} - \mathbf{B}\|_2 \leq \frac{4C\lambda}{1-\lambda} \lambda^{s_0}$$

where $s = 2s_0 + 1$.

Proof. Define $\mathbf{E} = \hat{\mathbf{B}} - \mathbf{B}$. We get

$$\|\hat{\mathbf{B}} - \mathbf{B}\|_2 = \|\mathbf{E}\|_2 \leq \sqrt{\|\mathbf{E}\|_1 \|\mathbf{E}\|_\infty}$$

using a special case of Hölder’s inequality [12, §2.3.3]. We now bound $\|\mathbf{E}\|_1$ and $\|\mathbf{E}\|_\infty$ using (13). By the construction of $\hat{\mathbf{B}}$, the i th row of \mathbf{E} satisfies

$$\begin{aligned} \|E(i, :)\|_1 &\leq \sum_{\ell=-s_0}^{s_0} |\hat{\mathbf{B}}_{i,i+\ell} - \mathbf{B}_{i,i+\ell}| + \sum_{\ell=s_0+1}^{n-i} |\mathbf{B}_{i,i+\ell}| + \sum_{\ell=s_0+1}^{i-1} |\mathbf{B}_{i,i-\ell}| \\ &\leq 2 \sum_{\ell=s_0+1}^{n-i} |\mathbf{B}_{i,i+\ell}| + 2 \sum_{\ell=s_0+1}^{i-1} |\mathbf{B}_{i,i-\ell}| \\ &< 2C \sum_{\ell=s_0+1}^{n-i} \lambda^\ell + 2C \sum_{\ell=s_0+1}^{i-1} \lambda^\ell \\ &< 4C\lambda^{s_0+1} \sum_{\ell=0}^{\infty} \lambda^\ell = \frac{4C\lambda}{1-\lambda} \lambda^{s_0}. \end{aligned}$$

Similarly, the j th column of \mathbf{E} satisfies

$$\|E(:, j)\|_1 < \frac{4C\lambda}{1-\lambda} \lambda^{s_0}.$$

Therefore

$$\|\mathbf{E}\|_1 = \max_{1 \leq j \leq n} \|E(:, j)\|_1 < \frac{4C\lambda}{1-\lambda} \lambda^{s_0}, \quad \|\mathbf{E}\|_\infty = \max_{1 \leq i \leq n} \|E(i, :)\|_1 < \frac{4C\lambda}{1-\lambda} \lambda^{s_0}.$$

We conclude that

$$\|\widehat{\mathbf{B}} - \mathbf{B}\|_2 \leq \frac{4C\lambda}{1-\lambda} \lambda^{s_0}.$$

□

Remark 2.

1. When $\mathbf{B} = f(\mathbf{A})$ for a k -banded matrix \mathbf{A} , we can use $s = 2ck + 1$ for some integer constant $c > 0$. This choice of s is motivated by the fact that if f is analytic then $f(\mathbf{A})$ can be well approximated by a low degree polynomial $p(\mathbf{A})$ which has bandwidth $\deg(p)k$. It is also useful to note the dependence of the decay factor λ on the bandwidth. If the entries of a 1-banded matrix decay away from the main diagonal at rate λ , then for a k -banded matrix with comparable spectral properties, the decay rate becomes $\lambda^{1/k}$ [20]. This behaviour aligns with the graph distance $d(i, j)$ for general sparse matrices discussed in Section 2, and the choice $s = 2ck + 1$.
2. In Theorem 2, if $\lambda \leq 0.9$ then for ϵ accuracy it suffices to take

$$s_0 \geq \frac{\log(36C) - \log \epsilon}{\log(\lambda^{-1})}.$$

3. In many cases, the constants C, λ and s_0 are either unknown, implicit or too expensive to compute in applications. For matrix functions, Theorem 2 relies on the off-diagonal decay bound in [15], which contains constants that are implicit as they depend on the function f and the spectral property of the matrix \mathbf{A} . Therefore, without decay bounds that include explicit constants, Theorem 2 cannot serve as a priori estimate for Algorithm 2. However, the analysis in Theorem 2 can still serve as a heuristic for when Algorithm 2 will be useful.

3.2.4. Different decay rates

A similar analysis can be performed if the decay rate is different on each side of the diagonal, for example, if the matrix satisfies (9) and (10). The argument is the same with a different definition for t_{ij} , which is

$$t_{ij} = \arg \min_{\substack{1 \leq j+st \leq n \\ t \in \mathbb{Z}}} \left| j + st - i - \frac{s-1}{2} \frac{\log(\lambda_1^{-1}) - \log(\lambda_2^{-1})}{\log(\lambda_1^{-1}) + \log(\lambda_2^{-1})} \right| \tag{18}$$

where λ_1 is the decay rate above the diagonal and λ_2 is that below the diagonal. This expression for t_{ij} accounts for the different decay rates and coincides with our analysis in the previous section when $\lambda_1 = \lambda_2$. The algorithm will be the same as BaMRAM with t_{ij} defined as (18). For $s = \lfloor c/\log(\lambda_1^{-1}) \rfloor + \lfloor c/\log(\lambda_2^{-1}) \rfloor + 1$ for some constant $c > 0$, the algorithm will output $\widehat{\mathbf{B}}$ with upper bandwidth $\lfloor c/\log(\lambda_1^{-1}) \rfloor$ and lower bandwidth $\lfloor c/\log(\lambda_2^{-1}) \rfloor$ and $\widehat{\mathbf{B}}$ will satisfy

$$\|\widehat{\mathbf{B}} - \mathbf{B}\|_2 \lesssim e^{-c}$$

which can easily be shown by following the proof of Theorem 2.

3.3. Extensions to kronecker sum

Some of the results for banded matrices and matrices with exponential off-diagonal decay can be naturally extended to Kronecker sums, which arises, for example in the finite difference discretization of the two-dimensional Laplace operator [52]. In particular, decay bounds for functions of Hermitian matrices with Kronecker structure have been studied in [50]. Let $\mathbf{A}^{(1)}, \mathbf{A}^{(2)} \in \mathbb{R}^{n \times n}$. The Kronecker sum $\mathcal{A} \in \mathbb{R}^{n^2 \times n^2}$ of $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ is defined by

$$\mathcal{A} = \mathbf{A}^{(1)} \oplus \mathbf{A}^{(2)} = \mathbf{A}^{(1)} \otimes \mathbf{I}_n + \mathbf{I}_n \otimes \mathbf{A}^{(2)}$$

where \otimes is the Kronecker product, defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \dots & A_{1n}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \dots & A_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}\mathbf{B} & A_{n2}\mathbf{B} & \dots & A_{nn}\mathbf{B} \end{bmatrix} \in \mathbb{R}^{n^2 \times n^2}.$$

Suppose $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are both banded matrices. Then $\mathbf{I}_n \otimes \mathbf{A}^{(2)} \in \mathbb{R}^{n^2 \times n^2}$ is a banded matrix with the same bandwidth as $\mathbf{A}^{(2)}$ and $\mathbf{A}^{(1)} \otimes \mathbf{I}_n \in \mathbb{R}^{n^2 \times n^2}$ is permutation equivalent to a banded matrix with the same bandwidth as $\mathbf{A}^{(1)}$ because $\mathbf{P}(\mathbf{A}^{(1)} \otimes \mathbf{I}_n)\mathbf{P}^T = \mathbf{I}_n \otimes \mathbf{A}^{(1)}$ where \mathbf{P} is the perfect shuffle matrix [53]. With this observation, we can use the recovery technique used for banded matrices in Section 3.1 to recover \mathcal{A} for banded matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$. Suppose that $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ have bandwidth $k^{(1)}$ and $k^{(2)}$ respectively. Notice that to recover \mathcal{A} , we need to recover the off-diagonal entries of $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ to recover the off-diagonal entries of \mathcal{A} and recover $A_{ii}^{(1)} + A_{jj}^{(2)}$ for all i, j to recover the diagonal entries of \mathcal{A} . We use the following $(2 + 2k^{(1)} + 2k^{(2)})$ matrix-vector products

$$\left[\mathbf{P}^T \begin{bmatrix} \mathbf{I}_n^{(1+2k^{(1)})} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_n^{(1+2k^{(2)})} \\ \mathbf{0} \end{bmatrix} \right] \in \mathbb{R}^{n^2 \times (2+2k^{(1)}+2k^{(2)})}$$

to recover \mathcal{A} . Since the top left $n \times n$ block of \mathcal{A} equals $A_{11}^{(1)}I_n + \mathbf{A}^{(2)}$, which is $k^{(2)}$ -banded, the first n rows of the matrix-vector product with $\begin{bmatrix} I_n^{(1+2k^{(2)})} \\ \mathbf{0} \end{bmatrix}$ recover $A_{11}^{(1)}I_n + \mathbf{A}^{(2)}$ exactly (Proposition 1), which in turn recovers all the off-diagonal entries of $\mathbf{A}^{(2)}$. Similarly, $\mathbf{P}^T \begin{bmatrix} I_n^{(1+2k^{(1)})} \\ \mathbf{0} \end{bmatrix}$ recovers $\mathbf{A}^{(1)} + A_{11}^{(2)}I_n$ exactly using $\mathbf{P}(\mathbf{A}^{(1)} \otimes I_n)\mathbf{P}^T = I_n \otimes \mathbf{A}^{(1)}$. Therefore we get all the off-diagonal entries of \mathcal{A} from recovering $A_{11}^{(1)}I_n + \mathbf{A}^{(2)}$ and $\mathbf{A}^{(1)} + A_{11}^{(2)}I_n$. The diagonal entries of \mathcal{A} can be recovered from the diagonal entries of $A_{11}^{(1)}I_n + \mathbf{A}^{(2)}$ and $\mathbf{A}^{(1)} + A_{11}^{(2)}I_n$ by noting

$$A_{ii}^{(1)} + A_{jj}^{(2)} = (A_{ii}^{(1)} + A_{11}^{(2)}) + (A_{11}^{(1)} + A_{jj}^{(2)}) - (A_{11}^{(1)} + A_{11}^{(2)})$$

for all i, j .

For matrices with rapid off-diagonal decay, we can extend the above idea in certain cases. Focusing on the matrix exponential, we have a special relation [11, §10]

$$\exp(\mathbf{A}^{(1)} \oplus \mathbf{A}^{(2)}) = \exp(\mathbf{A}^{(1)}) \otimes \exp(\mathbf{A}^{(2)}).$$

When $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are both banded matrices we can use the same procedure above by applying matrix-vector products with

$$\left[\mathbf{P}^T \begin{bmatrix} I_n^{(s_1)} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} I_n^{(s_2)} \\ \mathbf{0} \end{bmatrix} \right] \in \mathbb{R}^{n^2 \times (s_1 + s_2)}.$$

for sufficiently large s_1 and s_2 that captures the dominant entries of $\exp(\mathbf{A}^{(1)})$ and $\exp(\mathbf{A}^{(2)})$ near the main diagonal respectively. This extension is natural and analogous to the extension from recovering banded matrices to recovering matrices with exponential off-diagonal decay earlier in this section.

The strategy for recovering banded matrices and approximating matrices with rapid off-diagonal decay can also be used as laid out above in a similar setting involving Kronecker product or sum. Examples include $f(\mathbf{I} \otimes \mathbf{A}) = \mathbf{I} \otimes f(\mathbf{A})$ and the formulae for the cosine and the sine of the Kronecker sum between two matrices [11, §12].

4. Numerical illustrations

Here we illustrate BaMRAM and SpaMRAM using numerical experiments. In all experiments, we use 20 iterations of a polynomial Krylov method and 50 sample points for the contour integration from method 2 of [26] to evaluate the matrix-vector products with $f(\mathbf{A})$. The contour integration requires the knowledge of the largest and smallest eigenvalues of \mathbf{A} . We assume that these eigenvalues are known a priori. In all of the plots, the x -axis denotes the number of matrix-vector products s , i.e. $\mathbf{x} \mapsto f(\mathbf{A})\mathbf{x}$. Note that this refers to the matrix-vector products with $f(\mathbf{A})$, not with \mathbf{A} . If computing $\mathbf{x} \mapsto f(\mathbf{A})\mathbf{x}$ requires m matrix-vector products with \mathbf{A} , then approximating $f(\mathbf{A})$ in our experiments requires a total of ms matrix-vector products with \mathbf{A} . The y -axis denotes the relative error in the 2-norm with respect to $f(\mathbf{A})$ obtained by dense arithmetic in MATLAB, i.e., `expm`, `sqrtm` and `logm`. Here, we plot the relative error $\|f(\mathbf{A}) - \tilde{f}(\mathbf{A})\|_2 / \|f(\mathbf{A})\|_2$ instead of the absolute error $\|f(\mathbf{A}) - \tilde{f}(\mathbf{A})\|_2$ to provide a clearer, scale-invariant view of our numerical results. For SpaMRAM, the sparsity parameter k is chosen to be 8 times smaller than the number of matrix-vector products, i.e., $s = 8k$ and the sensing operator \mathbf{Y} is Gaussian. The experiments were conducted in MATLAB version 2021a using double precision arithmetic.

4.1. Synthetic examples

We illustrate synthetic examples for BaMRAM and SpaMRAM using the functions e^x , $\sqrt{1+x}$ and $\log(1+x)$. We consider a synthetic banded matrix and a synthetic sparse matrix given below.

- *Banded case:* symmetric 2-banded matrix $\mathbf{A}_B \in \mathbb{R}^{1024 \times 1024}$ with its nonzero entries drawn i.i.d. from $\mathcal{N}(0, 1)$. \mathbf{A}_B has been rescaled to $\|\mathbf{A}_B\|_2 = 0.5$.
- *General sparse case:* symmetric sparse matrix $\mathbf{A}_S \in \mathbb{R}^{1024 \times 1024}$ created using the MATLAB command `sprandsym` with density $1/1024$. \mathbf{A}_S has been rescaled to $\|\mathbf{A}_S\|_2 = 0.5$.

Symmetry was enforced to ensure that the eigenvalues are real, and \mathbf{A}_B and \mathbf{A}_S have been rescaled to ensure that their eigenvalues are sufficiently away from -1 , which is where $\sqrt{1+x}$ and $\log(1+x)$ have singularities. This was done to ensure that the matrix-vector products $f(\mathbf{A})\mathbf{b}$ can be computed with high accuracy with 20 Krylov steps.

Fig. 2 shows the accuracy as well as how well the error estimates $\delta_{RE}^{(S)}$ and $\delta_{RE}^{(B)}$ describe the behaviour of the two algorithms. For the banded case in Fig. 2a, we see that the error decays exponentially until the error is dominated by either the machine precision or the error from computing the matrix-vector products. This exponential decay rate is consistent with Theorem 2. The dotted lines indicate the error estimate $\delta_{RE}^{(B)}$. We observe that $\delta_{RE}^{(B)}$, which makes 5 extra Gaussian measurements, estimate the error well. Therefore, we recommend computing $\delta_{RE}^{(B)}$ in practice, especially in cases when either the bandwidth of the underlying matrix or the matrix function is unknown.

For the general sparse case in Fig. 2b, the error decays quickly with the number of nonzero entries recovered in each row. The sparsity pattern influences the decay rate of SpaMRAM, making the behaviour of the decay rate more unpredictable and irregular.

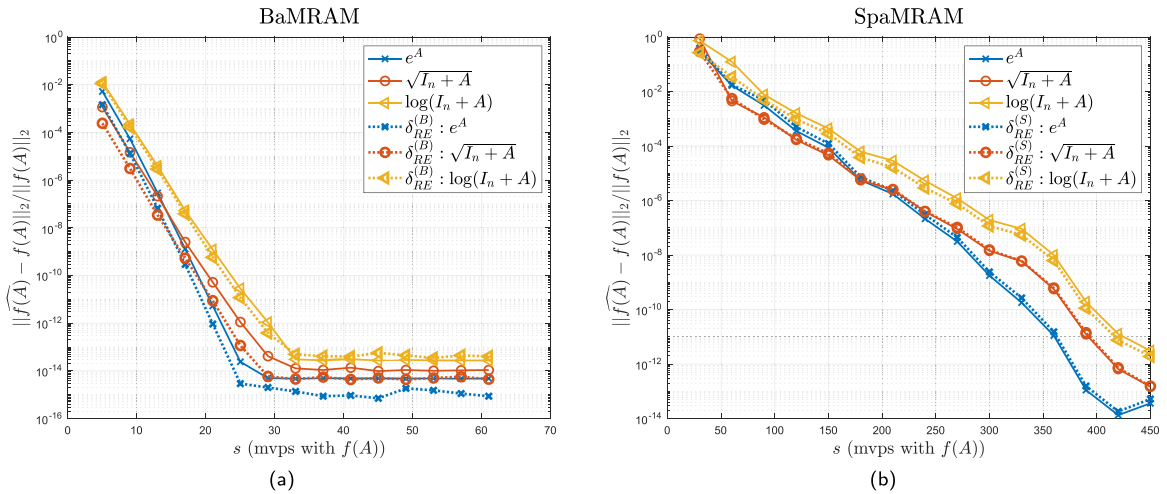


Fig. 2. The accuracy of BaMRRAM and SpaMRRAM using two 1024×1024 synthetic symmetric matrices: a random 2-banded matrix (left) and a random sparse matrix with nonzero density $1/1024$ (right). The dotted lines are the error estimates $\delta_{RE}^{(S)}$ and $\delta_{RE}^{(B)}$ discussed in Sections 2.1.3 and 3.2.2 respectively. The error estimates can be used to determine the behaviour of SpaMRRAM and BaMRRAM.

This has been shown partially in Theorem 1 where the bound depends on the sparsity pattern or more specifically the cardinality of the union of the combined support locations of I_n, A, \dots, A^d in each row for d large enough. Although the behavior of SpaMRRAM is irregular, $\delta_{RE}^{(S)}$ serves as an excellent error estimate for how SpaMRRAM performs.

Finally, we note that there are regimes in which the proposed method does not convincingly amortize its computational cost. For example, in Fig. 2b for the 1024×1024 problem, the SpaMRRAM requires 55 matrix-vector products with $f(A)$ —corresponding to $55 \times 19 = 1045$ matrix-vector products with A —to achieve only 10^{-3} accuracy. In contrast, naive dense methods that explicitly form the matrix and compute a full eigenvalue decomposition achieve full accuracy using fewer matrix-vector products with A . This highlights that the advantages of the proposed approach are most pronounced in large-scale settings where forming the matrix is infeasible, rather than in moderate-size problems where dense methods remain competitive. See Section 4.4 for experiments at larger scale.

4.2. Toy examples

In this section, we manufacture a matrix A such that $f(A)$ has a sparse structure and use our algorithms to approximate $f(A)$ using matrix-vector products with A only. We explore two functions, namely \sqrt{A} and $\log(A)$. For the square root function, we set \sqrt{A} equal to a sparse positive definite matrix and use our algorithms to recover \sqrt{A} via matrix-vector products with $A = (\sqrt{A})^2$ (which here is not very sparse). We follow a similar procedure for the matrix logarithm using the formula $\exp(\log(A)) = A$. The aim of this experiment is to show that if $f(A)$ is sparse then we can efficiently recover $f(A)$ to high accuracy using our algorithms. We choose a sparse positive definite matrix and a banded positive definite matrix from the SuiteSparse Matrix Collection [54].

- `gr_30_30`: 900×900 31-banded symmetric positive definite matrix with 7744 nonzero entries. This matrix has eigenvalues with $\lambda(A) \in [0.06, 11.96]$.
- `Trefethen_700`: 700×700 sparse symmetric positive definite matrix with 12654 nonzero entries and at most 19 nonzero entries per row and column. This matrix has eigenvalues with $\lambda(A) \in [1.12, 5280]$.

In Fig. 3, after an initial period of low accuracy—caused by the unrecovered entries of both matrices being $\mathcal{O}(1)$ —we obtain high accuracy as the number of nonzero entries recovered per row increases sufficiently for $f(A)$, which is either sparse or banded, to be recovered exactly. In Fig. 3a, the matrix `gr_30_30` is recovered to high accuracy using about 66 matrix-vector products, which is consistent with the fact that `gr_30_30` is 31-banded, and hence has at most 63 nonzero entries per row and column. This corresponds to a total of $66 \times 19 = 1254$ matrix-vector products with A . In Fig. 3b, the matrix `Trefethen_700` has been recovered to high accuracy using about 180 matrix-vector products. This is fairly consistent with the fact that with $s = 180$ matrix-vector products, we recover about 22 nonzero entries in each row and `Trefethen_700` has at most 19 nonzero entries per row. This corresponds to $180 \times 19 = 3420$ matrix-vector products with A .

4.3. Sparse dataset examples

We now illustrate BaMRRAM and SpaMRRAM using the SuiteSparse Matrix Collection [54], using the four matrices below. The first three matrices are sparse and the last matrix is banded.

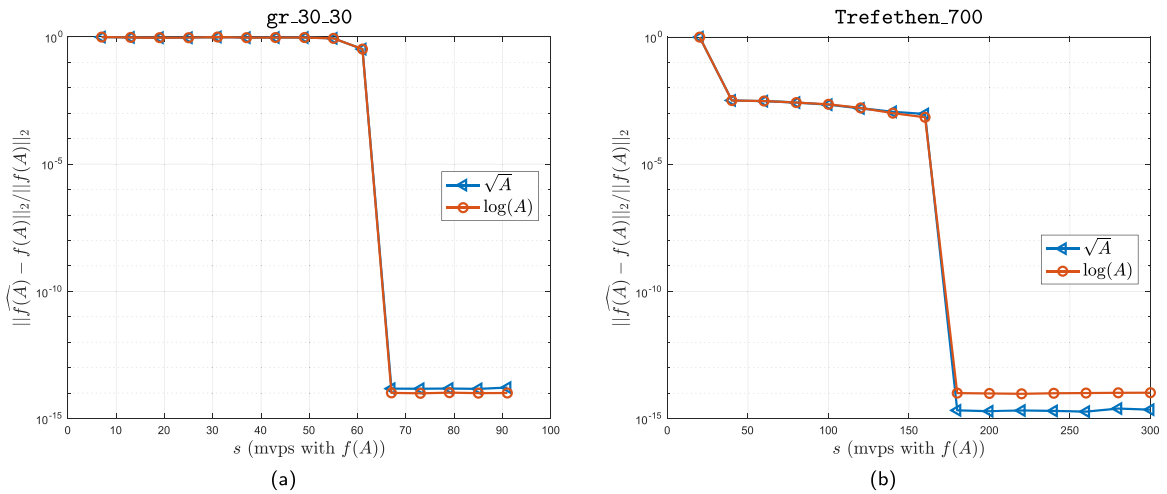


Fig. 3. The accuracy of BaMRAM and SpaMRAM for functions of matrices when $f(A)$ itself is banded or sparse.

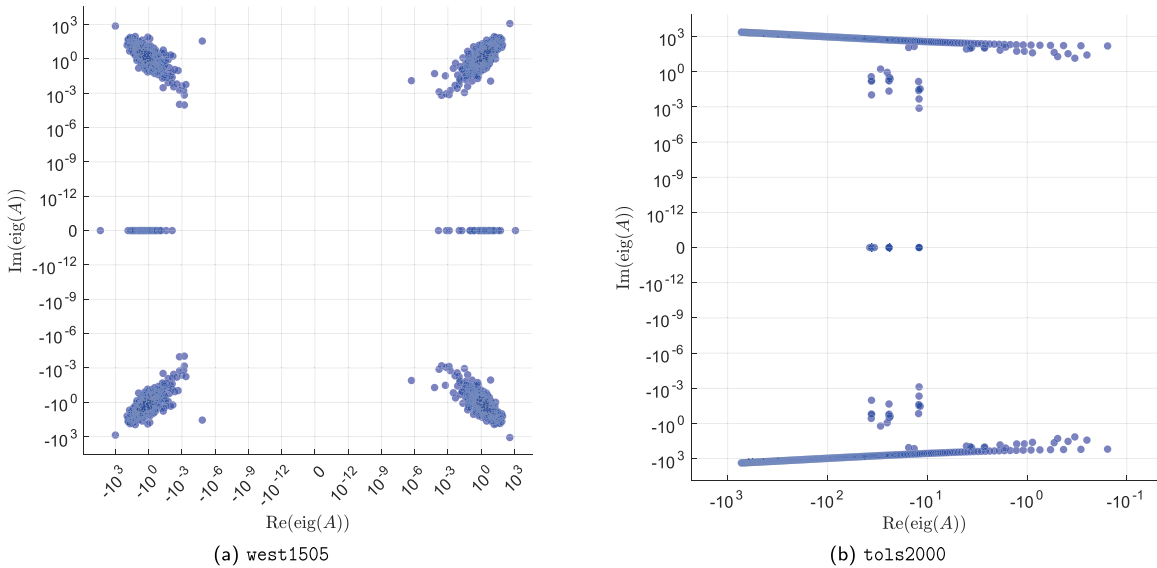


Fig. 4. Eigenvalue spectrum of west1505 and to1s2000.

- west1505: 1505×1505 sparse matrix with 5414 nonzero entries and at most 12 nonzero entries per row and at most 27 nonzero entries per column. This matrix has complex eigenvalues with $|\lambda(A)| \in [1.41 \cdot 10^{-4}, 2.29 \cdot 10^4]$.
- CSphd: 1882×1882 sparse matrix with 1740 nonzero entries and at most 45 nonzero entries per row and at most 3 nonzero entries per column. This matrix has all of its eigenvalues equal to zero.
- to1s2000: 2000×2000 sparse matrix with 5184 nonzero entries and at most 90 nonzero entries per row and at most 22 nonzero entries per column. This matrix has complex eigenvalues with $|\lambda(A)| \in [11.79, 2442]$.
- mhd3200b: 3200×3200 18-banded symmetric positive definite matrix with 18316 nonzero entries. This matrix has eigenvalues with $\lambda(A) \in [1.37 \cdot 10^{-13}, 2.2]$.

The eigenvalue spectrum of west1505 and to1s2000 are depicted in Fig. 4. For the sparse matrices west1505, CSphd and to1s2000, we use SpaMRAM to recover their matrix exponential. For the banded matrix mhd3200b, we use BaMRAM to recover its matrix exponential, matrix square root and matrix logarithm. The results are illustrated in Fig. 5.

In Fig. 5, we observe that our algorithms are able to approximate $f(A)$ with good accuracy. In Fig. 5a, we notice that the error decreases exponentially until the error stagnates at some point. The stagnation is due to the error incurred in the computation of matrix-vector products $f(A)b$, and can be improved by taking more than 20 Krylov steps. For banded matrices, we consistently obtain exponential convergence rate whose convergence rate is dictated by the smoothness of f . In Fig. 5b, we note that the error decreases at a rapid speed until it stagnates due to machine precision for CSphd or slows down due to the sparsity pattern in the case of west1505 and to1s2000. As observed in prior experiments, the behavior is more irregular when the matrix has a more general sparsity pattern.

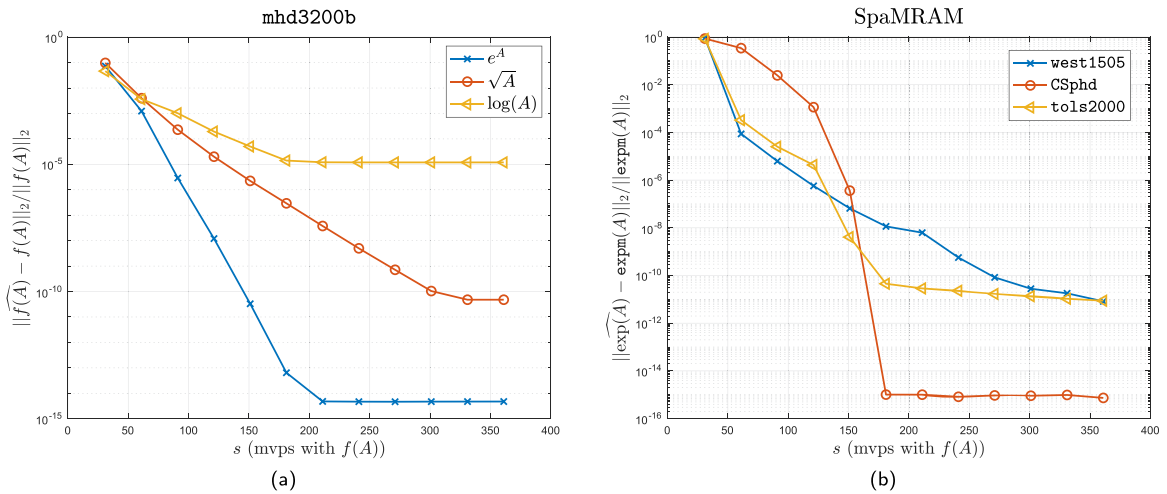


Fig. 5. The accuracy of BaMRAM (left) and SpaMRAM (right) using sparse matrices from [54].

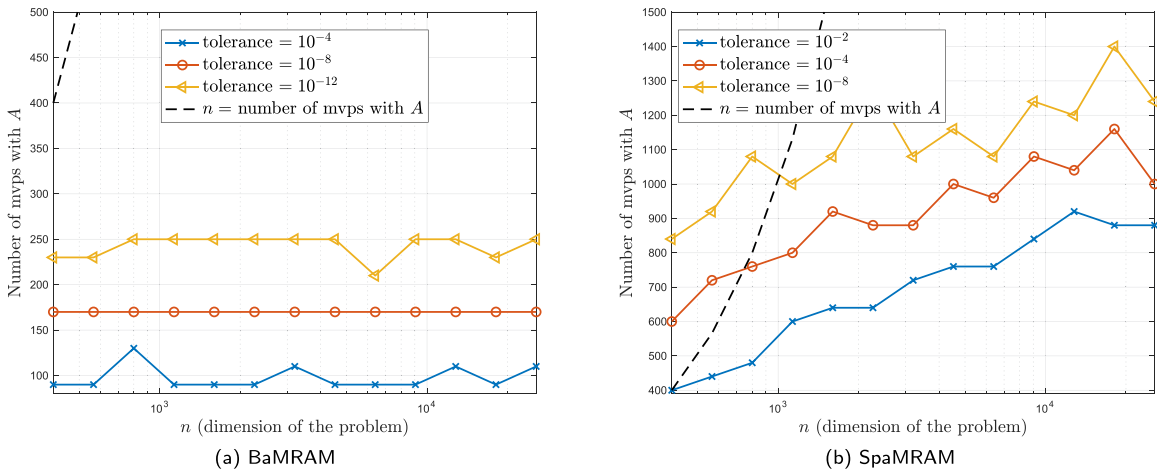


Fig. 6. Scalability of BaMRAM (left) and SpaMRAM (right) for matrices with roughly $\mathcal{O}(1)$ nonzeros per row and column.

The approximation to $f(A)$ in the above experiments can be highly accurate when the bandwidth of the matrix is small or the matrix has a sparsity pattern that makes $f(A)$ approximately sparse, in the sense that only few entries in each row/column are large in magnitude. Our algorithms are particularly useful when A can only be accessed through matrix-vector products and we require an approximation to the entire matrix function. An example is in network analysis where we are given an adjacency matrix A and the off-diagonal entries of e^A describe the communicability between two vertices and the diagonal entries of e^A describe the subgraph centrality of a vertex [5,55,56]. The sparse matrix CSphd in the above experiment is an adjacency matrix for some underlying graph, and with about 180 matrix-vector products we can approximate its matrix exponential to 10^{-15} accuracy. This corresponds to a total of $180 \times 19 = 3420$ matrix-vector products with A . This means that our algorithm can approximate the communicability between any two vertices and the subgraph centrality of any vertex of CSphd with 10^{-15} accuracy.

4.4. Scalability test

This section investigate the asymptotic scalability of BaMRAM and SpaMRAM on matrices larger than those considered in the previous sections. We consider the following classes of synthetic matrices, similar to those in Section 4.1:

- **Banded case:** $n \times n$ symmetric 2-banded matrix with its nonzero entries drawn i.i.d. from $\mathcal{N}(0, 1)$. The matrix is rescaled so that its 2-norm equals 0.5.
- **Sparse case:** $n \times n$ symmetric sparse matrix created using the MATLAB command `sprandsym` with density $\frac{1}{2n}$. This matrix is also rescaled to have 2-norm equal to 0.5.

We vary the matrix dimension n from 400 to 25600 and record the number of matrix-vector products with A required for the methods to meet a prescribed tolerance. The dotted line in the figures corresponds to the number of matrix-vector products with A equal to

n , which is the number required to recover a general dense matrix. Throughout, we use the matrix exponential, where the action $x \mapsto f(\mathbf{A})x$ is approximated using 10 Krylov iterations. All other parameters are identical to those used in the previous experiments.

The results are shown in Fig. 6. Both matrix families have $\mathcal{O}(1)$ nonzero entries per row and column, which suggests that the corresponding matrix exponential may contain only a small number of dominant entries.

In the banded case (Fig. 6a), we observe that the number of matrix-vector products with \mathbf{A} required by BaMRAM is essentially independent of the matrix dimension for all tolerances considered. This suggests that, for matrices with bounded bandwidth, BaMRAM can effectively recover the dominant entries of $f(\mathbf{A})$ using a number of matrix-vector products with \mathbf{A} that does not grow with n .

For the sparse case, SpaMRAM exhibits a more irregular behavior as the matrix dimension increases, which can be attributed to the unknown sparsity pattern of the original matrix \mathbf{A} . Nevertheless, the number of matrix-vector products grows slowly with n and remains below the dotted line for large n and moderate tolerances. Although the curves exhibit mild fluctuations, the overall scaling is sublinear for this test matrix, indicating that the SpaMRAM effectively exploits the localized structure of the underlying matrix function. In both cases, tighter tolerances lead to an increased number of matrix-vector products, while still maintaining a favorable complexity compared to recovering a general dense matrix.

5. Discussion and extensions

In this paper, we presented two algorithms for approximating functions of sparse matrices and matrices with similar structures using matrix-vector products only. We exploited the decay bound from [15] to recover the entries of $f(\mathbf{A})$ that are large in magnitude. The task of approximating functions of matrices is not the only application for SpaMRAM and BaMRAM. These two algorithms can be used more generally in applications where we are able to perform matrix-vector products with the matrix in question and the matrix we want to approximate satisfies a decay bound similar to (5) or (13). An example is the decay bound for spectral projectors [57], computed, for example by a discretized contour integral. Therefore the algorithms can be extended naturally to other similar applications.

In many applications, we also want to approximate quantities related to $f(\mathbf{A})$ such as the trace of $f(\mathbf{A})$ and the Schatten p -norm of $f(\mathbf{A})$, which for the trace of functions of sparse symmetric matrices has been studied in [58]. Since SpaMRAM and BaMRAM provide an approximation to $f(\mathbf{A})$, we can use $\tilde{f}(\mathbf{A})$ to estimate the quantity of interest. For example, the trace of a matrix function is an important quantity appearing in network analysis through the Estrada index [5] and statistical learning through the log-determinant estimation [8], among others. Therefore the algorithms can also be used to estimate quantities related to matrix functions.

SpaMRAM uses compressed sensing to recover the matrix function row by row. This algorithm is simple, however we may have access to extra information such as the sparsity pattern of the matrix. This information can potentially be used to speed up the algorithm and increase the accuracy of the algorithm. There are techniques in compressed sensing such as variance density sampling [48], which improves signal reconstruction using prior information. The improvement of SpaMRAM and the use of SpaMRAM and BaMRAM in different applications are left for future work.

Data availability

Data will be made available on request.

Acknowledgements

The authors are grateful to Jared Tanner for his helpful discussions and for pointing out useful literature in compressed sensing.

References

- [1] V. Druskin, S. Güttel, L. Knizhnerman, Near-Optimal perfectly matched layers for indefinite helmholtz problems, *SIAM Rev.* 58 (1) (2016) 90–116. <https://doi.org/10.1137/140966927>
- [2] R. Garrappa, M. Popolizio, On the use of matrix functions for fractional partial differential equations, *Math. Comput. Simul.* 81 (5) (2011) 1045–1056. <https://doi.org/10.1016/j.matcom.2010.10.009>
- [3] V. Grimm, M. Hochbruck, Rational approximation to trigonometric operators, *BIT* 48 (2) (2008) 215–229. <https://doi.org/10.1007/s10543-008-0185-9>
- [4] M. Benzi, P. Boito, Matrix functions in network analysis, *GAMM-Mitt.* 43 (3) (2020) e20200012. <https://doi.org/10.1002/gamm.202000012>
- [5] E. Estrada, D.J. Higham, Network properties revealed through matrix functions, *SIAM Rev.* 52 (4) (2010) 696–714. <https://doi.org/10.1137/090761070>
- [6] S. Goedecker, Linear scaling electronic structure methods, *Rev. Mod. Phys.* 71 (1999) 1085–1123. <https://doi.org/10.1103/RevModPhys.71.1085>
- [7] S. Mohr, W. Dawson, M. Wagner, D. Caliste, T. Nakajima, L. Genovese, Efficient computation of sparse matrix functions for large-scale electronic structure calculations: the chES library, *J. Chem. Theory Comput.* 13 (10) (2017) 4684–4698. <https://doi.org/10.1021/acs.jctc.7b00348>
- [8] A. Cortinovis, D. Kressner, On randomized trace estimates for indefinite matrices with an application to determinants, *Found. Comput. Math.* 22 (3) (2021) 875–903. <https://doi.org/10.1007/s10208-021-09525-9>
- [9] J. Wenger, G. Pleiss, P. Hennig, J.P. Cunningham, J.R. Gardner, Preconditioning for scalable gaussian process hyperparameter optimization, in: *International Conference on Machine Learning (ICML)*, 2022.
- [10] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices, *Phys. Rev. E* 74 (2006) 036104. <https://doi.org/10.1103/PhysRevE.74.036104>
- [11] N.J. Higham, *Functions of Matrices*, SIAM, 2008. <https://doi.org/10.1137/1.9780898717778>
- [12] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 4 edition, 2013.
- [13] S. Güttel, Rational krylov approximation of matrix functions: numerical methods and optimal pole selection, *GAMM-Mitt.* 36 (1) (2013) 8–31. <https://doi.org/10.1002/gamm.201310002>
- [14] S. Güttel, D. Kressner, K. Lund, Limited-memory polynomial methods for large-scale matrix functions, *GAMM-Mitt.* 43 (3) (2020) e20200019. <https://doi.org/10.1002/gamm.202000019>

- [15] M. Benzi, N. Razouk, Decay bounds and $\mathcal{O}(n)$ algorithms for approximating functions of sparse matrices, *ETNA* 28 (2007) 16–39. <http://eudml.org/doc/130625>.
- [16] M. Benzi, Localization in matrix computations: theory and applications, in: *Exploiting Hidden Structure in Matrix Computations: Algorithms and Applications*, Springer International Publishing, 2016, pp. 211–317. https://doi.org/10.1007/978-3-319-49887-4_4
- [17] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* 52 (2) (2006) 489–509. <https://doi.org/10.1109/TIT.2005.862083>
- [18] E.J. Candès, T. Tao, Near-Optimal signal recovery from random projections: universal encoding strategies?, *IEEE Trans. Inf. Theory* 52 (12) (2006) 5406–5425. <https://doi.org/10.1109/TIT.2006.885507>
- [19] D.L. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* 52 (4) (2006) 1289–1306. <https://doi.org/10.1109/TIT.2006.871582>
- [20] S. Demko, W.F. Moss, P.W. Smith, Decay rates for inverses of band matrices, *Math. Comp.* 43 (168) (1984) 491–499. <https://doi.org/10.1090/s0025-5718-1984-0758197-9>
- [21] M. Benzi, G.H. Golub, Bounds for the entries of matrix functions with applications to preconditioning, *BIT* 39 (3) (1999) 417–438. <https://doi.org/10.1023/A:1022362401426>
- [22] A. Frommer, C. Schimmel, M. Schweitzer, Analysis of probing techniques for sparse approximation and trace estimation of decaying matrix functions, *SIAM J. Matrix Anal. Appl.* 42 (3) (2021) 1290–1318. <https://doi.org/10.1137/20M1364461>
- [23] T.F. Coleman, J.J. Moré, Estimation of sparse jacobian matrices and graph coloring blems, *SIAM J. Numer. Anal.* 20 (1) (1983) 187–209. <https://doi.org/10.1137/0720013>
- [24] A.R. Curtis, M.J.D. Powell, J.K. Reid, On the estimation of sparse jacobian matrices, *IMA J. Appl. Math.* 13 (1) (1974) 117–119. <https://academic.oup.com/imaamat/article-pdf/13/1/117/2141139/13-1-117.pdf>. <https://doi.org/10.1093/imaamat/13.1.117>
- [25] Y. Saad, Analysis of some krylov subspace approximations to the matrix exponential operator, *SIAM J. Numer. Anal.* 29 (1) (1992) 209–228. <https://doi.org/10.1137/0729014>
- [26] N. Hale, N.J. Higham, L.N. Trefethen, Computing A^* , $\log(A)$, and related matrix functions by contour integrals, *SIAM J. Numer. Anal.* 46 (5) (2008) 2505–2523. <https://doi.org/10.1137/070700607>
- [27] A. Frommer, K. Lund, D.B. Szyld, Block krylov subspace methods for functions of matrices, *ETNA* 27 (2018) 100–126. https://doi.org/10.1553/etna_vol47s100
- [28] A. Cortinovis, D. Kressner, S. Massei, Divide-and-Conquer methods for functions of matrices with banded or hierarchical low-Rank structure, *SIAM J. Matrix Anal. Appl.* 43 (1) (2022) 151–177. <https://doi.org/10.1137/21M1432594>
- [29] D. Halikias, A. Townsend, Structured matrix recovery from matrix-vector products, *Numer. Linear Algebra Appl.* 31 (1) (2024) e2531. <https://doi.org/10.1002/nla.2531>
- [30] J. Levitt, P.-G. Martinsson, Linear-Complexity black-Box randomized compression of rank-Structured matrices, *SIAM J. Sci. Comput.* 46 (3) (2024) A1747–A1763. <https://doi.org/10.1137/22M1528574>
- [31] X. Sun, D.P. Woodruff, G. Yang, J. Zhang, Querying a matrix through matrix-Vector products, *ACM Trans. Algorithms* 17 (4) (2021). <https://doi.org/10.1145/3470566>
- [32] C. Bekas, E. Kokiopoulou, Y. Saad, An estimator for the diagonal of a matrix, *Appl. Numer. Math.* 57 (11) (2007) 1214–1229. *Numerical Algorithms, Parallelism and Applications* (2). <https://doi.org/10.1016/j.apnum.2007.01.003>
- [33] M.A. Herman, T. Strohmer, High-Resolution radar via compressed sensing, *IEEE Trans. Signal Process.* 57 (6) (2009) 2275–2284. <https://doi.org/10.1109/TSP.2009.2014277>
- [34] G.E. Pfander, H. Rauhut, J. Tanner, Identification of matrices having a sparse representation, *IEEE Trans. Signal Process.* 56 (11) (2008) 5376–5388. <https://doi.org/10.1109/TSP.2008.928503>
- [35] G. Dasarathy, P. Shah, B.N. Bhaskar, R.D. Nowak, Sketching sparse matrices, covariances, and graphs via tensor products, *IEEE Trans. Inf. Theory* 61 (3) (2015) 1373–1388. <https://doi.org/10.1109/TIT.2015.2391251>
- [36] E.J. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* 51 (12) (2005) 4203–4215. <https://doi.org/10.1109/TIT.2005.858979>
- [37] T. Blumensath, M.E. Davies, G. Rilling, Greedy algorithms for compressed sensing, in: Y.C. Eldar, G. Kutyniok (Eds.), *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012, p. 348–393. <https://doi.org/10.1017/CBO9780511794308.009>
- [38] J.D. Blanchard, J. Tanner, Performance comparisons of greedy algorithms in compressed sensing, *Numer. Linear Algebra Appl.* 22 (2) (2015) 254–282. <https://doi.org/10.1002/nla.1948>
- [39] T. Blumensath, M.E. Davies, Normalized iterative hard thresholding: guaranteed stability and performance, *IEEE J. Sel. Topics Signal Process.* 4 (2) (2010) 298–309. <https://doi.org/10.1109/JSTSP.2010.2042411>
- [40] S. Foucart, Hard thresholding pursuit: an algorithm for compressive sensing, *SIAM J. Numer. Anal.* 49 (6) (2011) 2543–2563. <https://doi.org/10.1137/100806278>
- [41] D. Needell, J.A. Tropp, CospaMP: iterative signal recovery from incomplete and inaccurate samples, *Appl. Comput. Harmon. Anal.* 26 (3) (2009) 301–321. <https://doi.org/10.1016/j.acha.2008.07.002>
- [42] S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing, Applied and Numerical Harmonic Analysis*, Springer New York, 2013. <https://books.google.co.uk/books?id=zb28BAAQBAJ>.
- [43] V. Rokhlin, M. Tygert, A fast randomized algorithm for overdetermined linear least-squares regression, *Proc. Natl. Acad. Sci. USA* 105 (36) (2008) 13212–13217.
- [44] S. Grattan, D. Tittley-Peloquin, Improved bounds for small-Sample estimation, *SIAM J. Matrix Anal. Appl.* 39 (2) (2018) 922–931. <https://doi.org/10.1137/17M1137541>
- [45] P.-G. Martinsson, J.A. Tropp, Randomized numerical linear algebra: foundations and algorithms, *Acta Numer.* 29 (2020) 403–572. <https://doi.org/10.1017/s0962492920000021>
- [46] F.L. Bayisa, Z. Zhou, O. Cronie, J. Yu, Adaptive algorithm for sparse signal recovery, *Digit Signal Process* 87 (2019) 10–18. <https://doi.org/10.1016/j.dsp.2019.01.002>
- [47] P. Indyk, E. Price, D.P. Woodruff, On the power of adaptivity in sparse recovery, in: 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science, 2011, pp. 285–294. <https://doi.org/10.1109/FOCS.2011.83>
- [48] F. Kraher, R. Ward, Stable and robust sampling strategies for compressive imaging, *IEEE Trans. Image Process.* 23 (2) (2014) 612–622. <https://doi.org/10.1109/TIP.2013.2288004>
- [49] V. Nakos, X. Shi, D.P. Woodruff, H. Zhang, Improved algorithms for adaptive compressed sensing, in: I. Chatzigiannakis, C. Kakkamanis, D. Marx, D. Sannella (Eds.), 45th International Colloquium on Automata, Languages, and Programming (ICALP 2018), 107 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2018, pp. 90:1–90:14. <https://doi.org/10.4230/LIPIcs.ICALP.2018.90>
- [50] M. Benzi, V. Simoncini, Decay bounds for functions of hermitian matrices with banded or kronecker structure, *SIAM J. Matrix Anal. Appl.* 36 (3) (2015) 1263–1282. <https://doi.org/10.1137/151006159>
- [51] S. Pozza, V. Simoncini, Inexact arnoldi residual estimates and decay properties for functions of non-Hermitian matrices, *BIT* 59 (4) (2019) 969–986. <https://doi.org/10.1007/s10543-019-00763-6>
- [52] R.J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations. Steady State and Time Dependent Problems*, SIAM, 2007. <https://doi.org/10.1137/1.9780898717839>
- [53] C.F.V. Loan, The ubiquitous kronecker product, *J. Comput. Appl. Math.* 123 (1) (2000) 85–100. *Numerical Analysis 2000. Vol. III: Linear Algebra*. [https://doi.org/10.1016/S0377-0427\(00\)00393-9](https://doi.org/10.1016/S0377-0427(00)00393-9)
- [54] T.A. Davis, Y. Hu, The university of florida sparse matrix collection, *ACM Trans. Math. Softw.* 38 (1) (2011). <https://doi.org/10.1145/2049662.2049663>
- [55] M. Benzi, E. Estrada, C. Klymko, Ranking hubs and authorities using matrix functions, *Linear Algebra Appl.* 438 (5) (2013) 2447–2474. <https://doi.org/10.1016/j.laa.2012.10.022>
- [56] E. Estrada, N. Hatano, Communicability in complex networks, *Phys. Rev. E* 77 (2008) 036111. <https://doi.org/10.1103/PhysRevE.77.036111>

- [57] M. Benzi, P. Boito, N. Razouk, Decay properties of spectral projectors with applications to electronic structure, *SIAM Rev.* 55 (1) (2013) 3–64. <https://doi.org/10.1137/100814019>
- [58] A. Frommer, M. Rinelli, M. Schweitzer, Analysis of stochastic probing methods for estimating the trace of functions of sparse symmetric matrices, *arXiv preprint arXiv:2009.11392* (2023). 2308.07722