

Sublinear Time Eigenvalue Approximation via Random Sampling

Rajarshi Bhattacharjee ✉ 🏠

Manning College of Information and Computer Sciences,
University of Massachusetts, Amherst, MA, USA

Gregory Dexter ✉ 🏠

Department of Computer Science, Purdue University, West Lafayette, IN, USA

Petros Drineas ✉ 🏠

Department of Computer Science, Purdue University, West Lafayette, IN, USA

Cameron Musco ✉ 🏠

Manning College of Information and Computer Sciences,
University of Massachusetts, Amherst, MA, USA

Archan Ray ✉ 🏠

Manning College of Information and Computer Sciences,
University of Massachusetts, Amherst, MA, USA

Abstract

We study the problem of approximating the eigenspectrum of a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with bounded entries (i.e., $\|\mathbf{A}\|_\infty \leq 1$). We present a simple sublinear time algorithm that approximates all eigenvalues of \mathbf{A} up to additive error $\pm \epsilon n$ using those of a randomly sampled $\tilde{O}\left(\frac{\log^3 n}{\epsilon^3}\right) \times \tilde{O}\left(\frac{\log^3 n}{\epsilon^3}\right)$ principal submatrix. Our result can be viewed as a concentration bound on the complete eigenspectrum of a random submatrix, significantly extending known bounds on just the singular values (the magnitudes of the eigenvalues). We give improved error bounds of $\pm \epsilon \sqrt{\text{nnz}(\mathbf{A})}$ and $\pm \epsilon \|\mathbf{A}\|_F$ when the rows of \mathbf{A} can be sampled with probabilities proportional to their sparsities or their squared ℓ_2 norms respectively. Here $\text{nnz}(\mathbf{A})$ is the number of non-zero entries in \mathbf{A} and $\|\mathbf{A}\|_F$ is its Frobenius norm. Even for the strictly easier problems of approximating the singular values or testing the existence of large negative eigenvalues (Bakshi, Chepurko, and Jayaram, FOCS '20), our results are the first that take advantage of non-uniform sampling to give improved error bounds. From a technical perspective, our results require several new eigenvalue concentration and perturbation bounds for matrices with bounded entries. Our non-uniform sampling bounds require a new algorithmic approach, which judiciously zeroes out entries of a randomly sampled submatrix to reduce variance, before computing the eigenvalues of that submatrix as estimates for those of \mathbf{A} . We complement our theoretical results with numerical simulations, which demonstrate the effectiveness of our algorithms in practice.

2012 ACM Subject Classification Theory of computation \rightarrow Sketching and sampling; Mathematics of computing \rightarrow Computations on matrices

Keywords and phrases sublinear algorithms, eigenvalue approximation, randomized linear algebra

Digital Object Identifier 10.4230/LIPIcs.ICALP.2023.21

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version:* <https://arxiv.org/abs/2109.07647> [10]

Funding RB, CM and AR were partially supported by an Adobe Research grant, along with NSF Grants 2046235 and 1763618. PD and GD were partially supported by NSF AF 1814041, NSF FRG 1760353, and DOE-SC0022085.

Acknowledgements We thank Ainesh Bakshi, Rajesh Jayaram, Anil Damle, Nicholas Monath and Christopher Musco for helpful conversations about this work.



© Rajarshi Bhattacharjee, Gregory Dexter, Petros Drineas, Cameron Musco, and Archan Ray;

licensed under Creative Commons License CC-BY 4.0

50th International Colloquium on Automata, Languages, and Programming (ICALP 2023).

Editors: Kousha Etessami, Uriel Feige, and Gabriele Puppis; Article No. 21; pp. 21:1–21:18

Leibniz International Proceedings in Informatics



Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



1 Introduction

Approximating the eigenvalues of a symmetric matrix is a fundamental problem – with applications in engineering, optimization, data analysis, spectral graph theory, and beyond. For an $n \times n$ matrix, all eigenvalues can be computed to high accuracy using direct eigen-decomposition in $O(n^\omega)$ time, where $\omega \approx 2.37$ is the exponent of matrix multiplication [14, 2]. When just a few of the largest magnitude eigenvalues are of interest, the power method and other iterative Krylov methods can be applied [39]. These methods repeatedly multiply the matrix of interest by query vectors, requiring $O(n^2)$ time per multiplication when the matrix is dense and unstructured.

For large n , it is desirable to have even faster eigenvalue approximation algorithms, running in $o(n^2)$ time – i.e., sublinear in the size of the input matrix. Unfortunately, for general matrices, no non-trivial approximation can be computed in $o(n^2)$ time: without reading $\Omega(n^2)$ entries, it is impossible to distinguish with reasonable probability if all entries (and hence all eigenvalues) are equal to zero, or if there is a single pair of arbitrarily large entries at positions (i, j) and (j, i) , leading to a pair of arbitrarily large eigenvalues. Given this, we seek to address the following question:

Under what assumptions on a symmetric $n \times n$ input matrix, can we compute non-trivial approximations to its eigenvalues in $o(n^2)$ time?

It is well known that $o(n^2)$ time eigenvalue computation is possible for highly structured inputs, like tridiagonal or Toeplitz matrices [26]. For sparse or structured matrices that admit fast matrix vector multiplication, one can compute a small number of the largest in magnitude eigenvalues in $o(n^2)$ time using iterative methods. Through the use of robust iterative methods, fast top eigenvalue estimation is also possible for matrices that admit fast *approximate* matrix-vector multiplication, such as kernel similarity matrices [25, 27, 4]. Our goal is to study simple, sampling-based sublinear time algorithms that work under much weaker assumptions on the input matrix.

1.1 Our Contributions

Our main contribution is to show that a very simple algorithm can be used to approximate *all eigenvalues* of any symmetric matrix with *bounded entries*. In particular, for any $\mathbf{A} \in \mathbb{R}^{n \times n}$ with maximum entry magnitude $\|\mathbf{A}\|_\infty \leq 1$, sampling an $s \times s$ principal submatrix \mathbf{A}_S of \mathbf{A} with $s = \tilde{O}\left(\frac{\log^3 n}{\epsilon^3}\right)$ and scaling its eigenvalues by n/s yields a $\pm \epsilon n$ additive error approximation to all eigenvalues of \mathbf{A} with good probability.¹ This result is formally stated below, where $[n] \stackrel{\text{def}}{=} \{1, \dots, n\}$.

► **Theorem 1** (Sublinear Time Eigenvalue Approximation). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric with $\|\mathbf{A}\|_\infty \leq 1$ and eigenvalues $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$. Let $S \subseteq [n]$ be formed by including each index independently with probability s/n as in Algorithm 1. Let \mathbf{A}_S be the corresponding principal submatrix of \mathbf{A} , with eigenvalues $\lambda_1(\mathbf{A}_S) \geq \dots \geq \lambda_{|S|}(\mathbf{A}_S)$.*

For all $i \in [|S|]$ with $\lambda_i(\mathbf{A}_S) \geq 0$, let $\tilde{\lambda}_i(\mathbf{A}) = \frac{n}{s} \cdot \lambda_i(\mathbf{A}_S)$. For all $i \in [|S|]$ with $\lambda_i(\mathbf{A}_S) < 0$, let $\tilde{\lambda}_{n-(|S|-i)}(\mathbf{A}) = \frac{n}{s} \cdot \lambda_i(\mathbf{A}_S)$. For all other $i \in [n]$, let $\tilde{\lambda}_i(\mathbf{A}) = 0$. If $s \geq \frac{c \log(1/(\epsilon\delta)) \cdot \log^3 n}{\epsilon^3 \delta}$, for large enough constant c , then with probability $\geq 1 - \delta$, for all $i \in [n]$, $\lambda_i(\mathbf{A}) - \epsilon n \leq \tilde{\lambda}_i(\mathbf{A}) \leq \lambda_i(\mathbf{A}) + \epsilon n$.

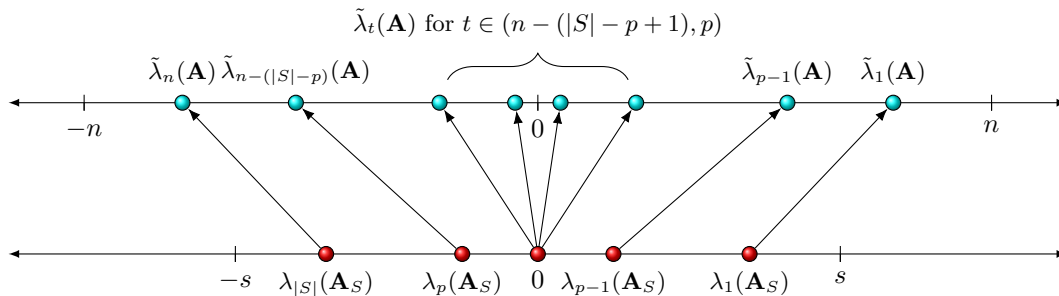
¹ Here and throughout, $\tilde{O}(\cdot)$ hides logarithmic factors in the argument. Note that by scaling, our algorithm gives a $\pm \epsilon n \cdot \|\mathbf{A}\|_\infty$ approximation for any \mathbf{A} .

See Figure 1 for an illustration of how the $|S|$ eigenvalues of \mathbf{A}_S are mapped to estimates for all eigenvalues of \mathbf{A} . Note that the principal submatrix \mathbf{A}_S sampled in Theorem 1 will have $O(s) = \tilde{O}\left(\frac{\log^3 n}{\epsilon^3 \delta}\right)$ rows/columns with high probability. Thus, with high probability, the algorithm reads just $\tilde{O}\left(\frac{\log^6 n}{\epsilon^6 \delta^2}\right)$ entries of \mathbf{A} and runs in $\text{poly}(\log n, 1/\epsilon, 1/\delta)$ time. Standard matrix concentration bounds imply that one can sample $O\left(\frac{s \log(1/\delta)}{\epsilon^2}\right)$ random entries from the $O(s) \times O(s)$ random submatrix \mathbf{A}_S and preserve its eigenvalues to error $\pm \epsilon s$ with probability $1 - \delta$ [1]. See Appendix F of [10] for a proof. This can be directly combined with Theorem 1 to give improved sample complexity:

► **Corollary 2 (Improved Sample Complexity via Entrywise Sampling).** *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric with $\|\mathbf{A}\|_\infty \leq 1$ and eigenvalues $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$. For any $\epsilon, \delta \in (0, 1)$, there is an algorithm that reads $\tilde{O}\left(\frac{\log^3 n}{\epsilon^3 \delta}\right)$ entries of \mathbf{A} and returns, with probability at least $1 - \delta$, $\tilde{\lambda}_i(\mathbf{A})$ for each $i \in [n]$ satisfying $|\tilde{\lambda}_i(\mathbf{A}) - \lambda_i(\mathbf{A})| \leq \epsilon n$.*

Observe that the dependence on δ in Theorem 1 and Corollary 2 can be improved via standard arguments: running the algorithm with failure probability $\delta' = 2/3$, repeating $O(\log(1/\delta))$ times, and taking the median estimate for each $\lambda_i(\mathbf{A})$. This guarantees that the algorithm will succeed with probability at most $1 - \delta$ at the expense of a $\log(1/\delta)$ dependence in the complexity.

We note that our $\pm \epsilon n$ error guarantee is particularly useful in applications where the matrix \mathbf{A} has low stable rank and the top eigenvalues have magnitude scaling roughly with n . Low stable rank is a common feature of real-life data matrices [47], including classes of bounded entry matrices, such as kernel similarity matrices [18] and adjacency matrices of power law graphs [36].



■ **Figure 1 Alignment of eigenvalues in Thm. 1 and Algo. 1.** We illustrate how the eigenvalues of \mathbf{A}_S , scaled by $\frac{n}{s}$, are used to approximate all eigenvalues of \mathbf{A} . If \mathbf{A}_S has $p - 1$ positive eigenvalues, they are set to the top $p - 1$ eigenvalue estimates. Its $|S| - p + 1$ negative eigenvalues are set to the bottom eigenvalue estimates. All remaining eigenvalues are simply approximated as zero.

Comparison to known bounds. Theorem 1 can be viewed as a concentration inequality on the full eigenspectrum of a random principal submatrix \mathbf{A}_S of \mathbf{A} . This significantly extends prior work, which was able to bound just the spectral norm (i.e., the magnitude of the top eigenvalue) of a random principal submatrix [38, 44]. Bakshi, Chepurko, and Jayaram [5] recently identified developing such full eigenspectrum concentration inequalities as an important step in expanding our knowledge of sublinear time property testing algorithms for bounded entry matrices.

21:4 Sublinear Time Eigenvalue Approximation via Random Sampling

Standard matrix concentration bounds [22] can be used to show that the *singular values* of \mathbf{A} (i.e., the magnitudes of its eigenvalues) are approximated by those of a $O\left(\frac{\log n}{\epsilon^2}\right) \times O\left(\frac{\log n}{\epsilon^2}\right)$ random submatrix (see Appendix G of [10]) with independently sampled rows and columns. However, such a random matrix will not be symmetric or even have real eigenvalues in general, and thus no analogous bounds were previously known for the eigenvalues themselves.

Lower Bounds. Recently, Bakshi, Chepurko, and Jayaram [5] studied the closely related problem of testing positive semidefiniteness in the bounded entry model. They show how to test whether the minimum eigenvalue of \mathbf{A} is either greater than 0 or smaller than $-\epsilon n$ by reading just $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$ entries. They show that this result is optimal in terms of query complexity, up to logarithmic factors. Like our approach, their algorithm is based on random principal submatrix sampling. Our eigenvalue approximation guarantee strictly strengthens the testing guarantee – given $\pm\epsilon n$ approximations to all eigenvalues, we immediately solve the testing problem. Thus, our query complexity is tight up to a $\text{poly}(\log n, 1/\epsilon)$ factor. It is open if our higher sample complexity is necessary to solve the harder full eigenspectrum estimation problem. See Section 1.4 for further discussion.

Improved bounds for non-uniform sampling. Our second main contribution is to show that, when it is possible to efficiently sample rows/columns of \mathbf{A} with probabilities proportional to their sparsities or their squared ℓ_2 norms, significantly stronger eigenvalue estimates can be obtained. In particular, letting $\text{nnz}(\mathbf{A})$ denote the number of nonzero entries in \mathbf{A} and $\|\mathbf{A}\|_F$ denote its Frobenius norm, we show that sparsity-based sampling yields eigenvalue estimates with error $\pm\epsilon\sqrt{\text{nnz}(\mathbf{A})}$ and norm-based sampling gives error $\pm\epsilon\|\mathbf{A}\|_F$. See Theorems 3 and 4 for formal statements. Observe that when $\|\mathbf{A}\|_\infty \leq 1$, its eigenvalues are bounded in magnitude by $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{\text{nnz}(\mathbf{A})} \leq n$. Thus, Theorems 3 and 4 are natural strengthenings of Theorem 1. Row norm-based sampling (Theorem 4) additionally removes the bounded entry requirement of Theorems 1 and 3.

As discussed in Section 1.3.1, sparsity-based sampling can be performed in sublinear time when \mathbf{A} is stored in a slightly augmented sparse matrix format, or when \mathbf{A} is the adjacency matrix of a graph accessed in the standard graph query model of the sublinear algorithms literature [23]. Norm-based sampling can also be performed efficiently with an augmented matrix format, and is commonly studied in randomized and “quantum-inspired” algorithms for linear algebra [19, 43].

► **Theorem 3 (Sparse Matrix Eigenvalue Approximation).** *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric with $\|\mathbf{A}\|_\infty \leq 1$ and eigenvalues $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$. Let $S \subseteq [n]$ be formed by including the i^{th} index independently with probability $p_i = \min\left(1, \frac{s \text{nnz}(\mathbf{A}_i)}{\text{nnz}(\mathbf{A})}\right)$ as in Algorithm 2 of [10]. Here $\text{nnz}(\mathbf{A}_i)$ is the number of non-zero entries in the i^{th} row of \mathbf{A} . Let \mathbf{A}_S be the corresponding principal submatrix of \mathbf{A} , and let $\tilde{\lambda}_i(\mathbf{A})$ be the estimate of $\lambda_i(\mathbf{A})$ computed from \mathbf{A}_S as in Algorithm 2 of citebhattacharjee2021sublinear. If $s \geq \frac{c \log^8 n}{\epsilon^8 \delta^4}$, for large enough constant c , then with probability $\geq 1 - \delta$, for all $i \in [n]$, $|\tilde{\lambda}_i(\mathbf{A}) - \lambda_i(\mathbf{A})| \leq \epsilon\sqrt{\text{nnz}(\mathbf{A})}$.*

► **Theorem 4 (Row Norm Based Matrix Eigenvalue Approximation).** *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric and eigenvalues $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$. Let $S \subseteq [n]$ be formed by including the i^{th} index independently with probability $p_i = \min\left(1, \frac{s\|\mathbf{A}_i\|_2^2}{\|\mathbf{A}\|_F^2} + \frac{1}{n^2}\right)$ as in Algorithm 3 of [10]. Here $\|\mathbf{A}_i\|_2$ is the ℓ_2 norm of the i^{th} row of \mathbf{A} . Let \mathbf{A}_S be the corresponding principal submatrix of \mathbf{A} , and let $\tilde{\lambda}_i(\mathbf{A})$ be the estimate of $\lambda_i(\mathbf{A})$ computed from \mathbf{A}_S as in Algorithm 3 of [10]. If $s \geq \frac{c \log^{10} n}{\epsilon^8 \delta^4}$, for large enough constant c , then with probability $\geq 1 - \delta$, for all $i \in [n]$, $|\tilde{\lambda}_i(\mathbf{A}) - \lambda_i(\mathbf{A})| \leq \epsilon\|\mathbf{A}\|_F$.*

The above non-uniform sampling theorems immediately yield algorithms for testing the presence of a negative eigenvalue with magnitude at least $\epsilon\sqrt{\text{nnz}(\mathbf{A})}$ or $\epsilon\|\mathbf{A}\|_F$ respectively, strengthening the results of [5], which require eigenvalue magnitude at least ϵn . In the graph property testing literature, there is a rich line of work exploring the testing of bounded degree or sparse graphs [23, 7]. Theorem 3 can be thought of as first step in establishing a related theory of sublinear time approximation algorithms and property testers for sparse matrices. Due to lack of space, we defer the proofs of Theorems 3 and 4 to Section 4 and Appendix E of [10] respectively.

Surprisingly, in the non-uniform sampling case, the eigenvalue estimates derived from \mathbf{A}_S cannot simply be its scaled eigenvalues, as in Theorem 1. E.g., when \mathbf{A} is the identity, our row sampling probabilities are uniform in all cases. However, the scaled submatrix $\frac{n}{s} \cdot \mathbf{A}_S$ will be a scaled identity, and have eigenvalues equal to n/s – failing to give a $\pm\epsilon\sqrt{\text{nnz}(\mathbf{A})} = \pm\epsilon\|\mathbf{A}\|_F = \pm\epsilon\sqrt{n}$ approximation to the true eigenvalues (all of which are 1) unless $s \gtrsim \frac{\sqrt{n}}{\epsilon}$. To handle this, and related cases, we must argue that selectively zeroing out entries in sufficiently low probability rows/columns of \mathbf{A} (see Algorithms 2 and 3 of [10]) does not significantly change the spectrum, and ensures concentration of the submatrix eigenvalues. It is not hard to see that simple random submatrix sampling fails even for the easier problem of singular value estimation. Theorems 3 and 4 give the first results of their kinds for this problem as well.

1.2 Related Work

Eigenspectrum estimation is a key primitive in numerical linear algebra, typically known as *spectral density estimation*. The eigenspectrum is viewed as a distribution with mass $1/n$ at each of the n eigenvalues, and the goal is to approximate this distribution [49, 35]. Applications include identifying motifs in social networks [15], studying Hessian and weight matrix spectra in deep learning [40, 51, 21], “spectrum splitting” in parallel eigensolvers [31], and the study of many systems in experimental physics and chemistry [48, 41, 28].

Recent work has studied sublinear time spectral density estimation for graph structured matrices – Braverman, Krishnan, and Musco [11] show that the spectral density of a normalized graph adjacency or Laplacian matrix can be estimated to ϵ error in the Wasserstein distance in $\tilde{O}(n/\text{poly}(\epsilon))$ time. Cohen-Steiner, Kong, Sohler, and Valiant study a similar setting, giving runtime $2^{O(1/\epsilon)}$ [13]. We note that the additive error eigenvalue approximation result of Theorem 1 (analogously Theorems 3 and 4) directly gives an ϵn approximation to the spectral density in the Wasserstein distance – extending the above results to a much broader class of matrices. When $\|\mathbf{A}\|_\infty \leq 1$, \mathbf{A} can have eigenvalues as large as n , while the normalized adjacency matrices studied in [13, 11] have eigenvalues in $[-1, 1]$. So, while the results are not directly comparable, our Wasserstein error can be thought as on order of their error of ϵ after scaling.

Our work is also closely related to a line of work on sublinear time property testing for bounded entry matrices, initiated by Balcan et al. [6]. In that work, they study testing of rank, Schatten- p norms, and several other global spectral properties. Sublinear time testing algorithms for the rank and other properties have also been studied under low-rank and bounded row norm assumptions on the input matrix [30, 33]. Recent work studies positive semidefiniteness testing and eigenvalue estimation in the matrix-vector query model, where each query computes $\mathbf{A}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^{n \times n}$. As in Theorem 4, $\pm\epsilon\|\mathbf{A}\|_F$ eigenvalue estimation can be achieved with $\text{poly}(\log n, 1/\epsilon)$ queries in this model [37]. Finally, several works study streaming algorithms for eigenspectrum approximation [3, 32, 34]. These algorithms are not sublinear time – they require at least linear time to process the input matrix. However, they

use sublinear working memory. Note that Theorem 1 immediately gives a sublinear space streaming algorithm for eigenvalue estimation. We can simply store the sampled submatrix \mathbf{A}_S as its entries are updated.

1.3 Technical Overview

In this section, we overview the main techniques used to prove Theorems 1, and then how these techniques are extended to prove Theorems 3 and 4. We start by defining a decomposition of any symmetric \mathbf{A} into the sum of two matrices containing its large and small magnitude eigendirections.

► **Definition 5** (Eigenvalue Split). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric. For any $\epsilon, \delta \in (0, 1)$, let $\mathbf{A}_o = \mathbf{V}_o \mathbf{\Lambda}_o \mathbf{V}_o^T$ where $\mathbf{\Lambda}_o$ is diagonal, with the eigenvalues of \mathbf{A} with magnitude $\geq \epsilon\sqrt{\delta}n$ on its diagonal, and \mathbf{V}_o has the corresponding eigenvectors as columns. Similarly, let $\mathbf{A}_m = \mathbf{V}_m \mathbf{\Lambda}_m \mathbf{V}_m^T$ where $\mathbf{\Lambda}_m$ has the eigenvalues of \mathbf{A} with magnitude $< \epsilon\sqrt{\delta}n$ on its diagonal and \mathbf{V}_m has the corresponding eigenvectors as columns. Then, \mathbf{A} can be decomposed as*

$$\mathbf{A} = \mathbf{A}_o + \mathbf{A}_m = \mathbf{V}_o \mathbf{\Lambda}_o \mathbf{V}_o^T + \mathbf{V}_m \mathbf{\Lambda}_m \mathbf{V}_m^T.$$

Any principal submatrix of \mathbf{A} , \mathbf{A}_S , can be similarly written as

$$\mathbf{A}_S = \mathbf{A}_{o,S} + \mathbf{A}_{m,S} = \mathbf{V}_{o,S} \mathbf{\Lambda}_o \mathbf{V}_{o,S}^T + \mathbf{V}_{m,S} \mathbf{\Lambda}_m \mathbf{V}_{m,S}^T,$$

where $\mathbf{V}_{o,S}, \mathbf{V}_{m,S}$ are the corresponding submatrices obtained by sampling rows of $\mathbf{V}_o, \mathbf{V}_m$.

Since $\mathbf{A}_S, \mathbf{A}_{m,S}$ and $\mathbf{A}_{o,S}$ are all symmetric, we can use Weyl's eigenvalue perturbation theorem [50] to show that for all eigenvalues of \mathbf{A}_S ,

$$|\lambda_i(\mathbf{A}_S) - \lambda_i(\mathbf{A}_{o,S})| \leq \|\mathbf{A}_{m,S}\|_2. \quad (1)$$

We will argue that the eigenvalues of $\mathbf{A}_{o,S}$ approximate those of \mathbf{A}_o – i.e. all eigenvalues of \mathbf{A} with magnitude $\geq \epsilon\sqrt{\delta}n$. Further, we will show that $\|\mathbf{A}_{m,S}\|_2$ is small with good probability. Thus, via (1), the eigenvalues of \mathbf{A}_S approximate those of \mathbf{A}_o . In the estimation procedure of Theorem 1, all other small magnitude eigenvalues of \mathbf{A} are estimated to be 0, which will immediately give our $\pm\epsilon n$ approximation bound when the original eigenvalue has magnitude $\leq \epsilon n$.

Bounding the eigenvalues of $\mathbf{A}_{o,S}$. The first step is to show that the eigenvalues of $\mathbf{A}_{o,S}$ well-approximate those of \mathbf{A}_o . As in [5], we critically use that the eigenvectors corresponding to large eigenvalues are *incoherent* – intuitively, since $\|\mathbf{A}\|_\infty$ is bounded, their mass must be spread out in order to witness a large eigenvalue. Specifically, [5] shows that for any eigenvector \mathbf{v} of \mathbf{A} with corresponding eigenvalue $\geq \epsilon\sqrt{\delta}n$, $\|\mathbf{v}\|_\infty \leq \frac{1}{\epsilon\sqrt{\delta}\sqrt{n}}$. We give related bounds on the Euclidean norms of the rows of \mathbf{V}_o (the *leverage scores* of \mathbf{A}_o), and on these rows after weighting by $\mathbf{\Lambda}_o$.

Using these incoherence bounds, we argue that the eigenvalues of $\mathbf{A}_{o,S}$ approximate those of \mathbf{A}_o up to $\pm\epsilon n$ error. A key idea is to bound the eigenvalues of $\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,S}^T \mathbf{V}_{o,S} \mathbf{\Lambda}_o^{1/2}$, which are identical to the non-zero eigenvalues of $\mathbf{A}_{o,S} = \mathbf{V}_{o,S} \mathbf{\Lambda}_o \mathbf{V}_{o,S}^T$. Via a matrix Bernstein bound and our incoherence bounds on \mathbf{V}_o , we show that this matrix is close to $\mathbf{\Lambda}_o$ with high probability. However, since $\mathbf{\Lambda}_o^{1/2}$ may be complex, the matrix is *not necessarily Hermitian* and standard perturbation bounds [42, 29] do not apply. Thus, to derive an eigenvalue

bound, we apply a perturbation bound of Bhatia [9], which generalizes Weyl’s inequality to the non-Hermitian case, with a $\log n$ factor loss. To the best of our knowledge, this is the first time that perturbation theory bounds for non-Hermitian matrices have been used to prove improved algorithmic results in the theoretical computer science literature.

We note that in Appendix B of [10], we give an alternate bound, which instead analyzes the Hermitian matrix $(\mathbf{V}_{o,S}^T \mathbf{V}_{o,S})^{1/2} \mathbf{A}_o (\mathbf{V}_{o,S}^T \mathbf{V}_{o,S})^{1/2}$, whose eigenvalues are again identical to those of $\mathbf{A}_{o,S}$. This approach only requires Weyl’s inequality, and yields an overall bound of $s = O\left(\frac{\log n}{\epsilon^4 \delta}\right)$, improving the $\log n$ factors of Theorem 1 at the cost of worse ϵ dependence.

Bounding the spectral norm of $\mathbf{A}_{m,S}$. The next step is to show that all eigenvalues of $\mathbf{A}_{m,S}$ are small provided a sufficiently large submatrix is sampled. This means that the “middle” eigenvalues of \mathbf{A} , *i.e.* those with magnitude $\leq \epsilon\sqrt{\delta}n$ do not contribute much to any eigenvalue $\lambda_i(\mathbf{A}_S)$. To do so, we apply a theorem of [38, 44] which shows concentration of the spectral norm of a uniformly random submatrix of an entrywise bounded matrix. Observe that while $\|\mathbf{A}\|_\infty \leq 1$, such a bound will not in general hold for $\|\mathbf{A}_m\|_\infty$. Nevertheless, we can use the incoherence of \mathbf{V}_o to show that $\|\mathbf{A}_o\|_\infty$ is bounded, which via triangle inequality, yields a bound on $\|\mathbf{A}_m\|_\infty \leq \|\mathbf{A}\|_\infty + \|\mathbf{A}_o\|_\infty$. In the end, we show that if $s \geq O\left(\frac{\log n}{\epsilon^2 \delta}\right)$, with probability at least $1 - \delta$, $\|\mathbf{A}_{m,S}\|_2 \leq \epsilon s$. After the n/s scaling in the estimation procedure of Theorem 1, this spectral norm bound translates into an additive ϵn error in approximating the eigenvalues of \mathbf{A} .

Completing the argument. Once we establish the above bounds on $\mathbf{A}_{o,S}$ and $\mathbf{A}_{m,S}$, Theorem 1 is essentially complete. Any eigenvalue in \mathbf{A} with magnitude $\geq \epsilon n$ will correspond to a nearby eigenvalue in $\frac{n}{s} \cdot \mathbf{A}_{o,S}$ and in turn, $\frac{n}{s} \cdot \mathbf{A}_S$ given our spectral norm bound on $\mathbf{A}_{m,S}$. An eigenvalue in \mathbf{A} with magnitude $\leq \epsilon n$ may or may not correspond to a nearby eigenvalue in $\mathbf{A}_{o,S}$ (it will only if it lies in the range $[\epsilon\sqrt{\delta}n, \epsilon n]$). In any case, in the estimation procedure of Theorem 1, such an eigenvalue will either be estimated using a small eigenvalue of \mathbf{A}_S , or be estimated as 0. In both instances, the estimate will give $\pm \epsilon n$ error.

Can we beat additive error? It is natural to ask if our approach can be improved to yield sublinear time algorithms with stronger relative error approximation guarantees for \mathbf{A} ’s eigenvalues. Unfortunately, this is not possible – consider a matrix with just a single pair of entries $\mathbf{A}_{i,j}, \mathbf{A}_{j,i}$ set to 1. To obtain relative error approximations to the two non-zero eigenvalues, we must find the pair (i, j) , as otherwise we cannot distinguish \mathbf{A} from the all zeros matrix. This requires reading a $\Omega(n^2)$ of \mathbf{A} ’s entries. More generally, consider \mathbf{A} with a random $n/t \times n/t$ principal submatrix populated by all 1s, and with all other entries equal to 0. \mathbf{A} has largest eigenvalue n/t . However, if we read $s \ll t^2$ entries of \mathbf{A} , with good probability, we will not see even a single one, and thus we will not be able to distinguish \mathbf{A} from the all zeros matrix. This example establishes that any sublinear time algorithm with query complexity s must incur additive error at least $\Omega(n/\sqrt{s})$.

1.3.1 Improved Bounds via Non-Uniform Sampling

We now discuss how to give improved approximation bounds via non-uniform sampling. We focus on the $\pm \epsilon \sqrt{\text{nnz}(\mathbf{A})}$ bound of Theorem 3 using sparsity-based sampling. Theorem 4’s proof (for row norm sampling) follows the same general ideas, but with some additional complications.

Theorem 3 requires sampling a submatrix \mathbf{A}_S , where each index i is included in S with probability $p_i = \min(1, \frac{s \operatorname{nnz}(\mathbf{A}_i)}{\operatorname{nnz}(\mathbf{A})})$. We reweight each sampled row by $\frac{1}{\sqrt{p_i}}$. Thus, if entry \mathbf{A}_{ij} is sampled, it is scaled by $\frac{1}{\sqrt{p_i p_j}}$. When the rows have uniform sparsity (so all $p_i = s/n$), this ensures that the full submatrix is scaled by n/s , as in Theorem 1.

The proof of Theorem 3 follows the same outline as that of Theorem 1: we first argue that the outlying eigenvectors in \mathbf{V}_o are incoherent, giving a bound on the norm of each row of \mathbf{V}_o in terms of $\operatorname{nnz}(\mathbf{A}_i)$. We then apply a matrix Bernstein bound and Bhatia's non-Hermitian eigenvalue perturbation bound to show that the eigenvalues of $\mathbf{A}_{o,S}$ approximate those of \mathbf{A}_o up to $\pm \epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}$.

Bounding the spectral norm of $\mathbf{A}_{m,S}$. The major challenge is showing that the subsampled middle eigendirections do not significantly increase the approximation error by bounding the $\|\mathbf{A}_{m,S}\|_2$ by $\epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}$. This is difficult since the indices in $\mathbf{A}_{m,S}$ are sampled nonuniformly, so existing bounds [44] on the spectral norm of uniformly random submatrices do not apply. We extend these bounds to the non-uniform sampling case, but still face an issue due to the rescaling of entries by $\frac{1}{\sqrt{p_i p_j}}$. In fact, without additional algorithmic modifications, $\|\mathbf{A}_{m,S}\|_2$ is simply not bounded by $\epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}$! For example, as already discussed, if $\mathbf{A} = \mathbf{I}$ is the identity matrix, we get $\mathbf{A}_{m,S} = \frac{n}{s} \cdot \mathbf{I}$ and so $\|\mathbf{A}_{m,S}\|_2 = \frac{n}{s} > \epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}$, assuming $s < \frac{\sqrt{n}}{\epsilon}$. Relatedly, suppose that \mathbf{A} is tridiagonal, with zeros on the diagonal and ones on the first diagonals above and below the main diagonal. Then, if $s \geq \sqrt{n}$, with constant probability, one of the ones will be sampled and scaled by $\frac{n}{s}$. Thus, we will again have $\|\mathbf{A}_{m,S}\|_2 \geq \frac{n}{s} \geq \epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}$, assuming $s < \frac{\sqrt{n}}{2\epsilon}$. Observe that this issue arises even when trying to approximate just the singular values (the eigenvalue magnitudes). Thus, while an analogous bound to the uniform sampling result of Theorem 1 can easily be given for singular value estimation via matrix concentration inequalities (see Appendix G of [10]), to the best of our knowledge, Theorems 3 and 4 are the first of their kind even for singular value estimation.

Zeroing out entries in sparse rows/columns. To handle the above cases, we prove a novel perturbation bound, arguing that the eigenvalues of \mathbf{A} are not perturbed by more than $\epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}$ if we zero out any entry \mathbf{A}_{ij} of \mathbf{A} where $\sqrt{\operatorname{nnz}(\mathbf{A}_i) \cdot \operatorname{nnz}(\mathbf{A}_j)} \leq \frac{\epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}}{c \log n}$. This can be thought of as a strengthening of Girshgorin's circle theorem, which would ensure that zeroing out entries in rows/columns with $\operatorname{nnz}(\mathbf{A}_i) \leq \epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}$ does not perturb the eigenvalues by more than $\epsilon \sqrt{\operatorname{nnz}(\mathbf{A})}$. Armed with this perturbation bound, we argue that if we zero out the appropriate entries of \mathbf{A}_S before computing its eigenvalues, then since we have removed entries in very sparse rows and columns which would be scaled by a large $\frac{1}{\sqrt{p_i p_j}}$ factor in \mathbf{A}_S , we can bound $\|\mathbf{A}_{m,S}\|_2$. This requires relating the magnitudes of the entries in $\mathbf{A}_{m,S}$ to those in \mathbf{A}_S using the incoherence of the top eigenvectors, which gives bounds on the entries of $\mathbf{A}_{o,S} = \mathbf{A}_S - \mathbf{A}_{m,S}$.

Sampling model. We note that the sparsity-based sampling of Theorem 3 can be efficiently implemented in several natural settings. Given a matrix stored in sparse format, i.e., as a list of nonzero entries, we can easily sample a row with probability $\frac{\operatorname{nnz}(\mathbf{A}_i)}{\operatorname{nnz}(\mathbf{A})}$ by sampling a uniformly random non-zero entry and looking at its corresponding row. Via standard techniques, we can convert several such samples into a sampled set S close in distribution to having each $i \in [n]$ included independently with probability $\min\left(1, \frac{s \operatorname{nnz}(\mathbf{A}_i)}{\operatorname{nnz}(\mathbf{A})}\right)$. If we store the values of $\operatorname{nnz}(\mathbf{A}), \operatorname{nnz}(\mathbf{A}_1), \dots, \operatorname{nnz}(\mathbf{A}_n)$, we can also efficiently access each p_i , which is

needed for rescaling and zeroing out entries. Also observe that if \mathbf{A} is the adjacency matrix of a graph, in the standard graph query model [23], it is well known how to approximately count edges and sample them uniformly at random, i.e., compute $\text{nnz}(\mathbf{A})$ and sample its nonzero entries, in sublinear time [24, 17]. Further, it is typically assumed that one has access to the node degrees, i.e., $\text{nnz}(\mathbf{A}_1), \dots, \text{nnz}(\mathbf{A}_n)$. Thus, our algorithm can naturally be used to estimate spectral graph properties in sublinear time.

The ℓ_2 norm-based sampling of Theorem 4 can also be performed efficiently using an augmented data structure for storing \mathbf{A} . Such data structures have been used extensively in the literature on quantum-inspired algorithms, and require just $O(\text{nnz}(\mathbf{A}))$ time to construct, $O(\text{nnz}(\mathbf{A}))$ space, and $O(\log n)$ time to update give an update to an entry of \mathbf{A} [43, 12].

1.4 Towards Optimal Query Complexity

As discussed, Bakshi et al. [5] show that any algorithm which can test with good probability whether \mathbf{A} has an eigenvalue $\leq -\epsilon n$ or else has all non-negative eigenvalues must read $\tilde{\Omega}\left(\frac{1}{\epsilon^2}\right)$ entries of \mathbf{A} . This testing problem is strictly easier than outputting $\pm \epsilon n$ error estimates of all eigenvalues, so gives a lower bound for our setting. If the queried entries are restricted to fall in a submatrix, [5] shows that this submatrix must have dimensions $\Omega\left(\frac{1}{\epsilon^2}\right) \times \Omega\left(\frac{1}{\epsilon^2}\right)$, giving total query complexity $\Omega\left(\frac{1}{\epsilon^4}\right)$. Closing the gap between our upper bound of $\tilde{O}\left(\frac{\log^3 n}{\epsilon^3}\right) \times \tilde{O}\left(\frac{\log^3 n}{\epsilon^3}\right)$ and the lower bound of $\Omega\left(\frac{1}{\epsilon^2}\right) \times \Omega\left(\frac{1}{\epsilon^2}\right)$ for submatrix queries is an intriguing open question.

Closing the gap. We show in Appendix A of [10] that this gap can be easily closed via a surprisingly simple argument if \mathbf{A} is positive semidefinite (PSD). In that case, $\mathbf{A} = \mathbf{B}\mathbf{B}^T$ with $\mathbf{B} \in \mathbb{R}^{n \times n}$. Writing $\mathbf{A}_S = \mathbf{S}^T \mathbf{A} \mathbf{S}$ for a sampling matrix $\mathbf{S} \in \mathbb{R}^{n \times |S|}$, the non-zero eigenvalues of \mathbf{A}_S are identical to those of $\mathbf{B}\mathbf{S}\mathbf{S}^T\mathbf{B}^T$. Via a standard approximate matrix multiplication analysis [16], one can then show that, for $s \geq \frac{1}{\epsilon^2 \delta}$, with probability at least $1 - \delta$, $\|\mathbf{B}\mathbf{B}^T - \mathbf{B}\mathbf{S}\mathbf{S}^T\mathbf{B}\|_F \leq \epsilon n$. Via Weyl's inequality, this shows that the eigenvalues of $\mathbf{B}\mathbf{S}\mathbf{S}^T\mathbf{B}$, and hence \mathbf{A}_S , approximate those of \mathbf{A} up to $\pm \epsilon n$ error.²

Unfortunately, this approach breaks down when \mathbf{A} has negative eigenvalues, and so cannot be factored as $\mathbf{B}\mathbf{B}^T$ for real $\mathbf{B} \in \mathbb{R}^{n \times n}$. This is more than a technical issue: observe that when \mathbf{A} is PSD and has $\|\mathbf{A}\|_\infty \leq 1$, it can have at most $1/\epsilon$ eigenvalues larger than ϵn – since its trace, which is equal to the sum of its eigenvalues, is bounded by n , and since all eigenvalues are non-negative. When \mathbf{A} is not PSD, it can have $\Omega(1/\epsilon^2)$ eigenvalues with magnitude larger than ϵn . In particular, if \mathbf{A} is the tensor product of a $1/\epsilon^2 \times 1/\epsilon^2$ random ± 1 matrix and the $\epsilon^2 n \times \epsilon^2 n$ all ones matrix, the bulk of its eigenvalues (of which there are $1/\epsilon^2$) will concentrate around $1/\epsilon \cdot \epsilon^2 n = \epsilon n$. As a result it remains unclear whether we can match the $1/\epsilon^2$ dependence of the PSD case, or if a stronger lower bound can be shown for indefinite matrices.

Outside the ϵ dependence, it is unknown if full eigenspectrum approximation can be performed with sample complexity independent of the matrix size n . [5] achieve this for the easier positive semidefiniteness testing problem, giving sample complexity $\tilde{O}(1/\epsilon^2)$. However our bounds have additional $\log n$ factors. As discussed, in Appendix B of [10] we give an alternate analysis for Theorem 1, which shows that sampling a $O\left(\frac{\log n}{\epsilon^4 \delta}\right) \times O\left(\frac{\log n}{\epsilon^4 \delta}\right)$ submatrix suffices for $\pm \epsilon n$ eigenvalue approximation, saving a $\log^2 n$ factor at the cost of

² In fact, via more refined eigenvalue perturbation bounds [9] one can show an ℓ_2 norm bound on the eigenvalue approximation errors, which can be much stronger than the ℓ_∞ norm bound of Theorem 1.

worse ϵ dependence. However, removing the final $\log n$ seems difficult – it arises when bounding $\|\mathbf{A}_{m,S}\|_2$ via bounds on the spectral norms of random principal submatrices [38]. Removing it seems as though it would require either improving such bounds, or taking a different algorithmic approach, as simple modifications such as using bounds depending on the intrinsic dimension do not seem to help.

Also note that our $\log n$ and ϵ dependencies for non-uniform sampling (Theorems 3 and 4) are likely not tight. It is not hard to check that the lower bounds of [5] still hold in these settings. For example, in the sparsity-based sampling setting, by simply having the matrix entirely supported on a $\sqrt{\text{nnz}(\mathbf{A})} \times \sqrt{\text{nnz}(\mathbf{A})}$ submatrix, the lower bounds of [5] directly carry over. Giving tight query complexity bounds here would also be interesting. Finally, it would be interesting to go beyond principal submatrix based algorithms, to achieve improved query complexity, as in Corollary 2. Finding an algorithm matching the $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$ overall query complexity lower bound of [5] is open even in the much simpler PSD setting.

2 Notation and Preliminaries

We now define notation and foundational results that we use throughout our work. For any integer n , let $[n]$ denote the set $\{1, 2, \dots, n\}$. We write matrices and vectors in bold literals – e.g., \mathbf{A} or \mathbf{x} . For a vector \mathbf{x} , we let $\|\mathbf{x}\|_2$ denote its Euclidean norm. We denote the eigenvalues of a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ by $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$, in decreasing order. A symmetric matrix is positive semidefinite if all its eigenvalues are non-negative. For two matrices \mathbf{A}, \mathbf{B} , we let $\mathbf{A} \succeq \mathbf{B}$ denote that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. For any matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $i \in [n]$, we let \mathbf{A}_i denote the i^{th} row of \mathbf{A} . We let $\text{nnz}(\mathbf{A})$ denote the total number of non-zero elements in \mathbf{A} , $\|\mathbf{A}\|_\infty$ denote the largest magnitude of an entry, and $\|\mathbf{A}\|_2 = \max_{\mathbf{x}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$ denote the spectral norm. We let $\|\mathbf{A}\|_F = (\sum_{i,j} \mathbf{A}_{ij}^2)^{1/2}$ denote the Frobenius norm, and $\|\mathbf{A}\|_{1 \rightarrow 2}$ denote the maximum Euclidean norm of a column. For $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $S \subseteq [n]$ we let \mathbf{A}_S denote the principal submatrix corresponding to S . We let \mathbb{E}_2 denote the L_2 norm of a random variable, $\mathbb{E}_2[X] = (\mathbb{E}[X^2])^{1/2}$, where $\mathbb{E}[\cdot]$ denotes expectation.

We use the following basic facts and identities on eigenvalues throughout our proofs.

► **Fact 1** (Eigenvalue of Matrix Product). *For any two matrices $\mathbf{A} \in \mathbb{C}^{n \times m}$, $\mathbf{B} \in \mathbb{C}^{m \times n}$, the non-zero eigenvalues of \mathbf{AB} are identical to those of \mathbf{BA} .*

► **Fact 2** (Gershgorin's circle theorem [20]). *Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ with entries \mathbf{A}_{ij} . For $i \in [n]$, let \mathbf{R}_i be the sum of absolute values of non-diagonal entries in the i^{th} row. Let $D(\mathbf{A}_{ii}, \mathbf{R}_i)$ be the closed disc centered at \mathbf{A}_{ii} with radius \mathbf{R}_i . Then every eigenvalue of \mathbf{A} lies within one of the discs $D(\mathbf{A}_{ii}, \mathbf{R}_i)$.*

► **Fact 3** (Weyl's Inequality [50]). *For any two Hermitian matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n \times n}$ with $\mathbf{A} - \mathbf{B} = \mathbf{E}$, $\max_i |\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{B})| \leq \|\mathbf{E}\|_2$.*

Weyl's inequality ensures that a small Hermitian perturbation of a Hermitian matrix will not significantly change its eigenvalues. The bound can be extended to the case when the perturbation is not Hermitian, with a loss of an $O(\log n)$ factor; to the best of our knowledge this loss is necessary:

► **Fact 4** (Non-Hermitian perturbation bound [9]). *Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be Hermitian and $\mathbf{B} \in \mathbb{C}^{n \times n}$ be any matrix whose eigenvalues are $\lambda_1(\mathbf{B}), \dots, \lambda_n(\mathbf{B})$ such that $\text{Re}(\lambda_1(\mathbf{B})) \geq \dots \geq \text{Re}(\lambda_n(\mathbf{B}))$ (where $\text{Re}(\lambda_i(\mathbf{B}))$ denotes the real part of $\lambda_i(\mathbf{B})$). Let $\mathbf{A} - \mathbf{B} = \mathbf{E}$. For some universal constant C , $\max_i |\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{B})| \leq C \log n \|\mathbf{E}\|_2$*

Beyond the above facts, we use several theorems to obtain eigenvalue concentration bounds. We first state a theorem from [44], which bounds the spectral norm of a principal submatrix sampled uniformly at random from a bounded entry matrix. We build on this to prove the full eigenspectrum concentration result of Theorem 1.

► **Theorem 6** (Random principal submatrix spectral norm bound [38, 44]). *Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be Hermitian, decomposed into diagonal and off-diagonal parts: $\mathbf{A} = \mathbf{D} + \mathbf{H}$. Let $\mathbf{S} \in \mathbb{R}^{n \times n}$ be a diagonal sampling matrix with the j^{th} diagonal entry set to 1 independently with probability s/n and 0 otherwise. Then, for some universal constant C ,*

$$\mathbb{E}_2 \|\mathbf{SAS}\|_2 \leq C \left[\log n \cdot \mathbb{E}_2 \|\mathbf{SHS}\|_\infty + \sqrt{\frac{s \log n}{n}} \cdot \mathbb{E}_2 \|\mathbf{HS}\|_{1 \rightarrow 2} + \frac{s}{n} \cdot \|\mathbf{H}\|_2 \right] + \mathbb{E}_2 \|\mathbf{SDS}\|_2.$$

For Theorems 3 and 4, we need an extension of Theorem 6 to the setting where rows are sampled non-uniformly. We will use two bounds here. The first is a decoupling and recoupling result for matrix norms. One can prove this lemma following an analogous result in [44] for sampling rows/columns uniformly. The proof is almost identical so we omit it.

► **Lemma 7** (Decoupling and recoupling). *Let \mathbf{H} be a Hermitian matrix with zero diagonal. Let δ_j be a sequence of independent random variables such that $\delta_j = \frac{1}{\sqrt{p_j}}$ with probability p_j and 0 otherwise. Let \mathbf{S} be a square diagonal sampling matrix with j^{th} diagonal entry set to δ_j . Then:*

$$\mathbb{E}_2 \|\mathbf{SHS}\|_2 \leq 2\mathbb{E}_2 \|\mathbf{SH}\hat{\mathbf{S}}\|_2 \quad \text{and} \quad \mathbb{E}_2 \|\mathbf{SH}\hat{\mathbf{S}}\|_\infty \leq 4\mathbb{E}_2 \|\mathbf{SHS}\|_\infty,$$

where $\hat{\mathbf{S}}$ is an independent diagonal sampling matrix drawn from the same distribution as \mathbf{S} .

The second theorem bounds the spectral norm of a non-uniform random column sample of a matrix. We give a proof for uniform sampling in Appendix D of [10], following the results of [45].

► **Theorem 8** (Non-uniform column sampling – spectral norm bound). *Let \mathbf{A} be an $m \times n$ matrix with rank r . Let δ_j be a sequence of independent random variables such that $\delta_j = \frac{1}{\sqrt{p_j}}$ with probability p_j and 0 otherwise. Let \mathbf{S} be a square diagonal sampling matrix with j^{th} diagonal entry set to δ_j .*

$$\mathbb{E}_2 \|\mathbf{AS}\|_2 \leq 5\sqrt{\log r} \cdot \mathbb{E}_2 \|\mathbf{AS}\|_{1 \rightarrow 2} + \|\mathbf{A}\|_2$$

We use a standard Matrix Bernstein inequality to bound the spectral norm of random submatrices.

► **Theorem 9** (Matrix Bernstein [46]). *Consider a finite sequence $\{\mathbf{S}_k\}$ of random matrices in $\mathbb{R}^{d \times d}$. Assume that for all k , $\mathbb{E}[\mathbf{S}_k] = \mathbf{0}$ and $\|\mathbf{S}_k\|_2 \leq L$. Let $\mathbf{Z} = \sum_k \mathbf{S}_k$ and let $\mathbf{V}_1, \mathbf{V}_2$ be semidefinite upper-bounds for the matrix valued variances $\mathbf{Var}_1(\mathbf{Z})$ and $\mathbf{Var}_2(\mathbf{Z})$:*

$$\mathbf{V}_1 \succeq \mathbf{Var}_1(\mathbf{Z}) \stackrel{\text{def}}{=} \mathbb{E} \left(\mathbf{ZZ}^T \right) = \sum_k \mathbb{E} \left(\mathbf{S}_k \mathbf{S}_k^T \right), \quad \text{and}$$

$$\mathbf{V}_2 \succeq \mathbf{Var}_2(\mathbf{Z}) \stackrel{\text{def}}{=} \mathbb{E} \left(\mathbf{Z}^T \mathbf{Z} \right) = \sum_k \mathbb{E} \left(\mathbf{S}_k^T \mathbf{S}_k \right).$$

Then, letting $v = \max(\|\mathbf{V}_1\|_2, \|\mathbf{V}_2\|_2)$, for any $t \geq 0$,

$$\mathbb{P}(\|\mathbf{Z}\|_2 \geq t) \leq 2d \cdot \exp \left(\frac{-t^2/2}{v + Lt/3} \right).$$

21:12 Sublinear Time Eigenvalue Approximation via Random Sampling

For real valued random variables, we use the standard Bernstein inequality.

► **Theorem 10** (Bernstein inequality [8]). *Let $\{z_j\}$ for $j \in [n]$ be independent random variables with zero mean such that $|z_j| \leq M$ for all j . Then for all positive t ,*

$$\mathbb{P} \left(\left| \sum_{j=1}^n z_j \right| \geq t \right) \leq \exp \left(\frac{-t^2/2}{\sum_{i=1}^n \mathbb{E}[z_i^2] + Mt/3} \right).$$

3 Sublinear Time Eigenvalue Estimation using Uniform Sampling

We now prove our main eigenvalue estimation result – Theorem 1. We give the pseudocode for our principal submatrix based estimation procedure in Algorithm 1. We will show that any positive or negative eigenvalue of \mathbf{A} with magnitude $\geq \epsilon n$ will appear as an approximate eigenvalue in \mathbf{A}_S with good probability. Thus, in step 5 of Algorithm 1, the positive and negative eigenvalues of \mathbf{A}_S are used to estimate the outlying largest and smallest eigenvalues of \mathbf{A} . All other interior eigenvalues of \mathbf{A} are estimated to be 0, which will immediately give our $\pm \epsilon n$ approximation bound when the original eigenvalue has magnitude $\leq \epsilon n$.

■ **Algorithm 1** Eigenvalue estimator using uniform sampling.

-
- 1: **Input:** Symmetric $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\|\mathbf{A}\|_\infty \leq 1$, Accuracy $\epsilon \in (0, 1)$, failure prob. $\delta \in (0, 1)$.
 - 2: Fix $s = \frac{c \log(1/(\epsilon\delta)) \cdot \log^3 n}{\epsilon^3 \delta}$ where c is a sufficiently large constant.
 - 3: Add each index $i \in [n]$ to the sample set S independently with probability $\frac{s}{n}$. Let the principal submatrix of \mathbf{A} corresponding S be \mathbf{A}_S .
 - 4: Compute the eigenvalues of \mathbf{A}_S : $\lambda_1(\mathbf{A}_S) \geq \dots \geq \lambda_{|S|}(\mathbf{A}_S)$.
 - 5: For all $i \in [|S|]$ with $\lambda_i(\mathbf{A}_S) \geq 0$, let $\tilde{\lambda}_i(\mathbf{A}) = \frac{n}{s} \cdot \lambda_i(\mathbf{A}_S)$. For all $i \in [|S|]$ with $\lambda_i(\mathbf{A}_S) < 0$, let $\tilde{\lambda}_{n-(|S|-i)}(\mathbf{A}) = \frac{n}{s} \cdot \lambda_i(\mathbf{A}_S)$. For all remaining $i \in [n]$, let $\tilde{\lambda}_i(\mathbf{A}) = 0$.
 - 6: **Return:** Eigenvalue estimates $\tilde{\lambda}_1(\mathbf{A}) \geq \dots \geq \tilde{\lambda}_n(\mathbf{A})$.
-

Running time. Observe that the expected number of indices chosen by Algorithm 1 is $s = \frac{c \log(1/(\epsilon\delta)) \cdot \log^3 n}{\epsilon^3 \delta}$. A standard concentration bound can be used to show that with high probability $(1 - 1/\text{poly}(n))$, the number of sampled entries is $O(s)$. Thus, the algorithm reads a total of $O(s^2)$ entries of \mathbf{A} and runs in $O(s^\omega)$ time – the time to compute a full eigendecomposition of \mathbf{A}_S .

3.1 Outer and Middle Eigenvalue Bounds

Recall that we will split \mathbf{A} into two symmetric matrices (Definition 5): $\mathbf{A}_o = \mathbf{V}_o \mathbf{\Lambda}_o \mathbf{V}_o^T$ which contains its large magnitude (outlying) eigendirections with eigenvalue magnitudes $\geq \epsilon \sqrt{\delta} n$ and $\mathbf{A}_m = \mathbf{V}_m \mathbf{\Lambda}_m \mathbf{V}_m^T$ which contains its small magnitude (middle) eigendirections.

We first show that the eigenvectors in \mathbf{V}_o are *incoherent*. I.e., that their (eigenvalue weighted) squared row norms are bounded. This ensures that the outlying eigenspace of \mathbf{A} is well-approximated via uniform sampling.

► **Lemma 11** (Incoherence of outlying eigenvectors). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric with $\|\mathbf{A}\|_\infty \leq 1$. Let \mathbf{V}_o be as in Definition 5. Let $\mathbf{V}_{o,i}$ denote the i^{th} row of \mathbf{V}_o . Then,*

$$\|\mathbf{A}_o^{1/2} \mathbf{V}_{o,i}\|_2^2 \leq \frac{1}{\epsilon \sqrt{\delta}} \quad \text{and} \quad \|\mathbf{V}_{o,i}\|_2^2 \leq \frac{1}{\epsilon^2 \delta n}.$$

Proof. Observe that $\mathbf{AV}_o = \mathbf{V}_o\mathbf{\Lambda}_o$. Let $[\mathbf{AV}_o]_i$ denote the i^{th} row of the \mathbf{AV}_o . Then we have

$$\|[\mathbf{AV}_o]_i\|_2^2 = \|[\mathbf{V}_o\mathbf{\Lambda}_o]_i\|_2^2 = \sum_{j=1}^r \lambda_j^2 \cdot \mathbf{V}_{o,i,j}^2, \quad (2)$$

where $r = \text{rank}(\mathbf{A}_o)$, $\mathbf{V}_{o,i,j}$ is the $(i, j)^{\text{th}}$ element of \mathbf{V}_o and $\lambda_j = \mathbf{\Lambda}_o(j, j)$. $\|\mathbf{A}\|_\infty \leq 1$ by assumption and since \mathbf{V}_o has orthonormal columns, its spectral norm is bounded by 1, thus we have $\|[\mathbf{AV}_o]_i\|_2^2 = \|[\mathbf{A}]_i\mathbf{V}_o\|_2^2 \leq \|[\mathbf{A}]_i\|_2^2 \cdot \|\mathbf{V}_o\|_2^2 \leq n$. Therefore, by (2), we have:

$$\sum_{j=1}^r \lambda_j^2 \cdot \mathbf{V}_{o,i,j}^2 \leq n. \quad (3)$$

Since by definition of $\mathbf{\Lambda}_o$, $|\lambda_j| \geq \epsilon\sqrt{\delta}n$ for all j , we finally have $\|\mathbf{\Lambda}_o^{1/2}\mathbf{V}_{o,i}\|_2^2 = \sum_{j=1}^r \lambda_j \cdot \mathbf{V}_{o,i,j}^2 \leq \frac{n}{\epsilon\sqrt{\delta}n} = \frac{1}{\epsilon\sqrt{\delta}}$ and $\|\mathbf{V}_{o,i}\|_2^2 = \sum_{j=1}^r \mathbf{V}_{o,i,j}^2 \leq \frac{n}{\epsilon^2\delta n^2} = \frac{1}{\epsilon^2\delta n}$. \blacktriangleleft

Let $\bar{\mathbf{S}} \in \mathbb{R}^{n \times |S|}$ be the scaled sampling matrix satisfying $\bar{\mathbf{S}}^T \mathbf{A} \bar{\mathbf{S}} = \frac{n}{s} \cdot \mathbf{A}_S$. We next apply Lemma 11 in conjunction with a matrix Bernstein bound to show that $\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2}$ concentrates around its expectation, $\mathbf{\Lambda}_o$. Since by Fact 1, this matrix has identical eigenvalues to $\frac{n}{s} \cdot \mathbf{A}_{o,S} = \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o \mathbf{V}_o^T \bar{\mathbf{S}}$, this allows us to argue that the eigenvalues of $\frac{n}{s} \cdot \mathbf{A}_{o,S}$ approximate those of $\mathbf{\Lambda}_o$.

► Lemma 12 (Concentration of outlying eigenvalues). *Let $S \subseteq [n]$ be sampled as in Algorithm 1 for $s \geq \frac{c \log(1/(\epsilon\delta))}{\epsilon^3\sqrt{\delta}}$ where c is a sufficiently large constant. Let $\bar{\mathbf{S}} \in \mathbb{R}^{n \times |S|}$ be the scaled sampling matrix satisfying $\bar{\mathbf{S}}^T \mathbf{A} \bar{\mathbf{S}} = \frac{n}{s} \cdot \mathbf{A}_S$. Letting $\mathbf{\Lambda}_o, \mathbf{V}_o$ be as in Definition 5, with probability at least $1 - \delta$,*

$$\|\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2} - \mathbf{\Lambda}_o\|_2 \leq \epsilon n.$$

Proof. Define $\mathbf{E} = \mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2} - \mathbf{\Lambda}_o$. For all $i \in [n]$, let $\mathbf{V}_{o,i}$ be the i^{th} row of \mathbf{V}_o and define the matrix valued random variable

$$\mathbf{Y}_i = \begin{cases} \frac{n}{s} \mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i} \mathbf{V}_{o,i}^T \mathbf{\Lambda}_o^{1/2}, & \text{with probability } s/n \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Define $\mathbf{Q}_i = \mathbf{Y}_i - \mathbb{E}[\mathbf{Y}_i]$. Observe that $\mathbf{Q}_1, \dots, \mathbf{Q}_n$ are independent random variables and that $\sum_{i=1}^n \mathbf{Q}_i = \mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2} - \mathbf{\Lambda}_o = \mathbf{E}$. Further, observe that $\|\mathbf{Q}_i\|_2 \leq \max(1, \frac{n}{s} - 1) \cdot \|\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i} \mathbf{V}_{o,i}^T \mathbf{\Lambda}_o^{1/2}\|_2 \leq \max(1, \frac{n}{s} - 1) \cdot \|\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i}\|_2^2$. Now, $\|\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i}\|_2^2 \leq \frac{1}{\epsilon\sqrt{\delta}}$ by Lemma 11. Thus, $\|\mathbf{Q}_i\|_2 \leq \frac{n}{\epsilon\sqrt{\delta}s}$. The variance $\text{Var}(\mathbf{E}) \stackrel{\text{def}}{=} \mathbb{E}(\mathbf{E}\mathbf{E}^T) = \mathbb{E}(\mathbf{E}^T \mathbf{E}) = \sum_{i=1}^n \mathbb{E}[\mathbf{Q}_i^2]$ can be bounded as:

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[\mathbf{Q}_i^2] &= \sum_{i=1}^n \left[\frac{s}{n} \cdot \left(\frac{n}{s} - 1\right)^2 + \left(1 - \frac{s}{n}\right) \right] \cdot (\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i} \mathbf{V}_{o,i}^T \mathbf{\Lambda}_o \mathbf{V}_{o,i} \mathbf{V}_{o,i}^T \mathbf{\Lambda}_o^{1/2}) \\ &\leq \sum_{i=1}^n \frac{n}{s} \cdot \|\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i}\|_2^2 \cdot (\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i} \mathbf{V}_{o,i}^T \mathbf{\Lambda}_o^{1/2}). \end{aligned} \quad (5)$$

Again by Lemma 11, $\|\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i}\|_2^2 \leq \frac{1}{\epsilon\sqrt{\delta}}$. Plugging back into (5) we can bound,

$$\sum_{i=1}^n \mathbb{E}[\mathbf{Q}_i^2] \leq \sum_{i=1}^n \frac{n}{s} \cdot \frac{1}{\epsilon\sqrt{\delta}} \cdot (\mathbf{\Lambda}_o^{1/2} \mathbf{V}_{o,i} \mathbf{V}_{o,i}^T \mathbf{\Lambda}_o^{1/2}) = \frac{n}{s\epsilon\sqrt{\delta}} \mathbf{\Lambda}_o \preceq \frac{n^2}{s\epsilon\sqrt{\delta}} \cdot \mathbf{I}.$$

21:14 Sublinear Time Eigenvalue Approximation via Random Sampling

Since \mathbf{Q}_i^2 is PSD, this establishes that $\|\mathbf{Var}(\mathbf{E})\|_2 \leq \frac{n^2}{s\epsilon\sqrt{\delta}}$. We then apply Theorem 9 (the matrix Bernstein inequality) with $L = \frac{n}{s\epsilon\sqrt{\delta}}$, $v = \frac{n^2}{s\epsilon\sqrt{\delta}}$, and $d \leq \frac{1}{\epsilon^2\delta}$ since there are at most $\frac{\|\mathbf{A}\|_F^2}{\delta\epsilon^2n^2} \leq \frac{1}{\epsilon^2\delta}$ outlying eigenvalues with magnitude $\geq \sqrt{\delta}\epsilon n$ in \mathbf{A}_o . This gives:

$$\begin{aligned} \mathbb{P}(\|\mathbf{E}\|_2 \geq \epsilon n) &\leq \frac{2}{\epsilon^2\delta} \cdot \exp\left(\frac{-\epsilon^2n^2/2}{v + L\epsilon n/3}\right) \leq \frac{2}{\epsilon^2\delta} \cdot \exp\left(\frac{-\epsilon^2n^2/2}{\frac{n^2}{s\epsilon\sqrt{\delta}} + \frac{\epsilon n^2}{3s\epsilon\sqrt{\delta}}}\right) \\ &\leq \frac{2}{\epsilon^2\delta} \cdot \exp\left(\frac{-s\epsilon^3\sqrt{\delta}}{4}\right). \end{aligned}$$

Thus, if we set $s \geq \frac{c \log(1/(\epsilon\delta))}{\epsilon^3\sqrt{\delta}}$ for large enough c , then the probability is bounded above by δ , completing the proof. \blacktriangleleft

We cannot prove an analogous leverage score bound to Lemma 11 for the interior eigenvectors of \mathbf{A} appearing in \mathbf{V}_m . Thus we cannot apply a matrix Bernstein bound as in Lemma 12. However, we can use Theorem 6 to show that the spectral norm of the random principal submatrix $\mathbf{A}_{m,S}$ is not too large, and thus that the eigenvalues of $\mathbf{A}_S = \mathbf{A}_{o,S} + \mathbf{A}_{m,S}$ are close to those of $\mathbf{A}_{o,S}$.

► Lemma 13 (Spectral norm bound – sampled middle eigenvalues). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric with $\|\mathbf{A}\|_\infty \leq 1$. Let \mathbf{A}_m be as in Definition 5. Let S be sampled as in Algorithm 1. If $s \geq \frac{c \log n}{\epsilon^2\delta}$ for some sufficiently large constant c , then with probability at least $1 - \delta$, $\|\mathbf{A}_{m,S}\|_2 \leq \epsilon s$.*

Proof. Let $\mathbf{A}_m = \mathbf{D}_m + \mathbf{H}_m$ where \mathbf{D}_m is the matrix of diagonal elements and \mathbf{H}_m the matrix of off-diagonal elements. Let $\mathbf{S} \in \mathbb{R}^{n \times |S|}$ be the binary sampling matrix with $\mathbf{A}_{m,S} = \mathbf{S}^T \mathbf{A}_m \mathbf{S}$. From Theorem 6, we have for some constant C ,

$$\mathbb{E}_2[\|\mathbf{A}_{m,S}\|_2] \leq C \left[\log n \cdot \mathbb{E}_2[\|\mathbf{S}^T \mathbf{H}_m \mathbf{S}\|_\infty] + \sqrt{\frac{s \log n}{n}} \mathbb{E}_2[\|\mathbf{H}_m \mathbf{S}\|_{1 \rightarrow 2}] + \frac{s}{n} \|\mathbf{H}_m\|_2 \right] + \mathbb{E}_2[\|\mathbf{S}^T \mathbf{D}_m \mathbf{S}\|]. \quad (6)$$

Considering the various terms in (6), we have $\|\mathbf{S}^T \mathbf{H}_m \mathbf{S}\|_\infty \leq \|\mathbf{A}_m\|_\infty$ and $\|\mathbf{S}^T \mathbf{D}_m \mathbf{S}\|_2 = \|\mathbf{S}^T \mathbf{D}_m \mathbf{S}\|_\infty \leq \|\mathbf{A}_m\|_\infty$. We also have

$$\|\mathbf{H}_m\|_2 \leq \|\mathbf{A}_m\|_2 + \|\mathbf{D}_m\|_2 \leq \|\mathbf{A}_m\|_2 + \|\mathbf{A}_m\|_\infty \leq \epsilon\delta^{1/2}n + \|\mathbf{A}_m\|_\infty$$

and

$$\|\mathbf{H}_m \mathbf{S}\|_{1 \rightarrow 2} \leq \|\mathbf{A}_m \mathbf{S}\|_{1 \rightarrow 2} \leq \|\mathbf{A}_m\|_{1 \rightarrow 2} \leq \sqrt{n}.$$

The final bound follows since $\mathbf{A}_m = \mathbf{V}_m \mathbf{V}_m^T \mathbf{A}$, where $\mathbf{V}_m \mathbf{V}_m^T$ is an orthogonal projection matrix. Thus, $\|\mathbf{A}_m\|_{1 \rightarrow 2} \leq \|\mathbf{A}\|_{1 \rightarrow 2} \leq \sqrt{n}$ by our assumption that $\|\mathbf{A}\|_\infty \leq 1$. Plugging all these bounds into (6) we have, for some constant C ,

$$\mathbb{E}_2[\|\mathbf{A}_{m,S}\|_2] \leq C \left[\log n \cdot \|\mathbf{A}_m\|_\infty + \sqrt{\log n \cdot s} + s \cdot \epsilon\delta^{1/2} \right]. \quad (7)$$

It remains to bound $\|\mathbf{A}_m\|_\infty$. We have $\mathbf{A} = \mathbf{A}_m + \mathbf{A}_o$ and thus by triangle inequality,

$$\|\mathbf{A}_m\|_\infty \leq \|\mathbf{A}\|_\infty + \|\mathbf{A}_o\|_\infty = 1 + \|\mathbf{A}_o\|_\infty. \quad (8)$$

Writing $\mathbf{A}_o = \mathbf{V}_o \mathbf{\Lambda}_o \mathbf{V}_o^T$ (see Definition 5), and letting $\mathbf{V}_{o,i}$ denote the i^{th} row of \mathbf{V}_o , the $(i, j)^{\text{th}}$ element of \mathbf{A}_o has magnitude

$$|\mathbf{A}_{o,i,j}| = |\mathbf{V}_{o,i} \mathbf{\Lambda}_o \mathbf{V}_{o,j}^T| \leq \|\mathbf{V}_{o,i}\|_2 \cdot \|\mathbf{\Lambda}_o \mathbf{V}_{o,j}^T\|_2,$$

by Cauchy-Schwarz. From Lemma 11, we have $\|\mathbf{V}_{o,i}\|_2 \leq \frac{1}{\epsilon \delta^{1/2} \sqrt{n}}$. Also, from (2), $\|\mathbf{\Lambda}_o \mathbf{V}_{o,j}^T\|_2 = \|[\mathbf{A} \mathbf{V}_o]_j\|_2 \leq \sqrt{n}$. Overall, for all i, j we have $\mathbf{A}_{o,i,j} \leq \frac{1}{\epsilon \delta^{1/2} \sqrt{n}} \cdot \sqrt{n} = \frac{1}{\epsilon \delta^{1/2}}$, giving $\|\mathbf{A}_o\|_\infty \leq \frac{1}{\epsilon \delta^{1/2}}$. Plugging back into (8) and in turn (7), we have for some constant C ,

$$\mathbb{E}_2[\|\mathbf{A}_{m,S}\|_2] \leq C \left[\frac{\log n}{\epsilon \delta^{1/2}} + \sqrt{s \log n} + s \epsilon \delta^{1/2} \right].$$

Setting $s \geq \frac{c \log n}{\epsilon^2 \delta}$ for sufficiently large c , all terms in the right hand side of the above equation are bounded by $\epsilon \sqrt{\delta} s$ and so

$$\mathbb{E}_2[\|\mathbf{A}_{m,S}\|_2] \leq 3\epsilon \sqrt{\delta} s$$

Thus, by Markov's inequality, with probability at least $1 - \delta$, we have $\|\mathbf{A}_{m,S}\|_2 \leq 3\epsilon s$. We can adjust ϵ by a constant to obtain the required bound. \blacktriangleleft

3.2 Main Accuracy Bounds

We now restate our main result, and give its proof via Lemmas 12 and 13.

► Theorem 1 (Sublinear Time Eigenvalue Approximation). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric with $\|\mathbf{A}\|_\infty \leq 1$ and eigenvalues $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$. Let $S \subseteq [n]$ be formed by including each index independently with probability s/n as in Algorithm 1. Let \mathbf{A}_S be the corresponding principal submatrix of \mathbf{A} , with eigenvalues $\lambda_1(\mathbf{A}_S) \geq \dots \geq \lambda_{|S|}(\mathbf{A}_S)$.*

For all $i \in [|S|]$ with $\lambda_i(\mathbf{A}_S) \geq 0$, let $\tilde{\lambda}_i(\mathbf{A}) = \frac{n}{s} \cdot \lambda_i(\mathbf{A}_S)$. For all $i \in [|S|]$ with $\lambda_i(\mathbf{A}_S) < 0$, let $\tilde{\lambda}_{n-(|S|-i)}(\mathbf{A}) = \frac{n}{s} \cdot \lambda_i(\mathbf{A}_S)$. For all other $i \in [n]$, let $\tilde{\lambda}_i(\mathbf{A}) = 0$. If $s \geq \frac{c \log(1/(\epsilon \delta)) \cdot \log^3 n}{\epsilon^3 \delta}$, for large enough constant c , then with probability $\geq 1 - \delta$, for all $i \in [n]$, $\lambda_i(\mathbf{A}) - \epsilon n \leq \tilde{\lambda}_i(\mathbf{A}) \leq \lambda_i(\mathbf{A}) + \epsilon n$.

Proof. Let $\mathbf{S} \in \mathbb{R}^{n \times |S|}$ be the binary sampling matrix with a single one in each column such that $\mathbf{S}^T \mathbf{A} \mathbf{S} = \mathbf{A}_S$. Let $\bar{\mathbf{S}} = \sqrt{n/s} \cdot \mathbf{S}$. Following Definition 5, we write $\mathbf{A} = \mathbf{A}_o + \mathbf{A}_m$. By Fact 1 we have that the nonzero eigenvalues of $\frac{n}{s} \cdot \mathbf{A}_{o,S} = \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o \mathbf{V}_o^T \bar{\mathbf{S}}$ are identical to those of $\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2}$ where $\mathbf{\Lambda}_o^{1/2}$ is the square root matrix of $\mathbf{\Lambda}_o$ such that $\mathbf{\Lambda}_o^{1/2} \mathbf{\Lambda}_o^{1/2} = \mathbf{\Lambda}_o$.

Note that $\mathbf{\Lambda}_o$ is Hermitian. However $\mathbf{\Lambda}_o^{1/2}$ may be complex, and hence $\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2}$ is *not necessarily Hermitian*, although it does have real eigenvalues. Thus, we can apply the perturbation bound of Fact 4 to $\mathbf{\Lambda}_o$ and $\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2}$ to claim for all $i \in [n]$, and some constant C ,

$$|\lambda_i(\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2}) - \lambda_i(\mathbf{\Lambda}_o)| \leq C \log n \|\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2} - \mathbf{\Lambda}_o\|_2.$$

By Lemma 12 applied with error $\frac{\epsilon}{2C \log n}$, with probability at least $1 - \delta$, for any $s \geq \frac{c \log(1/(\epsilon \delta)) \cdot \log^3 n}{\epsilon^3 \sqrt{\delta}}$ (for a large enough constant c) we have $\|\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2} - \mathbf{\Lambda}_o\|_2 \leq \frac{\epsilon n}{2C \log n}$. Thus, for all i ,

$$\left| \lambda_i(\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2}) - \lambda_i(\mathbf{\Lambda}_o) \right| < \frac{\epsilon n}{2}. \quad (9)$$

We note that the conceptual part of the proof is essentially complete: the nonzero eigenvalues of $\frac{n}{s} \cdot \mathbf{A}_{o,S}$ are identical to those of $\mathbf{\Lambda}_o^{1/2} \mathbf{V}_o^T \bar{\mathbf{S}} \bar{\mathbf{S}}^T \mathbf{V}_o \mathbf{\Lambda}_o^{1/2}$, which we have shown well approximate those of $\mathbf{\Lambda}_o$ and in turn \mathbf{A}_o . i.e., the non-zero eigenvalues of $\frac{n}{s} \cdot \mathbf{A}_{o,S}$ approximate all outlying

eigenvalues of \mathbf{A} . It remains to carefully argue how these approximations should be “lined up” given the presence of zero eigenvalues in the spectrum of these matrices. We also must account for the impact of the interior eigenvalues in $\mathbf{A}_{m,S}$, which is limited by the spectral norm bound of Lemma 13. The rest of the argument is completed in Theorem 1 of [10]. ◀

► **Remark.** The proof of Lemma 12 and consequently, Theorem 1 can be modified to give better bounds for the case when the eigenvalues of \mathbf{A}_o lie in a bounded range – between $\epsilon^a \sqrt{\delta} n$ and $\epsilon^b n$ where $0 \leq b \leq a \leq 1$. See Theorem 9 in Appendix C of [10] for details. For example, if all the top eigenvalues are equal, one can show that $s = \tilde{O}\left(\frac{\log^2 n}{\epsilon^2}\right)$ suffices to give $\pm \epsilon n$ error, nearly matching the lower bound of [5]. This indicates that improving Theorem 1 in general requires tackling the case when the outlying eigenvalues in \mathbf{A}_o have a wide range.

4 Conclusion

We present efficient algorithms for estimating all eigenvalues of a symmetric matrix with bounded entries up to additive error ϵn , by reading just a $\text{poly}(\log n, 1/\epsilon) \times \text{poly}(\log n, 1/\epsilon)$ random principal submatrix. We give improved error bounds of $\epsilon \sqrt{\text{nnz}(\mathbf{A})}$ and $\epsilon \|\mathbf{A}\|_F$ when the rows/columns are sampled with probabilities proportional to their sparsities or squared ℓ_2 norms, respectively (see Section 4 and Appendix E of [10]). We also perform numerical simulations which demonstrate the effectiveness of our algorithms in practice (see Section 5 of [10]).

Our work leaves several open questions. In particular, it is open if our query complexity for $\pm \epsilon n$ approximation can be improved, possibly to $\tilde{O}(\log^c n / \epsilon^4)$ total entries using principal submatrix queries or $\tilde{O}(\log^c / \epsilon^2)$ entries using general queries. The later bound is open even when \mathbf{A} is PSD, a setting where we know that sampling a $O(1/\epsilon^2) \times O(1/\epsilon^2)$ principal submatrix (with $O(1/\epsilon^4)$ total entries) does suffice. Additionally, it is open if we can achieve sample complexity independent of n , by removing all $\log n$ factors, as have been done for the easier problem of testing positive semidefiniteness [5]. See Section 1.4 for more details. Finally, it would be interesting to identify additional assumptions on \mathbf{A} or on the sampling model where stronger approximation guarantees (e.g., relative error) can be achieved in sublinear time.

References

- 1 Dimitris Achlioptas and Frank McSherry. Fast computation of low-rank matrix approximations. *Journal of the ACM (JACM)*, 54(2):9–es, 2007.
- 2 Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2021.
- 3 Alexandr Andoni and Huy L Nguyễn. Eigenvalues of a matrix in the streaming model. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2013.
- 4 Arturs Backurs, Piotr Indyk, Cameron Musco, and Tal Wagner. Faster kernel matrix algebra via density estimation. *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.
- 5 Ainesh Bakshi, Nadiia Chepurko, and Rajesh Jayaram. Testing positive semi-definiteness via random submatrices. *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2020.
- 6 Maria-Florina Balcan, Yi Li, David P Woodruff, and Hongyang Zhang. Testing matrix rank, optimally. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2019.

- 7 Itai Benjamini, Oded Schramm, and Asaf Shapira. Every minor-closed property of sparse graphs is testable. *Advances in Mathematics*, 223(6):2200–2218, 2010.
- 8 Serge Bernstein. Sur l’extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, 97(1):1–59, 1927.
- 9 Rajendra Bhatia. *Matrix analysis*. Springer Science & Business Media, 2013.
- 10 Rajarshi Bhattacharjee, Gregory Dexter, Petros Drineas, Cameron Musco, and Archan Ray. Sublinear time eigenvalue approximation via random sampling. *arXiv preprint*, 2021. [arXiv:2109.07647](#).
- 11 Vladimir Braverman, Aditya Krishnan, and Christopher Musco. Linear and sublinear time spectral density estimation. *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC)*, 2022.
- 12 Nadiia Chepurko, Kenneth L Clarkson, Lior Horesh, Honghao Lin, and David P Woodruff. Quantum-inspired algorithms from randomized numerical linear algebra. *arXiv*, 2020. [arXiv:2011.04125](#).
- 13 David Cohen-Steiner, Weihao Kong, Christian Sohler, and Gregory Valiant. Approximating the spectrum of a graph. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2018.
- 14 James Demmel, Ioana Dumitriu, Olga Holtz, and Robert Kleinberg. Fast matrix multiplication is stable. *Numerische Mathematik*, 2007.
- 15 Kun Dong, Austin R Benson, and David Bindel. Network density of states. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2019.
- 16 Petros Drineas and Ravi Kannan. Fast monte-carlo algorithms for approximate matrix multiplication. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2001.
- 17 Talya Eden and Will Rosenbaum. On sampling edges almost uniformly. *SIAM Symposium on Simplicity in Algorithms (SOSA)*, 2018.
- 18 Noureddine El Karoui. The spectrum of kernel random matrices. *The Annals of Statistics*, 38(1):1–50, 2010.
- 19 Alan Frieze, Ravi Kannan, and Santosh Vempala. Fast Monte-Carlo algorithms for finding low-rank approximations. *Journal of the ACM (JACM)*, 51(6):1025–1041, 2004.
- 20 Semyon Aranovich Gershgorin. Über die abgrenzung der eigenwerte einer matrix. *Izvestiya Rossiyskoy akademii nauk. Seriya matematicheskaya*, 6:749–754, 1931.
- 21 Behrooz Ghorbani, Shankar Krishnan, and Ying Xiao. An investigation into neural net optimization via hessian eigenvalue density. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019.
- 22 Alex Gittens and Joel A Tropp. Tail bounds for all eigenvalues of a sum of random matrices. *arXiv*, 2011. [arXiv:1104.4513](#).
- 23 Oded Goldreich and Dana Ron. Property testing in bounded degree graphs. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, 1997.
- 24 Oded Goldreich and Dana Ron. Approximating average parameters of graphs. *Random Structures & Algorithms*, 32(4):473–493, 2008.
- 25 Leslie Greengard and John Strain. The fast gauss transform. *SIAM Journal on Scientific and Statistical Computing*, 1991.
- 26 Ming Gu and Stanley C Eisenstat. A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem. *SIAM Journal on Matrix Analysis and Applications*, 1995.
- 27 Moritz Hardt and Eric Price. The noisy power method: A meta algorithm with applications. *Advances in Neural Information Processing Systems 27 (NIPS)*, 2014.
- 28 Jonas Helsen, Francesco Battistel, and Barbara M Terhal. Spectral quantum tomography. *Quantum Information*, 2019.
- 29 Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, USA, 2nd edition, 2012.

- 30 Robert Krauthgamer and Ori Sasson. Property testing of data dimensionality. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2003.
- 31 Ruipeng Li, Yuanzhe Xi, Lucas Erlandson, and Yousef Saad. The eigenvalues slicing library (EVSL): Algorithms, implementation, and software. *SIAM Journal on Scientific Computing*, 2019.
- 32 Yi Li, Huy L Nguyễn, and David P Woodruff. On sketching matrix norms and the top singular vector. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.
- 33 Yi Li, Zhengyu Wang, and David P Woodruff. Improved testing of low rank matrices. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2014.
- 34 Yi Li and David P Woodruff. On approximating functions of the singular values in a stream. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, 2016.
- 35 Lin Lin, Yousef Saad, and Chao Yang. Approximating spectral densities of large matrices. *SIAM Review*, 2016.
- 36 Milena Mihail and Christos Papadimitriou. On the eigenvalue power law. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 254–262. Springer, 2002.
- 37 Deanna Needell, William Swartworth, and David P Woodruff. Testing positive semidefiniteness using linear measurements. In *Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2022.
- 38 Mark Rudelson and Roman Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *Journal of the ACM (JACM)*, 2007.
- 39 Yousef Saad. *Numerical methods for large eigenvalue problems: revised edition*. SIAM, 2011.
- 40 Levent Sagun, Leon Bottou, and Yann LeCun. Eigenvalues of the hessian in deep learning: Singularity and beyond. *arXiv*, 2016. [arXiv:1611.07476](https://arxiv.org/abs/1611.07476).
- 41 RN Silver and H Röder. Densities of states of mega-dimensional Hamiltonian matrices. *International Journal of Modern Physics C*, 1994.
- 42 G. W. Stewart and Ji guang Sun. *Matrix Perturbation Theory*. Academic Press, 1990.
- 43 Ewin Tang. Quantum-inspired classical algorithms for principal component analysis and supervised clustering. *arXiv*, 2018. [arXiv:1811.00414](https://arxiv.org/abs/1811.00414).
- 44 Joel A Tropp. Norms of random submatrices and sparse approximation. *Comptes Rendus Mathématique*, 2008.
- 45 Joel A. Tropp. The random paving property for uniformly bounded matrices. *Studia Mathematica*, 185:67–82, 2008.
- 46 Joel A Tropp. An introduction to matrix concentration inequalities. *arXiv*, 2015. [arXiv:1501.01571](https://arxiv.org/abs/1501.01571).
- 47 Madeleine Udell and Alex Townsend. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, 2019.
- 48 Lin-Wang Wang. Calculating the density of states and optical-absorption spectra of large quantum systems by the plane-wave moments method. *Physical Review B*, 1994.
- 49 Alexander Weiße, Gerhard Wellein, Andreas Alvermann, and Holger Fehske. The kernel polynomial method. *Reviews of Modern Physics*, 2006.
- 50 Hermann Weyl. The asymptotic distribution law of the eigenvalues of linear partial differential equations (with an application to the theory of cavity radiation). *Mathematical Annals*, 1912.
- 51 Zhewei Yao, Amir Gholami, Qi Lei, Kurt Keutzer, and Michael W Mahoney. Hessian-based analysis of large batch training and robustness to adversaries. *arXiv*, 2018. [arXiv:1802.08241](https://arxiv.org/abs/1802.08241).