Testing Positive Semi-Definiteness via Random Submatrices

Ainesh Bakshi
Carnegie Mellon University
Pittsburgh, PA
abakshi@cs.cmu.edu

Nadiia Chepurko
Massachusetts Institute of Technology
Cambridge, MA
nadiia@mit.edu

Rajesh Jayaram
Carnegie Mellon University
Pittsburgh, PA
rkjayara@cs.cmu.edu

***Abstract*—We study the problem of testing whether a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with bounded entries ($\|\mathbf{A}\|_\infty \leq 1$) is positive semi-definite (PSD), or ϵ-far in Euclidean distance from the PSD cone, meaning that $\min_{\mathbf{B} \succeq 0} \|\mathbf{A} - \mathbf{B}\|_F^2 > \epsilon n^2$, where $\mathbf{B} \succeq 0$ denotes that $\mathbf{B}$ is PSD. Our main algorithmic contribution is a non-adaptive tester which distinguishes between these cases using only $\tilde{O}(1/\epsilon^4)$ queries to the entries of $\mathbf{A}$.[1] If instead of the Eucledian norm we considered the distance in spectral norm, we obtain the "ℓ_∞-gap problem", where $\mathbf{A}$ is either PSD or satisfies $\min_{\mathbf{B} \succeq 0} \|\mathbf{A} - \mathbf{B}\|_2 > \epsilon n$. For this related problem, we give a $\tilde{O}(1/\epsilon^2)$ query tester, which we show is optimal up to $\log(1/\epsilon)$ factors. Both our testers randomly sample a collection of principal sub-matrices and check whether these sub-matrices are PSD. Consequentially, our algorithms achieve *one-sided error*: whenever they output that $\mathbf{A}$ is not PSD, they return a certificate that $\mathbf{A}$ has negative eigenvalues.**

We complement our upper bound for PSD testing with Eucledian norm distance by giving a $\tilde{\Omega}(1/\epsilon^2)$ lower bound for any non-adaptive algorithm. Our lower bound construction is general, and can be used to derive lower bounds for a number of spectral testing problems. As an example of the applicability of our construction, we obtain a new $\tilde{\Omega}(1/\epsilon^4)$ sampling lower bound for testing the Schatten-1 norm with a $\epsilon n^{1.5}$ gap, extending a result of Balcan, Li, Woodruff, and Zhang [11]. In addition, our hard instance results in new sampling lower bounds for estimating the Ky-Fan Norm, and the cost of rank-k approximations, i.e. $\|\mathbf{A} - \mathbf{A}_k\|_F^2 = \sum_{i>k} \sigma_i^2(\mathbf{A})$.

***Keywords*-Property Testing; Positive Semi-Definiteness; Sampling; Random Matrices**

For a full version of this paper see https://arxiv.org/abs/2005.06441

I. Introduction

Positive Semi-Definite (PSD) matrices are central objects of interest in algorithm design, and continue to be studied extensively in optimization, spectral graph theory, numerical linear algebra, statistics, and dynamical systems, among many others [6], [7], [21], [22], [29], [63], [65], [67], [72], [74]. Specifically, a real-valued matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be PSD if it defines a non-negative quadratic form: namely if $x^\top \mathbf{A} x \geq 0$ for all $x \in \mathbb{R}^n$. If $\mathbf{A}$ is symmetric, this is equivalent to the eigenvalues of $\mathbf{A}$ being non-negative. Certifying whether a matrix is PSD often provides crucial insights into the structure of metric spaces [62], arises as a separation oracles in Semi-Definite Programming (SDP) [72], leads to faster algorithms for solving linear systems and linear algebra problems [8], [42], [56], [65] detects existence of community structure in random graphs [61], and is used to ascertain local convexity of functions. Furthermore, testing if a matrix is PSD is also used when studying the rate of dissipation in the heat equation [1] and the behavior of non-oscillatory, exponentially stable modes of linear differential equations [28]. For these applications, in addition to testing the existence of negative eigenvalues, it is often important to provide a *certificate* that the matrix is not PSD, by exhibiting a direction in which the quadratic form is negative.

While efficient, numerically stable algorithms for computing the spectrum of a matrix have been known since Turing [71], such algorithms require reading the entire matrix and incur a cubic running time in practice. Computing the eigenvalues of a matrix is often the bottleneck in applications, especially when just determining the existence of negative eigenvalues suffices. For instance, checking embeddability of a finite metric into Euclidean space, feasibility of a SDP, convexity of a function, and if specialized solvers are applicable for linear algebraic problems, all only require knowledge of whether a given matrix is PSD. The focus of this work is to study when the property of being PSD can be tested *sublinear* time and queries, without reading the entire matrix.

We approach the problem from the perspective of property testing [31], [32], where the input matrix $\mathbf{A}$ is promised to be either a PSD matrix, or "ϵ-far" from PSD under an appropriate notion of distance (discussed below). Specifically, we work in the *bounded-entry model*, proposed by Balcan, Li, Woodruff, and Zhang [11], where the input matrix has bounded entries: $\|\mathbf{A}\|_\infty \leq 1$. Boundedness is often a natural assumption in practice, and has numerous real world applications, such as recommender systems as in the Netflix Challenge [43], unweighted or bounded weight graphs [30], [32], correlation matrices, distance matrices with bounded radius, and others [11], [41], [49]. Further, the boundedness of entries avoids degenerate instances where an arbitrarily large entry is hidden in $\mathbf{A}$, thereby drastically changing the spectrum of $\mathbf{A}$, while being impossible to test without reading the entire matrix.

Our starting point is a simple fact: a matrix $\mathbf{A}$ is PSD

Ainesh Bakshi and Rajesh Jayaram would like to thank the partial support from the Office of Naval Research (ONR) grant N00014-18-1-2562, and the National Science Foundation (NSF) under Grant No. CCF-1815840.

[1]Throughout the paper, $\tilde{O}(\cdot)$ hides $\log(1/\epsilon)$ factors.

if and only if all *principal*[2] submatrices of $\mathbf{A}$ are PSD. However, a much more interesting direction is: if $\mathbf{A}$ is not PSD, what can be said about the eigenvalues of the submatrices of $\mathbf{A}$? Specifically, if $\mathbf{A}$ is far from PSD, how large of a submatrix must one sample in order to find a negative eigenvalue? Note that given a principal submatrix $\mathbf{A}_{T\times T}$ with $x^\top \mathbf{A}_{T\times T} x < 0$ for some $x \in \mathbb{R}^{|T|}$, this direction x can be used as a certificate that the input matrix is not PSD, since $y^\top \mathbf{A} y = x^\top \mathbf{A}_{T\times T} x < 0$, where y is the result of padding x with 0's. Further, it leads us to a natural algorithm to test definiteness: sample multiple principal submatrices and compute their eigenvalues. If any are negative, then $\mathbf{A}$ must not be PSD. Determining the query complexity of this task is the principal focus of this paper. Specifically, we ask:

Can the positive semi-definiteness of a bounded matrix be tested via the semi-definiteness of a small random submatrix?

The Testing Models.: The distance from $\mathbf{A}$ to the PSD cone is given by $\min_{\mathbf{B}\succeq 0} \|\mathbf{A}-\mathbf{B}\|$, where $\|\cdot\|$ is a norm, and $\mathbf{B} \succeq 0$ denotes that $\mathbf{B}$ is PSD. To instantiate $\|\cdot\|$, we consider two natural norms over $n \times n$ matrices: the spectral norm ($\|\cdot\|_2$) and the Euclidean norm ($\|\cdot\|_F$) . Perhaps surprisingly, the distance of a symmetric matrix $\mathbf{A}$ to the PSD cone under these norms can be characterized in terms of the eigenvalues of $\mathbf{A}$. In particular, let $\lambda \in \mathbb{R}^n$ be the vector of eigenvalues of $\mathbf{A}$. Then, the spectral norm distance corresponds to the ℓ_∞ distance between λ and the positive orthant. Similarly, the squared Frobenius distance corresponds to the ℓ_2^2 distance between λ and the positive orthant.

Therefore, we will refer to the two resulting gap problems as the ℓ_∞-gap and the ℓ_2^2-gap, respectively. This connection between matrix norms of $\mathbf{A}$ and vector norms of eigenvalues λ will be highly useful for the analysis of random submatrices. Next, we formally define the testing problems:

Problem 1 (PSD Testing with Spectral norm/ℓ_∞-gap)**.** *Given $\epsilon \in (0,1]$ and a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ such that $\|\mathbf{A}\|_\infty \le 1$, distinguish whether $\mathbf{A}$ satisfies:*

(1) $\mathbf{A}$ is PSD.

(2) $\mathbf{A}$ is ϵ-far from the PSD cone in Spectral norm: $\min_{\mathbf{B}\succeq 0} \|\mathbf{A}-\mathbf{B}\|_2 = \max_{i:\lambda_i<0} |\lambda_i(\mathbf{A})| \ge \epsilon n$.

The fact that the spectral norm distance from $\mathbf{A}$ to the PSD cone ($\min_{\mathbf{B}\succeq 0} \|\mathbf{A}-\mathbf{B}\|_2$) is equivalent to the magnitude of the smallest negative eigenvalue of $\mathbf{A}$ is a consequence of the variational principle for eigenvalues. For general non-symmetric matrices $\mathbf{A}$, one can replace (2) above with the condition $x^\top \mathbf{A} x < -\epsilon n$ for some unit vector $x \in \mathbb{R}^n$, which is equivalent to (2) if $\mathbf{A}$ is symmetric (again by the variational principle). We note that our results for the ℓ_∞-gap hold in this more general setting.[3]

Next, if we instantiate $\|\cdot\|$ with the (squared) Euclidean norm, we obtain the ℓ_2^2 gap problem.

Problem 2 (PSD Testing with ℓ_2^2-gap)**.** *Given $\epsilon \in (0,1]$ and a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ such that $\|\mathbf{A}\|_\infty \le 1$, distinguish whether $\mathbf{A}$ satisfies:*

(1) $\mathbf{A}$ is PSD.

(2) $\mathbf{A}$ is ϵ-far from the PSD cone in squared Euclidean norm: $\min_{\mathbf{B}\succeq 0} \|\mathbf{A}-\mathbf{B}\|_F^2 = \sum_{i:\lambda_i(\mathbf{A})<0} \lambda_i^2(\mathbf{A}) \ge \epsilon n^2$.

Note that the identity $\min_{\mathbf{B}\succeq 0} \|\mathbf{A}-\mathbf{B}\|_F^2 = \sum_{i:\lambda_i(\mathbf{A})<0} \lambda_i^2(\mathbf{A})$ also follows from the variational principle. Similarly to the ℓ_∞-gap, if $\mathbf{A}$ is not symmetric one can always run a tester on the symmetrization $(\mathbf{A}+\mathbf{A}^\top)/2$. Also observe that $\|\mathbf{A}\|_F^2 \le n^2$ and $\|\mathbf{A}\|_2 \le n$ for bounded entries matrices, hence the respective scales of n, n^2 in the two gap instances above. Notice by definition, if a symmetric matrix $\mathbf{A}$ is ϵ-far from PSD in ℓ_∞ then $\mathbf{A}$ is ϵ^2-far from PSD in ℓ_2^2. However, the converse is clearly not true, and as we will see the complexity of PSD testing with ϵ^2-ℓ_2^2 gap is strictly harder than testing with ϵ-ℓ_∞ gap.[4]

In fact, there are several important examples of matrices which are far from the PSD cone in ℓ_2^2, but which are not far in ℓ_∞. For instance, if $\mathbf{A}$ is a random matrix with bounded moments, such as a matrix with i.i.d. Rademacher ($\{1,-1\}$) or Gaussian entries, then as a consequence of Wigner's Semicircle Law $\mathbf{A}$ will be $\Omega(1)$-far in ℓ_2^2 distance. However, $\|\mathbf{A}\|_2 = O(\sqrt{n})$ with high probability, so $\mathbf{A}$ will only be $O(1/\sqrt{n})$-far in ℓ_∞ distance. Intuitively, such random instances should be very "far" from being PSD, and the ℓ_2^2 distance captures this fact.

A. Our Contributions

We now introduce our main contributions. Our algorithms for PSD testing randomly sample principal submatrices and check if they are PSD. Thus, all our algorithms have one-sided error; when $\mathbf{A}$ is PSD, they always return PSD, and whenever our algorithms return Not PSD, they output a certificate in the form of a principal submatrix which is not PSD. In what follows, $\omega < 2.373$ is the exponent of matrix multiplication, and $\tilde{O}, \tilde{\Omega}$ notation only hide $\log(1/\epsilon)$ factors (and $\log(s)$ factors for Ky-Fan-s and residual error bounds), thus our bounds have no direct dependency on the input size n. We first state our result for the ℓ_∞ gap problem in its most general form, which is equivalent to Problem 1 in the special case when $\mathbf{A}$ is symmetric.

[2]Recall that a principal submatrix $\mathbf{A}_{T\times T}$ for $T \subseteq [n]$ is the restriction of $\mathbf{A}$ to the rows and columns indexed by T.

[3]Also note that given query access to any $\mathbf{A} \in \mathbb{R}^{n\times n}$, one can always run a tester on the symmetrization $\mathbf{B} = (\mathbf{A}+\mathbf{A}^\top)/2$, which satisfies $x^\top \mathbf{A} x = x^\top \mathbf{B} x$ for all x, with at most a factor of 2 increase in query complexity.

[4]The difference in scaling of ϵ between the ℓ_∞ and ℓ_2^2 gap definitions (ϵ is squared in the latter) is chosen for the sake of convenience, as it will become clear the two problems are naturally studied in these respective paramaterizations.

Theorem 3 (ℓ_∞-gap Upper Bound)**.** *There is a non-adaptive sampling algorithm which, given $\mathbf{A} \in \mathbb{R}^{n\times n}$ with $\|\mathbf{A}\|_\infty \leq 1$ and $\epsilon \in (0,1)$, returns PSD if $x^\top \mathbf{A} x \geq 0$ for all $x \in \mathbb{R}^n$, and with probability $2/3$ returns Not PSD if $x^\top \mathbf{A} x \leq -\epsilon n$ for some unit vector $x \in \mathbb{R}^n$. The algorithm make $\tilde{O}(1/\epsilon^2)$ queries to the entries of $\mathbf{A}$, and runs in time $\tilde{O}(1/\epsilon^\omega)$.*

We demonstrate that the algorithm of Theorem 3 is optimal up to $\log(1/\epsilon)$ factors, even for adaptive algorithms with two-sided error. Formally, we show:

Theorem 4 (ℓ_∞-gap Lower Bound)**.** *Any adaptive or non-adaptive algorithm which solves the PSD testing problem with ϵ-ℓ_∞ gap with probability at least $2/3$, even with two-sided error and if $\mathbf{A}$ is promised to be symmetric, must query $\widetilde{\Omega}(1/\epsilon^2)$ entries of $\mathbf{A}$.*

Next, we present our algorithm for the ℓ_2^2-gap problem. Our algorithm crucially relies on first running our tester for the ℓ_∞-gap problem, which allows us to demonstrate that if $\mathbf{A}$ is far from PSD in ℓ_2^2 but close in ℓ_∞, then it must be far, under other notions of distance such as Schatten norms or residual tail error, from any bounded entry PSD matrix.

Theorem 5 (ℓ_2^2-gap Upper Bound)**.** *There is a non-adaptive sampling algorithm which, given a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with $\|\mathbf{A}\|_\infty \leq 1$ and $\epsilon \in (0,1)$, returns PSD if $\mathbf{A}$ is PSD, and with probability $2/3$ returns Not PSD if $\sum_{i:\lambda_i(\mathbf{A})<0} \lambda_i^2(\mathbf{A}) \geq \epsilon n^2$. The algorithm make a $\tilde{O}(1/\epsilon^4)$ queries to $\mathbf{A}$, and runs in time $\tilde{O}(1/\epsilon^{2\omega})$.*

We complement our upper bound by a $\widetilde{\Omega}(\frac{1}{\epsilon^2})$ lower bound for PSD-testing with ϵ-ℓ_2^2 gap, which holds even for algorithms with two sided error. Our lower bound demonstrates a separation between the complexity of PSD testing with $\sqrt{\epsilon}$-ℓ_∞ gap and PSD testing with ϵ-ℓ_2^2-gap, showing that the concentration of negative mass in large eigenvalues makes PSD testing a strictly easier problem.

Theorem 6 (ℓ_2^2-gap Lower Bound)**.** *Any non-adaptive algorithm which solves the PSD testing problem with ϵ-ℓ_2^2 gap with probability at least $2/3$, even with two-sided error, must query $\widetilde{\Omega}(1/\epsilon^2)$ entries of $\mathbf{A}$.*

Our lower bound is built on discrete hard instances which are "locally indistinguishable", in the sense that the distribution of any small set of samples is completely identical between the PSD and ϵ-far cases. At the heart of the lower bound is a key combinatorial Lemma about arrangements of paths on cycle graphs (see discussion in Section II-C). Our construction is highly general, and we believe will likely be useful for proving other lower bounds for matrix and graph property testing problems. Exemplifying the applicability of our construction, we obtain as an immediate corollary a new lower bound for testing the Schatten-1 norm of $\mathbf{A}$. Recall, that the Schatten-1 norm is defined via $\|\mathbf{A}\|_{\mathcal{S}_1} = \sum_i \sigma_i(\mathbf{A})$, where $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_n(\mathbf{A})$ are the singular values of $\mathbf{A}$.

Theorem 7 (Schatten-1 Lower Bound)**.** *Fix any $1/\sqrt{n} \leq \epsilon \leq 1$. Then any non-adaptive algorithm in the bounded entry model that distinguishes between*

1) $\|\mathbf{A}\|_{\mathcal{S}_1} \geq \epsilon n^{1.5}$,
2) $\|\mathbf{A}\|_{\mathcal{S}_1} \leq (1-\epsilon_0)\epsilon n^{1.5}$

with probability $2/3$, where $\epsilon_0 = \log^{O(1)}(1/\epsilon)$, must make at least $\tilde{\Omega}(1/\epsilon^4)$ queries to $\mathbf{A}$.

Note that one always has $\|\mathbf{A}\|_{\mathcal{S}_1} \leq n^{1.5}$ in the bounded entry model ($\|\mathbf{A}\|_\infty \leq 1$), which accounts for the above scaling. Theorem 7 extends a lower bound of Balcan et. al. [11], which is $\Omega(n)$ for the special case of $\epsilon, \epsilon_0 = \Theta(1)$. Thus, for the range $\epsilon = \tilde{O}(n^{-1/4})$, our lower bound is an improvement. To the best of our knowledge, Theorem 7 gives the first $\tilde{\Omega}(n^2)$ sampling lower bound for testing Schatten-1 in a non-degenerate range (i.e., for $\|\mathbf{A}\|_{\mathcal{S}_1} \geq n$).

Remark 8. *We note that the lower bound of [11] is stated for a slightly different version of gap (a "ϵ-ℓ_0"-gap), where either $\|\mathbf{A}\|_{\mathcal{S}_1} \geq c_1 n^{1.5}$ for a constant c_1, or at least ϵn^2 of the entries of $\mathbf{A}$ must be changed (while respecting $\|\mathbf{A}\|_\infty \leq 1$) so that the Schatten-1 is larger than $c_1 n^{1.5}$. However, their lower bound construction itself satisfies the "Schatten-gap" version as stated in Theorem 7, where either $\|\mathbf{A}\|_{\mathcal{S}_1} \geq c_1 n^{1.5}$, or $\|\mathbf{A}\|_{\mathcal{S}_1} \leq c_2 n^{1.5}$ and $c_1 > c_2$ are constants. From here, it is easy to see that this gap actually implies the ℓ_0-gap (and this is used to obtain the ℓ_0-gap lower bound in [11]), since if $\|\mathbf{A}\|_{\mathcal{S}_1} \leq c_2 n^{1.5}$ then for any $\mathbf{E}$ with $\|\mathbf{E}\|_\infty \leq 2$ and $\|\mathbf{E}\|_0 \leq \epsilon n^2$ for a small enough constant $\epsilon < c_2^2$, we have $\|\mathbf{A}+\mathbf{E}\|_{\mathcal{S}_1} \leq \|\mathbf{A}\|_{\mathcal{S}_1} + \|\mathbf{E}\|_{\mathcal{S}_1} \leq n^{1.5}(c_2 + 2\sqrt{\epsilon}) < c_1 n^{1.5}$. So Theorem 7 implies a lower bound of $\tilde{\Omega}(1/\epsilon^2)$ for distinguishing $\|\mathbf{A}\|_{\mathcal{S}_1} \geq \sqrt{\epsilon} n^{1.5}$ from the case of needing to change at least $\Omega(\epsilon n^2)$ entries of $\mathbf{A}$ so that $\|\mathbf{A}\|_{\mathcal{S}_1} \geq \sqrt{\epsilon} n^{1.5}$. Thus, our lower bound also extends the ℓ_0-gap version of the results of [11] for the range $\epsilon = \tilde{O}(1/\sqrt{n})$.*

In addition to Schatten-1 testing, the same lower bound construction and techniques from Theorem 6 also result in new lower bounds for testing the Ky-Fan s norm $\|\mathbf{A}\|_{\mathsf{KF}(s)} = \sum_{i=1}^s \sigma_i(\mathbf{A})$, as well as the cost of the best rank-s approximation $\|\mathbf{A} - \mathbf{A}_s\|_F^2 = \sum_{i>s} \sigma_i^2(\mathbf{A})$, stated below. In the following, s is any value $1 \leq s \leq n/\mathrm{poly}(\log n)$, and c is a fixed constant.

Theorem 9 (Ky-Fan Lower Bound)**.** *Any non-adaptive algorithm in the bounded entry model which distinguishes between*

1) $\|\mathbf{A}\|_{KF(s)} > (c/\log s)n$,
2) $\|\mathbf{A}\|_{KF(s)} < (1-\epsilon_0)(c/\log s)n$

with probability $2/3$, where $\epsilon_0 = \Theta(1/\log^2(s))$, must query at least $\tilde{\Omega}(s^2)$ entries of $\mathbf{A}$.

Theorem 10 (Residual Error Lower Bound)**.** *Any non-adaptive algorithm in the bounded entry model which distinguishes between*

1) $\|\mathbf{A}-\mathbf{A}_s\|_F^2 > cn/(s\log s)$,
2) $\|\mathbf{A}-\mathbf{A}_s\|_F^2 < (1-\epsilon_0)\cdot cn/(s\log s)$
with probability $2/3$*, where* $\epsilon_0 = \log^{O(1)}(s)$*, must query at least* $\tilde{\Omega}(s^2)$ *entries of* $\mathbf{A}$.

Our lower bound for the Ky-Fan norm complements a Ky-Fan testing lower bound of [51], which is $\Omega(n^2/s^2)$ for distinguishing **1)** $\|\mathbf{A}\|_{\mathsf{KF}(s)} < 2.1s\sqrt{n}$ from **1)** $\|\mathbf{A}\|_{\mathsf{KF}(s)} > 2.4s\sqrt{n}$ when $s = O(\sqrt{n})$. Note their bound decreases with s, whereas ours increases, thus the two bounds are incomparable (although they match up to $\log(s)$ factors at $s = \Theta(\sqrt{n})$).[5] We also point out that there are (not quite matching) upper bounds for both the problems of Ky-Fan norm and s-residual error testing in the bounded entry model, just based on a standard application of the Matrix Bernstein Inequality.[6] We leave the exact query complexity of these and related testing problems for functions of singular values in the bounded entry model as subject for future work.

A Remark on the ℓ_2^2-Gap: We note that there appear to be several key barriers to improving the query complexity of PSD testing with ℓ_2^2-gap beyond $O(1/\epsilon^4)$, which we briefly discuss here. First, in general, to preserve functions of the squared singular values of $\mathbf{A}$ up to error ϵn^2, such as $\|\mathbf{A}\|_F^2 = \sum_i \sigma_i^2(\mathbf{A})$ or $\|\mathbf{A}\|_2^2 = \sigma_1^2(\mathbf{A})$, any algorithm which samples a submatrix must make $\Omega(1/\epsilon^4)$ queries. Specifically, we show

Lemma 11. *Fix any* $1 \le k \le n$*, and fix any* $\frac{100}{\sqrt{n}} \le \epsilon \le 1/4$*. Any algorithm that queries a submatrix* $\mathbf{A}_{S\times T}$ *of* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *with* $\|\mathbf{A}\|_\infty \le 1$ *and distinguishes with probability at least* $4/5$ *between the case that either:*

- $\sum_{i=1}^k \sigma_i^2(\mathbf{A}) > n^2/2 + \epsilon n^2$.
- $\sum_{i=1}^k \sigma_i^2(\mathbf{A}) \le n^2/2$

must make $|S|\cdot|T| = \Omega(1/\epsilon^4)$ *queries to the matrix* $\mathbf{A}$.

In other words, Lemma 11 indicates detecting ϵn^2-sized perturbations in the spectrum of a matrix in general requires $\Omega(1/\epsilon^4)$ sized submatrices. This rules out improving the query complexity by detecting the ϵn^2 negative mass in $\mathbf{A}$ via, for instance, testing if the sum of squares of top $k = 1/\epsilon$ singular values has $\Theta(\epsilon n^2)$ less mass than it should if $\mathbf{A}$ were PSD (even this may require $\Omega(k^2/\epsilon^4)$ queries, see the discussion in Section II-B). The key issue at play here appears to be the requirement of sampling submatrices. Indeed, notice for the simplest case of $\|\mathbf{A}\|_F^2$, we can easily estimate $\|\mathbf{A}\|_F^2$ to additive ϵn^2 via $O(1/\epsilon^2)$ queries to random entries of $\mathbf{A}$. On the other hand, if these queries must form a submatrix, then it is easy to see that $\Omega(1/\epsilon^4)$ queries are necessary, simply from the problem of estimating $\|\mathbf{A}\|_F^2$ whose rows (or columns) have values determined by a coin flip with bias either equal to $1/2$ or $1/2+\epsilon$. On the other hand, for testing positive semi-definiteness, especially with one-sided error, the requirement of sampling a principal submatrix seems unavoidable.

In addition, a typical approach when studying spectral properties of submatrices is to first pass to a random row submatrix $\mathbf{A}_{S\times[n]}$, argue that it preserves the desired property (up to scaling), and then iterate the process on a column submatrix $\mathbf{A}_{S\times T}$. Unfortunately, these types of arguments are not appropriate when dealing with eigenvalues of $\mathbf{A}$, since after passing to the rectangular matrix $\mathbf{A}_{S\times[n]}$, any notion of negativity of the eigenvalues has now been lost. This forces one to argue indirectly about functions of the singular values of $\mathbf{A}_{S\times[n]}$, returning to the original difficulty described above. We leave it as an open problem to determine the exact non-adaptive query complexity of PSD testing with ℓ_2^2-gap. For a further discussion of these barriers and open problems, see Section III.

B. Connections to Optimization, Euclidean Metrics and Linear Algebra

We now describe some explicit instances where our algorithms may be useful for testing positive semi-definiteness. We emphasize that in general, the distance between $\mathbf{A}$ and the PSD cone may be too small to verify via our testers. However, when the input matrices satisfy a non-trivial gap from the PSD cone, we can speed up some basic algorithmic primitives. The first is testing feasibility of the PSD constraint in a Semi-Definite Program (SDP) with sublinear queries and time, so long as the variable matrix has bounded entries. Importantly, our algorithms also output a separating hyperplane to the PSD cone.

Corollary 12 (Feasibility and Separating Hyperplanes for SDPs)**.** *Given a SDP* $\mathcal{S}$*, let* $\mathbf{X} \in \mathbb{R}^{n\times n}$ *be a symmetric matrix that violates the PSD constraint for* $\mathcal{S}$*. Further, suppose* $\|\mathbf{X}\|_\infty \le 1$ *and* $\mathbf{X}$ *is* ϵn^2*-far in entry-wise* ℓ_2^2 *distance to the PSD cone. Then, there exists an algorithm that queries* $\widetilde{O}(1/\epsilon^4)$ *entries in* $\mathbf{X}$ *and runs in* $\widetilde{O}(1/\epsilon^{2\omega})$ *time, and with probability* $9/10$*, outputs a vector* $\tilde{v}$ *such that* $\tilde{v}^T\mathbf{X}\tilde{v} < 0$*. Moreover, if* $\lambda_{\min}(\mathbf{X}) < -\epsilon n$*, then there is an algorithm yielding the same guarantee, that queries* $\widetilde{O}(1/\epsilon^2)$ *entries in* $\mathbf{X}$ *and runs in* $\widetilde{O}(1/\epsilon^{\omega})$ *time.*

While in the worst case, our algorithm may need to read the whole matrix to exactly test if $\mathbf{X}$ is PSD, there may be applications where relaxing the PSD constraint to the convex set of matrices which are close to the PSD cone in Euclidean distance is acceptable. Moreover, our algorithm may be run as a preliminary step at each iteration of an SDP solver to check if the PSD constraint is badly violated, resulting in speed-ups by avoiding an expensive

[5]The bound from [51] is stated in the sketching model, however the entries of the instance are bounded, thus it also applies to the sampling model considered here.

[6]See Theorem 6.1.1 of [69], applied to $S_k = a_{(k)}(a_{(k)})^\top$, where $a_{(k)}$ is the k-th row sampled in $\mathbf{A}$; for the case of residual error, one equivalently applies matrix Bernstein inequality to estimate the head $\sum_{i\le k}\sigma_i^2(\mathbf{A})$. These bounds can be tightened via the usage of interior Chernoff bounds [27].

eigendecomposition of $\mathbf{X}$ whenever our algorithm outputs a separating hyperplane [72].

Next, we consider the problem of testing whether an arbitrary finite metric d over n points, $x_1, \ldots x_n \in \mathbb{R}^d$ is embeddable into Euclidean Space. Testing if a metric is Euclidean has a myriad of applications, such as determining whether dimensionality reduction techniques such as Johnson-Lindenstrauss can be used [57], checking if efficient Euclidean TSP solvers can be applied [5], and more recently, computing a low-rank approximation in sublinear time [9], [35]. It is well known (Schoenberg's criterion [20]) that given a distance matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ such that $\mathbf{D}_{i,j} = d(x_i, x_j)$, the points are isometrically embeddable into Euclidean space if and only If $\mathbf{G} = \mathbf{1} \cdot \mathbf{D}_{1,*} + \mathbf{D}_{1,*}^\top \cdot \mathbf{1}^\top - \mathbf{D} \succeq 0$, where $\mathbf{D}_{1,*}$ is the first row of $\mathbf{D}$. Notice that embeddability is scale invariant, allowing one to scale distances to ensure boundedness. Furthermore, since our algorithms sample submatrices and check for non-positive semi-definiteness, the tester need not know this scaling in advance, and gives guarantees for distinguishing definiteness if the necessary gap is satisfied after hypothetically scaling the entries.

Corollary 13 (Euclidean Embeddability of Finite Metrics)**.** *Given a finite metric d on n points $\{x_1, \ldots, x_n\}$, let $\mathbf{D} \in \mathbb{R}^{n \times n}$ be the corresponding distance matrix, scaled so that $\|\mathbf{D}\|_\infty \leq 1/3$, and let $\mathbf{G} = \mathbf{1}\mathbf{D}_{1,*} + \mathbf{D}_{1,*}^\top \mathbf{1}^\top - \mathbf{D}$. Then if $\min_{\mathbf{B} \succeq 0} \|\mathbf{G} - \mathbf{B}\|_F^2 \geq \epsilon n^2$, there exists an algorithm that queries $\widetilde{O}(1/\epsilon^4)$ entries in $\mathbf{A}$ and with probability $9/10$, determines the non-embeddability of $\{x_1, \ldots, x_n\}$ into Euclidean space. Further, the algorithm runs in time $\widetilde{O}(1/\epsilon^{2\omega})$.*

Remark 14. *An intriguing question is to characterized geometric properties of finite metrics based on the ℓ_2^2-distance of the Schoenberg matrix $\mathbf{G}$ from the PSD cone. For instance, given a finite metric with Schoenberg matrix $\mathbf{G}$ that is close to being PSD in ℓ_2^2-distance, can we conclude that the metric has a low worst or average case distortion embedding into Euclidean space?*

Remark 15. *Since rescaling entries to be bounded only affects the gap parameter ϵ, in both of the above cases, so long as the magnitude of the entries in $\mathbf{X}, \mathbf{D}$ do not scale with n, the running time of our algorithms is still sublinear in the input.*

Finally, several recent works have focused on obtaining sublinear time algorithms for low-rank approximation when the input matrix is PSD [8], [56]. However, such algorithms only succeed when the input is PSD or close to PSD (in ℓ_2^2), and it is unknown how to verify whether these algorithm succeeded in sublinear time. Therefore, our tester can be used as a pre-processing step to determine input instances where the aforementioned algorithms provably will (or will not) succeed.

C. Related work

Property testing in the bounded entry model was first considered in [11] to study the query complexity of testing spectral properties of matrices, such as stable rank (the value $\|\mathbf{A}\|_F^2 / \|\mathbf{A}\|_2^2$) and Schatten p norms. A related model, known as the *bounded row model*, where rows instead of entries are required to be bounded, was studied by Li, Wang, and Woodruff [49], who gave tight bounds for testing stable rank in this model. In addition, the problem of testing the rank of a matrix from a small number of queries has been well studied [13], [44], [49], [57], as well the problem of estimating the rank via a random submatrix [10], [12]. Notice that since rank is not a smooth spectral property, hiding an unbounded value in a single entry of $\mathbf{A}$ cannot drastically alter the rank. Thus, for testing rank, the condition of boundedness is not required.

More generally, the bounded entry model is the natural sampling analogue for the *linear sketching model*, where the algorithm gets to choose a matrix $\mathbf{S} \in \mathbb{R}^{t \times n^2}$, where t is the number of "queries", and then observes the product $\mathbf{S} \cdot \mathsf{vec}(\mathbf{A})$, where $\mathsf{vec}(\mathbf{A})$ is the vectorization of $\mathbf{A}$ [16], [17], [47]–[52]. The model has important applications to streaming and distributed algorithms. Understanding the query complexity of sketching problems, such as estimating spectral norms and the top singular values [4], [47], [51], estimating Schatten and Ky-Fan norms [17], [50]–[52], estimating ℓ_p norms [2], [15], [34], [38], [40], and ℓ_p sampling [36], [37], [39], [55], has been a topic of intense study. For the problem of sketching *eigenvalues* (with their signs), perhaps the most related result is [4], which gives pointwise estimates of the top eigenvalues. Notice that linear sketching can simulate sampling by setting the rows of $\mathbf{S}$ to be standard basis vectors, however sketching is in general a much stronger query model. Note that to apply a linear sketch, unlike in sampling, one must read all the entries of $\mathbf{A}$, which does not yield sublinear algorithms.

A special case of the sketching model is the *matrix-vector product* model, which has been studied extensively in the context of compressed sensing [18], [24] and sparse recovery [26]. Here, one chooses vectors $v_1, \ldots, v_k$ and observes the products $\mathbf{A}v_1, \ldots, \mathbf{A}v_k$. Like sketching, matrix-vector product queries are a much stronger access model than sampling. Recently, in the matrix-vector product model, Sun et. al. considered testing various graph and matrix properties [68], and Han et. al. considered approximating spectral sums and testing positive semi-definiteness [33].

Lastly, while there has been considerable work on understanding concentration of norms and singular values of random matrices, not as much is known about their *eigenvalues*. Progress in understanding the behavior of singular values of random matrices includes concentration bounds for spectral norms of submatrices [60], [70], concentration bounds for extreme singular values [25], [27], [46], [69], [73], non-

commutative Khintchine inequalities for Schatten-p norms [53], [58], [59], as well as Kadison-Singer type discrepancy bounds [45], [54], [64]. These random matrix concentration bounds have resulted in improved algorithms for many fundemental problems, such as low-rank approximation and regression [9], [19], [23], [35], [56] and spectral sparsification [3], [14], [66]. However, in general, understanding behavior of negative eigenvalues of random matrices and submatrices remains largely an open problem.

II. Technical Overview

In this section, we describe the techniques used in our non-adaptive testing algorithms for the ℓ_∞ and more general ℓ_2^2 gap problem, as well as the techniques involved in our lower bound construction for the ℓ_2^2-gap. We defer the proofs of these results to the full version of this article.

A. PSD Testing with ℓ_∞ Gap

Recall that in the general statement of the ℓ_∞-gap problem, our task is to distinguish between $\mathbf{A} \in \mathbb{R}^{n\times n}$ satisfying $x^\top \mathbf{A} x \geq 0$ for all $x \in \mathbb{R}^n$, or $x^\top \mathbf{A} x \leq -\epsilon n$ for some unit vector $x \in \mathbb{R}^n$. Since if $x^\top \mathbf{A} x \geq 0$ for all $x \in \mathbb{R}^n$ the same holds true for all principal submatrices of $\mathbf{A}$, it suffices to show that in the ϵ-far case we can find a $k \times k$ principal submatrix $\mathbf{A}_{T\times T}$ such that $y^\top \mathbf{A}_{T\times T} y < 0$ for some $y \in \mathbb{R}^k$.[7]

Warmup: A $O(1/\epsilon^3)$ query algorithm: Since we know $x^\top \mathbf{A} x \leq -\epsilon n$ for some fixed x, a natural approach would be to show that the quadratic form with the *same* vector x, projected onto to a random subset $T \subset [n]$ of its coordinates, is still negative. Specifically, we would like to show that the quadratic form $\mathcal{Q}_T(x) = x_T^\top \mathbf{A}_{T\times T} x_T$, of x with a random principal submatrix $\mathbf{A}_{T\times T}$ for $T \subset [n]$ will continue to be negative. If $\mathcal{Q}_T(x) < 0$, then clearly $\mathbf{A}_{T\times T}$ is not PSD. Now while our algorithm does not know the target vector x, we can still analyze the concentration of the scalar random variable $\mathcal{Q}_T(x)$ over the choice of T, and show that it is negative with good probability.

Proposition 16. *Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$ satisfies $\|\mathbf{A}\|_\infty \leq 1$ and $x^\top \mathbf{A} x \leq -\epsilon n$ where $\|x\|_2 \leq 1$. Then if $k \geq 6/\epsilon$, and if $T \subset [n]$ is a random sample of expected size k, we have $\mathbb{E}[\mathcal{Q}_T(x)] \leq -\frac{\epsilon k^2}{4n}$ and Var$(\mathcal{Q}_T(x)) \leq O(\frac{k^3}{n^2})$.*

By Proposition 16, after setting $k = \Theta(1/\epsilon^2)$, we have that $|\mathbb{E}[\mathcal{Q}_T(x)]|^2 = \Omega(\text{Var}(\mathcal{Q}_T(x))$, and so by Chebyshev's inequality, with constant probability we will have $\mathcal{Q}_T(x) < 0$. This results in a $k^2 = O(1/\epsilon^4)$ query tester. To improve the complexity, we could instead set $k = \Theta(1/\epsilon)$ and re-sample T for k times independently to reduce the variance. Namely, one can sample submatrices $T_1, T_2, \ldots, T_k$, and analyze $\frac{1}{k}\sum_{i=1}^k \mathcal{Q}_{T_i}(x)$. The variance of this sum goes down to $O(\frac{k^2}{n^2})$, so, again by Chebyshev's inequality, the average of these quadratic forms will be negative with constant probability. If this occurs, then at least one of the quadratic forms must be negative, from which we can conclude that at least one of $\mathbf{A}_{T_i\times T_i}$ will fail to be PSD, now using only $O(1/\epsilon^3)$ queries.

[7] This can be efficiently checked by computing the eigenvalues of $\mathbf{A}_{T\times T} + \mathbf{A}_{T\times T}^\top$.

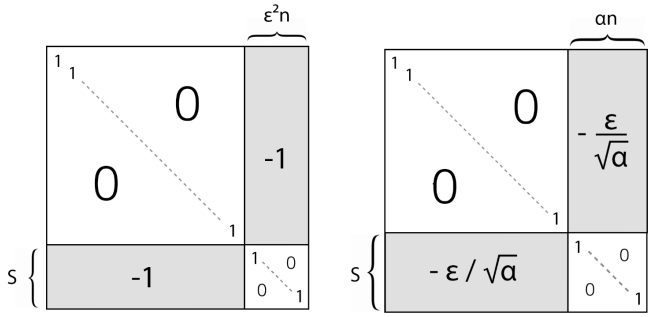

Figure 1: Hard instances for ℓ_∞ testing. On the left, the negative mass is highly concentrated in $|S| = \epsilon^2 n$ rows and columns, and on the right it more spread out over $|S| = \alpha n$, where $\epsilon^2 \leq \alpha \leq \epsilon$.

A Family of Hard Instances: One could now hope for an even tighter analysis of the concentration of $\mathcal{Q}_T(x)$, so that $O(1/\epsilon^2)$ total queries would be sufficient. Unfortunately, the situation is not so simple, and in fact the two aforementioned testers are tight in the query complexity for the matrix dimensions they sample. Consider the hard instance $\mathbf{A}$ in the left of Figure 1, which is equal to the identity on the diagonal, and is zero elsewhere except for a small subset $S \subset [n]$ of $|S| = \epsilon^2 n$ rows and columns, where we have $\mathbf{A}_{S\times\overline{S}} = \mathbf{A}_{\overline{S}\times S} = -\mathbf{1}$, where $\overline{S}$ is the complement of S. Notice that if we set $x_i^2 = 1/(2n)$ for $i \notin S$ and $x_i^2 = 1/(2\epsilon^2 n)$ for $i \in S$, then $x^\top \mathbf{A} x \leq -\epsilon n/4$. However, in order to even see a single entry from S, one must sample from at least $\Omega(1/\epsilon^2)$ rows or columns. In fact, this instance itself gives rise to a $\Omega(1/\epsilon^2)$ lower bound for any testing algorithm, even for adaptive algorithms (Theorem 4).

The difficulty of the above instance is that the negative mass of $x^\top \mathbf{A} x$ is hidden in only a ϵ^2-fraction of $\mathbf{A}$. On the other hand, since the negative entries are so large and concentrated, one need only sample $O(1)$ entries from a single row $i \in S$ in order for $\mathbf{A}_{T\times T}$ to be non-PSD in the prior example. Thus, an algorithm for such instances would be to sample $O(1/\epsilon^2)$ principal submatrices, each of *constant* size. On the other hand, the set S could also be more spread out; namely, we could have $|S| = \alpha n$ for any $\epsilon^2 \leq \alpha \leq \epsilon$, but where each entry in $\mathbf{A}_{S\times\overline{S}}$ is set to $-\epsilon/\sqrt{\alpha}$ (see the matrix in the right side of Figure 1). If instead, we define $x_i^2 = 1/(2\alpha n)$ for $i \in S$, we still have $x^\top \mathbf{A} x < -\epsilon n/4$. However, now any submatrix $\mathbf{A}_{T\times T}$ with $|T \cap S| = 1$ must have at least $|T| \geq \alpha/\epsilon^2$ rows and columns, otherwise $\mathbf{A}_{T\times T}$ would be PSD due to the identity on the

diagonal.

The aforementioned instances suggest the following approach: query matrices at $O(\log \frac{1}{\epsilon})$ different scales of subsampling. Specifically, for each $\epsilon^2 \leq \alpha = 2^i \leq \epsilon$, we sample $\tilde{O}(\frac{\epsilon^2}{\alpha^2})$ independent $k \times k$ submatrices, each of size $k = \tilde{O}(\alpha/\epsilon^2)$, giving a total complexity of $\tilde{O}(\frac{1}{\epsilon^2})$. The analysis now proceeds by a complete characterization of the ways in which $x^\top \mathbf{A} x$ can be negative. Specifically, we prove the following: either a substantial fraction of the negative mass is hidden inside of a small set of rows and columns S with $|S| < \epsilon n$, or it is the case that $\mathrm{Var}(\mathcal{Q}_T(x))$ is small enough so that a single $k \times k$ submatrix will already be non-PSD with good probability when $k \geq \Omega(1/\epsilon)$. Given this classification, it suffices to demonstrate a level of subsampling which will find a non-PSD submatrix when the negative mass is concentrated inside inside a small set S.

Eigenvector Switching: To analyze this case, ideally, one would like to demonstrate that conditioned on T intersecting S at some level of subsampling, we will have $\mathcal{Q}_T(x) < 0$ with good probability. Unfortunately, the approach of analyzing the quadratic form with respects to x will no longer be possible; in fact, $\mathcal{Q}_T(x)$ may never be negative conditioned on $|T \cap S| = 1$ (unless $|T| > 1/\epsilon$, which we cannot afford in this case). The complication arises from the fact that the coordinates of x_i in the small set S can be extremely large, and thus the diagonal contribution of $x_i^2 \mathbf{A}_{i,i}$ will dominate the quadratic form of a small submatrix. For instance, if $\mathbf{A}_{T \times T}$ is a sample with $k = |T| = O(1)$ which intersects the set S in the leftmost matrix in Figure 1, where $x_i = 1/(\epsilon\sqrt{n})$ for $i \in S$ and $x_i = 1/\sqrt{n}$ otherwise, then $\mathcal{Q}_T(x) \approx k/n - (k/\sqrt{n})x_i + \mathbf{A}_{i,i} x_i^2$, which is dominated by the diagonal term $\mathbf{A}_{i,i} x_i^2 = 1/(\epsilon^2 n)$. Thus, while $\mathbf{A}_{T \times T}$ itself is not PSD, we have that $\mathcal{Q}_T(x) > 0$.

To handle this, we must and analyze the quadratic form $\mathcal{Q}_T(\cdot)$ with respect to *another* direction y. The vector y may not even satisfy $y^\top \mathbf{A} y < 0$, however conditioned on $|T \cap S| \geq 1$, we will have $\mathcal{Q}_T(y) < 0$ with good probability. Clearly, we must scale down the large coordinates x_i for $i \in S$. However, one cannot scale too low, otherwise the negative contribution of the rows $i \in S$ would become too small. The correct scaling is then a careful balancing act between the contributions of the different portions of $\mathbf{A}_{T \times T}$. Informally, since the x_i's for $i \in S$ make up a $|S|/n$ fraction of all coordinates, they can be as large as $x_i^2 \geq (n/|S|) \cdot (1/n)$. However, inside of the smaller submatrix $\mathbf{A}_{T \times T}$, then conditioned on $i \in T$, since $|T|$ is small x_i now makes up a larger $1/|T|$ fraction of the submatrix, thus we should scale down x_i to only be $x_i^2 \approx |T|/n$. With this scaling in mind, we (very roughly) set $y_i^2 = (|S|/n) \cdot (|T|/n)$ if $i \in S$, and set $y_i = x_i$ otherwise. The remaining argument then requires a careful analysis of the contribution of entries of $\mathbf{A}$ outside of S to show that the target vector y indeed satisfies $\mathcal{Q}_T(y) < 0$ with good probability conditioned on T intersecting S.

B. PSD Testing with ℓ_2 Gap

Recall in the ℓ_2 gap problem, our task is to distinguish between $\mathbf{A}$ being PSD, and $\mathbf{A}$ being ϵ-far in ℓ_2^2 distance from any PSD matrix, namely that $\sum_{i:\lambda_i(\mathbf{A})<0} \lambda_i^2(\mathbf{A}) > \epsilon n^2$. In what follows, we refer to the quantity $\sum_{i:\lambda_i(\mathbf{A})<0} \lambda_i^2(\mathbf{A})$ as the *negative mass* of $\mathbf{A}$. First observe that in the special case that we had a "large" negative eigenvalue, say $\lambda_{\min}(\mathbf{A}) = -\epsilon n$, then by applying our testing algorithm for ℓ_∞-gap, we could find a non-PSD submatrix with only $\tilde{O}(1/\epsilon^2)$ queries. However, in general the negative mass of $\mathbf{A}$ may be spread out over many smaller eigenvalues. Thus, we cannot hope to apply our earlier approach for the ℓ_∞-gap, which preserved the quadratic form $\mathcal{Q}_T(x) = x_T^\top \mathbf{A}_{T \times T} x_T$ with respects to a fixed direction x. Instead, our approach will be to show that if $\mathbf{A}$ is ϵ-far from PSD in ℓ_2^2, then the singular values of $\mathbf{A}$ must be "far" from PSD, in some other notion of distance, allowing us to indirectly infer the existence of negative eigenvalues in submatrices.

PSD matrices are top-heavy: Our first step is to show that if $\mathbf{A} \in \mathbb{R}^{n \times n}$ is PSD, then the t-"tail" of $\mathbf{A}$, defined as $\sum_{i>t} \sigma_i^2(\mathbf{A})$, cannot be too large. This can be derived from the following fact: if $\mathbf{A}$ is PSD and $\|\mathbf{A}\|_\infty \leq 1$, then we can bound the Schatten-1 norm of $\mathbf{A}$ by $\|\mathbf{A}\|_{\mathcal{S}_1} = \sum_i \sigma_i(\mathbf{A}) = \mathrm{Tr}(\mathbf{A}) \leq n$. This simple fact will prove highly useful, since whenever we can demonstrate that the Schatten-1 norm of a submatrix $\mathbf{A}_{T \times T}$ is larger than $|T|$, we may immediately conclude the that $\mathbf{A}_{T \times T}$ is not PSD. In addition, it implies:

Proposition 17 (PSD matrices are top-heavy)**.** *Fix any $n \in \mathbb{N}$, $1 \leq t \leq n$, and $\mathbf{D} \in \mathbb{R}^{n \times n}$ such that $\|\mathbf{D}\|_\infty \leq 1$, and let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ be its singular values. Then if $\mathbf{D}$ is PSD, we have*

$$\sum_{i>t} \sigma_i^2(\mathbf{D}) \leq t^{-1} n^2$$

On the other hand, suppose that $\mathbf{A}$ is ϵ-far from PSD, and let $t > 10/\epsilon$. Then if no eigenvalue is smaller than $-\epsilon n/100$, a condition which can be checked with $\tilde{O}(1/\epsilon^2)$ queries by first running our ℓ_∞-gap tester, then the negative mass must be spread out, and it must be the case that a substantial fraction of the negative mass of $\mathbf{A}$ is contained in the bottom $n - t$ singular values. Specifically, we must have $\sum_{i>t} \sigma_i(\mathbf{A})^2 > (\epsilon/2) n^2$, whereas any PSD matrix $\mathbf{D}$ would have to satisfy $\sum_{i>t} \sigma_i^2(\mathbf{D}) \leq (\epsilon/10) n^2$ by the above Proposition. Thus, after first running our ℓ_∞ tester, it will suffices to estimate the tail $\sum_{i>t} \sigma_i^2(\mathbf{A})$. Equivelantly, since $\|\mathbf{A}\|_F^2 = \sum_i \sigma_i^2(\mathbf{A})$ can be efficiently estimated, it also suffices to estimate the "head" $\sum_{i \leq t} \sigma_i^2(\mathbf{A})$ to additive $O(\epsilon n^2)$.

In order to accomplish this, one could utilize the tools from random matrix concentration, such as Matrix Bernstein's inequality [69], which allows one to estimate each σ_i^2 to error ηn^2 by taking a random rectangular

$O(1/\eta^2) \times O(1/\eta^2)$ sized submatrix. The error in estimating $\sum_{i \leq t} \sigma_i^2(\mathbf{A})$ is then $t\eta n^2$, thus one needs to set $\eta = O(\epsilon/t)$, giving a $O(1/\epsilon^8)$ tester with two-sided error. Using a careful bucketing analysis on the error, along with the more powerful Interior Matrix Chernoff bounds of Gittens and Tropp [27], one can improve this to $O(t^2/\epsilon^4) = O(1/\epsilon^6)$. However, substantial improvements on unconditional estimation of $\sum_{i \leq t} \sigma_i^2(\mathbf{A})$ seem unlikely. In fact, we demonstrate that event for $t = 1$ (spectral norm estimation), tools such as matrix concentration inequalities which sample submatrices of $\mathbf{A}$, must make $\Omega(1/\epsilon^4)$ queries (Lemma 11), which rules out, for instance, a $o(t^2/\epsilon^4)$ upper bound for general t. Thus, instead of unconditional estimation, our main insight is to demonstrate conditions under which $\sum_{i \leq t} \sigma_i^2(\mathbf{A})$ can be efficiently estimated. When these conditions do not hold, we show that it is because the Schatten-1 norm of our sampled submatrix must be too large, from which we can deduce the existence of negative eigenvalues in our query.

In the first case, if the t-th singular value is not too large, say $\sigma_{t+1}(\mathbf{A}) \leq 10n/t$, we show that the (re-scaled) tail $\frac{n^2}{k^2} \sum_{i>t} \sigma_i^2(\mathbf{A}_{S \times T})$ of a random rectangular matrix, where $|S| = |T| = k = O(1/\epsilon^2)$, approximates the tail of $\mathbf{A}$ to error $O(\epsilon n^2)$. Our argument relies on splitting $\mathbf{A}$ into head and tail pieces $\mathbf{A} = \mathbf{A}_t + \mathbf{A}_{-t}$, where $\mathbf{A}_t$ is $\mathbf{A}$ projected onto the top-t eigenvectors of $\mathbf{A}$. We demonstrate that the spectral mass of each is preserved after passing to a random row submatrix, and additionally demonstrate that $\sigma_{\max}(\mathbf{A}_{-t}) = \sigma_{t+1}(\mathbf{A})$ does not grow too much using spectral decay inequalities for random submatrices [60]. This forces the spectrum of $(\mathbf{A}_{-t})_{S \times [n]}$ to be well spread out, allowing us to apply interlacing inequalities to demonstrate that after adding $(\mathbf{A}_t)_{S \times [n]}$ back in, the resulting tail is still sufficiently large, and then iterate the argument when sampling columns to obtain $\mathbf{A}_{S \times T}$.

On the other hand, if $\sigma_{t+1}(\mathbf{A})$ is too large, then after moving to a random row submatrix the spectral norm of $\mathbf{A}_{-t}$ can concentrate highly in its top eigenvalues, which can then be absorbed by the top t eigenvalues of $\mathbf{A}_t$, stealing too much mass from the tail. Instead, note that if $\sigma_{t+1}(\mathbf{A}) \geq 10n/t$, then the Schatten norm of $\mathbf{A}$ must be large, namely $\sum_i \sigma_i(\mathbf{A}) > 10n$, which cannot occur if $\mathbf{A}$ is PSD. We show that by applying Interior Eigenvalue Matrix Chernoff bounds (mentioned above), we can preserve this fact, obtaining $\frac{n}{k}\sigma_{t+1}(\mathbf{A}_{S \times T}) > 10n/t$ with good probability when $k = \Omega(1/\epsilon^2)$. If this is the case, then the Schatten norm of the submatrix will be too large: $\|\mathbf{A}_{S \times T}\|_{\mathcal{S}_1} \geq t(10k/t) > 10k$. To obtain a certificate from this fact, we move to the larger principal submatrix $\mathbf{A}_{(S \cup T) \times (S \cup T)}$, which we show must still have large Schatten norm, from which we can infer the existence of negative eigenvalues. Similarly, in the earlier case, we show that the large tail of $\mathbf{A}_{S \times T}$ implies that the principal submatrix $\mathbf{A}_{(S \cup T) \times (S \cup T)}$ also has too large of a tail, meaning it must not be PSD.

C. Lower Bounds

As seen above, the distribution of negative mass in the matrix $\mathbf{A}$ plays an important role in the complexity of testing if $\mathbf{A}$ is PSD. Specifically, the problem becomes easier the more concentrated the negative mass is within a few eigenvalues. So in order to avoid a $o(1/\epsilon^2)$ upper bound from the ℓ_∞-testing algorithm, our hard instance must have $|\lambda_{\min}(\mathbf{A})| = O(\epsilon n)$ in the ϵ-far case. On the other hand, we cannot allow the negative mass to be extremely spread out, otherwise we would have to add many more *positive* eigenvalues to avoid violating the trace constraint $|\mathrm{Tr}(\mathbf{A})| = |\sum_i \lambda_i(\mathbf{A})| \leq n$ implied by the boundedness, creating further spectral differences between the instances. With this in mind, our hard distribution will have $1/\epsilon$ negative eigenvalues, each roughly equal to $\lambda_i(\mathbf{A}) = -\epsilon n$.

The Hard Instance: Our first insight is to construct a discrete instance, with the property that the distribution induced by observing a small sample of the "meaningful" entries of $\mathbf{A}$ is *identical* in both cases. Specifically, we construct two distribtuions: $\mathcal{D}_{\text{YES}}$ and $\mathcal{D}_{\text{NO}}$ over $n \times n$ matrices. In both cases, $\mathbf{A}$ will be block diagonal, with k disjoint blocks $B_1, B_2, \ldots, B_k \subset [n]$, each of size $|B_i| = n/k$, for some parameter k; we will later set $k = \Theta(1/\epsilon)$, so our target lower bound is $\Omega(k^2)$. In $\mathcal{D}_{\text{YES}}$, each block $\mathbf{A}_{B_i \times B_i}$ will be PSD, whereas in $\mathcal{D}_{\text{NO}}$ we will have $\lambda_{\min}(\mathbf{A}_{B_i \times B_i}) = -\tilde{\Theta}(n/k) \approx -\epsilon n$. The partition $B_1 \cup B_2 \cup \cdots \cup B_k = [n]$ is chosen randomly, so that for any fixed set of samples, only a small fraction them will be contained inside any block $\mathbf{A}_{B_i \times B_i}$. The diagonal entries will always be fixed to 1, and all off-diagonal entries are either $\{0, 1, -1\}$. The samples $a_1, a_2, \ldots, a_s \in [n] \times [n]$ of any algorithm can then be interpreted as a graph H (possibly with self-loops), where for each edge $a_r = (i, j) \in E(H)$, the algorithm learns the value $\mathbf{A}_{i,j} \in \{0, 1, -1\}$.

Now consider the algorithm which just samples a $t \times t$ principal submatrix $T \subset [n]$, so that H is a t-clique. Now in expectation $\mathbb{E}[|T \cap B_i|] = \frac{t}{k}$ for each i, however, by a balls and bins argument, as t approaches k we will obtain some blocks i with $|T \cap B_i| = \Omega(\log k / \log\log k)$. Thus, to fool this query, we must be able to "fool" cliques of size roughly $\log k$ within a block B_i. On the other hand, an algorithm could find many more entries in a block by lopsided sampling: for instance, it could sample k^2 entries in a single column of $\mathbf{A}$ (H is a k^2-star), getting k entries inside a column of a block $\mathbf{A}_{B_i \times B_i}$. Thus we must also fool large star queries. It turns out that the right property to consider is the *matching number* $\nu(H)$ of the query graph H, i.e. the size of a maximum matching. Notice for a star H, we have $\nu(H) = 1$. We prove (roughly) that if within each block B_i, one can "fool" every query graph H inside $\mathbf{A}_{B_i \times B_i}$ with matching number $\nu(H) < \ell$, one would obtain a lower

bound of $\Omega(k^{\frac{2(\ell-1)}{\ell}})$. Thus, it will suffice to fool all query graphs H within a block B_i with $\nu(H) \leq \log k$.

For a first step towards this, suppose that in $\mathcal{D}_{\text{YES}}$, we set each block independently to $\mathbf{A}_{B_i \times B_i} = vv^\top$, where $v \in \{1, -1\}^{|B_i|}$ is a random sign vector, and in $\mathcal{D}_{\text{NO}}$, we set $\mathbf{A}_{B_i \times B_i} = -vv^\top$ (except we fix the diagonal to be 1 in both cases). Now notice that the distribution of any individual entry $(\mathbf{A}_{B_i \times B_i})_{a,b}$ is symmetric, and identical in both $\mathcal{D}_{\text{YES}}$ and $\mathcal{D}_{\text{NO}}$. Furthermore, it is not difficult to check that the distribution of a path or star query H inside of $\mathbf{A}_{B_i \times B_i}$ is also identical in both cases. On the other hand, if H contained a *triangle*, then this would not be the case, since in $\mathcal{D}_{\text{YES}}$ one could never have a negative cycle (x, y, z) where $v_x v_y = v_y v_z = v_z v_x = -1$, whereas this could occur in $\mathcal{D}_{\text{NO}}$, since we could have that $-v_x v_y = -v_y v_z = -v_z v_x = -1$. Thus, roughly, to distinguish between these distributions $\mathcal{D}_{\text{YES}}$ from $\mathcal{D}_{\text{NO}}$, an algorithm must sample a triangle within one of the blocks $\mathbf{A}_{B_i \times B_i}$, which one can show requires $\Omega(k^{4/3})$ queries, yielding a first lower bound.[8]

Boosting to $\Omega(k^2)$: Given the above example, we would now like to construct instances which fool H with larger and larger $\nu(H)$. In fact, our next insight is to have an even simpler structure on $\mathcal{D}_{\text{YES}}$ and $\mathcal{D}_{\text{NO}}$: each of them will be a random permutation of one of two *fixed* matrices $\mathbf{D}_1, \mathbf{D}_2$ respectively. We now formalize the "fooling" condition we need. For a matrix $\mathbf{B}$ and a query graph H, let $(\mathbf{B})_H$ denote the result of setting all entries of $\mathbf{B}$ not in H equal to zero. Then the matrices $\mathbf{D}_1, \mathbf{D}_2$ must have the property that for any graph H with $\nu(H) \leq \log k$, if $\sigma : [m] \to [m]$ is a random permutation and $\mathbf{P}_\sigma \in \mathbb{R}^{m \times m}$ is the row permutation matrix corresponding to σ, then the distribution of $(\mathbf{P}_\sigma \mathbf{D}_1 \mathbf{P}_\sigma^\top)_H$ is identical to the distribution $(\mathbf{P}_\sigma \mathbf{D}_2 \mathbf{P}_\sigma^\top)_H$. We call this property *H-subgraph equivalence*. This implies that any algorithm which queries the edges in H inside of $\mathbf{P}_\sigma \mathbf{D}_1 \mathbf{P}_\sigma^\top$ or $\mathbf{P}_\sigma \mathbf{D}_2 \mathbf{P}_\sigma^\top$ will be unable to distinguish between them with any advantage. To obtain a lower bound, we must also have a gap between $\lambda_{\min}(\mathbf{D}_1)$ and $\lambda_{\min}(\mathbf{D}_2)$, so that their spectrum can be shifted to make one PSD and the other far. Furthermore, neither $\lambda_{\min}(\mathbf{D}_1)$ or $\lambda_{\min}(\mathbf{D}_2)$ can be too negative, otherwise by shifting we would lose boundedness of the entries.

A priori, it is not even clear that such matrices $\mathbf{D}_1, \mathbf{D}_2$ exist, even for constant values of $\nu(H)$. Our main contribution now is to demonstrate their existence for every $\nu(H)$. Our construction is simple, but perhaps surprisingly so. Both $\mathbf{D}_1, \mathbf{D}_2$ will be adjacency matrices; in the PSD case, we set $\mathbf{D}_1$ to be the cycle graph C_{2m+1} on $2m + 1 = \Theta(\log k)$ vertices, and in the ϵ-far case we set $\mathbf{D}_2$ to be the disjoint union of two cycles $C_{m+1} \oplus C_m$. Since one of m and $m + 1$ is even, while $2m + 1$ is odd, we will have that $\lambda_{\min}(C_{m+1} \oplus C_m) = -2$, but $\lambda_{\min}(C_{2m+1}) > -2$.[9] To show subgraph equivalence, it suffices to show a slightly more general version of the following: for any graph H with $\nu(H) < m/4$, the number of subgraphs of C_{2m+1} isomorphic to H is the same as the number of subgraphs of $C_{m+1} \oplus C_m$ isomorphic to H.[10] Note that if $\nu(H) < m/4$, then H is just a disjoint collection of paths.

Our proof of this fact is by a construction of a bijection from arrangements of H in C_{2m+1} to H in $C_{m+1} \oplus C_m$. While a seemingly simple property, some care must be taken when designing a bijection. Our mapping involves first "swapping" two paths (whose length depends on H) in C_{2m+1}, before "splitting" C_{2m+1} into two cycles of length m and $m+1$. We direct the reader to the full version of this work for further details.

Amplifying the Gap: The subgraph equivalence between C_{2m+1} and $C_{m+1} \oplus C_m$ prevents any algorithm from distinguishing between them with a small number of samples, however the gap in the minimum eigenvalue shrinks at the rate of $\Theta(1/m^2)$. Meaning, if we set $\gamma = \lambda_{\min}(C_{2m+1}) = 2 - \Theta(1/m^2)$, while the matrix $\gamma \mathbf{I} + C_{2m+1}$ is PSD and has constant sized entries, we only have $\lambda_{\min}(\gamma \mathbf{I} + C_{m+1} \oplus C_m) = -\Theta(1/m^2)$, which is not far enough from PSD. Instead, recall that we only need $m = \Omega(\log k)$ to fool all H with $\nu(H) \leq \log k$, but the block size which we must fill is much larger: $\mathbf{A}_{B_i \times B_i}$ has size $|B_i| = n/k$. Thus, instead of setting $m = \Theta(n/k)$ and filling all of $\mathbf{A}_{B_i \times B_i}$ with the cycles, we set $m = \Theta(\log k)$, and we amplify the spectral gap by taking the tensor product of the small graphs C_{2m+1} and $C_{m+1} \oplus C_m$ with a large, fixed matrix $\mathbf{M}$, so that $(\gamma \mathbf{I} + C_{2m+1}) \otimes \mathbf{M}$ has $|B_i|$ rows and columns.

We prove that taking the tensor product with any fixed $\mathbf{M}$ preserves the subgraph equivalence properties of the original matrices. From here, our lower bounds for testing PSD with ℓ_2 gap, Schatten norms, Ky fan, and the cost of the best rank-k approximation, all follow by a proper choice of $\mathbf{M}$. For PSD testing, we can choose $\mathbf{M} = \mathbf{1}$ to be the all 1's matrix, and to amplify the gap in Schatten 1 norm, we can choose $\mathbf{M}$ to be a random Rademacher matrix. Since $\mathbf{M} = \mathbf{1}$ is PSD and $\|\mathbf{M}\|_2 = \widetilde{\Omega}(n/k)$, the gap is amplified to the desired $-\widetilde{\Omega}(n/k)$. Finally, we remark that to obtain a lower bound for another norm, any matrix $\mathbf{M}$ which is large in that norm may be suitable, so long as the original sub-graph equivalent matrices also have a gap in that norm. We pose it as an interesting open problem to design other pairs of matrices $\mathbf{D}_1, \mathbf{D}_2$ with different spectral gaps which have good sub-graph equivalence properties.

[8] Note that $\nu(H) = 1$ for a triangle H, so the $\Omega(k^{2(\ell-1)/\ell})$ lower bound when $\nu(H) < \ell$ is actually loose here.

[9] To intuitively see why this is true, note that if m is even and $v \in \{-1, 1\}^m$ is the vector that assigns opposite signs to adjacent vertices of C_m, then we have $C_m v = -2v$. However, if m is odd, this assignment v is no longer possible.

[10] A more general statement is needed since H can also query for edges which do not exist in C_{2m+1}.

III. Conclusion

In this work, we gave an optimal (up to $\log(1/\epsilon)$ factors) algorithm for testing if a matrix was PSD, or was far in spectral norm distance from the PSD cone. In addition, we gave a query efficient algorithm for testing if a matrix was PSD, or was ϵn^2 far from the PSD-cone in ℓ_2^2 distance. Furthermore, we established a new technique for proving lower bounds based on designing "subgraph-equivelant" matrices. We believe that this technique is quite general, as shown by its immediate application to lower bounds for the Schatten-1 norm, Ky-Fan norm, and tail error testing. Our construction could also likely be useful for proving lower bounds against testing of *graph properties*, which is a well studied area [30]. We pose the open problem to design (or demonstrate the non-existence of) additional subgraph-equivalent matrices beyond the cycle graph construction utilized in this work, which have gaps in their spectral or graph-theoretic properties.

Additionally, we pose the open problem of determining the exact non-adaptive query complexity of PSD testing with ℓ_2^2 gap. As discussed in Section I-A, there appear to be several key barriers to improving the complexity beyond $O(1/\epsilon^4)$. Indeed, it seems that perhaps the main tool that is lacking is a concentration inequality for the eigenvalues of random principal submatrices. Since most such decay results apply only to norms [60], [70], progress in this direction would likely result in important insights into eigenvalues of random matrices.

Finally, we note that the complexity of the testing problems for several matrix norms, specifically the Schatten p and Ky-Fan norms, are still open in the bounded entry model. In particular, for the Schatten 1 norm, to the best of our knowledge no non-trivial algorithms exist even for estimation with additive error $\Theta(n^{1.5})$, thus any improvements would be quite interesting.

Acknowledgment

We thank Erik Waingarten for many useful suggestions, being closely involved in the early stages of this project, and for feedback on early drafts of this manuscript. We also thank Roie Levin, Ryan O'Donnell, Pedro Paredes, Nicolas Resch, and Goran Zuzic for illuminating discussions related to this project.

References

[1] M Necati Ã-zisik, M Necati Özısık, and M Necati Özışık. *Heat conduction*. John Wiley & Sons, 1993.

[2] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 20–29. ACM, 1996.

[3] Alexandr Andoni, Jiecao Chen, Robert Krauthgamer, Bo Qin, David P Woodruff, and Qin Zhang. On sketching quadratic forms. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 311–319. ACM, 2016.

[4] Alexandr Andoni and Huy L. Nguyen. Eigenvalues of a matrix in the streaming model. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1729–1737. Society for Industrial and Applied Mathematics, 2013.

[5] Sanjeev Arora. Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems. *Journal of the ACM (JACM)*, 45(5):753–782, 1998.

[6] Sanjeev Arora, Elad Hazan, and Satyen Kale. Fast algorithms for approximate semidefinite programming using the multiplicative weights update method. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 339–348. IEEE, 2005.

[7] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):1–37, 2009.

[8] Ainesh Bakshi, Nadiia Chepurko, and David P Woodruff. Robust and sample optimal algorithms for psd low-rank approximation. *arXiv preprint arXiv:1912.04177*, 2019.

[9] Ainesh Bakshi and David Woodruff. Sublinear time low-rank approximation of distance matrices. In *Advances in Neural Information Processing Systems*, pages 3782–3792, 2018.

[10] Maria-Florina Balcan and Nicholas JA Harvey. Learning submodular functions. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 793–802, 2011.

[11] Maria-Florina Balcan, Yi Li, David P Woodruff, and Hongyang Zhang. Testing matrix rank, optimally. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 727–746. SIAM, 2019.

[12] Maria-Florina F Balcan and Hongyang Zhang. Noise-tolerant life-long matrix completion via adaptive sampling. In *Advances in Neural Information Processing Systems*, pages 2955–2963, 2016.

[13] Siddharth Barman, Arnab Bhattacharyya, and Suprovat Ghoshal. Testing sparsity over known and unknown bases. In *International Conference on Machine Learning*, pages 491–500, 2018.

[14] Joshua Batson, Daniel A Spielman, Nikhil Srivastava, and Shang-Hua Teng. Spectral sparsification of graphs: theory and algorithms. *Communications of the ACM*, 56(8):87–94, 2013.

[15] Omri Ben-Eliezer, Rajesh Jayaram, David P Woodruff, and Eylon Yogev. A framework for adversarially robust streaming algorithms. In *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pages 63–80, 2020.

[16] Vladimir Braverman, Stephen Chestnut, Robert Krauthgamer, Yi Li, David Woodruff, and Lin Yang. Matrix norms in data streams: Faster, multi-pass and row-order. In *International Conference on Machine Learning*, pages 649–658, 2018.

[17] Vladimir Braverman, Robert Krauthgamer, Aditya Krishnan, and Roi Sinoff. Schatten norms in matrix streams: Hello sparsity, goodbye dimension. *arXiv preprint arXiv:1907.05457*, 2019.

[18] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

[19] Kenneth L Clarkson and David P Woodruff. Low-rank approximation and regression in input sparsity time. *Journal of the ACM (JACM)*, 63(6):54, 2017.

[20] Jon Dattorro. *Convex optimization & Euclidean distance geometry*. Lulu. com, 2010.

[21] Michel Marie Deza and Monique Laurent. *Geometry of cuts and metrics*, volume 15. Springer, 2009.

[22] Ilias Diakonikolas and Daniel M Kane. Recent advances in algorithmic high-dimensional robust statistics. *arXiv preprint arXiv:1911.05911*, 2019.

[23] Huaian Diao, Rajesh Jayaram, Zhao Song, Wen Sun, and David Woodruff. Optimal sketching for kronecker product regression and low rank approximation. In *Advances in Neural Information Processing Systems*, pages 4737–4748, 2019.

[24] Yonina C Eldar and Gitta Kutyniok. *Compressed sensing: theory and applications*. Cambridge university press, 2012.

[25] Ankit Garg, Yin Tat Lee, Zhao Song, and Nikhil Srivastava. A matrix expander chernoff bound. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1102–1114, 2018.

[26] Anna Gilbert and Piotr Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 98(6):937–947, 2010.

[27] Alex Gittens and Joel A Tropp. Tail bounds for all eigenvalues of a sum of random matrices. *arXiv preprint arXiv:1104.4513*, 2011.

[28] Paul Glendinning. *Stability, instability and chaos: an introduction to the theory of nonlinear differential equations*, volume 11. Cambridge university press, 1994.

[29] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

[30] Oded Goldreich. Introduction to testing graph properties. In *Property testing*, pages 105–141. Springer, 2010.

[31] Oded Goldreich. *Introduction to property testing*. Cambridge University Press, 2017.

[32] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM (JACM)*, 45(4):653–750, 1998.

[33] Insu Han, Dmitry Malioutov, Haim Avron, and Jinwoo Shin. Approximating spectral sums of large-scale matrices using stochastic chebyshev approximations. *SIAM Journal on Scientific Computing*, 39(4):A1558–A1585, 2017.

[34] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM (JACM)*, 53(3):307–323, 2006.

[35] Piotr Indyk, Ali Vakilian, Tal Wagner, and David Woodruff. Sample-optimal low-rank approximation of distance matrices. *arXiv preprint arXiv:1906.00339*, 2019.

[36] Rajesh Jayaram, Gokarna Sharma, Srikanta Tirthapura, and David P Woodruff. Weighted reservoir sampling from distributed streams. In *Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pages 218–235, 2019.

[37] Rajesh Jayaram and David P Woodruff. Perfect lp sampling in a data stream. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 544–555. IEEE, 2018.

[38] Rajesh Jayaram and David P Woodruff. Towards optimal moment estimation in streaming and distributed models. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[39] Hossein Jowhari, Mert Sağlam, and Gábor Tardos. Tight bounds for lp samplers, finding duplicates in streams, and related problems. In *Proceedings of the Thirtieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '11, pages 49–58, New York, NY, USA, 2011. ACM. URL: http://doi.acm.org/10.1145/1989284.1989289, doi:10.1145/1989284.1989289.

[40] Daniel M Kane, Jelani Nelson, and David P Woodruff. On the exact space complexity of sketching and streaming small norms. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 1161–1178. SIAM, 2010.

[41] Ramakrishnan Kannan, Mariya Ishteva, Barry Drake, and Haesun Park. Bounded matrix low rank approximation. In *Non-negative Matrix Factorization Techniques*, pages 89–118. Springer, 2016.

[42] Jonathan A Kelner, Lorenzo Orecchia, Aaron Sidford, and Zeyuan Allen Zhu. A simple, combinatorial algorithm for solving sdd systems in nearly-linear time. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 911–920, 2013.

[43] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[44] Robert Krauthgamer and Ori Sasson. Property testing of data dimensionality. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 18–27. Society for Industrial and Applied Mathematics, 2003.

[45] Rasmus Kyng, Kyle Luh, and Zhao Song. Four deviations suffice for rank 1 matrices. *Advances in Mathematics*, 375:107366, 2020.

[46] Rasmus Kyng and Zhao Song. A matrix chernoff bound for strongly rayleigh distributions and spectral sparsifiers from a few random spanning trees. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 373–384. IEEE, 2018.

[47] Yi Li, Huy L Nguyen, and David P Woodruff. On sketching matrix norms and the top singular vector. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1562–1581. SIAM, 2014.

[48] Yi Li, Huy L Nguyen, and David P Woodruff. On approximating matrix norms in data streams. *SIAM Journal on Computing*, 48(6):1643–1697, 2019.

[49] 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*, pages 691–700, 2014.

[50] Yi Li and David P Woodruff. On approximating functions of the singular values in a stream. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 726–739, 2016.

[51] Yi Li and David P Woodruff. Tight bounds for sketching the operator norm, schatten norms, and subspace embeddings. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[52] Yi Li and David P Woodruff. Embeddings of schatten norms with applications to data streams. In *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[53] Françoise Lust-Piquard and Gilles Pisier. Non commutative khintchine and paley inequalities. *Arkiv för matematik*, 29(1-2):241–260, 1991.

[54] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison—singer problem. *Annals of Mathematics*, pages 327–350, 2015.

[55] Morteza Monemizadeh and David P Woodruff. 1-pass relative-error lp-sampling with applications. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 1143–1160. SIAM, 2010.

[56] Cameron Musco and David P Woodruff. Sublinear time low-rank approximation of positive semidefinite matrices. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 672–683. IEEE, 2017.

[57] Michal Parnas and Dana Ron. Testing metric properties. *Information and Computation*, 187(2):155–195, 2003.

[58] Gilles Pisier. Remarks on the non-commutative khintchine inequalities for 0¡ p¡ 2. *Journal of Functional Analysis*, 256(12):4128–4161, 2009.

[59] Gilles Pisier and Éric Ricard. The non-commutative khintchine inequalities for $0 < p < 1$. *Journal of the Institute of Mathematics of Jussieu*, 16(5):1103–1123, 2017.

[60] Mark Rudelson and Roman Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *Journal of the ACM (JACM)*, 54(4):21–es, 2007.

[61] Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral clustering of graphs with the bethe hessian. In *Advances in Neural Information Processing Systems*, pages 406–414, 2014.

[62] Isaac J Schoenberg. Remarks to maurice frechet's article"sur la definition axiomatique d'une classe d'espace distances vectoriellement applicable sur l'espace de hilbert. *Annals of Mathematics*, pages 724–732, 1935.

[63] Jean-Jacques E Slotine, Weiping Li, et al. *Applied nonlinear control*, volume 199. Prentice hall Englewood Cliffs, NJ, 1991.

[64] Zhao Song and Ruizhe Zhang. Hyperbolic polynomials i: Concentration and discrepancy. *arXiv preprint arXiv:2008.09593*, 2020.

[65] Daniel A Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90, 2004.

[66] Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

[67] David Steurer. Fast sdp algorithms for constraint satisfaction problems. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 684–697. SIAM, 2010.

[68] Xiaoming Sun, David P Woodruff, Guang Yang, and Jialin Zhang. Querying a matrix through matrix-vector products. *arXiv preprint arXiv:1906.05736*, 2019.

[69] Joel Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[70] Joel A Tropp. Norms of random submatrices and sparse approximation. *Comptes Rendus Mathematique*, 346(23-24):1271–1274, 2008.

[71] Alan M Turing. Rounding-off errors in matrix processes. *The Quarterly Journal of Mechanics and Applied Mathematics*, 1(1):287–308, 1948.

[72] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

[73] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[74] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.