The Approximate Rank of a Matrix and its Algorithmic Applications *

Noga Alon † Troy Lee ‡ Adi Shraibman § Santosh Vempala ¶

December 21, 2013

Abstract

We study the ϵ-rank of a real matrix A, defined for any $\epsilon > 0$ as the minimum rank over matrices that approximate every entry of A to within an additive ϵ. This parameter is connected to other notions of approximate rank and is motivated by problems from various topics including communication complexity, combinatorial optimization, game theory, computational geometry and learning theory. Here we give bounds on the ϵ-rank and use them for algorithmic applications. Our main algorithmic results are (a) polynomial-time additive approximation schemes for Nash equilibria for 2-player games when the payoff matrices are positive semidefinite or have logarithmic rank and (b) an additive PTAS for the densest subgraph problem for similar classes of weighted graphs. We use combinatorial, geometric and spectral techniques; our main new tool is an efficient algorithm for the following problem: given a convex body A and a symmetric convex body B, find a covering a A with translates of B.

1 Introduction

1.1 Background

A large body of work in theoretical computer science deals with various ways of approximating a matrix by a simpler one. The motivation from the design of approximation algorithms is clear. When the input to a computational problem is a matrix (that may represent a weighted graph, a payoff matrix in a two-person game or a weighted constraint satisfaction problem), the hope is that it is easier to solve or approximately solve the computational problem for the approximating matrix, which is simpler. If the notion of approximation is suitable for the problem at hand, then the solution will be an approximate solution for the original input matrix as well.

A typical example of this reasoning is the application of cut decomposition and that of regular decomposition of matrices. The cut-norm of a matrix A with a set of rows R and a set of columns C is

$$\max_{S \subset R, T \subset T} |\sum_{i \in S, j \in T} A_{ij}|.$$

*A preliminary version of this paper appeared in Proc. STOC 2013

†Schools of Mathematics and Computer Science, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by an ERC Advanced grant, by a USA-Israeli BSF grant, and by the Hermann Minkowski Minerva Center for Geometry in Tel Aviv University.

‡Centre for Quantum Technologies, Singapore. Email: troyjlee@gmail.com

§School of Computer Science, Academic College of Tel-Aviv Yaffo, Israel. Email: adish@mta.ac.il

¶School of Computer Science and ARC, Georgia Tech, Atlanta 30332. Email: vempala@gatech.edu. Supported in part by the NSF and by a Raytheon fellowship.

A cut matrix $B(S,T;r)$ is a matrix B for which $B_{ij} = r$ iff $i \in S, j \in T$ and $B_{ij} = 0$ otherwise. Frieze and Kannan showed that any n by m matrix B with entries in $[-1,1]$ can be approximated by a sum of at most $\frac{1}{\epsilon^2}$ cut matrices, in the sense that the cut norm of the difference between B and this sum is at most ϵmn. Such an approximation can be found efficiently and leads to several approximation algorithms for dense graphs—see [19]. Similar approximations of matrices can be given using variants of the regularity lemma of Szemerédi. These provide a more powerful approximation at the expense of increasing the complexity of the approximating matrix, and supply approximation algorithms for additional problems see [6, 18, 7].

All these methods, however, deal with global properties, as the approximation obtained by all these variants of the regularity lemma are not sensitive to local changes in the matrix. In particular, these methods cannot provide approximate solutions to problems like that of finding an approximate Nash equilibrium in a two person game, or that of approximating the maximum possible density of a subgraph on, say, $\sqrt{n}$ vertices in a given n vertex weighted graph. Motivated by applications of this type, it is natural to consider a stronger notion of approximation of a matrix, an approximation in the infinity norm, by a matrix of low rank. This motivates the following definition.

Definition 1.1 *For a real $n \times n$ matrix A, the* ϵ-rank *of A is defined as follows:*

$$\epsilon\text{-}rank(A) = \min\{\text{rank}(B) : B \in \Re^{n\times n}, \|A - B\|_\infty \leq \epsilon\}.$$

We will usually assume that the matrix A has entries in $[-1,1]$, but the definition holds for any real matrix. Define the density norm of a matrix A to be

$$\text{den}(A) = \max_{x,y\in\Re^n_+} \frac{|x^T A y|}{\|x\|_1 \|y\|_1}.$$

It is easy to verify that the following definition of ϵ-rank is equivalent to Definition 1.1:

$$\epsilon\text{-rank}(A) = \{\min \text{rank}(B) : B \in \Re^{n\times n}, \text{den}(A - B) \leq \epsilon\}.$$

The investigation of notions of simple matrices that approximate given ones is motivated not only by algorithmic applications, but by applications in complexity theory as well. Following Valiant [42] call a matrix A (r,s)-rigid if for any matrix B of the same dimensions as A and rank at most r, $A - B$ contains a row with at least s nonzero entries. Here the notion of simple matrix is thus a matrix of low rank, and the notion of approximation is to allow a limited number of changes in each row. Valiant proved that if an n by n matrix is $(\Omega(n), n^{\Omega(1)})$-rigid, then there is no arithmetic circuit of linear size and logarithmic depth that computes Ax for any given input x. Therefore, the main problem in this context (which is still wide open) is to give an explicit construction of such a rigid matrix.

Another notion that received a considerable amount of attention is the sign-rank of a real matrix. For a matrix A, $\text{rank}_\pm(A)$ is defined as the minimum rank over matrices each of whose entries has the same sign as the corresponding entry in the original matrix. The notion of approximation here refers to keeping the signs of the entries, while the simplicity of the approximating matrix is measured by its rank. The sign rank has played a useful role in the study of the unbounded error communication complexity of Boolean functions (see, e.g., [5, 20] and their references), in establishing lower bounds in learning theory and in providing lower bounds for the size of threshold-of-majority circuits computing a function in AC^0 (see [35]). It is clear that for $-1, 1$ matrices or matrices with entries of absolute value exceeding ϵ, ϵ-rank$(A) \geq \text{rank}_\pm(A)$, and simple examples show that in many cases the ϵ-rank is far larger than the sign-rank.

Another line of work on ϵ-rank is motivated by communication complexity [15, 30, 34]. The ϵ-*error private-coin* communication complexity of a Boolean function is bounded from below by the logarithm of the ϵ-rank of the corresponding communication matrix (see [27]). In addition, the ϵ-error private-coin *quantum* communication complexity of a Boolean function is bounded from below by $\log \epsilon\text{-rank}(A)/2$ where A is the communication matrix of the function. Using this connection, it has been proven that for fixed ϵ the 2^n by 2^n set-disjointness matrix has ϵ-rank $2^{\Theta(\sqrt{n})}$, using the quantum protocol of Aaronson and Ambainis [1] for disjointness, and Razborov's lower bound for the problem (see [34], where there is a lower bound for the trace-norm of any matrix that ϵ-approximates the disjointness matrix restricted to the sets of size $n/4$.) In another work, Lee and Shraibman [28] have given algorithmic bounds on ϵ-rank via the γ_2 norm. The γ_2-norm of a real matrix A, denoted by $\gamma_2(A)$, is the minimum possible value of the product $c(X)c(Y)$, where $c(Z)$ is the maximum ℓ_2-norm of a column of a matrix Z, and the minimum is taken over all factorizations of A of the form $A = X^tY$. For a sign matrix A and for $\alpha \geq 1$, let $\gamma_2^\alpha(A)$ denote the minimum possible value of $\gamma_2(B)$, where B ranges over all matrices of the same dimension as A that satisfy $1 \leq A_{ij} \cdot B_{ij} \leq \alpha$ for all admissible i, j. Let $rank_\alpha(A)$ denote the minimum possible rank of such a matrix B. Note that for $\alpha = 1 + \epsilon$ this is (roughly) the $\epsilon/2$-rank of A. In [28] it is shown that $rank_\alpha(A)$ and $\gamma_2^\alpha(A)$ are polynomially related for any sign matrix A, up to a poly-logaritmic factor in the dimension of A. Since $\gamma_2^\alpha(A)$ can be computed efficiently using semi-definite programming, this provides a (rough) approximation algorithm for the ϵ-rank of a given sign matrix.

For the special case of the n by n identity matrix the ϵ-rank has been studied and provided several applications. In [2] it is shown that it is at least $\Omega(\frac{\log n}{\epsilon^2 \log(1/\epsilon)})$ and at most $O(\frac{\log n}{\epsilon^2})$. This is used in [2] to derive several applications in geometry, coding theory, extremal finite set theory and the study of sample spaces supporting nearly independent random variables. See also [11] for a more recent application of the lower bound (for the special case $\epsilon < 1/\sqrt{n}$) in combinatorial geometry and in the study of locally correctable codes over real and complex numbers.

The notion of the ϵ-rank of a matrix is also related to learning and to computational geometry. Indeed, the problem of computing the ϵ-rank of a given n by n matrix A is equivalent to the geometric problem of finding the minimum possible dimension of a linear subspace of R^n that intersects the aligned cubes of edge length 2ϵ centered at the columns of A. In learning, the problem of learning with margins, the fat-shattering dimension of a family of functions and the problem of learning functions approximately are all related to this notion (see e.g., [4] and the references therein).

The focus of our paper is algorithmic applications of ϵ-rank; we state our results in the next section.

1.2 Results

Our results are divided into three related topics: bounds on the ϵ-rank, algorithmic applications, and efficient covering of one convex body by another.

Bounds on approximate rank. We begin with bounds on the ϵ-rank. A well known result of Forster [20] asserts that the sign-rank of any n by n Hadamard matrix H is at least $\Omega(\sqrt{n})$. This clearly implies the same lower bound for the ϵ-rank of any such matrix for any $\epsilon < 1$. Using the approximate γ_2 norm, Linial et al. [29] further show that $\epsilon\text{-rank}(H) \geq (1 - 2\epsilon)n$. The following gives a slightly stronger estimate, which is tight.

Theorem 1.2 *For any $n \times n$ Hadamard matrix H and any $0 < \epsilon < 1$, ϵ-rank$(H) \geq (1 - \epsilon^2)n$.*

Next we show a lower bound on the approximate rank of a random d-regular graph. Let A_G denote the adjacency matrix of a graph G, and $\bar{A}_G$ denote the "signed" adjacency matrix where the (i, j)

entry is 1 for an edge and -1 for a non-edge. We show, in fact, a stronger statement that for a random d-regular graph the sign rank of $\bar{A}_G$ is $\Omega(d)$.

A closely related result was shown by Linial and Shraibman [30]. Similarly to the sign rank, define $\gamma_2^\infty(A)$ as the minimum γ_2 norm of a matrix that has the same sign pattern as A and all entries at least 1 in magnitude. It is known that $\mathrm{rank}_\pm(A) = O(\gamma_2^\infty(A)\log(mn))$ [12] and also that the sign rank can be exponentially smaller than γ_2^∞ [14, 40]. Linial and Shraibman show that $\gamma_2^\infty(A_G) = \Omega(\sqrt{d})$ for a random d-regular graph, which also implies an $\Omega(d)$ lower bound on the ϵ-approximate rank for any constant $\epsilon < 1/2$.

Our proof is different from these γ_2 techniques and relies, as in previous lower bounds on the sign rank [5], on Warren's theorem from real algebraic geometry [44].

Theorem 1.3 *For almost all d-regular graphs G on n vertices $\mathrm{rank}_\pm(\bar{A}_G) = \Omega(d)$ for the adjacency matrix of G.*

By "almost all" we mean that the fraction of d-regular graphs on n vertices for which the statement holds tends to 1 as n tends to infinity. Note that as $\bar{A}_G = 2A_G - J$, where J is the all ones matrix, this result also implies that $\epsilon\text{-rank}(A_G) = \Omega(d)$ for any $\epsilon < 1/2$. The lower bound here is tight for ϵ-rank up to a $\log n$ factor: it follows from results in [29, 28] that for any fixed ϵ bounded away from zero the ϵ-rank of A_G for every d-regular graph on n vertices is $O(d \log n)$. The lower bound is tight for sign rank, as a result of [5] implies that the sign rank is bounded by the maximal number of sign changes in each row.

The ϵ-rank of any positive semidefinite matrix can be bounded from above as stated in the next theorem. The theorem follows via a direct application of the Johnson-Lindenstrauss lemma [25].

Theorem 1.4 *For a symmetric positive semi-definite $n \times n$ matrix A with $|A_{ij}| \leq 1$, we have*

$$\epsilon\text{-rank}(A) \leq \frac{9 \log n}{\epsilon^2 - \epsilon^3}.$$

Note that this is nearly tight, by the above mentioned lower bound for the ϵ-rank of the identity matrix.

The last theorem can be extended to linear combinations of positive semi-definite (=PSD) matrices.

Corollary 1.5 *Let $A = \sum_{i=1}^m \alpha_i B_i$ where $|\alpha_i| \leq 1$ are scalars and B_i are $n \times n$ PSD matrices with entries at most 1 in magnitude. Then*

$$\epsilon\text{-rank}(A) \leq C m^2 \frac{\log n}{\epsilon^2}$$

for an absolute constant C.

Algorithmic applications. The notion of ϵ-rank and the bounds above have algorithmic applications. The high-level idea is that we can replace the input matrix of a given instance by another that approximates it in every entry and has lower rank.

For a 2-player game specified by two payoff matrices A, B, a fundamental problem is computing an approximate ϵ-Nash equilibrium of the game (we define this precisely in Section 4.1). Since changing any entry of a payoff matrix can affect the payoff to a player by at most ϵ, the ϵ-rank is a very natural notion in this context. Lipton et al. [31] showed that an ϵ-Nash for any 2-player game can be computed in time $n^{O(\log n/\epsilon^2)}$ and it has been an important open question to determine whether

this problem has a PTAS (i.e., an algorithm of complexity of $n^{f(\epsilon)}$). Our result establishes a PTAS when $A+B$ is PSD or when $A+B$ has ϵ-rank $O(\log n)$ (and we are given an ϵ-approximating matrix C of $A+B$ with rank $O(\log n)$), where A and B are the payoff matrices of the game. Note that the special case when $A+B=0$ corresponds to zero-sum games, a class for which the exact Nash equilibrium can be computed using linear programming. The setting of $A+B$ having *constant* rank was investigated by Kannan and Theobald [26], who gave a PTAS for the case when the rank is a constant. Their algorithm has running time $n^{\mathrm{poly}(d,1/\epsilon)}$; ours has complexity $[O(1/\epsilon)]^d poly(n)$, giving a PTAS for ϵ-rank $d = O(\log n)$.

Theorem 1.6 *Let $A, B \in [-1,1]^{n\times n}$ be the payoff matrices of a 2-player game. If $A+B$ is positive semidefinite, then an ϵ-Nash equilibrium of the game can be computed by a deterministic algorithm using $poly(n)$ space and time*

$$n^{O(\log(1/\epsilon)/\epsilon^2)}$$

i.e., there is a PTAS to compute an ϵ-Nash equilibrium.

The above theorem can be recovered by a similar algorithm using the γ_2-norm approach.

The next theorem is more general (the γ_2 approach seems to achieve only a weaker bound of $(n/\epsilon)^{O(d)}$ here).

Theorem 1.7 *Let $A, B \in [-1,1]^{n\times n}$ be the payoff matrices of a 2-player game. Suppose $(\epsilon/2)$-rank$(A+B) = d$ and suppose we have a matrix C of rank d satisfying $\|A+B-C\|_\infty \le \epsilon/2$. Then, an ϵ-Nash equilibrium of the game can be computed in time*

$$\left(\frac{1}{\epsilon}\right)^{O(d)} poly(n).$$

Note that in particular if the rank of $A+B$ is $O(\log n)$ we can simply take $C = A+B$ and get a polynomial-time algorithm.

Our second application is to finding an approximately densest bipartite subgraph, a problem that has thus far evaded a PTAS even in the dense setting. In fact, there are hardness results indicating that even the dense case is hard to approximate to within any constant factor, see [3]. Here we observe that we can efficiently compute a good additive approximation for the special case that the input matrix has a small ϵ-rank (assuming we are given an approximating matrix of low rank). For a matrix A with entries in $[0,1]$ and subsets S, T of rows and columns, let $A_{S,T}$ be the submatrix induced by S and T. The density of the submatrix $A_{S,T}$ is

$$\mathrm{density}(A_{S,T}) = \frac{\sum_{i\in S, j\in T} A_{ij}}{|S||T|}$$

the average of the entries of the submatrix.

Theorem 1.8 *Let A be an $n\times n$ real matrix with entries in $[0,1]$. Then for any integer $1 \le k \le n$, there is a deterministic algorithm to find subsets S, T of rows and columns with $|S| = |T| = k$ s.t.,*

$$\mathrm{density}(A_{S,T}) \ge \max_{|U|=|V|=k} \mathrm{density}(A_{U,V}) - \epsilon.$$

Its time complexity is bounded by $n^{O(\log(1/\epsilon)/\epsilon^2)}$ if A is PSD and by $\left(\frac{1}{\epsilon}\right)^{O(d)} poly(n)$ where $d = (\epsilon/2)$-rank(A) (and we are given an $\epsilon/2$-approximation of rank d).

Note that this is a bipartite version of the usual densest subgraph problem in which the objective is to find the density of the densest subgraph on (say) $2k$ vertices in a a given (possibly weighted) input graph. It is easy to see that the answers to these two problems can differ by at most a factor of 2, and as the best known polynomial time approximation algorithm for this problem, given in [13], only provides an $O(n^{1/4+o(1)})$-approximation for an n-vertex graph, this bipartite version also appears to be very difficult for general graphs.

Covering convex bodies. Our results for general matrices are based on an algorithm to efficiently find a near-optimal cover of a convex body A by translates of another convex body B. We state this result here as it seems to be of independent interest. Let $N(A, B)$ denote the minimum number of translates of B required to cover A. It suffices for the algorithm that an input convex body is specified by a membership oracle, along with bounds on its inscribing and circumscribing ball radii and a point inside the body [23]. In the applications in this paper, the bodies will in fact be explicit polytopes.

Theorem 1.9 *For any convex body A and any centrally symmetric convex body B in $\Re^d$, a cover of A using translates of B of size $N(A, B)2^{O(d)}$ can be enumerated by a deterministic algorithm using either*

1. *$N(A, B)2^{O(d)}$ time and $2^{O(d)}$ space, or*
2. *$N(A, B)O(d)^d$ time and poly(d) space.*

Although we do not prove it here, we expect that the theorem can be extended to asymmetric convex bodies[1].

1.3 Organization

The rest of the paper is organized as follows. In Section 2 we describe lower and upper bounds for the ϵ-rank of a matrix, presenting the proofs of Theorems 1.2, 1.3, 1.4 and Corollary 1.5. In Section 3 we describe efficient constructions of ϵ-nets which are required to derive the algorithmic applications, and prove Theorem 1.9. Section 4 contains the algorithmic applications including the proofs of Theorems 1.6, 1.7 and 1.8. The final Section 5 contains some concluding remarks, open problems and plans for future extensions.

2 Bounds on ϵ-rank

2.1 Lower bounds

We start with the proof of Theorem 1.2, which is based on some spectral techniques. For a symmetric m by m matrix A let $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_m(A)$ denote its eigenvalues, ordered as above. We need the following simple lemma.

Lemma 2.1 *(i) Let A be an n by n real matrix, then the $2n$ eigenvalues of the symmetric $2n$ by $2n$ matrix*

$$B = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} \tag{1}$$

[1] In an earlier version of this paper, we incorrectly claimed a p$oly(d)$ space bound while achieving time $N(A, B)2^{O(d)}$. Fortunately, this does not affect any of the time complexity bounds, of the covering theorem or of the applications.

appear in pairs λ and $-\lambda$.
(ii) For any two symmetric m by m matrices B and C and all admissible values of i and j

$$\lambda_{i+j-1}(B+C) \leq \lambda_i(B) + \lambda_j(C).$$

Proof.
(i) Let (x, y) be the eigenvector of an eigenvalue λ of B, where x and y are real vectors of length n. Then $Ay = \lambda x$ and $A^T x = \lambda y$. It is easy to check that the vector $(x, -y)$ is an eigenvector of the eigenvalue $-\lambda$ of B. This proves part (i).
(ii) The proof is similar to that of the Weyl Inequalities - c.f., e.g., [22], and follows from the variational characterization of the eigenvalues. It is easy and well known that

$$\lambda_i(B) = \min_{U,\ dim(U)=m-i+1} \max_{x \in U, \|x\|=1} x^T B x$$

where the minimum is taken over all subspaces U of dimension $m - i + 1$. Similarly,

$$\lambda_j(C) = \min_{W,\ dim(W)=m-j+1} \max_{x \in W, \|x\|=1} x^T C x.$$

Put $V = U \cap W$. Clearly, the dimension of V is at least $m - i - j + 2$ and for any $x \in V, \|x\| = 1$,

$$x^T(B+C)x = x^T B x + x^T C x \leq \lambda_i(B) + \lambda_j(C).$$

Therefore, $\lambda_{i+j-1}(B+C)$ is equal to

$$\min_{Z,\ dim(Z)=m-i-j+2} \max_{x \in Z, \|x\|=1} x^T(B+C)x \leq \lambda_i(B) + \lambda_j(C).$$

□

Proof. (of Theorem 1.2). Let E be an n by n matrix, $\|E\|_\infty \leq \epsilon$ so that the rank of $H + E$ is the ϵ-rank of H. Let B and C be the following two symmetric $2n$ by $2n$ matrices

$$B = \begin{pmatrix} 0 & H \\ H^T & 0 \end{pmatrix}, \tag{2}$$

$$C = \begin{pmatrix} 0 & E \\ E^T & 0 \end{pmatrix}. \tag{3}$$

Since H is a Hadamard matrix, $B^T B = B^2$ is n times the $2n$ by $2n$ identity matrix. It follows that $\lambda_i^2(B) = n$ for all i, and by Lemma 2.1, part (i), exactly n eigenvalues of B are $\sqrt{n}$ and exactly n are $-\sqrt{n}$. In particular $\lambda_{n+1}(B) = -\sqrt{n}$.

The absolute value of every entry of E is at most ϵ, and thus the square of the Frobenuis norm of C is at most $2n^2\epsilon^2$. As this is the trace of $C^T C$, that is, the sum of squares of eigenvalues of C, it follows that C has at most $2n\epsilon^2$ eigenvalues of absolute value at least $\sqrt{n}$. By Lemma 2.1, part (i), this implies that there are at most $\epsilon^2 n$ eigenvalues of C of value at least $\sqrt{n}$ and thus $\lambda_{\lfloor \epsilon^2 n \rfloor + 1}(C) < \sqrt{n}$. By Lemma 2.1, part (ii) we conclude that $\lambda_{n+\lfloor \epsilon^2 n \rfloor + 1}(B + C) < 0$. Therefore $B + C$ has at least $n - \lfloor \epsilon^2 n \rfloor$ negative eigenvalues and hence also at least $n - \lfloor \epsilon^2 n \rfloor$ positive eigenvalues. Therefore its rank, which is exactly twice the rank of $H + E$, is at least $2(n - \lfloor \epsilon^2 n \rfloor)$, completing the proof. □

Remark: By the last theorem, if $\epsilon < \frac{1}{\sqrt{n}}$ then the ϵ-rank of any n by n Hadamard matrix is n. This is tight in the sense that for any n which is a power of 4 there is an n by n Hadamard matrix with ϵ-rank $n-1$ for $\epsilon = \frac{1}{\sqrt{n}}$. Indeed, the matrix

$$H_1 = \begin{pmatrix} +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & +1 \\ +1 & +1 & -1 & +1 \\ +1 & -1 & -1 & -1 \end{pmatrix} \tag{4}$$

is a 4 by 4 Hadamard matrix in which the sum of elements in every row is of absolute value 2. Thus, the tensor product of k copies of this matrix is a Hadamard matrix of order $n = 4^k$ in which the sum of elements in every row is in absolute value 2^k. We can thus either add or subtract $\frac{1}{2^k} = \frac{1}{\sqrt{n}}$ to each element of this matrix and get a matrix in which the sum of elements in every row is 0, that is, a singular matrix.

In a similar vein, we can show that the ϵ-rank of any n-by-n matrix A with entries in $[-1,1]$ is at most $n-1$ for $\epsilon = \frac{6}{\sqrt{n}}$. Indeed, for any such $A = (a_{ij})$ there are, by the main result of [41], $\delta_j \in \{-1,1\}$ so that for every i, $|\sum_{j=1}^n a_{ij}\delta_j| \leq 6\sqrt{n}$. Fix such δ_j and define, for each i, $\epsilon_i = (\sum_{j=1}^n a_{ij}\delta_j)/n$. Therefore, for every i, $|\epsilon_i| \leq \frac{6}{\sqrt{n}}$ and $\sum_{j=1}^n (\delta_j a_{ij} - \epsilon_i) = 0$, implying that the inner product of the vector $(\delta_1, \delta_2, \ldots, \delta_n)$ with any row of the matrix $(a_{ij} - \delta_j \epsilon_i)$ is zero. Since $|\delta_j \epsilon_i| \leq \frac{6}{\sqrt{n}}$ for all admissible i, j, this shows that the ϵ-rank of A is at most $n-1$, for $\epsilon = \frac{6}{\sqrt{n}}$.

A similar reasoning shows that for any n-by-n matrix A with entries $a_{ij} \in [-1,1]$, the ϵ-rank is at most $n - \Omega(\frac{\epsilon^2 n}{\log n})$. The argument here proceeds as follows. Break the columns of A into $k = \frac{\epsilon^2 n}{2 \log n}$ disjoint subsets of nearly equal size, thus breaking A into k submatrices, each having n rows and roughly $\frac{2 \log n}{\epsilon^2}$ columns. It suffices to show that the ϵ-rank of each of these submatrices is at most one less than the number of its columns. Let A' be such a submatrix and let p be the number of its columns. Let $\delta = (\delta_1, \ldots, \delta_p)$ be a random vector in $\{-1,1\}^p$. By the the Chernoff-Hoeffding Inequality, with positive (and in fact high) probability the inner product of each row of A' with δ is of absolute value at most ϵp. We can now proceed as before to get a matrix B' satisfying $\|A' - B'\|_\infty \leq \epsilon$ so that each row of B' is orthogonal to the vector δ, completing the proof.

Recall that the sign rank, $\mathrm{rank}_\pm(A)$, is the minimal rank of a matrix with the same sign pattern as A. In [5] it is shown that there are n-by-n sign matrices A for which $\mathrm{rank}_\pm(A) \geq n/32$, thus for larger ϵ (close to 1) we cannot expect that ϵ-rank$(A) \leq (1-\epsilon^2)n$ for all A.

The proof of Theorem 1.3 is a simple consequence of a known bound of Warren for the number of sign pattern of real polynomials. For a sequence $P_1, P_2, \ldots, P_m$ of real polynomials in ℓ variables, and for $x \in R^\ell$ for which $P_i(x) \neq 0$ for any i, the sign pattern of the polynomials P_i at x is the vector

$$(sign(P_1(x)), sign(P_2(x)), \ldots, sign(P_m(x))) \in \{-1,1\}^m.$$

Let $s(P_1, P_2, \ldots, P_m)$ denote the total number of sign patterns of the polynomials P_i as x ranges over all points of R^ℓ in which no P_i vanishes. Note that this is bounded by the number of connected components of the semi-variety $V = \{x \in R^\ell : P_i(x) \neq 0 \text{ for all } 1 \leq i \leq m\}$. In this notation, the theorem of Warren is the following.

Theorem 2.2 (Warren [44]) *Let $P_1, P_2, \ldots, P_m$ be real polynomials in ℓ variables, each of degree at most k. Then the number of connected components of $V = \{x \in R^\ell : P_i(x) \neq 0 \text{ for all } 1 \leq i \leq m\}$ is at most $(4ekm/\ell)^\ell$. Therefore, $s(P_1, P_2, \ldots, P_m) \leq (4ekm/\ell)^\ell$.*

Proof. [of Theorem 1.3] It is easy and well known that for any admissible n and d (with nd even) the number of labelled d regular graphs on n vertices is $[\Theta(n/d)]^{nd/2}$. In particular this number is at least $(cn/d)^{nd/2}$ for some absolute positive constant c. Let us estimate the number of d-regular graphs G for which $\mathrm{rank}_{\pm}(\bar{A}_G) \leq r$. For any such matrix there are two matrices B and C where B is an n by r matrix and C is an r by n matrix, so that $\mathrm{sgn}(BC) = \bar{A}_G$, where $\mathrm{sgn}(A)$ is the sign function applied entrywise to A. Thinking about the $2nr$ entries of the matrices B and C as variables, every entry of their product BC is a degree 2 polynomial in these variables. This means that the number of such adjacency matrices that have sign rank bounded by r is at most the number of sign patterns of n^2 polynomials, each of degree 2, in $\ell = 2nr$ variables. (It actually suffices to look at half of the entries, as the matrix is symmetric, but this will not lead to any substantial change in the estimate obtained). By Theorem 2.2 this number is at most

$$\left(\frac{4e2n^2}{2nr}\right)^{2nr}.$$

For $r < c'd$ for an appropriate absolute positive constant c' this is exponentially smaller than the number of labelled d-regular graphs on n vertices, completing the proof. □

2.2 Upper bounds

We will use the Johnson-Lindenstrauss (JL, for short) Lemma [25]. The following version is from [9, 43].

Lemma 2.3 *Let R be an $n \times k$ matrix, $1 \leq k \leq n$, with i.i.d. entries from $N(0, 1/k)$. For any $x, y \in \Re^n$ with $\|x\|, \|y\| \leq 1$,*

$$\Pr(|(R^T x)^T (R^T y) - x^T y| \geq \epsilon) < 2e^{-(\epsilon^2 - \epsilon^3)k/4}.$$

Proof. (of Theorem 1.4). Since A is PSD there is a matrix B so that $A = BB^T$. Let R be a random $n \times k$ matrix with entries from $N(0, 1/k)$. Consider $\tilde{A} = BRR^T B^T$.

For two vectors $x, y \in \Re^n$,

$$\mathrm{E}(x^T RR^T y) = x^T y$$

and by the JL Lemma,

$$\Pr(|x^T RR^T y - x^T y| \geq \epsilon) \leq 2e^{-(\epsilon^2 - \epsilon^3)k/4}.$$

Setting, say, $k = 9 \ln n/(\epsilon^2 - \epsilon^3)$ we conclude that whp

$$\|A - \tilde{A}\|_\infty \leq \epsilon.$$

□

Proof. (of Corollary 1.5). Put

$$A_1 = \frac{1}{\sum_{\alpha_i > 0} \alpha_i} \sum_{\alpha_i > 0} \alpha_i B_i$$

and

$$A_2 = \frac{1}{\sum_{\alpha_i < 0} |\alpha_i|} \sum_{\alpha_i < 0} |\alpha_i| B_i.$$

Then A_1 and A_2 are positive semi-definite with entries of magnitude at most 1. We can thus apply Theorem 1.4 to approximate the entries of each of them to within $\frac{\epsilon}{m}$ and obtain the desired result by expressing A as a linear combination of A_1 and A_2 with coefficients whose sum of magnitudes is at most m. □

We can sometimes combine the Johnson-Lindenstrauss lemma with some information about the negative eigenvalues and eigenvectors of a matrix to derive nontrivial bounds for its ϵ-rank. A more elaborate use of the lemma arises when trying to estimate the ϵ-rank of the n by n matrix $A = (a_{ij})$ in which $a_{ij} = +1$ for all $i \geq j$ and $a_{ij} = -1$ otherwise. This matrix, also known as the *half-graph* matrix or the *greater-than* matrix, is motivated by applications in computational complexity as well as by regularity partitions. It is known that its ϵ-rank is $\Omega(\log n/(\epsilon^2 \log(1/\epsilon)))$ [2] and is also $\Omega(\log^2 n)$ for any $\epsilon < 0.99$. The latter inequality follows from the result of [21] that $\gamma_2^\alpha(A) \geq \Omega(\log n)$ for any α, and the known relation between the ϵ-rank of A and $\gamma_2^\alpha(A)$ for $\alpha = 1 + \Theta(\epsilon)$.

The best-known upper bound is that for any ϵ the ϵ-rank of A is $O(\log^3 n/\epsilon^2)$ where here one combines the result of [32] that $\gamma_2(A) = O(\log n)$ with the the upper bound on approximate rank in terms of approximate γ_2 norm [28].

To see this directly, consider a decomposition of A first into an upper-triangular matrix and a complementary lower-triangular matrix. For the upper-triangular 0-1 matrix, we extract the large block of all 1 entries, roughly the upper right quarter of the matrix and let A_1 be the matrix of the same dimensions as A with this block as its support. The remaining matrix can be viewed as two upper-triangular matrices put together, we recursively extract a block of 1's from each of them and let A_2 be the matrix (of the same dimensions as A) with support equal to these two blocks. This process clearly terminates after $\lceil \log n \rceil$ steps, giving a decomposition into $l = O(\log n)$ matrices. Now we notice that for each matrix in the decomposition, upon removing the all-zero rows and columns, we are left with a block-diagonal matrix, with each block consisting entirely of 1's. It is easy to see that using the JL lemma, any such block-diagonal matrix has ϵ-rank $O(\log m/\epsilon^2)$ where m is the number of blocks; all rows of the same block are mapped to the same random vector and so the entry in the approximate matrix depends only on the block numbers of the indices. We assign a single random vector to each index $i \in [n]$ formed by concatenating the random vectors for each element of the decomposition, i.e., a random vector from $N(0, 1/k)^{kl}$. Then the expectation of the inner product of two such indices is 0 or 1 corresponding to the entry, and the probability of deviation from the expectation of any single entry by an additive tl is at most $e^{-kl(t^2-t^3)/4}$. Using $l = O(\log n)$ and setting $t = \epsilon/l$, we get a probability of failure bounded by $e^{-k(\epsilon^2-\epsilon^3)/l}$. Thus setting $k = O(l \log n/\epsilon^2)$, we ensure that with high probability, every entry deviates from its expectation by at most ϵ. The total dimension of the embedding is $kl = O(\log^3 n/\epsilon^2)$.

3 Constructing ϵ-nets

For a matrix A that has small ϵ-rank, we will be able to construct small ϵ-nets to approximate the quadratic form $x^T Ay$ for any x, y of ℓ_1-norm 1 to within additive ϵ. We describe the construction of these nets in this section, then apply them to some algorithmic problems in what follows. The construction in the PSD case is more explicit and independent of the input matrix; we describe this first.

Theorem 3.1 *Let $A = BB^T$, where A is an $n \times n$ positive semidefinite matrix with entries in $[-1, 1]$ and B is $n \times d$. Let $\Delta = \Delta_n = \{x \in R^n, \|x\|_1 = 1, x \geq 0\}$. There is a finite set $S \subset \Re^d$ independent*

of A, B such that

$$\forall x \in \Delta, \exists \tilde{x} \in S : \|B^T x - \tilde{x}\|_\infty \leq \frac{\epsilon}{\sqrt{d}}$$

with $|S| = O(1/\epsilon)^d$. Moreover, S can be computed in time $O(1/\epsilon)^d poly(n)$.

Proof. We note that since the diagonal entries of A are at most 1, every column of B^T has 2-norm at most 1. For $x \in \Delta$, we have $y = B^T x \in [-1, 1]^d$. Let us classify the entries of y into buckets based on their magnitude. Let $m = \lceil \log(\sqrt{d}) \rceil$ and

$$b_j = \left| \{i : \frac{1}{2^j} \leq |y_i| \leq \frac{1}{2^{j-1}}\} \right|$$

for $j = 1, \ldots m-1$, and $b_m = |\{i : |y_i| \leq 1/2^{m-1}\}|$. Then

$$\sum_{j=1}^{m-1} b_j 2^{-2j} \leq 1.$$

We call the vector $(b_1, b_2, \ldots, b_m)$ the *profile* of a vector y. Thus, $b_j \leq 2^{2j}$ for all $j \leq m$.

We will now construct an ϵ-net by discretizing each coordinate to multiples of $\epsilon/\sqrt{d}$. If we replace each coordinate of a vector y by its nearest multiple of $\epsilon/\sqrt{d}$, then the resulting vector $\tilde{y}$ satisfies

$$\|y - \tilde{y}\|_\infty \leq \frac{\epsilon}{\sqrt{d}} \text{ and thus } \|y - \tilde{y}\| \leq \epsilon.$$

We now bound the size of this ϵ-net. The total number of distinct profiles is

$$\binom{d+m-1}{m-1} \; (\; < 2^{2d}). \tag{5}$$

For a fixed profile, the number of ways to realize the profile (by assigning coordinates to each bucket) is

$$\binom{d}{b_1}\binom{d-b_1}{b_2}\cdots\binom{d-\sum_{j=1}^{m-1} b_j}{b_m} < \prod_{i=1}^{\log_2 d} \binom{d}{d/2^i}. \tag{6}$$

This last product is bounded by

$$2^{d(H(1/2)+H(1/4)+H(1/8)+\ldots+H(1/d))} = 2^{O(d)},$$

where $H(x) = -x \log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function.

For each realization of a profile, the maximum size of the ϵ-net can be bounded as

$$\Pi_{j=1}^m \left(\frac{\sqrt{d}}{2^{j-2}\epsilon}\right)^{b_j} \leq (\frac{1}{\epsilon})^d \Pi_{j=1}^m \left(\frac{\sqrt{d}}{2^{j-2}}\right)^{2^{2j}} \leq (\frac{1}{\epsilon})^d 2^{O(d)}. \tag{7}$$

The product of (5), (6) and (7) is an upper bound on the size of S. Thus, the size of the net is bounded by $[O(1/\epsilon)]^d$ as claimed. □

We prove a similar algorithmic bound for the general case of a rank k matrix, with a set that depends on the input matrix.

Theorem 3.2 *Let A be an $n \times n$ matrix with entries in $[-1,1]$ and $(\epsilon/2)$-rank$(A) = d$. There is a finite set $S \subset \Re^n$ s.t.*

$$\forall x \in \Delta, \exists \tilde{x} \in S : \|Ax - A\tilde{x}\|_\infty \leq \epsilon$$

and $|S| = O(1/\epsilon)^d$. Moreover a set of size $O(1/\epsilon)^d$ can be computed in time $O(1/\epsilon)^d poly(n)$ given an $\epsilon/2$-approximating matrix of rank d.

This theorem will be proved using a more general statement, Theorem 1.9. For compact sets A, B, let $N(A,B)$ be the minimum number of translates of B required to cover A. The following bounds are well-known via volume arguments (see e.g., Pisier's book [33]).

Lemma 3.3 *1. For a centrally-symmetric convex body $K \in \Re^d$, $N(K, \epsilon K) \leq \left(1 + \frac{2}{\epsilon}\right)^d$.*

2. For a convex body A and a centrally-symmetric convex body B, both in $\Re^d$,

$$\frac{\text{vol}(A)}{\text{vol}(A \cap B)} \leq N(A,B) \leq 3^d \frac{\text{vol}(A)}{\text{vol}(B)}.$$

3. For compact sets $A, B, C \in \Re^d$, $\text{vol}(A + C) \leq N(A,B)\text{vol}(B + C)$.

Proof. (of Theorem 3.2). We assume that A has rank d (replacing it with its $\epsilon/2$ approximation, if needed, and rescaling to assume that in this approximation too every entry lies in $[-1,1]$). Let K be the intersection of the span of A with $[-1,1]^n$. Thus K is a d-dimensional convex body. Now let L be the intersection of the span of A with $[-\epsilon/2, \epsilon/2]^n$. A cover of K by copies of L would achieve the property we need for S. Note that $(\epsilon/2)K = L$. Therefore,

$$N(K,L) = N(K, \frac{\epsilon}{2}K) \leq \left(1 + \frac{4}{\epsilon}\right)^d$$

using Lemma 3.3, part (1). We can now apply Theorem 1.9 (proved in the next subsection) to complete the proof. □

3.1 Constructing nearly optimal ϵ-nets

In this section we prove Theorem 1.9. We begin with the second part, for which the standard lattice $\mathbb{Z}^n$ can be used. For the near optimal cover size and time complexity of the first part, we will need some additional tools.

Proof. [of Theorem 1.9(2).] Start by finding a parallelopiped P s.t.

$$P \subseteq A \cap B \subseteq \alpha P$$

i.e., a sandwiching of $A \cap B$ by parallelopipeds. It is shown in [8] that $\alpha = 4d$ can be achieved in polynomial time. Now apply an affine transformation that makes P a unit cube. Tile the resulting A with unit cubes so that every point of A is covered. This covering can be carried out by a simple depth-first search enumeration starting with any copy of P that intersects A and only exploring copies of P that intersect $A + P$. We prove two facts about this procedure:

1. The set of cubes intersecting A is connected via paths that use only cubes that intersect $A + P$.

2. The total number of cubes explored in the enumeration is bounded by $2d$ times the number of cubes that intersect $A + P$, which is bounded by $O(d)^d N(A, B)$.

To prove the first part, consider the centers of two cubes P_1 and P_2 that intersect A and the straight line joining these centers. To go from P_1 to P_2 it suffices to traverse a sequence of cubes that intersect this line. Since P_1, P_2 intersect A, they lie in $A + P$, and so do their centers. Thus, by convexity, the line joining the centers lies entirely in $A + P$ and therefore the cubes traversed in the path between P_1 and P_2 all intersect $A + P$.

For the second part, we note that the set of cubes intersecting $A + P$ are all contained in $A + 2P$. Therefore, using Lemma 3.3(2), their number is at most

$$\begin{aligned} \frac{\text{vol}(A+2P)}{\text{vol}(P)} \leq \frac{\text{vol}(A+2A)}{\text{vol}(P)} &\leq 3^d (4d)^d \frac{\text{vol}(A)}{\text{vol}(A \cap B)} \\ &\leq (12d)^d \, N(A, B). \end{aligned}$$

This proves the correctness of the DFS enumeration and bounds its time. Finally, we note that the space required is bounded by $O(d)$ times the logarithm of the diameter of $A + P$ (after making P a cube), which is polynomial in the input specification. □

Our main tool for the near-optimal covering is an extremal lattice construction by Rogers [36] from 1950. The norm defined by a symmetric convex body K is

$$\|x\|_K = \inf\{s \,:\, x \in sK\}.$$

For a lattice $L \in \Re^d$, two standard parameters are its *shortest nonzero vector* and its *covering radius*. We recall their definitions:

$$\lambda_1(K, L) = \min_{x \in L \setminus \{0\}} \|x\|_K$$

and

$$\mu(K, L) = \inf\{s \,:\, L + sK = \Re^d\}.$$

Theorem 3.4 *[36] For any centrally-symmetric convex body $K \subset \Re^d$, there is an d-dimensional lattice L with the following properties:*

1. $\mu(K, L) \leq 1$, *i.e.,* $L + K = \Re^d$.
2. $\lambda_1(K, L) \geq 2/3$.

Using Minkowski's first theorem on lattices, we have $\lambda_1(K, L)^d \leq 2^d \det(L)/\text{vol}(K))$ and therefore $\det(L) \geq 3^{-d}\text{vol}(K)$ for Rogers' extremal lattice. Lattices with covering radius 1 and large determinant were investigated in the 1950's in a series of papers by Rogers [36, 37, 38, 39]. The simplest of these consists of choosing a random lattice with a fixed determinant. Whp, this already satisfies $\frac{\text{vol}(K)}{\det(L)} = 2^{O(d)}$. The highest possible determinant for a covering lattice is an open question. The current best existential result [39] is a covering lattice L for any convex body K with

$$\frac{\text{vol}(K)}{\det(L)} = d^{O(\log \log d)}.$$

Rogers' proof of Theorem 3.4 is algorithmic and can be used to compute a basis of an extremal lattice. We include a proof here for completeness along with an explicit algorithm and complexity analysis.

Theorem 3.5 *[36] For any convex body K in $\Re^d$ given by a membership oracle, there is a deterministic algorithm that outputs a basis for a lattice satisfying the conditions of Rogers' extremal lattice theorem and has complexity $2^{O(d)}$.*

We will use this lattice in our covering algorithm. In addition to this lattice, we need an enumeration algorithm for lattice points in a convex body from [16]. An M-ellipsoid (Milman ellipsoid) of a convex body K is an ellipsoid that achieves $N(K,E)N(E,K) = 2^{O(d)}$. This object is a powerful tool in convex geometry. It can be computed deterministically in time $2^{O(d)}$ [17].

Theorem 3.6 *[16] Given any convex body $K \in \Re^d$, along with an M-ellipsoid E of K and any d-dimensoinal lattice $L \in \Re^d$, the set $K \cap L$ can be computed by a deterministic algorithm in time $G(K,L)2^{O(d)}$ where $G(K,L) = \sup_{x\in\Re^d} |L \cap (x+K)|$.*

Using the above two results, we can prove the first part of Theorem 1.9.

Proof. [of Theorem 1.9(1).] To cover A with B, we compute a Rogers lattice L for B and an M-ellipsoid E of $A+B$. This can be done deterministically in time $2^{O(d)}$ as shown in [17]. Using the M-ellipsoid, we can apply Theorem 3.6 to enumerate the set $L \cap (A+B)$ in time $2^{O(d)}G(A+B,L)$. It remains to bound $G(A+B,L)$, which we do as follows:

$$\begin{aligned} G(A+B,L) &= \sup_{x\in\Re^d} |L \cap (x+A+B)| \\ &\le N(A+B,B)G(B,L) \\ &\le 3^d \frac{\mathrm{vol}(A+B)}{\mathrm{vol}(B)} \\ &\le 3^d N(A,B)\frac{\mathrm{vol}(2B)}{\mathrm{vol}(B)}G(B,L) \\ &\le 6^d N(A,B)G(B,L). \end{aligned}$$

Here we used Lemma 3.3(2),(3), the latter with $C = B$.

We now bound $G(B,L) = \sup_{x\in\Re^d} |L\cap(x+B)|$ using the fact that L is extremal for B. Fix $x \in \Re^d$ and let $S = L \cap (x+B)$. Using the fact that $\lambda_1(B,L) \ge 2/3$, we have that $y + B/3$ and $z + B/3$ for any two $y, z \in L$ are disjoint. Hence,

$$\bigcup_{x\in S} x + \frac{B}{3} \subseteq x + B + \frac{B}{3} = x + \frac{4B}{3}.$$

Therefore

$$|S| \le \frac{\mathrm{vol}(4B/3)}{\mathrm{vol}(B/3)} \le 4^d.$$

As a consequence, we have $G(A+B,L) \le 24^d N(A,B)$. The running time of the enumeration algorithm is thus $2^{O(d)}N(A,B)$ as claimed. □

We now describe the algorithm implicit in Rogers' deterministic construction. The main subroutine, computing $\lambda_1(K,L)$, the length of shortest lattice vector, can be done using the deterministic algorithm of [16] in time $2^{O(d)}$. We will need the following fact from [24] (Lemma 4.7 in that paper).

Lemma 3.7 *[24] For any integer $M > 0$, lattice L and symmetric convex body K,*

$$\mu(K,L) \le \frac{M}{M-1} \sup_{x\in\frac{1}{3}L} \min_{y\in L} \|x-y\|_K.$$

Input: Convex body $K \in \Re^d$.

1. Let L be a lattice of determinant 1 with basis $b = (b_1, \ldots, b_d)$. Compute $\lambda_1(K, L)$.
2. For each choice of $\xi = (\xi_1, \ldots, \xi_d) \in \{-1, 0, 1\}^d$, with at least one $\xi_i \neq 0$,
   (a) Let $a = \sum_{i=1}^d \frac{\xi_i}{3} b_i$.
   (b) Check if the refined lattice $L' = L + a\mathbb{Z}$ with basis $(b_1, \ldots, b_d, a)$ has $\lambda_1(K, L') = \lambda_1(K, L)$.
   (c) If yes, set $L := L'$, compute a new basis; restart Step (2).
3. Scale L to make $\lambda_1(K, L) = 2/3$. Output the corresponding $b_1, \ldots, b_d$.

Figure 1: Rogers' algorithm for the basis of an extremal lattice

Proof. [of Theorem 3.5.] The proof of correctness is directly from [36]. When the algorithm terminates, the lattice L has the property that for any $x \in \frac{1}{3}L \setminus L$, $\lambda_1(K, L + x\mathbb{Z}) < \lambda_1(K, L)$. Thus, if $w \in L + x\mathbb{Z}$ achieves $\lambda_1(K, L + x\mathbb{Z})$, we write $w = y + z$ with $y \in L$, and $z \in \{x, -x\}$, since with $z = 0$ we would get $\|w\|_K \geq \lambda_1(K, L) > \lambda_1(K, L + x\mathbb{Z})$. Thus,

$$\lambda_1(K, L + x\mathbb{Z}) = \|w\|_K = \min_{y \in L}\{\|y - x\|_K, \|y + x\|_K\} = \min_{y \in L} \|y - x\|_K.$$

Using Lemma 3.7 with $M = 3$, we have

$$\mu(K, L) \leq \frac{3}{2} \sup_{x \in \frac{1}{3}L} \min_{y \in L} \|y - x\|_K = \sup_{x \in \frac{L}{3}} \lambda_1(K, L + x\mathbb{Z}) < \lambda_1(K, L).$$

In the last step of the algorithm, when we scale to make $\lambda_1(K, L) = 2/3$, we get $\mu(K, L) < 1$, i.e., $L + K = \Re^d$ as required.

To bound the time complexity of the algorithm, we note that each iteration of Step (2) takes at most $3^d - 1$ shortest vector computations, each of which has complexity $2^{O(d)}$. Moreover, whenever an iteration succeeds and refines the current lattice, its determinant decreases by a factor of $1/3$. Let L_0 be the starting lattice with $|\det(L_0)| = \mathrm{vol}(K)$. After t iterations, the lattice L_t obtained has determinant $|\det(L_t)| = 3^{-t}|\det(L_0)|$. On the other hand, by design, the shortest nonzero vector of any L_t has the same length as the the original lattice L_0. Thus, using Minkowski's first theorem,

$$\lambda_1(K, L_0)^d = \lambda_1(K, L_t)^d \leq 2\frac{|\det(L_t)|}{\mathrm{vol}(K)} \leq \frac{2}{3^t}\frac{|\det(L_0)|}{\mathrm{vol}(K)} \leq \frac{2}{3^t}.$$

Hence, $t \leq d \log_3(2/\lambda_1(K, L_0))$, and the overall bound on the complexity is $2^{O(d)}$ as claimed. □

4 Algorithmic applications

4.1 Approximate Nash equilibria

Let $A, B \in [-1, 1]^{n \times n}$ be the payoff matrices of the row and column players of a 2-player game. A Nash equilibrium is a pair of strategies $x, y \in \Delta_n = \{x \in \Re^n : \|x\|_1 = 1, x \geq 0\}$ s.t.

$$\begin{aligned} x^T A y &\geq e_i^T A y \quad \forall i \in \{1, \ldots, n\} \\ x^T B y &\geq x^T B e_j \quad \forall j \in \{1, \ldots, n\} \end{aligned}$$

Alternatively, a Nash equilibrium is a solution to the following optimization problem:

$$\min \quad \max_i A^i y + \max_j x^T B_j - x^T(A+B)y \tag{8}$$

$$x, y \in \Delta_n. \tag{9}$$

An ϵ-Nash equilibrium is a pair of strategies with the property that each player's payoff cannot improve by more than ϵ by moving to a different strategy, i.e.,

$$\begin{aligned} x^T A y &\geq e_i^T A y - \epsilon \quad \forall i \in \{1, \ldots, n\} \\ x^T B y &\geq x^T B e_j - \epsilon \quad \forall j \in \{1, \ldots, n\} \end{aligned}$$

Lemma 4.1 *Any $x, y \in \Delta_n$ that achieve an objective value of at most ϵ for (8) form an ϵ-Nash equilibrium.*

The algorithm for the case when $A + B$ is PSD is described in Figure 2.

1. Let $d = 9 \log n/(\epsilon/6)^2 - (\epsilon/6)^3)$ and R be an $n \times d$ random matrix with iid entries from $N(0, 1/k)$.
2. Write $A + B = UU^T$ and let $V = UR$.
3. Let S be an $(\epsilon/6\sqrt{d})$-net in the L_∞-norm for $\{V^T y : y \in \Delta_n\}$.
4. For each $\tilde{y} \in S$, solve the following convex program:

$$\begin{aligned} \min \quad & \max_i A^i y + \max_j x^T B_j - x^T V \tilde{y} \\ \text{s.t.} \quad & y \in \Delta_n \\ & \|V^T y - \tilde{y}\|_\infty \leq \frac{\epsilon}{6\sqrt{d}}. \end{aligned}$$

5. Output x, y that achieve an objective value of at most $\frac{\epsilon}{2}$.

Figure 2: Finding ϵ-Nash when $A + B$ is PSD

We are ready to prove Theorem 1.6.

Proof. Using Lemma 2.3, we have that for any $x, y \in \Delta_n$

$$\|x^T(A+B)y - x^T V V^T y\| \leq \frac{\epsilon}{6}.$$

Using Theorem 3.1, we can find a set S in $\Re^d$ of size $O(1/\epsilon)^d$ with the property that for every $y \in \Delta_n$, there is a $\tilde{y} \in S$ s.t.

$$\|V^T y - \tilde{y}\|_\infty \leq \frac{\epsilon}{6\sqrt{d}}.$$

The algorithm enumerates over $\tilde{y} \in S$ and looks for a pair with a corresponding $x, y \in \Delta_n$ such that the objective function (8) is at most $\epsilon/2$, i.e., solves a convex program for each $\tilde{y} \in S$. We note that since $\tilde{y}$ is fixed, the resulting program has a convex objective function subject to linear constraints and thus can be solved in polynomial time.

A solution of small objective value exists: any x, y that form a Nash equilibrium will have a corresponding $\tilde{y}$ with objective value at most $\epsilon/2$. Similarly any solution of objective value at most $\epsilon/2$ for the program used in the algorithm implies a solution of the original quadratic program of value at most ϵ, and thus satisfies the ϵ-Nash conditions, completing the proof. □

Next we give a variant of the algorithm that works for any $A + B$, not necessarily PSD, with complexity depending on the $(\epsilon/2)$-rank of $A + B$. This appears in Figure 3.

1. Let C be a rank d matrix that satisfies $\|A + B - C\|_\infty \leq \epsilon/2$.
2. Let S be an $\epsilon/2$-net in the L_∞-norm for $\{Cx : x \in \Delta_n\}$.
3. For each $\tilde{y} \in S$, solve the following convex program:

$$\begin{aligned} \min \quad & \max_i A^i y + \max_j x^T B_j - x^T \tilde{y} \\ \text{s.t.} \quad & x, y \in \Delta_n \\ & \|Cy - \tilde{y}\|_\infty \leq \frac{\epsilon}{4}. \end{aligned}$$

4. Output x, y that achieve an objective value of at most $\frac{\epsilon}{2}$.

Figure 3: Finding ϵ-Nash when $A + B$ has $(\epsilon/2)$-rank d

The proof of Theorem 1.7 follows from this algorithm and Theorem 3.2.

4.2 Densest subgraph

Our strategy for approximating the densest (bipartite) subgraph is similar to the one described in the previous subsection. We observe that an ϵ-approximate solution to the following optimization problem suffices.

$$\begin{aligned} \max \quad & x^T A y \\ & x, y \in \Delta_n \\ & x_i, y_i \leq \frac{1}{k} \quad \forall i \in \{1, \ldots, n\} \end{aligned}$$

Indeed, in a solution here no x_i, y_j exceeds $1/k$. Since for fixed y the function $x^T A y$ is linear in x and vice versa, it is easy to replace any solution by one of at least the same value in which each x_i and each y_j is either 0 or $1/k$, and this corresponds to the problem of maximizing the quantity density$(A_{S,T})$ over all sets S of k rows and T of k columns.

To solve this optimization problem, we replace the given adjacency matrix A by an approximating matrix $\tilde{A}$ of rank equal to the $(\epsilon/2)$-rank of A. Then we enumerate an $\epsilon/2$-net in the ∞-norm for $\tilde{A}y$ and solve a convex program for each element. This establishes Theorem 1.8.

5 Concluding remarks

We have studied the ϵ-rank of a real matrix A, and exhibited classes of matrices for which it is large (Hadamard matrices, random matrices) and ones for which it is small (positive semi-definite matrices and linear combinations of those). This leads to several approximation algorithms, mainly to problems in which the input can be approximated by a positive semi-definite matrix or a low-rank matrix. Essentially all the discussion here applies to rectangular matrices as well, and we have considered square matrices just for the sake of simplicity.

It will be interesting to investigate how difficult the problem of determining or estimating the ϵ-rank of a given real matrix is. As mentioned in the introduction, a rough approximation is given in [28].

One of our main motivating problems was the complexity of finding ϵ-Nash equilibria. This remains open for general 2-player games. Our methods do not work when the sum of the payoff matrices has high rank with many negative eigenvalues, e.g., for random matrices. However, in the latter case, there is a known simple algorithm even for exact Nash equilibria, based on the existence of *small support* equilibria [10]. It would be interesting to find a common generalization of this class with the classes considered here.

Recall that much of the earlier work regarding several notions dealing with the approximation of matrices by simpler ones revealed interesting applications of lower bounds, that is, of results stating that certain matrices cannot be well approximated by very simple ones with respect to the appropriate notions of simplicity and approximation. This is the case with the results regarding matrix rigidity and the sign-rank of matrices as well as with some of the applications in [2, 11]. It is likely to expect that lower bounds for the ϵ-rank of certain matrices can yield additional interesting applications. As a simple, not too natural and yet interesting example we mention the following.

Claim 5.1 *Let p be a prime, and let X be a sample space supporting $2p$ random variables*

$$X_0, X_1, \ldots, X_{p-1} \text{ and } Y_0, Y_1, \ldots, Y_{p-1}.$$

Assume, further that for each i and j, the covariance $Cov(X_i, Y_j)$ satisfies $|Cov(X_i, y_j) - \chi(i-j)| < 1/2$, where the difference $i - j$ is computed modulo p and $\chi(z)$ is the quadratic character which is 1 iff z is a quadratic residue modulo p (or 0) and -1 otherwise. Then the size of the sample space X satisfies $|X| \geq \Omega(p)$.

This can be proved using the fact that for $\epsilon = 1/2$ the ϵ-rank of the matrix $(\chi(i-j))_{i,j \in Z_p}$ is $\Omega(p)$, as its eigenvalues are very similar to those of a Hadamard matrix. We omit the details but note that it will be interesting to find more natural applications of lower bounds for the ϵ-rank of matrices.

Finally, note that the main component in our algorithmic applications is an efficient procedure for generating, for a given input matrix A, a collection of not too many ϵ-cubes in the ℓ_∞ norm whose union covers the convex hull $\mathrm{conv}(A)$ of the columns of A. This motivates the study of $N_\epsilon(A)$, the minimum possible size of such a collection. Thus, $N_\epsilon(A)$ is the minimum size of a set T such that for all $z \in \mathrm{conv}(A)$ there is a $t \in T$ such that $\|z - t\|_\infty \leq \epsilon$. Call a cover of $\mathrm{conv}(A)$ by ϵ-cubes an ϵ-net for A. The applications motivate the study of $N_\epsilon(A)$ and that of finding efficient means of constructing ϵ-nets for A. The focus in the present paper is to do so using the rank or approximate rank of A, and as we have mentioned one can also use the γ_2 norm of A for this purpose. It turns out that there is another complexity measure of A, its VC-dimension, that can be used here.

For a real matrix A, and a subset C of its columns, we say that A shatters C if there are real numbers $\{t_c,\ c \in C\}$, such that for any $D \subseteq C$ there is a row i with $A(i,c) < t_c$ for all $c \in D$

and $A(i,c) > t_c$ for all $c \in C \setminus D$. The VC-dimension of A, denoted by $VC(A)$ is the maximum cardinality of a shattered set of columns. We can prove that for any real m by n matrix A with entries in $[-1,1]$ and VC-dimension d, $N_\epsilon(A) \leq n^{O(d/\epsilon^2)}$. Moreover, an ϵ-net of this size can be constructed deterministically in time proportional to its size times a polynomial factor. Note that by definition the VC-dimension of any matrix with m rows cannot exceed $\log m$, and thus the corresponding covers are always of size at most quasi-polynomial. We can also show that this quasi-polynomial behavior is tight in general. Therefore, while the covers obtained using this approach do not suffice to reproduce our polynomial time approximation algorithms obtained for matrices of logarithmic rank, they do provide improved bounds in many cases. In particular, this approach supplies a quasi-polynomial time additive approximation scheme for the densest bipartite subgraph problem for any weighted graph with bounded weights. We do not include the proofs of these results here, but plan to investigate the approach as well as some additional aspects of the γ_2 approach in a subsequent work.

Acknowledgements. We are grateful to Ronald de Wolf for several helpful comments regarding the relevant work in communication complexity. We thank Daniel Dadush for bringing Rogers' greedy construction paper to our attention, and for finding an error in a previous version of this paper. Very recently, Dadush found an algorithm to construct Rogers' lattice using only poly(d) space with the same $2^{O(d)}$ time complexity. Thus, while his algorithm does not improve the time complexity of our enumeration algorithm or its applications, it achieves these bounds using polynomial space.

References

[1] S. Aaronson and A. Ambainis, Quantum Search of Spatial Regions, Proc. FOCS 2003, 200-209.

[2] N. Alon, Perturbed identity matrices have high rank: proof and applications, Combinatorics, Probability and Computing 18 (2009), 3-15.

[3] N. Alon, S. Arora, R. Manokaran, D. Moshkovitz and O. Weinstein, On the inapproximability of the densest k-subgraph problem. Manuscript, 2011.

[4] N. Alon, S. Ben-David, N. Cesa-Bianchi and D. Haussler, Scale-sensitive dimensions, uniform convergence, and learnability, J. ACM 44(4) (1997), 615-631.

[5] N. Alon, P. Frankl and V. Rödl, Geometrical realization of set systems and probabilistic communication complexity, Proc. 26^{th} Annual Symp. on Foundations of Computer Science (FOCS), Portland, Oregon, IEEE(1985),277-280.

[6] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski, Random sampling and approximation of MAX-CSPs. J. Comput. Syst. Sci. 67(2), 2003, 212-243. (Also in STOC 2002).

[7] N. Alon, A. Shapira and B. Sudakov, Additive approximation for Edge-deletion problems, Proc. 46^{th} IEEE FOCS, IEEE (2005), 419-428. Also: Annals of Mathematics 170 (2009), 371-411.

[8] D. Applegate and R. Kannan, Sampling and Integration of Near Log-Concave functions, STOC 1991, 156-163.

[9] R. I. Arriaga, S. Vempala, An algorithmic theory of learning: Robust concepts and random projection. Machine Learning 63(2), 2006, 161-182 (also in FOCS 1999).

[10] I. Bárány, S. Vempala and A. Vetta, Nash equilibria in random games, Random Struct. Algorithms 31(4), 2007, 391-405.

[11] B. Barak, Z. Dvir, A. Wigderson and A. Yehudayoff, Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes, Proc. of the 43rd annual STOC, ACM Press (2011), 519–528.

[12] S. Ben-David, N. Eiron, and H. Simon. Limitations of learning via embeddings in Euclidean half spaces. *Journal of Machine Learning Research*, 3:441–461, 2002.

[13] A. Bhaskara, M. Charikar, E. Chlamtac, U. Feige and A. Vijayaraghavan, Detecting high log-densities: an $n^{1/4}$ approximation for densest - subgraph, Proc. STOC 2010, 201–210.

[14] H. Buhrman, N. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*, pages 24–32. IEEE, 2007.

[15] H. Buhrman, R. de Wolf. Communication Complexity Lower Bounds by Polynomials, Proc. of the 16th IEEE Annual Conference on Computational Complexity, 2001, 120–130.

[16] D. Dadush, C. Peikert and S. Vempala, Enumerative Lattice Algorithms in any Norm Via M-ellipsoid Coverings. Proc. FOCS 2011, 580-589.

[17] D. Dadush and S. Vempala, Near-optimal deterministic algorithms for volume computation via M-ellipsoids. Proc. of the National Academy of Sciences, September 2013.

[18] W. Fernandez de la Vega, M. Karpinski, R. Kannan and S. Vempala: Tensor decomposition and approximation schemes for constraint satisfaction problems, Proc. of STOC 2005, 747-754.

[19] A. Frieze and R. Kanan, Quick approximation to matrices and applications, Combinatorica, 19 (1999), pp. 175–220.

[20] J. Forster, A linear lower bound on the unbounded error probabilistic communication complexity, in SCT: Annual Conference on Structure in Complexity Theory, 2001. Also: JCSS, 65 (2002), 612–625.

[21] J. Forster, N. Schmitt, and H. U. Simon. Estimating the optimal margins of embeddings in euclidean half spaces. Technical Report NC2-TR-2001-094, NeuroCOLT2 Technical Report Series, 2000.

[22] Joel N. Franklin, *Matrix Theory*, Dover Publications, 1993.

[23] M. Grötschel, L. Lovász and A. Schrijver, Geometric Algorithms and Combinatorial Optimization. Springer, 1988.

[24] V. Guruswami, D. Micciancio and O. Regev, The complexity of the covering radius problem. Computational Complexity, 14(2), pp. 90-121, 2005.

[25] W. B. Johnson and J. Lindenstrauss, Extensions of Lipschitz maps into a Hilbert space, Contemp Math 26, 1984, 189–206.

[26] R. Kannan and T, Theobald. Games of fixed rank: A hierarchy of bimatrix games, Econom. Theory 42:157-173, 2010 (also in SODA 2007).

[27] M. Krause, Geometric arguments yield better bounds for threshold circuits and distributed computing, Theoretical Computer Science, 156 (1996), 99–117.

[28] T. Lee, A. Shraibman. An approximation algorithm for approximation rank, Proc. of the 24th IEEE Conference on Computational Complexity, 2009, 351–357.

[29] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. Combinatorica, 27 (2007), 439–463.

[30] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures and Algorithms*, 34:368–394, 2009.

[31] R. J. Lipton, E. Markakis, A. Mehta: Playing large games using simple strategies. ACM Conference on Electronic Commerce 2003: 36-41.

[32] R. Mathias. The Hadamard operator norm of a circulant and applications. SIAM journal on matrix analysis and applications, 14(4):1152–1167, 1993.

[33] G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry, Second Ed.*, Cambridge University Press, 1999.

[34] A. A. Razborov, Quantum communication complexity of symmetric predicates, arXiv:quant-ph/0204025, 2002.

[35] A. A. Razborov and A. A. Sherstov, The Sign-Rank of AC0, SIAM J. Comput. 39(5), 1833-1855 (2010).

[36] C. A. Rogers, A note on coverings and packings. J. London Math. Soc. 25, (1950). 327–331.

[37] C. A. Rogers, Lattice coverings of space with convex bodies. J. London Math. Soc. 33, 1958, 208–212.

[38] C. A. Rogers, Lattice covering of space: The Minkowski-Hlawka theorem. Proc. London Math. Soc. (3) 8, 1958, 447–465.

[39] C. A. Rogers, Lattice coverings of space. Mathematika 6, 1959, 33–39.

[40] A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[41] J. H. Spencer, Six standard deviations suffice, Trans. Amer. Math. Soc. 289 (1985), 679–706.

[42] L. G. Valiant, Graph-theoretic arguments in low-level complexity, Lecture notes in Computer Science 53 (1977), Springer, 162-176.

[43] S. Vempala, *The Random Projection Method*, AMS, 2004.

[44] H. E. Warren, Lower Bounds for approximation by nonlinear manifolds, Trans. Amer. Math. Soc. 133 (1968), 167-178.