Electronic Colloquium on Computational Complexity, Report No. 169 (2012)

The Approximate Rank of a Matrix and its Algorithmic Applications

Noga Alon * Santosh Vempala †

Abstract

We introduce and 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 combinatorial optimization, game theory, computational geometry and learning theory. Here we give bounds on the ϵ-rank and use them to derive (a) polynomial-time approximation schemes for Nash equilibria for substantially larger classes of 2-player games than previously known and (b) an additive PTAS for the densest subgraph problem on inputs having small ϵ-rank. We use combinatorial, geometric and spectral techniques; our main new tool is an algorithm for efficiently covering a convex body with translates of another convex body.

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 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 [16]. Similar approximations of matrices can be given using variants of the regularity lemma of Szemerédi. These provide a more powerful approximation at the expense

*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.

†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.

ISSN 1433-8092

of increasing the complexity of the approximating matrix, and supply approximation algorithms for additional problems see [5], [15], [6].

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, we consider a stronger notion of approximation of a matrix, an approximation in the infinity norm. Our notion for the simpler approximating matrix is 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 [32] 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, $signrank(A)$ 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., [4], [17] 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 [26]). It is clear that for $-1, 1$ matrices or matrices with entries of absolute value exceeding ϵ, ϵ-rank$(A) \geq signrank(A)$, and simple examples show that in many cases the ϵ-rank is far larger than the sign-rank.

For the special case of the n by n identity matrix the ϵ-rank has been studied and provided several applications. In [1] 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 [1] 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 [12] 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., [3] and the references therein).

1.2 Results

We begin with bounds on the ϵ-rank. A well known result of Forster [17] asserts that the sign-rank of any n by n Hadamard matrix is at least $\Omega(\sqrt{n})$. This clearly implies the same lower bound for the ϵ-rank of any such matrix for any $\epsilon < 1$. The following gives a much stronger estimate.

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

A similar lower bound holds for random binary matrices, and even for sparse random matrices. These follow from known bounds on the sign rank estbalished by applying, as in [4], Warren's theorem from real algebraic geometry [34].

Theorem 1.3 *For any $\epsilon < 1/2$ and almost all d-regular graphs G on n vertices, ϵ-rank is $\Omega(d)$ for the adjacency matrix of G.*

The lower bound here is tight up to a $\log n$ factor: for any fixed ϵ bounded away from zero the ϵ-rank of almost all d-regular graphs on n vertices is $O(d \log n)$. By "almost all" we mean here and in the theorem above that the fraction of d-regular graphs on n vertices for which the statement holds tends to 1 as n tends to infinity.

The ϵ-rank of any positive semidefinite matrix can be bounded from above as stated in the next theorem.

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.

The results above provide several algorithmic applications. Our first application is finding approximate Nash equilibria in 2-player games. Lipton et al. [23] 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 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 [22], 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 Las Vegas randomized algorithm using $poly(n)$ space and expected time*

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

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

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$(A+B) = d$. Then, an ϵ-Nash equilibrium of the game can be computed by a Las Vegas randomized algorithm using $poly(n)$ space and expected time*

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

Our second application is to finding an approximately densest subgraph (submatrix), 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 [2]. Here we observe that we can get efficiently a good additive approximation for the special case that the input matrix has a small ϵ-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 Las Vegas randomized 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 expected 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 its space complexity is $poly(n)$.

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.

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.

Theorem 1.9 *For any two centrally symmetric convex bodies A, 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 Las Vegas randomized algorithm with expected running time $N(A, B)2^{O(d)}$ and using space poly(n).*

In essence, this theorem allows us to find and use covers of size $(1/\epsilon)^{O(d)}$ rather than $(d/\epsilon)^{O(d)}$ that can constructed more easily. Although we do not prove it here, we expect that the theorem can be extended to asymmetric convex bodies.

1.3 Organization

The rest of the paper is organised 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 and open problems.

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*

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

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., [18], 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.

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 [34]) *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. Fix an $\epsilon < 1/2$, and let us estimate the number of d-regular graphs on n vertices whose adjacency matrix A has ϵ-rank at most 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 $\|A - BC\|_\infty \leq \epsilon$. 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. Moreover, if we subtract $1/2$ from any such polynomial corresponding to an entry of A, then the resulting value must be positive if the entry is 1, and negative if it is 0. This means that the number of such adjacency matrices that have ϵ-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

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

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. □

Remark: As mentioned in the paragraph following Theorem 1.3 the lower bound stated in the theorem is tight up to a $\log n$ factor. Indeed, it is known that for most d-regular graphs on n vertices the most negative eigenvalue of the adjacency matrix is in absolute value smaller than $2\sqrt{d}$ (see [19]). Therefore, for almost all adjacency matrices A of d-regular graphs on n vertices the matrix $I + \frac{A}{2\sqrt{d}}$ is positive semi-definite. The result now follows from Corollary 1.5 (with $\epsilon' = \frac{\epsilon}{2\sqrt{d}}$ and $m = 2$) as $\frac{A}{2\sqrt{d}}$ is the difference between the two positive semi-definite matrices $I + \frac{A}{2\sqrt{d}}$ and I.

2.2 Upper bounds

We will use the Johnson-Lindenstrauss Lemma [21]. The following version is from [8, 33].

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 $x, y \in \Re^n$,*

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

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 $\tilde{A} = BRR^TB^T$.

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 simple example appears in the remark at the end of the previous subsection. A more complicated example 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. Since $A + A^T = 2I$ the ϵ-rank of A is at least $\Omega(\frac{\log n}{\epsilon^2 \log(1/\epsilon)})$, by the known lower bound of [1] for the ϵ-rank of the identity I. We do not have a matching upper bound but can prove that for any ϵ the ϵ-rank of A is at most $O(n^{2/3} \log n/\epsilon^2)$. This is done by defining a circulant $2n$ by $2n$ matrix B with $-1, 1$ entries which contains A as its right upper n by n block. As B is circulant its eigenvalues λ_i are character sums that can be computed explicitly and the eigenvectors v_i are the characters of the cyclic group. By subtracting from B the rank one matrices $\lambda_i v_i v_i^T$ corresponding to the $n^{2/3}$ most negative eigenvalues and by adding to the matrix obtained an appropriate product of the identity we get a positive semi-definite matrix with diagonal entries that are at most $O(n^{1/3})$, and can then apply Johnson Lindenstrauss to get the claimed upper bound. We omit the details.

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 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 $O(1/\epsilon)^d poly(n)$.

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 next two bounds are well-known via volume arguments (see e.g., Pisier's book [25]).

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

Lemma 3.4 *For centrally symmetric convex bodies, A, B,*

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

We prove Theorem 1.9 via an algorithm in the next section. Here, to warm up, we give a simpler proof of a net of size $N(A,B)O(d)^d$.

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 [7] 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. Then, using Lemma 3.4, the number of translates used is at most

$$\begin{aligned} \frac{\text{vol}(A+P)}{\text{vol}(P)} \leq \frac{\text{vol}(A+A)}{\text{vol}(P)} &\leq 2^d (4d)^d \frac{\text{vol}(A)}{\text{vol}(A \cap B)} \\ &\leq (8d)^d N(A,B). \end{aligned}$$

Proof. (of Theorem 3.2). We assume that A has rank d (we replace A with its $\epsilon/2$ approximation). 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. 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 by algorithmically finding a nearly optimal ϵ-net. More generally, given two centrally-symmetric convex bodies $A, B \in \Re^d$, we give a Las Vegas algorithm that finds a set of points T such that:

1. $|T| \leq N(A,B)2^{O(d)}$.

2. $A \subseteq B + T$.

In other words, for any point x in A, there is a point y in T such that $\|x - y\|_B \leq 1$.

The algorithm is based on classical proofs of the existence of nearly optimal lattice coverings. For a convex body K in $\Re^d$, a *covering lattice* L has the property that $K + L$ covers $\Re^d$, i.e., lattice translates of K cover all of space. Our goal is to find a covering lattice of large determinant, comparable to the volume of K. The next lemma shows how such a lattice allows us to find a near-optimal ϵ-net.

Lemma 3.5 *Let A, B be centrally-symmetric convex bodies in $\Re^d$ and let L be a covering lattice with respect to $A \cap B$. Then there is a set $S \subset L$ of size $2^d N(A,B)\mathrm{vol}(A \cap B)/\det(L)$ with the property that for any $x \in A$, there is a point $y \in S$ s.t. $\|x - y\|_B \leq 1$ and this set can be enumerated in time $2^{O(d)} N(A,B)\mathrm{vol}(A \cap B)/\det(L)$.*

Proof. Let

$$S = \{x \in L : (x + (A \cap B) \cap A \neq \phi\}.$$

We can enumerate S by a depth-first traversal over translates of the fundamental parallelopiped of L. By the definitions of S and L, we have $A \subseteq S + B$, and the running time is at most $|S|$ times a polynomial in d (to check for intersections). Next we bound the size of S, using Lemma 3.4.

$$\begin{aligned} |S| &\leq \frac{\mathrm{vol}(A + (A \cap B))}{\det(L)} \\ &\leq \frac{\mathrm{vol}(2A)}{\mathrm{vol}(A \cap B)} \frac{\mathrm{vol}(A \cap B)}{\det(L)} \\ &\leq 2^d \frac{\mathrm{vol}(A)}{\mathrm{vol}(A \cap B)} \frac{\mathrm{vol}(A \cap B)}{\det(L)} \\ &\leq 2^d N(A,B) \frac{\mathrm{vol}(A \cap B)}{\det(L)}. \end{aligned}$$

□

We now turn to the question of algorithmically finding covering lattices of large determinant. The existence of such lattices was investigated in the 1950's in a series of papers by Rogers [27, 28, 29, 30]. The simplest of these consists of choosing a random lattice with a fixed determinant. This already works with high probability and gives

$$\frac{\mathrm{vol}(K)}{\det(L)} = 2^{O(d)}.$$

The algorithm below describes this in detail along with a verification step to make it a Las Vegas algorithm, i.e., the output is verified deterministically and the expected running time of the algorithm is $2^{O(d)}$.

The current best existential result [30] is a covering lattice L for any convex body K with

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

i.e., the "density" of the resulting covering of space by translates of K is bounded by $d^{O(\log\log d)}$. Such a lattice would directly give an approximation to the volume of K with the same factor. We note, however, that the results of Bárány and Füredi [9, 10] imply that for any deterministic algorithm running in $2^{O(d)}$ time the approximation factor is $2^{\Omega(d)}$ (in the general oracle model for convex bodies).

Rogers also gives another proof based on a deterministic greedy construction [27] and this too can be made algorithmic in a straightforward manner. However this variant uses exponential *space* unlike our randomized algorithm, which needs only polynomial space for enumerating the cover[1].

Input: Convex bodies $A, B \in \Re^d$.

1. Let $K = A \cap B$ and compute V such that $\text{vol}(K) \leq V \leq 2\text{vol}(K)$ using a volume estimation algorithm.
2. Let L be a random lattice of determinant V/c_1^d.
3. Compute the set of points $S = \{x \in L : (x + K) \cap A \neq \phi\}$, via a depth-first traversal, starting with any $x \in L$ for which $x + K$ intersects A; if the size of this set exceeds c_2^d at any point, stop.
4. Output the set of translates of B that cover A.

Figure 1: Las Vegas algorithm for an ϵ-net

To estimate the volume, we can use a randomized volume algorithm [14, 24]. To pick a random lattice of fixed determinant, we can use the algorithm of Goldstein and Mayer [20], which shows how to do this with a discrete algorithm and approximate the continuous probability distribution with respect to the integral of any bounded function.

For a convex body K and lattice K, let $\delta(K, L)$ be the density of points not covered by the union of sets $x + K$ for $x \in L$.

Lemma 3.6 *[29] Let K be a convex body of volume $V \leq \frac{1}{4}d\log(27/16) - 3\log d$ and L be a random lattice of determinant 1. Then,*

$$\left|\mathrm{E}(\delta(K, L)) - e^{-V}\right| \leq cd^3 \left(\frac{16}{27}\right)^{\frac{d}{4}}$$

where the expectation is over the choice of the random lattice and c is an absolute constant.

The lemma is based on a fundamental theorem of Siegel about random lattices [31].

Proof. (of Theorem 1.9.) Let $K = A \cap B$. By Lemma 3.6, with probability at least 3/4, a random lattice L_1 with

$$\det(L_1) = \frac{4\text{vol}(K)}{d\log(27/16) - 12\log d}$$

has the property that the set $K + L_1$ leaves uncovered a subset of density at most $p = (4c+1)d^3\left(\frac{16}{27}\right)^{\frac{d}{4}}$. We now argue (following Rogers) that $2K + L_1$ covers $\Re^d$, i.e., $K + \frac{1}{2}L_1$ covers $\Re^d$.

With probability at least 3/4, the density of $K + L_1$ is at least $1 - p$. So, with this probability, for every $x \in \Re^d$, the density of $x - K + L_1$ is at least $1 - p$. Fix a point x. Since $p < 1/2$, these set systems intersect and we have a point $z \in K + L_1 \cap (x - K + L_1)$, i.e., there are points $a, b \in K$ and $a', b' \in L_1$ such that

$$z = a + a' = x - b + b'$$

[1]We thank Daniel Dadush for bringing Rogers' greedy construction paper to our attention.

implying that $x = a+b+a'-b' \in 2K+L_1$. Since this holds for every x, it follows that L_1 is a covering lattice for $2K$, and therefore $L = L_1/2$ is a covering lattice for K with determinant at least $\frac{2\text{vol}(K)}{d2^d}$.

We can now apply Lemma 3.5 with this lattice L to get a cover of A with translates of B of size $N(A,B)2^{O(d)}$. □

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.

$$\begin{aligned} x^T Ay &\geq e_i^T Ay \quad \forall i \in \{1,\ldots,n\} \\ x^T By &\geq x^T Be_j \quad \forall j \in \{1,\ldots,n\} \end{aligned}$$

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.,

$$\begin{aligned} x^T Ay &\geq e_i^T Ay - \epsilon \quad \forall i \in \{1,\ldots,n\} \\ x^T By &\geq x^T Be_j - \epsilon \quad \forall j \in \{1,\ldots,n\} \end{aligned}$$

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.

We are ready to prove Theorem 1.6.

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

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

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.

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

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

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

4.2 Densest subgraph

Our strategy for approximating the densest 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 introduced and studied the notion of 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.

It will be interesting to investigate how difficult the problem of determining or estimating the ϵ-rank of a given real matrix is. Determining the ϵ-rank or giving asymptotically tight bounds for it seems to be interesting even for some explicit families of matrices. One intriguing family is the one mentioned in the end of subsection 2.2, whose members are essentially the adjacency matrices of the *half-graphs*. We currently know that for fixed ϵ its ϵ-rank lies between roughly $\log n$ and $n^{2/3} \log n$. Determining the precise dependence on n might reveal new embedding techniques.

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 [11]. It would be interesting to find a common generalization of this class with the classes considered here.

Finally we 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 [1], [12]. 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.

References

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

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

[3] 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.

[4] 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.

[5] 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).

[6] 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.

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

[8] 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).

[9] I. Bárány and Z. Füredi, Computing the volume is difficult, Discrete and Computational Geometry 2, 1987, 319-326.

[10] I. Bárány and Z. Füredi, Approximation of the sphere by polytopes having few vertices, Proc. of the AMS 102, 1988, 651-659.

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

[12] 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.

[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] M. Dyer, A. Frieze and R. Kannan, A random polynomial-time algorithm for approximating the volume of convex bodies, Journal of the ACM 38 (1), 1991, 1–17.

[15] 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.

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

[17] 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.

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

[19] J. Friedman, A proof of Alon's second eigenvalue conjecture and related problems. Mem. Amer. Math. Soc. 195 (2008), no. 910, viii+100 pp.

[20] D. Goldstein and A. Mayer, On the equidistribution of Hecke points. Forum Mathematicum, 15:2, pp. 165–189, 2003.

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

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

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

[24] L. Lovász and S. Vempala: Simulated annealing in convex bodies and an $O*(n^4)$ volume algorithm, J. Comput. Syst. Sci. 72(2), 2006, 392-417.

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

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

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

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

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

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

[31] C. L. Siegel, A mean value theorem in the geometry of numbers, Ann. of Math., 45(2), 1945, 340-347.

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

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

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

ECCC ISSN 1433-8092

http://eccc.hpi-web.de