

The Cover Number of a Matrix and its Algorithmic Applications

Noga Alon^{1,2}, Troy Lee³, and Adi Shraibman⁴

- 1 Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel
nogaa@tau.ac.il
- 2 Institute for Advanced Study, Princeton, New Jersey, 08540, USA
- 3 School of Physics and Mathematical Sciences, Nanyang Technological University and Centre for Quantum Technologies, Singapore
troyjlee@gmail.com
- 4 School of Computer Science. Academic College of Tel-Aviv Yaffo, Israel
adish@mta.ac.il

Abstract

Given a matrix A , we study how many ϵ -cubes are required to cover the convex hull of the columns of A . We show bounds on this cover number in terms of VC dimension and the γ_2 norm and give algorithms for enumerating elements of a cover. This leads to algorithms for computing approximate Nash equilibria that unify and extend several previous results in the literature. Moreover, our approximation algorithms can be applied quite generally to a family of quadratic optimization problems that also includes finding the densest k -by- k combinatorial rectangle of a matrix. In particular, for this problem we give the first quasi-polynomial time additive approximation algorithm that works for any matrix $A \in [0, 1]^{m \times n}$.

1998 ACM Subject Classification G.1.2 Approximation

Keywords and phrases Approximation algorithms, Approximate Nash equilibria, Cover number, VC dimension

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2014.34

1 Introduction

Consider a quadratic optimization problem where we wish to maximize $p^T A q$ over probability distributions p, q , subject to linear constraints. Examples of problems of this type include Nash equilibrium and the densest combinatorial rectangle problem. A general scheme for finding an approximately optimal solution is based on the following notion of an ϵ -net for an m -by- n matrix A . Denote by $\text{conv}(A)$ the convex hull of the columns of A . We call a set of vectors $S \subseteq \mathbb{R}^m$ an ϵ -net for A if for all $v \in \text{conv}(A)$ there is a vector $u \in S$ such that $\|v - u\|_\infty \leq \epsilon$. An efficient means to enumerate elements of an ϵ -net S for A gives an efficient means for finding a near optimal solution to the original quadratic optimization problem: for each $u \in S$ solve the linear program to maximize $p^T u$ over probability distributions p, q , subject to the original linear constraints on p and q and the additional constraint $\|u - Aq\|_\infty \leq \epsilon$. The largest such value will be within 2ϵ of the optimal and the running time of this approximation algorithm will be a polynomial factor times the time required to enumerate an ϵ -net for A . This approximation algorithm motivates the study of ϵ -nets and efficient algorithms for enumerating them.

Say that $A \in [-1, 1]^{m \times n}$. Denote by $N_\epsilon(A)$ the minimal size of an ϵ -net for A , which we will also informally refer to as the *cover number* of A . An obvious upper bound on $N_\epsilon(A)$ is



© Noga Alon, Troy Lee, and Adi Shraibman;

licensed under Creative Commons License CC-BY

17th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'14) / 18th Int'l Workshop on Randomization and Computation (RANDOM'14).

Editors: Klaus Jansen, José Rolim, Nikhil Devanur, and Cristopher Moore; pp. 34–47



Leibniz International Proceedings in Informatics

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

$(1/\epsilon)^m$. This naive bound can be improved by realizing that the convex hull of the columns of A actually lives in a space of dimension $\text{rank}(A)$, which allows an improvement to $N_\epsilon(A) = [O(1/\epsilon)]^{\text{rank}(A)}$. Recently, [7] made this bound algorithmic, showing that an ϵ -net for A can be enumerated by a randomized Las Vegas algorithm in time $[O(1/\epsilon)]^{\text{rank}(A)} \text{poly}(mn)$. Following the above approximation paradigm, this led to polynomial time additive approximation schemes for two-player Nash Equilibrium when the sum of the payoff matrices has logarithmic rank, improving work of Kannan and Theobald [17] who showed the same when the sum of the payoff matrices has constant rank. The [7] bound on the cover number combined with the above approximation paradigm also gave an efficient approximation algorithm for finding the densest k -by- k combinatorial rectangle provided the associated matrix has rank at most logarithmic in the dimension.

In this paper, we continue the study of $N_\epsilon(A)$ and its relation to other complexity measures of A , like VC dimension, γ_2 norm, and communication complexity (these measures are formally defined in the sequel). In particular, we show that $N_\epsilon(A) = n^{O(\text{VC}(A)/\epsilon^2)}$ and that an ϵ -net can be enumerated deterministically in the same time. As $\text{VC}(A) \leq \log(m)$ for any matrix with m rows, this recovers the quasi-polynomial time approximation for Nash equilibrium shown by Lipton et al. [19], and also gives a quasi-polynomial time additive approximation algorithm for the densest k -by- k combinatorial rectangle problem.

By the triangle inequality it is easy to see that an $\epsilon/2$ -net for a matrix B satisfying $\|A - B\|_\infty \leq \epsilon/2$ gives an ϵ -net for A (here $\|X\|_\infty$ denotes the largest absolute value of an entry of X). Thus to construct ϵ -nets for A , it suffices to look for “simpler” matrices that are entrywise close to A . Define the ϵ -approximate rank of A as $\text{rank}_\epsilon(A) = \min_{B: \|A-B\|_\infty \leq \epsilon} \text{rank}(B)$. Existentially $N_\epsilon(A) \leq [O(1/\epsilon)]^{\text{rank}_{\epsilon/2}(A)}$, but for the algorithm of [7] to enumerate such a cover, it explicitly needs to find an approximating matrix B whose rank is equal to $\text{rank}_{\epsilon/2}(A)$. We currently do not know an algorithm to do this working in time $[O(1/\epsilon)]^{\text{rank}_{\epsilon/2}(A)}$, or even $(n/\epsilon)^{\text{rank}_{\epsilon/2}(A)}$ for that matter.

For a sign matrix A and any $\epsilon < 1$, it is easy and known that $\text{VC}(A) \leq \text{rank}_\epsilon(A)$. Thus the results in this paper give a way to enumerate an ϵ -net for a sign matrix A in deterministic time $n^{O(\text{rank}_\epsilon(A)/\epsilon^2)}$. We present a similar result in terms of the γ_2 norm. The γ_2 norm, also known as the Hadamard product operator norm, has recently seen many applications in communication complexity and learning theory [20, 24, 25, 23]. Part of its usefulness is that the approximate γ_2 norm serves as a proxy for the approximate rank and can be computed efficiently via semidefinite programming. Based on the γ_2 norm, we show a Las Vegas randomized algorithm for enumerating an ϵ -net for A in time $(1/\epsilon)^{r \log(r) \log(mn)/\epsilon^2}$ where $r = \text{rank}_{\epsilon/4}(A)$. While being a slightly weaker result than the one using the VC dimension, this has the benefit of having a simple self-contained proof.

2 Algorithmic Applications

We first show how efficient constructions of an ϵ -net for A lead to approximation algorithms for Nash equilibria and finding a densest combinatorial rectangle. This idea was already presented in [7] for ϵ -nets constructed from low rank decompositions of A . We present the proof again here in a slightly more general form for completeness.

2.1 Approximate Nash Equilibria

Let $A, B \in [-1, 1]^{m \times n}$ be the payoff matrices of the row and column players of a 2-player game. In other words, $A(i, j)$ is the payoff to Alice when she plays strategy i and Bob plays strategy j , and similarly $B(i, j)$ is the payoff to Bob when Alice plays strategy i and he plays

1. Create an $\epsilon/2$ -net S for $A + B$. For each $u \in S$, solve the following linear program:

$$\begin{aligned} & \text{maximize} && p^T u - \max_i e_i^T A q - \max_j p^T B e_j \\ & \text{subject to} && \|(A + B)q - u\|_\infty \leq \epsilon/2. \end{aligned}$$
2. Output p, q that achieve an objective value at least $-\epsilon$.

■ **Figure 1** Finding ϵ -Nash equilibrium for payoff matrices A, B given an $\epsilon/2$ -net for $A + B$.

strategy j . Let $\Delta_n = \{p \in \mathbb{R}^n : \|p\|_1 = 1, p \geq 0\}$ be the set of n -dimensional probability vectors. A Nash equilibrium is a pair of strategies (p, q) for $p \in \Delta_m, q \in \Delta_n$ satisfying

$$\begin{aligned} p^T A q &\geq e_i^T A q \quad \forall i \in \{1, \dots, m\} \\ p^T B q &\geq p^T B e_j \quad \forall j \in \{1, \dots, n\} \end{aligned}$$

Here e_i denotes the vector with a 1 in the i th position and zeros elsewhere.

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

$$\max_{p \in \Delta_m, q \in \Delta_n} p^T (A + B)q - \max_i e_i^T A q - \max_j p^T B e_j \quad (1)$$

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.,

$$\begin{aligned} p^T A q &\geq e_i^T A q - \epsilon \quad \forall i \in \{1, \dots, m\} \\ p^T B q &\geq p^T B e_j - \epsilon \quad \forall j \in \{1, \dots, n\} \end{aligned}$$

► **Lemma 1.** *Any $p \in \Delta_m, q \in \Delta_n$ that achieve an objective value at least $-\epsilon$ for (1) form an ϵ -Nash equilibrium.*

We now show how to find an ϵ -Nash equilibrium for a game with payoff matrices A, B given an ϵ -net for $A + B$. The algorithm is described in Figure 1.

► **Theorem 2.** *Let $A, B \in [-1, +1]^{m \times n}$. Suppose there is a deterministic (or Las Vegas randomized) algorithm running in time t for enumerating an $\epsilon/2$ -net for $A + B$. Then an ϵ -Nash equilibrium for the game with payoff matrices A, B can be found by a deterministic (or Las Vegas randomized) algorithm in time $t \cdot \text{poly}(mn)$.*

Proof. The algorithm to find an ϵ -Nash equilibrium enumerates all vectors u in an $(\epsilon/2)$ -net for $A + B$. For each of these vectors the algorithm solves the following program:

$$\begin{aligned} & \text{maximize} && p^T u - \max_i e_i^T A q - \max_j p^T B e_j \\ & \text{subject to} && \|(A + B)q - u\|_\infty \leq \epsilon/2. \end{aligned}$$

Let $p_* \in \Delta_m, q_* \in \Delta_n$ be a Nash equilibrium, and so $0 = p_*^T (A + B)q_* - \max_i e_i^T A q_* - \max_j p_*^T B e_j$. For u in the $\epsilon/2$ -net satisfying $\|(A + B)q_* - u\|_\infty \leq \epsilon/2$ we then have

$$\begin{aligned} \max_{p, q} p^T u - \max_i e_i^T A q - \max_j p^T B e_j &\geq p_*^T u - \max_i e_i^T A q_* - \max_j p_*^T B e_j \\ p_*^T (A + B)q_* - \max_i e_i^T A q_* - \max_j p_*^T B e_j - \epsilon/2 &\geq -\epsilon/2 \end{aligned}$$

Thus the algorithm finds a pair p, q such that the optimal value is at least $-\epsilon/2$. By going via the $\epsilon/2$ -net again, and using Lemma 1 we see that p, q are an ϵ -Nash equilibrium. ◀

2.2 Densest Combinatorial Rectangle

For a matrix $A \in [0, 1]^{m \times n}$ 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

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

that is, the average of the entries in $A_{S,T}$.

► **Definition 3** (Densest k -by- k combinatorial rectangle). Let $A \in [0, 1]^{m \times n}$. The densest k -by- k combinatorial rectangle problem is to find sets S_*, T_* , each of size k , such that

$$\text{density}(A_{S_*, T_*}) = \max_{S, T: |S|=|T|=k} \text{density}(A_{S,T}).$$

Sets S, T which achieve the maximum up to an additive ϵ we call an ϵ -approximate densest k -by- k combinatorial rectangle.

A closely related problem is the densest k -subgraph problem. Here the goal is to find a set S_* realizing $\max_{S: |S|=k} \text{density}(A_{S,S})$. This problem is NP-hard and Khot has also shown that it does not have a PTAS unless $\text{NP} \subseteq \text{BPTIME}(2^{n^\epsilon})$ [16]. The best known polynomial time algorithm guarantees an optimal solution within a multiplicative factor of $n^{1/4+\epsilon}$ of the optimal density [9]. For dense graphs (with at least an ϵ -fraction of edges), Arora, Karger, and Karpinski give a polynomial time approximation scheme for $k = \Omega(n)$ [4]. This also follows from the results in [2].

It is straightforward to see that the density of the densest k -by- k combinatorial rectangle and the densest $2k$ -subgraph differ by at most a factor of 2. Thus hardness results for densest k -subgraph carry over to densest k -by- k combinatorial rectangle. We note that as shown in [1], assuming that the *Hidden Clique Problem*, that is, the problem of finding a planted clique of size $n^{1/3}$ in the random graph $\mathcal{G}(n, 1/2)$ is hard, then so is approximating the Densest k -Subgraph to within any constant factor or to within any additive error bounded away from 1, for a subgraph of size $k = n^{1-\epsilon}$ for any $2/3 \geq \epsilon > 0$ in an n vertex graph. Moreover, any algorithm that solves the above approximation problem in $n^{o(\log n)}$ time would yield an algorithm with essentially the same running time for the hidden clique problem, and there is no such known result despite the extensive study of the hidden clique problem (see [14, 5, 13, 12]).

Our strategy for approximating the densest k -by- k combinatorial rectangle follows the paradigm outlined in the introduction. First, note that the problem can be equivalently reformulated as follows.

► **Lemma 4.** Let $A \in [0, 1]^{m \times n}$. Then

$$\begin{aligned} \max_{S, T: |S|=|T|=k} \text{density}(A_{S,T}) &= \max_{\substack{x \in \Delta_m, y \in \Delta_n \\ \|x\|_\infty \leq 1/k, \|y\|_\infty \leq 1/k}} x^T A y \end{aligned}$$

Proof. For fixed y , the function $x^T A y$ is linear in x and vice versa, thus 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 $1/k$. This corresponds to the problem of maximizing the quantity $\text{density}(A_{S,T})$ over all sets S of k rows and T of k columns. ◀

By Lemma 4 it can be seen that the densest k -by- k combinatorial rectangle fits into the general class of problems of our approximation algorithm. Thus, as described above, by iterating over elements of the cover and sequentially solving the associated linear programs, we obtain the following theorem.

► **Theorem 5.** *Let $A \in [0, 1]^{m \times n}$. Suppose that there is a deterministic (or Las Vegas randomized) algorithm to enumerate an $\epsilon/2$ -net for A in time t . Then a solution to the k -by- k densest combinatorial rectangle within an additive ϵ of the optimal can be found in time $t \cdot \text{poly}(mn)$.*

3 γ_2 Bounds on the Cover Number

Results of [7] show that an ϵ -net for A can be constructed by a randomized algorithm in time $(1/\epsilon)^{O(d)}$ given a matrix B of rank d that is an $\epsilon/2$ -approximation of A . A drawback to this result is that it requires finding such a low rank approximation B .

We address this issue here by considering the γ_2 norm. As we describe next, the (approximate) γ_2 norm characterizes the approximate rank up to a logarithmic factor in the size of the matrix and small change in the error parameter [23]. Moreover, the approximate γ_2 norm can be computed in polynomial time via semidefinite programming, and thus also gives a polynomial time randomized Las Vegas algorithm to find an approximation B whose rank is within a logarithmic factor of the optimal. Combining this with the result of [7] gives a randomized Las Vegas algorithm for constructing an ϵ -net for an m -by- n matrix A of approximate rank d working in time $(1/\epsilon)^{O(d \log(mn))}$. We also give a simple and direct proof of a weaker result solely in terms of the γ_2 norm. Namely, if $\gamma = \min_{B: \|A-B\|_\infty \leq \epsilon/4} \gamma_2(B)$ then there is a randomized algorithm constructing an ϵ -net for A in time $(\gamma/\epsilon)^{\gamma^2 \log(mn)/\epsilon^2}$.

3.1 Factorization Norm

For a m -by- n matrix A of rank d , let $\sigma_1(A) \geq \dots \geq \sigma_d(A) \geq 0$ denote the non-zero singular values of A . The trace norm $\|A\|_{tr} = \sum_{i=1}^d \sigma_i(A)$ is the sum of the singular values of A . A simple bound on the rank of A can be given in terms of the trace norm,

$$\|A\|_{tr} = \sum_{i=1}^d \sigma_i(A) \leq d^{1/2} \left(\sum_{i=1}^d \sigma_i^2(A) \right)^{1/2}.$$

This gives

$$\text{rank}(A) \geq \left(\frac{\|A\|_{tr}}{\|A\|_F} \right)^2, \quad (2)$$

where $\|A\|_F = \sqrt{\sum_i \sigma_i(A)^2} = \sqrt{\text{Tr}(AA^*)}$ is the Frobenius norm of A .

A drawback to this bound is that it is non-monotone in the sense that it can give a better bound on a submatrix of A than on A itself. We can remedy this in the following way. Let $A \circ B$ denote the entrywise product of A and B . As $\text{rank}(A) \geq \text{rank}(A \circ uv^*)$ for any vectors u, v we can maximize the above bound on $A \circ uv^*$ over all vectors u, v . This motivates the definition of γ_2 . Here $\|u\|$ denotes the ℓ_2 norm of u .

► **Definition 6.**

$$\gamma_2(A) = \max_{\substack{u, v \\ \|u\| = \|v\| = 1}} \|A \circ uv^*\|_{tr}$$

In a similar way to rank, we can define an approximate version of γ_2 . Originally this was defined in a multiplicative sense [25], but for consistency with approximate rank we define it in an additive way here.

► **Definition 7.** Let A be a matrix and $\epsilon \geq 0$.

$$\gamma_2^\epsilon(A) = \min_{\substack{B \\ \|A-B\|_\infty \leq \epsilon}} \gamma_2(B).$$

Exactly as in (2) we obtain that γ_2^ϵ gives the following lower bound on approximate rank.

► **Theorem 8.** Let A be a matrix and $\epsilon \geq 0$. Then

$$\text{rank}_\epsilon(A) \geq \left(\frac{\gamma_2^\epsilon(A)}{\|A\|_\infty + \epsilon} \right)^2.$$

To show that γ_2^ϵ is also not too much smaller than the approximate rank it is useful to work with an alternative characterization of γ_2 as a factorization norm. Let $\|v\|_p$ denote the ℓ_p norm of v . For a m -by- n matrix A and non-negative integers p, q (possibly ∞) define the norm

$$\|A\|_{p \rightarrow q} = \max_{\|y\|_p=1} \|Ay\|_q.$$

By writing γ_2 as a semidefinite program and taking the dual, one arrives at the following formulation (see, for example, [10] or [26]).

► **Lemma 9.** Let A be an m -by- n matrix. Then

$$\gamma_2(A) = \min_{\substack{X, Y \\ XY=A}} \|X\|_{2 \rightarrow \infty} \|Y\|_{1 \rightarrow 2}.$$

Notice that $\|X\|_{2 \rightarrow \infty}$ is equal to the largest ℓ_2 norm of a row of X . Similarly $\|Y\|_{1 \rightarrow 2}$ is equal to the largest ℓ_2 norm of a column of Y .

Using the Johnson-Lindenstrauss [15] dimension reduction lemma, [23] show that the approximate γ_2 norm in fact characterizes the approximate rank, up to a logarithmic factor and small change in the approximation parameter.

► **Theorem 10** ([23]). Let A be an m -by- n matrix with $\gamma_2^\epsilon(A) = \gamma$ witnessed by a factorization $A = XY$ where X is an m -by- k matrix and Y is k -by- n . For any $\delta > 0$ let $r = 8\gamma^2 \ln(4mn)/\delta^2$. Then

$$\Pr_R[\|A - XRR^TY\|_\infty \leq \delta] \geq \frac{1}{2},$$

where the probability is taken over R a random k -by- r matrix with entries independent and identically distributed according to the normal distribution with mean 0 and variance 1. In particular,

$$\text{rank}_{\delta+\epsilon}(A) \leq 8 \ln(4mn) \frac{\gamma_2^\epsilon(A)^2}{\delta^2}$$

The logarithmic factor in this theorem is in fact necessary, as can be seen with the identity matrix. The identity matrix I_n of size n has $\gamma_2(I_n) = 1$, but Alon [6] shows that $\text{rank}_\epsilon(I_n) = \Omega\left(\frac{\log(n)}{\epsilon^2 \log(1/\epsilon)}\right)$ for $\frac{1}{2\sqrt{n}} \leq \epsilon \leq \frac{1}{4}$.

Theorem 10 combined with the results in [7] gives the following corollary.

► **Corollary 11.** Let A be an $m \times n$ matrix with entries in $[-1, 1]$ and $\text{rank}_{\epsilon/4}(A) = d$. Then an ϵ -net for A can be constructed by a Las Vegas randomized algorithm in time $(1/\epsilon)^{O(d \ln(mn))}$.

3.2 Constructing ϵ -Nets via γ_2

In this section we prove from scratch an upper bound on the covering number in terms of the γ_2 norm. This gives weaker bounds than Corollary 11 but has the advantage of having a direct and simple proof.

► **Theorem 12.** *Let A be an m -by- n matrix. Suppose that $A = XY$ where X is m -by- d , Y is d -by- n , and $\|X\|_{2 \rightarrow \infty} \|Y\|_{1 \rightarrow 2} = \gamma$. Then $N_\epsilon(A) = O(\gamma/\epsilon)^d$. Moreover, an ϵ -net of this size can be constructed in time $O(\gamma/\epsilon)^d \text{poly}(mn)$.*

Proof. We can assume without loss of generality that $\|X\|_{2 \rightarrow \infty} = \gamma$ and $\|Y\|_{1 \rightarrow 2} = 1$. Then by definition $\|Yx\|_2 \leq 1$ for any x with $\|x\|_1 \leq 1$.

Let S be an ϵ/γ -net for the unit ball in \mathbb{R}^d of size $O(\gamma/\epsilon)^d$. There are standard explicit constructions for such nets that work in time $O(\gamma/\epsilon)^d$, for example by taking a tiling by cubes of size $\epsilon/(\gamma\sqrt{d})$. Then

$$\forall x : \|x\|_1 = 1, \exists \tilde{y} \in S : \|Yx - \tilde{y}\|_2 \leq \frac{\epsilon}{\gamma}.$$

Now, apply X to the vector $Yx - \tilde{y}$. Since $\|X\|_{2 \rightarrow \infty} = \gamma$, it holds that

$$\|XYx - X\tilde{y}\|_\infty \leq \gamma \cdot \frac{\epsilon}{\gamma} = \epsilon.$$

Thus we can take the set $T = \{X\tilde{y} : \tilde{y} \in S\}$. This can be constructed from S in time $O(\gamma/\epsilon)^d \text{poly}(m, n)$. ◀

► **Corollary 13.** *Let A be an m -by- n matrix and $\epsilon > 0$. Let $\gamma_2^{\epsilon/4}(A) = \gamma$. Then $N_\epsilon(A) = (\gamma/\epsilon)^{O(\gamma^2 \ln(mn)/\epsilon^2)}$. Moreover, an ϵ -net of this size can be constructed by a Las Vegas randomized algorithm in time $(\gamma/\epsilon)^{O(\gamma^2 \ln(mn)/\epsilon^2)} \text{poly}(m, n)$.*

Proof. First we solve the semidefinite program for $\gamma_2^{\epsilon/4}$ to obtain matrices U, V such that $\|UV - A\|_\infty \leq \epsilon/4$ and $\|U\|_{2 \rightarrow \infty} \|V\|_{1 \rightarrow 2} = \gamma$. Then let $X = UR$ and $Y = R^T V$ for a random d -by- d matrix R with $d = O(\frac{\gamma^2}{\epsilon^2} \ln(mn))$. By Theorem 10 with high probability we have $\|A - XY\|_\infty \leq \epsilon/2$. Applying Theorem 12 to XY gives a set T of size $O(\frac{\gamma}{\epsilon})^d$ such that

$$\forall x \in \Delta_n, \exists \tilde{x} \in T : \|XYx - \tilde{x}\|_\infty \leq \frac{\epsilon}{2}.$$

Thus

$$\begin{aligned} \|Ax - \tilde{x}\|_\infty &= \|Ax - XYx + XYx - \tilde{x}\|_\infty \\ &\leq \|(A - XY)x\|_\infty + \|XYx - \tilde{x}\|_\infty \\ &\leq \epsilon. \end{aligned}$$

The identity matrix again shows that the logarithmic factor in the statement of Corollary 13 is necessary. ◀

► **Lemma 14.** *Fix a natural number $k > 0$. Then $N_\epsilon(I_n) \geq \binom{n}{k}$, for every $\epsilon < \frac{1}{2k}$.*

Proof. For a subset $S \subseteq [n]$ of size k denote by v_S the vector $v = (v_1, v_2, \dots, v_n)$ satisfying $v_i = 1/k$ if $i \in S$ and $v_i = 0$ otherwise. Then, for every pair of subsets $S \neq T \subseteq [n]$ of size k , we have that $\|v_S - v_T\|_\infty = 1/k$. If $\epsilon < \frac{1}{2k}$ this implies that v_S and v_T must have distinct representatives, which implies the lemma. ◀

4 A Quasi-polynomial Upper Bound and VC dimension

Considering the upper bounds on the cover number in terms of approximate rank or approximate γ_2 described above, one might build the expectation that these bounds characterize the cover number well. This is actually true for some ranges of error, as we will see in Section 5.1. But for fixed ϵ this is far from the truth. In this case, by the results in [3] the bound via approximate rank is at least as large as $2^{\Omega(n)}$ for almost all $n \times n$ sign matrices, and the same holds for the bound via approximate γ_2 , while on the other hand, the next theorem states that the cover number is at most $n^{O(\log n)}$ for every such matrix.

► **Theorem 15.** For any $A \in [-1, 1]^{m \times n}$,

$$N_\epsilon(A) \leq \binom{n + \frac{2 \ln(2m)}{\epsilon^2}}{\frac{2 \ln(2m)}{\epsilon^2}} < n^{\frac{2 \ln(2m)}{\epsilon^2}}.$$

Theorem 15 can be derived as a special case of Maurey’s Lemma ([27]), see also [28, Lemma 13] for a related result. For completeness we include here a short proof.

Proof. Put $k = \frac{2 \ln(2m)}{\epsilon^2}$, and let A be as in the theorem. Let A_1, A_2, \dots, A_n denote the columns of A . It suffices to prove that for any vector $y = (y_1, y_2, \dots, y_m)$ in the convex hull of the columns of A there is a (multi)-subset $S = (A_{i_1}, A_{i_2}, \dots, A_{i_k})$ of k (not necessarily distinct) columns of A so that

$$\left\| \frac{1}{k} \sum_{j=1}^k A_{i_j} - y \right\|_\infty \leq \epsilon, \tag{3}$$

since the number of these subsets is at most

$$\binom{n + \frac{2 \ln(2m)}{\epsilon^2}}{\frac{2 \ln(2m)}{\epsilon^2}}.$$

To prove this fact suppose $y = \sum_{j=1}^m A_j p_j$. Choose the elements of the subset S randomly and independently among the columns of A (with repetitions), where each A_{i_j} is obtained by picking one of the columns, where A_j is chosen with probability p_j . The coordinate number i of the random sum $\sum_{j=1}^k A_{i_j}$ obtained is thus a sum of k independent identically distributed random variables, each having expectation y_i/k and each being bounded in absolute value by $1/k$. It thus follows by the standard Chernoff-Hoeffding-Azuma Inequality (c.f., e.g., [8]) that the probability this coordinate differs from y_i by more than ϵ is smaller than $1/m$, and hence with positive probability (3) holds. ◀

As we show in Section 5, the assertion of Theorem 15 is essentially tight. But it can still be improved if we have some extra information about the matrix A . One way to do it is in terms of the VC-dimension of A , defined next.

► **Definition 16.** Let $A \in \mathbb{R}^{m \times n}$ be a matrix. Let $C = \{c_1, \dots, c_k\} \subseteq [n]$ be a subset of columns of A . We say that A shatters C if there are real numbers $(t_{c_1}, \dots, t_{c_k})$ 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 $c \in C \setminus D$. Let $\text{VC}(A)$ be the maximal size of a set of columns shattered by A .

Note that $\text{VC}(A) \leq \log(m)$ for any m -by- n matrix. Sometimes the quantity in Definition 16 is referred to as pseudo-dimension, and VC dimension is reserved for sign or boolean matrices

where the choice of thresholds is not needed. For convenience we will use VC dimension for this more general definition as well.

The key to our upper bound is the following lemma. This was originally shown, with an additional logarithmic factor, in the original paper of Vapnik and Chervonenkis defining VC dimension [30]. The logarithmic factor was later removed by Talagrand [29] (see also [18] for a simpler proof).

► **Lemma 17.** [30, 29, 18] *Let $A \in [-1, 1]^{m \times n}$ be a matrix with $\text{VC}(A) = d$. For $S \subseteq [n]$ let $\chi_S \in \{0, 1\}^n$ denote its characteristic vector. For any $\epsilon > 0$ and $S \subseteq [n]$ there is a set $T \subset S$ of size $|T| = O\left(\frac{d}{\epsilon^2}\right)$ such that*

$$\left\| \frac{A\chi_S}{|S|} - \frac{A\chi_T}{|T|} \right\|_{\infty} \leq \epsilon.$$

This lemma says that every uniform combination of columns of A can be ϵ approximated by a uniform combination of about $\text{VC}(A)/\epsilon^2$ many columns. In the next theorem we obtain an upper bound on the cover number in terms of VC dimension by reducing the case of arbitrary probability distributions to that of uniform distributions and applying Lemma 17.

► **Theorem 18.** *Let $A \in [-1, 1]^{m \times n}$ be a matrix with $\text{VC}(A) = d$. Then*

$$N_{\epsilon}(A) \leq n^{O(d/\epsilon^2)}.$$

Proof. We will use Lemma 17 to show that every element of the convex hull of the columns of A can be ϵ -approximated by the average of some $O(d/\epsilon^2)$ columns (with repetition) of A .

Let $N = 2n/\epsilon$. Let A' be the matrix where every column of A is repeated N times. As duplicating columns does not change the VC dimension we have $\text{VC}(A') = d$. Let $p \in [0, 1]^n$ be a probability vector. Define $p' \in [0, 1]^{Nn}$ as

$$p'(jN + i) = \begin{cases} \frac{1}{N} & \text{for } j = 0, \dots, n-1 \text{ and } i = 1, \dots, \lfloor p(j)N \rfloor \\ 0 & \text{for } j = 0, \dots, n-1 \text{ and } i = \lfloor p(j)N \rfloor + 1, \dots, N. \end{cases}$$

Note that $\|Ap - A'p'\|_{\infty} \leq \frac{n}{N} \leq \epsilon/2$. As p' is a normalized characteristic vector, by Lemma 17 we have that there is a set D of size $O\left(\frac{d}{\epsilon^2}\right)$ such that

$$\left\| A'p' - \frac{A'\chi_D}{|D|} \right\|_{\infty} \leq \epsilon/2.$$

Thus to obtain an ϵ -net for A it suffices to take all uniform combinations of $O\left(\frac{d}{\epsilon^2}\right)$ columns, taken with repetition. This gives the theorem. ◀

5 Lower Bounds

In this section, we show some lower bounds on the covering number in terms of approximate rank, one-way communication complexity and VC dimension. Some of the lower bounds we prove match the corresponding upper bounds shown earlier.

5.1 Tight Lower Bounds in the General Case

In this section, we show that $N_{0.99}(A) = n^{\Omega(\log n)}$ for a random sign matrix A , and thus that our upper bounds in terms of VC dimension is tight in this case. We also show a lower bound of $N_{2/7}(A) = 2^{\Omega(\text{VC}(A))}$ for any sign matrix A . Our lower bounds follow from bounds

on the closely related *packing number* of A . Let $C_\delta(A)$ be the maximal number of δ -size ℓ_∞ balls that can be packed into the convex hull of the columns of A . Then

$$C_{2\epsilon}(A) \leq N_\epsilon(A) \leq C_\epsilon(A).$$

We obtain our bounds via the next simple lemma, together with the existence of appropriate nearly disjoint families of sets.

► **Lemma 19.** *Let A be an m -by- n sign matrix and \mathcal{F} a family of subsets of $[n]$ such that*

1. *for every $F, F' \in \mathcal{F}$ the columns of A in $F \cup F'$ are shattered.*
2. *$|F \cap F'| \leq (1 - \delta/2)|F|$ for all distinct $F, F' \in \mathcal{F}$.*

Then $N_{\delta/2}(A) \geq |\mathcal{F}|$.

Proof. Let A_j denote the j th column of A . For any $F \in \mathcal{F}$, the vector

$$v_F = \frac{1}{|F|} \sum_{j \in F} A_j$$

lies in the convex hull of the columns of A . Now consider $\|v_F - v_{F'}\|_\infty$ for distinct $F, F' \in \mathcal{F}$. As $F \cup F'$ is shattered, there is a row i such that $A(i, j) = 1$ for all $j \in F$ and $A(i, j) = -1$ for all $j \in F' \setminus F$. Thus

$$\|v_F - v_{F'}\|_\infty = 1 - \frac{1}{|F'|} (|F \cap F'| - |F' \setminus F|) \geq \delta.$$

◀

► **Claim 20.** *There is a family \mathcal{F} of subsets of $[d]$ such that*

1. *$|F| \geq \frac{7}{16}d$ for all $F \in \mathcal{F}$*
2. *$|F \cap F'| \leq \frac{5}{16}d$ for all distinct $F, F' \in \mathcal{F}$*
3. *$|\mathcal{F}| \geq 2^{.001d}$*

► **Claim 21.** *There is a family \mathcal{F} of subsets of $[n]$ such that*

1. *$|F| = 0.49 \log_2 n$ for all $F \in \mathcal{F}$*
2. *$|F \cap F'| \leq 0.0001 \log_2 n$ for all distinct $F, F' \in \mathcal{F}$*
3. *$|\mathcal{F}| \geq n^{\Omega(\log n)}$*

Proof. The existence of such \mathcal{F} follows either by a simple probabilistic argument, or by using known bounds for constant weight codes, or by an explicit constructions using polynomials.

◀

► **Lemma 22.** *Let A be a sign matrix. Then $N_{2/7}(A) \geq 2^{\Omega(\text{VC}(A))}$.*

Proof. This follows from Lemma 19 together with the set family from Claim 20.

◀

► **Lemma 23.** *For almost all n -by- n sign matrices A ,*

$$N_{.99}(A) = n^{\Omega(\log n)}.$$

Proof. Let $A = (a_{ij})$ be a random n -by- n sign matrix, where each entry $a_{ij} \in \{-1, 1\}$ is chosen randomly, independently and uniformly in $\{-1, 1\}$. We show that with high probability A shatters every subset $J \subseteq [n]$ of columns with $|J| \leq 0.98 \log n$. Indeed, for a fixed J and sign pattern $s \in \{-1, +1\}^{|J|}$, the probability that no row of A restricted to J is equal to s is

$$(1 - 2^{-|J|})^n < e^{-n^{0.02}}.$$

The result thus follows from the union bound.

Therefore, the VC dimension of a random n -by- n sign matrix is greater than $0.98 \log n$ with high probability. The proof of the Lemma now follows from Lemma 19 using the set family from Claim 21.

◀

5.1.1 An Explicit Example

In addition to the lower bound $N_{99}(A) = n^{\Omega(\log(n))}$ for a random sign matrix A , we can also show explicit examples where the cover number is this large. We show next that a simple Hadamard matrix requires covers of quasi-polynomial size. Consider the 2^t -by- 2^t Hadamard matrix, $H = (h_{v,M})$, whose columns are indexed by monomials $M = \prod_{i \in I} x_i$ with $I \subset [t]$ and whose rows are indexed by vectors $v \in \{-1, 1\}^t$, where $h_{v,M} = M(v)$. In this matrix, for any choice of monomials M_1, M_2, \dots, M_k in which no product of a subset is identically 1, the polynomial

$$\frac{1 + M_1}{2} \frac{1 + M_2}{2} \dots \frac{1 + M_k}{2}$$

is the average of 2^k monomials. Its value on a vector v is 1 if $M_j(v) = 1$ for all j , and is 0 otherwise. This gives, if we shift to an additive rather than multiplicative notation, for every subspace of dimension $t/2$ of Z_2^t , a vector in the convex hull of the columns of H which is 1 on the members of the subspace and 0 outside it. Therefore, this example is an $n = 2^t$ by $n = 2^t$ sign matrix H for which $N_\epsilon(A) \geq n^{(1+o(1)) \log n/4}$ for all $\epsilon < 1/2$.

5.2 Lower Bounds on the Cover Number in Terms of Approximation Rank

For a vector v and a linear subspace U we also define

$$d(v, U) = \min_{u \in U} \|v - u\|_\infty.$$

► **Lemma 24.** *Let A be a real matrix and fix $0 < \epsilon$. Let d be the ϵ -approximate rank of A . Then there are d columns of A , $a_{i_1}, a_{i_2}, \dots, a_{i_d}$, such that*

$$d(a_{i_j}, \text{span}\{a_{i_1}, \dots, a_{i_{j-1}}\}) \geq \epsilon,$$

for every $1 \leq j \leq d$.

Proof. We construct the set of columns inductively. We choose the first column as any nonzero column (such a column must exist if $d > 0$). If we have constructed d columns already, we are done. Otherwise we have $a_{i_1}, a_{i_2}, \dots, a_{i_t}$ for $t < d$. By definition of approximate rank, and since $d_\epsilon(A) = d > t$, there must be a column that is ϵ -far from $\text{span}(a_{i_1}, a_{i_2}, \dots, a_{i_t})$. We add this column to the set. ◀

► **Theorem 25.** *Let A be a real matrix and fix $0 < \epsilon$. Let d be the ϵ -approximate rank of A . Then*

$$N_{\epsilon/d}(A) = \Omega\left(\frac{2^d}{\sqrt{2d}}\right).$$

Proof. By Lemma 24 there are d columns of A , $a_{i_1}, a_{i_2}, \dots, a_{i_d}$, such that

$$d(a_{i_j}, \text{span}\{a_{i_1}, \dots, a_{i_{j-1}}\}) \geq \epsilon,$$

for every $1 \leq j \leq d$. Assume w.l.o.g that these are the first d columns of A , a_1, a_2, \dots, a_d . For simplicity we assume that d is even; if not the argument below can be done with $d - 1$ which only changes the bound by a constant factor.

Consider the set of vectors $S = \{\frac{2}{d} \sum_{i \in T} a_i : T \subseteq [d], |T| = d/2\}$. We claim that for every two vectors $v, u \in S$ it holds that $\|v - u\|_\infty \geq \epsilon/d$: Let $u = \frac{2}{d} \sum_{i \in T_1} a_i$ and $v = \frac{2}{d} \sum_{i \in T_2} a_i$

for $T_1 \neq T_2$. Then

$$u - v = \frac{2}{d} \left(\sum_{i \in T_1} a_i - \sum_{i \in T_2} a_i \right) = \frac{2}{d} \left(\sum_{i \in T_1 \setminus T_2} a_i - \sum_{i \in T_2 \setminus T_1} a_i \right).$$

Let j be the largest index in $T_1 \Delta T_2$. Assume w.l.o.g that $j \in T_1$, we have

$$\|u - v\|_\infty = \frac{2}{d} \left\| a_j - \left(\sum_{i \in T_2 \setminus T_1} a_i - \sum_{j \neq i \in T_1 \setminus T_2} a_i \right) \right\|_\infty \geq \frac{2\epsilon}{d}.$$

The last inequality is because $\sum_{i \in T_2 \setminus T_1} a_i - \sum_{j \neq i \in T_1 \setminus T_2} a_i$ is contained in the linear subspace spanned by a_1, a_2, \dots, a_{j-1} .

Since S is in the convex hull of the columns of A and $|S| \geq 2^d / \sqrt{2d}$, we get that

$$\frac{2^d}{\sqrt{2d}} \leq C_{2\epsilon/d}(A) \leq N_{\epsilon/d}(A).$$

◀

5.3 Lower Bounds via Communication Complexity

► **Lemma 26.** *Let A be a sign matrix, and denote by $cc(A)$ the one-way (from Bob to Alice) deterministic communication complexity of A . Then*

$$cc(A) \leq \log(N_\epsilon(A))$$

for every ϵ in $(0, 1)$.

Proof. Let $k = cc(A)$, then there are 2^k distinct columns in A . Since the ℓ_∞ distance between every two distinct sign vectors is at least 2, we have

$$2^k \leq N_\epsilon(A)$$

for every $\epsilon \in (0, 1)$.

◀

The log rank conjecture, formulated by Lovász and Saks [22] is a long-standing open problem in communication complexity. A simple argument shows that $\log \text{rank}(A)$ is a lower bound on the deterministic communication complexity $D(A)$ of A . The log rank conjecture states that this bound is polynomially tight, $D(A) = \log(\text{rank}(A))^{O(1)}$. The best upper bound on communication complexity in terms of rank was recently improved to show $D(A) = O(\sqrt{\text{rank}(A)} \log(\text{rank}(A)))$ [21].

Combining Lemma 26 with the relation between the covering number and approximate rank proved in [7], we get an upper bound on the one-way communication complexity in terms of the *approximate* rank.

► **Corollary 27.** *Let A be a sign matrix, and denote by $cc(A)$ the one-way deterministic communication complexity of A . Then*

$$cc(A) \leq 3\text{rank}_{1/2}(A) + O(1).$$

6 Conclusion and Open Problems

Efficiently enumerable covers of the convex hull of a matrix lead to efficient approximation algorithms for a broad class of optimization problems including Nash equilibrium and densest k -by- k combinatorial rectangle. We have shown that $N_\epsilon(A) \leq n^{O(\text{VC}(A)/\epsilon^2)}$ and moreover that such covers can be deterministically enumerated in about the same time. This result unifies many previous approximation algorithms for Nash equilibrium in the literature, including the quasi-polynomial time approximation algorithm of Lipton et al. [19] and the approximation algorithm of Kannan and Theobald for game matrices A, B such that $A + B$ has constant rank [17]. For the densest k -by- k combinatorial rectangle problem this gives for the first time a $n^{O(\log(n)/\epsilon^2)}$ time algorithm to obtain an additive ϵ -approximation.

The central open problem if Nash Equilibrium has a polynomial time approximation scheme remains open. One avenue to make progress on this question may be to find a common generalization of the cover based approximation algorithms given here and in [7], with the approximation algorithm for random games of Bárány et al. [11].

Acknowledgements. We thank Gideon Schechtman, Shai Shalev-Shwartz and Santosh Vempala for helpful discussions. Noga Alon is supported in part by an ERC Advanced grant, a USA-Israeli BSF grant, an ISF grant, the Israeli I-Core program, and by the Simonyi Fund. Troy Lee is supported in part by the Singapore National Research Foundation under NRF RF Award No. NRF-NRFF2013-13.

References

- 1 N. Alon, S. Arora, R. Manokaran, D. Moshkovitz, and O. Weinstien. On the inapproximability of the densest k -subgraph problem. unpublished manuscript, 2011.
- 2 N. Alon, R. Duke, H. Lefmann, V. Rödl, and R. Yuster. The algorithmic aspects of the regularity lemma. *J. Algorithms*, 16(1):80–109, 1994.
- 3 N. Alon, P. Frankl, and V. Rödl. Geometric realization of set systems and probabilistic communication complexity. In *Proceedings of the 26th IEEE Symposium on Foundations of Computer Science*, pages 277–280. IEEE, 1985.
- 4 S. Arora, D. Karger, and M. Karpinski. Polynomial time approximation schemes for dense instances of np-hard problems. *Journal of Computer and System Sciences*, 58:193–210, 1999.
- 5 Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *SODA*, pages 594–598, 1998.
- 6 N. Alon. Perturbed identity matrices have high rank: proof and applications. *Combinatorics, Probability, and Computing*, 18:3–15, 2009.
- 7 N. Alon, T. Lee, A. Shraibman, and S. Vempala. The approximate rank of a matrix and its algorithmic applications. In *Proceedings of the 45th ACM Symposium on the Theory of Computing*. ACM, 2013.
- 8 N. Alon and J. Spencer. *The Probabilistic Method*. John Wiley and Sons, third edition, 2008.
- 9 A. Bhaskara, M. Charikar, E. Chlamtac, U. Feige, and A. Vijayaraghavan. Detecting high log-densities: an $n^{1/4}$ approximation for densest subgraph. In *Proceedings of the 42nd ACM Symposium on the Theory of Computing*, pages 201–210, 2010.
- 10 R. Bhatia. *Positive Definite Matrices*. Princeton University Press, 2007.
- 11 Imre Bárány, Santosh Vempala, and Adrian Vetta. Nash equilibria in random games. *Random Struct. Algorithms*, 31(4):391–405, 2007.

- 12 Yael Dekel, Ori Gurel-Gurevich, and Yuval Peres. Finding hidden cliques in linear time with high probability. *CoRR*, abs/1010.2997, 2010.
- 13 Uriel Feige and Dorit Ron. Finding hidden cliques in linear time. In *AOFA*, 2010.
- 14 Mark Jerrum. Large cliques elude the metropolis process. *Random Struct. Algorithms*, 3(4):347–360, 1992.
- 15 W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mapping into Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.
- 16 S. Khot. Ruling out PTAS for graph min-bisection, densest subgraph and bipartite clique. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*. IEEE, 2004.
- 17 R. Kannan and T. Theobald. Games of fixed rank: A hierarchy of bimatrix games. *Econom. Theory*, 42:157–173, 2010. appeared in SODA 2007.
- 18 Y. Li, P. M. Long, and A. Srinivasan. Improved bounds on the sample complexity of learning. *Journal of Computer and System Sciences*, 62:516–527, 2001.
- 19 R. J. Lipton, E. Markakis, and A. Mehta. Playing large games using simple strategies. In *ACM Conference on Electronic Commerce*, pages 36–41, 2003.
- 20 N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.
- 21 S. Lovett. Communication is bounded by root of rank. Technical Report arXiv:1306.1877, arXiv, 2013.
- 22 L. Lovász and M. Saks. Möbius functions and communication complexity. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 81–90. IEEE, 1988.
- 23 T. Lee and A. Shraibman. An approximation algorithm for approximation rank. In *Proceedings of the 24th IEEE Conference on Computational Complexity*, pages 351–357. IEEE, 2008. arXiv:0809.2093 [cs.CC].
- 24 N. Linial and A. Shraibman. Learning complexity versus communication complexity. *Combinatorics, Probability, and Computing*, 18:227–245, 2009.
- 25 N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures and Algorithms*, 34:368–394, 2009.
- 26 T. Lee, A. Shraibman, and R. Špalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 71–80. IEEE, 2008.
- 27 B. Maurey. Théorèmes de factorisation pour les opérateurs linéaires à valeurs dans les espaces l_p . *Astérisque*, (11):1–163, 1974.
- 28 S. Sabato, S. Shalev-Shwartz, N. Srebro, D. Hsu and T. Zhang. Learning Sparse Low-Threshold Linear Classifiers *CoRR*, abs/1212.3276, 2012.
- 29 M. Talagrand. Sharper bounds for gaussian and empirical processes. *Annals Probab.*, 22:28–76, 1994.
- 30 V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, 16:264–280, 1971.