

Complexity Measures of Sign Matrices

Nati Linial*

School of Computer Science and Engineering
Hebrew University
Jerusalem, Israel
e-mail: nati@cs.huji.ac.il

Shahar Mendelson†

Research School of Information Sciences and Engineering,
The Australian National University,
Canberra, ACT 0200, Australia
e-mail: shahar.mendelson@anu.edu.au

Gideon Schechtman‡

Department of Mathematics,
The Weizmann Institute,
Rehovot, Israel
e-mail: gideon.schechtman@weizmann.ac.il

Adi Shraibman

School of Computer Science and Engineering
Hebrew University
Jerusalem, Israel
e-mail: adidan@cs.huji.ac.il

August 2, 2004

Abstract

In this paper we consider four previously known parameters of sign matrices from a complexity-theoretic perspective. The main technical contributions are tight (or nearly tight) inequalities that we establish among these parameters. Several new open problems are raised as well.

*Supported by the ISF

†Supported by the ARC

‡Supported by the ISF

1 Introduction

What is *complexity*, and how should it be studied mathematically? In the interpretation that we adopt, there are several underlying common themes to complexity theories. The basic ground rules are these: There is a family \mathcal{F} of some mathematical objects under consideration. The elements of some subset $\mathcal{S} \subseteq \mathcal{F}$ are deemed *simple*. Also, there are certain composition rules that allow one to put together objects in order to generate other objects in \mathcal{F} . The complexity of an object is determined by the length of the shortest chain of steps to generate it from simple objects. In full generality one would want to get good estimates for all or many objects in the family \mathcal{F} . Specifically, a major challenge is to be able to point out specific concrete objects that have high complexity. That is, elements that cannot be generated from simple objects using only a small number of composition steps.

Arguably the currently most developed mathematical theory of complexity, is to be found in the field of computational complexity. Typically (but not exclusively), \mathcal{F} consists of all boolean functions $f : \{0,1\}^m \rightarrow \{0,1\}$. The class \mathcal{S} of simple objects contains the constant functions, and the functions $x \rightarrow x_i$ (the i -th coordinate). Functions can be composed using the basic logical operations (**or**, **and**, **not**). Thus, one possible formulation of the P vs. NP problem within this framework goes as follows: Suppose that $m = \binom{n}{2}$ and so each $x \in \{0,1\}^m$ can be viewed as a graph G on n vertices (each coordinate of x indicates whether a given pair of vertices is connected by an edge or not). We define $f(x)$ to be 0 or 1 according to whether or not G has a Hamiltonian path (a path that visits every vertex in G exactly once). It is conjectured that in order to generate the function f , exponentially many composition steps must be taken. The lamentable state of affairs is that we are at present unable to prove even any super linear lower bound for this number.

In view of the fundamental importance and the apparent great difficulty of the problems of computational complexity we suggest to address issues of complexity in other mathematical fields. Aside from the inherent interest in understanding complexity in general, insights gained from such investigations are likely to help in speeding up progress in computational complexity. This paper is a small step in this direction. We seek to develop a complexity theory for sign matrices (matrices all of whose entries are ± 1). There are several good reasons why this should be a good place to start. First, a number of hard and concrete problems in computational complexity proper can be stated in this language. The two most notable examples are (i) The log-rank conjecture and (ii) The matrix rigidity problem, to be explained in the sequel. Also, matrices come with a complexity measure that we all know, namely, the *rank*. To see why, let us declare the class \mathcal{S} of simple matrices to be those matrices (not necessarily with ± 1 entries) that have rank one. Suppose, furthermore, that the composition rule is matrix sum. We recall a theorem from linear algebra that the rank of a matrix A equals the least number of rank-one matrices whose sum is A . This shows that rank indeed fits the definition of a complexity measure for matrices.

One important lesson from the experience gathered in computational complexity is that it is beneficial to study a variety of complexity measures in order to understand the behavior of the main quantities of interest. Thus, aside from circuit complexity (the “real” thing), people are investigating communication complexity, proof complexity, decision tree models etc. This is the direction we take here, and our main work here is a comparative study of several measures of complexity for sign matrices.

We turn to the two conjectures mentioned above. The log-rank conjecture arose in the subfield of computational complexity known as *communication complexity*. A few words about this area will be said below, and the interested reader should consult the beautiful monograph [15]. In purely matrix-theoretic terms, here is the conjecture:

Conjecture 1 [20] *Let A be an $n \times n$ sign matrix. Denote by M the largest area of a monochromatic rectangle of A , then*

$$M \geq n^2 / 2^{(\text{rank}(A))^{O(1)}}.$$

One recurring theme in computational complexity is that in many important situations, *random elements in \mathcal{F} have the highest possible complexity* (or nearly that). Thus a random sign matrix tends to have full rank (we will soon elaborate on this point). From this perspective, the log-rank conjecture probes the situation away from that realm, and asks whether low rank imposes strong structural restrictions on the matrix?

Indeed, ranks of sign matrices have attracted much attention over the years. The most famous open problem about them is this: What is the probability that a random $n \times n$ sign matrix be singular? In its strongest form, the conjecture says that singularity comes mostly from one of the following four events - Two rows (columns) that are equal (opposite). This would mean that the probability for being singular is $(1 + o(1)) \frac{n(n-1)}{2^n - 1}$. This conjecture still seems beyond reach and in a celebrated work Kahn, Komlos and Szemerédi [11] were able to give an upper bound of $(1 - \epsilon_0)^n$ for some $\epsilon_0 > 0$. Be it as it may, the result from [11] says that if \mathcal{F} consists of all $n \times n$ sign matrices, and if our complexity measure is the rank, then random objects in \mathcal{F} have the highest possible complexity and the exceptional set is exponentially small. Such phenomena are often encountered in complexity.

The *rigidity* problem (first posed by Valiant [27]) highlights another prevalent phenomenon in computational complexity. Namely, while *most* objects in \mathcal{F} have (almost) the largest possible complexity, finding *explicit* members in \mathcal{F} that have high complexity is a different matter altogether. Some of the hardest problems in computational complexity are instances of this general phenomenon. Of course, finding explicit matrices of full rank is very easy. But as it was proved by Valiant for real matrices and in [23] for ± 1 matrices, high rank is not only very common, it is also very rigid. Namely, when you draw a random sign matrix, even when you are allowed to *arbitrarily* change a constant fraction of the entries in the matrix the rank will remain high. The problem is to construct

explicit matrices which with this property. It is conjectured that the Sylvester matrices are rigid and in spite significant effort [14] this problem remains open. Other variants of rigidity were also studied [16, 24], with some success, but the implications for computational complexity are not very strong.

This paper revolves around four matrix parameters. All four had been studied before, but not necessarily as complexity measures. Let us introduce these parameters in view of the following definition for the rank. We observe that the rank of a real $m \times n$ matrix A is the smallest d , such that it's possible to express $A = XY$, where X is a real $m \times d$ matrix and Y a $d \times n$ real matrix. In general, all four complexity measures that we consider will be derived as various parameters optimized over all possible ways to express A as $A \simeq XY$ for some real matrices X, Y . We consider two interpretations for \simeq . It will either mean matrix equality, or it will mean that A is the sign matrix of XY , and that all entries in XY have absolute values ≥ 1 (the latter is a necessary normalization condition).

The other ingredient in our definitions is that we'd like X and Y to have “short” rows and columns respectively. Here short may be interpreted in two ways: either meaning few coordinates or having small ℓ_2 norm. We are thus led to four distinct definitions.

	equality	sign
num. of rows	$r = \text{rank}$	$d =$ randomized comm. compl.
length of rows	γ_2 - normed spaces theory	$mc =$ margin complexity

Figure 1: complexity measures

Of the four parameters that appear in Figure 1, the rank needs no introduction, of course. The parameter γ_2 originates from the theory of normed spaces and will be discussed below. The other two parameters were first introduced in computational contexts. Margin complexity mc is a notion that comes from the field of machine learning. The fourth and last of the parameters comes from the field of communication complexity.

The main results of this paper concern these four parameters. We establish inequalities among them, and determine almost completely how tight these inequalities are. Besides, we prove concentration-of-measure results for them. It turns out that for comparison purposes, it's better to speak of γ_2^2 and mc^2 , rather than γ_2 and mc . Specifically, let $m \geq n$, we show for every $m \times n$ sign matrix A that:

- $\text{rank}(A) \geq \gamma_2^2(A)$.

The gap here can be arbitrarily large. For example, the “identity” matrix $2I_n - J_n$ has rank n and $\gamma_2 = O(1)$ (J_n is the $n \times n$ all 1's matrix).

- $\gamma_2(A) \geq mc(A)$.

Again the gap can be almost arbitrarily large. Specifically, we exhibit $n \times n$ sign matrices A for which

$$mc(A) = \log n \quad \text{and} \quad \gamma_2(A) = \Theta\left(\frac{\sqrt{n}}{\log n}\right).$$

We can construct certain specific matrices with $mc = O(1)$ for which we conjecture that γ_2 is still $\Theta\left(\frac{\sqrt{n}}{\log n}\right)$.

- $d(A), mc(A) \geq \Omega\left(\frac{nm}{\|A\|_{\infty \rightarrow 1}}\right)$.

We prove that for random sign matrices the right hand side is almost always $\Omega(\sqrt{n})$.

- $d(A) \leq O(mc(A)^2 \log(n + m))$.

- We show that the parameter γ_2 for $m \times n$ sign matrices is concentrated.

$$\Pr(|\gamma_2(A) - m_\gamma| \geq c) \leq 2e^{-c^2/16},$$

where m_γ denotes the median of γ_2 .

A one-sided inequality of a similar nature is:

$$\Pr(\gamma_2(A) \leq m_M - c/\sqrt{m}) \leq 2e^{-c^2/16},$$

where M denotes the median of $\gamma_2^*(A)$, and $m_M = nm/M$.

2 Definitions of the Complexity Measures

We turn to discuss the complexity measures under consideration here. The rank is, of course well known, and we introduce the three remaining measures.

2.1 γ_2 and operator norms

Denote by $M_{m,n}(\mathbb{R})$ the space of $m \times n$ matrices over the reals and set $\|\cdot\|_{\ell_1^n}$ and $\|\cdot\|_{\ell_2^n}$ the ℓ_1^n and ℓ_2^n norms on \mathbb{R}^n , respectively.

Let us recall the notion of a dual norm. If $\|\cdot\|$ is a norm on \mathbb{R}^n , the dual norm $\|\cdot\|^*$ is defined for every $x \in \mathbb{R}^n$ by

$$\|x\|^* = \max_{\|y\|=1} \langle x, y \rangle,$$

where \langle, \rangle denotes the (usual) inner product. An easy consequence of the definition is that for every $x, y \in \mathbb{R}^n$ and every norm on \mathbb{R}^n , $\|x\|\|y\|^* \geq |\langle x, y \rangle|$.

Given two norms E_1 and E_2 , on \mathbb{R}^m and \mathbb{R}^n respectively, the corresponding operator norm $\|\cdot\|_{E_1 \rightarrow E_2}$ is defined on $M_{m,n}(\mathbb{R})$ by

$$\|A\|_{E_1 \rightarrow E_2} = \sup_{\|x\|_{E_1}=1} \|Ax\|_{E_2}.$$

When the dimensions of the underlying normed spaces are evident from the context, we use the notation $\|\cdot\|_{p \rightarrow q}$ to denote the operator norm between the spaces ℓ_p^n and ℓ_q^m . An easy but useful property of operator norms is that:

$$\|BC\|_{E_1 \rightarrow E_2} \leq \|C\|_{E_1 \rightarrow E_3} \|B\|_{E_3 \rightarrow E_2}.$$

for every two matrices $B \in M_{m,k}(\mathbb{R})$ and $C \in M_{k,n}(\mathbb{R})$ and for every three norms E_1, E_2, E_3 , on $\mathbb{R}^m, \mathbb{R}^n$ and \mathbb{R}^k respectively.

Factorization of operators plays a central role in our discussion. This concept has been extensively studied in Banach spaces theory, see for example [26]. Given three normed spaces W_1, W_2 and Z and an operator $T : W_1 \rightarrow W_2$, the factorization problem deals with representations of the operator T as $T = uv$, where $v : W_1 \rightarrow Z$ and $u : Z \rightarrow W_2$, such that v and u have small norms. For fixed spaces W_1 and W_2 and $T : W_1 \rightarrow W_2$, define the factorization constant $\gamma_Z(T) = \inf \|v\|_{W_1 \rightarrow Z} \|u\|_{Z \rightarrow W_2}$, where the infimum is over all representations $T = uv$.

Factorization constants reflect the geometry of the three spaces involved. For example, if W_1, W_2 and Z are n -dimensional and if T is the identity operator, the factorization constant $\gamma = \gamma_Z(\text{Id})$ of this operator through Z corresponds to finding an image of the unit ball of Z (denoted by B_Z) which is contained in B_{W_2} and contains $1/\gamma \cdot B_{W_1}$. It is possible to show [26] that if Z is a Hilbert space, then for any W_1 and W_2 the factorization constant is a norm on the space of operators between W_1 and W_2 .

In the case of greatest interest for us, $W_1 = \ell_1^n, W_2 = \ell_\infty^m$ and $Z = \ell_2$. Then, denoting here and in the sequel $\gamma_2 = \gamma_{\ell_2}$,

$$\gamma_2(A) = \min_{XY=A} \|X\|_{\ell_2 \rightarrow \ell_\infty^m} \|Y\|_{\ell_1^n \rightarrow \ell_2}$$

which is one of the four complexity measures we investigate in this paper. It is not hard to check that if A is an $m \times n$ matrix then $\|A\|_{\ell_1^n \rightarrow \ell_2^m}$ is the largest ℓ_2^m norm of a column in A , and $\|A\|_{\ell_2^n \rightarrow \ell_\infty^m}$ is equal to the largest ℓ_2^n norm of a row in A . Thus

$$\gamma_2(A) = \min_{XY=A} \max_{i,j} \|x_i\|_{\ell_2} \|y_j\|_{\ell_2},$$

where $\{x_i\}_{i=1}^m$ are the rows of X , and $\{y_j\}_{j=1}^n$ are the columns of Y . Notice that $\gamma_2(A) = \gamma_2(A^t)$ and thus $\gamma_2^*(A) = \gamma_2^*(A^t)$, for every real matrix A .

We need a fundamental result from Banach spaces theory, known as Grothendieck's inequality, see e.g. [22], page 64.

Theorem 2 *There is an absolute constant $1 < K_G < 1.8$ such that the following holds: Let a_{ij} be a real matrix, and suppose that $|\sum_{i,j} a_{ij} s_i t_j| \leq 1$ for every choice of reals with $|s_i|, |t_j| \leq 1$ for all i, j . Then*

$$\left| \sum_{i,j} a_{ij} \langle x_i, y_j \rangle \right| \leq K_G,$$

for every choice of unit vectors x_i, y_j in a real Hilbert space.

Using duality, it is possible to restate Grothendieck's inequality as follows: For every matrix $A \in M_{m,n}(\mathbb{R})$

$$\gamma_2^*(A^t) \leq K_G \|A^t\|_{\ell_\infty^m \rightarrow \ell_1^n},$$

where γ_2^* is the dual norm to γ_2 .

On the other hand, it is easy to verify that if $A \in M_{m,n}$ then $\|A^t\|_{\ell_\infty^m \rightarrow \ell_1^n} \leq \gamma_2^*(A^t)$, implying that up to a small multiplicative constant γ_2^* is equivalent as a norm on $M_{n,m}$ to the norm $\|\cdot\|_{\ell_\infty^m \rightarrow \ell_1^n}$.

2.2 Margin complexity and machine learning

We turn now to define the margin of a concept class, an important quantity in modern machine learning (see, e.g. [6, 28]). A *concept class* is an $m \times n$ sign matrix, where the rows of the matrix represent points in the (finite) domain and columns correspond to concepts, i.e. $\{-1, 1\}$ -valued functions. The value of the j -th function on the i -th point is a_{ij} . The idea behind *margin based bounds* is to try and represent the function class as a class of linear functionals on an inner product space, namely to find vectors $y_1, \dots, y_n \in \ell_2$ to represent the functions in the class and vectors x_1, \dots, x_m corresponding to the points in the domain. This choice is a *realization* of the concept class if $\text{sign}(\langle x_i, y_j \rangle) = a_{ij}$ for every $1 \leq i \leq m$ and $1 \leq j \leq n$. In matrix terms, a realization of A , is a pair of matrices X, Y such that the matrix XY has the same sign pattern as A . The *margin* of this realization is defined as

$$\min_{i,j} \frac{|\langle x_i, y_j \rangle|}{\|x_i\| \|y_j\|}.$$

Hence, the closer the margin is to 1 the closer the representation (using elements of norm 1) is to be a completely accurate rendition. The margin provides a measure to the difficulty of a learning problem - at least to some extent, the larger the margin is, the simpler the concept class is, and more amenable to description with linear functionals.

The *margin* of a sign matrix A is defined as the largest possible margin of a realization of A , denoted $m(A)$. Observe that

$$m(A) = \sup \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{\|x_i\| \|y_j\|}, \quad (1)$$

where the supremum is over all matrices X, Y with $\text{sign}(\langle x_i, y_j \rangle) = a_{ij}$. It will be convenient to denote $mc(A) = m(A)^{-1}$, the *margin complexity* of A .

2.3 A few words on communication complexity

In table (1) we define $d(A)$ of an $m \times n$ sign matrix A as follows: This is the smallest dimension d such that it's possible to find vectors x_1, \dots, x_m and y_1, \dots, y_n in \mathbb{R}^d for which $\text{sign}(\langle x_i, y_j \rangle) = a_{ij}$ for all i, j . We also say that the matrix A can be realized in \mathbb{R}^d .

The first occurrence of this parameter was in *communication complexity*, a subfield of computational complexity mentioned above. We restrict ourselves to very brief remarks about it, and the reader is referred to the standard reference on this subject, the book by Kushilevitz and Nisan [15].

The basic scenario includes an $m \times n$ sign matrix A and two players, Alice the row player and Bob the column player. Prior to the game, they meet and agree on a common strategy, i.e. a communication protocol \mathcal{P} . When the game begins, Alice receives an index $1 \leq i \leq m$ and Bob receives an index $1 \leq j \leq n$ each known only to the respective player. They communicate by alternately sending bits to each other. By the end of the game, they should both know the value of a_{ij} . The *cost* of a protocol \mathcal{P} is the number of bits being exchanged between Alice and Bob for the worst pair ij . The communication complexity of the matrix A is defined as

$$cc(A) := \min_{\mathcal{P}} \text{cost}(\mathcal{P}).$$

In this definition, the protocol \mathcal{P} is deterministic. However, there are also several similar definitions where Alice and Bob are allowed to use randomization. As usual in randomized computation, chance can affect the computational process in two main ways. (i) The number of bits that are exchanged becomes a random variable, whose distribution we investigate. (ii) The process need not always succeed, and we consider the probability of a correct computation.

The version of communication we now consider is this: The matrix A , and the players Alice and Bob are as above. However, Alice, upon receiving the index i randomly selects a string σ_i of r bits and sends them over to Bob. He inspects σ_i and his index j and chooses at random with some probability $1 \geq p_j \geq 0$ to declare the outcome to be 1. We define, r_0 as the least integer r for which this procedure can be carried out with success probability $> 1/2$. (The emphasis here is on the “greater than” - This is an *unbounded error* computational model). It turns out that $2^{r_0-2} \leq d(A) \leq 2^{r_0}$ [21].

3 Margin complexity and γ_2

3.1 An equivalent definition of margin complexity

Our first step is to find a relation between margin complexity and factorization property - and in particular, the relation between margin complexity and γ_2 . Define the *sign pattern* of a matrix $B \in M_{m,n}(\mathbb{R})$ (denoted by $sp(B)$) as the sign matrix ($sign(b_{ij})$). For a sign matrix A , let $SP(A)$ be the family of matrices B satisfying $b_{ij}a_{ij} \geq 1$ for all i and j . In other words, $SP(A)$ consists of matrices $B = (b_{ij})$ for which $sp(B) = A$ and $|b_{ij}| \geq 1$ for all i, j .

The following lemma gives a simple alternative characterization of the margin complexity of sign matrices.

Lemma 3 *For every $m \times n$ sign matrix A ,*

$$mc(A) = \min_{XY \in SP(A)} \|X\|_{\ell_2 \rightarrow \ell_\infty^m} \|Y\|_{\ell_1^n \rightarrow \ell_2}.$$

Proof Equation (1), and the definition $mc(A) = m(A)^{-1}$, imply that

$$\begin{aligned} mc(A) &= \min_{XY: sp(XY)=A} \max_{i,j} \frac{\|x_i\| \|y_j\|}{|\langle x_i, y_j \rangle|} \\ &= \min_{XY: sp(XY)=A} \max_{i,j} \frac{1}{\left| \langle \frac{x_i}{\|x_i\|}, \frac{y_j}{\|y_j\|} \rangle \right|}, \end{aligned}$$

which is equivalent to

$$mc(A) = \min_{i,j} \max \frac{1}{|\langle x_i, y_j \rangle|},$$

where the minimum is over all pairs of matrices X, Y such that

1. A is the sign pattern of XY , i.e. $sp(XY) = A$.
2. The rows of X and the columns of Y are unit vectors.

Given such X and Y , let us define \tilde{Y} to be $\frac{1}{\min_{i,j} |\langle x_i, y_j \rangle|} Y$ (so that all entries in $X\tilde{Y}$ have absolute value ≥ 1). We can now interpret the above definition as saying that $mc(A)$ is the smallest α for which there exist matrices X and \tilde{Y} such that

1. $XY \in SP(A)$,
2. All rows in X are unit vectors,
3. All columns in \tilde{Y} have length α .

In other words,

$$mc(A) = \min \gamma_2(XY),$$

where the minimum is over all pairs of matrices X, Y such that $XY \in SP(A)$ and the rows of X and the columns of Y have equal length. It is easy to see that when the dimension of the vectors is not bounded, the restriction on the vectors' lengths does not affect the optimum and thus

$$mc(A) = \min_{B \in SP(A)} \gamma_2(B),$$

which is equivalent to the assertion of the lemma. ■

Since $A \in SP(A)$, we can easily conclude:

Corollary 4 $mc(A) \leq \gamma_2(A)$.

3.2 An improved lower bound on margin complexity

The following corollary is a simple application of duality and the equivalent definition of margin complexity given in the last section.

Corollary 5 *Let A be an $m \times n$ sign matrix. Then,*

$$mc(A) \geq \frac{nm}{\gamma_2^*(A^t)}, \quad (2)$$

and in particular,

$$mc(A) \geq \frac{nm}{K_G \|A\|_{\ell_\infty^n \rightarrow \ell_1^m}}.$$

Proof Let B be a matrix in $SP(A)$ such that $mc(A) = \gamma_2(B)$. Then,

$$mc(A)\gamma_2^*(A^t) = \gamma_2(B)\gamma_2^*(A^t) \geq \langle A, B \rangle \geq nm.$$

Hence $mc(A) \geq nm/\gamma_2^*(A^t)$. By Grothendieck's inequality,

$$\gamma_2^*(A^t) \leq K_G \|A^t\|_{\ell_\infty^m \rightarrow \ell_1^n} = K_G \|A\|_{\ell_\infty^n \rightarrow \ell_1^m}.$$

■

To compare our bound on the margin complexity with known results, we need to understand the relationship between γ_2 and the trace norm $\|A\|_{tr}$. This is the sum of A 's singular values, which are the roots of the eigenvalues of AA^t . We prove:

Lemma 6 *For every $m \times n$ matrix A , $\frac{1}{\sqrt{mn}}\|A\|_{tr} \leq \gamma_2(A)$.*

Since the trace norm and the ℓ_2 operator norm are dual (see e.g. [26]), this is equivalent to:

Lemma 7 *For every $m \times n$ matrix A , $\gamma_2^*(A) \leq \sqrt{mn}\|A\|_{2 \rightarrow 2}$.*

Proof Let B be a real $m \times n$ matrix satisfying $\gamma_2(B) \leq 1$. Let $XY = B$ be a factorization of B such that $\|Y\|_{1 \rightarrow 2} \leq 1$ and $\|X\|_{2 \rightarrow \infty} \leq 1$. Denote by x_i the i th column of X and by y_i^t the i th row of Y . For every matrix A

$$\begin{aligned} \langle B, A \rangle &= \langle XY, A \rangle = \left\langle \sum x_i y_i^t, A \right\rangle \\ &= \sum x_i^t A y_i = \sum \|x_i\| \|y_i\| \frac{x_i^t}{\|x_i\|} A \frac{y_i}{\|y_i\|} \\ &\leq \|A\|_{2 \rightarrow 2} \sum \|x_i\| \|y_i\| \leq \|A\|_{2 \rightarrow 2} \sqrt{\left(\sum \|x_i\|^2\right) \left(\sum \|y_i\|^2\right)} \leq \sqrt{mn} \|A\|_{2 \rightarrow 2}. \end{aligned}$$

It follows that

$$\gamma_2^*(A) = \max_{\gamma_2(B) \leq 1} \langle B, A \rangle \leq \sqrt{mn} \|A\|_{2 \rightarrow 2}.$$

■

Corollary 5 improves a bound proved by Forster [7]. Forster proved that for any $m \times n$ matrix,

$$mc(A) \geq \frac{\sqrt{mn}}{\|A\|_{\ell_2^n \rightarrow \ell_2^m}}.$$

That this is indeed an improvement follows from Lemma 7. It may sometime yield an asymptotically better bound, since we next exhibit $n \times n$ sign matrices A for which $\|A\|_{\ell_\infty^n \rightarrow \ell_1^n} \ll n \|A\|_{\ell_2^n \rightarrow \ell_2^n}$. For other extensions of Forster's bound see [9].

Consider the matrix A where in the upper left block of $n^{3/4} \times n^{3/4}$ all entries are one. All other entries of A are ± 1 chosen uniformly and independently. Let B be the matrix with an $n^{3/4} \times n^{3/4}$ block of ones in the upper left corner and zeros elsewhere. Then

$$\|A\|_{2 \rightarrow 2} \geq \|B\|_{2 \rightarrow 2} = n^{3/4}.$$

Now let $C = A - B$. It is not hard to see that with high probability $\|C\|_{\infty \rightarrow 1} \leq O(n^{3/2})$. Indeed, this easily follows from Lemma 12 below. Also, $\|B\|_{\infty \rightarrow 1} = n^{3/2}$. By the triangle inequality

$$\|A\|_{\infty \rightarrow 1} \leq \|C\|_{\infty \rightarrow 1} + \|B\|_{\infty \rightarrow 1} \leq O(n^{3/2}).$$

Thus $\|A\|_{\infty \rightarrow 1} \leq O(n^{3/2})$ whereas $n\|A\|_{2 \rightarrow 2} \geq \Omega(n^{7/4})$.

3.3 Computing the Optimal Margin

In this section we observe that the margin complexity and γ_2 can be computed in polynomial time, using semi-definite programming. As we show later, this has some nice theoretical consequences as well.

We start with the semi-definite programs for the margin complexity and for γ_2 . To that end, it is often convenient to identify the vector space of all $n \times n$ symmetric matrices with the Euclidean space \mathbb{R}^m where $m = n(n+1)/2$. Denote the cone of all $n \times n$ positive semi-definite matrices by PSD_n . Let A be an $n \times N$ $\{\pm 1\}$ -valued matrix, and let E_{ij} be the $(n+N) \times (n+N)$ symmetric matrix with $e_{i,(n+j)} = e_{(n+j),i} = a_{ij}$ for $i = 1, \dots, n$ and $j = 1, \dots, N$, and all other entries zero. Observe that the optimum of the following optimization problem is $mc(A)$:

$$\begin{aligned} & \text{minimize} && \eta \\ & \forall i && \eta \geq X_{ii} \\ & \forall i, j && \langle E_{ij}, X \rangle \geq 1 \\ & && X \in PSD_{n+N} \end{aligned} \tag{3}$$

Indeed, since X is positive semi-definite, it can be expressed as $X = YY^t$ for some matrix Y . Express Y in block form as $\begin{pmatrix} B \\ C \end{pmatrix}$, where B has n rows and C has N rows. The constraints of type (3) state that $sp(BC^t) = A$ and that all the entries of BC^t are at least 1 in absolute value. The diagonal entries of X are the squared lengths of the rows in B and C , from which the claim about the optimum follows.

Likewise, consider a slight modification of this program, by replacing Condition (3) with $\langle E_{ij}, X \rangle = 1$ for all i, j . The optimum of the modified program is $\gamma_2(A)$.

Recall that an appropriate adaptation of the ellipsoid algorithm solves positive semidefinite programs to any desirable accuracy in polynomial time. Consequently, the margin complexity and γ_2 of any matrix can be approximated to any degree in polynomial time.

Aside from the algorithmic implications, there is more to be gained by expressing margin complexity as the optimum of a positive definite program, by incorporating SDP duality. Specifically, duality yields the following equivalent definition for margin complexity:

$$mc(A) = \max_{\gamma_2^*(X)=1, sp(X)=A} \langle X, A \rangle. \quad (4)$$

This definition entails that equality in (2) is always achieved by some matrix.

4 Relations with Rank and the Minimal Dimension of a Realization

It is obvious that for every $m \times n$ sign matrix $d(A) \leq \text{rank}(A)$. On the other hand, the gap can be arbitrarily large as we now observe. For a sign matrix A , denote by $s(A)$, the maximum number of sign-changes in a row of A . (The number of sign-changes in a sign vector (a_1, \dots, a_n) , is the number of indices i such that $a_i = -a_{i+1}$, $1 \leq i \leq n-1$.)

Theorem 8 ([2]) *For any sign matrix A , $d(A) \leq s(A) + 1$.*

Thus, for example the matrix $2I_n - J_n$ has rank n and can be realized in \mathbb{R}^2 . Also, it follows easily from the Johnson-Lindenstrauss lemma [13] that $d(A) \leq O(mc(A)^2 \log(n+m))$, see e.g. [3] for details.

What may be more surprising is that $\gamma_2^2(A) \leq \text{rank}(A)$. The proof of this inequality is well known to Banach spaces theorists, and we include it for the sake of completeness.

Lemma 9 *For every matrix $A \in M_{m,n}(\mathbb{R})$,*

$$\gamma_2^2(A) \leq \|A\|_{\ell_1^n \rightarrow \ell_\infty^m} \text{rank}(A).$$

In particular, if A is a sign matrix then $\gamma_2^2(A) \leq \text{rank}(A)$.

Proof Consider factorizations of A of the form $A = XYA$, where $XY = I$, the identity $m \times m$ matrix, then

$$\begin{aligned} \gamma_2(A) &\leq \|X\|_{\ell_2 \rightarrow \ell_\infty^m} \|YA\|_{\ell_1^n \rightarrow \ell_2} \\ &\leq \|X\|_{\ell_2 \rightarrow \ell_\infty^m} \|Y\|_{\ell_\infty^m \rightarrow \ell_2} \|A\|_{\ell_1^n \rightarrow \ell_\infty^m}. \end{aligned}$$

In particular, $\frac{\gamma_2(A)}{\|A\|_{\ell_1^n \rightarrow \ell_\infty^m}} \leq \min_{XY=I} \|X\|_{\ell_2 \rightarrow \ell_\infty^m} \|Y\|_{\ell_\infty^m \rightarrow \ell_2}$. A formulation of the well-known John's theorem ([12]), states that for any d -dimensional norm E , it is possible to find two matrices X and Y with $XY = I$, and

$$\|X\|_{\ell_2 \rightarrow E} \|Y\|_{E \rightarrow \ell_2} \leq \sqrt{d}.$$

If we consider $E \subseteq \ell_\infty^m$ given by $\text{range}(A)$ - that is, the vector space $\text{range}(A)$ endowed with the norm whose unit ball is $[-1, 1]^m \cap \text{range}(A)$ - then by John's theorem our assertion holds. \blacksquare

It is known that $d(A) = \Omega(n)$ for almost all $n \times n$ sign matrices [1]. This is in line with the principle that random instances tend to have high complexity. We also encounter here the other side of the complexity coin in that it is a challenging open problem to construct explicit $n \times n$ sign matrices A with $d(A) \geq \Omega(n)$. In a recent work, Forster [7] shows that $d \geq \Omega(\sqrt{n})$ for Sylvester matrices. This follows from the following lemma.

Lemma 10 ([7]) *For every $n \times m$ sign matrix A*

$$d(A) \geq \frac{\sqrt{nm}}{\|A\|_{2 \rightarrow 2}}.$$

We prove the following improvement of Forster's bound. As we saw in section 3.2, this improvement can be significant.

Lemma 11 *For every $m \times n$ sign matrix A*

$$d(A) \geq \frac{nm}{\gamma_2^*(A)},$$

and in particular,

$$d(A) \geq \frac{nm}{K_G \|A\|_{\infty \rightarrow 1}}.$$

Proof It was shown in [7] that for any $m \times n$ sign matrix A there exists a matrix B such that $sp(B) = A$, and $\sum_{i,j} |b_{ij}| \geq \frac{nm\gamma_2(B)}{d(A)}$. Thus

$$\begin{aligned} \gamma_2^*(A) &= \max_{B: \gamma_2(B)=1} \langle A, B \rangle \\ &\geq \max_{B: \gamma_2(B)=1, sp(B)=A} \langle A, B \rangle \\ &= \max_{B: \gamma_2(B)=1, sp(B)=A} \sum_{i,j} |b_{ij}| \\ &\geq nm/d(A). \end{aligned}$$

Applying Grothendieck's inequality it follows that

$$d(A) \geq \frac{nm}{K_G \|A\|_{\infty \rightarrow 1}},$$

as claimed. \blacksquare

Other variant of the above bound can also be proved using the same line of proof. For example, by starting with $\gamma_2^*(\tilde{A})$ where \tilde{A} is any matrix such that $sp(\tilde{A}) = A$, it follows that

$$d(A) \geq \frac{nm}{\gamma_2^*(\tilde{A})} \min_{i,j} |\tilde{A}_{ij}|.$$

This improves a bound from [8].

5 Typical Values and Concentration of Measure

As we have already mentioned, almost every $n \times n$ sign matrix has rank n [11], and cannot be realized in dimension $o(n)$. In [4], it was shown that the margin complexity of a sign matrix is almost always $\Omega(\sqrt{n/\log n})$. Here we improve this result and show that the margin complexity and γ_2 of sign matrices are almost always $\Theta(\sqrt{n})$. We also prove that γ_2 is concentrated around its mean. The following lemma is well known, and we include its proof for completeness.

Lemma 12

$$\Pr(\|A\|_{\ell_\infty^n \rightarrow \ell_1^m} \leq 2mn^{1/2}) \geq 1 - (e/2)^{-2m},$$

where $m \geq n$, and the matrix A is drawn uniformly from among the $m \times n$ sign matrices.

Proof Recall that $\|A\|_{\ell_\infty^n \rightarrow \ell_1^m} = \max_{\|x\|_{\ell_\infty^n} = 1} \|Ax\|_{\ell_1^m}$, and the maximum is attained at the extreme points of the ℓ_∞^n unit ball, i.e., the vectors in $\{-1, 1\}^n$. For $x \in \{-1, 1\}^n$ and $y \in \{-1, 1\}^m$ let

$$Z_{x,y} = \sum_{i=1}^m \sum_{j=1}^n A_{ij} x_j y_i.$$

The distribution of $Z_{x,y}$ is clearly independent of x and y , therefore we can take $x = (1, \dots, 1)$ and $y = (1, \dots, 1)$, and conclude

$$\Pr(Z_{x,y} \geq t) = \Pr\left(\sum_{i,j} A_{i,j} \geq t\right) \leq \exp\left(-\frac{t^2}{2mn}\right).$$

Hence,

$$\Pr(\|A\|_{\ell_\infty^n \rightarrow \ell_1^m} \geq t) \leq 2^{m+n} \exp\left(-\frac{t^2}{2mn}\right),$$

and taking $t = 2m\sqrt{n}$ completes the proof. \blacksquare

Combining this lemma and the connection between $\gamma_2(A)$ and $\|A\|_{\ell_\infty^n \rightarrow \ell_1^m}$ (see Corollary 5) we obtain the following

Corollary 13

$$\Pr(\gamma_2(A) \geq c\sqrt{n}) \geq 1 - (e/2)^{-2m}$$

and

$$\Pr(mc(A) \geq c\sqrt{n}) \geq 1 - (e/2)^{-2m}.$$

Here $c > 0$ is an absolute constant, $m \geq n$, and the matrix A is drawn uniformly from among the $m \times n$ sign matrices.

For the concentration of measure we use the following theorem by Talagrand [25] (see also [2], Chapter 7).

Theorem 14 Let $\Omega_1, \Omega_2, \dots, \Omega_n$ be probability spaces. Denote by Ω their product space, and let $\mathcal{A}, \mathcal{B} \subset \Omega$. Suppose that for each $B \in \mathcal{B}$, there is a real vector $\alpha \in \mathbb{R}^n$ such that

$$\sum_{i: A_i \neq B_i} \alpha_i \geq c \|\alpha\|_2$$

for every $A \in \mathcal{A}$. Then $\Pr(\mathcal{A}) \Pr(\mathcal{B}) \leq e^{-c^2/4}$.

Lemma 15 Let $m \geq n$ and let A be a random $m \times n$ sign matrix. Denote by m_γ the median of γ_2 , then

$$\Pr(|\gamma_2(A) - m_\gamma| \geq c) \leq 4e^{-c^2/16}$$

Proof Consider the sets $\mathcal{A} = \{A : \gamma_2(A) \leq m_\gamma - c\}$ and $\mathcal{B} = \{B : \gamma_2(B) \geq m_\gamma\}$ of $m \times n$ sign matrices. For each $B \in \mathcal{B}$, there is a matrix β such that $\gamma_2^*(\beta) = 1$ and $\langle \beta, B \rangle = m_\gamma$, whereas $\langle \beta, A \rangle \leq m_\gamma - c$ for every $A \in \mathcal{A}$. Also, $1 = \gamma_2^*(\beta) \geq \|\beta\|_{\ell_2}$, where $\|\cdot\|_{\ell_2}$ is the ℓ_2 norm in \mathbb{R}^{nm} also known as the Hilbert-Schmidt norm. It follows that

$$c \leq \langle \beta, B - A \rangle \leq 2 \sum_{i,j: A_{ij} \neq B_{ij}} |\beta_{ij}|.$$

In order to apply Theorem 14, define the matrix α via $\alpha_{ij} = |\beta_{ij}|$. Then $\|\alpha\|_{\ell_2} = \|\beta\|_{\ell_2} \leq 1$, as needed. It follows that

$$\Pr(\mathcal{A}) \leq 2e^{-c^2/16}.$$

That $\Pr(\{A : \gamma_2(A) \geq m_\gamma + c\}) \leq 2e^{-c^2/16}$ is proved equivalently by taking $\mathcal{A} = \{A : \gamma_2(A) \leq m_\gamma\}$, and $\mathcal{B} = \{B : \gamma_2(B) \geq m_\gamma + c\}$. \blacksquare

We are also able to give a measure concentration estimate for the left tail of γ_2 as follows.

Lemma 16 Let $m \geq n$ and let A be a random $m \times n$ sign matrix. Denote by M the median of $\gamma_2^*(A)$, and let $m_M = nm/M$, then

$$\Pr(\gamma_2(A) \leq m_M - c/\sqrt{m}) \leq 2e^{-c^2/16}$$

Proof Let $\mathcal{A} = \{A : \gamma_2(A) \leq m_M - c\}$ and $\mathcal{B} = \{B : \gamma_2^*(B) \leq M\}$ be sets of $m \times n$ sign matrices. Pick $B \in \mathcal{B}$, and let $\beta = B/\gamma_2^*(B)$, then $\langle \beta, B \rangle = \langle B, B \rangle / \gamma_2^*(B) \geq nm/M = m_M$, whereas $\langle \beta, A \rangle \leq m_M - c$ for every $A \in \mathcal{A}$. It follows that

$$c \leq \langle \beta, B - A \rangle \leq 2 \sum_{i,j: A_{ij} \neq B_{ij}} |\beta_{ij}|.$$

In addition, it is known that $\gamma_2^*(B) \geq m\sqrt{n}$, which implies that

$$1/\sqrt{m} \geq \sqrt{mn}/\gamma_2^*(B) = \|B\|_{\ell_2}/\gamma_2^*(B) = \|\beta\|_{\ell_2}.$$

Define the matrix $\alpha_{ij} = |\beta_{ij}|$, then $\|\alpha\|_{\ell_2} = \|\beta\|_{\ell_2} \leq 1/\sqrt{m}$. It follows that

$$\sum_{i,j: A_{ij} \neq B_{ij}} \alpha_{ij} \geq c/2 \geq \frac{c\sqrt{m}\|\alpha\|_{\ell_2}}{2}.$$

Applying Theorem 14, we get

$$\Pr(\mathcal{A}) \Pr(\mathcal{B}) \leq e^{-c^2 m/16}.$$

Since $\Pr(\mathcal{B}) = 1/2$ the statement follows. \blacksquare

It is not clear how good the last estimate is. Note that both m_γ and m_M are of order \sqrt{n} so obviously, for relatively large c , the last lemma gives stronger lower tail estimates than Lemma 15. What is the exact relation between m_γ and m_M is not clear to us. Also note that since the trace norm is \sqrt{n} -Lipschitz it can be shown that $\Pr(|\frac{1}{\sqrt{nm}}\|A\|_{tr} - m_{tr}| \geq c/\sqrt{n}) \leq 4e^{-c^2/16}$ using the same method as in the proof of Lemma 15, where m_{tr} is the median of $\frac{1}{\sqrt{nm}}\|A\|_{tr}$. Thus, if $m_M \leq m_{tr}$ the estimate in Lemma 16 is trivial.

In light of the above discussion it would be interesting to know where the medians of γ_2 , γ_2^* and the trace norm lie. Moreover our choice of \mathcal{B} in the proof of Lemma 16 may not be optimal, and it is interesting what the best choice is. More related questions are raised in Section 8.

6 Specific Examples

As usual, much of our insight for such a new set of parameters stems from an acquaintance with specific examples. Examples also suggest interesting challenges to the development of new methods and bounds. An interesting case in point is the determination in [9], of the exact margin complexity of the identity matrix and the triangular matrix. Here we determine the complexity of several more families of matrices.

6.1 Hadamard matrices, and highly imbalanced matrices

Consider $m \times n$ sign matrices, with $m \geq n$. It is easy to see that in this case $1 \leq \gamma_2(A) \leq \sqrt{n}$ and by duality that also $m\sqrt{n} \leq \gamma_2^*(A) \leq nm$. It follows from Corollary 5 and Lemma 7 that a matrix whose columns are orthogonal have the largest possible margin complexity, \sqrt{n} . In particular Hadamard matrices have the largest possible margin complexity. At the other extreme, a sign matrix A satisfies $\gamma_2(A) < \sqrt{2}$ if and only if it has rank 1. This is because a sign matrix has rank 1 if and only if it does not contain a 2×2 Hadamard matrix. Next we prove a useful upper bound on γ_2 for sign matrices.

For a real valued $m \times n$ matrix A it is easy to show that $\gamma_2(A) \leq \|A\|_{1 \rightarrow 2}$ as well as $\gamma_2(A) \leq \|A\|_{2 \rightarrow \infty}$. These follow from the trivial factorizations $A = IA$ (resp. $A = AI$), with I the $m \times m$ (resp. $n \times n$) identity matrix. This is not particularly helpful for sign matrices where it yields the same trivial bound

$\gamma_2(A) \leq \min(\sqrt{m}, \sqrt{n})$ for all sign matrices. This bound does provide useful estimates for 0, 1 matrices with only few 1s in every column. These bounds can, in turn, be applied to sign matrices as well, as we now show. Let $J_{m,n}$ be the all-ones $m \times n$ matrix. It has rank 1, so $\gamma_2(J_{m,n}) = 1$. For A a real $m \times n$ matrix, let $T(A) = (A + J_{m,n})/2$. Clearly, T maps sign matrices to 0, 1 matrices. Also, the inverse of T is $T^{-1}(B) = 2B - J_{m,n}$. Since γ_2 is a norm on $M_{m \times n}(\mathbb{R})$, the following holds for every sign matrix A ,

$$\begin{aligned} \gamma_2(A) &= \gamma_2(T^{-1}(T(A))) = \gamma_2(2T(A) - J_{m,n}) \\ &\leq 2\gamma_2(T(A)) + \gamma_2(J_{m,n}) = 2\gamma_2(T(A)) + 1. \end{aligned}$$

Thus, if we denote by $N_c(A)/N_r(A)$, the largest number of 1s in any column/row of a sign matrix A , then

$$mc(A) \leq \gamma_2(A) \leq 2 \min\{\sqrt{N_c(A)}, \sqrt{N_r(A)}\} + 1. \quad (5)$$

Notice that all the complexity measures under consideration here are invariant under sign reversals of rows or columns of a matrix. This can sometimes be incorporated to the above argument.

We can now determine the margin complexity of the following matrix up to a factor of 2. For $n \geq 2d$, let D be the $n \times \binom{n}{d}$ sign matrix whose columns are all the sign vectors with exactly d 1s. Inequality (5) implies $mc(D) \leq \gamma_2(D) \leq 2\sqrt{d} + 1$. On the other hand, the margin complexity of a matrix is at least as large as the margin complexity of any of its submatrices. D contains as a submatrix the $d \times 2^d$ matrix in which every sign vector of length d appear as a column. Since the rows in this matrix are orthogonal, $mc(D) \geq \sqrt{d}$.

6.2 Adjacency matrices of highly expanding graphs

We show that $\gamma_2(A) = \Theta(\sqrt{d})$, when A is the adjacency matrix of d -regular highly expanding (or “nearly Ramanujan”) graphs. Let $G(V, E)$ be a graph with vertex set $V = \{v_1, \dots, v_n\}$. The adjacency matrix $A = (a_{ij})$ of G is the symmetric 0, 1 matrix with $a_{ij} = a_{ji} = 1$ if and only if $(v_i, v_j) \in E$. A graph in which every vertex has exactly d neighbors is called d -regular. In this case, there are exactly d 1’s in every row and column of the adjacency matrix A . Let us denote the singular values of A (i.e. the absolute value of its eigenvalues) by $s_1 \geq \dots \geq s_n$. It is easy to verify that $s_1 = d$, and an inequality of Alon and Boppana [19] says that $s_2 \geq 2\sqrt{d-1} - o(1)$. It was recently shown by Friedman [10] that for every $\epsilon > 0$, almost all d -regular graphs satisfy $s_2 \leq 2\sqrt{d-1} + \epsilon$. Graphs with $s_2 \leq 2\sqrt{d-1}$ exist [17, 18] and are called Ramanujan graphs. By Inequality (5) $\gamma_2(A) \leq \sqrt{d}$ for the adjacency matrix of every d -regular graph. We observe that for nearly Ramanujan graphs, the reverse inequality holds.

Claim 17 *Let G be a d -regular graph on n vertices, and let A be its adjacency matrix. If $s_2 \leq c\sqrt{d}$, then*

$$\|A\|_{tr} \geq c^{-1} \left(n\sqrt{d} - d^{3/2} \right).$$

Proof $nd = \text{tr}(AA^t) = \sum_{i=1}^n s_i^2$. Therefore, $\sum_{i=2}^n s_i^2 = nd - d^2$. It follows that

$$\sum_{i=1}^n |s_i| \geq \frac{1}{s_2} \sum_{i=2}^n s_i^2 \geq c^{-1} (n\sqrt{d} - d^{3/2}).$$

■

The following is an immediate corollary of the above claims and of Lemma 6.

Corollary 18 *Let A be the adjacency matrix of a d -regular graph on n vertices, with $d \leq n/2$. If $s_2 \leq c\sqrt{d}$ then $\gamma_2(A) = \Theta(\sqrt{d})$.*

7 A Gap Between the Margin and γ_2

Let $m = 3^k$ and $n = 2^k$, an example of an $m \times n$ matrix with a large gap between the margin complexity and the trace norm normalized by \sqrt{nm} was given in [9]. The fact that γ_2 may be significantly larger than the margin complexity for square matrices, was, at least for some of us somewhat unexpected. In this section we present such examples. We begin with a specific example, and then we present a technique to generate many matrices with a large gap.

7.1 An Example

Let n be an odd integer and let K be the $n \times 2^n$ sign matrix with no repeated columns. Denote $A = \text{sign}(K^t K)$. We consider the rows and columns of A as indexed by vectors in $\{\pm 1\}^n$, and interpret the rows of A as functions from $\{\pm 1\}^n$ to $\{\pm 1\}$. The row indexed by the vector $(1, 1, \dots, 1)$ corresponds to the majority function, which we denote by f . The row indexed by $s \in \{\pm 1\}^n$ corresponds to the function f_s , given by $f_s(x) = f(x \circ s)$ for all $x \in \{\pm 1\}^n$. Here $s \circ x$, is the Hadamard (or Schur) product of s and x , i.e. the vector $(s_1 x_1, s_2 x_2, \dots, s_n x_n)$. We now show how to express the eigenvalues of A by the Fourier coefficients of the majority function. This is subsequently used to estimate the trace norm of A . As we will see the trace norm of A is large, and thus $\gamma_2(A)$ is large as a consequence from Lemma 6.

Claim 19 *Denote by H_n the $2^n \times 2^n$ Sylvester-Hadamard¹ matrix. Then $AH_n = SH_n$ where S is a diagonal matrix with diagonal entries $2^n \hat{f}(t)$, $t \in \{\pm 1\}^n$.*

Proof Denote by χ_s a character.

$$\begin{aligned} \sum_{\eta} f_s(\eta) \chi_t(\eta) &= \sum_{\eta} f(s \circ \eta) \chi_t(\eta) \\ &= \sum_{\bar{\eta} \circ s} f(\bar{\eta}) \chi_t(\bar{\eta} \circ s) = \sum_{\bar{\eta} \circ s} f(\bar{\eta}) \chi_t(\bar{\eta}) \chi_t(s) \\ &= \chi_t(s) \sum_{\eta} f(\eta) \chi_t(\eta) = \chi_t(s) 2^n \hat{f}(t). \end{aligned}$$

¹ Also known as the Walsh matrix.

Thus χ_t is an eigenvector of A to the eigenvalue $2^n \hat{f}(t)$. ■

Claim 20 *Let $n = 2k + 1$ and let the vector $t \in \{\pm 1\}^n$ have $m - 1$ entries. If m is even then $\hat{f}(t) = 0$, and if $m = 2r + 1$ then*

$$2^n \hat{f}(t) = 2 \sum_{i=0}^{2r} (-1)^i \binom{2r}{i} \binom{2k - 2r}{k - i}$$

Proof Denote by S_q the sum $S_q = \sum_{i=0}^q \binom{n-m}{i}$, and its complement by $\overline{S}_q = 2^{n-m} - S_q$. Denote by Φ_i the subset of vectors in $\{\pm 1\}^n$ with exactly i -1 s that agree with the -1 s in t .

$$\begin{aligned} 2^n \hat{f}(t) &= \sum_{\eta} f(\eta) \chi_t(\eta) = \sum_{i=0}^m \sum_{\eta \in \Phi_i} f(\eta) (-1)^i \\ &= \sum_{i=0}^m \binom{m}{i} (-1)^i \sum_{\eta \in \Phi_i} f(\eta) = \sum_{i=0}^m \binom{m}{i} (-1)^i (S_{k-i} - \overline{S}_{k-i}) \\ &= \sum_{i=0}^m \binom{m}{i} (-1)^i (S_{k-i} + S_{k-i} - 2^{m-n}) \\ &= \sum_{i=0}^m \binom{m}{i} (-1)^i 2S_{k-i} \end{aligned}$$

The last equality follows from $\sum_{i=0}^m (-1)^i \binom{m}{i} = 0$. Using the identity $\sum_{i=0}^t \binom{m}{i} (-1)^i = \binom{m-1}{t}$, we can write the last formula as follows

$$2 \sum_{i=0}^{m-1} (-1)^i \binom{m-1}{i} \binom{n-m}{k-i}.$$

Notice that $(k - (m - 1) + i) + (k - i) = 2k + 1 - m = n - m$, so if m is even the sum is 0. If $m = 2r + 1$ is odd we can write the expression as

$$2 \sum_{i=0}^{2r} (-1)^i \binom{2r}{i} \binom{2k - 2r}{k - i},$$

which concludes the proof. ■

Claim 21 *Let r and k be integers then,*

$$\sum_{i=0}^{2r} (-1)^i \binom{2r}{i} \binom{2k - 2r}{k - i} = (-1)^r (k - r)! (2r)! / r! k! \binom{2k - 2r}{k - r}$$

Proof Dividing both sides by the right side we get

$$\sum_{i=0}^{2r} (-1)^{r-i} \binom{k}{i} \binom{k}{2r-i} = \binom{k}{r}$$

We prove this by induction on k . If $k = 0$ it is trivial. For the induction step we write

$$\begin{aligned} \sum_{i=0}^{2r} (-1)^{r-i} \binom{k}{i} \binom{k}{2r-i} &= \sum_{i=0}^{2r} (-1)^{r-i} \binom{k-1}{i} \binom{k-1}{2r-i} \\ &+ \sum_{i=0}^{2r} (-1)^{r-i} \binom{k-1}{i} \binom{k-1}{2r-i-1} \\ &+ \sum_{i=0}^{2r} (-1)^{r-i} \binom{k-1}{i-1} \binom{k-1}{2r-i} \\ &+ \sum_{i=0}^{2r} (-1)^{r-i} \binom{k-1}{i-1} \binom{k-1}{2r-i-1} \end{aligned}$$

The second term is equal to $\sum_{i=0}^{2r-1} (-1)^{r-i} \binom{k-1}{i} \binom{k-1}{2r-i-1}$. By substituting $j = i-1$ the third term is $-\sum_{j=0}^{2r-1} (-1)^{r-i} \binom{k-1}{j} \binom{k-1}{2r-1-j}$. So the second and the third terms cancel each other. By the induction hypothesis the first term is $\binom{k-1}{r}$, and the fourth term is $\binom{k-1}{r-1}$ (again by substituting $j = i-1$). Summing the four terms we get $\binom{k}{r}$. ■

The trace norm of A is thus given by

$$2 \sum_{r=0}^k \binom{n}{2r+1} (k-r)!(2r)!/r!k! \binom{2k-2r}{k-r}$$

which is equal to

$$2 \sum_{r=0}^k \frac{(2k+1)!}{(2r+1)k!r!(k-r)!}$$

For a lower bound on this sum, we estimate the $\frac{k}{2}$ th term

$$\begin{aligned} \frac{(2k+1)!}{(k+1)k! \frac{k}{2}! \frac{k}{2}!} &= \frac{(2k+1)!k!}{(k+1)!k! \frac{k}{2}! \frac{k}{2}!} \\ &= \binom{2k+1}{k} \binom{k}{k/2} \\ &= \Omega(8^k/k) \end{aligned}$$

We conclude that for every n there is an $n \times n$ sign matrix A for which $mc(A) = \log n$ and $\gamma_2(A) = \Theta(\frac{\sqrt{n}}{\log n})$.

7.2 Using Tensor Products

It is also possible to construct matrices with a gap between γ_2 and the margin, using tensor products. Suppose we can find a matrix A with $mc(A)$ strictly smaller than $\|A\|_{tr}$, where $\|\cdot\|_{tr}$ denotes the trace norm of A . Given such a $k \times k$ matrix A , consider the n -fold tensor product of A with itself, i.e. consider the matrix $\bar{A} = \otimes^n A$. By Lemma 22, $mc(\bar{A})$ is significantly smaller than $\frac{1}{k^n} \|(\bar{A})\|_{tr}$, which is smaller than $\gamma_2(\bar{A})$.

Lemma 22 *For every two matrices A and B ,*

$$mc(A \otimes B) \leq mc(A) \cdot mc(B) \quad (6)$$

$$\|A \otimes B\|_{tr} = \|A\|_{tr} \|B\|_{tr} \quad (7)$$

Proof The proof is based on the simple observation that $(A \otimes B)(C \otimes D) = AC \otimes BD$ (see e.g. [5]). To prove the inequality (6), consider optimal factorizations $X_1 Y_1$ and $X_2 Y_2$, for A and B respectively. Then $(X_1 \otimes X_2)(Y_1 \otimes Y_2) = X_1 Y_1 \otimes X_2 Y_2$ is a factorization for $A \otimes B$, which proves the inequality.

To prove the identity (7), observe that $(A \otimes B)(A \otimes B)^t = (A \otimes B)(A^t \otimes B^t) = AA^t \otimes BB^t$. Now let a be an eigenvector of AA^t with eigenvalue μ_1 and b an eigenvector of BB^t with eigenvalue μ_2 , then $a \otimes b$ is an eigenvector of $AA^t \otimes BB^t$ with eigenvalue $\mu_1 \mu_2$. ■

8 Other problems

It should be clear that this is only the beginning of a new reserach direction, and the unresolved questions are numerous. Here are some problems which are directly related to the content of this paper.

- Is the log factor in $d(A) \leq (mc(A))^2 \log n$ necessary ?
- What can be said about the distribution of γ_2 and γ_2^* ?
 - In particular, estimates for their medians are crucial for our discussion in Section 5.
 - Is it true that $\gamma_2 > \sqrt{n-1}$ for almost every $n \times n$ sign matrix? Note that an affirmative answer would imply that almost every $n \times n$ sign matrix is nonsingular. In view of Inequality 5 and the discussion around it, the best consequence would be that the probability of being singular is $\leq 2^{-O(\sqrt{n})}$, which is weaker than the exponential bound in [11].
- Is there an efficient algorithm to factorize a given sign matrix $A = XY$ with $\|X\|_{2 \rightarrow \infty} \|Y\|_{1 \rightarrow 2} \leq \sqrt{rank(A)}$?
- Compare $\|\cdot\|_{tr}$ as well as $\frac{mn}{\gamma_2^*}$ with the complexity measures in the paper.
- Is there a polynomial-time algorithm to determine $d(A)$ of a sign matrix?

References

- [1] N. Alon, P. Frankl, and V. Rödl. Geometrical realizations of set systems and probabilistic communication complexity. In *Proceedings of the 26th Symposium on Foundations of Computer Science*, pages 277–280. IEEE Computer Society Press, 1985.
- [2] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley, New York, second edition, 2000.
- [3] R. I. Arriaga and S. Vempala. An algorithmic theory of learning: Robust concepts and random projection. In *IEEE Symposium on Foundations of Computer Science*, pages 616–623, 1999.
- [4] S. Ben-David, N. Eiron, and H.U. Simon. Limitations of learning via embeddings in euclidean half-spaces. In *14th Annual Conference on Computational Learning Theory, COLT 2001 and 5th European Conference on Computational Learning Theory, EuroCOLT 2001, Amsterdam, The Netherlands, July 2001, Proceedings*, volume 2111, pages 385–401. Springer, Berlin, 2001.
- [5] R. Bhatia. *Matrix Analysis*. Springer-Verlag, New York, 1997.
- [6] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- [7] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. In *SCT: Annual Conference on Structure in Complexity Theory*, 2001.
- [8] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H.U. Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *Proceedings of the 21st Conference on Foundations of Software Technology and Theoretical Computer Science*, pages 171–182, 2001.
- [9] J. Forster, N. Schmitt, and H.U. Simon. Estimating the optimal margins of embeddings in euclidean half spaces. In *14th Annual Conference on Computational Learning Theory, COLT 2001 and 5th European Conference on Computational Learning Theory, EuroCOLT 2001, Amsterdam, The Netherlands, July 2001, Proceedings*, volume 2111, pages 402–415. Springer, Berlin, 2001.
- [10] J. Friedman. A proof of alon’s second eigenvalue conjecture. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 720–724. ACM Press, 2003.
- [11] J. Komlos, J. Kahn and E. Szemerédi. On the probability that a random ± 1 -matrix is singular. *Journal of the American Mathematical Society*, 8(1):223–240, January 1995.

- [12] F. John. Extremum problems with inequalities as subsidiary conditions. *Studies and essays presented to R. Courant in his 60th birthday*, pages 187–204, 1948.
- [13] W. B. Johnson and J. Lindenstrauss. Extensions of lipshitz mappings into a hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.
- [14] B. Kashin and A. Razborov. Improved lower bounds on the rigidity of hadamard matrices. *Mathematical Notes*, 63(4):471–475, 1998.
- [15] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambride University Press, 1997.
- [16] S. V. Lokam. Spectral methods for matrix rigidity with applications to size-depth tradeoffs and communication complexity. In *IEEE Symposium on Foundations of Computer Science*, pages 6–15, 1995.
- [17] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.
- [18] G. A. Margulis. Explicit constructions of expanders. *Problemy Peredaci Informacii*, 9(4):71–80, 1973.
- [19] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991.
- [20] N. Nisan and A. Wigderson. On rank vs. communication complexity. In *IEEE Symposium on Foundations of Computer Science*, pages 831–836, 1994.
- [21] R. Paturi and J. Simon. Probabilistic communication complexity. *Journal of Computer and System Sciences*, 33:106–123, 1986.
- [22] G. Pisier. *Factorization of linear operators and geometry of Banach spaces*, volume 60 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1986.
- [23] P. Pudlak and V. Rödl. Some combinatorial-algebraic problems from complexity theory. In *Discrete Mathematics*, volume 136, pages 253–279, 1994.
- [24] M. A. Shokrollahi, D. A. Spielman, and V. Stemann. A remark on matrix rigidity. *Information Processing Letters*, 64(6):283–285, 1997.
- [25] M. Talagrand. Concentration of measures and isoperimetric inequalities in product spaces. *Publications Mathematiques de l’I.H.E.S.*, 81:73–205, 1996.

- [26] N. Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals*, volume 38 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Longman Scientific & Technical, Harlow, 1989.
- [27] L.G. Valiant. Graph-theoretic arguments in low level complexity. In *Proc. 6th MFCS*, volume 53, pages 162–176. Springer-Verlag LNCS, 1977.
- [28] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1999.