Learning Complexity vs. Communication Complexity

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

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

Abstract

This paper has two main focal points. We first consider an important class of machine learning algorithms - large margin classifiers, such as Support Vector Machines. The notion of margin complexity *quantifies the extent to which a given class of functions can be learned by large margin classifiers. We prove that up to a small multiplicative constant, margin complexity is equal to the inverse of discrepancy. This establishes a strong tie between seemingly very different notions from two distinct areas.*

In the same way that matrix rigidity is related to rank, we introduce the notion of rigidity of margin complexity. We prove that sign matrices with small margin complexity rigidity are very rare. This leads to the question of proving lower bounds on the rigidity of margin complexity. Quite surprisingly, this question turns out to be closely related to basic open problems in communication complexity, e.g., whether $PSPACE$ can be separated from the polynomial hierarchy in communication complexity.

There are numerous known relations between the field of learning theory and that of communication complexity [6, 9, 25, 15], as one might expect since communication is an inherent aspect of learning. The results of this paper constitute another link in this rich web of relations. This link has already proved significant as it was used in the solution of a few open problems in communication complexity [19, 17, 28].

1 Introduction

Many papers and results link between learning theoretic quantities and communication complexity counterparts. Examples are the characterization of unbounded error communication complexity in terms of dimension complexity of [25], and the equivalence between VC-dimension and one-way distributional complexity with respect to product distributions proved in [15].

We study here the margin complexity of sign matrices. A primary motivation for this study is the desire to understand the strengths and weaknesses of large margin classifiers, such as Support Vector Machines (aka SVM). But as it turned out communication complexity is also an inherent subject of this study, as was the case with previous work on margin complexity and related issues, e.g. [6, 9].

We first describe the learning theoretic point of view, and define margin complexity and then review the relevant background and explain our new results.

A *classification algorithm* receives as input a *sample* $(z_1, f(z_1)), \ldots, (z_m, f(z_m))$ which is a sequence of points $\{z_i\}$ from a set $\mathcal{D}$ (the *domain*) and the corresponding evaluations of some unknown function $f : \mathcal{D} \to \{\pm 1\}$. The output of the algorithm is a function $h : \mathcal{D} \to \{\pm 1\}$, which should be close to f. Here we think of f as chosen by an adversary from a predefined class $\mathcal{F}$ (the so-called *concept class*). (In practical situations the choice of the class $\mathcal{F}$ represents our prior knowledge of the situation.)

Large margin classifiers take the following route to the solution of classification problems: The domain $\mathcal{D}$ is mapped into $\mathbb{R}^t$ (this map is usually called a *feature map*). If z_i is mapped to x_i for each i, our sample points are now $\{x_i\} \subset \mathbb{R}^t$. The algorithm then seeks a linear functional (i.e., a vector) y that maximizes

$$m_f(\{x_i\}, y) = \min_i \frac{|\langle x_i, y\rangle|}{\|x_i\|_2 \|y\|_2}.$$

under the constraint that $sign(\langle x_j, y\rangle) = f(x_j)$, for all j. We denote this maximum by $m_f(\{x_i\})$.

Clearly, an acceptable linear functional y defines a hyperplane H that separates the points (above and below H) as dictated by the function f. What determines the performance of the classifier associated with y is the distances of the points x_i from H, i.e., the *margin* $m_f(\{x_i\}, y)$. For

1093-0159/08 $25.00 © 2008 IEEE
DOI 10.1109/CCC.2008.28

more on classifiers and margins, see [32]. Thus, the margin captures the extent to which the family $\mathcal{F}$ can be described by the sign of a linear functional. We study the margin, in quest of those properties of a concept class that determine how well suited it is for such a description. Large margin classifiers occupy a central place in present-day machine learning in both theory and practice. As we show here, there is an interesting connection between margins of concept classes and complexity theory, in particular with the study of communication complexity.

These considerations lead us to define the *margin of a class of functions*. But before we do that, some words about the feature map are in order. The theory and the practice of the choice of a feature map is at present a subtle art. Making the proper choice of a feature map can have a major impact on the performance of the classification algorithm. Our intention here is to avoid this delicate issue and concentrate instead on the concept class *per se*. In order to bypass the dependence of our analysis on the choice of a feature map, we consider the *best possible* choice. This explains the supremum in the definition of margin below

$$m(\mathcal{F}) = \sup_{\{x_i\}} \inf_{f \in \mathcal{F}} m_f(\{x_i\}).$$

How should we model this setup? For every set of m samples there is only a finite number, say n, of possible classifications by functions from the relevant concept class. Consequently, we can represent a concept class by an $m \times n$ sign matrix, each column of which represents a function $f : [m] \to \{\pm 1\}$. It should be clear then, that the margin of a sign matrix A is

$$m(A) = \sup \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{\|x_i\|_2 \|y_j\|_2}, \qquad (1)$$

where the supremum is over all choices of $x_1, \ldots, x_m, y_1, \ldots, y_n \in \mathbb{R}^{m+n}$ such that $sign(\langle x_i, y_j \rangle) = a_{ij}$, for all i,j. (It is not hard to show that there is no advantage in working in any higher dimensional Euclidean space.)
It is also convenient to define $mc(A) = m(A)^{-1}$, the *margin complexity* of A.

We mention below some previous results and some simple observations on margin complexity. We begin with some very rough bounds:

Observation 1 *For every $m \times n$ sign matrix A,*

$$1 \leq mc(A) \leq \min\{\sqrt{m}, \sqrt{n}\}.$$

The lower bound follows from Cauchy-Schwartz. For the upper bound, assume w.l.o.g that $m \geq n$ and let x_i be the i-th row of A and y_j the j-th vector in the standard basis.

The first paper on margin complexity [6] mainly concerns the case of random matrices. Among other things they proved:

Theorem 2 (Ben-David, Eiron and Simon [6]) *Almost* [1] *every $n \times n$ sign matrix has margin complexity at least $\Omega(\sqrt{\frac{n}{\log n}})$.*

This theorem illustrates the general principle that random elements are complex. A main goal in that paper is to show that $VC - dimension$ and margin complexity are very distinct measures of complexity. E.g.,

Theorem 3 (Ben-David, Eiron and Simon [6]) *Let $d \geq 2$. Almost every matrix with VC-dimension at most $2d$ has margin complexity larger than*

$$\Omega\left(n^{\frac{1}{2} - \frac{1}{2d} - \frac{1}{2^{d+1}}}\right).$$

If $A : U \to V$ is a linear map between two normed spaces, we denote its operator norm by $\|A\|_{U \to V} = \max_{x:\|x\|_U = 1} \|Ax\|_V$, with the shorthand $\|\cdot\|_{p \to q}$ to denote $\|\cdot\|_{\ell_p \to \ell_q}$. A particularly useful instance of this is $\|A\|_{2 \to 2}$ that is equal to the largest singular value of A which can be computed efficiently. Forster [8] proved the following lower bound on margin complexity.

Claim 4 (Forster [8]) *For every $m \times n$ sign matrix A*

$$mc(A) \geq \frac{\sqrt{nm}}{\|A\|_{2 \to 2}}.$$

This result has several nice consequences. For example, it implies that almost every $n \times n$ sign matrix has margin complexity $\Omega(\sqrt{n})$. Also, together with Observation 1 it yields that the margin complexity of an $n \times n$ Hadamard matrix is $\sqrt{n}$. Forster's proof is of interest too, and provides more insight than earlier proofs which were based on counting arguments.

Subsequent papers [9, 10, 11] following [8], improved Forster's bound in different ways. Connections were shown between margin complexity and other complexity measures. These papers also determine exactly the margin complexity of some specific families of matrices.

In [18] we noticed the relation between margin complexity and factorization norms. Given an operator $A : U \to V$ and a normed space W, the factorization problem seeks to express A as $A = XY$, where $Y : U \to W$ and $X : W \to V$, such that X and Y have small operator norms. Of special interest is the case $U = \ell_1^n$, $V = \ell_\infty^m$ and $W = \ell_2$. We denote

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

It is not hard to see that $\|B\|_{1 \to 2}$ is the largest ℓ_2 norm of a column of B, and $\|B\|_{2 \to \infty}$ is the largest ℓ_2 norm of a row

[1] Here and below we adopt a common abuse of language and use the shorthand "almost every" to mean "asymptotically almost every".

of B.
It is proved in [18] that for every $m \times n$ sign matrix A,

$$mc(A) = \min_{B:\ b_{ij}a_{ij} \geq 1\ \forall i,j} \gamma_2(B). \tag{2}$$

This identity turns out to be very useful in the study of margin complexity. Some consequences drawn in [18] are: For every $m \times n$ sign matrix A,

- $mc(A) = \max_{B:sign(B)=A, \gamma_2^*(B) \leq 1} \langle A, B \rangle$.
- $mc(A) \leq \gamma_2(A) \leq \sqrt{rank(A)}$.
- Let $RC(A)$ be the randomized or quantum communication complexity of A, then
$$\log mc(A) \leq RC(A) \leq mc(A).$$

In this paper, we derive another consequence of the relation between margin complexity and γ_2. *Discrepancy* is a combinatorial notion that comes up in many contexts, see e.g. [21, 7]. We prove here that the margin and the discrepancy are equivalent up to a constant factor for every sign matrix. Let A be a sign matrix, and P a probability measure on its entries. We define

$$disc_P(A) = \max_{S,T \subset [n]} \left| \sum_{i \in S, j \in T} p_{ij} a_{ij} \right|.$$

The discrepancy of A is then defined by

$$disc(A) = \min_P disc_P(A).$$

Then

Theorem 5 *For every sign matrix A*

$$\frac{1}{8} m(A) \leq disc(A) \leq 8m(A).$$

Discrepancy is used to derive lower bounds on communication complexity in different models [33, 5], and Theorem 5 constitutes another link in the rich web of relations between the study of margins and the field of communication complexity, see e.g. [6, 9, 19]. As described below, we find in this paper new relations to communication complexity, specifically to questions about separation of communication complexity classes.

It is very natural to consider as well classification algorithms that tolerate a certain probability of error but achieve larger margins. Namely, we are led to consider the following complexity measure

$$mc_r(A, l) = min_{B:h(B,A) \leq l}\ mc(B),$$

where $h(A, B)$ is the Hamming distance between the two matrices. We call this quantity *mc-rigidity*. The relation between this complexity measure and margin complexity is analogous to the relation between *rank-rigidity* and rank. $Rank$-rigidity (usually simply called rigidity) was first defined in [31] and has attracted considerable interest, e.g. [20, 29, 14]. A main motivation to study $rank$-rigidity is that, as shown in [31], the construction of explicit examples of sign matrices with high $rank$-rigidity would have very interesting consequences in computational complexity.

It transpires that mc-rigidity behaves similarly. To begin, it does not seem easy to construct sign matrices with high mc-rigidity (where 'high' means close to the expected complexity of a random matrix). Furthermore, we are able to establish interesting relations between the construction of sign matrices of high mc-rigidity and complexity classes in communication complexity, as introduced and studied in [5, 20].

The mc-rigidity of random matrices is considered in [23, 24]. It is shown there that there is an absolute constant $1 > c > 0$ so that for almost every $n \times n$ sign matrix

$$mc_r(A, cn^2)) \geq \Omega(\sqrt{n}).$$

We give a bound on the number of sign matrices with small mc-rigidity that is much stronger than that of [23, 24]. Our proof is also significantly simpler.

Regarding explicit bounds, we prove the following lower bounds on mc-rigidity

Claim 6 *Every $m \times n$ sign matrix A satisfies*

$$mc_r(A, \frac{mn}{8g}) \geq g,$$

provided that $g < \frac{mn}{2K_G \|A\|_{\infty \to 1}}$.

and

Claim 7 *Every $n \times n$ sign matrix A with $\gamma_2(A) \geq \Omega(\sqrt{n})$ (this is a condition satisfied by almost every sign matrix) satisfies*

$$mc_r(A, cn^2) \geq \Omega(\sqrt{\log n}),$$

for some constant $c > 0$.

In a 1986 paper [5], Babai *et al.*, took a complexity theoretic approach to communication complexity. They defined communication complexity classes analogous to computational complexity classes. For example, the polynomial hierarchy is defined as follows: We define the following classes of $2^m \times 2^m$ $0-1$ matrices. We begin with Σ_0^{cc} the set of combinatorial rectangles and with $\Pi_0^{cc} = co\Sigma_0^{cc}$. From here we proceed to define

$$\begin{aligned}
\Sigma_i^{cc} &= \left\{ A | A = \bigvee_{j=1}^{2^{polylog(m)}} A_j, A_j \in \Pi_{i-1}^{cc} \right\} \\
\Pi_i^{cc} &= \left\{ A | A = \bigwedge_{j=1}^{2^{polylog(m)}} A_j, A_j \in \Sigma_{i-1}^{cc} \right\}.
\end{aligned}$$

For more on communication complexity classes see [5, 20, 16].

Some communication complexity classes were implicitly defined prior to [5], e.g. the communication complexity classes analogous of P, NP and $coNP$. For example, it is known that $P^{cc} = NP^{cc} \cap coNP^{cc}$ [1].

It remains a major open question in this area whether the hierarchy can be separated. We approach this problem using results of Lokam [20] and Tarui [30]. (In our statement of Theorems 8 and 9 we adopt a common abuse of language and speak of individual matrices where we should refer to an infinite family of sign matrices of growing dimensions).

Theorem 8 *Let A be an $n \times n$ sign matrix. If there exists a constant $c \geq 0$ such that for every $c_1 \geq 0$*

$$mc_r(A, n^2/2^{(\log\log n)^c}) \geq 2^{(\log\log n)^{c_1}},$$

then A is not in PH^{cc}.

and

Theorem 9 *An $n \times n$ sign matrix A that satisfies*

$$mc_r(A, n^2/2^{(\log\log n)^c}) \geq 2^{(\log\log n)^{c_1}}$$

for every $c, c_1 \geq 0$, is outside AM^{cc}.

As mentioned, questions about rigidity tend to be difficult, and mc-rigidity seems to follow this pattern as well. However, the following conjecture, if true, would shed some light on the mystery surrounding mc-rigidity:

Conjecture 10 *For every constant c_1 there are constants c_2, c_3 such that every $n \times n$ sign matrix A satisfying $mc(A) \geq c_1\sqrt{n}$ also satisfies:*

$$mc_r(A, c_2 n^2) \geq c_3\sqrt{n}.$$

What the conjecture says is that every matrix with high margin complexity has a high mc-rigidity as well. In particular, explicit examples are known for matrices of high margin complexity e.g. Hadamard matrices. It would follow that such matrices have high mc-rigidity as well.

The rest of this paper is organized as follows. We start with relevant background and notations in Section 2. In Section 3 we prove the equivalence of discrepancy and margin. Section 4 contains the definition of mc-rigidity, the mc-rigidity of random matrices, and applications to the theory of communication complexity classes. In Section 5 we prove lower bounds on mc-rigidity. Open questions are discussed in Section 6.

2 Background and notations

Basic notations Let A and B be two real matrices. We use the following notations:

- The inner product of A and B is denoted $\langle A, B\rangle = \sum_{ij} a_{ij} b_{ij}$.
- Matrix norms: $\|B\|_1 = \sum |b_{ij}|$ is B's ℓ_1 norm, $\|B\|_2 = \sum b_{ij}^2$ is its ℓ_2 (Frobenius) norm, and $\|B\|_\infty = \max_{ij} |b_{ij}|$ is its ℓ_∞ norm.
- If A and B are sign matrices then $h(A, B) = \frac{1}{2}\|A - B\|_1$ denotes the Hamming distance between A and B.

The minimal dimension in which a sign matrix A *can be realized* is defined as

$$d(A) = \min_{B: A = sign(B)} rank(B).$$

(For more about this complexity measure see [25, 8, 6, 11, 9, 18].)

Definition 11 (Discrepancy) *Let A be a sign matrix, and let P be a probability measure on the entries of A. The P-discrepancy of A, denoted $disc_P(A)$, is defined as the maximum over all combinatorial rectangles R in A of $|P^+(R) - P^-(R)|$, where P^+ $[P^-]$ is the measure of the positive entries [negative entries].*

The discrepancy of a sign matrix A, denoted $disc(A)$, is the minimum of $disc_P(A)$ over all probability measures P on the entries of A.

We make substantial use of *Grothendieck's inequality* (see e.g. [26, pg. 64]), which we now recall.

Theorem 12 (Grothendieck's inequality) *There is a universal constant*
$1.5 \leq K_G \leq 1.8$ such that for every real matrix B and every $k \geq 1$

$$\max \sum b_{ij}\langle u_i, v_j\rangle \leq K_G \max \sum b_{ij}\epsilon_i\delta_j. \qquad (3)$$

where the max are over the choice of $u_1, \ldots, u_m, v_1, \ldots, v_n$ as unit vectors in $\mathbb{R}^k$ and $\epsilon_1, \ldots, \epsilon_m, \delta_1, \ldots, \delta_n \in \{\pm 1\}$.

We denote by γ_2^* the dual norm of γ_2, i.e. for every real matrix B

$$\gamma_2^*(B) = \max_{C: \gamma_2(C) \leq 1} \langle B, C\rangle.$$

We note that for any real matrix γ_2^* and $\|\cdot\|_{\infty\to 1}$ are equivalent up to a small multiplicative factor, viz.

$$\|B\|_{\infty\to 1} \leq \gamma_2^*(B) \leq K_G\|B\|_{\infty\to 1}. \qquad (4)$$

The left inequality is easy, and the right inequality is a reformulation of Grothendieck's inequality. Both use the observation that the left hand side of (3) equals $\gamma_2^*(B)$, and the max term on the right hand side is $K_G\|B\|_{\infty\to 1}$.

The norm dual to $\|\cdot\|_{\infty\to 1}$ is the *nuclear norm* from l_1 to l_∞. The nuclear norm of a real matrix B, is defined as follows

$$\nu(B) = \min\{\sum |w_i| \text{ such that } \sum w_i x_i y_i^t = B$$

$$\text{for some choice of sign vectors } x_1, x_2, \ldots, y_1, y_2 \ldots\}.$$

See [12] for more details.

It is a simple consequence of the definition of duality and (4) that for every real matrix B

$$\gamma_2(B) \le \nu(B) \le K_G \cdot \gamma_2(B). \tag{5}$$

3 Margin and discrepancy are equivalent

Here we prove (Recall that $mc(A) = m(A)^{-1}$):

Theorem 13 *For every sign matrix A*

$$\frac{1}{8} m(A) \le disc(A) \le 8m(A).$$

We first define a variant of margin:

Margin with sign vectors: Given an $m \times n$ sign matrix A, denote by $\Lambda = \Lambda(A)$ the set of all pairs of *sign matrices* X, Y such that the sign pattern of XY equals A, i.e., $A = \text{sign}(XY)$ and let

$$m_\nu(A) = \max_{(X,Y)\in\Lambda} \min_{i,j} \frac{|\langle x_i, y_j\rangle|}{\|x_i\|_2 \|y_j\|_2}. \tag{6}$$

Here x_i is the i-th row of X, and y_j is the j-th column of Y. The definitions of m_ν is almost the same as that of margin (Equation 1), except that in defining m_ν we consider only pairs of *sign matrices* X, Y and not arbitrary matrices. It is therefore clear that $m_\nu(A) \le m(A)$ for every sign matrix A. As we see next, the two parameters are equivalent up to a small multiplicative constant.

3.1 Proof of Theorem 13

First we prove that margin and m_ν are equivalent up to a constant factor $K_G < 1.8$, the Grothendieck constant. Then we show that m_ν is equivalent to discrepancy up to a multiplicative factor of at most 4.

Lemma 14 *For every sign matrix A,*

$$K_G^{-1} \cdot m(A) \le m_\nu(A) \le m(A),$$

where K_G is the Grothendieck constant.

Proof The right inequality is an easy consequence of the definitions of m and m_ν, so we focus on the left one. Let $\mathcal{B}_\nu$ be the convex hull of rank one sign matrices. The norm induced by $\mathcal{B}_\nu$ is the nuclear norm ν, which is dual to the operator norm from ℓ_∞ to ℓ_1. With this terminology we can express $m_\nu(A)$ as

$$m_\nu(A) = \max_{B\in\mathcal{B}_\nu} \min_{ij} a_{ij} b_{ij}. \tag{7}$$

It is not hard to check (Using Equation 2) that $m(A)$ can be equivalently expressed as

$$m(A) = \max_{B\in\mathcal{B}_{\gamma_2}} \min_{ij} a_{ij} b_{ij}.$$

Equation (5) can be restated as

$$\mathcal{B}_\nu \subset \mathcal{B}_{\gamma_2} \subset K_G \cdot \mathcal{B}_\nu.$$

Now let $B \in \mathcal{B}_{\gamma_2}$ be a real matrix satisfying $m(A) = \min_{ij} a_{ij} b_{ij}$. The matrix $K_G^{-1} B$ is in $\mathcal{B}_\nu$ and therefore

$$m_\nu(A) \ge K_G^{-1} \min_{ij} a_{ij} b_{ij} = K_G^{-1} m(A).$$

■

Remark 15 *Grothendieck's inequality has an interesting consequence in the study of large margin classifiers. As mentioned above, such classifiers map the sample points into $\mathbb{R}^t$ and then seek an optimal linear classifier (a linear functional, i.e. a real vector). Grothendieck's inequality implies that if we map our points into $\{\pm 1\}^k$ rather than to real space, the loss in margin is at worst a factor of K_G.*

We return to prove the equivalence between m_ν and discrepancy. The following relation between discrepancy and the $\infty \to 1$ norm is fairly simple (e.g. [4]):

$$disc(A) \le min_P \|P \circ A\|_{\infty\to 1} \le 4 \cdot disc(A)$$

where $P \circ A$ denotes, as usual, the Hadamard (entry-wise) product of the two matrices.

Lemma 16 *Denote by $\mathcal{P}$ the set of matrices whose elements are nonnegative and sum up to 1. For every sign matrix A,*

$$m_\nu(A) = \min_{P\in\mathcal{P}} \|P \circ A\|_{\infty\to 1}. \tag{8}$$

Proof We express m_ν as the optimum of some linear program and observe that the right hand side of Equation (8) is the optimum for the dual program. The statement then follows from LP duality.

Equation (7) allows us to express m_ν as the optimum of a linear program. The variables of this program correspond to a probability measure q on the vertices of the polytope $\mathcal{B}_\nu$, and an auxiliary variable δ is used to express $\min_{ij}\ a_{ij}b_{ij}$. The vertices of $\mathcal{B}_\nu$ are in 1:1 correspondence with all $m \times n$ sign matrices of rank one. We denote this collection of matrices by $\{X_i | i \in I\}$. The linear program is

$$\begin{array}{lccc} & \textbf{maximize } \delta & & \\ \textbf{s.t.}: & & & \\ & \sum_{i\in I} q_i\,(X_i \circ A) - \delta J & \geq & 0 \\ \forall\, i \in I & q_i \geq 0 & & \\ & \sum_i q_i = 1. & & \end{array}$$

Here J is the all-ones matrix. It is not hard to see that the dual of this linear program is

$$\begin{array}{lccc} & \textbf{minimize } \Delta & & \\ \textbf{s.t.}: & & & \\ \forall\, i \in I & \langle P \circ A, X_i \rangle = \langle P, X_i \circ A \rangle & \leq & \Delta \\ \forall i,j & p_{ij} \geq 0 & & \\ & \sum_{i,j} p_{ij} = 1, & & \end{array}$$

where $P = (p_{ij})$. The optimum of the dual program is equal to the right hand side of Equation (8) by definition of $\|\cdot\|_{\infty\to 1}$. The statement of the lemma follows from LP duality. ■

To conclude, we have proved the following:

Theorem 17 *The ratio between any two of the following four parameters is at most* 8 *for any sign matrix* A*,*

- $m(A) = mc(A)^{-1}$
- $m_\nu(A)$
- $disc(A)$*,*
- $\min_{P\in\mathcal{P}} \|P \circ A\|_{\infty\to 1}$*, where* $\mathcal{P}$ *is the set of matrices with nonnegative entries that sum up to* 1*.*

4 Soft margin complexity, or mc-rigidity

As mentioned in the introduction, some classification algorithms allow the classifier to make a few mistakes, in search of a better margin. Such algorithms are called *soft margin algorithms*. The complexity measure associated with these algorithms is what we call mc-rigidity. The mc-rigidity of a sign matrix A is defined

$$mc_r(A,l) = min_{B:h(B,A)\leq l}\ mc(B),$$

We prove that low mc-rigidity is rare.

Theorem 18 *There is a constant* $c > 0$ *such that the number of* $n \times n$ *sign matrices* A *that satisfy* $mc_r(A,l) \leq \sqrt{\frac{l}{n}}$ *is at most*

$$\left(\frac{n^2}{l}\right)^{c\cdot l\cdot \log\frac{n^2}{l}},$$

for every $0 < l \leq n^2/2$*.*
In particular, there exist $\epsilon > 0$ *such that almost every* $n \times n$ *sign matrix* A *satisfies:*

$$\Pr(mc_r(A,\epsilon n^2) > \sqrt{\epsilon n}) = 1 - o(1).$$

The first part of the theorem is significantly better than previous bounds [6, 18, 23, 24]. Note that using Theorems 13 and 18 we get an upper bound on the number of sign matrices with small discrepancy. We don't know of a direct method to show that low-discrepancy matrices are so rare.

Theorem 18 is reminiscent of bounds found in [3, 27] on the number of sign matrices that are realizable in a low dimensional space.

To prove Theorem 18 we use the following theorem by Warren (see [2] for a comprehensive discussion), and the lemma below it.

Theorem 19 (Warren (1968)) *Let* $P_1, \ldots, P_m$ *be real polynomials in* $t \leq m$ *variables, of total degree* $\leq k$ *each. Let* $s(P_1, \ldots, P_m)$ *be the total number of sign patterns of the vectors* $(P_1(x), \ldots, P_m(x))$*, over* $x \in \mathbb{R}^t$*. Then*

$$s(P_1, \ldots, P_m) \leq (4ekm/t)^t\,.$$

In the next lemma we consider the relation between the margin complexity of a sign matrix A and the minimal dimension at which it can be realized (see Section 2). This relation makes it possible to use Warren's theorem in the proof of Theorem 18.

Lemma 20 *Let* B *be a* $n \times n$ *sign matrix, and let* $0 < \rho < 1$*. There exists a matrix* $\tilde{B}$ *with Hamming distance* $h(B, \tilde{B}) < \rho n^2$*, such that*

$$d(\tilde{B}) \leq O(\log \rho^{-1} \cdot mc(B)^2).$$

Proof We use the following known fact (e.g. [13, 22]): Let $x, y \in \mathbb{R}^n$ be two unit vectors with $|\langle x, y\rangle| \geq \epsilon$ then

$$\Pr_L\left(sign(\langle P(x), P(y)\rangle) \neq sign(\langle x, y\rangle)\right) \leq 4e^{-k\epsilon^2/8}.$$

Where the probability is over k-dimensional subspaces L, and where $P : \mathbb{R}^n \to L$ is the projection onto L.

By definition of the margin complexity, there are two $n \times n$ matrices X and Y such that

- $B = \text{sign}(XY)$

- Every entry in XY has absolute value ≥ 1
- $\|X\|_{2\to\infty} = \|Y\|_{1\to 2} = \sqrt{mc(B)}$.

Denote by $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ the rows of X and columns of Y respectively. Take C such that $4e^{-C/8} \leq \rho$, then by the above fact, for $k = Cmc(B)^2$ there is a k-dimensional linear subspace L, such that projecting the points onto L preserves at least $(1-\rho)n^2$ signs of the n^2 inner products $\{\langle x_i, y_j\rangle\}$. ■

To complete the proof of Theorem 18 let A be a sign matrix with $mc_r(A, l) \leq \mu$. Namely, it is possible to flip at most l entries in A to obtain a sign matrix B with $mc(B) \leq \mu$. Let $\rho = l/n^2$ and apply Lemma 20 to B. This yields a matrix E, such that the Hamming distance $h(sign(E), B) \leq l$ and E has rank $O(\log \rho^{-1} \cdot \mu^2)$ (In the terminology of Lemma 20 $\tilde{B} = sign(E)$). To sum up, we change at most l entries in A to obtain B and then at most l more entries to obtain $sign(E)$, a matrix realizable in dimension $O(\log \rho^{-1} \cdot \mu^2)$. Therefore $A = sign(E + F_1 + F_2)$, where F_1, F_2 have support of size at most l each (corresponding to the entries where sign flips were made).

Now E can be expressed as $E = UV^t$ for some $n \times r$ matrices U and V with $r \leq c_1 \log \rho^{-1} \cdot \mu^2$ (c_1 is a constant). Let us fix one of the $\leq \binom{n^2}{l}^2$ choices for the supports of the matrices F_1, F_2 and consider the entries of U, V and the nonzero entries in F_1, F_2 as formal variables. Each entry in A is the sign of a polynomial of degree 2 in these variables. We apply Warren's theorem (Theorem 19) with these parameters to conclude that the number of $n \times n$ sign matrices A with $mc_r(A, l) \leq \mu$ is at most

$$\binom{n^2}{l}^2 \cdot \left(8en^2/(2c_1 \log \rho^{-1} \cdot \mu^2 \cdot n + 2l)\right)^{2c_1 \log \rho^{-1} \cdot \mu^2 \cdot n + 2l}.$$

Recall that $\rho = l/n^2$ and substitute $\mu = \sqrt{\frac{l}{n}}$, to get

$$\binom{n^2}{l}^2 \cdot \left(8en^2/(2c_1 \cdot l \cdot \log \frac{n^2}{l} + 2l)\right)^{2c_1 \cdot l \cdot \log \frac{n^2}{l} + 2l}$$
$$= \left(\frac{n^2}{l}\right)^{O(l \cdot \log \frac{n^2}{l})}.$$

4.1 Communication complexity classes

Surprisingly, mc-rigidity is related to questions about separating communication complexity classes. A major open problem, from [5] is to separate the polynomial hierarchy. Lokam [20] has raised the question of explicitly construction matrices outside AM^{cc}, the class of bounded round interactive proof systems. We tie these questions to mc-rigidity.

Theorem 21 *Let A be an $n \times n$ sign matrix. If there exists a constant $c \geq 0$ such that for every $c_1 \geq 0$*

$$mc_r(A, n^2/2^{(\log\log n)^c}) \geq 2^{(\log\log n)^{c_1}},$$

then A is not in PH^{cc}.

and

Theorem 22 *An $n \times n$ sign matrix A that satisfies*

$$mc_r(A, n^2/2^{(\log\log n)^c}) \geq 2^{(\log\log n)^{c_1}}$$

for every $c, c_1 \geq 0$, is outside AM^{cc}.

Following are the proofs of Theorems 21 and 22

Proof [of Theorem 21] The theorem is a consequence of the definition of mc_r and the following claim: For every $2^m \times 2^m$ sign matrix $A \in PH^{cc}$ and every constant $c \geq 0$ there is a constant $c_1 \geq 0$ and a matrix B such that:

1. The entries of B are nonzero integers.
2. $\gamma_2(B) \leq 2^{(\log\log n)^{c_1}}$.
3. $h(A, sign(B)) \leq n^2/2^{(\log\log n)^c}$.

The proof of this claim is based on a theorem of Tarui [30] (see also Lokam [20]).

It should be clear how boolean gates operate on $0-1$ matrices. By definition of the polynomial hierarchy in communication complexity, every boolean function f in Σ_k can be computed by an AC^0 circuit of polynomial size whose inputs are $0-1$ matrices of size $2^m \times 2^m$ and rank 1. Namely,

$$f(x, y) = C(X_1, \ldots, X_s),$$

where C is an AC^0 circuit, $\{X_i\}_{i=1}^s$ are $0-1$ rank 1 matrices, and $s \leq 2^{polylog(m)}$.

Now, AC^0 circuits are well approximated by low degree polynomials, as proved by Tarui [30]. Let C be an AC^0 circuit of size $2^{polylog(m)}$ acting on $2^m \times 2^m$ $0-1$ matrices $\phi_1, \ldots, \phi_s$. Fix $0 < \delta = 2^{(\log m)^c}$ for some constant $c \geq 0$. Then there exists a polynomial $\Phi \in \mathbb{Z}[X_1, \ldots, X_s]$ such that

1. The sum of absolute values of the coefficients of Φ is at most $2^{polylog(m)}$.
2. The fraction of entries where the matrices $C(\phi_1, \ldots, \phi_s)$ and $\Phi(\phi_1, \ldots, \phi_s)$ differ is at most δ. Here and below, when we evaluate $\Phi(\phi_1, \ldots, \phi_s)$, products are pointwise matrix products.
3. Where Φ and C differ, $\Phi(\phi_1, \ldots, \phi_s) \geq 2$.

Let us apply Tarui's theorem on the $0-1$ version of A, and let $\Phi = \sum_{T \in \{0,1\}^s} a_T \Pi_{i \in T} X_i$ be the polynomial given by the theorem. Notice that $Y_T = \Pi_{i \in T} X_i$ has rank 1. Let

$$B = (\sum_{T \in \{0,1\}^s} a_T Y_T) - J.$$

Then

1. The entries of B are nonzero integers,
2. $\gamma_2(B) \leq 1 + \sum_{T \in \{0,1\}^s} |a_T| \leq 2^{polylog(m)}$,
3. $h(A, sign(B)) \leq \delta n^2$,

as claimed. ∎

Proof [of Theorem 22] We first recall some background from [20]. A family $\mathcal{G} \subset 2^{[n]}$, is said to generate a family $\mathcal{F} \subset 2^{[n]}$ if every $F \in \mathcal{F}$ can be expressed as the union of sets from $\mathcal{G}$. We denote by $g(\mathcal{F})$ the smallest cardinality of a family $\mathcal{G}$ that generates $\mathcal{F}$. Each column in a $0-1$ matrix Z is considered as the characteristic vector of a set and $\Phi(\mathrm{Z})$ is the family of all such sets. If A is an $n \times n$ sign matrix, we define $\mathcal{F}(\mathrm{A})$ as $\Phi(\bar{\mathrm{A}})$ where $\bar{A}$ is obtained by replacing each -1 entry in A by zero. We denote $g(\mathcal{F}(A))$ by $g(A)$. Finally there is the rigidity variant of $g(A)$:

$$g(A, l) = min_{B: h(B,A) \leq l} \ g(B),$$

Lokam [20, Lemma 6.3] proved that if $g(A, n^2/2^{(\log\log n)^c}) \geq 2^{(\log\log n)^{\omega(1)}}$ for every $c > 0$ then $A \notin AM^{cc}$. We conclude the proof by showing that

$$g(A) \geq (mc(A) - 1)/2$$

for every $n \times n$ sign matrix A.

Let $g = g(A)$ and let $\mathcal{G} = \{G_1, \ldots, G_g\}$ be a minimal family that generates $\mathcal{F}(A)$. Let X be the $n \times g$ $0-1$ matrix whose i-th column is the characteristic vector of G_i. Denote by Y a $g \times n$ $0-1$ matrix that specifies how to express the columns of $\bar{A}$ by unions of sets in $\mathcal{G}$. Namely, if we choose to express the i-th column in $\bar{A}$ as $\cup_{t \in T} G_t$, then the i-th column in Y is the characteristic vector of T. Clearly XY is a nonnegative matrix whose zero pattern is given by $\bar{A}$. Consequently, the matrix $B = XY - \frac{J}{2}$ satisfies

1. $sign(B) = A$,
2. $|b_{ij}| \geq 1/2$ and
3. $\gamma_2(B) \leq \gamma_2(XY) + \gamma_2(\frac{J}{2}) \leq g + 1/2$.

It follows that

$$mc(A) \leq \gamma_2(2B) \leq 2g + 1,$$

as claimed. ∎

5 Lower bounds on mc-rigidity

To provide some perspective for our discussion of lower bounds on mc-rigidity, it is worthwhile to recall first some of the known results about $rank$-rigidity. The best known explicit lower bound for $rank$-rigidity is for the $n \times n$ Sylvester-Hadamard matrix H_n [14], and has the following form: For every $r > 0$, at least $\Omega(\frac{n^2}{r})$ changes have to be made in H_n to reach a matrix with rank at most r. Our first lower bound has a similar flavor. For example, since $\|H_n\|_{\infty \to 1} = \Theta(n^{3/2})$ (e.g. Lindsey lemma), Theorem 23 below implies that at least $\Omega(\frac{n^2}{g})$ sign flips in H_n are required to reach a matrix with margin complexity $\leq g$. (This applies for all relevant values of g, since we only have to consider $g \leq O(\sqrt{n})$.)

Theorem 23 *Every $m \times n$ sign matrix A satisfies*

$$mc_r(A, \frac{mn}{8g}) \geq g,$$

provided that $g < \frac{mn}{2K_G \|A\|_{\infty \to 1}}$.

We conjecture that there is an absolute constant $\epsilon_0 > 0$ such that for every sign matrix A with $mc(A) \geq \Omega(\sqrt{n})$ at least $\Omega(n^2)$ sign flips are needed in A to reach a sign matrix with margin complexity $\leq \epsilon_0 \cdot mc(A)$. Theorem 23 yields this conclusion only when $\epsilon_0 \leq O(\frac{1}{\sqrt{n}})$. The next theorem offers a slight improvement and yields a similar conclusion already for $\epsilon_0 \leq O(\sqrt{\frac{\log n}{n}})$. (Recall that $mc(A) \leq \gamma_2(A)$ for every sign matrix A. Thus $mc(A) \geq \Omega(\sqrt{n})$ entails the assumption of Theorem 24.)

Theorem 24 *Every $n \times n$ sign matrix A with $\gamma_2(A) \geq \Omega(\sqrt{n})$ satisfies*

$$mc_r(A, \delta n^2) \geq \Omega(\sqrt{\log n}),$$

for some $\delta > 0$.

The proofs of Theorems 23, 24 use some information about the Lipschitz constants of two of our complexity measures:

Lemma 25 *The Hamming distance of two sign matrices A, B is at least*

$$h(A, B) \geq \frac{1}{2} \left(\|B\|_{\infty \to 1} - \|A\|_{\infty \to 1} \right).$$

Proof [of Lemma 25] Let x and y be two sign vectors satisfying $\sum b_{i,j}x_iy_j = \|B\|_{\infty\to 1}$. If M is the Hamming distance between A and B, then

$$\begin{aligned} \|A\|_{\infty\to 1} &\geq \sum a_{ij}x_iy_j \\ &= \sum b_{i,j}x_iy_j + \sum (a_{ij}-b_{i,j})x_iy_j \\ &\geq \sum b_{i,j}x_iy_j - \sum |a_{ij}-b_{i,j}| \\ &= \|B\|_{\infty\to 1} - 2M \end{aligned}$$

■

We next need a similar result for γ_2:

Lemma 26 *For every pair of sign matrices A and B*

$$h(A,B) \geq \Omega(|\gamma_2(A)-\gamma_2(B)|^4).$$

In the proof of Lemma 26 we need a bound on the γ_2 of sparse $(-1,0,1)$-matrices given by the following lemma.

Lemma 27 *Let A be a $(-1,0,1)$-matrix with N non-zero entries, then $\gamma_2(A) \leq 2(N)^{1/4}$.*

Proof We find matrices B and C such that $A = B + C$ and

$$\gamma_2(B), \gamma_2(C) \leq (N)^{1/4},$$

since γ_2 is convex, $\gamma_2(A) \leq 2(N)^{1/4}$.

Let I be the set of rows of A with more than $(N)^{1/2}$ non-zero entries, we define the matrices B and C by:

$$\begin{aligned} b_{ij} &= \begin{cases} a_{ij} & \text{if } i \in I \\ 0 & \text{otherwise} \end{cases} \\ c_{ij} &= \begin{cases} a_{ij} & \text{if } i \notin I \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

The matrix B has at most $(N)^{1/2}$ non zero rows, and each row in C has at most $(N)^{1/2}$ non-zero entries, thus by considering the trivial factorizations ($IX = XI = X$) we conclude that $\gamma_2(B), \gamma_2(C) \leq (N)^{1/4}$. Obviously $A = B + C$, which concludes the proof. ■

Proof [of Lemma 26] Let A and B be two sign matrices. The matrix $\frac{1}{2}(A-B)$ is a $(-1,0,1)$-matrix with $h(A,B)$ non-zero entries, thus by Lemma 27

$$\gamma_2(A-B) \leq 4h(A,B)^{1/4}$$

Since γ_2 is a norm

$$\gamma_2(A-B) \geq |\gamma_2(A)-\gamma_2(B)|.$$

The claim follows by combining the above two inequalities. ■

We can now complete the proof of Theorems 23 and 24:

Proof of Theorem 23 It is proved in [18] that for every $m \times n$ sign matrix Z

$$\|Z\|_{\infty\to 1} \geq \frac{mn}{K_G \cdot mc(Z)}.$$

We apply this to a matrix B with $mc(B) = g$ and conclude that $\|B\|_{\infty\to 1} \geq \frac{mn}{gK_G}$. On the other hand, by assumption, $\|A\|_{\infty\to 1} \leq \frac{mn}{2gK_G}$, so by Lemma 25, $h(A,B) \geq \frac{mn}{4gK_G} \geq \frac{mn}{8g}$.

Proof of Theorem 24 Let A be an $n \times n$ sign matrix with $\gamma_2(A) \geq \epsilon\sqrt{n}$, for some constant ϵ. By Lemma 26 there is a constant $\delta > 0$ such that every sign matrix B with $h(A,B) \leq \delta n^2$ satisfies $\gamma_2(B) \geq \frac{\epsilon}{2}n^2$. As observed in the Discussion Section in [19] every $n \times n$ sign matrix B with $\gamma_2(B) \geq \Omega(\sqrt{n})$ also satisfies $mc(A) \geq \Omega(\sqrt{\log n})$. It follows that $mc_r(A, cn^2) \geq \Omega(\sqrt{\log n})$.

5.1 Relations with $rank$-rigidity

We discuss relations mc-rigidity has with $rank$-rigidity. First we prove the following lower bound on $rank$-rigidity, which compares favorably to the best known bounds (see e.g. [14]). This lower bound is related to mc-rigidity in that its method of proof is the same as for the proof of Theorem 23. We then prove lower bounds in terms of mc-rigidity on a variant of $rank$-rigidity.

Claim 28 *Every $n \times n$ sign matrix A requires at least $\Omega(\frac{n^2}{r})$ sign reversals to reach a matrix of rank r, provided that $r < \frac{n^2}{2K_G\|A\|_{\infty\to 1}}$.*

Proof Let $\tilde{B}$ be a matrix of rank r obtained by changing entries in A, and let $B = \text{sign}(\tilde{B})$ be its sign matrix. Then

$$r \geq d(B) \geq \frac{n^2}{K_G\|B\|_{\infty\to 1}}.$$

The first inequality follows from the definition of d and the latter from a general bound proved in [18]. It follows that

$$\|B\|_{\infty\to 1} \geq \frac{n^2}{rK_G}.$$

By assumption, $\|A\|_{\infty\to 1} \leq \frac{n^2}{2rK_G}$, and so Lemma 25 implies that the sign matrices A and B differ in at least $\Omega(\frac{n^2}{r})$ places. ■

We turn to discuss $rank$-rigidity when only bounded changes are allowed.

Definition 29 ([20]) *Let A be a sign matrix, and $\theta \geq 0$. For a matrix B denote by $wt(B)$ the number of nonzero entries in B. Define*

$$R^+_A(r,\theta) = \min_B \{wt(A-B) : \text{rank}(B) \leq r,\ 1 \leq |b_{ij}| \leq \theta\}$$

Claim 30 *For every sign matrix A,*

$$R_A^+(mc_r^2(A,l)/\theta^2, \theta) \geq l.$$

Proof Let A be a sign matrix, and B a real matrix with $\theta \geq |b_{ij}| \geq 1$ for all i, j, and $rank(B) \leq mc_r^2(A,l)/\theta^2$. Denote by $\tilde{A}$ the sign matrix of B, i.e. $\tilde{a}_{ij} = sign(b_{ij})$, then $wt(A - \tilde{A}) \leq wt(A - B)$. Also, it holds that

$$mc(\tilde{A}) \leq \gamma_2(B) \leq \|B\|_\infty \sqrt{rank(B)} \leq mc_r(A,l).$$

The first inequality follows from Equation 2 and the second is since $\gamma_2 \leq \sqrt{rank(B)}$ for every real matrix B (This inequality is well known to Banach spaces theorists, see e.g. [18] for a proof). We conclude that $wt(A - B) \geq wt(A - \tilde{A}) \geq l$. Since this is true for every matrix B satisfying the assumptions, $R_A^+(mc_r^2(A,l)/\theta^2, \theta) \geq l$. ■

6 Discussion and open problems

It remains a major open problem to derive lower bounds on mc-rigidity. In particular the following conjecture seems interesting and challenging:

Conjecture 31 *For every constant c_1 there are constants c_2, c_3 such that every $n \times n$ sign matrix A satisfying $mc(A) \geq c_1\sqrt{n}$ also satisfies:*

$$mc_r(A, c_2 n^2) \geq c_3\sqrt{n}.$$

This conjecture says that every matrix with high margin complexity has a high mc-rigidity as well. This is helpful since we do have general techniques for proving lower bounds on margin complexity, e.g. [18]. In particular, an $n \times n$ Hadamard matrix has margin complexity $\sqrt{n}$ ([18]). Thus, Conjecture 31 combined with Theorems 8 and 9, implies that $PH^{cc} \neq PSPACE^{cc}$ and $AM^{cc} \neq IP^{cc}$, since Sylvester-Hadamard matrices are in $IP^{cc} \cap AM^{cc}$. The relation between margin complexity and discrepancy (Theorem 5) adds another interesting angle to these statements.

References

[1] A.V. Aho, J.D. Ullman, and M. Yannakakis. On notions of information transfer in vlsi circuits. In *Proceedings of the 15th ACM STOC*, pages 133–139, 1983.

[2] N. Alon. Tools from higher algebra. *Handbook of combinatorics*, 1:1749–1783, 1995.

[3] 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.

[4] N. Alon and A. Naor. Approximating the cut-norm via grothendieck's inequality. In *Proceedings of the 36th ACM STOC*, pages 72–80, 2004.

[5] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity. In *Proceedings of the 27th IEEE FOCS*, pages 337–347, 1986.

[6] 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.

[7] B. Chazelle. *The Discrepancy Method.* Randomness and complexity. Cambridge University Press, 2000.

[8] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. In *SCT: Annual Conference on Structure in Complexity Theory*, 2001.

[9] 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.

[10] 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.

[11] J. Forster and H. U. Simon. *On the smallest possible dimension and the largest possible margin of linear arrangements representing given concept classes uniform distribution*, volume 2533/2002 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2002.

[12] G. J. O. Jameson. *Summing and nuclear norms in Banach space theory*. London mathematical society student texts. Cambridge university press, 1987.

[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] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. In *Proceedings of the 35th IEEE FOCS*, 1994.

[16] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambride University Press, 1997.

[17] T. Lee, A. Shraibman, and R. Špalek. A direct product theorem for discrepancy. 2008. Accepted to CCC"08.

[18] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

[19] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Manuscript*, 2006.

[20] 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.

[21] J. Matousek. *Geometric Discrepancy. An illustrated guide*, volume 18 of *Algorithms and Combinatorics*. Springer-Verlag, 1999.

[22] J. Matousek. *Lectures on Discrete Geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, 2002.

[23] S. Mendelson. Embeddings with a lipschitz function. *Random Structures and Algorithms*, 27(1):25–45, 2005.

[24] S. Mendelson. On the limitations of embedding methods. In *Proceedings of the 18th annual conference on Learning Theory COLT05, Peter Auer, Ron Meir (Eds.)*, pages 353–365. Lecture notes in Computer Sciences, Springer 3559, 2005.

[25] R. Paturi and J. Simon. Probabilistic communication complexity. *Journal of Computer and System Sciences*, 33:106–123, 1986.

[26] 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.

[27] P. Pudlak and V. Rödl. Some combinatorial-algebraic problems from complexity theory. In *Discrete Mathematics*, volume 136, pages 253–279, 1994.

[28] A. A. Sherstov. Communication complexity under product and nonproduct distributions, 2008. Accepted to CCC"08.

[29] M. A. Shokrollahi, D. A. Spielman, and V. Stemann. A remark on matrix rigidity. *Information Processing Letters*, 64(6):283–285, 1997.

[30] J. Tarui. Randomized polynomials, threshold circuits and polynomial hierarchy. *Theoret. Comput. Sci.*, 113:167–183, 1993.

[31] L.G. Valiant. Graph-theoretic arguments in low level complexity. In *Proc. 6th MFCS*, volume 53, pages 162–176. Springer-Verlag LNCS, 1977.

[32] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1999.

[33] A. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the 15th ACM STOC*, pages 420–428, 1983.