Electronic Colloquium on Computational Complexity, Report No. 66 (2021)

Dimension-free Bounds and Structural Results in Communication Complexity

Lianna Hambardzumyan * Hamed Hatami † Pooya Hatami ‡

Abstract

The purpose of this article is to initiate a systematic study of dimension-free relations between basic communication and query complexity measures and various matrix norms. In other words, our goal is to obtain inequalities that bound a parameter solely as a function of another parameter. This is in contrast to perhaps the more common framework in communication complexity where poly-logarithmic dependencies on the number of input bits are tolerated.

Dimension-free bounds are also closely related to structural results, where one seeks to describe the structure of Boolean matrices and functions that have low complexity. We prove such theorems for several communication and query complexity measures as well as various matrix and operator norms. In several other cases we show that such bounds do not exist.

We propose several conjectures, and establish that, in addition to applications in complexity theory, these problems are central to characterization of the idempotents of the algebra of Schur multipliers, and could lead to new extensions of Cohen's celebrated idempotent theorem regarding the Fourier algebra.

*School of Computer Science, McGill University. lianna.hambardzumyan@mail.mcgill.ca

†School of Computer Science, McGill University. hatami@cs.mcgill.ca. Supported by an NSERC grant.

‡Department of Computer Science and Engineering, The Ohio State University. pooyahat@gmail.com. Supported by NSF grant CCF-1947546

ISSN 1433-8092

Contents

1 Introduction

A matrix is called *Boolean* if its entries are either 0 or 1, and similarly, a function is called *Boolean* if it takes only 0 and 1 values. Our goal in this article is to study whether dimension-free relations exist between basic communication and query complexity measures and various matrix norms for Boolean matrices and functions.

The field of communication complexity, formally defined in 1979 in a paper by Yao [Yao77], studies the communication costs of computing Boolean functions whose input is split between two or more parties. Developed by complexity theorists, this field has been naturally influenced by the more classical areas of complexity theory such as computational complexity where the main challenges lie in separation of complexity classes. The communication complexity classes are defined in [BFS86] as the set of problems that can be solved using protocols with communication costs $\log^c(n)$ in the corresponding model. As a result, a major part of the literature of communication complexity is focused on finding explicit instances (e.g. set-disjointness [She14], Hadamard matrix [For02], gap Hamming distance [CR12]) that require communication cost $\log^c(n)$ in one model (e.g. non-deterministic), whereas they require a much higher communication cost in a different model (e.g. randomized), ideally $\Omega(n)$, where n is the number of input bits. However, a $O(\log(n))$ versus $\Omega(n)$ separation unfortunately does not overrule the existence of dimension-free relations, as for instance, it is possible that one parameter is bounded by an exponential function in the other parameter.

Dimension-free bounds are also often closely related to structural results. For instance, it is well-known that if the deterministic communication complexity of a Boolean matrix is bounded by a constant c, then the matrix is highly structured. Namely, its rank is bounded by 2^c, and it can be partitioned into a constant number of all-zero or all-one submatrices. In other words, its *partition number* is bounded by 2^c.

The simple example of the identity matrix, often called the *equality function* in the context of communication complexity, shows that having small *randomized* communication complexity does not imply a small partition number, or equivalently a small rank. While this, and a handful of other known examples show that the rank of a matrix with bounded randomized communication complexity can be arbitrarily high, they do not overrule the possibility that such matrices might be structured in a different way, or at least contain highly structured parts. Investigating such structures is another focus of this article.

All the known examples of matrices with small randomized communication complexity contain a large all-zero or all-one submatrix. The following conjecture in [CLV19], speculates that this structure holds in general.

Conjecture I. *If the randomized communication complexity of an $n \times n$ Boolean matrix M is bounded by c, then it contains an all-zero or all-one $\delta_c n \times \delta_c n$ submatrix, where $\delta_c > 0$ is a constant that only depends on c.*

In fact [CLV19] conjectures that one can take $\delta_c = 2^{-O(c)}$ in the above statement.

It is well-known that the so-called *approximate trace norm* provides a lower bound for the randomized communication complexity [LS07, Theorem 44]. Hence, one way to establish Conjecture I would be to show that every Boolean matrix with small approximate trace norm contains a large constant submatrix. This motivates us to ask the following tantalizing question about the trace norm itself.

Conjecture II. *If an $n \times n$ Boolean matrix M satisfies $\|M\|_{\mathrm{tr}} \leq cn$, then it contains an all-zero or all-one $\delta_c n \times \delta_c n$ submatrix, where $\delta_c > 0$ is a constant that only depends on c.*

This conjecture is interesting also from the point of view of graph theory. The trace norm of the adjacency matrix of a graph is considered an important graph parameter, and is often called *graph energy* [LSG12] in that context. Furthermore, there is an extensive body of research that investigates graph theoretic [Chu14] or spectral conditions [GN08, BN07, Nik06, LLT07, Nik09] that guarantee the existence of large complete bipartite subgraphs in a graph or its complement. Conjecture II, if true, provides a very natural condition based on graph energy.

The motivation behind the subject of this article goes beyond communication complexity and combinatorics. Several of the problems considered in this article are basic questions about Boolean matrices, and unsurprisingly, they also arise naturally in other areas of mathematics such as *operator theory*, and *Harmonic analysis*.

Let $\mathcal{X}$ and $\mathcal{Y}$ be fixed countable sets, finite or infinite, and consider the set of $\mathcal{X} \times \mathcal{Y}$ Boolean matrices $M : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$. We shall think of rank-one Boolean matrices as the most structured of those. Every such matrix is of the form $\mathbf{1}_{\mathcal{X}_0} \otimes \mathbf{1}_{\mathcal{Y}_0}$ for some $\mathcal{X}_0 \subseteq \mathcal{X}$ and $\mathcal{Y}_0 \subseteq \mathcal{Y}$. These matrices, which correspond to *combinatorial rectangles* $\mathcal{X}_0 \times \mathcal{Y}_0 \subseteq \mathcal{X} \times \mathcal{Y}$, are the building blocks of communication complexity. We denote by

$$\mathscr{Rect} = \{M : \mathcal{X} \times \mathcal{Y} \to \{0,1\} \mid \mathrm{rk}(M) = 1\},$$

the set of all rank-one Boolean matrices.

The next important class of structured Boolean matrices for the purposes of this article is defined as follows. We call a matrix $M : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ *blocky* if there exist, possibly infinitely many, disjoint sets $\mathcal{X}_i \subseteq \mathcal{X}$ and disjoint sets $\mathcal{Y}_i \subseteq \mathcal{Y}$ such that the support of M is

$$\bigcup_i \mathcal{X}_i \times \mathcal{Y}_i.$$

A simple example of a blocky matrix is the *identity matrix*. We denote by $\mathscr{Blocky}$ the set of all blocky matrices. Figure 1 demonstrates examples of a combinatorial rectangle, and blocky matrices.

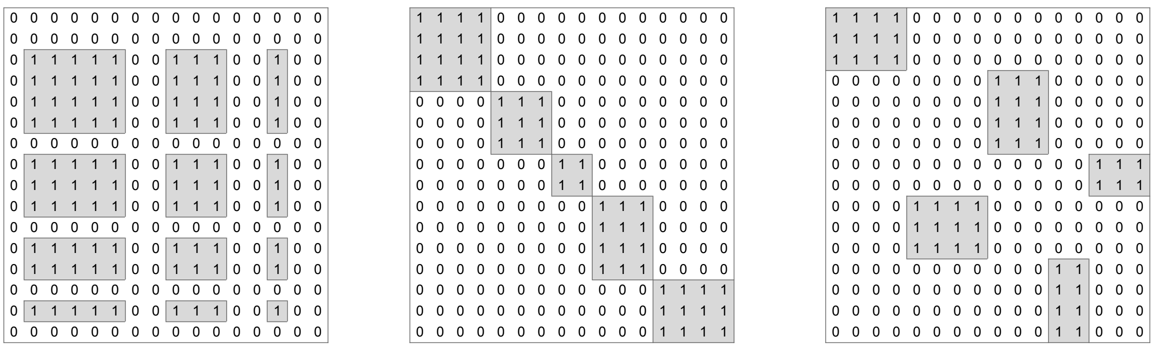

Figure 1: A combinatorial rectangle on the left, and a blocky matrix on the middle and on the right.

These basic matrices appear naturally in different contexts, including those related to the topic of this article, and have been given different names. In graph theory, blocky matrices correspond to equivalence relations on vertex set of a graph, and thus they have been called *equivalence graphs* [Duc79, Fra82, Alo86, BK95]. In complexity theory, blocky matrices have found applications in proving bounds against circuits and branching programs [PR94, Juk06].

A blocky matrix is essentially a blow-up of the identity matrix, obtained by duplicating rows and columns, and then permuting them. Hence, similar to the identity matrix, the randomized communication complexity of every finite blocky matrix is bounded by a fixed constant.

Blocky matrices also arise in the context of Schur multipliers. Recall that the *Schur product* (also called the Hadamard product) of two $\mathcal{X} \times \mathcal{Y}$ matrices M_1 and M_2, denoted by $M_1 \circ M_2$, is their entry-wise product. Let $B(\mathcal{X}, \mathcal{Y})$ denote the space of bounded linear operators $A : \ell_2(\mathcal{X}) \to \ell_2(\mathcal{Y})$ endowed with the operator norm. A matrix $M_{\mathcal{X} \times \mathcal{Y}}$ is called a *Schur multiplier* if for every $A \in B(\mathcal{X}, \mathcal{Y})$, we have $M \circ A \in B(\mathcal{X}, \mathcal{Y})$. Every Schur multiplier M defines a map $B(\mathcal{X}, \mathcal{Y}) \to B(\mathcal{X}, \mathcal{Y})$ via $A \mapsto M \circ A$, which assigns an operator norm to it:

$$\|M\|_m := \|M\|_{B(\mathcal{X},\mathcal{Y}) \to B(\mathcal{X},\mathcal{Y})} = \sup\{\|M \circ A\|_{\ell_2(\mathcal{X}) \to \ell_2(\mathcal{Y})} : \|A\|_{\ell_2(\mathcal{X}) \to \ell_2(\mathcal{Y})} \leq 1\}.$$

Note that Schur multipliers form a Banach algebra via Schur product:

$$\|M_1 \circ M_2\|_m \leq \|M_1\|_m \|M_2\|_m.$$

The following question arises naturally.

What are the *idempotents* of the algebra of Schur multipliers?

Obviously, every idempotent of this algebra must satisfy $M = M \circ M$, and thus is a Boolean matrix. However, not every (infinite) Boolean matrix is a bounded Schur multiplier, as it is possible to have $\|M\|_m = \infty$ for a Boolean matrix M. It is shown in [Liv95] that blocky matrices are exactly the set of all *contractive* idempotents . In other words, an idempotent Schur multiplier satisfies $\|M\|_m \le 1$ if and only if it is a blocky matrix. Livshits's characterization of idempotent Schur multipliers has been extended to other related settings [BH04, Neu06, KP05, Lev14, MP16]. An important question in this area (see e.g. [ELT16]) is whether idempotent Schur multipliers are exactly those Boolean matrices that can be written as a linear combination of *finitely* many contractive idempotents, or equivalently blocky matrices. A simple compactness argument, as outlined in Theorem 5, shows that this problem is equivalent to the following basic question about Boolean matrices.

Conjecture III. *For every $c > 0$, there exists $k_c \in \mathbb{N}$ such that the following holds. If a finite Boolean matrix M is a linear combination of rank-one Boolean matrices with coefficients λ_i satisfying $\sum |\lambda_i| \le c$, then M is a ± 1-linear combination of at most k_c blocky matrices.*

On the other hand, it is not difficult to see that if M is a ± 1-linear combination of at most k_c blocky matrices, then M can be written as linear combination of rank-one Boolean matrices with coefficients whose absolute values sum to at most $O(k_c)$.

By Grothendieck's inequality, the assumption in Conjecture III can be equivalently replaced with the bound $\|M\|_{\gamma_2} = O(1)$, where

$$\|M\|_{\gamma_2} \coloneqq \min\{\|B\|_{2\to\infty}\|C\|_{1\to 2} \ : \ M = BC\}.$$

The connection to Schur multipliers is due to the fact, stated in Theorem 1, that γ_2 norm coincides with the norm of M as a Schur multiplier.

Next, let us state the connection to Harmonic analysis. Let G be a locally compact Abelian group with dual group $\widehat{G}$. Let $\mathbf{M}(G)$ denote the measure algebra of G, that is to say the algebra of bounded, regular, complex-valued measures on G with the convolution operator as multiplication. Note that every idempotent of this algebra satisfies $\mu * \mu = \mu$, and this is equivalent to the statement that the Fourier transform $\widehat{\mu}$ satisfies $\widehat{\mu}^2 = \widehat{\mu}$, and thus is Boolean. Paul Cohen, in a celebrated article [Coh60], proved that μ is an idempotent if and only if $\widehat{\mu}$ can be expressed as a ± 1-linear combination of the indicator functions of a finite number of cosets of $\widehat{G}$. More recently, Green and Sanders [GS08], and Sanders [San20] have proven effective bounds on the required number of cosets as a function of $\|\mu\|$ when G is finite.

As we will explain below, Cohen's idempotent theorem is closely related to Conjecture III. Consider a *finite* Abelian group G. In this case, since $G \cong \widehat{G}$, and $\mathbf{M}(G) = L^1(G)$, by switching the roles of G and $\widehat{G}$, one can state Cohen's idempotent theorem as follows. For every $c > 0$, there exists $k_c > 0$ such that the following holds. If $f : G \to \{0, 1\}$ satisfies

$$\|f\|_A \coloneqq \sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)| \le c, \tag{1}$$

then

$$f = \sum_{i=1}^{k_c} \pm \mathbf{1}_{H_i + a_i}, \tag{2}$$

where each $H_i \le G$ is a subgroup, and each $a_i \in G$. The norm $\|\cdot\|_A$ is called the Fourier algebra norm, and for finite Abelian groups, it is equal to the sum of absolute values of Fourier coefficients of the function.

Note that $\|\mathbf{1}_{H_i+a_i}\|_A = 1$, and furthermore it is not difficult to prove that the indicator functions of cosets $\mathbf{1}_{H+a}$ are the only non-zero idempotents of the Fourier algebra. This is called the Kawada-Itô theorem [KI40, Theorem 3] and dates back to 1940. In other words, if $f : G \to \{0,1\}$ satisfies $\|f\|_A = 1$, then $f = \mathbf{1}_{H+a}$ for some coset $H + a$. Hence, Cohen's idempotent theorem says that every idempotent of the Fourier algebra of G can be expressed as a linear combination of $\kappa(\|f\|_A)$ many *contractive* idempotents for some function $\kappa(\cdot)$. This is precisely what Conjecture III is trying to establish regarding the idempotents of the algebra of Schur mutlipliers. As we explain below, this connection is more than just a verbal analogy.

Let G be a finite Abelian group. Consider a Boolean $f : G \to \{0,1\}$ satisfying (1), and let the Boolean matrix $F : G \times G \to \{0,1\}$ be defined as $F(x,y) = f(x-y)$. It is well-known [LS09, Lemma 36] that

$$\|F\|_{\gamma_2} = \frac{\|F\|_{\mathrm{tr}}}{|G|} = \sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)| = \|f\|_A. \tag{3}$$

Hence if $\|f\|_A \leq c$, then the assumption of Conjecture III holds, and if the conjecture is true, one should be able to express F as a linear combination of a bounded number (as a function of c) of blocky matrices. Indeed in this case, Conjecture III follows from Cohen's idempotent theorem, since a coset $\mathbf{1}_{H_i+a_i}$ in (2) corresponds to the blocky matrix supported on the entries in

$$\bigcup_{b \in G/H} (H_i + b) \times (H_i + b - a_i).$$

Thus Cohen's idempotent theorem implies that both Conjecture II and Conjecture III are true for matrices of the form $F(x,y) = f(x-y)$. In this regard, Conjecture III can be thought of as an extension, or more accurately, an analogue of Cohen's idempotent theorem for the algebra of Schur multipliers. Obviously due to lack of group structure, one cannot hope to find cosets—instead Conjecture III promises blocky matrices.

Finally, let us discuss the *approximate* version of Cohen's idempotent theorem, significant to us due to connections to randomized query and communication complexity. Let G be an Abelian group, and let $f : G \to \{0,1\}$ be a Boolean function. Now, instead of assuming that $\|f\|_A$ is small, let us assume a weaker condition that f has an approximator with small algebra norm. More precisely, there exists a function $g : G \to \mathbb{R}$, not necessarily Boolean, such that $\|f - g\|_\infty \leq \epsilon$ and $\|g\|_A \leq c$. Such functions have been studied by Méla [Mé82] and Host, Méla, and Parreau [HMP86] under the name ϵ-*quasi-idempotent*. In [Mé82] Méla shows that in general, a structure similar to Cohen's idempotent theorem does not necessary hold for such functions. However, in the spirit of Conjecture I, we conjecture that for $G = \mathbb{Z}_2^n$, every ϵ-*quasi-idempotent* contains a highly structured part.

Conjecture IV. *Let $f, g : \mathbb{Z}_2^n \to \mathbb{R}$ be such that f is Boolean, $\|f - g\|_\infty \leq \frac{1}{3}$, and $\|g\|_A \leq c$. There exists a coset $V = H + a \subseteq \mathbb{Z}_2^n$ such that $\frac{|V|}{|\mathbb{Z}_2^n|} \geq \delta_c > 0$, and f is constant on V.*

The constant $\frac{1}{3}$ in the statement is not important and can be replaced by any fixed constant $\epsilon \in (0, 1/2)$, as it is not difficult to see that all such statements will be equivalent.

Note also that by Equation (3), and the fact that randomized communication complexity bounds the approximate trace norm, Conjecture IV, if true, would imply Conjecture I for matrices of the form $F(x,y) = f(x-y)$ where $f : \mathbb{Z}_2^n \to \{0,1\}$.

Public-coin versus private-coin randomness: We caution the reader that in this article, randomized communication complexity always refers to the *public-coin model* where randomness is shared between the players. We also reserve the notation $\mathrm{R}(M)$ to denote the public-coin randomized communication complexity of a Boolean matrix M. See Section 2.2.2 for formal definitions.

Qualitative versus quantitative, and dimension-free-ness: In this article we are interested in *dimension-free* results. In other words, we call two parameters *qualitatively equivalent* if each can be bounded as a function of solely the other one. Furthermore, since the main purpose of this article is establishing dimension-free dependencies, we will not be concerned with quantitative effectiveness of these bounds.

For example, the well-known relations

$$\log \mathrm{rk}(M) \leq \mathrm{D}(M) \leq \mathrm{rk}(M),$$

between rank and deterministic communication complexity, show that insofar as this article is concerned, they are qualitatively equivalent. In contrast, despite Newman's theorem [New91], which states that for $n \times n$ matrices,

$$\mathrm{R}(M) \leq \mathrm{R}^{\mathrm{private}}(M) \leq O(\mathrm{R}(M) + \log\log(n)),$$

due to the $\log\log(n)$ term (which is necessary), public and private randomized communication complexities are not qualitatively equivalent.

In fact, the private-coin model is not interesting from our standpoint: For every Boolean matrix M,

$$\Omega(\log \mathrm{D}(M)) = \mathrm{R}^{\mathrm{private}}(M) \leq \mathrm{D}(M),$$

and thus, as far as this article is concerned, the private-coin randomized communication complexity is qualitatively equivalent to the deterministic communication complexity [KN97, Lemma 3.8].

1.1 Our contributions

In this section, we summarize some of the results proven in this article.

- In Section 3.1 we prove that the deterministic communication complexity with access to an equality oracle is qualitatively equivalent to the smallest k such that the matrix can be written as a linear combination of k blocky matrices.
- In Section 3.2, we show that *zero-error* randomized communication complexity and rank are qualitatively equivalent. Consequently, combining this with a recent result of Gál and Syed [GS19] establishes qualitative equivalence between approximate rank, zero-error randomized communication complexity, deterministic communication complexity, and rank.
- In Section 3.3, we establish Conjecture I for one-sided error randomized communication complexity.
- In Section 3.4, in Theorem 5 we use a compactness argument to show that Conjecture III is equivalent to the statement that every idempotent of the algebra of Schur multipliers is a linear combination of finitely many contractive idempotents.
- In Section 3.5, we consider matrices that are constructed from functions on finite groups. Cohen's idempotent theorem has been generalized to hold for non-Abelian groups as well by Lefranc [Lef72], and effective bounds were given by Sanders [San11]. We use these results, in conjunction with a theorem of Davidson and Donsig [DD07] to verify Conjecture II and Conjecture III for matrices of the form $F(x, y) = f(y^{-1}x)$, where $f : G \to \{0, 1\}$ and G is any finite group.

- In Section 4, we consider XOR-lifts $F_\oplus(x,y) = f(x_1 \oplus y_1, \ldots, x_n \oplus y_n)$, where $f : \{0,1\}^n \to \{0,1\}$. Note that XOR-lift is a special case of $F(x,y) = f(y^{-1}x)$, where $G = \mathbb{Z}_2^n$, and thus, as we mentioned above, Conjecture II and Conjecture III are true for these matrices. We further discuss the analogue of Conjecture I for the $\oplus$-query model, i.e. for *parity decision trees*. In other words, we consider Conjecture IV in relation to randomized $\oplus$-query complexity. Furthermore, we show that the deterministic and zero-error randomized $\oplus$-query complexities are both qualitatively equivalent to the number of nonzero Fourier coefficients.

- In Section 5, we consider AND-lifts $F_\wedge(x,y) = f(x_1 \wedge y_1, \ldots, x_n \wedge y_n)$ for $f : \{0,1\}^n \to \{0,1\}$. We prove that the analogue of Conjecture IV is true in the $\wedge$-query model. Namely, in Theorem 8, we prove that if the randomized AND-decision tree of $f : \{0,1\}^n \to \{0,1\}$ is small, then there is a small set J of coordinates such that f is constant on $\{x : x_j = 0 \ \forall j \in J\}$.

  We remark that Conjecture I, Conjecture II and Conjecture III all remain unresolved for AND-lifts.

- In Section 6, we explain our failure in proving Conjecture I, Conjecture II and Conjecture III by providing an example which shows that the common technique used in proving Cohen's idempotent theorem, and several similar theorems, including some of our results in this article, are inherently inadequate for establishing these conjectures.

2 Preliminaries

Let $\mathbb{D}$ denote the complex unit disk $\{z \in \mathbb{C} \mid |z| \leq 1\}$. For a positive integer n, we use $[n]$ to denote $\{1, \ldots, n\}$. For a set S we denote by $\mathbf{1}_S$ the indicator function of S. For a vector $x \in \{0,1\}^n$, and $S \subseteq [n]$, we denote by $x_S \in \{0,1\}^S$ the restriction of x to the coordinates in S. The Hamming weight of x is defined as $|x| \coloneqq \sum x_i$.

All logarithms in this article are in base 2.

For two functions $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$, we use the following asymptotic notations:

- $f(n) = O(g(n))$, if $\lim_{n\to\infty} \sup \frac{|f(n)|}{|g(n)|} < \infty$.
- $f(n) = \Omega(g(n))$, if and only if $g(n) = O(f(n))$.
- $f(n) = \Theta(g(n))$, if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.
- $f(n) = o(g(n))$, if $\lim_{n\to\infty} \frac{|f(n)|}{|g(n)|} = 0$.
- $f(n) = \omega(g(n))$, if $\lim_{n\to\infty} \frac{|f(n)|}{|g(n)|} = \infty$.

We sometimes identify $\{0,1\}^n$ or $\mathbb{Z}_2^n$ with the vector space $\mathbb{F}_2^n$ over $\mathbb{F}_2$. In this context, we refer to cosets $H + a \subseteq \mathbb{Z}_2^n$ as affine subspaces, which naturally assign a dimensions and a codimension to them.

For sets $\mathcal{X}$ and $\mathcal{Y}$, we will often identify a function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{C}$ with its corresponding matrix $[f(x,y)]_{x\in\mathcal{X}, y\in\mathcal{Y}}$.

For a measure space (Ω, μ), and $p \in [1, \infty)$, we denote by $L^p(\mu)$ the normed space of functions $f : \Omega \to \mathbb{C}$ with $\int |f|^p d\mu < \infty$, together with the norm

$$\|f\|_{L^p(\mu)} \coloneqq \left(\int |f|^p d\mu \right)^{1/p},$$

and $\|f\|_{L^\infty(\mu)}$ is defined as the essential supremum of $|f|$.

For a *finite* set Ω, we write μ_Ω to denote the uniform probability measure on Ω, and we shorthand $\|f\|_{L^p(\mu_\Omega)}$ to $\|f\|_{L^p(\Omega)}$. When Ω is a countable set, we define the normed space $\ell_p(\Omega)$ according to the counting measure:

$$\|f\|_{\ell_p(\Omega)} = \left(\sum_{x \in \Omega} |f(x)|^p \right)^{1/p}.$$

There are several natural norms on the space of $m \times n$ matrices. Considering an $m \times n$ matrix M as a linear operator $M : \mathbb{C}^n \to \mathbb{C}^m$ endows the space with operator norms: For $p, q \in [1, \infty]$, we use the notation $\|M\|_{p \to q}$ to denote its operator norm from ℓ_p to ℓ_q. That is

$$\|M\|_{p \to q} = \sup_{x \in \mathbb{C}^n, \|x\|_{\ell_p} \le 1} \|Mx\|_{\ell_q},$$

It is easy to see that

$$\|M\|_{2 \to 2} = \sigma_{\max},$$

where $\sigma_{\max}$ is the largest singular value of M.

We shall need the following well-known inequality.

Lemma 2.1 (Hoeffding's inequality)**.** *For $i = 1, \ldots, n$, let X_i be independent random variables taking values from range $[a_i, b_i]$ and let $X = \sum_{i=1}^n X_i$. Then,*

$$\Pr[|X - \mathbb{E}[X]| \ge t] < 2 \exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2} \right).$$

2.1 Matrix norms and ranks

In this section we describe some well-known as well as some new matrix parameters which arise from representations of general matrices in terms of more structured matrices. Allowing $\mathcal{S}$ to be various sets of structured matrices (for example, $\mathcal{S} = \mathcal{Rect}$ or $\mathcal{S} = \mathcal{Blocky}$) we define, in a generic way, the matrix parameters that come up in this article. This also makes it easier to see how some of these parameters relate to each other. For a fixed set $\mathcal{S}$ of structured matrices, we introduce a notion of matrix *rank* in terms of $\mathcal{S}$, which we call $\mathcal{S}$-rank, and a matrix *norm* in terms of $\mathcal{S}$, which we call $\mathcal{S}$-norm analogously.

Definition 2.2. *Let $\mathcal{Z}$ be a* finite *set, and let $\mathcal{S}$ be a* spanning *subset of the vector space $\{f : \mathcal{Z} \to \mathbb{C}\}$.*

- *Define the $\mathcal{S}$-*rank *of a function f, denoted by $\mathrm{rk}(\mathcal{S}, f)$, to be the smallest k such that f can be expressed as a linear combination of at most k functions in $\mathcal{S}$ over $\mathbb{C}$.*
- *Define $\|f\|_{\mathcal{S}}$ as*

$$\|f\|_{\mathcal{S}} = \inf \left\{ \sum_{i=1}^r |\lambda_i| \ : \ f = \sum_i^r \lambda_i g_i, \ \text{for } g_i \in \mathcal{S}, \lambda_i \in \mathbb{C}, r \in \mathbb{N} \right\}.$$

It is easy to verify that $\|\cdot\|_{\mathcal{S}}$ is always a semi-norm. By considering different $\mathcal{S}$ we can recover many of the norms and parameters related to this article.

- (*Normalized trace norm*) The *trace norm* of an $m \times n$ matrix M is defined as the sum of its singular values $\sigma_{\max} \coloneqq \sigma_1 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$, namely

  $$\|M\|_{\mathrm{tr}} = \sum_{i=1}^{\min(m,n)} \sigma_i.$$

  In this article, it is more convenient to work with the following normalized version of this norm, which we call the *normalized trace norm*:

  $$\|M\|_{\mathrm{ntr}} = \frac{\|M\|_{\mathrm{tr}}}{\sqrt{mn}}.$$

  When $\mathcal{S}$ is the set of all $m \times n$ matrices of the form $\mathbf{a} \otimes \mathbf{b}$, where $\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^n$ satisfy

  $$\|\mathbf{a}\|_{L^2(m)} \coloneqq \left(\sum_{i=1}^{m} \frac{|\mathbf{a}_i|^2}{m}\right)^{1/2} \leq 1, \text{ and } \|\mathbf{b}\|_{L^2(n)} \coloneqq \left(\sum_{i=1}^{n} \frac{|\mathbf{b}_i|^2}{n}\right)^{1/2} \leq 1,$$

  then $\mathrm{rk}(\mathcal{S}, M)$ coincides with $\mathrm{rk}(M)$ over $\mathbb{C}$, and it follows from the singular value decomposition that

  $$\|M\|_{\mathcal{S}} = \|M\|_{\mathrm{ntr}}.$$

- (*μ-norm*) If $\mathcal{S} = \mathscr{Rect}$, that is the set of rank-one Boolean matrices $\mathbf{a} \otimes \mathbf{b}$, where $\mathbf{a} \in \{0,1\}^m$ and $\mathbf{b} \in \{0,1\}^n$, then $\|\cdot\|_{\mathscr{Rect}}$ is commonly known as the $\|\cdot\|_\mu$ norm. Note that to define $\|\cdot\|_\mu$ one could equivalently take $\mathbf{a} \otimes \mathbf{b}$, where $\mathbf{a} \in [0,1]^m$ and $\mathbf{b} \in [0,1]^n$.

- (*ν-norm*) If $\mathcal{S}$ is the set of all $m \times n$ matrices of the form $\mathbf{a} \otimes \mathbf{b}$, where $\mathbf{a} \in \{-1,1\}^m$ and $\mathbf{b} \in \{-1,1\}^n$, then $\|\cdot\|_{\mathcal{S}}$ is commonly known as the $\|\cdot\|_\nu$ norm. Again to define $\|\cdot\|_\nu$ one could equivalently take $\mathbf{a} \otimes \mathbf{b}$, where $\mathbf{a} \in [-1,1]^m$ and $\mathbf{b} \in [-1,1]^n$.

  It immediately follows that $\|\cdot\|_\nu \leq \|\cdot\|_\mu$, but in fact the two norms are equivalent, since every $\{-1,1\}$-valued vector can be written as the difference of two Boolean vectors:

  $$\|\cdot\|_\nu \leq \|\cdot\|_\mu \leq 4\|\cdot\|_\nu. \tag{4}$$

  Note that for the identity matrix, we have

  $$\mathtt{I}_n(x,y) = \frac{1}{2^n} \sum_{S \subseteq [n]} (-1)^{\mathbf{1}_{x \in S}} (-1)^{\mathbf{1}_{y \in S}},$$

  and thus $\|\mathtt{I}_n\|_\nu = 1$.

- (*γ_2-norm*) We can relax the ν-norm further. Let $\mathcal{S}$ be the set of all $m \times n$ matrices with ij-entries $\langle \mathbf{a}_i, \mathbf{b}_j \rangle$, where $\mathbf{a}_i$ and $\mathbf{b}_j$ are *unit vectors* in any Hilbert space $\mathcal{H}$.

  Taking $\mathcal{H}$ to be $\mathbb{R}$, we have only two unit vectors ± 1 and thus we recover ν norm. Hence $\|\cdot\|_{\gamma_2} \leq \|\cdot\|_\nu$. It turns out that γ_2-norm is also equivalent to the ν norm. This is in fact the well-known Grothendieck inequality (see Theorem 1):

  $$\|\cdot\|_{\gamma_2} \leq \|\cdot\|_\nu \leq \frac{\pi}{2\ln\left(1+\sqrt{2}\right)} \|\cdot\|_{\gamma_2}.$$

  The constant $\frac{\pi}{2\ln\left(1+\sqrt{2}\right)}$ is due to Krivine [Kri79], and it holds for both real and complex Hilbert spaces. Note also that the unit ball of $\|\cdot\|_{\gamma_2}$ is the set of $m \times n$ matrices with ij-entries $\langle \mathbf{a}_i, \mathbf{b}_j \rangle$, where $\|\mathbf{a}_i\| \leq 1$ and $\|\mathbf{b}_j\| \leq 1$ in some Hilbert space $\mathcal{H}$.

- (*Blocky-rank and norm*) For $\mathcal{S} = \mathcal{Blocky}$, we study $\mathrm{rk}(\mathcal{Blocky}, f)$, which we prove is qualitatively equivalent to the deterministic communication complexity with access to equality oracle (see Proposition 3.1). We refer to $\|\cdot\|_{\mathcal{Blocky}}$ as *blocky-norm*. Blocky matrices are the blow-ups of the identity matrix, and thus every non-zero blocky matrix B satisfies

  $$\|B\|_{\gamma_2} = \|B\|_\nu = 1.$$

  On the other hand every $\mathbf{a} \otimes \mathbf{b}$, where $\mathbf{a} \in \{-1,1\}^m$ and $\mathbf{b} \in \{-1,1\}^n$, can be written as the difference of two blocky matrices, and thus satisfies $\|\mathbf{a} \otimes \mathbf{b}\|_{\mathcal{Blocky}} \le 2$. We conclude

  $$\|\cdot\|_\nu \le \|\cdot\|_{\mathcal{Blocky}} \le 2\|\cdot\|_\nu. \tag{5}$$

- (*Fourier rank and algebra norm*) Let G be a finite Abelian group with dual $\widehat{G}$. Then for $f : G \to \mathbb{C}$,

  $$\mathrm{rk}(\widehat{G}, f)$$

  corresponds to the so-called *Fourier rank* of f, which is the number of non-zero Fourier coefficients of f. In this case, the corresponding norm coincides with Fourier algebra norm

  $$\|f\|_{\widehat{G}} = \|f\|_A.$$

- (*Monomial rank and norm*) Consider the space of functions $f : \{0,1\}^n \to \mathbb{C}$, and let

  $$\mathcal{Mon} := \left\{ x \mapsto \prod_{i \in S} x_i \mid S \subseteq [n] \right\}$$

  be the set of all monomials where every variable appears with degree at most 1. Then, for a function $f : \{0,1\}^n \to \mathbb{C}$,

  $$\mathrm{rk}(\mathcal{Mon}, f)$$

  corresponds to the number of non-zero coefficients in the (unique) polynomial representation of f. This is often called the *sparsity* of f in the literature of computer science. Note also that $\|f\|_{\mathcal{Mon}}$ coefficients in the unique polynomial representation of f in the ring $\mathbb{C}[x_1, \ldots, x_n]/(x_1^2 = x_1, \ldots, x_n^2 = x_n)$.

Schur Multipliers Let $\mathcal{X}$ and $\mathcal{Y}$ be two countable sets. The *Schur product* of two $\mathcal{X} \times \mathcal{Y}$ matrices $A = [a_{xy}]$ and $B = [b_{xy}]$, denoted by $A \circ B$, is their entry-wise product $[a_{xy} b_{xy}]$.

Consider the two Hilbert spaces $\mathcal{H}_1 = \ell_2(\mathcal{Y})$ and $\mathcal{H}_2 = \ell_2(\mathcal{X})$, and let $B(\mathcal{H}_1, \mathcal{H}_2)$ be the space of all *bounded* linear operators $A : \mathcal{H}_1 \to \mathcal{H}_2$ together with the operator norm $\|A\|_{\mathcal{H}_1 \to \mathcal{H}_2}$. A matrix $M_{\mathcal{X} \times \mathcal{Y}}$ is called a *Schur multiplier* if for every $A \in B(\mathcal{H}_1, \mathcal{H}_2)$, the matrix $M \circ A \in B(\mathcal{H}_1, \mathcal{H}_2)$. Every Schur multiplier defines a map $B(\mathcal{H}_1, \mathcal{H}_2) \to B(\mathcal{H}_1, \mathcal{H}_2)$ via $A \mapsto M \circ A$.

To distinguish from the norm on bounded operators, we will write $\|M\|_m$ for the norm of a Schur multiplier:

$$\|M\|_m = \sup\{\|M \circ A\|_{\mathcal{H}_1 \to \mathcal{H}_2} \ : \ \|A\|_{\mathcal{H}_1 \to \mathcal{H}_2} \le 1\}.$$

It turns out that $\|\cdot\|_m$ coincides with γ_2 norm defined above. The following relations are essentially due to Grothendieck (see also [LS07, Pis12]).

Theorem 1 (Grothendieck [Gro52]). *For every matrix M,*

$$\|M\|_m = \|M\|_{\gamma_2} \le \|M\|_\nu \le \frac{\pi}{2\ln(1+\sqrt{2})}\|M\|_{\gamma_2}.$$

In other words, $\|\cdot\|_m$, $\|\cdot\|_\mu$, $\|\cdot\|_\nu$, and $\|\cdot\|_{\gamma_2}$ are all within constant factors of each other. Let us also mention the following common property of these norms, which is straightforward to verify.

Proposition 2.3. *Let $\|\cdot\|$ be any of the norms $\|\cdot\|_m$, $\|\cdot\|_\mu$, $\|\cdot\|_\nu$, or $\|\cdot\|_{\gamma_2}$. Then*

$$\|\oplus_{i=1}^{\infty} M_i\| = \sup_i \|M_i\|.$$

Idempotents and Boolean matrices Schur multipliers on $B(\mathcal{H}_1, \mathcal{H}_2)$ form a Banach algebra via the Schur product, since

$$\|M_1 \circ M_2\|_m \leq \|M_1\|_m \|M_2\|_m.$$

When $\mathcal{H}_1$ and $\mathcal{H}_2$ are finite dimensional, Boolean matrices and *idempotents* of this algebra coincide: $M \circ M = M$ if and only if M is a Boolean matrix. However, in the infinite dimensions, not every Boolean matrix is a bounded Schur multiplier.

We will be interested in characterizing the idempotents of the algebra of Schur multipliers. As we shall see in Theorem 5, this reduces to characterizing the structure of finite Boolean matrices M with a uniform bound on $\|M\|_m$.

First let us consider the contractive idempotents. Note that every rank-one Boolean matrix is a contraction. As a result, by Proposition 2.3, the identity matrix and, more generally, all blocky matrices are contractions.

Note that the Schur multiplier norm is monotone in the sense that the norm of a submatrix cannot be larger than the original matrix. Since $\|1\|_m = 1$, it follows that every non-zero Boolean matrix satisfies $\|M\|_m \geq 1$. Livshits [Liv95] showed that the 2×2 matrix with three 1's is not contractive.

Lemma 2.4 ([Liv95])**.** *We have*

$$\left\| \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right\|_m = \frac{2}{\sqrt{3}} > 1.$$

Since $\|\cdot\|_m$ norm is invariant under row and column permutations, it follows that a contractive idempotent M cannot have any 2×2 submatrices with exactly 3 ones. In this context, the property is often called the 3-*of*-4 *property*, which fully characterizes such matrices as being the same as the set of *blocky*-matrices.

Theorem 2 ([Liv95])**.** *M is a contractive idempotent of the algebra of Schur multipliers if and only if $M \in \mathcal{Blocky}$. More generally, this is true for idempotents that satisfy $\|M\|_m < \frac{2}{\sqrt{3}}$.*

Relation to the Normalized Trace Norm As we saw above $\|\cdot\|_{\gamma_2} = \|\cdot\|_m$, $\|\cdot\|_\mu$, and $\|\cdot\|_\nu$, are all equivalent. Furthermore, it is easy to see [LS07, Section 2.3.2] that

$$\|\cdot\|_{\mathrm{ntr}} \leq \|\cdot\|_{\gamma_2}. \tag{6}$$

However, $\|\cdot\|_{\mathrm{ntr}}$ could be much smaller than the above norms since adding all-zero rows or columns would decrease the normalized trace norm, while other norms would remain intact.

2.1.1 The Fourier algebra norm

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Identifying $\{0,1\}^n$ with the finite Abelian group $G = \mathbb{Z}_2^n$ allows us to consider the Fourier expansion of $f = \sum_{\chi \in \widehat{G}} \hat{f}(\chi)\chi$, where $\widehat{G}$ is the dual of G. It is common in theoretical computer science to represent this expansion as

$$f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S,$$

by representing the characters of $\mathbb{Z}_2^n$ as

$$\chi_S : x \mapsto \prod_{i \in S} (-1)^{x_i}.$$

The Fourier algebra norm of f, denoted by $\|f\|_A$, is the sum of absolute values of Fourier coefficients:

$$\|f\|_A = \sum_S |\widehat{f}(S)|.$$

The name comes from the fact that it satisfies $\|f_1 f_2\|_A \le \|f_1\|_A \|f_2\|_A$ for any $f_1, f_2 : G \to \mathbb{C}$. In the literature of theoretical computer science, this norm is sometimes called the *spectral norm* of f, but in order to avoid confusion with spectral norm of matrices, we will use the harmonic analysis term, *Fourier algebra norm*.

The above definition immediately generalizes to every finite Abelian group G, namely the Fourier algebra norm of $f : G \to \mathbb{C}$ is the sum of absolute values of Fourier coefficients. This can be further generalized to every locally compact Abelian group, and in fact Eymard in [Eym64] generalized the definition of the Fourier algebra to every locally compact group. In this article, we are only concerned with finite groups. Suppose that G is a finite group and $f, g \in L^1(\mu_G)$, where μ_G denotes the unique Haar probability measure on G, which is the uniform probability measure on G, since G is finite. The convolution $f * g$ of f and g is then defined point-wise by

$$f * g(x) \coloneqq \int f(y) g(y^{-1}x) d\mu_G(y) = \mathbb{E}_{y \in G} \left[f(y) g(y^{-1}x) \right].$$

This can be used to introduce the convolution operator: given $h \in L^1(G)$, define $L_h : L^2(G) \to L^2(G)$ via $L_h : \nu \mapsto \nu * h$. The Fourier algebra norm of f is then defined as

$$\|f\|_A \coloneqq \sup \left\{ \langle f, h \rangle \ : \|L_h\|_{L^2(G) \to L^2(G)} \le 1 \right\}.$$

When G is an *Abelian* group, it is not difficult to see that this coincides with the sum of absolute values of Fourier coefficients of f:

$$\|f\|_A = \sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|.$$

2.2 Communication complexity

2.2.1 Deterministic communication complexity

The field of communication complexity studies the amount of communication required to solve a problem of computing discrete functions when the input is split between two parties. Every Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ defines a communication problem. An input $x \in \mathcal{X}$ is given to Alice, and an input $y \in Y$ is given to Bob. Together, they should both compute the entry $f(x,y)$ by exchanging bits of information in turn, according to a previously agreed-on protocol. There is no restriction on their computational power; the only measure we care to minimize is the number of exchanged bits.

A deterministic protocol π specifies for each of the two players, the bit to send next, as a function of their input and history of the communication so far. A protocol naturally corresponds to a binary tree as follows. Every internal node is associated with either Alice or Bob. If an internal node v is associated with Alice, then it is labeled with a function $a_v : \mathcal{X} \to \{0,1\}$, which prescribes the bit sent by Alice at this node as a function of her input. Similarly, Bob's nodes are labeled with Boolean functions on $\mathcal{Y}$. Each leaf is labeled by 0 or 1 which corresponds to the output of the protocol. We denote the number of bits exchanged on the input (x, y) by $\mathrm{cost}_\pi(x, y)$. This is exactly the length of the path from the root to the corresponding leaf. The *communication cost* of the protocol is simply the depth of the protocol tree, which is the maximum of $\mathrm{cost}_\pi(x, y)$ over all inputs (x, y).

$$\mathrm{CC}(\pi) \coloneqq \max_{x,y} \mathrm{cost}_\pi(x, y).$$

Every such protocol π computes a function $\mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, which we also denote by π. Namely $\pi(x, y)$ is the label of the leaf reached by the path corresponding to the players' communication on the input (x, y). We say that π computes f if $\pi(x, y) = f(x, y)$ for all x, y. The *deterministic communication complexity* of f, denoted by $\mathrm{D}(f)$, is the smallest communication cost of a protocol that computes f.

A useful insight is that a bit sent by Alice at a node v corresponds to a partition of the rows into two parts $a_v^{-1}(0)$ and $a_v^{-1}(1)$, and every bit sent by Bob corresponds to a partition of the columns. Every time Alice sends a bit, we restrict to a subset of the rows, and proceed with the created submatrix. Similarly Bob's communicated bits restrict the columns. As this process continues, we see that every c-bit protocol induces a partition of the matrix f into at most 2^c *submatrices*. In the context of the communication complexity, submatrices are often called *combinatorial rectangles* or simply *rectangles*. If the protocol computes f, then all submatrices in this partition are *monochromatic*, namely, labeled by a unique element 0 or 1.

Note that every rank-one Boolean matrix is of the form $\mathbf{1}_{\mathcal{X}_0} \cdot \mathbf{1}_{\mathcal{Y}_0}^T$ for subsets $\mathcal{X}_0 \subseteq \mathcal{X}$ and $\mathcal{Y}_0 \subseteq \mathcal{Y}$. Thus rank-one Boolean matrices are essentially the same as 1-monochromatic rectangles. We conclude the following proposition.

Proposition 2.5 ([KN97]). *For every Boolean matrix F, we have*

$$\log \mathrm{rk}(F) \leq \mathrm{D}(F) \leq \mathrm{rk}(F) \leq \mathrm{rk}(\mathcal{R}ect, F) \leq c \leq 2^{\mathrm{rk}(F)},$$

where c is the partition number *of f, which is the smallest $c > 0$ such that f can be partitioned into c constant submatrices. In particular, all the above parameters are qualitatively equivalent.*

To the extent that we are concerned with qualitative results, Proposition 2.5 provides a satisfactory description of the structure of Boolean matrices whose deterministic communication complexities are uniformly bounded. However, quantitatively, closing the exponential gap between $\mathrm{D}(f)$ and $\log \mathrm{rk}(f)$ into a polynomial dependency is called the log-rank conjecture, and is perhaps the most famous open problem in communication complexity [Lov14].

2.2.2 Randomized communication complexity

In this article, we use the public coin model, where a *probabilistic protocol* π_R is simply a distribution over deterministic protocols. In this notation R is a random variable, and every fixation of R to a particular value r leads to a deterministic protocol π_r. We define the communication cost of a probabilistic protocol π_R as the maximum cost of any protocol π_r in the support of this distribution:

$$\mathrm{CC}(\pi_R) = \max_r \mathrm{CC}(\pi_r) = \max_r \max_{x,y} \mathrm{cost}_{\pi_r}(x, y).$$

We also define the average cost of such a protocol as the expected number of exchanged bits over the worst input (x, y):

$$\mathrm{CC}^{avg}(\pi_R) = \max_{x,y} \mathbb{E}_R[\mathrm{cost}_{\pi_R}(x, y)].$$

In the probabilistic models of computation, three types of error are often considered.

- **Two-sided error:** This is the most important notion of randomized communication complexity. For every x, y, we require

$$\Pr_R[\pi_R(x,y) \neq f(x,y)] \leq \epsilon,$$

  where ϵ is a fixed constant that is strictly less than $1/2$. Note that $\epsilon = 1/2$ can be easily achieved by outputting a random bit; hence it is crucial that ϵ in the definition is strictly less than $1/2$. It is common to take $\epsilon = \frac{1}{3}$. Indeed, the choice of ϵ is not important so long as $\epsilon \in (0, 1/2)$, since the probability of error can be reduced to any constant $\epsilon' > 0$ by repeating the same protocol independently for some $O(1)$ times, and outputting the most frequent output.

  The two-sided error communication complexity is simply called the *randomized communication complexity*. It is denoted by $\mathrm{R}_\epsilon(f)$ and is defined as the smallest communication cost $\mathrm{CC}(\pi_R)$ of a probabilistic protocol that computes f with two-sided error at most ϵ. We set $\epsilon = 1/3$ as the standard error, and denote

$$\mathrm{R}(f) = \mathrm{R}_{\frac{1}{3}}(f).$$

- **One-sided error:** In this setting the protocol is only allowed to make an error if $f(x,y) = 1$. In other words, for every x, y with $f(x,y) = 0$, we have

$$\Pr_R[\pi_R(x,y) = 0] = 1,$$

  and for every x, y with with $f(x,y) = 1$, we have

$$\Pr_R[\pi_R(x,y) \neq f(x,y)] \leq \epsilon.$$

  Again the choice of ϵ is not important so long as $\epsilon \in (0,1)$ because the probability of error can be reduced from ϵ to ϵ^k by repeating the same protocol independently k times and outputing 1 only when at least one of the repetitions outputs 1. We denote by $\mathrm{R}^1_\epsilon(f)$ the smallest $\mathrm{CC}(\pi_R)$ over all protocols π_R with one-sided error of at most ϵ. We set $\epsilon = 1/3$ as the standard error, and denote

$$\mathrm{R}^1(f) = \mathrm{R}^1_{\frac{1}{3}}(f).$$

- **Zero error:** In this case the protocol is not allowed to make any errors. For every x, y, we must have $\Pr_R[\pi_R(x,y) \neq f(x,y)] = 0$. In this setting, $\mathrm{CC}^{avg}(\cdot)$ is considered, as $\mathrm{CC}(\cdot)$ leads to the same notion of complexity as the deterministic communication complexity. We denote

$$\mathrm{R}_0(f) = \inf \mathrm{CC}^{avg}(\pi_R),$$

  over all such protocols.

Note that one can convert a zero-error protocol π with average cost c to a one-sided error protocol π' with cost $3c$, by terminating the protocol after at most $3c$ steps, and outputting 0 in the case where the protocol is terminated prematurely. The protocol π' clearly does not make any errors on 0-inputs. Furthermore, since the average cost of π is c, by Markov's inequality, the probability that the protocol π' is terminated prematurely is at most $\frac{1}{3}$. We conclude

$$\mathrm{R}(f) \leq \mathrm{R}^1(f) \leq 3\,\mathrm{R}_0(f).$$

Obviously, $\mathrm{R}(f), \mathrm{R}^1(f), \mathrm{R}_0(f)$ are all bounded by $\mathrm{D}(f)$.

2.3 Query complexity

In Section 2.2, we introduced various models of communication complexity. In this section we discuss query complexity. Let $\mathcal{X}$ be a finite set, often endowed with a product structure, most commonly $\mathcal{X} = \{0,1\}^n$. In query complexity, a function $f : \mathcal{X} \to \{0,1\}$ is fixed, and a player, who does not know the input x, wants to find out the value of $f(x)$ by making queries about x. The goal is to minimize the number of queries. Depending on what type of queries are allowed, we arrive at different models of query complexity. The most natural setting is to have $f : \{0,1\}^n \to \{0,1\}$. Denoting the input $x = (x_1, \ldots, x_n) \in \{0,1\}^n$, we consider three important types of queries, each leading to a different model of query complexity.

- The *coordinate queries* x_i for $i \in \{1, \ldots, n\}$.
- The *parity queries* $\oplus_{i\in S} x_i$, which are the XOR of the coordinates in S, for $S \subseteq [n]$.
- The AND *queries* $\prod_{i\in S} x_i$, for $S \subseteq [n]$.

Note that, similar to communication complexity, a protocol in each of these models corresponds to a binary tree where each internal node is labeled with a query, and the computation branches according to the output of these queries. The leaves are labeled with the output of the protocol. When only coordinate-queries are allowed, these trees are simply called *decision trees*. The *parity decision trees*, and AND-*decision trees*, respectively correspond to parity queries and AND queries.

The cost of such a protocol is the maximum number of queries made on an input, which is equal to the *depth* of the tree. Such trees naturally correspond to Boolean functions, and the *decision tree complexity* $\mathrm{dt}(f)$, the *parity decision tree complexity* $\mathrm{dt}^{\oplus}(f)$, and the AND-*decision tree complexity* $\mathrm{dt}^{\wedge}(f)$ are defined as the smallest depth required for the function f.

A randomized protocol is simply a distribution over deterministic protocols, and the notions of cost, average cost, zero-error, one-sided error, and two-sided error are defined analogous to communication complexity. The complexity measures corresponding to zero-error, one-sided error, and two-sided error are denoted respectively by rdt_0, rdt^1, rdt.

In the AND-query model, we denote these by $\mathrm{rdt}_0^{\wedge}$, $\mathrm{rdt}^{\wedge 1}$, $\mathrm{rdt}^{\wedge}$, and in the parity query model by $\mathrm{rdt}_0^{\oplus}$, $\mathrm{rdt}^{\oplus 1}$, $\mathrm{rdt}^{\oplus}$.

In the simple decision tree model of coordinate queries, a theorem of Nisan [Nis91] shows that all these parameters are qualitatively equivalent, in fact with polynomial dependencies.

Proposition 2.6 (Coordinate Query Equivalencies [Nis91])**.** *For every Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\mathrm{rdt}(f) \le \mathrm{rdt}^1(f) \le 3\,\mathrm{rdt}_0(f) \le 3\,\mathrm{dt}(f) \le 81\,\mathrm{rdt}(f)^2.$$

In light of Proposition 2.6, from the point of view of this article, the case of the coordinate query has been completely resolved. However, as we shall see later, in both the XOR and AND models, there are examples for which the randomized query complexity is $O(1)$, while the deterministic query complexity is $\Omega(n)$. We discuss the XOR-model in Section 4, and the AND-model in Section 5.

2.4 Lifting theorems

Let G be a finite group. Every function $f : G \to \mathbb{C}$ defines a matrix

$$F : G \times G \to \mathbb{C}, \qquad F : (x,y) \mapsto f(y^{-1}x). \tag{7}$$

These constructions sometimes allow us to *lift* lower-bounds on the query complexity to lower-bounds on the communication complexity. Similarly, one can relate results regarding the function spaces on G to the setting of the matrix spaces on $G \times G$.

The study of lifting theorems have been a very active and successful area of theoretical computer science, particularly in the past two decades [RM97, CKLM19, HHL18, GPW18, GLM$^+$16, GPW17, GKPW17]. Not all these lifting theorems follow the above $f(y^{-1}x)$ framework , nevertheless they generally fit the theme of translating a query complexity result regarding functions $f : \mathcal{X} \to \{0,1\}$ to the communication complexity bounds on the matrices F that are constructed from f.

The XOR lift. The case of $G = \mathbb{Z}_2^n$ in (7) is closely related to the parity query complexity. The group operation on $\mathbb{Z}_2^n$ corresponds to the point-wise XOR operation on $\{0,1\}^n$, and hence for a given function $f : \{0,1\}^n \to \{0,1\}$, Equation (7) translates to $F_\oplus(x,y) = f(x \oplus y)$. The Fourier transform of f carries important information about the matrix $F_\oplus$. Indeed Fourier characters are the eigenvectors of $F_\oplus$, Fourier coefficients of f are their corresponding eigenvalues, and as a result

$$\mathrm{rk}(F_\oplus) = \mathrm{rk}_\oplus(f), \tag{8}$$

where $\mathrm{rk}_\oplus(f)$ denotes the number of non-zero Fourier coefficients of f.

The relation between parity query complexity parameters of f and their corresponding communication complexity parameters of $F_\oplus$ has been studied extensively [HHL18, TWXZ13, Zha14, ZS10, MS20, MO09].

Note that for $x, y \in \{0,1\}^n$,

$$\oplus_{i\in S}(x \oplus y)_i = (\oplus_{i\in S} x_i) \oplus (\oplus_{i\in S} y_i)\,,$$

which in particular allows one to translate every party decision tree to a communication protocol. Namely, every time that a query $\oplus_{i\in S}$ has been made in the parity decision tree, in the communication setting, the players can individually compute the two bits $\oplus_{i\in S} x_i$ and $\oplus_{i\in S} y_i$ and exchange them to find out the answer to the query on $x \oplus y$. It follows that $\mathrm{D}(F_\oplus), \mathrm{R}_0(F_\oplus), \mathrm{R}^1(F_\oplus)$, $\mathrm{R}(F_\oplus)$ are bounded respectively by $2\,\mathrm{dt}^\oplus(f), 2\,\mathrm{rdt}_0^\oplus(f), 2\,\mathrm{rdt}^{\oplus 1}(f)$, $2\,\mathrm{rdt}^\oplus(f)$.

The difficult part of establishing a lifting theorem is indeed bounding the query complexity in terms of the communication complexity. We will discuss these in Section 4.

The AND lift. In this case, we will work with the semigroup $(\{0,1\}^n, \wedge)$ where $\wedge$ corresponds to the pointwise product. Namely,

$$x \wedge y = (x_1 y_1, \ldots, x_n y_n),$$

and the lifted function is defined as

$$F_\wedge(x,y) = f(x \wedge y).$$

Similar to the XOR setting, one easily shows that $\mathrm{D}(F_\wedge), \mathrm{R}_0(F_\wedge), \mathrm{R}^1(F_\wedge)$, $\mathrm{R}(F_\wedge)$ are bounded respectively by $2\,\mathrm{dt}^\wedge(f), 2\,\mathrm{rdt}_0^\wedge(f), 2\,\mathrm{rdt}^{\wedge 1}(f)$, $2\,\mathrm{rdt}^\wedge(f)$. We will discuss the AND-lift in detail in Section 5.

2.5 Approximate norms and randomized complexity, a general approach

The study of randomized complexity classes is often naturally linked to approximate norms. For every matrix norm $\|\cdot\|$ and every $\epsilon > 0$, we define a corresponding ϵ-approximate norm for *real* matrices M as

$$\|M\|_\epsilon = \inf\{\|N\| \;:\; |M(x,y) - N(x,y)| \le \epsilon \;\; \forall x, y\},$$

where in the infimumm N is a real matrix of the same dimensions as M.

Similarly, for every norm $\|\cdot\|$ on the space of *real-valued* functions $f : \mathcal{X} \to \mathbb{R}$, we define the ϵ-approximate version of the norm as

$$\|f\|_\epsilon = \inf\{\|g\| \ : \|f-g\|_\infty \leq \epsilon,\ g : \mathcal{X} \to \mathbb{R}\}.$$

We also define the notion of the approximate $\mathcal{S}$-rank similarly:

$$\mathrm{rk}_\epsilon(\mathcal{S}, f) = \min\{\mathrm{rk}(\mathcal{S}, g) \ : \|f-g\|_\infty \leq \epsilon,\ g : \mathcal{X} \to \mathbb{R}\},$$

where we are using the notation of Definition 2.2.

We use $\mathrm{rk}_\epsilon(M)$ to denote the ϵ-rank of a real matrix M, which is the minimum rank over real matrices that approximate every entry of M to within an additive ϵ. Similar to randomized complexity measures, the choice of ϵ is not very important, as changing ϵ could only affect the value of the approximate-rank of a Boolean matrix polynomially [KS07].

Approximate norms and randomized protocols, a general approach. Suppose we are given a function $f : \mathcal{Z} \to \{0,1\}$, and we are interested in complexity of f in a randomized model of computation $\mathcal{M}$. Here $\mathcal{M}$ could be the communication complexity model, in which case we think of $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, or any of the query complexity models discussed above, in which case $\mathcal{Z} = \{0,1\}^n$.

Consider also the set of all the deterministic (query or communication) protocols π, each computing a corresponding function $\pi : \mathcal{Z} \to \{0,1\}$. Furthermore, the cost of every deterministic protocol π, denoted by $\mathrm{cost}(\pi) \in \mathbb{N}$, is the worst-case number of queries or communicated bits used by the protocol over the set of all inputs. This defines the deterministic complexity of a function f as

$$\mathrm{D}^{\mathcal{M}}(f) := \inf\{\mathrm{cost}(\pi) : \pi(z) = f(z)\ \forall z \in \mathcal{Z}\}.$$

A randomized protocol π_R is a probability distribution over deterministic protocols π_r, and the cost of a randomized protocol is defined to be the maximum cost of a deterministic protocol in its support. This leads to the notion of the randomized complexity of a function f:

$$\mathrm{R}^{\mathcal{M}}_\epsilon(f) := \inf\{\mathrm{cost}(\pi_R) : \Pr_R[\pi_R(z) \neq f(z)] \leq \epsilon\ \forall z \in \mathcal{Z}\}.$$

The following lemma provides a connection between the randomized complexity and a suitable notion of approximate norm.

Lemma 2.7 (Equivalence of $\mathrm{R}^{\mathcal{M}}_\epsilon(f)$ and $\|f\|_{\mathcal{S},\epsilon}$)**.** *Consider the setting described above. Let $\mathcal{S}$ be a spanning subset of functions $\mathcal{Z} \to \mathbb{D}$, and $\epsilon \in (0, \frac{1}{2})$ be a parameter.*

(i) If there exists an increasing function $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ such that for every function $f : \mathcal{Z} \to \{0,1\}$,

$$\|f\|_{\mathcal{S}} \leq \kappa(\mathrm{D}^{\mathcal{M}}(f)),$$

then

$$\|f\|_{\mathcal{S},\epsilon} \leq \kappa(\mathrm{R}^{\mathcal{M}}_\epsilon(f)).$$

(ii) If every $h \in \mathcal{S}$ satisfies

$$\mathrm{D}^{\mathcal{M}}(h) \leq c,$$

then

$$\mathrm{R}^{\mathcal{M}}_\epsilon(f) \leq \frac{32c\log(2/\epsilon)}{(1-2\epsilon)^2}\|f\|^2_{\mathcal{S},\epsilon}.$$

Proof. (i) Consider a randomized protocol π_R of cost at most c that computes f with two-sided error at most ϵ. Then

$$\| \mathbb{E}_R[\pi_R] - f\|_\infty \leq \epsilon,$$

while by convexity

$$\|f\|_{\mathcal{S},\varepsilon} \leq \| \mathbb{E}_R[\pi_R]\|_{\mathcal{S}} \leq \mathbb{E}_R\left[\|\pi_R\|_{\mathcal{S}}\right] \leq \max_r \|\pi_r\|_{\mathcal{S}} \leq \max_r \kappa(\mathrm{D}^{\mathcal{M}}(\pi_r)) \leq \max_r \kappa(\mathrm{cost}(\pi_r)) = \kappa(\mathrm{R}_\epsilon^{\mathcal{M}}(f)),$$

as desired.

(ii) Let $\delta = \frac{1-2\epsilon}{4}$, and let $\lambda_i \in \mathbb{C}$ and $h_i \in \mathcal{S}$ be such that for $f' = \sum_{i=1}^k \lambda_i h_i$, we have $\|f - f'\|_\infty \leq \epsilon + \delta$, and

$$L := \sum_{i=1}^k |\lambda_i| \leq \|f\|_{\mathcal{S},\epsilon}.$$

We will convert this to a randomized protocol.

For every i, define $\lambda_i' := \frac{\lambda_i}{|\lambda_i|}$, so that $|\lambda_i'| = 1$. Pick g randomly from $\{\lambda_1' h_1, \ldots, \lambda_k' h_k\}$ according to the probability distribution

$$\Pr\big[g = \lambda_i' h_i\big] = \frac{|\lambda_i|}{\sum_{i=1}^k |\lambda_i|}.$$

Note that $\mathbb{E}[g] = f'/L$, and furthermore $\|g\|_\infty \leq 1$ by our assumption about $\mathcal{S}$. Let $N = 2\delta^{-2}L^2\log(2/\epsilon) = \frac{32L^2\log(2/\epsilon)}{(1-2\epsilon)^2}$, and $g_1, \ldots, g_N$ be i.i.d. copies of g, and define $\widetilde{G} = \frac{L}{N}\sum_{i=1}^N g_i$. For every $z \in \mathcal{Z}$, by applying Hoeffding's inequality (Lemma 2.1) to the real part of $\widetilde{G}$, we have

$$\Pr\Big[|\operatorname{re}(\widetilde{G}(z)) - \operatorname{re}(f'(z))| \geq \delta\Big] < 2\exp\left(-\frac{2\delta^2}{4N\cdot(L/N)^2}\right) \leq \epsilon,$$

where the last inequality is by the choice of N. Next, let G be the Boolean rounding of $\widetilde{G}$, that is $G(z) = 1$ if and only $\operatorname{re}(\widetilde{G}(z)) \geq 1/2$. Noting that $|\operatorname{re}(f'(z)) - f(z)| \leq \epsilon + \delta$, we have

$$\Pr[G(z) \neq f(z)] \leq \Pr\left[|\operatorname{re}(\tilde{G}(z)) - \operatorname{re}(f'(z))| \geq \frac{1}{2} - \epsilon - \delta\right] \leq \Pr\Big[|\operatorname{re}(\widetilde{G}(z)) - \operatorname{re}(f'(z))| \geq \delta\Big] \leq \epsilon.$$

Note that by our assumption each h_i can be computed at cost at most c. Since $\tilde{G}(z)$ can be computed by rounding a linear combination of N such h_i's, it can be computed at cost cN. This concludes the statement. $\square$

Next we apply Lemma 2.7 to specific models of query and communication complexity.

Corollary 2.8. *For $\epsilon > 0$, let $c_\epsilon = \frac{\log(1/\varepsilon)}{(1-2\epsilon)^2}$. We have*

(a) AND*-query model:*

$$\log_3 \|f\|_{\mathcal{M}on,\varepsilon} \leq \mathrm{rdt}_\varepsilon^{\wedge}(f) \leq O\left(c_\epsilon \cdot \|f\|_{\mathcal{M}on,\varepsilon}^2\right).$$

(b) XOR*-query model:*

$$\log_2 \|f\|_{A,\varepsilon} \leq \mathrm{rdt}_\epsilon^{\oplus}(f) \leq O\left(c_\epsilon \cdot \|f\|_{A,\varepsilon}^2\right).$$

(c) Randomized communication complexity:

$$\log_2 \|F\|_{\mu,\varepsilon} \leq \mathrm{R}_\epsilon(F) \leq O\left(c_\epsilon \cdot \|F\|_{\mu,\varepsilon}^2\right),$$

which, in particular, implies

$$\log_2 \|F\|_{\gamma_2,\varepsilon} \leq \mathrm{R}_\epsilon(F) \leq O\left(c_\epsilon \cdot \|F\|_{\gamma_2,\varepsilon}^2\right),$$

Proof. (a) AND-query model: $\mathcal{Z} = \{0,1\}^n$, and $\mathcal{S} = \mathcal{M}on$.

Later in Proposition 5.1, we will prove that $\|f\|_{\mathcal{M}on} \leq 3^{\mathrm{dt}^{\wedge}(f)}$. Hence the lower bounds follows from Lemma 2.7 (i).

The upper bound follows directly from Lemma 2.7 (ii), as for every $h_S := \prod_{i \in S} x_i \in \mathcal{M}on$, $\mathrm{dt}^{\wedge}(h_S) = 1$.

(b) XOR-query model: $\mathcal{Z} = \{0,1\}^n$, and $\mathcal{S} = \{\chi_S\}_{S \subseteq [n]}$, the set of characters of $\mathbb{Z}_2^n$.

By Cauchy-Schwarz inequality $\|f\|_A \leq \sqrt{\mathrm{rk}_{\oplus}(f)} \cdot \|f\|_{L^2(\mathcal{Z})} \leq \sqrt{\mathrm{rk}_{\oplus}(f)}$, which combined with Proposition 4.1 below, gives $\|f\|_A \leq 2^{\mathrm{dt}^{\oplus}(f)}$. Now Lemma 2.7 (i) yields the lower bound.

The upper bound follows from Lemma 2.7 (ii), noting that $\mathrm{dt}^{\oplus}(\chi_S) = 1$ for all $S \subseteq [n]$.

(c) Randomized Communication Complexity: $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, $\mathcal{S} = \mathcal{R}ect$.

A communication protocol of cost c provides a partition of F into at most 2^c monochromatic rectangles, and thus $\|F\|_{\mu} \leq 2^{\mathrm{D}(F)}$. Now the lower bound follows from Lemma 2.7 (i).

The upper bound follows from Lemma 2.7 (ii) by noting that $\mathrm{D}(h) = O(1)$ for every $h \in \mathcal{R}ect$. □

2.6 Important examples: Equality, Greater-Than, Threshold Functions

In this section, we review the properties of some specific examples of matrices and functions. These will be used in later sections.

As usual denote by $\mathtt{J}_n$ the $n \times n$ all-one matrix. We start from the identity matrix.

Example 2.9 (Identity Matrix, Equality Function)**.** *The $n \times n$ identity matrix $\mathtt{I}_n$, and its complement $\overline{\mathtt{I}}_n := \mathtt{J}_n - \mathtt{I}_n$ satisfy the following.*

(i) See [KN97, Example 3.9]:

$$\mathrm{R}_0(\mathtt{I}_n) = \mathrm{R}_0(\overline{\mathtt{I}}_n) = \Theta(\log(n)).$$

(ii) See [KN97, Example 3.9]:

$$\mathrm{R}^1(\mathtt{I}_n) = \Theta(\log(n)), \text{ and } \mathrm{R}^1(\overline{\mathtt{I}}_n) = O(1),$$

In particular, $\mathrm{R}(\mathtt{I}_n) = O(1)$.

Next, we consider the *greater-than* matrix, where all the entries on the diagonal and below it are 0, and all the entries above the diagonal are 1.

Example 2.10 (Greater-than)**.** *The $n \times n$ greater-than matrix GT_n, defined as $\mathrm{GT}_n(i,j) = 1$ if and only if $i < j$, and its complement $\overline{\mathrm{GT}}_n := \mathtt{J}_n - \mathrm{GT}_n$ satisfy the following.*

(i) See [KN97, Exercise 3.10]:

$$\mathrm{R}^1(\mathrm{GT}_n) = \Omega(\log(n)), \text{ and } \mathrm{R}^1(\overline{\mathrm{GT}}_n) = \Omega(\log(n)).$$

In particular

$$\mathrm{R}_0(\mathrm{GT}_n) = \mathrm{R}_0(\overline{\mathrm{GT}}_n) = \Omega(\log(n)).$$

(ii) See [Vio15, RS15]:

$$\mathrm{R}(\mathrm{GT}_n) = \Omega(\log\log(n)).$$

Finally, we turn to threshold functions. For an integer $k \geq 0$, define the *threshold function* $\mathrm{thr}_k : \{0,1\}^n \to \{0,1\}$ as $\mathrm{thr}_k(x) = 1$ if and only if $\sum_{i=1}^n x_i \geq k$. We will also write $\overline{\mathrm{thr}}_k = 1 - \mathrm{thr}_k$.

Denote the XOR and AND-lifts of thr_k as $\mathrm{Thr}_k^{\oplus}(x,y) = \mathrm{thr}_k(x \oplus y)$ and $\mathrm{Thr}_k^{\wedge}(x,y) = \mathrm{thr}_k(x \wedge y)$, respectively. Recall that $\mathrm{rk}_{\oplus}(f)$ denotes the number of non-zero Fourier coefficients of a function $f : \{0,1\}^n \to \{0,1\}$, which is also equal to the rank of $F^{\oplus}(x,y) \coloneqq f(x \oplus y)$.

Lemma 2.11 (Threshold function in the XOR-model)**.** *For every* $0 \leq k \leq n$*, we have*

(i) $\mathrm{rdt}^{\oplus}(\mathrm{thr}_k) \leq \mathrm{rdt}^{\oplus 1}(\mathrm{thr}_k) = 2^{O(k)}$*. In particular,* $\mathrm{R}(\mathrm{Thr}_k^{\oplus}) = 2^{O(k)}$*.*

(ii) We have $\mathrm{rk}_{\oplus}(\mathrm{thr}_k) = \mathrm{rk}(\mathrm{Thr}_k^{\oplus}) \geq 2^{n/2}$*, and consequently* $\mathrm{dt}^{\oplus}(\mathrm{thr}_k) = \Omega(n)$*.*

Proof. (i) The randomized protocol will first randomly partition $\{1,\ldots,n\}$ into sets $S_1,\ldots,S_k$, where each element $j \in [n]$ is uniformly and independently assigned to one of the k sets. Next, for each $i \in [k]$, pick a subset $T_i \subseteq S_i$ uniformly at random, and query $\oplus_{j \in T_i} x_j$. Output 1 if all the queries are 1, and output 0 otherwise.

If $\mathrm{thr}_k(x) = 0$, then we will always correctly output 0, as in this case there always exists i such that $x|_{S_i}$ is all zeros. On the other hand, if $\mathrm{thr}_k(x) = 1$, with probability at least $\frac{k!}{k^k} \geq e^{-k}$, every S_i will contain at least one 1. Conditioned on the prior event, with probability at least 2^{-k} every query satisfies $\oplus_{j \in T_i} x_j = 1$, in which case the protocol correctly outputs 1. Thus, the probability of error is at most $1 - (2e)^{-k}$. Finally, by standard error-reduction, repeating this procedure $2^{O(k)}$ times can reduce the error to at most $1/3$. We conclude that there is a constant $c_k = 2^{O(k)}$ such that $\mathrm{rdt}^{\oplus 1}(\mathrm{thr}_k) = c_k$.

(ii) First note that fixing the values of variables can only decrease the size of the support of the Fourier transform. Now if $k \leq n/2$, then setting $k-1$ of the variables to 1 will result in the function that is 1 everywhere except on $\mathbf{0}$. This restricted function has a full Fourier support, which is of size $2^{n-k+1} \geq 2^{n/2}$. Similarly, if $k \geq n/2$, then setting $n-k$ of the variables to 0 yields a function which is 0 everywhere except on $\mathbf{1}$. Hence this function has a full Fourier support, which is of size $2^k \geq 2^{n/2}$.

Next, Proposition 4.1 from below implies

$$\mathrm{dt}^{\oplus}(\mathrm{thr}_k) \geq \frac{1}{2} \log \mathrm{rk}_{\oplus}(\mathrm{thr}_k) \geq \frac{n}{4}.$$

□

The threshold functions are also important instances for the AND-query model.

Lemma 2.12 (Threshold functions in AND-model [KLMY20, Example 6.3])**.** *We have*

(i) $\mathrm{dt}^{\wedge}(\mathrm{thr}_k) \geq \log \binom{n}{k} \sim n \cdot \mathrm{H}(\frac{k}{n})$*, where* H *is the binary entropy function.*

(ii) $\mathrm{rdt}^{\wedge}(\mathrm{thr}_{n-k}) = \mathrm{rdt}^{\wedge}(\overline{\mathrm{thr}}_{n-k}) \leq \mathrm{rdt}^{\wedge 1}(\overline{\mathrm{thr}}_{n-k}) = 2^{O(k)}$.
In particular, $\mathrm{R}(\mathrm{Thr}_{n-k}^{\wedge}) = 2^{O(k)}$*.*

Proof. (i) Consider an AND-decision tree T computing thr_k. It suffices to show that T has at least $\binom{n}{k}$ leaves. Let $\binom{[n]}{k}$ denote the set of all elements of Hamming weight exactly k. Note that if the output of a query $\wedge_{i \in S}$ is the same for two elements $x, y \in \{0,1\}^n$, then the query will also return the same value for $x \wedge y$. This shows that the computation in T for two distinct $x, y \in \binom{[n]}{k}$ cannot lead to the same leaf, as then $x \wedge y$ must also lead to the same leaf, but $1 = \mathrm{thr}_k(x) \neq \mathrm{thr}_k(x \wedge y) = 0$.

(ii) Note that $\overline{\mathrm{thr}}_{n-k}(x) = 1$ if and only if $x \in \{0,1\}^n$ contains at least $k+1$ 0's. We partition $[n]$ uniformly at random into $k+1$ sets $S_1,\ldots,S_{k+1}$, and query $\wedge_{j \in S_i} x_j$ for $i \in [k+1]$. If all of the queries return 0, we output 1, and otherwise we output 0. This protocol is always correct on inputs x with $\overline{\mathrm{thr}}_{n-k}(x) = 0$, and furthermore for inputs with $\overline{\mathrm{thr}}_{n-k}(x) = 1$, the probability of error is at most $1 - \frac{(k+1)!}{(k+1)^{k+1}} \leq 1 - e^{k+1}$. The claim now follows from standard error reduction. $\square$

Finally, we prove a lower-bound on the Fourier algebra norm of threshold functions.

Lemma 2.13 (Fourier algebra norm of threshold functions). *For $k \leq n/2$, we have*

$$e^{-(k-1)} \sqrt{\sum_{i=0}^{k-1} \binom{n}{i}} \leq \|\overline{\mathrm{thr}}_k\|_A \leq \sqrt{\sum_{i=0}^{k-1} \binom{n}{i}}.$$

In particular, by Corollary 3.9, the same bounds hold for $\|\overline{\mathrm{Thr}}_k^{\oplus}\|_{\mathrm{ntr}} = \|\overline{\mathrm{Thr}}_k^{\oplus}\|_{\gamma_2}$.

Proof. Define $p : \{-1,1\}^n \to \mathbb{R}$ as

$$p(y) = \sum_{\substack{S \subseteq [n] \\ |S| \leq k-1}} \prod_{i \in S} y_i,$$

and note that $p(y) = \sum_{x \in \{0,1\}^n} \overline{\mathrm{thr}}_k(x) \chi_{T_y}(x) = 2^n \widehat{\overline{\mathrm{thr}}}_k(T_y)$, where $T_y = \{i : y_i = -1\}$. Hence,

$$\|\overline{\mathrm{thr}}_k\|_A = \frac{1}{2^n} \sum_y |p(y)| = \|p\|_{L^1(\{-1,1\}^n)}.$$

By Parseval

$$\|p\|_{L^2(\{-1,1\}^n)} = \sqrt{\sum_{i=0}^{k-1} \binom{n}{i}},$$

and furthermore, since $\deg(p) \leq k-1$, by generalization of Khintchine's inequality to degree $k-1$ polynomials ([O'D14, Theorem 9.22]), we have

$$e^{-(k-1)} \|p\|_{L^2(\{-1,1\}^n)} \leq \|p\|_{L^1(\{-1,1\}^n)} \leq \|p\|_{L^2(\{-1,1\}^n)}.$$

$\square$

3 Main results: General matrices

We start by proving the results that apply to general Boolean matrices. Later, in Section 4 and Section 5, we study special classes of XOR and AND-matrices.

3.1 Blocky matrices and blocky-rank

As we have discussed earlier, EQ provides a separation between deterministic communication complexity and randomized communication complexity, in both one-sided and two-sided error models. Now suppose that we equip the players, Alice and Bob, with an *equality oracle*. To be more precise, we allow these protocols to have query nodes v, on which the players map their inputs to strings $\alpha_v(x)$ and $\beta_v(y)$, respectively, and the oracle will broadcast the value of $\mathrm{EQ}(\alpha_v(x), \beta_v(y))$ to both players. This will contribute only 1 to the communication cost. Note that the usual communicated bits can also be simulated by oracle queries. For example, if it is Alice's turn to send a bit $a_v(x)$, then they can use the query $\mathrm{EQ}(a_v(x), 1)$ to transmit this bit to Bob. Hence, in this model, we can assume that all the communication is done through oracle queries.

Obviously, having access to an equality oracle, Alice and Bob can solve EQ deterministically at cost $O(1)$, namely by querying the oracle for $\mathrm{EQ}(x, y)$.

Let $\mathrm{D}^{\mathrm{EQ}}(M)$ denote the smallest cost of a deterministic protocol with equality oracle for the matrix M.

Proposition 3.1. *Let $M : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a matrix. Then*

$$\frac{1}{2} \log \mathrm{rk}(\mathcal{Blocky}, M) \leq \mathrm{D}^{\mathrm{EQ}}(M) \leq \mathrm{rk}(\mathcal{Blocky}, M),$$

and

$$\frac{1}{2} \log \|M\|_{\mathcal{Blocky}} \leq \mathrm{D}^{\mathrm{EQ}}(M).$$

Proof. We first prove $\mathrm{D}^{\mathrm{EQ}}(M) \leq \mathrm{rk}(\mathcal{Blocky}, M)$. Let $k = \mathrm{rk}(\mathcal{Blocky}, M)$. We construct an EQ-oracle protocol for f. In advance, Alice and Bob agree on a decomposition $M = \sum_{i=1}^{k} \lambda_i M_i$, where M_i is a blocky matrix and $\lambda_i \in \mathbb{R}$ for $i \in [k]$. Since each blocky matrix M_i corresponds to an EQ query, for an input (x, y) Alice and Bob make k queries to the oracle to determine $M_1(x, y), \ldots, M_k(x, y)$. At this point both Alice and Bob can compute $M(x, y) = \sum_{i=1}^{k} \lambda_i M_i(x, y)$.

For the lower bounds, let $d = \mathrm{D}^{\mathrm{EQ}}(M)$. Consider a leaf ℓ in the EQ-oracle protocol tree computing M and let P_ℓ denote the path of length $k_\ell \leq d$ from the root to ℓ. Note that each non-leaf node v in the tree corresponds to a query to the equality oracle, and each such query corresponds to a blocky matrix B_v. For the matrix M_v, define $B_v^1 = B_v$ and $B_v^0 = \overline{B}_v = \mathrm{J}_{\mathcal{X} \times \mathcal{Y}} - B_v$.

Suppose $P_\ell = v_1, v_2, \ldots, v_{k_\ell}, \ell$, and consider the matrix

$$M_{P_\ell} \coloneqq B_{v_1}^{\sigma_{v_1}} \circ B_{v_2}^{\sigma_{v_2}} \circ \ldots \circ B_{v_{k_\ell}}^{\sigma_{v_{k_\ell}}},$$

where $\sigma_{v_i} \in \{0, 1\}$ and $\sigma_{v_i} = 1$ if and only if the edge (v_{i-1}, v_i) is labeled by 1. Hence, after simplification, M_{P_ℓ} can be written as a sum of at most 2^d summands with ± 1 coefficients, where each summand is a Schur product of at most k_l blocky matrices. Observe that the Schur product of two blocky matrices is a blocky matrix. Thus, M_{P_ℓ} can be written as a sum of at most 2^d blocky matrices with ± 1 coefficients.

Summing over all the leaves that are labeled by 1, we get

$$M = \sum_{\ell \text{ is a 1-leaf}} M_{P_\ell}.$$

As the number of leaves is bounded by 2^d, and each M_{P_ℓ} is a ± 1 linear combination of at most 2^d blocky matrices, it follows that $\mathrm{rk}(\mathcal{Blocky}, M) \leq 2^{2d}$ and $\|M\|_{\mathcal{Blocky}} \leq 2^{2d}$. $\square$

Combining the two inequalities, we have the following useful relation

$$\frac{1}{2} \log \|M\|_{\mathcal{Blocky}} \leq \mathrm{rk}(\mathcal{Blocky}, M). \tag{9}$$

The opposite direction turns out to be equivalent to Conjecture III.

Conjecture 3.2. *There exists $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ such that for a Boolean matrix M,*

$$\mathrm{rk}(\mathcal{Blocky}, M) \leq \kappa(\|M\|_{\mathcal{Blocky}}).$$

Proposition 3.3. *Conjecture 3.2 and Conjecture III are equivalent.*

Proof. Conjecture III $\implies$ Conjecture 3.2: Conjecture III implies that there is a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ such that M can be written as a sum of $\tau(\|M\|_\mu)$ blocky matrices with ± 1 coefficients. By Equation (4) and Equation (5), $\|M\|_\mu \le 4\|M\|_\nu \le 8\|M\|_{\mathcal{Blocky}}$. Hence,

$$\mathrm{rk}(\mathcal{Blocky}, M) \le \tau(8 \cdot \|M\|_{\mathcal{Blocky}}).$$

Conjecture 3.2 $\implies$ Conjecture III: By the proof of Proposition 3.1, M can be written as a sum of $2^{2\,\mathrm{D}^{\mathrm{EQ}}(M)}$ blocky matrices with ± 1 coefficients. If Conjecture 3.2 is true, then for some $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$,

$$\mathrm{D}^{\mathrm{EQ}}(M) \le \mathrm{rk}(\mathcal{Blocky}, M) \le \kappa(\|M\|_{\mathcal{Blocky}}). \tag{10}$$

Now, by the assumption of Conjecture III, $\|M\|_\mu \le c$ for some constant c. Recall from Equation (5) that $\|M\|_{\mathcal{Blocky}} \le 2\|M\|_\nu \le 2\|M\|_\mu$, so $\|M\|_{\mathcal{Blocky}} \le 2c$. Combining this with Equation (10), we conclude that M can be written as a sum of $k_c := 2^{2\kappa(c)}$ blocky matrices with ± 1 coefficients. $\square$

3.1.1 Relation to randomized communication complexity and Conjecture I

Proposition 3.4. *For a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$,*

$$\mathrm{R}(f) \le O(\mathrm{D}^{\mathrm{EQ}}(f) \cdot \log \mathrm{D}^{\mathrm{EQ}}(f)).$$

Proof. Suppose $d := \mathrm{D}^{\mathrm{EQ}}(f)$. An EQ oracle protocol tree of depth d can be used to design a randomized protocol for f: The parties simply simulate the tree, where at each node the equality oracles are simulated (up to some error probability) via an efficient randomized communication protocol for EQ. By a simple union bound, to ensure that the final error is bounded by $1/3$, it suffices to use randomized equality protocols with error at most $\frac{1}{3d}$. Recall that by Example 2.9, $\mathrm{R}(\mathrm{EQ}) = O(1)$, and thus $\mathrm{R}_{\frac{1}{2^c}}(\mathrm{EQ}) \le O(c)$. As a result, $\mathrm{R}_{\frac{1}{3d}}(\mathrm{EQ}) \le O(\log d)$ and $\mathrm{R}(f) \le O(d \log d)$. $\square$

It follows from this, and Proposition 3.1 that

$$\mathrm{R}(f) \le O(\mathrm{rk}(\mathcal{Blocky}, f) \cdot \log \mathrm{rk}(\mathcal{Blocky}, f)). \tag{11}$$

The function $\overline{\mathrm{Thr}_2^\oplus}$ from Lemma 2.11 demonstrates that the opposite relation is not true – small randomized communication does not imply having a small $\mathrm{rk}(\mathcal{Blocky}, \cdot)$. Indeed, by Lemma 2.11 (i), $\mathrm{R}(\overline{\mathrm{Thr}_2^\oplus}) = \mathrm{R}(\mathrm{Thr}_2^\oplus) = O(1)$. On the other hand, since the γ_2 norm of every blocky matrix is at most 1, by Equation (9), we have

$$\mathrm{rk}(\mathcal{Blocky}, \overline{\mathrm{Thr}_2^\oplus}) \ge \frac{1}{2} \log \|\overline{\mathrm{Thr}_2^\oplus}\|_{\mathcal{Blocky}} \ge \frac{1}{2} \log \|\overline{\mathrm{Thr}_2^\oplus}\|_{\gamma_2},$$

and by Lemma 2.13, we have

$$\log \|\overline{\mathrm{Thr}_2^\oplus}\|_{\gamma_2} \ge \Omega(\log n).$$

Remark. By the above discussion, $\overline{\mathrm{Thr}_2^\oplus}$ witnesses a gap of $O(1)$ vs. $\Omega(\log(n))$ between randomized communication complexity and deterministic communication complexity with access to equality oracle. The difference between these two parameters had also been studied in [CLV19], where a function on n bits with $\mathrm{R}(f) = O(\log n)$ and $\mathrm{D}^{\mathrm{EQ}}(f) = \Omega(n)$ is exhibited. However, the separation of [CLV19] was not ruling out a dimension-free relation between these parameters.

As Equation (11) shows, randomized communication complexity can be bounded by a function of blocky-rank, and thus it is natural to wonder whether a relaxation of Conjecture I holds for matrices bounded blocky-rank, or equivalently $\mathrm{D}^{\mathrm{EQ}}(\cdot) = O(1)$. It is not hard to see that this is indeed true.

Lemma 3.5. *If an $n \times n$ matrix M satisfies $\mathrm{rk}(\mathcal{Blocky}, M) \leq c$, then M has a monochromatic rectangle of size $\delta_c n \times \delta_c n$, where $\delta_c > 0$ only depends on c.*

Proof. We prove by induction on c that the statement is true with $\delta_c \geq 3^{-c}$. As the base case we first show that every $n \times n$ blocky matrix has an $n/3 \times n/3$ monochromatic rectangle. Suppose B is a blocky matrix with blocks $X_1 \times Y_1, \ldots, X_t \times Y_t$. We assume without loss of generality that $|\cup_i X_i| \geq 2n/3$, as otherwise $([n] \setminus \cup_i X_i) \times [n]$ contains an $n/3 \times n/3$ all-zero rectangle. Moreover, note that if for some $i \in [t]$, $|X_i| \geq n/3$, then one of $X_i \times Y_i$ or $X_i \times [n] \setminus Y_i$ contains an $n/3 \times n/3$ monochromatic rectangle. Now, suppose that for all i, $|X_i| < n/3$. This implies that there is k such that $\sum_{i=1}^{k} |X_i| \in (n/3, 2n/3)$. Note that both $(\cup_{i \leq k} X_i) \times ([n] \setminus \cup_{i \leq k} Y_i)$ and $([n] \setminus \cup_{i \leq k} X_i) \times (\cup_{i \leq k} Y_i)$ are monochromatic rectangles, and furthermore one of them contains an $n/3 \times n/3$ monochromatic rectangle.

Now suppose that M is an $n \times n$ matrix such that $M = \sum_{i=1}^{m} \lambda_i B_i$, where B_i are blocky matrices. By the base case, B_m has an $n/3 \times n/3$ monochromatic rectangle $X \times Y$. Then

$$M' \coloneqq (M - \lambda_m B_m)|_{X \times Y} = \sum_{i=1}^{m-1} \lambda_i B_i|_{X \times Y},$$

which shows $\mathrm{rk}(\mathcal{Blocky}, M') \leq c - 1$. Consequently, M' has an $\frac{|X|}{3^{c-1}} \times \frac{|Y|}{3^{c-1}}$ monochromatic rectangle, which translates to an $\frac{n}{3^c} \times \frac{n}{3^c}$ monochromatic rectangle in M. $\square$

Lemma 3.5 combined with the lower bound from Proposition 3.1 implies that a weaker version of Conjecture I holds where instead of assuming bounded randomized communication complexity, one makes the stronger assumption that $\mathrm{D}^{\mathrm{EQ}}(\cdot) = O(1)$.

3.2 Zero-error complexity and approximate-rank are qualitatively equivalent to rank

In this section, we prove that both approximate-rank, and zero-error randomized communication complexity are qualitatively equivalent to the rank, and deterministic communicating complexity.

It is known that, allowing a loss of $O(\log\log(n))$, the gap between the zero-error randomized communication complexity, and the deterministic communication complexity of an $n \times n$ matrix M can be at most quadratic [KN97, Exercise 3.15]:

$$\Omega(\sqrt{\mathrm{D}(M)} - \log\log(n)) \leq \mathrm{R}_0(M) \leq \mathrm{D}(M).$$

The above bound does not provide a dimension-free equivalence between $\mathrm{D}(M)$ and $\mathrm{R}_0(M)$ due to the $O(\log\log(n))$ term which is from applying Newman's lemma to convert zero-error private randomness to zero-error public randomness. To obtain a dimension-free equivalence, we use a different method.

Our approach is to find copies of submatrices that have large zero-error randomized communication complexity in every high-rank Boolean matrix. The following key lemma states that if the rank of a Boolean matrix is sufficiently large, then it must contain, as a submatrix, a large copy of at least on of the four matrices: the identity matrix $\mathtt{I}_k$, its complement $\overline{\mathtt{I}}_k$, greater-than function GT_k, or its complement $\overline{\mathrm{GT}}_k$.

Lemma 3.6 (Key lemma for zero-error and approximate-rank)**.** *Let M be a Boolean matrix of rank r, and let $k = \log_5(r)/4$. Then M contains a copy of at least one of $\mathtt{I}_k$, $\overline{\mathtt{I}}_k$, GT_k, or $\overline{\mathrm{GT}}_k$ as a submatrix.*

Proof. The proof is similar to the proof of the existence of Ramsey numbers. Let $R(k_1, k_2, k_3, k_4)$ be the smallest r such that every Boolean matrix of rank r, contains a copy of at least one of $\mathtt{I}_{k_1}$, $\overline{\mathtt{I}}_{k_2}$, GT_{k_3}, or $\overline{\mathrm{GT}}_{k_4}$. We will show by induction that

$$R(k_1, k_2, k_3, k_4) \le 5^{k_1+k_2+k_3+k_5}. \tag{12}$$

The base cases are when $k_i = 1$ for some $i \in \{1, \dots, 4\}$, in which case $R(k_1, k_2, k_3, k_4) \le 2$, as any matrix of rank 2 must contain both 0 and 1 entries, and thus must contain, as a submatrix, a copy of each of $\mathtt{I}_1, \overline{\mathtt{I}}_1, \mathrm{GT}_1, \overline{\mathrm{GT}}_1$.

To prove the induction step, assume $k_i \ge 2$ for all $i \in [4]$, and consider a Boolean matrix $M = [a_{ij}]_{m\times n}$ of rank at least $5^{k_1+k_2+k_3+k_4}$. Since M contains both 0's and 1's, we may assume without loss of generality that the n-th column contains both 0's and 1's. This partitions the rows of the matrix into two non-empty sets:

$$R_0 = \{i \in [m] : a_{in} = 0\} \text{ and } R_1 = \{i \in [m] : a_{in} = 1\}.$$

Let $a \in \{0, 1\}$ be chosen such that $R_a \times [n]$ corresponds to the submatrix with the larger rank, that is

$$\mathrm{rk}(M|_{R_a\times[n]}) \ge \mathrm{rk}(M)/2.$$

By permuting the rows if necessary, we can assume that $m \notin R_a$, or equivalently $a_{mn} \ne a$. Define

$$C_0 = \{j \in [n] : a_{mj} = 0\} \text{ and } C_1 = \{j \in [n] : a_{mj} = 1\}.$$

Let M_{00} be the submatrix of M on $(R_0 \cap [m-1]) \times (C_0 \cap [n-1])$, and define M_{01}, M_{10}, M_{11} similarly (see Figure 2).

For a matrix N, let $m_{\mathtt{I}}(N)$ denote the largest k such that N contains a copy of $\mathtt{I}_k$. Define $m_{\overline{\mathtt{I}}}(N)$, $m_{\mathrm{GT}}(N)$, and $m_{\overline{\mathrm{GT}}}(N)$ similarly.

If $a_{mn} = 1$, then

$$m_{\mathtt{I}}(M) \ge m_{\mathtt{I}}(M_{00}) + 1, \qquad \text{and} \qquad m_{\overline{\mathrm{GT}}}(M) \ge m_{\overline{\mathrm{GT}}}(M_{01}) + 1,$$

since one can use the last row and the last column to extend those submatrices in M_{00} and M_{01} to larger ones in M. Note also that in this case, since $a = 0$,

$$\mathrm{rk}(M_{00}) + \mathrm{rk}(M_{01}) \ge \mathrm{rk}(M|_{R_0\times[n]}) \ge \mathrm{rk}(M)/2,$$

which implies that either

$$\mathrm{rk}(M_{00}) \ge 5^{k_1+k_2+k_3+k_4-1} \ge R(k_1 - 1, k_2, k_3, k_4),$$

or

$$\mathrm{rk}(M_{01}) \ge 5^{k_1+k_2+k_3+k_4-1} \ge R(k_1, k_2, k_3, k_4 - 1).$$

In both cases, the induction hypothesis yields the desired bound Equation (12).

Similarly if $a_{mn} = 0$, then

$$m_{\overline{\mathtt{I}}}(M) \ge m_{\overline{\mathtt{I}}}(M_{11}) + 1, \qquad \text{and} \qquad m_{\mathrm{GT}}(M) \ge m_{\mathrm{GT}}(M_{10}) + 1,$$

and in this case, since $a = 1$, we obtain

$$\mathrm{rk}(M_{10}) + \mathrm{rk}(M_{11}) + 1 \geq \mathrm{rk}(M|_{R_1 \times [n]}) \geq \mathrm{rk}(M)/2,$$

which implies

$$\mathrm{rk}(M_{10}) \geq 5^{k_1+k_2+k_3+k_4-1} \geq R(k_1, k_2, k_3 - 1, k_4),$$

or

$$\mathrm{rk}(M_{11}) \geq 5^{k_1+k_2+k_3+k_4-1} \geq R(k_1, k_2 - 1, k_3, k_4).$$

Again in both cases, the induction hypothesis implies Equation (12) as desired. □

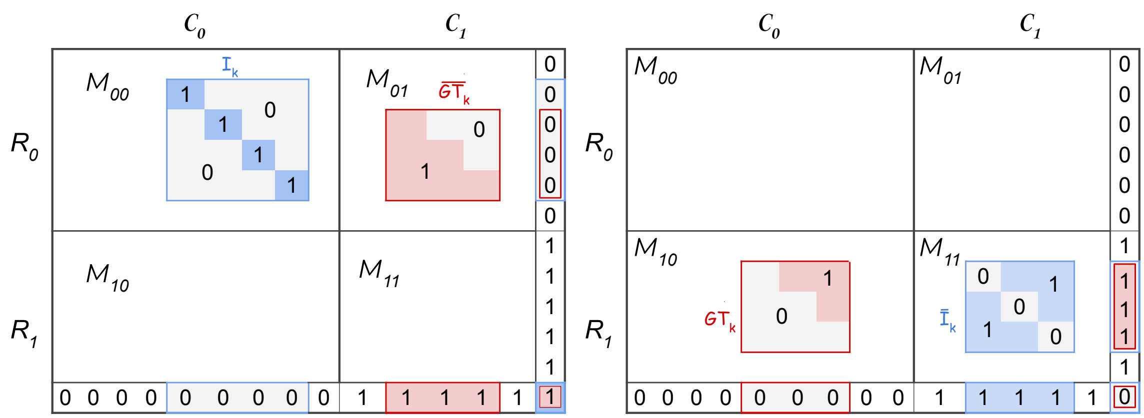

Figure 2: The matrix M with the row partitions R_0 and R_1, the column partitions C_0 and C_1, and the respective submatrices M_{00}, M_{01}, M_{10} and M_{11}. When $a_{mn} = 1$, as shown in the left figure, a copy of $\mathtt{I}_k$ in M_{00} can be extended to $\mathtt{I}_{k+1}$, and a copy of $\overline{\mathrm{GT}}_k$ in M_{01} to $\overline{\mathrm{GT}}_{k+1}$. When $a_{mn} = 0$, as in the right figure, a copy of $\overline{\mathtt{I}}_k$ in M_{11} can be extended to $\overline{\mathtt{I}}_{k+1}$, and a copy of GT_k in M_{10} to GT_{k+1}.

It was proved in [GS19] that for every Boolean matrix M, $\mathrm{rk}_\varepsilon(M) = \Omega(\log(\mathrm{rk}(M)))$. This combined with Lemma 3.6 shows that zero-error randomized communication complexity, approximate rank, and rank are all qualitatively equivalent.

Theorem 3 (Equivalence between zero-error, rank, and approximate rank)**.** *There exist a constant $c > 0$, such that for every Boolean matrix M, we have*

$$c \log \mathrm{rk}(M) \leq \mathrm{R}_0(M) \leq \mathrm{rk}(M), \tag{13}$$

and furthermore for every $\epsilon < 1/2$, there exists a constant $c_\epsilon > 0$ such that

$$c_\epsilon \log \mathrm{rk}(M) \leq \mathrm{rk}_\epsilon(M) \leq \mathrm{rk}(M). \tag{14}$$

Proof. Equation (14) is due to [GS19].

The upper bound in (13) follows from $\mathrm{R}_0(M) \leq \mathrm{D}(M)$. It remains to prove the lower-bound in (13). By Lemma 3.6, we are guaranteed to find a copy of $\mathtt{I}_k$, $\overline{\mathtt{I}}_k$, GT_k, or $\overline{\mathrm{GT}}_k$ as a submatrix in M, where $k = \frac{1}{4}\log_5 \mathrm{rk}(M)$. By Example 2.9 and Example 2.10, all the four matrices $\mathtt{I}_k$, $\overline{\mathtt{I}}_k$, GT_k, $\overline{\mathrm{GT}}_k$ have zero-error randomized communication complexity $\Omega(\log k)$, which yields the lower-bound of (13). □

3.3 One-sided error complexity

In this section, we consider one-sided error randomized protocols, and study the structure of matrices M that satisfy $\mathrm{R}^1(M) = O(1)$. As in the case of two-sided error randomized communication, the identity matrix (Example 2.9) shows that there is a gap between rank and one-sided error randomized communication complexity. The XOR lift of the threshold function also witnesses such a gap; for a constant k, we have $\mathrm{R}^1(\mathrm{Thr}_k^{\oplus}) = O(1)$ and $\mathrm{rk}(\mathrm{Thr}_k^{\oplus}) \geq 2^{\Omega(n)}$ by Lemma 2.11. These examples demonstrate that even for matrices with uniformly bounded one-sided error randomized communication complexity we cannot hope to obtain a full structure through bounded rank. Therefore, similar to the theme of Conjecture I, we focus on finding a highly structured object in such matrices.

Theorem 4 (Conjecture I for one-sided error)**.** *For every $c > 0$, there exists a constant $\delta_c > 0$ such that if the* one-sided error *randomized communication complexity $\mathrm{R}^1(M)$ of an $n \times n$ Boolean matrix M is bounded by c, then it contains an all-zero or all-one $\delta_c n \times \delta_c n$ submatrix.*

Proof. Let t be a constant to be determined later. Assume $n > 2^{\frac{c}{t}+1}$, as otherwise the claim is trivial with $\delta_c = 2^{-\frac{c}{t}-1}$. Fix a small constant $0 < \varepsilon < 2^{-\frac{2c}{t}-4}$. We will assume $|\mathrm{supp}(M)| < \varepsilon n^2$, as otherwise we can find a large all-one submatrix as follows: Given a one-sided error randomized protocol π_R for M with communication at most c, there is a fixing of the randomness r, so that $S = \{(x, y) \mid \pi_r(x, y) = 1\}$ satisfies $|S| \geq \epsilon n^2/3$, where π_r is a deterministic protocol. As π_R is a one-sided error protocol, we have $S \subseteq \mathrm{supp}(M)$. Since π_r is deterministic, then it provides a partitioning of S into at most 2^c all-one submatrices. As a result, M has an all-one submatrix of size at least $\frac{\varepsilon n^2}{3 \cdot 2^c}$.

Let S be the maximal subset of $\mathrm{supp}(M)$ such that for any distinct pairs $(x_1, y_1), (x_2, y_2) \in S$, $x_1 \neq x_2$ and $y_1 \neq y_2$. Let $r = |S|$, and note that if $r \leq 2^{\frac{c}{t}}$, then from the maximality of S it follows that deleting all the rows and columns involved in S from M will remove all the 1 entries from M. So the resulting submatrix of M will be all-zero and will have size at least $(n - 2^{\frac{c}{t}}) \times (n - 2^{\frac{c}{t}}) \geq \frac{1}{4} \cdot n^2$. Thus, we may assume $r > 2^{\frac{c}{t}}$.

Denote $k = 2^{\frac{c}{t}}$. By Example 2.9, the identity matrix is hard for one-sided randomized communication, more precisely $\mathrm{R}^1(\mathtt{I}_k) > \tau \log k$ for some constant $\tau > 0$. Fixing $t = \tau$, we get $\mathrm{R}^1(\mathtt{I}_k) > c$.

This means that M cannot contain a copy of the $k \times k$ identity matrix as a submatrix. Thus, every $k \times k$ submatrix of M that contains k entries from S must also have at least one 1-entry outside of S – call such entries off-diagonal 1's. Let m be the number of such off-diagonal 1's in M. The number of $k \times k$ submatrices of M that have k entries from S is $\binom{r}{k}$, and each of these submatrices have at least one off-diagonal 1. In this process, each off-diagonal 1 in M is counted in at most $\binom{r-2}{k-2}$ many submatrices. Hence,

$$m \geq \frac{\binom{r}{k}}{\binom{r-2}{k-2}} \geq \frac{r^2}{4k^2}.$$

Now, if $r \geq 2\sqrt{\varepsilon} k \cdot n$, then $m \geq \varepsilon n^2$, hence $|\mathrm{supp}(M)| \geq \varepsilon n^2$, which is a contradiction to our assumption of $|\mathrm{supp}(M)| < \varepsilon n^2$. So, $r < 2\sqrt{\varepsilon} k \cdot n$. In this case, by deleting all the rows and columns of S from M, we obtain an all-zero rectangle of size at least $(n - 2\sqrt{\varepsilon} k \cdot n)^2 = (1 - 2\sqrt{\varepsilon} k)^2 \cdot n^2$. To sum up, by taking $\delta_c = 1 - 2\sqrt{\varepsilon} \cdot 2^{c/t}$, we get that there is an all-zero rectangle of size at least $\delta_c^2 n^2$. $\square$

3.4 Idempotent Schur multipliers. An infinite version of Conjecture III

Let $\mathcal{X}$ and $\mathcal{Y}$ be two countable sets. Recall that a matrix $M_{\mathcal{X} \times \mathcal{Y}}$ is a Schur multiplier, if $A \mapsto M \circ A$ defines a map $B(\mathcal{H}_1, \mathcal{H}_2) \to B(\mathcal{H}_1, \mathcal{H}_2)$. In Theorem 2, we saw that M is a *contractive idempotent* of the algebra of Schur multipliers if and only if $M \in \mathcal{Blocky}$.

Consequently, if a Boolean matrix $M_{\mathcal{X}\times\mathcal{Y}}$ can be written as a linear combination of finitely many contractive idempotent Schur multipliers, then by the triangle inequality it is a Schur multiplier. More precisely, if $M = \sum_{i=1}^{t} \lambda_i M_i$ is Boolean valued and each M_i is contractive, then M is an idempotent Schur multiplier as $M \circ M = M$, and $\|M\|_m \leq \sum_{i=1}^{t} |\lambda_i|$. This leads to the following conjecture.

Conjecture 3.7. *A matrix M, finite or infinite, is an idempotent Schur multiplier if and only if M is Boolean and can be written as a linear combination of finitely many* contractive *idempotent Schur multipliers.*

A simple compactness argument shows that Conjecture 3.7 is equivalent to Conjecture III.

Theorem 5. *Conjecture 3.7 and Conjecture III are equivalent.*

Proof. By the equivalence of the norms $\|\cdot\|_\mu$ and $\|\cdot\|_m$, Conjecture III can be rephrased as follows:

For every constant c, there exists a constant k_c such that if a finite Boolean matrix M satisfies $\|M\|_m \leq c$, then there exists k_c blocky matrices B_i and signs $\sigma_i \in \{-1, 1\}$ such that

$$M = \sum_{i=1}^{k_c} \sigma_i B_i.$$

Conjecture 3.7 $\Longrightarrow$ Conjecture III: If Conjecture III is not true, then there must exist an infinite sequence of finite Boolean matrices $\{M_i\}_{i\in\mathbb{N}}$ with $\|M_i\|_m \leq k$ for all i, such that M_i cannot be expressed as a ±1-linear combination of at most i *contractive* idempotent Schur multipliers. Then $M = \oplus_{i\in\mathbb{N}} M_i$ would be an idempotent Schur multiplier, but for every $i \in \mathbb{N}$ it cannot be expressed as a ±1-linear combination of i idempotent contractions. Since M is Boolean, it follows that M cannot be expressed as a linear combination of at most a finite number of idempotent contractions.

Conjecture III $\Longrightarrow$ Conjecture 3.7: Let M be an idempotent Schur multiplier on $B(\ell_2(\mathcal{X}), \ell_2(\mathcal{Y}))$, and consider a nested sequence $X_1 \subseteq X_2 \subseteq X_3 \ldots$ of *finite* subsets of $\mathcal{X}$, and a nested sequence $Y_1 \subseteq Y_2 \subseteq Y_3 \ldots$ of *finite* subsets of $\mathcal{Y}$ such that $\mathcal{X} \times \mathcal{Y} = \bigcup \mathcal{X}_i \times \mathcal{Y}_i$. Let $M_i = \mathbf{1}_{\mathcal{X}_i\times\mathcal{Y}_i} \circ M$, which can be interpreted as a Schur multiplier on $B(\ell_2(X_i), \ell_2(Y_i))$. Since our sequences are nested, for every $i < j$, we have

$$\mathbf{1}_{X_i\times Y_i} \circ M_j = M_i. \tag{15}$$

Furthermore, $\|M_i\|_m \leq \|\mathbf{1}_{X_i\times Y_i}\|_m \cdot \|M\|_m \leq \|M\|_m$, and thus by Conjecture III, there is a constant t, depending only on $\|M\|_m$, such that $M_i = \sum_{k=1}^{t} \sigma_{i,k} N_{i,k}$ for idempotent contractions $N_{i,k}$. Furthermore by (15) for every $j > i$,

$$M_i = \sum_{k=1}^{t} \sigma_{j,k} \left(\mathbf{1}_{X_i\times Y_i} \circ N_{j,k}\right).$$

For a fixed i and k, since $N_{i,k}$, and $\mathbf{1}_{X_i\times Y_i} \circ N_{j,k}$ for all j, are supported on the finite set $X_i \times Y_i$, by restricting to a sub-sequence $i_1 < i_2 < i_3 < \ldots$, we can assume without loss of generality that for every $j \geq i$ we have

$$\mathbf{1}_{X_i\times Y_i} \circ N_{j,k} = N_{i,k}.$$

By restricting to further sub-sequences we can assume this is true for all i, and furthermore for every k, there exists a $\sigma_k \in \{-1, 1\}$ such that $\sigma_{j,k} = \sigma_k$ for all j. To summarize: for all k, and $j > i$,

$$\mathbf{1}_{X_i\times Y_i} \circ N_{j,k} = N_{i,k}, \tag{16}$$

and moreover $\sigma_{j,k} = \sigma_k$ for all j, k.

For $k \in \{1, \ldots, t\}$, define the matrix $N_k = [N_k(x, y)]_{x \in \mathcal{X}, y \in \mathcal{Y}}$ as

$$N_k(x, y) = N_{i,k}(x, y),$$

where i is any index such that $(x, y) \in X_i \times Y_i$. This is well-defined since $\mathcal{X} \times \mathcal{Y} = \bigcup X_i \times Y_i$, and (16).

Note that N_k is an idempotent contractive Schur multiplier, since, for example, it obviously does not contain any 2×2 submatrix with exactly three 1's. Moreover $M = \sum_{k=1}^{t} \sigma_k N_k$, which finishes the proof. □

3.5 Group lifts

In this section we focus on the matrices of the form $F(x, y) = f(y^{-1}x)$, where $f : G \to \{0, 1\}$, and G is a finite group. We start by showing that for any finite group G, the Fourier algebra norm of f coincides with the normalized trace norm of its lift $F(x, y) = f(y^{-1}x)$.

Proposition 3.8. *Let G be a finite group, and $f : G \to \mathbb{C}$. Let the matrix $F : G \times G \to \mathbb{C}$ be defined as $F(x, y) = f(y^{-1}x)$. We have*

$$\|f\|_A = \|F\|_{\mathrm{ntr}} := \frac{1}{|G|} \|F\|_{\mathrm{tr}}.$$

Proof. Note that the Fourier algebra norm is defined through its dual. The proof will rely on the fact that the dual of the trace norm is the operator norm $\|\cdot\|_{L^2(G) \to L^2(G)}$.

Let $h : G \to \mathbb{C}$, and the matrix H be its lift $H(x, y) = h(y^{-1}x)$. For $\nu : G \to \mathbb{C}$, note that

$$L_h \nu(x) = \frac{1}{|G|} \sum_{y \in G} h(y^{-1}x)\nu(y) = \frac{1}{|G|} \sum_{y \in G} H(x, y)\nu(y) = \frac{1}{|G|} H\nu(x).$$

Hence,

$$\frac{\|L_h \nu\|_{L^2(G)}}{\|\nu\|_{L^2(G)}} = \frac{\|L_h \nu\|_{\ell_2(G)}}{\|\nu\|_{\ell_2(G)}} = \frac{\|H\nu\|_{\ell_2(G)}/|G|}{\|\nu\|_{\ell_2(G)}},$$

which shows

$$\|L_h\|_{L^2(G) \to L^2(G)} = \frac{1}{|G|} \|H\|_{\ell_2(G) \to \ell_2(G)}.$$

Now note that

$$\langle f, h \rangle_{L^2(G)} = \frac{1}{|G|^2} \langle F, H \rangle \le \frac{1}{|G|^2} \|F\|_{\mathrm{tr}} \|H\|_{\ell_2(G) \to \ell_2(G)} = \|F\|_{\mathrm{ntr}} \|L_h\|_{L^2(G) \to L^2(G)},$$

which shows that

$$\|f\|_A = \sup \left\{ \langle f, h \rangle \ : \|L_h\|_{L^2(G) \to L^2(G)} \le 1 \right\} \le \|F\|_{\mathrm{ntr}}.$$

On the other hand, let $H : G \times G \to \mathbb{C}$ be such that

$$\|H\|_{\ell_2(G) \to \ell_2(G)} = 1 \qquad \text{and} \qquad \|F\|_{\mathrm{tr}} = \langle F, H \rangle,$$

and let $\widetilde{H} : G \times G \to \mathbb{C}$ be the following symmetrization of H:

$$\widetilde{H}(x, y) = \mathbb{E}_{z \sim G}\, H(zx, zy).$$

By convexity
$$\|\widetilde{H}\|_{\ell_2(G)\to\ell_2(G)} \leq \|H\|_{\ell_2(G)\to\ell_2(G)} = 1.$$
Define $h : G \to \mathbb{C}$ by $h(x) = \widetilde{H}(x,1)$, and note that for every y and x, $h(y^{-1}x) = \widetilde{H}(y^{-1}x,1) = \widetilde{H}(x,y)$. Since $F(zx,zy) = F(x,y) = f(y^{-1}x)$ for all z, we have

$$\begin{aligned}\langle F,H\rangle = \langle F,\widetilde{H}\rangle &= |G|^2\langle f,h\rangle_{L^2(G)}\\ &\leq |G|^2\|f\|_A\|L_h\|_{L^2(G)\to L^2(G)}\\ &= |G|\|f\|_A\|\widetilde{H}\|_{\ell_2(G)\to\ell_2(G)}\\ &\leq |G|\|f\|_A,\end{aligned}$$

this shows $\|F\|_{\mathrm{ntr}} \leq \|f\|_A$ and completes the proof. □

Davidson and Donsig [DD07] by applying a theorem of Mathias [Mat93] showed that $\|M\|_{\mathrm{ntr}} = \|M\|_m$ if the entries of M are invariant under a transitive group action.

Theorem 6 ([DD07])**.** *Let $\mathcal{X}$ be a* finite *set with a* transitive group action *G on $\mathcal{X}$. Suppose that the matrix $M_{\mathcal{X}\times\mathcal{X}}$ belongs to the commutant of the action G, or equivalently $M(x,y) = M(gx,gy)$ for all $g \in G$. Then*
$$\|M\|_{\mathrm{ntr}} = \|M\|_m = \|M\|_{\gamma_2}.$$

Combining Proposition 3.8 and Theorem 6, we obtain the following corollary.

Corollary 3.9. *Let G be a finite group, $f : G \to \mathbb{C}$, and $F : G \times G \to \mathbb{C}$ be its lift defined as $F(x,y) = f(y^{-1}x)$. We have*
$$\|F\|_m = \|F\|_{\gamma_2} = \|F\|_{\mathrm{ntr}} = \|f\|_A.$$

This corollary combined with the non-Abelian version of Cohen's idempotent theorem settles Conjecture II and Conjecture III for matrices of the form $F(x,y) = f(y^{-1}x)$.

Theorem 7. *Conjecture II and Conjecture III are true for for the class of functions $F : G \times G \to \{0,1\}$ of the form $F(x,y) = f(y^{-1}x)$, where G is a finite group, and $f : G \to \{0,1\}$.*

Proof. By Corollary 3.9,
$$\|F\|_m = \|F\|_{\gamma_2} = \|F\|_{\mathrm{ntr}} = \|f\|_A.$$
Suppose that $\|f\|_A < c$. By the general version of Cohen's idempotent theorem [San11, Theorem 1.2], there is some constant $k = k_c$, subgroups $H_1,\dots,H_k \subseteq G$, elements $a_1,\dots,a_k \in G$, and signs $\sigma_1,\dots,\sigma_k \in \{-1,1\}$ such that
$$f = \sum_{i=1}^{k} \sigma_i \mathbf{1}_{H_i a_i}.$$
Then
$$F(x,y) = \sum_{i=1}^{k} \sigma_i \times \left(\sum_{b\in H_i\backslash G} \mathbf{1}_{Hb}(x)\mathbf{1}_{a_i^{-1}Hb}(y)\right),$$
and note that each $B_i(x,y) \coloneqq \sum_{b\in H_i\backslash G} \mathbf{1}_{bHa_i}(x)\mathbf{1}_{bH}(y)$ is a blocky matrix as desired. □

4 XOR-functions

Recall that the XOR-lift of a function $f : \{0,1\}^n \to \{0,1\}$ is defined as $F_\oplus : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with $F_\oplus : (x, y) \mapsto f(x \oplus y)$.

Since XOR-lift is special case of the group lift for $G = \mathbb{Z}_2^n$, by Theorem 7, both Conjecture II, and Conjecture III are true for XOR functions.

4.1 Structure for bounded query complexity

Let $f : \{0,1\}^n \to \{0,1\}$, and consider the complexity measures

$$\mathrm{rdt}^{\oplus}(f) \le \mathrm{rdt}^{\oplus 1}(f) \le 3\,\mathrm{rdt}_0^{\oplus}(f),$$

and $\mathrm{dt}^{\oplus}(f)$. We shall study the structure of the function if we assume a uniform bound on each of these measures.

Deterministic and zero-error randomized case. The Fourier spectrum of a Boolean function plays an important role in understanding these parameters. The *Fourier rank* of f, denoted $\mathrm{rk}_\oplus(f)$, is simply the number of non-zero Fourier coefficients of f. The Fourier rank is also commonly referred to as Fourier sparsity in literature. Note that denoting $G = \mathbb{Z}_2^n$, using the notation of Definition 2.2, we have

$$\mathrm{rk}_\oplus(f) = \mathrm{rk}(\widehat{G}, f).$$

Proposition 4.1 (Equivalence between zero-error and deterministic complexities)**.** *For $f : \{0,1\}^n \to \{0,1\}$, $\mathrm{D}(F_\oplus)$, $\mathrm{rk}(F_\oplus)$, $\mathrm{R}_0(F_\oplus)$, $\mathrm{dt}^{\oplus}(f)$, $\mathrm{rk}_\oplus(f)$, and $\mathrm{rdt}_0^{\oplus}(f)$ are qualitatively equivalent. More precisely, we have*

$$\frac{1}{2}\log \mathrm{rk}_\oplus(f) \le \mathrm{dt}^{\oplus}(f) \le \mathrm{rk}_\oplus(f), \tag{17}$$

and there are constants $c_1, c_2, c_3 > 0$ such that

$$\mathrm{D}(F_\oplus) \le 2\,\mathrm{dt}^{\oplus}(f) \le c_1 \cdot \mathrm{D}(F_\oplus)^6 \le c_2 \cdot \mathrm{rk}(F_\oplus)^6 \le 2^{c_3 \cdot \mathrm{R}_0(F_\oplus)} \le 2^{2c_3\,\mathrm{rdt}_0^{\oplus}(f)} \le 2^{2c_3\,\mathrm{dt}^{\oplus}(f)}. \tag{18}$$

Proof. Equation (17): Each parity query $\oplus_{i\in S} x_i$ corresponds to querying the value of the corresponding character $\chi_S(x)$. In particular, if the Fourier spectrum of f is supported on at most c characters, then the value of $f(x)$ will be determined from the value of these characters, and thus $\mathrm{dt}^{\oplus}(f) \le \mathrm{rk}_\oplus(f)$.

For the other direction, the indicator function of every leaf of a depth d parity decision tree is determined by the value of d characters and thus has Fourier rank at most 2^d. Since the number of leaves is bounded by 2^d, we obtain $\mathrm{rk}_\oplus(f) \le 2^{2d}$.

Equation (18): The first inequality is the straightforward simulation of a parity decision tree by a communication protocol as discussed in Section 2.4, namely the fact that Alice and Bob can simulate an XOR-query $\oplus_S(x \oplus y)$ by two bits of communication $\oplus_S(x)$ and $\oplus_S(y)$. The second inequality is the parity lifting theorem of [HHL18], and the third inequality is a property of deterministic communication complexity Proposition 2.5. The fourth inequality is Theorem 3. The fifth inequality is again the simulation of parity decision trees by communication protocols. The final inequality is trivial since $\mathrm{rdt}_0^{\oplus}(f) \le \mathrm{dt}^{\oplus}(f)$. □

Remark. To prove the equivalences stated in Proposition 4.1, instead of $\mathrm{dt}^{\oplus}(f) \le c_1 \cdot \mathrm{D}(F_\oplus)^6$, it would have sufficed to use the weaker but trivial inequality $\mathrm{dt}^{\oplus}(f) \le \mathrm{rk}_\oplus(f) = \mathrm{rk}(F_\oplus) \le 2^{\mathrm{D}(F_\oplus)}$. However, the lifting theorem of [HHL18] provides stronger bounds.

One-sided randomized case. In Lemma 2.11 we saw that for a fixed integer k, the threshold function thr_k satisfies $\mathrm{rdt}^{\oplus 1}(\mathrm{thr}_k) \le c_k$ for some constant c_k depending on parameter k, while $\mathrm{dt}^{\oplus}(\mathrm{thr}_k) = \Omega(n)$. This shows that for XOR-query model the one-sided error case is not qualitatively equivalent to the zero-error and the deterministic case.

Proposition 4.2. *For every Boolean function $f : \{0,1\}^n \to \{0,1\}$, there exists an affine subspace V of co-dimension $\mathrm{rdt}^{\oplus 1}(f)$ such that f is constant on V.*

Proof. Consider a one-sided randomized parity decision tree A_R with randomness R that could only make errors when $f(x) = 1$. Suppose that $f \not\equiv 0$, as otherwise we can take $V = \{0,1\}^n$. Pick $x \in f^{-1}(1)$. Since $\Pr_R[A_R(x) = 1] > 0$, there is a fixing of randomness $R = r$, such that A_r is a deterministic parity decision tree satisfying $A_r(x) = 1$. That is, x leads to a leaf of A_r labeled with 1, and the leaf corresponds to an affine subspace V of codimension $\le \mathrm{rdt}^{\oplus 1}(f)$. Moreover, since A_r does not make errors on $f^{-1}(0)$, then $V \cap f^{-1}(0) = \emptyset$ or, equivalently, $f|_V \equiv 1$. □

Two-sided error case. Next we turn to two-sided error. We saw in Corollary 2.8 that the randomized parity decision tree complexity and the approximate Fourier algebra norm of f are qualitatively equivalent. These parameters are also qualitatively equivalent to the randomized communication complexity of the parity lift.

Proposition 4.3. *For $f : \{0,1\}^n \to \{0,1\}$ and $\epsilon \in (0, \frac{1}{2})$, $\mathrm{R}_\varepsilon(F_\oplus)$, $\mathrm{rdt}^{\oplus}_\epsilon(f)$, and $\|f\|_{A,\varepsilon}$ are qualitatively equivalent. More precisely,*

$$\log \|f\|_{A,\varepsilon} \le \mathrm{rdt}^{\oplus}_\varepsilon(f) \le O\left(c_\epsilon \|f\|^2_{A,\varepsilon}\right), \tag{19}$$

$$\frac{1}{2}\log \|f\|_{A,\varepsilon} \le \mathrm{R}_\epsilon(F_\oplus) \le O\left(c_\epsilon \|f\|^2_{A,\varepsilon}\right), \tag{20}$$

where $c_\epsilon = \frac{\log(1/\varepsilon)}{(1-2\epsilon)^2}$, and

$$\mathrm{R}_\varepsilon(F_\oplus) \le 2\,\mathrm{rdt}^{\oplus}_\varepsilon(f) \le O\left(c_\epsilon 2^{4\,\mathrm{R}_\varepsilon(F_\oplus)}\right). \tag{21}$$

Proof. Observe that a parity lift is a $y^{-1}x$-group lift for $G = \mathbb{Z}_2^n$, and thus by Corollary 3.9, we have $\|F_\oplus\|_{\gamma_2,\varepsilon} = \|f\|_{A,\varepsilon}$. Hence Equation (19) and Equation (20) have already been proven in Corollary 2.8.

The first inequality in Equation (21) is the standard simulation of a parity decision tree by a communication protocol. The second inequality in Equation (21) is a direct consequence of the upper-bound in Equation (19) and the lower-bound in Equation (20).

□

Remark. Note that Equation (19) provides an exponential lifting theorem for the randomized parity decision tree model. It is conjectured in [HHL18] that this can be improved to $\mathrm{rdt}^{\oplus}(f) \le \mathrm{R}(F_\oplus)^{O(1)}$, which remains an intriguing open problem.

Next, we observe that for the class of XOR-functions, Conjecture IV would imply Conjecture I.

Proposition 4.4. *For the class of* XOR *functions,*

$$\textit{Conjecture IV} \Rightarrow \textit{Conjecture I}.$$

Proof. Suppose that $\mathrm{R}(F_\oplus) \le c$. It follows then from Equation (20) that

$$\|f\|_{A,\epsilon} \le 2^{2c}.$$

Now if Conjecture IV is true, then f would be constant on a large subspace $V \subseteq \mathbb{Z}_2^n$. Then $V \times V$ would be a large monochromatic rectangle in $F_\oplus$. □

5 AND-functions

In this section we focus on AND-functions $F_\wedge(x,y) \coloneqq f(x \wedge y)$. As we saw in Section 4, investigating the Fourier expansion of $f : \{0,1\}^n \to \{0,1\}$ was extremely useful for understanding the properties of their XOR -lifts. This is chiefly because Fourier characters are multiplicative with respect to the XOR operation, and thus the Fourier transform naturally translates to an expansion of the matrix $F_\oplus$ as a linear combination of rank-one matrices. When studying the AND-lifts, the representation of f as a multilinear polynomial over the reals plays a similar role since monomials are multiplicative with respect to the AND operation. More precisely, using the notation $x^S = \prod_{i \in S} x_i$, the polynomial representation

$$f(x) = \sum_{S \subseteq [n]} \lambda_S x^S,$$

translates to

$$F_\wedge(x,y) = f(x \wedge y) = \sum_{S \subseteq [n]} \lambda_S x^S y^S.$$

Equivalently,

$$F_\wedge = \sum_{S \subseteq [n]} \lambda_S m_S m_S^t,$$

where $m_S(x) = x^S$. Since for each S, $m_S m_S^t$ is a rank-1 matrix, and m_S for $S \subseteq [n]$ are linearly independent, then $\mathrm{rk}(F_\wedge)$ is equal to the number of non-zero coefficients λ_S, which by the notation of Section 2.1 is denoted by $\mathrm{rk}(\mathcal{M}on, f)$. In other words,

$$\mathrm{rk}(F_\wedge) = \mathrm{rk}(\mathcal{M}on, f). \tag{22}$$

We obtain the following simple proposition, which establishes the equivalence of several parameters related to the AND-lift.

Proposition 5.1 (Equivalence between zero-error and deterministic complexities)**.** *For $f : \{0,1\}^n \to \{0,1\}$, the parameters $\mathrm{dt}^\wedge(f)$, $\mathrm{rdt}_0^\wedge(f)$, $\mathrm{rk}(\mathcal{M}on, f)$, $\|f\|_{\mathcal{M}on}$, $\mathrm{rk}(F_\wedge)$, $\mathrm{D}(F_\wedge)$, and $\mathrm{R}_0(F_\wedge)$ are all qualitatively equivalent. More precisely, there exists a constant $c > 0$ such that*

$$\log \mathrm{rk}(\mathcal{M}on, f) \le \mathrm{D}(F_\wedge) \le 2\,\mathrm{dt}^\wedge(f) \le 2\mathrm{rk}(\mathcal{M}on, f) = 2\mathrm{rk}(F_\wedge) \le 2^{c\,\mathrm{R}_0(F_\wedge)} \le 2^{2c\cdot\mathrm{rdt}_0^\wedge(f)} \le 2^{2c\cdot\mathrm{rk}(\mathcal{M}on,f)},$$

and

$$\mathrm{rk}(\mathcal{M}on, f) \le \|f\|_{\mathcal{M}on} \le 3^{\mathrm{dt}^\wedge(f)}.$$

Proof. Recall $\mathrm{rk}(F_\wedge) = \mathrm{rk}(\mathcal{M}on, f)$. Thus the inequality $\log \mathrm{rk}(\mathcal{M}on, f) \le \mathrm{D}(F_\wedge)$ is the well-known rank lower bound of Proposition 2.5, and the inequality $\mathrm{D}(F_\wedge) \le 2\,\mathrm{dt}^\wedge(f)$ is the straightforward simulation of an AND-decision tree by a communication protocol, discussed in Section 2.4.

The inequality $\mathrm{dt}^\wedge(f) \le \mathrm{rk}(\mathcal{M}on, f)$ follows from the fact that the value of a monomial can be determined by making one AND-query.

By Theorem 3, there exists a constant $c > 0$ such that

$$\mathrm{rk}(F_\wedge) \le 2^{c\,\mathrm{R}_0(F_\wedge)} \le 2^{2c\,\mathrm{rdt}_0^\wedge(f)},$$

and the last inequality in the first equation follows from $\mathrm{R}_0(F_\wedge) \le 2\,\mathrm{rdt}_0^\wedge(f) \le 2\,\mathrm{dt}^\wedge(f) \le 2\mathrm{rk}(\mathcal{M}on, f)$.

The inequality $\mathrm{rk}(\mathcal{M}on, f) \le \|f\|_{\mathcal{M}on}$ follows from the easy and well-known fact that the coefficients in the polynomial representation of f are all integers.

It remains to prove $\|f\|_{\mathcal{Mon}} \leq 3^{\mathrm{dt}^{\wedge}(f)}$. We use induction on $d = \mathrm{dt}^{\wedge}(f)$. The base case for $d = 0$ is trivial, as $\|f\|_{\mathcal{Mon}}$ is at most 1 for every constant Boolean function f. For the induction step, consider an AND-decision tree of depth d computing f, and suppose that the top node of the tree queries x^S, and branches accordingly to compute f_1 and f_2. Now

$$f(x) = x^S \cdot f_1(x) + (1 - x^S) \cdot f_2(x),$$

and since $\mathrm{dt}^{\wedge}(f_1), \mathrm{dt}^{\wedge}(f_2) \leq d - 1$, we have

$$\|f\|_{\mathcal{Mon}} \leq \|f_1\|_{\mathcal{Mon}} + \|x^S f_2\|_{\mathcal{Mon}} + \|f_2\|_{\mathcal{Mon}} \leq 3 \cdot 3^{d-1} = 3^d.$$

□

We conjecture that the exponential equivalence between $\mathrm{D}(F_\wedge)$ and $\mathrm{dt}^{\wedge}(f)$ in Proposition 5.1 can be improved to a polynomial equivalency. Recently, [KLMY20] proved $\mathrm{dt}^{\wedge}(f) = O(\mathrm{D}(f_\wedge)^3 \log n)$, but due to the $\log(n)$ factor, their statement comes short of establishing this conjecture.

Now, let us turn to randomized communication complexity and its related matrix parameters such as the trace and the γ_2 norm. Unlike Fourier characters, the monomials in the polynomial representation are not orthogonal, and thus the coefficients in the polynomial representation of f do not correspond to the eigenvalues of $F_\wedge$. This makes relating the spectral properties of $F_\wedge$ to similar properties of f difficult. For example, unlike the $F_\oplus$ case, we do not know how to verify Conjecture II or Conjecture III for matrices of the form $F_\wedge$. Similarly, we do not know how to relate the randomized communication complexity assumption of Conjecture I to an assumption about $\mathrm{rdt}^{\wedge}$. Contrast this with the XOR case where we have established that $\mathrm{R}(F_\oplus)$, $\|F_\oplus\|_{\gamma_2,\epsilon}$, $\|f\|_{A,\epsilon}$, and $\mathrm{rdt}_\oplus(f)$ are all qualitatively equivalent. We conjecture however that a similar statement is true for the AND-functions.

Conjecture 5.2. *There exist an increasing function $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ such that for every $f : \{0,1\}^n \to \{0,1\}$,*

$$\mathrm{rdt}^{\wedge}(f) \leq \kappa(\mathrm{R}(F_\wedge)).$$

Interestingly in the case of the AND-functions, we know how to establish the analogue of Conjecture IV.

Theorem 8. *Suppose $f : \{0,1\}^n \to \{0,1\}$ satisfies $\mathrm{rdt}^{\wedge}(f) \leq d$. Then, there exists a set $J \subseteq [n]$ of size at most 3^{d+1}, such that f is constant on $\{x : x_J = \mathbf{0}\}$.*

We will prove Theorem 8 in Section 5.1, but first, let us state the following corollary.

Corollary 5.3. *Conjecture 5.2, if true, would imply that Conjecture I is true for $F_\wedge$ matrices.*

Proof. It would follow from Conjecture 5.2 that if $\mathrm{R}(F_\wedge) \leq c$, then $\mathrm{rdt}^{\wedge}(f) \leq \kappa(c)$. Then by Theorem 8, f is constant on $V = \{x : x_J = \mathbf{0}\}$, where $|J| \leq 3^{\kappa(c)+1}$. Consequently, $F_\wedge$ is constant on $V \times V$, which is a $\delta 2^n \times \delta 2^n$ combinatorial rectangle with $\delta = 2^{-|J|} \geq 2^{-3^{\kappa(c)+1}}$. □

To summarize, in the case of $F_\wedge$, the missing step for establishing Conjecture I is a dimension-free lifting theorem for randomized communication complexity (i.e. Conjecture 5.2), since we know how to deduce structure from a uniform bound on randomized query complexity. In contrast, in the case of $F_\oplus$ such a lifting theorem is known, but we do not know how to establish structure from a uniform bound on randomized query complexity (i.e. Conjecture IV).

5.1 Proof of Theorem 8

By Corollary 2.8,

$$\log_3 \|f\|_{\mathcal{Mon},\varepsilon} \le \mathrm{rdt}^{\wedge}_{\varepsilon}(f) \le O\left(\|f\|^2_{\mathcal{Mon},\varepsilon} \cdot \frac{\log(1/\varepsilon)}{(1-2\epsilon)^2}\right). \tag{23}$$

Theorem 8 now follows from the first inequality and the following lemma.

Lemma 5.4. *For every $f : \{0,1\}^n \to \{0,1\}$, there exists a set $J \subseteq [n]$ of size at most $3\|f\|_{\mathcal{Mon},1/3}$, such that f is constant on $\{x : x_J = \mathbf{0}\}$.*

Proof. Let $p = \sum_{S\subseteq[n]} \lambda_S x^S$ be a multilinear polynomial satisfying $\|p - f\|_\infty \le \frac{1}{3}$ and $\|p\|_{\mathcal{Mon}} = d$.

Consider the partial ordering on the Boolean cube where $x \preceq y$ if for every i, $x_i \le y_i$. Under this ordering, pick a minimal $w \in \{0,1\}^n$ such that $f(\mathbf{0}) \ne f(w)$. This means that for every $v \prec w$, $f(v) = f(\mathbf{0})$. Pick an arbitrary j such that $w_j = 1$, and let $v = w - \mathbf{e}_j$, where $\mathbf{e}_j$ denotes the jth standard vector. Note that $|f(w) - f(v)| = 1$, and as a result $|p(w) - p(v)| \ge 1/3$, which means that

$$\sum_{S\subseteq w : S\ni j} |\lambda_S| \ge \frac{1}{3},$$

where $S \subseteq w$ means $S \subseteq \{i : w_i = 1\}$. Consequently, $\|p|_{x_j=0}\|_{\mathcal{Mon}} \le \|p\|_{\mathcal{Mon}} - \frac{1}{3}$. Thus

$$\|f|_{x_j=0}\|_{\mathcal{Mon},1/3} \le \|f\|_{\mathcal{Mon},1/3} - \frac{1}{3}.$$

We include j in J and repeat the above process, replacing f with $f|_{x_j=0}$. Since $\|\cdot\|_{\mathcal{Mon},1/3} \ge 0$, this process can be repeated for at most $3\|f\|_{\mathcal{Mon},1/3}$ times, after which we will end up with a constant function. □

5.2 Randomized AND-decision trees: One-sided and two-sided error

Let us briefly discuss $\mathrm{rdt}^{\wedge 1}$ and $\mathrm{rdt}^{\wedge}$. The example of the threshold function, as discussed in Example 2.10, shows that the one-sided and the two-sided error case are not qualitatively equivalent to the deterministic case. In particular, for $f = \overline{\mathrm{thr}}_{n-1}$, Example 2.10 shows that

$$\mathrm{R}(F_\wedge) \le 2\,\mathrm{rdt}^{\wedge}(f) \le 2\,\mathrm{rdt}^{\wedge 1}(f) = O(1), \qquad \text{while} \qquad \mathrm{dt}^{\wedge}(f) = \mathrm{dt}^{\wedge}(\overline{f}) = \Omega(\log(n)).$$

On the other hand, in Theorem 8, we showed that if $\mathrm{rdt}^{\wedge}(f) \le d$, then there exists a set $J \subseteq [n]$ of size at most 3^{d+1}, such that f is constant on $\{x : x_J = \mathbf{0}\}$. Thus for AND-functions we know how to prove the analogue of Proposition 4.2, even for two-sided error.

6 Forbidden substructures: A proof-barrier for Conjectures I, II, III

In this section, we discuss a proof barrier, which shows that the techniques used for proving Cohen's idempotent theorem, as well as many similar structural results cannot establish Conjectures I, II, and III. Such proofs are based on forbidding substructures. For instance, to prove Cohen's idempotent theorem for $f : \mathbb{Z}_2^n \to \{0,1\}$, one uses the fact that the function $g_r : \mathbb{Z}_2^r \to \{0,1\}$, defined as $g_r(x) = 1$ iff $|x| = 1$, satisfies $\|g_r\|_A = \Omega(\sqrt{r})$. Consequently, if $\|f\|_A \le c$, then no restriction of f to any affine subspace of dimension $k = k_c = O(c^2)$ can be isomorphic to g_k. One then uses the fact that f does not have a copy of this forbidden substructure to obtain general structural results about f. The proof of Cohen's theorem, even for more general groups, follows the same approach.

Similarly, in Lemma 3.6, we showed that every Boolean matrix of high rank must contain as a submatrix one of the four matrices $\mathtt{I}_k$, $\overline{\mathtt{I}}_k$, GT_k, or $\overline{\mathrm{GT}}_k$, each with large zero-error randomized communication complexity. In other words, we used these four matrices as forbidden substructures for matrices that have small zero-error randomized communication complexity. For one-sided error, in Theorem 4 we used the forbidden matrix $\mathtt{I}_k$. Note that even Sherstov's pattern-matrix method [She11], which has been used successfully to lower-bound several complexity measures of various important matrices, is based on finding certain highly symmetric patterns in them.

One may suspect that a similar approach could also be used to establish Conjectures I, II, and III. Namely, one needs to find a suitable list of matrices with high randomized communication complexity, high trace norm, or high γ_2 norm, and show that if a Boolean matrix M does not contain any of them as a submatrix, then it must have the desired structure. We prove that this approach fails as there are matrices that cannot be handled by this proof technique.

Theorem 9. . *For every sufficiently large n, there exists an $n \times n$ Boolean matrix M with the following properties.*

(i) Every $n^{1/4} \times n^{1/4}$ submatrix F of M satisfies

$$\|F\|_{\mathrm{ntr}} \le \|F\|_{\gamma_2} \le 4, \text{ and } \quad \mathrm{R}(F) = O(1).$$

(ii) M does not contain any monochromatic rectangles of size $n^{0.99} \times n^{0.99}$.

One interesting related proof that does not follow the forbidden substructure approach is the purely spectral proof of Shpilka, Tal, and Volk [STV17] for the fact that every $f : \mathbb{Z}_2^n \to \{0,1\}$ with $\|f\|_A \le c$ is constant on an affine subspace of co-dimension k_c. This obviously follows from Cohen's theorem, but [STV17] obtained stronger bounds on k_c.

Before stating the proof of Theorem 9, we will set up and prove an auxiliary lemma on the blocky-rank of matrices that correspond to forests. A matrix $M : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ naturally corresponds to a bipartite graph G_M with bipartition $\mathcal{X} \cup \mathcal{Y}$, where there is an edge between vertices $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ if and only if $M(x,y) = 1$. Note that the bipartite graph corresponding to a blocky matrix M is an edge-disjoint union of vertex-disjoint complete bipartite graphs.

Recall that a graph is called a *forest* if it does not contain any cycles. A connected forest is called a *tree*.

Lemma 6.1. *Let M be a finite Boolean matrix corresponding to a forest. Then M is a sum of two blocky matrices.*

Proof. As mentioned above, a blocky matrix corresponds to an edge-disjoint union of vertex-disjoint complete bipartite graphs. Hence it suffices to show that the edges of every forest can be partitioned into two sets, each forming a disjoint union of complete bipartite graphs. Obviously, it suffices to prove this for a tree as a forest is a disjoint union of trees. Let v be an arbitrary vertex of the tree, and for $i = 0, 1, \ldots$, let L_i be the set of the vertices that are in distance i from v. To complete the proof note that the edges between L_i and L_{i+1} for even values of i form one blocky matrix, and similarly the edges between L_i and L_{i+1} for odd values of i form the other blocky matrix. $\square$

Proof of Theorem 9. Set $p = \frac{n^{0.05}}{n}$, and select a random $n \times n$ matrix $M = [m_{ij}]$ by setting each entry to 1 with probability p and independently of other entries. It suffices to show that with probability $1 - o(1)$ both (i) and (ii) hold.

(i) Let $k = n^{1/4}$. We will show that every $n^k \times n^k$ submatrix F of M can be written as a sum of four blocky matrices. Then $\mathrm{R}(F) = O(1)$ immediately follows from Equation (11), and $\|F\|_{\mathrm{ntr}} \leq \|F\|_{\gamma_2} \leq 4$ follows from the fact that the γ_2-norm of a blocky matrix is at most 1.

We first prove that with probability $1 - o(1)$, for every $r, t \leq k$, every $r \times t$ submatrix of M contains a row or a column with at most two 1's. Note that the statement is trivial when $\min(r, t) \leq 2$. Fix $r, t > 2$, and assume without loss of generality that $r \leq t$. The probability that there is an $r \times t$ submatrix such that each of its t columns contains at least three 1's is bounded by

$$\binom{N}{r}\binom{N}{t}\left(\binom{r}{3}p^3\right)^t \leq n^r n^t (r^3 p^3)^t \leq (n^2 p^3 t^3)^t \leq \left(\frac{n^{0.15}}{n^{1/4}}\right)^t \leq o(n^{-1/2}).$$

Thus by a union bound over all choices of $r, t \leq k$, the probability that there is $r, t \in [k]$ and an $r \times t$ submatrix where every column contains at least three 1's is bounded by $o(k^2 n^{-1/2})$ which is $o(1)$ as desired.

Now suppose that every $r \times t$ submatrix F of M contains a row or a column with at most two 1's. We will show that in this case, every such F is a disjoint union of two forests, and by Lemma 6.1 M is a sum of four blocky matrices. Consider a row (or a column) with at most two 1's, and let e_1 and e_2 be the edges corresponding to these (at most) two entries. Removing this row from F will result in a smaller submatrix, which by induction hypothesis, can be written as the union of two forests $\mathcal{F}_1$ and $\mathcal{F}_2$. Now F can be decomposed into the union of two forests $\mathcal{F}_1 \cup \{e_1\}$ and $\mathcal{F}_2 \cup \{e_2\}$.

(ii) Let $K = n^{0.99}$. The expected number of monochromatic rectangles of size $K \times K$ is at most

$$2^n \times 2^n \times \left(p^{K^2} + (1-p)^{K^2}\right) \leq 2^{2N}(2e^{-pK^2}) \leq 2^{1+2n-pK^2+1} = 2^{2+2n-n^{0.98+0.05}} = o(1).$$

□

References

[Alo86] Noga Alon. Covering graphs by the minimum number of equivalence relations. *Combinatorica*, 6(3):201–206, 1986.

[BFS86] László Babai, Péter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *FOCS*, pages 337–347. IEEE Computer Society, 1986.

[BH04] William D. Banks and Asma Harcharras. On the norm of an idempotent Schur multiplier on the Schatten class. *Proc. Amer. Math. Soc.*, 132(7):2121–2125, 2004.

[BK95] Aart Blokhuis and Ton Kloks. On the equivalence covering number of splitgraphs. *Information Processing Letters*, 54(5):301–304, 1995.

[BN07] Béla Bollobás and Vladimir Nikiforov. Cliques and the spectral radius. *Journal of Combinatorial Theory, Series B*, 97(5):859–865, 2007.

[Chu14] Maria Chudnovsky. The Erdös-Hajnal conjecture—a survey. *J. Graph Theory*, 75(2):178–190, 2014.

[CKLM19] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudo-random properties. *computational complexity*, 28(4):617–659, 2019.

[CLV19] Arkadev Chattopadhyay, Shachar Lovett, and Marc Vinyals. Equality alone does not simulate randomness. In *34th Computational Complexity Conference (CCC 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[Coh60] Paul J. Cohen. On a conjecture of Littlewood and idempotent measures. *Amer. J. Math.*, 82:191–212, 1960.

[CR12] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-Hamming-distance. *SIAM J. Comput.*, 41(5):1299–1317, 2012.

[DD07] Kenneth R. Davidson and Allan P. Donsig. Norms of Schur multipliers. *Illinois J. Math.*, 51(3):743–766, 2007.

[Duc79] Pierre Duchet. *Représentations, noyaux en théorie des graphes et hypergraphes.* PhD thesis, 1979.

[ELT16] G. K. Eleftherakis, R. H. Levene, and I. G. Todorov. Schur idempotents and hyperreflexivity. *Israel J. Math.*, 215(1):317–337, 2016.

[Eym64] Pierre Eymard. L'algèbre de Fourier d'un groupe localement compact. *Bull. Soc. Math. France*, 92:181–236, 1964.

[For02] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. volume 65, pages 612–625. 2002. Special issue on complexity, 2001 (Chicago, IL).

[Fra82] Péter Frankl. Covering graphs by equivalence relations. In *North-Holland Mathematics Studies*, volume 60, pages 125–127. Elsevier, 1982.

[GKPW17] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for P^{NP}. 79:Art. No. 12, 16, 2017.

[GLM+16] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM Journal on Computing*, 45(5):1835–1869, 2016.

[GN08] C. D. Godsil and M. W. Newman. Eigenvalue bounds for independent sets. *J. Combin. Theory Ser. B*, 98(4):721–734, 2008.

[GPW17] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. pages 132–143, 2017.

[GPW18] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. *SIAM Journal on Computing*, 47(6):2435–2450, 2018.

[Gro52] A. Grothendieck. Résumé des résultats essentiels dans la théorie des produits tensoriels topologiques et des espaces nucléaires. *Ann. Inst. Fourier (Grenoble)*, 4:73–112 (1954), 1952.

[GS08] Ben Green and Tom Sanders. A quantitative version of the idempotent theorem in harmonic analysis. *Ann. of Math. (2)*, 168(3):1025–1054, 2008.

[GS19] Anna Gál and Ridwan Syed. Upper bounds on communication in terms of approximate rank. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 26, page 6, 2019.

[HHL18] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. *SIAM Journal on Computing*, 47(1):208–217, 2018.

[HMP86] B. Host, J.-F. Méla, and F. Parreau. Analyse harmonique des mesures. *Astérisque*, (135-136):261, 1986.

[Juk06] Stasys Jukna. On graph complexity. *Comb. Probab. Comput.*, 15(6):855–876, 2006.

[KI40] Yukiyosi Kawada and Kiyosi Itô. On the probability distribution on a compact group. I. *Proc. Phys.-Math. Soc. Japan (3)*, 22:977–998, 1940.

[KLMY20] Alexander Knop, Shachar Lovett, Sam McGuire, and Weiqiang Yuan. Log-rank and lifting for AND-functions, 2020.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, Cambridge, 1997.

[KP05] Aristides Katavolos and Vern I. Paulsen. On the ranges of bimodule projections. *Canad. Math. Bull.*, 48(1):97–111, 2005.

[Kri79] J.-L. Krivine. Constantes de Grothendieck et fonctions de type positif sur les sphères. *Adv. in Math.*, 31(1):16–30, 1979.

[KS07] Adam R Klivans and Alexander A Sherstov. A lower bound for agnostically learning disjunctions. In *International Conference on Computational Learning Theory*, pages 409–423. Springer, 2007.

[Lef72] Marcel Lefranc. Sur certaines algèbres de fonctions sur un groupe. *C. R. Acad. Sci. Paris Sér. A-B*, 274:A1882–A1883, 1972.

[Lev14] Rupert H. Levene. Norms of idempotent Schur multipliers. *New York J. Math.*, 20:325–352, 2014.

[Liv95] Leo Livshits. A note on 0-1 Schur multipliers. *Linear Algebra Appl.*, 222:15–22, 1995.

[LLT07] Mei Lu, Huiqing Liu, and Feng Tian. Laplacian spectral bounds for clique and independence numbers of graphs. *Journal of Combinatorial Theory, Series B*, 97(5):726–732, 2007.

[Lov14] Shachar Lovett. Recent advances on the log-rank conjecture in communication complexity. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (112):18–35, 2014.

[LS07] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Found. Trends Theor. Comput. Sci.*, 3(4):front matter, 263–399 (2009), 2007.

[LS09] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures & Algorithms*, 34(3):368–394, 2009.

[LSG12] Xueliang Li, Yongtang Shi, and Ivan Gutman. *Graph energy.* Springer, New York, 2012.

[Mé82] J.-F. Méla. Mesures ε-idempotentes de norme bornée. *Studia Math.*, 72(2):131–149, 1982.

[Mat93] Roy Mathias. The Hadamard operator norm of a circulant and applications. *SIAM J. Matrix Anal. Appl.*, 14(4):1152–1167, 1993.

[MO09] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *arXiv preprint arXiv:0909.3392*, 2009.

[MP16] Jayden Mudge and Hung Le Pham. Idempotents with small norms. *J. Funct. Anal.*, 270(12):4597–4603, 2016.

[MS20] Nikhil S Mande and Swagato Sanyal. On parity decision trees for Fourier-sparse Boolean functions. *arXiv preprint arXiv:2008.00266*, 2020.

[Neu06] Stefan Neuwirth. Cycles and 1-unconditional matrices. *Proc. London Math. Soc. (3)*, 93(3):761–790, 2006.

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Inform. Process. Lett.*, 39(2):67–71, 1991.

[Nik06] Vladimir Nikiforov. The smallest eigenvalue of K_r-free graphs. *Discrete Math.*, 306(6):612–616, 2006.

[Nik09] Vladimir Nikiforov. More spectral bounds on the clique and independence numbers. *Journal of Combinatorial Theory, Series B*, 99(6):819–826, 2009.

[Nis91] Noam Nisan. CREW PRAMs and decision trees. *SIAM J. Comput.*, 20(6):999–1007, 1991.

[O'D14] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, New York, 2014.

[Pis12] Gilles Pisier. Grothendieck's theorem, past and present. *Bull. Amer. Math. Soc. (N.S.)*, 49(2):237–323, 2012.

[PR94] Pavel Pudlák and Vojtech Rödl. Some combinatorial-algebraic problems from complexity theory. *Discrete Mathematics*, 1(136):253–279, 1994.

[RM97] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 234–243. IEEE, 1997.

[RS15] Sivaramakrishnan Natarajan Ramamoorthy and Makrand Sinha. On the communication complexity of greater-than. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 442–444. IEEE, 2015.

[San11] Tom Sanders. A quantitative version of the non-abelian idempotent theorem. *Geom. Funct. Anal.*, 21(1):141–221, 2011.

[San20] Tom Sanders. Bounds in Cohen's idempotent theorem. *J. Fourier Anal. Appl.*, 26(2):Paper No. 25, 64, 2020.

[She11] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

[She14] Alexander A. Sherstov. Communication complexity theory: thirty-five years of set disjointness. In *Mathematical foundations of computer science 2014. Part I*, volume 8634 of *Lecture Notes in Comput. Sci.*, pages 24–43. Springer, Heidelberg, 2014.

[STV17] Amir Shpilka, Avishay Tal, and Ben Lee Volk. On the structure of Boolean functions with small spectral norm. *Comput. Complex.*, 26(1):229–273, March 2017.

[TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, FOCS '13, pages 658–667, Washington, DC, USA, 2013. IEEE Computer Society.

[Vio15] Emanuele Viola. The communication complexity of addition. *Combinatorica*, 35(6):703–747, 2015.

[Yao77] Andrew Chi Chih Yao. Probabilistic computations: toward a unified measure of complexity (extended abstract). In *18th Annual Symposium on Foundations of Computer Science (Providence, R.I., 1977)*, pages 222–227. 1977.

[Zha14] Shengyu Zhang. Efficient quantum protocols for XOR-functions. In *Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pages 1878–1885, Philadelphia, PA, USA, 2014. Society for Industrial and Applied Mathematics.

[ZS10] Zhiqiang Zhang and Yaoyun Shi. On the parity complexity measures of Boolean functions. *Theoretical Computer Science*, 411(26-28):2612–2618, 2010.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il