Foundations and Trends® in
sample
Vol. xx, No xx (xxxx) 1–135

DOI: xxxxxx

now
the essence of knowledge

Lower Bounds in Communication Complexity

Troy Lee[1] and Adi Shraibman[2]

[1] *Columbia University New York 10027, United States, troyjlee@gmail.com*
[2] *Weizmann Institute Rehovot 76100, Israel, adi.shribman@gmail.com*

Abstract

The communication complexity of a function $f(x, y)$ measures the number of bits that two players, one who knows x and the other who knows y, must exchange to determine the value $f(x, y)$. Communication complexity is a fundamental measure of complexity of functions. Lower bounds on this measure lead to lower bounds on many other measures of computational complexity. This monograph survey lower bounds in the field of communication complexity. Our focus is on lower bounds that work by first representing the communication complexity measure in Euclidean space. That is to say, the first step in these lower bound techniques is to find a geometric complexity measure, such as rank or trace norm, that serves as a lower bound to the underlying communication complexity measure. Lower bounds on this geometric complexity measure are then found using algebraic and geometric tools.

Contents

1

Introduction

Communication complexity studies how much communication is needed in order to evaluate a function whose output depends on information distributed amongst two or more parties. Yao [Yao79] introduced an elegant mathematical framework for the study of communication complexity, applicable in numerous situations, from an email conversation between two people, to processors communicating on a chip. Indeed, the applicability of communication complexity to other areas, including circuit and formula complexity, VLSI design, proof complexity, and streaming algorithms, is one reason why it has attracted so much study. See the excellent book of Kushilevitz and Nisan [KN97] for more details on these applications and communication complexity in general.

Another reason why communication complexity is a popular model for study is simply that it is an interesting mathematical model. Moreover, it has that rare combination in complexity theory of a model for which we can actually hope to show tight lower bounds, yet these bounds often require the development of nontrivial techniques and sometimes are only obtained after several years of sustained effort.

In the basic setting of communication complexity, two players Alice and Bob wish to compute a function $f : X \times Y \to \{T, F\}$ where X, Y

are arbitrary finite sets. Alice holds an input $x \in X$, Bob $y \in Y$, and they wish to evaluate $f(x, y)$ while minimizing the number of bits communicated. We let Alice and Bob have arbitrary computational power as we are really interested in how much information must be exchanged in order to compute the function, not issues of running time or space complexity.

Formally, a communication protocol is a binary tree where each internal node v is labeled either by a function $a_v : X \to \{0, 1\}$ or a function $b_v : Y \to \{0, 1\}$. Intuitively each node corresponds to a turn of either Alice or Bob to speak. The function a_v indicates, for every possible input x, how Alice will speak if the communication arrives at that node, and similarly for b_v. The leaves are labeled by an element from $\{T, F\}$. On input x, y the computation traces a path through the tree as indicated by the functions a_v, b_v. The computation proceeds to the left child of a node v if $a_v(x) = 0$ and the right child if $a_v(x) = 1$, and similarly when the node is labeled by b_v. The protocol correctly computes f if for every input x, y, the computation arrives at a leaf ℓ labeled by $f(x, y)$.

The cost of a protocol is the height of the protocol tree. The deterministic communication complexity of a function f, denoted $D(f)$, is the minimum cost of a protocol correctly computing f. Notice that, as we have defined things, the transcript of the communication defines the output, thus both parties "know" the answer at the end of the protocol. One could alternatively define a correct protocol where only one party needs to know the answer at the end, but this would only make a difference of one bit in the communication complexity.

If we let $n = \min\{\lceil \log |X| \rceil, \lceil \log |Y| \rceil\}$ then clearly $D(f) \leq n + 1$ as either Alice or Bob can simply send their entire input to the other, who can then compute the function and send the answer back. We refer to this as the trivial protocol. Thus the communication complexity of f will be a natural number between 1 and $n + 1$, and our goal is to determine this number. This can be done by showing a lower bound on how much communication is needed, and giving a protocol of matching complexity.

The main focus of this survey is on showing lower bounds on the communication complexity of explicit functions. We treat different vari-

ants of communication complexity, including randomized, quantum, and multiparty models. Many tools have been developed for this purpose from a diverse set of fields including linear algebra, Fourier analysis, and information theory. As is often the case in complexity theory, demonstrating a lower bound is usually the more difficult task.

One of the most important lower bound techniques in communication complexity is based on matrix rank. In fact, it is not too much of an exaggeration to say that a large part of communication complexity is the study of different variants of matrix rank. To explain the rank bound, we must first introduce the *communication matrix*, a very useful and common way of representing a function $f : X \times Y \to \{T, F\}$. We will consider both a *Boolean* and a *sign* version of the communication matrix, the difference being in the particular integer representation of $\{T, F\}$. A Boolean matrix has all entries from $\{0, 1\}$, whereas a sign matrix has entries from $\{-1, +1\}$. The Boolean communication matrix for f, denoted B_f, is a $|X|$-by-$|Y|$ matrix where $B_f[x, y] = 1$ if $f(x, y) = T$ and $B_f[x, y] = 0$ if $f(x, y) = F$. The sign communication matrix for f, denoted A_f, is a $\{-1, +1\}$-valued matrix where $A_f[x, y] = -1$ if $f(x, y) = T$ and $A_f[x, y] = +1$ if $f(x, y) = F$. Depending on the particular situation, it can be more convenient to reason about one representation or the other, and we will use both versions throughout this survey. Fortunately, this choice is usually simply a matter of convenience and not of great consequence—it can be seen that they are related as $B_f = (J - A_f)/2$, where J is the all-ones matrix. Thus the matrix rank of the two versions, for example, will differ by at most one.

Throughout this survey we identify a function $f : X \times Y \to \{T, F\}$ with its corresponding (sign or Boolean) communication matrix. The representation of a function as a matrix immediately puts tools from linear algebra at our disposal. Indeed, Mehlhorn and Schmidt [MS82] showed how matrix rank can be used to lower bound deterministic communication complexity. This lower bound follows quite simply from the properties of a deterministic protocol, but we delay a proof until Chapter 2.

Theorem 1.1 (Mehlhorn and Schmidt [MS82]). For every sign

matrix A,

$$\log \operatorname{rank}(A) \leq D(A).$$

The rank bound has nearly everything one could hope for in a lower bound technique. From a complexity point of view it can be efficiently computed, i.e. computed in time polynomial in the size of the matrix. Furthermore, it frees us from thinking about communication protocols and lets us just consider the properties of A as a linear operator between Euclidean spaces, with all the attendant tools of linear algebra to help in doing this. Finally, it is even conjectured that one can always show polynomially tight bounds via the rank method. This log rank conjecture is one of the greatest open problems in communication complexity.

Conjecture 1.1 (Lovász and Saks [LS88]). There is a constant c such that for every sign matrix A

$$D(A) \leq (\log \operatorname{rank}(A))^c + 2.$$

The additive term is needed because a rank-one sign matrix can require two bits of communication. Thus far the largest known separation between log rank and deterministic communication, due to Nisan and Wigderson [NW95], shows that in Conjecture 1.1 the constant c must be at least $1.63\ldots$

The problems begin, however, when we start to study other models of communication complexity such as randomized, quantum, or multiparty variants. Here one can still give a lower bound in terms of an appropriate variation of rank, but the bounds now can become very difficult to evaluate. In the case of multiparty complexity, for example, the communication matrix becomes a communication tensor, and one must study tensor rank. Unlike matrix rank, the problem of computing tensor rank is NP-hard [Hås90], and even basic questions like the largest possible rank of an n-by-n-by-n real tensor remain open.

For randomized or quantum variants of communication complexity, as shown by Krause [Kra96] and Buhrman and de Wolf [BW01] respectively, the relevant rank bound turns out to be *approximate rank*.

Definition 1.1. Let A be a sign matrix. The approximate rank of A with approximation factor α, denoted $\text{rank}^{\alpha}(A)$, is

$$\text{rank}^{\alpha}(A) = \min_{B:1\leq A[i,j]B[i,j]\leq\alpha} \text{rank}(B).$$

As we shall see in Chapter 4 and Chapter 5, the logarithm of approximate rank is a lower bound on randomized and quantum communication complexity, where the approximation factor α relates to the success probability of the protocol. In analogy with the log rank conjecture, it is also reasonable to conjecture here that this bound is polynomially tight.

Approximate rank, however, can be quite difficult to compute. While we do not know if it is NP-hard, similar rank minimization problems subject to linear constraints are NP-hard, see for example section 7.3 of [VB96]. Part of this difficulty stems from the fact that approximate rank is an optimization problem over a nonconvex function.

This brings us to the main theme of our survey. We focus on lower bound techniques which are real-valued functions and ideally possess some "nice" properties, such as being convex. The development and application of these techniques follows a three-step approach which we now describe. This approach can be applied in much the same way for different models, be they randomized, quantum, or multiparty.

Say that we are interested in a complexity measure CC, a mapping from functions to the natural numbers, which could represent any one of the above models.

(1) Embed the problem in $\mathbb{R}^{m\times n}$. That is, find a function $\mathcal{G} : \mathbb{R}^{m\times n} \to \mathbb{R}$ such that

$$\mathcal{G}(A) \leq \text{CC}(A),$$

for every sign matrix A. As is the case with rank and approximate rank, often $\mathcal{G}$ will itself be naturally phrased as a minimization problem.

(2) Find an equivalent formulation of $\mathcal{G}$ in terms of a maximization problem. This will of course not always be possible, as

in the case of approximate rank. This can be done, however, for rank and for a broad class of optimization problems over convex functions.

(3) Prove lower bounds on $\mathcal{G}$ by exhibiting an element of the feasible set for which the objective function is large. We call such an element a *witness* as it witnesses that $\mathcal{G}$ is at least as large as a certain value.

We will delay most of the technical details of this approach to the main body of the survey, in particular to Chapter 6 where we discuss the use of duality to perform the key step 2 to go from a "min" formulation to a "max" formulation. Here we limit ourselves to more general comments, providing some intuition as to why and in what circumstances this approach is useful.

Step 1 We are all familiar with the idea that it can be easier to find the extrema of a smooth real-valued function than a discrete valued function. For example, for smooth functions the powerful tools of calculus are available. To illustrate, think of integer programming vs. linear programming. The latter problem can be solved in polynomial time, while even simple instances of integer programming are known to be NP-hard.

The intuition behind the first step is the same. The complexity of a protocol is a discrete valued function, so in determining communication complexity we are faced with an optimization problem over a discrete valued function. By working instead with a real valued lower bound $\mathcal{G}$ we will have more tools at our disposal to evaluate $\mathcal{G}$. Moreover, if $\mathcal{G}$ is "nice"—for example being an optimization problem over a convex function—then the set of tools available to us is particularly rich. For instance, we can use duality to enact step 2.

We do potentially pay a price in performing Step 1 and working with a "nicer" function $\mathcal{G}$. It could be the case that $\mathcal{G}(A)$ is much smaller than $\mathrm{CC}(A)$ for some sign matrices A. Just as in approximation algorithms, we seek a bound that is not only easier to compute but also approximates $\mathrm{CC}(A)$ well. We will say that a representation $\mathcal{G}(A)$ is *faithful* if there is some constant k such that $\mathrm{CC}(A) \leq \mathcal{G}(A)^k$ for all sign matrices

A.

Step 2 A communication complexity measure $\mathrm{CC}(A)$ is naturally phrased as a minimization problem—looking for a protocol of minimum cost. Often times, as with the case of approximate rank, our lower bound $\mathcal{G}$ is also naturally phrased as a minimization problem.

The difficulty, of course, is that to lower bound a minimization problem one has to deal with the universal quantifier $\forall$—we have to show that every possible protocol requires a certain amount of communication.

When our complexity measure $\mathcal{G}$ is of a nice form, however, such as a minimization problem of a convex function, we can hope to find an *equivalent* formulation of $\mathcal{G}$ in terms of a maximization problem. A maximization problem is much easier to lower bound since we simply have to demonstrate a particular feasible instance for which the target function is large. In some sense this can be thought of as an "algorithmic approach" to lower bounds. In Chapter 6 we will show how this can be done for a large class of complexity measures known as *approximate norms*.

This is an instance of a more general phenomena: showing a statement about *existence* is often easier than proving a statement about *nonexistence*. The former can be certified by a witness, which we do not always expect for the latter. Take the example of graph planarity, i.e. the question of whether a graph can be drawn in the plane in such a way that its edges intersect only at their endpoints. While it can be tricky to find such a drawing, at least we know what form the answer will take. To show that a graph is nonplanar, however, seems like a much more daunting task unless one has heard of Kuratowski's Theorem or Wagner's Theorem. These theorems reduce the problem of nonexistence to that of existence: for example, Wagner's theorem states that a graph is nonplanar if and only if it contains K_5, the complete graph on five vertices, or $K_{3,3}$ the complete three-by-three bipartite graph, as a minor. Not surprisingly, theorems of this flavor are key in efficient algorithmic solutions to planarity and nonplanarity testing.

Step 3 Now that we have our complexity measure $\mathcal{G}$ phrased in terms of a maximization problem, we are in much better shape. Any element from the feasible set can be used to show a lower bound, albeit not necessarily a good one. As a simple example, going back to the rank lower bound, observe that a natural way to prove a lower bound on rank is to find a large set of columns (or rows) that are independent.

Finding a good witness to prove a lower bound for a certain complexity measure $\mathcal{G}$ can still be a very difficult task. This is the subject we take up in Chapter 7. There are still only a few situations where we know how to choose a good witness, but this topic has recently seen a lot of exciting progress and more is certainly still waiting to be discovered.

Approximate norms The main example of the three-step approach we study in this survey is for *approximate norms*. We now give a more technical description of this case; the reader can skip this section at first reading, or simply take it as an "impression" of what is to come.

Let Φ be any norm on $\mathbb{R}^{m \times n}$, and let $\alpha \geq 1$ be a real number. The α-*approximate norm* of an $m \times n$ sign matrix A is

$$\Phi^{\alpha}(A) = \min_{B : 1 \leq A[i,j]B[i,j] \leq \alpha} \Phi(B).$$

The limit as $\alpha \to \infty$ motivates the definition

$$\Phi^{\infty}(A) = \min_{B : 1 \leq A[i,j]B[i,j]} \Phi(B).$$

In Step 1 of the framework described above we will usually take $\mathcal{G}(A) = \Phi^{\alpha}(A)$ for an appropriate norm Φ. We will see that the familiar matrix trace norm is very useful for showing communication complexity lower bounds, and develop some more exotic norms as well. We discuss this step in each of the model specific chapters, showing which norms can be used to give lower bounds on deterministic (Chapter 2), nondeterministic (Chapter 3), randomized (Chapter 4), quantum (Chapter 5), and multiparty (Chapter 8) models.

The nice thing about taking $\mathcal{G}$ to be an approximate norm is that we can implement Step 2 of this framework in a general way. As described in Chapter 6, duality can be applied to yield an *equivalent* formulation for any approximate norm Φ^{α} in terms of a maximization. Namely, for

a sign matrix A

$$\Phi^{\alpha}(A) = \max_{W} \frac{(1+\alpha)\langle A, W\rangle + (1-\alpha)\|W\|_1}{2\Phi^*(W)} \tag{1.1}$$

Here Φ^* is the dual norm:

$$\Phi^*(W) = \max_{X} \frac{\langle W, X\rangle}{\Phi(X)}$$

We have progressed to Step 3. We need to find a witness matrix W that makes the bound from Equation 1.1 large. As any matrix W at all gives a lower bound, we can start with an educated guess and modify it according to the difficulties that arise. This is similar to the case discussed earlier of trying to prove that a graph is planar—one can simply start drawing and see how it goes. The first choice of a witness that comes to mind is the target matrix A itself. This gives the lower bound

$$\Phi^{\alpha}(A) \geq \frac{(1+\alpha)\langle A, A\rangle + (1-\alpha)\|A\|_1}{2\Phi^*(A)} = \frac{mn}{\Phi^*(A)}. \tag{1.2}$$

This is actually not such a bad guess; for many interesting norms this lower bound is tight with high probability for a random matrix. But it is not always a good witness, and there can be a very large gap between the two sides of the inequality (1.2). One reason that the matrix A might be a bad witness, for example, is that it contains a large submatrix S for which $\Phi^*(S)$ is relatively large.

A way to fix this deficiency is to take instead of A any matrix $P \circ A$, where P is a real matrix with nonnegative entries that sum up to 1. Here $\circ$ denotes the entry-wise product. This yields a better lower bound

$$\Phi^{\alpha}(A) \geq \max_{\substack{P:P\geq 0 \\ \|P\|_1=1}} \frac{1}{\Phi^*(P \circ A)}. \tag{1.3}$$

Now, by a clever choice of P, we can for example give more weight to a good submatrix of A and less or zero weight to submatrices that attain large values on the dual norm. Although this new lower bound is indeed better, it is still possible to exhibit an exponential gap between the two sides of (1.3). This is nicely explained by the following characterization given in Chapter 7.

Theorem 1.2. For every sign matrix A

$$\Phi^{\infty}(A) = \max_{\substack{P:P\geq 0 \\ \|P\|_1=1}} \frac{1}{\Phi^{*}(P \circ A)}.$$

The best value a witness matrix W which has the same sign as A in each entry can provide, therefore, is equal to $\Phi^{\infty}(A)$. It can be expected that there are matrices A for which $\Phi^{\infty}(A)$ is significantly smaller than $\Phi^{\alpha}(A)$ for say $\alpha = 2$ [1]. This is indeed the case for some interesting communication complexity problems such as the SET INTERSECTION problem where $f(x, y) = \bigvee_i (x_i \wedge y_i)$, which will be a running example throughout the survey.

When $\Phi^{\infty}(A)$ is not a good lower bound on $\Phi^{\alpha}(A)$ for bounded α, there are only a few situations where we know how to choose a good witness. One case is where A is the sign matrix of a so-called *block composed function*, that is, a function of the form $(f \bullet g^n)(x, y) = f(g(x^1, y^1), \ldots, g(x^n, y^n))$ where $x = (x^1, \ldots, x^n)$ and $y = (y^1, \ldots, y^n)$. This case has recently seen exciting progress [She09, She08c, SZ09b]. These works showed a lower bound on the complexity of a block composed function in terms of the approximate degree of f, subject to the inner function g satisfying some technical conditions. The strength of this approach is that the approximate degree of $f : \{0,1\}^n \to \{-1,+1\}$ is often easier to understand than its communication complexity. In particular, in the case where f is symmetric, i.e. only depends on the Hamming weight of the input, the approximate polynomial degree has been completely characterized [Pat92]. These results are described in detail in Chapter 7.2.

Historical context The "three-step approach" to proving communication complexity lower bounds has already been used in the first papers studying communication complexity. In 1983, Yao [Yao83] gave an equivalent "max" formulation of randomized communication complexity using von Neumann's minimax theorem. He showed that the

[1] Notice that $\Phi^{\alpha}(A)$ is a decreasing function of α.

1/3-error randomized communication complexity is equal to the maximum over all probability distributions P, of the minimum cost of a deterministic protocol which errs with probability at most 1/3 with respect to P. Thus one can show lower bounds on randomized communication complexity by exhibiting a probability distribution which is hard for deterministic protocols. This principle is the starting point for many lower bound results on randomized complexity.

A second notable result using the "three-step approach" is a characterization by Karchmer, Kushilevitz, and Nisan [KKN95] of nondeterministic communication complexity. Using results from approximation theory, they show that a certain linear program characterizes nondeterministic communication complexity, up to small factors. By then looking at the dual of this program, they obtain a "max" quantity which can always show near optimal lower bounds on nondeterministic communication complexity.

The study of quantum communication complexity has greatly contributed to our understanding of the role of convexity in communication complexity lower bounds, and these more recent developments occupy a large portion of this survey. The above two examples are remarkable in that they implement the "three-step approach" with a (near) exact representation of the communication model. For quantum communication complexity, however, we do not yet have such a characterization which is convenient for showing lower bounds. The search for good representations to approximate quantum communication complexity led in particular to the development of approximate norms [Kla01, Raz03, LS09c]. Klauck (Lemma 3.1) introduced what we refer to in this survey as the μ^α approximate norm, also known as the *generalized discrepancy method*. While implicit in Klauck and Razborov, the use of Steps 2 and 3 of the three-step approach becomes explicit in later works [LS09c, She08c, SZ09b].

What is not covered In the thirty years since its inception, communication complexity has become a vital area of theoretical computer science, and there are many topics which we will not have the opportunity to address in this survey. We mention some of these here.

Much work has been done on protocols of a restricted form, for example one-way communication complexity where information only flows from Alice to Bob, or simultaneous message passing where Alice and Bob send a message to a referee who then outputs the function value. A nice introduction to some of these results can be found in [KN97]. In this survey we focus only on general protocols.

For the most part, we stick to lower bound methods that fit into the general framework described earlier. As we shall see, these methods do encompass many techniques proposed in the literature, but not all. In particular, a very nice approach which we do not discuss are lower bounds based on information theory. These methods, for example, can give an elegant proof of the optimal $\Omega(n)$ lower bound on the SET INTERSECTION problem. We refer the reader to [BYJKS04] for more details.

We also restrict ourselves to the case where Alice and Bob want to compute a Boolean function. The study of the communication complexity of relations is very interesting and has nice connections to circuit depth and formula size lower bounds. More details on this topic can be found in Kushilevitz and Nisan [KN97].

Finally, there are some models of communication complexity which we do not discuss. Perhaps the most notable of these is the model of unbounded-error communication complexity. This is a randomized model where Alice and Bob only have to succeed on every input with probability strictly greater than $1/2$. We refer the reader to [For02, She08d, RS08] for interesting recent developments on this model.

2

Deterministic communication complexity

In this chapter, we look at the simplest variant of communication complexity, where the two parties act deterministically and are not allowed to err. As we shall see, many of the lower bound techniques we develop for this model can be fairly naturally extended to more powerful models later on.

Say that Alice and Bob wish to arrange a meeting, and want to know if there is a common free slot in their busy schedules. How much might Alice and Bob have to communicate to figure this out? We will shortly see that, in the worst case, Alice may have to send her entire agenda to Bob.

We can describe this scenario as a function $f : \{0,1\}^n \times \{0,1\}^n \to \{T, F\}$ where the ones in Alice and Bob's input represent the free time slots. This function is one of the recurrent examples of our survey, the SET INTERSECTION function. In general, the two binary inputs $x, y \in \{0,1\}^n$ are thought of as characteristic vectors of subsets of $[n] = \{1 \dots n\}$. Alice and Bob wish to decide whether these subsets intersect.

We informally described a deterministic protocol in the introduction; let us now make this formal.

Definition 2.1. A deterministic protocol for a function $f : X \times Y \to \{T, F\}$ is a binary tree $\mathcal{T}$ with internal nodes labeled either by a function $a_v : X \to \{0, 1\}$ or $b_v : Y \to \{0, 1\}$, and leaves labeled by elements from $\{T, F\}$. An input (x, y) defines a path in $\mathcal{T}$ from the root to a leaf as follows: beginning at the root, at an internal node v move to the left child of v if $a_v(x) = 0$ or $b_v(y) = 0$ and otherwise move to the right child of v, until arriving at a leaf. A protocol correctly computes a function f if for every input (x, y) the path defined by (x, y) in $\mathcal{T}$ arrives at a leaf labeled by $f(x, y)$. The cost of a protocol is the number of edges in a longest path from the root to a leaf. The deterministic communication complexity of a function f, denoted $D(f)$ is the minimum cost of a protocol which correctly computes f.

One of the most fundamental concepts in deterministic communication complexity is that of a *combinatorial rectangle*. This is a subset $C \subseteq X \times Y$ which can be written in the form $C = X' \times Y'$ for some $X' \subseteq X$ and $Y' \subseteq Y$. There is a bijection between combinatorial rectangles and Boolean rank-one $|X|$-by-$|Y|$ matrices—namely, we associate to a rectangle C the Boolean matrix R where $R[x, y] = 1$ if $(x, y) \in C$ and $R[x, y] = 0$ otherwise. We identify a combinatorial rectangle with its Boolean matrix representation. We say that a combinatorial rectangle C is *monochromatic with respect to* f if $f(x, y) = f(x', y')$ for all pairs $(x, y), (x', y') \in C$.

A basic and very useful fact is that a correct deterministic communication protocol for a function f partitions the set of inputs $X \times Y$ into combinatorial rectangles which are monochromatic with respect to f.

Definition 2.2 (partition number). Let X, Y be two finite sets and $f : X \times Y \to \{T, F\}$. Define the partition number, $C^D(f)$ as the minimal size of a partition of $X \times Y$ into combinatorial rectangles which are monochromatic with respect to f.

Theorem 2.1 (partition bound). Let $f : X \times Y \to \{T, F\}$. Then

$$D(f) \geq \log C^D(f).$$

Proof. Let $\mathcal{T}$ be a protocol of cost c which correctly computes f. Recall from the definition of a protocol that we describe $\mathcal{T}$ as a binary tree of height c. As the height of this tree is c, it has at most 2^c many leaves ℓ. For each leaf ℓ, define the set $C_\ell = \{(x, y) : (x, y) \in X \times Y, \mathcal{T}(x, y) \to \ell\}$. By the notation $\mathcal{T}(x, y) \to \ell$ we mean that the path defined by (x, y) arrives at leaf ℓ.

We have now defined at most 2^c many sets $\{C_\ell\}$ for each leaf of the protocol. Since the protocol is correct, it is clear that each set C_ℓ is monochromatic with respect to f, and because the functions $a_v(x), b_v(y)$ are deterministic, the sets $\{C_\ell\}$ form a partition of $X \times Y$. It remains to show that each C_ℓ is a combinatorial rectangle.

Suppose that $(x, y'), (x', y) \in C_\ell$. This means that the paths described by $(x, y'), (x', y)$ in $\mathcal{T}$ coincide. We will show by induction that (x, y) follows this same path and so $(x, y) \in C_\ell$ as well. This is clearly true after 0 steps as all paths begin at the root. Suppose that after k steps the path described by $(x, y), (x, y'), (x', y)$ have all arrived at a node v. If this is an Alice node then both (x, y) and (x, y') will move to the child of v indicated by $a_v(x)$; if this is a Bob node, then both (x, y) and (x', y) will move to the child of v indicated by $b_v(y)$. In either case, the paths described by $(x, y), (x, y'), (x', y)$ still agree after $k + 1$ steps, finishing the proof. The reader can verify that a set for which $(x, y'), (x', y) \in C_\ell$ implies $(x, y) \in C_\ell$, is a combinatorial rectangle. □

The partition bound is a relaxation of deterministic communication complexity. A correct deterministic protocol leads to a "tree-like" partition of f into monochromatic combinatorial rectangles, whereas the partition bound allows an arbitrary partition. This relaxation, however, remains relatively tight.

Theorem 2.2 (Aho, Ullman, Yannakakis [AUY83]). Let $f : X \times$

$Y \to \{T, F\}$ be a function, then

$$D(f) \leq (\log(C^D(f)) + 1)^2.$$

We will see a proof of a stronger version of this theorem in Chapter 3. Kushilevitz et al. [KLO96] exhibited a function for which $D(f) \geq 2 \log C^D(f)$, currently the largest such gap known.

The partition bound is a relaxation of communication complexity which is guaranteed to give relatively tight bounds. On the negative side, the partition bound is hard to compute. Counting the number of 1-rectangles in a smallest monochromatic partition is equivalent to the biclique partition problem, which is NP-hard [JR93]. While this says nothing about the ability of humans to compute the partition bound for communication problems of interest, experience demonstrates the partition bound is also difficult to compute in practice.

We now look at further easier-to-compute relaxations of the partition bound coming from linear algebra. Consider the Boolean communication matrix corresponding to $f : X \times Y \to \{T, F\}$, denoted B_f. This is a $|X|$-by-$|Y|$ matrix where $B_f[x, y] = 1$ if $f(x, y) = T$ and $A_f[x, y] = 0$ if $f(x, y) = F$. Denote by $\bar{B}_f$ the communication matrix corresponding to the negation of f.

The partition bound leads us immediately to one of the most important lower bound techniques in deterministic communication complexity, the log rank bound, originally developed by Mehlhorn and Schmidt [MS82].

Theorem 2.3 (log rank bound). Let $f : X \times Y \to \{T, F\}$ be a function and B_f the Boolean communication matrix for f. Then

$$D(f) \geq \log(\text{rank}(B_f) + \text{rank}(\bar{B}_f)).$$

Proof. By Theorem 2.1, if $D(f) = c$, then there exists a partition of $|X| \times |Y|$ into at most 2^c many combinatorial rectangles which are monochromatic with respect to f. Consider such an optimal partition P.

A monochromatic rectangle of f either has all entries equal to zero or all entries equal to one— say that there are Z all zero rectangles and O all-one rectangles in the partition P. We clearly have $O + Z \leq 2^c$.

With each all-one rectangle R_i in P we associate a rank-one Boolean matrix $u_i v_i^t$. Naturally

$$B_f = \sum_{i=1}^{O} u_i v_i^t,$$

where we sum over the all-one rectangles in P. By subadditivity of rank, we find that $\text{rank}(B_f) \leq O$. A similar argument shows that $\text{rank}(\bar{B}_f) \leq Z$, giving the theorem. □

Remark 2.1. We can also consider the sign matrix A_f corresponding to the function f, where $A_f[x, y] = -1$ if $f(x, y) = T$ and $A_f[x, y] = 1$ if $f(x, y) = F$. By a very similar argument one can show that $D(f) \geq \log \text{rank}(A_f)$.

Recall that the trivial protocol for a function $f : \{0,1\}^n \times \{0,1\}^n \to \{T, F\}$ requires $n + 1$ many bits. Whereas the log rank bound with a sign matrix can show bounds of size at most n, the Boolean form of the log rank bound can sometimes give bounds of size $n + 1$, satisfyingly showing that the trivial protocol is absolutely optimal.

We see that the log rank bound relaxes the bound from Theorem 2.1 in two ways. First, rank is the smallest k such that $A = \sum_{i=1}^{k} x_i y_i^t$, where x_i, y_i are allowed to be arbitrary real vectors, not Boolean vectors as is required in the partition bound; second, the rank one matrices $x_i y_i^t, x_j y_j^t$ are allowed to overlap, whereas the partition bound looks at the size of a smallest partition. What we give up in strength of the bound, we gain in ease of application as matrix rank can be computed in time polynomial in the size of the matrix.

Let us see an example of the log rank bound in action. We return to the problem of Alice and Bob arranging a meeting, the SET INTERSECTION problem. It will actually be more convenient to study the complement of this problem, the DISJOINTNESS function. It is clear that in the deterministic case, a problem and its complement have

the same complexity. The Boolean communication matrix for the DISJOINTNESS function on one bit is

$$\mathrm{DISJ}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

The columns and rows of DISJ_1 are labeled by the two possible inputs $0, 1$ in that order. Now consider the communication matrix for the problem on two bits

$$\mathrm{DISJ}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

We see that $\mathrm{DISJ}_2 = \mathrm{DISJ}_1 \otimes \mathrm{DISJ}_1$ is the tensor product of DISJ_1 with itself. This is because taking the tensor product of two Boolean matrices evaluates the AND of their respective inputs. We can easily check that the matrix DISJ_1 has full rank. Thus the rank of the disjointness function on k bits, $\mathrm{DISJ}_k = \mathrm{DISJ}_1^{\otimes k}$ is 2^k as rank is multiplicative under tensor product. Now applying Theorem 2.3, we find that the deterministic communication complexity of the disjointness problem on k inputs is at least $k + 1$. This shows that the trivial protocol is optimal in the case of SET INTERSECTION.

2.1 Log rank conjecture

One of the most notorious open problems in communication complexity is the log rank conjecture. This conjecture asserts that the log rank bound is faithful, i.e. that it is polynomially related to deterministic communication complexity.

Conjecture 2.1 (Lovász and Saks [LS88]). There exists a constant c such that for any function f

$$D(f) \leq (\log \operatorname{rank}(A_f))^c + 2.$$

The log rank conjecture actually has its origins in graph theory. Let $G = (V, E)$ be a graph, the *chromatic number* $\chi(G)$ of G is the size of a

smallest partition of the vertices of G into "color" classes such that there is no edge between vertices in the same class. The *adjacency matrix* A_G of G is a $|V|$-by-$|V|$ Boolean matrix where $A_G[v,w] = 1$ if $(v,w) \in E$ and $A_G[v,w] = 0$ otherwise.

It was conjectured that the rank of the adjacency matrix of a simple graph is an upper bound on its chromatic number, that is $\chi(G) \leq \text{rank}(A_G)$. This conjecture was made independently by van Nuffelen [Nuf76] and by the conjecture generating computer program Graffiti [Faj88]. This conjecture was disproved by Alon and Seymour [AS89], who gave an example of a graph on 64 vertices with chromatic number 32 and rank 29.

Lovász and Saks [LS88] made the log rank conjecture in the above form, and showed that it is equivalent to the statement that $(\log \text{rank}(A_G))^c$ is an upper bound on $\log \chi(G)$, for some constant c.

Several examples that separate communication complexity and log rank have been given. Raz and Spieker [RS95] gave an example with a super-linear separation between communication complexity and log rank, and Nisan and Wigderson [NW95] showed the largest separation currently known: a function f where $D(f) \geq (\log \text{rank}(A_f))^{1.63}$ (this constant was obtained via a slight improvement due to Kushilevitz, see [NW95]).

On the positive side, we know that $D(f) \leq \text{rank}(A_f)$, but this is exponentially far from the goal of the conjecture. In fact, we can also upper bound communication complexity by the rank over $GF(2)$. This is sometimes tight, e.g. for the inner product function.

Theorem 2.4. Let B be a Boolean matrix and $\text{rank}_2(B)$ be the rank of B over $GF(2)$. Then

$$D(f) \leq \text{rank}_2(B_f) + 1$$

Proof. Let $r = \text{rank}_2(B_f)$. We can factor $B_f = X^t Y$ where X, Y are Boolean valued matrices with r many rows. On input i, Alice can simply send X_i, the column of X labeled by i, to Bob with r many bits. Bob can then take the inner product of X_i with Y_j—the column labeled by

his input j—to determine $B_f[i,j]$. With one more bit Bob sends the value of $B_f[i,j]$ to Alice. □

2.2 Nonnegative rank

Yannakakis [Yan91] introduced to communication complexity the notion of nonnegative rank.

Definition 2.3. Let M be a nonnegative matrix. The *nonnegative rank* of M, denoted $\text{rank}^+(A)$ is the least r such that

$$M = \sum_{i=1}^{r} x_i y_i^t$$

for *nonnegative* vectors x_i, y_i.

We clearly have $\text{rank}(M) \leq \text{rank}^+(M)$ for a nonnegative matrix M. For a Boolean matrix B, notice that the rank-one decomposition of B induced by a successful protocol as in Theorem 2.3 only uses nonnegative (in fact Boolean) matrices, and so $\log \text{rank}^+(B_f) \leq D(f)$. While nonnegative rank gives stronger lower bounds, it is also NP-hard to compute [Vav07] and we are not aware of any lower bounds which actually use nonnegative rank in practice.

Lovász shows that $\max\{\log \text{rank}^+(B_f), \log \text{rank}^+(J - B_f)\}$, where J is the all-ones matrix, faithfully represents deterministic communication complexity. In fact, he gives the following stronger bound.

Theorem 2.5 (Lovász [Lov90], Corollary 3.7).

$$D(f) \leq (\log(\text{rank}^+(B_f)) + 1)(\log(\text{rank}(J - B_f)) + 1).$$

We will see the proof of this theorem in Section 3.2 in the chapter on nondeterministic communication complexity.

Theorem 2.5 suggests the following equivalent formulation of the log rank conjecture as a purely mathematical question:

Theorem 2.6. The log rank conjecture holds if and only if there is a constant c such that for all Boolean matrices B

$$\log \operatorname{rank}^+(B) \leq (\log \operatorname{rank}(B))^c.$$

2.3 Norm based methods

We have seen the partition bound which is a lower bound on deterministic communication complexity, and the rank lower bound which in turn relaxes the partition bound. We now survey another family of lower bound methods based on (matrix and vector) norms.

Although sometimes requiring nontrivial arguments, it turns out that all the norm based methods discussed in this section in fact lower bound matrix rank. This prompts the question: why study these norm based techniques if the rank method gives a better lower bound and is, at least from a theoretical perspective, easy to compute? The real advantage of these norm based methods will only be seen later on in the study of randomized and multiparty models. While the rank method can also be appropriately extended to these models, it becomes much more difficult to compute; the convexity properties of norm based methods make their extensions to these models more tractable. We go ahead and introduce the norm based methods in the context of deterministic communication complexity where the situation is simpler, and later discuss how they can be adapted to more powerful models.

We repeatedly use the ℓ_1, ℓ_2 and ℓ_∞ norms, hence we recall their definition here: for a vector $v \in \mathbb{R}^n$ and a real number $p \geq 1$ the ℓ_p norm of v, denoted $\|v\|_p$, is defined as $\|v\|_p = (\sum_{i=1}^n |v[i]|^p)^{1/p}$. We see then that $\|v\|_1 = \sum_i |v[i]|$, and $\|v\|_2 = (\sum_i |v[i]|^2)^{1/2}$. The limiting case is $\|v\|_\infty = \max_i |v[i]|$. The reader should keep in mind that matrices and functions to $\mathbb{R}$ can be naturally viewed as vectors, whence taking their ℓ_p norms is likewise natural.

2.3.1 Trace norm

We begin with a simple lower bound on the rank. Let A be a m-by-n real matrix. The matrix AA^t is positive semidefinite, hence it has

nonnegative real eigenvalues. We denote these eigenvalues by $\lambda_1(AA^t) \geq \cdots \geq \lambda_m(AA^t)$. The i^{th} *singular value* of A, denoted $\sigma_i(A)$, is defined by $\sigma_i(A) = \sqrt{\lambda_i(AA^t)}$. In many ways, singular values can be seen as a generalization of eigenvalues to non-square matrices. If A is symmetric, then its singular values are just the absolute values of its eigenvalues. While not every matrix can be diagonalized, every matrix A can be factored as $A = U\Sigma V$ where U is a m-by-m unitary matrix, V is a n-by-n matrix and Σ is a diagonal matrix with the singular values of A on its diagonal. This shows in particular that the rank of A is equal to the number of nonzero singular values.

Definition 2.4 (trace norm). Let A be a m-by-n matrix, and let $\sigma = (\sigma_1, \ldots, \sigma_{\text{rank}(A)})$ be the vector of nonzero singular values of A. The *trace norm* of A, denoted $\|A\|_{tr}$ is

$$\|A\|_{tr} = \|\sigma\|_1$$

We also make use of the Frobenius norm.

Definition 2.5 (Frobenius norm). Let A be a m-by-n matrix, and let $\sigma = (\sigma_1, \ldots, \sigma_{\text{rank}(A)})$ be the vector of nonzero singular values of A. The Frobenius norm of A, denoted $\|A\|_F$ is

$$\|A\|_F = \|\sigma\|_2.$$

As the number of nonzero singular values is equal to the rank of A, the Cauchy-Schwarz inequality gives

$$\|A\|_{tr} = \sum_{i=1}^{\text{rank}(A)} \sigma_i(A) \leq \sqrt{\text{rank}(A)} \sqrt{\sum_{i=1} \sigma_i^2(A)}.$$

Rearranging, this gives us the following lower bound on the rank.

$$\text{rank}(A) \geq \frac{\|A\|_{tr}^2}{\|A\|_F^2}.$$

In this section, it is convenient to consider the sign version of the communication matrix. The reason is that, for a sign matrix A, the

Frobenius norm simplifies very nicely. Notice that as the trace of a symmetric matrix is the sum of its eigenvalues, we have $\|A\|_F^2 = \mathrm{Tr}(AA^t)$. By explicitly writing the diagonal elements of AA^t, we also see that $\mathrm{Tr}(AA^t) = \sum_{i,j} |A[i,j]|^2 = \|A\|_2^2$. Hence for a m-by-n sign matrix A, we have $\|A\|_F^2 = mn$. This gives the following lower bound, which we call the *trace norm method*.

Theorem 2.7 (trace norm method). Let A be a m-by-n sign matrix. Then

$$D(A) \geq \log \mathrm{rank}(A) \geq \log \frac{\|A\|_{tr}^2}{mn}$$

As an example, let us compute the trace norm bound for the INNER PRODUCT function. Recall that in this case Alice and Bob wish to evaluate the parity of the number of positions for which $x_i = y_i = 1$. The sign matrix of this function turns out to be the familiar Sylvester construction of Hadamard matrices [1]. If we look at the INNER PRODUCT function on just one bit we have the matrix

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Since taking the tensor product of sign matrices corresponds to taking the parity of their inputs, the communication matrix of the INNER PRODUCT function on k bits is $H_k = H_1^{\otimes k}$. It is not hard to prove that the matrix H_k is *orthogonal*, i.e. satisfies $H_k H_k^t = 2^k I_k$, where I_k is the 2^k-by-2^k identity matrix. One can simply verify that H_1 is orthogonal, and that the tensor product of two orthogonal matrices is also orthogonal. It follows then that H_k has 2^k singular values, all equal to $2^{k/2}$, and so its trace norm is $2^{3k/2}$. Thus, applying the trace norm method, we obtain a bound of k, implying that the trivial protocol is essentially optimal for the INNER PRODUCT function.[2]

[1] A *Hadamard matrix* is an orthogonal sign matrix. That is, a sign matrix whose rows are pairwise orthogonal.

[2] For this example it is even easier to argue about the rank as $H_k H_k^t = 2^k I_k$ means that H_k is full rank. The advantage of the trace norm argument is that it can be easily extended to the randomized case, see Example 7.1.

2.3.2 γ_2 norm

As a complexity measure, the trace norm method suffers from one drawback: it is not monotone with respect to function restriction. In other words, it sometimes gives a worse bound on a restriction of a function than on the function itself. An example of this is the matrix

$$\begin{pmatrix} H_k & J_k \\ J_k & J_k \end{pmatrix},$$

where J_k is the 2^k-by-2^k matrix whose entries are all equal to one. The trace norm of this matrix is at most $2^{3k/2}+3\cdot 2^k$. Since in the trace norm method we normalize by the matrix size, in this case 2^{2k+2}, this method gives a smaller bound on the above matrix than on the submatrix H_k.

To remedy this, we seek a way to focus on the "difficult" part of the matrix. We do this by putting weights on the entries of the matrix in the form of a rank-one matrix uv^t. The weighted matrix is then the entrywise product of A with uv^t, denoted $A \circ uv^t$. It is easy to check that $\mathrm{rank}(A \circ uv^t) \leq \mathrm{rank}(A)$, and so

$$\mathrm{rank}(A) \geq \max_{u,v} \frac{\|A \circ uv^t\|_{tr}^2}{\|A \circ uv^t\|_F^2} \tag{2.1}$$

for any u, v. This new bound is monotone with respect to function restriction. If A is a sign matrix, then it is particularly nice to choose u, v to be unit vectors, for then $\|A \circ uv^t\|_F = \|u\|_2\|v\|_2 = 1$. This motivates the following definition

---

Definition 2.6 (γ_2 norm). For a matrix A we define

$$\gamma_2(A) = \max_{u,v:\|u\|_2=\|v\|_2=1} \|A \circ uv^t\|_{tr}$$

The connection of γ_2 to communication complexity is given by the following theorem which follows from Equation 2.1.

Theorem 2.8. Let A be a sign matrix. Then

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

While the γ_2 norm has been introduced relatively recently to complexity theory [LMSS07, LS09c], it has been around in matrix analysis for a while. Tracing its heritage is somewhat difficult because of its many different names: in the matrix analysis community it has been called by various combinations of "Hadamard/Schur operator/trace norm." Schur in 1911 [Sch11] showed that if a matrix A is positive semidefinite, then $\gamma_2(A) = \max_i A[i,i]$. It should be noted that unlike the trace norm, γ_2 is not a matrix norm as it is not true in general that $\gamma_2(AB) \leq \gamma_2(A)\gamma_2(B)$.

That γ_2 is a norm, namely that it satisfies $\gamma_2(A+B) \leq \gamma_2(A) + \gamma_2(B)$, can be seen most easily by the following equivalent formulation as a minimization problem. A proof of the equivalence of this definition with the one given above can be found in [LSŠ08].

Theorem 2.9. For any real m-by-n matrix M

$$\gamma_2(M) = \min_{X,Y:XY^t=M} r(X)r(Y) \tag{2.2}$$

where $r(X)$ is the largest ℓ_2 norm of a row of X.

Let us see that the optimization problem of Theorem 2.9 can be written as a semidefinite program as follows.

$$\begin{aligned} \gamma_2(M) = \min\, & c \\ & cI \succeq Z \circ I \\ & Z \circ \begin{pmatrix} 0 & J_{m,n} \\ J_{n,m} & 0 \end{pmatrix} = \begin{pmatrix} 0 & M \\ M^t & 0 \end{pmatrix} \\ & Z \succeq 0. \end{aligned}$$

Here $J_{m,n}$ indicates the m-by-n all-ones matrix, and I the $(m+n)$-by-$(m+n)$ identity matrix.

Let X, Y be an optimal solution to Equation 2.2. Notice that by multiplying X by a suitable constant and dividing Y by the same constant, we may assume that $r(X) = r(Y)$. Now let

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}^t.$$

By construction this matrix is positive semidefinite, and is equal to M on the off diagonal blocks as $M = XY^t$. Furthermore, the diagonal entries of Z are at most $\max\{r(X)^2, r(Y)^2\} \leq \gamma_2(M)$.

For the other direction, given an optimal solution Z to the above semidefinite program, we can factor Z as

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}^t,$$

for some m-by-k matrix X and n-by-k matrix Y. Then the constraints of the semidefinite program give that $XY^t = M$ and that $\max\{r(X)^2, r(Y)^2\} \leq c$, the value of the program.

Finally, from this semidefinite programming formulation of γ_2, it is easy to verify that $\gamma_2(M_1 + M_2) \leq \gamma_2(M_1) + \gamma_2(M_2)$. Let X_1 and X_2 be semidefinite matrices attaining the optimal value in this program for the matrices M_1 and M_2 respectively. Recall that the sum of two semidefinite matrices is also a semidefinite matrix. It is easy to see that $X_1 + X_2$ is a feasible instance of the above program for $M_1 + M_2$, achieving the value $\gamma_2(M_1) + \gamma_2(M_2)$. The semidefinite programming formulation also shows that $\gamma_2(M)$ can be computed up to error ϵ in time polynomial in the size of the matrix and $\log(1/\epsilon)$.

2.3.3 μ norm

In this section we introduce another norm useful for communication complexity, which we denote by μ. While the motivation for this norm is somewhat different from that of γ_2, it surprisingly turns out that μ and γ_2 are equal up to a small multiplicative constant.

Recall the "partition bound" (Theorem 2.3) and the "log rank lower bound" (Theorem 2.1). A basic property we used is that a sign matrix A can be written as

$$A = \sum_{i=1}^{2^{D(A)}} \alpha_i R_i$$

where each $\alpha_i \in \{-1, 1\}$ and each R_i is a rank-one Boolean matrix, which we also call a combinatorial rectangle. As each $\alpha_i \in \{-1, 1\}$ we of course have $\sum_i |\alpha_i| = 2^{D(A)}$.

The log rank lower bound is a relaxation of the partition bound where we express A as the sum of arbitrary real rank-one matrices instead of just Boolean matrices. The following definition considers a different relaxation of the partition bound, where we still consider a decomposition in terms of Boolean matrices, but count their "weight" instead of their number.

Definition 2.7 (μ norm). Let M be a real matrix.

$$\mu(M) = \min_{\alpha_i \in \mathbb{R}} \{ \sum_i |\alpha_i| : M = \sum_i \alpha_i R_i \},$$

where each R_i is a combinatorial rectangle.

It is not hard to check that μ is a norm. Notice that $\mu(M) \geq \gamma_2(M)$ as any combinatorial rectangle R_i satisfies $\gamma_2(R_i) \leq 1$.

Applying similar reasoning as for the log rank bound, we get

Theorem 2.10. Let A be a sign matrix, then

$$D(A) \geq \log \mu(A).$$

2.3.4 The nuclear norm

It is sometimes useful to consider a slight variant of the norm μ where instead of rank-one Boolean matrices we consider rank-one sign matrices.

Definition 2.8 (ν norm). Let M be a real matrix,

$$\nu(M) = \min_{\alpha_i \in \mathbb{R}} \{ \sum_i |\alpha_i| : M = \sum_i \alpha_i x_i y_i^t, \text{ for } x_i, y_i \text{ sign vectors} \}$$

For readers with some background on norms, the norm ν is a *nuclear norm* [Jam87]. Nuclear norms are dual to operator norms which we also encounter later on. The norms μ and ν are closely related.

Theorem 2.11. For every real matrix M,

$$\nu(M) \leq \mu(M) \leq 4\nu(M).$$

Proof. Since both μ and ν are norms, it is enough to show that

(1) $\nu(sr^t) \leq 1$ for every pair of Boolean vectors s and r.
(2) $\mu(xy^t) \leq 4$ for every pair of sign vectors x and y.

For the first inequality we consider the following correspondence between sign vectors and Boolean vectors. Given a Boolean vector s we denote $\bar{s} = 2s - \mathbf{1}$ (Here $\mathbf{1}$ is a vector of ones). Note that $\bar{s}$ is a sign vector.

Now, $s = \frac{1}{2}(\bar{s} + \mathbf{1})$, and therefore

$$\nu(sr^t) = \frac{1}{4}\nu((\bar{s} + \mathbf{1})(\bar{r} + \mathbf{1})^t) \leq 1.$$

To prove the second inequality simply split the sign vector x to two Boolean vectors depending on whether x_i is equal to 1 or -1, and similarly for y. The rank-one sign matrix xy^t can be written this way as the linear combination of 4 combinatorial rectangles with coefficients 1 and -1. □

2.3.5 Dual norms

We have now introduced three norms γ_2, μ, ν and seen that they give lower bounds on deterministic communication complexity. For actually proving lower bounds via these methods, a key role is played by their *dual* norms. We go ahead and define these dual norms here to collect all the definitions in one place and as this is also the easiest way to see that γ_2 and ν are related by a constant factor. As we have already seen that μ and ν are equivalent up to a factor of 4, this means that all three norms are related by a constant factor.

For any arbitrary norm Φ, the dual norm, denoted Φ^*, is defined as

$$\Phi^*(M) = \max_{Z:\Phi(Z)\leq 1} \langle M, Z \rangle.$$

Let us first consider ν^*, the dual norm of ν. If a matrix Z satisfies $\nu(Z) \leq 1$, then it can be written as $Z = \alpha_1 Z_1 + \ldots + \alpha_p Z_p$ where each Z_i is a rank-one sign matrix and $\sum |\alpha_i| \leq 1$. Thus

$$\nu^*(M) = \max_{\substack{\{\alpha_i\} \\ \sum_i |\alpha_i| \leq 1}} \sum_i \alpha_i \langle M, Z_i \rangle.$$

The maximum will be achieved by placing weight 1 on the rank-one sign matrix Z_i which maximizes $\langle M, Z_i \rangle$. Thus we have

$$\nu^*(M) = \max_{\substack{x \in \{-1,+1\}^m \\ y \in \{-1,+1\}^n}} \sum_{i,j} M[i,j]\, x[i] \cdot y[j]. \tag{2.3}$$

The dual norm $\nu^*(M)$ is also known as the *infinity-to-one* norm $\|M\|_{\infty \to 1}$. This is because one can argue that

$$\nu^*(M) = \max_{\substack{x : \|x\|_\infty \leq 1 \\ y : \|y\|_\infty \leq 1}} \sum_{i,j} M[i,j]\, x[i] \cdot y[j] = \max_{y : \|y\|_\infty \leq 1} \|My\|_1.$$

In words, the first equality says that the optimal x, y will have each entry in $\{-1, +1\}$.

A similar argument can be used to see that

$$\mu^*(M) = \max_{\substack{x \in \{0,1\}^m \\ y \in \{0,1\}^n}} \sum_{i,j} M[i,j]\, x[i] \cdot y[j]. \tag{2.4}$$

This norm is also known as the *cut norm* [FK99, AN06], as $x^t M y$ represents the weight of edges between the sets with characteristic vectors given by x and y.

Both the ν^* and μ^* norms are NP-hard to compute [AN06]. The dual norm γ_2^* of γ_2 can be viewed as a natural semidefinite relaxation of these norms. In fact, γ_2^* is exactly the quantity studied by Alon and Naor [AN06] to give an efficient approximation algorithm to the cut norm, and is also closely related to the semidefinite relaxation of MAX CUT considered by Goemans and Williamson [GW95].

We can derive a convenient expression for γ_2^* as we did with ν^*. If a matrix Z satisfies $\gamma_2(Z) = 1$, then we can write $Z_i[k, \ell] = \langle x_k, y_\ell \rangle$ for a collection of unit vectors $\{x_k\}, \{y_\ell\}$. Thus

$$\gamma_2^*(M) = \max_{\substack{\{x_i\}, \|x_i\|_2 \leq 1 \\ \{y_i\}, \|y_i\|_2 \leq 1}} \sum_{i,j} M[i,j] \langle x_i, y_j \rangle. \tag{2.5}$$

It is clear that $\gamma_2^*(M) \geq \nu^*(M)$ as the maximization is taken over a larger set. Grothendieck's famous inequality says that γ_2^* cannot be too much larger.

Theorem 2.12 (Grothendieck's Inequality). There is a constant K_G such that for any matrix M

$$\gamma_2^*(M) \leq K_G \, \nu^*(M).$$

The current best bounds on K_G show that $1.67\ldots \leq K_G \leq 1.78\ldots$ [Ree91, Kri79].

It is not hard to check that two norms are equivalent up to a constant factor if and only if the corresponding dual norms are. Thus summarizing, we have the following relationship between γ_2, ν, μ:

Corollary 2.1 ([LS09b]). For any matrix M

$$\gamma_2(M) \leq \nu(M) \leq \mu(M) \leq 4K_G\gamma_2(M).$$

One consequence of this relationship is that $\mathrm{rank}(A) = \Omega(\mu(A)^2)$ for a sign matrix A, a fact which is not obvious from the definition of μ.

We will discuss all these norms in more detail in Chapter 4 on randomized communication complexity.

2.4 Summary

Complexity measures Let A be a sign matrix, and denote by M a real matrix.

- $\mathrm{rank}(M)$ - the rank of a the matrix M over the real field.
- $\mathrm{rank}_2(A)$ - the rank of the matrix A over $GF(2)$.
- $\mathrm{rank}^+(M)$ - the nonnegative rank of M.
- $D(A)$ - the deterministic communication complexity of A (equivalently, the communication complexity of the corresponding function).
- $\|M\|_p = (\sum_{i,j} |M[i,j]|^p)^{1/p}$ - the ℓ_p norm.

- $\|M\|_{tr}$ - the trace norm of M.
- $\|M\|_F$ - the Frobenius norm of M.
- $\gamma_2(M) = \max_{u,v:\|u\|_2=\|v\|_2=1} \|A \circ uv^t\|_{tr}$ - the γ_2 norm.
- $\mu(M)$ - the dual norm to the cut norm.
- $\nu(M)$ - the nuclear norm of M.

Relations For every $m \times n$ sign matrix A and a real matrix M

- $\text{rank}_2(A) \leq \text{rank}(A) \leq \text{rank}^+(A)$.
- $\log \text{rank}^+(A) \leq D(A) \leq \text{rank}_2(A)$.
- $D(A) \leq \log \text{rank}(A) \log \text{rank}^+(A)$.
- $\text{rank}(A) \geq \gamma_2(M)^2 \geq \frac{\|A\|_{tr}^2}{mn}$.
- $\|M\|_F = \|M\|_2$.
- $\gamma_2(M) \leq \nu(M) \leq \mu(M) \leq 4K_G\gamma_2(M)$.

3

Nondeterministic communication complexity

As with standard complexity classes, we can also consider a nondeterministic version of communication complexity. It turns out that nondeterministic communication complexity has a very nice combinatorial characterization and in some ways is better understood than deterministic communication complexity. In particular, nondeterministic communication complexity provides a prime example of the implementation of the "three-step approach" outlined in the introduction.

We now formally define nondeterministic communication complexity. Let $f : X \times Y \to \{T, F\}$ be a function, and let $L = \{(x, y) : f(x, y) = T\}$. A successful nondeterministic protocol for f consists of functions $A : X \times \{0,1\}^k \to \{0,1\}$ and $B : Y \times \{0,1\}^k \to \{0,1\}$ such that

(1) For every $(x, y) \in L$ there is $z \in \{0,1\}^k$ such that $A(x, z) \wedge B(y, z) = 1$.
(2) For every $(x, y) \notin L$, for all $z \in \{0,1\}^k$, it holds $A(x, z) \wedge B(y, z) = 0$.

The cost of such a protocol is k, the length of the message z. We define $N^1(f)$ to be the minimal cost of a successful nondeterministic protocol

for f. We let $N^0(f) = N^1(\neg f)$ and $N(f) = \max\{N^0(f), N^1(f)\}$.

One can equivalently think of a nondeterministic protocol as consisting of two stages—in the first stage both players receive a message z, and in the second stage they carry out a deterministic protocol—the cost now being the length of z plus the communication in the deterministic protocol. The definition we have given is equivalent to this one as the transcript of the deterministic protocol can always be appended to the witness z with each party accepting if the transcript agrees with what they would have said in the protocol, given what the other party said.

Let us revisit the SET INTERSECTION problem. If someone with knowledge of both Alice's and Bob's schedule tells them they can meet Thursday at noon, they can easily verify this is true. On the other hand, if they have no time in common to meet, every suggestion of the prover will lead to a conflict. Thus for the SET INTERSECTION problem f we have $N^1(f) \leq \lceil \log n \rceil$.

Besides the analogy with nondeterministic complexity classes, another motivation for studying nondeterministic complexity is that it has a very natural combinatorial characterization.

Definition 3.1. Let $f : X \times Y \to \{0,1\}$ be a function, and let $b \in \{0,1\}$. We denote by $C^b(f)$ the smallest cardinality of a covering of $f^{-1}(b)$ by combinatorial rectangles. To illustrate, a *covering* of $f^{-1}(1)$ is a set of rank-one Boolean matrices $\{R_i\}_{i \in I}$ such that: if $f(x,y) = 1$ then $R_i[x,y] = 1$ for some $i \in I$, and if $f(x,y) = 0$ then $R_i[x,y] = 0$ for all $i \in I$.

We can also view this definition in graph theoretic terms. An arbitrary Boolean matrix B can be thought as the "reduced" adjacency matrix of a bipartite graph, where rows are labeled by vertices of one color class and columns by vertices of the other color class. Then $C^1(B)$ is the size of a smallest covering of the edges of B by bipartite cliques. It is known that this is NP-hard to compute in the size of the graph [Orl77], and even NP-hard to approximate within a factor of n^δ for some $\delta > 0$ for a n-vertex graph [LY94].

Theorem 3.1. Let $f : X \times Y \to \{0,1\}$ be a function,

$$N^1(f) = \lceil \log C^1(f) \rceil$$

Proof. First we show that $N^1(f) \leq \lceil \log C^1(f) \rceil$. Let $\{R_i\}$ be a covering of $f^{-1}(1)$ of minimal cardinality. If $f(x,y) = 1$, the players receive the name i of a rectangle $R_i = X' \times Y'$ such that $R_i[x,y] = 1$. Alice then checks that $x \in X'$ and Bob checks that $y \in Y'$. If this is indeed the case then they accept. If $f(x,y) = 0$ then by definition of a covering no such index exists.

Now consider the opposite direction. Let $k = N^1(f)$ and let $f_A : X \times \{0,1\}^k \to \{0,1\}$, $f_B : Y \times \{0,1\}^k \to \{0,1\}$ be functions realizing that the nondeterministic complexity of f is k. Let $R_z = \{(x,y) : f_A(x,z) \wedge f_B(y,z) = 1\}$. This is a rectangle, and by definition of success of a protocol if $f(x,y) = 1$ then $(x,y) \in R_z$ for some z and if $f(x,y) = 0$ then $(x,y) \notin R_z$ for all z. This gives a covering of size 2^k. □

3.1 Relation between deterministic and nondeterministic complexity

Many open questions in communication complexity revolve around showing that the combinatorial and linear-algebraic lower bound techniques we have developed are faithful. The next theorem, due to Aho, Ullman, and Yannakakis, is one of the deepest results we know in this regard. It shows how to turn a monochromatic rectangle covering of a matrix into a deterministic algorithm. Alternatively, it says that if a function has both an efficient nondeterministic protocol and an efficient co-nondeterministic protocol, then it has an efficient deterministic protocol. This is quite different from what we expect in terms of traditional complexity classes.

Theorem 3.2 (Aho, Ullman, Yannakakis [AUY83]). Let $f : X \times Y \to \{0,1\}$ be a function, then

$$D(f) \leq (N^0(f) + 1)(N^1(f) + 1).$$

We will present a tighter version of this theorem due to Lovász [Lov90], which follows the same basic outline but replaces $N^1(f)$ with a smaller quantity. For a Boolean matrix B, define $\rho_1(B)$ such that $\rho_1(B) - 1$ is equal to the largest (after possibly permuting rows and columns) submatrix of B with ones on the diagonal and zeros below the diagonal. Notice that $\rho_1(B)$ is at most $N^1(B) + 1$, and also is at most $\text{rank}(B) + 1$.

Theorem 3.3 (Lovász). Let B be a Boolean matrix. Then

$$D(B) \leq (\log(\rho_1(B)) + 1)(N^0(B) + 1).$$

Proof. The proof is by induction on $\rho_1(B)$. Suppose that $\rho_1(B) = 1$. In this case, B does not contain any ones at all and so is the all zeros matrix and has $D(B) = 1$. This finishes the base case of the induction.

Let $R_1, \ldots, R_M$ be a covering of the zeros of B which realizes $N^0(B) = \lceil \log M \rceil$. Let S_i denote the submatrix of rows of B which are incident with R_i, and similarly T_i the submatrix of columns of B incident with R_i.

The key observation to make is that $\rho_1(S_i) + \rho_1(T_i) \leq \rho_1(B)$ for each i. Thus we may assume without loss of generality that $\rho_i(S_i) \leq \rho(B)/2$ for $i = 1, \ldots, N$ with $N \geq M/2$ and $\rho_i(T_i) \leq \rho(B)/2$ for $i = N + 1, \ldots, M$.

The protocol goes as follows. On input x, Alice looks for a rectangle R_i for $i = 1, \ldots, N$ which intersects with the x^{th} row of B. If such a rectangle exists she sends to Bob the bit '0' followed by the index of the rectangle and the round ends. If no such rectangle exists, she sends '1.'

If Bob receives the message '1' from Alice, he looks for a rectangle R_i for $i = N + 1, \ldots, M$ which intersects the column labeled by his input y. If yes, he sends a '0' followed by the name of such a rectangle, otherwise he gives the answer '1.'

We analyze the cost of this protocol based on the outcome of three cases. In each case, the protocol will either output the answer, or reduce the search to a submatrix where the induction hypothesis may be applied.

Case 1: Both Alice and Bob output '1.' In this case, (x, y) does not lie in any zero rectangle of the covering, thus the answer must actually be '1.'

Case 2: Alice outputs '0' and the name of a rectangle R_i. In this case, the protocol continues in the submatrix A_i. As $\rho(A_i) \leq \rho(B_i)/2$ we can apply the induction hypothesis to see that the communication complexity of A_i is at most

$$(\log(\rho(A_i)) + 1)(N^0(A_i) + 1) \leq \log(\rho(B))(N^0(B) + 1).$$

Adding in the length of Alice's initial message of $1 + N^0(B)$ we get that the total communication complexity of B is at most $(\log(\rho(B)) + 1)(N^0(B) + 1)$ as desired.

Case 3: Alice outputs '1' and Bob outputs '0' and the name of a rectangle. This case is works in the same way as case 2. Notice that the communication for the initial round is again $N^0(B) + 1$. We have one bit from Alice's message and $1 + (N^0(B) - 1)$ for Bob's message, as there are less than $M/2$ possible rectangle labels he might output. □

This theorem can be tight. Jayram, Kumar, and Sivakumar [JKS03] give an example of a function f where $N^0(f) = N^1(f) = O(\sqrt{n})$ and $D(f) = \Omega(n)$.

3.2 Relation with nonnegative rank

Recall the notion of nonnegative rank introduced earlier. Yannakakis [Yan91] showed that nondeterministic communication complexity is a lower bound on the logarithm of nonnegative rank.

Theorem 3.4 (Yannakakis [Yan91]). Let $f : X \times Y \to \{0, 1\}$ be a function, and let B be a Boolean matrix where $B[x, y] = f(x, y)$. Then

$$N^1(f) \leq \log \operatorname{rank}^+(B)$$

Proof. Suppose that $\text{rank}^+(B) = r$. Then we have a decomposition of B as

$$B = \sum_{i=1}^{r} x_i y_i^t$$

where x_i, y_i are nonnegative vectors. Define new vectors x_i', y_i' by

$$x_i'[j] = \begin{cases} 1 & \text{if } x_i[j] > 0 \\ 0 & \text{otherwise} \end{cases}$$

and similarly for y_i'. Then $\sum_i x_i'(y_i')^t$ will be zero whenever B is, and will be at least 1 whenever B is equal to 1. □

This result, together with Theorem 3.3, implies that $D(f) \leq (\log(\text{rank}(B_f)) + 1))(\log(\text{rank}^+(J - B_f)) + 1)$, where J is the all-ones matrix.

3.3 Fractional cover

As in the case of the rectangle bound, we can similarly write the cover number as an integer program. Let $f : X \times Y \to \{0, 1\}$ be a function. Then we see that

$$\begin{aligned} C^1(f) = \min_{\alpha_i} & \sum \alpha_i \\ & \sum_i \alpha_i R_i[x, y] \geq 1 \text{ for all } (x, y) \in f^{-1}(1) \\ & \sum_i \alpha_i R_i[x, y] = 0 \text{ for all } (x, y) \in f^{-1}(0) \\ & \alpha_i \in \{0, 1\}, \end{aligned}$$

where each R_i is a combinatorial rectangle. Karchmer, Kushilevitz, and Nisan [KKN95] considered showing lower bounds on nondeterministic communication complexity by the linear programming relaxation of the cover number.

Definition 3.2 (Fractional cover). Let $f : X \times Y \to \{0, 1\}$ be a

function, the *fractional cover* of $f^{-1}(1)$ is

$$\begin{aligned}\bar{C}^1(f) = \min_{\alpha_i} & \sum \alpha_i \\ & \sum_i \alpha_i R_i[x,y] \geq 1 \text{ for all } (x,y) \in f^{-1}(1) \\ & \sum_i \alpha_i R_i[x,y] = 0 \text{ for all } (x,y) \in f^{-1}(0) \\ & \alpha_i \geq 0,\end{aligned}$$

where each R_i is a combinatorial rectangle.

Notice that the size of this program can be exponential in $|X| + |Y|$ since there is a variable α_i for every combinatorial rectangle.

The fractional cover implements the first step of the "three-step approach." It is a convex real-valued function and clearly $\bar{C}^1(f) \leq C^1(f)$. Moreover, $\bar{C}^1(f)$ turns out to give a tight bound on $C^1(f)$. Lovász [Lov75] has shown quite generally that the linear program relaxation of set cover gives a good approximation to the integral problem. Karchmer, Kushilevitz, and Nisan [KKN95] observed that this has the following application for nondeterministic communication complexity.

Theorem 3.5 (Karchmer, Kushilevitz, and Nisan [KKN95]). Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. Then

$$\log \bar{C}^1(f) \leq N^1(f) \leq \log \bar{C}^1(f) + \log n + O(1).$$

The fact that nondeterministic communication complexity is tightly characterized by a linear program implies other nice properties as well. Lovász [Lov75] also showed that the fractional cover program obeys a product property: $\bar{C}^1(A \otimes B) = \bar{C}^1(A)\bar{C}^1(B)$ for every pair of Boolean matrices A and B. Karchmer, Kushilevitz, and Nisan used this result to obtain the following consequence for nondeterministic communication complexity.

Theorem 3.6 (Karchmer, Kushilevitz, and Nisan [KKN95]). Let A, B be any two Boolean matrices of size 2^n-by-2^n. Then

$$N^1(A \otimes B) \geq N^1(A) + N^1(B) - 2\log n - O(1).$$

Karchmer, Kushilevitz, and Nisan also carried out the second step of the three-step formulation to give an equivalent "max" formulation of $\bar{C}^1(f)$ convenient for showing lower bounds. To see this, let us first write the program for $\bar{C}^1(f)$ more compactly in matrix notation. Let $\mathcal{A}$ be a Boolean matrix with rows labeled by elements of $f^{-1}(1)$ and columns labeled by 1-monochromatic rectangles, and where $\mathcal{A}[(x,y),R_i] = 1$ if and only if $(x,y) \in R_i$. Then the program from Definition 3.2 can be written

$$\begin{aligned} \bar{C}^1(f) = \min_{\alpha} \ & \mathbf{1}^t\alpha \\ & \mathcal{A}\alpha \geq \mathbf{1} \\ & \alpha \geq \mathbf{0}. \end{aligned}$$

Here $\mathbf{1}, \mathbf{0}$ stand for the all-one vector and all zero vector respectively, where the dimension can be inferred from the context.

Now consider the following maximization problem over a vector μ of dimension $|f^{-1}(1)|$.

$$\begin{aligned} \underline{C}^1(f) = \max_{\mu} \ & \mathbf{1}^t\mu \\ & \mathcal{A}^t\mu \leq \mathbf{1} \\ & \mu \geq \mathbf{0}. \end{aligned}$$

In this program one assigns nonnegative weight $\mu_{x,y}$ to every input (x,y) with $f(x,y) = 1$, such that the total weight given to any 1 monochromatic rectangle is at most 1.

It is easy to see that $\underline{C}^1(f) \leq \bar{C}^1(f)$. Indeed, let μ satisfy the constraints of the former and α satisfy the constraints of the latter. Then

$$\mathbf{1}^t\mu \leq (\alpha^t\mathcal{A}^t)\mu = \alpha^t(\mathcal{A}^t\mu) \leq \mathbf{1}^t\alpha.$$

As we will discuss in Chapter 6, using linear programming duality one can show that in fact $\underline{C}^1(f) = \bar{C}^1(f)$.

3.4 Summary

Complexity measures Let A be a Boolean matrix.

- $N^1(A)$ - the nondeterministic communication complexity of A.
- $N^0(A) = N^1(J - A)$, where J is the all-ones matrix.
- $N(A) = \max\{N^0(A), N^1(A)\}$.
- $C^1(A)$ - the size of a minimal cover of the ones of A by monochromatic rectangles.
- $\bar{C}^1(A)$ - fractional cover.

Relations Let A and B be a pair of Boolean matrices of size 2^n-by-2^n

- $D(A) \leq (N^0(A) + 2)(N^1(A) + 2)$.
- $N^1(A) = \lceil \log C^1(A) \rceil$.
- $\log \bar{C}^1(A) \leq N^1(A) \leq \log \bar{C}^1(A) + \log n + O(1)$.
- $N^1(A \otimes B) \geq N^1(A) + N^1(B) - 2 \log n - O(1)$.

4

Randomized communication complexity

Suppose that Alice and Bob have access to a source of randomness—that is, the next message they send is no longer a deterministic function of their input and prior messages, but can also depend on the outcome of a coin flip. In this case the output of a protocol is no longer fixed by the input, but can depend on the sequence of coin flips. Say that we also relax the requirement that the output of the protocol is always correct, and only require this with high probability. As we shall see in this chapter, for some problems randomness allows vastly more efficient protocols, and the task of proving lower bounds on randomized protocols is correspondingly more challenging.

There are several different versions of randomized communication complexity one can consider. First, one can vary the output conditions, and study protocols with no error, one-sided error, or two-sided error. In this survey, we will just consider the two-sided error case. Secondly, one can vary the players' mode of access to randomness, namely whether randomness is shared or private.

In the shared randomness or *public coin* model, at the beginning of the protocol Alice and Bob receive a common random string r. There is no bound on how long this random string can be, and its length has

no effect on the cost of the protocol. Alice and Bob then perform a deterministic protocol $\mathcal{P}_r$ where their messages can depend on the fixed string r. Thus in the public coin model, a randomized communication protocol is simply a probability distribution over deterministic protocols. We say that a randomized protocol $\mathcal{P}$ computes a function f with error at most ε if $\Pr_r[\mathcal{P}_r(x,y) = f(x,y)] \geq 1-\varepsilon$ for every input (x,y). The cost of a public coin protocol is the maximal cost of any of the deterministic protocols $\mathcal{P}_r$. We let $R_\varepsilon(f)$ denote the minimal cost of a public coin protocol which computes f with error at most ε.

The next theorem gives a nice equivalent definition of public coin randomized communication complexity, in matrix language.

Theorem 4.1. A Boolean matrix A has randomized communication complexity $R_\varepsilon(A) = h$ in the public coin model, if and only if there is a probability distribution $p_1, \ldots, p_m$ and Boolean matrices $B_1, \ldots, B_m$ such that

(1) $D(B_i) \leq h$ for every $i = 1, \ldots, m$.
(2) $\|A - \sum_{i=1}^m p_i B_i\|_\infty \leq \varepsilon$.

We wrote Theorem 4.1 with Boolean matrices, but we also use the corresponding steatement for sign matrices. The reader can easily verify that, in the corresponding statement for sign matrices, one only needs to change ε to 2ε in the last inequality.

In the *private coin* model, Alice and Bob independently receive a random string which is not seen by the other player. Rather than initially receiving a random string at the beginning of the protocol, it will be more convenient to imagine that Alice and Bob flip coins "as they go." More precisely, let R be an arbitrary finite set. A private coin protocol is a binary tree where each internal node is either labeled by a function $a_v : X \times R \to \{0,1\}$ or $b_v : X \times R \to \{0,1\}$. Leaves are labeled by elements of $\{0,1\}$. If a node v is labeled by a_v, then on input x Alice outputs 1 with probability $\Pr_{r \in R}[a_v(x,r) = 1]$ and outputs 0 with probability $\Pr_{r \in R}[a_v(x,r) = 0]$, and similarly for nodes labeled by b_v. We say that a private coin protocol $\mathcal{P}$ computes a function $f : X \times Y \to \{0,1\}$ with error at most ε if for every input (x,y)

the probability that the path traced through the tree arrives at a leaf labeled by $f(x, y)$ is at least $1 - \varepsilon$. The cost of a protocol is the height of the tree, and we denote $R_\varepsilon^{\text{pri}}(f)$ the minimal cost of a private coin protocol which computes f with error at most ε.

The first observation about the relative power of these two models is that $R_\varepsilon(f) \leq R_\varepsilon^{\text{pri}}(f)$. This is because in a public coin protocol Alice can simply use the left half of the common random string and Bob the right half to simulate a private coin protocol.

A very nice result by Newman [New91] shows that, at least for the bounded-error case we are interested in, private coin protocols can also simulate public coin protocols without too much overhead.

Theorem 4.2 (Newman [New91]). Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. For every $\varepsilon, \delta > 0$ it holds that

$$R_{\varepsilon+\delta}^{\text{pri}}(f) \leq R_\varepsilon(f) + O(\log n - \log \delta).$$

We will be concerned with the case where the error probability is a small constant, and in this case the theorem essentially says that public coin and private coin communication complexity are the same up to an additive $O(\log n)$ term. [1] For this reason, and as we are primarily concerned with lower bounds, we will focus on the public coin model in this chapter. In particular, when we refer to "randomized communication complexity" without specifying public or private coin, we mean the public coin model.

Before moving on to lower bounds for randomized communication complexity, it is first insightful to see an example where randomized protocols can be significantly more powerful than deterministic protocols. Luckily there is a simple example which shows the largest possible gap between deterministic and randomized complexity, and incidentally

[1] Notice that the smaller δ is, the larger the potential gap between these models. In the unbounded-error model, where the players need only succeed with probability strictly greater than 1/2, the public coin model becomes trivial—every function can be computed with constant communication. In the unbounded-error private coin model, on the other hand, the inner product function still has complexity $\Omega(n)$ [For02].

also shows an $\Omega(\log n)$ additive gap between the public and private coin models, also implying that Newman's theorem can be tight.

Consider the $2^n \times 2^n$ identity matrix I_{2^n}, the Boolean communication matrix for the EQUALITY problem. That is, $I_{2^n}[x, y] = 1$ if $x = y$ and $I_{2^n}[x, y] = 0$ otherwise. As this matrix is full rank, the log rank lower bound of Theorem 2.3 thus implies that $D(I_{2^n}) \geq n + 1$. The public coin randomized communication complexity of this problem on the other hand is $\Theta(1)$, and in the private coin model the complexity is $\Theta(\log n)$.

A randomized protocol that achieves this constant complexity is as follows. Alice and Bob interpret their inputs x and y respectively as elements of $\mathbb{Z}_2^n$, and also interpret the shared random string as a tuple of d strings $z_1, \ldots, z_d \in \mathbb{Z}_2^n$. Using d bits of communication Alice sends to Bob the values $\langle x, z_i \rangle \in \mathbb{Z}_2$ for $i = 1 \ldots d$. If the corresponding inner products $\langle y, z_i \rangle$ all agree, Bob reports that $x = y$, and otherwise he announces that x and y are distinct. If x and y are equal this protocol is always correct. If x and y are distinct the output of this protocol is correct with probability $1 - 2^{-d}$.

It is not hard to similarly devise an $O(\log n)$ complexity private coin protocol for equality, or one can simply invoke Newman's theorem to see this. A matching $\Omega(\log n)$ lower bound is given by the following relation between deterministic communication complexity and private coin randomized communication complexity.

Lemma 4.1 ([KN97]). For every function $f : X \times Y \to \{0, 1\}$

$$R^{\text{pri}}_{1/3}(f) \geq \Omega(\log D(f)).$$

The proof of this lemma works by presenting a deterministic simulation of a private coin randomized protocol.

4.1 Approximate rank

Our first lower bound technique for randomized communication complexity, as it was with deterministic communication complexity, will be based on matrix rank. For randomized complexity the relevant rank

bound becomes *approximate rank*, which is the minimal rank of a matrix "close" to the target matrix.

Definition 4.1 (Approximate rank). Let A be a sign matrix. The *approximate rank* of A with approximation factor α, denoted $\mathrm{rank}^{\alpha}(A)$, is

$$\mathrm{rank}^{\alpha}(A) = \min_{B:1\leq A[i,j]B[i,j]\leq\alpha} \mathrm{rank}(B).$$

Motivated by the limit as $\alpha \to \infty$, define

$$\mathrm{rank}^{\infty}(A) = \min_{B:1\leq A[i,j]B[i,j]} \mathrm{rank}(B).$$

This latter quantity is also known as the *sign rank* of A.

Note that approximate rank is defined as the optimum of an optimization problem. In practice, this optimum can be quite difficult to compute. While we do not know if computing approximate rank is NP-hard, the general class of problems of minimizing rank subject to linear constraints does contain NP-hard problems, see for example Section 7.3 of [VB96].

Krause [Kra96] showed that approximate rank can be used to lower bound randomized communication complexity.

Theorem 4.3 (Krause [Kra96]). Let A be a sign matrix and $0 \leq \varepsilon < 1/2$. Then

$$R^{\mathrm{pri}}_{\varepsilon}(A) \geq \log \mathrm{rank}^{\alpha}(A)$$

where $\alpha = 1/(1-2\varepsilon)$.

Before we go into the proof, let us mention its application to the previous example of the equality function. Alon [Alo09] shows a bound on the approximate rank of the identity matrix with approximation factor 2 of the form $\mathrm{rank}^{2}(I_{2^n}) = \Omega(n)$. Combined with Theorem 4.3, this implies that $R^{\mathrm{pri}}_{1/4}(I_{2^n}) \geq \log n - O(1)$, giving an alternative proof of this lower bound.

Proof. [Theorem 4.3] For this proof, it will be easier to work with Boolean matrices, thus let $A_0 = (J - A)/2$ be the Boolean version

of the sign matrix A, where J is the all-ones matrix. Consider an optimal private coin protocol $\mathcal{P}$ for A with success probability $1-\varepsilon$ and communication complexity c. Let P be a matrix where

$$P[x,y] = \Pr[\mathcal{P}(x,y)=1].$$

Notice that by assumption $\|A_0 - P\|_\infty \leq \varepsilon$. We will now show that the rank of P is at most 2^c. This will finish the proof by setting $P^* = -2P + J$, that is by taking the inverse of the transformation from Boolean to sign matrices above.

As described above, we can represent a randomized protocol as a binary tree with internal nodes labeled by functions $a_v : X \times R \to \{0,1\}$ or $b_v : Y \times R \to \{0,1\}$. Consider a leaf ℓ in this tree, and let $v_1, \ldots, v_k, \ell$ be the path from the root v_1 to ℓ. We can alternatively indicate this path by labels $s_1 \ldots s_k \in \{0,1\}^k$, specifying the values of a_{v_i} or b_{v_i}. For the sake of notational simplicity we assume that Alice and Bob strictly alternate speaking on this path. On input x, y, we can write the probability that the protocol arrives at leaf ℓ and outputs 1 as the product

$$\Pr_r[a_{v_1}(x,r) = s_1] \Pr_r[b_{v_2}(y,r) = s_2] \cdots \Pr_r[a_{v_k}(x,r) = s_k] \Pr_r[b_\ell(y,r) = 1]$$

By grouping together the terms that depend on x and separately those that depend on y, we can write this more succinctly as $U_\ell(x)V_\ell(y)$. Then we have

$$P[x,y] = \sum_\ell U_\ell(x)V_\ell(y).$$

Define the matrix U with rows labeled by elements from X and columns labeled by leaves ℓ, and where $U[x,\ell] = U_\ell(x)$. Define V similarly by $V[y,\ell] = V_\ell(y)$. Now both U and V have at most 2^c many columns and $P = UV^t$. This demonstrates that $\operatorname{rank}(P) \leq 2^c$. □

By Newman's theorem, we obtain the following corollary for public coin randomized complexity

---

Corollary 4.1. Let A be a 2^n-by-2^n sign matrix. For every $\varepsilon, \delta > 0$ such that $\varepsilon + \delta < 1/2$ set $\alpha = (1 - 2\varepsilon - 2\delta)^{-1}$. Then

$$R_\varepsilon(A) \geq \log(\operatorname{rank}^\alpha(A)) - O(\log n - \log \delta).$$

In analogy with the log rank conjecture for deterministic complexity it is natural to conjecture that approximate rank similarly gives a polynomially tight lower bound on randomized complexity.

Conjecture 4.1 (log approximate rank conjecture). Fix $0 \leq \varepsilon < 1/2$. There is a constant c such that for every sign matrix A

$$R_{\varepsilon}^{\mathrm{pri}}(A) \leq (\log \mathrm{rank}^{\alpha}(A))^{c} + 2$$

where $\alpha = 1/(1-2\varepsilon)$.

The limiting case as $\varepsilon \to 1/2$ of this conjecture is true as Paturi and Simon [PS86] show that $\log \mathrm{rank}^{\infty}(A)$ characterizes the unbounded-error complexity of A. In this model, the players simply have to output the correct answer with probability strictly greater than $1/2$. On the other hand, in the bounded-error case a larger gap is known here than for the deterministic log rank conjecture—for the DISJOINTNESS problem on n bits $R_{1/4}^{\mathrm{pri}}(\mathrm{DISJ}) = \Omega(n)$ [KS87] yet $\mathrm{rank}^{2}(\mathrm{DISJ}) = 2^{O(\sqrt{n})}$. The upper bound on the approximate rank of DISJOINTNESS follows from an upper bound on quantum communication complexity [BCW98, AA05], see also Chapter 5.

The main drawback to the approximate rank technique is that in practice it can be very difficult to bound. One of the goals of the framework of approximate norms, discussed in Section 4.2, is to develop easier to compute convex relaxations of approximate rank.

4.2 Approximate norms

Like matrix rank for deterministic communication complexity, approximate rank is one of the strongest lower bound techniques available for randomized communication complexity. In Chapter 2 we saw several norms that can be used to lower bound matrix rank, for example $\gamma_2(\cdot), \mu(\cdot), \|\cdot\|_{tr}$. We can naturally adapt these techniques to lower bound approximate rank by considering *approximate norms*.

Definition 4.2 (approximate norm). Fix a general norm Φ, which could be any of the above. For $\alpha \geq 1$ a real number and a sign matrix A, we define

$$\Phi^\alpha(A) = \min_{\substack{B \\ 1 \leq A[i,j]B[i,j] \leq \alpha}} \Phi(B).$$

Motivated by the limit as $\alpha \to \infty$ define

$$\Phi^\infty(A) = \min_{\substack{B \\ 1 \leq A[i,j]B[i,j]}} \Phi(B).$$

Note that an approximate norm is not in general itself a norm.

Recall Equation 2.1 which shows that for any real matrix A

$$\text{rank}(A) \geq \max_{u,v} \frac{\|A \circ uv^t\|_{tr}^2}{\|A \circ uv^t\|_F^2},$$

This immediately implies that

$$\text{rank}^\alpha(A) \geq \frac{\gamma_2^\alpha(A)^2}{\alpha^2}, \tag{4.1}$$

as any matrix B that approximates the sign matrix A with factor α satisfies $\|B \circ uv^t\|_F^2 \leq \alpha^2$, for arbitrary unit vectors u, v. The advantage of studying γ_2^α is that it is an optimization problem over a convex function, and it turns out that it can be computed with arbitrary accuracy in time polynomial in the size of the matrix by reduction to semidefinite programming.

As $\gamma_2^\alpha(A)$ provides a lower bound on approximate rank, Corollary 4.1 immediately implies that it, as well as $\mu^\alpha(A)$ and $\|A\|_{tr}^\alpha$, can be used to lower bound randomized communication complexity.

The characterization of public coin randomized complexity as a probability distribution over deterministic protocols (Theorem 4.1), however, leads to a much more natural proof based on convexity. We now give this proof which shows in general that for any norm Φ which gives a lower bound on deterministic communication complexity, the approximate norm Φ^α can be used to give a lower bound on randomized complexity.

Theorem 4.4. Let Φ be a norm. Suppose there is a constant c such that $\Phi(A) \leq 2^{cD(A)}$ for every sign matrix A. Then

$$cR_\varepsilon(A) \geq \log \Phi^\alpha(A) - \log \alpha,$$

for every sign matrix A, where $\alpha = \alpha(\varepsilon) = \frac{1}{1-2\varepsilon}$.

Proof. Since a public coin randomized communication protocol is equivalent to a probability distribution over deterministic protocols (Theorem 4.1), there are sign matrices $B_1, \ldots, B_m$ and a probability distribution $p_1, \ldots, p_m$ such that

(1) $\|A - \sum_{i=1}^m p_i B_i\|_\infty \leq 2\varepsilon$.
(2) $D(B_i) \leq R_\varepsilon(A)$ for $1 \leq i \leq m$.

Let $B = \frac{1}{1-2\varepsilon} \sum_i p_i B_i$. Chosen in this way, by item (1) we have that $1 \leq A[x,y]B[x,y] \leq \alpha$ for every entry (x,y). Therefore $\Phi^\alpha(A) \leq \Phi(B)$ by definition of an approximate norm. Since Φ is a norm we have

$$\begin{aligned}
\Phi(B) &= \Phi\left(\frac{1}{1-2\varepsilon} \sum_i p_i B_i\right) \\
&\leq \frac{1}{1-2\varepsilon} \sum_i p_i \Phi(B_i) \\
&\leq \frac{1}{1-2\varepsilon} \max_i \Phi(B_i).
\end{aligned}$$

Combining this with our assumption that $\Phi(B_i) \leq 2^{cD(B_i)}$ and with item (2), we obtain $\Phi(B) \leq \frac{1}{1-2\varepsilon} 2^{cR_\varepsilon(A)}$. This implies $\log \Phi^\alpha(A) - \log \alpha \leq cR_\varepsilon(A)$. □

As a corollary of this theorem, the approximate versions of the norms discussed in the chapter on deterministic complexity can all be used to lower bound randomized complexity. We just state this corollary for γ_2, which dominates the trace norm method and μ norm, up to a constant factor.

Corollary 4.2. For every sign matrix A and $0 \leq \varepsilon < \frac{1}{2}$

$$R_\varepsilon(A) \geq 2 \log \gamma_2^\alpha(A) - 2 \log \alpha,$$

where $\alpha = \alpha(\varepsilon) = \frac{1}{1-2\varepsilon}$.

Proof. Use the previous theorem together with the fact $\gamma_2(A) \leq 2^{D(A)/2}$. □

Up to an additive constant term, $2 \log \gamma_2^\alpha(A)$ gives tight bounds on the randomized communication complexity for almost every sign matrix [LMSS07].

Now γ_2^α is a convex relaxation of approximate rank which can be computed with high precision in polynomial time. The question remains: How much have we lost in considering this relaxation? Is there a sign matrix A for which $\gamma_2^\alpha(A)$ can be much smaller than $\mathrm{rank}^\alpha(A)$? The next theorem shows that, for constant α larger than 1, the answer is no.

Theorem 4.5 (Lee and Shraibman [LS08]). Let $1 < \alpha < \infty$. Then for any m-by-n sign matrix A

$$\mathrm{rank}^\alpha(A) = O\left(\frac{\alpha^6}{(\alpha-1)^6} \ln^3(4mn) \gamma_2^\alpha(A)^6\right).$$

Before giving a proof sketch, let us make a few remarks. First, for the case $\alpha = \infty$, Ben-David et al. [BES02] show the tighter result $\mathrm{rank}^\infty(A) = O(\ln(4mn)\gamma_2^\infty(A)^2)$. The interesting thing is that in this case, the *lower bound* fails. Notice from Equation 4.1 that the lower bound deteriorates as α grows. Buhrman et al. [BVW07] and independently Sherstov [She08b] have shown that this is *necessary*, giving an example of a sign matrix A where $\gamma_2^\infty(A)$ is exponentially larger than $\mathrm{rank}^\infty(A)$.

Secondly, we mention that the logarithmic term in this theorem is necessary as demonstrated by the equality example studied earlier.

Indeed, as $\log \gamma_2^\alpha$ is a lower bound on public coin communication complexity, $\gamma_2^\alpha(I_{2^n})$ is constant while Alon shows that $\mathrm{rank}^\alpha(I_{2^n}) = \Omega(n)$, for $\alpha = 2$. This example also demonstrates that the assumption $\alpha > 1$ cannot be removed, since $\mathrm{rank}(I_{2^n}) = 2^n$.

Proof. [sketch of Theorem 4.5] The proof combines two ideas. The first, observed by Alon and used by several authors [Alo03, AKM+06, KS07], is that approximate rank changes relatively slowly as a function of α, for $\alpha > 1$. Quantitatively, one can show

$$\mathrm{rank}^\alpha(A) \leq 2(\mathrm{rank}^{2\alpha-1}(A))^3.$$

This reduces the problem to showing

$$\mathrm{rank}^{2\alpha-1}(A) \leq O\left(\frac{\alpha^2}{(\alpha-1)^2}\ln(4mn)\gamma_2^\alpha(A)^2\right).$$

For this part, we use *dimension reduction*. Notice that both matrix rank and γ_2 are statements about matrix factorization. If a m-by-n matrix M has rank k, then there are vectors $x_1, \ldots, x_m \in \mathbb{R}^k$ and $y_1, \ldots, y_n \in \mathbb{R}^k$ such that $M[i,j] = \langle x_i, y_j \rangle$ for all $1 \leq i \leq m, 1 \leq j \leq n$. Similarly, if $\gamma_2(M) \leq \gamma$, then there are vectors $x_1, \ldots, x_m \in \mathbb{R}^r$ and $y_1, \ldots, y_n \in \mathbb{R}^r$ such that $M[i,j] = \langle x_i, y_j \rangle$ and $\|x_i\|^2, \|y_j\|^2 \leq \gamma$ for every i, j. In other words, in the case of rank, we have a bound on the *dimension* of the vectors x_i, y_j, while in the case of γ_2 we have a bound on their *length*. If the dimension r is much larger than the length of these vectors, however, then intuitively it seems plausible that they can be compressed to a smaller dimension without affecting the inner products $\langle x_i, y_j \rangle$ too much. This intuition can be made precise by the Johnson-Lindenstrauss lemma [JL01]. We use a variant of this lemma from [BES02] which addresses exactly this situation.

Lemma 4.2 (Corollary 19, [BES02]). Let $x, y \in \mathbb{R}^r$. Let R be a random k-by-r matrix with entries independent and identically distributed according to the normal distribution with mean 0 and variance 1. Then for every $\delta > 0$

$$\Pr_R\left[|\langle Rx, Ry\rangle - \langle x, y\rangle| \geq \frac{\delta}{2}\left(\|x\|_2^2 + \|y\|_2^2\right)\right] \leq 4\exp(-\delta^2 k/8).$$

Now let $\{x_i\}, \{y_j\}$ be a factorization which realizes $\gamma_2^\alpha(A) = \gamma$. That is, $\langle x_i, y_j \rangle = B[i,j]$ where $1 \leq A[i,j]B[i,j] \leq \alpha$ and $\|x_i\|_2^2, \|y_j\|_2^2 \leq \gamma$ for all i, j. By taking $k = 8(\frac{\alpha-1}{\alpha}\gamma)^2 \ln(8mn)$ we can ensure by a union bound that $|\langle Rx_i, Ry_j \rangle - \langle x_i, y_j \rangle| \leq \frac{\alpha-1}{2\alpha}$ simultaneously for all i, j with probability at least $1/2$ over the choice of R. Setting $x'_i = \frac{\alpha+1}{2\alpha} Rx_i$ and $y'_j = \frac{\alpha+1}{2\alpha} Ry_j$, the matrix whose (i,j) entry is $\langle x'_i, y'_j \rangle$ gives a $2\alpha - 1$ approximation to A and has rank at most k. □

4.3 Diagonal Fourier coefficients

We have just seen that the approximate norm γ_2^α is polynomially related to approximate rank, for constant α. This seems to indicate that one can generally show good lower bounds by this technique—indeed if the log approximate rank conjecture is true, then one can always show polynomially tight lower bounds by studying γ_2^α.

More concretely, Linial and Shraibman [LS09c] show that the γ_2^α bound subsumes many other techniques in the literature. Here we see that γ_2^α gives bounds at least as large as those given by a technique based on diagonal Fourier coefficients, developed for randomized communication complexity by Raz [Raz95] and later extended to the quantum case by Klauck [Kla01].

We identify vectors in $\mathbb{R}^{2^n}$ with functions $f : \mathbb{Z}_2^n \to \mathbb{R}$. Similarly we identify real $2^m \times 2^n$ matrices with functions $A : \mathbb{Z}_2^m \times \mathbb{Z}_2^n \to \mathbb{R}$.

Corresponding to every $z \in \mathbb{Z}_2^n$, is a character of $\mathbb{Z}_2^n$ denoted χ_z. It is defined as

$$\chi_z(x) = (-1)^{\langle z,x \rangle},$$

for every $x \in \mathbb{Z}_2^n$. The Fourier coefficients of a function $f : \mathbb{Z}_2^n \to \mathbb{R}$ are $\hat{f}_z = \frac{1}{2^n}\langle f, \chi_z \rangle$ for all $z \in \mathbb{Z}_2^n$. The characters of $\mathbb{Z}_2^m \times \mathbb{Z}_2^n$ are denoted $\chi_{z,z'}$ for $(z, z') \in \mathbb{Z}_2^m \times \mathbb{Z}_2^n$. They satisfy

$$\chi_{z,z'}(x, y) = (-1)^{\langle z,x \rangle + \langle z',y \rangle},$$

for every $x \in \mathbb{Z}_2^m$ and $y \in \mathbb{Z}_2^n$. We denote by $\hat{B}_{z,z'}$ the corresponding Fourier coefficient of a real matrix B.

Let A be a $2^n \times 2^n$ sign matrix. The subject of this section is the

sum of diagonal Fourier coefficients of A, i.e.

$$\|A\|_D = \sum_{z \in \mathbb{Z}_2^n} |\hat{A}_{z,z}|.$$

By our notation one can get the impression that the sum of diagonal Fourier coefficients is a norm. It is actually not a norm, because it can achieve the value 0 on nonzero matrices. On the other hand, $\|\cdot\|_D$ satisfies all the other properties of a norm: nonnegativity, homogeneity, and subadditivity. A function that satisfy these properties is called a *seminorm*.

Raz [Raz95] derived the following lower bound in terms of diagonal Fourier coefficients.

Theorem 4.6. For every sign matrix A and $\varepsilon < \frac{1}{2}$

$$R_\varepsilon(A) \geq \log(\|A\|_D^\alpha) - \log \alpha - O(1),$$

where $\alpha = \alpha(\varepsilon) = \frac{1}{1-2\varepsilon}$.

The lower bound in Theorem 4.6 follows from Corollary 4.2 using the following relation between γ_2 and the sum of diagonal Fourier coefficients.

Theorem 4.7. For every sign matrix A and $\alpha \geq 1$

$$\gamma_2^\alpha(A) \geq \frac{1}{2}\|A\|_D^\alpha.$$

We make use of the following simple fact:

Fact 4.1. Let $z \in \mathbb{Z}_2^m$ and $z' \in \mathbb{Z}_2^n$. The Fourier coefficient of the rank-one matrix fg^t with respect to z, z' is equal to $\hat{f}_z \hat{g}_{z'}$.

Proof. [of Theorem 4.7] Recall that $\gamma_2(M) \geq \frac{1}{2}\nu(M)$ for every real matrix M. We show next that $\nu(M) \geq \|M\|_D$. Combining these inequalities we get that $\gamma_2(M) \geq \frac{1}{2}\|M\|_D$ for every real matrix M. The

inequality between the corresponding approximate norms then easily follows.

We turn to prove the inequality $\nu(M) \geq \|M\|_D$. Since the unit ball of ν is the convex hull of rank-one sign matrices, it is enough to prove that for every rank-one sign matrix fg^t it holds that $\|fg^t\|_D \leq 1$. This follows because if $M = \sum_i \beta_i f_i g_i^t$ and $\nu(M) = \sum |\beta_i|$, then

$$\|M\|_D = \|\sum_i \beta_i f_i g_i^t\|_D \leq \sum_i |\beta_i| \|f_i g_i^t\|_D \leq \sum_i |\beta_i|.$$

To prove that $\|fg^t\|_D \leq 1$ we use Fact 4.1. We get

$$\begin{aligned} \|fg^t\|_D &= \sum_z \hat{f}_z \hat{g}_z \\ &\leq \left(\sum_z \hat{f}_z^2\right)^{1/2} \left(\sum_z \hat{g}_z^2\right)^{1/2} \\ &= \frac{1}{2^{n/2}}\|f\|_2 \frac{1}{2^{n/2}}\|g\|_2 \\ &= 1. \end{aligned}$$

The first inequality is an instance of the Cauchy-Schwarz inequality. The next identity holds because the characters form an orthogonal basis. The last equality uses the fact that f and g are sign vectors. □

Remarks

(1) The lower bounds proved in [Raz95, Kla01] in terms of diagonal Fourier coefficients, are different than what we have described above. But these bounds can be derived from the bounds we showed in terms of $\|\cdot\|_D^\alpha$, using simple properties of approximate norms and error amplification for the randomized communication complexity, see e.g. [LS09c].

(2) As observed by Raz, instead of diagonal Fourier coefficients one can consider other sets of Fourier coefficients. One can prove a similar statement (repeating the same arguments) for any subset F of $\mathbb{Z}_2^n \times \mathbb{Z}_2^n$ of the form $F = \{(z, \pi(z)) : z \in \mathbb{Z}_2^n\}$, where π is some permutation over $\mathbb{Z}_2^n$.

4.4 Distributional complexity and discrepancy

We turn to an alternative characterization of randomized complexity in terms of *distributional complexity* given by Yao [Yao83]. Distributional complexity considers deterministic algorithms, but allows them to answer incorrectly on some inputs (x, y). The requirement now is that the total measure of incorrect answers should be small, according to a probability distribution fixed in advance. This alternative characterization is very useful for proving lower bounds.

Recall that a public coin randomized protocol of complexity t is simply a probability distribution over deterministic protocols of cost at most t. This leads us to consider a geometric object, the *convex hull* of sign matrices with deterministic communication complexity at most t. Denote this convex body by $\mathcal{B}(D, t)$. A matrix B is in $\mathcal{B}(D, t)$ if and only if it can be written as the *convex combination* of sign matrices with deterministic communication complexity at most t. That is, if there is a probability distribution $p_1, \ldots, p_m$ and matrices $B_1, \ldots, B_m$ such that $D(B_i) \leq t$ for $i = 1 \ldots m$ and $B = \sum_{i=1}^{m} p_i B_i$. Obviously,

$$\mathcal{B}(D, 1) \subseteq \mathcal{B}(D, 2) \subseteq \mathcal{B}(D, 3) \ldots$$

As observed in Theorem 4.1, the randomized communication complexity $R_\varepsilon(A)$ is equal to the smallest number t such that A is 2ε close in ℓ_∞ distance to $\mathcal{B}(D, t)$. Formally, $R_\varepsilon(A)$ is equal to the minimal t such that

$$\min_{B \in \mathcal{B}(D,t)} \|A - B\|_\infty \leq 2\varepsilon.$$

This geometric definition shows that randomized complexity is an optimization problem over a convex set. As we shall see in Chapter 6 this allows the use of duality to obtain an equivalent characterization of randomized complexity in terms of a maximization problem. For the moment we simply state this result, and defer the proof to Section 6.2.1

Fix a distribution P on the entries of A. The distributional complexity of A with respect to P and ε, denoted $D_{P,\varepsilon}(A)$, is the minimal number t such that there exists a sign matrix B satisfying $D(B) \leq t$ and $P(\{(i, j) : A[i, j] \neq B[i, j]\}) \leq \varepsilon$. The *distributional* complexity of A with error parameter ε is defined

$$D_\varepsilon(A) = \max_P D_{P,\varepsilon}(A).$$

Using the von Neumann minimax theorem, Yao [Yao83] showed the following.

Theorem 4.8. Let A be a sign matrix.

$$R_\varepsilon(A) = D_\varepsilon(A).$$

This theorem is the starting point for many lower bound proofs on randomized complexity.

Discrepancy One early lower bound technique developed for showing lower bounds on randomized complexity based on Theorem 4.8 is the *discrepancy method*. This is a very general technique, which applies even as the error probability becomes very close to $1/2$.

The discrepancy bound can be derived as follows. Let A be a sign matrix, and P a probability distribution over its entries. We think of P as a nonnegative matrix whose entries sum to one and which is of the same dimensions as A. Let $t = D_{P,\varepsilon}(A)$ and let B be a sign matrix which "realizes" this—in other words B agrees with A with probability at least $1-\varepsilon$ under the distribution P and $D(B) = t$. As A and B are different with probability at most ε with respect to P we have

$$\langle A \circ P, B \rangle \geq 1 - 2\varepsilon \tag{4.2}$$

On the other hand, by Theorem 2.1, the matrix B can be partitioned into at most 2^t monochromatic combinatorial rectangles $\{R_\ell\}$. Let $c(R_\ell) \in \{-1, +1\}$ be the "color" of R_ℓ. We can rewrite Equation 4.2 as

$$\langle A \circ P, B \rangle = \sum_\ell c(R_\ell) \sum_{(i,j)\in R_\ell} A[i,j]P[i,j] \tag{4.3}$$

$$\leq \sum_\ell \left| \sum_{(i,j)\in R_\ell} A[i,j]P[i,j] \right|. \tag{4.4}$$

Denote

$$\mathrm{disc}_P(A) = \max_R \left| \sum_{(i,j)\in R} A[i,j]P[i,j] \right|,$$

where the maximum is over all combinatorial rectangles. This is simply the cut norm of the matrix $A \circ P$, so we can alternatively say $\mathrm{disc}_P(A) = \mu^*(A \circ P)$. It then follows from the Equations (4.2), (4.4) that $2^t \mathrm{disc}_P(A) \geq 1 - 2\varepsilon$, or equivalently

$$D_{P,\varepsilon}(A) \geq \log \frac{1-2\varepsilon}{\mathrm{disc}_P(A)}.$$

Let $\mathrm{disc}(A) = \min_P \mathrm{disc}_P(A)$ be the *discrepancy* of A. As this derivation holds for an arbitrary probability distribution P we get

---

Theorem 4.9 (discrepancy method). For every sign matrix A and every $0 \leq \varepsilon < 1/2$

$$R_\varepsilon(A) \geq \log \frac{1-2\varepsilon}{\mathrm{disc}(A)}.$$

Before moving on, let us connect the discrepancy bound with the linear programming relaxation of nondeterministic communication complexity we saw in Chapter 3. Both $\mathrm{disc}(A)$ and $\bar{C}^1(A)$ are optimization problems over probability distributions. The dual formulation of $\bar{C}^1(A)$ can be expressed as

$$\frac{1}{\bar{C}^1(A)} = \min_P \max_R P(R) = \min_P \max_R \sum_{(i,j)\in R} A[i,j]P[i,j]$$

where P is a probability distribution over $\{(i,j) : A[i,j] = 1\}$ and R is a 1-monochromatic rectangle. On the other hand, in discrepancy the maximization is over *all* rectangles. Thus as long as the optimal distribution in the discrepancy bound puts substantial weight on the set $\{(i,j) : A[i,j] = 1\}$, we will have $\bar{C}^1(A) = \Omega(1/\mathrm{disc}(A))$. If the optimal distribution in the discrepancy bound does not put substantial weight on $\{(i,j) : A[i,j] = 1\}$, then the discrepancy method does not show a good lower bound anyway. This gives the following theorem.

Theorem 4.10. Let A be a sign matrix. Then

$$\bar{C}^1(A) \geq \frac{1}{3\,\mathrm{disc}(A)}.$$

Proof. Let P be a probability distribution such that $\mathrm{disc}(A) = \mathrm{disc}_P(A)$. If P puts weight less than $1/3$ on $\{(i,j) : A[i,j] = 1\}$ then clearly $\mathrm{disc}(A) \geq 1/3$. Otherwise, there is a 1-monochromatic rectangle R such that $3P(R) \geq 1/\bar{C}^1(A)$. As $\mathrm{disc}(A) \geq P(R)$, we get the theorem. □

In particular, the discrepancy method also gives a lower bound on nondeterministic communication complexity.

4.5 Corruption bound

The last lower bound technique we discuss for randomized communication complexity is the *corruption bound*. This is a powerful lower bound technique which is particularly interesting because for some functions it gives larger lower bounds than approximate rank. Indeed, the corruption bound can be used to show a tight $\Omega(n)$ lower bound on the complexity of DISJOINTNESS [Raz92a], while approximate rank can only show a lower bound of $\Theta(\sqrt{n})$. In particular, as the quantum communication complexity of DISJOINTNESS is $O(\sqrt{n})$ [BCW98, AA05] the corruption bound is noteworthy as one of the few lower bound techniques we know that can potentially separate quantum and randomized communication complexity for total functions. It is a major open question if these measures are polynomially related for all total functions. It is also an open problem whether the corruption bound and approximate rank are polynomially related.

The corruption bound does not really fall into the framework of lower bounds described in this survey. The definition of this bound is purely combinatorial, and it is not known if there is an equivalent definition which involves a representation of the function in Euclidean space. Still we feel it is insightful to describe this lower bound here and encourage further study of it and its relation with the other bounds in this survey.

The corruption bound was first introduced in a paper of Yao [Yao83], and was derived using the characterization of randomized complexity in terms of distributional complexity given in that same paper. Over the years, this technique has been variously called *one-sided discrepancy*,

ε-error rectangle bound [Kla03], and *corruption bound* [BPSW06]. We will follow the latter terminology.

The original application of Yao gave the first explicit example of a function whose randomized two-party complexity is $\Theta(n)$. Let p be an n-bit prime. The function he considered is $f : \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_2$ defined as $f(x, y) = ((x \cdot y) \mod p) \mod 2$. In other words, $f(x, y)$ is the parity of $(x \cdot y) \mod p$. To prove a lower bound on this function, Yao proved the following general statement.

Let $f : X \times Y \to Z$ be a function and say $|X| = |Y| = N$. For each $x \in X$ and $z \in Z$, denote $S_x(z) = \{y : f(x, y) = z\}$. We write $f^{-1}(z)$ for the set of indices (x, y) for which $f(x, y) = z$.

Fix constants $0 < \mu < 1/2$ and $0 < \lambda$. We need the notions of *moderate* and *anticorrelated*. For $z \in Z$ we say that f is z-moderate if $\mu < |f^{-1}(z)|/N^2 < 1 - \mu$. We say that f is z-anticorrelated if

$$|S_x(z) \cap S_{x'}(z)| \leq \frac{1}{N}|S_x(z)| \cdot |S_{x'}(z)|(1 + O(\frac{1}{N^\lambda})).$$

for every $x, x' \in X$.

Theorem 4.11 (Yao [Yao83]). Let $\varepsilon < 1/2$ be a fixed constant. If $f : X \times Y \to Z$ is both z-moderate and z-anticorrelated for some $z \in Z$, then $R_\varepsilon(f) = \Theta(\log N)$, where $|X| = |Y| = N$.

The above lower bound is very elegant. It implies in particular that a pseudorandom sign matrix has high randomized communication complexity.

A key lemma in the proof of this theorem is the corruption bound technique, which has since been used and reformulated by several authors [BFS86, Raz92a, Kla03, BPSW06]. Intuitively, the corruption bound interpolates between the linear programming relaxation of nondeterministic communication complexity $\bar{C}^1(f)$ and the discrepancy method.

$$\frac{1}{\bar{C}^1(A)} = \min_P \max_{R:\chi(R)=1} \sum_{i,j \in R} A[i, j]P[i, j]$$

$$\mathrm{disc}(A) = \min_P \max_R \sum_{i,j \in R} A[i, j]P[i, j].$$

In the first case, the probability distribution P is only over $\{(i,j) : A[i,j] = 1\}$ and the maximization is over 1-monochromatic rectangles; in the second case the probability distribution is arbitrary but the maximization is over all rectangles. In the corruption bound, we minimize over probability distributions that put sufficient weight on $\{(i,j) : A[i,j] = 1\}$, and maximize over rectangles that are *nearly* 1-monochromatic.

For a probability distribution μ and function f, say that a rectangle R is ε-monochromatic if $\mu(R \cap f^{-1}(0)) \leq \varepsilon\mu(R)$. Further, let $\varepsilon\text{-mono}_\mu(f) = \max\{\mu(R) : R \text{ is } \varepsilon\text{-monochromatic}\}$ be the largest weight given to an ε-monochromatic rectangle under μ. The corruption bound gives the following.

Theorem 4.12 (corruption bound [Yao83, BPSW06]). Let B be a Boolean matrix and μ a probability distribution on the entries of B. Let $\varepsilon' > \varepsilon \geq 0$. Then

$$R_\varepsilon(B) \geq \frac{\langle B, \mu\rangle - \varepsilon - \varepsilon/\varepsilon'}{\varepsilon\text{-mono}_\mu(f)}.$$

Proof. We first use Yao's principle that $R_\varepsilon(B) = D_\varepsilon(B)$. Suppose that $D_\varepsilon(B) = c$. Fix a distribution μ. Then here is a deterministic c-bit protocol $\mathcal{P}$ which errs on B with probability at most ε with respect to μ. Let $\{R_i\}$ be a partition of B into 2^c many rectangles given by this protocol, and let I be the set of indices i for which the protocol outputs 1 on R_i. Note that this partition satisfies the following two properties.

(1) The probability with respect to μ that R_i contains a zero is at most ϵ:
$$\sum_{i\in I} \mu(R_i \cap f^{-1}(0)) \leq \epsilon.$$

(2) The rectangles R_i cover nearly all the ones of B, with respect to μ:
$$\sum_{i\in I} \mu(R_i) \geq \langle B, \mu\rangle - \epsilon.$$

Set a cut-off value $\varepsilon' > \varepsilon$. Let $I' \subseteq I$ be the set of indices of ε'-monochromatic rectangles. By item (1), the total weight of rectangles which are not ε'-monochromatic is at most ε/ε'. We then have

$$\sum_{i \in I'} \mu(R_i) \geq \langle B, \mu \rangle - \varepsilon - \varepsilon/\varepsilon'$$
$$\sum_{i \in I'} \mu(R_i) \leq 2^c \max_{\substack{R \\ \varepsilon'-\text{monochromatic}}} \mu(R).$$

Rearranging gives the theorem. □

It is an open problem if the corruption bound is polynomially related to randomized communication complexity (open problem 3.23 from [KN97]). Klauck [Kla03] has shown interesting connections between the corruption bound and the communication models corresponding to the complexity classes MA and AM. He shows that the square root of the corruption bound is a lower bound on MA complexity, thereby giving an $\Omega(\sqrt{n})$ lower bound on the MA complexity of disjointness. Surprisingly, this bound turns out to be tight by a beautiful protocol of Aaronson and Wigderson [AW09]. Klauck also shows that the corruption bound gives an *upper bound* on AM communication complexity.

4.6 Summary

Complexity measures Let A be a sign matrix and let M be a real matrix

- $R_\varepsilon(A)$ - the randomized communication complexity of A with error bound $0 \leq \varepsilon < \frac{1}{2}$.
- $\text{rank}^\alpha(A)$ - the approximate rank of A, with approximation factor $\alpha \geq 1$.
- $D_\varepsilon(A)$ - the distributional complexity of A with error at most $0 \leq \varepsilon < \frac{1}{2}$.
- $\text{disc}(A)$ - the discrepancy of A.
- $\Phi^\alpha(A)$ - the approximate Φ-norm of A, with approximation factor $\alpha \geq 1$.
- $\|M\|_D$ - the sum of diagonal Fourier coefficients of M.

Relations Let $m \leq n$. For every $m \times n$ sign matrix A, a real matrix M and $0 \leq \varepsilon < \frac{1}{2}$

- $R_\varepsilon(A) = D_\varepsilon(A)$.
- $R_\varepsilon(A) \geq \log \frac{1-2\varepsilon}{\mathrm{disc}(A)}$.
- $R_\varepsilon(A) \geq 2 \log \gamma_2^\alpha(A) - 2 \log \alpha$, where $\alpha = \alpha(\varepsilon) = \frac{1}{1-2\varepsilon}$.
- $R_\varepsilon(A) \geq \log \mathrm{rank}^\alpha(A) - O(\log n)$, where $\alpha = \alpha(\varepsilon) = \frac{1}{1-2\varepsilon}$.
- $\mathrm{rank}^\alpha(A) \geq \frac{\gamma_2^\alpha(A)^2}{\alpha^2}$.
- $\mathrm{rank}^\alpha(A) = O_\alpha\left(\ln^3(4mn)\gamma_2^\alpha(A)^6\right)$, for every $1 < \alpha < \infty$.
- $\gamma_2^\alpha(A) \leq \nu^\alpha(A) \leq \mu^\alpha(A) \leq 4K_G\gamma_2^\alpha(A)$.
- $\nu(M) \geq \|M\|_D$.

5

Quantum communication complexity

In this chapter, we look at lower bounds on quantum communication complexity. We will encounter familiar faces from the chapter on randomized communication complexity. Indeed, it is a major open question whether or not quantum communication complexity and randomized complexity are polynomially related for all total functions. For a partial function, however, an exponential separation between these models is known [Raz99].

A major obstacle to showing larger separations between these models for total functions—other than the fact that perhaps no such separation exists—is that nearly all lower bound techniques developed for the randomized model also work in the quantum model, even in the strongest quantum model where Alice and Bob share prior entanglement.

One outstanding exception to this is the corruption bound from Section 4.5 which as mentioned can be used to show a lower bound of $\Omega(n)$ on the randomized communication complexity of disjointness [KS87, Raz92b]. On the other hand, the quantum communication complexity of disjointness is $\Theta(\sqrt{n})$ [Raz03, AA05], giving a quadratic gap between these models. This is the largest gap known for a total function.

5.1 Definition of the model

Intuitively, quantum communication complexity models are defined similarly to the corresponding classical ones, with qubits replacing classical bits. Formally the protocols are described by unitary transformations operating on vectors in a Hilbert space. The *state* of a quantum communication protocol is represented as vector in a Hilbert space $H_A \otimes C \otimes H_B$. Here H_A, H_B are Hilbert spaces of arbitrary finite dimension representing the "workspace" of Alice and Bob respectively. The Hilbert space C is 2-dimensional and it stands for a one-qubit channel. We assume that H_A and H_B each contain a register to hold the input. Thus we have $H_A = H_{I_A} \otimes H_{W_A}$ and $H_B = H_{I_B} \otimes H_{W_B}$ where H_I is a register and H_W represents workspace that is used arbitrarily.

In the model without entanglement, the initial state of a quantum protocol on input (x, y) is the vector $|\Psi^0_{x,y}\rangle = |x, 0\rangle|0\rangle|y, 0\rangle$, where in each case $|0\rangle$ is an arbitrary unit vector independent of the input. Thus informally, the workspaces and channel are initially "clear."

With entanglement, the initial state is a unit vector of the form $|\Psi^0_{x,y}\rangle = \sum_w \alpha_w |x, w\rangle|0\rangle|y, w\rangle$, where the coefficients α_w are arbitrary real numbers satisfying $\sum_w \alpha_w^2 = 1$. This difference in allowed initial state is the only change between the models with and without entanglement.

Unlike other models of communication, here we assume that the speaking order of the players strictly alternates [1]. Alice and Bob "speak" by applying a unitary transformation to the current state of the protocol. On Alice's turn, she applies an arbitrary unitary transformation of the form $U \otimes I_B$ which acts as the identity on H_B. Similarly, on a turn of Bob he applies a transformation of the form $I_A \otimes U$ which acts as the identity on H_A. Thus after $2t$ rounds, the state of the protocol is

$$|\Psi^{2t}_{x,y}\rangle = (I_A \otimes U_{2t}) \cdots (I_A \otimes U_2)(U_1 \otimes I_B)|\Psi^0_{x,y}\rangle$$

At the end of a t-round protocol, we project the final state $|\Psi^t_{x,y}\rangle$ onto the subspace $H_A \otimes |1\rangle \otimes H_B$. Denoting the length of this projection by p, the protocol outputs 1 with probability p^2, and outputs 0

[1] Note that this requirement makes a difference of at most a multiplicative factor of two in the communication complexity.

otherwise. As the state of a quantum protocol is a unit vector, this is equivalent to outputting 0 with probability the norm squared of the projection of the final state onto $H_A \otimes |0\rangle \otimes H_B$. A quantum protocol computes a Boolean matrix B with probability at least $1-\varepsilon$, if its output is in $\{0,1\}$ and the output is equal to $B[x,y]$ with probability at least $1-\varepsilon$ for every input (x,y). The *cost* of a protocol is simply the number of rounds. We define the *quantum communication complexity* $Q_\varepsilon(B)$ of a Boolean matrix B, to be the minimal cost of a quantum protocol for B which succeeds with probability at least $1-\varepsilon$ in the model without entanglement. We define $Q^*_\varepsilon(B)$ similarly for the model with entanglement. We will also refer to these measures for a sign matrix A by the usual identification with the Boolean matrix $B = (A-J)/2$, where J is the all-ones matrix.

5.2 Approximate rank

As we have already been accustomed, we begin with a lower bound based on matrix rank. Buhrman and de Wolf [BW01] showed that approximate rank provides a lower bound on quantum communication without entanglement.

Theorem 5.1 (Buhrman and de Wolf [BW01]). Let A be a sign matrix, and $0 \leq \varepsilon < 1/2$, then

$$Q_\varepsilon(A) \geq \frac{\log \operatorname{rank}^\alpha(A)}{2},$$

where $\alpha = 1/(1-2\varepsilon)$.

Notice that this bound is the same as the approximate rank bound for private coin randomized protocols, up to the factor of two. Similarly to public coin randomized protocols, this bound does not hold as is for the model of quantum communication with entanglement—with entanglement Alice and Bob can simulate a public coin.

Instead of proving Theorem 5.1, in the next section we prove a lower bound on quantum communication complexity with entanglement in terms of γ_2^α. As γ_2^α and approximate rank are polynomially related Theorem 4.5, this implies that the logarithm of approximate rank also

lower bounds this model, up to a slightly larger multiplicative factor and logarithmic fudge terms.

Where there is a rank bound, so goes a log rank conjecture. The quantum model is no exception.

Conjecture 5.1 (log rank conjecture, quantum version). Fix $0 \leq \varepsilon < 1/2$. There is a constant c such that for every sign matrix A

$$Q_\varepsilon(A) \leq (\log \operatorname{rank}^\alpha(A))^c + 2,$$

where $\alpha = 1/(1-2\varepsilon)$.

In our opinion, the quantum version of the log rank conjecture is the most plausible of them all. Indeed, here we are not aware of any example where quantum communication complexity is significantly larger than the logarithm of approximate rank.

Question 5.1. Give an example of a sign matrix A such that

$$Q_{1/3}(A) \geq (\log \operatorname{rank}^2(A))^c + 2$$

for a constant $c > 1$.

5.3 A lower bound via γ_2^α

As approximate rank is a lower bound on quantum communication complexity, all the methods subsumed by the approximate rank technique will also lower bound quantum complexity. In particular, the approximate norm γ_2^α can be used to lower bound quantum communication complexity. We present a separate proof for this fact however, which gives a sharper bound and also works in the stronger model with entanglement.

Previously, we proved lower bounds on communication complexity by identifying a set of "simple" objects and seeing how a correct protocol decomposes the communication matrix as a linear combination of these simple objects. For example, in the case of deterministic protocols these simpler objects were combinatorial rectangles, or equivalently rank-one Boolean matrices. The μ norm measures the lowest "weight" decomposition of a matrix as a linear combination of combinatorial rectangles.

Combinatorial rectangles came up naturally in the context of deterministic communication protocols, as the latter partition the matrix into monochromatic combinatorial rectangles. This property is not shared by quantum communication protocols in any apparent way. Remarkably, an efficient quantum protocol *still* leads to an efficient (approximate) description of the underlying matrix in terms of rank-one Boolean matrices, just as in the randomized case. This fact was first observed by Klauck [Kla01] who showed that $Q_\varepsilon(A) = \Omega(\log \mu^\alpha(A))$.

Klauck's work builds on earlier work of Yao and Kremer [Yao93, Kre95] which began to formalize the mathematical properties of a sign matrix with small bounded-error quantum communication complexity. We present here what is currently the sharpest analysis of this form, which is in terms of the γ_2 norm due to Linial and Shraibman [LS09c]. Roughly speaking, this result says that if a sign matrix A has bounded-error quantum communication complexity c, even in the model with entanglement, then there is a matrix B close to A which satisfies $\gamma_2(B) \leq 2^c$. The relation to combinatorial rectangles is then nicely explained by Grothendieck's inequality which says that γ_2 and μ are closely related (See also Section 2.3).

The proof of Linial and Shraibman proceeds by explicitly constructing a factorization to imply that γ_2 is small. For variety, we present here a different proof of this result communicated to us by Harry Buhrman [Buh07] that proceeds via the dual norm of γ_2 and exploits the connection of γ_2^* with so-called XOR games. A similar approach is used by Degorre et al. [DKLR08] to show a lower bound not just on functions, but also on the complexity of simulating probability distributions. To explain this approach, let us first introduce XOR games.

5.3.1 XOR games

A XOR game is played by three parties, provers Alice and Bob and a verifier V. Let A be a $|X|$-by-$|Y|$ sign matrix for two finite sets X, Y. The verifier chooses $(x, y) \in X \times Y$ according to some probability distribution $\pi(x, y)$ and sends x to Alice and y to Bob. The provers then answer by $a_x, b_y \in \{-1, +1\}$ with the goal that $a_x \cdot b_y = A[x, y]$. In other words, the provers want the XOR of their answers to agree with

A, explaining the name "XOR game." The provers are allowed to have access to a shared random string, but a convexity argument shows that they can perform as well without it.

We will be interested in the maximal correlation the provers are able to achieve with A under the distribution π. By definition, this is the probability that the provers answer correctly minus the probability that they answer incorrectly. One can see that this is exactly given by

$$\begin{aligned}\mathrm{Corr}_\pi(A,P) &= \max_{\substack{a\in\{-1,+1\}^{|X|}\\ b\in\{-1,+1\}^{|Y|}}} \sum_{x,y} A[x,y]\pi(x,y)a_x \cdot b_y \\ &= \max_{\substack{a\in\{-1,+1\}^{|X|}\\ b\in\{-1,+1\}^{|Y|}}} a^T(A\circ\pi)b.\end{aligned}$$

This last quantity is exactly the infinity-to-one norm introduced in Section 2.3.5. Thus we can equivalently say

$$\mathrm{Corr}_\pi(A,P) = \|A\circ\pi\|_{\infty\to 1}.$$

We have just described a classical XOR game. We will actually be interested in the case where Alice and Bob share entanglement. Formally, a XOR protocol with entanglement is described by a shared quantum state $|\psi\rangle \in \mathbb{C}^{d\times d}$ for some $d \geq 1$ and a choice of ± 1 valued measurement observables A_x for every $x \in X$ and similarly B_y for every $y \in Y$. On input (x, y) the protocol outputs 1 with probability

$$\langle\psi|A_x \otimes B_y|\psi\rangle.$$

Thus for a protocol P with shared entanglement and sign matrix A we have

$$\mathrm{Corr}_\pi(A,P) = \sum_{x,y} A[x,y]\pi(x,y)\cdot\langle\psi|A_x\otimes B_y|\psi\rangle.$$

Tsirelson [Tsi87] has given a very elegant characterization of the maximal correlation achievable by the provers in this situation.

Theorem 5.2 (Tsirelson [Tsi87]). Let A be a sign matrix. For any probability distribution π over A

$$\max_P \mathrm{Corr}_\pi(A,P) = \gamma_2^*(A\circ\pi).$$

where the maximum is taken over XOR protocols P with shared entanglement.

A very accessible presentation of this result can be found in the appendix of Unger's thesis [Ung08]. Note that as the maximization in the definition of γ_2^* is over a larger set than that of $\|A\|_{\infty\to 1}$ we have $\|A \circ \pi\|_{\infty\to 1} \leq \gamma_2^*(A \circ \pi)$. The famous CHSH game [CHSH69] is a simple XOR game where provers sharing entanglement can do better than those without, under the uniform distribution. The CHSH game is described by the matrix

$$\mathrm{CHSH} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

As this matrix represents the truth table of the AND function (where -1 represents true), in words the goal of the provers in a CHSH game on input x, y is to output a, b such that $a \oplus b = x \wedge y$. It is not difficult to show that $1/2 = (1/4)\|\mathrm{CHSH}\|_{\infty\to 1} < (1/4)\gamma_2^*(\mathrm{CHSH}) = 1/\sqrt{2}$, thus provers sharing entanglement can achieve strictly better correlation in the CHSH game under the uniform distribution.

The gap between $\|A \circ \pi\|_{\infty\to 1}$ and $\gamma_2^*(A \circ \pi)$ cannot be much larger. Grothendieck's inequality (Theorem 2.12) ensures that $\gamma_2^*(M) \leq K_G\|M\|_{\infty\to 1}$ for every real matrix M, where $K_G \leq 1.78\ldots$ is Grothendieck's constant. Thus the $\sqrt{2}$ gap exhibited by the CHSH game example is not far from optimal.

5.3.2 A communication protocol gives a XOR protocol

As observed by Harry Buhrman, Theorem 5.2 can be used to give an alternative proof of a result of Linial and Shraibman. See also [DKLR08] for a proof of a more general result along similar lines.

Theorem 5.3 (Linial and Shraibman [LS09c]). Let A be a sign matrix and $0 \leq \varepsilon < 1/2$, then

$$\begin{aligned} 2^{Q_\varepsilon^*(A)} &\geq (1-2\varepsilon) \cdot \gamma_2^{1/(1-2\varepsilon)}(A) \\ &= \max_M \frac{(1-\varepsilon)\langle A, M\rangle - \varepsilon\|M\|_1}{\gamma_2^*(M)}. \end{aligned}$$

The equality here follows by duality as proved in 6.2.2.

Buhrman's proof actually gives a slightly weaker bound

$$2^{2Q^*_\varepsilon(A)} \geq \max_M \frac{\langle A, M\rangle - 2\varepsilon\|M\|_1}{\gamma_2^*(M)}. \tag{5.1}$$

The better bound in Theorem 5.3 makes use of the assumption that the players strictly alternate in speaking order, which the proof given here does not. For comparison, recall that we know that the square of the right hand side in Theorem 5.3 is a lower bound on $2^{R_\varepsilon(A)}$.

Proof. [of inequality (5.1)] Say that $Q^*_\varepsilon(A) = c$. By using *teleportation* two players who share entanglement can encode a quantum bit with two classical bits. We refer the reader to the textbook of Nielsen and Chuang for details [NC00]. This allows the transformation of the original protocol into one with the same properties using at most $2c$ classical bits. Let $R[x, y]$ denote the expectation of the output of this protocol on input x, y. Note that, by assumption of the correctness of the protocol, if $A[x, y] = 1$ then $1 - 2\varepsilon \leq R[x, y] \leq 1$ and if $A[x, y] = -1$ then $-1 \leq R[x, y] \leq -1 + 2\varepsilon$.

Fix a probability distribution π and let B be an arbitrary sign matrix of the same dimensions as A. We will see how the communication protocol for A can be used to design a XOR protocol for B. The bias of this protocol will be related to the amount of communication c and the correlation $\langle A, B \circ \pi\rangle$.

The strategy in the XOR game is as follows: As the provers share entanglement, we may also assume they share a random string r of length $2c$. Essentially, the players will simulate their actions in the communication protocol but instead of actually communicating will take the responses from the other player to be given by r. More explicitly, say that Alice speaks first in the communication protocol—she makes a measurement on the entangled state and communicates a t-bit string b to Bob. In the XOR protocol, she does the exact same thing but instead of sending the message to Bob, she checks that what she would send agrees with the first t bits of r. If at any point in the protocol r does not agree with the communication Alice would have sent, we say

that r is inconsistent with Alice. Similarly, Bob treats the first t bits of r as if this was the message he received from Alice, and performs the same action he would then do in the communication protocol. If Bob ever sees that r does not agree with the message he would send in the communication protocol, we say that r is inconsistent with Bob.

Now we define the output conditions

- If the random string r is inconsistent with Alice, then she outputs a random bit in $\{-1, +1\}$. Otherwise, she outputs a bit $\{-1, +1\}$ with expectation $R[x, y]$.
- If r is inconsistent with Bob, then he outputs a random bit. Otherwise, he outputs 1.

Let $P(x, y)$ be the expected output of this protocol on input x, y. Let us now compute the correlation of this protocol with B under π:

$$\begin{aligned}\gamma_2^*(B \circ \pi) &\geq \mathrm{Corr}_\pi(B, P) \\ &= \frac{1}{2^{2c}} \sum_{x,y} \pi(x, y) B[x, y] R[x, y] \\ &\geq \frac{1}{2^{2c}} \left(\sum_{x,y} \pi(x, y) B[x, y] A[x, y] - 2\varepsilon \right)\end{aligned}$$

Rearranging, this gives the desired result:

$$2^{2c} \geq \max_{B,\pi} \frac{\langle A, B \circ \pi \rangle - 2\varepsilon}{\gamma_2^*(B \circ \pi)} = \max_M \frac{\langle A, M \rangle - 2\varepsilon \|M\|_1}{\gamma_2^*(M)}$$

□

5.4 Summary: equivalent representations

Recall that this survey revolves around a "three-step approach." We focus on lower bounds that generally speaking follow the three steps

(1) **Finding a representation.**
(2) **Quantifier switching.**
(3) **Finding a witness.**

The previous chapters were mainly concerned with the first step of finding a representation of the communication complexity measure in Euclidean space. We have seen various representations for several communication complexity measures, e.g. rank, approximate rank, and approximate norms. Different communication complexity measures give rise to different representations, reflecting their nature. For example, many representations for randomized complexity were simply "approximate" versions of deterministic complexity measures. These approximate versions, such as rank^α and γ_2^α, tend to be fairly robust in the variation of $\alpha > 1$, just as randomized complexity itself is robust in the allowable error parameter ϵ. On the other hand, $\text{rank}(A)$ and $\text{rank}^\alpha(A)$ for $\alpha > 1$ can be vastly different, as can $D(f)$ and $R_\epsilon(f)$ In this way, the difference between deterministic and randomized complexity is well expressed by their representations.

One might expect to similarly see a difference in the representations for randomized and quantum communication complexity. The surprising result of this section is that this is not the case. Natural representations for randomized communication complexity, like approximate rank and approximate norms such as μ^α turn out to also be representations for quantum communication complexity. Thus, at least when considering this approach, we cannot distinguish between classical and quantum complexities after the first step of the three-step approach.

We again stress that there are randomized lower bound techniques, such as the corruption bound and information theoretic techniques, which do not always give lower bounds on quantum communication complexity. These bounds, however, are more combinatorial and less geometrical. It seems a difficult task to find other representations which distinguish between randomized and quantum communication complexity. We refer the reader to [LS09c, LSS09] for further reading on the equivalence between the natural representations for randomized and quantum communication complexity in different models.

6

The role of duality in proving lower bounds

This chapter deals with the second step of our three-step approach, that is quantifier-switching. This is, of course, a very broad subject, and we only discuss one generic tool to address this problem, *duality*. We consider principles of duality and study their close relation to quantifier switching. Recall that at this point, step 2 of the three-step approach, we have some complexity measure $\mathcal{G}$ which is phrased in terms of a minimization, e.g. an approximate norm. Duality provides a means to obtain an equivalent formulation of $\mathcal{G}$ in terms of a maximization.

Finding a dual formulation (in terms of a maximization) is the subject of Section 6.2. The preliminary Section 6.1 details the exact principles of duality that we need, including basic definitions and the statement of the separation theorem.

Although our main focus is in communication complexity, we take the opportunity in this section to discuss duality and its application in a broader context.

6.1 Duality and the separation theorem

Duality has a large number of meanings in mathematics. It appears in many areas e.g. linear algebra, geometry, group theory, graph theory, logic and order theory. Even if we restrict ourselves to the area we are considering, i.e. geometry, we are still left with many different notions. Examples of such notions of duality are linear programming and semidefinite programming duality, the dual space (of linear functionals), norms duality and a list of duality transforms.

The notion of duality we have in mind is the linear algebraic duality between a vector space and the space of linear functionals operating on that space. This notion of duality is related to other notions such as a dual norm and the duality transform. These concepts are discussed in more detail in Section 6.1.1.

The following theorem (and similar theorems) is the main engine behind the use of duality for quantifier-switching.

Theorem 6.1 (Separation theorem). Let X_1 and X_2 be convex subsets of $\mathbb{R}^n$. If $X_1 \cap X_2 = \emptyset$ then there is a vector $y \neq 0$ and a scalar b such that

(1) $\langle x, y\rangle \leq b$ for all $x \in X_1$,
(2) $\langle x, y\rangle \geq b$ for all $x \in X_2$.

If X_1 and X_2 are closed and at least one of them is bounded, then the separation above can be made strict.

An equivalent way to state Theorem 6.1 is: Given two convex subsets X_1 and X_2 of $\mathbb{R}^n$, if X_1 and X_2 are disjoint then there exists an affine hyperplane H such that X_1 is contained in the half-space on one side of H and X_2 is contained in the complementing half-space (i.e., H separates X_1 from X_2).

Note that the assumption $X_1 \cap X_2 = \emptyset$ in Theorem 6.1 is about *nonexistence* (of a point in both X_1 and X_2), while the consequence is about *existence* (of a separating hyperplane). This way the separation theorem, when applicable, provides a way to transform a statement about nonexistence into a statement about existence. In particular The-

orem 6.1 provides a tool to switch a universal quantifier with an existential quantifier, which is what we need.

Illustrative examples of how Theorem 6.1 is used for switching quantifiers in the context of communication complexity are provided in Section 6.2. We prove there the equivalence between randomized communication complexity and distributional complexity, and also provide an equivalent formulation for approximate norms.

6.1.1 Dual space, dual norms and the duality transform

The aim of this section is to introduce the notions of a dual space, dual norm and the duality transform. We also study basic properties of these definitions and examine them from a complexity theoretic point of view.

We start with the dual space of linear functionals. Given a vector space V over a field $\mathbb{F}$, the set of all linear functionals form V to $\mathbb{F}$ is called the *dual space*. When the vector space is $\mathbb{R}^n$, the dual space is also isomorphic to $\mathbb{R}^n$. A natural isomorphism maps a vector $y \in \mathbb{R}^n$ to the linear functional $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = \langle y, x \rangle$.

Observe that the consequence in Theorem 6.1 is really a statement about the dual space, its subject is a separating *hyperplane* or equivalently a linear functional. Therefore duality is inherent in any application of the separation theorem.

Banach space theory studies vector spaces equipped with a norm. For example the space $\ell_2^n = (\mathbb{R}^n, \|\cdot\|_2)$ is the vector space $\mathbb{R}^n$ with the ℓ_2 (Frobenius) norm. Now the dual space of linear operators is also a Banach space (with a norm defined on it). The dual space of a Banach space $(V, \|\cdot\|)$ is the pair $(V^*, \|\cdot\|^*)$. Here V^* is the dual space of linear functionals and $\|\cdot\|^*$ is the *dual norm*, defined for every $f \in V^*$ by

$$\|f\|^* = \max_{x \in V : \|x\| \leq 1} f(x).$$

In words, the dual norm of f is the maximal value given by f to a vector with unit norm. In the special case $V = \mathbb{R}^n$ we are interested in, this takes the form

$$\|y\|^* = \max_{x \in \mathbb{R}^n : \|x\| \leq 1} \langle y, x \rangle.$$

for every $y \in \mathbb{R}^n$.

Another, more combinatorial way, to reach the definition of a dual norm is as follows: It starts with the (geometric) *duality transform*. To a nonzero point $a \in \mathbb{R}^n \backslash \{0\}$ the duality transform assigns the hyperplane $\{x \in \mathbb{R}^n : \langle a, x \rangle = 1\}$, and to a hyperplane H that does not pass trough the origin and can be uniquely written as $H = \{x \in \mathbb{R}^n : \langle a, x \rangle = 1\}$ it assigns the point a.

From the duality transform that assigns hyperplanes to points and vice versa, one can derive a transformation on sets of points. For a set $X \subset \mathbb{R}^n$ we define the set *dual* to X, denoted by X^*, as follows:

$$X^* = \{y \in \mathbb{R}^n : \langle x, y \rangle \leq 1 \text{ for all } x \in X\}.$$

In other words

$$X^* = \{y \in \mathbb{R}^n : \sup_{x \in X} \langle x, y \rangle \leq 1\}.$$

A simple, but important, property of dual sets is

Fact 6.1. For any set $X \subset \mathbb{R}^n$

(1) The set X^* is closed and convex and contains the origin.
(2) The set $(X^*)^*$ is the closure of $\text{conv}(X \cup \{0\})$.

The first part of Fact 6.1 is easy, and the second is proved using Theorem 6.1. It follows that for every closed and convex set $X \subset \mathbb{R}^n$ that contains the origin we have that $(X^*)^* = X$.

Convexity is a prerequisite for Theorem 6.1 to apply. Without the assumption that X_1 and X_2 are convex the statement of this theorem is false. This is one reason why working with norms is very convenient. We describe norms and duality of norms in a little more detail next.

6.1.1.1 Norms

A *norm* on $\mathbb{R}^n$ is a function $\Phi : \mathbb{R}^n \to \mathbb{R}$ such that

(1) $\Phi(x) > 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
(2) $\Phi(a \cdot x) = |a| \Phi(x)$ for every $x \in \mathbb{R}^n$ and $a \in \mathbb{R}$.

(3) $\Phi(x+y) \leq \Phi(x) + \Phi(y)$ for every $x, y \in \mathbb{R}^n$.

Let Φ be a norm on $\mathbb{R}^n$, the set $B_\Phi = \{x \in \mathbb{R}^n : \Phi(x) \leq 1\}$ is called the *unit ball* of Φ. It is not hard to verify that B_Φ is convex and symmetric around the origin. On the other hand, given a convex and 0-symmetric body $B \subset \mathbb{R}^n$, it naturally induces a norm Φ_B by the Minkowski functional. Namely, $\Phi_B(x)$ is equal to c^{-1}, where c is the minimal scalar such that $c \cdot x \in B$. Norms can therefore be identified with their unit balls.

Consider B_Φ^*, the dual set to B_Φ. We have that

$$B_\Phi^* = \{y \in \mathbb{R}^n : \sup_{x:\Phi(x)\leq 1} \langle x, y\rangle \leq 1\}.$$

Therefore, for every pair of vectors $x \in \mathbb{R}^n$ and $y \in B_\Phi^*$

$$\langle x, y\rangle \leq \Phi(x).$$

This inequality is sometimes called *weak duality*. The separation theorem implies that *strong duality* also holds, i.e. that for every $x \in \mathbb{R}^n$ there is a $y \in B_\Phi^*$ such that $\langle x, y\rangle = \Phi(x)$. Observe that $B_\Phi^* = B_{\Phi^*}$, i.e. the dual norm Φ^* is the norm induced by the dual set B_Φ^*.

Duality, combined with the separation theorem thus provides an equivalent formulation of any norm Φ in terms of a maximization:

$$\Phi(x) = \max_{y:\Phi^*(y)\leq 1} |\langle x, y\rangle|. \tag{6.1}$$

Let us now interpret this from a complexity theoretic point of view, which is our main objective. Assume we have a vector x for which we want to prove that $\Phi(x)$ is large. Any vector $y \in B_\Phi^*$ provides a proof of some lower bound on $\Phi(x)$, and strong duality implies that this proof system is optimal, i.e. there is always a proof of this type that gives the value of $\Phi(x)$.

6.2 Applying the separation theorem - finding a dual formulation

6.2.1 Distributional complexity

The next theorem is the equivalence between randomized communication complexity and distributional complexity, proved by Yao

[Yao83, KN97] using a minimax theorem. We apply Theorem 6.1 directly to prove this equivalence.

Theorem 6.2. For every sign matrix A, $R_\varepsilon(A)$ is equal to

$$\max_{P} \min_{\substack{B \\ P(B[i,j] \neq A[i,j]) \leq \varepsilon}} D(B),$$

where the maximum is over all probability distributions P on the entries of A, and B ranges over sign matrices.

Proof. Recall from Section 4.4 that for a sign matrix A, $R_\varepsilon(A)$ is the minimal number t such that there exists a matrix $\tilde{E} \in \mathcal{B}(D,t)$ satisfying $\|A - \tilde{E}\|_\infty \leq 2\varepsilon$.

First let us see that $R_\varepsilon(A)$ is larger than the expression in the theorem (i.e., distributional complexity of A). Say $R_\varepsilon(A) = t$ and assume

$$\|A - \sum_k p_k E_k\|_\infty \leq 2\varepsilon,$$

where $p_1, \ldots, p_m$ are a probability distribution and $E_1, \ldots, E_m$ are sign matrices with deterministic communication complexity at most t. For every matrix Y with $\|Y\|_1 = 1$ we have

$$\sum_{i,j} Y[i,j](A[i,j] - \sum_k p_k E_k[i,j]) \leq 2\varepsilon.$$

Rearranging, we get

$$\sum_k p_k \sum_{i,j} Y[i,j](A[i,j] - E_k[i,j]) \leq 2\varepsilon.$$

Therefore there is some $k \in [m]$ such that

$$\sum_{i,j} Y[i,j](A[i,j] - E_k[i,j]) \leq 2\varepsilon$$

Considering matrices Y of the form $Y = A \circ P$ for an arbitrary probability distribution P we can reinterpret the above inequality as $P(A[i,j] \neq E_k[i,j]) \leq \varepsilon$. Since we work with an arbitrary distribution P, and $D(E_k) \leq t$, we get that the distributional complexity of A is smaller than $R_\varepsilon(A)$.

For the opposite inequality, fix a number t and assume that $R_\varepsilon(A) > t$. Denote by $\mathcal{A}_\varepsilon$ the following set

$$\mathcal{A}_\varepsilon = \{\tilde{A} : \|A - \tilde{A}\|_\infty \leq 2\varepsilon\} = \{A + \Delta : \|\Delta\|_\infty \leq 2\varepsilon\}.$$

Both $\mathcal{B}(D,t)$ and $\mathcal{A}_\varepsilon$ are closed and bounded convex sets, and our assumption that $R_\varepsilon > t$ implies that they are disjoint. By Theorem 6.1 therefore, there is a matrix Y and a scalar b such that

(1) $\langle \tilde{E}, Y\rangle \leq b$ for all $\tilde{E} \in \mathcal{B}(D,t)$,
(2) $\langle \tilde{A}, Y\rangle > b$ for all $\tilde{A} \in \mathcal{A}_\varepsilon$.

It is not hard to check that the above two conditions are equivalent to:

(1) $\langle E, Y\rangle \leq b$ for all E with $D(E) \leq t$,
(2) $\langle A + \Delta, Y\rangle > b$ for all Δ satisfying $\|\Delta\|_\infty \leq 2\varepsilon$.

We can assume without loss of generality that $\|Y\|_1 = 1$. Let $S = \text{sign}(Y)$ be the sign matrix whose (i,j) entry is the sign of $Y[i,j]$, and denote $P = S \circ Y$. Observe that P is a probability distribution. We can rewrite our two conditions as

(1) $\langle E \circ S, P\rangle \leq b$ for all E with $D(E) \leq t$,
(2) $\langle (A + \Delta) \circ S, P\rangle > b$ for all Δ satisfying $\|\Delta\|_\infty \leq 2\varepsilon$.

By choosing $\Delta = -2\varepsilon S$ in the second condition we get that $\langle A \circ S, P\rangle > b + 2\varepsilon$, which together with the first condition implies that the probability with respect to P that A is different from E is larger than ε for every sign matrix E with $D(E) \leq t$. The distributional complexity of A is therefore also larger than t, concluding the proof. □

6.2.2 Approximate norms

We have defined approximate norms in Section 4.2, to serve as a representing measure for randomized communication complexity in different models. In this section we apply duality to find an equivalent formulation for approximate norms.

As it will require no additional effort in the proof, we consider the following slightly more general definition of an approximate norm. Let

Φ and ξ be two norms on $\mathbb{R}^n$ and let $\alpha \geq 1$ be a real number. The (α, Φ, ξ)-approximate norm is defined by

$$\Phi_\xi^\alpha(x) = \min_{y:\xi(y-\frac{1+\alpha}{2}x)\leq\frac{\alpha-1}{2}} \Phi(y)$$

for every $x \in \{\pm 1\}^n$. Using duality we get

Theorem 6.3. For any two norms Φ and ξ on $\mathbb{R}^n$, a real number $\alpha \geq 1$, and every $x \in \{\pm 1\}^n$

$$\Phi_\xi^\alpha(x) = \max_{\substack{w \\ \Phi^*(w)\leq 1}} \frac{(1+\alpha)}{2}\langle x, w\rangle + \frac{(1-\alpha)}{2}\xi^*(w)$$

Proof. Fix a sign vector $x \in \{\pm 1\}^n$, and let t be a real number such that $\Phi_\xi^\alpha(x) > t$. Consider the following two convex sets

$$Y = \left\{ y | \xi(y - \frac{1+\alpha}{2}x) \leq \frac{\alpha-1}{2} \right\}$$

and

$$Z = \{z | \Phi(z) \leq t\}.$$

Y and Z are closed, convex, and bounded subsets of $\mathbb{R}^n$. Our assumption that $\Phi_\xi^\alpha(x) > t$ implies that $Y \cap Z = \emptyset$. By the separation theorem therefore, there is a vector w and a scalar b such that

(1) $\langle w, y\rangle > b$ for all $y \in Y$,
(2) $\langle w, z\rangle \leq b$ for all $z \in Z$.

Note that $Z = t \cdot B_\Phi$ and hence, the second condition is equivalent to saying that $\Phi(w)^* \leq b/t$. As for the first condition, observe that one can rewrite Y as

$$Y = \left\{ \frac{1+\alpha}{2}x + \Delta | \xi(\Delta) \leq \frac{\alpha-1}{2} \right\}$$

Rewriting the first condition accordingly we get that

$$\frac{1+\alpha}{2}\langle w, x\rangle - \max_\Delta \langle w, \Delta\rangle > b, \tag{6.2}$$

where the maximum is over all Δ such that $\xi(\Delta) \leq (\alpha - 1)/2$. By the definition of a dual norm the maximum in Equation (6.2) is equal to $(\alpha - 1)\xi^*(w)/2$. We conclude then that there exists a vector w such that $\Phi^*(w) \leq b/t$ and

$$\frac{1+\alpha}{2}\langle w, x\rangle + \frac{1-\alpha}{2}\xi^*(w) > b.$$

Normalizing w so that $\Phi^*(w) \leq 1$ we get that the expression in the theorem is at least as large as $\Phi^\alpha_\xi(x)$. It is not hard to verify that the converse inequality is also true, which concludes the theorem. □

Comparing the dual expression for norms and approximate norms we see that for approximate norms the functional we are optimizing on is more cumbersome, involving $\xi^*(w)$ and not only the inner product we had before, and a weighting that depends on α. Nevertheless the set of witnesses we need to consider, i.e. B_{Φ^*}, remains the same. The problems of finding a good witness for some norm or any approximate version of it share much in common.

7

Choosing a witness

We have now seen several examples of approximate norms and how they can give lower bounds on communication complexity. We have also seen how approximate norms have an equivalent formulation in terms of a maximization problem. We are left with the third step of the three-step approach, the problem of finding a good witness. In this section we focus on this problem for approximate norms where $\xi = \ell_\infty$. Recall from Theorem 6.3 that the equivalent maximization formulation for an approximate norm in this case is

$$\Phi^\alpha(x) = \max_{w:\Phi^*(w)\leq 1} \frac{(1+\alpha)}{2}\langle x, w\rangle + \frac{(1-\alpha)}{2}\|w\|_1.$$

Thus, we need to find a vector w satisfying $\Phi^*(w) \leq 1$ for which the target function is large. This problem can be quite difficult to solve in general; we will restrict ourselves to two specific known cases: a general heuristic for choosing w as a weighted version of x, described in Section 7.1, and a family of structured vectors x for which we know how to optimally choose w, described in Section 7.2.

7.1 Nonnegative weighting

For a given sign vector $x \in \{\pm 1\}^N$ we seek a vector w that maximizes

$$\max_w \frac{(1+\alpha)\langle x, w\rangle + (1-\alpha)\|w\|_1}{2\Phi^*(w)}. \tag{7.1}$$

A first choice that comes to mind is to take $w = x$. Since $x \in \{\pm 1\}^N$ this gives

$$\Phi^\alpha(x) \geq \frac{N}{\Phi^*(x)}. \tag{7.2}$$

The quality of this lower bound depends on the value of $\Phi^*(x)$. Sometimes this simple choice provides a tight lower bound for $\Phi^\alpha(x)$, as illustrated by the next example.

Example 7.1. Consider real 2^n-by-2^n matrices with the trace norm $\|\cdot\|_{tr}$. The operator norm, or spectral norm, $\|\cdot\|$, is dual to the trace norm.

Recall from Theorem 2.7 that for a 2^n-by-2^n sign matrix A

$$2^{D(A)} \geq \frac{\|A\|_{tr}^2}{2^{2n}}.$$

As $D(A) \leq n+1$, this shows that $\|A\|_{tr} \leq 2^{(3n+1)/2}$. It follows that

$$\|A\|_{tr}^\alpha \leq \|A\|_{tr} \leq 2^{(3n+1)/2},$$

for every $2^n \times 2^n$ sign matrix A and every real number $\alpha \geq 1$.

Taking A to be the Hadamard matrix H_n, we see that this upper bound can be tight. Simply take the witness w to be H_n as well. As $H_n H_n^t = 2^n I$ we have $\|H_n\| = 2^{n/2}$, and so by Equation (7.2),

$$\|H_n\|_{tr}^\alpha \geq \frac{2^{2n}}{\|H_n\|} = 2^{3n/2}.$$

In communication complexity terms, by Corollary 4.2 this means that the inner product function on n bits has randomized communication complexity $n - O(\log n)$ bits even when the allowed error is $\varepsilon = \frac{1}{2} - \frac{1}{n^{O(1)}}$.

The above example shows that x itself can sometimes be a good witness. But this is not always so:

Example 7.2. Consider the settings of Example 7.1. Denote by J_n the $2^n \times 2^n$ sign matrix all of whose elements are 1, and H_n is again the Hadamard matrix of the same size. Take the matrix

$$A = \begin{pmatrix} H_{n-1} & J_{n-1} \\ J_{n-1} & J_{n-1} \end{pmatrix}$$

The approximate trace norm of a sign matrix is at least the approximate trace norm of any of its submatrices, thus by the previous example

$$\|A\|_{tr}^{\alpha} \geq 2^{3(n-1)/2}$$

On the other hand, the operator norm of A is larger than $\frac{2^n}{\sqrt{2}}$. Therefore, taking A itself as a witness only gives $\|A\|_{tr} \geq 2^{n+1/2}$, which leads to a trivial communication complexity bound.

A natural remedy to overcome the pitfalls of the previous example is to choose a witness of the form $w = p \circ x$ where p is a probability vector (i.e., a nonnegative vector whose entries sum up to 1). The numerator of Equation 7.1 is still easy to handle, and we get

$$\Phi^{\alpha}(x) \geq \max_{\substack{p:p\geq 0 \\ \|p\|_1=1}} \frac{1}{\Phi^*(p \circ x)}. \tag{7.3}$$

Note that taking $w = x$ is equivalent to taking p above to be the uniform probability distribution.

Taking a weighted version of x can improve the lower bounds we can get. For example, consider the sign matrix A from Example 7.2. By taking A weighted with the following probability distribution P we can get a nearly optimal bound.

$$P = \frac{4}{2^{2n}} \begin{pmatrix} J_{n-1} & 0 \\ 0 & 0 \end{pmatrix}$$

With $w = P \circ A$ as a witness we get a bound of $\|A\|_{tr}^{\alpha} \geq 2^{3(n-1)/2}$, which is close to optimal.

The heuristic of taking a witness of the form $p \circ x$ for a probability distribution p has a very nice characterization in terms of approximate norms. Using duality we see that it is actually equivalent to the ∞-approximate norm.

Theorem 7.1. Let Φ be a norm on $\mathbb{R}^N$, then for every $x \in \{-1,+1\}^N$

$$\Phi^\infty(x) = \max_{\substack{p:p\geq 0 \\ \|p\|_1=1}} \frac{1}{\Phi^*(p \circ x)}.$$

Proof. The proof follows similar lines to that of Theorem 6.3. Intuitively, the above expression for Φ^∞ is what we expect as this is what one gets by taking the limit as $\alpha \to \infty$ in the expression for Φ^α in Theorem 6.3. □

Consider the special case $\Phi = \mu$. Recall the definition of discrepancy from Chapter 4. It is not hard to check that a reformulation of this definition is

$$\mathrm{disc}(A) = \min_{\substack{P:P\geq 0 \\ \|P\|_1=1}} \mu^*(P \circ A).$$

Theorem 7.1 implies that the inverse of discrepancy is equal to μ^∞. Note that the approximate norm $\mu^\alpha(A)$ is a decreasing function of α, achieving its minimal value when $\alpha = \infty$. Thus, the lower bound on randomized communication complexity in terms of the α-approximate μ norm, given in Chapter 4, is a generalization of the discrepancy lower bound and has also been called the *generalized discrepancy method.*

Dependence of Φ^α on α We saw in Example 7.1 that the α-approximate trace norm $\|\cdot\|_{tr}^\alpha$ can remain constant as α varies, since $\|H_n\|_{tr}^\alpha = 2^{3n/2}$ for every $\alpha \geq 1$. On the other hand, there are also sign matrices A for which the value of $\mu^\alpha(A)$ decreases very rapidly. In particular $\mu^\infty(A)$ can be exponentially smaller than $\mu^\alpha(A)$ for $\alpha = 2$. For such matrices A, the discrepancy method fails to give good lower bounds on $R_\varepsilon(A)$, while the lower bound via the bounded α-approximate norm does much better. In particular, this means that for some functions one cannot show a good lower bound via Equation 7.1 simply by choosing a witness of the form $w = p \circ x$ for a probability distribution p.

A famous example where μ^α decreases rapidly is the disjointness function.

Theorem 7.2. Consider the $2^n \times 2^n$ sign matrix A_n that corresponds to the disjointness function. The disjointness function DISJ : $\{0,1\}^n \times \{0,1\}^n \to \{T,F\}$ is defined by $\mathrm{DISJ}(x,y) = T$ if $x \cap y = \emptyset$ and $\mathrm{DISJ}(x,y) = F$ otherwise. Then

$$\mu(A_n) \geq 2\left(\frac{\sqrt{5}}{2}\right)^n - 1$$
$$\mu^\infty(A_n) \leq 16n + 8.$$

Proof. Recall from Corollary 2.1 that $\gamma_2(M) \leq \mu(M) \leq 8\gamma_2(M)$ for every real matrix M. We show next that $\gamma_2^\infty(A_n) \leq 2n+1$.

Denote by L_n the $n \times 2^n$ Boolean matrix whose columns are the 2^n Boolean strings of length n. Then the matrix $2L_nL_n^t - J$ has the same sign pattern as A_n and its entries are all larger than 1. Therefore

$$\gamma_2^\infty(A_n) \leq \gamma_2(2L_nL_n^t - J) \leq 2n+1.$$

Now we show a lower bound on $\gamma_2(A_n)$. Recall that

$$\gamma_2(M) \geq \frac{\|M\|_{tr}}{\sqrt{\mathrm{size}(M)}}$$

for any sign matrix M. We actually lower bound the trace norm of A_n. It will be more convenient to work with the Boolean valued matrix $B_n = (J - A_n)/2$. This is the Boolean communication matrix for the disjointness function. The matrix B_n is nice to work with because of its tensor product structure; indeed, if we let

$$B_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

then $B_n = (B_1)^{\otimes n}$.

As $\|M \otimes M\|_{tr} = \|M\|_{tr}^2$, and a simple calculation shows $\|B_1\|_{tr} = \sqrt{5}$, we find

$$\|B_n\|_{tr} = 5^{n/2},$$

which gives the theorem. □

7.2 Block composed functions

Many functions of interest in communication complexity can be viewed as the composition of an inner function and an outer function. More precisely, the input x to Alice can be broken up into blocks $x = (x^1, \ldots, x^n)$ and similarly with the input to Bob $y = (y^1, \ldots, y^n)$, so that the communication function h can be expressed as

$$h(x, y) = (f \bullet g^n)(x, y) = f(g(x^1, y^1), \ldots, g(x^n, y^n)).$$

Here f and g can be arbitrary Boolean functions of the appropriate size. We call a function h of this form a *block composed function*, and refer to f as the *outer* function and g as the *inner* function.

A well studied class of functions arises when we take the inner function g to be the AND function on one bit. Then if f is the PARITY function, for example, we arrive at the INNER PRODUCT function, and if f is OR we get the SET INTERSECTION function.

Before we discuss lower bounds for block composed functions, let us first think about what we should expect their complexity to be. A fundamental idea going back to Buhrman, Cleve, and Wigderson [BCW98], is that the complexity of $f \bullet g^n$ can be related to the *query* complexity of f and the communication complexity of g. Let $RQ(f)$ indicate the randomized query complexity of f with error probability at most $1/3$. This is the number of queries of the form $x_i =?$ needed by a randomized algorithm to evaluate $f(x)$ with probability at least $2/3$ on the worst case input. Similarly let $QQ(f)$ be the number of queries needed by a quantum query algorithm to evaluate $f(x)$ with success probability at least $2/3$. For formal definitions and a survey of query complexity we recommend Buhrman and de Wolf [BW02].

Theorem 7.3 (Buhrman, Cleve, and Wigderson [BCW98]). For any two functions $f : \{-1, +1\}^n \to \{-1, +1\}$ and $g : X \times Y \to \{-1, +1\}$,

$$R(f \bullet g^n) = O(RQ(f)R(g) \log RQ(f))$$
$$Q(f \bullet g^n) = O(QQ(f)Q(g) \log n).$$

Proof. We just treat the randomized case here. Let $\mathcal{P}$ be a randomized query algorithm for f with success probability $2/3$ and making $RQ(f)$ many queries. When $\mathcal{P}$ queries the i^{th} bit we run a randomized communication protocol computing $g(x^i, y^i)$. By repeating this protocol $O(\log RQ(f))$ many times, we can reduce the error probability to be at most $(10RQ(f))^{-1}$. Then by a union bound, with probability at least $9/10$ all such queries will be made correctly, and the decision tree algorithm will answer correctly with probability at least $3/5$. By repeating a constant number of times, we can boost this up to $2/3$. □

Unlike randomized communication complexity, for randomized and quantum query complexity we know a polynomially tight characterization in terms of a natural mathematical quantity, the approximate polynomial degree.

Definition 7.1 (approximate degree). Let $f : \{-1,+1\}^n \to \{-1,+1\}$. The *(polynomial) degree* of f is the degree of the unique multilinear polynomial representing f. For $\alpha \geq 1$ we say that a function f' gives an α-approximate to f if $1 \leq f'(x)f(x) \leq \alpha$ for all $x \in \{-1,+1\}^n$. The *α-approximate degree* of f, denoted $\deg_\alpha(f)$, is the smallest degree of a function f' which gives an α-approximate to f.

Remark 7.1. In a different scenario, one can consider a Boolean valued function f and define the approximate degree as $\min\{\deg(f') : \|f - f'\|_\infty \leq \varepsilon\}$. Letting $f_\pm$ be the sign representation of f, one can see that this definition with error parameter $0 \leq \varepsilon < 1/2$ is equivalent to $\deg_{\alpha_\varepsilon}(f_\pm)$ where $\alpha_\varepsilon = \frac{1+2\varepsilon}{1-2\varepsilon}$.

A polynomial relationship between query complexity and approximate degree was first shown by Nisan and Szegedy [NS94] and later improved by Buhrman et al. [BBC+01].

Theorem 7.4 ([NS94, BBC+01]). Let $f : \{-1,+1\}^n \to \{-1,+1\}$. Then

$$RQ(f) \leq \deg_2(f)^6.$$

This bound holds even with $RQ(f)$ replaced by the deterministic query complexity of f. Using this result together with Theorem 7.3 gives the following corollary:

Corollary 7.1. For any two functions $f : \{-1,+1\}^n \to \{-1,+1\}$ and $g : X \times Y \to \{-1,+1\}$,

$$R(f \bullet g^n) = O(\deg_2(f)^6 R(g) \log \deg_2(f)),$$
$$Q(f \bullet g^n) = O(\deg_2(f)^6 Q(g) \log n).$$

Our goal, then, in showing lower bounds on the complexity of a block composed function $f \bullet g^n$ is to get something at least in the ballpark of this upper bound. Of course, this is not always possible—the protocol given by Theorem 7.3 is not always optimal. For example, when f is the PARITY function on n bits, and g is the two bit XOR function, this protocol just gives an upper bound of n bits, when the true complexity is constant.

For a broad class of functions, however, we do expect that a lower bound resembling the upper bound of Corollary 7.1 should hold. In a seminal paper, Razborov [Raz03] showed a theorem of this type. He showed that whenever f is a symmetric function, and g is the AND function the complexity of $f \bullet g^n$ is $\Omega(\deg_2(f))$ by using the approximate trace norm method. This lower bound is polynomially related to the upper bound in Corollary 7.1.

More recently, very nice frameworks have been developed by Sherstov [She09, She08c] and in independent work by Shi and Zhu [SZ09b] which can show a lower bound on a block composed function $f \bullet g^n$ in terms of the approximate degree of f for an arbitrary function f, provided that g satisfies certain technical conditions. Both of these papers work again with the approximate trace norm, more precisely with the dual formulation of the approximate trace norm as in Equation 7.1. Using this dual form, the problem of showing lower bounds on communication complexity reduces to finding a good witness. A key idea of Sherstov and Shi–Zhu is to define a witness for the approximate trace norm bound in terms of the *dual polynomial*, described next.

A degree d polynomial which approximates a function f provides a certificate that the approximate degree is at most some value. Similarly, a dual polynomial for f provides a certificate that the approximate polynomial degree of f is *at least* a certain value. More precisely, the dual polynomial has the following properties.

Lemma 7.1 (Sherstov [She08c], Shi-Zhu [SZ09b]). [1] Let $f : \{-1,+1\}^n \to \{-1,+1\}$ and let $d = \deg_\alpha(f)$ for $\alpha \geq 1$. Then there exists a function $v : \{-1,+1\}^n \to \mathbb{R}$ such that

(1) $\langle v, f\rangle \geq \frac{\alpha-1}{\alpha+1}$.
(2) $\|v\|_1 = 1$.
(3) $\langle v, g\rangle = 0$ for any function g of degree $\leq d$.

Furthermore, when $\alpha = \infty$, there is a function $v : \{-1,+1\}^n \to \mathbb{R}$ satisfying items (2), (3), and such that $v(x)f(x) \geq 0$ for all $x \in \{-1,+1\}^n$.

A function v as above is called a *dual polynomial* for f. The properties of the dual polynomial mesh very well with the properties required by a good witness in the dual formulation of approximate trace norm. Namely, if v is a dual polynomial for f, to show that the approximate trace norm of $f \bullet g^n$ is large we choose as a witness the function $v \bullet g^n$, up to a normalization factor. Item (1) is then used to show that $\langle f \bullet g^n, v \bullet g^n\rangle$ is large; item (2) is used to bound $\|v \bullet g^n\|_1$; and, in the most difficult step, item (3) is used to upper bound $\|v \bullet g^n\|$.

The function g serves as a mediator in the transference of properties of the dual polynomial to the trace norm witness. All that is required of g for the transference of properties (1), (2) is that it is *balanced*—it outputs the value $+1$ as often as -1. When g is balanced, the inner product $\langle v, f\rangle$ and the ℓ_1 norm $\|v\|_1$ are proportional to $\langle v \bullet g^n, f \bullet g^n\rangle$ and $\|v \bullet g^n\|_1$, respectively.

Item (3), however, is more difficult, and this is where the papers of Sherstov [She08c] and Shi and Zhu [SZ09b] diverge. Sherstov considers the case where g is a fixed function of a particularly nice form, which

[1] This is Lemma 3.3.1 of [She08c] and can be found in Section 3.1 of [SZ09b].

leads $f \bullet g^n$ to be what he calls a *pattern matrix*. The structure of a pattern matrix allows one to compute the spectral norm of $v \bullet g^n$ precisely. The pattern matrix framework has proven extremely useful and has found many further applications in unbounded-error communication complexity [She08d, RS08] and multiparty communication complexity [Cha07, LS09a, CA08, BHN08] which will be described in Chapter 8.

Shi and Zhu take a different approach, and are able to get a bound on $f \bullet g^n$ in terms of the approximate degree of f whenever g is sufficiently "hard." This approach simply uses the triangle inequality to upper bound the spectral norm of $v \bullet g^n$. We refer the reader to the survey of Sherstov [She08a] for a detailed comparison of these two methods.

We present both the approaches of Sherstov and Shi-Zhu in the next two sections.

7.2.1 Strongly balanced inner function

We begin by describing the pattern matrix method of Sherstov. A pattern matrix can be seen as a block composed function where the inner function g is of the form $g : \{-1,+1\}^k \times ([k] \times \{-1,+1\})$. On input $(x,(i,b))$ define $g(x,(i,b)) = x_i \cdot b$. In other words, the inner function simply selects a bit of x or its negation. If we let A be a 2^k-by-k matrix where $A[x,i] = x_i$ we can see that A_g is of the form

$$A_g = \begin{bmatrix} A & -A \end{bmatrix}. \tag{7.4}$$

For a function $f : \{-1,+1\}^n \to \{-1,+1\}$, the pattern matrix corresponding to f is a block composed function $f(g(x^1,y^1),\ldots,g(x^n,y^n))$. Notice that the role of g here is to select a bit from each block of $x = (x^1,\ldots,x^n)$, possibly negating some bits, and then applying f to the resulting n-bit substring. Sherstov shows the following theorem.

Theorem 7.5 (Sherstov, Theorem 5.1 [She08c]). Let f be an arbitrary function, and let $g : \{-1,+1\}^k \times ([k] \times \{-1,+1\})$ be as described above. Then

$$Q^*_{1/4}(f \bullet g^n) \geq \frac{1}{2} \deg_3(f) \log(k) - 2.$$

A key property of pattern matrices which allows for a precise calculation of the spectral norm of $v \bullet g^n$ is that the inner function is *strongly balanced*, as defined next.

Definition 7.2 (Strongly balanced). Let A be a matrix, and J be the all-ones matrix of the same dimensions as A. We say that A is *balanced* if $\mathrm{Tr}(AJ^t) = 0$. We say that A is *strongly balanced* if $AJ^t = JA^t = 0$. That is, a matrix is strongly balanced if the sum over each row is zero, and similarly for each column. We say that a function is balanced or strongly balanced if its sign matrix representation is.

Notice in Equation 7.4 that columns of A correspond to the characters of degree one, and hence are balanced. The purpose of b in the construction of a pattern matrix is to also make the rows of A_g also balanced, and thus A_g strongly balanced overall.

Lee, Shraibman, and Zhang [LSZ09] observe that Sherstov's proof of Theorem 7.5 works essentially unchanged whenever the inner function g is strongly balanced.

Theorem 7.6 (Lee-Shraibman-Zhang [LSZ09]). Let $f : \{-1,+1\}^n \to \{-1,+1\}$ be an arbitrary function, and let g be a strongly balanced function. Fix $0 < \varepsilon < 1/2$, let $\alpha = 1/(1-2\varepsilon)$, and $\alpha_0 > \alpha$. Then

$$Q^*_\varepsilon(f \bullet g^n) \geq \deg_{\alpha_0}(f) \log\left(\frac{\sqrt{\mathrm{size}(A_g)}}{\|A_g\|}\right) + \log\left(\frac{\alpha_0 - \alpha}{\alpha(\alpha_0+1)}\right).$$

Proof. Let $d = \deg_{\alpha_0}(f)$ and let v be a dual polynomial for f with properties as in Lemma 7.1. By Theorem 5.3 and the definition of γ_2 given in 2.3.2, the approximate trace norm of $A_{f \bullet g^n}$ gives a lower bound on $Q^*_\varepsilon(f \bullet g^n)$. We therefore use the dual expression of the approximate trace norm of $A_{f \bullet g^n}$ to prove a lower bound. For this purpose we define a witness matrix B by

$$B[x,y] = \frac{2^n}{\mathrm{size}(A_g)^n} v(g(x^1,y^1),\ldots,g(x^n,y^n)).$$

Let us first lower bound the inner product $\langle A_{f \bullet g^n}, B \rangle$. Notice that as g is strongly balanced, it is in particular balanced, and so

$$\begin{aligned}
\langle A_{f \bullet g^n}, B \rangle &= \frac{2^n}{\mathrm{size}(A_g)^n} \sum_{x,y} (f \bullet g^n)(x,y) \cdot (v \bullet g^n)(x,y) \\
&= \frac{2^n}{\mathrm{size}(A_g)^n} \sum_{z \in \{-1,+1\}^n} f(z) v(z) \cdot |\{x,y : g^n(x,y) = z\}| \\
&= \langle f, v \rangle \\
&\geq \frac{\alpha_0 - 1}{\alpha_0 + 1}.
\end{aligned}$$

The third equality is because $|\{x,y : g^n(x,y) = z\}| = \mathrm{size}(A_g)/2$ for balanced functions g, and the last inequality is by Lemma 7.1. A similar argument shows that $\|B\|_1 = 1$ as $\|v\|_1 = 1$.

Now we turn to evaluate $\|B\|$. To do this, we expand B in terms of the Fourier coefficients of v. We write

$$\begin{aligned}
B[x,y] &= \frac{2^n}{\mathrm{size}(A_g)^n} \sum_{T \subseteq [n]} \hat{v}_T \, \chi_T(g(x^1,y^1), \ldots, g(x^n,y^n)) \\
&= \frac{2^n}{\mathrm{size}(A_g)^n} \sum_{T \subseteq [n]} \hat{v}_T \prod_{i \in [n]} g(x^i,y^i)^{T[i]}
\end{aligned}$$

where $T[i] = 1$ if $i \in T$ and $T[i] = 0$ otherwise. We can write this more compactly in matrix notation using tensor product

$$\begin{aligned}
B &= \frac{2^n}{\mathrm{size}(A_g)^n} \sum_{T \subseteq [n]} \hat{v}_T A_{\chi_T \bullet g^n} \\
&= \frac{2^n}{\mathrm{size}(A_g)^n} \sum_{T \subseteq [n]} \hat{v}_T \bigotimes_i A_g^{T[i]}
\end{aligned}$$

where $A_g^1 = A_g$ and $A_g^0 = J$ the all-ones matrix of appropriate size.

Now observe that the product $A_{\chi_T \bullet g^n} A_{\chi_S \bullet g^n}$ is equal to zero when-

ever $S \neq T$. Indeed,

$$\begin{aligned} A_{\chi_T \bullet g^n} A^t_{\chi_S \bullet g^n} &= \left(\bigotimes_i A_g^{T[i]} \right) \left(\bigotimes_i A_g^{S[i]} \right)^t \\ &= \bigotimes_i \left(A_g^{T[i]} (A_g^{S[i]})^t \right) \\ &= 0. \end{aligned}$$

When $T \neq S$, there is some i for which $T[i] \neq S[i]$. The corresponding term is thus either $A_g J^t$ or $J A_g^t$. Either way the product nullifies as g is strongly balanced.

We can therefore use the following fact to analyze $\|B\|$.

Fact 7.1 (Lemma 4.2 [She08c]). If $AB^t = BA^t = 0$ then $\|A+B\| = \max\{\|A\|, \|B\|\}$.

This is the key fact which makes Sherstov's analysis of the spectral norm of pattern matrices so clean. As we shall see in the next section, Shi-Zhu do not assume that the inner function is strongly balanced, and thus simply use $\|A + B\| \leq \|A\| + \|B\|$ the triangle inequality at this stage.

We get

$$\begin{aligned} \|B\| &= \frac{2^n}{\mathrm{size}(A_g)^n} \| \sum_{T \subseteq [n]} \hat{v}_T A_{\chi_T \bullet g^n} \| \\ &= \frac{2^n}{\mathrm{size}(A_g)^n} \max_T |\hat{v}_T| \| A_{\chi_T \bullet g^n} \| \\ &= \max_T \ 2^n |\hat{v}_T| \prod_i \frac{\|A_g^{T[i]}\|}{\mathrm{size}(A_g)} \\ &\leq \max_{T : \hat{v}_T \neq 0} \prod_i \frac{\|A_g^{T[i]}\|}{\mathrm{size}(A_g)} \\ &\leq \left(\frac{\|A_g\|}{\sqrt{\mathrm{size}(A_g)}} \right)^d \left(\frac{1}{\mathrm{size}(A_g)} \right)^{n/2} \end{aligned}$$

In the second to last step we have used that $|\hat{v}_T| \leq 1/2^n$ as $\|v\|_1 = 1$, and in the last step we have used the fact that $\|J\| = \sqrt{\text{size}(A_g)}$.
We conclude that

$$\frac{\|A_{f \bullet g^n}\|_{tr}^{\alpha}}{\sqrt{\text{size}(A_{f \bullet g^n})}} \geq \frac{\alpha_0 - \alpha}{(\alpha_0 + 1)} \left(\frac{\sqrt{\text{size}(A_g)}}{\|A_g\|} \right)^d.$$

□

While it may look strange at first, the expression for the complexity of g in Theorem 7.6 is closely related to the discrepancy of g with respect to the uniform distribution. Let A be a m-by-n sign matrix. The discrepancy of A with respect to the uniform distribution, $\text{disc}_U(A)$, can be written as

$$\text{disc}_U(A) = \frac{1}{\text{size}(A)} \max_{\substack{x \in \{0,1\}^m \\ y \in \{0,1\}^n}} |x^t A y|.$$

It is easy to see from this expression that

$$\text{disc}_U(A) \leq \frac{\|A\|}{\sqrt{\text{size}(A)}}.$$

Shaltiel [Sha03] has shown the deeper result that this bound is in fact polynomially tight:

Theorem 7.7 (Shaltiel). Let A be a sign matrix.

$$\text{disc}_U(A) = \Omega\left(\frac{\|A\|}{\sqrt{\text{size}(A)}} \right)^3.$$

Using this characterization, we get the following corollary:

Corollary 7.2. In the settings of Theorem 7.6.

$$Q_{\varepsilon}^*(f \bullet g^n) = \deg_{\alpha_0}(f) \left(\frac{1}{3} \log\left(\frac{1}{\text{disc}_U(A_g)} \right) - O(1) \right) + \log\left(\frac{\alpha_0 - \alpha}{\alpha(\alpha_0 + 1)} \right).$$

7.2.2 Triangle Inequality

The method of Shi-Zhu does not restrict the form of the inner function g, but rather works for any g which is sufficiently "hard." The hardness condition they require is phrased in terms of a somewhat awkward measure they term spectral discrepancy.

Definition 7.3 (spectral discrepancy). Let A be a m-by-n sign matrix. The spectral discrepancy of A, denoted $\rho(A)$, is the smallest r such that there is a submatrix A' of A and a probability distribution μ on the entries of A' satisfying:

(1) A' is balanced with respect to μ, i.e. the distribution which gives equal weight to -1 entries and $+1$ entries of A'.
(2) The spectral norm of $A' \circ \mu$ is small:

$$\|A' \circ \mu\| \leq \frac{r}{\sqrt{\text{size}(A')}}$$

(3) The entrywise absolute value of the matrix $A' \circ \mu$ should also have a bound on its spectral norm in terms of r:

$$\||A' \circ \mu|\| \leq \frac{1+r}{\sqrt{\text{size}(A')}}$$

While conditions (1),(2) in the definition of spectral discrepancy are quite natural, condition (3) can be complicated to verify. Note that condition (3) will always be satisfied when μ is taken to be the uniform distribution. Using this notion of spectral discrepancy, Shi-Zhu show the following theorem.

Theorem 7.8 (Shi-Zhu [SZ09b]). Let $f : \{-1,+1\}^n \to \{-1,+1\}$, and $g : \{-1,+1\}^{m_1} \times \{-1,+1\}^{m_2} \to \{-1,+1\}$. If $\rho(g) \leq \frac{\deg_3(f)}{2en}$ then

$$Q_{1/4}(f \bullet g^n) \geq \Omega(\deg_3(f)).$$

Here $e = 2.718\ldots$ is Euler's number.

Chattopadhyay [Cha08] extended the technique of Shi-Zhu to the case of multiparty communication complexity, answering an open question of Sherstov [She08a]. In doing so, he gave a more natural condition on the hardness of g in terms of an upper bound on discrepancy we shall encounter in Chapter 8, Theorem 8.5. Following this proof, Lee-Shraibman-Zhang further relaxed the requirement on the inner function g to simply having small discrepancy under a balancing distribution.

Theorem 7.9 (Lee-Shraibman-Zhang [LSZ09]). Let $f : \{-1,+1\}^n \to \{-1,+1\}$, and $g : \{-1,+1\}^{m_1} \times \{-1,+1\}^{m_2} \to \{-1,+1\}$. Fix $0 < \varepsilon < 1/2$, let $\alpha = 1/(1-2\varepsilon)$, and $\alpha_0 > \alpha$. Then

$$Q^*_\varepsilon(f \bullet g^n) \geq \deg_{\alpha_0}(f) + \log\left(\frac{\alpha_0 - \alpha}{\alpha(\alpha_0+1)}\right).$$

provided there is a distribution μ which is balanced with respect to g and for which $\gamma_2^*(g \circ \mu) \leq \frac{\deg_{\alpha_0}(f)}{2en}$.

Proof. As before we use Theorem 5.3, and thus we really lower bound $\gamma_2^\alpha(A_{f \bullet g^n})$. By the dual expression for γ_2^α (Theorem 6.3) we have

$$\gamma_2^\alpha(A_{f\bullet g^n}) = \max_B \frac{(1+\alpha)\langle A_{f\bullet g^n}, B\rangle + (1-\alpha)\|B\|_1}{2\gamma_2^*(B)}.$$

To prove a lower bound we choose a witness matrix B as follows

$$B[x,y] = 2^n \cdot v(g(x^1,y^1),\ldots,g(x^n,y^n)) \cdot \prod_{i=1}^n \mu(x^i,y^i).$$

where v witnesses that f has approximate degree at least $d = \deg_{\alpha_0}(f)$. This definition is the same as in the previous section where μ was simply the uniform distribution. As argued before, we have $\langle A_{f\bullet g^n}, B\rangle \geq \frac{\alpha_0-1}{\alpha_0+1}$ and $\|B\|_1 = 1$ because $A_g \circ \mu$ is balanced.

We again expand B as

$$B = 2^n \sum_{T:|T|\geq d} \hat{v}_T \bigotimes_{i=1}^n (A_g \circ \mu)^{T(i)},$$

where $(A_g \circ \mu)^1 = A_g \circ \mu$ and $(A_g \circ \mu)^0 = \mu$.

Now comes the difference with the previous proof. As we do not have special knowledge of the function g, we simply bound $\gamma_2^*(B)$ using the triangle inequality.

$$\begin{aligned}\gamma_2^*(B) &\leq 2^n \sum_{T:|T|\geq d} |\hat{v}_T|\ \gamma_2^*\left(\bigotimes_{i=1}^{n}(A_g \circ \mu)^{T(i)}\right) \\ &= 2^n \sum_{T:|T|\geq d} |\hat{v}_T|\ \gamma_2^*(A_g \circ \mu)^{|T|}\gamma_2^*(\mu)^{n-|T|} \\ &\leq \sum_{T:|T|\geq d} \gamma_2^*(A_g \circ \mu)^{|T|},\end{aligned}$$

where in the last step we have used that $\gamma_2^*(\mu) \leq 1$ as μ is a probability distribution and that $|\hat{v}_T| \leq 2^{-n}$. In the second step (equality) we used the fact that γ_2^* is multiplicative with respect to tensor product, a property proved in [LSŠ08]. We continue with simple arithmetic:

$$\begin{aligned}\gamma_2^*(B) &\leq \sum_{i=d}^{n} \binom{n}{i} \gamma_2^*(A_g \circ \mu)^i \\ &\leq \sum_{i=d}^{n} \left(\frac{en\gamma_2^*(A_g \circ \mu)}{d}\right)^i \\ &\leq 2^{-d}\end{aligned}$$

provided that $\gamma_2^*(g \circ \mu) \leq \frac{d}{2en}$. □

An interesting open question, raised by Sherstov [She08a], is if there are cases where Theorem 7.9 can be applied where Theorem 7.6 cannot. While the condition that the inner function g be strongly balanced may seem rather restrictive, as communication complexity is nonincreasing under function restriction, this theorem can also be applied with respect to a strongly balanced submatrix of g. In this light, the question informally becomes: does every function g for which there is a distribution μ such that $g \circ \mu$ is balanced and $\gamma_2^*(g \circ \mu)$ is small contain a large strongly balanced submatrix?

7.2.3 Examples

In this section we give some examples of the application of Theorem 7.6.

Disjointness Disjointness can be written in the following way.

$$\mathrm{DISJ}(x,y) = \bigvee_{i=1}^{n} (x_i \wedge y_i).$$

In other words, in this case f is the n-bit OR function, and g is the AND function on one bit. Notice that the sign matrix of the AND function on one bit looks as follows.

$$H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

This matrix is not strongly balanced. Sherstov gets around this problem in the following way. We can alternatively think of the disjointness function on n bits (assuming that four divides n) as

$$\mathrm{DISJ}(x,y) = \bigvee_{i=1}^{n/4} \left(\vee_{j=0}^{3} (x_{4i+j} \wedge y_{4i+j})\right)$$

Here we take f to be the OR function on $n/4$ bits and g to be the OR-AND function on 4 bits.

Now g does contain a strongly balanced submatrix. Namely,

	0001	0010	1000	0100
0011	-1	-1	1	1
0101	-1	1	1	-1
1100	1	1	-1	-1
1010	1	-1	-1	1

One can check that the spectral norm of this matrix is $2\sqrt{2}$. Indeed, we can write the spectral norm of this matrix as

$$\left\| \begin{bmatrix} -H & H \\ H & -H \end{bmatrix} \right\| \leq \left\| \begin{bmatrix} -H & 0 \\ 0 & -H \end{bmatrix} \right\| + \left\| \begin{bmatrix} 0 & H \\ H & 0 \end{bmatrix} \right\| = 2\sqrt{2}.$$

This gives us

$$\frac{\sqrt{\mathrm{size}(A_{g'})}}{\|A_{g'}\|} = \sqrt{2}$$

Nisan and Szegedy [NS94] show that the bounded-error degree of the OR function on n bits is $\Omega(\sqrt{n})$, and so we get a $\Omega(\sqrt{n})$ lower bound for the disjointness function.

When g is inner product When g is the inner product function on k-bits, the sign matrix A_g is a 2^k-by-2^k Hadamard (Sylvester) matrix H_k. This matrix can be defined recursively by $H_0 = [1]$ and

$$H_k = \begin{bmatrix} H_{k-1} & H_{k-1} \\ H_{k-1} & -H_{k-1} \end{bmatrix}$$

From this definition we can see that H_k contains a strongly balanced submatrix of size 2^{k-1}-by-2^{k-1} for all $k > 1$. For $k > 1$ we can write H_k as

$$H_k = \begin{bmatrix} H_{k-2} & H_{k-2} & H_{k-2} & H_{k-2} \\ H_{k-2} & -H_{k-2} & H_{k-2} & -H_{k-2} \\ H_{k-2} & H_{k-2} & -H_{k-2} & -H_{k-2} \\ H_{k-2} & -H_{k-2} & -H_{k-2} & H_{k-2} \end{bmatrix}$$

The submatrix sitting in the middle

$$\begin{bmatrix} -H_{k-2} & H_{k-2} \\ H_{k-2} & -H_{k-2} \end{bmatrix}$$

is clearly strongly balanced and is of dimension 2^{k-1}-by-2^{k-1}. This matrix has spectral norm $2\sqrt{2^{k-1}}$ and so we obtain

$$Q^*_{1/4}(f \bullet g^n) \geq \frac{(k-1)}{2} \deg_3(f) - O(1)$$

for any function f, when g is the inner product function on k bits.

7.2.4 XOR functions

As we have hopefully demonstrated, the dual polynomial of f can be a very powerful technique for choosing a witness when working with a block composed function. In this section, however, we will consider a case when the lower bound in terms of approximate degree, Theorem 7.6, does not show a good lower bound.

Say that the inner function $g = \oplus(x, y)$ is the XOR function on one bit. This is a strongly balanced function, so we can apply Theorem 7.6. In this case, however, the theorem gives nothing as A_g is a rank-one matrix and $\|A_g\| = \sqrt{\text{size}(A_g)}$. Indeed, we should expect this as when f is the PARITY function, the complexity of $f \bullet g^n$ is constant.

When g is the XOR function, it turns out that the complexity of $f \bullet g^n$ is not related to the approximate degree of f, but to the approximate ℓ_1 norm of the Fourier coefficients of f. Let us define this next.

Definition 7.4. Let $f : \{-1,+1\}^n \to \{-1,+1\}$ be a Boolean function. Denote by $\|\hat{f}\|_1$ the ℓ_1 norm of the Fourier coefficients of f. In the usual way we also define the approximate version of this norm: for $\alpha \geq 1$, let

$$\|\hat{f}\|_1^\alpha = \min_{g: 1 \leq g(x)f(x) \leq \alpha} \|\hat{g}\|_1$$

Theorem 7.10. Let $f : \{-1,+1\}^n \to \{-1,+1\}$ be an arbitrary function, and $g(x,y) = \oplus(x,y)$ be the XOR function on one bit. Let A be the sign matrix representation of $f \bullet g^n$. Then

$$\frac{\|A\|_{tr}^\alpha}{\sqrt{\text{size}(A)}} = \|\hat{f}\|_1^\alpha.$$

Proof. We show the lower bound first. For this, we will use the general formulation of the dual of an approximate norm from Theorem 6.3. The dual norm of $\|\hat{f}\|_1$ is given by

$$\begin{aligned}\|\hat{f}\|_1^* &= \max_{u: \|\hat{u}\|_1 = 1} \langle f, u \rangle \\ &= \max_{u: \|\hat{u}\|_1 = 1} \sum_T \hat{f}_T \langle \chi_T, u \rangle = 2^n \max_T |\hat{f}_T|.\end{aligned}$$

Thus we see that $\|\hat{f}\|_1^* = 2^n \|\hat{f}\|_\infty$.

By the general formulation of the dual of an approximate norm, this means that there is a function v such that

(1) $\frac{(1+\alpha)\langle v, f\rangle + (1-\alpha)\|v\|_1}{2} = \|\hat{f}\|_1^\alpha$.
(2) $2^n \|\hat{v}\|_\infty \leq 1$,

We now define a witness to show that the approximate trace norm of A is large as

$$B[x,y] = \frac{2^n}{\text{size}(A_g)^n} v(g(x^1,y^1), \ldots, g(x^n,y^n)).$$

As argued in the proof of Theorem 7.6, item (1) can be used to show that

$$\frac{(1+\alpha)\langle A, B\rangle + (1-\alpha)\|B\|_1}{2} = \|\hat{f}\|_1^{\alpha}.$$

Now we use item (2) to show $\|B\| \leq 1$. We will again use the fact that g is strongly balanced and so the matrices $A_{\chi_T \bullet g^n}$ corresponding to distinct T are orthogonal.

$$\begin{aligned}
\|B\| &= \frac{2^n}{\text{size}(A_g)^n} \| \sum_{T \subseteq [n]} \hat{v}_T \bigotimes A_g^{T[i]} \| \\
&= \frac{2^n}{\text{size}(A_g)^n} \max_{T \subseteq n} |\hat{v}_T| \prod_i \|A_g^{T[i]}\| \\
&= \max_{T \subseteq n} |\hat{v}_T|.
\end{aligned}$$

The last step follows since when g is the XOR function on one bit we have $\|A_g\| = \|J\| = \sqrt{\text{size}(A_g)} = \text{size}(A_g)/2$. This gives the desired lower bound as by property (2)

$$\max_{T \subseteq n} |\hat{v}_T| \leq \frac{1}{2^n} = \frac{1}{\sqrt{\text{size}(A)}}.$$

Now we turn to the upper bound. Let w be a function such that $1 \leq w(x)f(x) \leq \alpha$ and $\|\hat{w}\|_1 = \|\hat{f}\|_1^{\alpha}$. Define

$$B[x, y] = w(g(x^1, y^1), \ldots, g(x^n, y^n)).$$

Then clearly $1 \leq A[x, y]B[x, y] \leq \alpha$. To bound the trace norm of B we will use the fact that if $XY^t = X^tY = 0$ then $\|X + Y\|_{tr} = \|X\|_{tr} + \|Y\|_{tr}$. This gives

$$\begin{aligned}
\|B\|_{tr} &= \| \sum_{T \subseteq [n]} \hat{w}_T \bigotimes A_g^{T[i]} \| \\
&= \sum_{T \subseteq n} \hat{w}_T \prod_i \|A_g^{T[i]}\| \\
&= \sqrt{\text{size}(A)} \|\hat{f}\|_1^{\alpha}.
\end{aligned}$$

□

Remark 7.2. This theorem can be alternatively proven as follows. Let $f : \{-1,+1\}^n \to \{-1,+1\}$ and A be the sign matrix where $A[x,y] = f(x \oplus y)$. It is easy to see that

$$\frac{\|A\|_{tr}}{\sqrt{\text{size}(A)}} = \ell_1(\hat{f})$$

as the (unnormalized) eigenvectors of A are given by the characters $\chi_S(x) = \prod_{i \in S} x_i$ for $S \subseteq \{0,1\}^n$.

By a symmetrization argument one can then show that, without loss of generality, the matrix B of minimal trace norm which gives an α-approximation to A is of the form $B[x,y] = g(x \oplus y)$ for a function g which gives an α-approximation to f.

Shi and Zhang [SZ09a] are able to completely determine the randomized and quantum complexities of XOR functions in the case of symmetric f. For the lower bound they do not use the above approach, but rather a reduction to the case of $f(x \wedge y)$. We record their result here.

Theorem 7.11 (Shi and Zhang [SZ09a]). Let $S : [n] \to \{-1,+1\}$ and define a sign matrix $A[x,y] = S(|x \oplus y|)$. Let $r_0, r_1 \leq n/2$ be the minimum integers such that $S(k) = S(k+2)$ for all $k \in [r_0, n - r_1)$. Then

$$R_{1/4}(A) = O(r_1 + r_2)$$
$$Q_{1/4}(A) = \Omega(r_1 + r_2).$$

8

Multiparty communication complexity

In this chapter we look at communication complexity where the input is distributed over k-many players who wish to evaluate a function $f(x_1, \ldots, x_k)$. There are two common ways of distributing the inputs among the players, leading to the number-in-hand and number-on-the-forehead models. In the case of two players, both of these models specialize to the usual model of two-party complexity. In the number-in-hand model, player i receives input x_i. In the number-on-the-forehead model, player i receives $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$, i.e. the entire input except for x_i. Here one can picture the input x_i sitting on the forehead of player i.

Deterministic and randomized communication complexity in the multiparty model are defined similarly as in the two player model. We will only deal with the public coin model of multiparty complexity, thus a randomized multiparty protocol is simply a probability distribution over deterministic protocols. It is common in multiparty communication complexity to take a generous model of communication and assume that the players write their messages on a blackboard that is seen by everyone. We denote k-party deterministic communication complexity by D^k. Randomized communication complexity with error bound ε is

denoted by R^k_ε. The superscripts are omitted when k is clear from the context. If the subscript is omitted it is assumed that $\varepsilon = 1/3$.

The study of both the number-in-hand and number-on-the-forehead models have interesting applications. As shown by a seminal paper of Alon, Matias, and Szegedy [AMS99], lower bounds on disjointness in the number-in-hand model have applications to showing lower bounds on the memory requirements of streaming algorithms. In the number-on-the-forehead model as well, Beame, Pitassi, and Segerlind [BPS06] have shown that lower bounds on disjointness have interesting applications to lower bounds on proof complexity. Finally, and perhaps the greatest motivation to the study of the number-on-the-forehead model, is that lower bounds here imply circuit complexity lower bounds. In particular, showing an explicit function which requires super-polylogarithmic complexity in the number-on-the-forehead model with super-polylogarithmic many players would give an explicit function outside of the circuit complexity class ACC^0, currently a major open problem. Currently the best lower bounds in the number-on-the-forehead for explicit functions are of the form $n/2^k$ for k-players. The applications of multiparty complexity are discussed in more detail in Chapter 9.

To give a flavor of the difference between the number-in-hand and number-on-the-forehead models, let us consider the equality problem EQ_n. This is defined as $EQ_n(x_1, \ldots, x_k) = T$ if all of the n-bit strings $x_1, \ldots, x_k$ are all equal, and F otherwise. In the number-in-hand model of communication complexity a lower bound of n can be proved for EQ_n by a reduction to the 2-player case. On the other hand there is a 2-bit protocol for the equality problem in the number-on-the-forehead model for $k \geq 3$ players. In this protocol player 1 and player 2 check if all the inputs they observe are equal, and output T if they are and F otherwise. It is not hard to check that this protocol is correct.

8.1 Protocol decomposition

Let X be a finite set and consider a multiparty function $f : X^k \to \{-1, +1\}$. We will associate this function with a k-dimensional tensor A_f, known as the communication tensor, where $A_f[x_1, \ldots, x_k] = f(x_1, \ldots, x_k)$.

As in the case of two-party deterministic communication complexity, we first look at the decomposition of the communication tensor into simpler objects which is induced by a correct communication protocol. Recall that in the case of two-party deterministic complexity, these simpler objects were monochromatic combinatorial rectangles.

In the number-in-hand model, the analog of combinatorial rectangles are the natural higher dimension version of these objects. In other words, we now consider combinatorial cubes of the form $I_1 \times I_2 \times \cdots \times I_k$, where each $I_j \subseteq X$.

In the number-on-the-forehead model, the analog of combinatorial rectangles are more complicated objects known as cylinder intersections. A cylinder in the i^{th} dimension is a set $C_i \subseteq X^k$ which does not depend on the i^{th} coordinate. In other words, if $(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k) \in C_i$ then also $(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_k) \in C_i$ for every $x'_i \in X$. One can see how such sets might be relevant to number-on-the-forehead complexity as the message of the i^{th} player does not depend on the i^{th} input. A cylinder intersection C is a set which can be written as the intersection of cylinders: $C = C_1 \cap \ldots \cap C_k$.

Theorem 8.1 (NIH partition bound). A successful c-bit deterministic number-in-hand communication protocol for f partitions the communication tensor into at most 2^c combinatorial cubes which are monochromatic with respect to f.

Theorem 8.2 (NOF partition bound). A successful c-bit deterministic number-on-the-forehead communication protocol for f partitions the communication matrix into at most 2^c cylinder intersections which are monochromatic with respect to f.

As in the two-party case, we can use the partition bound to define an appropriate norm to bound multiparty communication complexity. From now on, we focus on the number-on-the-forehead model. Usually, it will be straightforward to transfer the elements of the discussion to the number-in-hand model as well. Also we should point out that for the

number-in-hand model information theoretic lower bounds often give better lower bounds [CKS03, BYJKS04] than the norm based approach we outline here. On the other hand, it is not known how to extend these information theoretic techniques to the number-on-the-forehead model.

A *characteristic tensor* of a set $F \subseteq X^k$ is the $(0,1)$ tensor χ_F satisfying $\chi_F[x_1, \ldots, x_k] = 1$ if and only if $(x_1, \ldots, x_k) \in F$. By a slight abuse of notation we identify a set with its characteristic tensor.

Definition 8.1 (Cylinder intersection norm). Let M be a k-tensor. Define

$$\mu(M) = \min \left\{ \sum_i |\alpha_i| : M = \sum_i \alpha_i C_i \right\}$$

where each C_i is a cylinder intersection.

Notice that when M is a matrix, this reduces to the μ norm we have seen before.

As in the two-party case, the following theorem is immediate from the partition bound.

Theorem 8.3. Let A be a sign k-tensor. Then

$$D(A) \geq \log(\mu(A)).$$

Also in direct analogy with the two-party case Theorem 4.4, we can show that the approximate version of μ can be used to lower bound randomized number-on-the-forehead complexity.

Theorem 8.4. Let A be a sign k-tensor, and $0 \leq \varepsilon < 1/2$. Then

$$R_\varepsilon(A) \geq \log(\mu^\alpha(A)) - \log(\alpha_\varepsilon)$$

where $\alpha_\varepsilon = 1/(1-2\varepsilon)$ and $\alpha \geq \alpha_\varepsilon$.

Finally, discrepancy of cylinder intersections (sometimes called multiparty discrepancy) is also defined as in the two player case. By Theorem 7.1 the bound given by the multiparty discrepancy method of a sign tensor A is equal to $\mu^\infty(A)$.

8.2 Bounding number-on-the-forehead discrepancy

Cylinder intersections are difficult combinatorial objects. After much effort, there is still essentially just one technique available for bounding multiparty discrepancy. This technique appeared in the early paper of Babai, Nisan, and Szegedy [BNS89]. Moreover, this technique is inherently limited to showing lower bounds of the form $n/2^k$. Thus one of the major obstacles to showing non-trivial bounds on number-on-the-forehead complexity for more than $\log n$ players is finding alternative lower bound approaches.

The next statement is from [BNS89], see also [Chu90, Raz00] for similar formulations.

Theorem 8.5. Let M be a k-tensor. Then

$$\left(\frac{\mu^*(M)}{\text{size}(M)}\right)^{2^k} \leq \mathbb{E}_{\bar{x}^0,\bar{x}^1}\left[\prod_{\ell\in\{0,1\}^k} M[x_1^{\ell_1}, x_2^{\ell_2}, \ldots, x_k^{\ell_k}]\right]$$

where $\bar{x}^0 = (x_1^0, \ldots, x_k^0)$ and similarly $\bar{x}^1 = (x_1^1, \ldots, x_k^1)$.

The following variant of the bound of [BNS89] first appeared in [Cha07] for use with the pattern tensor framework.

Theorem 8.6. Let M be a k-tensor of dimensions $n_1, n_2, \ldots, n_k$, and let $\text{size}(M) = n_1 n_2 \cdots n_k$ be the number of entries in M. Denote by $(M \bullet_1 M)$ the $(2k-2)$-tensor defined by

$$(M \bullet_1 M)[y_1^0, y_1^1, \ldots, y_{k-1}^0, y_{k-1}^1] = \mathbb{E}_x \prod_{\ell\in\{0,1\}^{k-1}} M[x, y_1^{\ell_1}, \ldots, y_{k-1}^{\ell_{k-1}}].$$

Then

$$\left(\frac{\mu^*(M)}{\text{size}(M)}\right)^{2^{k-1}} \leq \mathbb{E}_{\bar{y}^0,\bar{y}^1} \left|(M \bullet_1 M)\right|,$$

where $\bar{y}^0 = (y_1^0, \ldots, y_{k-1}^0)$ and similarly $\bar{y}^1 = (y_1^1, \ldots, y_{k-1}^1)$.

Both theorems are proved using repeated application of Cauchy-Schwarz inequality—it is applied k times for the first theorem, and $k-1$ many times for the second.

8.2.1 Example: Hadamard tensors

We give an example to show how Theorem 8.6 can be used in conjunction with duality. This example is similar to Example 7.1. Let H be a N-by-N Hadamard matrix. We show that $\mu^\infty(H) \geq \sqrt{N}$. Indeed, simply let the witness matrix be H itself. With this choice

$$\mu^\infty(H) \geq \frac{\langle H, H \rangle}{\mu^*(H)} = \frac{N^2}{\mu^*(H)}$$

Now we bound $\mu^*(H)$ using Theorem 8.6 which gives:

$$\mu^*(H)^2 \leq N^4 \, \mathbb{E} \, |H \bullet_1 H| = N^3$$

as $H \bullet_1 H$ has nonzero entries only on the diagonal, and these entries are of magnitude one.

Ford and Gál [FG05] extend the notion of matrix orthogonality to tensors, defining what they call Hadamard tensors.

Definition 8.2 (Hadamard tensor). Let H be a sign k-tensor. We say that H is a Hadamard tensor if

$$(H \bullet_1 H)[x_2^0, x_2^1, \ldots, x_k^0, x_k^1] = 0$$

whenever $x_i^0 \neq x_i^1$ for all $i = 2, \ldots, k$.

The "orthogonality" property of Hadamard tensors combined with Theorem 8.6 imply

$$\mu^*(H)^{2^{k-1}} \leq (k-1) \frac{N^{k 2^{k-1}}}{N}.$$

The simple proof above for Hadamard matrices can now be easily extended to Hadamard tensors:

Theorem 8.7 (Ford and Gál [FG05]). Let H be a Hadamard k-tensor of side length N. Then

$$\mu^\infty(H) \geq \left(\frac{N}{k-1} \right)^{1/2^{k-1}}$$

Remark 8.1. By doing a more careful inductive analysis, Ford and Gál obtain this result without the $k-1$ term in the denominator. They also construct explicit examples of Hadamard tensors.

8.3 Pattern Tensors

We now define a natural generalization of the pattern matrices of Sherstov [She09, She08c] to the tensor case. This generalization was first defined, in a slightly different form, by Chattopadhyay [Cha07]. Like a pattern matrix, a pattern tensor can be thought of as a block composed function. The outer function $f : \{-1,+1\}^n \to \{-1,+1\}$ can be arbitrary. Let us now describe the inner function g. The inner function takes k arguments—the first argument can be thought of as a $k-1$ dimensional tensor, and the other $k-1$ arguments as indices into the sides of this tensor. Thus

$$g(x, y_1, \ldots, y_{k-1}) = x[y_1, \ldots, y_{k-1}].$$

Fix an integer N. As a whole, the pattern tensor $A_{f,N}$ is then defined such that $A_{f,N}[x, y_1, \ldots, y_{k-1}]$ is equal to

$$f(g(x_1, y_1[1], \ldots, y_{k-1}[1]), \ldots, g(x_n, y_1[n], \ldots, y_{k-1}[n])).$$

Here x is a k dimensional sign tensor with dimensions $n \times N \times \cdots \times N$, x_i is the $k-1$ dimensional tensor achieved by constraining the first index of x to be equal to i. The y_j's are vectors of indices in $[N]^n$, and $y_j[i]$ is the i^{th} element of the j^{th} vector.

Notice that here the definition of the inner function g is less sophisticated than that used for pattern matrices. For pattern matrices, remember that a key feature is that they are strongly balanced. In the matrix case, having g be strongly balanced aided substantially in evaluating the spectral norm of a pattern matrix. In the tensor case, however, instead of using spectral norm we are using the more blunt tool of the theorem of [BNS89]. Here it is not clear that being strongly balanced is of any help, so we stick to this simpler definition of g.

Pattern tensors satisfy the following.

Theorem 8.8 ([LS09a, CA08]). Let $f : \{-1,+1\}^n \to \{-1,+1\}$ be a boolean function, then

$$\log \mu^{\alpha}(A_{f,N}) \geq \frac{\deg_{\alpha_0}(f)}{2^{k-1}} + \log \frac{\alpha_0 - \alpha}{\alpha_0 + 1},$$

for every $1 \leq \alpha < \alpha_0 < \infty$, provided $N \geq \frac{2e(k-1)2^{2^{k-1}}n}{\deg_{\alpha_0}(f)}$.

By now the reader should be able to guess the sketch of the proof of Theorem 8.8. We use the dual formulation of μ^{α} and take as a witness the pattern tensor of the dual polynomial. Namely, if v is the dual polynomial as in Lemma 7.1 we take as our witness the pattern tensor $Q = Q_{v,N}$. That is, $Q[x, y_1, \ldots, y_{k-1}]$ is equal to

$$v(g(x_1, y_1[1], \ldots, y_{k-1}[1]), \ldots, g(x_n, y_1[n], \ldots, y_{k-1}[n])).$$

Just as we saw in Section 7.2, the dual formulations of approximate degree and approximate norm fit well together. This makes handling the numerator in the max formulation of μ^{α} very easy. The inner product $\langle A_{f,N}, Q_{v,N} \rangle$ is proportional to $\langle f, v \rangle$ and similarly $\|Q_{v,N}\|_1$ is proportional to $\|v\|_1$, with the same ratio $\frac{\text{size}(Q)}{2^n}$. It is therefore left to bound $\mu^*(Q)$ in terms of something proportional to $\deg_{\alpha_0}(f)$. Here we first prove such a relation for the case $k = 2$, and then extend it to general k.

Lemma 8.1 ([Cha07, LS09a, CA08]). Let $v : \{-1,+1\}^n \to \mathbb{R}$ be a function satisfying:

(1) $\|v\|_1 \leq 1$,
(2) $\hat{v}_T = 0$ for every $T \subseteq [n]$ with cardinality $|T| \leq d$.

Take $Q = Q_{v,N}$ be the pattern matrix corresponding to v. Then

$$\mu^*(Q) \leq \frac{\text{size}(Q)}{2^{n+d/2}},$$

provided that $N \geq 2en/d$.

Proof. Consider the definition of a pattern tensor. In the two dimensional (matrix) case, the input x is an $n \times N$ sign matrix, and y is a vector in $[N]^n$. By Theorem 8.6 we have

$$\left(\frac{\mu^*(Q)}{\text{size}(Q)}\right)^2 \leq \mathbb{E}_{y^0,y^1} \left|\mathbb{E}_x Q[x,y^0]Q[x,y^1]\right|. \tag{8.1}$$

To estimate the inner expectations over x, we use the Fourier representation $v = \sum_T \hat{v}_T \chi_T$ of v.

We can express Q as a linear combination $Q = \sum_T \hat{v}_T \chi_{T,N}$, where $\chi_{T,N}$ is the pattern matrix corresponding to the character χ_T. Now the right hand side of (8.1) becomes

$$\mathbb{E}_{y^0,y^1} \left| \mathbb{E}_x \sum_{T,T'} \hat{v}_T \hat{v}_{T'} \chi_{T,N}[x,y^0]\chi_{T',N}[x,y^1] \right|.$$

By linearity of expectation and the triangle inequality this is bounded by

$$\sum_{T,T'} |\hat{v}_T \hat{v}_{T'}| \, \mathbb{E}_{y^0,y^1} \left|\mathbb{E}_x \chi_{T,N}[x,y^0]\chi_{T',N}[x,y^1]\right|.$$

We now use the properties of v. First, $\|v\|_1 \leq 1$ and therefore $|\hat{v}_T| \leq \frac{1}{2^n}$. In addition $\hat{v}_T = 0$ for every set T with $|T| \leq d$. We therefore arrive at the following expression

$$\frac{1}{2^{2n}} \sum_{T,T':|T|,|T'|>d} \mathbb{E}_{y^0,y^1} \left|\mathbb{E}_x \chi_{T,N}[x,y^0]\chi_{T',N}[x,y^1]\right|.$$

This is equal to

$$\frac{1}{2^{2n}} \sum_{T,T':|T|,|T'|>d} \mathbb{E}_{y^0,y^1} \left| \mathbb{E}_x \prod_{i \in T} x[i,y^0[i]] \prod_{j \in T'} x[j,y^1[j]] \right|.$$

The expectation inside the absolute value is equal to 0 if $T \neq T'$, and also if $T = T'$ but there is an element $i \in T$ such that $y^0[i] \neq y^1[i]$. The value of this expectation is 1 in all other cases. Our expression is therefore equal to

$$\frac{1}{2^{2n}} \sum_{T:|T|>d} \Pr_{y^0,y^1} \left[\forall i \in T \ , \ y^0[i] = y^1[i]\right]$$

For a set T of cardinality $|T| = t$, we have

$$\Pr_{y^0,y^1}\left[\forall i \in T\ ,\ y^0[i] = y^1[i]\right] \leq N^{-t}.$$

Therefore, the sum of these probabilities above can be bounded as follows

$$\begin{aligned}\frac{1}{2^{2n}}\sum_{t=d+1}^{n}\binom{n}{t}N^{-t} &\leq \frac{1}{2^{2n}}\sum_{t=d+1}^{n}\left(\frac{en}{dN}\right)^t \\ &\leq \frac{1}{2^{2n+d}},\end{aligned}$$

assuming $N \geq 2en/d$. Plugging this back into (8.1) we get the desired result. □

The extension of Lemma 8.1 to general k is:

Lemma 8.2 ([Cha07, LS09a, CA08]). Let $v : \{-1,+1\}^n \to \mathbb{R}$ be a function satisfying:

(1) $\|v\|_1 \leq 1$,
(2) $\hat{v}_T = 0$ for every $T \subset [n]$ with cardinality $|T| \leq d$.

Take $Q = Q_{v,N}$ be the pattern k-tensor corresponding to v. Then

$$\mu^*(Q) \leq \frac{\mathrm{size}(Q)}{2^{n+d/2^{k-1}}},$$

provided that $N \geq \frac{2e(k-1)2^{2^{k-1}}n}{d}$.

Proof. The proof starts as the proof of Lemma 8.1. First, Theorem 8.6 is applied

$$\left(\frac{\mu^*(Q)}{\mathrm{size}(Q)}\right)^{2^{k-1}} \leq \mathbb{E}_{\bar{y}^0,\bar{y}^1}\left|\mathbb{E}_x \prod_{l\in\{0,1\}^{k-1}} Q[x, y_1^{l[1]}, \ldots, y_{k-1}^{l[k-1]}]\right|,$$

where $\bar{y}^0 = (y_1^0, \ldots, y_{k-1}^0)$ and similarly $\bar{y}^1 = (y_1^1, \ldots, y_{k-1}^1)$. Now we use the Fourier representation of v. Denote by $\binom{[n]}{>d}$ the family of subsets

of $[n]$ whose size is larger than d, and let $\mathcal{T} = \binom{[n]}{>d}^{2^{k-1}}$ be the family of 2^{k-1}-tuples of subsets from $\binom{[n]}{>d}$. Then,

$$\frac{1}{2^{2^{k-1}n}} \sum_{(T_1,\ldots,T_{2^{k-1}})\in\mathcal{T}} \mathbb{E}_{\bar{y}^0,\bar{y}^1} \left| \mathbb{E}_x \prod_{l\in\{0,1\}^{k-1}} \chi_{T_l,N}[x, y_1^{l[1]}, \ldots, y_{k-1}^{l[k-1]}] \right|.$$

By definition of $\chi_{T_l,N}$ this is equivalent to

$$\frac{1}{2^{2^{k-1}n}} \sum_{(T_1,\ldots,T_{2^{k-1}})\in\mathcal{T}} \mathbb{E}_{\bar{y}^0,\bar{y}^1} \left| \mathbb{E}_x \prod_{l\in\{0,1\}^{k-1}} \prod_{i\in T_l} x[i, y_1^{l[1]}[i], \ldots, y_{k-1}^{l[k-1]}[i]] \right|.$$

As we have seen in Lemma 8.1, the inner expectation can be equal to either 0 or 1. The value 1 is obtained if and only if every vector of indices $(i, y_1^{l[1]}[i], \ldots, y_{k-1}^{l[k-1]}[i])$ that appears in the expression inside the expectation, appears there an even number of times. Thus, to complete the proof, we bound the probability of this event. Clearly, a sufficient condition for this event not to hold is that there is some $i \in \bigcup_l T_l$ such that the 2^{k-1} possible vectors $(y_1^{l[1]}[i], \ldots, y_{k-1}^{l[k-1]}[i])$ for $l \in \{0,1\}^{k-1}$, are all distinct. Namely, for every $m = 1 \ldots k-1$, $y_m^0[i] \neq y_m^1[i]$. If this condition holds we call the index i, *nondegenerate*, and i is called *degenerate* otherwise.

Suppose that for some choice of $y_1^0, \ldots, y_{k-1}^0, y_1^1, \ldots, y_{k-1}^1$ there are g many degenerate indices $i \in [n]$. By the above reasoning the number of sets $(T_1, \ldots, T_{2^{k-1}}) \in \mathcal{T}$ which lead to a nonzero expectation is at most

$$\left(\sum_{r=d+1}^{g} \binom{g}{r} \right)^{2^{k-1}} \leq 2^{g2^{k-1}}.$$

Now we bound the probability that for $y_1^0, \ldots, y_{k-1}^0, y_1^1, \ldots, y_{k-1}^1$ there are g many degenerate indices. The probability that $y_m^0[i] \neq y_m^1[i]$ is $1/N$. Thus by a union bound, the probability that a single index is degenerate is at most $(k-1)/N$. Finally, as each index is chosen independently, the probability of g many degenerate cubes is at most

$$\binom{n}{g} \left(\frac{k-1}{N} \right)^g.$$

Putting everything together we have

$$\begin{aligned}\mu^*(Q)^{2^{k-1}} &\leq \frac{\text{size}(Q)^{2^{k-1}}}{2^{2^{k-1}n}} \sum_{g=d+1}^{m} \binom{n}{g} \left(\frac{k-1}{N}\right)^g 2^{g2^{k-1}} \\ &\leq \frac{\text{size}(Q)^{2^{k-1}}}{2^{2^{k-1}n}} \sum_{g=d+1}^{m} \left(\frac{e(k-1)2^{2^{k-1}}n}{dN}\right)^g \\ &\leq \frac{\text{size}(Q)^{2^{k-1}}}{2^{2^{k-1}n+d}}.\end{aligned}$$

The last step holds, provided that $N \geq 2e(k-1)2^{2^{k-1}}n/d$.

□

8.4 Applications

8.4.1 The complexity of disjointness

As mentioned, the best lower bound proved so far on the NOF communication complexity of a k-tensor is of the form $\frac{n}{2^{k-1}}$, where n is the size of the input [FG05] (see also Section 8.2). This bound is proved using discrepancy, and as long as we use Theorem 8.6 to bound discrepancy, this 2^{k-1} factor will remain. Improving on this is a big open problem in multiparty communication complexity. In particular, an explicit function which requires super-polylogarithmic complexity in the number-on-the-forehead model with super-polylogarithmic many players would give an explicit function outside of the circuit complexity class ACC^0 (see Chapter 9). But even within this limit, lower bounds on the communication complexity of explicit functions in the number-on-the-forehead model have strong applications.

Beame et al [BPS06] showed that an $\omega(\log^4 n)$ lower bound for 3-party communication complexity of the set-disjointness function, implies superpolynomial lower bounds on tree-like Lovász-Schrijver proof systems that refute unsatisfiable CNF's. More generally, a strong lower bound on $(k+1)$-party number-on-the-forehead communication complexity of set-disjointness implies a strong lower bound for all tree-like proof systems whose formulas are degree k polynomial inequalities.

Let $\text{OR}_n : \{0,1\}^n \to \{-1,+1\}$ be the OR function on n bits, then the k-part set-disjointness function $\text{DISJ}_{k,n} : (\{0,1\}^n)^k \to \{-1,+1\}$ is

defined as $\mathrm{DISJ}_{k,n}(x_1, \ldots, x_k) = \mathrm{OR}_n(x_1 \wedge x_2 \ldots \wedge x_k)$.

Recall the discussion from Section 8.3, about the dependency of μ^α on α. We saw that $\mu^\alpha(\mathrm{DISJ}_{2,n})$ decreases very rapidly as α tends to infinity. In particular in Lemma 7.2 we proved that $\mu^\infty(\mathrm{DISJ}_{2,n}) = O(n)$, while $\mu(\mathrm{DISJ}_{2,n}) \geq \left(\frac{\sqrt{5}}{2}\right)^n$. Thus the discrepancy bound yields weak results for this function. In Section 7.2.3 we proved that for bounded approximation it holds that $\mu^2(A_n) \geq \Omega(2^{\sqrt{n}})$, by reduction to approximate degree, and we will see in Chapter 5 that this is tight.

The situation with the k-party set-disjointness function $\mathrm{DISJ}_{k,n}$, for $k \geq 3$, seems similar. The extension of the upper bound on $\mu^\infty(\mathrm{DISJ}_{k,n})$ is quite straightforward. And exponentially better bounds can be proved for $\mu^2(\mathrm{DISJ}_{k,n})$:

Theorem 8.9 ([LS09a, CA08]). For every k and n

$$R(\mathrm{DISJ}_{k,n}) \geq \Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right).$$

Observe that the lower bound on $\mathrm{DISJ}_{k,n}$ becomes weaker as k grows. It is reasonable to assume that even for $k = 3$ the lower bound is no longer tight. The lower bounds on k-party set-disjointness are proved by a reduction to pattern tensors of Section 8.3.

Proof. For integers m, M, let $A_{m,M}$ be the pattern tensor that corresponds to OR_m. We show that $A_{m,M}$ is a subtensor of $\mathrm{DISJ}_{k,n}$. Recall that $A_{m,M}[x, y_1, \ldots, y_{k-1}]$ is equal to

$$\mathrm{OR}(x[1, y_1[1], \ldots, y_{k-1}[1]], \ldots, x[m, y_1[m], \ldots, y_{k-1}[m]]).$$

Here we consider x as a k-dimensional Boolean tensor with dimensions $m \times M \times \cdots \times M$, and the y_i's are vectors of m indices in $[M]$. Thus the $k-1$ inputs $y_1, \ldots, y_{k-1}$ serve as pointers into the first input x. Then the OR function is invoked on the m bits of x which are selected by the inputs $\{y_i\}$. Let us interpret the disjointness function in a similar manner, namely

$$\mathrm{DISJ}[x, z_1, \ldots, z_{k-1}] = \mathrm{OR}(x \wedge (z_1 \wedge \cdots \wedge z_{k-1})).$$

Since the OR function only cares about the 1 inputs in x, we can think of disjointness as the OR function invoked on the bits of x where the vector $(z_1 \wedge \cdots \wedge z_{k-1})$ is equal to 1. In other words, first the z_i's select the set of bits where they are all equal to 1, then OR is called upon those bits in x.

Thus, the only obstacle in relating the pattern tensor $A_{m,M}$ with disjointness is in the manner in which the last inputs point into the first. This difference can be bridged though. Let $x, y_1, \ldots, y_{k-1}$ be inputs to $A_{m,M}$. We associate each y_i with a Boolean tensor z_i of dimensions $m \times M \times \cdots \times M$ defined as follows: z_i is equal to 1 on inputs of the form $(t, s_1, \ldots, s_{i-1}, y_i[t], s_{i+1}, \ldots, s_{k-1})$, and is equal to zero on all other inputs. With this definition it is not hard to check that the vector $(z_1 \wedge \cdots \wedge z_{k-1})$ points into x exactly the same way as $y_1, \ldots, y_{k-1}$. We see therefore that $A_{m,M}$ is a subtensor of $\mathrm{DISJ}_{k,n}$ where $n = mM^{k-1}$.

Now, back to the lower bound. By Theorem 8.8

$$\log \mu^2(A_{m,M}) \geq \frac{\deg_3(\mathrm{OR}_m)}{2^{k-1}} - O(1),$$

provided $M \geq \frac{2e(k-1)2^{2^{k-1}}m}{\deg_3(\mathrm{OR}_m)}$. Nisan and Szegedy [NS94] show that $\deg_3(\mathrm{OR}_m) = \Omega(\sqrt{m})$, thus

$$\log \mu^2(A_{m,M}) \geq \Omega\left(\frac{\sqrt{m}}{2^{k-1}}\right),$$

provided $M \geq \Omega\left(2^{2^k}\sqrt{m}\right)$. Theorem 8.4 implies that this lower bound holds for the randomized communication complexity of $A_{m,M}$ as well. Since $A_{m,M}$ is a subtensor of $\mathrm{DISJ}_{k,n}$ for $n = mM^{k-1}$, or alternatively for $\sqrt{m} = \Theta(n^{\frac{1}{k+1}}/2^{2^k})$, we get

$$R(\mathrm{DISJ}_{k,n}) \geq \Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right).$$

□

Observe that Theorem 8.9 provides a nontrivial lower bound on $\mathrm{DISJ}_{k,n}$ only as long as the number of players is at most $\log \log n$. Beame and Huynh-Ngoc obtained a lower bound of the form $\Omega(2^{\sqrt{\log n}/\sqrt{k}-k})$

on k-party number-on-the-forehead randomized communication complexity of set-disjointness, which is nontrivial up to $\Theta(\log^{1/3} n)$ players. Beame and Huynh-Ngoc show their results by reducing the communication complexity problem to a stronger notion of approximate degree which they call (ε, α)-approximate degree. This notion of degree is not easy to handle, but they give a general technique to convert functions of large approximate degree to functions of large (ε, α)-approximate degree. This technique cannot be applied directly to OR, thus their lower bounds on set-disjointness is proved via reduction. In addition to the new notion of approximate degree, they also use pattern tensors with other classes of inner functions g, than the one presented in Section 8.3.

8.4.2 Separating communication complexity classes

It is interesting to compare the strength of different types of communication protocols, deterministic vs. nondeterministic, deterministic vs. randomized etc. The example of the identity function from Chapter 4 demonstrates that randomized protocols can be much more powerful than deterministic ones. As we saw, for this simple function there is a randomized protocol which requires constant number of bits, while there is no deterministic protocol that is more efficient than the trivial one. Randomized protocols are known to be more powerful than deterministic ones also for k-party number-on-the-forehead multiparty communication complexity with k up to $n^{O(1)}$ [BDPW07]. When $k \geq 3$, though, this separation is not proved for an explicit function, but rather via a counting argument. Note that an explicit separation for $k \geq 3$ seems to require new ideas, as there is currently no lower bound technique for deterministic number-on-the-forehead communication complexity that does not also hold for randomized communication.

The set-disjointness function of Section 8.4.1 demonstrates that nondeterministic protocols can be much stronger than randomized protocols, and thus also deterministic ones. There is a simple nondeterministic protocol for the set-disjointness, as we have seen in Chapter 3 for two players. It is more convenient to specify this protocol for the complement function *set-intersection*: The players nondeterministically guess a coordinate, then the players output 1 if their input is equal to 1

on that coordinate. The output of the protocol is 1 if all players wrote 1 on the blackboard, and the output is 0 otherwise. Certainly, if the player's subsets intersect there is a guess which leads to an output of 1. While if the sets do not intersect then all guesses leads to an output of 0. Thus the nondeterministic communication complexity of $D_{k,n}$ is $O(\log n)$. On the other hand, by results from Section 8.4.1 for 2-players, and Section 8.4.1 for up to $\log \log n$ players, the randomized communication complexity of set-intersection is exponentially larger. Using similar techniques as in Section 8.3 David, Pitassi, and Viola [DPV08] give an explicit function which separates nondeterministic and randomized number-on-the-forehead communication complexity for up to $\Omega(\log n)$ players. They are also able, for any constant c to give a function computable in AC^0 which separates them for up to $c \log \log n$ players. (Note that disjointness can be also computed in AC^0.)

In fact the above simple nondeterministic protocol for set-intersection provides another explanation for why the discrepancy method yields bad bounds for set-disjointness, as discrepancy is also a lower bound on nondeterministic communication complexity:

Theorem 8.10. For any sign k-tensor A

$$N^k(A) \geq \log \mu^\infty(A) - O(1).$$

9

Upper bounds on multiparty communication complexity

The study of multiparty communication complexity has applications to many other models of communication. Often these results take on the following form: an efficient algorithm for a problem f in some computational model results in an efficient multiparty communication protocol for some related function g. Of course this means that by showing *lower bounds* on the multiparty communication complexity of g one can obtain lower bounds in the desired computational model. Some examples of results of this type are:

(1) Streaming algorithms is a burgeoning field designed to compute properties of massive data sets too large to be stored in memory in their entirety. A seminal paper in this field [AMS99] shows that algorithms to compute some important features of a data stream lead to efficient algorithms for a promise version of DISJOINTNESS in the multiparty number-in-hand model.

(2) Beame, Pitassi, and Segerlind [BPS06] show that efficient refutations of certain unsatisfiable formulas in a broad class of proof systems give rise to efficient algorithms for DISJOINT-

NESS in the multiparty number-on-the-forehead model.

(3) Work of [All89, Yao90, BT94, HG91] has given a huge motivation to the study of the number-on-the-forehead model by showing that functions with small ACC^0 circuits have efficient number-on-the-forehead protocols. ACC^0 is the class of languages which can be computed by polynomial size circuits of constant depth using AND, OR, NOT and MOD m gates for any m.

Fortunately, the reductions involved for items (1),(3) are relatively easy and we will present them below. The argument for item (2) is fairly involved and we refer the reader to the paper of Beame, Pitassi, and Segerlind for more details.

We will also discuss an upper bound on multiparty communication complexity in the more traditional sense. This is a clever protocol of Grolmusz [Gro94] which shows that any problem of the form $f(x_1 \wedge \ldots \wedge x_k)$ for a symmetric function f has a number-on-the-forehead protocol of complexity roughly $k^2 n/2^k$. Showing a lower bound on an explicit function for more than $k = \log n$ players is a major open problem, with serious implications by item (3) above. This protocol illuminates the surprising power of the number-on-the-forehead model and part of the difficulty to break the $k = \log n$ barrier.

9.1 Streaming lower bounds

We are increasingly exposed to large amounts of data. Sometimes we would like to compute some features of this data, but storing it in its entirety is infeasible. Imagine instead that the data is streaming by and we only have a small amount of memory to use to take notes as it passes. Can we still compute meaningful properties of the data in this scenario?

Many important properties of the data can be learned from its *frequency moments* Consider a stream $a = a_1 \ldots a_m$ consists of m elements from a universe $[n]$. For $i \in [n]$, let c_i be the number of occurrences of i in the string a. Then the k^{th} frequency moment, F_k is

$$F_k = \sum_i c_i^k.$$

Notice that F_0 is simply the number of distinct characters in the string and F_1 is the length of the string. We define F_∞ to be

$$F_\infty = \max_i c_i.$$

In a seminal paper, Alon, Matias, and Szegedy [AMS99] consider the complexity of streaming algorithms computing frequency moments. Somewhat surprisingly, they show that one can approximate F_0, F_1, F_2 by a randomized algorithm using logarithmic space. On the other hand, approximately computing F_k for $k \geq 6$ requires space of order $n^{\Omega(1)}$. This lower bound was later shown to also hold for all $k \geq 3$ [BYJKS04], even for multiple pass algorithms [CKS03], thus giving essentially a complete picture of the complexity of computing frequency moments.

All of these lower bounds work by showing that an efficient algorithm for computing F_k leads to an efficient protocol for a promise version of DISJOINTNESS in the number-in-the-hand model of multiparty communication complexity, and then showing lower bounds on this communication problem. We will now sketch how this reduction works.

Let us first see how to get a lower bound on the space required to compute F_∞ from the two-party lower bound on DISJOINTNESS. This case is easier and will illustrate the basic idea. Say that we have a streaming algorithm which uses space c and which for any stream of at most $2n$ elements outputs a value in $(1 \pm 1/4)F_\infty$ with probability at least $2/3$. We will use this streaming algorithm to construct a randomized protocol for disjointness using c bits of communication and success probability at least $2/3$.

On input $x \in \{0,1\}^n$, Alice forms a string a_x consisting of the indices where x has a one. Bob does the same with y to create a_y. Alice then runs the streaming algorithm on a_x and sends the contents of the memory to Bob with c bits of communication. Bob finishes the computation of the streaming algorithm on a_y. If the computation of F_∞ returns a value of at most $5/4$, then they say that the strings do not intersect, otherwise that they do. Notice that if the strings do not intersect the true value of F_∞ will be one, whereas if they do intersect, it will be at least two. Thus by the success guarantee of the streaming algorithm, the communication protocol will be correct with probability

at least 2/3. Finally, as Kalyanasundaram and Schnitger [KS87] have shown that any randomized protocol for disjointness requires communication $\Omega(n)$ it follows that any streaming algorithm approximating F_∞ with high probability requires memory linear in the length of the stream. Notice that the algorithm we constructed was in fact one-way; as the Kalyanasundaram and Schnitger lower bound works for general protocols, this shows that even streaming algorithms which make multiple passes over the data in the same order must use linear space.

The general case reduces computing F_k to a promise version of disjointness, UNIQUE DISJOINTNESS, in the k-party number-in-hand model. This problem is defined as $\mathrm{UDISJ}_k(x_1, \ldots, x_k) = T$ if $x_1 \cap \cdots \cap x_k = \emptyset$ and F if they intersect in exactly one element. Otherwise the problem is undefined.

Theorem 9.1 (Alon, Matias, and Szegedy [AMS99]). Fix a natural number k. Suppose there is a randomized streaming algorithm which for every stream of length n uses $c(n)$ bits of memory and outputs a value in $[\frac{9}{10}F_k, \frac{11}{10}F_k]$ with probability at least 2/3. Then in the $n^{1/k}$-party number-in-hand model

$$R(\mathrm{UDISJ}_{n^{1/k}}) = O(c(n)n^{1/k})$$

Proof. We will assume that the total number of ones in the inputs to the players is exactly n. As in the F_∞ case described above, each player converts her input x_i into a stream a_i consisting of the indices where x_i has a one. The first player then runs the streaming algorithm for F_k on a_1 and communicates with c bits the contents of the memory to the second player, who does the same thing and reports the contents of the memory to the third player, etc. The last player sees the final output of the streaming algorithm and says that the strings do intersect if the answer is at least $\frac{12}{10}n$, and otherwise says that they do not intersect. The total communication is thus c and is one-way. Notice that if the inputs are mutually disjoint, then F_k is exactly n. On the other hand, if the sets uniquely intersect, then F_k will be $n + n - n^{1/k} \geq 3n/2$. □

The best known results on the number-in-the-hand multiparty complexity of disjointness show

Theorem 9.2 (Chakrabarti, Khot, and Sun [CKS03]). In the number-in-the-hand multiparty model, s players must communicate at least $\Omega(n/s \log(s))$ bits to determine if their sets from the universe $[n]$ are mutually disjoint or uniquely intersecting with probability at least $2/3$.

By the above reduction, this theorem has the following application:

Theorem 9.3. Any streaming algorithm which computes F_k within a multiplicative factor of 1.1 with high probability must use space at least $\Omega(n^{1-2/k}/\log n)$, even if a constant number of passes are permitted.

9.2 NOF upper bounds

A major open problem is to show a lower bound in the number-on-the-forehead model of multiparty communication complexity which remains nontrivial for more than $\log n$ players. We have surveyed the existing lower bound techniques in the number-on-the-forehead model, and also seen that they are inherently limited to bounds of the form $n/2^k$ for k-many players. The results we discuss in this section both give a lot of motivation to showing lower bounds for more than $\log n$ players, and show that this is not possible for a class of well studied functions.

9.2.1 Protocol of Grolmusz

Let $f : \{0,1\}^n \to \{0,1\}$ be a symmetric function. In other words, $f(x) = f(y)$ whenever x and y have the same number of ones. Let $x_1, \ldots, x_k \in \{0,1\}^n$ be the inputs to the k-players. We will think of the input as being described by a k-by-n matrix X whose i^{th} row is x_i. The key step in the protocol of Grolmusz is the following lemma:

Lemma 9.1 (Grolmusz [Gro94]). Let f be a symmetric function. Suppose the players know that some string $r \in \{0,1\}^k$ does not appear in the input matrix X. Then they can evaluate $f(x_1 \wedge \ldots \wedge x_k)$ with $k \log n$ bits of communication.

Proof. Notice that if the players can count the number of all-one columns in X then they can compute f. By rearranging the rows of X as necessary, we may assume that the missing column r is of the form $0^\ell 1^{k-\ell}$ where $\ell \in \{1, \ldots, k\}$. If $\ell = 0$ then the all-one column does not appear, and the players can immediately evaluate f.

More generally, let $e_i = 0^i 1^{k-i}$ for $i \in \{0, \ldots, k\}$. Thus the players want to count the number of times e_0 appears as a column of X. Let E_i be the number of times the string e_i appears as a column of X.

Although the first player cannot distinguish between a column of the form e_0 or e_1 as he does not see the first bit, he can exactly compute $E_0 + E_1$. Player 1 announces this number with $\log n$ bits. Similarly, player 2 announces $E_1 + E_2$. The players continue this way until they reach player ℓ, who will announce $E_{\ell-1}$ as by assumption e_ℓ does not appear as a column of X. With this knowledge, the players can then solve for E_0 and evaluate f. □

Theorem 9.4 (Grolmusz [Gro94]). Let f be a symmetric function. Then

$$D_k(f(x_1 \wedge \ldots \wedge x_k)) \leq k^2 \log(n) \left\lceil \frac{n}{2^{k-1} - 1} \right\rceil .$$

Proof. The first player will play a special role in the protocol. He (mentally) partitions the input matrix X into $\left\lceil \frac{n}{2^{k-1}-1} \right\rceil$ many blocks of columns of size $2^{k-1}-1$. By counting, in each of these blocks of columns, there is some $k-1$ bit string which does not appear, and can be identified by the first player. The first player announces these strings, and then the players perform the protocol given in the lemma. The total communication is

$$k^2 \log(n) \left\lceil \frac{n}{2^{k-1} - 1} \right\rceil .$$

□

Remark 9.1. In the special case where f is the parity function, this communication can be reduced to

$$k \left\lceil \frac{n}{2^{k-1}-1} \right\rceil$$

as in the lemma the players do not need to say $E_i + E_{i+1}$ but just the parity of this number.

Remark 9.2. Notice that the protocol of Grolmusz is nearly simultaneous, but not quite as the first player must announce the missing columns to the other players. Babai et al. [BGKL03] have shown that any function $f(x_1 \wedge \ldots \wedge x_k)$ for symmetric f indeed has a simultaneous protocol with $O(\log^{O(1)}(n))$ bits of communication whenever the number of players is larger than $\log n + 1$.

9.2.2 Protocol for small circuits

One of the principal motivations for studying multiparty number-on-the-forehead communication complexity is that lower bounds in this model imply circuit complexity lower bounds. A key observation in this connection is due to Håstad and Goldmann.

Lemma 9.2 (Håstad and Goldmann [HG91]). Suppose that a function f can be computed by a depth-2 circuit whose top gate is an arbitrary symmetric function of fan-in s and whose bottom gates compute arbitrary functions of fan-in at most $k-1$. Then, under any partition of the input variables, the k-party number-on-the-forehead complexity of f is at most $k \log(s)$. Furthermore, this can be achieved by a simultaneous protocol.

Proof. As each bottom gate has fan-in at most $k-1$, under any partition of the input variables, there is some player who sees the entire input to this gate. By a scheme arranged beforehand, the players partition these gates among themselves so that each gate is computed by some

player. Each player then announces the number of gates assigned to him which evaluate to true. This takes $\log s$ bits of communication. Once the players know the total number of bottom gates which evaluate to true, they can compute f. The total communication is $k \log(s)$. □

The functions which can be computed by quasipolynomial size depth-2 circuits with a symmetric top gate and bottom gates of polylogarithmic size fan-in is a surprisingly rich class. Indeed, Allender [All89] shows that this class can compute all of AC^0. Further work by Yao [Yao90] shows that the probabilistic version of this class can compute all of ACC^0 and Beigel and Tarui [BT94] improve this to a deterministic simulation. We record this statement for reference.

Theorem 9.5 (Beigel and Tarui). Any language in ACC^0 can be computed by a depth-2 circuit of size $2^{\log^{O(1)}(n)}$ with a symmetric gate at the top and AND gates of fan-in $\log^{O(1)} n$ at the bottom.

As a consequence, showing that a function f requires super-polylogarithmic communication for super-polylogarithmic many players in the simultaneous number-on-the-forehead model will show that f is not in ACC^0. We currently do not know of any explicit function which is outside of ACC^0.

Acknowledgements

We would like to thank Eitan Bachmat, Harry Buhrman, Arkadev Chattopadhyay, Nati Linial, Gideon Schechtman, Sasha Sherstov, Ronald de Wolf, and the anonymous referee for many helpful comments which improved this survey.

References

[AA05] S. Aaronson and A. Ambainis. Quantum search of spatial regions. *Theory of Computing*, 1:47–79, 2005.

[AKM+06] N. Alon, Y. Kohayakawa, C. Mauduit, C. Moreira, and V. Rödl. Measures of pseudorandomness for finite sequences: minimal values. *Combinatorics, Probability, and Computing*, 15:1–29, 2006.

[All89] E. Allender. A note on the power of threshold circuits. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 580–584, 1989.

[Alo03] N. Alon. Problems and results in extremal combinatorics, part i. *Discrete Mathematics*, 273:31–53, 2003.

[Alo09] N. Alon. Perturbed identity matrices have high rank: proof and applications. *Combinatorics, Probability, and Computing*, 18:3–15, 2009.

[AMS99] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

[AN06] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. *SIAM Journal on Computing*, 35:787–803, 2006.

[AS89] N. Alon and P. Seymour. A counterexample to the rank-coloring conjecture. *Journal of Graph Theory*, 13(4):523–525, 1989.

[AUY83] A. Aho, J. Ullman, and M. Yannakakis. On notions of information transfer in VLSI circuits. In *Proceedings of the 15th ACM Symposium on the Theory of Computing*, pages 133–139. ACM, 1983.

[AW09] S. Aaronson and A. Wigderson. Algebrization: A new barrier in complexity theory. *ACM Transactions on Computing Theory*, 1, 2009.

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS'98.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th ACM Symposium on the Theory of Computing*, pages 63–68. ACM, 1998.

[BDPW07] P. Beame, M. David, T. Pitassi, and P. Woelfel. Separating deterministic from randomized nof multiparty communication complexity. In *Proceedings of the 34th International Colloquium On Automata, Languages and Programming*, Lecture Notes in Computer Science. Springer-Verlag, 2007.

[BES02] S. Ben-David, N. Eiron, and H. Simon. Limitations of learning via embeddings in Euclidean half spaces. *Journal of Machine Learning Research*, 3:441–461, 2002.

[BFS86] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th IEEE Symposium on Foundations of Computer Science*. IEEE, 1986.

[BGKL03] L. Babai, A. Gál, P. Kimmel, and S. Lokam. Simultaneous messages vs. communication. *SIAM Journal on Computing*, 33(1):137–166, 2003.

[BHN08] P. Beame and D. Huynh-Ngoc. Multiparty communication complexity of AC^0. Technical Report TR-08-082, ECCC, 2008.

[BNS89] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols and Logspace-hard pseudorandom sequences. In *Proceedings of the 21st ACM Symposium on the Theory of Computing*, pages 1–11. ACM, 1989.

[BPS06] P. Beame, T. Pitassi, and N. Segerlind. Lower bounds for Lovász-Schrijver systems and beyond follow from multiparty communication complexity. *SIAM Journal on Computing*, 37(3):845–869, 2006.

[BPSW06] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product lemma for corruption and the NOF complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006.

[BT94] R. Beigel and J. Tarui. On ACC. *Computational Complexity*, 4:350–366, 1994.

[Buh07] Harry Buhrman, December 2007. Personal communication.

[BVW07] H. Buhrman, N. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*, pages 24–32. IEEE, 2007.

[BW01] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th IEEE Conference on Computational Complexity*, pages 120–130, 2001.

[BW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288:21–43, 2002.

[BYJKS04] Z. Bar-Yossef, T. Jayram, R. Kumar, and D. Sivakumar. Information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[CA08] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. Technical Report TR-08-002, ECCC, 2008.

[Cha07] A. Chattopadhyay. Discrepancy and the power of bottom fan-in depth-three circuits. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 449–458. IEEE, 2007.

[Cha08] A. Chattopadhyay. PhD thesis, McGill University, 2008.

[CHSH69] J. Clauser, M. Horne, A. Shimony, and R. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969.

[Chu90] F. Chung. Quasi-random classes of hypergraphs. *Random Structures and Algorithms*, 1:363–382, 1990.

[CKS03] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal lower bounds on the multi-party communication complexity of set-disjointness. In *Proceedings of the 18th IEEE Conference on Computational Complexity*. IEEE, 2003.

[DKLR08] J. Degorre, M. Kaplan, S. Laplante, and J. Roland. The communication complexity of non-signaling distributions. Technical Report 0804.4859, arXiv, 2008.

[DPV08] M. David, T. Pitassi, and E. Viola. Improved separations between nondeterministic and randomized multiparty communication. In *APPROX-RANDOM*, volume 5171 of *Lecture Notes in Computer Science*, pages 371–384. Springer, 2008.

[Faj88] S. Fajtlowicz. On conjectures of graffiti. *Discrete Mathematics*, 72:113–118, 1988.

[FG05] J. Ford and A. Gál. Hadamard tensors and lower bounds on multiparty communication complexity. In *Proceedings of the 32th International Colloquium On Automata, Languages and Programming*, pages 1163–1175, 2005.

[FK99] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19:175–220, 1999.

[For02] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *Journal of Computer and System Sciences*, 65:612–625, 2002.

[Gro94] V. Grolmusz. The BNS lower bound for multi-party protocols is nearly optimal. *Information and computation*, 112(1):51–54, 1994.

[GW95] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.

[Hås90] J. Håstad. Tensorrank is NP-complete. *Journal of Algorithms*, 11:644–654, 1990.

[HG91] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1:113–129, 1991.

[Jam87] G. J. O. Jameson. *Summing and nuclear norms in banach space theory*. Cambridge University Press, 1987.

[JKS03] T. Jayram, R. Kumar, and D. Sivakumar. Two applications of information complexity. In *Proceedings of the 35th ACM Symposium on the Theory of Computing*, pages 673–682. ACM, 2003.

[JL01] W. Johnson and J. Lindenstrauss. Basic concepts in the geometry of Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 1–84. North-Holland, Amsterdam, 2001.

[JR93] T. Jiang and B. Ravikumar. Minimal NFA problems are hard. *SIAM Journal on Computing*, 22:1117–1141, 1993.

[KKN95] M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. *SIAM Journal on Discrete Mathematics*, 8(1):76–92, 1995.

[Kla01] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*. IEEE, 2001.

[Kla03] H. Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proceedings of the 18th IEEE Conference on Computational Complexity*. IEEE, 2003.

[KLO96] E. Kushilevitz, N. Linial, and R. Ostrovsky. The linear array conjecture of communication complexity is false. In *Proceedings of the 28th ACM Symposium on the Theory of Computing*. ACM, 1996.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[Kra96] M. Krause. Geometric arguments yield better bounds for threshold circuits and distributed computing. *Theoretical Computer Science*, 156:99–117, 1996.

[Kre95] I. Kremer. Quantum communication. Technical report, Hebrew University of Jerusalem, 1995.

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

[KS87] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. In *Proceedings of the 2nd Annual Conference on Structure in Complexity Theory*, pages 41–49, 1987.

[KS07] A. Klivans and A. Sherstov. A lower bound for agnostically learning disjunctions. In *Proceedings of the 20th Conference on Learning Theory*, 2007.

[LMSS07] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

[Lov75] L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13:383–390, 1975.

[Lov90] L. Lovász. Communication complexity: A survey. In B. Korte, L. Lovász, H. Prömel, and A. Schrijver, editors, *Paths, flows, and VLSI-layout*, pages 235–265. Springer-Verlag, 1990.

[LS88] L. Lovász and M. Saks. Möbius functions and communication complexity. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 81–90. IEEE, 1988.

[LS08] T. Lee and A. Shraibman. An approximation algorithm for approximation rank. In *Proceedings of the 24th IEEE Conference on Computational Complexity*. IEEE, 2008. arXiv:0809.2093 [cs.CC].

[LS09a] T. Lee and A. Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[LS09b] N. Linial and A. Shraibman. Learning complexity versus communication complexity. *Combinatorics, Probability, and Computing*, 18:227–245, 2009.

[LS09c] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures and Algorithms*, 34:368–394, 2009.

[LSŠ08] T. Lee, A. Shraibman, and R. Špalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 71–80. IEEE, 2008.

[LSS09] T. Lee, G. Schechtman, and A. Shraibman. Lower bounds on quantum multiparty communication complexity. In *Proceedings of the 24th IEEE Conference on Computational Complexity*. IEEE, 2009.

[LSZ09] T. Lee, A. Shraibman, and S. Zhang. Composition theorems in communication complexity. Manuscript, 2009.

[LY94] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5):960–981, 1994.

[MS82] K. Mehlhorn and E. Schmidt. Las Vegas is better than determinism in VLSI and distributed computing. In *Proceedings of the 14th ACM Symposium on the Theory of Computing*, pages 330–337. ACM, 1982.

[NC00] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[New91] I. Newman. Private versus common random bits in communication complexity. *Information Processing Letters*, 39:67–71, 1991.

[NS94] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[Nuf76] C. van Nuffelen. A bound for the chromatic number of a graph. *American Mathematical Monthly*, 83:265–266, 1976.

[NW95] N. Nisan and A. Wigderson. A note on rank vs. communication complexity. *Combinatorica*, 15(4):557–566, 1995.

[Orl77] J. Orlin. Contentment in graph theory: covering graphs with cliques. *Indigationes Mathematicae*, 80:406–424, 1977.

[Pat92] R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions (preliminary version). In *Proceedings of the 24th ACM Symposium on the Theory of Computing*, pages 468–474. ACM, 1992.

[PS86] R. Paturi and J. Simon. Probabilistic communication complexity. *Journal of Computer and System Sciences*, 33(1):106–123, 1986.

[Raz92a] A. Razborov. On submodular complexity measures. In M. Paterson, editor, *Boolean function complexity*, volume 169 of *London Math. Soc. Lecture Notes Series*, pages 76–83. Cambridge University Press, 1992.

[Raz92b] A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106:385–390, 1992.

[Raz95] R. Raz. Fourier analysis for probabilistic communication complexity. *Computational Complexity*, 5(3/4):205–221, 1995.

[Raz99] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the 31st ACM Symposium on the Theory of Computing*, pages 358–367. ACM, 1999.

[Raz00] R. Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[Raz03] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

[Ree91] J. Reeds. A new lower bound on the real Grothendieck constant. Available at `http://www.dtc.umn.edu/~reedsj`, 1991.

[RS95] R. Raz and B. Spieker. On the log rank conjecture in communication complexity. *Combinatorica*, 15(4):567–588, 1995.

[RS08] A. Razborov and A. Sherstov. The sign rank of AC^0. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, pages 57–66. IEEE, 2008.

[Sch11] I. Schur. Bemerkungen zur Theorie beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *J. Reine Angew. Math.*, 140:1–28, 1911.

[Sha03] R. Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1–2):1–22, 2003.

[She08a] A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008.

[She08b] A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[She08c] A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the 40th ACM Symposium on the Theory of Computing*, pages 85–94. ACM, 2008.

[She08d] A. Sherstov. The unbounded-error communication complexity of symmetric functions. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*. IEEE, 2008.

[She09] A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM Journal on Computing*, 38(6):2113–2129, 2009.

[SZ09a] Y. Shi and Z. Zhang. Communication complexities of XOR functions. *Quantum information and computation*, 2009.

[SZ09b] Y. Shi and Y. Zhu. Quantum communication complexity of block-composed functions. *Quantum information and computation*, 9(5,6):444–460, 2009.

[Tsi87] B. Tsirelson. Quantum analouges of the Bell inequalities: the case of two spatially separated domains. *Journal of Soviet Mathematics*, 36:557–570, 1987.

[Ung08] F. Unger. *Noise in classical and quantum computation and non-locality*. PhD thesis, University of Amsterdam, 2008.

[Vav07] S. Vavasis. On the complexity of nonnegative matrix factorization. Technical Report arXiv:0708.4149 [cs.NA], ArXiv, 2007.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.

[Yan91] M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991.

[Yao79] A. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th ACM Symposium on the Theory of Computing*, pages 209–213. ACM, 1979.

[Yao83] A. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the 24th IEEE Symposium on Foundations of Computer Science*, pages 420–428, 1983.

[Yao90] A. Yao. On ACC and threshold circuits. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 619–627. IEEE, 1990.

[Yao93] A. Yao. Quantum circuit complexity. In *Proceedings of the 34th IEEE Symposium on Foundations of Computer Science*, pages 352–360. IEEE, 1993.