Kolmogorov complexity and combinatorial methods in communication complexity

Marc Kaplan* Sophie Laplante*

Abstract

A very important problem in quantum communication complexity is to show that there is, or isn't, an exponential gap between randomized and quantum complexity for a total function. There are currently no clear candidate functions for such a separation; and there are fewer and fewer randomized lower bound techniques that are not known to extend to the quantum setting. Among these are some information theoretic proofs, such as the one of [BYJKS04] and more recently, the subdistribution bounds [JKN08]. We introduce a new technique to this family, based on Kolmogorov complexity.

In order to gain a better understanding of how these techniques differ from the family of techniques that follow from Linial and Shraibman's recent work on factorization norms [LS07], all of which extend to the quantum setting, we take another look at a few of these information theoretic proofs. Our main tool is Kolmogorov complexity.

We use Kolmogorov complexity for three different things:

- give a general lower bound in terms of Kolmogorov mutual information
- prove an alternative to Yao's minmax principle based on Kolmogorov complexity
- identify worst case inputs.

We show that our method implies the rectangle and corruption bounds [BPSW06], known to be closely related to the subdistribution bound [JKN08]. We apply our method to the hidden matching problem, a relation introduced in [BYJK08] to prove an exponential gap between quantum and classical communication. We then show that our method generalizes the VC dimension [KNR99] and shatter coefficient lower bound [BYJKS02]. In the second part, we compare one-way communication and simultaneous communication in the case of distributional communication complexity and improve the previous known result [BYJKS02].

In these proofs, the intuition is mostly the same as the original proofs of these results. Nevertheless, we give what we believe to be more elementary proofs, and this allows us to improve some of the previous results.

1 Introduction

Yao introduced the model of communication complexity in 1979 [Yao79]. It has since become a central model of computation, studied for itself as well as for its numerous applications. The model addresses problems whose inputs are shared among different players, who have to communicate in order to solve it.

In complexity theory, a central question is to prove lower bounds. In this context, we wish to determine, for a given communication problem, how many bits the players have to exchange in order to solve it. A simple answer is that the messages should contain at least enough information to solve the problem. For

*LRI, Université Paris-Sud XI, 91405 Orsay CEDEX, FRANCE. Email: {kaplan,laplante}@lri.fr

example, they need to distinguish between inputs that produce different outputs. This idea has led to many lower bound techniques, and in particular to proofs involving information theory. Shannon's information theory's original purpose was to study communication problems [Sha48], so it seems natural that these techniques found many applications in the field of communication complexity.

Information complexity is a general lower bound method [BYJKS04], but in many other cases, ad hoc proofs have been given for specific problems [BYJKS02, JKS03, BYJK08]. One of the appealing features of these proofs is the way they capture the intuition of the hardness of the problem very naturally. However, by using elaborate results in information theory which in turn are based on statistics and probability, the essential mechanics of the proof is not always so readily apparent. More recently, the subdistribution method [JKN08] was introduced as a relaxation of the rectangle or corruption bound [BPSW06].

The intuition that arises from both Kolmogorov complexity and information theory is often close, but they differ in their underlying mechanics: where information theory uses statistics, Kolmogorov complexity uses combinatorics. It is not clear that there is a general way to rephrase the elements from statistics into the language of combinatorics. One of our goals is to use Kolmogorov complexity to capture the intuition of the information theoretic approach, while bringing out the combinatorial nature of these proofs.

The use of Kolmogorov complexity in lower bound techniques has proven to be useful in randomized and quantum query complexity lower bounds [LM08]. In a surprising turn of events, this information-based approach turned out [ŠS06] to be dual (in the sense of semi-definite programming) to the spectral lower bound method of Barnum, Saks, and Szegedy [BSS03], and hence equivalent to a wide family of adversary lower bounds, which imply lower bounds on randomized as well as quantum query complexity.

Duality of linear programming has also been a cornerstone of lower bounds in randomized communication complexity, in the form of Yao's minmax theorem [Yao83]. Since the randomized case is not easy to handle, Yao's theorem reduces it to the deterministic case. Most lower bounds in randomized communication complexity use this theorem. This is true of most combinatorial lower bounds, as well as many proofs using information theory. A secondary goal in this work is to use Kolmogorov complexity for this reduction.

The main tool from Kolmogorov complexity that we use is incompressibility. It allows us to identify worst case inputs, those that require a large amount of communication. The second is mutual information. It gives us a general expression on the amount of information that the player must exchange. The formulation of our general lower bound is very similar to the information complexity method [BYJKS04]. Most of the techniques from Kolmogorov complexity that we use here are proved using elementary counting arguments. Moreover, our general method is an extension to communication complexity of the well known incompressibility technique, used to prove lower bounds for various resources, including time complexity [LV93], average-case complexity, and communication complexity [BJLV00, KS92].

We prove a version of Yao's theorem based on Kolmogorov complexity, which allows us to restrict the random choices to a single incompressible string. By choosing both the worst case input and the random choices of the algorithm to be incompressible, we get the advantage of having them be independent of one another, which tends to simplify the proofs.

One of the main open problems in quantum communication complexity is to show an exponential gap between classical and quantum communication complexity, for a total function. The hidden matching relation was introduced in [BYJK08] to exhibit such a gap. However, this problem falls short of this goal since the problem is a relation. More recently, it has been proved that the gap holds for a partial function [GKK$^+$07], but the question remains open for total functions.

Linial and Shraibman's work on randomized and quantum communication complexity [LS07], and recent subsequent work, such as [LS08, LSS08], can be viewed as mounting evidence that there is at most a polynomial gap between classical and quantum communication complexity, for total functions. Indeed,

their method generalizes most of the previously known lower bound techniques, including discrepancy, trace norm [Raz03], and some Fourier based techniques [Raz95]; and these techniques all extend to the quantum setting.

A fundamental question is where information theoretic techniques stand with respect to the factorization norm and related bounds, and whether they yield lower bounds for quantum communication. Up until now, it remains open whether these techniques are related. We show that our method based on Kolmogorov complexity is related to the corruption [BPSW06] and subdistribution bounds [JKN08].

Perhaps a key to strong separations lies in the essential use of Yao's minmax principle. Although some efforts have been made, no equivalent of Yao's principle is known for quantum complexity [dGdW02]. It is natural to surmise that an approach using Yao's principle, or its Kolmogorov alternative, could be a good candidate to prove large separations between randomized and quantum communication complexity.

We apply our method to the hidden matching problem [BYJK08]. We believe our proof is simpler, since by choosing the protocol's random choices and hard instance to be Kolmogorov random and independent, we avoid the iterative construction of the worst case distribution on the inputs of the original proof.

We also prove that our technique generalizes the VC dimension [KNR99] and shatter coefficient lower bound [BYJKS02]. Kolmogorov complexity turns out to be a very good tool in this case, since it highlights very nicely the combinatorial nature of the proof. We go further and give a second proof entirely without Kolmogorov complexity, which only requires a simple counting argument.

Finally, we use combinatorial techniques to compare one-way and simultaneous communication in the multi-player setting. Again, by first casting the proof in the language of Kolmogorov complexity, a purely combinatorial argument emerged. The result was previously known [BYJKS02], but we significantly improve the error dependence.

2 Preliminaries

2.1 Communication complexity

Let X, Y and Z be finite sets and fix a function $f : X \times Y \to Z$. In the communication complexity model, two players, Alice and Bob, each receive one input, and their goal is to compute f. Neither of them sees the other player's input. To perform this task, they have to communicate. At the end, Bob has to output the value of the function.

Messages are sent according to a shared communication protocol. The cost of the protocol is measured by the sum of messages' length (in the worst case). The communication complexity of the function f is the cost of the best protocol that computes f. We denote it $D(f)$. Notice that, as we are only interested in the communication between the players, we can assume that they have unlimited computational power.

This model has many variations, and among all, we will consider these in particular:

- *One way communication:* In this model, Alice sends a single message to Bob. We denote the one-way communication complexity $D^{A\to B}(f)$.

- *Simultaneous messages:* In this model, Alice and Bob each send a single message to a referee, who outputs $f(x, y)$. We denote the simultaneous messages communication complexity $D^{A||B}(f)$.

- *Communication complexity of relations:* The problem Alice and Bob are solving is relational. Let $\mathcal{R} \subseteq X \times Y \times Z$. Alice receives $x \in X$ and Bob $y \in Y$. Bob has to output any z such that $(x, y, z) \in \mathcal{R}$.

One important notion is the *transcript* of the protocol on input (x, y). This is the concatenation of the messages sent by Alice and Bob when they receive inputs x and y. We assume for simplicity that the protocol has the property that the length of the messages in each round depends only on the round, and the length of the inputs. (We may always pad the messages so that this holds.) For one-way communication, the transcript is just the message sent by Alice to Bob.

An monochromatic rectangle for f is a set $R = S \times T$ with $S \subseteq X$ and $T \subseteq Y$ such that there exists $b \in \{0, 1\}$ and for all $(x, y) \in R$, $f(x, y) = b$. A classical result in communication complexity states that a deterministic protocol partitions the set of inputs into monochromatic rectangles, each rectangle corresponding to a transcript of the protocol [KN97]. We will also consider approximately monochromatic rectangles. Let μ be a probability distribution over $X \times Y$ and $\varepsilon > 0$. A rectangle is called (μ, ε)-monochromatic for f if there exists a $b \in \{0, 1\}$ such that $\mu(\{(x, y) \in R | f(x, y) = b\}) \geq (1 - \varepsilon)\mu(R)$.

More importantly, we will be interested in the probabilistic version of communication complexity. In this model, Alice and Bob can toss coins. The output of the protocol is now probabilistic, and we restrict ourselves to protocols that make few errors.

Definition 1. *Let* $0<\varepsilon<1$. *A probabilistic communication protocol* $\mathcal{P}$ *is* ε*-correct if for all* $(x, y) \in X \times Y$,

$$Prob(\mathcal{P}(x, y) \neq f(x, y)) < \varepsilon$$

where the probability is taken over the randomness of $\mathcal{P}$.

The randomized communication complexity is the cost of the best probabilistic ε-correct protocol that computes f, and is denoted by $R_\varepsilon(f)$. Usually, we will need to consider the randomness used in communication protocols explicitly. More precisely, we will assume that before the execution of the protocol, each player receives a random string r_A and r_B from sets R_A, $R_B \subseteq \{0, 1\}^*$. If the randomness is shared, then $R_A = R_B$ and $r_A = r_B$. We denote by $R_\varepsilon^{pub}(f)$ the randomized communication complexity with shared randomness.

We also consider the distributional model. In this model, protocols are deterministic, but they can make errors on some inputs.

Definition 2. *Let* $0<\varepsilon<1$ *and* μ *a distribution over the inputs* $X \times Y$. *A distributional communication protocol* $\mathcal{P}$ *is* (μ, ε)*-correct if:*

$$Prob_\mu(\mathcal{P}(x, y) \neq f(x, y)) < \varepsilon$$

We denote by $D_\varepsilon^\mu(f)$ the cost of best distributional (μ, ε)-correct protocol that makes error on at most a fraction ε of the inputs (according to μ). The distributional communication complexity $D_\varepsilon(f)$ is $max_\mu D_\varepsilon^\mu(f)$. We will consider the special case where μ ranges over rectangular (or product) distributions only. These are distributions μ over $X \times Y$ such that $\mu = \mu_1 \otimes \mu_2$ where μ_1 is a distribution over X and μ_2 a distribution over Y. In this special case, we denote the communication complexity by $D_\varepsilon^{\square}(f)$. In the general case, Yao's minmax theorem states that distributional communication complexity is equivalent to randomized communication complexity with shared randomness. The proof of the minmax theorem rests essentially on the duality of linear programming.

Theorem 1. *[Yao83] for any function* $f : X \times Y \to \{0, 1\}$ *and* $\varepsilon > 0$*:*

$$D_\varepsilon(f) = R_\varepsilon^{pub}(f).$$

2.2 Kolmogorov Complexity

We first recall the basic definitions of Kolmogorov complexity. We give the definition of prefix free complexity, as we won't use plain complexity [LV93].

Definition 3. *A set of strings is called prefix free if no string in the set is a prefix of another.*

Definition 4. *Let φ be a universal Turing machine and $\mathbb{P}$ a prefix free set. The prefix free Kolmogorov complexity of a string x given y with respect to $\varphi, \mathbb{P}$ is $K_\varphi(x|y) = \min\{|p| : p \in \mathbb{P} \, and \, \varphi(p, y) = x\}$.*

If θ is the empty string, we just write $K_\varphi(x)$ for $K_\varphi(x|\theta)$. Also, in the rest of the paper, we fix a universal Turing machine φ, a prefix free set $\mathbb{P}$, and write K instead of K_φ.

We now recall some properties that we use extensively in the rest of the paper. In the first proposition, we use the following basic program to compute any $x \in X$: give the index of x in X.

Proposition 1. *[LV93] For any computable set X and string σ, there exists a constant c such that for all $x \in X$ $K(x|\sigma) \leq \log |X| + c$.*

The next proposition shows that this coding is almost optimal.

Proposition 2. *[LV93] For any set X and string σ, there exists an element $x \in X$ such that $K(x|\sigma) \geq \log |X|$. Such elements are called* incompressible.

A basic manipulation leads to following proposition:

Proposition 3. *There exists a constant c such that, for all x,y, σ:*

$$K(x|\sigma) \leq K(x|y, \sigma) + K(y) + c$$

Proof. On the right hand side, we consider the following program to compute x. First compute y and then use y to compute x. It is certainly at least as long as the shortest program that computes x (the left hand side). The constant c is the cost of the simulation of the program by our universal Turing machine φ. However, this constant can be made equal to zero by "hiding" the instructions in the string σ. Therefore, in the rest of this paper, we will omit the constant. □

Corollary 1. *Let X and Y be two finite sets. For $x \in X$ and $y \in Y$ and for all σ, if $K(x, y|\sigma) \geq \log |X| + \log |Y|$ then both $K(x|y, \sigma) \geq \log |X|$ and $K(y|x, \sigma) \geq \log |Y|$.*

Strings x, y verifying the premise of the above corollary are said to be independent Kolmogorov-incompressible strings. We will need the following proposition to analyze the asymptotic behavior of some components of our proofs.

Proposition 4. *[KN97] Let $n \in \mathbb{N}$ and $\varepsilon \in]0, 1[$.*

$$\log \binom{n}{\lceil \varepsilon n \rceil} \sim n H_2(\varepsilon)$$

where $H_2(\varepsilon)$ is the entropy of a random variable following a Bernoulli distribution with parameter ε.

Given a distribution on strings, they can be coded using the Shannon-Fano code. The next proposition shows how this translates to Kolmogorov complexity. Proposition 6 shows that this coding is also essentially optimal.

Proposition 5. *[LV93] Fix a finite set X and a probability distribution μ over X. There exists a constant c such that for all $\sigma \in \{0,1\}^*$, and for all $x \in X$ such that $\mu(x) \neq 0$, $K(x|\sigma) \leq \log(\frac{1}{\mu(x)}) + c$.*

Proposition 6. *[LV93] Fix a finite set X and a probability distribution μ over X. For all $\sigma \in \{0,1\}^*$, there exists $x \in X$ such that $\mu(x) \neq 0$ and $K(x|\sigma) \geq \log(\frac{1}{\mu(x)})$.*

3 Lower Bounds

3.1 Main theorem in the deterministic case

In this section, we give a general lower bound on communication complexity. Although we do not use this form directly, it provides the main intuition. The lower bound is in terms of Kolmogorov complexity. More precisely, it uses mutual information [LV93] between an input of the problem and the transcript of the communication protocol. The mutual information between x and y is $K(x) - K(x|y)$, which can be interpreted as how much information about x is gained when y is given, compared to when it is not given. It is a measure of the information that y contains on x.

Theorem 2. *Fix $f : X \times Y \to \{0,1\}$ and $\mathcal{P}$ an optimal deterministic protocol for f. Denote by $T(x,y)$ the transcript of $\mathcal{P}$ on input (x,y).*

$$\forall \sigma \in \{0,1\}^*,\ D(f) \geq \max_{(x,y)\in X\times Y} K(x,y|\sigma) - K(x,y|T(x,y),\sigma)$$

Proof. Fix $(x,y) \in X \times Y$. Using proposition 3: $K(x,y|\sigma) \leq K(x,y|T(x,y),\sigma) + K(T(x,y))$. Moreover, using Proposition 1, $K(T(x,y)) \leq |T(x,y)| \leq D(f)$. Finally, we get $K(x,y|\sigma) - K((x,y)|T(x,y),\sigma) \leq D(f)$. □

As an application, we prove that Theorem 2 generalizes the corruption lower bound [BPSW06]. This bound is a lower bound on distributional complexity. Nevertheless, as a distributional protocol is deterministic, Theorem 2 suffices to handle this setting.

Definition 5. *For a function $f : X \times Y \to \{0,1\}$, a distribution μ over $X \times Y$ and $\varepsilon > 0$, define:*

$$mono_\mu(f,\varepsilon) = max\{\mu(S) \mid S \text{ is a } (\mu,\varepsilon)\text{-monochromatic rectangle for } f\}$$

Theorem 3 ([BPSW06])**.** *For a function $f : X \times Y \to \{0,1\}$, a distribution μ over $X \times Y$ and $1/2 > \varepsilon > 0$:*

$$D^\mu_\varepsilon(f) \geq \log \frac{1}{mono_\mu(f, 2\varepsilon)}$$

Proof. Fix an (ε,μ)-correct protocol $\mathcal{P}$ for f, and according to Proposition 6, let (x^*,y^*) be a pair of inputs such that $K(x^*,y^*|\mu,\mathcal{P},f) \geq \log \frac{1}{\mu(x^*,y^*)}$. Recall that $\mathcal{P}$ induces a partition $\mathcal{R}$ of the input into rectangle [KN97]. For $S \subseteq X \times Y$, define $Err(S) = \{(x,y) \in S | \mathcal{P}(x,y) \neq f(x,y)\}$. Denote $\tilde{\mathcal{R}} = \{S \in \mathcal{R} | \mu(Err(S)) > 2\varepsilon\mu(S)\}$. Let $E = \bigcup_{S\in\tilde{\mathcal{R}}} S$ be the inputs in non-2ϵ monochromatic rectangles.

First, we prove that $(x^*,y^*) \notin E$. Notice that $\mu(E) = \sum_{S\in\tilde{\mathcal{R}}} \mu(S) \leq 1/2$, otherwise $\mu(Err(X \times Y)) > \varepsilon$, which contradicts the correctness of the protocol. Suppose that $(x^*,y^*) \in E$. One can encode (x^*,y^*) by giving an index in E. Using the probability distribution induced by μ on E and a Sannon-Fano code as defined in Proposition 5, we get:

$$\log \frac{1}{\mu(x^*,y^*)} \leq K(x^*,y^*) \leq \log \frac{\mu(E)}{\mu(x^*,y^*)} \leq \log \frac{1}{2\mu(x^*,y^*)},$$

a contradiction.

Since $(x^*, y^*) \notin E$, the transcript $T = T(x^*, y^*)$ determines a rectangle R such that $\mu(Err(R)) < 2\varepsilon$. By definition, $\mu(R) \leq mono_\mu(f, 2\varepsilon)$. Given T, one can encode (x^*, y^*) by giving its index in R, using the probability distribution induced by μ on R and the Sannon-Fano code, from Proposition 5. Therefore,

$$K(x^*, y^* | \mu, \mathcal{P}, f, T) \leq \log \frac{\mu(R)}{\mu(x^*, y^*)} \leq \log \frac{mono_\mu(f, 2\varepsilon)}{\mu(x^*, y^*)}.$$

Using Theorem 2 with $\sigma = (\mu, \mathcal{P}, f)$, we get $D^\mu_\varepsilon(f) \geq \log \frac{1}{mono_\mu(f, 2\varepsilon)}$, as claimed. □

3.2 The randomized case: a Kolmogorov alternative to Yao's min-max principle

In the following lemma, we show how to derive a deterministic protocol from a randomized one, with the same complexity and performance in terms of errors. Recall that in the communication complexity model, Alice and Bob have full computational power. In particular, the players can choose an incompressible string in advance (which is in general not computable), and simulate a randomized protocol $\mathcal{P}$ using this string for randomness.

We denote by $\mathcal{P}^{r_A, r_B}$ the deterministic protocol obtained by executing a randomized protocol $\mathcal{P}$ with fixed random strings (r_A, r_B). This protocol certainly makes errors for some inputs, but the next lemma shows that using incompressible strings, the distribution of errors in the resulting protocol has good properties. The existence of a Kolmogorov random string with these good properties is proved using the pigeonhole principle.

Lemma 1. *Let $\mathcal{P}$ be an ε-correct randomized protocol for $f : X \times Y \to \{0, 1\}$ and μ a probability distribution on $X \times Y$. For all $S \subseteq X \times Y$, we define $Err_{r_A, r_B}(S) = \{(x, y) \in S : \mathcal{P}^{r_A, r_B}(x, y) \neq f(x, y)\}$. Fix r^*_Aet r^*_B such that $K(r^*_A, r^*_B | \mu, P, S) \geq \log(|R_A||R_B|)$. Then $\mu(Err_{r^*_A, r^*_B}(S)) \leq 2\varepsilon\mu(S)$.*

Proof. Let $\tilde{R}$ denote the bad random strings: $\tilde{R} = \{\mu(r_A, r_B) : \mu(Err_{r_A, r_B}(S)) > 2\varepsilon\mu(S)\}$. We will prove that $|\tilde{R}| < \frac{|R_A||R_B|}{2}$. This is sufficient to conclude that $(r^*_A, r^*_B) \notin \tilde{R}$; otherwise, one could compute it by giving an index in $\tilde{R}$, which contradicts the assumption $K(r^*_A, r^*_B | \mu, P, S) \geq \log(|R_A||R_B|)$.

$\mathcal{P}$ being ε-correct, by summing over $R_A \times R_B$:

$$\sum_{r_A, r_B} \mu(Err_{r_A, r_B}(S)) \leq |R_A||R_B|\varepsilon\mu(S)$$

on the other hand:

$$\sum_{r_A, r_B} \mu(Err_{r_A, r_B}(S)) \geq \sum_{\tilde{R}} \mu(Err_{r_A, r_B}(S)) > 2\varepsilon\mu(S)|\tilde{R}|.$$

Combining the two inequalities, we have: $|\tilde{R}| < \frac{|R_A||R_B|}{2}$. □

This proof is (not surprisingly) very similar to the proof of the upper bound in Yao's minmax theorem. What we gain by using the Kolmogorov alternative is that we do not require distributional complexity. In distributional complexity, we have to analyse the behavior of deterministic protocols with respect to a distribution μ over the inputs. In our case, by choosing a single random string, and an independent Kolmogorov random hard instance, we analyze a deterministic algorithm acting on a single input. The hard

instance can be chosen as incompressible with respect to some hard distribution by applying optimality of the Shannon Fano Code 6; but we are free choose it in any other way.

By definition, a randomized protocol is a distribution over deterministic protocols. Moreover, the cost of a randomized protocol is the cost of the most expensive protocol in the support of this distribution. We can now state the randomized version of Theorem 2.

Theorem 4. *Fix $f : X \times Y \to \{0,1\}$ and $\mathcal{P}$ an optimal randomized ε-correct protocol for f. If $T(x,y,r_A,r_B)$ is the transcript of $\mathcal{P}^{r_A,r_B}$ on input (x,y), then for all $S \subseteq X \times Y$, $(r_A,r_B) \in R_A \times R_B$ and $\sigma \in \{0,1\}^*$:*

$$R_\varepsilon(f) \geq \max_{(x,y)\in S} K(x,y|\sigma) - K(x,y|T(x,y,r_A,r_B),\sigma)$$

Proof. Fix r_A and r_B in protocol $\mathcal{P}$. By definition, $R_\varepsilon(f) \geq |T(x,y,r_A,r_B)|$. Using Proposition 1, we get $|T(x,y,r_A,r_B)| \geq K((Tx,y,r_A,r_B)|\sigma)$. Using Proposition 3, we get $K(x,y|\sigma) \leq K(x,y|T(x,y,r_A,r_B),\sigma)+ K(T(x,y,r_A,r_B))$. As in Theorem 2, combining both: $R_\varepsilon(f) \geq K(x,y|\sigma) - K(x,y|T(x,y,r_A,r_B),\sigma)$. □

4 Applications

4.1 The Hidden matching problem

In this section, we study the communication complexity of the hidden matching problem. This relation was introduced to show a gap between randomized and quantum communication complexity [BYJK08]. The following theorem is the randomized lower bound for the hidden matching problem.

Definition 6. *The Hidden Matching problem $HM_n(x,M)$ is defined as follows:*

- *Alice receives a string $x \in \{0,1\}^n$.*
- *Bob receives a matching M on n vertices.*
- *Bob outputs a triple (i,j,b) such that $x_i \oplus x_j = b$ and $(i,j) \in M$*

Theorem 5. *[BYJK08] For one-way randomized communication complexity:*

$$R_\varepsilon^{A\to B}(HM_n) \geq \Omega(\sqrt{n})$$

Proof. Fix an ε-correct protocol $\mathcal{P}$ for HM_n. Let $\mathcal{M}$ be a set of n pairwise independent matchings. For example $M_i = \{(k, k+i \mod n), k = 0 \dots n-1\}$ for $i = 0 \dots n-1$. We will pick an incompressible subset of $\mathcal{M}$ of size $\sqrt{n}$ among all subsets of size $\sqrt{n}$. By definition, the complexity of such a subset is at least $\log \binom{n}{\sqrt{n}}$. More precisely, pick $x^* \in X$, r_A^*, r_B^* and $\mathcal{M}'$ such that: $K(x^*,\mathcal{M}',r_A^*,r_B^*|f,P) \geq \log|X| + \log\binom{n}{\sqrt{n}} + \log|R_A| + \log|R_B|$.

By Corollary 1 and Lemma 1, we have $|Err_{r_A^*,r_B^*}(\{x^*\}\times\mathcal{M}')| < 2\varepsilon\sqrt{n}$. By definition, Bob's output on input x and M yields an equation $x_i \oplus x_j = b$. We will prove that the set of equations obtained by running the protocol on each matching of $\mathcal{M}'$ has dimension at least $\Omega(\sqrt{n})$.

First, denote by $E(\mathcal{M},x)$ the set of edges obtained by running the protocol $\mathcal{P}^{r_A^*,r_B^*}$ on x^* and each $M \in \mathcal{M}$. Denote by G the graph generated by $E(\mathcal{M},x)$. As G has n edges, it has a cycle free subgraph with at least $\sqrt{n}$ edges, since it has at least $\sqrt{n}$ non-isolated vertices and it suffices to take a spanning forest over these vertices. Therefore, running $\mathcal{P}$ on x^* and each $M \in \mathcal{M}$ yields at least $\sqrt{n}$ independent equations. But to conclude, we have to prove that $\mathcal{M}'$ also yields at least $\sqrt{n}$ independent equations. It is sufficient

to prove that $E(\mathcal{M}', x^*)$ generates a graph with at least $\sqrt{n}$ vertices. We prove in the appendix a stronger statement that implies the following proposition. Fix a graph with n edges. Pick $\sqrt{n}$ edges at random, uniformly among all sets of $\sqrt{n}$ edges, and call H the subgraph generated by these edges. Then there exists a constant c such that for a sufficiently large n, $Prob(|V(H)| < c\sqrt{n}) < 1/2$.

Returning to our graph G, let $\mathcal{B} = \{H \subset G : |E(H)| = \sqrt{n} \text{ and } |V(H)| < c\sqrt{n}\}$. By the previous argument, we know that $|\mathcal{B}| < \binom{n}{\sqrt{n}}/2$. Let H be the graph generated by $E(\mathcal{M}', x^*)$. We can conclude that $H \notin \mathcal{B}$, otherwise one could describe H by giving its index in $\mathcal{B}$, which contradicts the incompressibility assumption.

Notice that since the protocol is one-way, the transcript only depends on Alice's input and randomness. Let $T(x, r_A)$ be the transcript of protocol $\mathcal{P}^{r_A, r_B}$ on input x. Choosing x^* incompressible, we have a lower bound on $K(x^*|r_A^*, r_B^*, \mathcal{M}', \mathcal{P}, f)$. Now we need an upper bound on the complexity of x^* knowing its transcript. To do that, we design an algorithm that computes x^* knowing $T(x^*, r_A^*), \mathcal{M}', r_B^*$:

1. Simulate $\mathcal{P}^{r_A^*, r_B^*}$ with message $T(x^*, r_A^*)$ for every $M \in \mathcal{M}'$.
2. Correct the errors in $\{x^*\} \times \mathcal{M}'$. The set of errors is given as an auxiliary input.
3. Solve the system of equations. As this system is of dimension at least $c\sqrt{n}$, this gives at least $c\sqrt{n}$ coordinates of x^*.
4. The remaining $n - c\sqrt{n}$ coordinates of x^* are given as an input of the program.

The program uses $\log \binom{\sqrt{n}}{2\varepsilon\sqrt{n}}$ input bits to correct the errors. By Proposition 4, for any $\delta > 0$ and n sufficiently large, this is less than $(1+\delta)\sqrt{n}H_2(2\varepsilon)$. We also need $n - c\sqrt{n}$ bits to solve the whole system. This implies $K(x^*|T(x^*, r_A^*), r_A^*, r_B^*, \mathcal{M}', \mathcal{P}, f,) < n - (c - (1+\delta)H_2(2\varepsilon))\sqrt{n}$. Finally, using theorem 4, we get:

$$R_\varepsilon^{A\to B}(HM_n) \geq (c - (1+\delta)H_2(2\varepsilon))\sqrt{n},$$

which concludes the proof.

□

4.2 VC dimension and shatter coefficients lower bounds

In this section, we consider a general lower bound on one-way communication complexity. This lower bound was previously proved using combinatorial techniques [KNR99] and later re-proved and extended using information theory techniques [BYJKS02]. Our proof uses only elementary counting arguments.

To any function $f : X \times Y \to \{0, 1\}$, we associate the communication matrix M_f. Its size is $|X||Y|$ and it verifies $M_f(x, y) = f(x, y)$. We identify M_f or any sub-matrix and the set of its rows, which we think of as boolean strings. For M_f, it is a set of $|X|$ strings of length $|Y|$. The VC dimension of M_f is the size of the largest set $Y_0 \subseteq Y$ such that there exist some $X_0 \subseteq X$ of size $2^{|Y_0|}$ and $M_f|_{X_0,Y_0}$ is the set of all strings of length $|Y_0|$.

The following definition generalizes the VC dimension. For $l \geq VC(M_f)$, the l-th shatter coefficient of M_f, denoted by $SC(l, M_f)$, is the size of the largest set $X_0 \subseteq X$ such that there exists some $Y_0 \subseteq Y$ of size l and all rows of $M_f|_{X_0,Y_0}$ are different. A witness for $SC(l, M_F)$ is a set $S \subseteq X \times Y$ such that if $S{=}U \times V$, $|V|{=}l$ and $|U|{=}SC(l, M_f)$ and all rows in S are different.

Theorem 6. *[KNR99, BYJKS02] For every function $f : X \times Y \to \{0, 1\}$, there exists a constants $c > 0$ such that:*

$$R_\varepsilon^{X\to Y}(f) \geq VC(M_f)(1 - (1+c)H_2(2\varepsilon))$$

and for every $l > VC(M_f)$

$$R_\varepsilon^{X\to Y}(f) \geq \log(SC(M_f, l)) - l(1+c)H_2(2\varepsilon)$$

Kolmogorov complexity version. Notice that $\log(SC(l, M_f))$=$VC(M_f)$ for l=$VC(M_f)$. Therefore, we just have to prove the second point. Fix an optimal ε-correct randomized one-way protocol $\mathcal{P}$ for f. Let $S = U \times V$ be a witness for $SC(l, M_f)$. Pick $x^* \in U$, and $(r_A^*, r_B^*) \in R_A \times R_B$ such that $K(x^*, r_A^*, r_B^* | f, \mathcal{P}, S) \geq \log|U| + \log|R_A| + \log|R_B|$. By Corollary 1, $K(r_A^*, r_B^* | f, \mathcal{P}, S, x^*) \geq \log|R_A| + \log|R_B|$ and $K(x^* | r_A^*, r_B^* f, \mathcal{P}, S) \geq \log|U|$. Let $S' = \{x^*\} \times V$. As this set is computable knowing x^* and S, we apply Lemma 1 to get $|Err_{r_A^*, r_B^*}(\{x^*\} \times V)| < 2\varepsilon|\{x^*\} \times V|$.

Let $T(x, r_A)$ denote the transcript of the protocol $\mathcal{P}^{r_A, r_B}$ on input x. We define an algorithm that computes any $x \in U$ knowing $T(x, r_A^*)$ and r_B^*.

1. Simulate $\mathcal{P}^{r_A^*, r_B^*}(x, y)$ using $T(x, r_A^*)$ for every $y \in Y$.
2. Correct the errors in $\{x\} \times V$. The set of errors is given as an auxiliary input of the program.
3. Compare the obtained row with every row of S. As they are all different, only one corresponds to x.

This program uses $\log\binom{l}{2\varepsilon l}$ input bits to describe the set of errors. By Proposition 4, this is asymptotically equivalent to $lH_2(2\varepsilon)$, and there exists a constant c such that $K(x^* | T(x^*, r_A^*), r_B^*) \leq l(1+c)H_2(2\varepsilon)$. Finally, applying Theorem 4:

$$R_\varepsilon^{X\to Y}(f) \geq \log|U| - l(1+c)H_2(2\varepsilon)$$

Notice that $|U| = SC(l, M_f)$ and this concludes the proof. □

From this Kolmogorov based proof, we may now derive a purely combinatorial proof. We use Yao's minmax theorem to reduce the problem to the deterministic case.

Counting version. Let $S = U \times V$ be a witness for $SC(l, M_f)$. By definition, S has $SC(l, M_f)$ rows and l columns. Let μ be the uniform distribution on S:

$$\mu(x, y) = \begin{cases} 1/|S| & \text{if } (x, y) \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Using Yao's minmax theorem (Theorem 1), we know that $R_\varepsilon(f) \geq D_\varepsilon^\mu(f)$. Let $\mathcal{P}$ be an optimal one-way (μ, ε)-correct protocol for f. Let $U' \subseteq U$ be the set of rows of S on which the protocol makes at most $2\varepsilon l$ errors. Notice that since $\mathcal{P}$ makes at most $\varepsilon|S| = \varepsilon l 2^l$ errors on inputs in S, we get by summing over all rows of S that $|U'| \geq |U|/2$. Define $S' = U' \times V$.

Let $D(\mathcal{P})$ be the cost of $\mathcal{P}$ and $\binom{V}{\leq k}$ be the set of subsets of V of size at most k. We now define the following mapping γ:

$$\begin{array}{rcl} \gamma : U' & \to & \{0,1\}^{D(\mathcal{P})} \times \dbinom{V}{\leq 2\varepsilon l} \\ x & \mapsto & (\sigma, E) \end{array}$$

γ maps an input $x \in U'$ to its transcript according to protocol $\mathcal{P}$ and the set of errors made by $\mathcal{P}$. We claim that γ is injective, which implies that $|U'| \leq 2^{D(\mathcal{P})} 2\varepsilon l \binom{l}{2\varepsilon l}$. Applying $\log$ on both sides, and using Proposition 4, there exists a constant c such that:

$$D_\varepsilon^\mu(f_S) = D(\mathcal{P}) \geq \log|U'| - l(1+c)H_2(2\varepsilon) \geq \log|U| - l(1+c)H_2(2\varepsilon) - 1$$

Notice that $|U| = SC(l, M_f)$ and this concludes the proof. It only remains to prove that γ is injective. Fix x_1 and x_2 in U' and suppose $\gamma(x_1) = \gamma(x_2) = (\sigma, E)$. Using σ, we can simulate $\mathcal{P}$ for each input $y \in V$. Now, flip the result for all $y \in E$. As E is the set of errors made by $\mathcal{P}$, we get a row of the matrix $M_f|_S$. As all rows of $M_f|_S$ are different, we conclude $x_1 = x_2$. □

5 One way versus simultaneous messages

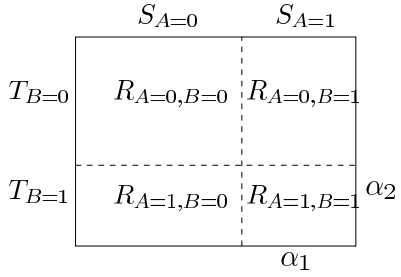
In this section, we consider multiparty number in hand communication complexity. In this model, n players have to compute an n-variable function. A simultaneous message protocol works as follows: every player receives an input and sends a message to a referee who has to output the value of the function. Fix $f : X_1 \times \cdots \times X_n \to \{0,1\}$, then $D^{X_1||\cdots||X_n}(f)$ denotes the simultaneous communication complexity of f.

The next result compares this model to the complexity of one-way protocols. For $f : X_1 \times \cdots \times X_k \to \{0,1\}$, we consider two-player protocols in which one player receives one variable, say $x_i \in X_i$, and the other one receives all the other variables. The communication complexity for such protocols is denoted by $D^{X_i \to X_{-i}}(f)$. Of course, these protocols can be deterministic, randomized or distributional. Here, we compare the distributional versions of simultaneous and one-way communication, restricted to rectangular distributions. We improve the previous bound [BYJKS02] on the error probability, from $\sum_i H_2(\varepsilon_i)$ to $\sum_i \varepsilon_i$.

Theorem 7. *[BYJKS02] Fix $f : X_1 \times \cdots \times X_n \to \{0,1\}$. Then for $\varepsilon \geq (1 + 1/n) \sum_{i=1}^n \varepsilon_i$:*

$$D^{[],X_1||\cdots||X_n}_{\varepsilon}(f) \leq \sum_{i=1}^{n} D^{[],X_i \to X_{-i}}_{\varepsilon_i}(f).$$

Proof. The proof is by induction on the number on player. Let $X_1 = X$ and $X_2 = Y$, and fix a product distribution $\mu = \mu_1 \otimes \mu_2$ over $X \times Y$, and two one-way protocols $\mathcal{P}_1$ and $\mathcal{P}_2$. In the first protocol, player 1 sends a single message to player 2, and conversely in the second protocol. The messages sent by players in these protocols are denoted $m_1(x)$ and $m_2(y)$ for inputs $x \in X$ and $y \in Y$. These messages correspond to sets $S \in X$ and $T \in Y$; S is the set of inputs on which the message $m_A(x)$ is sent and T on which the message is $m_B(y)$. The following diagram shows how we decompose the rectangle $R = S \times T$.



- $S_{A=c}$=$\{x \in S : \mathcal{P}_2$ outputs c on input x and message $m_B(y)\}$
- $T_{B=d}$=$\{y \in t : \mathcal{P}_1$ outputs d on input y and message $m_A(x)\}$
- $R_{A=c,B=d} = S_{A=c} \times T_{B=d}$

The simultaneous protocol works as follows: Alice sends $m_A(x)$ and Bob $m_B(y)$ to the referee. These messages correspond to a rectangular set of inputs $R = S \times T$. The referee simulates the one-way protocols for every input in R, and answers 0 if $\mu(R_{A=0,B=0}) > \mu(R_{A=1,B=1})$ and 1 otherwise. The complexity of this protocol is at most the sum of the costs of $\mathcal{P}_1$ and $\mathcal{P}_2$, so we only have to analyze the errors of the simultaneous protocol. Suppose without loss of generality that the referee outputs 0 on R.

- For inputs in $R_{A=0,B=0}$, $\mathcal{P}_1$ and $\mathcal{P}_2$ output the same answer. If the simultaneous protocol makes an error for some input, then both $\mathcal{P}_1$ and $\mathcal{P}_2$ were wrong.

- In $R_{A=0,B=1}$ and $R_{A=1,B=0}$, $\mathcal{P}_1$ and $\mathcal{P}_2$ output different answers, such that for each input, exactly one protocol makes an error.
- In $R_{A=1,B=1}$, both $\mathcal{P}_1$ and $\mathcal{P}_2$ output the answer 1. As the simultaneous protocol outputs the value 0, it can make an error where neither of the one-way protocols did.

Except in the last set, if the simultaneous protocol is wrong for some input, then at least one of the one-way protocols was already making an error. The last set is the only one in which new errors appear. Therefore, to conclude in the two-player case, it is sufficient to prove that $\mu(R_{A=1,B=1})$ is at most the average of $\mu(R_{A=0,B=1})$ and $\mu(R_{A=1,B=0})$.

Let μ' be the measure induced by μ on R. As μ' is also rectangular, let $\mu' = \mu'_1 \otimes \mu'_2$ and $\alpha_1 = \mu'_1(T_{B=1})$ and $\alpha_2 = \mu'_2(S_{A=1})$. We claim that $\alpha_1\alpha_2 < \frac{1}{2}[\alpha_1(1-\alpha_2)+\alpha_2(1-\alpha_1)]$. The left hand side is the measure of $R_{A=1,B=1}$, whereas the right hand side is the average of measures of $R_{A=1,B=0}$ and $R_{A=0,B=1}$. Summing over all rectangles, this proves that the total measure of errors created in the simultaneous protocol is at most the average of errors made by the one-way protocols, which concludes the proof. We now prove our claim.

By definition of the simultaneous protocol, we have $\alpha_1\alpha_2 \leq (1-\alpha_1)(1-\alpha_2)$. We can suppose without loss of generality that $\alpha_1 \leq (1-\alpha_1)$ (otherwise $\alpha_2 \leq (1-\alpha_2)$). We may now examine two cases:

- If $\alpha_2 \leq 1-\alpha_2$, then multiplying by α_1: $\alpha_1\alpha_2 \leq \alpha_1(1-\alpha_2)$. In the other hand, we have $\alpha_1 \leq (1-\alpha_1)$. Multiplying by α_2: $\alpha_1\alpha_2 \leq (1-\alpha_1)\alpha_2$. Summing both results, we get $2\alpha_1\alpha_2 \leq \alpha_1(1-\alpha_2)+\alpha_2(1-\alpha_1)$.
- If $\alpha_2 > 1-\alpha_2$, then notice that $(1-\alpha_1)(1-\alpha_2)-\alpha_1(1-\alpha_2) = (1-2\alpha_1)(1-\alpha_2) < (1-2\alpha_1)\alpha_2$. Therefore,

$$\begin{aligned} \alpha_1\alpha_2 &< (1-\alpha_1)(1-\alpha_2) \\ &< \alpha_2(1-2\alpha_1)+\alpha_1(1-\alpha_2) \\ 2\alpha_1\alpha_2 &< \alpha_2(1-\alpha_1)+\alpha_1(1-\alpha_2). \end{aligned}$$

We now consider the multiparty case. Let $n \geq 2$ be the number of players. Fix n one-way protocols $\mathcal{P}_i$ and a product distribution $\mu = \mu_1 \otimes \ldots \otimes \mu_n$. Remember that in the one-way protocol $\mathcal{P}_i$, one player receives an input x_i and the second all other inputs. We denote the second player's input by x_{-i}. We extend all notations introduced in the two-player case.

The simultaneous protocol works like in the two-player case: each player sends to the referee the message he would have sent in the one-way protocol. Denote these messages by $m_i(x_i)$. Each message defines a set $S^i \in X_i$ of inputs for which the same message is sent. Notice that $R = S^1 \times \ldots \times S^n$ is a communication rectangle of the simultaneous protocol. Let $S_{-i} = S^1 \times \ldots \times S^{i-1} \times S^{i+1} \times \ldots \times S^n$ and $S^i_a = \{x_{-i} \in S_{-i} : \mathcal{P}_i \text{ outputs } a \text{ on input } x_{-i} \text{ and message } m_i(x_i)\}$. The rule for the referee is he outputs 0 if $\mu(\prod_i S^i_0) > \mu(\prod_i S^i_1)$ and 1 otherwise. We now analyze the errors made by this protocol.

Suppose without loss of generality that the referee outputs 0 for all inputs in R. Then, as in the two-player case, all the new errors created in the simultaneous protocol are in $\prod_i S^i_1$. Let μ' be the measure induced by μ on R. μ' is also rectangular, so let $\mu' = \mu'_1 \times \ldots \times \mu'_n$. Finally, call $\alpha_i = \mu'_i(S^i_1)$.

Using our notations, it is sufficient to prove that:

$$n\prod_{i=1}^{n}\alpha_i \leq \sum_{\substack{S\in[n]\\ S\neq 0, S\neq[n]}} \prod_{i\in S}\alpha_i \prod_{i\notin S}(1-\alpha_i).$$

The left hand side is precisely the measure of $\prod_i S_1^i$, whereas the right hand side is the sum of the measures of all sets of inputs for which at least one protocol outputs 0, except the set of inputs for which all protocol outputs 0. Following this idea, notice that:

$$\sum_{\substack{S\in[n]\\ S\neq 0, S\neq[n]}} \prod_{i\in S}\alpha_i \prod_{i\notin S}(1-\alpha_i) = 1 - \prod_{i=1}^{n}\alpha_i - \prod_{i=1}^{n}(1-\alpha_i)$$

The simultaneous protocol is defined such that $\prod_{i=1}^{n}\alpha_i \leq \prod_{i=1}^{n}(1-\alpha_i)$. By induction, we assume

$$(n-1)\prod_{i=1}^{n-1}\alpha_i \leq 1 - \prod_{i=1}^{n-1}\alpha_i - \prod_{i=1}^{n-1}(1-\alpha_i) \tag{1}$$

When not specified, products range from 1 to n. There are again two cases:

- $\prod_{i=1}^{n-1}\alpha_i \leq \prod_{i=1}^{n-1}(1-\alpha_i)$. Then using (1) and multiplying by α_n.

$$\begin{aligned} n\prod \alpha_i &= (n-1)\alpha_n \prod_{i=1}^{n-1}\alpha_i + \alpha_n \prod_{i=1}^{n-1}\alpha_i \\ &\leq \alpha_n \sum_{\substack{S\subset[n-1]\\ S\neq 0, S\neq[n-1]}} \prod_{i\in S}\alpha_i \prod_{i\notin S}(1-\alpha_i) + \alpha_n \prod_{i=1}^{n-1}(1-\alpha_i) \\ &\leq \sum_{\substack{S\subset[n]\\ S\neq 0, S\neq[n]}} \prod_{i\in S}\alpha_i \prod_{i\notin S}(1-\alpha_i) \end{aligned}$$

- $\alpha_n \leq 1-\alpha_n$. Then $\prod \alpha_i < (1-x_n)\prod_{i=1}^{n-1}\alpha_i$. Using this together with induction hypothesis:

$$\begin{aligned} n\prod \alpha_i &= \prod \alpha_i + (n-1)x_n \prod_{i=1}^{n-1}\alpha_i \\ &\leq (1-\alpha_n)\prod_{i=1}^{n-1}\alpha_i + \alpha_n\Big(\sum_{\substack{S\subset[n-1]\\ S\neq 0, S\neq[n-1]}} \prod_{S}\alpha_i \prod_{\overline{S}}(1-\alpha_i)\Big) \\ &\leq \sum_{\substack{S\subset[n]\\ S\neq 0, S\neq[n]}} \prod_{S}\alpha_i \prod_{\overline{S}}(1-\alpha_i) \end{aligned}$$

□

6 Conclusion

Our proofs share a similar framework. We start by giving a general lower bound based on Kolmogorov mutual information between the inputs and the transcript of the protocol. Then, we use incompressibility to identify worst case inputs. Finally, in an alternative to Yao's minmax theorem, we use incompressibility

again to fix the randomness in the protocols. We illustrate our method with two applications, providing new proofs which we believe are simpler, in that they only use elementary techniques. For the VC dimension and shatter coefficient lower bound, we also go one step further and provide a fully combinatorial proof, removing Kolmogorov complexity altogether.

The first application we give is the hidden matching problem. In a recent work [GKK$^+$07], it has been proved that the gap holds for a partial function. The lower bound is based on Fourier analysis, and uses Yao's minmax theorem. Trying to give a combinatorial proof of this lower bound seems to be an interesting problem, as we believe it would give better understanding of the general case, and possibly lead to a separation for total functions.

We also compared one-way and simultaneous communication complexity in the distributional case, for product distributions. The result is known to be false in general. But for reasonable classes of functions, these models may still be equivalent, in particular, [BYJKS02] suggest that it could hold in the special case of symmetric functions, thought they consider it to be a "considerably difficult" problem. We hope our approach might help to analyse this case for symmetric or other natural classes of functions.

Acknowledgments

We wish to thank Troy Lee for many useful discussions. The research was supported by the EU 5th framework program QAP, and by the ANR Blanc AlgoQP.

References

[BJLV00] H. Buhrman, T. Jiang, M. Li, and P. Vitanyi. New applications of the incompressibility method: Part ii. *Theoretical Computer Science*, 235(1):59–70, 2000.

[BPSW06] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006.

[BSS03] Howard Barnum, Michael Saks, and Mario Szegedy. Quantum query complexity and semidefinite programming. In *Proc. of the 18th Annual IEEE Conference on Computational Complexity (CCC)*, pages 179–193. IEEE Computer Society, 2003.

[BYJK08] Ziv Bar-Yossef, T. S. Jayram, and Iordanis Kerenidis. Exponential separation of quantum and classical one-way communication complexity. *SIAM J. Comput.*, 38(1):366–384, 2008.

[BYJKS02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *Proc. of the 17th Annual IEEE Conference on Computational Complexity (CCC)*, pages 93–102. IEEE Computer Society, 2002.

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[dGdW02] Mart de Graaf and Ronald de Wolf. On quantum versions of the yao principle. In *Proc. of the 19th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 347–358, London, UK, 2002. Springer-Verlag.

[GKK+07] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *Proc. of the 39th Annual ACM symposium on Theory of Computing (STOC)*, pages 516–525. ACM, 2007.

[JKN08] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds: extended abstract. In *Proc. of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 599–608, 2008.

[JKS03] T. S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In *Proc. of the 35th Annual ACM Symposium on Theory of Computing (STOC)*, pages 673–682, New York, NY, USA, 2003. ACM.

[JLR00] S. Janson, T. Luczak, and A. Ruci. *Random Graphs*. John Wiley and Sons, 2000.

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

[KNR99] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

[KS92] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[LM08] Sophie Laplante and Frédéric Magniez. Lower bounds for randomized and quantum query complexity using kolmogorov arguments. *SIAM J. Comput.*, 38(1):46–62, 2008.

[LS07] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proc. of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 699–708. ACM, 2007.

[LS08] Troy Lee and Adi Shraibman. Disjointness is hard in the multi-party number-on-the-forehead model. In *Proc. of the 23th Annual IEEE Conference on Computational Complexity (CCC)*, pages 81–91. IEEE Computer Society, 2008.

[LSS08] Troy Lee, Adi Shraibman, and Robert Spalek. A direct product theorem for discrepancy. In *Proc. of the 23th Annual IEEE Conference on Computational Complexity (CCC)*, pages 71–80. IEEE Computer Society, 2008.

[LV93] Ming Li and Paul M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, Berlin, 1993.

[Raz95] Ran Raz. Fourier analysis for probabilistic communication complexity. *Computational Complexity*, 5(3/4):205–221, 1995.

[Raz03] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

[Sha48] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[ŠS06] R. Špalek and M. Szegedy. All quantum adversary methods are equivalent. *Theory of Computing*, 2(1):1–18, 2006.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proc. of the 11h Annual ACM Symposium on Theory of Computing (STOC)*, pages 209–213. ACM, 1979.

[Yao83] Andrew Chi-Chih Yao. Lower bounds by probabilistic arguments (extended abstract). In *Proc. of the 24th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 420–428. IEEE, 1983.

A A random graph theorem

Let $G = (V(G), E(G))$ be a bipartite graph such that $E(G) = n$. There are two procedures to define a random subgraph H of G:

- For each edge $e \in E(G)$, let $e \in E(H)$ with probability p. Let H_p denote the random graph built this way.
- Pick $E(H)$ of size m at random in $E(G)$. Denote H_m the random graph built this way.

For our application, we are interested in properties of H_m with $m = \sqrt{n}$. But we know that with $m = pn$, these models are very similar. Therefore, we can begin by studying H_p with $p = \frac{1}{\sqrt{n}}$, and transpose it to H_m. The following theorem makes precise the similarity between these two models:

Theorem 8. *[JLR00] Let $\mathbb{Q}$ be any set of subgraphs of G. Then for $p = \frac{1}{\sqrt{n}}$, $m = \sqrt{n}$, and n sufficiently large:*

$$Prob(H_m \in \mathbb{Q}) \leq \sqrt{2\pi} n^{1/4} Prob(H_p \in \mathbb{Q})$$

Proof.

$$\begin{aligned} Prob(H_p \in \mathbb{Q}) &= \sum_{k=0}^{n} Prob(H_p \in \mathbb{Q} \,|\, |E(H_p)| = k) Prob(|E(H_p)| = k) \\ &= \sum_{k=0}^{n} Prob(H_k \in \mathbb{Q}) Prob(|E(H_p)| = k) \\ &\geq Prob(H_m \in \mathbb{Q}) Prob(|E(H_p)| = m) \end{aligned}$$

Note that $Prob(|E(H_p)| = m) = \binom{n}{m} p^m (1-p)^{n-m}$. One can easily prove that $Prob(|E(H_p)| = m) \sim \frac{1}{\sqrt{2\pi} n^{1/4}}$, which finishes the proof. □

Theorem 9. *Fix a bipartite graph G, and pick a random subgraph $H = H_{\sqrt{n}}$. Let $k_H = \#\{v \in V(H) : deg_H(v) > 0\}$. There exists a constant $c > 0$ such that:*

$$\lim_{n \to +\infty} Prob_H(k_H \geq c\sqrt{n}) = 1$$

Proof. We only have to prove the theorem for a random subgraph $H = H_{1/\sqrt{n}}$ and use Theorem 8 to conclude that it holds for $H = H_{\sqrt{n}}$. Let $V(G) = X_1 \cup X_2$, such that $E(G) \subset X_1 \times X_2$. Call $X_i^+ = \{v \in X_i : deg(v) > 2\sqrt{n}\}$ and $X_i^- = \overline{X_i^+}$. We shall start the proof with the following structural fact.

Claim 1. *For at least one index i, we have $|X_i| > \sqrt{n}$ and $\sum_{X_i^-} deg(v) > 3n/8$.*

First, notice that at least one side of the graph has at least $\sqrt{n}$ vertices. Then, if one side of the graph has strictly less than $2\sqrt{n}$ vertices, every vertex in the other side has degree strictly less than $2\sqrt{n}$. Summing over all X_i^- proves the claim. Therefore, assume both sides have more than $2\sqrt{n}$ vertices.

Call $E^{\delta,\varphi} = E(G) \cap X_1^\delta \times X_2^\varphi$ for $\delta, \varphi \in \{+,-\}$. With this decomposition, the total number of edges is $n = |E^{++}| + |E^{+-}| + |E^{-+}| + |E^{--}|$. Moreover: $\sum_{X_1^+} deg(v) = |E^{++}| + |E^{+-}|$ and $\sum_{X_2^+} deg(v) = |E^{++}| + |E^{-+}|$. Now notice that since $\sum_{v \in X_i} deg(v) = n$ (for $i = 1, 2$), there are at most $\frac{\sqrt{n}}{2}$ vertices of degree larger than $2\sqrt{n}$. This implies $|E^{++}| < \frac{n}{4}$. Finally:

$$\begin{aligned} n &= |E^{++}| + |E^{+-}| + |E^{-+}| + |E^{--}| \\ &= \sum_{X_1^+} deg(v) + \sum_{X_2^+} deg(v) - |E^{++}| + |E^{--}| \\ &> \sum_{X_1^+} deg(v) + \sum_{X_2^+} deg(v) - \frac{n}{4} \\ \frac{5n}{4} &> \sum_{X_1^+} deg(v) + \sum_{X_2^+} deg(v). \end{aligned}$$

We can conclude that for at least one i, $\sum_{X_i^+} deg(v) < 5n/8$, which on the other side implies $\sum_{X_i^-} deg(v) > 3n/8$.

Let $v_0 = \#\{v : deg_H(v) = 0\}$. We prove an upper bound on $\mathbf{E}v_0$. Let X_i verify the previous claim. For each vertex $v \in X_i$, $Prob(deg(v) = 0) = (1 - \frac{1}{\sqrt{n}})^{deg(v)} \sim e^{-deg(v)/\sqrt{n}}$. Notice that for vertices in the same part of G, degrees in H are independent random variables. Therefore, $\mathbf{E}v_0 = \sum_v e^{-deg(v)/\sqrt{n}}$.

Let $\alpha_1 = \frac{1-e^{-2}}{2}$. Notice that for all x such that $0 < x < 2$, $e^{-x} < 1 - \alpha_1 x$. Therefore , we have for low degree vertices:

$$\sum_{v \in X_i^-} e^{-deg(v)/\sqrt{n}} < \sum_{X_i^-} 1 - \alpha_1 \frac{deg(v)}{\sqrt{n}} < |X_i| - \frac{3\alpha_1}{8}\sqrt{n}.$$

On the other hand, for high degree vertices:

$$\sum_{X_i^+} e^{-deg(v)/\sqrt{n}} < e^{-2}|X_i^+| < \frac{e^{-2}}{2}\sqrt{n}$$

Letting $c = (\frac{3\alpha_1}{8} - \frac{e^{-2}}{2})$

$$\mathbf{E}v_0 < |X_i| - c\sqrt{n}$$
$$\mathbf{E}k_H > c\sqrt{n}$$

One can verify that $c > 1$.

It was noticed that v_0 is a sum of independent indicator random variables, such that k_H is also a sum of independent random variables. Therefore, we can apply Chernov bounds directly to derive:

$$Prob[k_H < \frac{c}{2}\sqrt{n}] < e^{-\frac{c\sqrt{n}}{8}}$$

□