An Information Complexity Approach to the Inner Product Problem

William Henderson-Frost

Advisor: Amit Chakrabarti

Senior Honors Thesis
Submitted to the faculty in partial fulfillment of the
requirements for a Major in Computer Science
Dartmouth College

· June 2009 ·

Dartmouth Computer Science Technical Report TR2009-646

Chapter 1

Introduction

1.1 Communication Complexity

In a communication complexity problem, multiple parties compute a function on private inputs by communicating. Variations on this framework are numerous, but most share the same core features. The only resource of concern is the amount of communication between the parties. Individual parties are allowed infinite processing power and infinite memory. The parties generate a single answer to the problem. This answer may be known only to a single party. The number of parties involved is often two. In this case, it is common to refer to the players as Alice and Bob. This is the framework introduced by Yao in 1979, and is examined in considerable depth in the definitive textbook [5]. Though the communication complexity problem framework may appear narrow, communication complexity results have proven useful in various areas such as data stream algorithms, boolean circuits, and data structures [1], [5].

A communication scheme employed by Alice and Bob to solve the problem is called a protocol. An individual conversation between Alice and Bob is called a transcript. The cost of a protocol is the length of the longest transcript it produces. The communication complexity of a problem is the minimum cost of a protocol that adequately solves the problem. Researchers seek to devise protocols with small cost, as well as construct proofs that lower bound communication complexity.

The two forms of the communication complexity problem framework are the deterministic model and the randomized model. In the deterministic case, the communication between the parties is entirely dependent upon the inputs, and will always be the same for a given input pair. In the randomized case, the protocol may take into account random events. These random events are known as coin tosses, and may be either private or public. Public coin tosses are known to all parties. Private coin tosses, however, are visible to a single party. In this paper, unless otherwise noted, we focus upon private-coin randomized protocols. The randomized model allows Alice and Bob to make errors. The error of a randomized protocol is the maximum error of the protocol on any input distribution. Normally, deterministic protocols are not allowed to err. Occasionally, however, it is useful to reason about the cost

of a deterministic protocol which is allowed some error on a given input distribution.

When studying the communication complexity of a function, it is helpful to visualize the function's truth table. Generally, the rows of the truth table represent Bob's inputs, and the columns represent Alice's inputs. Each entry in the table indicates the function's value on a specific input pair. Recall that in the deterministic case, the transcript chosen by the protocol on a given input pair is fixed. We assert the following useful claim: The set of inputs that yield a particular transcript form a combinatorial rectangle in the truth table. We refer the reader to [5] for a proof. In the randomized setting, the transcript produced by the protocol is a result of player inputs as well as random coin tosses. These transcripts form combinatorial rectangles in the larger table whose rows and columns represent input/private-coin-toss combinations. Proving deterministic communication complexity lower bounds is substantially easier than proving randomized communication complexity lower bounds. This is because it is usually a simple task to lower bound the minimum number of monochromatic rectangles required to partition the truth table of a function.

The following is a naturally arising communication complexity problem: How much additional communication does it require to solve multiple instances of a function at the same time? We require that protocols solving multiple instances of a problem provide answers that are simultaneously correct with high probability. Since the techniques used within this paper to prove a lower bound on the communication complexity of the inner product function also produce a lower bound on the communication complexity of the direct sum of inner product function problems, we include this proof as well.

1.2 Communication Complexity Problems

For illustrative purposes, we examine three two-party communication complexity problems. Alice and Bob both receive n bits of input in each of these problems.

- **GT**, the greater than function. Alice and Bob are given numbers $x, y \in [2^n]$, and must output 1 if $x > y$, and 0 otherwise.
- **DISJ**, the disjointedness function. Alice and Bob are given sets $A, B \subseteq [n]$ and they must output 0 if $A \cap B = \emptyset$, and 1 otherwise.
- **IP**, the inner product function. Alice is given binary string $x = x_1x_2...x_n$, Bob is given binary string $y = y_1y_2..y_n$, and they must output $\sum_{i=1}^{n} x_iy_i \pmod 2$

If Alice sends her entire input to Bob, Bob may perform the calculation and output the correct answer. Thus communication complexity cannot be greater than the length of the inputs. In the deterministic setting, none of the problems above can be solved with fewer than n bits of communication. However, under the randomized private-coin model, the communication complexity of GT is reduced to $\Omega(\log(n))$ [KN].

.

The inner product function is very hard even in the randomized model. We use $R_\delta(\text{IP})$ to indicate the randomized communication complexity of the inner product function with δ error. The two following bounds are known:

- $R_\delta(\text{IP}) \geq n - O(\log(1/\delta))$ [5]
- $R_\delta(\text{IP}) \geq \frac{n}{4}(1 - 2\sqrt{\delta})$ [1]

The former bound is proven using a concept called discrepancy, which requires showing that all nearly-monochromatic rectangles in the truth table are small. This strategy involves hefty linear algebra. The second bound, though not proven explicitly in [1], can be derived from the proof of the disjointedness function lower bound. This proof uses subtle information complexity techniques on a distribution over inputs which are disjoint. Alice's and Bob's n bit strings represent disjoint sets if for every index i, the i-th bits of the players' strings are not both one. Inputs from this distribution have the decomposing property that if the bits at any index become ones, the solution to the disjointedness function changes. On this input distribution, the inner product function behaves similarly. The inner product is always zero, and if the bits of both strings at any index become ones, the function's value changes. In this paper, we present a lower bound for the randomized communication complexity of the inner product function using information complexity techniques on the uniform distribution.

Chapter 2

Preliminaries

2.1 Probabilistic Inequalities

Markov's Inequality

Let X be any random variable. For positive values of a:

$$\Pr[|X| \geq a] \leq \frac{\mathbb{E}[|X|]}{a}$$

Hoeffding's Inequality

Suppose that $X_1, X_2, ..., X_n$ is a set of mutually independent random variables, such that each X_i takes values in the range $[a_i, b_i]$, and $S = \sum_{i=1}^{n} X_i$. Then for positive values of t:

$$\Pr[\mathbb{E}[S] - S \geq nt] \leq \exp\left(\frac{-2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

2.2 Communication Complexity

The random variable $\pi(X, Y)$ represents the transcript chosen by protocol π when the players' inputs are decided by random variables X and Y. The random variable $\pi(x, y)$ represents the transcript chosen by protocol π on the inputs x, y. If the transcript τ occurs, τ_{out} denotes the output submitted by Alice and Bob. The event that τ_{out} doesn't match $f(x, y)$ will be denoted by $\epsilon(x, y, \tau)$. That is:

$$\epsilon(x, y, \tau) = \begin{cases} 0, \text{ if } \tau_{\text{out}} = f(x, y) \\ 1, \text{ otherwise} \end{cases}$$

The random variable $\epsilon(X, Y, \pi(X, Y))$ indicates if π's choice of transcript results in an error when the players' inputs are decided by random variables X and Y. Accordingly, the random variable $\epsilon(x, y, \pi(x, y))$ indicates if π's choice of transcript for the input pair x, y results in an error. We say that a protocol has error δ on distribution $(X, Y) \sim \mu$ if $\mathbb{E}[\epsilon(X, Y, \pi(X, Y))] \leq \delta$. A protocol is said to a δ-error protocol if it has error δ on all input distributions.

Definition 1. The cost of a protocol π is the maximum length of a transcript chosen by π.

Definition 2. $\mathrm{D}^{\mu}_{\delta}(f)$ is the minimum cost of a deterministic protocol which has error less than δ when computing f on inputs from the distribution μ.

Definition 3. $\mathrm{R}_{\delta}(f)$, the randomized communication complexity of f, is the minimum cost of a δ-error protocol which computes f.

Definition 4. If f is a function and ℓ is an integer, f^{ℓ} is defined to be:

$$f^{\ell}((x_1, x_2, ..., x_{\ell}), (y_1, y_2, ..., y_{\ell})) = (f(x_1, y_1), f(x_2, y_2), ..., f(x_{\ell}, y_{\ell}))$$

Definition 5. $\mathrm{D}^{\mu^{\ell}}_{\delta}(f^{\ell})$ is the minimum cost of a deterministic protocol which computes all ℓ copies of f on inputs each from the distribution μ with probability at least $1 - \delta$ that all outputs are simultaneously correct.

Definition 6. $\mathrm{R}_{\delta}(f^{\ell})$ is the minimum cost of a randomized protocol which solves all ℓ copies of f with probability at least $1 - \delta$ that all outputs are simultaneously correct.

Yao's Min-Max Principle

For any function f, $R^{pub}_{\delta}(f) = \max_{\mu} \mathrm{D}^{\mu}_{\delta}(f)$, where $R^{pub}_{\delta}(f)$ is the public-coin randomized communication complexity of f. Since $R^{pub}_{\delta}(f) \leq R_{\delta}(f)$, as any public-coin protocol may be simulated by a private-coin protocol, we have:

$$R_{\delta}(f) \geq \max_{\mu} \mathrm{D}^{\mu}_{\delta}(f)$$

2.3 Information Theory

Information theory, a field established in 1948 by Claude Shannon, examines the quantification of information. Of central importance to information theory is the notion of entropy. The entropy of a random variable is a quantification of the random variable's uncertainty. Mathematically, we define entropy as:

$$\mathrm{H}(X) = -\sum_{x \in X} \Pr[x = X] \cdot \log\left(\Pr[x = X]\right)$$

The conditional entropy of Y conditioned on X can be thought of as uncertainty of Y after X has been revealed. It is defined to be:

$$\mathrm{H}(X \mid Y) = \mathrm{H}(X, Y) - \mathrm{H}(Y)$$

Two properties of entropy are the subadditivity of entropy and the subadditivity of conditional entropy.

$$\mathrm{H}(X, Y) \leq \mathrm{H}(X) + \mathrm{H}(Y)$$
$$\mathrm{H}(X, Y \mid Z) \leq \mathrm{H}(X \mid Z) + \mathrm{H}(Y \mid Z)$$

From this subadditivity property and the definition of conditional entropy, it is evident that $\mathrm{H}(X \mid Y) \leq \mathrm{H}(X)$. Another useful information theoretic concept is mutual information.

The mutual information between two random variables is the amount of uncertainty that they share. Formally,

$$\mathrm{I}(X : Y) = \mathrm{H}(X) - \mathrm{H}(X \mid Y)$$

The conditional mutual information between X and Y conditioned upon Z is defined to be:

$$\mathrm{I}(X : Y \mid Z) = \sum_{z \in Z} \Pr[Z = z] \cdot I(X : Y \mid Z = z)$$

By the definition of mutual information and the entropy subaddivity property:

$$\mathrm{H}(X) \geq \mathrm{I}(X : Y)$$
$$\mathrm{H}(X \mid Z) \geq \mathrm{I}(X : Y \mid Z)$$

2.4 Information Complexity

At the intersection of communication complexity and information theory lies information complexity. The information complexity of a function is the amount of information the players solving the function must reveal about their inputs. This concept was introduced in [3]. The measure of information complexity of a protocol used in this paper is called *information content* and is similar to the measure used in [2].

Definition 7. Suppose X and Y are random variables distributed according to μ, the *information content* of π with respect to μ, denoted $\mathrm{IC}_\mu(\pi)$, is defined to be:

$$\mathrm{IC}_\mu(\pi) = \mathrm{Max}\{\mathrm{I}(X : \pi(X, Y) \mid Y),\ \mathrm{I}(Y : \pi(X, Y) \mid X)\}$$

Chapter 3

Inner Product Information Complexity

3.1 Information Content Derived Bounds

The following two theorems relate information content lower bounds to communication complexity lower bounds.

Theorem 1. *Suppose that any protocol π computing f on inputs from distribution μ makes error less than δ and has the property that $IC_\mu(\pi) \geq c$. Then $R_\delta(f) \geq c$.*

Proof. Let π be the δ-error protocol for f with the least communication, let X, Y be input random variables distributed according to μ, and let $|\pi|$ denote the length of the longest transcript of π. There are at most $2^{|\pi|}$ possible transcripts generated by π, thus $|\pi| \geq \mathrm{H}(\pi(X,Y))$. Using this and the entropy properties described above:

$$\begin{aligned} \mathrm{R}_\delta(f) = |\pi| &\geq \mathrm{H}(\pi(X,Y)) \\ &\geq \mathrm{H}(\pi(X,Y) \mid Y) \\ &\geq \mathrm{I}(X : \pi(X,Y) \mid Y) \end{aligned}$$

The same holds for the opposite use of X and Y, showing $\mathrm{R}_\delta(f) \geq \mathrm{I}(Y : \pi(X,Y) \mid X)$. It follows that $\mathrm{R}_\delta(f) \geq \mathrm{Max}\{\mathrm{I}(X : \pi(X,Y) \mid Y),\ \mathrm{I}(Y : \pi(X,Y) \mid X)\} \geq c$. □

Theorem 2. *Suppose that any protocol π computing f on inputs from distribution μ makes error less than δ and has the property that $IC_\mu(\pi) \geq c$. Then $R_\delta(f^\ell) \geq \frac{c \cdot \ell}{2}$.*

Proof. This is a direct application of results from [2]. We state Theorem 1.9 from [2] here explicitly: For every μ, f, δ, there exists a protocol π computing f on inputs drawn from $(X,Y) \sim \mu$ with probability of error at most δ and communication at most $\mathrm{D}_\delta^{\mu^\ell}(f^\ell)$ such that

$$\mathrm{I}(X : \pi(X,Y) \mid Y) + \mathrm{I}(Y : \pi(X,Y) \mid X) \leq \frac{2\mathrm{D}_\delta^{\mu^\ell}(f^\ell)}{\ell}$$

Since $\mathrm{IC}_\mu(\pi) \leq \mathrm{I}(X : \pi(X,Y) \mid Y) + \mathrm{I}(Y : \pi(X,Y) \mid X)$ it follows that:

$$\frac{c \cdot \ell}{2} \leq \mathrm{D}_\delta^{\mu^\ell}(f^\ell)$$

By Yao's principle:

$$\mathrm{R}_\delta(f^\ell) \geq \frac{c \cdot \ell}{2}$$

□

3.2 Inner Product Information Content

Let μ be the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$. If a protocol π computes IP making $\delta < \frac{3}{640}$ error on distribution μ, then:

$$\mathrm{IC}_\mu(\pi) > \frac{n}{32}$$

By Theorem 1 we have:

$$\mathrm{R}_\delta(\mathrm{IP}) > \frac{n}{32}$$

By Theorem 2 we have:

$$\mathrm{R}_\delta(\mathrm{IP}^\ell) > \frac{n \cdot \ell}{64}$$

This is a consequence of the following more delicate statement. Let $(X,Y) \sim \mu$ be the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$. If a protocol π computes IP, $k_2 > \frac{3}{4}$, and:

$$\begin{aligned}
\mathbb{E}\left[\epsilon(X,Y,\pi(X,Y))\right] &\leq k_1 \\
\mathrm{I}(X : \pi(X,Y) \mid Y) &\leq \frac{1}{4}(1-k_2)n \\
\mathrm{I}(Y : \pi(X,Y) \mid X) &\leq \frac{1}{4}(1-k_2)n
\end{aligned}$$

Then:

$$k_1 > \left(k_2 - \frac{3}{4}\right)^2 \cdot \frac{3}{10}$$

Proof

The strategy used here is heavily derived from that used in [5].

First we identify a transcript, which when conditioned upon, the protocol errs infrequently and the entropies of both X and Y are great. Using the high entropy of X and the low error, we prove the existence of many X values on which the transcript makes little error. From these X values we chose a subset with the following property: The transcript's error on one value is independent of the transcript's error on the other values. Using a uniform distribution on these particular X values and the nature of the inner product function, we prove the transcript makes considerable error on most Y inputs and thus makes considerable error.

We proceed by proving the existence of a transcript with the desired properties.

Since $\mathbb{E}[\epsilon(X, Y, \pi(X, Y))] \leq k_1$, an application of Markov's Inequality reveals:

$$\Pr_{\tau}[\mathbb{E}[\epsilon(X, Y, \tau) \mid \pi(X, Y) = \tau] \geq 4k_1] \leq \frac{1}{4}$$

Using the property that $\mathrm{I}(A, B) = \mathrm{H}(A) - \mathrm{H}(A \mid B)$ and that $\mathrm{H}(X) = n$ (as X is uniformly distributed over 2^n strings), we may show:

$$\begin{aligned}\frac{1}{4}(1 - k_2)n \;\geq\; \mathrm{I}(X : \pi(X, Y) \mid Y) &= \mathrm{H}(X \mid Y) - \mathrm{H}(X \mid \pi(X, Y), Y) \\ &= n - \mathrm{H}(X \mid \pi(X, Y), Y) \\ &= n - \mathrm{H}(X \mid \pi(X, Y)) \qquad \text{by rectangle property}\end{aligned}$$

Another application of Markov's Inequality shows:

$$\Pr_{\tau}[n - \mathrm{H}(X \mid \pi(X, Y) = \tau) \geq (1 - k_2)n] \leq \frac{1}{4}$$

Performing the same manipulation with $\mathrm{I}(Y : \pi(X, Y))$ reveals:

$$\Pr_{\tau}[n - \mathrm{H}(Y \mid \pi(X, Y) = \tau) \geq (1 - k_2)n] \leq \frac{1}{4}$$

Together, these three inequalities show:

$$\Pr_{\tau}\begin{bmatrix} \mathrm{H}(X \mid \pi(X, Y) = \tau) \geq k_2 n \;\wedge \\ \mathrm{H}(Y \mid \pi(X, Y) = \tau) \geq k_2 n \;\wedge \\ \mathbb{E}[\epsilon(X, Y, \tau) \mid \pi(X, Y) = \tau] \leq 4k_1 \end{bmatrix} \geq 1 - \frac{1}{4} - \frac{1}{4} - \frac{1}{4}$$

Thus there exists a transcript $\hat{\tau}$ such that the following conditions hold.

$$\mathbb{E}[\epsilon(X, Y, \hat{\tau}) \;\mid\; \pi(X, Y) = \hat{\tau}] \leq 4k_1 \tag{3.1}$$

$$H(X \;\mid\; \pi(X, Y) = \hat{\tau}) \geq k_2 n \tag{3.2}$$

$$H(Y \;\mid\; \pi(X, Y) = \hat{\tau}) \geq k_2 n \tag{3.3}$$

To simplify notation, let X' be X conditioned upon $\pi(X, Y) = \hat{\tau}$, and let Y' be Y conditioned upon $\pi(X, Y) = \hat{\tau}$. Since $\mathrm{H}(X')$ is close to $\mathrm{H}(X)$, the probability that X' takes a value which occurs infrequently must be high.

Definition 8. $X^* = \{x \in X' \;\mid\; \Pr[X' = x] \leq \frac{1}{2^{\frac{3}{4}n}}\}$

Claim 1. $\Pr[X' \in X^*] \geq 4(k_2 - \frac{3}{4})$

Proof. We use the property that for an event E and a random variable X:

$$H[X] = \Pr[E] \cdot \mathrm{H}(X \mid E) + \Pr[\bar{E}] \cdot \mathrm{H}(X \mid \bar{E}) + \mathrm{H}_{\mathrm{B}}(\Pr[E])$$

Using X' as the random variable and $X' \in X^*$ as the event we have:

$$\mathrm{H}(X') = \Pr[X' \in X^*] \cdot \mathrm{H}(X' \mid X' \in X^*) + \Pr[X' \notin X^*] \cdot \mathrm{H}(X' \mid X' \notin X^*) + \mathrm{H}_{\mathbf{B}}(\Pr[X' \in X^*])$$

Trivially, $\mathrm{H}_{\mathbf{B}}(\Pr[X' \in X^*]) \leq 1$ and $\mathrm{H}(X' \mid X' \notin X^*) \leq n$. Since all values of $X' \notin X^*$ occur with probability at least $\frac{1}{2^{(3/4)n}}$, there are at most $2^{(3/4)n}$ distinct values in $X' \notin X^*$. Thus $\mathrm{H}(X' \mid X' \notin X^*) \leq (3/4)n$. Therefore:

$$\mathrm{H}(X') \leq \Pr[X' \in X^*] \cdot n + \Pr[X' \notin X^*] \cdot \frac{3}{4}n + 1$$
$$\mathrm{H}(X') \leq \left(\frac{1}{4}\Pr[X' \in X^*] + \frac{3}{4}\right) \cdot n + 1$$

If $\Pr[X' \in X^*] < 4(k_2 - \frac{3}{4})$ then for sufficiently large n, $\mathrm{H}(X') < k_2 n$, which contradicts (2). □

Definition 9. $\hat{X} = \left[x \in X^* \mid \mathbb{E}[\epsilon(x, Y', \hat{\tau})] \leq \frac{2k_1}{k_2 - \frac{3}{4}}\right]$

Claim 2. $|\hat{X}| \geq (k_2 - \frac{3}{4})2^{\frac{3}{4}n+1}$

Proof.

$$\begin{aligned}
\mathbb{E}[\epsilon(X', Y', \hat{\tau})] &= \Pr[X' \in X^*] \cdot \mathbb{E}[\epsilon(X', Y', \hat{\tau}) \mid X' \in X^*] + \\
&\quad \Pr[X' \notin X^*] \cdot \mathbb{E}[\epsilon(X', Y', \hat{\tau}) \mid X' \notin X^*] \\
\mathbb{E}[\epsilon(X', Y', \hat{\tau})] &\geq \Pr[X' \in X^*] \cdot \mathbb{E}[\epsilon(X', Y', \hat{\tau}) \mid X' \in X^*] \\
4k_1 &\geq \Pr[X' \in X^*] \cdot \mathbb{E}[\epsilon(X', Y', \hat{\tau}) \mid X' \in X^*] && \text{by (3.1)} \\
\frac{4k_1}{\Pr[X' \in X^*]} &\geq \mathbb{E}[\epsilon(X', Y', \hat{\tau}) \mid X' \in X^*] \\
\frac{k_1}{k_2 - \frac{3}{4}} &\geq \mathbb{E}[\epsilon(X', Y', \hat{\tau}) \mid X' \in X^*] && \text{by Claim 1}
\end{aligned}$$

Now by an application of Markov's Inequality:

$$\Pr_{x \in X' | X' \in X^*}\left[\mathbb{E}[\epsilon(x, Y', \hat{\tau})] \leq \frac{2k_1}{k_2 - \frac{3}{4}}\right] \geq \frac{1}{2} \tag{3.4}$$

$$\begin{aligned}
\Pr\left[X' \in \hat{X}\right] &\geq \Pr[X' \in X^*] \cdot \Pr[X' \in \hat{X} \mid X' \in X^*] \\
&\geq 4(k_2 - \frac{3}{4}) \cdot \frac{1}{2} = 2(k_2 - \frac{3}{4}) && \text{by (3.4) and Claim 1}
\end{aligned}$$

Since individual values in $\hat{X}$ occur with probability at most $\frac{1}{2^{\frac{3}{4}n}}$ (because $\hat{X} \subseteq X^*$), the number of distinct values in $\hat{X}$ is at least:

$$\frac{2(k_2 - \frac{3}{4})}{\frac{1}{2^{\frac{3}{4}n}}} = (k_2 - \frac{3}{4})2^{\frac{3}{4}n+1} \tag{3.5}$$

□

Claim 3. *There exists a subset of $M \subseteq \hat{X}$, with the following properties: $\{\epsilon(x, Y, \hat{\tau}) \mid x \in M\}$ is a set of mutually independent random variables, and $|M| \geq log_2(|\hat{X}|)$.*

In order to prove Claim 3, we will make use of two lemmas:

Lemma 1. *Suppose $\mathbb{F}$ is a field and $X \subseteq \mathbb{F}^n$ then $\exists M \subseteq X$ s.t. $|M| \geq \log_{|\mathbb{F}|}(|X|)$ and M is a linearly independent set.*

Proof. Suppose that M is a linearly independent subset of X and $|M| = m$. The span of M has size $|\mathbb{F}|^m$. Therefore, if $m < \log_{|\mathbb{F}|}(|X|)$ then $|\text{Span}(M)| < |X|$. This implies that there exists $x \in X$ which is linearly independent of the vectors in M. □

Lemma 2. *Suppose M is a linearly independent subset of $\mathbb{F}^n$. If Y is a random variable uniformly distributed over $\mathbb{F}^n$ then $\{Y \cdot x \mid x \in M\}$ is a set of mutually independent random variables.*

Proof. For any $x \in M$, $f \in \mathbb{F}$, $x \neq 0$, $\Pr[Y \cdot x = f] = \frac{1}{|\mathbb{F}|}$, as $0 \notin M$. Accordingly, to prove that $\{Y \cdot x \mid x \in M\}$ is a set of mutually independent random variables it it suffices to show that for any $\{x_1, x_2, ..., x_k\} \subseteq M$, $f_1, f_2, ..., f_k \in \mathbb{F}$, $\Pr[Y \cdot x_1 = f_1 \wedge Y \cdot x_2 = f_2 \wedge ... \wedge Y \cdot x_k = f_k] = \frac{1}{|\mathbb{F}|^k}$. Select some such $\{x_1, x_2, ..., x_k\} \subseteq M$, $f_1, f_2, ..., f_k \in \mathbb{F}$. now consider the linear map $T : \mathbb{F}^n \to \mathbb{F}^k$

$$T(y) = y \cdot x_1, y \cdot x_2, ..., y \cdot x_k$$

We note now that $T(y) = \mathbf{A}y$, where $\mathbf{A}$ is the matrix whose rows are $x_1, x_2, ..., x_k$. Since $x_1, x_2, ..., x_k$ are linearly independent, the row rank of $\mathbf{A}$ is k, and thus the dimension of the image of T is k. Therefore T is a linear map onto $\mathbb{F}^k$. It follows that for any $f \in \mathbb{F}^k$

$$|T^{-1}(f)| = \frac{|\mathbb{F}^n|}{|\mathbb{F}^k|}$$

Since Y is a random variable uniformly distributed over $\mathbb{F}^n$,

$$\begin{aligned}\Pr[Y \cdot x_1 = f_1 \wedge Y \cdot x_2 = f_2 \wedge ... \wedge Y \cdot x_k = f_k] &= \Pr[T(Y) = f_1, f_2, ..., f_k] \\ &= \Pr[Y \in T^{-1}(f_1, f_2, ..., f_k)] \\ &= \frac{\mathbb{F}^n}{\mathbb{F}^k} \cdot \frac{1}{\mathbb{F}^n} = \frac{1}{\mathbb{F}^k}\end{aligned}$$

□

Proof of Claim 3. Since $\hat{X} \subseteq \mathbb{F}_2^n$, by Lemma 1, there exists $M \subseteq \hat{X}$ s.t. $|M| \geq \log_2(|\hat{X}|)$ and M is a linearly independent set. By Lemma 2, $\{x \cdot Y \mid x \in M\}$ is a set of mutually independent random variables. If $\hat{\tau}_{\text{out}} = 0$, $\epsilon(x, Y, \hat{\tau}) = x \cdot Y$. Otherwise $\hat{\tau}_{\text{out}} = 1$ and $\epsilon(x, Y, \hat{\tau}) = 1 - x \cdot Y$. Thus if $\{x \cdot Y \mid x \in M\}$ is a set of mutually independent random variables, $\{\epsilon(x, Y, \hat{\tau}) \mid x \in M\}$ is a set of mutually independent random variables. □

By (3.5), $|\hat{X}| \geq (k_2 - \frac{3}{4})2^{\frac{3}{4}n+1}$. By Claim 3, we know there exists $M \subseteq \hat{X}$ where $|M| \geq \frac{3}{4}n + \log(2(k_2 - \frac{3}{4}))$ and $\{\epsilon(x, Y, \hat{\tau}) \mid x \in M\}$ is a set of mutually independent random

variables. Pick some small $\delta > 0$ and let X'' be a uniform distribution on $(\frac{3}{4} - \delta)n$ values of M. Since $M \subseteq \hat{X}$ and $\forall\ x \in \hat{X}\ \ \mathbb{E}[\epsilon(x, Y', \hat{\tau})] \leq \frac{2k_1}{k_2 - \frac{3}{4}}$, we have:

$$\mathbb{E}[\epsilon(X'', Y', \hat{\tau})] \leq \frac{2k_1}{k_2 - \frac{3}{4}} \tag{3.6}$$

Definition 10. $g(y) = \mathbb{E}[\epsilon(X'', y, \hat{\tau})]$

The function g is of interest as the error of $\hat{\tau}$ on the distribution $X'' \times Y'$ is equal to $\mathbb{E}[g(Y')]$. In order to bound $\mathbb{E}[g(Y')]$, we instead calculate $\mathbb{E}[g(Y)]$, and then show $g(Y)$ is concentrated around $\mathbb{E}[g(Y)]$. It follows that $\mathbb{E}[g(Y')]$ is close to $\mathbb{E}[g(Y)]$, because $\mathrm{H}(Y')$ is large.

$\mathbb{E}[\epsilon(x, Y, \hat{\tau})] = \frac{1}{2}\ \forall\ x \in X$. Accordingly:

$$\mathbb{E}[g(Y)] = \mathbb{E}[\epsilon(X'', Y, \hat{\tau})] = \mathop{\mathbb{E}}_{x \in X''}[\mathbb{E}[\epsilon(x, Y, \hat{\tau})]] = \frac{1}{2}$$

Also note that:

$$g(y) = \mathbb{E}[\epsilon(X'', y, \hat{\tau})] = \sum_{x \in X} \Pr[X'' = x] \cdot \epsilon(x, y, \hat{\tau}) = \sum_{x \in X''} \frac{\epsilon(x, y, \hat{\tau})}{\left(\frac{3}{4} - \delta\right) n}$$

$$g(Y) = \sum_{x \in X} \Pr[X'' = x] \cdot \epsilon(x, Y, \hat{\tau}) = \sum_{x \in X''} \frac{\epsilon(x, Y, \hat{\tau})}{\left(\frac{3}{4} - \delta\right) n}$$

Since $g(Y)$ is the sum of a set of mutually independent random variables, each of which is bounded, we may use Hoeffding's Inequality to bound the probability that $g(Y)$ is significantly less than its expectation.

$$\Pr\left[\mathbb{E}[g(Y)] - g(Y) \geq c\right] \ \leq\ \exp\left(\frac{-2c^2}{(\frac{3}{4} - \delta)n\left(\frac{1}{(\frac{3}{4} - \delta)n}\right)^2}\right)$$

$$\Pr\left[g(Y) \leq \frac{1}{2} - c\right] \ \leq\ \exp\left(-2c^2\left(\frac{3}{4} - \delta\right)n\right)$$

$$\text{Let } c = \frac{1}{\sqrt{8\left(\frac{3}{4} - \delta\right)\log(e)}}$$

$$\Pr\left[g(Y) \leq \frac{3}{20}\right] \ \leq\ \exp\left(\frac{-n}{4 \cdot \log(e)}\right) \tag{3.7}$$

Claim 4. $\Pr[g(Y') > \frac{3}{20}] \geq 4(k_2 - \frac{3}{4})$

Proof. Y is a uniform distribution over 2^n values, and therefore by (3.7) the number of Y values such that $g(Y) \leq \frac{3}{20}$ is at most:

$$\frac{2^n}{e^{\left(\frac{n}{4 \cdot \log(e)}\right)}}$$

Since the number of such values is small, the probability Y' takes such a value must be small, as otherwise the entropy of Y' would also be small. To show this, we decompose $\mathrm{H}(Y')$ using the event $g(Y') \leq \frac{3}{20}$.

$$\begin{aligned}\mathrm{H}(Y') \;=\; &\Pr\left[g(Y') \leq \frac{3}{20}\right] \cdot \mathrm{H}\left(Y' \;\middle|\; g(Y') \leq \frac{3}{20}\right) + \\ &\Pr\left[g(Y') > \frac{3}{20}\right] \cdot \mathrm{H}\left(Y' \;\middle|\; g(Y') > \frac{3}{20}\right) + \mathrm{H}_{\mathbf{B}}\left(\Pr[g(Y') \leq \frac{3}{20}]\right)\end{aligned}$$

Trivially, $\mathrm{H}_{\mathbf{B}}\left(\Pr[g(Y') \leq \frac{3}{20}]\right) \leq 1$ and $\mathrm{H}\left(Y' \;\middle|\; g(Y') > \frac{3}{20}\right) \leq n$. Additionally:

$$\mathrm{H}\left(Y' \;\middle|\; g(Y') \leq \frac{3}{20}\right) \leq \log\left(\frac{2^n}{e^{\left(\frac{n}{4 \cdot \log(e)}\right)}}\right) = \frac{3}{4} \cdot n$$

Altogether, this implies:

$$\mathrm{H}(Y') \leq \Pr\left[g(Y') \leq \frac{3}{20}\right] \cdot \frac{3}{4}n + \Pr\left[g(Y') > \frac{3}{20}\right] \cdot n + 1$$

$$\mathrm{H}(Y') \leq \left(\frac{1}{4}\Pr\left[g(Y') > \frac{3}{20}\right] + \frac{3}{4}\right) \cdot n + 1$$

If $\Pr[g(Y') > \frac{3}{20}] < 4(k_2 - \frac{3}{4})$ then for sufficiently large n we have $\mathrm{H}(Y') < k_2 n$, which contradicts (3.3). □

Now let us reason about $\mathbb{E}\left[\epsilon(X'', Y', \hat{\tau})\right]$. By Claim 4 we have that:

$$\mathbb{E}\left[\epsilon(X'', Y', \hat{\tau})\right] = \mathbb{E}[g(Y')] > \Pr\left[g(Y') > \frac{3}{20}\right] \cdot \frac{3}{20} \geq 4 \cdot \left(k_2 - \frac{3}{4}\right) \cdot \frac{3}{20}$$

By equation (3.6) we have that:

$$\mathbb{E}[\epsilon(X'', Y', \hat{\tau})] \leq \frac{2k_1}{k_2 - \frac{3}{4}}$$

Putting these together:

$$\frac{2k_1}{k_2 - \frac{3}{4}} \geq \mathbb{E}\left[\epsilon(X'', Y', \hat{\tau})\right] > 4 \cdot \left(k_2 - \frac{3}{4}\right) \cdot \frac{3}{20}$$

Thus:

$$k_1 > \left(k_2 - \frac{3}{4}\right)^2 \cdot \frac{3}{10} \quad \square$$

Chapter 4

Conclusion

4.1 Conclusions and Open Problems

The randomized communication complexity lower bound for the inner product function proved here is not the strongest known bound, nor even the strongest bound proven using information complexity techniques. Nonetheless, it is still of interest. The proof's information complexity approach provides bounds on the direct sum of problem instances, which is not possible for discrepancy based methods. This proof differs from the information complexity proof implicit in [1] in its use of the natural uniform distribution, and notably different inner product function properties. We contend that our proof is the most straightforward of the three.

The immediate question is how this proof can be improved. Since it was finished very recently and particular constants were chosen for the sake of simplicity, an extended review may reveal substantial simplifications and improvements. Hopefully this technique may be applied to other problems for which previous methods are ineffective.

4.2 Acknowledgments

This work is attributed to Amit Chakrabarti, and Ranganath Kondapally, and myself.

My foremost thanks go to my advisor Amit Chakrabarti for his guidance, advice, and patience. I also heartily thank Ranganath Kondapally for his devoted assistance.

Bibliography

[1] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702-732, 2004.

[2] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. Direct Sums in Randomized Communication Complexity. *ECCC Report TR09-044*, 2009.

[3] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity. *In Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science* 270-278, 2001.

[4] Tomàs Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM Journal on Computing*, 24(4):736-750, 1995.

[5] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, 1997.

[6] Mihai Pătraçscu. *Randomized Lower Bounds for Lopsided Set Disjointness*, 2009.