On Randomized One-Round Communication Complexity

Ilan Kremer Noam Nisan Dana Ron
Institute of Computer Science
Hebrew University, Jerusalem 91904, Israel
Email: {pilan,noam,danar}@cs.huji.ac.il

November 30, 1994

Abstract

We present several results regarding randomized one-round communication complexity. These include a connection to the VC dimension, a study of the problem of computing the inner product of two real valued vectors, and a relation between "simultaneous" protocols and one-round protocols.

1 Introduction

In this paper we are concerned with randomized two-party communication complexity as defined by Yao [14]: Alice holds an input x, Bob holds an input y, and they wish to compute a given function $f(x, y)$, to which end they communicate with each other via a randomized protocol. We allow them bounded, two-sided error.

We study very simple types of protocols which include only one round of communication. These protocols were also studied by Ablayev [1]. In a one-round protocol, Alice is allowed to send a single message (depending upon her input x and upon her random coin flips) to Bob who must then be able to compute the answer $f(x, y)$ (using the message sent by Alice, his input, y, and his random coin flips). We denote the communication complexity of f (i.e. the number of bits of communication that need to be transmitted by such a protocol that computes f) by $R^{A\to B}(f)$. We also consider the situation where f is not symmetric and the roles of Alice and Bob are reversed, and denote this complexity by $R^{B\to A}(f)$. Finally, we consider an even more restricted scenario, the "simultaneous" one, where Both Alice and Bob transmit a single message to a "referee", Caroll, who sees no part of the input but who must then be able to compute $f(x, y)$ just from these two messages. This complexity we denote by $R^{||}(f)$.

While one-round communication complexity is quite trivial in the deterministic case, it turns out to be more interesting in the randomized case. In fact, even the simultaneous case poses difficulties. The reader may try to analyse the complexity of the following problem: in the equality problem, EQ, Alice and Bob each get an n-bit string. Caroll should return "true" *iff* the two strings are equal.

Open Problem: What is $R^{||}(EQ)$?

While we cannot believe that this question is very difficult, we have not been able to answer it. We suspect that the complexity is $\Omega(n)$. Let us remark that if public coins are allowed then $O(1)$ bits suffice.

Our results consider three main questions.

1.1 The VC Dimension and related Combinatorial Dimensions

We first observe the (surprising) fact that the VC dimension of a function class, f_X, defined by the set of rows of the matrix associated with f (denoted by $VC\text{-}dim(f_X)$), provides a lower bound on $R^{A\to B}(f)$. This has been independently observed by Ablayev [1]. We then study the power of this lower bound technique and show that it essentially characterizes the distributional complexity of the worst-case *product* distribution (denoted by $R^{A\to B,\Box}(f)$.)

Theorem: $R^{A\to B}(f) \geq R^{A\to B,\Box}(f) = \Theta(VC\text{-}dim(f_X))$.

For the case of non-boolean functions, we get similar bounds on $R^{A\to B}(f)$ and $R^{A\to B,\Box}(f)$ in terms of the *combinatorial* dimension which is a generalization of the VC dimension.

We also present several applications of this theorem. These include an example (the "greater than" function, GT) for which there is a large gap between $R^{A\to B}(f) = \Omega(n)$ and $R^{A\to B,\Box}(f) = O(1)$, and an example (the "index function", IND) for which there is an exponential gap between $R^{A\to B}(f)$ and $R^{B\to A}(f)$.

1.2 Real Inner Product

In this problem Alice and Bob each get an n-dimensional real valued vector and they wish to compute the inner product $f(\vec{x}, \vec{y}) = (\vec{x}, \vec{y}) = \sum_i x_i y_i$. We limit ourselves to cases where there is an apriory bound on certain norms of $\vec{x}$ and $\vec{y}$, a bound which will ensure that $|(\vec{x}, \vec{y})| \leq 1$. We consider two variants of these problems. In the first we ask that the inner product be computed within some given additive error, ϵ. In the second we define a *partial* boolean function whose value is 1 if the value of the inner product is above

2/3, and is 0 if it is below 1/3, and ask that this function be computed. Clearly, the second variant is easier than the first.

We first consider the case in which $\|\vec{x}\|_2 \leq 1$ and $\|\vec{y}\|_2 \leq 1$, where $\|\vec{x}\|_2 = \sqrt{\sum_i x_i{}^2}$. (This of course ensures that $|(\vec{x}, \vec{y})| \leq 1$.) If we denote this problem by $INP_{2,2}$, then we show:

Theorem: *$R^{A\to B,pub}(INP_{2,2}) = \Theta(1)$, and $R^{A\to B}(INP_{2,2}) = \Theta(\log n)$.*

The second case we consider is one in which $\|\vec{x}\|_1 \leq 1$ and $\|\vec{y}\|_\infty \leq 1$, where $\|\vec{x}\|_1 = \sum_i |x_i|$ and $\|y\|_\infty = \max_i |y_i|$. Again, this ensures that $|(\vec{x}, \vec{y})| \leq 1$. If we denote this function by $INP_{1,\infty}$, and the related partial boolean function by $INP^{par}_{1,\infty}$, then we obtain:

Theorem: *$R^{A\to B}(INP_{1,\infty}) = O(logn)$, but, $R^{B\to A}(INP^{par}_{1,\infty}) = \Omega(n)$. Furthermore, $INP^{par}_{1,\infty}$ is complete for the class of (boolean) functions whose one-round randomized complexity is polylog(n).*

Here completeness is under rectangular reductions as introduced by Babai, Frankl, and Simon [3].

We also present an interesting application of the above theorems. Let B^n_1, B^n_2, and B^n_∞, be the sets of n-dimensional real vectors whose L_1, L_2, and L_∞ norm, respectively, is bounded by 1. Then we prove the following theorem.

Theorem: *(1) For every constant $0 < \epsilon < 1$, there exist mappings, $g_{2,1} : B^n_2 \to B^{n'}_1$, and $g_{2,\infty} : B^n_2 \to B^{n'}_\infty$, where $n' = poly(n)$, such that for every pair of vectors, $\vec{u}, \vec{v} \in B^n_2$, $|(\vec{u}, \vec{v}) - (g_{2,1}(\vec{u}), g_{2,\infty}(\vec{v}))| \leq \epsilon$. (2) For any given integer n', there do* not *exist mappings, $g_{1,2} : B^n_1 \to B^{n'}_2$, and $g_{\infty,2} : B^n_\infty \to B^{n'}_2$, such that for every pair of vectors, $\vec{u} \in B^n_1$, $v \in B^n_\infty$, $|(\vec{u}, \vec{v}) - (g_{1,2}(\vec{u}), g_{\infty,2}(\vec{v}))| \leq 1/6$.*

1.3 One-Round vs. Simultaneous

It is clear that $R^{||}(f) \geq R^{A\to B}(f) + R^{B\to A}(f)$. Can this inequality be sharp? It turns out that the analogous question for deterministic communication is no: If f has a k-bit one-round deterministic protocol in which Alice sends a message to Bob and another l-bit one-round deterministic protocols in which Bob sends a message to Alice, then f has a $k + l$-bit simultaneous deterministic protocol.

We can prove a similar statement for the randomized case if Alice and Bob are allowed shared, public coins.

Theorem: *$R^{||,pub}(f) = \Theta(R^{A\to B,pub}(f) + R^{B\to A,pub}(f))$.*

We do not know whether the same is true for the private coins model. We also do not know how big can the gap between $R^{||}(f)$ and $R^{||,pub}(f)$ be.

2 Preliminaries

A two-party *communication problem* is a problem in which two players, *Alice* and *Bob*, wish to compute the value of a function $f : X \times Y \to Z$, on a given pair of inputs, $x \in X$ and $y \in Y$, where X, Y, and Z are arbitrary sets. The difficulty in this problem is that only Alice knows x, and only Bob knows y. However, they are allowed to *communicate* by sending messages (bits, or strings of bits) between themselves according to some *protocol* P. The *cost* of P on a given input (x, y) is the number of bits sent by Alice and Bob when given that input. The cost of P is the worst case cost over all inputs. The *communication complexity* of a function f is the minimum cost over all protocols that compute f.

In this work we are primarily interested in *1-round protocols* (or *1-way protocols*) which are composed of one round of communication. Namely, Alice sends a single message to Bob, and then Bob computes the output of the protocol. If f is not symmetric, then we also study the case in which the communication is in the opposite direction, i.e., from Bob to Alice. We are interested both in *deterministic* protocols, and in the following two variants of *randomized* protocols. In the first variant, Alice and Bob each have their own *private-coin* as a source of randomness, and they do not have access to the random string which is the outcome of the coin flips of the other player. In the second variant, Alice and Bob have a common *public-coin*. In other words, they have access to the same random string.

We also study randomized *simultaneous* protocols. In a simultaneous protocol we have three players: Alice, Bob, and Caroll. As in 1-round protocols, Alice and Bob each get a part of the input to the function. Caroll does not receive any input, and does not have access neither to Alice's nor to Bob's input. Instead of communicating among themselves, Alice and Bob each send a single message to Caroll and she computes the output of the protocol.

For a function $f : X \times Y \rightarrow \{0,1\}$, and $0 < \epsilon < 1$, we use the following notations (exact definition are given in Appendix A).

- $R_\epsilon^{A\rightarrow B}(f)$ denotes the randomized private-coin 1-round communication complexity of f with error probability ϵ.
- $R_\epsilon^{A\rightarrow B,pub}(f)$ denotes the randomized public-coin 1-round communication complexity of f with error probability ϵ.
- $R_\epsilon^{||}(f)$ denotes the randomized simultaneous communication complexity of f with error probability ϵ.
- For a probability distribution μ on $X \times Y$, $D_\epsilon^{A\rightarrow B,\mu}(f)$ denotes the 1-round **μ**-distributional complexity of f with error probability ϵ (where the error is with respect to μ).

In the case where the communication is required to be in the opposite direction (i.e., from Bob to Alice) we simply change the superscript in the notations from $A \rightarrow B$ to $B \rightarrow A$.

We usually assume that ϵ is a constant (smaller than $1/3$). In this case we are not interested in the exact dependence of the communication complexity on ϵ, and it is omitted from our notations (e.g., $R_\epsilon^{A\rightarrow B}(f)$ with constant ϵ, is simply denoted by $R^{A\rightarrow B}(f)$).

When f is a real valued function, then we also allow the protocol an *approximation* error ϵ_2. In this case we simply adapt the notations above by adding ϵ_2 as a subscript (and referring to ϵ as ϵ_1, e.g., $R_{\epsilon_1,\epsilon_2}^{A\rightarrow B}(f)$). We usually assume that ϵ_2 is a constant (smaller than $1/6$), and it is omitted from our notations.

The following fundamental theorem proven by Yao [15], characterizes public-coin complexity in terms of distributional (deterministic) complexity.

Theorem 1 ([15]) *For every function $f : X \times Y \rightarrow \{0,1\}$, and for every $0 < \epsilon < 1$,*

$$R_\epsilon^{A\rightarrow B,pub}(f) = max_\mu D_\epsilon^{A\rightarrow B,\mu}(f),$$

where μ ranges over all distributions on $X \times Y$.

We are sometimes interested in the following special case of distributional complexity.

DEFINITION 2.1 *A distribution μ over $X \times Y$ is called a* product distribution *(or a* rectangular distribution*) if for some distributions μ_X over X and μ_Y over Y, $\mu(x,y) = \mu_X(x)\mu_Y(y)$, for every $x \in X$, $y \in Y$. Let $R_\epsilon^{A\rightarrow B,[]}(f)$ denote $max_\mu D_\epsilon^{A\rightarrow B,\mu}$, where the maximum is taken over all product distributions.*

Finally, we shall need the following theorem of Newman [10], which gives an upper bound on private-coin complexity in terms of public-coin complexity.

Theorem 2 ([10]) *Let $f : X \times Y \rightarrow \{0,1\}$, where X and Y are arbitrary finite sets. For every $\delta > 0$ and every $0 < \epsilon < 1$,*

$$R_{\epsilon+\delta,}^{A\rightarrow B}(f) \leq R_\epsilon^{A\rightarrow B,pub}(f) + O(\log\log(|X| \cdot |Y|) + \log(1/\delta)).$$

3 1-Round Random Communication Complexity and Combinatorial Dimensions

Let $f : X \times Y \rightarrow Z$ be a function whose random communication complexity we would like to study. For $x \in X$ we define $f_x : Y \rightarrow Z$ as follows: for every $y \in Y$, $f_x(y) \stackrel{\text{def}}{=} f(x,y)$. Let $f_X \stackrel{\text{def}}{=} \{f_x \mid x \in X\}$. Similarly, for $y \in Y$ let $f_y(x) = f(x,y)$, and $f_Y \stackrel{\text{def}}{=} \{f_y \mid y \in Y\}$. In this section we show that if f is a boolean function then the *Vapnik Chervonenkis* (VC) *dimension* of f_X gives a lower bound on $R^{A\rightarrow B}(f)$, and an upper bound on $R^{A\rightarrow B,\square}(f)$. If f is not a boolean function, then a related combinatorial dimension gives similar bounds.

3.1 Boolean Functions

We start by recalling the definition of the VC dimension [13, 4].

DEFINITION 3.1 *Let H be class of boolean functions over a domain Y. We say that a set $S \subseteq Y$ is* shattered *by H if for* every *subset $R \subseteq S$ there exists a function $h_R \in H$ such that $\forall y \in S$, $h_R(y) = 1$ iff $y \in R$. The largest value d for which there exists a set S of size d that is shattered by H is the* VC dimension *of H, denoted by $VC\text{-}dim(H)$. If arbitrarily large finite sets can be shattered by H, then $VC\text{-}dim(H) = \infty$.*

Then we have the following theorem.

Theorem 3 *For every function $f : X \times Y \rightarrow \{0,1\}$, and for every constant error $\epsilon \leq 1/8$,*

$$R_\epsilon^{A\rightarrow B}(f) \;\geq\; R_\epsilon^{A\rightarrow B,\square}(f) \;=\; \Theta(VC\text{-}dim(f_X)).$$

Before proving this theorem, we discuss a simple example which illustrates how this theorem can be applied.

Example 1 *For $x,y \subseteq [1..n]$, $DISJ(x,y)$ is defined to be* 1 iff *$x \bigcap y = \emptyset$. We apply Theorem 3 to show that $R^{A\rightarrow B}(DISJ) = \Omega(n)$. Let S be the set of all singleton sets. It is easy to see that S is shattered by $DISJ_X$, since for every subset R of the singletons, there exists a set x such that for each $y \in S$, $DISJ_x(y) = 1$* iff *$y \in R$, where x is simply the complement of the union of all these singletons. It is interesting to note, that though it is known that even for multi-round communication complexity $R(DISJ) = \Omega(n)$, [7, 12], the upper bound on multi-round rectangular distributional complexity is lower [3]: $R^{\square}(DISJ) = O(\sqrt{n}\log(n))$. Thus, our example shows a quadratic gap between $R^{A\rightarrow B,\square}(f)$, and $R^{\square}(f)$.*

Proof of Theorem 3 : According to Yao's theorem quoted in Theorem 1, $R_\epsilon^{A\rightarrow B,pub}(f) = max_\mu D_\epsilon^{A\rightarrow B,\mu}(f)$, where the maximum is taken over *all* distributions μ. $R_\epsilon^{A\rightarrow B,\square}(f)$, on the other hand, is defined to be the maximum of $D_\epsilon^{A\rightarrow B,\mu}(f)$ taken over all *product* distributions μ. Clearly, $R_\epsilon^{A\rightarrow B}(f) \geq R_\epsilon^{A\rightarrow B,pub}(f)$ and hence $R_\epsilon^{A\rightarrow B}(f) \geq R_\epsilon^{A\rightarrow B,\square}(f)$. It thus remains to prove the upper and lower bounds on $R_\epsilon^{A\rightarrow B,\square}(f)$.

$\underline{\boldsymbol{R_\epsilon^{A\rightarrow B,\square}(f) = \Omega(d)}}$**:** In order to prove this lower bound we describe a product distribution μ for which $D_\epsilon^{A\rightarrow B,\mu}(f) = \Omega(d)$, where $d = VC\text{-}dim(f_X)$. By definition of the VC dimension, there exists a set $S \subseteq Y$ of size d which is shattered by f_X. Namely, for every subset $R \subseteq S$ there exists $x_R \in X$, such that $\forall y \in S$, $f_{x_R}(y) = 1$ *iff* $y \in R$. Clearly, there may be many such x's, so we choose one arbitrarily. Let μ be the uniform distribution over the set of pairs $\{(x_R, y) \mid R \subseteq S,\ y \in S\}$.

Let P be a single round deterministic protocol for computing $f(\cdot,\cdot)$ whose communication complexity is at most $d/15$. Thus, P defines two mappings. $P_1 : \{0,1\}^d \rightarrow \{0,1\}^{d/15}$, determines which $d/15$ bits should

Alice send to Bob for every given x_R, and $P_2 : \{0,1\}^{d/15} \rightarrow \{0,1\}^d$ determines the value of f computed by Bob for every $y \in S$, given the $d/15$ bits sent by Alice. Combining these two mappings together, P defines a mapping $P_{1,2} \stackrel{\text{def}}{=} P_1 \cdot P_2$ from $\{0,1\}^d$ into a set $U \subset \{0,1\}^d$, where $|U| = 2^{d/15}$. The expected error of P is $1/(d2^d)\sum_{z\in\{0,1\}^d} dist(z, P_{1,2}(z))$, where $dist(\cdot,\cdot)$ denotes the hamming distance between the vectors. If we define $dist(z, U)$ to be the minimum hamming distance between z and a vector $u \in U$, then the following lemma gives us the lower bound stated in the theorem.

Lemma 3.1 *For every $d \geq 15$, and for every set $U \subset \{0,1\}^d$, $|U| = 2^{d/15}$,*

$$E[dist(z, U)] > d/8,$$

where the expectation is taken with respect to the uniform distribution over $z \in \{0,1\}^d$.

Proof: For each $u \in U$, let $N_u = \{z \mid dist(z, u) \leq d/4\}$. We show that $|\cup_u N_u| \leq \sum_u |N_u| \leq 2^{d-1}$, and hence $E[dist(z, U)] > d/8$. For each u,

$$|N_u| = \sum_{i=0}^{d/4} \binom{d}{i} \;\leq\; (ed/(d/4))^{d/4} \;=\; 2^d(\log(4e)/4) \;<\; 2^{.861d}\ .$$

Thus,

$$\sum_u |N_u| \;<\; 2^{d/15+.861d} \;<\; 2^d \cdot 2^{-d/15}\ . \tag{1}$$

If $d \geq 15$ then the claim follows. ■ (Lemma 3.1 and $R_\epsilon^{A\rightarrow B,\square}(f) = \Omega(d)$).

$\boldsymbol{R_\epsilon^{A\rightarrow B,\square}(f) = O(d)}$**:** For this claim we need the following theorem [4], which is one of the most fundamental theorems in computational learning theory.

Theorem [4] *Let H be class of boolean functions over a domain Y with VC dimension d, let π be an arbitrary probability distribution over Y, and let $0 < \epsilon, \delta < 1$. Let L be any algorithm that takes as input a set $S \in Y^m$ of m examples labeled according to an unknown function $h \in H$, and outputs a hypothesis function $h' \in H$ that is* consistent *with h on the sample S. If L receives a random sample of size $m \geq m_0(d, \epsilon, \delta)$ distributed according to π^m, where*

$$m_0(d, \epsilon, \delta) = c_0\left(\frac{1}{\epsilon}\log\frac{1}{\delta} + \frac{d}{\epsilon}\log\frac{1}{\epsilon}\right)$$

for some constant $c_0 > 0$, then with probability at least $1-\delta$ over the random samples, $Pr_\pi[h'(y) \neq h(y)] \leq \epsilon$.

We next show that for every rectangular distribution μ, there exists a deterministic protocol whose (μ, ϵ)-distributional complexity is $O(\log(1/\epsilon) \cdot d/\epsilon)$.

Let $\mu : X \times Y \rightarrow \{0,1\}$ be a product distribution over $X \times Y$, where for every $x \in X$, $y \in Y$, $\mu(x, y) = \mu_X(x)\mu_Y(y)$. Consider the following family of deterministic protocols. For every set $S = (y_1, \ldots, y_m)$, of $m = m_0(d, \epsilon/2, \epsilon/2)$ points, where $y_i \in Y$, and m_0 is as defined in the theorem above. let P_S be the following protocol. For a given input (x, y), Alice sends Bob the value of $f_x(\cdot)$ on each point in S; Bob then finds a function $f_{x'} \in f_X$ that is consistent with the labeling sent by Alice, and outputs $f_{x'}(y)$. We define the following probability distribution, Π, on this family of deterministic protocols: $\Pi(P_S) = \mu^m(S)$. Then we have that

$$\forall x \in X \quad Pr_\Pi[Pr_{\mu_Y}[P_S(x, y) \neq f(x, y)] > \epsilon/2] < \epsilon/2\ .$$

It directly follows that

$$Pr_\Pi[Pr_\mu[P_S(x, y) \neq f(x, y)] > \epsilon/2] < \epsilon/2\ ,$$

and hence

$$E_\Pi[Pr_\mu[P_S(x, y) \neq f(x, y)]] < \epsilon.$$

Therefore, there exists at least one deterministic protocol whose error probability with respect to μ is bounded by ϵ, as required. ■ ($R_\epsilon^{A\rightarrow B,\square}(f) = O(d)$ and Theorem 3)

3.2 Non-Boolean Functions

In the case where the range of f (and f_X) is not $\{0,1\}$, we show that a related combinatorial dimension gives similar bounds to those obtained in the boolean case. Details are given in Appendix B.

3.3 Applications

3.3.1 $R^{A\to B}(f)$ vs. $R^{B\to A}(f)$

In our first application we use Theorem 3 to prove the following gap between $R^{A\to B}(f)$ and $R^{B\to A}(f)$, which seems to be folklore.

Theorem 4 *There exists a function f for which $R^{A\to B}(f) = \Theta(\log n)$, while $R^{B\to A}(f) = \Theta(n)$.*

Proof: Let $IND : [1..n] \times \{0,1\}^n \to \{0,1\}$, the "index" function, be defined as follows: $IND(i, \vec{y}) = y_i$. The upper bounds on $R^{A\to B}(IND)$ and $R^{B\to A}(IND)$ are clear since each party can simply send the other party its input.

In order to prove the lower bound on $R^{B\to A}(IND)$, we show that the (whole) set $[1..n]$ is shattered by IND_Y. For every subset R of $[1..n]$, let $\vec{y}(R)$ be defined as follows: $y_i(R) = 1$ *iff* $i \in R$. Then by definition $IND_{\vec{y}(R)}(i) = 1$ *iff* $i \in R$, and $[1..n]$ is shattered by IND_Y.

To see that the upper bound on $R^{A\to B}(IND)$ is also tight we define the following set $S \subset \{0,1\}^n$, $S = \{\vec{y}^1, \ldots, \vec{y}^{\log n}\}$. For every $1 \le j \le \log n$, and for every $1 \le i \le n$, $y_i^j = 1$, *iff* the j'th coordinate in the binary representation of i is 1. It is not hard to verify that S is shattered by IND_X. ■

3.3.2 $R^{A\to B}(f)$ vs. $R^{A\to B,\Box}(f)$

We next show the existence of a function f for which there is a large gap between $R^{A\to B}(f)$ and $R^{A\to B,\Box}(f)$. It is interesting to note that in the case of multi rounds communication complexity there are only examples showing a polynomial (quadratic) gap between $R^{\Box}(f)$ and $R(f)$ [3, 7, 12].

Theorem 5 *There exists a function f for which $R^{A\to B,\Box}(f) = O(1)$, while $R^{A\to B}(f) = \Omega(n)$.*

Proof: Let $GT : [1..2^n] \times [1..2^n] \to \{0,1\}$, the "greater than" function, be defined as follows: $GT(x,y) = 1$ *iff* $x > y$. The VC dimension of GT_X is exactly 1, since almost every set $\{y\}$ of size one can be shattered by GT_X, while for every set $\{y_1, y_2\}$, $y_1 < y_2$, there is no x satisfying $x > y_2$ and $x \le y_1$. Applying Theorem 3, we get the upper bound on $R^{A\to B,\Box}(GT)$.

The proof of the lower bound on $R^{A\to B}(GT)$ is given in [9]. We describe a simpler proof to a slightly weaker bound of $\Omega(n/\log n)$. Let P be a randomized private-coin protocol for GT whose cost is $R^{A\to B}(GT)$. Assume that for a given x and y, Alice and Bob execute P, $k \log n$ times for some constant k, and Bob outputs the majority outcome. Then, for any given y, the probability of error decreases to $n^{-\Omega(1)}$. Furthermore, with high probability, for a given x and for any *set* $\{y_1, \ldots, y_n\}$, Bob can compute the correct value of $GT(x, y_i)$ for *every* y_i in the set. It follows that Bob can use this protocol to perform a binary search and find the *exact* value of x with high probability. In particular, he can compute each coordinate in the binary representation of x. But we have just shown in the proof of Theorem 4 that $R^{B\to A}(IND) = \Omega(n)$, and hence, $R^{A\to B}(GT) = \Omega(n/\log n)$. ■

4 Real Inner Products

In this section we study problems in which Alice and Bob each receive an n-dimensional real-valued vector, and they wish to compute the inner product between the two vectors. We assume that there are known bounds on certain norms of the vectors. We consider two variants of these problems. In the first we ask that the inner product be computed within some small given additive error ϵ_2. In the second we define a *partial* boolean function whose value is 1 if the value of the inner product is above 2/3, and is 0 if it is below 1/3, and ask that this partial function be computed. Our upper bounds all apply to the first variant with *any* constant error ϵ_2.

4.1 The bounded L_2 norm case

We start with the problem of computing the inner product of two n-dimensional vectors whose L_2 norm is bounded by 1. More precisely, we define:

$$INP_{2,2}(\vec{u}, \vec{v}) \stackrel{\text{def}}{=} (\vec{u}, \vec{v}) \stackrel{\text{def}}{=} \sum_i u_i v_i,$$

where $\|\vec{u}\|_2 (= \sqrt{\sum u_i^2}) \leq 1$, and $\|\vec{v}\|_2 \leq 1$.

Theorem 6

$$R^{A \to B, pub}(INP_{2,2}) = \Theta(1) \quad \textit{and} \quad R^{A \to B}(INP_{2,2}) = \Theta(\log n).$$

Proof: Without loss of generality, we assume that the L_2 norm of both $\vec{u}$ and $\vec{v}$. is exactly 1. If this is not the case then Alice can simply send bob the L_2 norm of $\vec{u}$ within a small additive error, using a constant number of bits. In the public-coin protocol we are about to describe, we apply a technique presented in [5]. In particular, we need the following lemma:

Lemma [5] *Let $\vec{u}$ and $\vec{v}$ be two n-dimensional real vectors whose L_2 norm is 1. Let $\vec{r}$ be a random n-dimensional real vector whose L_2 norm is 1. Then:*

$$Pr[sgn(\vec{u}, \vec{r}) \neq sgn(\vec{v}, \vec{r})] = \frac{1}{\pi} \arccos(\vec{u}, \vec{v}) \ ,$$

where $sgn(x) = 1$, if $x \geq 0$, and 0 otherwise.

Given this lemma, the public-coin protocol is quite obvious. Alice and Bob use their common random string to choose k random vectors $\vec{r}_1, \ldots, \vec{r}_k$, for some constant k, such that $\forall i \ \|\vec{r}_i\|_2 = 1$. Alice sends Bob $(\vec{u}, \vec{r}_1), \ldots, (\vec{u}, \vec{r}_k)$, and Bob estimates $\arccos(\vec{u}, \vec{v})$ by $\frac{\pi}{k} |\{\vec{r}_i \mid sgn(\vec{u}, \vec{r}_i) \neq sgn(\vec{v}, \vec{r}_i)\}|$. Since the absolute value of the derivative of the *cosine* function is bounded by 1, Bob can use this estimate to compute, with high probability, the value of $(\vec{u}, \vec{v})$ within an additive error $\epsilon_2 = O(1/\sqrt{k})$.

In order to get the upper bound on $R^{A \to B}(INP_{2,2})$, we apply Newman's theorem [10], quoted in Theorem 2. Though the theorem is not directly applicable in our case since it applies to functions defined on finite domains X and Y, we need only make the following observation. Assume that the public-coin protocol described above is run on bounded precision representations of $\vec{u}$ and $\vec{v}$, where each coordinate is written in $O(n)$ bits. Then we incur only an exponentially small additional error, while the size of our effective domain is $2^{poly(n)}$. In Appendix C we give an alternative proof to a slightly weaker claim ($R^{A \to B}(INP_{2,2}) = polylog(n)$), in which we *explicitly* describe a private-coin protocol for the function (the application of Newman's theorem only proves the *existence* of such a protocol).

It remains to show that $R^{A \to B}(INP_{2,2}) = \Omega(\log n)$ (clearly $R^{A \to B, pub}(INP_{2,2}) = \Omega(1)$). Assume, contrary to the claim, that there exists a protocol, P, for $INP_{2,2}$ whose cost is $o(\log n)$, and let the approximation error of P, ϵ_2, be bounded by 1/5. Consider the following "intersection" function, INT: given two sets

$S, T \in [1..n]$, $|S| = |T| = n/2$, $INT(S,T)$ is the size of $S \bigcap T$. Then we can use P to compute $INT(S,T)$ (within an additive error of $n/10$) as follows. Let $u_i(S) = 1/\sqrt{S}$, if $i \in S$, and 0 otherwise. Define $\vec{v}(T)$ similarly. Thus, $(\vec{u}(S), \vec{v}(T)) = |S \bigcap T|/\sqrt{|S||T|} = 2|S \bigcap T|/n$. Since $\|\vec{u}(S)\|_2 = \|\vec{v}(T)\|_2 = 1$, we can run P to compute $(\vec{u}(S), \vec{v}(T))$, from which we can derive the size of $S \bigcap T$ within an additive error of $n/10$.

The protocol for INT gives us a randomized protocol whose cost is $o(\log n)$ for computing the "disjointness" function, $DISJ$, when the sets, S and T are know to be of size exactly $n/2$, and their intersection is either empty or of size at least $n/5$. It is well known that $D(f) = 2^{O(R(f))}$, and hence we get an upper bound of $o(n)$ on the deterministic complexity of the restricted version of $DISJ$ described above. However, it can be shown that the known linear lower $D(DISJ) = \Omega(n)$, holds in our restricted case as well. This can be done by proving the existance of an exponential size *fooling set*[1] $\mathcal{S}$ of pairs $\{(S, \bar{S})\}$, such that $|S| = n/2$, and for every two pairs $(S, \bar{S})$, and $(T, \bar{T})$ in $\mathcal{S}$, if $S \neq T$, then $|S \bigcap \bar{T}| = |\bar{S} \bigcap T| \geq n/5$. This existence can be shown using a simple probabilistic argument based on the fact that the expected size of the intersection of two random sets of size $n/2$, is $n/4$. ■

4.2 A Complete Problem for 1-Round Communication Complexity

In this section we present a fairly simple function which is complete for the class of boolean functions whose 1-round communication complexity is *polylog*(n). First we need to define *completeness* in this context. This is done using *rectangular reductions*, which were introduced in [3].

DEFINITION 4.1 *Let X, X', Y, Y' and Z be arbitrary sets. Let $f : X \times Y \to Z$, and let $f' : X' \times Y' \to Z$. A* rectangular reduction *from f' to f is a pair of functions $g_1 : X' \to X$, $g_2 : Y' \to Y$, which satisfy the following conditions:*

1. $\forall x' \in X'$, $y' \in Y'$, $f'(x', y') = f(g_1(x'), g_2(y'))$.
2. $\forall x' \in X'$, $|g_1(x')| = 2^{polylog(|x'|)}$.
3. $\forall y' \in Y'$, $|g_2(y')| = 2^{polylog(|y'|)}$.

If there exists a rectangular reduction from f' to f then we say that f' reduces to *f and we denote this by $f' \propto f$.*

Functions can be classified according to their communication complexity, similarly to the way in which they are classified according to their computational complexity. A *complete* function for a given communication complexity class is defined as follows.

DEFINITION 4.2 *We say that a function f is* complete *for a communication complexity class $\mathcal{C}$ if the following two conditions hold:*

1. $f \in \mathcal{C}$;
2. $\forall f' \in C$, $f' \propto f$.

Let $INP_{1,\infty}(\cdot,\cdot)$ be the following inner product function. $INP_{1,\infty}(\vec{p}, \vec{q}) = (\vec{p}, \vec{q})$, where $\vec{p}$ and $\vec{q}$ are n-dimensional vectors which have the following properties: $\|\vec{p}\|_1 \leq 1$ ($\sum_i |p_i| \leq 1$), and $\|\vec{q}\|_\infty \leq 1$ ($\max_i |y_i| \leq 1$). Let the *partial* boolean function $INP^{par}_{1,\infty}(\cdot,\cdot)$ be defined as follows:

$$INP^{par}_{1,\infty}(\vec{p}, \vec{q}) \stackrel{\text{def}}{=} \begin{cases} 1 & INP_{1,\infty}(\vec{p}, \vec{q}) \geq 2/3 \\ 0 & INP_{1,\infty}(\vec{p}, \vec{q}) \leq 1/3 \end{cases}$$

[1] A set $\mathcal{S} \subset X \times Y$ is called a fooling set for a function $f : X \times Y \to \{0,1\}$, if there exists $z \in \{0,1\}$ such that for every $(x, y) \in \mathcal{S}$, $f(x, y) = z$, while for every $(x_1, y_1) \neq (x_2, y_2)$, either $f(x_1, y_2) \neq z$ or $f(x_1, y_2) \neq z$. It is not hard to verify that if f has a fooling set of size t, then $D(f) \geq \log t$.

Theorem 7 *$R^{A\to B}(INP_{1,\infty}) = O(\log n)$, but $R^{B\to A}(INP^{par}_{1,\infty}) = \Omega(n)$. Furthermore, $INP^{par}_{1,\infty}$ is complete for the class of boolean functions whose 1-round communication complexity is polylog(n).*

Proof: Without loss of generality, we assume that $\vec{q}$ is non-negative, and that $\vec{p}$ is a probability vector, i.e., $\vec{p}$ is non-negative, and $\|\vec{p}\|_1 = 1$. We start by describing a 1-round randomized protocol for computing $INP_{1,\infty}$, whose cost is $O(\log n)$. Clearly, it follows that $INP^{par}_{1,\infty}$ belongs to the class of boolean functions whose 1-round communication complexity is $polylog(n)$. Alice repeats the following process k times, where k is a constant. She chooses an index i with probability p_i and sends it to Bob. For the ℓ'th repetition of this process, let X_ℓ be the value of the q_i corresponding to the index i sent by Alice. Bob then outputs the average of the X_ℓ's. The X_ℓ's are random variables which take values in $[0,1]$, and whose expect value is $\sum_{j=1}^{n} p_j \cdot q_j$, the inner product between $\vec{p}$ and $\vec{q}$. Applying Chernoff bounds, if $k = O(1/\epsilon_2^2 \log(1/\epsilon_1))$, then with probability at least $1-\epsilon_1$, the absolute value of the difference between the average of the X_ℓ's, and $(\vec{p},\vec{q})$ is at most ϵ_2. For constant ϵ_1 and ϵ_2, the cost of the protocol is thus $O(\log n)$.

Next, we describe a rectangular reduction from any given $f : X \times Y \to \{0,1\}$ for which $R^{A\to B}(f) = polylog(n)$, to $INP^{par}_{1,\infty}$. If $R^{A\to B}(f) = polylog(n)$ then there exists a 1-round communication protocol P for f that has the following properties. For every $x \in X$, Alice's side of the protocol defines a probability distribution over all messages of length c, where $c = polylog(n)$, and for each such message, and for every $y \in Y$, Bob's side of the protocols determines a probability of outputting 1. For $1 \le i \le 2^c$, let M_i denote the i'th message in some arbitrary enumeration of the messages Alice can send Bob. Let $p_i(x)$ be the probability that Alice sends the message M_i to Bob given that her input is x, and let $q_i(y)$ be the probability that Bob outputs 1 given that he received the input y, and that Alice sent him the message M_i. Thus, using the notations from Definition 4.1, we define $g_1(x)$ to be $\vec{p}(x)$, and $g_2(y) = \vec{q}(y)$. The dimension of both vectors is 2^c which is $2^{polylog(n)}$, and we let each coordinate be written with exponential precision, using $O(n)$ bits. It remains to be shown that $f(x,y) = INP^{par}_{1,\infty}(\vec{p}(x), \vec{q}(y))$. By definition of $\vec{p}$ and $\vec{q}$,

$$Pr[P(x,y)=1] \quad = \quad \sum_{i=1}^{2^c} p_i \cdot q_i \pm o(2^{-n}). \tag{2}$$

Since $Pr[P(x,y)=1]$ should be greater than 2/3 if $f(x,y)=1$ and smaller than 1/3 otherwise, the claim follows.

Finally, the lower bound on $R^{B\to A}(INP^{par}_{1,\infty})$ directly follows from lower bound on the index function, IND, proved in Theorem 4. ■

4.3 An Application – Inner Product preserving mappings

Let B_1^n, B_2^n, and B_∞^n, be the sets of n-dimensional real vectors whose L_1, L_2, and L_∞ norm, respectively, is bounded by 1. Then we have the following theorem.

Theorem 8

- *For every constant $0 < \epsilon < 1$, there exist mappings, $g_{2,1} : B_2^n \to B_1^{n'}$, and $g_{2,\infty} : B_2^n \to B_\infty^{n'}$, where $n' = poly(n)$, such that for every pair of vectors, $\vec{u}, \vec{v} \in B_2^n$, $|(\vec{u},\vec{v}) - (g_{2,1}(\vec{u}), g_{2,\infty}(\vec{v}))| \le \epsilon$.*
- *For any given integer n', there do* not *exist mappings, $g_{1,2} : B_1^n \to B_2^{n'}$, and $g_{\infty,2} : B_\infty^n \to B_2^{n'}$, such that for every pair of vectors, $\vec{u} \in B_1^n$, $v \in B_\infty^n$, $|(\vec{u},\vec{v}) - (g_{1,2}(\vec{u}), g_{\infty,2}(\vec{v}))| \le 1/6$.*

Proof sketch : The construction of the mappings for the first claim is very similar to the one described in Theorem 7. Let P be the private-coin protocol for computing $INP_{2,2}$, whose existence is proven in Theorem 6, (and whose cost is $O(\log n)$). For a vector $\vec{u} \in B_2^n$, let $g_{2,1}(\vec{u})$, be the probability vector over the messages Alice may send Bob according to P, given that she received $\vec{u}$ as input. Let $q_{i,j}(\vec{v})$ be the probability that Bob outputs $j \in [0,1]$, given that Alice sent the ith message, and he received $\vec{v}$ as input, and let $q_i(\vec{v}) = \sum_j (q_{i,j}(\vec{v}) \cdot j)$. We note that it can be shown that in P Bob is deterministic, [2] and hence $q_i(\vec{v})$ is simply the value Bob outputs given the input $\vec{v}$ and the ith message. Let $g_{2,\infty}(\vec{v}) = \vec{q}(\vec{v})$. Then $(g_{2,1}(\vec{u}), g_{2,\infty}(\vec{v}))$ is the expected value of Bob's output. But we know that for some constants $0 < \epsilon_1, \epsilon_2 < 1$, with probability at least $1 - \epsilon_1$, Bob's output is at most ϵ_2 away from the correct value of the inner product. Hence the expected value is at most $\epsilon_1 + \epsilon_2$ away from the correct value, and the claim follows.

The second claim follows from the upper bound $R^{A \to B, pub}(INP_{2,2}) = R^{B \to A, pub}(INP_{2,2}) = O(1)$ given in Theorem 6, and the lower bound $R^{B \to A}(INP^{par}_{1,\infty}) = \Omega(n)$ given in Theorem 7. □

5 Simultaneous Protocols

As stated in the introduction, we do not know bounds on $R^{||}(f)$ even for very simple functions such as the equality function, EQ. ¿From Theorem 4, we only know that $R^{B \to A}(f)$, and hence $R^{||}(f)$ may be exponentially larger than $R^{A \to B}(f)$. We do not know if $R^{||}(f)$ can be much larger than the sum of $R^{A \to B}(f)$ and $R^{B \to A}(f)$. However, if we allow public coins, then we have the following theorem.

Theorem 9 *For every boolean function f,*

$$R^{||, pub}(f) = \theta \left(R^{A \to B, pub}(f) + R^{B \to A, pub}(f) \right) \quad .$$

Proof sketch (Complete proof appears in the appendix)**:** It is clear that $R^{||, pub}(f) \geq R^{A \to B, pub}(f) + R^{B \to A, pub}(f)$, since both Alice and Bob can simulate Caroll's side of the protocol. We show that for every probability distribution μ on $X \times Y$, $D^{||, \mu}(f) = O\left(D^{A \to B, \mu}(f) + D^{B \to A, \mu}(f)\right)$, where $D^{||, \mu}(f)$ is the simultaneous μ-distributional complexity of f. The theorem follows since, similarly to the statement in Theorem 1, it holds that $R^{||, pub}(f) = max_\mu D^{||, \mu}(f)$. Let $P^{A \to B}$ be a deterministic protocol for f in which Alice sends Bob a single message of length $c_1 = D^{A \to B, \mu}_\epsilon(f)$, and whose error probability with respect to μ is ϵ. Let $P^{B \to A}$ be a protocol whose cost is c_2, and which is chosen analogously. We prove that there exists a deterministic simultaneous protocol $P^{||}$ whose cost is $c_1 + c_2$, and whose error with respect to μ is $O(\epsilon^{1/4})$. Furthermore, in $P^{||}$, Alice simply simulates her side of the protocol in $P^{A \to B}$, and Bob simulates his side in $P^{B \to A}$.

Let M_f be the "truth table" of f, i.e., it is an $|X|$ by $|Y|$ matrix where each row is labeled by some $x \in X$, each column is labeled by some $y \in Y$, and $M_f(x, y) = f(x, y)$. Since $P^{A \to B}$ induces a partition on the rows of M_f, and $P^{B \to A}$ induces a partition on the columns of M_f, $P^{||}$, according to our definition, induced a partition on pairs (x, y) into rectangles in M_f. We show that most rectangles are almost monochromatic, where "most" and "almost" are with respect to μ. □

References

[1] F. Ablayev. Private communication. 1994.

[2] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. Scale-sensetive dimensions, uniform convergence, and learnability. In *34th Annual Symposium on Foundations of Computer Science*, 1993.

[3] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, 1986.

[2] Actually, the proof of Newman's theorem (Theorem 2) implies that every randomized 1-round private-coin protocol can be transformed into a protocol in which Bob is deterministic while incurring small additional error and cost.

[4] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the Association for Computing Machinery*, 36(4):929–965, October 1989.

[5] M. X. Goemans and D. P. Williamson. .878-approximation algorithms for max cut and max 2sat. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on the Theory of Computing*, pages 422–431, 1994.

[6] D. Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100(1):78–150, 1992.

[7] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. In *2nd Structure in Complexity Theory Conference*, pages 41–49, 1987.

[8] M. J. Kearns and R. E. Schapire. Efficient distribution-free learning of probabilistic concepts. In *31st Annual Symposium on Foundations of Computer Science*, pages 382–391, October 1990. To appear, *Journal of Computer and System Sciences.*

[9] P. Bro Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. Manuscript, 1994.

[10] I. Newman. Private vs. common random bits in communication complexity. *Information processing letters*, 39:67–71, 1991.

[11] D. Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, 1984.

[12] A. Razborov. On the distributional complexity of disjointness. In *Proceedings of the ICALP*, pages 249–253, 1990. (to appear in *Theoretical Computer Science*).

[13] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its applications*, XVI(2):264–280, 1971.

[14] A. Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.

[15] A. Yao. Lower bounds by probabilistic arguments. In *24th Annual Symposium on Foundations of Computer Science*, pages 420–428, 1983.

A Definitions

Definition A.1 *Let X, Y, and Z be arbitrary sets, and let $f : X \times Y \to Z$ be an arbitrary function.*

- *A* deterministic 1-round communication protocol *P for f is a pair of functions $P_A : X \to \{0,1\}^c$, and $P_B : \{0,1\}^c \times Y \to Z$. The output of P on input (x,y) is $P(x,y) = P_B(P_A(x), y)$. The cost of P is c. Let μ be a probability distribution on $X \times Y$, and let $0 \le \epsilon_1, \epsilon_2 \le 1$. The* 1-round $(\boldsymbol{\mu}, \boldsymbol{\epsilon_1}, \boldsymbol{\epsilon_2})$-distributional complexity *of f, $D^{A \to B, \mu}_{\epsilon_1, \epsilon_2}(f)$, is defined to be the cost of the best deterministic protocol P for which $Pr_\mu[|P(x,y) - f(x,y)| > \epsilon_2] < \epsilon_1$. When $Z = \{0,1\}$, we omit ϵ_2, and $D^{A \to B, \mu}_{\epsilon}(f)$, is defined to be the cost of the best protocol P for which $Pr_\mu[P(x,y) \neq f(x,y)] < \epsilon$.*

- *A* randomized private-coin 1-round communication protocol *is a pair of functions $P_A : X \times \{0,1\}^{\rho_A} \to \{0,1\}^c$, and $P_B : \{0,1\}^c \times Y \times \{0,1\}^{\rho_B} \to Z$. The output of P on input (x,y), the (private) random coin tosses of Alice, $r_A \in \{0,1\}^{\rho_A}$, and the (private) random coin tosses of Bob, $r_B \in \{0,1\}^{\rho_B}$, is $P(x,y,r_A,r_B) = P_B(P_A(x,r_A), y, r_B)$. The cost of P is c. The* randomized private-coin 1-round communication complexity *of f, $R^{A \to B}_{\epsilon_1,\epsilon_2}(f)$, is defined to be the cost of the best randomized private-coin 1-round communication protocol P for which $Pr[|P(x,y,r_A,r_B) - f(x,y)| > \epsilon_2] < \epsilon_1$, where the probability is taken over the random coin tosses r_A and r_B. When $Z = \{0,1\}$, we omit ϵ_2, and $R^{A \to B}_{\epsilon}(f)$, is defined to be the cost of the best protocol P for which $Pr[P(x,y,r_A,r_B) \neq f(x,y)] < \epsilon$. A* randomized public-coin 1-round communication protocol *is defined similarly, only that the functions P_A and P_B are defined on a* common *random string r. The* randomized public-coin 1-round communication complexity *of a f is denoted by $R^{A \to B, pub}_{\epsilon_1,\epsilon_2}(f)$.*

- *A* randomized simultaneous communication protocol *is a triplet of functions $P_A : X \times \{0,1\}^{\rho_A} \to \{0,1\}^{c_1}$, $P_B : Y \times \{0,1\}^{\rho_B} \to \{0,1\}^{c_2}$, and $P_C : \{0,1\}^{c_1} \times \{0,1\}^{c_2} \times \{0,1\}^{\rho_C} \to Z$. The output of P on input (x,y), the random coin tosses of Alice, $r_A \in \{0,1\}^{\rho_A}$, the random coin tosses of Bob, $r_B \in \{0,1\}^{\rho_B}$, and the random coin tosses of Caroll, $r_C \in \{0,1\}^{\rho_C}$, is $P(x,y,r_A,r_B,r_C) = P_C(P_A(x,r_A), P_B(y,r_B), r_C)$. The cost of P is $c_1 + c_2$. The* randomized simultaneous communication complexity *of f, $R^{||}_{\epsilon_1,\epsilon_2}(f)$, is defined to be the cost of the best randomized simultaneous communication protocol P for which $Pr[|P(x,y,r_A,r_B,r_C) - f(x,y)| > \epsilon_2] < \epsilon_1$, where the probability is taken over the random coin tosses r_A, r_B and r_C. The small variations in notations for the cases where $Z = \{0,1\}$, and where ϵ_1 and ϵ_2 are constants smaller than $1/3$, is similar to those introduced for 1-way protocols.* Deterministic simultaneous communication protocols, *and* Randomized simultaneous public-coin communication protocols *can be defined similarly, and the related notations can be adapted as well, as in the 1-round case.*

B Bounds on the Communication Complexity of Non-Boolean Functions

In the case where the range of f (and f_X) is not $\{0,1\}$, we must consider the following generalization of the VC dimension, named the *combinatorial dimension*. This definition (as well as the next one), follow works of Pollard [11], Haussler [6], Kearns and Schapire [8], and Alon *et. al* [2].

Definition B.1 *Let H be class of functions over a domain Y into a range Z, and let $\gamma \ge 0$. We say that a set $S = \{y_1, \ldots, y_k\} \subseteq Y$, is* γ-shattered *by H if there exists a vector $\vec{w} = (w_1, \ldots w_k) \in Z^k$ of dimension $k = |S|$ for which the following holds. For* every *subset $R \subseteq S$, there exists a function $h_R \in H$ such that $\forall y_i \in S$, if $y_i \in R$ then $h_R > w_i + \gamma$, and if $y_i \notin R$ then $h_R < w_i - \gamma$. The largest value d for which there exists a set S of size d that is γ-shattered by H is the* γ-combinatorial dimension *of H and is denoted by C_γ-$dim(H)$. If arbitrarily large finite sets can be shattered by H, then C_γ-$dim(H) = \infty$*

We also need the following definition and theorem.

Definition B.2 *Let H be class of functions over a domain Y into a range Z, let π be a distribution over Y, and let $0 \leq \epsilon \leq 1$. A subclass $H_{\epsilon,\pi} \subseteq H$ is an* **ϵ-cover** *of H with respect to π if for every function $h \in H$, there exists a function $h' \in H_{\epsilon,\pi}$ such that*

$$E_\pi[|h(y) - h'(y)|] \leq \epsilon.$$

Lemma B.1 ([6] following [11]) *Let H be class of functions from a domain Y, into the interval $[0, M]$, and let $d = C_0\text{-}dim(H)$. For any given $0 \leq \epsilon \leq 1/2d$, and for any distribution π over Y, there exists an ϵ-cover $H_{\epsilon,\pi}$ of H with respect to π which is of size at most $2(2eM/\epsilon \ln(2eM/\epsilon))^d$.*

Similarly to the Boolean case, we have the following lower and upper bounds.

Theorem 10 *For every function $f : X \times Y \to [0, M]$, and every $\epsilon_1 \leq 1/8$,*

$$R^{A\to B}_{\epsilon_1,\epsilon_2}(f) \;\geq\; R^{A\to B,\square}_{\epsilon_1,\epsilon_2}(f) \;=\; \Omega(C_{\epsilon_2}\text{-}dim(f_X)).$$

Theorem 11 *For every function $f : X \times Y \to [0, M]$,*

$$R^{A\to B,\square}_{\epsilon_1,\epsilon_2}(f) = O(d \ln(M \cdot max(d, 1/(\epsilon_1\epsilon_2)))),$$

where $d = C_0\text{-}dim(f_X)$.

It should be noted that in some cases there might exist a large gap between the upper and lower bounds stated above due to the possible gap between $C_0\text{-}dim(f_X)$ and $C_\gamma\text{-}dim(f_X)$ for $\gamma > 0$. Not only might this gap be exponential, but there exist function classes H for which C_0-dim(H) is infinite, while C_γ-dim$(H) = O(1/\gamma)$ [2]. As described below, we prove the upper bound in Theorem 11 using the bound given in Lemma B.1 on the *ϵ-cover* of f_X in terms of C_0-dim(f_X). Thus, an interesting question is if there exist bounds on the size of the ϵ-cover of a function class H in terms of C_γ-dim(H), which can be applied to achieve a tighter upper bound. Alon *et. al.* [2] give a bound on the size of the ϵ-cover (which they define with respect to the l^∞ distance) of a function class H in terms of C_γ-dim(H), but their bound is not applicable to our problem since it also depends polynomially on the size of the domain.

Proof of Theorem 10 : Similarly to the proof of Theorem 3, we use the fact that $R^{A\to B,pub}_{\epsilon_1,\epsilon_2}(f) = max_\mu D^{A\to B,\mu}_{\epsilon_1,\epsilon_2}(f)$, and describe a rectangular distribution μ on $X\times Y$ for which $D^{A\to B,\mu}_{\epsilon_1,\epsilon_2}(f) = \Omega(C_{\epsilon_2}\text{-}dim(f_X))$.

Let $d = C_{\epsilon_2}\text{-}dim(f_X)$, and let $S = \{y_1, \ldots, y_d\} \subseteq Y$ be an ϵ_2-shattered set with respect to $\vec{w} = (w_1, \ldots, w_d)$. For every $R \subseteq S$, let $x_R \in X$ satisfy the following requirement. For every $y_i \in S$, if $y_i \in R$ then $f_{x_R}(y_i) > w_i + \epsilon_2$, and if $y_i \notin R$ then $f_{x_R}(y_i) < w_i - \epsilon_2$.

We define a boolean function $f' : X \times S \to \{0, 1\}$ as follows: $f'(x, y_i) = sign(f_x(y_i) - w_i)$. By this definition, for every $R \subseteq S$, there exists $x_R \in X$ such that for every $y_i \in S$, $f'_{x_R}(y_i) = 1$ *iff* $y_i \in R$, and hence, the VC dimension of f' is d. Let μ be the uniform distribution on $\{x_R \mid R \subseteq S\} \times S$. Clearly, $D^{A\to B,\mu}_{\epsilon_1,\epsilon_2}(f) \geq D^{A\to B,\mu}_{\epsilon_1}(f')$. But from the proof of Theorem 3 we know that $D^{A\to B,\mu}_{\epsilon_1}(f') = \Omega(d)$, which completes the proof. ■

Proof of Theorem 11 : Let $\mu : X \times Y \to \{0, 1\}$ be a product distribution over $X \times Y$, where for every $x \in X$, $y \in Y$, $\mu(x, y) = \mu_X(x)\mu_Y(y)$. We next describe a deterministic protocol P satisfying $Pr_\mu[|P(x, y) - f(x, y)| > \epsilon_2] < \epsilon_1$, whose cost is $O(d \ln(M \cdot max(d, 1/(\epsilon_1\epsilon_2))))$. Let $\epsilon = \epsilon_1\epsilon_2$. Alice and Bob choose in advance an ϵ-cover of f_X with respect to μ_Y. Given x, Alice sends to Bob a coding of a function $f_{x'}$ in the cover for which $E[|f_x(y) - f_{x'}(y)|]$ is minimized, where the expectation is taken with respect to μ_Y. According to Lemma B.1, the number of bits needed is $O(d \ln(M \cdot max(d, 1/\epsilon)))$. Bob then outputs $f_{x'}(y)$.

What remain to be shown is that since $E[|f_x(y) - f_{x'}(y)|] \leq \epsilon$, then $Pr[|f_x(y) - f_{x'}(y)| > \epsilon_2] < \epsilon_1$. Assume to the contrary that $Pr[|f_x(y) - f_{x'}(y)| > \epsilon_2] \geq \epsilon_1$. But then $E[|f_x(y) - f_{x'}(y)|] > \epsilon_1\epsilon_2 = \epsilon$, in contradiction to our choice of $f_{x'}$, and the properties of the cover. ■

B.1 An Application – Linear Halfspaces

We apply Theorem 10 to get an upper bound on the combinatorial dimension of a function class that is closely related to the class of *linear halfspaces*. The latter is a class of *threshold* functions, where each function is defined by a vector $\vec{x} \in [-1,1]^n$ whose L_2 norm equals 1. The value of the function on an input $\vec{y} \in \mathcal{R}^n$ is the *sign* of $(\vec{x}, \vec{y})$ – the sign of the angle between the vectors (or the *side* of the halfspace perpendicular to $\vec{x}$ on which $\vec{y}$ falls). This class has received much attention in the learning theory literature. We consider the related class of real valued functions whose value on $\vec{y}$ is the *angle* between $\vec{x}$ and $\vec{y}$, or the inner product between $\vec{x}$ and $(1/\sqrt{\|\vec{y}\|_2}) \cdot \vec{y}$. This class is simply $(INP_{2,2})_X$ For brevity we denote it by INP_X.

It is well known that the VC dimension of the class of linear halfspaces is $n+1$. In contrast, by applying Theorem 10, and following the proof of Theorem 6 to obtain the dependence on ϵ_2 (and subsequently on γ), we get the following corollary. We do not know of any direct way to obtain this upper bound.

Corollary B.2

$$C_\gamma(INP_X) = O(log(n)/\gamma^2) \ .$$

C An additional version of Theorem 6

The following is a slightly weaker version of Theorem 6 in which we explicitly describe a private-coin protocol for $INP_{2,2}$.

Theorem 12

$$R^{A\rightarrow B}(INP_{2,2}) = polylog(n).$$

Proof: We describe a 1-round randomized (private coin) communication protocol for $INP_{2,2}$. Without loss of generality, we may assume that the vectors $\vec{u}$ and $\vec{v}$ are positive. [3] Let ℓ denote the length of the representation of each coordinate in $\vec{u}$ and $\vec{v}$, where we may assume that $\ell = polylog(n)$. Alice and Bob transform their vectors into sums of binary vectors. More explicitly, let

$$\vec{u} = \sum_{j=1}^{\ell} 2^{-j+1}\vec{\mu}^j \ \ , \ \ \vec{v} = \sum_{j=1}^{\ell} 2^{-j+1}\vec{\nu}^j \qquad \forall i,j \ \vec{\mu}_i^j, \vec{\nu}_i^j \in \{0,1\} \ , \tag{3}$$

where $\vec{\mu}_i^j$ equals u_{ij}, the j'th bit in the binary representation of the i'th coordinate of $\vec{u}$, u_i, and $\vec{\nu}_i^j$ is defined similarly.

We next make the following two key observations.

1. $(\vec{u}, \vec{v}) = \sum_{j,k=1}^{l} 2^{-(j+k)+2}(\mu^j, \nu^k)$;
2. $\forall j,\ 1 \le j \le l,\ sup(\vec{\mu}^j), sup(\vec{\nu}^j) \le 2^{2j-2}$, where for a vector $\vec{r}$, $sup(\vec{r})$, (the *support* of $\vec{r}$), is the number of non-zero coordinates in $\vec{r}$.

The first observation can be verified through the following sequence of equalities.

$$(\vec{u}, \vec{v}) \ = \ \sum_{i=1}^{n} u_i \cdot v_i \tag{4}$$

[3] If this is not the case then let $\vec{u} = \vec{u}^+ + \vec{u}^-$, and $\vec{v} = \vec{v}^+ + \vec{v}^-$, where $\vec{u}^+$ and $\vec{v}^+$ are positive vectors and $\vec{u}^-$ and $\vec{v}^-$ are negative vectors. Then $(\vec{u}, \vec{v}) = (\vec{u}^+, \vec{v}^+) + (-\vec{u}^-, -\vec{v}^-) + (\vec{u}^+, -\vec{v}^-) + (-\vec{u}^-, \vec{v}^+)$.

$$= \sum_{i=1}^{n}\sum_{j=1}^{l}\sum_{k=1}^{l} 2^{-j+1}u_{ij} \cdot 2^{-k+1}u_{ik} \tag{5}$$

$$= \sum_{j,k=1}^{l} 2^{-(j+k)+2}\sum_{i=1}^{n} u_{ij} \cdot v_{ik} \tag{6}$$

$$= \sum_{j,k=1}^{l} 2^{-(j+k)+2} \cdot (\mu^j, \nu^k). \tag{7}$$

The second observation is true since following the first observation:

$$(\vec{u}, \vec{u}) = \sum_{j,k} 2^{-(j+k)+2} \cdot (\vec{\mu}^j, \vec{\nu}^k) \tag{8}$$

$$\geq \sum_{j} 2^{-2j+2} \cdot sup(\vec{\mu}^j) \ , \tag{9}$$

but $(\vec{u}, \vec{u}) \leq 1$ and hence in particular $\forall j,\ 2^{-2j+2} \cdot sup(\vec{\mu}^j) \leq 1$.

Let $a_{j,k} \stackrel{\text{def}}{=} 2^{-(j+k)+2}(\vec{\mu}^j, \vec{\nu}^j)$. Alice and Bob compute each $a_{j,k}$ separately, and then sum them all up. The number of pairs (j,k) is $polylog(n)$, and hence it remains to show how these values can be computed each with $polylog(n)$ bits of communication, $1/polylog(n)$ accuracy, and with confidence at least $1 - 1/polylog(n)$.

We separate the discussion into two cases: $j \leq k$, and $j > k$.

- $\boldsymbol{j \leq k}$: Alice repeats the following process $polylog(n)$ times. She picks a coordinate i in $sup(\vec{\mu}^j)$, uniformly, and at random, and sends i to Bob. Clearly, the corresponding coordinate of $\vec{\nu}^k$, ν_i^k, is a $\{0,1\}$ random variable whose expectation is:

$$\frac{(\vec{\mu}^j, \vec{\nu}^k)}{|sup(\vec{\mu}^j)|} \ .$$

  Since this process is repeated $polylog(n)$ times, if we apply Chernoff bounds, we get that with high probability, the average value of the ν_i^k's approximates $(\vec{\mu}^j, \vec{\nu}^k)/|sup(\vec{\mu}^j|)$ within an additive error of $1/polylog(n)$. Alice also sends Bob the size of $sup(\vec{\mu}^j)$, and Bob then multiplies the average of the ν_i^k's by $sup(\vec{\mu}^j)$ and by $2^{-(j+k)+2}$ to get an approximation of $a_{j,k}$. With high probability $(1 - 2^{-polylog(n)})$, the error of this approximation is

$$\frac{2^{-(j+k)+2} \cdot |sup(\vec{\mu}^j|}{polylog(n)} \ .$$

  Since $k \geq j$, and using the second observation above, we get that the error is bounded by $1/polylog(n)$, as required. The communication complexity is clearly $polylog(n)$.

- $\boldsymbol{j > k}$: This case can be divided into two sub-cases:

  1. $2^{j-k} \leq polylog(n)$: Alice and Bob essentially follow the same protocol as described above for the case $j \leq k$. Since

$$2^{-(j+k)+2} \cdot |sup(\vec{\mu}^j)| \leq 2^{j-k},$$

     the claim follows.

2. $2^{j-k} > polylog(n)$: in this case

$$
\begin{aligned}
a_{j,k} &= 2^{-(j+k)+2}(\vec{\mu}^k, \vec{\nu}^l) && (10)\\
&\leq 2^{-(j+k)+2}|sup(\vec{\nu}^k)| && (11)\\
&\leq 2^{k-j} \quad \leq \quad \frac{1}{polylog(n)}\ , && (12)
\end{aligned}
$$

and we can ignore all such pairs by assuming the corresponding $a_{j,k}$'s are 0.

■

D Theorem 9

Proof of Theorem 9 : It is clear that $R^{||,pub}(f) \geq R^{A\to B,pub}(f) + R^{B\to A,pub}(f)$, since both Alice and Bob can simulate Caroll's side of the protocol. We show that for every probability distribution μ on $X \times Y$, $D^{||,\mu}(f) = O\left(D^{A\to B,\mu}(f) + D^{B\to A,\mu}(f)\right)$, where $D^{||,\mu}(f)$ is the simultaneous μ-distributional complexity of f. The theorem follows since, similarly to the statement in Theorem 1, it holds that $R^{||,pub}(f) = max_\mu D^{||,\mu}(f)$. Let $P^{A\to B}$ be a deterministic protocol for f in which Alice sends Bob a single message of length $c_1 = D_\epsilon^{A\to B,\mu}(f)$, and whose error probability with respect to μ is ϵ for $\epsilon < 3^{-8}$. Let $P^{B\to A}$ be a protocol whose cost is c_2, and which is chosen analogously. In what follows we show that there exists a deterministic simultaneous protocol $P^{||}$ whose cost is $c_1 + c_2$, and whose error with respect to μ is $O(\epsilon^{1/4})$. Furthermore, in $P^{||}$, Alice simply simulates her side of the protocol in $P^{A\to B}$, and Bob simulates his side in $P^{B\to A}$.

Let M_f be the "truth table" of f, i.e., it is an $|X|$ by $|Y|$ matrix where each row is labeled by some $x \in X$, each column is labeled by some $y \in Y$, and $M_f(x,y) = f(x,y)$. Then $P^{A\to B}$ induces a partition of the rows in M_f into 2^{c_1} classes such that for every x_1 and x_2 which are in the same class X', and for every y, $P^{A\to B}(x_1,y) = P^{A\to B}(x_2,y)$. Let us denote this value by $P^{A\to B}(X',y)$. Similarly, $P^{B\to A}$ induces a partition of the columns in M_f into 2^{c_2} classes. For a class $X' \subset X$ in the partition induced by $P^{A\to B}$, and for a class $Y' \subset Y$ in the partition induced by $P^{B\to A}$, let $R_{X',Y'} = X' \times Y'$ be the corresponding rectangle in M_f. Then we ask that $P^{||}$ have a constant value, $P^{||}(X',Y')$, on the entries in $R_{X',Y'}$. Let this value be the majority value, with respect to μ, of the entries in $R_{X',Y'}$. We would like to show that the total weight (with respect to μ), of the rectangles which are almost monochromatic (with respect to μ), is high.

We may assume that for every class X', and for every y, the value of $P^{A\to B}(X',y)$, is the majority value (with respect to μ) of the entries in the sub-column of M_f corresponding to y and X'. We say that an entry (x,y) in M_f is column-bad (with respect to $P^{A\to B}$) *iff* $M_f(x,y) \neq P^{A\to B}(X',y)$, where $X' \ni x$. The bound on the error probability of $P^{A\to B}$ ensures that the total weight of the column-bad entries is at most ϵ. Based on a similar assumption on $P^{B\to A}$, we can define row-bad entries, and bound their total weight in M_f.

We need three more definitions. For a class $X' \subset X$ ($Y' \subset Y$) and $y \in Y$ ($x \in X$), we say that the corresponding sub-column (sub-row) is bad, *iff* the relative weight of column-bad (row-bad) entries in it is higher than $\epsilon^{1/4}$. For a class $X' \subset X$, and a class $Y' \subset Y$, we say that the corresponding rectangle $R_{X',Y'}$ in M_f is bad-in-columns (bad-in-rows), *iff* the relative weight of bad sub-columns (sub-rows) in R is higher than $\epsilon^{1/4}$. A rectangle R is bad, if it is either bad-in-columns or bad-in-rows. Otherwise it is good. Finally we say that an entry (x,y) is rectangle-bad, if $M_f(x,y) \neq P^{||}(X',Y')$, where $x \in X'$, $y \in Y'$, and $P^{||}(X',Y')$ is as defined previously.

CLAIM D.1 *The total weight of* bad *rectangles in M_f is at most $2\epsilon^{1/2}$.*

Proof: By definition, in a bad-in-columns (bad-in-rows) rectangle, the relative weight of the column-bad (row-bad) entries is at least $\epsilon^{1/2}$. Since the total weight of the column-bad (row-bad) entries is at most ϵ, the claim follows. ■

We now prove that in every good rectangle R, the relative weight of the rectangle-bad entries is at most $4\epsilon^{1/4}$. It follows that the total weight of rectangle-bad entries is bounded by $2\epsilon^{1/2} + 4\epsilon^{1/4} = O(\epsilon^{1/4})$.

Let R be a good rectangle. We partition the (sub-)columns in R into three sets: A_1, A_0, and A_b. A_b are the bad columns. Their relative weight is at most $\epsilon^{1/4}$. A_1 (A_0) are good columns whose majority value is 1 (0). Let the relative weight of A_1 (A_0) be α_1 (α_0). Without loss of generality, $\alpha_1 \geq \alpha_0$. The relative weight of rectangle-bad entries in R is hence at most $\epsilon^{1/4} + \alpha_1\epsilon^{1/4} + \alpha_0$. It remains to bound α_0. We use the following claim whose proof is similar to the proof of Claim D.1.

CLAIM D.2 *The relative weight of sub-rows in A_1 (A_0) in which the relative weight of entries (x, y) for which $M_f(x, y) = 0$ (1) is larger than $\epsilon^{1/8}$, is at most $\epsilon^{1/8}$.*

It follows from this claim, that in at least $1 - 2\epsilon^{1/8}$ of the sub-rows in $A_1 \cup A_0$, the relative weight of entries with value 0 is at least $\alpha_0(1 - \epsilon^{1/8})$ Since R is good, in particular it is good-in-rows. For $\epsilon < 3^{-8}$, $1 - 2\epsilon^{1/8} > \epsilon^{1/4}$, and thus necessarily $\alpha_0 < \epsilon^{1/4}/(1 - \epsilon^{1/8})$ which is less than $2\epsilon^{1/4}$. ■