On Randomized One-Round Communication Complexity

Ilan Kremer
Managerial Economics and Decision Science Dept.
Kellogg Business School
Northwestern University
Evanston IL USA
idk112@casbah.acns.nwu.edu

Noam Nisan*
Institute of Computer Science
Hebrew University
Jerusalem Israel
noam@cs.huji.ac.il

Dana Ron†
Laboratory of Computer Science
MIT
Cambridge MA USA
danar@theory.lcs.mit

Abstract

We present several results regarding randomized one-round communication complexity. Our results 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 [Yao79]: 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 introduced by Yao in his original communication complexity paper [Yao79] and were later studied by several authors (*cf.* [PM84, Abl93, NS96]; For a general survey on communication complexity see [KN95]). 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

*This work was supported by BSF grant 92-00043 and by a Wolfeson award administered by the Israeli Academy of Sciences.

†Supported by a National Science Foundation Postdoctoral Research Fellowship, Grant No. DMS-9508963

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 case where Alice and Bob have access to a *public* (*common*) source of random coin flips, and call such protocols *public-coin*. We denote the public-coin communication complexity of f by $R^{A\to B,pub}(f)$. When f is not symmetric we also consider the case where 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", Carol, who sees no part of the input but must be able to compute $f(x,y)$ just from these two messages. This complexity we denote by $R^{||}(f)$.

In this paper we consider three main questions.

1.1 The VC-Dimension and a related Combinatorial Dimension

We first observe the (surprising) fact that the VC-dimension [VC71, BEHW89] 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 $R^{A\to B}(f)$. 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 $R^{A\to B,[]}(f)$.)

Theorem 3 $R^{A\to B}(f) \geq R^{A\to B,[]}(f) = \Theta(VC\text{-}dim(f_X))$.

For more general real valued functions we get similar bounds on $R^{A\to B}(f)$ and $R^{A\to B,[]}(f)$ in terms of another combinatorial dimension which is a parameterized version of the *pseudo-dimension* (*cf.* [Pol84, Hau92, KS94]).

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 $R^{A\to B,[]}(f) = O(1)$, 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 $INP(\vec{x},\vec{y}) = \vec{x}\cdot\vec{y} = \sum_i x_i y_i$. We limit ourselves to cases where there is an a priori bound on certain norms of $\vec{x}$ and $\vec{y}$, a bound which will ensure that $|\vec{x}\cdot\vec{y}| \leq 1$. 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 no harder than the first.

We first study the case in which $\|\vec{x}\|_1 \leq 1$ and $\|\vec{y}\|_\infty \leq 1$, where $\|\vec{x}\|_1 = \sum_i |x_i|$ and $\|y\|_\infty = \max_i |y_i|$. (This of course ensures that $|\vec{x}\cdot\vec{y}| \leq 1$.) In his original paper [Yao79], Yao actually defined the one-round randomized communication complexity of functions in terms of the inner product of such pairs of vectors. If we denote the corresponding inner product function by $INP_{1,\infty}$, and the related partial boolean function by $INP^{par}_{1,\infty}$, then we obtain:

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 one-round randomized complexity is polylog(n).

Here completeness is under rectangular reductions as introduced by Babai, Frankl, and Simon [BFS86]. It should be noted that the reduction we use was implicit in [Yao79].

The second case we consider is one in which $\|\vec{x}\|_2 \leq 1$ and $\|\vec{y}\|_2 \leq 1$, where $\|\vec{x}\|_2 = \sqrt{\sum_i {x_i}^2}$. Again, this ensures that $|\vec{x} \cdot \vec{y}| \leq 1$. If we denote this problem by $INP_{2,2}$, then we show:

Theorem 8 $R^{A \rightarrow B,pub}(INP_{2,2}) = \Theta(1)$, *and* $R^{A \rightarrow B}(INP_{2,2}) = \Theta(\log(n))$.

The upper bound on $R^{A \rightarrow B,pub}(INP_{2,2})$ also holds for $R^{\|,pub}(INP_{2,2})$. This follows from Theorem 8 stated above, and Theorem 10 which is stated in the next subsection. However, when only private coins are allowed, the communication complexity of the problem changes dramatically. Recent results [NS96, BW96, BK96] (that are briefly discussed in Subsection 1.3) imply that $R^{\|}(INP_{2,2}) = \Omega(\sqrt{n})$ (and this bound is tight [Amb94, Nao94, New94, NS96]).

We also present an interesting application of the above theorems. 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 prove the following theorem.

Theorem 9 *(1) For every constant* $0 < \epsilon < 1$, *there exist mappings,* $g_{2,1} : B_2^n \rightarrow B_1^{n'}$, *and* $g_{2,\infty} : B_2^n \rightarrow B_\infty^{n'}$, *where* $n' = \text{poly}(n)$, *such that for every pair of vectors,* $\vec{u}, \vec{v} \in B_2^n$, $|\vec{u} \cdot \vec{v} - g_{2,1}(\vec{u}) \cdot g_{2,\infty}(\vec{v})| \leq \epsilon$.
(2) For any given integer n', *there do* not *exist mappings,* $g_{1,2} : B_1^n \rightarrow B_2^{n'}$, *and* $g_{\infty,2} : B_\infty^n \rightarrow B_2^{n'}$, *such that for every pair of vectors,* $\vec{u} \in B_1^n$, $\vec{v} \in B_\infty^n$, $|\vec{u} \cdot \vec{v} - g_{1,2}(\vec{u}) \cdot g_{\infty,2}(\vec{v})| \leq 1/6$.

1.3 One-Round vs. Simultaneous

It is clear that $R^{\|}(f) \geq R^{A \rightarrow B}(f) + R^{B \rightarrow A}(f)$. 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 prove a similar statement for the randomized case if Alice and Bob are allowed shared, public coins.

Theorem 10 $R^{\|,pub}(f) = \Theta(R^{A \rightarrow B,pub}(f) + R^{B \rightarrow A,pub}(f))$.

Unfortunately, when only private coins are allowed, then the situation is grimmer. Bourgain and Wigderson [BW96] and Babai and Kimmel [BK96] show that $R^{\|}(f) = \Omega\left(\sqrt{D^{\|}(f)}\right)$, where $D^{\|}(f)$ is the deterministic simultaneous complexity of f. In particular this implies a gap of $\Theta(\sqrt{n})$ between $R^{\|,pub}(f)$ and $R^{\|}(f)$ for any function f having $D^{\|}(f) = \Omega(n)$ and $R^{\|,pub}(f) = O(1)$. An example of a function which portrays such a gap is the equality function[1]. As noted previously, the same gap holds for $INP_{2,2}$.

[1] The lower bound of $\Omega(\sqrt{n})$ on the private coin simultaneous communication complexity of the equality function was also proved by Newman and Szegedy [NS96] (prior to the results of Bourgain and Wigderson [BW96] and Babai and Kimmel [BK96]) using different techniques.

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 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 Carol. As in 1-round protocols, Alice and Bob each get a part of the input to the function. Carol does not receive any input, and has access neither to Alice's nor to Bob's input. Instead of communicating among themselves, Alice and Bob each send a single message to Carol and she computes the output of the protocol.

For a function $f : X \times Y \to \{0, 1\}$, and $0 < \epsilon < 1$, we use the following notations (exact definitions are given in Appendix A). All error probabilities referred to below are of *two-sided* error.

- $R_\epsilon^{A \to B}(f)$ denotes the randomized private-coin 1-round communication complexity of f with error probability ϵ.
- $R_\epsilon^{A \to B, pub}(f)$ 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 $X \times Y$, $D_\epsilon^{A \to B, \mu}(f)$ denotes the 1-round μ-distributional complexity of f with error probability ϵ (i.e., $D_\epsilon^{A \to B, \mu}(f)$ is the minimum cost taken over all deterministic protocols P for which $\Pr_\mu[P(x, y) \neq f(x, y)] < \epsilon$).

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 $A \to B$ to $B \to A$.

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., $R_\epsilon^{A \to B}(f)$ with constant ϵ, is simply denoted by $R^{A \to B}(f)$).

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., $R^{A\to B}_{\epsilon_1,\epsilon_2}(f)$). 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 [Yao83], characterizes public-coin complexity in terms of distributional (deterministic) complexity.

Theorem 1 ([Yao83]) *For every function $f : X \times Y \to \{0,1\}$, and for every $0 < \epsilon < 1$,*

$$R^{A\to B,pub}_{\epsilon}(f) = \max_{\mu} D^{A\to B,\mu}_{\epsilon}(f),$$

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 $R^{A\to B,[]}_{\epsilon}(f)$ denote $\max_\mu D^{A\to B,\mu}_{\epsilon}(f)$, where the maximum is taken over all product distributions.*

Finally, we shall need the following (slight) variation of a theorem of Newman [New91], which gives an upper bound on private-coin complexity in terms of public-coin complexity.

Theorem 2 ([New91]) *Let $f : X \times Y \to Z$, where X and Y are arbitrary finite sets, and Z is an arbitrary (and not necessarily finite) set. For every $0 < \delta \leq 1$ and every $0 \leq \epsilon_1, \epsilon_2 < 1/2$,*

$$R^{A\to B}_{\epsilon_1+\delta,\epsilon_2}(f) \leq R^{A\to B,pub}_{\epsilon_1,\epsilon_2}(f) + O(\log\log(|X| \cdot |Y|) + \log(1/\delta)) .$$

Proof Sketch: Newman's theorem as stated in [New91] assumes the function f is boolean. However, since we are essentially only varying the definition of the correctness of a communication protocol on a given input $(x,y) \in X \times Y$ (see details below), while the domain of the function remains finite, his proof, with only slight variations, can be used to prove Theorem 2. More precisely, Newman's proof works by showing that if (a boolean function) f has a public-coin protocol P whose probability of error is ϵ_1 (and which uses any amount of randomness), then f has a public-coin protocol P' with error $\epsilon_1 + \delta$ which uses only $O(\log(n/\delta))$ random bits. The existence of a private-coin protocol with error $\epsilon_1 + \delta$ whose cost is $O(\log(n/\delta))$ larger than the cost of P, directly follows: Define a corresponding private-coin protocol in which Alice first selects all $O(\log(n/\delta))$ random bits needed for executing P', sends them to Bob, and then Alice and Bob execute P', using the "effectively public" random bits .

To prove the existence of such a protocol P', Newman views the public-coin protocol P as a distribution over deterministic protocols, where each deterministic protocol is defined by some fixed random string. He then shows that there is a small set L of such deterministic protocols so that if we define a randomized protocol P' which chooses a protocol randomly from L (using only $O(\log(n/\delta))$ random bits), and executes it, then the error of P' is at most $\epsilon_1 + \delta$, and its cost is the same as the cost of P. Thus the only modification needed is in the definition of the set L of "good" deterministic protocols. Instead of requiring that for every input $(x,y) \in X \times Y$, the fraction of deterministic protocols D in L for which $D(x,y) \neq f(x,y)$, is small, we need to require that for every input (x,y), the fraction of deterministic protocols D in L for which $|D(x,y) - f(x,y)| \leq \epsilon_2$, is small. With this and one additional very similar modification (in the definition of the sets $A(x,y)$ of deterministic protocols which err on (x,y)), the proof proceeds as in [New91]. □

3 1-Round Randomized Communication Complexity and Combinatorial Dimensions

Let $f : X \times Y \to Z$ be a function whose randomized communication complexity we would like to study. For $x \in X$ we define $f_x : Y \to Z$ 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 both an upper and a lower bound on $R^{A\to B,[]}(f)$, and hence a lower bound on $R^{A\to B}(f)$. If f is a real valued function then a related combinatorial dimension which is a parameterized version of the *pseudo-dimension* gives similar bounds.

3.1 Boolean Functions

We start by recalling the definition of the *VC-dimension* [VC71, BEHW89].

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 $f : X \times Y \to \{0,1\}$, and for every constant error $\epsilon \leq 1/8$,*

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

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

Example 1 *For $x, y \subseteq \{1,\ldots,n\}$, $DISJ(x,y)$ is defined to be 1 iff $x \bigcap y = \emptyset$. We apply Theorem 3 to show that $R^{A\to B}(DISJ) = \Omega(n)$. 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 $R(DISJ) = \Omega(n)$, [KS92, Raz90], the upper bound on multi-round rectangular distributional complexity is lower [BFS86]: $R^{[]}(DISJ) = O(\sqrt{n}\log(n))$. Thus, our example shows a quadratic gap between $R^{A\to B,[]}(f)$, and $R^{[]}(f)$ (for $f = DISJ$).*

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

• $\boldsymbol{R_\epsilon^{A\to B,[]}(f) = \Omega(d)}$**:** In order to prove this lower bound we describe a product distribution μ for which $D_\epsilon^{A\to B,\mu}(f) = \Omega(d)$, 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* $y \in R$. For each $R \subseteq S$, fix such an x_R. 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 cost is at most $d/15$. Thus, P induces two mappings. $P_1 : \{0,1\}^d \to \{0,1\}^{d/15}$, determines which $d/15$ bits Alice should send to Bob for every given x_R, and $P_2 : \{0,1\}^{d/15} \to \{0,1\}^d$ 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 induces a mapping $P_{1,2} \stackrel{\text{def}}{=} P_1 \circ P_2$ from $\{0,1\}^d$ into a set $U \subset \{0,1\}^d$, where $|U| \leq 2^{d/15}$. 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 $U \subset \{0,1\}^d$, $|U| \leq 2^{d/15}$, $E[dist(z,U)] > d/8$, 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 $|\bigcup_u N_u| \leq \sum_u |N_u| \leq 2^{d-1}$, and hence $E[dist(z,U)] > d/8$. For each u,

$$|N_u| = \sum_{i=0}^{d/4} \binom{d}{i} \leq (ed/(d/4))^{d/4} \quad (1)$$

$$= 2^{d(\log(4e)/4)} < 2^{.861d}. \quad (2)$$

Thus,

$$\sum_u |N_u| < 2^{d/15+.861d} < 2^d \cdot 2^{-d/15}. \quad (3)$$

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

• $\boldsymbol{R_\epsilon^{A\to B,[]}(f) = O(d)}$**:** For this claim we need the following theorem [BEHW89], which is one of the most fundamental theorems in computational learning theory.

Theorem [BEHW89] *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 $O(d/\epsilon \cdot \log(1/\epsilon))$.

Let $\mu : X \times Y \to \{0,1\}$ be a product distribution over $X \times Y$, where for every $x \in X$, $y \in Y$, $\mu(x,y) = \mu_X(x)\mu_Y(y)$. 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

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

It directly follows that

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

and hence

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

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

3.2 Non-Boolean Functions

In the case where the range of f (and f_X) is not $\{0,1\}$, we must consider the following generalization of the VC-dimension, which is a parameterized version of what is known as the *pseudo-dimension*. This definition follows works of Pollard [Pol84], Haussler [Hau92], Kearns and Schapire [KS94], and Alon *et. al* [ABCH93].

DEFINITION 3.2 *Let H be class of functions over a domain Y into a range $Z \subseteq \Re$, and let $\gamma \geq 0$. We say that a set $S = \{y_1, \ldots, y_k\} \subseteq Y$, is* γ-shattered *by H if there exists a vector $\vec{w} = \langle w_1, \ldots w_k \rangle \in Z^k$ 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 $h_R(y_i) > w_i + \gamma$, and if $y_i \notin R$ then $h_R(y_i) < w_i - \gamma$. The largest value d for which there exists a set S of size d that is γ-shattered by H is the* γ-pseudo-dimension *of H and is denoted by $\mathcal{P}_\gamma$-$dim(H)$. If arbitrarily large finite sets can be γ-shattered by H, then $\mathcal{P}_\gamma$-$dim(H) = \infty$.*

Similarly to the Boolean case, we have the following theorem for $[0,1]$ valued functions. As we shall see in the proof, the lower bound holds for functions with any range, and we later discuss how to generalize the upper bound.

Theorem 4 *For every function $f : X \times Y \to [0,1]$, every constant $\epsilon_1 \leq 1/8$, and every constant $\epsilon_2 \leq 1/6$,*

$$R_{\epsilon_1,\epsilon_2}^{A\to B}(f) \ \geq \ R_{\epsilon_1,\epsilon_2}^{A\to B,[]}(f) \ = \ \tilde{\Theta}(\mathcal{P}_{\Theta(1)}\text{-}dim(f_X)).$$

More precisely,

$$R_{\epsilon_1,\epsilon_2}^{A\to B,[]}(f) \ = \ \Omega(\mathcal{P}_{\epsilon_2}\text{-}dim(f_X))$$

and

$$R_{\epsilon_1,\epsilon_2}^{A\to B,[]}(f) \ = \ O\left(\mathcal{P}_{\Theta((\epsilon_1\epsilon_2)^2)}\text{-}dim(f_X) \cdot \log^2(\mathcal{P}_{\Theta((\epsilon_1\epsilon_2)^2)}\text{-}dim(f_X))\right)$$

Proof: The inequality $R_{\epsilon_1,\epsilon_2}^{A\to B}(f) \geq R_{\epsilon_1,\epsilon_2}^{A\to B,[]}(f)$ follows the same argument used for proving that $R_{\epsilon}^{A\to B}(f) \geq R_{\epsilon}^{A\to B,[]}(f)$ in Theorem 3. Here we apply a straightforward generalization of Theorem 1, namely that $R_{\epsilon_1,\epsilon_2}^{A\to B,pub}(f) = max_{\mu} D_{\epsilon_1,\epsilon_2}^{A\to B,\mu}(f)$. Though Yao's theorem was stated only for the boolean case, its proof remains correct in the more general real valued case. The reason is that in the proof bounds are given on the probability that protocols succeeds, and whether success

is defined as obtaining the exact value of the function or a good approximation of it is irrelevant to the proof.

In order to prove the lower bound $R^{A\to B,[]}_{\epsilon_1,\epsilon_2}(f) = \Omega(\mathcal{P}_{\Theta(1)}\text{-}dim(f_X))$, we essentially reduce the problem to the $\{0,1\}$-valued case. In particular, we show that there exists a product distribution μ and a $\{0,1\}$-valued function f' such that $D^{A\to B,\mu}_{\epsilon_1,\epsilon_2}(f) \geq D^{A\to B,\mu}_{\epsilon_1}(f')$, and $D^{A\to B,\mu}_{\epsilon_1}(f') = \Omega\left(\mathcal{P}_{\epsilon_2}\text{-}dim(f_X)\right)$. By definition of $\mathcal{P}_{\epsilon_2}\text{-}dim(f_X)$, there exists a set $S \subseteq Y$ of size $d = \mathcal{P}_{\epsilon_2}\text{-}dim(f_X)$ which is ϵ_2-shattered by f_X. Namely, there exists a vector $\vec{w} = \langle w_1, \ldots, w_d \rangle \in Z^d$ such that for every subset $R \subseteq S$ there exists $x_R \in X$, such that $\forall y_i \in S$, if $y_i \in R$ then $f_{x_R}(y_i) > w_i + \epsilon_2$, and if $y_i \notin R$, then $f_{x_R}(y_i) < w_i - \epsilon_2$. Let f' be the partial boolean function defined as follows on the pairs $\{(x_R, y_i) \,|R \subseteq S,\ y_i \in S\}$: $f'(x_R, y_i) = 1$ if $f_{x_R}(y_i) > w_i + \epsilon_2$, and $f'(x_R, y_i) = 0$ if $f_{x_R}(y_i) < w_i - \epsilon_2$. Let μ be the uniform distribution on all the such pairs (x_R, y_i).

Clearly, $D^{A\to B,\mu}_{\epsilon_1,\epsilon_2}(f) \geq D^{A\to B,\mu}_{\epsilon_1}(f')$, since in order to compute f' on a given pair (x_R, y_i), Alice and Bob can run the protocol for computing f and Bob can output 1 if the answer is greater than w_i, and 0 otherwise. Since for each x_R and y_i, $f(x_R, y_i)$ is either greater than $w_i + \epsilon_2$, or smaller than $w_i - \epsilon_2$, and with probability $1 - \epsilon_1$ the protocol for computing f has error at most ϵ_2, the resulting protocol for f' errs with probability at most ϵ_1, as required. As for the lower bound on $D^{A\to B,\mu}_{\epsilon_1}(f')$, since S is shattered by $\{f'_{x_R}\}$, we can directly apply the proof of the lower bound in Theorem 3 to get that $D^{A\to B,\mu}_{\epsilon_1}(f') = \Omega(d)$ (where $d = |S| = \mathcal{P}_{\epsilon_2}\text{-}dim(f_X)$).

It remains to prove the upper bound on $R^{A\to B,[]}_{\epsilon_1,\epsilon_2}(f)$. We shall need the following theorem of Bartlett, Long and Williamson [BLW94].

Theorem [BLW94] *Let H be a class of functions over a domain Y into the range $[0,1]$, let π be an arbitrary probability distribution over Y, and let $0 < \epsilon < 1/2$, $0 < \delta < 1$. Let $d = \mathcal{P}_{\frac{\epsilon^2}{576}}\text{-}dim(H)$. Then there exists a (deterministic) learning algorithm L which has the following property. Given as input a set $S \in Y^m$ of m examples chosen according to π^m and labeled according to an unknown function $h \in H$, L outputs a hypothesis $h' \in H$ such that if $m \geq m_0(d, \epsilon, \delta)$ where*

$$m_0(d, \epsilon, \delta) = c_0 \left(\frac{1}{\epsilon^4} \log \frac{1}{\delta} + \frac{d}{\epsilon^4} \log^2 \frac{d}{\epsilon} \right)$$

for some constant $c_0 > 0$, then with probability at least $1 - \delta$ over the random samples,

$$\int_{y \in Y} |h'(y) - h(y)| \, d\pi(y) \leq \epsilon \ .$$

Given the above theorem, the proof of the upper bound on $R^{A\to B,[]}_{\epsilon_1,\epsilon_2}(f)$ is very similar to the proof of the upper bound on $R^{A\to B,[]}_{\epsilon}(f)$ in Theorem 3. Let $m_0(\cdot,\cdot,\cdot)$ be as defined in the theorem. Here we show that for every rectangular distribution μ, there exists a deterministic protocol whose $(\mu, \epsilon_1, \epsilon_2)$-distributional complexity is bounded by $\log(2/\epsilon_2)$ times $m = m_0(d, \epsilon, \delta)$, where $\epsilon = (\epsilon_1 \cdot \epsilon_2)/4$, $\delta = \epsilon_1/2$, and $d = \mathcal{P}_{\frac{\epsilon^2}{576}}\text{-}dim(f_X)$. Note that this complexity is $O(d \log^2(d))$ for constant ϵ_1, ϵ_2.

We first consider a bounded precision variant of f which we denote by $\bar{f}$. Namely, for each $(x, y) \in X \times Y$, $\bar{f}(x, y)$ is the value of $f(x, y)$ truncated after the $\log(2/\epsilon_2)$ bit of precision. Clearly, for every (x, y), $|\bar{f}(x, y) - f(x, y)| \leq \epsilon_2/2$. Moreover, since $\bar{f}_X \subseteq f_X$, we have that for every γ, $\mathcal{P}_\gamma\text{-}dim(\bar{f}_X) \leq \mathcal{P}_\gamma\text{-}dim(f_X)$. Let $\mu : X \times Y \to [0,1]$ be a product distribution over $X \times Y$, where for every $x \in X$, $y \in Y$, $\mu(x, y) = \mu_X(x)\mu_Y(y)$. Consider the following family of deterministic protocols. For every set $S = (y_1, \ldots, y_m)$, of m points where m was defined above and $y_i \in Y$,

let P_S be the following protocol. For a given input (x, y), Alice sends Bob the value of $\bar{f}_x(\cdot)$ on each point in S (using a constant number ($\log(2/\epsilon_2)$) of bits). Bob runs the learning algorithm L (referred to in the theorem of [BLW94]) for learning the class $\bar{f}_X$, giving it as input the set S together with the labels sent by Alice. Bob then outputs $\bar{f}_{x'}(y)$, where $\bar{f}_{x'} \in \bar{f}_X$ is the hypothesis output by L.

We define the following probability distribution, Π, on this family of deterministic protocols: $\Pi(P_S) = \mu^m(S)$. Then from the theorem of [BLW94], our choice of m, and the definition of $\bar{f}$, we have that

$$\forall x \in X,\ \Pr_{\Pi}\left[\Pr_{\mu_Y}[|P_S(x,y) - \bar{f}(x,y)| > \epsilon_2/2] > \epsilon_1/2\right] < \epsilon_1/2.$$

It directly follows that

$$\Pr_{\Pi}\left[\Pr_{\mu}[|P_S(x,y) - \bar{f}(x,y)| > \epsilon_2/2] > \epsilon_1/2\right] < \epsilon_1/2 \ ,$$

and hence

$$E_{\Pi}\left[\Pr_{\mu}[|P_S(x,y) - f(x,y)| > \epsilon_2]\right] < \epsilon_1.$$

Therefore, there exists at least one deterministic protocol that with probability at least $1 - \epsilon_1$ over the choice of (x, y), outputs a value which differs by at most ϵ_2 from $f(x, y)$. ■

If the range of f is not $[0, 1]$ but rather in $[-B, B]$ for some integer B, Alice and Bob can execute the protocol for the $[0, 1]$-valued function $f^{[0,1]}(\cdot,\cdot) \stackrel{\text{def}}{=} f(\cdot,\cdot)/(2B) + 1/2$ with $\epsilon_2^{[0,1]}$ set to $\epsilon_2/(2B)$, and translate the value received back to the original range. In such a case, the upper bound we get on the running time is $O(d^{[0,1]} \log^2(d^{[0,1]}))$, where $d^{[0,1]} = \mathcal{P}_{\frac{(\epsilon_1\epsilon_2)^2}{2^{16}B^2}}\text{-}dim(f_X^{[0,1]})$. By definition of the pseudo-dimension, we have that for every γ, $\mathcal{P}_{\frac{\gamma}{(2B)}}\text{-}dim(f_X^{[0,1]}) \leq \mathcal{P}_{\gamma}\text{-}dim(f_X)$, and hence for constant B, the bound has the same form as the bound for $[0, 1]$ valued function.

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 5 *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 $IND : \{1, \ldots, n\} \times \{0, 1\}^n \to \{0, 1\}$, 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 $\{1, \ldots, n\}$ is shattered by IND_Y. For every subset R of $\{1, \ldots, n\}$, 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 $\{1, \ldots, n\}$ 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 $1 \leq j \leq \log(n)$, and for every $1 \leq i \leq n$, $y_i^j = 1$ *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. $R^{A\to B,[]}(f)$

We next show the existence of a function f for which there is a large gap between $R^{A\to B}(f)$ and $R^{A\to B,[]}(f)$. It is interesting to note that in the case of multi-round communication complexity there are only examples showing a polynomial (quadratic) gap between $R^{[]}(f)$ and $R(f)$ [BFS86, KS92, Raz90].

Theorem 6 *There exists a function f for which $R^{A\to B,[]}(f) = O(1)$, while $R^{A\to B}(f) = \Omega(n)$.*

Proof: Let $GT : \{1,\ldots,2^n\} \times \{1,\ldots,2^n\} \to \{0,1\}$, 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 $x \leq y_1$. Applying Theorem 3, we get the upper bound on $R^{A\to B,[]}(GT)$.

The proof of the lower bound on $R^{A\to B}(GT)$ is given in [MNSW95]. We describe a simpler proof to a slightly weaker bound of $\Omega(n/\log(n))$. Let IND' be the same as the index function IND except that $IND' : \{0,1\}^n \times \{1,\ldots,n\} \to \{0,1\}$. That is, Alice gets a vector $\vec{x} \in \{0,1\}^n$, Bob gets an index $i \in \{1,\ldots,n\}$, and they want to compute x_i. From Theorem 5 we have the $R^{A\to B}(IND') = \Omega(n)$. Let P be a randomized private-coin protocol for GT whose cost is $c = R^{A\to B}(GT)$. We now show how $O(\log(n))$ executions of P (at a cost of $O(c\log(n))$ can be used to compute IND'. In fact, we show that $O(\log(n))$ executions of P can be used to exactly find *every* coordinate x_i of $\vec{x}$. The lower bound of $\Omega(n/\log(n))$ on $R^{A\to B}(GT)$ directly follows.

Assume that for a fixed $\vec{x}, \vec{y} \in \{0,1\}^n$, Alice and Bob execute P, $k\log(n)$ times for some constant k, and Bob outputs the majority outcome. Since the probability that the outcome of a single execution is incorrect, is at most $1/3$, the probability that the majority outcome, among $k\log(n)$ executions, is incorrect, is $\exp(-\Omega(\log(n)) = n^{-\Omega(1)}$. Furthermore, with high probability, for a fixed $\vec{x}$ and for any *set* $\{\vec{y}^1,\ldots,\vec{y}^n\}$, since the set has only n elements, Bob can compute the correct value of $GT(\vec{x},\vec{y}^i)$ for *every* $\vec{y}^i$ in the set with probability at least $1 - n^{-\Omega(1)}$. It follows that Bob can use this protocol (while ignoring his input $i \in \{1,\ldots,n\}$) to perform a binary search and find the *exact* value of $\vec{x}$ with high probability. In other words, he can compute each coordinate in the binary representation of $\vec{x}$, and in particular, x_i. ■

3.3.3 Linear Halfspaces

We apply Theorem 4 to get an upper bound on the (parameterized) pseudo-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 $\vec{y} \in \Re^n$ is the *sign* of $\vec{x}\cdot\vec{y}$ – the sign of the cosine 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 $(1/\|\vec{y}\|_2)\cdot\vec{y}$. 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, from Theorem 4 we have that for constant ϵ_1, $R^{A\to B}_{\epsilon_1,\epsilon_2}(f) = \Omega(\mathcal{P}_{\epsilon_2}\text{-}dim(f_X))$, and from the proof of Theorem 8 (appearing in Section 4.2) we have that (for constant ϵ_1) $R^{A\to B}_{\epsilon_1,\epsilon_2}(INP_{2,2}) = O(\log(n) + 1/\epsilon_2^2)$. By Combining the two bounds we get the following corollary. We do not know of any direct way to obtain this upper bound.

Corollary 3.2 $\mathcal{P}_\gamma\text{-}dim(INP_X) = O(\log(n) + 1/\gamma^2)$.

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 A Complete Problem for 1-Round Communication Complexity

We start by presenting a fairly simple inner product 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 [BFS86].

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

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. $INP_{1,\infty}(\vec{p}, \vec{q}) = \vec{p} \cdot \vec{q}$, where $\vec{p} = \langle p_1, \ldots, p_n \rangle$ and $\vec{q} = \langle q_1, \ldots, q_n \rangle$ are n-dimensional vectors which have the following properties:

$\|\vec{p}\|_1 \leq 1$ and $\|\vec{q}\|_\infty \leq 1$, where $\|\vec{p}\|_1 \stackrel{\text{def}}{=} \sum_i |p_i|$, and $\|\vec{q}\|_\infty \stackrel{\text{def}}{=} \max_i |q_i|$. Let the *partial* boolean function $INP^{par}_{1,\infty}(\cdot,\cdot)$ be defined as follows:

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

The domain of $INP^{par}_{1,\infty}(\cdot,\cdot)$ is hence all $(\vec{p},\vec{q})$ such that either $INP_{1,\infty}(\vec{p},\vec{q}) \geq 2/3$, or $INP_{1,\infty}(\vec{p},\vec{q}) \leq 1/3$, and $INP^{par}_{1,\infty}(\cdot,\cdot)$ is undefined on inputs $(\vec{p},\vec{q})$, such that $1/3 < INP_{1,\infty}(\vec{p},\vec{q}) < 2/3$. A protocol for computing $INP^{par}_{1,\infty}(\cdot,\cdot)$, must be correct (with high probability) only when given inputs which belong to the domain of the function, and may output either 0 or 1 (or any other value) when given inputs not in the domain of the function. Such a protocol, however, does not need to distinguish between inputs in the domain, and inputs outside the domain.

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 $\vec{p}$ is a probability vector, i.e., $\vec{p}$ is non-negative, and $\|\vec{p}\|_1 = 1$.[2] 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 $[-1,1]$ 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 $\vec{p}\cdot\vec{q}$ 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 $INP^{par}_{1,\infty}$. As noted in the introduction, this reduction was implicit in [Yao79]. If $R^{A\to B}(f) =$ polylog(n) then there exists a 1-round communication protocol P for f with error $1/4$ 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 $1 \leq i \leq 2^c$, 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 $2^{\text{polylog}(n)}$, and we let each coordinate be written with exponential precision, using $\Theta(n)$ bits. It remains to be shown that $f(x,y) = INP^{par}_{1,\infty}(\vec{p}(x),\vec{q}(y))$. By the definition of $\vec{p}(x)$ and $\vec{q}(y)$,

$$\Pr[P(x,y)=1] \quad = \quad \sum_{i=1}^{2^c} p_i(x)\cdot q_i(y) \pm o(2^{-n}). \tag{4}$$

[2] The first assumption can be removed as follows. Let $\vec{p} = \vec{p}^{+} + \vec{p}^{-}$ where $\vec{p}^{+}$ is a non-negative vector, and $\vec{p}^{-}$ is a negative vector. $\vec{p}^{+}\cdot\vec{q}$ and $(-\vec{p}^{-})\cdot\vec{q}$ can then be computed separately, and summed together. The second assumption can be removed simply by normalizing.

Since $\Pr[P(x,y)=1]$ should be greater than $3/4$ if $f(x,y)=1$ and smaller than $1/4$ 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 5. ■

4.2 The bounded L_2 norm case

In this subsection we study the problem of computing the inner product of two n-dimensional vectors whose L_2 norm is bounded by 1. More precisely, we define:

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

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

Theorem 8 *$R^{A\to B,pub}(INP_{2,2}) = \Theta(1)$ and $R^{A\to B}(INP_{2,2}) = \Theta(\log(n))$.*

Proof: Without loss of generality, we assume that the L_2 norm of both $\vec{u}$ and $\vec{v}$ is exactly 1. In the public-coin protocol we are about to describe, we apply a technique presented in [GW94]. In particular, we need the following lemma:

Lemma [GW94] *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:*

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

where $\mathrm{sgn}(x)=1$, 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 $\mathrm{sgn}(\vec{u}\cdot\vec{r}_1),\ldots,\mathrm{sgn}(\vec{u}\cdot\vec{r}_k)$, and Bob estimates $\arccos(\vec{u}\cdot\vec{v})$ by $\frac{\pi}{k}|\{\vec{r}_i \mid \mathrm{sgn}(\vec{u}\cdot\vec{r}_i) \neq \mathrm{sgn}(\vec{v}\cdot\vec{r}_i)\}|$. 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 $\vec{u}\cdot\vec{v}$ within an additive error $\epsilon_2 = O(1/\sqrt{k})$. Stated slightly differently, for every given ϵ_1 and ϵ_2, if $k=\Theta(\log(1/\epsilon_1)/\epsilon_2^2)$, then with probability at least $1-\epsilon_1$, Bob's estimate is within ϵ_2 from the correct value.

In order to get the upper bound on $R^{A\to B}(INP_{2,2})$, we apply Newman's theorem [New91], 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 $\Theta(n)$ bits. Then we incur only an exponentially small additional error, while the size of our effective domain is $2^{\mathrm{poly}(n)}$. Thus, $R^{A\to B}_{\epsilon_1,\epsilon_2}(INP_{2,2}) \leq O(\log(1/\epsilon_1)/\epsilon_2^2)) + O(\log(\mathrm{poly}(n)/\epsilon_1))$, which for constant ϵ_1 and ϵ_2 is $O(\log(n))$. In Appendix B we give an alternative proof to a slightly weaker claim ($R^{A\to B}(INP_{2,2}) = \mathrm{polylog}(n)$), 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 $S, T \in \{1, \ldots, n\}$, $|S| = |T| = n/2$, $INT(S,T)$ 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, $\vec{u}(S) \cdot \vec{v}(T) = |S \bigcap T|/\sqrt{|S||T|} = 2|S \bigcap T|/n$. Since $\|\vec{u}(S)\|_2 = \|\vec{v}(T)\|_2 = 1$, we can run P to compute $\vec{u}(S) \cdot \vec{v}(T)$, from which we can derive the size of $S \bigcap T$ within an additive error of $n/10$.

Therefore, under our counter assumption on $INP_{2,2}$, 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 for every f, $D(f) = 2^{O(R(f))}$, where $D(f)$ ($R(f)$) denotes the deterministic (randomized) *multi-round* communication complexity of f. We next show that the known linear lower bound $D(DISJ) = \Omega(n)$ holds for our restricted version of $DISJ$ as well. It then follows that $R^{A \to B}(DISJ) \geq R(DISJ) = \Omega(\log(n))$, contradicting our assumption on the 1-round randomized communication complexity of $INP_{2,2}$. For this we need the following definition of a *fooling set* [Yao79, LS81].

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 $(x_1,y_1) \neq (x_2,y_2)$ (both in $\mathcal{S}$), either $f(x_1,y_2) \neq z$ or $f(x_2,y_1) \neq z$. It is not hard to verify that if f has a fooling set of size t then $D(f) \geq \log t$. We would thus like to show that there exists an exponential size fooling set $\mathcal{S}$ for $DISJ$ (when the domain of the function is restricted as defined above). In particular let $\mathcal{S}$ consist 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 $|S \bigcap \bar{T}| = |\bar{S} \bigcap T| \geq n/5$. Clearly such a set is a fooling set since $DISJ(S, \bar{S}) = 1$ while $DISJ(S, \bar{T}) = DISJ(\bar{S}, T) = 0$. In order to show that such a (large) set exists we simply upper bound the probability that a *random* choice of pairs $(S, \bar{S})$ does not have the desired property. Let S and T be two sets of size $n/2$ chosen independently and at random. Then the expected size of $S \bigcap T$ (or of $S \bigcap \bar{T}$) is $n/4$. By applying a simple Chernoff bound we have that the probability that their intersection is smaller than $n/5$ is less than $e^{-n/200}$. If we now choose $2^{n/400}$ such sets, independently and at random, then the probability that some pair of sets have an intersection of size smaller than $n/5$ is less than 1. Hence there exists a choice of $2^{O(n)}$ sets which defines a fooling set, as required. ■

As a corollary of Theorem 8 and Theorem 10 (which is proved in Section 5) we have that $R^{||,pub}(INP_{2,2}) = \Theta(1)$. In contrast, in the private-coin simultaneous model we have the following proposition.

Proposition 4.1 $R^{||}(INP_{2,2}) = \Theta(\sqrt{n})$.

Proof: The upper bound follows from the application of a general upper bound [Amb94, Nao94, New94] (see [NS96]) which asserts that $R^{||}(f) = O(\sqrt{n} \cdot R^{||,pub}(f))$, where we get from Theorem 8 and Theorem 10 that $R^{||,pub}(INP_{2,2}) = O(1)$. (Though the upper bound in [NS96] was stated only for boolean functions, it also holds in our case, since we are only varying the definition of correctness (or success) of a protocol on inputs, and this is irrelevant to the proof.)

In order to obtain the lower bound we only need to show that Alice, Bob and Carol can use a private coin simultaneous protocol that computes $INP_{2,2}$ for computing the equality function EQ. The lower bound will then follow from the $\Omega(\sqrt{n})$ lower bound on $R^{||}(EQ)$ [NS96, BW96, BK96]. This can be done simply as follows. Let $g : \{0,1\}^n \to \{-1,+1\}^m$, $m = O(n)$, be an error correcting code with linear distance. Namely, there exists a constant α such that for every $x, y \in \{0,1\}^n$, if $x \neq y$, then $g(x)$ and $g(y)$ differ on at least αm bits. It is well known that such codes exist

and in particular that a random linear code will have this property with high probability. Let $g'(x) = \frac{1}{\sqrt{m}} g(x)$. Then we have that for every x, $INP_{2,2}(g'(x), g'(x)) = 1$, while for every $x \neq y$, $INP_{2,2}(g'(x), g'(y)) \leq 1 - 2\alpha$. This gap ensures that if Alice, Bob and Carol execute a protocol for $INP_{2,2}$ with $g'(x)$ and $g'(y)$ as inputs, and with any constant approximation parameter smaller than α, then they can use the resulting value to decide $EQ(x, y)$. ■

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 valued vectors whose L_1, L_2, and L_∞ norm, respectively, is bounded by 1. Then we have the following theorem.

Theorem 9 *(1) For every constant $0 < \epsilon < 1$, there exist mappings, $g_{2,1} : B_2^n \rightarrow B_1^{n'}$, and $g_{2,\infty} : B_2^n \rightarrow B_\infty^{n'}$, where $n' = \text{poly}(n)$, such that for every pair of vectors, $\vec{u}, \vec{v} \in B_2^n$, $|\vec{u} \cdot \vec{v} - g_{2,1}(\vec{u}) \cdot g_{2,\infty}(\vec{v})| \leq \epsilon$.*

(2) For any given integer n', there do not *exist mappings, $g_{1,2} : B_1^n \rightarrow B_2^{n'}$, and $g_{\infty,2} : B_\infty^n \rightarrow B_2^{n'}$, such that for every pair of vectors, $\vec{u} \in B_1^n$, $v \in B_\infty^n$, $|\vec{u} \cdot \vec{v} - g_{1,2}(\vec{u}) \cdot g_{\infty,2}(\vec{v})| \leq 1/6$.*

Proof: The construction of the mappings for the first claim is very similar to the one described in Theorem 7. Let $0 < \epsilon_1, \epsilon_2 < 1$ be such that $\epsilon_1 + \epsilon_2 < \epsilon$, and let P be the private-coin protocol for computing $INP_{2,2}$ within ϵ_2 and with probability $1 - \epsilon_1$, whose cost is $O(\log(n))$ and which is guaranteed in Theorem 8. 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[3] 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 $g_{2,1}(\vec{u}) \cdot g_{2,\infty}(\vec{v})$ is the expected value of Bob's output. But we know that 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: (1) The upper bound $R^{A \rightarrow B,pub}(INP_{2,2}) = R^{B \rightarrow A,pub}(INP_{2,2}) = O(1)$ stated in Theorem 8; (2) The lower bound $R^{B \rightarrow A}(INP_{1,\infty}^{par}) = \Omega(n)$ stated in Theorem 7 combined with Newman's theorem [New91], (stated in Theorem 2) which imply that

$$R^{B \rightarrow A,pub}(INP_{1,\infty}^{par}) \geq R^{B \rightarrow A}(INP_{1,\infty}^{par}) - O(\log(n)) \ = \Omega(n)$$

Given the constant upper bound on $INP_{2,2}$, such a pair of mappings would clearly contradict the linear lower bound on $INP_{1,\infty}^{par}$. ■

5 Simultaneous Protocols

In this section we show that the public-coin simultaneous communication complexity of any function f is bounded by the sum of the two corresponding one-round public-coin protocols. As mentioned

[3] 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.

previously, there is a general transformation from simultaneous public-coin protocols to simultaneous private-coin protocols at a multiplicative cost of $O(\sqrt{n})$ [Amb94, Nao94, New94, NS96]. While it may be the case that for certain functions there is a better upper bound on $R^{||}(f)$, the results of Bourgain and Wigderson [BW96] and Babai and Kimmel [BK96] imply that this bound is tight for every function f such that $D^{||}(f) = \Omega(n)$ and $R^{||,pub}(f) = O(1)$ (as is the case for the equality function). The existence of such a multiplicative $\Omega(\sqrt{n})$ gap between public-coin and private-coin communication complexity in the simultaneous model should be contrasted with the worst case $O(\log(n))$ *additive* difference between public-coin and private-coin one-round communication complexity [New91].

Theorem 10 *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: 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 Carol'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 for $\epsilon < 2^{-8}$. 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 larger 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 5.1 *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. ■ (Claim 5.1)

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 5.1.

CLAIM 5.2 *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 $1 - \epsilon^{1/8}$ 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 $\epsilon < 2^{-8}$, $1 - \epsilon^{1/8} > \epsilon^{1/4}$, and thus necessarily $\alpha_0 < \epsilon^{1/4}/(1 - \epsilon^{1/8})$ which is less than $2\epsilon^{1/4}$. ■ (Theorem 10)

Acknowledgments

We would like to thank Oded Goldreich, Nati Linial, Amnon Ta'shama and Avi Wigderson for many helpful discussions. We would also like to thank an anonymous referee for very detailed and helpful comments.

References

[ABCH93] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. Scale-sensetive dimensions, uniform convergence, and learnability. In *Proceedings of the Thirty-Fourth Annual Symposium on Foundations of Computer Science*, pages 292–301, 1993.

[Abl93] F. Ablayev. Lower bounds for one-way probabilistic communication complexity. In *Automata, Languages and Programming: 20th International Colloquium*, pages 241–252, 1993.

[Amb94] A. Ambainis. Manuscript. 1994.

[BEHW89] 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.

[BFS86] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the Twenty-Seventh Annual Symposium on Foundations of Computer Science*, pages 337–347, 1986.

[BK96] L. Babai and P. Kimmel. Randomized simultaneous messages. Manuscript, 1996.

[BLW94] P. L. Bartlett, P. M. Long, and R. C. Williamson. Fat-shattering and the learnability of real-valued functions. In *Proceedings of the Seventh Annual ACM Conference on Computational Learning Theory*, pages 299–310, 1994.

[BW96] J. Bourgain and A. Wigderson. Private Communications, 1996.

[GW94] 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.

[Hau92] D. Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100(1):78–150, 1992.

[KN95] E. Kushilevitz and N. Nisan. Communication Complexity. Preprint, 1995.

[KS92] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Computing*, 13(4):547–557, 1992.

[KS94] M. J. Kearns and R. E. Schapire. Efficient distribution-free learning of probabilistic concepts. *Journal of Computer and System Sciences*, 48(3):464–497, 1994.

[LS81] R. J. Lipton and R. Sedgewick. Lower bounds for VLSI. In *Proceedings of the Thirteenth Annual ACM Symposium on Theory of Computing*, pages 300–307, 1981.

[MNSW95] P. B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on the Theory of Computing*, pages 103–111, 1995.

[Nao94] M. Naor. Private communication. 1994.

[New91] I. Newman. Private vs. common random bits in communication complexity. *Information processing letters*, 39:67–71, 1991.

[New94] I. Newman. Private communication. 1994.

[NS96] I. Newman and M. Szegedy. Public vs. private coin flips in one round communication games. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, pages 561–570, 1996.

[PM84] C. Papadimitriou and M.Sipser. Communication complexity. *Journal of Computer and System Sciences*, 28(1):106–123, 1984.

[Pol84] D. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, 1984.

[Raz90] A. Razborov. On the distributional complexity of disjointness. In *Proceedings of the ICALP*, pages 249–253, 1990. (To appear in *Theoretical Computer Science*).

[VC71] 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*, 17(2):264–280, 1971.

[Yao79] 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.

[Yao83] A. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the Twenty-Fourth 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 $0 \leq \epsilon_1, \epsilon_2 \leq 1$. The* 1-round $(\boldsymbol{\mu, \epsilon_1, \epsilon_2})$-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 Carol, $r_C \in \{0,1\}^{\rho_C}$, 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 both constants, $\epsilon_1 < 1/3$, and $\epsilon_2 < 1/6$, 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 An additional version of Theorem 6

The following is a slightly weaker version of Theorem 8 in which we explicitly describe a private-coin protocol for $INP_{2,2}$.

Theorem 11

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

Proof: We describe a 1-round randomized (private-coin) communication protocol for $INP_{2,2}$. Without loss of generality we assume that the vectors $\vec{u} = \langle u_1, \ldots, u_n\rangle$ and $\vec{v} = \langle v_1, \ldots, v_n\rangle$ are both positive (see Footnote 2). Let ℓ denote the length of the representation of each coordinate in $\vec{u}$ and $\vec{v}$, where we may assume that $\ell = \text{polylog}(n)$ (since we are allowed constant error). Alice and Bob transform their vectors into sums of binary vectors. More explicitly, for each $j \in \{1, \ldots, \ell\}$, let $\vec{u}^j = \langle u_1^j, \ldots, u_n^j\rangle$, and let $\vec{v}^j = \langle v_1^j, \ldots, v_n^j\rangle$ be defined as follows: for every $i \in \{1, \ldots, n\}$, u_i^j is the j^{th} bit in the binary representation of u_i, and v_i^j is the j^{th} bit in the binary representation of v_i. Then $\vec{u}$ and $\vec{v}$ can be represented as follows:

$$\vec{u} = \sum_{j=1}^{\ell} 2^{-j+1}\vec{u}^j\ ,\quad \vec{v} = \sum_{j=1}^{\ell} 2^{-j+1}\vec{v}^j\ , \tag{5}$$

We next make the following two key observations.

1. $\vec{u}\cdot\vec{v} = \sum_{j,k=1}^{\ell} 2^{-(j+k)+2}(\vec{u}^j\cdot\vec{v}^k)$;
2. $\forall j \in \{1, \ldots, \ell\}$, $|\sup(\vec{u}^j)|, |\sup(\vec{v}^j)| \le 2^{2j-2}$, where for a vector $\vec{r}$, $\sup(\vec{r}) \subseteq \{1, \ldots, n\}$, (the *support* of $\vec{r}$), is the set of non-zero coordinates in $\vec{r}$.

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

$$\begin{aligned}
\vec{u}\cdot\vec{v} &= \sum_{i=1}^{n} u_i\cdot v_i \\
&= \sum_{i=1}^{n}\sum_{j=1}^{\ell}\sum_{k=1}^{\ell} 2^{-j+1}u_i^j\cdot 2^{-k+1}v_i^k \\
&= \sum_{j,k=1}^{\ell} 2^{-(j+k)+2}\sum_{i=1}^{n} u_i^j\cdot v_i^k \\
&= \sum_{j,k=1}^{\ell} 2^{-(j+k)+2}(\vec{u}^j\cdot\vec{v}^k).
\end{aligned}$$

The second observation is true since following the first observation:

$$\begin{aligned}
\vec{u}\cdot\vec{u} &= \sum_{j,k} 2^{-(j+k)+2}(\vec{u}^j\cdot\vec{u}^k) \\
&\ge \sum_{j} 2^{-2j+2}\cdot|\sup(\vec{u}^j)|\ ,
\end{aligned}$$

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

Let $a_{j,k} \stackrel{\text{def}}{=} 2^{-(j+k)+2}(\vec{u}^j \cdot \vec{v}^j)$. 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, $1/\text{polylog}(n)$ accuracy, and with confidence at least $1 - 1/\text{polylog}(n)$.

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 $\sup(\vec{u}^j)$, uniformly, and at random, and sends i to Bob. Clearly, the corresponding coordinate of $\vec{v}^k$, v_i^k, is a $\{0,1\}$ random variable which is 1 with probability

$$\frac{\vec{u}^j \cdot \vec{v}^k}{|\sup(\vec{u}^j)|} \quad .$$

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

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

  Since $k \geq j$, and using the second observation above, we get that the error is bounded by $1/\text{polylog}(n)$, 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

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

     the claim follows.

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

$$\begin{aligned} a_{j,k} &= 2^{-(j+k)+2}(\vec{u}^j \cdot \vec{v}^k) \\ &\leq 2^{-(j+k)+2}|\sup(\vec{v}^k)| \\ &\leq 2^{k-j} \quad \leq \quad \frac{1}{\text{polylog}(n)} \quad , \end{aligned}$$

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

■