

Hellinger Strikes Back: A Note on the Multi-Party Information Complexity of AND

T.S. Jayram

IBM Almaden Research Center

June 30, 2009

Abstract

The AND problem on t bits is a promise decision problem where either at most one bit of the input is set to 1 (NO instance) or all t bits are set to 1 (YES instance). In this note, I will give a new proof of an $\Omega(1/t)$ lower bound on the information complexity of AND in the number-in-hand model of communication. This was recently established by Gronemeier, STACS 2009. The proof exploits the information geometry of communication protocols via Hellinger distance in a novel manner and avoids the analytic approach inherent in previous work. As previously known, this bound implies an $\Omega(n/t)$ lower bound on the communication complexity of multiparty disjointness and consequently a $\Omega(n^{1-2/k})$ space lower bound on estimating the k -th frequency moment F_k .

1 Introduction

Welcome to the magical world of *Hellinger distance*!¹ In this note, I will describe a short proof of an $\Omega(1/t)$ lower bound for the information complexity of the AND function in the number-in-hand model of communication. I should mention at the forefront that the result is not new (perhaps for the constants involved) as was shown by Gronemeier [Gro09] recently. My focus, however, is show the power and beauty of Hellinger distance when applied to communication protocols, and in particular, the light that it sheds on the *information geometry* inherent in the structure of communication protocols. To describe this problem and its motivation, I must first take you on a detour into the world of communication complexity and data streams.

1.1 Data stream space complexity of frequency moments

Space... the Final Frontier.

Star Trek

¹This is a metric between probability distributions μ and σ whose *square* is given by $\frac{1}{2} \sum_x (\sqrt{\mu(x)} - \sqrt{\sigma(x)})^2$. A different viewpoint of this definition is given in Section 3.

A major influence on the foundations of massive data sets as well as a pioneer of novel techniques has been the frequency moments problem and its variants [FM85, AMS99, Ind06, BJK⁺02, CCFC04, IW05]. In the k -th frequency moment problem F_k , the goal is to estimate the sum of the k -th power of the frequencies of items in a stream, presented as a sequence of non-negative updates. This paper deals with the case $k \geq 2$. The best space upper bound for this problem is $O(n^{1-2/k})$ up to polylogarithmic factors [AMS99, IW05] (a better dependence on the polylogarithmic term is in [BGKS06, MW]). But what about space lower bounds? It is here that communication complexity enters the picture.

Communication complexity [Yao79], one of the crown jewels of complexity theory, measures the necessary amount of communication in a distributed setting. It is often the case that computation problems are too complex or the models are too fine-grained to be amenable to analysis. Communication models judiciously abstract away details of the original problem and perhaps even weaken some of the restrictions of the original model. Their power resides in their simplicity, in which one can hope to gain traction for solving difficult problems. The complexity theorist's life, being hard as it is, sees some glimmer of hope in such things!

In this note, I will consider the *number-in-hand multiparty communication model*. Loosely speaking (see Section 2 for formal details), the input is partitioned amongst several players, and their goal is to compute some function of the input by exchanging messages via a shared blackboard. The communication cost of a protocol is the maximum length of this shared communication over all inputs. In a randomized protocol the players also have private access to random coins. The protocol solves the communication problem if the answer equals the value of the function to some desired confidence.

In order to show tight bounds for space, Alon, Matias and Szegedy [AMS99] introduced a generalization of set-disjointness in the t -party communication model. Each of the t players is given a subset of $[n]$ with the following promise: either the sets are pairwise disjoint (No instance) or they have a unique common element but are otherwise disjoint (Yes instance). They proved a communication complexity lower bound of $\Omega(n/t^4)$, which implies a space lower bound of $\Omega(n^{1-5/k})$ for estimating the frequency moment F_k , and thus is non-trivial only when $k > 5$. They left open the problem of getting a $\Omega(n/t)$ communication complexity lower bound on t -party set-disjointness in order to close this gap.

1.2 Information complexity ...

Bar-Yossef, Jayram, Ravi Kumar and Sivakumar [BJKS04] tackled this problem in an *information complexity* paradigm, hoping to prove the result via a direct sum argument. Information theoretic arguments have been used in previous work [Ab196, SS02, BCKO93], but information complexity was given first-class status as a resource measure first by Chakrabarti, Shi, Wirth, and Yao [CSWY01] for two-party simultaneous protocols in the context of proving direct sum theorems. Briefly speaking, information complexity of a communication problem f characterizes how much information about the inputs the players must reveal in a correct protocol for f . Bar-Yossef *et al.* considered a powerful generalization of this measure to general communication protocols. In particular, they introduced *conditional information complexity* as a

means to handle non-product distributions that are essential for proving tight lower bounds for multiparty set-disjointness.

By proving a direct-sum theorem, they reduced the problem to giving an $\Omega(1/t)$ information complexity bound on the multiparty AND problem: the t players each have a single bit with the promise that either at most one bit is set to 1 (NO instance) or all t bits are set to 1 (YES instance). Proving sub-constant lower bounds for information theoretic measures is somewhat unusual in that domain. By translating this problem to the domain of statistical divergences, especially using Hellinger distances, they obtained a non-optimal $\Omega(1/t^2)$ lower bound for general protocols. On the other hand, by using analytic techniques involving Rényi divergences, they obtained a near- $\Omega(1/t)$ optimal lower bound for the restricted one-way protocols. This created a gap between the two models even though it yielded near-optimal $\Omega(n^{1-2/k})$ space bounds for one-pass F_k estimation for all $k > 2$.

The situation was somewhat remedied by Chakrabarti, Khot and Sun [CKS03] who proved an $\Omega(1/t \log t)$ information complexity lower bound for AND. Recently, Gronemeier [Gro09] closed the gap to $\Omega(1/t)$. A common thread to both these papers is that they expand the information theory expressions directly in terms of analytic expressions (via Kullback-Liebler distance) since Rényi divergences seem to offer no advantage while dealing with general protocols. By using analytic techniques on the logarithm and associated functions, they manage to avoid the loss incurred by Bar-Yossef *et al.* in taking the Hellinger distance route.

1.3 ... to Hellinger distance

A thing of beauty is a joy for ever:
Its loveliness increases; it will never
Pass into nothingness; but still will keep

J. Keats

This brings me to the main thrust of this paper—proving an optimal $\Omega(1/t)$ information complexity lower bound for AND using Hellinger distance. An immediate concern is whether Hellinger distance is too weak to yield such an optimal bound, as perhaps has been the impression created in previous work either explicitly or implicitly. It is true that Hellinger distance can be arbitrarily smaller than Kullback-Liebler distance. On the other hand, expressing the information complexity of AND as a distance measure results in the *Jensen-Shannon* distance. Although this measure can be expressed using Kullback-Leibler distances, the form is quite restricted. Indeed, both Hellinger and Jensen-Shannon distances are within small constants of each other, so the *a priori* loss in transitioning to Hellinger distance is not significant. The *real* weakness in the Bar-Yossef *et al.* approach to using Hellinger distance amounted to the following: the expressions to be bounded involved the square of Hellinger distance and therefore, since Hellinger distance is a metric, but not its square, only a weak form of triangle inequality could be used. Unfortunately, that loss was significant and only yielded a sub-optimal $\Omega(1/t^2)$ bound.

I will demonstrate in this paper that Hellinger distance exposes the rich *geometric* structure of communication protocols. Since Hellinger distance is just a scaled Euclidean distance, its square is not a metric. Nevertheless, it has been studied extensively in the theory of metric spaces [DL97] under the area of *negative-type distances*. I will show that a simple geometric negative-type inequality suffices to overcome the lossy triangle inequality of the previous approach, thereby yielding an optimal bound for AND.

The inductive argument used in the paper is perhaps more intuitive because it explicitly shows where the protocol must *create* distances in order for it to be a valid communication protocol. Growing enough of these distances results in a (squared) Hellinger distance between a YES and NO instance of AND. For a correct protocol this must be constant, yielding the desired lower bound. As further evidence to the power of this geometric structure, Jayram and Woodruff [JW09] have shown that estimating the product norm $\ell_2 \circ \ell_0$ requires communication $\Omega(\sqrt{n})$, and Andoni, Jayram, and Patrascu [AJP09] have shown improved lower bounds for the communication complexity of edit distance.

A central message promoted in this paper is that transcript distributions have a natural place in the Euclidean space; taking square-roots of probabilities puts them in the unit sphere of ℓ_2 . Since pure states in a quantum system are naturally described this way, it would be interesting to explore the applicability of the techniques in the paper to quantum communication.

Section 2 contains the preliminaries including a review of information complexity notions. In Section 3, I will describe the key properties of Hellinger distance including the new ingredient needed in the proof, namely a negative-type inequality. These ingredients are combined in Section 4 in order to prove the main result.

2 Preliminaries

Suppose there are $t \geq 2$ players jointly holding an input $x = (x_1, x_2, \dots, x_t) \in \mathcal{X}^t$, where player i has x_i , for $i \in [t]$. Their goal is to solve some communication problem $f(x_1, x_2, \dots, x_t)$, defined on a *subset* of \mathcal{X}^t , by sending messages to each other. In this paper, the standard *blackboard* model will be used where the messages are all written on a shared medium. A *protocol* \mathcal{P} on \mathcal{X}^t specifies the rules for the players to write their messages on the blackboard when the inputs come from (all of) \mathcal{X}^t . The resulting sequence of messages is called the *transcript*. The maximum length of the transcript (in bits) over all inputs is the *communication cost* of the protocol \mathcal{P} . For technical reasons, it will be convenient not to require that the transcript also contain the answer. Instead, there is some referee who outputs an answer by looking only at the transcript and not the inputs. The protocol is allowed to be randomized in which each player, as well as the referee, has *private* access to an unlimited supply of random coins. The protocol solves the communication problem if the answer equals $f(x_1, x_2, \dots, x_t)$ with probability at least $1 - \delta$. Throughout this paper, δ will be a small constant and such protocols will be called as *correct* protocols. Note that the protocol itself is legally defined for all inputs in \mathcal{X}^t although no restriction is placed on the answer of the protocol outside the domain of f .

A family of sets $S_1, S_2, \dots, S_t \subseteq [n]$ is called a *sunflower with kernel* T if for every $i \neq j$, $S_i \cap S_j = T$. (These are also known as delta-systems.) In other words, if an element belongs to any

distinct pair of sets then it belongs to all of them, so in fact, the kernel equals $\bigcap_i S_i$. The *multi-party set-disjointness* communication problem, $\text{DISJ}_{t,n}$, with t players on a universe of size n is a (promise) decision problem where the input $S_1, S_2, \dots, S_t \subseteq [n]$ to the players is a sunflower whose kernel is either *empty* (No instance) or a *singleton* (Yes instance). A randomized private-coin communication protocol \mathcal{P} that solves $\text{DISJ}_{t,n}$ should accept YES instances and reject No instances with error probability at most δ .

To describe $\text{DISJ}_{t,n}$ as a valid Boolean formula over promise instances, encode the input of the players as bits as follows. Let $x = (x_{ij})$ denote a $t \times n$ array of bits. The i -th row of x is the characteristic vector of the set S_i . Then,

$$\text{DISJ}_{t,n}(x) = \bigvee_{j=1}^n \bigwedge_{i=1}^t x_{ij}.$$

Define

$$\text{AND}_t(u_1, u_2, \dots, u_t) \triangleq \bigwedge_{i=1}^t u_i,$$

with the promise that either at most one input bit is set to 1 (No instance) or all input bits are set to 1 (Yes instance). Letting $x^j \in \{0, 1\}^t$ denote the j -th column of x ,

$$\text{DISJ}_{t,n}(x^1, x^2, \dots, x^n) \triangleq \text{DISJ}_{t,n}(x) = \bigvee_{j=1}^n \text{AND}_t(x^j).$$

This way of splitting the input highlights the fact that the set-disjointness problem is an OR of n instances of the AND_t problem on t bits. It therefore suggests a direct-sum argument for proving communication lower bounds for $\text{DISJ}_{t,n}$.

I will now briefly review the *information complexity* paradigm for proving communication lower bounds via direct sum arguments, as developed in [BJKS04], for multi-party number-in-hand communication protocols. Information complexity of a communication problem f characterizes how much information about the inputs the players must reveal in a correct protocol for f . The underlying distribution on the inputs to the players can be *independent* across the players but in many cases the tight bounds are obtained by requiring dependent input distributions. This causes some complications which are overcome by introducing *conditional independence* on the inputs via auxiliary random variables. This is formalized below.

Notation. Random variables will be denoted by upper case Roman or Greek letters, and the values they take by corresponding lower case letters. Probability distributions will be denoted by lower case Greek letters. A random variable X with distribution μ is denoted by $X \sim \mu$. If μ is the uniform distribution over a set \mathcal{W} , then this is also denoted as $X \in_R \mathcal{W}$.

Definition 1. A distribution μ over \mathcal{X}^t is *partitioned* by η if there exists a joint probability space $(X_1, X_2, \dots, X_t, F)$ such that $(X_1, X_2, \dots, X_t) \sim \mu$, $F \sim \eta$, and X_1, X_2, \dots, X_t are jointly independent conditioned on F i.e. $P(X_1, X_2, \dots, X_t | F) = \prod_i P(X_i | F)$ \square

Definition 2 (Information Complexity). Let \mathcal{P} be a t -party randomized private-coin protocol on the input domain \mathcal{X}^t and let its random coins be denoted by the random variable R . Suppose μ is a distribution over \mathcal{X}^t partitioned by η in some joint probability space where $X = (X_1, X_2, \dots, X_t) \sim \mu$ and $F \sim \eta$. Extend this to a joint probability space over (X, F, R) such that (X, F) is independent of R . Now, let $\Pi = \Pi(X, R)$ be the random variable denoting the transcript of the protocol, where the randomness is *both* over the input distribution and the random coins of the protocol \mathcal{P} . The (*conditional*) *information cost* of \mathcal{P} under (μ, η) is defined to be $I(X : \Pi | F)$, i.e., the (Shannon) conditional mutual information between X and Π conditioned on F .

The *information complexity* of a communication problem f , denoted by $\text{IC}_\mu(f | \eta)$, is defined to be the minimum information cost of a correct protocol for f under (μ, η) . \square

Since $I(X : \Pi | D) \leq H(\Pi) \leq |\Pi|$, it suffices to prove lower bounds on the information cost of a correct protocol.

For the problem $\text{DISJ}_{t,n} = \sqrt{\text{AND}_t}$, I will first define a distribution (ν, ζ) for AND_t . Let $(U_1, U_2, \dots, U_t) \sim \nu$ and $G \sim \zeta$ be such that

1. $G \in_R [t]$. G picks a player whose bit will vary while the rest are fixed to 0.
2. Conditioned on the event $G = i$, let $(U_1, U_2, \dots, U_t) \in_R \{0, e_i\}$. Here, e_i is the standard basis vector with a 1 in the i -th position and 0 elsewhere.

The distribution for $\text{DISJ}_{t,n}$ is defined by letting $\mu = \nu^n$ and $\eta = \zeta^n$. In other words, if $X = (X^1, X^2, \dots, X^n)$ is the input and $F = (F^1, F^2, \dots, F^n)$ is the auxiliary random variable, then independently for each $j \in [n]$, $F^j \sim \zeta$ and $X^j \sim \nu$.

Proposition 3 (Direct Sum for Information Complexity [BJKS04]).

$$\text{CC}(\text{DISJ}_{t,n}) \geq \text{IC}_\mu(\text{DISJ}_{t,n} | \eta) \geq n \cdot \text{IC}_\nu(\text{AND}_t | \zeta). \quad \square$$

Consequently, I will show an $\Omega(1/t)$ lower bound on the information complexity of AND_t .

3 Hellinger Distance

Notation. Let $\|\cdot\|$ denote the standard ℓ_2 norm and $\|\cdot\|_1$ denote the standard ℓ_1 norm.

Let u be an input to a protocol \mathcal{P} . Let $\pi(u)$ denote the probability distribution over the transcripts induced by \mathcal{P} on input u , where the randomness is over the private coins of \mathcal{P} . Let $\pi(u)_\tau$ denote the probability that the transcript equals τ . Viewing $\pi(u)$ as an element of ℓ_1 , note that $\|\pi(u)\|_1 = \sum_\tau \pi(u)_\tau = 1$.

The following switch in viewpoint is the perhaps most important notion in this paper. Consider the element $\psi(u) \in \ell_2$ obtained via the square-root map $\pi(u) \mapsto \psi(u) = \sqrt{\pi(u)}$. This means $\psi(u)_\tau = \sqrt{\pi(u)_\tau}$ for all τ . The central tenet is that $\psi(u)$ is an object that deserves real attention on its own right from the standpoint of information complexity. Now, $\|\psi(u)\| =$

$\|\pi(u)\|_1 = 1$, and so $\psi(u) \in \mathbb{S}_+$, where \mathbb{S}_+ denotes the unit sphere in ℓ_2 restricted to the non-negative orthant. In analogy with quantum physics, call $\psi(u)$ the *transcript wave function* of u in \mathcal{P} .

Definition 4 (Hellinger Distance). The *Hellinger distance* between $\psi_1, \psi_2 \in \mathbb{S}_+$ is a scaled Euclidean distance defined as $h(\psi_1, \psi_2) \triangleq \frac{1}{\sqrt{2}}\|\psi_1 - \psi_2\|$. \square

Since $\|\psi_1 - \psi_2\|^2 \leq \|\psi_1\|^2 + \|\psi_2\|^2 = 2$, the scaling ensures that Hellinger distance is always between 0 and 1. To emphasize the geometric nature of Hellinger distance, I will almost exclusively use the norm notation to refer to Hellinger distance.

The following properties of Hellinger distance are well-known (see [BJKS04]):

Proposition 5 (Hellinger distance and communication protocols). *Let \mathcal{P} be a randomized t -party private-coin protocol on the input domain \mathcal{X}^t . Let g be a decision problem defined on a subset of \mathcal{X}^t . Let $u, v \in \mathcal{X}^t$ be two distinct inputs whose transcript wave functions in \mathcal{P} are denoted by $\psi(u)$ and $\psi(v)$, respectively.*

1. **Mutual information to Hellinger distance:** *Suppose $U \in_R \{u, v\}$. If Π denotes the transcript random variable, then*

$$I(U : \Pi) \geq \frac{1}{2}\|\psi(u) - \psi(v)\|^2.$$

2. **Soundness:** *If \mathcal{P} is a correct protocol for g , and $g(u) \neq g(v)$, then*

$$\frac{1}{2}\|\psi(u) - \psi(v)\|^2 \geq 1 - 2\sqrt{\delta}.$$

3. **Cut-and-paste:** *Let u' and v' denote the inputs obtained by performing some cut-and-paste on u and v . In other words for each $1 \leq i \leq t$, either (a) $u'_i = u_i$ and $v'_i = v_i$ or (b) $u'_i = v_i$ and $v'_i = u_i$. Then*

$$\|\psi(u) - \psi(v)\| = \|\psi(u') - \psi(v')\|.$$

Consequently, suppose the inputs to \mathcal{P} are such that each player holds a single bit, i.e., $\mathcal{X} = \{0, 1\}$. Identify the input $u \in \mathcal{X}^t$ with the subset $A = \{i \mid u_i = 1\} \subseteq [t]$. Similarly, identify v with $B \subseteq [t]$. Then

$$\|\psi(A) - \psi(B)\| = \|\psi(A \cup B) - \psi(A \cap B)\|. \quad \square$$

Property 1 in the above proposition is just a restatement of the fact that the Jensen-Shannon distance between $\psi(u)$ and $\psi(v)$ is bounded from below by their Hellinger distance. Property 2 follows by relating Hellinger to variational distance and then invoking the correctness of the protocol. Property 3 generalizes the rectangle property of deterministic communication protocols to randomized protocols. The corollary to this property, where each player holds a single bit, follows by letting $u'_i = u_i \vee v_i$ and $v'_i = u_i \wedge v_i$ for all i .

The next inequality is the new key ingredient that enables the tight lower bound for AND_t via Hellinger distance:

Proposition 6. For any $v_0, v_1, v_2, \dots, v_s \in \ell_2$,

$$\sum_{i=1}^s \|v_0 - v_i\|^2 \geq \frac{1}{s} \sum_{1 \leq i < j \leq s} \|v_i - v_j\|^2$$

Proof. This is a special case of a general class of negative-type inequalities [DL97] satisfied by the square of the ℓ_2 -distance: for any set of real numbers b_0, b_1, \dots, b_s such that $\sum_{i=0}^s b_i = 0$, it holds that

$$\sum_{\substack{0 \leq i \leq s \\ 0 \leq j \leq s}} b_i b_j \|v_i - v_j\|^2 \leq 0.$$

The above inequality is simple to derive and I will show this below for the sake of completeness. Setting $b_0 = s$ and $b_1 = b_2 = \dots = b_s = -1$ yields the statement of the proposition.

Observe that:

$$\begin{aligned} \sum_{\substack{0 \leq i \leq s \\ 0 \leq j \leq s}} b_i b_j \|v_i - v_j\|^2 &= \sum_{\substack{0 \leq i \leq s \\ 0 \leq j \leq s}} b_i b_j (\|v_i\|^2 + \|v_j\|^2 - 2\langle v_i, v_j \rangle) \\ &= \left(\sum_{0 \leq i \leq s} b_i \|v_i\|^2 \sum_{0 \leq j \leq s} b_j \right) + \left(\sum_{0 \leq j \leq s} b_j \|v_j\|^2 \sum_{0 \leq i \leq s} b_i \right) \\ &\quad - 2 \left(\sum_{0 \leq i \leq s} b_i v_i \right) \cdot \left(\sum_{0 \leq j \leq s} b_j v_j \right) \\ &= 0 + 0 - 2 \left\| \sum_{0 \leq i \leq s} b_i v_i \right\|^2 \\ &\leq 0, \end{aligned}$$

proving the inequality. □

4 Information Complexity of AND_t

Beauty is the first test: there is no permanent place in the world for ugly mathematics.

G.H. Hardy

The following is the main technical result of this note.

Theorem 7. Let \mathcal{D} be a t -party protocol on the input domain $\{0, 1\}^t$. Identify every subset of $[t]$ with its characteristic vector in $\{0, 1\}^t$. Let $\psi(A)$ denote the transcript wave function of input $A \subseteq [t]$ in \mathcal{D} . Suppose A_1, A_2, \dots, A_s are a pairwise disjoint collection of $s = 2^k$ subsets of $[t]$, where $k \geq 0$. Set $A \triangleq \bigcup_i A_i$. Then,

$$\sum_{i=1}^s \|\psi(\emptyset) - \psi(A_i)\|^2 \geq \|\psi(\emptyset) - \psi(A)\|^2 \cdot \prod_{\ell=1}^k \left(1 - \frac{1}{2^\ell}\right)$$

Proof. By induction on k . The base case $k = 0$ (i.e., $s = 1$) follows trivially with equality. For the induction step, let $k \geq 1$ so that $s = 2^k$ is even. Now,

$$\begin{aligned}
& \sum_{i=1}^s \|\psi(\emptyset) - \psi(A_i)\|^2 \\
& \geq \frac{1}{s} \sum_{1 \leq i < j \leq s} \|\psi(A_i) - \psi(A_j)\|^2 && \text{(Proposition 6)} \\
& = \frac{1}{s} \sum_{1 \leq i < j \leq s} \|\psi(\emptyset) - \psi(A_i \cup A_j)\|^2 && \text{(Proposition 5, cut-and-paste)} \tag{1}
\end{aligned}$$

Associate $\{(i, j) \mid 1 \leq i < j \leq s\}$ with the edges of the complete graph K_s . Since s is even, K_s can be decomposed into an edge-disjoint union of $s - 1$ perfect matchings, $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_{s-1}$, each having $s/2$ edges. Using this, rewrite the expression within the sum in (1) as follows:

$$\sum_{1 \leq i < j \leq s} \|\psi(\emptyset) - \psi(A_i \cup A_j)\|^2 = \sum_{p=1}^{s-1} \sum_{\{i, j\} \in \mathcal{M}_p} \|\psi(\emptyset) - \psi(A_i \cup A_j)\|^2 \tag{2}$$

Fix a p within the sum. The sets $A_i \cup A_j$, for $\{i, j\} \in \mathcal{M}_p$, are a pairwise disjoint collection of $s/2 = 2^{k-1}$ sets. By the induction hypothesis,

$$\sum_{\{i, j\} \in \mathcal{M}_p} \|\psi(\emptyset) - \psi(A_i \cup A_j)\|^2 \geq \|\psi(\emptyset) - \psi(A)\|^2 \cdot \prod_{\ell=1}^{k-1} \left(1 - \frac{1}{2^\ell}\right)$$

Substitute this bound in (2) for every p , and then combine it with (1) to get:

$$\begin{aligned}
\sum_{i=1}^s \|\psi(\emptyset) - \psi(A_i)\|^2 & \geq \frac{1}{s} (s-1) \cdot \|\psi(\emptyset) - \psi(A)\|^2 \cdot \prod_{\ell=1}^{k-1} \left(1 - \frac{1}{2^\ell}\right) \\
& = \left(1 - \frac{1}{2^k}\right) \cdot \|\psi(\emptyset) - \psi(A)\|^2 \cdot \prod_{\ell=1}^{k-1} \left(1 - \frac{1}{2^\ell}\right) \\
& = \|\psi(\emptyset) - \psi(A)\|^2 \cdot \prod_{\ell=1}^k \left(1 - \frac{1}{2^\ell}\right),
\end{aligned}$$

proving the theorem. □

Corollary 8. *The information complexity of AND_t is $\Omega(1/t)$.*

Proof. Let $U \sim \nu$ and $G \sim \zeta$. Let \mathcal{P} be a correct protocol for AND_t whose information cost under (ν, ζ) equals C . If Π denotes the transcript, then

$$C = I(U : \Pi \mid G) = \frac{1}{t} \sum_{i=1}^t I(U : \Pi \mid G = i)$$

Conditioned on $G = i$, $U \in_R \{0, e_i\}$. Applying the Mutual-information-to-Hellinger-distance property in Proposition 5,

$$C \geq \frac{1}{t} \sum_{i=1}^t \frac{1}{2} \|\psi(0) - \psi(e_i)\|^2 = \frac{1}{t} \sum_{i=1}^t \frac{1}{2} \|\psi(\emptyset) - \psi(\{i\})\|^2$$

Suppose for the moment that $t = 2^k$ is a power of 2 with $k \geq 1$. Applying Theorem 7 with $s = t$ and $A_i = \{i\}$, for $1 \leq i \leq t$, to the RHS above,

$$C \geq \frac{1}{t} \cdot \left(\frac{1}{2} \|\psi(\emptyset) - \psi([t])\|^2\right) \cdot \prod_{\ell=1}^k \left(1 - \frac{1}{2^\ell}\right) \quad (3)$$

Since $\text{AND}_t(\emptyset) \neq \text{AND}_t([t])$, the soundness property in Proposition 5 applied to \mathcal{P} implies the following:

$$\frac{1}{2} \|\psi(\emptyset) - \psi([t])\|^2 \geq 1 - 2\sqrt{\delta} \quad (4)$$

For the product term in (3),

$$\prod_{\ell=1}^k \left(1 - \frac{1}{2^\ell}\right) \geq \prod_{\ell=1}^{\infty} \left(1 - \frac{1}{2^\ell}\right) = 0.288788\dots^2 \quad (5)$$

Substituting the bounds in (4) and (5) into (3) shows that the information cost of \mathcal{P} is $\Omega(1/t)$.

For arbitrary values of t , a minor modification yields the same asymptotic bound. Let t' be the largest power of 2 that is at most t . Partition $[t]$ in some arbitrary manner into a collection of t' sets $A_1, A_2, \dots, A_{t'}$ of sizes 1 and 2 that are pairwise disjoint. For each set $\{i, j\}$ of size 2, apply Theorem 7 with $k = 1$ to bound

$$\|\psi(\emptyset) - \psi(\{i\})\|^2 + \|\psi(\emptyset) - \psi(\{j\})\|^2 \geq \frac{1}{2} \|\psi(\emptyset) - \psi(\{i, j\})\|^2.$$

Thus,

$$\sum_{i=1}^t \|\psi(\emptyset) - \psi(\{i\})\|^2 \geq \frac{1}{2} \cdot \sum_{k=1}^{t'} \|\psi(\emptyset) - \psi(A_k)\|^2.$$

Then proceed with the same argument. □

Remark. The proof shows that the constant in $\Omega(1/t)$ is $c(1 - 2\sqrt{\delta})$ where c equals $0.288788\dots$, the digital search tree constant, if t is a power of 2, and half that value otherwise.

²This constant is known as the digital search tree constant (see Sloane's A048651 [Slo]). It also has connections to random binary matrices. Euler studied this in the context of generating functions for integer partitions, and gave methods to compute the infinite product that converge fairly rapidly. Thanks to Laurens Gunnarsen for discussions on this topic.

References

- [Abl96] F. Ablyev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoretical Computer Science*, 157(2):139–159, 1996.
- [AJP09] Alexandr Andoni, T.S. Jayram, and Mihai Patrascu. Non-embeddability and sketching complexity via information geometry, 2009.
- [AMS99] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.
- [BCKO93] R. Bar-Yehuda, B. Chor, E. Kushilevitz, and A. Orlitsky. Privacy, additional information, and communication. *IEEE Transactions on Information Theory*, 39(6):1930–1943, 1993.
- [BGKS06] Lakshminath Bhuvanagiri, Sumit Ganguly, Deepanjan Kesh, and Chandan Saha. Simpler algorithm for estimating frequency moments of data streams. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006, Miami, Florida, USA, January 22-26, 2006*, pages 708–713. ACM Press, 2006.
- [BJK⁺02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, D. Sivakumar, and Luca Trevisan. Counting distinct elements in a data stream. In José D. P. Rolim and Salil P. Vadhan, editors, *RANDOM*, volume 2483 of *Lecture Notes in Computer Science*, pages 1–10. Springer, 2002.
- [BJKS04] Ziv Bar-Yossef, T.S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.
- [CCFC04] Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. *Theor. Comput. Sci.*, 312(1):3–15, 2004.
- [CKS03] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal lower bounds on the multiparty communication complexity of set-disjointness. In *Proceedings of the 18th Annual IEEE Conference on Computational Complexity*, pages 107–117, 2003.
- [CSWY01] A. Chakrabarti, Y. Shi, A. Wirth, and A. C-C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 270–278, 2001.
- [DL97] M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Springer, 1997.
- [FM85] Philippe Flajolet and G. Nigel Martin. Probabilistic counting algorithms for data base applications. *J. Comput. Syst. Sci.*, 31(2):182–209, 1985.

- [Gro09] Andre Gronemeier. Asymptotically optimal lower bounds on the nih-multi-party information complexity of the and-function and disjointness. In Susanne Albers and Jean-Yves Marion, editors, *STACS*, volume 09001 of *Dagstuhl Seminar Proceedings*, pages 505–516. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2009.
- [Ind06] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006.
- [IW05] Piotr Indyk and David P. Woodruff. Optimal approximations of the frequency moments of data streams. In *STOC*, pages 202–208, 2005.
- [JW09] T.S. Jayram and David Woodruff. The data stream space complexity of cascaded norms. Submitted, 2009.
- [MW] M. Monemizadeh and D. Woodruff. l_p -sampling with applications. Manuscript.
- [Slo] N. Sloane. The on-line encyclopedia of integer sequences! <http://www.research.att.com/njas/sequences/A048651>.
- [SS02] M. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 360–369, 2002.
- [Yao79] A. C-C. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th ACM Symposium on Theory of Computing (STOC)*, pages 209–213, 1979.