Asymptotically Optimal Lower Bounds on the NIH-Multi-Party Information Complexity of the AND-Function and Disjointness

André Gronemeier

Symposium on Theoretical Aspects of Computer Science 2009 (Freiburg), pp. 505–516
www.stacs-conf.org

ASYMPTOTICALLY OPTIMAL LOWER BOUNDS ON THE NIH-MULTI-PARTY INFORMATION COMPLEXITY OF THE AND-FUNCTION AND DISJOINTNESS

ANDRÉ GRONEMEIER [1]

[1] Technische Universität Dortmund, Lehrstuhl Informatik 2, 44227 Dortmund, Germany
E-mail address: andre.gronemeier@cs.uni-dortmund.de

Abstract. Here we prove an asymptotically optimal lower bound on the information complexity of the k-party disjointness function with the unique intersection promise, an important special case of the well known disjointness problem, and the AND_k-function in the number in the hand model. Our $\Omega(n/k)$ bound for disjointness improves on an earlier $\Omega(n/(k \log k))$ bound by Chakrabarti *et al.* (2003), who obtained an asymptotically tight lower bound for one-way protocols, but failed to do so for the general case. Our result eliminates both the gap between the upper and the lower bound for unrestricted protocols and the gap between the lower bounds for one-way protocols and unrestricted protocols.

1. Introduction

Primarily, communication complexity, introduced by Yao [10], deals with the amount of communication that is needed in distributed computation, but apart from distributed computation, nowadays communication complexity has found applications in virtually all fields of complexity theory. The book by Kushilevitz and Nisan [9] gives a comprehensive introduction to communication complexity and its applications.

Suppose that k players, each of them knowing exactly one argument of a function $f(x_1, \ldots, x_k)$ with k arguments, want to evaluate the function for the input that is distributed among them. Clearly, to succeed at this task the players need to communicate. Here we consider the case that the players communicate by writing to a blackboard that is shared by all players. The rules that determine who writes which message to the blackboard are usually called a *protocol*. The protocol terminates if the value of the function can be inferred from the contents of the blackboard, the so-called *transcript* of the protocol. Then the communication complexity of the function is the minimum number of bits that the players need to write to the blackboard in the worst case to jointly compute the result. This setting is usually called the *number in the hand model* since each part of the input is exclusively known to a single player who figuratively hides the input in his hand. In the randomized version of this model each player has access to a private source of unbiased independent random bits and his actions may depend on his input and his random bits. For a *randomized ε-error protocol* the output of the protocol may be different from the value

Key words and phrases: computational complexity, communication complexity.

of the function f with probability at most ε. The ε-error randomized communication complexity of a function is defined in the obvious way. A formal definition of k-party protocols can be found in [9]. Note that there are also other models of multi-party communication, but these models are not the topic of this paper.

In recent publications [5, 2, 3, 4] lower bounds on the communication complexity of functions have been obtained by using information theoretical methods. In this context communication complexity is supplemented by an information theoretical counterpart, the information complexity of a function. Roughly, the information complexity of a function f is the minimal amount of information that the transcript of a protocol for f must reveal about the input. Besides being a lower bound for the communication complexity, information complexity has additional nice properties with respect to so-called direct sum problems.

1.1. Our Result

In this paper we will prove an asymptotically optimal lower bound on the communication complexity of the multi-party set disjointness problem with the unique intersection promise.

Definition 1.1. In the k-party set disjointness problem each of the players is given the characteristic vector of a subset of an n-element set. It is promised that the subsets are either pairwise disjoint or that there is a single element that is contained in all subsets and that the subsets are disjoint otherwise. The players have to distinguish these two cases, the output of a protocol for set disjointness should be 0 in the first case and 1 in the second case. If the promise is broken, then the players may give an arbitrary answer.

Here we will prove the following result about the randomized communication complexity of the multi-party set disjointness problem in the number in the hand model.

Theorem 1.2. *For every sufficiently small constant $\varepsilon > 0$ the randomized ε-error communication complexity of the k-party set disjointness problem with the unique intersection promise is bounded from below by $\Omega(n/k)$.*

By the upper bound shown in [4] this result is asymptotically optimal with respect to the number of players k and the size of the inputs n. An important application of this problem is the proof of a lower bound for the memory requirements of certain data stream algorithms [1]. Our improvement of the lower bound for disjointness does not have a significant impact on this application. But we think that the disjointness problem is interesting and important on its own since it is a well-known basic problem in communication complexity theory [1, 3, 4, 9]. Up to now the best known lower bound was $\Omega(n/(k \log k))$ by Chakrabarti, Khot, and Sun [4], who also proved an asymptotically optimal lower bound for one-way protocols. This result left a gap both between the upper and the lower bound and between the lower bounds for one-way protocols and unrestricted protocols. Our result closes these gaps.

Like the earlier results, our lower bound is based on an information theoretical approach. The main ingredient of this approach is a lower bound on the information complexity of the AND_k-function, the Boolean conjunction of k bits. Since Theorem 1.2 will be a simple corollary of this result, and more importantly, since AND_k is a basic building block of any computation, the lower bound on the information complexity of AND_k is the main result of this paper. We postpone the precise statement of this result to Theorem 3.2 in

Section 3 because some preparing definitions are needed beforehand. But we stress here that our result also closes the gap between the upper and lower bound on the conditional information complexity of AND_k for unrestricted protocols and the gap between the lower bounds on the information complexity of AND_k for one-way protocols and unrestricted protocols that was left open in [4].

1.2. Related Work

The general disjointness problem without the unique intersection promise has a long history in communication complexity theory. Here we focus only on recent results for the multi-party set disjointness problem with the unique intersection promise, and especially on lower bounds that rely on information complexity arguments. For older results we refer the reader to the book by Kushilevitz and Nisan [9] and the references therein.

Alon, Matias, and Szegedy [1] proved an $\Omega(n/k^4)$ lower bound for multi-party set disjointness and applied this bound to prove lower bounds for the memory requirements of data stream algorithms. Bar-Yossef, Jayram, Kumar, and Sivakumar [3] improved this to a lower bound of $\Omega(n/k^2)$. They introduced the direct sum approach on which later results, including our result, are based and proved that the information complexity of AND_k is bounded from below by $\Omega(1/k^2)$. Chakrabarti, Khot, and Sun [4] improved the lower bound for the information complexity of AND_k to $\Omega(1/(k \log k))$ and thereby improved the lower bound for multi-party set disjointness to $\Omega(n/(k \log k))$. They also proved an asymptotically optimal lower bound for one-way protocols, a restricted model in which the players communicate in a predetermined order. Our result improves on these results, but furthermore we think that our proof technique is a useful contribution to the framework for which Bar-Yossef *et al.* [3] coined the term "information statistics". Bar-Yossef *et al.* use this term for the combination of information theory and other statistical metrics on probability spaces. We use the direct sum approach from [3], but instead of the Hellinger distance that is used in [3] we use the Kullback Leibler distance. Since the Kullback Leibler distance is closely related to mutual information, we do not loose precision in the transition from information theory to statistical distance measures. By this, we are able to prove sharper bounds. Like Chakrabarti *et al.* [4], we take a closer look at the analytical properties of the functions that are involved. Our improvements on this result are also due to the fact that our Kullback Leibler distance based arguments are very close to the information theory domain.

2. Preliminaries

2.1. Notation

We use lower case letters for constants and variables and upper case letters for random variables. If the random variables X and Y have the same distribution, we briefly write $X \sim Y$. For vector-valued variables we use a boldface font. For example, $\mathbf{X} = (X_1, \ldots, X_k)$ is a random vector whose components are the random variables X_i for $i = 1, \ldots, k$. In this case let $\mathbf{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$ denote the vector $\mathbf{X}$ without the ith component. A boldface zero $\mathbf{0}$ and boldface one $\mathbf{1}$ denote the all-zero vector and all-one vector of appropriate size, respectively. Thus $\mathbf{X}_{-i} = \mathbf{0}$ says that $X_j = 0$ for all $j \in \{1, \ldots, k\} - \{i\}$. For sums like $\sum_{i=0}^{n} a_i$ we sometimes do not explicitly specify the bounds of summation and

just write $\sum_i a_i$. In this case the sum is taken over the set of all values of i for which a_i is meaningful. This set must be derived from context. For example, the sum $\sum_v f(\Pr\{X=v\})$ should be taken over all values v in the range of X. All logarithms, denoted by log, are with respect to base 2.

2.2. Information Theory

Here we can merely define our notation for the basic quantities from information theory and cite some results that are needed in this paper. For a proper introduction to information theory we refer the reader to the book by Cover and Thomas [6]. In the following let h_2 denote the binary entropy function $h_2(p) = -p\log p - (1-p)\log(1-p)$ for $p \in [0,1]$. Let X, Y, and Z be random variables and let E be an event, for example the event $Y = y$. Then $\mathrm{H}(X)$ denotes the entropy of the random variable X and $\mathrm{H}(X|E)$ denotes the entropy of X with respect to the conditional distribution of X given that the event E occurred. If there are several events separated by commas, then we analogously use the conditional distribution of X given that all of the events occurred. Let $\mathrm{H}(X|Y)$ denote the conditional entropy of X given Y. Recall that $\mathrm{H}(X|Y) = \sum_y \Pr\{Y=y\}\,\mathrm{H}(X|Y=y)$. If we condition on several variables, we separate the variables by commas. If we mix events and variables in the condition, we first list the variables, after that we list the events, for example $\mathrm{H}(X|Y, Z=z)$. The mutual information of X and Y is $\mathrm{I}(X:Y) = \mathrm{H}(X) - \mathrm{H}(X|Y)$ and $\mathrm{I}(X:Y|E) = \mathrm{H}(X|E) - \mathrm{H}(X|Y,E)$ is the mutual information of X and Y with respect to the conditional distribution of X and Y given that the event E occurred. The conditional mutual information of X and Y given Z is $\mathrm{I}(X:Y|Z) = \mathrm{H}(X|Z) - \mathrm{H}(X|Y,Z)$. Recall that $\mathrm{I}(X:Y|Z) = \sum_z \Pr\{Z=z\}\,\mathrm{I}(X:Y|Z=z)$.

Suppose that the random variables X and Y have the same range. Then the Kullback Leibler distance of their distributions is $\mathrm{D}(X,Y) = \sum_v \Pr\{X=v\}\log\frac{\Pr\{X=v\}}{\Pr\{Y=v\}}$. If $\Pr\{X=v\} = 0$ in the above sum, then the corresponding term is 0 independently of the value of $\Pr\{Y=v\}$, by continuity arguments. If $\Pr\{X=v\} \neq 0$ and $\Pr\{Y=v\} = 0$ for some v, then the whole sum is defined to be equal to ∞. If E is an event, then $(X|E)$ denotes the conditional distribution of X given that the event E occurred, for example $\mathrm{D}((X|E), X)$ is the Kullback Leibler distance of the conditional distribution of X given that the event E occurred and the distribution of X. Recall that the mutual information of X and Y is the Kullback Leibler distance of the joint distribution (X,Y) and the product distribution of the marginal distributions:

$$\mathrm{I}(X:Y) = \sum_{x,y} \Pr\{X=x, Y=y\} \cdot \log \frac{\Pr\{X=x, Y=y\}}{\Pr\{X=x\}\cdot\Pr\{Y=y\}}\ .$$

The following lemma is a useful tool for the proof of lower bounds on the Kullback Leibler distance of distributions. A proof of the log sum inequality can be found in [6].

Lemma 2.1 (Log sum inequality)**.** *For nonnegative numbers a_i and b_i, where $i = 1, \ldots, n$,*

$$\sum_i a_i \log\frac{a_i}{b_i} \geq \left(\sum_i a_i\right)\log\frac{\sum_i a_i}{\sum_i b_i}\ .$$

Suppose that the random variables X and Y have the same finite range R. Then the total variation distance of their distributions is $\mathrm{V}(X,Y) = \frac{1}{2}\sum_v |\Pr\{X=v\} - \Pr\{Y=v\}|$. It is a well-known fact (see e.g. [7]) that $\mathrm{V}(X,Y) = \max_{S\subseteq R} |\Pr\{X\in S\} - \Pr\{Y\in S\}|$.

The following lemma by Kullback relates the Kullback Leibler distance of distributions to their total variation distance.

Lemma 2.2 (Kullback [8])**.** *Suppose that X and Y are random variables that have the same finite range. Then* $\mathrm{D}(X,Y) \geq 2 \cdot \mathrm{V}(X,Y)^2$.

2.3. Information Complexity

The notion of the information cost of a protocol was introduced by Chakrabarti, Shi, Wirth, and Yao [5]. The information cost of a randomized protocol is the mutual information of the input and the transcript of the protocol. Then the information complexity of a function can be defined in the canonical way. Here we will use the conditional information complexity of a function, a refinement that was introduced by Bar-Yossef, Jayram, Kumar, and Sivakumar [3].

Definition 2.3. Let B be a set, let $f\colon B^k \longrightarrow \{0,1\}$ be a function, and let $\mathbf{X} \in B^k$ and D be random variables. Suppose that P is a randomized k-party protocol for f and that $M(\mathbf{X})$ is the transcript of P for the input $\mathbf{X}$. Then the conditional information cost of P with respect to $\mathbf{X}$ and D is defined by

$$\mathrm{icost}(P;\mathbf{X}|D) = \mathrm{I}(M(\mathbf{X}) : \mathbf{X}|D) \; .$$

The conditional ε-error information complexity $\mathrm{IC}_\varepsilon(f;\mathbf{X}|D)$ of f w.r.t. $\mathbf{X}$ and D is the minimal conditional information cost of a communication protocol for $f(\mathbf{X})$ where the minimum is taken over all randomized ε-error protocols for f.

The information complexity of a function is a lower bound for the communication complexity. A proof of the next theorem can be found in [3].

Theorem 2.4. *Let B be a set, let $f\colon B^k \longrightarrow \{0,1\}$ be a function, and let $\mathbf{X} \in B^k$ and D be random variables. Then the ε-error communication complexity of f is bounded from below by* $\mathrm{IC}_\varepsilon(f;\mathbf{X}|D)$.

2.4. The Direct Sum Paradigm

Information complexity has very nice properties with respect to direct sum problems. In this section we summarize the approach of Bar-Yossef, Jayram, Kumar and Sivakumar [3] using a slightly different terminology. We call a problem f a direct sum problem if it can be decomposed into simpler problems of smaller size.

Definition 2.5. Let $f\colon (B^n)^k \longrightarrow \{0,1\}$ be a function and let $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,n}) \in B^n$ for $i = 1, \ldots, k$. If there are functions $g\colon \{0,1\}^n \longrightarrow \{0,1\}$ and $h\colon B^k \longrightarrow \{0,1\}$ such that

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_k) = g\left(\, h(x_{1,1}, x_{2,1}, \ldots, x_{k,1}) \,, \; \ldots \; , \, h(x_{1,n}, x_{2,n}, \ldots, x_{k,n}) \,\right)$$

then the function f is called a g-h-direct sum.

Here the goal is to express a lower bound on the conditional information complexity of f in terms of the conditional information complexity of the simpler function h and the parameter n. In order for this approach to work, the joint distribution of the inputs of h and the condition must have certain properties. As a first requirement, the condition must partition the distribution of the inputs into product distributions.

Definition 2.6. Let B be a set and let $\mathbf{X} = (X_1, \ldots, X_k) \in B^k$ and D be random variables. The variable D partitions $\mathbf{X}$, if for every d in the support of D the conditional distribution $(\mathbf{X}|D{=}d)$ is the product distribution of the distributions $(X_i|D{=}d)$ for $i = 1, \ldots, k$.

The function f can be decomposed into instances of the function h if the distribution of the inputs of f satisfies our second requirement.

Definition 2.7. Let B be a set, let $g\colon \{0,1\}^n \longrightarrow \{0,1\}$ and $h\colon B^k \longrightarrow \{0,1\}$ be functions, and let $\mathbf{X} \in B^k$ be a random variable. If for every $i \in \{1, \ldots, n\}$, for every $a \in B^k$, and for every $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in (B^k)^n$ such that $\mathbf{x}_j \in \text{support}(\mathbf{X})$ for all j

$$g\left(h(x_1), \ldots, h(x_{i-1}), h(a), h(x_{i+1}), \ldots, h(x_n)\right) = h(a)$$

then the distribution of $\mathbf{X}$ is called collapsing for g and h.

If these two requirements are met, then the conditional information complexity of f can be expressed in terms of the conditional information complexity of h and the parameter n.

Theorem 2.8 (Bar-Yossef *et al.* [3])**.** *Suppose that $f\colon (B^n)^k \longrightarrow \{0,1\}$ is a g-h-direct sum and that $\mathbf{X} \in B^k$ and D are random variables such that the distribution of $\mathbf{X}$ is collapsing for g and h and D partitions $\mathbf{X}$. Let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_k) \in (B^n)^k$ and $\mathbf{E} \in \text{support}(D)^n$ be random variables and let Y_i^j and E^j denote the projection of $\mathbf{Y}_i$ and $\mathbf{E}$ to the jth coordinate, respectively. If the random variables $\mathbf{V}_j = ((Y_1^j, \ldots, Y_k^j), E^j)$ for $j = 1, \ldots, n$ are independent and $\mathbf{V}_j \sim (\mathbf{X}, D)$ for all j, then $\text{IC}_\varepsilon(f; \mathbf{Y}|\mathbf{E}) \geq n \cdot \text{IC}_\varepsilon(h; \mathbf{X}|D)$.*

This direct sum approach can be applied to the k-party set disjointness problem.

Observation 2.9. Let AND_ℓ and OR_ℓ denote the Boolean conjunction and disjunction of ℓ bits, respectively. Then the k-party set disjointness problem is a OR_n-AND_k-direct sum.

Consequently, for the proof of Theorem 1.2 it is sufficient to prove a lower bound on the conditional information complexity of AND_k for a distribution that satisfies the requirements of Theorem 2.8 and, in addition, honors the unique intersection promise. A distribution with these properties is defined in the following section. This approach was already used in [3] and [4].

3. The Information Complexity of AND_k

For the following distribution of D and the input $\mathbf{Z} = (Z_1, \ldots, Z_k)$ of AND_k the variable D partitions $\mathbf{Z}$ and the distribution of $\mathbf{Z}$ is collapsing for OR_n and AND_k. Additionally, there is at most a single i such that $Z_i = 1$.

Definition 3.1. From here on let $\mathbf{Z} = (Z_1, \ldots, Z_k) \in \{0,1\}^k$ and $D \in \{1, \ldots, k\}$ be random variables such that the joint distribution of $\mathbf{Z}$ and D has the following properties: D is uniformly distributed in $\{1, \ldots, k\}$. For all $i \in \{1, \ldots, k\}$ we have $\Pr\{Z_j{=}0|D{=}i\} = 1$ for $j \neq i$ and $\Pr\{Z_i{=}0|D{=}i\} = \Pr\{Z_i{=}1|D{=}i\} = \frac{1}{2}$.

Now we can state the main result of this paper, an asymptotically optimal lower bound on the information complexity of the AND_k-function for inputs that are distributed according to the last definition.

Theorem 3.2. *Let $\varepsilon < \frac{3}{10}\left(1 - \sqrt{\frac{1}{2}\log\frac{4}{3}}\right)$ be a constant. Then there is a constant $c(\varepsilon) > 0$ that does only depend on ε such that $\text{IC}_\varepsilon(\text{AND}_k; \mathbf{Z}|D) \geq c(\varepsilon)/k$.*

It is easy to see that $\mathrm{icost}(P;\mathbf{Z}|D) = 1/k$ for a trivial deterministic protocol P for AND_k where each player in turn writes his input to the blackboard until the first 0 is written. Therefore our lower bound is optimal. As we have seen, this result immediately implies Theorem 1.2, the other main result of this paper. In the rest of the paper we will outline the proof of Theorem 3.2.

3.1. Some Basic Observations

We start with some basic observations about the joint distribution of the inputs and the transcript of a protocol for AND_k with independent, uniformly distributed inputs.

Definition 3.3. From now on, let P be a fixed randomized k-player protocol that computes AND_k with error at most ε and for $\mathbf{x} \in \{0,1\}^k$ let $M(\mathbf{x})$ denote the transcript of P for the input $\mathbf{x}$. Let $\mathbf{X} = (X_1, \ldots, X_k)$ be a random variable that is uniformly distributed in $\{0,1\}^k$ and let $T = M(\mathbf{X})$ denote the transcript of P for the the input $\mathbf{X}$.

Note that the transcript $M(\mathbf{x})$ does depend on $\mathbf{x}$ and the random inputs of the players. Thus even for a fixed input $\mathbf{x}$ the transcript is a random variable whose value depends on the random bits used in the protocol.

A randomized k-party protocol can be seen as a deterministic protocol in which the ith player has two inputs: The input to the randomized protocol, in our case X_i, and as a second input the random bits that are used by the ith player. Then the first observation is a restatement of the fact that the set of the inputs (real inputs and random bits) that correspond to a fixed transcript is a combinatorial rectangle (see [9] for a definition of combinatorial rectangles).

Observation 3.4 ([3, 4])**.** Let $\mathbf{x} = (x_1, \ldots, x_k) \in \{0,1\}^k$ and let t be an element from the support of T. Then $\Pr\{\mathbf{X}=\mathbf{x}|T=t\} = \prod_i \Pr\{X_i=x_i|T=t\}$.

We omit the simple combinatorial proof of this observation because this basic property of k-party protocols was already used in [3] and [4]. The following observation is an immediate, but very useful consequence of the previous one.

Observation 3.5. Let $\mathbf{x} = (x_1, \ldots, x_k) \in \{0,1\}^k$ and let t be an element from the support of T. Then $\Pr\{X_i=x_i|T=t, \mathbf{X}_{-i}=\mathbf{x}_{-i}\} = \Pr\{X_i=x_i|T=t\}$ for all $i \in \{1, \ldots, k\}$.

Proof. This observation follows immediately from Observation 3.4: By adding the equality from Observation 3.4 for $(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_k)$ and $(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_k)$ we obtain

$$\Pr\{\mathbf{X}_{-i}=\mathbf{x}_{-i}|T=t\} = \prod_{j\neq i} \Pr\{X_j=x_j|T=t\} \; .$$

Using this and Observation 3.4 verbatim yields

$$\begin{aligned}\Pr\{X_i=x_i|T=t, \mathbf{X}_{-i}=\mathbf{x}_{-i}\} &= \frac{\Pr\{X_i=x_i, \mathbf{X}_{-i}=\mathbf{x}_{-i}|T=t\}}{\Pr\{\mathbf{X}_{-i}=\mathbf{x}_{-i}|T=t\}} \\ &= \frac{\prod_j \Pr\{X_j=x_j|T=t\}}{\prod_{j\neq i} \Pr\{X_j=x_j|T=t\}} = \Pr\{X_i=x_i|T=t\} \; .\end{aligned}$$

■

The next observation relates the joint distribution of Z_i and $M(\mathbf{Z})$ given that $D = i$ to the joint distribution of X_i and $T = M(\mathbf{X})$ given that $\mathbf{X}_{-i} = \mathbf{0}$. Combined with the previous observations, this will be the basis for the proof of the main result.

Observation 3.6. Let $i \in \{1, \ldots, k\}$. Then $\mathrm{I}(M(\mathbf{Z}) : Z_i | D = i) = \mathrm{I}(T : X_i | \mathbf{X}_{-i} = \mathbf{0})$.

Proof. First observe that $\Pr\{\mathbf{Z} = \mathbf{v}, M(\mathbf{Z}) = t | D = i\} = \Pr\{\mathbf{X} = \mathbf{v}, T = t | \mathbf{X}_{-i} = \mathbf{0}\}$ for every $v \in \{0,1\}^k$ and every t in the support of $M(X)$ and $M(Z)$. This follows from the fact that the conditional distribution of $\mathbf{X}$ given that $\mathbf{X}_{-i} = \mathbf{0}$ is the same as the conditional distribution of $\mathbf{Z}$ given that $D = i$, the fact that the random inputs of P are independent of $\mathbf{X}$ and $\mathbf{Z}$, and the fact that the transcript is a function of the inputs and the random inputs. Then the claim of the lemma is an immediate consequence of the initial observation. ∎

3.2. Main Idea of the Proof

Like the approach of Bar-Yossef *et al.* [3], our approach is based on the observation that the distribution of the transcripts of a randomized protocol for AND_k with small error must at least be very different for the inputs $\mathbf{X} = \mathbf{0}$ and $\mathbf{X} = \mathbf{1}$. The difference is expressed using some appropriate metric on probability spaces. Then, by using Observations 3.4 and 3.5, this result is decomposed into results about the distributions of $(X_i, M(\mathbf{X}) | \mathbf{X}_{-i} = \mathbf{0})$ which are finally used to bound the conditional mutual information of $\mathbf{Z}$ and $M(\mathbf{Z})$ given D by using Observation 3.6. The result from [3] mainly uses the Hellinger distance (see [7]) to carry out this very rough outline of the proof. We will stick to the rough outline, but our result will use the Kullback Leibler distance instead of the Hellinger distance. Due to the limited space in the STACS-proceedings we can only present proof-sketches of the technical lemmas in this section. A version of this paper with full proofs can be found on the authors homepage [1].

We will first decompose the Kullback Leibler distance of the distributions $(T | \mathbf{X} = \mathbf{0})$ and $(T | \mathbf{X} = \mathbf{1})$ into results about the joint distributions of X_i and T for $i = 1, \ldots, k$. The result will be expressed in terms of the following function.

Definition 3.7. From now on, let $g(x) = x \log \frac{x}{1-x}$.

Note that the left hand side of the equation in the following lemma is the Kullback Leibler distance of $(T | \mathbf{X} = \mathbf{0})$ and $(T | \mathbf{X} = \mathbf{1})$ if S is the set of all possible transcripts.

Lemma 3.8. *Let S be a subset of the set of all possible transcripts. Then*

$$\sum_{t \in S} \Pr\{T = t | \mathbf{X} = \mathbf{0}\} \cdot \log \frac{\Pr\{T = t | \mathbf{X} = \mathbf{0}\}}{\Pr\{T = t | \mathbf{X} = \mathbf{1}\}} = 2 \sum_i \sum_{t \in S} \Pr\{T = t | \mathbf{X}_{-i} = \mathbf{0}\} \cdot g(\Pr\{X_i = 0 | T = t\}) \; .$$

Proof Sketch. The proof of this lemma is mainly based on the fact that

$$\frac{\Pr\{T = t | \mathbf{X} = \mathbf{0}\}}{\Pr\{T = t | \mathbf{X} = \mathbf{1}\}} = \frac{\Pr\{\mathbf{X} = \mathbf{0} | T = t\}}{\Pr\{\mathbf{X} = \mathbf{1} | T = t\}} \; .$$

Then Observation 3.4 can be applied to decompose the log-function into a sum. Finally, we use that $\Pr\{T = t | \mathbf{X} = \mathbf{0}\} = 2 \Pr\{T = t | \mathbf{X}_{-i} = \mathbf{0}\} \cdot \Pr\{X_i = 0 | T = t\}$ by Observation 3.5. ∎

[1] `http://ls2-www.cs.uni-dortmund.de/~gronemeier/`

Next, we will express a lower bound on $\mathrm{I}(M(\mathbf{Z}):Z|D)$ in terms of the following function f and set $B(\alpha)$.

Definition 3.9. From now on, let $f(x) = x\log 2x + \frac{1-x}{2}\log 2(1-x)$.

Definition 3.10. Let $B(\alpha)$ denotes the set of all transcripts t such that $\Pr\{X_i=0|T=t\} < \alpha$ for all $i \in \{1,\ldots,k\}$.

The role of the parameter α will become apparent later. The only property that is needed for the proof of the following lemma is that $\alpha > 1/2$.

Lemma 3.11. *Let $\alpha > \frac{1}{2}$ be a constant. Then*

$$\mathrm{I}(M(\mathbf{Z}):\mathbf{Z}|D) \geq \frac{1}{k}\sum_i \sum_{t\in B(\alpha)} \Pr\{T=t|\mathbf{X}_{-i}=\mathbf{0}\}\cdot f(\Pr\{X_i=0|T=t\}) \ .$$

Proof Sketch. This lemma can be proved by using that $f(x) = \frac{1}{2}(f_1(x)+f_2(x))$ where $f_1(x) = x\log 2x + (1-x)\log 2(1-x)$ and $f_2(x) = x\log 2x$. It is sufficient to prove that the lower bound holds for f_1 and f_2 instead of f. To this end one can show that

$$\mathrm{I}(M(\mathbf{Z}):\mathbf{Z}|D) = \frac{1}{k}\sum_i \sum_t \Pr\{T=t|\mathbf{X}_{-i}=\mathbf{0}\}\cdot f_1(\Pr\{X_i=0|T=t\}) \ .$$

Then the bound for f_1 is obvious since $f_1(x)$ is nonnegative for all $x \in [0,1]$. The bound for f_2 use the fact that $f_1(x) = f_2(x) + f_2(1-x)$, that $f_2(x) \geq 0$ for $x \in [1/2,1]$, and that

$$\sum_t \Pr\{T=t|\mathbf{X}_{-i}=\mathbf{0}\}\cdot f_2(\Pr\{X_i=1|T=t\})$$

is nonnegative. ∎

The right hand sides of the equation in Lemma 3.8 and the inequality in Lemma 3.11 look very similar. In fact, if there was a positive constant c such that $c\cdot f(x) \geq g(x)$ for all $x \in [0,1]$, then for a complete proof of Theorem 3.2 it would be sufficient to show that the Kullback Leibler distance of $(T|\mathbf{X}=\mathbf{0})$ and $(T|\mathbf{X}=\mathbf{1})$ is bounded from below by a constant $c(\varepsilon)$ if the error of the protocol P is bounded by ε. Unfortunately $f(x) \leq 1$ for $x \in [0,1]$ while $g(x)$ is not bounded from above for $x \in [0,1]$. So this naive first idea does not work. But the function $g(x)$ is bounded in every interval $[0,\beta]$ where $\beta < 1$. The following Lemma shows that we can easily bound $f(x)$ from below in terms of $g(x)$ if we restrict x to an appropriate interval $[0,\beta]$.

Lemma 3.12. *There is a constant $\beta > \frac{1}{2}$ such that $4\cdot f(x) \geq g(x)$ for all $x \in [0,\beta]$.*

This lemma can probably be proved in many ways. By inspection and numeric computations it is easy to verify that it holds for $\beta \approx 0.829$. Here it is more important to note that our choice of the function f is one of the crucial points of our proof: The function $g(x)$ is negative for $x \in [0,\frac{1}{2})$ and nonnegative and increasing for $x \in [\frac{1}{2},1]$. Furthermore $g(\frac{1}{2}) = 0$ and in the interval $[\frac{1}{2},1]$ the slope of $g(x)$ is bounded from below by a positive constant. It will become clear in Lemma 3.14 that we have to lower bound $f(x)$ in terms of $g(x)$ for $x \approx \frac{1}{2} + O(\frac{1}{k})$ where k is the number of players. Recall that $f(x) = \frac{1}{2}(f_1(x)+f_2(x))$ where $f_1(x) = x\log 2x + (1-x)\log 2(1-x)$ and $f_2(x) = x\log 2x$ and that we prove Lemma 3.11 by lower bounding the mutual information of $M(\mathbf{Z})$ and $\mathbf{Z}$ in terms of $f_1(x)$ and $f_2(x)$. Thus $f_1(x)$ and $f_2(x)$ would be natural candidates for the function $f(x)$. Unfortunately, neither $f_1(x)$ nor $f_2(x)$ alone does work in our proof. The function $f_1(x)$ is nonnegative for

$x \in [0,1]$, therefore $f_1(x) \geq g(x)$ for $x \in [0, \frac{1}{2}]$, but the slope of $f_1(x)$ is too small in the interval $[\frac{1}{2}, 1]$. It turns out that $f_1(\frac{1}{2} + \frac{1}{k}) \approx 1/k^2$. If we used the function $f_1(x)$ instead of $f(x)$ in our proof, we could only obtain an $\Omega(1/k^2)$ lower bound for the information complexity of AND_k. The function $f_2(x)$ does not suffer from this problem since the slope of $f_2(x)$ in $[\frac{1}{2}, 1]$ is bounded from below by a constant. But here we have the problem that $f_2(x)$ is too small for $x \in [0, \frac{1}{2})$. For every constant $c > 0$ such that $c \cdot f(x) \geq g(x)$ in the interval $x \in [\frac{1}{2}, 1]$ we have $g(x) > c \cdot f(x)$ in the interval $x \in [0, \frac{1}{2})$. Luckily, for the average $f(x)$ of $f_1(x)$ and $f_2(x)$ the good properties of the functions are preserved while the bad properties "cancel out". The bounded slope for $x \in [\frac{1}{2}, 1]$ of $f(x)$ is inherited from $f_2(x)$. The fact that $f(x)$ is not to small for $x \in [0, \frac{1}{2})$ is inherited from $f_1(x)$.

We can use the set $B(\alpha)$ in Lemma 3.11 and the set S in Lemma 3.8 to restrict t to the transcripts that satisfy $\Pr\{X_i = 0 | T = t\} \leq \beta$ for all $i \in \{1, \ldots, k\}$. Then, by our previous observations, it is easy to lower bound $f(\Pr\{X_i = 0 | T = t\})$ in terms of $g(\Pr\{X_i = 0 | T = t\})$.

Definition 3.13. Let β be the constant from Lemma 3.12. recall that $B(\alpha)$ denotes the the set of all transcripts t such that $\Pr\{X_i = 0 | T = t\} < \alpha$ for all $i \in \{1, \ldots, k\}$. Then B is a shorthand notation for the set $B(\beta)$.

Unfortunately, the restriction of t to the set $S = B$ complicates the proof of a lower bound for the left hand sum in Lemma 3.8 since we remove the largest terms from the sum. For example, we will see in the proof of Corollary 3.17 that for zero-error protocols the set B does only contain transcripts for the output 1. Therefore, by the zero-error property, $\Pr\{T \in B | \mathbf{X} = \mathbf{0}\} = 0$ for zero error protocols and the left hand sum in Lemma 3.8 is equal to 0. Consequently, without further assumptions that do not hold in general it is impossible to prove large lower bounds on the sum in Lemma 3.8 for the set $S = B$. However, the next Lemma shows that we can lower bound the sum, if we assume that $\Pr\{T \in B | \mathbf{X} = \mathbf{0}\}$ is sufficiently large.

Lemma 3.14. *Suppose that* $\Pr\{T \in B | \mathbf{X} = \mathbf{0}\} \geq \frac{3}{4}$ *and that the error* ε *of the protocol* P *is bounded by* $\varepsilon < \frac{3}{10}\left(1 - \sqrt{\frac{1}{2}\log\frac{4}{3}}\right)$. *Then*

$$\sum_{t \in B} \frac{\Pr\{T = t | \mathbf{X} = \mathbf{0}\}}{\Pr\{T \in B | \mathbf{X} = \mathbf{0}\}} \cdot \log \frac{\Pr\{T = t | \mathbf{X} = \mathbf{0}\}}{\Pr\{T = t | \mathbf{X} = \mathbf{1}\}} \geq \min\left\{\log\frac{3}{2}, 2\left(1 - \frac{10}{3}\varepsilon\right)^2 - \log\frac{4}{3}\right\} > 0 \,.$$

Proof Sketch. For the proof of this lemma we consider two cases: If $\Pr\{T \in B | \mathbf{X} = \mathbf{1}\} < \frac{1}{2}$ then we can use the log sum inequality (Lemma 2.1) to lower bound the sum on the left hand side. If $\Pr\{T \in B | \mathbf{X} = \mathbf{1}\} \geq \frac{1}{2}$ then the error of the protocol P under the condition that $T \in B$ must be small both for the input $\mathbf{X} = \mathbf{0}$ and the input $\mathbf{X} = \mathbf{1}$. With this assumption we can lower bound the left hand side using Lemma 2.2 since in this case the total variation distance of $(T | \mathbf{X} = \mathbf{0}, T \in B)$ and $(T | \mathbf{X} = \mathbf{1}, T \in B)$ is large. ∎

Note that, by Lemma 3.8 and the fact that the slope of $g(x)$ is bounded from below by a positive constant for $x \in [1/2, 1]$, this lower bound can be met if $\Pr\{X_i = 0 | T = t\} = \frac{1}{2} + \Theta(\frac{1}{k})$ for all $i \in \{1, \ldots, k\}$ and every $t \in B$.

By Lemma 3.14, under the condition that $\Pr\{T \in B | \mathbf{X} = \mathbf{0}\} \geq \frac{3}{4}$ our initial naive plan of bounding f in terms of g does work. The details of this idea are elaborated on in the proof of Theorem 3.16. Next, we look at the case that $\Pr\{T \in B | \mathbf{X} = \mathbf{0}\}$ is small. It turns out that this assumption alone already leads to a large lower bound on $\mathrm{I}(M(\mathbf{Z}) : \mathbf{Z} | D)$.

Lemma 3.15. *Let α be a constant subject to $1/2 < \alpha \leq 1$. Then*

$$\mathrm{I}(M(\mathbf{Z}):\mathbf{Z}|D) \geq \frac{1}{2k} \cdot \Pr\{T \notin B(\alpha)|\mathbf{X}=\mathbf{0}\} \cdot (1 - h_2(\alpha)).$$

Proof Sketch. The proof of this lemma is based on the fact that, by the definition of $B(\alpha)$, under the condition that $T = t \notin B(\alpha)$ the entropy of X_i is bounded by $h_2(\alpha) < 1$ for at least one i. ∎

Now all prerequisites for a full proof of Theorem 3.2 are in place. It is implied by the following theorem because P was assumed to be an arbitrary ε-error protocol for AND_k.

Theorem 3.16. *Let $\varepsilon < \frac{3}{10}\left(1 - \sqrt{\frac{1}{2}\log\frac{4}{3}}\right)$ be a constant. If the error of the protocol P is bounded by ε, then there is a constant $c(\varepsilon) > 0$ that does only depend on ε such that*

$$\mathrm{I}(M(\mathbf{Z}):\mathbf{Z}|D) \geq \frac{c(\varepsilon)}{k}.$$

Proof. Recall that B is the set of all transcripts t such that $\Pr\{X_i=0|T=t\} < \beta$ for all $i \in \{1,\ldots,k\}$, where β is the constant from Lemma 3.12. For the proof of the lemma we will consider two cases.

For the first case, assume that $\Pr\{T \in B|\mathbf{X}=\mathbf{0}\} \leq \frac{3}{4}$. In this case we can apply Lemma 3.15 with $\alpha = \beta$ and we get

$$\mathrm{I}(M(\mathbf{Z}):\mathbf{Z}|D) \geq \frac{1}{2k}\Pr\{T \notin B|\mathbf{X}=\mathbf{0}\}(1 - h_2(\beta)) \geq \frac{1}{8k}(1 - h_2(\beta)) \ .$$

Note that in this case the lower bound does not depend on ε and that, since $\beta > 1/2$, there is a constant $c_1 > 0$ such that the right hand side of the last inequality is bounded from below by c_1/k.

For the second case, assume that $\Pr\{T \in B|\mathbf{X}=\mathbf{0}\} > \frac{3}{4}$. In this case we first apply Lemma 3.11 for $\alpha = \beta$, thus $B(\alpha) = B$, then Lemma 3.12, and finally Lemma 3.8 for the subset $S = B$ to get

$$\begin{aligned}
\mathrm{I}(M(\mathbf{Z}):\mathbf{Z}|D) &\geq \frac{1}{k}\sum_i \sum_{t\in B} \Pr\{T=t|\mathbf{X}_{-i}=\mathbf{0}\} \cdot f(\Pr\{X_i=0|T=t\}) \\
&\geq \frac{1}{4k}\sum_i \sum_{t\in B} \Pr\{T=t|\mathbf{X}_{-i}=\mathbf{0}\} \cdot g(\Pr\{X_i=0|T=t\}) \\
&= \frac{1}{8k}\sum_{t\in B} \Pr\{T=t|\mathbf{X}=\mathbf{0}\} \cdot \log\frac{\Pr\{T=t|\mathbf{X}=\mathbf{0}\}}{\Pr\{T=t|\mathbf{X}=\mathbf{1}\}} \ .
\end{aligned}$$

Then, by the assumption $\Pr\{T \in B|\mathbf{X}=\mathbf{0}\} > \frac{3}{4}$, we can apply Lemma 3.14 to obtain

$$\begin{aligned}
\mathrm{I}(M(\mathbf{Z}):\mathbf{Z}|D) &\geq \frac{1}{8k} \cdot \Pr\{T \in B|\mathbf{X}=\mathbf{0}\} \cdot \min\left\{\log\frac{3}{2}, 2\left(1 - \frac{10}{3}\varepsilon\right)^2 - \log\frac{4}{3}\right\} \\
&\geq \frac{3}{32k} \cdot \min\left\{\log\frac{3}{2}, 2\left(1 - \frac{10}{3}\varepsilon\right)^2 - \log\frac{4}{3}\right\} \ .
\end{aligned}$$

For $\varepsilon < \frac{3}{10}\left(1 - \sqrt{\frac{1}{2}\log\frac{4}{3}}\right)$ the minimum in the last inequality is a positive constant that does only depend on the constant ε. Hence, there is a constant $c_2(\varepsilon) > 0$ that does only

depend on the constant ε such that the right hand side is bounded from below by $c_2(\varepsilon)/k$. The claim of the Lemma follows from the two cases if we choose $c(\varepsilon) = \min\{c_1, c_2(\varepsilon)\}$. ∎

3.3. A Simple Lower Bound for Zero-Error Protocols

For zero-error protocols a lower bound can be proved by using only Lemma 3.15.

Corollary 3.17. *For every randomized k-player zero-error protocol with input $\mathbf{Z}$ and transcript $M(\mathbf{Z})$ the conditional information cost satisfies* $\mathrm{I}(M(\mathbf{Z}):\mathbf{Z}|D) \geq 1/(2k)$.

Proof. Consider the transcript T of the protocol P for the input $\mathbf{X}$. Then the corollary follows immediately from Lemma 3.15 if we set $\alpha = 1$: Recall that the output of the protocol can be inferred from the transcript and let $P(t)$ denote the output of the protocol P for transcript t. Suppose that $P(t) = 0$. Then $\Pr\{X_i = 0|T = t\} = 1$ for at least one i since otherwise, by Observation 3.4, $\Pr\{\mathbf{X} = \mathbf{1}|T = t\} > 0$ and under the condition $T = t$ the output of P would be wrong with a nonzero probability. Clearly this is not possible for zero-error protocols, hence $\Pr\{T \notin B(1)|P(T) = 0\} = 1$. Under the condition $\mathbf{X} = \mathbf{0}$ the output of P is 0 with probability 1, again by the zero-error property, therefore the last observation implies that $\Pr\{T \notin B(1)|\mathbf{X} = \mathbf{0}\} = 1$ and obviously $1 - h_2(1) = 1$. ∎

Acknowledgments

Thanks to Martin Sauerhoff for helpful discussions and proofreading.

References

[1] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999.

[2] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *Proc. of 17th CCC*, pages 93–102, 2002.

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

[4] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *Proc. of 18th CCC*, pages 107–117, 2003.

[5] A. Chakrabarti, Y. Shi, A. Wirth, and A. C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. of 42nd FOCS*, pages 270–278, 2001.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[7] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70:419, 2002.

[8] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory*, 4:126–127, 1967.

[9] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[10] A. C. Yao. Some complexity questions related to distributive computing (preliminary report). In *Proc. of 11th STOC*, pages 209–213, 1979.