Electronic Colloquium on Computational Complexity, Report No. 151 (2016)

Pointer chasing via triangular discrimination

Amir Yehudayoff*

Abstract

We prove an essentially sharp $\tilde{\Omega}(n/k)$ lower bound on the k-round distributional complexity of the k-step pointer chasing problem under the uniform distribution, when Bob speaks first. This is an improvement over Nisan and Wigderson's $\tilde{\Omega}(n/k^2)$ lower bound. A key part of the proof is using triangular discrimination instead of total variation distance; this idea may be useful elsewhere.

1 Introduction

Pointer chasing is a natural and well-known problem that captures the importance of interaction. In its two-player bit version, Alice gets as input a map $f_A : A \to B$ and Bob gets as input $f_B : B \to A$, where $A = \{1, 2, \ldots, n\}$ and $B = \{n+1, n+2, \ldots, 2n\}$. The pointers $z_0, z_1, \ldots$ are defined inductively as

$$z_0 = 1,\ z_1 = f_A(z_0),\ z_2 = f_B(z_1),\ z_3 = f_A(z_2),\ z_4 = f_B(z_3),\ \ldots \tag{1}$$

The k-step pointer chasing function PC_k is defined as[1]

$$PC_k(f_A, f_B) = z_k \mod 2.$$

This problem was suggested by Papadimitriou and Sipser to study the number of rounds and the order in which the players talk in communication protocols [15]. Its communication complexity was consequently studied in many works (e.g. [7, 14, 6, 10, 16]).

*Department of mathematics, Technion-IIT. amir.yehudayoff@gmail.com. Research supported by ISF.

[1] One may replace parity with some other balanced boolean function.

ISSN 1433-8092

Pointer chasing is also known to be related to other models and questions. Nisan and Wigderson showed that it is a "complete" problem for monotone constant-depth boolean circuits [14], and that it can be used to prove the monotone constant-depth hierarchy that was proved by Klawe, Paul, Pippenger and Yannakakis [11]. It was further used for proving lower bounds on the time complexity of distributed computation [13], and for proving lower bounds on the space complexity of streaming algorithms [8].

This work studies the communication complexity of the pointer chasing problem. We start with a survey of known results, and then state our result and discuss its proof.

Communication complexity

Upper bounds. There is an obvious k-round deterministic protocol for computing PC_k with communication $O(k \log n)$ in which Alice speaks first. Nisan and Wigderson [14] described a randomized $(k-1)$-round protocol for PC_k with communication $O((k+(n/k)) \log n)$. Damm, Jukna and Sgall [6] described a k-round deterministic protocol with communication at most $O(n \log^{(k-1)} n)$ for PC_k when Bob speaks first (see [16]).

Lower bounds. Papadimitriou and Sipser [15] conjectured that $(k-1)$-round protocols for PC_k must use $\Omega(n)$ bits of communication for constant k, and proved it for $k = 2$. Duris, Galil and Schnitger [7] showed that this conjecture is true; they proved that the $(k-1)$-round deterministic communication complexity of PC_k is at least $\Omega(n/k^2)$. Later on, Nisan and Wigderson [14] improved this deterministic lower bound to $\Omega(n - k \log n)$, and also proved an $\Omega((n/k^2) - k \log n)$ lower bound on its k-round randomized communication complexity when Bob speaks first. Ponzio, Radhakrishnan and Venkatesh [16] proved that the protocol from [6] is tight; they proved an $\Omega(n \log^{(k-1)} n)$ on the $(k-1)$-round randomized communication complexity of PC_k for constant k. Finally, Klauck [10] observed that the proof of the deterministic lower bound from [14] actually implies an essentially sharp $\Omega(k+(n/k))$ lower bound on the k-round randomized communication complexity of PC_k when Bob speaks first.

This work. Here we focus on the distributional complexity of the pointer chasing problem, under the uniform distribution (i.e. f_A, f_B are chosen independently and uniformly at random); the uniform distribution seems to be the most natural distribution on inputs. Previously, the only known lower bound on the k-round distributional complexity of PC_k under the uniform distribution when Bob speaks first was

Nisan and Wigderson's $\Omega((n/k^2) - k \log n)$ lower bound. Klauck's observation in [10] together with von Neumann minimax theorem (Yao's principle) show that there is some distribution for which an $\Omega(k + (n/k))$ lower bound holds. This distribution is, however, not explicit, and e.g. prior to this work the best lower bound that was known for any product distribution was Nisan and Wigderson's.

The main result of this work is a tight (up to poly $\log n$ factors) lower bound on the distributional complexity over the uniform distribution.

Theorem 1.1. *The length of every k-round protocol in which Bob speaks first that computes the k-step pointer chasing function with average-case error at most* 1/3 *under the uniform distribution is at least* $\frac{n}{1000k} - k \log n$.

Theorem 1.1 is proved in Section 4. In a nutshell, the idea is to keep track – round by round – of the amount of information revealed by the protocol (the proof in [14] can be stated in such a way as well). The goal is to prove that if the protocol is short then after the protocol terminates the inputs are still pretty random, which is impossible when the protocol achieves its goal.

The proof uses a measure of distance between distributions that is new in this context: the triangular discrimination. Roughly speaking, triangular discrimination replaces total variation distance in a way that allows to avoid the square-root loss that Pinsker's inequality yields.

This square-root loss appears in many works, and is directly related to fundamental questions. For example, it appears in the parallel repetition theorem, and is connected to the "strong parallel repetition" conjecture which is motivated by Khot's unique games conjecture [9]. The "strong parallel repetition" conjecture was falsified by Raz [17]; showing that this square-root loss is necessary for parallel repetition. This loss also appears in direct sums and products in communication complexity [1, 3], where it is related to the question of optimal compression of protocols. It is still unclear if this square-root loss is necessary in this case. Finally, this loss appears in Nisan and Wigderson's aforementioned lower bound [14]. The argument here shows that this loss in [14] is not necessary. This argument may yield better quantitative bounds in other cases as well. For this reason, in Section 3, we provide a clean example that demonstrates the main new technical idea.

Triangular discrimination

Measures of distance between probability distributions are extremely useful tools in many areas of research. A specific family of such measures is f-divergences (also known as Csiszár-Morimoto or Ali-Silvey divergences). These are measures of the form

$$D_f(p||q) = \sum_{\omega \in \Omega} q(\omega) f\left(\frac{p(\omega)}{q(\omega)}\right)$$

for a convex function f so that $f(1) = 0$ (where some conventions like $0f(0/0) = 0$ are used). For more background, see [5] and references within.

Some well-known examples are the ℓ_1 distance $|p-q|_1 = D_{f_1}(p||q)$ where $f_1(\xi) = |1-\xi|$, the Kullback-Leibler divergence $D(p||q) = D_{f_{KL}}(p||q)$ where $f_{KL}(\xi) = \xi \log \xi$, and the Jensen-Shannon divergence $JS(p||q) = D(p||(p+q)/2) + D(q||(p+q)/2)$.

Each of these measures has unique properties, which make it useful in different contexts. For example, ℓ_1 is useful due to its statistical meaning, and the Kullback-Leibler divergence is useful due to its tight relation to information theory (and properties like the chain rule).

Here we use the triangular discrimination [18] defined as $\Delta(p,q) = D_{f_\Delta}(p||q)$ with $f_\Delta(\xi) = \frac{(\xi-1)^2}{\xi+1}$. Stated differently:

$$\Delta(p,q) = \sum_{\omega \in \Omega} \frac{(p(\omega) - q(\omega))^2}{p(\omega) + q(\omega)},$$

where by convention $0/0 = 0$.

Since Δ is not so well-known in this context, we start with listing some of its properties (for more details see [18, 5]). As all f-divergences, it is non-negative, it is convex in (p,q), it satisfies a data processing inequality (also known as a lumping property), and more. It is also equivalent to the Jensen-Shannon divergence:

$$\Delta/2 \leq JS \leq 2\Delta.$$

It is, however, sometimes easier to work with than JS since its formula is simpler. It satisfies the following "improvement" over Pinsker's inequality (which states that $|p-q|_1^2 \leq 2D(p||q)$).

Lemma 1.2 ([18])**.** *$|p-q|_1^2/2 \leq \Delta(p,q) \leq 2D(p||q)$.*

Another interesting ("operational") interpretation of Δ, which is implicit in [2], is that "Δ is to ℓ_2 what ℓ_1 is to ℓ_∞" in the following sense: It is well-known that

$$|p-q|_1 = \max\left\{\frac{p.g - q.g}{\|g\|_\infty} \ : \ g \in \mathbb{R}^\Omega\right\},$$

where $p.g = \sum_{\omega \in \Omega} p(\omega) g(\omega)$. This property of ℓ_1 is related to that ℓ_1 is equivalent to total variation distance. For Δ we have that

Lemma 1.3. $\Delta(p,q) = \max\left\{\frac{(p.g-q.g)^2}{p.g^2+q.g^2} \ : \ g \in \mathbb{R}^\Omega\right\}$.

Proof. If $g(\omega) = \frac{p(\omega)-q(\omega)}{p(\omega)+q(\omega)}$ then

$$\Delta(p,q) = p.g - q.g = p.g^2 + q.g^2$$

and so

$$\Delta(p,q) \le \max\left\{\frac{(p.g-q.g)^2}{p.g^2+q.g^2} \ : \ g \in \mathbb{R}^\Omega\right\}.$$

On the other hand, for every g, by Cauchy-Schwartz,

$$p.g - q.g = \sum_\omega \frac{p(\omega)-q(\omega)}{\sqrt{p(\omega)+q(\omega)}}\sqrt{p(\omega)+q(\omega)}g(\omega) \le \sqrt{\Delta(p,q)}\sqrt{p.g^2+q.g^2}.$$

□

2 Preliminaries

Probability. We consider only random variables with finite support. We denote random variables by capital letters $(X, Y, \ldots)$ and the values they attain by small letters $(x, y, \ldots)$. We denote by $p_{X|y}$ the probability distribution of X conditioned on $Y = y$. We denote by $\mathbb{E}_X f(x)$ the expectation of $f(X)$, and by $\mathbb{E}_{X|y}$ the expectation of $f(X)$ conditioned on $Y = y$.

Communication complexity. We use standard communication complexity terminology. For formal definitions see e.g. the textbook [12].

KL-divergence. We state two lemmas that will be useful later on[2] (see e.g. the textbook [4]).

Lemma 2.1 (Subadditivity)**.** *If X, Y are random variables taking values in S^n for some finite set S, and the n coordinates of Y are independent, then*

$$D(p_X||p_Y) \ge \sum_{i\in[n]} D(p_{X_i}||p_{Y_i}).$$

Lemma 2.2 (Information is at most bit length)**.** *If X, Y are jointly distributed, and Y takes values in a set of size at most 2^h, then*

$$\mathop{\mathbb{E}}_Y D(p_{X|y}||p_X) \le h.$$

[2]The lemmas can be stated in terms of mutual information, but since it seems more natural to use KL-divergence in this text, we state them in this form.

3 An example

Before proving the lower bound for pointer chasing, we describe a cleaner example that demonstrates how can one use Δ instead of ℓ_1 to get quantitively better bounds.

Let X be a random vector in $\{0,1\}^n$. Assume that it has high entropy:

$$D(p_X||u_n) \leq k,$$

where u_n is the uniform distribution on $\{0,1\}^n$. Also assume that I is chosen uniformly in $[n]$ and independently of X. Lemma 2.1 implies that

$$\mathop{\mathbb{E}}_I D(p_{X_i}||u_1) \leq \frac{1}{n} D(p_X||u_n) \leq \frac{k}{n}. \tag{2}$$

That is, on average, the marginal distribution of X_I is close to uniform in KL-divergence, when $k \ll n$. Pinsker's inequality allows to deduce that the distribution of X_I is close to uniform in ℓ_1 distance as well.

It is natural to ask what happens when I is not uniform but only close to uniform. Let J be a random element of $[n]$, chosen independently of X, I, with very high entropy:

$$D(p_J||p_I) \leq \epsilon.$$

Pinsker's inequality implies that $|p_J - p_I|_1 \leq \sqrt{2\epsilon}$, which in turn allows to prove that

$$\mathop{\mathbb{E}}_J |p_{X_j} - u_1|_1 \leq |p_J - p_I|_1 + \mathop{\mathbb{E}}_I |p_{X_i} - u_1|_1 \leq \sqrt{2\epsilon} + \sqrt{2k/n}.$$

This square-root dependence is often too expensive to pay, especially when we apply such an argument several times. Triangular discrimination allows to remove this square-root dependence.

Theorem 3.1. *$\mathbb{E}_J \Delta(p_{X_j}, u_1) \leq 4\epsilon + 10k/n$.*

For the rest of this section, we prove Theorem 3.1. We start with the following simple claim.

Claim 3.2. *If $|\xi| \leq \sqrt{a(b+\xi)}$ with $a, b \geq 0$ then $\xi \leq a + 2b$.*

Proof. Assume without loss of generality that $a > 0$. If $\xi^2 - a\xi - ab \leq 0$ then

$$\xi \leq \frac{a + \sqrt{a^2 + 4ab}}{2} = \frac{a}{2}\left(1 + \sqrt{1 + 4b/a}\right) \leq \frac{a}{2}(1 + 1 + 4b/a).$$

□

For $s \in [n]$, let $g(s) = \Delta(p_{X_s}, u_1)$. Write

$$\mathop{\mathbb{E}}_{J} \Delta(p_{X_j}, u_1) = p_J.g = p_I.g + (p_J.g - p_I.g).$$

Lemma 1.2 and (2) allow to bound the left term:

$$p_I.g \leq 2 \mathop{\mathbb{E}}_{I} D(p_{X_i}||u_1) \leq \frac{2k}{n}. \tag{3}$$

It thus remains to upper bound

$$\xi = p_J.g - p_I.g.$$

This is done as follows:

$$\begin{aligned}
|\xi| &\leq \sqrt{\sum_s \frac{(p_J(s) - p_I(s))^2}{p_J(s) + p_I(s)} g(s)} \sqrt{\sum_s (p_J(s) + p_I(s)) g(s)} && \text{(Cauchy-Schwartz)} \\
&\leq \sqrt{2 \sum_s \frac{(p_J(s) - p_I(s))^2}{p_J(s) + p_I(s)}} \sqrt{\sum_s (p_J(s) + p_I(s)) g(s)} && (\Delta \leq 2) \\
&= \sqrt{2\Delta(p_J, p_I)} \sqrt{\xi + 2p_I.g}.
\end{aligned}$$

Use Claim 3.2, together with (3) and

$$\Delta(p_J, p_I) \leq 2D(p_J||p_I) \leq 2\epsilon,$$

to deduce that

$$\xi \leq 4\epsilon + 8k/n.$$

Together with (3), the theorem is proved.

4 The lower bound

We now prove the lower bound for pointer chasing, Theorem 1.1. We shall use the following variant of Δ:

$$\Lambda(p, q) = \sum_{\omega : p(\omega) \geq q(\omega)} \frac{(p(\omega) - q(\omega))^2}{p(\omega) + q(\omega)} \leq \Delta(p, q).$$

Note that Δ is symmetric but Λ is not. One important property of Λ is that

$$\frac{|p-q|_1}{2} = \sum_{\omega : p(\omega) \geq q(\omega)} \frac{p(\omega)-q(\omega)}{\sqrt{p(\omega)+q(\omega)}} \sqrt{p(\omega)+q(\omega)}$$
$$\leq \sqrt{2\Lambda(p,q)}. \qquad \text{(Cauchy-Schwartz)}$$

Another important property of Λ is that

$$\Lambda(p,q) \leq \frac{|p-q|_1}{2} \leq 1$$

(Δ can take the value two).

Proof of Theorem 1.1. Denote by ℓ the length of the protocol (which we assume to be deterministic). Denote by $M_1, \ldots, M_t$ the messages sent in the first t rounds of the protocol. Recall that $Z_0, Z_1, \ldots$ are defined in (1).

We shall show that if ℓ is small then Z_k is close to being uniform, even conditioned on the transcript of the protocol. This implies that ℓ must be large, if the protocol achieves it goal.

We prove, by induction on $t = 0, 1, \ldots, k$, that the following holds. Let R_t denote the random variable

$$R_t = (M_1, \ldots, M_t, Z_1, \ldots, Z_{t-1})$$

(where R_0 is empty and $R_1 = M_1$). Roughly speaking, the expression $\mathbb{E}_{R_t} \Lambda(p_{Z_t|r_t}, p_{Z_t})$ measures how much did we learn on Z_t from the first t rounds of the protocol. We shall prove that

$$\mathop{\mathbb{E}}_{R_t} \Lambda(p_{Z_t|r_t}, p_{Z_t}) \leq 6t\delta, \tag{4}$$

where

$$\delta = 2\frac{\ell + k\log n}{n}.$$

Before proving (4), we explain why it completes the proof. Since the fraction of even numbers in $[n]$ is at least $\frac{1}{2} - \frac{1}{n}$, the error of the protocol conditioned on $R_k = r_k$ is at least

$$err_{r_k} \geq \frac{1}{2} - \frac{1}{n} - \frac{|p_{Z_k|r_{k-1}} - p_{Z_k}|_1}{2}.$$

Hence, since the protocol has error 1/3,

$$\begin{aligned}
\left(\frac{1}{9} - \frac{2}{3n}\right)^2 &\leq \left(\mathop{\mathbb{E}}_{R_k} \frac{|p_{Z_k|r_k} - p_{Z_k}|_1}{3}\right)^2 \\
&\leq \mathop{\mathbb{E}}_{R_k} \frac{|p_{Z_k|r_k} - p_{Z_k}|_1^2}{8} && \text{(by convexity)} \\
&\leq \mathop{\mathbb{E}}_{R_k} \Lambda(p_{Z_k|r_k}, p_{Z_k}) && (\ell_1^2 \leq 8\Lambda) \\
&\leq 12k \frac{\ell + k \log n}{n}.
\end{aligned}$$

The lower bound on ℓ thus follows (we may assume $n \geq 1000$).

It thus remains to prove (4). When $t = 0$ it indeed holds (R_0 is empty). Suppose $t \geq 1$. There are two cases to consider, depending on the parity of t. We consider the case when t is odd, and Bob sends the message M_t. When t is even, the argument is similar due to the symmetry between Alice and Bob.

By induction, we have

$$\mathop{\mathbb{E}}_{R_{t-1}} \Lambda(p_{Z_{t-1}|r_{t-1}}, p_{Z_{t-1}}) \leq 6(t-1)\delta. \tag{5}$$

We want to bound $\mathbb{E}_{R_t} \Lambda(p_{Z_t|r_t}, p_{Z_t})$ from above. We start by simplifying it.

The following two independence properties are crucial: Denote by X the vector that represents Alice's input ($X_s = f_A(s)$ for each s), and denote by Y the vector that represents Bob's input ($Y_s = f_B(n+s)$ for each s).

(A) Conditioned on $(R_{t-1}, Z_{t-1}) = (r_{t-1}, z_{t-1})$, we know that $Z_t = X_{z_{t-1}}$ is independent of Y, and therefore also of M_t which is a function of $(Y, m_1, \ldots, m_{t-1})$.

(B) Conditioned on $R_{t-1} = r_{t-1}$, we know that X and Z_{t-1} are independent (when $t = 1$ we have $Z_{t-1} = 1$ and when $t > 1$ we have $Z_{t-1} = Y_{z_{t-2}}$).

These properties hold since (i) the distribution of (X, Y) conditioned on the values of $Z_0, Z_1, \ldots, Z_{t'}$ is a product distribution, (ii) conditioning on the value of $M_1, \ldots, M_t$ means focusing on some rectangle (i.e. a product set) in the input space, and (iii) the conditional distribution of a product distribution on a rectangle is again a product distribution.

We are therefore interested in

$$\begin{aligned}
\mathop{\mathbb{E}}_{R_t} \Lambda(p_{Z_t|r_t}, p_{Z_t}) &= \mathop{\mathbb{E}}_{R_{t-1},Z_{t-1},M_t} \Lambda(p_{Z_t|r_{t-1},z_{t-1}}, p_{Z_t}) && \text{(A)} \\
&= \mathop{\mathbb{E}}_{R_{t-1},Z_{t-1}} \Lambda(p_{X_{z_{t-1}}|r_{t-1},z_{t-1}}, p_{Z_t}) \\
&= \mathop{\mathbb{E}}_{R_{t-1},Z_{t-1}} \Lambda(p_{X_{z_{t-1}}|r_{t-1}}, p_{Z_t}) && \text{(B)} \\
&= \mathop{\mathbb{E}}_{R_{t-1}} \mathop{\mathbb{E}}_{Z_{t-1}|r_{t-1}} \Lambda_{z_{t-1}},
\end{aligned} \tag{6}$$

where $\Lambda_s = \Lambda_s(r_{t-1})$ is

$$\Lambda_s = \Lambda(p_{X_s|r_{t-1}}, p_{Z_t}).$$

Intuitively, by induction we know that $p_{Z_{t-1}|r_{t-1}}$ is close to uniform, so we start by checking what happens if we replace $Z_{t-1}|r_{t-1}$ by a truly uniform variable. Let I be chosen uniformly at random in $[n]$, and independently of all other choices. Since the coordinates in X are uniform and independent, and p_{Z_t} is uniform,

$$\begin{aligned}
\mathop{\mathbb{E}}_{R_{t-1}} \mathop{\mathbb{E}}_{I} \Lambda_i &\le \mathop{\mathbb{E}}_{R_{t-1}} \mathop{\mathbb{E}}_{I} \Delta(p_{X_i|r_{t-1}}, p_{X_i}) && (\Lambda \le \Delta) \\
&\le 2 \mathop{\mathbb{E}}_{R_{t-1}} \mathop{\mathbb{E}}_{I} D(p_{X_i|r_{t-1}} || p_{X_i}) && \text{(Lemma 1.2)} \\
&\le \tfrac{2}{n} \mathop{\mathbb{E}}_{R_{t-1}} D(p_{X|r_{t-1}} || p_X) && \text{(Lemma 2.1)} \\
&\le \delta. && \text{(Lemma 2.2)}
\end{aligned} \tag{7}$$

Now, consider the difference

$$\mathop{\mathbb{E}}_{R_{t-1}} \left[\mathop{\mathbb{E}}_{Z_{t-1}|r_{t-1}} \Lambda_{z_{t-1}} \right] - \mathop{\mathbb{E}}_{R_{t-1}} \left[\mathop{\mathbb{E}}_{I} \Lambda_i \right] = \mathop{\mathbb{E}}_{R_{t-1}} \left[\mathop{\mathbb{E}}_{Z_{t-1}|r_{t-1}} [\Lambda_{z_{t-1}}] - \mathop{\mathbb{E}}_{I} [\Lambda_i] \right].$$

Start by fixing r_{t-1}. Let $q = p_{Z_{t-1}|r_{t-1}}$. The difference inside the expectation on the right hand side above equals

$$\xi = \xi(r_{t-1}) = \sum_s (q(s) - p_I(s)) \Lambda_s.$$

Bound it from above as follows

$$\begin{aligned}
|\xi| &= \left| \sum_s \frac{q(s) - p_I(s)}{\sqrt{q(s) + p_I(s)}} \sqrt{\Lambda_s} \cdot \sqrt{(q(s) + p_I(s)) \Lambda_s} \right| \\
&\le \sqrt{\sum_s \frac{(q(s) - p_I(s))^2}{q(s) + p_I(s)} \Lambda_s} \sqrt{\sum_s (q(s) + p_I(s)) \Lambda_s} && \text{(Cauchy-Schwartz)} \\
&\le \sqrt{\Lambda(q, p_I) + \sum_{s: q(s) < p_I(s)} \frac{(q(s) - p_I(s))^2}{q(s) + p_I(s)} \Lambda_s} \sqrt{\xi + 2 \mathop{\mathbb{E}}_{I} \Lambda_i}. && (\Lambda \le 1)
\end{aligned}$$

Since

$$\sum_{s:q(s)<p_I(s)} \frac{(q(s)-p_I(s))^2}{q(s)+p_I(s)}\Lambda_s \le \sum_s \frac{(p_I(s))^2}{p_I(s)}\Lambda_s = \mathop{\mathbb{E}}_I \Lambda_i,$$

by Claim 3.2 we have

$$\xi \le \Lambda(q,p_I) + 5\mathop{\mathbb{E}}_I \Lambda_i.$$

Now, taking expectation over R_{t-1}, by (5) and (7), since $p_I = p_{Z_{t-1}}$,

$$\mathop{\mathbb{E}}_{R_{t-1}}\left[\mathop{\mathbb{E}}_{Z_{t-1}|r_{t-1}}[\Lambda_{z_{t-1}}] - \mathop{\mathbb{E}}_I[\Lambda_i]\right] \le \mathop{\mathbb{E}}_{R_{t-1}}\left[\Lambda(p_{Z_{t-1}|r_{t-1}}, p_I) + 5\mathop{\mathbb{E}}_I \Lambda_i\right] \le 6(t-1)\delta + 5\delta.$$

Finally, by (6) and (7), the inductive claim is proved. □

Acknowledgments

I wish to thank Anup Rao and Avi Wigderson for helpful comments.

References

[1] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[2] I. Benjamini, H. Duminil-Copin, G. Kozma, and A. Yadin. Disorder, entropy and harmonic functions. *The Annals of Probability*, 43(5):2332–2373, 2015.

[3] M. Braverman, A. Rao, O. Weinstein, and A. Yehudayoff. Direct products in communication complexity. In *FOCS*, pages 746–755. IEEE, 2013.

[4] T. M. Cover and J. A. Thomas. *Elements of information theory.* John Wiley & Sons, 2012.

[5] I. Csiszár and P. C. Shields. *Information theory and statistics: A tutorial.* Now Publishers Inc, 2004.

[6] C. Damm, S. Jukna, and J. Sgall. Some bounds on multiparty communication complexity of pointer jumping. *Computational complexity*, 7(2):109–127, 1998.

[7] P. Duris, Z. Galil, and G. Schnitger. Lower bounds on communication complexity. In *STOC*, pages 81–91, 1984.

[8] V. Guruswami and K. Onak. Superlinear lower bounds for multipass graph processing. In *CCC*, pages 287–298, 2013.

[9] S. Khot. On the power of unique 2-prover 1-round games. In *STOC*, pages 767–775, 2002.

[10] H. Klauck. On quantum and probabilistic communication: Las vegas and one-way protocols. In *STOC*, pages 644–651, 2000.

[11] M. Klawe, W. J. Paul, N. Pippenger, and M. Yannakakis. On monotone formulae with restricted depth. In *STOC*, pages 480–487, 1984.

[12] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 2006.

[13] D. Nanongkai, A. Das Sarma, and G. Pandurangan. A tight unconditional lower bound on distributed randomwalk computation. In *PODC*, pages 257–266, 2011.

[14] N. Nisan and A. Widgerson. Rounds in communication complexity revisited. In *STOC*, pages 419–429, 1991.

[15] C. H. Papadimitriou and M. Sipser. Communication complexity. *Journal of Computer and System Sciences*, 28(2):260–269, 1984.

[16] S. J. Ponzio, J. Radhakrishnan, and S. Venkatesh. The communication complexity of pointer chasing. *Journal of Computer and System Sciences*, 62(2):323–355, 2001.

[17] R. Raz. A counterexample to strong parallel repetition. *SIAM Journal on Computing*, 40(3):771–777, 2011.

[18] F. Topsøe. Some inequalities for information divergence and related measures of discrimination. *IEEE Transactions on Information Theory*, 46(4):1602–1609, 2000.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de