Combinatorics, Probability and Computing (2020), **29**, pp. 485–494
doi:10.1017/S0963548320000085

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Pointer chasing via triangular discrimination

Amir Yehudayoff†

Department of Mathematics, Technion – Israel Institute of Technology, Haifa 32000, Israel
Email: amir.yehudayoff@gmail.com

(Received 5 February 2018; revised 16 December 2019; first published online 15 May 2020)

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, and essentially matches the randomized lower bound proved by Klauck. The proof is information-theoretic, and a key part of it is using asymmetric triangular discrimination instead of total variation distance; this idea may be useful elsewhere.

2020 MSC Codes: Primary 68Q17; Secondary 94A05

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 $f_A\colon A \to B$ and Bob gets as input $f_B\colon B \to A$, 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

$$z_0 = 1, \quad z_1 = f_A(z_0), \quad z_2 = f_B(z_1), \quad z_3 = f_A(z_2), \quad z_4 = f_B(z_3), \ \ldots. \tag{1.1}$$

The k-step pointer chasing function $PC_{n,k}$ is defined as[1]

$$PC_{n,k}(f_A, f_B) = z_k \mod 2.$$

Equivalently, Alice sees the rightward edges and Bob sees the leftward edges in a directed balanced bipartite graph where the out-degree of each vertex is one. Their goal is to output the parity of the endpoint of a directed path of length k in the graph.

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 [19]. Its communication complexity was subsequently studied in many works (*e.g.* [6, 7, 11, 17, 20]).

The pointer chasing problem 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 [17], and that it can be used to prove the monotone constant-depth hierarchy that was proved by Klawe, Paul, Pippenger and Yannakakis [13]. It was further used for proving lower bounds on the time complexity of distributed computation [16], and for proving lower bounds on the space

†Research supported by ISF.

[1] Parity is just a balanced Boolean function.

complexity of streaming algorithms [9]. It is also related to the round elimination lemma, and to the complexity of predecessor search in the cell probe model [15, 23].

This work continues the study of the communication complexity of the pointer chasing problem. As discussed above, the results here may be helpful in other models as well. We start with a survey of known related results, and then state our main result and discuss its proof. We use standard communication complexity terminology. For formal definitions, see for example the textbook [14].

Background

Upper bounds. There is an obvious k-round deterministic protocol for computing $PC_{n,k}$ with communication $O(k \log n)$ in which Alice speaks first. Nisan and Wigderson [17] described a randomized $(k-1)$-round protocol for $PC_{n,k}$ 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 $PC_{n,k}$ in which Bob speaks first (see [20]).

Lower bounds. Papadimitriou and Sipser [19] conjectured that $(k-1)$-round protocols for $PC_{n,k}$ 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 $PC_{n,k}$ is at least $\Omega(n/k^2)$. Later on, Nisan and Wigderson [17] 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 $(k-1)$-round randomized communication complexity. Ponzio, Radhakrishnan and Venkatesh [20] 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 $PC_{n,k}$ for constant k. Klauck [11] observed that the proof of the deterministic lower bound from [17] actually implies an essentially sharp $\Omega(k+(n/k))$ lower bound on the $(k-1)$-round randomized communication complexity of $PC_{n,k}$. Finally, Klauck, Nayak, Ta-Shma and Zuckerman [12] proved an $\Omega(n/\exp(k) - k \log n)$ lower bound on the $(k-1)$-round quantum communication complexity of $PC_{n,k}$.

This work. Here we focus on the distributional communication complexity of the pointer chasing problem. We consider the uniform distribution on inputs (*e.g.* f_A, f_B are chosen independently and uniformly at random) which seems like the most natural choice. Previously, the only known lower bound on the k-round distributional complexity of $PC_{n,k}$ 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 [11] together with von Neumann's minimax theorem (Yao's principle) shows that there is some distribution for which an $\Omega(k+(n/k))$ lower bound holds. However, this distribution is not explicit[2] and, for example, prior to this work the best lower bound that was known for any product distribution was Nisan and Wigderson's. Finding simple and explicit distributions for central problems proved to be useful, for example, for the disjointness function [22].

Main result. The main result of this work is a tight (up to polylog(n) factors) lower bound on the distributional communication complexity of pointer chasing under the uniform distribution.

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

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 the protocol reveals on the inputs (the proof in [17] can be stated in such a way as well). The goal is to prove that if the protocol is short, then after the protocol

[2] We are not aware of any method that is based on the minimax theorem for making it explicit.

terminates the inputs are still fairly 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 us to avoid the square-root loss that Pinsker's inequality yields.

This square-root loss appears in many works, and is directly related to several 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 [10]. The 'strong parallel repetition' conjecture was falsified by Raz [21]; showing 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 the square-root loss is necessary for the direct sum question.

Coming back to pointer chasing, the square-root loss also appears in Nisan and Wigderson's lower bound [17]. This work shows that we can circumvent this loss by using triangular discrimination instead of Kullback–Leibler divergence. We are not aware of any other metric or divergence that can replace triangular discrimination in this respect. We believe that using triangular discrimination can yield better quantitative bounds in other cases as well. For this reason, in Section 2.1, we provide a clean example that demonstrates the main new technical idea.

2. 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 real convex function f so that $f(1) = 0$ (where some conventions such as $0f(0/0) = 0$ are used). For more background, see [5] and the references within.

Some well-known examples of f-divergences 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 such as the chain rule).

Here we use the triangular discrimination [24] 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 briefly discuss its properties (for more details see [24, 5]). Like 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: $\Delta/2 \leqslant JS \leqslant 2\Delta$. It is, however, sometimes easier to work with Δ than JS since its formula is 'simpler'. It satisfies the following 'improvement' over Pinsker's inequality (which states that $|p - q|_1^2 \leqslant 2D(p||q)$).

https://doi.org/10.1017/S0963548320000085 Published online by Cambridge University Press

Lemma 2.1 ([24]). $|p-q|_1^2/2 \leqslant \Delta(p,q) \leqslant 2D(p||q)$.

Another interesting ('operational' or 'dual') interpretation of Δ 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 the fact that ℓ_1 is equivalent to total variation distance. For Δ we have the following.

Lemma 2.2.

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

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

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

On the other hand, for every g, by Cauchy–Schwarz,

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

As a final remark, we mention that recently Δ was implicitly used in information-theoretic proofs in group theory; it was used to construct group homomorphisms [8], it was used to study harmonic functions on groups [2], and it was used in a functional analytic proof of Gromov's theorem on groups of polynomial growth [18]. It is therefore reasonable that Δ will find more applications in computer science as well.

2.1 An example

Before proving the lower bound for pointer chasing, we describe a clean example that demonstrates how can one use Δ instead of ℓ_1 to get quantitatively better bounds. Let X be a random vector in $\{0,1\}^n$. Assume that it has high entropy:

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

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 3.1 implies that

$$\mathop{\mathbb{E}}_I D(p_{X_i}||u_1) \leqslant \frac{1}{n}D(p_X||u_n) \leqslant \frac{k}{n}. \tag{2.1}$$

That is, on average over I, the marginal distribution of X_I is close to uniform in Kullback–Leibler divergence, when $k \ll n$. Pinsker's inequality allows us 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:

$$D(p_J||p_I) \leqslant \varepsilon.$$

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

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

This square-root dependence is often too expensive, especially when we apply such an argument several times, as discussed after the statement of Theorem 1.1. Triangular discrimination allows us to remove this square-root dependence.

Theorem 2.3. $\mathbb{E}_J\, \Delta(p_{X_j}, u_1) \leqslant 4\varepsilon + 10k/n.$

For the rest of this subsection we prove Theorem 2.3. We start with the following simple claim.

Claim 1. *If* $|\xi| \leqslant \sqrt{a(b+\xi)}$ *with* $a, b \geqslant 0$, *then* $\xi \leqslant a + 2b$.

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

$$\xi \leqslant \frac{a + \sqrt{a^2 + 4ab}}{2} = \frac{a}{2}(1 + \sqrt{1 + 4b/a}) \leqslant \frac{a}{2}(1 + 1 + 4b/a). \qquad \square$$

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 2.1 and (2.1) allow us to bound the left term:

$$p_I.g \leqslant 2 \mathop{\mathbb{E}}_I D(p_{X_i}||u_1) \leqslant \frac{2k}{n}. \tag{2.2}$$

It thus remains to upper-bound

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

This is done as follows:

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

Use Claim 1, together with (2.2) and

$$\Delta(p_J, p_I) \leqslant 2D(p_J||p_I) \leqslant 2\varepsilon,$$

to deduce that

$$\xi \leqslant 4\varepsilon + 8k/n.$$

Together with (2.2), Theorem 2.3 is proved.

2.2 Asymmetric triangular discrimination

To prove the lower bound for pointer chasing, we shall actually use the following variant of Δ:

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

https://doi.org/10.1017/S0963548320000085 Published online by Cambridge University Press

Note that $\Lambda \leqslant \Delta$ and that Δ is symmetric in p, q while Λ is not.

The following lemma states important properties of Λ; it relates Λ to ℓ_1, and shows that Λ is at most one (Δ may take the value two).

Lemma 2.4.

$$\frac{1}{8}|p-q|_1^2 \leqslant \Lambda(p,q) \leqslant \frac{1}{2}|p-q|_1 \leqslant 1.$$

This difference between $\Lambda \leqslant 1$ and $\Delta \leqslant 2$ is useful when we iteratively bound the 'distance' between two distributions, as in the proof of Theorem 1.1, since $1^k = 1$ but 2^k grows quickly with k.

Proof. The left inequality holds by Cauchy–Schwarz:

$$\frac{1}{2}|p-q|_1 = \sum_{\omega:\, p(\omega)\geqslant q(\omega)} \frac{p(\omega)-q(\omega)}{\sqrt{p(\omega)+q(\omega)}}\sqrt{p(\omega)+q(\omega)} \leqslant \sqrt{2\Lambda(p,q)}.$$

The middle inequality holds by the first equality in the equation above, because

$$(p(\omega)-q(\omega))^2/(p(\omega)+q(\omega)) \leqslant |p(\omega)-q(\omega)| \quad \text{for all } \omega.$$

The right inequality holds since $|p-q|_1 \leqslant |p|_1 + |q|_1 = 2$. □

To explain the reason for using Λ instead of Δ, let us go back to Theorem 2.3. Although the theorem avoids the square-root loss, the coefficient of ε on the right-hand side is 4. When repeatedly applying this theorem, we get an exponential blowup, which is too costly to carry. The following theorem shows that Λ allows us to avoid this blowup; the coefficient on the right-hand side can be 1.

Theorem 2.5. *With the same notation as in Theorem 2.3,*

$$\mathop{\mathbb{E}}_J \Lambda(p_{X_j}, u_1) \leqslant \Lambda(p_J, p_I) + 10k/n.$$

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

$$\begin{aligned}
\mathop{\mathbb{E}}_J \Lambda(p_{X_j}, u_1) &= p_J.g \\
&= p_I.g + (p_J.g - p_I.g) \\
&\leqslant p_I.g + \sum_{s\in S}(p_J(s)-p_I(s))g(s), \quad (\Lambda \geqslant 0)
\end{aligned}$$

where $S = \{s\colon p_J(s) \geqslant p_I(s)\}$. Bound

$$\xi = \sum_{s\in S}(p_J(s)-p_I(s))g(s)$$

by writing

$$\begin{aligned}
|\xi| &\leqslant \sqrt{\sum_{s\in S}\frac{(p_J(s)-p_I(s))^2}{p_J(s)+p_I(s)}g(s)}\sqrt{\sum_{s\in S}(p_J(s)+p_I(s))g(s)} \quad \text{(Cauchy–Schwarz)} \\
&\leqslant \sqrt{\Lambda(p_J,p_I)}\sqrt{\sum_{s\in S}(p_J(s)+p_I(s))g(s)} \quad (\Lambda \leqslant 1) \\
&= \sqrt{\Lambda(p_J,p_I)}\sqrt{\xi + 2p_I.g}, \quad (\Lambda \geqslant 0)
\end{aligned}$$

https://doi.org/10.1017/S0963548320000085 Published online by Cambridge University Press

Claim 1 implies that

$$\xi \leqslant \Lambda(p_J, p_I) + 4p_I.g.$$

Because $\Lambda \leqslant \Delta$ and using (2.2), we get $p_I.g \leqslant 2k/n$. Overall,

$$\mathop{\mathbb{E}}_J \Lambda(p_{X_j}, u_1) \leqslant \Lambda(p_J, p_I) + 5p_I.g \leqslant \Lambda(p_J, p_I) + 10k/n. \qquad \square$$

3. 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 let $p_{X|y}$ denote the probability distribution of X conditioned on $Y = y$. We let $\mathbb{E}_X f(x)$ denote the expectation of $f(X)$, and let $\mathbb{E}_{X|y} f(x)$ denote the expectation of $f(X)$ conditioned on $Y = y$.

Kullback–Leibler divergence. We state two lemmas that will be useful later on[3] (see *e.g.* the textbook [4]).

Lemma 3.1 (superadditivity). *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*

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

Lemma 3.2 (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*

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

4. The lower bound

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

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 its 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 $R_1 = M_1$). We shall prove that

$$\mathop{\mathbb{E}}_{R_t} \Lambda(p_{Z_t|r_t}, p_{Z_t}) \leqslant 6t\delta, \tag{4.1}$$

where

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

Roughly speaking, the expression $\mathbb{E}_{R_t} \Lambda(p_{Z_t|r_t}, p_{Z_t})$ measures how much we learned on the value of Z_t from observing r_t; if this expression is small then we did not learn much on Z_t from the first t rounds of the protocol.

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

https://doi.org/10.1017/S0963548320000085 Published online by Cambridge University Press

Before proving (4.1), we explain why it completes the proof. Since the fraction of even numbers in $[n]$ is at least $1/2 - 1/n$, the error of the protocol conditioned on $R_k = r_k$ is at least

$$\mathrm{err}_{r_k} \geqslant \frac{1}{2} - \frac{1}{n} - \frac{|p_{Z_k|r_{k-1}} - p_{Z_k}|_1}{2}.$$

Hence, since the protocol has error $1/3$,

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

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

It thus remains to prove (4.1). When $t = 0$ it indeed holds (R_0 is empty). Suppose $t \geqslant 1$. 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

$$\mathop{\mathbb{E}}_{R_{t-1}} \Lambda(p_{Z_{t-1}|r_{t-1}}, p_{Z_{t-1}}) \leqslant 6(t-1)\delta. \tag{4.2}$$

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: let X denote the vector that represents Alice's input ($X_s = f_A(s)$ for each s), and let Y denote 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 $Z_{t-1} = Y_{z_{t-2}}$). This means that conditioned on $(R_{t-1}, Z_{t-1}) = (r_{t-1}, z_{t-1})$, the distribution $p_{X_{z_{t-1}}|r_{t-1}, z_{t-1}}$ is equal to $p_{X_{z_{t-1}}|r_{t-1}}$.

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 (*e.g.* 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

$$\begin{aligned}
\mathop{\mathbb{E}}_{R_t} \Lambda(p_{Z_t|r_t}, p_{Z_t}) &= \mathop{\mathbb{E}}_{R_{t-1}, Z_{t-1}, M_t} \Lambda(p_{Z_t|r_{t-1}, z_{t-1}}, p_{Z_t}) && \text{(A)} \\
&= \mathop{\mathbb{E}}_{R_{t-1}, Z_{t-1}} \Lambda(p_{X_{z_{t-1}}|r_{t-1}, z_{t-1}}, p_{Z_t}) \\
&= \mathop{\mathbb{E}}_{R_{t-1}, Z_{t-1}} \Lambda(p_{X_{z_{t-1}}|r_{t-1}}, p_{Z_t}) && \text{(B)} \\
&= \mathop{\mathbb{E}}_{R_{t-1}} \mathop{\mathbb{E}}_{Z_{t-1}|r_{t-1}} \Lambda_{z_{t-1}},
\end{aligned} \tag{4.3}$$

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

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

https://doi.org/10.1017/S0963548320000085 Published online by Cambridge University Press

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 $Z_{t-1}|r_{t-1}$ with 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,

$$\begin{aligned}
\mathop{\mathbb{E}}_{R_{t-1}} \mathop{\mathbb{E}}_{I} \Lambda_i &\leqslant \mathop{\mathbb{E}}_{R_{t-1}} \mathop{\mathbb{E}}_{I} \Delta(p_{X_i|r_{t-1}}, p_{X_i}) && (\Lambda \leqslant \Delta)\\
&\leqslant 2 \mathop{\mathbb{E}}_{R_{t-1}} \mathop{\mathbb{E}}_{I} D(p_{X_i|r_{t-1}} || p_{X_i}) && (\text{Lemma 2.1})\\
&\leqslant \frac{2}{n} \mathop{\mathbb{E}}_{R_{t-1}} D(p_{X|r_{t-1}} || p_X) && (\text{Lemma 3.1})\\
&\leqslant \delta. && (\text{Lemma 3.2})
\end{aligned} \tag{4.4}$$

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:

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

Since

$$\sum_{s\colon q(s) < p_I(s)} \frac{(q(s) - p_I(s))^2}{q(s) + p_I(s)} \Lambda_s \leqslant \sum_s \frac{(p_I(s))^2}{p_I(s)} \Lambda_s = \mathop{\mathbb{E}}_{I} \Lambda_i,$$

by Claim 1 we have

$$\xi \leqslant \Lambda(q, p_I) + 5 \mathop{\mathbb{E}}_{I} \Lambda_i.$$

Now, taking expectation over R_{t-1}, by (4.2) and (4.4), since $p_I = p_{Z_{t-1}}$,

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

Finally, by (4.3) and (4.4), the inductive claim is proved. □

Acknowledgements

I thank Anup Rao and Avi Wigderson for helpful comments.

https://doi.org/10.1017/S0963548320000085 Published online by Cambridge University Press

References

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

[2] Benjamini, I., Duminil-Copin, H., Kozma, G. and Yadin, A. (2015) Disorder, entropy and harmonic functions. *Ann. Probab.* **43** 2332–2373.

[3] Braverman, M., Rao, A., Weinstein, O. and Yehudayoff, A. (2013) Direct products in communication complexity. In *54th Annual Symposium on Foundations of Computer Science (FOCS)*, IEEE, pp. 746–755.

[4] Cover, T. M. and Thomas, J. A. (2012) *Elements of Information Theory*, Wiley.

[5] Csiszár, I. and Shields, P. C. (2004) *Information Theory and Statistics: A Tutorial*, Now Publishers.

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

[7] Duris, P., Galil, Z. and Schnitger, G. (1984) Lower bounds on communication complexity. In *16th Annual ACM Symposium on Theory of Computing (STOC)*, ACM, pp. 81–91.

[8] Erschler, A. and Karlsson, A. (2010) Homomorphisms to *r* constructed from random walks. *Ann. Inst. Fourier* **60** 2095–2113.

[9] Guruswami, V. and Onak, K. (2013) Superlinear lower bounds for multipass graph processing. In *2013 IEEE Conference on Computational Complexity (CCC)*, IEEE, pp. 287–298.

[10] Khot, S. (2002) On the power of unique 2-prover 1-round games. In *34th Annual ACM Symposium on Theory of Computing (STOC)*, ACM, pp. 767–775.

[11] Klauck, H. (2000) On quantum and probabilistic communication: Las Vegas and one-way protocols. In *32nd Annual ACM Symposium on Theory of Computing (STOC)*, ACM, pp. 644–651.

[12] Klauck, H., Nayak, A., Ta-Shma, A. and Zuckerman, D. (2007) Interaction in quantum communication. *IEEE Trans. Inform. Theory* **53** 1970–1982.

[13] Klawe, M., Paul, W. J., Pippenger, N. and Yannakakis, M. (1984) On monotone formulae with restricted depth. In *16th Annual ACM Symposium on Theory of Computing (STOC)*, ACM, pp. 480–487.

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

[15] Miltersen, P. B., Nisan, N., Safra, S. and Wigderson, A. (1995) On data structures and asymmetric communication complexity. In *27th Annual ACM Symposium on Theory of Computing (STOC)*, ACM, pp. 103–111.

[16] Nanongkai, D., Das Sarma, A. and Pandurangan, G. (2011) A tight unconditional lower bound on distributed random walk computation. In *30th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC)*, pp. 257–266.

[17] Nisan, N. and Wigderson, A. (1991) Rounds in communication complexity revisited. In *23rd Annual ACM Symposium on Theory of Computing (STOC)*, ACM, pp. 419–429.

[18] Ozawa, N. (2018) A functional analysis proof of Gromov's polynomial growth theorem. *Ann. Sci. De L'ENS* 549–556.

[19] Papadimitriou, C. H. and Sipser, M. (1984) Communication complexity. *J. Comput. System Sci.* **28** 260–269.

[20] Ponzio, S. J., Radhakrishnan, J. and Venkatesh, S. (2001) The communication complexity of pointer chasing. *J. Comput. System Sci.* **62** 323–355.

[21] Raz, R. (2011) A counterexample to strong parallel repetition. *SIAM J. Comput.* **40** 771–777.

[22] Razborov, A. A. (1992) On the distributional complexity of disjointness. *Theoret. Comput. Sci.* **106** 385–390.

[23] Sen, P. and Venkatesh, S. (2008) Lower bounds for predecessor searching in the cell probe model. *J. Comput. System Sci.* **74** 364–385.

[24] Topsøe, F. (2000) Some inequalities for information divergence and related measures of discrimination. *IEEE Trans. Inform. Theory* **46** 1602–1609.

Cite this article: Yehudayoff A (2020). Pointer chasing via triangular discrimination. *Combinatorics, Probability and Computing* **29**, 485–494. https://doi.org/10.1017/S0963548320000085