

# Relative Discrepancy does not separate Information and Communication Complexity

Lila Fontes<sup>1</sup>, Rahul Jain<sup>2</sup>, Iordanis Kerenidis<sup>3</sup>, Sophie Laplante<sup>1</sup>,  
Mathieu Laurière<sup>1</sup>, Jérémie Roland<sup>4</sup>

<sup>1</sup> Université Paris-Diderot, LIAFA

<sup>2</sup> CQT, National University of Singapore

<sup>3</sup> CNRS, Université Paris-Diderot, LIAFA

<sup>4</sup> ULB, QuIC, Ecole Polytechnique de Bruxelles

fontes, jkeren, mathieu.lauriere, laplante@liafa.univ-paris-diderot.fr,  
rahul@comp.nus.edu.sg, jroland@ulb.ac.be

**Abstract.** Does the information complexity of a function equal its communication complexity? We examine whether any currently known techniques might be used to show a separation between the two notions. Ganor *et al.* recently provided such a separation in the distributional case for a specific input distribution. We show that in the non-distributional setting, the relative discrepancy bound is smaller than the information complexity, hence it cannot separate information and communication complexity. In addition, in the distributional case, we provide a linear program formulation for relative discrepancy and relate it to variants of the partition bound, resolving also an open question regarding the relation of the partition bound and information complexity. Last, we prove the equivalence between the adaptive relative discrepancy and the public-coin partition, implying that the logarithm of the adaptive relative discrepancy bound is quadratically tight with respect to communication.

## 1 Introduction

The question of whether information complexity equals communication complexity is one of the most important outstanding questions in communication complexity. Communication complexity measures the amount of bits Alice and Bob need to communicate to each other in order to compute a function whose input is shared between them. On the other hand, information complexity measures the amount of information Alice and Bob must reveal about their inputs in order to compute the function. Equality between information and communication complexity is equivalent to a compression theorem in the interactive setting. It is known that a single message can be compressed to its information content [1,2,3,4] and here the question is whether such a compression is possible for an interactive conversation.

An important application of information complexity is to prove direct sum theorems for communication complexity, namely show that computing  $k$  instances of a function costs  $k$  times the communication of computing a single instance. This has been shown to be true in the simultaneous and one-way models [5,6], for bounded-round two-way protocols under product distributions [7,3] or non-product distributions [1], and also for specific functions like Disjointness [8]; non-trivial direct sum theorems have also been shown for general two-way randomized communication complexity [9]. Since the information complexity is equal to amortized communication complexity [1], the question of whether information and communication complexity are equal is equivalent to whether communication complexity has a direct sum property [10,1]. Note that in the case of deterministic, zero-error protocols, a separation between information and communication complexity is known for Equality [10].

Since information complexity deals with the information Alice and Bob transmit about their inputs, it is necessary to define a distribution on these inputs. For each fixed distribution  $\mu$ , we define the distributional information complexity of a function  $f$  (also known as the information cost) as the information Alice and Bob transmit about their inputs in any protocol that solves  $f$  with small error according to  $\mu$  [1,5]. The (non-distributional) information complexity of the function  $f$  is defined as its distributional information complexity for the worst distribution  $\mu$  [10]. In this paper we consider the internal information complexity.

Similarly, for communication complexity, one may also consider a model with a distribution  $\mu$  over the inputs, and the error probability of the protocol is taken over this distribution. This is called a distributional model, and Yao's minmax principle [11] states that the randomized communication complexity of  $f$  is equal to its distributional communication complexity for the worst distribution  $\mu$ , where the randomized communication complexity of a function  $f$  is defined as the minimum number of bits exchanged, in the worst case over the inputs, for a randomized protocol to compute the function with small error [12].

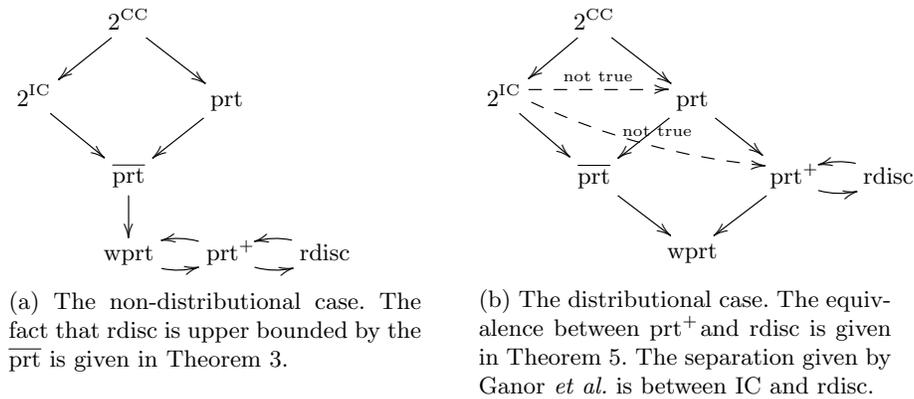
One can therefore ask whether the following stronger relation holds: is the distributional communication complexity equal to the distributional information complexity for all input distributions  $\mu$ ? A positive answer to this question would also imply a positive answer to the initial question, proving the equality of information and communication complexity.

In a recent breakthrough, Ganor *et al.* [13,14] defined a function  $f$  and a distribution  $\mu$ , for which there is an exponential separation between the distributional information and communication complexity. Does this

settle the question of communication versus information? First, let us note that the gap, although exponential, is very small compared to the input size: a  $\log \log(n)$  communication lower bound and a  $\log \log \log(n)$  information upper bound, for inputs of size  $n$ . More importantly, Ganor *et al.*'s results prove that the *distributional* information and communication complexities are not equal for all distributions  $\mu$ .

How could we settle the question in the non-distributional setting? To prove a separation it is necessary to show that the communication complexity of a specific function is large, while its information complexity is small. In other words, we need a lower bound technique which provides a lower bound for communication but *not* for information.

In previous work, Kerenidis *et al.* [15] showed that almost all known lower bound techniques for communication also provide lower bounds for information. More precisely, they studied the relaxed partition bound and proved that it subsumes all known lower bound techniques (except the partition bound [16]). In addition, they proved that for any distribution  $\mu$ , the distributional information complexity can be lower bounded by the relaxed partition bound. This also holds in the non-distributional setting. An open question was whether the partition bound remained a candidate for separating information and communication complexity.



(a) The non-distributional case. The fact that  $\text{rdisc}$  is upper bounded by the  $\overline{\text{prt}}$  is given in Theorem 3.

(b) The distributional case. The equivalence between  $\text{prt}^+$  and  $\text{rdisc}$  is given in Theorem 5. The separation given by Ganor *et al.* is between  $IC$  and  $\text{rdisc}$ .

Fig. 1: Definitions follow in Sections 3, 4, and 5. An arrow from one bound to another indicates that the former is at least as large as the latter.

The main question we ask is whether the techniques developed by Ganor *et al.* can help in proving, or disproving, the equality of information and communication complexity of a function  $f$  in the non-distributional

setting. For their separation, Ganor *et al.* introduced a new communication lower bound called relative discrepancy. They showed that for a specific function  $f$  and a specific distribution  $\mu$ , this quantity is high, while the distributional information complexity is low. We study how large this new bound is compared to the other known lower bound techniques, and whether it can be used to separate information and communication complexity in the non-distributional setting. Our main results are:

**Result 1:** In the non-distributional case, we show that relative discrepancy is bounded above by the relaxed partition bound (**Theorem 3**). By the results of [15], this means that relative discrepancy cannot be used to separate information and communication complexity.

**Result 2:** In the distributional case, we provide a clear relation between relative discrepancy, relaxed partition and partition bound. We give an equivalent linear program formulation for relative discrepancy (**Theorem 5**) and show how relative discrepancy and relaxed partition can be derived from the partition bound by imposing some simple extra constraints. This also answers negatively to the open question in [15] regarding whether the partition bound is a lower bound on information.

Recently, lower bound techniques that use partitions instead of considering just rectangles have been proposed. Jain *et al.* defined the public coin partition bound, and showed that its logarithm is quadratically related to communication complexity [17]. In addition, Ganor *et al.* introduced the adaptive relative discrepancy [14]. We study the relation between them and show the following:

**Result 3:** For any  $\mu$ , adaptive relative discrepancy and public-coin partition bound are equivalent (**Theorem 6**). Hence the logarithm of the adaptive relative discrepancy is quadratically tight to communication.

In addition to providing a linear program for relative and adaptive relative discrepancies, the different variants of the partition bound have several other advantages. They can be defined for a wider range of problems, including non-boolean functions; they have natural interpretations in terms of zero-communication protocols, a fact used for relating information complexity to these bounds [15] and for recent advances in the log rank conjecture [18].

In Section 2 we provide the necessary background and definitions. In Section 3 we prove that relative discrepancy is less than relaxed partition (in the non-distributional setting). In Section 4 we consider the setting with a fixed  $\mu$ , and compare the partition bound and its variants to the relative discrepancy bound. In Section 5, we consider the adaptive relative

discrepancy and compare it to the public coin partition bound. The full version of the paper appears in [19].

## 2 Preliminaries

Let  $\mathbf{X}$  and  $\mathbf{Y}$  be the sets of inputs to the two players, and  $\mathbf{Z}$  be the set of possible outputs. Since the discrepancy-based bounds studied in this paper apply naturally only to boolean functions,  $f$  will usually denote a (possibly partial) function over  $\mathbf{X} \times \mathbf{Y}$  taking values in  $\mathbf{Z} = \{0, 1\}$ , while  $\mu$  denotes a probability distribution over  $\mathbf{X} \times \mathbf{Y}$ .<sup>5</sup>

### 2.1 Information and communication complexity

For any (possibly partial) function  $f$  over inputs  $\mathbf{X} \times \mathbf{Y}$ , and any  $\epsilon \in (0, 1/2)$ , the communication cost of a protocol that computes  $f$  with error probability at most  $\epsilon$  is the number of bits sent for the worst case input.

**Definition 1.** *The (public-coin) communication complexity of  $f$ , denoted  $R_\epsilon(f)$ , is the best communication cost for any protocol that computes  $f$  using public coins with error at most  $\epsilon$  for any input  $(x, y)$ . For any distribution  $\mu$  over the inputs, the distributional (public-coin) communication complexity of  $f$ , denoted  $R_\epsilon(f, \mu)$ , is the cost of the best protocol that computes  $f$  with error at most  $\epsilon$ , where the error probability is taken over the input distribution.*

For information complexity, we are interested not in the number of bits exchanged, but the amount of information revealed about the inputs. We consider the internal information complexity in this paper. Here  $I(X; Y)$  denotes the mutual information between random variables  $X$  and  $Y$ , and  $I(X; Y|Z)$  is the mutual information conditioned on  $Z$ .

**Definition 2 (Information complexity).** *Fix  $f, \mu, \epsilon$ . Let  $(X, Y, \Pi)$  be the tuple distributed according to  $(X, Y)$  sampled from  $\mu$  and then  $\Pi$  being the transcript of the protocol  $\pi$  applied to  $X, Y$ . Then define:*

1.  $IC_\mu(\pi) = I(X; \Pi | Y) + I(Y; \Pi | X)$
2.  $IC_\mu(f, \epsilon) = \inf_\pi IC_\mu(\pi)$ , where  $\pi$  computes  $f$  with error at most  $\epsilon$
3.  $IC(f, \epsilon) = \max_\mu IC_\mu(f, \epsilon)$

---

<sup>5</sup> The partition-based definitions apply to non-boolean functions, relations, and bipartite distributions as well, but we do not give the full definitions in this paper for those settings.

## 2.2 Lower bound techniques

For any family of variables  $\{\beta_{x,y}\}_{(x,y)\in\mathbf{X}\times\mathbf{Y}}$  and any subset  $E \subseteq \mathbf{X} \times \mathbf{Y}$ , we will denote  $\beta(E) = \sum_{(x,y)\in E} \beta_{x,y}$ , and  $\beta = \beta(\mathbf{X} \times \mathbf{Y})$ . Unless otherwise specified “ $\forall x, y$ ” means “ $\forall x, y \in \mathbf{X} \times \mathbf{Y}$ ”, “ $\forall z$ ” means “ $\forall z \in \mathbf{Z}$ ”, “ $\forall R$ ” means “for all rectangles  $R$  in  $\mathbf{X} \times \mathbf{Y}$ ”, and “ $\forall P$ ” means “for all partitions  $P$  of  $\mathbf{X} \times \mathbf{Y}$  into labeled rectangles  $(R, z)$ ”. We also denote by  $|P|$  the size of the partition, that is, the number of rectangles  $(R, z)$  it contains.

Following Ganor *et al.* (with small changes that do not affect the value of the bound), we define the relative discrepancy bound  $\text{rdisc}_\varepsilon(f, \mu)$ , as follows. Without loss of generality, we assume  $\text{supp}(\mu) = \text{supp}(f)$ .

**Definition 3 (Relative discrepancy bound [14]).** *Let  $\mu$  be a distribution over  $\mathbf{X} \times \mathbf{Y}$  and let  $f : \text{supp}(\mu) \rightarrow \{0, 1\}$  be a function.*

$$\begin{aligned} \text{rdisc}_\varepsilon(f, \mu) = \sup_{\kappa, \delta, \rho_{xy}} & \frac{1}{\delta} \left( \frac{1}{2} - \kappa - \varepsilon \right) \\ \text{subject to} & \quad \left( \frac{1}{2} - \kappa \right) \cdot \rho(R) \leq \mu(R \cap f^{-1}(z)) \quad \forall R, z \text{ s.t. } \rho(R) \geq \delta \\ & \quad \sum_{xy} \rho_{xy} = 1, \quad 0 \leq \kappa < \frac{1}{2}, \quad 0 < \delta < 1, \quad \rho_{xy} \geq 0 \quad \forall (x, y). \end{aligned}$$

For the non-distributional case, we define  $\text{rdisc}_\varepsilon(f) = \max_\mu \text{rdisc}_\varepsilon(f, \mu)$ , where the maximum is over distributions  $\mu$  over  $\mathbf{X} \times \mathbf{Y}$  (which implicitly adds nonnegativity and normalization constraints on  $\mu$ ).

Note that neither the constraints nor the objective function are linear in the variables. Intuitively, the distribution  $\rho$  rebalances the weight of the 0-region and the 1-region of any rectangle  $R$  by putting weights on all  $(x, y)$  and not just the ones in the support of  $\mu$ . If this rebalancing is possible even for rectangles with very small weight (i.e.  $\delta$  is small), then the relative discrepancy increases.

Using this formulation, Ganor *et al.* show:

**Theorem 1 ([14]).** *Let  $f : \text{supp}(\mu) \rightarrow \{0, 1\}$  be a (possibly partial) function. Then  $\log(\text{rdisc}_\varepsilon(f, \mu)) \leq R_\varepsilon(f, \mu)$ .*

The relaxed partition bound was introduced by Kerenidis *et al.* [15] who proved that for any function, it is bounded above by its information complexity. Their result holds also relative to any input distribution.<sup>6</sup>

<sup>6</sup> Compared with the original formulation [15], there is an implicit change of variables: we use  $\beta_{x,y}$  here to denote what was  $\alpha_{x,y} - \beta_{x,y}$  in the original notation.

**Definition 4 (Relaxed partition bound [15]).** Let  $\mu$  be a distribution over  $\mathbf{X} \times \mathbf{Y}$  and let  $f : \text{supp}(\mu) \rightarrow \{0, 1\}$  be a function.

$$\begin{aligned} \overline{\text{prt}}_\epsilon(f, \mu) &= \max_{\alpha, \beta_{xy}} \beta - \alpha\epsilon \\ \text{subject to : } & \beta(R) - \alpha\mu(R \cap f^{-1}(z)) \leq 1 \quad \forall R, z \\ & \alpha \geq 0, \quad \alpha\mu_{xy} - \beta_{xy} \geq 0 \quad \forall (x, y), \end{aligned}$$

where  $R$  ranges over all rectangles,  $(x, y) \in \mathbf{X} \times \mathbf{Y}$  and  $z \in \{0, 1\}$ . The non-distributional relaxed partition bound is  $\overline{\text{prt}}_\epsilon(f) = \max_\mu \overline{\text{prt}}_\epsilon(f, \mu)$ . For the non-distributional case, we use  $\alpha_{x,y}$  instead of  $\alpha\mu_{x,y}$  (which is not linear if  $\mu$  is no longer fixed), with  $\alpha_{x,y}$  positive but not normalized.

Kerenidis *et al.* [15] provided both a primal and dual formulation of the relaxed partition bound. The above is the dual formulation. The corresponding primal formulation can be interpreted in terms of the highest non-abort probability of a zero-communication protocol for  $f$ .

**Theorem 2 ([15]).** For all  $\mu$ , boolean functions  $f$  over the support of  $\mu$  and all  $\epsilon \in (0, \frac{1}{4}]$ ,  $\Omega(\epsilon^2 \log \overline{\text{prt}}_{2\epsilon}(f, \mu)) = \text{IC}_\mu(f, \epsilon) \leq R_\epsilon(f, \mu)$ .

### 3 Relative discrepancy is bounded by relaxed partition

We show that the non-distributional relative discrepancy is bounded above by the relaxed partition, which implies that a stronger technique is necessary in order to separate information and communication complexity. (See Figure 1a).

**Theorem 3.** For any boolean  $f$ , and  $\epsilon \in (0, 1/3)$ ,  $\text{rdisc}_{\frac{3}{2}\epsilon}(f) \leq \overline{\text{prt}}_\epsilon(f)$ .

*Proof.* It suffices to show that for any feasible solution of  $\text{rdisc}$ , there exists a feasible solution for  $\overline{\text{prt}}$  whose objective value is at least as large. Let  $(\kappa, \delta, \{\rho_{x,y}\}_{x,y}, \{\mu_{x,y}\}_{x,y})$  be a feasible solution of relative discrepancy for  $f$ . Define for any  $(x, y) \in \mathbf{X} \times \mathbf{Y}$ ,  $\alpha_{x,y} = \frac{1}{\delta}(\frac{1}{2} - \kappa)\rho_{x,y} + \frac{1}{\delta}\mu_{x,y}$  and  $\beta_{x,y} = \frac{1}{\delta}(\frac{1}{2} - \kappa)\rho_{x,y}$ . We show that the relaxed partition constraints are satisfied. First, the sign constraints are satisfied. Moreover, for any  $R, z$ ,

$$\begin{aligned} & \beta(R) - \alpha(R \cap f^{-1}(z)) \\ &= \frac{1}{\delta}(\frac{1}{2} - \kappa)\rho(R) - \frac{1}{\delta}\mu(R \cap f^{-1}(z)) - \frac{1}{\delta}(\frac{1}{2} - \kappa)\rho(R \cap f^{-1}(z)) \\ &\leq \frac{1}{\delta}(\frac{1}{2} - \kappa)\rho(R) - \frac{1}{\delta}\mu(R \cap f^{-1}(z)) \quad (\text{since } \rho_{xy} \geq 0 \text{ for any } (x, y)) \end{aligned}$$

There are two cases: if  $\rho(R) \geq \delta$ , then  $\frac{1}{\delta}(\frac{1}{2} - \kappa)\rho(R) - \frac{1}{\delta}\mu(R \cap f^{-1}(z)) \leq 0 \leq 1$  by the relative discrepancy constraint; otherwise  $\rho(R) < \delta$  and  $\frac{1}{\delta}(\frac{1}{2} - \kappa)\rho(R) - \frac{1}{\delta}\mu(R \cap f^{-1}(z)) < (\frac{1}{2} - \kappa) - \frac{1}{\delta}\mu(R \cap f^{-1}(z)) \leq \frac{1}{2} \leq 1$ .

Finally we compare the objective values. Since  $\rho$  and  $\mu$  are distributions,  $\alpha = \frac{1}{\delta}(\frac{3}{2} - \kappa)$  and  $\beta = \frac{1}{\delta}(\frac{1}{2} - \kappa)$ , so  $\beta - \epsilon\alpha = \frac{1}{\delta}[\frac{1}{2} - \kappa - (\frac{3}{2} - \kappa)\epsilon] \geq \frac{1}{\delta}(\frac{1}{2} - \kappa - \frac{3}{2}\epsilon) = \text{rdisc}_{\frac{3}{2}\epsilon}(f)$ .  $\square$

Combining Theorem 2 and Theorem 3 gives us that relative discrepancy is a lower bound on information complexity.

**Corollary 1.** *For all functions  $f : \mathbf{X} \times \mathbf{Y} \rightarrow \{0, 1\}$  and all  $\epsilon \in (0, \frac{1}{6}]$ ,  $\Omega(\epsilon^2 \log(\text{rdisc}_{3\epsilon}(f))) = \text{IC}(f, \epsilon) \leq R_\epsilon(f)$ .*

*Remark 1.* Our change of variables satisfies an additional constraint :

$$\beta_{x,y} \geq 0 \text{ for any } (x, y) \in \mathbf{X} \times \mathbf{Y}. \quad (1)$$

since  $\rho_{x,y} \geq 0$ . We will examine the role of this constraint in Section 4. It turns out to be a key point in understanding how relative discrepancy relates to the partition bound and its variants. Also notice that  $\alpha_{x,y}$  is not proportional to  $\mu_{x,y}$ , so this change of variable does not carry over to the distributional case, since  $\alpha_{x,y}$  cannot be written as  $\alpha\mu_{xy}$ .

## 4 The distributional case

In this section we study how the various bounds relate, relative to a fixed distribution  $\mu$ , and uncover an elegant relationship between the bounds by adding simple positivity constraints to the partition bound.

We start with a fixed-distribution version of the partition bound [16], which we define below. It follows easily from the original proof that this is a lower bound on distributional communication complexity and that it equals the partition bound in the worst case distribution.

**Definition 5 (Partition bound).**

$$\begin{aligned} \text{prt}_\epsilon(f, \mu) &= \max_{\alpha, \beta_{xy}} && \beta - \epsilon\alpha \\ \text{subject to :} &&& \beta(R) - \alpha\mu(R \cap f^{-1}(z)) \leq 1 \quad \forall R, z \\ &&& \alpha \geq 0. \end{aligned}$$

*The non-distributional bound is  $\text{prt}_\epsilon(f) = \max_\mu \text{prt}_\epsilon(f, \mu)$ . Going from the non-distributional setting to a fixed distribution  $\mu$ ,  $\alpha_{x,y}$  is replaced by  $\alpha \cdot \mu_{x,y}$ , that is,  $\{\alpha_{x,y}\}$  is  $\{\mu_{x,y}\}$  scaled by a factor  $\alpha$ .*

**Theorem 4 ([16]).** *Let  $f : \text{supp}(\mu) \rightarrow \{0, 1\}$  be a (possibly partial) function. Then  $\log(\text{prt}_\epsilon(f, \mu)) \leq R_\epsilon(f, \mu)$ .*

Note that the relaxed partition bound (Definition 4) is obtained from the partition bound by adding the constraint  $\alpha\mu_{x,y} - \beta_{xy} \geq 0$  for all  $(x, y)$ .

As suggested in the proof of Theorem 3, we now consider the constraint  $\beta_{x,y} \geq 0$  for all  $x, y$ . Adding this constraint to the partition bound results in a new bound which we call the positive partition bound.

**Definition 6 (Positive partition bound).**

$$\begin{aligned} \text{prt}_\epsilon^+(f, \mu) = \max_{\alpha, \beta_{xy}} \quad & \beta - \epsilon\alpha \\ \text{subject to :} \quad & \beta(R) - \alpha\mu(R \cap f^{-1}(z)) \leq 1 \quad \forall R, z \\ & \alpha \geq 0, \quad \beta_{xy} \geq 0 \quad \forall (x, y). \end{aligned}$$

We also define  $\text{prt}_\epsilon^+(f) = \max_\mu \text{prt}_\epsilon^+(f, \mu)$ , and use  $\alpha_{x,y}$  instead of  $\alpha\mu_{x,y}$ .

The weak partition bound is obtained by adding both constraints.

**Definition 7 (Weak partition bound).**

$$\begin{aligned} \text{wprt}_\epsilon(f, \mu) = \max_{\alpha, \beta_{xy}} \quad & \beta - \epsilon\alpha \\ \text{subject to :} \quad & \beta(R) - \alpha\mu(R \cap f^{-1}(z)) \leq 1 \quad \forall R, z, \\ & \alpha \geq 0, \quad \beta_{xy} \geq 0, \quad \alpha\mu_{xy} - \beta_{xy} \geq 0 \quad \forall (x, y). \end{aligned}$$

We also define  $\text{wprt}_\epsilon(f) = \max_\mu \text{wprt}_\epsilon(f, \mu)$ .

Because we have added a constraint to a maximization problem, it is easy to see that the following holds (see Figure 1b).

**Proposition 1.** *For all  $f, \mu, \epsilon$ ,*

$$\text{wprt}_\epsilon(f, \mu) \leq \text{prt}_\epsilon^+(f, \mu) \leq \text{prt}_\epsilon(f, \mu) \text{ and } \text{wprt}_\epsilon(f, \mu) \leq \overline{\text{prt}}_\epsilon(f, \mu) \leq \text{prt}_\epsilon(f, \mu).$$

In [19], we show the following equivalence:

**Theorem 5.** *Let  $\mu$  be a distribution on  $\mathbf{X} \times \mathbf{Y}$  and  $f$  be a boolean function on the support of  $\mu$  such that either  $\text{rdisc}_\epsilon(f, \mu) \geq 1$  or  $\text{prt}_{4\epsilon}^+(f, \mu) > 2$ . Then for any  $\epsilon \in (0, 1/4)$ ,  $\frac{\epsilon}{2}\text{prt}_{4\epsilon}^+(f, \mu) \leq \text{rdisc}_\epsilon(f, \mu) \leq \text{prt}_\epsilon^+(f, \mu)$ .*

Each inequality is proven by a different change of variables. At a high level,  $\rho_{x,y}$  is proportional to  $\beta_{x,y}$  and  $\delta$  is a scaling factor.

*Revisiting the non-distributional case* For the change of variables in the proof of Theorem 3, we have noted that the constraint  $\beta_{xy} \geq 0$  holds  $\forall(x, y)$  (see Inequality 1). This shows that, in the non-distributional case, relative discrepancy is, in fact, no larger than the weak partition bound, i.e.  $\text{rdisc}_\epsilon(f) \leq \text{wprt}_{\frac{2}{3}\epsilon}(f)$ .

**Lemma 1.** *For any boolean  $f$ , and  $\epsilon \in (0, 1/2)$ ,  $\text{prt}_\epsilon^+(f) \leq \text{wprt}_{\frac{\epsilon}{2}}(f) + \frac{\epsilon}{2}$ .*

*Proof.* Let  $\alpha_{x,y}, \beta_{x,y}$  be a feasible solution for  $\text{prt}^+$ , and consider the following assignment for  $\text{wprt}$ :  $\alpha'_{x,y} = \alpha_{x,y} + \beta_{x,y}$ ,  $\beta'_{x,y} = \beta_{x,y}$ . The constraint on rectangles is still satisfied, and the added positivity constraint  $\alpha'_{x,y} - \beta'_{x,y} = \alpha_{x,y} \geq 0$  is also satisfied. Finally, the objective function for  $\text{wprt}$  with error  $\frac{\epsilon}{2}$  is  $\beta' - \frac{\epsilon}{2}\alpha' = \beta - \frac{\epsilon}{2}\beta - \frac{\epsilon}{2}\alpha \geq \beta - \epsilon\alpha - \frac{\epsilon}{2}$  (where we have used the constraint on  $R = \mathbf{X} \times \mathbf{Y}$ ), as claimed.  $\square$

The change of variables in the proof of Theorem 3 is just the composition of the two changes of variables in Theorem 5 and Lemma 1. It is also now clearer how the distributional and the non-distributional settings differ. It cannot be the case that  $\text{prt}_\epsilon^+(f, \mu) \leq \text{wprt}_\epsilon(f, \mu)$  for fixed distribution, since Ganor *et al.* provide a counterexample. We can also see that for this specific change of variable, by setting  $\alpha'_{x,y} = \alpha_{x,y} + \beta_{x,y}$ ,  $\alpha'_{x,y}$  cannot be written as  $\alpha_{x,y} = \alpha\mu_{x,y}$ , as we would need in the distributional case, since it is a combination of  $\alpha$  and  $\beta$ .

## 5 Adaptive relative discrepancy is equivalent to the public coin partition

In this section, we compare two lower bound techniques for communication complexity introduced recently. We give below a distributional version of the public-coin partition<sup>7</sup>.

**Definition 8 (Public coin partition bound [17]).**

$$\begin{aligned} \text{pprt}_\epsilon(f, \mu) &= \max_{\alpha, \beta} \beta - \epsilon\alpha \\ \text{subject to : } & \beta - \sum_{(R,z) \in P} \alpha\mu(R \cap f^{-1}(z)) \leq |P| \quad \forall P \\ & \alpha \geq 0, \beta \geq 0. \end{aligned}$$

<sup>7</sup> Note that this is a simplified definition with respect to the original one by means of removing redundant variables and constraints in the primal formulation, taking the dual of the resulting expression, and replacing  $\alpha_{x,y}$  by  $\alpha\mu_{x,y}$ , where the distribution  $\mu$  is fixed.

Ganor *et al.* introduced the following notion, which is not a linear program:

**Definition 9 (Adaptive relative discrepancy [14]).**

$$\text{ardisc}_\varepsilon(f, \mu) = \sup_{\kappa, \delta, \rho_{x,y}^P} \frac{1}{\delta} \left( \frac{1}{2} - \kappa - \varepsilon \right) \quad \text{subject to :}$$

$$\left( \frac{1}{2} - \kappa \right) \rho^P(R) \leq \mu(R \cap f^{-1}(z)), \quad \forall P, \forall (z, R) \in P : \rho^P(R) \geq \delta$$

$$0 \leq \kappa < \frac{1}{2}, \quad 0 < \delta < 1, \quad \rho^P = 1, \rho_{x,y}^P \geq 0, \quad \forall P, \forall (x, y).$$

Then  $\text{ardisc}_\varepsilon(f) = \max_\mu \text{ardisc}_\varepsilon(f, \mu)$ .

In [19], we prove the following result :

**Theorem 6.** *For any distribution  $\mu$ , any function  $f : \text{supp}(\mu) \rightarrow \{0, 1\}$  and  $\epsilon \in (0, \frac{1}{4})$  such that either  $\text{ardisc}_\epsilon(f, \mu) \geq 1$  or  $\text{pprt}_{4\epsilon}(f, \mu) > 2$ ,*

$$\frac{\epsilon}{2} \text{pprt}_{4\epsilon}(f, \mu) \leq \text{ardisc}_\epsilon(f, \mu) \leq \text{pprt}_\epsilon(f, \mu).$$

Since the logarithm of the public coin partition bound is polynomially related to randomized communication complexity [17], this tells us that the logarithm of the adaptive relative discrepancy is also polynomially related to communication complexity.

**Corollary 2.** *For any  $\mu$ ,  $f : \text{supp}(\mu) \rightarrow \{0, 1\}$  and  $\epsilon \in (0, \frac{1}{8})$ ,*

$$\log(\text{ardisc}_\epsilon(f, \mu)) \leq R_\epsilon(f, \mu) \leq \left( \log \text{ardisc}_{\epsilon/8}(f, \mu) + 2 \log \frac{1}{\epsilon} + 6 \right)^2.$$

## References

1. M. Braverman and A. Rao, “Information equals amortized communication,” *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6058–6069, 2014.
2. R. M. Fano, “The transmission of information,” Technical Report 65, Research Laboratory for Electronics, MIT, Cambridge, MA, USA, 1949.
3. R. Jain, J. Radhakrishnan, and P. Sen, “A direct sum theorem in communication complexity via message compression,” in *ICALP*, pp. 300–315, 2003.
4. C. E. Shannon, “A Mathematical Theory of Computation,” in *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
5. A. Chakrabarti, A. Wirth, A. Yao, and Y. Shi, “Informational complexity and the direct sum problem for simultaneous message complexity,” *FOCS*, pp. 270–278, 2001.
6. R. Jain, J. Radhakrishnan, and P. Sen, “Optimal direct sum and privacy trade-off results for quantum and classical communication complexity,” *CoRR*, vol. abs/0807.1267, pp. 285–296, 2008.

7. P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, “The communication complexity of correlation,” in *CCC*, pp. 10–23, 2007.
8. Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar, “An information statistics approach to data stream and communication complexity,” *Journal of Computer and System Sciences*, vol. 68, no. 4, pp. 702–732, 2004.
9. B. Barak, M. Braverman, X. Chen, and A. Rao, “How to compress interactive communication,” *STOC*, 2010.
10. M. Braverman, “Interactive information complexity,” in *STOC*, pp. 505–524, 2012.
11. A. C.-C. Yao, “Lower bounds by probabilistic arguments,” in *FoCS*, pp. 420–428, 1983.
12. A. C. C. Yao, “Some complexity questions related to distributive computing (preliminary report),” in *STOC*, pp. 209–213, 1979.
13. A. Ganor, G. Kol, and R. Raz, “Exponential separation of information and communication,” *ECCC*, vol. 21, p. 49, 2014.
14. A. Ganor, G. Kol, and R. Raz, “Exponential Separation of Information and Communication for Boolean Functions,” in *ECCC*, vol. 113, 2014.
15. I. Kerenidis, S. Laplante, V. Lerays, J. Roland, and D. Xiao, “Lower bounds on information complexity via zero-communication protocols and applications,” in *FoCS*, pp. 500–509, 2012.
16. R. Jain and H. Klauck, “The Partition Bound for Classical Communication Complexity and Query Complexity,” in *CCC*, pp. 1–28, 2010.
17. R. Jain, T. Lee, and N. Vishnoi, “A quadratically tight partition bound for classical communication complexity and query complexity,” *CoRR*, vol. abs/1401.4512, 2014.
18. D. Gavinsky and S. Lovett, “En route to the log-rank conjecture: New reductions and equivalent formulations,” in *ICALP*, pp. 514–524, 2014.
19. L. Fontes, R. Jain, I. Kerenidis, S. Laplante, M. Lauriere, and J. Roland, “Relative Discrepancy does not separate Information and Communication Complexity,” <http://eccc.hpi-web.de/report/2015/028/>, 2015.

**Acknowledgements.** We are grateful to Nikos Leonardos for useful discussions and we also thank Virginie Lerays for many fruitful discussions and for the simplification of the public coin partition bound. R.J. would like to thank Anurag Anshu, Prahladh Harsha, Priyanka Mukhopadhyay and Vankatesh Srinivasan for helpful discussions. L.F., I.K., S.L. and M.L. acknowledge support from the French ANR Blanc project RDAM ANR-12-BS02-005, European Union CHIST-ERA grant DIQIP, the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 600700 (QALGO) and the ERC project QCC. J.R. acknowledges support from the Belgian ARC project CPHYMA. The work of R.J. is supported by the Singapore Ministry of Education Tier 3 Grant, the National University of Singapore Young Researcher Award 2012 and the Core Grants of the Center for Quantum Technologies, Singapore. Part of the work done while visiting the Banff International Research Station for Mathematical Innovation and Discovery, Banff, Canada and while visiting the Simon’s Institute, U.C. Berkeley, USA.