Lower bounds on information complexity via zero-communication protocols and applications

Iordanis Kerenidis [*] Sophie Laplante [†] Virginie Lerays [‡] Jérémie Roland [§]
David Xiao [¶]

April 6, 2012

Abstract

We show that almost all known lower bound methods for communication complexity are also lower bounds for the information complexity. In particular, we define a relaxed version of the *partition bound* of Jain and Klauck [JK10] and prove that it lower bounds the information complexity of any function. Our relaxed partition bound subsumes all norm based methods (e.g. the γ_2 method) and rectangle-based methods (e.g. the rectangle/corruption bound, the smooth rectangle bound, and the discrepancy bound), except the partition bound.

Our result uses a new connection between rectangles and *zero-communication* protocols where the players can either output a value or abort. We prove the following compression lemma: given a protocol for a function f with information complexity I, one can construct a zero-communication protocol that has non-abort probability at least $2^{-O(I)}$ and that computes f correctly with high probability conditioned on not aborting. Then, we show how such a zero-communication protocol relates to the relaxed partition bound.

We use our main theorem to resolve three of the open questions raised by Braverman [Bra12]. First, we show that the information complexity of the Vector in Subspace Problem [KR11] is $\Omega(n^{1/3})$, which, in turn, implies that there exists an exponential separation between quantum communication complexity and classical information complexity. Moreover, we provide an $\Omega(n)$ lower bound on the information complexity of the Gap Hamming Distance Problem.

[*]CNRS, LIAFA, Université Paris 7 and CQT, NUS Singapore. jkeren@liafa.univ-paris-diderot.fr
[†]LRI, Université Paris-Sud 11. laplante@lri.fr
[‡]LRI, Université Paris-Sud 11. virginie.lerays@lri.fr
[§]Université Libre de Bruxelles, QuIC, Ecole Polytechnique de Bruxelles. jroland@ulb.ac.be
[¶]CNRS, LIAFA, Université Paris 7. dxiao@liafa.univ-paris-diderot.fr

1 Introduction

Information complexity is a way of measuring the amount of information Alice and Bob must reveal to each other in order to solve a distributed problem. The importance of this notion has been made apparent in recent years through a flurry of results that relate the information complexity of a function and its communication complexity. The first definition of the information cost was given by Chakrabarti, Shi, Wirth and Yao [CSWY01], where it was used to prove a direct sum theorem for simultaneous messages protocols (their notion is now usually called the external information cost). Then, Bar-Yossef et al. [BYJKS04], used the information cost in order to prove a linear lower bound on the two-way randomized communication complexity of Disjointness. More recently, information-theoretic techniques enabled the proof of the first direct sum result for general two-way randomized communication complexity: the randomized communication complexity of n copies of a function f is at least $\sqrt{n}$ times the randomized communication complexity of f [BBCR10]. Then, Braverman and Rao [BR11], showed a tight relation between the amortized distributional communication complexity of a function and its internal information cost. Braverman [Bra12], defined interactive information complexity, a notion which is independent of the prior distribution of the inputs and proved that it is equal to the amortized communication complexity of the function. Braverman and Weinstein [BW11] showed that the information complexity is lower bounded by discrepancy.

The main question pertaining to information complexity is its relation to communication complexity. On the one hand, the information complexity provides a lower bound on the communication complexity of the function, since there cannot be more information leaked than the length of the messages exchanged. However, it is still open, whether the information complexity of a function can be much smaller than its communication complexity or whether the two notions are basically equivalent. In order to make progress towards this question, it is imperative to provide strong lower bounds for information complexity, and more specifically to see whether the lower bound methods for communication complexity can be compared to the model of information complexity.

Lower bound methods in communication complexity can be seen to fall into three main categories: the norm based methods, such as the γ_2 method of Linial and Shraibman [LS09b] (see Lee and Shraibman's survey for an overview [LS09a]); the rectangle based methods, such as discrepancy and the rectangle bound; and, of course, the information theoretic methods, among which, information complexity. Recently, Jain and Klauck [JK10] introduced the smooth rectangle bound, as well as the stronger partition bound, and showed that they subsume both γ_2 and the rectangle bound [JK10].

The first lower bound on information complexity was proved by Braverman [Bra12], who showed that it is lower bounded by the logarithm of the communication complexity. Recently, Braverman and Weinstein showed that the discrepancy method lower bounds the information complexity [BW11]. Their result follows from a compression lemma for protocols: a protocol for a function f that leaks I bits of information implies the existence of a protocol with communication complexity $O(I)$ and advantage on computing f (over a random guess) of $2^{-O(I)}$.

1.1 Our results

In this paper, we show that all known lower bound methods for communication complexity, with the notable exception of the partition bound, generalize to information complexity. More precisely, we introduce the *relaxed partition bound* (in Definition 3.2) denoted by $\bar{\mathsf{prt}}^{\mu}_{\epsilon}(f)$, which depends on the function to be computed f, the input distribution μ, and the error parameter ϵ, and such that the distributional communication complexity $D^{\mu}_{\epsilon}(f) \geq \log(\bar{\mathsf{prt}}^{\mu}_{\epsilon}(f))$ for any f. We prove that the information complexity of a function f is bounded below by the relaxed partition bound:

Theorem 1.1. *There is a positive constant C such that for all functions $f : \mathcal{I} \to \mathcal{Z}$, all $\epsilon, \delta \in (0, \frac{1}{2}]$, and all distributions μ, we have* $\mathsf{IC}_{\mu}(f, \epsilon) \geq \frac{\delta^2}{C} \cdot \left(\log \bar{\mathsf{prt}}^{\mu}_{\epsilon+3\delta}(f) - \log |\mathcal{Z}|\right) - \delta$.

Since we show in Lemma 3.3 that the relaxed partition bound subsumes the norm based methods (e.g.

the γ_2 method) and the rectangle-based methods (e.g. the rectangle/corruption bound, the smooth rectangle bound, and the discrepancy bound), all of these bounds are also lower bounds on the information complexity.

Technique. The key idea of our result is a new connection between communication rectangles and zero-communication protocols, where the players can either output a value or abort but *without communicating*. A priori, it is surprising that protocols with no communication can actually provide some insight on the communication or information complexity of a function. However, this model, which has been extensively used in quantum information for the study of non-local games and Bell inequalities, turns out to be a very powerful tool for the study of classical communication and information complexity. The communication complexity of simulating distributions is known to be related to the probability of not aborting in zero-communication protocols that can abort [GG99, Mas02, BHMR03, BHMR06]. More recently connections have been shown for specific lower bound methods. It has been shown that zero-communication protocols with error give rise to the factorization norm method [DKLR11], and the connection between the partition bound and zero-communication protocols with abort was studied in [LLR12].

In a deterministic zero-communication protocol with abort, each of the two players looks at their input and decides either to abort the protocol or to output some value z. The output of the protocol is z if both players agree on z, or it aborts otherwise. It is easy to see that for any deterministic zero-communication protocol with abort, the set of inputs where both players choose to output z forms a rectangle, and so the protocol is characterized by a set of rectangles each labeled by an output. In a randomized protocol, we have instead a distribution over labeled rectangles.

This connection between rectangles and zero-communication protocols with abort allows us to obtain our lower bound for information complexity from a new compression lemma for protocols (Lemma 3.4): a protocol for a function f that leaks I bits of information implies the existence of a *zero*-communication protocol that has non-abort probability at least $2^{-O(I)}$ and that computes f correctly with high probability when not aborting. Our main theorem follows from this new compression.

The technical tools we use are drawn from Braverman [Bra12] and in particular Braverman and Weinstein [BW11]. We describe the difference between our compression and that of [BW11]. There, they take a protocol for computing a function f that has information cost I and compress it to a protocol with communication $O(I)$ and advantage of computing f of $2^{-O(I)}$ (*i.e.* the error increases considerably). Then, they apply the discrepancy method, which can handle such small advantage.

In our compression, we suppress the communication entirely, and, moreover, we only introduce an arbitrarily small error since the compressed protocol aborts when it does not believe it can correctly compute the output. This compression enables us to provide much sharper lower bounds on the information complexity and in particular, the lower bound in terms of the relaxed partition bound.

Applications. Our lower bound implies that for most functions for which there exists a lower bound on their communication complexity, the same bound extends to their information complexity. Specifically, we can apply our lower bound in order to resolve three of the open questions in [Bra12].

First, we show that there exists a function f, such that the quantum communication complexity of f is exponentially smaller than the information complexity of f (Open Problem 3 in [Bra12]).

Theorem 1.2. *There exists a function f, s.t. for all $\epsilon \in (0, \frac{1}{2}), Q(f, \epsilon) = O(\log(\mathsf{IC}(f, \epsilon)))$.*

In order to prove the above separation, we show that the proof of the lower bound on the randomized communication complexity of the Vector in Subspace Problem ($\widetilde{\mathrm{VSP}}$) [KR11] provides, in fact, a lower bound on the relaxed partition bound. By our lower bound, this implies that $\mathsf{IC}(\widetilde{\mathrm{VSP}}_{\theta,n}, 1/3) = n^{1/3}$ (Open Problem 7 in [Bra12]). Since the quantum communication complexity of $\widetilde{\mathrm{VSP}}_{\theta,n}$ is $O(\log n)$, we have the above theorem. Moreover, this implies an exponential separation between classical and quantum information complexity. We refrain from defining quantum information cost in this paper (see [JN10] for a definition),

but since the quantum information cost is always smaller than the quantum communication complexity, the separation follows trivially from the above theorem.

In addition, we resolve the question of the information complexity of the Gap Hamming Distance Problem (GHD) (Open Problem 6 in [Bra12]), since the lower bounds on the randomized communication complexity of this problem go through the rectangle/corruption bound [She11] or smooth rectangle bound [CR11, Vid11].

Theorem 1.3. $\mathsf{IC}(\mathrm{GHD}_n, 1/3) = \Omega(n)$.

Finally, regarding the central open question of whether or not it is possible to compress communication down to the information complexity, we note that our result says that if one hopes to prove a negative result and separate information complexity from communication complexity, then one must use a lower bound technique that is stronger than the relaxed partition bound. To the best of our knowledge, the only such technique in the literature is the (standard) partition bound. We note, however, that to the best of our knowledge there are no known problems whose communication complexity can be lower-bounded by the partition bound but not by the relaxed partition bound.

2 Preliminaries

2.1 Notation and information theory facts

Let μ be a probability distribution over a (finite) universe $\mathcal{U}$. We will often treat μ as a function $\mu : 2^{\mathcal{U}} \to [0,1]$. For $T, S \subseteq \mathcal{U}$, we let $\mu(T \mid S) = \Pr_{u \leftarrow \mu}[u \in T \mid S]$. For singletons $u \in U$, we write interchangeably $\mu_u = \mu(u) = \mu(\{u\})$. Random variables are written in uppercase and fixed values in lowercase. We sometimes abuse notation and write a random variable in place of the distribution followed by that random variable.

For two distributions μ, ν, we let $|\mu - \nu|$ denote their statistical distance, *i.e.* $|\mu - \nu| = \max_{T \subseteq \mathcal{U}}(\mu(T) - \nu(T))$. We let $D(\mu \parallel \nu) = \mathbb{E}_{U \sim \mu}[\log \frac{\mu(U)}{\nu(U)}]$ be the relative entropy (*i.e.* KL-divergence). For two random variables X, Y, the mutual information is defined as $I(X : Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$, where $H(\cdot)$ is the Shannon entropy.

A *rectangle* of $\mathcal{X} \times \mathcal{Y}$ is a product set $A \times B$ where $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$. We let R denote a rectangle in $\mathcal{X} \times \mathcal{Y}$. We let $(x, y) \in \mathcal{X} \times \mathcal{Y}$ denote a fixed input, and (X, Y) be random inputs sampled according to some distribution (specified from context and usually denoted by μ).

2.2 Information complexity

We study 2-player communication protocols for calculating a function $f : \mathcal{I} \to \mathcal{Z}$, where $\mathcal{I} \subseteq \mathcal{X} \times \mathcal{Y}$. Let π be a randomized protocol (allowing both public and private coins, unless otherwise specified). We denote the randomness used by the protocol π by r_π. Let $\pi(x, y)$ denote its output, *i.e.* the value in $\mathcal{Z}$ the two parties wish to compute.

The transcript of a protocol includes all messages exchanged, the output of the protocol (in fact we just need that both players can compute the output of the protocol from the transcript), as well as any public coins (but no private coins). The complexity of π is the maximum (over all random coins) of the number of bits exchanged.

Let μ be a distribution over $\mathcal{X} \times \mathcal{Y}$. Define $\mathsf{err}_f(\pi; x, y) = \Pr_{r_\pi}[f(x, y) \neq \pi(x, y)]$ if $(x, y) \in \mathcal{I}$ and 0 otherwise and $\mathsf{err}_f(\pi; \mu) = \mathbb{E}_{(X,Y) \sim \mu} \mathsf{err}_f(\pi; X, Y) = \Pr_{r_\pi, (X,Y) \sim \mu}[(X, Y) \in \mathcal{I} \wedge f(X, Y) \neq \pi(X, Y)]$.

Definition 2.1. Fix f, μ, ϵ. Let (X, Y, Π) be the tuple distributed according to (X, Y) sampled from μ and then Π being the transcript of the protocol π applied to X, Y. Then define:

1. $\mathsf{IC}_\mu(\pi) = I(X; \Pi \mid Y) + I(Y; \Pi \mid X)$
2. $\mathsf{IC}_\mu(f, \epsilon) = \inf_{\pi : \mathsf{err}_f(\pi; \mu) \leq \epsilon} \mathsf{IC}_\mu(\pi)$

3. $\mathsf{IC}_D(f,\epsilon) = \max_\mu \mathsf{IC}_\mu(f,\epsilon)$

Braverman [Bra12] also defined the non-distributional information cost IC, and all of our results extend to it trivially by the inequality $\mathsf{IC}_D \leq \mathsf{IC}$. (We do not require the reverse inequality $\mathsf{IC} \leq O(\mathsf{IC}_D)$, whose proof is non-trivial and was given in [Bra12]).

3 Zero-communication protocols and the relaxed partition bound

3.1 The zero-communication model and rectangles

Let us consider a (possibly partial) function f. We say that (x,y) is a valid input if $(x,y) \in \mathcal{I}$, that is, (x,y) satisfies the promise. In the zero-communication model with abort, the players either output a value $z \in \mathcal{Z}$ (they *accept* the run) or output $\perp$ (they *abort*).

Definition 3.1. The *zero-communication* model with abort is defined as follows:

Inputs Alice and Bob receive inputs x and y respectively.
Output Alice outputs $a \in \mathcal{Z} \cup \{\perp\}$ and Bob outputs $b \in \mathcal{Z} \cup \{\perp\}$. If both Alice and Bob output the same $z \in \mathcal{Z}$, then the output is z. Otherwise, the output is $\perp$.

We will study (public-coin) randomized *zero-communication* protocols for computing functions in this model.

3.2 Relaxed partition bound

The relaxed partition bound with error ϵ and input distribution μ, denoted by $\bar{\mathsf{prt}}^\mu_\epsilon(f)$, is defined as follows.

Definition 3.2. The distributional relaxed partition bound $\bar{\mathsf{prt}}^\mu_\epsilon(f)$ is the value of the following linear program. (The values of z ranges over $\mathcal{Z}$ and R over all rectangles, including the empty rectangle.)

$$\bar{\mathsf{prt}}^\mu_\epsilon(f) = \min_{\eta, p_{R,z} \geq 0} \frac{1}{\eta} \quad \text{subject to:}$$

$$\sum_{(x,y)\in\mathcal{I}} \mu_{x,y} \sum_{R:(x,y)\in R} p_{R,f(x,y)} + \sum_{(x,y)\notin\mathcal{I}} \mu_{x,y} \sum_{z,R:(x,y)\in R} p_{R,z} \geq (1-\epsilon)\eta \tag{1}$$

$$\forall (x,y) \in \mathcal{X}\times\mathcal{Y}, \quad \sum_{z,R:(x,y)\in R} p_{R,z} \leq \eta \tag{2}$$

$$\sum_{R,z} p_{R,z} = 1, \tag{3}$$

The relaxed partition bound is defined as $\bar{\mathsf{prt}}_\epsilon(f) = \max_\mu \bar{\mathsf{prt}}^\mu_\epsilon(f)$.

We can identify feasible solutions to the program in Definition 3.2 as a particular type of randomized zero-communication protocol: Alice and Bob sample (R,z) according to the distribution given by the $p_{R,z}$, and each individually sees if their inputs are in R and if so they output z, otherwise they abort. The parameter η is the *efficiency* of the protocol [LLR12], that is, the probability that the protocol does not abort, and ideally we want it to be as large as possible.

There is also a natural way to convert any zero-communication protocol π into a distribution over (R,z): sample z uniformly from $\mathcal{Z}$, sample random coins r_π for π, and let $R = A \times B$ be such that A is the set of inputs on which Alice outputs z in the protocol π using random coins r_π, and similarly for B. (The sampling of a random z incurs a loss of $|\mathcal{Z}|$ in the efficiency, which is why our bounds have a loss depending on $|\mathcal{Z}|$. See Section 3.4 for details.)

Relation to other bounds. The relaxed partition bound is, as its name implies, a relaxation of the partition bound $\mathsf{prt}_\epsilon(f)$ [JK10]. We also show that the relaxed partition bound is stronger than the smooth rectangle bound $\mathsf{srec}^z_\epsilon(f)$ (the proof is provided in Appendix A).

Lemma 3.3. *For all f, ϵ and $z \in \mathcal{Z}$, we have $\mathsf{srec}^z_\epsilon(f) \leq \bar{\mathsf{prt}}_\epsilon(f) \leq \mathsf{prt}_\epsilon(f)$.*

Since Jain and Klauck have shown in [JK10] that the smooth rectangle bound is stronger than the rectangle/corruption bound, the γ_2 method and the discrepancy method, this implies that the relaxed partition bound subsumes all these bounds as well. Therefore, our result implies that all these bounds are also lower bounds for information complexity.

We briefly explain the difference between the relaxed partition bound and the partition bound (more details appear in Appendix A). The partition bound includes two types of constraints. The first is a correctness constraint: on every input, the output of the protocol should be correct with probability at least $(1-\epsilon)\eta$. The second is a completeness constraint: on every input, the efficiency of the protocol (*i.e.* the probability it does not abort) should be *exactly* η. In the relaxed partition bound, we keep the same correctness constraint. Since in certain applications the function is partial (such as the Vector in Subspace Problem [KR11]), one also has to handle the inputs where the function is not defined. We make this explicit in our correctness constraint. On the other hand, we relax the completeness constraint so that the efficiency may lie anywhere between $(1-\epsilon)\eta$ and η. This relaxation seems to be crucial for our proof of the lower bound on information complexity, since we are unable to achieve efficiency exactly η.

3.3 Compression lemma

Lemma 3.4 (Main compression lemma)**.** *There exists a universal constant C such that for all distributions μ, communication protocols π and $\delta \in (0,1)$, there exists a zero-communication protocol π' and a real number $\lambda \geq 2^{-C(\mathsf{IC}_\mu(\pi)/\delta^2+1/\delta)}$ such that*

$$\left|(X,Y,\pi(X,Y)) - (X,Y,\pi'(X,Y)|\pi'(X,Y) \neq \perp)\right| \leq \delta \tag{4}$$

(in statistical distance) and

$$\forall (x,y) \quad \Pr_{r_{\pi'}}[\pi'(x,y) \neq \perp] \leq (1+\delta)\lambda \tag{5}$$

$$\Pr_{r_{\pi'},(X,Y)\sim\mu}[\pi'(X,Y) \neq \perp] \geq (1-\delta)\lambda \tag{6}$$

Our compression π' extends the strategy outlined by [BW11]. At a high level, the protocol π' does the following:

Sample transcripts Alice and Bob use their shared randomness to repeat T independent executions of an experiment to sample transcripts (Algorithm 4.1). Alice and Bob each decide whether the experiment is accepted (they may differ in their opinions).

Find common transcript Let $\mathcal{A}$ be the set of accepted experiments for Alice, and $\mathcal{B}$ the set of accepted experiments for Bob. They try to guess an element of $\mathcal{A} \cap \mathcal{B}$. If they find one, they output according to the transcript from this experiment.

We prove our compression lemma in Section 4.

3.4 Information cost is lower bounded by the relaxed partition bound

We show how our compression lemma implies the main theorem.

Proof of Theorem 1.1. Let π be a randomized communication protocol achieving $\mathsf{IC}_\mu(f,\epsilon)$ and let $\mathcal{R}$ be the following relation that naturally arises from the function f

$$\mathcal{R} = \{(x,y,f(x,y)) : (x,y) \in \mathcal{I}\} \cup \{(x,y,z) : (x,y) \notin \mathcal{I}, z \in \mathcal{Z}\}.$$

Let us now consider the zero-communication protocol π' from Lemma 3.4. As mentioned in Section 3.2, there is a natural way to identify π' with a distribution over labeled rectangles (R, z): sample z uniformly from $\mathcal{Z}$, sample r_π and let $R = A \times B$ where A is the set of inputs on which Alice outputs z, and similarly for B. The sampling of z incurs a loss of $|\mathcal{Z}|$ in the efficiency.

We make this formal: for any fixed randomness r occurring with probability p_r, we define the rectangle $R(z, r)$ as the set of (x, y) such that the protocol outputs z, and we let $p_{R,z} = \sum_{r:R=R(z,r)} p_r/|\mathcal{Z}|$.

We check the normalization constraint

$$\sum_{R,z} p_{R,z} = \frac{1}{|\mathcal{Z}|} \sum_{R,z} \sum_{r:R=R(z,r)} p_r = \frac{1}{|\mathcal{Z}|} \sum_r p_r \sum_{R,z:R=R(z,r)} 1 = \sum_r p_r = 1.$$

To see that Equation 2 is satisfied, we have by definition of $p_{R,z}$ that for any (x, y):

$$\sum_{z,R:(x,y)\in R} p_{R,z} = \frac{1}{|\mathcal{Z}|} \Pr_{r_{\pi'}}[\pi'(x, y) \neq \perp] \leq \frac{(1+\delta)\lambda}{|\mathcal{Z}|}$$

Finally, to see that Equation 1 is satisfied, we have

$$\begin{aligned}
&\sum_{(x,y)\in\mathcal{I}} \mu_{x,y} \sum_{R:(x,y)\in R} p_{R,f(x,y)} + \sum_{(x,y)\notin\mathcal{I}} \mu_{x,y} \sum_{z,R:(x,y)\in R} p_{R,z} \\
&= \frac{1}{|\mathcal{Z}|} \Pr_{r_{\pi'},(X,Y)\sim\mu}[(X, Y, \pi'(X, Y)) \in \mathcal{R}] \\
&= \frac{1}{|\mathcal{Z}|} \Pr_{r_{\pi'},(X,Y)\sim\mu}[\pi'(X, Y) \neq \perp] \Pr_{r_{\pi'},(X,Y)\sim\mu}[(X, Y, \pi'(X, Y)) \in \mathcal{R} \mid \pi'(X, Y) \neq \perp] \\
&\geq \frac{1}{|\mathcal{Z}|}(1-\delta)\,\lambda \left(\Pr_{r_{\pi'},(X,Y)\sim\mu}[(X, Y, \pi(X, Y)) \in \mathcal{R}] - \delta \right) \\
&\geq \frac{1}{|\mathcal{Z}|}(1-\delta)\,\lambda\,(1-\epsilon-\delta) \quad \geq \quad \tfrac{1}{|\mathcal{Z}|}\,\lambda\,(1-\epsilon-2\delta) \quad \geq \quad \tfrac{\lambda(1+\delta)}{|\mathcal{Z}|}(1-\epsilon-3\delta)
\end{aligned}$$

where for the last line we used the fact that π has error ϵ, and so $\Pr_{r_\pi,(X,Y)\sim\mu}[(X, Y, \pi(X, Y)) \in \mathcal{R}] \geq 1-\epsilon$. This satisfies the constraints in the linear program (Definition 3.2) for $\bar{\mathsf{prt}}^{\mu}_{\epsilon+3\delta}(f)$ with objective value $\eta = (1+\delta)\lambda/|\mathcal{Z}| \geq 2^{-C(\mathsf{IC}_\mu(\pi)/\delta^2+1/\delta)}/|\mathcal{Z}|$. □

By the definitions of the information complexity and the relaxed partition bound, we have immediately

Corollary 3.5. *There exists a universal constant C such that for all functions $f : \mathcal{I} \to \mathcal{Z}$, all $\epsilon, \delta \in (0, 1/2)$, we have $\mathsf{IC}_D(f, \epsilon) \geq \frac{\delta^2}{C}[\log \bar{\mathsf{prt}}_{\epsilon+3\delta}(f) - \log|\mathcal{Z}|] - \delta$.*

4 The zero-communication protocol

The zero-communication protocol consists of two stages. First, Alice and Bob use their shared randomness to come up with candidate transcripts, based on the a priori information they have on the distribution of the transcripts given by the information cost of the protocol. To do this, they run some sampling experiments and decide which ones to accept. Second, they use their shared randomness in order to choose an experiment that they have both accepted. If anything fails in the course of the protocol, they abort by outputting $\perp$.

4.1 Single sampling experiment

The single sampling experiment is described in Algorithm 4.1 and appeared first in [BW11] (variants also appeared in [Bra12] and [BR11]). Roughly, Algorithm 4.1 takes a distribution τ and two distributions $\nu_{\mathsf{A}}, \nu_{\mathsf{B}}$

Fix a finite universe $\mathcal{U}$. Let $p_{\mathsf{A}}, q_{\mathsf{A}}, p_{\mathsf{B}}, q_{\mathsf{B}} : \mathcal{U} \to [0,1]$ such that $\tau = p_{\mathsf{A}} p_{\mathsf{B}}$, $\nu_{\mathsf{A}} = p_{\mathsf{A}} q_{\mathsf{A}}$, $\nu_{\mathsf{B}} = p_{\mathsf{B}} q_{\mathsf{B}}$ are all probability distributions.
Alice's input: $p_{\mathsf{A}}, q_{\mathsf{A}}$. Bob's input: $p_{\mathsf{B}}, q_{\mathsf{B}}$. Common input: parameter $\Delta > 0$.

1. Using public coins, sample $u \leftarrow \mathcal{U}$, $\alpha, \beta \leftarrow [0, 2^{\Delta}]$.
2. Alice accepts the run if $\alpha \le p_{\mathsf{A}}(u)$ and $\beta \le 2^{\Delta} q_{\mathsf{A}}(u)$.
3. Bob accepts the run if $\alpha \le 2^{\Delta} q_{\mathsf{B}}(u)$ and $\beta \le p_{\mathsf{B}}(u)$.
4. If both Alice and Bob accept, then we say that the experiment is accepted and the output is u. Otherwise, the output is $\perp$.

Algorithm 4.1. Single sampling experiment

over a universe $\mathcal{U}$ such that $\nu_{\mathsf{A}}, \nu_{\mathsf{B}}$ are not too far from τ and tries to sample an element of $\mathcal{U}$ that is close to being distributed according to τ.

Let us informally describe the goal of this sampling experiment in our context. Alice knowing x and Bob y want to sample transcripts according to $\Pi_{x,y}$ which is the distribution over the transcripts of the protocol π applied to (x, y). When inputs x, y are fixed, the probability of a transcript u occurring is the product of the probabilities of each bit in the transcript. The product of the probabilities for Alice's bits is some function $p_{\mathsf{A}}(u)$ which depends on x and the product of the probabilities for Bob's bits is some function $p_{\mathsf{B}}(u)$ which depends on y and $\Pi_{x,y}(u) = p_{\mathsf{A}}(u) p_{\mathsf{B}}(u)$. Alice can also estimate $p_{\mathsf{B}}(u)$ by taking the average over y of $\Pi_y(u)$. Call this estimate $q_{\mathsf{A}}(u)$; similarly for Bob's estimate $q_{\mathsf{B}}(u)$. Set $\nu_{\mathsf{A}} = p_{\mathsf{A}} q_{\mathsf{A}}$ and $\nu_{\mathsf{B}} = q_{\mathsf{B}} p_{\mathsf{B}}$.

The challenge is that Alice and Bob know only $(p_{\mathsf{A}}, q_{\mathsf{A}})$ and $(p_{\mathsf{B}}, q_{\mathsf{B}})$ respectively and do not know τ (in our setting, $\tau = \Pi_{x,y}$). They use a variant of rejection sampling, in which Alice will overestimate q_{A} by a factor 2^{Δ}; likewise for Bob. Let us define the set of Δ-bad elements with respect to τ, ν as follows:

$$B_{\Delta}(\tau, \nu) = \{u \in \mathcal{U} \mid 2^{\Delta} \nu(u) < \tau(u)\}$$

Intuitively, u is bad if τ gives much more weight to it than ν. Observe that if $\tau = p_{\mathsf{A}} p_{\mathsf{B}}, \nu_{\mathsf{A}} = p_{\mathsf{A}} q_{\mathsf{A}}$, then $u \notin B_{\Delta}(\tau, \nu_{\mathsf{A}})$ implies that $2^{\Delta} q_{\mathsf{A}}(u) \ge p_{\mathsf{B}}(u)$.

To prove our compression lemma, we use the following claim about the single sampling experiment.

Claim 4.2. *Let $B = B_{\Delta}(\tau, \nu_{\mathsf{A}}) \cup B_{\Delta}(\tau, \nu_{\mathsf{B}})$. Let $\gamma = \tau(B)$. Then the following holds about Algorithm 4.1:*

1. *The probability that Alice accepts equals $\frac{1}{|\mathcal{U}| 2^{\Delta}}$ and the same for Bob.*
2. *The probability that the experiment is accepted is at most $\frac{1}{|\mathcal{U}| 2^{2\Delta}}$ and at least $\frac{1-\gamma}{|\mathcal{U}| 2^{2\Delta}}$.*
3. *Let τ' denote the distribution of the output of the experiment, conditioned on it being accepted. Then $|\tau - \tau'| \le \gamma$.*

Intuitively, this claim says that Alice accepts each single experiment with probability $\frac{1}{|\mathcal{U}| 2^{\Delta}}$, and also implies that conditioned on Alice accepting the i'th experiment, it is relatively likely that Bob accepts it. Therefore, by repeating this experiment enough times, there is reasonable probability of Alice and Bob both accepting the same execution of the experiment. Conditioned on the experiment accepting, the output of the experiment is distributed close to the original distribution τ. In the next section, we show how to use a hash function to select a common accepting execution of the experiment out of many executions.

We will use the following lemma that appears in [Bra12].

Lemma 4.3 ([Bra12])**.** *For all $\tau, \nu, \Delta, \epsilon$, it holds that $\tau(B_{\Delta}(\tau, \nu)) \le \frac{D(\tau \| \nu) + 1}{\Delta}$.*

Proof of Claim 4.2. We use the arguments first given in [BW11] and prove the items in order.

1. **Probability Alice/Bob accepts.** We do the analysis for Alice; the case for Bob is entirely symmetric. We may write:

$$\Pr[\text{Alice accepts}] = \sum_{u \in \mathcal{U}} \frac{1}{|\mathcal{U}|} \frac{p_{\mathsf{A}}(u)}{2^{\Delta}} q_{\mathsf{A}}(u) = \frac{1}{|\mathcal{U}| 2^{\Delta}} \sum_{u \in \mathcal{U}} \nu_{\mathsf{A}}(u) = \frac{1}{|\mathcal{U}| 2^{\Delta}}$$

2. **Probability of accepting.** First consider $u \notin B$. For such u, if $\alpha \leq p_{\mathsf{A}}(u)$ then $\alpha \leq 2^{\Delta} q_{\mathsf{B}}(u)$ and also if $\beta \leq p_{\mathsf{B}}(u)$ then $\beta \leq 2^{\Delta} q_{\mathsf{A}}(u)$. Therefore we may write

$$\Pr[\text{Experiment outputs } u] = \frac{1}{|\mathcal{U}|} \frac{p_{\mathsf{A}}(u) p_{\mathsf{B}}(u)}{2^{2\Delta}} = \frac{\tau(u)}{|\mathcal{U}| 2^{2\Delta}} \tag{7}$$

Furthermore, for any $u \in \mathcal{U}$, we may write

$$\Pr[\text{Experiment outputs } u] = \frac{1}{|\mathcal{U}| 2^{2\Delta}} \cdot \min\left\{p_{\mathsf{A}}(u), q_{\mathsf{B}}(u) 2^{\Delta}\right\} \cdot \min\left\{p_{\mathsf{B}}(u), q_{\mathsf{A}}(u) 2^{\Delta}\right\} \leq \frac{\tau(u)}{|\mathcal{U}| 2^{2\Delta}}$$

For the upper bound we have:

$$\Pr[\exists u \in \mathcal{U}, \text{ Experiment outputs } u] = \sum_{u \in \mathcal{U}} \Pr[\text{Experiment outputs } u] \leq \sum_{u \in \mathcal{U}} \frac{\tau(u)}{|\mathcal{U}| 2^{2\Delta}} = \frac{1}{|\mathcal{U}| 2^{2\Delta}}$$

For the lower bound we have

$$\begin{aligned} \sum_{u \in \mathcal{U}} \Pr[\text{Experiment outputs } u] &\geq \sum_{u \notin B} \Pr[\text{Experiment outputs } u] \\ &= \sum_{u \notin B} \frac{\tau(u)}{|\mathcal{U}| 2^{2\Delta}} \quad = \quad \frac{1 - \tau(B)}{|\mathcal{U}| 2^{2\Delta}} \quad = \quad \frac{1 - \gamma}{|\mathcal{U}| 2^{2\Delta}} \end{aligned}$$

3. **Statistical closeness of τ and τ'.** Let η denote the probability that the experiment is accepted. From the previous point, we have that $\eta \in [\frac{1-\gamma}{|\mathcal{U}| 2^{2\Delta}}, \frac{1}{|\mathcal{U}| 2^{2\Delta}}]$. By the definition of statistical distance, it suffices to prove that:

$$\forall S \subseteq \mathcal{U}, \quad \tau(S) - \tau'(S) \leq \gamma$$

We proceed by splitting the elements of S based on whether they intersect B.

$$\begin{aligned} \tau(S) - \tau'(S) &= \tau(S \cap B) - \tau'(S \cap B) + \tau(S \cap \overline{B}) - \tau'(S \cap \overline{B}) \\ &\leq \gamma + \tau(S \cap \overline{B}) - \tau'(S \cap \overline{B}) \end{aligned}$$

From Equation 7 we can deduce that $\tau'(S \cap \overline{B}) = \frac{\tau(S \cap \overline{B})}{|\mathcal{U}| 2^{2\Delta} \eta}$. Therefore:

$$\tau(S \cap \overline{B}) - \tau'(S \cap \overline{B}) = \tau(S \cap \overline{B})(1 - \tfrac{1}{|\mathcal{U}| 2^{2\Delta} \eta})$$

Since $\eta \leq \frac{1}{|\mathcal{U}| 2^{2\Delta}}$, we have that $(1 - \frac{1}{|\mathcal{U}| 2^{2\Delta} \eta}) \leq 0$, which concludes the proof.

□

Alice's input: x. Bob's input: y. Common inputs: $\delta > 0, I > 0$.
Set parameters: $\Delta = \frac{4}{\delta} \cdot (\frac{8 \cdot I}{\delta} + 1)$ and $T = |\mathcal{U}| 2^{\Delta} \ln(8/\delta)$ and $k = \Delta + \log(\frac{64}{\delta} \ln(8/\delta)^2)$.

1. Alice constructs functions $p_{\mathsf{A}}, q_{\mathsf{A}} : \mathcal{U} \to [0,1]$ and Bob constructs functions $p_{\mathsf{B}}, q_{\mathsf{B}} : \mathcal{U} \to [0,1]$, such that for all transcripts $u \in \mathcal{U}$, $\Pi_{x,y}(u) = p_{\mathsf{A}}(u) p_{\mathsf{B}}(u)$, $\Pi_x(u) = p_{\mathsf{A}}(u) q_{\mathsf{A}}(u)$, and $\Pi_y(u) = p_{\mathsf{B}}(u) q_{\mathsf{B}}(u)$.
2. (**Run experiments.**) Using public coins, Alice and Bob run Algorithm 4.1 T independent times with inputs $p_{\mathsf{A}}, q_{\mathsf{A}}, p_{\mathsf{B}}, q_{\mathsf{B}}$ and Δ.
3. Let $\mathcal{A} = \{i \in [T] : \text{Alice accepts experiment } i\}$ and similarly $\mathcal{B}$ for Bob. If either set is empty, that party outputs the abort symbol $\perp$.
4. (**Find intersection.**) Using public coins, Alice and Bob choose a random function $h : [T] \to \{0,1\}^k$ and a random string $r \in \{0,1\}^k$.
   (a) Alice finds the smallest $i \in \mathcal{A}$. If $h(i) \neq r$ then Alice outputs $\perp$. Otherwise, Alice outputs in accordance with the transcript of experiment i.
   (b) Bob finds the smallest $j \in \mathcal{B}$ such that $h(j) = r$. If no such j exists, he outputs $\perp$. Otherwise, Bob outputs in accordance with the transcript of experiment j.

Algorithm 4.4. Zero-communication protocol derived from π

4.2 Description and analysis of the zero-communication protocol

Let μ be any distribution on inputs and π be any protocol with information complexity $I = \mathsf{IC}_{\mu}(\pi)$. Let (X, Y, Π) be the joint random variables where X, Y are distributed according to μ and Π is the distribution of the transcript of the protocol π applied to X, Y (by slight abuse of notation we use the letter Π for both the transcript and its distribution). Let $\Pi_{x,y}$ be Π conditioned on $X = x, Y = y$, Π_x be Π conditioned $X = x$, and Π_y likewise.

Let $\mathcal{U}$ be the space of all possible transcripts. We assume that each transcript contains the output of the protocol. As shown in [Bra12] and described above, Alice can construct functions $p_{\mathsf{A}}, q_{\mathsf{A}} : \mathcal{U} \to [0,1]$ and Bob can construct functions $p_{\mathsf{B}}, q_{\mathsf{B}} : \mathcal{U} \to [0,1]$, such that for all $u \in \mathcal{U}$, $\Pi_{x,y}(u) = p_{\mathsf{A}}(u) p_{\mathsf{B}}(u)$, $\Pi_x(u) = p_{\mathsf{A}}(u) q_{\mathsf{A}}(u)$, and $\Pi_y(u) = p_{\mathsf{B}}(u) q_{\mathsf{B}}(u)$.

The zero-communication protocol π' is described in Algorithm 4.4. This protocol is an extension of the one in [BW11], where here Alice uses public coins to guess the hash function value instead of calculating and transmitting it to Bob and both players are allowed to abort when they do not believe they can output the correct value.

In order to analyze our protocol, we first define some events and give bounds on their probabilities.

Definition 4.5. We define the following events over the probability space of sampling (X, Y) according to μ and running π' on (X, Y) to produce a transcript Π:

1. **Large divergence.** B_D occurs if $(X, Y) = (x, y)$ such that $D(\Pi_{x,y} \parallel \Pi_x) > \frac{8 \mathsf{IC}_{\mu}(\pi)}{\delta}$ or $D(\Pi_{x,y} \parallel \Pi_y) > \frac{8 \mathsf{IC}_{\mu}(\pi)}{\delta}$. We will also let B_D denote the set of such (x, y).
2. **Collision.** B_C occurs if there exist distinct $i, j \in \mathcal{A} \cup \mathcal{B}$ such that $h(i) = h(j) = r$.
3. **Protocol outputs something.** H occurs if $\pi'(X, Y) \neq \perp$.

The proof of the main compression lemma (Lemma 3.4) uses the following claim.

Claim 4.6. *The probability of the above events are bounded as follows:*

1. *The inputs rarely have large divergence:* $\Pr_{(X,Y) \sim \mu}[B_D] \leq \delta/4$.
2. *For all (x, y), the hash function rarely has a collision:* $\Pr_{r_{\pi'}}[B_C] \leq \frac{\delta}{16} \cdot 2^{-(k+\Delta)}$.

3. *For all* $(x,y) \notin B_D$*, the probability of outputting something is not too small:* $\Pr_{r_{\pi'}}[H] \geq (1 - \frac{11\delta}{16})2^{-(k+\Delta)}$.
4. *For all* (x,y) *the probability of outputting something is not too large:* $\Pr_{r_{\pi'}}[H] \leq (1 + \frac{\delta}{16})2^{-(k+\Delta)}$.
5. *For all protocols* π*, input distributions* μ *and* $\delta > 0$*, the algorithm* π' *in Algorithm 4.4 satisfies: For all* $(x,y) \notin B_D$*, let* $\Pi'_{x,y,H}$ *be the distribution of* $\pi'(x,y)$ *conditioned on* H *(namely, on* $\pi'(x,y) \neq \perp$*). Then* $|\Pi_{x,y} - \Pi'_{x,y,H}| \leq 3\delta/4$.

Proof. In the following, we will frequently use the fact that for all $p, \alpha \in [0,1]$, it holds that $p(1-\alpha) \geq p - \alpha$. We extend the arguments given in [BW11] to prove the items of the claim in order.

1. By the definition of information complexity and the fact that mutual information is equal to the expectation of the divergence, we have that for (X,Y) distributed according to μ,

$$\mathsf{IC}_\mu(\pi) = I(X;\Pi \mid Y) + I(Y;\Pi \mid X) = \mathbb{E}_{(x,y)\leftarrow(X,Y)}[D(\Pi_{x,y} \parallel \Pi_y) + D(\Pi_{x,y} \parallel \Pi_x)]$$

This implies that $\mathbb{E}_{(x,y)\leftarrow(X,Y)}[D(\Pi_{x,y} \parallel \Pi_x)] \leq \mathsf{IC}_\mu(\pi)$, and since divergence is non-negative we have by Markov's inequality that

$$\Pr_{(x,y)\leftarrow(X,Y)}[D(\Pi_{x,y} \parallel \Pi_x) > 8\mathsf{IC}_\mu(\pi)/\delta] \leq \delta/8$$

The same argument holds for $D(\Pi_{x,y} \parallel \Pi_y)$ and by a union bound, we have $\Pr_{(X,Y)\sim\mu}[B_D] \leq \delta/4$.

2. We may write:

$$\begin{aligned}
&\Pr_{r_{\pi'}}[B_C] \\
&= \Pr[\exists i \neq j \in [T] \text{ s.t. } i \in (\mathcal{A} \cup \mathcal{B}), j \in (\mathcal{A} \cup \mathcal{B}), h(i) = h(j) = r] \\
&\leq \sum_{i \neq j \in [T]} \Pr[i \in (\mathcal{A} \cup \mathcal{B}) \wedge j \in (\mathcal{A} \cup \mathcal{B}) \wedge h(i) = h(j) = r] \\
&= \sum_{i \neq j} \Pr[i \in (\mathcal{A} \cup \mathcal{B})] \Pr[j \in (\mathcal{A} \cup \mathcal{B})] \Pr[h(i) = h(j) = r] \\
&\leq T^2 \frac{4}{(|\mathcal{U}|2^\Delta)^2} \cdot \frac{1}{2^{2k}} \\
&\leq \tfrac{\delta}{16} \cdot 2^{-(k+\Delta)}
\end{aligned}$$

where we have used the independence between the trials and independence of the h from the trials, as well as Item 1 of Claim 4.2.

3. Let us define G to be the event that the smallest $i \in \mathcal{A}$ satisfies $h(i) = r$, and also $i \in \mathcal{B}$. (Notice this implies that $\mathcal{A}, \mathcal{B}$ are both non-empty.) We have

$$\Pr_{r_{\pi'}}[G] = \Pr[\mathcal{A} \neq \varnothing] \cdot \Pr[G \mid \mathcal{A} \neq \varnothing].$$

Observe that an element i is in $\mathcal{A}$ if and only if experiment i is accepted by Alice. By Item 1 of Claim 4.2, the probability of Alice aborting each experiment i is $1 - \frac{1}{|\mathcal{U}|2^\Delta}$. Since the experiments are independent, the probability of Alice aborting all experiments is

$$\Pr[\mathcal{A} = \varnothing] = \left(1 - \tfrac{1}{|\mathcal{U}|2^\Delta}\right)^T \leq e^{-\frac{T}{|\mathcal{U}|2^\Delta}} \leq \delta/8$$

We assume now that $\mathcal{A}$ is non empty and we denote by i its first element.

$$\begin{aligned}&\Pr[G \mid \mathcal{A} \neq \varnothing]\\ &\quad = \Pr[h(i) = r \mid \mathcal{A} \neq \varnothing] \Pr[i \in \mathcal{B} \mid i \in \mathcal{A} \wedge h(i) = r \wedge \mathcal{A} \neq \varnothing]\end{aligned}$$

For all j, the probability that $h(j) = r$ is exactly 2^{-k}, in particular this holds for the first element of $\mathcal{A}$.

For any $(x, y) \notin B_D$, we have that $D(\Pi_{x,y} \parallel \Pi_x) \leq 8\mathsf{IC}_\mu(\pi)/\delta$. Let us say that a transcript is "bad for Alice" (resp. Bob) if it lies in the set $B_\Delta(\Pi_{x,y}, \Pi_x)$ (resp. in the set $B_\Delta(\Pi_{x,y}, \Pi_y)$. Using Lemma 4.3, this implies that

$$\begin{aligned}\gamma_A &= \Pr[\Pi_{x,y} \text{ bad for Alice}] \leq \frac{\frac{8}{\delta}\mathsf{IC}_\mu(\pi) + 1}{\Delta} \leq \delta/4\\ \gamma_B &= \Pr[\Pi_{x,y} \text{ bad for Bob}] \leq \delta/4\end{aligned}$$

It follows that $\gamma = \Pr[\Pi_{x,y} \text{ bad for Alice or Bob}] \leq \gamma_A + \gamma_B \leq \delta/2$.

By definition, for any j, experiment j is accepted if and only if $j \in \mathcal{A} \cap \mathcal{B}$. Therefore, $\forall j \in [T]$:

$$\begin{aligned}\Pr[j \in \mathcal{B} \mid j \in \mathcal{A}] &= \Pr[\text{experiment } j \text{ is accepted} \mid j \in \mathcal{A}]\\ &= \frac{\Pr[\text{experiment } j \text{ is accepted}]}{\Pr[j \in \mathcal{A}]}\\ &\geq \frac{1-\gamma}{2^\Delta}\\ &\geq \frac{1-\frac{\delta}{2}}{2^\Delta}\end{aligned}$$

where we used Item 1 and Item 2 of Claim 4.2, and the fact that $\neg B_D$ implies that $\gamma \leq \delta/4$.

Also, observe that by the definition of the protocol, the choice of h and r are completely independent of the experiments. Therefore we may add the condition that $h(j) = r$ without altering the probability. Since $j \in \mathcal{A}$ implies $\mathcal{A} \neq \varnothing$, we can add this condition too. We use this with $j = i$, so therefore we may write

$$\Pr_{r_{\pi'}}[G] \geq (1 - \delta/8)\frac{1-\frac{\delta}{4}}{2^{k+\Delta}} \geq \frac{1-\frac{5\delta}{8}}{2^{k+\Delta}}$$

Finally, observe that $H \setminus B_C = G \setminus B_C$, therefore we may conclude that:

$$\Pr[H] \geq \Pr[H \setminus B_C] = \Pr[G \setminus B_C] \geq \Pr[G] - \Pr[B_C] \geq (1 - \tfrac{11\delta}{16})2^{-k-\Delta}$$

4. We will again use the event G as defined in the previous point. We will again use the fact that:

$$\begin{aligned}\Pr_{r_{\pi'}}[G] &= \Pr[\mathcal{A} \neq \varnothing] \Pr[h(i) = r \mid \mathcal{A} \neq \varnothing] \Pr[i \in \mathcal{B} \mid i \in \mathcal{A} \wedge h(i) = r \wedge \mathcal{A} \neq \varnothing]\\ &\leq \Pr[h(i) = r \mid \mathcal{A} \neq \varnothing] \Pr[i \in \mathcal{B} \mid i \in \mathcal{A} \wedge h(i) = r \wedge \mathcal{A} \neq \varnothing]\end{aligned}$$

As before the first factor is exactly 2^{-k} for any i. We may also write:

$$\begin{aligned}\Pr[i \in \mathcal{B} \mid i \in \mathcal{A}] &= \Pr[\text{experiment } i \text{ is accepted} \mid i \in \mathcal{A}]\\ &= \frac{\Pr[\text{experiment } i \text{ is accepted}]}{\Pr[i \in \mathcal{A}]}\\ &\leq \frac{1}{2^\Delta},\end{aligned}$$

where we used Item 1 and Item 2 of Claim 4.2. As with the previous point, adding the conditions $h(i) = r$ and $\mathcal{A} \neq \varnothing$ does not affect the probabilities. Therefore, $\Pr[G] \leq 2^{-k-\Delta}$. Finally, observe that $H \subseteq G \cup B_C$, and therefore:

$$\Pr[H] \leq \Pr[G \cup B_C] \leq \Pr[G] + \Pr[B_C] \leq (1 + \tfrac{\delta}{16})2^{-k-\Delta}$$

5. The distribution of $\Pi'_{x,y}$ conditioned on not aborting *and* on no collision is simply the distribution of the output of a single experiment, and we know from the facts about the single experiment (Claim 4.2) that this is close to $\Pi_{x,y}$. We wish to conclude that $\Pi'_{x,y}$ conditioned *only* on not aborting is also close to $\Pi_{x,y}$. The following lemma allows us to do this by using the fact that the probability of collision is small:

Claim 4.7. *Let Π and Π' two distributions taking output in a common universe. Let F and H be two events in the underlying probability space of Π'. Finally, we let Π'_E denote the distribution Π' conditioned on $E = H \setminus F$, and assume that $|\Pi'_E - \Pi| \leq c$. Then it holds that $|\Pi'_H - \Pi| \leq c + \frac{\Pr_{\Pi'}[F]}{\Pr_{\Pi'}[H]}$.*

Proof. For shorthand, for any event E let us write $\Pi'_H(E) = \Pr_{\Pi'}[E \mid H]$, and similarly for $\Pi'(E)$ and $\Pi(E)$. For a set S in the support of Π' and Π, we let S also denote the event that the value of the random variable is in S.

It suffices to prove that, for all subsets S in the union of the supports of Π and Π'_H, it holds that $\Pi'_H(S) - \Pi(S) \leq c + \frac{\Pr_{\Pi'}[F]}{\Pr_{\Pi'}[H]}$. To show this, we may write:

$$\begin{aligned}
\Pi'_H(S) - \Pi(S) &= \frac{\Pi'(H \cap S)}{\Pi'(H)} - \Pi(S) \\
&\leq \frac{\Pi'(E \cap S) + \Pi'((H \setminus E) \cap S)}{\Pi'(H)} - \Pi(S) \\
&\leq \frac{\Pi'(E)(c + \Pi(S)) + \Pi'(F)}{\Pi'(H)} - \Pi(S)
\end{aligned}$$

since $(H \setminus E) \subseteq F$ and $|\Pi'_E - \Pi| \leq c$. Using the fact that $E \subset H$, we can conclude that:

$$\Pi'_H(S) - \Pi(S) \leq c + \frac{\Pi'(F)}{\Pi'(H)} = c + \frac{\Pr_{\Pi'}[F]}{\Pr_{\Pi'}[H]}.$$

□

We apply this lemma with $\Pi = \Pi_{x,y}$, $\Pi' = \Pi'_{x,y}$, $F = B_C$, and H the event that $\pi'(x,y) \neq \perp$. We note that $E = H \setminus B_C = G \setminus B_C$. We calculate c, $\Pr_{r_{\pi'}}[F]$ and $\Pr_{r_{\pi'}}[H]$:

- Since h and r are completely independent of the actual experiments themselves, it holds that the distribution of $\Pi'_{x,y}$ conditioned on $G \setminus B_C$ is identical to the output of a single experiment (Algorithm 4.1). Since $(x,y) \notin B_D$, the measure of the bad set $B_\Delta(\Pi_{x,y}, \Pi_y) \cup B_\Delta(\Pi_{x,y}, \Pi_x)$ is bounded by $\delta/2$. We apply Item 3 of Claim 4.2 to deduce that $|\Pi_{x,y} - \Pi'_{x,y,E}| \leq \delta/2 = c$.
- From Item 2 of Claim 4.6, we know that $\Pr_{r_{\pi'}}[F] = \Pr_{r_{\pi'}}[B_C] \leq \frac{\delta}{16} \cdot 2^{-(k+\Delta)}$.
- From Item 3 of Claim 4.6, we know $\Pr_{r_{\pi'}}[H] \geq (1 - \frac{11\delta}{16})2^{-(k+\Delta)}$ because $(x,y) \notin B_D$.

Therefore, Claim 4.7 implies that:

$$|\Pi_{x,y} - \Pi'_{x,y,\neq\perp}| \leq \delta/2 + \frac{\delta}{16(1-\frac{11\delta}{16})} \leq \delta/2 + \delta/5 < 3\delta/4$$

where we use the assumption that $\delta \leq 1$.

□

4.3 Proof of the Compression Lemma

Proof of Lemma 3.4. Set $\lambda = 2^{-(k+\Delta)}$. It holds that $\lambda \geq 2^{-C(\mathsf{IC}_\mu(\pi)/\delta^2+1/\delta)}$ for $C = 64$. Let $\mathcal{R}$ be any subset of the support of $(X, Y, \pi(X, Y))$. Then

$$\Pr_{r_\pi,(X,Y)\sim\mu}[(X, Y, \pi(X, Y)) \in \mathcal{R}] \leq \Pr_{(X,Y)\sim\mu}[B_D] + \Pr_{(X,Y)\sim\mu}[\neg B_D] \cdot \Pr_{r_\pi,(X,Y)\sim\mu}[(X, Y, \pi(X, Y)) \in \mathcal{R} \mid \neg B_D].$$

Applying Item 5 of Claim 4.6 and the fact that $\mathcal{R}$ is simply an event, it follows that for all $(x, y) \notin B_D$

$$\Pr_{r_\pi}[(x, y, \pi(x, y)) \in \mathcal{R}] \leq \Pr_{r_{\pi'}}[(x, y, \pi'(x, y)) \in \mathcal{R} \mid \pi'(x, y) \neq \perp] + \tfrac{3\delta}{4}.$$

Since $\Pr[B_D] \leq \delta/4$ (Item 1 of Claim 4.6),

$$\Pr_{r_\pi,(X,Y)\sim\mu}[(X, Y, \pi(X, Y)) \in \mathcal{R}] \leq \Pr_{r_{\pi'},(X,Y)\sim\mu}[(X, Y, \pi'(X, Y)) \in \mathcal{R} \mid \pi'(x, y) \neq \perp] + \delta.$$

This proves one direction of Equation 4. For the other direction, we have that

$$\begin{aligned}
&\Pr[(X, Y, \pi'(X, Y)) \in \mathcal{R} \mid \pi'(X, Y) \neq \perp] \\
&\quad\leq \Pr[B_D] + \Pr[\neg B_D] \cdot \Pr[(X, Y, \pi'(X, Y))) \in \mathcal{R} \mid \neg B_D, \pi'(X, Y) \neq \perp] \\
&\quad\leq \tfrac{\delta}{4} + \Pr[\neg B_D] \cdot \left(\Pr[(X, Y, \pi(X, Y)) \in \mathcal{R} \mid \neg B_D] + \tfrac{3\delta}{4}\right) \qquad (8) \\
&\quad\leq \Pr[(X, Y, \pi(X, Y)) \in \mathcal{R} \wedge \neg B_D] + \delta \\
&\quad\leq \Pr[(X, Y, \pi(X, Y)) \in \mathcal{R}] + \delta
\end{aligned}$$

where in Equation 8 we applied Item 5 of Claim 4.6. This proves Equation 4 of Lemma 3.4.

Equation 5 follows immediately from Item 4 of Claim 4.6.

Finally, for Equation 6, we may write:

$$\begin{aligned}
\Pr_{r_{\pi'},(X,Y)\sim\mu}[\pi'(X, Y) \neq \perp] &\geq \Pr_{(X,Y)\sim\mu}[\neg B_D] \Pr_{r_{\pi'},(X,Y)\sim\mu}[\pi'(X, Y) \neq \perp | \neg B_D] \\
&\geq (1 - \tfrac{\delta}{4})(\Pr_{r_{\pi'},(X,Y)\sim\mu}[H \mid \neg B_D]) \\
&\geq (1 - \tfrac{\delta}{4})(1 - \tfrac{11\delta}{16})\lambda \quad > \quad (1 - \delta)\lambda
\end{aligned}$$

where we used Item 3 of Claim 4.6. □

5 Applications

We can prove lower bounds on the information complexity of specific problems, by checking that their communication lower bounds were obtained by one of the methods subsumed by the relaxed partition bound, including the factorization norm, smooth rectangle, rectangle, or discrepancy. However, a bit of care is required to ascertain this. For example, while a paper may say it uses the "rectangle bound", we must still verify that the value of the linear program for prt (or one of the subsumed programs such as srec or rec) is at least the claimed bound, since different authors may use the term "rectangle bound" to mean different things. In particular what they call "rectangle bound" may not satisfy the constraints of the rectangle/smooth rectangle linear programs given by Jain and Klauck [JK10]. After we have verified that prt is appropriately bounded, then we can apply our main theorem (Theorem 1.1). We do this for the problems below.

5.1 Exponential separation of quantum communication and classical information complexity

We prove that the quantum communication complexity of the Vector in Subspace Problem is exponentially smaller than its classical information complexity (Theorem 1.2). In the Vector in Subspace Problem $\mathrm{VSP}_{0,n}$, Alice is given an $n/2$ dimensional subspace of an n dimensional space over $\mathbb{R}$, and Bob is given a vector. This is a partial function, and the promise is that either Bob's vector lies in the subspace, in which case the function evaluates to 1, or it lies in the orthogonal subspace, in which case the function evaluates to 0. Note that the input set of $\mathrm{VSP}_{0,n}$ is continuous, but it can be discretized by rounding, which leads to the problem $\widetilde{\mathrm{VSP}}_{\theta,n}$ (see [KR11] for details).

Klartag and Regev [KR11] show that the Vector in Subspace Problem can be solved with an $O(\log n)$ quantum protocol, but the randomized communication complexity of this problem is $\Omega(n^{1/3})$. Their lower bound uses a modified version of the rectangle bound, which we show is still weaker than our relaxed partition bound.

Lemma 5.1. *There exist universal constants C and γ such that for any ϵ,*

$$\bar{\mathsf{prt}}_\epsilon(\widetilde{\mathrm{VSP}}_{\theta,n}) \geq \frac{1}{C}(0.8 - 2.8\epsilon)\exp(\gamma n^{1/3}).$$

Klartag and Regev's lower bound is based on the following lemma:

Lemma 5.2 ([KR11])**.** *Let $f = \mathrm{VSP}_{0,n}$, $\mathcal{X}$ and $\mathcal{Y}$ denote Alice and Bob's input sets for f, σ be the uniform distribution over $\mathcal{X} \times \mathcal{Y}$ and σ_b be the uniform distribution over $f^{-1}(b)$. There exist universal constants C and γ such that, for any rectangle R and any $b \in \{0,1\}$, we have:*

$$\sigma_b(R \cap f^{-1}(b)) \geq 0.8\sigma(R) - C\exp(-\gamma n^{1/3}).$$

We show that this implies the required lower bound on the relaxed partition bound of $\widetilde{\mathrm{VSP}}_{\theta,n}$.

Proof of Lemma 5.1. Let us first consider $f = \mathrm{VSP}_{0,n}$ and show that $\log \bar{\mathsf{prt}}_\epsilon(\mathrm{VSP}_{0,n}) = \Omega(n^{1/3})$. Note that since the input set of $\mathrm{VSP}_{0,n}$ is continuous, we need to extend the definition of the relaxed partition bound to the case of a continuous input set, but this follows naturally by replacing summation symbols by integrals.

Let $\mu = \frac{\sigma_0+\sigma_1}{2}$, which satisfies $\mu(f^{-1}) = 1$, *i.e.* μ only has support on valid inputs. Then, Lemma 5.2 implies that for any rectangle R and any $b \in \{0,1\}$, we have

$$2\mu(R \cap f^{-1}(b \oplus 1)) \geq 0.8\sigma(R) - C\exp(-\gamma n^{1/3}).$$

Since μ only has support on f^{-1} and f is Boolean, we have $\mu(R) = \mu(R \cap f^{-1}(b)) + \mu(R \cap f^{-1}(b \oplus 1))$, so that the inequality can be rewritten as

$$2\mu(R) - 2\mu(R \cap f^{-1}(b)) \geq 0.8\sigma(R) - C\exp(-\gamma n^{1/3}).$$

Setting $t = \frac{\exp(\gamma n^{1/3})}{C}$, we have that:

$$t \cdot \left(2\mu(R \cap f^{-1}(b)) + 0.8\sigma(R) - 2\mu(R)\right) \leq 1$$

We now construct a feasible point for the dual formulation of $\bar{\mathsf{prt}}_\epsilon(f)$ given in Claim A.2 by setting $\alpha(x,y) = 2t\mu(x,y)$ for $(x,y) \in f^{-1}$ and $\alpha(x,y) = 0.8t\sigma(x,y)$ otherwise; and $\beta(x,y) = 2t\mu(x,y) - 0.8t\sigma(x,y)$ for $(x,y) \in f^{-1}$, and $\beta(x,y) = 0$ otherwise. Note that for $(x,y) \in f^{-1}$, we have $\mu(x,y) = \sigma(x,y)/\sigma(f^{-1}) \geq \sigma(x,y)$, hence all these values are positive. By the previous inequality, we also have

$$\int_{R \cap f^{-1}(b)} \alpha(x,y)dxdy + \int_{R \setminus f^{-1}} \alpha(x,y)dxdy - \int_R \beta(x,y)dxdy \leq 1$$

for any R, b, therefore this is a valid feasible point. Moreover, the corresponding objective value is

$$\begin{aligned}\int_{\mathcal{X}\times\mathcal{Y}}(1-\epsilon)\alpha(x,y)dxdy - \int_{\mathcal{X}\times\mathcal{Y}}\beta(x,y)dxdy \\ &= t\left((1-\epsilon)2\mu(f^{-1}) + (1-\epsilon)0.8(1-\sigma(f^{-1})) - 2\mu(f^{-1}) + 0.8\sigma(f^{-1})\right) \\ &= t(0.8 - 2.8\epsilon + 0.8\epsilon\sigma(f^{-1})) \\ &\geq t(0.8 - 2.8\epsilon).\end{aligned}$$

Finally, let us note that we can construct a zero-communication protocol for $\mathrm{VSP}_{0,n}$ by first rounding off the inputs, and then applying a zero-communication protocol for $\widetilde{\mathrm{VSP}}_{\theta,n}$. This means that we can turn a feasible point for the primal form of $\bar{\mathsf{prt}}_\epsilon(\widetilde{\mathrm{VSP}}_{\theta,n})$ (given in Claim A.2) into a feasible point for $\bar{\mathsf{prt}}_\epsilon(\mathrm{VSP}_{0,n})$ with the same objective value, so that $\bar{\mathsf{prt}}_\epsilon(\widetilde{\mathrm{VSP}}_{\theta,n}) \geq \bar{\mathsf{prt}}_\epsilon(\mathrm{VSP}_{0,n}) \geq t(0.8 - 2.8\epsilon)$. □

Finally, Lemma 5.1 together with Theorem 1.1 imply that $IC(\widetilde{\mathrm{VSP}}_{\theta,n}, \epsilon) = \Omega(n^{1/3})$ and also Theorem 1.2.

This allows us to conclude that the information complexity of this function is at least $\Omega(n^{1/3})$. This solves Braverman's Open Problem 3 (Are there problems for which $Q(f,\epsilon) = O(polylog(\mathsf{IC}(f,\epsilon)))$?) and Open Problem 7 (is it true that $\mathsf{IC}(\widetilde{\mathrm{VSP}}_{\theta,n}, 1/3) = n^{\Omega(1)}$?)

Moreover, our result implies an exponential separation between classical and quantum information complexity. We refrain from defining quantum information cost and complexity in this paper (see [JN10] for a definition), but since the quantum information complexity is always smaller than the quantum communication complexity, the separation follows trivially from Theorem 1.2.

5.2 Information complexity of the Gap Hamming Distance Problem

In the Gap Hamming Distance Problem (GHD_n), Alice and Bob each receive a string of length n and they need to determine whether their Hamming distance is at least $n/2 + \sqrt{n}$ or less than $n/2 - \sqrt{n}$. We prove that the information complexity of Gap Hamming Distance is $\Omega(n)$ (Theorem 1.3; Open Problem 6 in [Bra12]). The communication complexity of Gap Hamming Distance was shown to be $\Omega(n)$ by Chakrabarti and Regev [CR11]. The proof was subsequently simplified by Vidick [Vid11] and Sherstov [She11]. The first two proofs use the smooth rectangle bound, while Sherstov uses the rectangle/corruption bound.

The corruption bound used by Sherstov is a slight refinement of the rectangle bound as defined by Jain and Klauck [JK10], since it can handle distributions that put small weight on the set of inputs that map to some function value z. We verify that this bound is weaker than our relaxed partition bound, which implies Theorem 1.3.

Lemma 5.3. *There exist universal constants C and δ such that for any small enough ϵ, $\bar{\mathsf{prt}}_\epsilon(\mathrm{GHD}) \geq C2^{\delta n}$.*

We let the output set be $\mathcal{Z} = \{-1, 1\}$ rather than bits, to follow the notation of [She11]. Let us recall the corruption bound used by Sherstov.

Theorem 5.4. *For any function f with output set $\mathcal{Z} = \{-1,1\}$, and $\epsilon, \delta, \beta > 0$ if a distribution on the inputs μ is such that*

$$\mu(R) > \beta \implies \mu(R \cap f^{-1}(1)) > \delta\mu(R \cap f^{-1}(-1))$$

then

$$2^{R_\epsilon(f)} \geq \frac{1}{\beta}(\mu(f^{-1}(-1)) - \frac{\epsilon}{\delta}).$$

We can derive the general corruption bound as used by Sherstov by giving a feasible solution to the dual of the linear program by Jain and Klauck. In the dual form, $\mathsf{rec}_\epsilon^z(f)$ is defined for $z \in \mathcal{Z}$ as

$$\mathsf{rec}_\epsilon^z(f) = \max_{\alpha_{x,y} \geq 0} (1-\epsilon) \sum_{(x,y)\in f^{-1}(z)} \alpha_{x,y} - \epsilon \sum_{(x,y)\in f^{-1}\setminus f^{-1}(z)} \alpha_{x,y} \tag{9}$$

$$\forall R, \quad \sum_{(x,y)\in R\cap f^{-1}(z)} \alpha_{x,y} - \sum_{(x,y)\in R\cap(f^{-1}\setminus f^{-1}(z))} \alpha_{x,y} \leq 1. \tag{10}$$

Consider the following assignment for $\mathsf{rec}_\epsilon^{-1}$, where μ is the distribution over the inputs in the theorem. Letting $\alpha_{x,y} = \frac{1}{\beta}\mu(x,y)$ if $f(x,y) = -1$, and $\alpha_{x,y} = \frac{1}{\delta}\frac{1}{\beta}\mu(x,y)$ if $f(x,y) = 1$, we can verify the constraints and the objective value is greater than the corruption bound.

To conclude the bound on GHD, Sherstov gives a reduction to the Gap Orthogonality Problem and proves the following lemma.

Lemma 5.5 ([She11]). *Let f denote the Gap Orthogonality Problem. For a small enough constant $\delta < 1$ and for the uniform distribution μ on inputs, and any rectangle R such that $\mu(R) > 2^{-\delta n}$, $\mu(R \cap f^{-1}(1)) > \delta\mu(R)$. Furthermore, $\mu(f^{-1}(-1)) = \Theta(1)$.*

Putting all this together, we have that $\mathsf{rec}_\epsilon^{-1}(f) \geq C2^{-\delta n}$ for appropriate choices of ϵ, and the reduction gives us $\bar{\mathsf{prt}}_\epsilon(\mathrm{GHD}) \geq \mathsf{rec}_\epsilon^{-1}(\mathrm{GHD}) \geq C2^{-\delta n}$.

6 Conclusions and open problems

We have shown that the information complexity is lower bounded by a relaxed version of the partition bound. This subsumes all known algebraic and rectangle based methods, except the partition bound. It remains to see if the partition bound also provides a lower bound on the information complexity. Alternatively, if we would like to separate the communication and information complexities, then possible candidates could be functions whose partition bound is strictly larger than their relaxed partition bound.

Moreover, we have seen how the relaxed partition bound naturally relates to zero-communication protocols with abort. Actually, we can relate all other lower bound methods to different variants of zero-communication protocols [DKLR11, LLR12]. This provides new insight on the inherent differences between these bounds and may lead to new lower bound methods, coming from different versions of zero-communication protocols. Moreover, since these protocols have been extensively studied in the field of quantum information, it is intriguing to see what other powerful tools can be transferred to the model of classical communication complexity.

References

[BBCR10] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. In *Proc. 42nd STOC*, pages 67–76, 2010. doi:10.1145/1806689.1806701.

[BHMR03] H. Buhrman, P. Høyer, S. Massar, and H. Röhrig. Combinatorics and quantum nonlocality. *Phys. Rev. Lett.*, 91, 2003. arXiv:quant-ph/0209052, doi:10.1103/PhysRevLett.91.047903.

[BHMR06] H. Buhrman, P. Høyer, S. Massar, and H. Röhrig. Multipartite nonlocal quantum correlations resistant to imperfections. *Phys. Rev. A*, 73, 2006. doi:10.1103/PhysRevA.73.012321.

[BR11] M. Braverman and A. Rao. Information equals amortized communication. In *Proc. 52nd FOCS*, pages 748–757, 2011. arXiv:1106.3595, doi:10.1109/FOCS.2011.86.

[Bra12] M. Braverman. Interactive information complexity. In *Proc. 44th STOC*, 2012. To appear. Available from: http://eccc.hpi-web.de/report/2011/123/.

[BW11] M. Braverman and O. Weinstein. A discrepancy lower bound for information complexity. Technical Report 1112.2000, ECCC, 2011. Available from: http://eccc.hpi-web.de/report/2011/164/.

[BYJKS04] 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*, 68(4):702–732, 2004. doi:10.1016/j.jcss.2003.11.006.

[CR11] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. In *Proc. 43rd STOC*, pages 51–60, 2011. doi:10.1145/1993636.1993644.

[CSWY01] A. Chakrabarti, Y. Shi, A. Wirth, and A. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 42nd FOCS*, pages 270–278, 2001. doi:10.1109/SFCS.2001.959901.

[DKLR11] J. Degorre, M. Kaplan, S. Laplante, and J. Roland. The communication complexity of non-signaling distributions. *Quantum Information and Computation*, 11(7 –8):649–676, 2011. arXiv:0804.4859.

[GG99] B. Gisin and N. Gisin. A local hidden variable model of quantum correlation exploiting the detection loophole. *Phys. Lett. A*, 260:323–327, 1999. arXiv:quant-ph/9905018, doi:10.1016/S0375-9601(99)00519-8.

[JK10] R. Jain and H. Klauck. The partition bound for classical complexity and query complexity. In *Proc. 25th CCC*, pages 247–258, 2010. arXiv:0910.4266, doi:10.1109/CCC.2010.31.

[JN10] R. Jain and A. Nayak. The space complexity of recognizing well-parenthesized expressions in the streaming model: the Index function revisited. Technical Report 1004.3165, quant-ph, 2010.

[KR11] B. Klartag and O. Regev. Quantum one-way communication can be exponentially stronger than classical communication. In *Proc. 43rd STOC*, pages 31–40, 2011. arXiv:1009.3640, doi:10.1145/1993636.1993642.

[LLR12] S. Laplante, V. Lerays, and J. Roland. Classical and quantum partition bound and detector inefficiency. Technical Report 1203.4155, quant-ph, 2012.

[LS09a] T. Lee and A. Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–399, 2009. doi:10.1561/0400000040.

[LS09b] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures and Algorithms*, 34(3):368–394, 2009. doi:10.1002/rsa.20232.

[Mas02] S. Massar. Non locality, closing the detection loophole and communication complexity. *Phys. Rev. A*, 65, 2002. arXiv:quant-ph/0109008, doi:10.1103/PhysRevA.65.032121.

[She11] A. Sherstov. The communication complexity of Gap Hamming Distance. Technical Report TR11-063, ECCC, 2011. To appear, Theory of Computing. Available from: http://eccc.hpi-web.de/report/2011/063/.

[Vid11] T. Vidick. A concentration inequality for the overlap of a vector on a large set with application to the communication complexity of the Gap-Hamming-Distance problem. Technical Report TR11-051, ECCC, 2011. Available from: http://eccc.hpi-web.de/report/2011/051/.

A Relaxed partition bound and other bounds

Recall the definition of the partition bound $\mathsf{prt}_\epsilon(f)$ in [JK10]:

Definition A.1 ([JK10])**.** The partition bound $\mathsf{prt}_\epsilon(f)$ is defined as the value of the following linear program:

$$\mathsf{prt}_\epsilon(f) = \min_{w_{R,z}\geq 0} \sum_{R,z} w_{R,z} \quad \text{subject to:} \quad \forall (x,y)\in\mathcal{I}, \sum_{R:(x,y)\in R} w_{R,f(x,y)} \geq 1-\epsilon$$
$$\forall (x,y)\in\mathcal{X}\times\mathcal{Y}, \sum_{z,R:(x,y)\in R} w_{R,z} = 1$$

(To put this into a form similar to Definition 3.2, simply substitute $\frac{1}{\eta} = \sum_{R,z} w_{R,z}$ and $p_{R,z} = w_{R,z}\eta$.)

We first show that $\overline{\mathsf{prt}}_\epsilon(f)$ is indeed a relaxation of the partition bound by providing linear programming formulations for $\overline{\mathsf{prt}}^\mu_\epsilon(f)$ and $\overline{\mathsf{prt}}_\epsilon(f)$.

Claim A.2. *$\overline{\mathsf{prt}}^\mu_\epsilon(f)$ can be expressed as the following linear programs:*

1. *Primal form:*

$$\overline{\mathsf{prt}}^\mu_\epsilon(f) = \min_{w_{R,z}\geq 0} \sum_{R,z} w_{R,z} \quad \textit{subject to:}$$
$$\sum_{(x,y)\in\mathcal{I}} \mu_{x,y} \sum_{R:(x,y)\in R} w_{R,f(x,y)} + \sum_{(x,y)\notin\mathcal{I}} \mu_{x,y} \sum_{z,R:(x,y)\in R} w_{R,z} \geq 1-\epsilon$$
$$\forall (x,y)\in\mathcal{X}\times\mathcal{Y}, \sum_{z,R:(x,y)\in R} w_{R,z} \leq 1$$

2. *Dual form:*

$$\overline{\mathsf{prt}}^\mu_\epsilon(f) = \max_{\alpha\geq 0,\beta_{x,y}\geq 0} (1-\epsilon)\alpha + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \beta_{x,y} \textit{ subject to:}$$
$$\forall R, z, \sum_{(x,y)\in R\cap f^{-1}(z)} \alpha\mu_{x,y} + \sum_{(x,y)\in R\setminus\mathcal{I}} \alpha\mu_{x,y} - \sum_{(x,y)\in R} \beta_{x,y} \leq 1$$

Similarly, $\overline{\mathsf{prt}}_\epsilon(f)$ is given by the value of the following linear programs:

1. *Primal form:*

$$\overline{\mathsf{prt}}_\epsilon(f) = \min_{w_{R,z}\geq 0} \sum_{R,z} w_{R,z} \quad \textit{subject to:} \quad \forall (x,y)\in\mathcal{I}, \sum_{R:(x,y)\in R} w_{R,f(x,y)} \geq 1-\epsilon$$
$$\forall (x,y)\notin\mathcal{I}, \sum_{z,R:(x,y)\in R} w_{R,z} \geq 1-\epsilon$$
$$\forall (x,y)\in\mathcal{X}\times\mathcal{Y}, \sum_{z,R:(x,y)\in R} w_{R,z} \leq 1$$

2. *Dual form:*

$$\bar{\mathsf{prt}}_\epsilon(f) = \max_{\alpha_{x,y}\geq 0,\beta_{x,y}\geq 0} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (1-\epsilon)\alpha_{x,y} - \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \beta_{x,y} \text{ subject to:}$$
$$\forall R, z, \quad \sum_{(x,y)\in R\cap f^{-1}(z)} \alpha_{x,y} + \sum_{(x,y)\in R\setminus\mathcal{I}} \alpha_{x,y} - \sum_{(x,y)\in R} \beta_{x,y} \leq 1$$

Proof. The primal form of $\bar{\mathsf{prt}}^{\mu}_\epsilon(f)$ can be obtained from Definition 3.2 by using the change of variables $w_{R,z} = p_{R,z}/\eta$. The dual form then follows from standard linear programming duality.

As for $\bar{\mathsf{prt}}_\epsilon(f)$, by Definition 3.2 we have $\bar{\mathsf{prt}}_\epsilon(f) = \max_\mu \bar{\mathsf{prt}}^{\mu}_\epsilon(f)$. The dual form of $\bar{\mathsf{prt}}_\epsilon(f)$ then follows from the dual form of $\bar{\mathsf{prt}}^{\mu}_\epsilon$ via the change of variables $\alpha_{x,y} = \alpha\mu_{x,y}$. Finally, the primal form can be obtained via standard linear programming duality. □

Let us compare the primal form of $\bar{\mathsf{prt}}_\epsilon(f)$ with the definition of the partition bound $\mathsf{prt}_\epsilon(f)$. We see that the first two constraints in $\bar{\mathsf{prt}}_\epsilon(f)$ imply that $\sum_{z,R:(x,y)\in R} w_{R,z}$ lies between $1-\epsilon$ and 1, while for $\mathsf{prt}_\epsilon(f)$, this should be exactly equal to 1. In terms of zero-communication protocols, this difference can be interpreted as follows: for $\mathsf{prt}_\epsilon(f)$, the protocol should output anything but $\perp$ with constant probability η for any input (x,y), while for $\bar{\mathsf{prt}}_\epsilon(f)$, the probability of not outputting $\perp$ is allowed to fluctuate between $(1-\epsilon)\eta$ and η.

Just as the partition bound, the relaxed partition bound is stronger than the smooth rectangle bound, defined in [JK10] as follows:

Definition A.3 ([JK10])**.** The smooth rectangle bound $\mathsf{srec}^{z_0}_\epsilon(f)$ is the value of the following linear program:

$$\mathsf{srec}^{z_0}_\epsilon(f) = \min_{w'_R\geq 0} \sum_R w'_R \quad \text{subject to:} \quad \forall (x,y)\in f^{-1}(z_0), \quad \sum_{R:(x,y)\in R} w'_R \geq 1-\epsilon$$
$$\forall (x,y)\in f^{-1}(z_0), \quad \sum_{R:(x,y)\in R} w'_R \leq 1$$
$$\forall (x,y)\in \mathcal{I}\setminus f^{-1}(z_0), \quad \sum_{R:(x,y)\in R} w'_R \leq \epsilon.$$

Let us now prove Lemma 3.3, that is, $\mathsf{srec}^{z_0}_\epsilon(f) \leq \bar{\mathsf{prt}}_\epsilon(f) \leq \mathsf{prt}_\epsilon(f)$.

Proof of Lemma 3.3. The second inequality is immediate since the only difference between the linear programs is that the constraint $\sum_{z,R:(x,y)\in R} w_{R,z} = 1$ in the primal form of $\mathsf{prt}_\epsilon(f)$ has been relaxed to $1-\epsilon \leq \sum_{z,R:(x,y)\in R} w_{R,z} \leq 1$.

As for the first inequality, let $w_{R,z}$ be an optimal solution for the primal formulation of $\bar{\mathsf{prt}}_\epsilon(f)$ in Claim A.2. Then, it is straightforward to check that setting $w'_R = w_{R,z_0}$ leads to a feasible point for $\mathsf{srec}^{z_0}_\epsilon(f)$, with objective value $\sum_R w'_R = \sum_R w_{R,z_0} \leq \sum_{R,z} w_{R,z} = \bar{\mathsf{prt}}_\epsilon(f)$. □