Electronic Colloquium on Computational Complexity, Report No. 199 (2015)

Relaxed partition bound is quadratically tight for product distributions*

Prahladh Harsha† Rahul Jain‡ Jaikumar Radhakrishnan§

December 7, 2015

Abstract

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a 2-party function. For every product distribution μ on $\{0,1\}^n \times \{0,1\}^n$, we show that

$$\mathsf{CC}^{\mu}_{0.49}(f) = O\left(\left(\log \mathsf{rprt}_{1/4}(f) \cdot \log\log \mathsf{rprt}_{1/4}(f)\right)^2\right),$$

where $\mathsf{CC}^{\mu}_{\varepsilon}(f)$ is the distributional communication complexity with error at most ε under the distribution μ and $\mathsf{rprt}_{1/4}(f)$ is the *relaxed partition bound* of the function f, as introduced by Kerenidis et al. [8]. A similar upper bound for communication complexity for product distributions in terms of information complexity was recently (and independently) obtained by Kol [9].

We show a similar result for query complexity under product distributions. Let $g : \{0,1\}^n \to \{0,1\}$ be a function. For every bit-wise product distribution μ on $\{0,1\}^n$, we show that

$$\mathsf{QC}^{\mu}_{1/3}(g) = O\left(\left(\log \mathsf{rqprt}_{1/4}(g) \log\log \mathsf{rqprt}_{1/4}(g))\right)^2\right),$$

where $\mathsf{QC}^{\mu}_{1/3}(g)$ is the distributional query complexity with error at most 1/3 under the distribution μ and $\mathsf{rqprt}_{1/4}(g))$ is the *relaxed query partition bound* of the function g.

Recall that relaxed partition bounds were introduced (in both communication complexity and query complexity models) to provide LP-based lower bounds to randomized communication complexity and randomized query complexity. Our results demonstrate that these lower bounds are polynomially tight for *product* distributions.

*Work done when the three authors were visiting the Simons Institute for the Theory of Computing, Berkeley, USA.

†Tata Institute of Fundamental Research, Mumbai, India. prahladh@tifr.res.in. Research supported in part by ISF-UGC grant 1399/4 and Google India Fellowship.

‡Centre for Quantum Technologies and Department of Computer Science, National University of Singapore, Singapore. rahul@comp.nus.edu.sg. Partly supported by the Singapore Ministry of Education Tier 3 Grant and Young Researcher Award, National University of Singapore.

§Tata Institute of Fundamental Research, Mumbai, India. jaikumar@tifr.res.in.

ISSN 1433-8092

1 Introduction

Over the last decade, several lower bound techniques using linear programming formulations and information complexity methods have been developed for problems in communication complexity and query complexity. One of the central questions in communication complexity is to understand the tightness of these lower bound techniques. For instance, over the last few years, considerable effort has gone into understanding the *information complexity* measure. Informally speaking, (internal) information complexity is the amount of information the two parties reveal to each other about their respective inputs while computing the joint function. It is known that for product distributions, the internal information complexity not only lower bounds but also upper bounds the distributional communication complexity (up to logarithmic multiplicative factors) [1]. On the other hand, recent works due to Ganor, Kol and Raz [3, 5, 4] show that there exist non-product distributions which exhibit exponential separation between internal information complexity and distributional communication complexity[1]. However, it is still open if internal information complexity (or a polynomial of it) upper bounds the public-coin randomized communication complexity [2].

Jain and Klauck [6], using tools from linear programming, gave a uniform treatment to several of the existing lower bound techniques and proposed the *partition bound*. Kerenidis et al. [8] proposed a relaxation of the partition bound, called the *relaxed partition bound* and showed that the relaxed partition bound lower bounds the internal information complexity. This leads to another interesting conjecture: does relaxed partition bound (or a polynomial of it) yield an upper bound on the communication complexity? We are not aware of any counterexample to this conjecture.

We consider this question when the inputs to Alice and Bob are drawn from a product distribution and show the following.

Theorem 1.1. *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and let $\mathsf{rprt}_\varepsilon(f)$ be the relaxed partition bound of f with error at most ε. For a product distribution μ on $\{0,1\}^n \times \{0,1\}^n$, the distributional communication complexity of f under distribution μ with error at most 0.49, denoted by $\mathsf{CC}^\mu_{0.49}(f)$, can be bounded above as follows:*

$$\mathsf{CC}^\mu_{0.49}(f) = O\left(\left(\log \mathsf{rprt}_{1/4}(f) \cdot \log\log \mathsf{rprt}_{1/4}(f)\right)^2\right).$$

We remark that recently (and independently of this work) Kol [9] obtained the following similar upper bound for communication complexity over product distributions μ in terms of $\mathsf{IC}^\mu(f)$, the information complexity for f under μ. She showed that

$$\mathsf{CC}^\mu_{\delta+\varepsilon}(f) = O\left(\mathsf{IC}^\mu_\delta(f)^2 \cdot \text{poly}\log \mathsf{IC}^\mu_\delta(f)/\varepsilon^5\right), \tag{1.1}$$

and concluded that

$$\mathsf{CC}^\mu_{0.49}(f) = O\left(\mathsf{IC}_{1/8}(f)^2 \cdot \text{poly}\log \mathsf{IC}_{1/8}(f)\right), \tag{1.2}$$

where $\mathsf{IC}_{1/8}(f) = \max_\mu \mathsf{IC}^\mu_{1/8}(f)$. The above mentioned result of Kerenedis et al. [8] implies that $\mathsf{rprt}_{1/4}(f) = O(\mathsf{IC}_{1/8}(f)$. Thus, (1.2) follows from our Theorem 1.1; Kol's stronger result (1.1) and our Theorem 1.1 are otherwise incomparable.

We consider similar question in query complexity and show the following.

[1]The third result of Ganor, Kol and Raz [4] actually demonstrates an exponential separation between external information and communication complexity, albeit for a communication task.

Theorem 1.2. *Let $g : \{0,1\}^n \to \{0,1\}$ be a function and μ be a bit-wise product distribution on $\{0,1\}^n$. Let $\mathsf{rqprt}_\varepsilon(g)$ be the relaxed query partition bound for g with error ε. Then, the distributional query complexity with error at most $1/3$ under the distribution μ, denoted by $\mathsf{QC}^{\mu}_{1/3}(f)$, can be bounded above as follows:*

$$\mathsf{QC}^{\mu}_{1/3}(g) = O\left(\left(\log \mathsf{rqprt}_{1/4}(g) \cdot \log\log \mathsf{rqprt}_{1/4}(g)\right)^2\right).$$

A similar quadratic upper bound for query complexity for product distributions in terms of approximate certificate complexity was obtained by Smyth [12]. His proof uses Reimer's inequality while our proof technique is arguably more elementary.

The rest of the paper is devoted to the proof of these two theorems. In the respective sections, we provide an overview of our proof techniques and its relationship with earlier works.

2 Communication Complexity

2.1 Preliminaries

We work in Yao's two-party communication model [13] (see Kushilevitz and Nisan [10] for an excellent introduction to the area). Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be finite non-empty sets, and let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. A two-party protocol for computing f consists of two parties, Alice and Bob, who get inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively, and exchange messages in order to compute $f(x,y) \in \mathcal{Z}$ (using shared randomness if necessary).

For a distribution μ on $\mathcal{X} \times \mathcal{Y}$, let the ε-error distributional communication complexity of f under μ (denoted by $\mathsf{CC}^{\mu}_{\varepsilon}(f)$), be the number of bits communicated (for the worst-case input) by the best deterministic protocol for f with average error at most ε under μ. Let $\mathsf{CC}^{\mathrm{pub}}_{\varepsilon}(f)$, the public-coin randomized communication complexity of f with worst case error ε, be the number of bits communicated (for the worst-case input) by the best public-coin randomized protocol, that for each input (x,y) computes $f(x,y)$ correctly with probability at least $1-\varepsilon$. Randomized and distributional complexity are related by the following special case of von Neumann's minmax principle.

Theorem 2.1 (Yao's minmax principle [14])**.** $\mathsf{CC}^{\mathrm{pub}}_{\varepsilon}(f) = \max_\mu \mathsf{CC}^{\mu}_{\varepsilon}(f)$.

In this section, we will prove Theorem 1.1 by first showing an upper bound on communication complexity in terms of the smooth rectangle bound and then observing that the smooth rectangle bound is bounded above by the relaxed partition bound.

Smooth rectangle bound: The smooth rectangle bound was introduced by Jain and Klauck [6], as a generalization of the rectangle bound. Informally, the smooth rectangle bound for a function f under a distribution μ, is the maximum over all functions g , which are close to f under the distribution μ, of the rectangle bound of g. However, it will be more convenient for us to work with the following linear programming formulation. Please see [6, Lemma 2] and [7, Lemma 6] for the relations between the LP formulation and the more "natural" formulation in terms of rectangle bound. Another advantage of the LP formulation is that it is evident that the smooth rectangle bound is a further relaxation of the relaxed-partition bound. The results in this paper are best phrased in terms of a distributional version of the above smooth rectangle bound. First for some notation, For any function $\mu : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and any $z \in \mathcal{Z}$ and rectangle R, let $\mu_z(R) := \mu(R \cap$

$f^{-1}(z)$) and $\mu_{\bar{z}}(R) := \mu(R) - \mu_z(R)$. Furthermore, let $\mu_z := \mu_z(\mathcal{X} \times \mathcal{Y})$ and $\mu_{\bar{z}} := \mu_{\bar{z}}(\mathcal{X} \times \mathcal{Y})$. The smooth rectangle and its distributional version are defined below.

Definition 2.2 (Smooth rectangle bound)**.**

- *For a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and $\varepsilon \in (0,1)$, the (ε, δ)-smooth rectangle bound of f denoted $\mathsf{srec}_{\varepsilon,\delta}(f)$ is defined to be $\max\{\mathsf{srec}^z_{\varepsilon,\delta}(f) : z \in \mathcal{Z}\}$, where $\mathsf{srec}^z_{\varepsilon,\delta}(f)$ is given by the optimal value of the following linear program.*
- *For a distribution μ on $\mathcal{X} \times \mathcal{Y}$ and function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, the (ε, δ)-smooth rectangle bound of f with respect to μ denoted $\mathsf{srec}^{\mu}_{\varepsilon,\delta}(f)$ is defined to be $\max\{\mathsf{srec}^{z,\mu}_{\varepsilon,\delta}(f) : z \in \mathcal{Z}\}$, where $\mathsf{srec}^{z,\mu}_{\varepsilon,\delta}(f)$ is given by the optimal value of the following linear program.*

$$\underline{\mathsf{srec}^z_{\varepsilon,\delta}(f)}$$

$$\begin{aligned}
\min \quad & \sum_R w_R \\
& \sum_{R:(x,y)\in R} w_R \geqslant 1 - \varepsilon, && \forall (x,y) \in f^{-1}(z) \\
& \sum_{R:(x,y)\in R} w_R \leqslant \delta, && \forall (x,y) \in f^{-1} - f^{-1}(z) \\
& \sum_{R:(x,y)\in R} w_R \leqslant 1, && \forall (x,y) \\
& w_R \geqslant 0, && \forall R\ .
\end{aligned}$$

$$\underline{\mathsf{srec}^{z,\mu}_{\varepsilon,\delta}(f)}$$

$$\min \quad \sum_R w_R$$

$$\sum_{(x,y)\in f^{-1}(z)} \mu_{x,y} \sum_{R:(x,y)\in R} w_R \geqslant (1-\varepsilon)\cdot \mu_z \tag{2.1}$$

$$\sum_{R:(x,y)\in R} w_R \leqslant \delta, \qquad \forall (x,y) \in f^{-1} - f^{-1}(z) \tag{2.2}$$

$$\sum_{R:(x,y)\in R} w_R \leqslant 1, \qquad \forall (x,y) \tag{2.3}$$

$$w_R \geqslant 0, \qquad \forall R\ .$$

We will refer to the constraint in (2.1) as the covering constraint and the ones in (2.2) as the packing constraints. Note that while there is a single covering constraint (averaged over all the inputs (x,y) that satsify $f(x,y) = z$) there are packing constraints corresponding to each $(x,y) \notin f^{-1}(z)$.

Similar to Yao's minmax principle Theorem 2.1, we have the following proposition relating the distributional version of the smooth rectangle bound to the smooth rectangle bound.

Proposition 2.3. $\mathsf{srec}_{\varepsilon,\delta}(f) = \max_\mu \mathsf{srec}^{\mu}_{\varepsilon,\delta}(f)$.

The main result of this section is the following

Theorem 2.4. *For any Boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and any* product *distribution μ on $\{0,1\}^n \times \{0,1\}^n$, we have*

$$\mathsf{CC}^{\mu}_{0.49}(f) = O\left((\log \mathsf{srec}^{\mu}_{1/n^2,1/n^2}(f))^2 \cdot \log n\right).$$

Furthermore, if there exists $k \geqslant 20$ such that

$$\lceil 100 \log \mathsf{srec}^{\mu}_{\delta,\delta}(f) \rceil \leqslant k,$$

for $\delta \leqslant 1/(30 \cdot 100(k+1)^4)$, then

$$\mathsf{CC}^{\mu}_{0.49}(f) = O(k^2).$$

The smooth rectangle bound can in turn be bounded from above by the relaxed partition bound (see Appendix B for the definition of rprt).

Proposition 2.5. *For any Boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and any $\delta \in (0,1)$,*

$$\log \mathsf{srec}_{\delta,\delta}(f) \leqslant O\left(\log \frac{1}{\delta}\right) \cdot \log \mathsf{rprt}_{1/4}(f).$$

Using the above proposition, we can reduce the error (i.e., δ) to $1/n^2$ and show that $\mathsf{CC}^{\mu}_{0.49}(f) = O\left(\left(\log \mathsf{rprt}_{1/4}(f)\right)^2 \cdot (\log n)^3\right)$. However, we can also reduce the error to $1/\mathrm{poly}(\log \mathsf{rprt}_{1/4}(f))$ and show that there exists a $k = O\left(\log \mathsf{rprt}_{1/4}(f) \cdot \log\log \mathsf{rprt}_{1/4}(f)\right)$ that satisfies the hypothesis for the final conclusion of Theorem 2.4. Theorem 1.1 now follows by combining Propositions 2.3 and 2.5 and Theorem 2.4.

2.2 Proof of Theorem 2.4

In this section, we construct a communication protocol tree with a small number of leaves from the optimal solution to the LP corresponding to $\mathsf{srec}^{1,\mu}_{\varepsilon,\delta}$. The construction of the protocol tree with a small number of leaves is inspired by a construction due to Nisan and Wigderson, in the context of log-rank conjecture [11, Theorem 2] (see also [10, Combinatorial proof of Theorem 2.11]). Unlike these constructions, our protocol works for a distribution and allows for error. As a result, the decomposition into sub-problems needs to be performed more carefully. This step critically uses the product nature of the distribution μ.

The protocol is inductively obtained from the optimal LP solution to $\mathsf{srec}^{1,\mu}_{\varepsilon,\delta}$ by successively decomposing the communication matrix into appropriate rectangles. This decomposition is accomplised using an inductive argument. First, we show that a small srec^1 value implies the existence of a large 1-biased rectangle (see Lemma 2.6). Based on this large rectangle, the entire communication matrix is partitioned into three parts: (1) the large biased rectangle itself, (2) a rectangle whose corresponding sub-problem admits an LP solution leading to a smaller srec^1 value (the underlying product nature of the distribution μ is crucially used here) and (3) a rectangle where the total measure with respect to μ drops significantly (see Lemma 2.7).

Suppose f satisfies $\mathsf{srec}^{\mu}_{\varepsilon,\delta}(f) = \mathsf{srec}^{1,\mu}_{\varepsilon,\delta}(f) = D$ with respect to some *product* distribution μ.

Lemma 2.6 (large biased rectangle)**.** *If $\mathsf{srec}^{1,\mu}_{\varepsilon,\delta}(f) = D$, then for every $\rho \in (0,1)$ there exists a rectangle S such that $\mu_1(S) > \mu_0(S)/\rho$ (ie., S is biased towards 1) and*

$$\mu(S) \geqslant \mu_1(S) \geqslant \frac{1}{D} \cdot \left((1-\varepsilon)\cdot \mu_1 - \frac{\delta}{\rho}\cdot \mu_0\right).$$

By the above lemma there exists a large (approximately $1/D$-sized) $\sqrt{\delta}$-biased (towards 1) rectangle $S = X_0 \times Y_0$. Let $X_1 = \mathcal{X}\backslash X_0$ and $Y_1 = \mathcal{Y}\backslash Y_0$. For $i,j \in \{0,1\}$, define rectangles $R^{(ij)} := X_i \times Y_j$, $R^{(1*)} := X_1 \times \mathcal{Y}$, and $R^{(*1)} := \mathcal{X} \times Y_1$. (Note, $S = R^{(00)}$.) For $i,j \in \{0,1,*\}$, let $\mu^{(ij)}$ be the restriction of μ to the rectangle $R^{(ij)}$. We show in the lemma below that the function f when restricted to either $R^{(10)}$ or $R^{(01)}$ has the property that the corresponding srec^1 drops by a constant factor.Define

$$\varepsilon(f) := 1 - \frac{\left(\sum_{(x,y)\in f^{-1}(1)} \mu_{x,y} \sum_{R:(x,y)\in R} w_R\right)}{\mu_1},$$

$$\varepsilon^{(ij)}(f) := 1 - \frac{\left(\sum_{(x,y)\in f^{-1}(1)\cap R^{(ij)}} \mu_{x,y} \sum_{R:(x,y)\in R} w_R\right)}{\mu_1(R^{(ij)})}; \qquad \text{for } i,j \in \{0,1\}.$$

It follows from the covering constraint that $\varepsilon(f) \leqslant \varepsilon$. Furthermore, $\varepsilon(f)$ is an average of the $\varepsilon^{(ij)}$'s in the sense that $\varepsilon(f) = \left(\sum_{i,j\in\{0,1\}} \mu_1(R^{(ij)})\varepsilon^{(ij)}\right)/\mu_1$.

Lemma 2.7. *For a product distribution μ, there exists an $(ij) \in \{(01),(10)\}$ such that the following holds: either f is 1/2-biased towards 0 with respect to the distribution $\mu^{(ij)}$ or $\mathsf{srec}^{1,\mu^{(ij)}}_{\varepsilon^{(ij)}+30\sqrt[4]{\delta},\delta}(f) \leqslant 0.9D$ where $\varepsilon^{(ij)}$ is as defined above.*

We will defer the proof of these lemmas and first show how they imply a low cost communication protocol (claimed in Theorem 2.4).

Suppose $\mu^{(01)}$ satisfies $\mathsf{srec}^{1,\mu^{(01)}}_{\varepsilon^{(01)}+30\sqrt[4]{\delta},\delta}(f) \leqslant 0.9D$ as given by the above lemma. Consider the following decomposition of the space $\mathcal{X} \times \mathcal{Y}$ given by $(R^{(00)}, R^{(01)}, R^{(1*)} = R^{(10)} \cup R^{(11)})$. We note that $R^{(00)}$ is a large biased rectangle, $R^{(01)}$ has lower srec^1 value while $R^{(1*)}$ has lower μ value (since $R^{(00)}$ is large) and srec no larger than that of the entire space. The same is the case when $\mu^{(10)}$ satisfies $\mathsf{srec}^{1,\mu^{(10)}}_{\varepsilon^{(10)}+30\sqrt[4]{\delta},\delta}(f) \leqslant 0.9D$. This suggests a natural inductive protocol Π to compute the function f. This is formalized in the following lemma.

First, some notation. For our induction it will be convenient to work with μ that are not necessarily normalized. So, we will only assume $\mu : \mathcal{X} \times \mathcal{Y} \to [0,1]$ but not that $|\mu| = \mu(\mathcal{X} \times \mathcal{Y}) = 1$. For a protocol Π, let the advantage of Π be defined by

$$\mathsf{adv}_\mu(\Pi) = \sum_{(x,y):f(x,y)=\Pi(x,y)} \mu(x,y) - \sum_{(x,y):f(x,y)\neq\Pi(x,y)} \mu(x,y).$$

Let $L(\Pi)$ be the number of leaves in Π. We now formulate the induction hypothesis as follows.

Lemma 2.8. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and product distribution $\mu : \mathcal{X} \times \mathcal{Y} \to [0,1]$. Let $\varepsilon, \delta \in (0,1)$ and $\Delta \in (0,1)$. Let*

$$s = s(\mu,\varepsilon,\delta) := \left\lceil 100 \log \mathsf{srec}^{1,\mu}_{\varepsilon,\delta}(f) \right\rceil;$$
$$t = t(\mu,\varepsilon,\delta) := \lceil 1002^s \log(|\mu|/\Delta) \rceil.$$

Then, there is a protocol Π such that

$$L(\Pi) \leqslant 4\binom{s+t}{t} - 1;$$
$$\mathsf{adv}_\mu(\Pi) \geqslant \left(\frac{1}{10} - \varepsilon - 30(s+1)\sqrt[4]{\delta}\right)|\mu| - \Delta \cdot L(\Pi).$$

Proof. First, we observe that if $\max\{\mu_0, \mu_1\} \geqslant 2\min\{\mu_0, \mu_1\}$, then the protocol Π consisting of just one leaf, with the most popular value as label, meets the requirements: for, $\mathsf{adv}_\mu(\Pi) \geqslant \frac{1}{3}$ and $L(\Pi) = 1$, and our claim holds. So, we proceed by induction on $s + t$, assuming that μ is balanced: $\max\{\mu_0, \mu_1\} \leqslant 2\min\{\mu_0, \mu_1\}$.

Base case ($s = 0$)**:** That is, $\log \mathsf{srec}^{1,\mu}_{\varepsilon,\delta}(f) \leqslant \frac{1}{100}$. We will show a protocol Π where Alice sends one bit after which Bob announces the answer. Consider the optimal solution $\langle w_R : R \text{ a rectangle}\rangle$ to the LP corresponding to $\mathsf{srec}^{1,\mu}_{\varepsilon,\delta}(f)$; thus, $\mathsf{OPT} := \sum_R w_R = \mathsf{srec}^{1,\mu}_{\varepsilon,\delta}(f) \leqslant 2^{1/100} \leqslant 2$. Let $R = R_X \times R_Y$ be a random rectangle picked with probability proportional to w_R (using public coins). In the protocol Π, Alice tells Bob if $x \in R_X$, and Bob returns the answer 1 if $(x, y) \in R_Y$ and returns 0 otherwise. Let $p_{xy} := \Pr_R[(x, y) \in R]$. Then, $\sum_{(x,y)\in f^{-1}(1)} \mu(x, y)p_{xy} \geqslant (1 - \varepsilon)\mu_1/\mathsf{OPT}$ and $\sum_{(x,y)\in f^{-1}(0)} \mu(x, y)p_{xy} \leqslant \delta\mu_0/\mathsf{OPT}$. Thus,

$$\begin{aligned}
\mathbb{E}_R\left[\sum_{(x,y):\Pi(x,y)\neq f(x,y)} \mu(x,y)\right] &= \sum_{x,y\in f^{-1}(1)} \mu(x,y)(1-p_{xy}) + \sum_{x,y\in f^{-1}(0)} \mu(x,y)p_{x,y} \\
&\leqslant \mu_1 - (1-\varepsilon)\mu_1/\mathsf{OPT} + \delta\mu_0/\mathsf{OPT} \\
&\leqslant \mu_1 - ((1-\varepsilon)\mu_1 - \delta\mu_0)/\mathsf{OPT} \\
&\leqslant \frac{1}{2}(\mu_1 + \varepsilon\mu_1 + \delta\mu_0).
\end{aligned}$$

We fix a choice R such that the left hand side is at least this quantity. Then,

$$\begin{aligned}
\mathsf{adv}(\Pi) &= |\mu| - 2\sum_{(x,y):\Pi(x,y)\neq f(x,y)} \mu(x,y) \\
&\geqslant |\mu| - \mu_1 - \varepsilon\mu_1 - \delta\mu_0 \\
&\geqslant \left(\frac{1}{3} - \varepsilon - \delta\right)|\mu|.
\end{aligned}$$

Base case ($t = 0$)**:** In this case, $\mu \leqslant \Delta$, and the protocol Π that gives the most probable answer achieves $\mathsf{adv}(\Pi) \geqslant 0 \geqslant \mu - \Delta$. This protocol has a single leaf.

Induction step: We will use Lemma 2.6 to decompose the communication matrix into a small number of rectangles. After an exchange of a few bits to determine in which rectangle the input lies, Alice and Bob will be left with a problem where one of the parameters, s or t, is significantly smaller.

Formally, from Lemma 2.6, we obtain a large 1-biased rectangle $R^{(00)} = X_0 \times Y_0$ such that $\mu(R^{(00)}) \geqslant \frac{1}{3\cdot 2^{s/100}}(1 - \varepsilon - 2\sqrt{\delta})|\mu|$ (we take $\rho := \sqrt{\delta}$). Recall the definitions of the rectangles $R^{(10)}, R^{(01)}, R^{(11)}$ and the corresponding restrictions of μ: $\mu^{(01)}, \mu^{(10)}, \mu^{(11)}$. Now, one of the alternatives mentioned in Lemma 2.7 holds: suppose, we have

$$\mathsf{srec}^{1,\mu^{(01)}}_{\varepsilon^{(01)}+30\sqrt[4]{\delta},\delta}(f) \leqslant 0.9\mathsf{srec}^{1,\mu}_{\varepsilon,\delta}(f).$$

Recall $R^{(1*)} = R^{(10)} \cup R^{(11)}$ and $\mu^{(1*)}$ is the restriction of μ to $R^{(1*)}$. Observe that

$$|\mu^{(1*)}| \leqslant |\mu| - \mu(R^{(00)}) \leqslant |\mu| \left(1 - \frac{1}{3 \cdot 2^{s/100}}(1 - \varepsilon - 2\sqrt{\delta})\right) \leqslant |\mu| \left(1 - \frac{1}{3 \cdot 2^{s}}\right). \tag{2.4}$$

The protocol Π proceeds as follows: Alice informs Bob if $x \in X_0$. If $x \notin X_0$, Alice and Bob follow the protocol $\Pi^{(1*)}$ for $\mu^{(1*)}$ guaranteed by the induction hypothesis (by (2.4), we note that t for the subproblem $R^{(1*)}$ is at most the original t minus 1); otherwise, Bob informs Alice if $y \in Y_0$. If $(x, y) \in R^{(00)}$, they immediately output the most promising answer (i.e., 1); otherwise, Alice and Bob follow the protocol $\Pi^{(01)}$ promised by induction for $\mu^{(01)}$. We then, have the following:

$$\begin{aligned}
L(\Pi) &= 1 + L(\Pi^{(1*)}) + L(\Pi^{(01)}) \\
&\leqslant 1 + \left(4\binom{s + (t-1)}{t-1} - 1\right) + \left(4\binom{(s-1) + t}{t} - 1\right) \\
&= 4\binom{s+t}{t} - 1; \\
\mathsf{adv}(\Pi) &\geqslant |\mu^{(00)}| \cdot (1 - \sqrt{\delta}) \\
&\quad + |\mu^{(01)}| \cdot \left(\frac{1}{10} - (\varepsilon^{(01)} + 30\sqrt[4]{\delta}) - 30s\sqrt[4]{\delta}\right) - \Delta \cdot L(\Pi^{(01)}) \\
&\quad + |\mu^{(1*)}| \cdot \left(\frac{1}{10} - \varepsilon^{(1*)} - 30(s+1)\sqrt[4]{\delta}\right) - \Delta \cdot L(\Pi^{(1*)}) \\
&\geqslant \left(\frac{1}{10} - \varepsilon - 30(s+1)\sqrt[4]{\delta}\right) |\mu| - \Delta \cdot L(\Pi).
\end{aligned}$$

□

The above induction lemma yields a protocol whose corresponding protocol tree has bounded number of leaves, but not necessarily low depth. We can easily convert this protocol Π into another Π' with low depth using the following proposition.

Proposition 2.9 ([10, Lemma 2.8]). *If f has a determinstic communication protocol tree with l leaves, then f has a protocol tree with depth at most $O(\log l)$.*

With this, we are in a position to complete the proof of the main theorem of this section.

Proof of Theorem 2.4. To prove the first part of Theorem 2.4, we invoke Lemma 2.8 with $\Delta = 1/2^{4n}$ and $\varepsilon = \delta = 1/n^2$ to derive a protocol tree Π with at most

$$L(\Pi) = n^{O\left(\log \mathsf{srec}^{1,\mu}_{1/n^2, 1/n^2}(f)\right)^2}$$

leaves and advantage at least $1/20$. The first part now follows from Proposition 2.9.

To prove the final conclusion of Theorem 2.4, we invoke Lemma 2.8 with $\Delta = 1/2^{5k}$ and $\varepsilon = \delta = 1/(30 \cdot 100(k+1)^4)$ where k satisfies the hypothesis. With this setting of parameters $t = \lceil 500 \cdot 2^k k \rceil \leqslant 2^{2k}$ for $k \geqslant 20$. Lemma 2.8 implies a protocol tree Π with at most

$$L(\Pi) \leqslant (t+s)^s \leqslant t^{2s} \leqslant 2^{4k^2}$$

leaves and advantage at most $1/20$.The conclusion follows from Proposition 2.9. □

2.3 Proofs of Lemmas 2.6 and 2.7

Proof of Lemma 2.6. We will call a rectangle R *ρ-biased* (towards z) if $\mu_{\bar{z}}(R) < \rho \cdot \mu_z(R)$ and *ρ-unbiased* (towards z) otherwise.

Fiz an solution $\langle w_R : R \text{ is a rectangle } \rangle$ that achieves the optimum D. It follows from the packing constraints (2.2) that

$$\begin{aligned}
\sum_{R:\rho\text{-unbiased}} w_R \cdot \mu_z(R) &\leqslant \sum_{R:\rho\text{-unbiased}} w_R \cdot \frac{\mu_{\bar{z}}(R)}{\rho} & &\leqslant \frac{1}{\rho} \cdot \sum_{R:\rho\text{-unbiased}} w_R \cdot \mu_{\bar{z}}(R) \\
&\leqslant \frac{1}{\rho} \cdot \sum_R w_R \cdot \mu_{\bar{z}}(R) & &= \frac{1}{\rho} \sum_R w_R \sum_{(x,y) \in f^{-}(\bar{z}) \cap R} \mu(x,y) \\
&= \frac{1}{\rho} \sum_{(x,y)\in f^{-1}(\bar{z})} \mu(x,y) \sum_{R:(x,y)\in R} w_R & &\leqslant \frac{\delta}{\rho} \cdot \mu_{\bar{z}}.
\end{aligned}$$

We now use the covering constraints (2.1) to conclude that

$$\sum_{R:\rho\text{-biased}} w_R \cdot \mu_z(R) = \sum_R w_R \cdot \mu_z(R) - \sum_{R:\rho\text{-unbiased}} w_r \cdot \mu_z(R) \geqslant (1-\varepsilon)\cdot \mu_z - \frac{\delta}{\rho} \cdot \mu_{\bar{z}}. \tag{2.5}$$

Define a probability distribution on the rectangles R as follows $p(R) := w_R/D$. Then (2.5) can be rewritten as

$$\mathbb{E}_R \left[\mathbb{1}_{\rho\text{-biased}}(R) \cdot \mu_z(R) \right] \geqslant \frac{1}{D} \cdot \left((1-\varepsilon)\cdot \mu_z - \frac{\delta}{\rho} \cdot \mu_{\bar{z}} \right).$$

Hence, there exists a large biased rectangle $S = X_0 \times Y_0$ as stated in the claim. $\square$

Proof of Lemma 2.7. Since $R^{(00)}$ is $\sqrt{\delta}$-biased (towards 0), we have from the packing and covering constraints (2.2) and (2.1) that

$$\begin{aligned}
\sum_{(x,y)\in R^{(00)}} \mu_{x,y} \sum_{R \ni (x,y)} w_R &= \sum_{(x,y)\in R^{(00)} \cap f^{-1}(1)} \mu_{x,y} \sum_{R\ni(x,y)} w_R + \sum_{(x,y)\in R^{(00)} \cap f^{-1}(0)} \mu_{x,y} \sum_{R\ni(x,y)} w_R \\
&\leqslant \mu_1(R^{(00)}) + \delta\mu_0(R^{(00)}) \leqslant (\sqrt{\delta}+\delta)\mu_0(R^{(00)}) \leqslant 2\sqrt{\delta}\mu(R^{(00)}).
\end{aligned}$$

$$\text{Hence, } \sum_R w_R \left(\frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \right) \leqslant 2\sqrt{\delta}. \tag{2.6}$$

Define subsets of rectangles as follows:

$$B^{(01)} := \left\{ R : \frac{\mu(R^{(01)} \cap R)}{\mu(R^{(01)})} \geqslant \frac{10\sqrt[4]{\delta}}{D} \right\}, \qquad B^{(10)} := \left\{ R : \frac{\mu(R^{(10)} \cap R)}{\mu(R^{(10)})} \geqslant \frac{10\sqrt[4]{\delta}}{D} \right\},$$

$$B := \left\{ R : \frac{\mu(R^{(11)} \cap R)}{\mu(R^{(11)})} \geqslant \frac{10}{D} \right\}.$$

Using (2.3), we have

$$\sum_{(x,y)\in R^{(11)}} \mu_{x,y} \sum_{R\ni(x,y)} w_R \leqslant \sum_{(x,y)\in R^{(11)}} \mu_{x,y} = \mu(R^{(11)}).$$

Or equivalently,

$$\sum_R \frac{w_R}{D} \cdot \frac{\mu(R^{(11)} \cap R)}{\mu(R^{(11)})} \leqslant \frac{1}{D}.$$

Hence,

$$\sum_{R \in B} w_R \leqslant 0.1D.$$

We will now argue that either $\sum_{R \in B^{(01)}} w_R \leqslant 0.9D$ or $\sum_{R \in B^{(10)}} w_R \leqslant 0.9D$. Suppose neither is true. Then, combining with the above, we have that

$$\sum_{R \in B^{(01)} \cap B^{(10)} \cap \bar{B}} w_R \geqslant 0.7D. \tag{2.7}$$

We now use the fact that μ is a product distribution to infer that for all rectangles R, we have

$$\frac{\mu(R^{(01)} \cap R)}{\mu(R^{(01)})} \cdot \frac{\mu(R^{(10)} \cap R)}{\mu(R^{(10)})} = \frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \cdot \frac{\mu(R^{(11)} \cap R)}{\mu(R^{(11)})}.$$

Using the above we have

$$\sum_{R \in B^{(01)} \cap B^{(10)} \cap \bar{B}} w_R \left(\frac{10\sqrt[4]{\delta}}{D} \cdot \frac{10\sqrt[4]{\delta}}{D} \right) \leqslant \sum_{R \in B^{(01)} \cap B^{(10)} \cap \bar{B}} w_R \left(\frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \cdot \frac{10}{D} \right).$$

Combining with (2.7), we have

$$\sum_R w_R \left(\frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \right) \geqslant \sum_{R \in B^{(01)} \cap B^{(10)} \cap \bar{B}} w_R \left(\frac{\mu(R^{(00)} \cap R)}{\mu(R^{(00)})} \right) \geqslant \frac{10\sqrt{\delta}}{D} \cdot (0.7D) = 7\sqrt{\delta}.$$

This contradicts (2.6). Hence, either $\sum_{R \in B^{(01)}} w_R \leqslant 0.9D$ or $\sum_{R \in B^{(10)}} w_R \leqslant 0.9D$. Assume, wlog. that $\sum_{R \in B^{(01)}} w_R \leqslant 0.9D$. If f is $1/2$-biased towards 0 with respect to the distribution $\mu^{(01)}$, then we are done. Suppose otherwise, that is $0.5\mu_0(R^{(01)}) \leqslant \mu_1(R^{(01)})$ or equivalently $\mu(R^{(01)}) \leqslant 3\mu^{(01)}(R^{(01)})$. We will infer from this that $\mathsf{srec}^{1,\mu^{(01)}}_{\varepsilon^{(01)}+30\sqrt[4]{\delta},\delta}(f) \leqslant 0.9D$. Consider the primal solution given by

$$w'_R = \begin{cases} w_R, & \text{if } R \in B^{(01)} \\ 0, & \text{if } R \notin B^{(01)}. \end{cases}$$

Clearly, w'_R being a sub-solution satisifies (2.2) and (2.3) and has objective value at most $0.9D$. All we need to show is that it satisfies the covering constraint (2.1). For this, we first consider

$$\sum_{R \in \overline{B^{(01)}}} w_R \left(\frac{\mu_1(R^{(01)} \cap R)}{\mu(R^{(01)})} \right) \leqslant \sum_{R \in \overline{B^{(01)}}} w_R \left(\frac{\mu(R^{(01)} \cap R)}{\mu(R^{(01)})} \right) \leqslant \frac{10\sqrt[4]{\delta}}{D} \cdot D \leqslant 10\sqrt[4]{\delta} \tag{2.8}$$

Now,

$$
\begin{aligned}
&\sum_{(x,y)\in f^{-1}(1)\cap R^{(01)}} \mu_{x,y} \sum_{R\in(x,y)} w'_R \\
&= \sum_{(x,y)\in f^{-1}(1)\cap R^{(01)}} \mu_{x,y} \sum_{R\in(x,y),R\in B^{(01)}} w_R \\
&= \sum_{(x,y)\in f^{-1}(1)\cap R^{(01)}} \mu_{x,y} \left(\sum_{R\in(x,y)} w_R - \sum_{R\in(x,y),R\notin B^{(01)}} w_R \right) \\
&= (1-\varepsilon^{(01)})\mu_1(R^{(01)}) - \sum_{(x,y)\in f^{-1}(1)\cap R^{(01)}} \mu_{x,y} \sum_{R\in(x,y),R\notin B^{(01)}} w_R \\
&= (1-\varepsilon^{(01)})\mu_1(R^{(01)}) - \sum_{R\notin B^{(01)}} w_R \mu_1(R^{(01)}\cap R) \\
&\geqslant (1-\varepsilon^{(01)})\mu_1(R^{(01)}) - 10\sqrt[4]{\delta}\mu(R^{(01)}) && \text{[From (2.8)]} \\
&\geqslant (1-\varepsilon^{(01)})\mu_1(R^{(01)}) - 30\sqrt[4]{\delta}\mu_1(R^{(01)}) && \text{[Since } \mu(R^{(01)}) \leqslant 3\mu_1(R^{(01)})\text{]} \\
&= (1-\varepsilon^{(01)} - 30\sqrt[4]{\delta})\mu_1(R^{(01)})
\end{aligned}
$$

Thus, (2.1) holds with ε replaced by $(\varepsilon^{(01)} + 30\sqrt[4]{\delta})$.

□

3 Query Complexity

In the following, $f : \{0,1\}^n \to \{0,1\}$ is the function for which we wish to build a decision tree. Let $\mu : \{0,1\}^n \to [0,1]$ be a bit-wise distribution on the inputs to f. Our goal is to build an efficient decision tree $\mathcal{T}$ for f such that $\Pr_\mu[f(x) \neq \mathcal{T}(x)]$ is small.

3.1 Preliminaries

Definition 3.1 (product distribution)**.** *We say that μ is a (bit-wise) product distribution on $\{0,1\}^n$ if there exist positive reals $(p_i(0), p_i(1))$ (for $i = 1,2,\ldots,n$ satisfying $p_i(0) + p_i(1) = 1$) such that $\mu(x) = \prod_i p_i(x_i)$.*

Let $s \in \{0,1,\star\}^n$. The subcube of $\{0,1\}^n$ with support s is

$$\mathsf{subcube}(s) := \{x \in \{0,1\}^n : s_i \in \{0,1\}^n \Rightarrow x_i = s_i\}.$$

The size of such a subcube, denoted by $\mathsf{size}(s)$ is

$$\mathsf{size}(s) := |\{i : s_i \in \{0,1\}\}|.$$

Various linear programming bounds are known for the query complexity such as the partition bound, relaxed partition bound, smooth rectangle bound (see Appendix A for the exact definition of these bounds and their inter-relationships).

Let $\varepsilon > 0$. In the following, A represents a subcube of $\{0,1\}^n$.

Definition 3.2 (Relaxed query partition bound []). *The ε-relaxed query partition bound of f, denoted $\mathsf{rqprt}_\varepsilon(f)$, is given by the optimal value of the following linear program.*

$$\begin{aligned}
\min \quad & \sum_z \sum_A w_{z,A} \cdot 2^{|A|} & \\
& \sum_{A: x\in A} w_{f(x),A} \geqslant 1-\varepsilon, & \forall x \\
& \sum_{A: x\in A} \sum_z w_{z,A} \leqslant 1, & \forall x \\
& w_{z,A} \geqslant 0, & \forall (z,A)
\end{aligned}$$

The error in the above definition can be reduced using the following boosting claim (proved in the appendix).

Claim 3.3. *Let $\varepsilon > \delta > 0$. Then* $\log \mathsf{rqprt}_\delta(f) \leqslant \left(\frac{2}{(0.5-\varepsilon)^2}\log\frac{1}{\delta}\right)\log \mathsf{rqprt}_\varepsilon(f)$.

Definition 3.4. *We say μ is an $(\alpha_0, \beta_0, \alpha_1, \beta_1, a, b)$-feasible distribution for f if there exists a feasible solution to the following inequalities. The variables are (u_R : R is a subcube with support of size at most a) and (w_R : R is a subcube with support of size at most b).*

$$\sum_{R: x\in R} u_R \geqslant 1-\alpha_0 \qquad \forall x \in f^{-1}(0) \tag{3.1}$$

$$\sum_{R: x\in R} u_R \leqslant \beta_0 \qquad \forall x \in f^{-1}(1); \tag{3.2}$$

$$u_R \geqslant 0. \tag{3.3}$$

$$\sum_R \mu_1(R) w_R \geqslant (1-\alpha_1)\mu_1 \tag{3.4}$$

$$\sum_{R: x\in R} w_R \leqslant 1 \qquad \forall x \in \{0,1\}^n; \tag{3.5}$$

$$\sum_{R: x\in R} w_R \leqslant \beta_1 \qquad \forall x \in f^{-1}(0); \tag{3.6}$$

$$w_R \geqslant 0. \tag{3.7}$$

Remark: If the i-th bit of the input is fixed in μ (that is, $p_i(0) = 0$ or $p_i(0) = 1$), then we will assume i is not part of the support of any R in the above feasible solution.

The main result of this section is the following lemma.

Lemma 3.5. *Let $\delta > 0$ and μ be a product distribution that is $(\alpha_0, \beta_0, \alpha_1, \beta_1, a, b)$-feasible for f. Then, there is a decision tree for f of depth at most ab that errs with probability at most*

$$\frac{1}{4} + \alpha_1 + \beta_1 + 4b(\beta_1 + \delta) + \frac{\beta_0}{(1-\alpha_0)\delta}.$$

The query complexity bound stated in the introduction (Theorem 1.2) follows from the above lemma.

Proof of Theorem 1.2. Let $c := \log \mathsf{rqprt}_\varepsilon(f)$. Let $\delta = 1/c^8$. Then from Claim 3.3 we get that $d := \log \mathsf{rqprt}_\delta(f) = O(c \log c)$. Let $\{w_{z,A}\}$ be an optimal solution for the primal of $\mathsf{rqprt}_\delta(f)$. Let $\mathcal{B} := \{A : |A| > d + \log \frac{1}{\delta}\}$. Then $\sum_z \sum_{A \in \mathcal{B}} w_{z,A} < \delta$ since $\sum_z \sum_A w_{z,A} \cdot 2^{|A|} = 2^d$. This implies (by first boosting and then removing the $A \in \mathcal{B}$) that μ is an $(\alpha_0, \beta_0, \alpha_1, \beta_1, a, b)$-feasible distribution for f, with $\alpha_0 = \beta_0 = \alpha_1 = \beta_1 = 2\delta$ and $a = b = O(c \log c)$.

From Lemma 3.5 (by setting $\delta = 1/c^4$) we get that there is a decision tree for f of depth at most $O(c^2 \log^2 c)$ with error under μ at most $\frac{1}{4} + O(\frac{1}{c^2})$. □

3.2 Proof of Lemma 3.5

In this section, we show that if a product distribution μ is feasible, then the functions admits a decision tree of low complexity. This decision tree is obtained from the feasible solution of the LP as follows. We first show that feasibility implies the existence of a biased subcube of small support (see Claim 3.6). After querying the support of this subcube, one is left with several subproblems. One of the subproblems corresponds to the subcube itself, in which case we answer according to its bias. For each of the other subproblems, we observe that the induced distribution μ admits a feasible solution consisting of rectangles with a strictly smaller support size. This is proved by showing that the contribution of rectangles whose supports are disjoint to the original subcube is negligible (see Claim 3.7). This step crucially uses the product nature of the distribution μ.

For a set $A \subseteq \{0,1\}^n$, let $\mu_0(A) = \mu(A \cap f^{-1}(0))$ and $\mu_1(A) = \mu(A \cap f^{-1}(1))$; let $\mu_0 = \mu_0(\{0,1\}^n)$ and $\mu_1 = \mu_1(\{0,1\}^n)$.

Claim 3.6. *Suppose $\mu : \{0,1\}^n \to [0,1]$ is a product probability distribution satisfying (3.1) , (3.2) and (3.3). Further, suppose $\delta > 0$ is such that*

$$(1 - \alpha_0)\mu_0 - \left(\frac{\beta_0}{\delta}\right)\mu_1 > 0. \tag{3.8}$$

Then, there is subcube A with support of size at most a, such that $\mu_1(A) \leqslant \delta \mu_0(1)$.

Proof. We say A is biased if $\mu_1(A) \leqslant \delta\mu_0(A)$. From (3.2), we have

$$\begin{aligned}
\beta_0\mu_1 &\geqslant \sum_A \mu_1(A)u_A & (3.9)\\
&\geqslant \sum_{A \text{ not biased}} \mu_1(A)u_A & (3.10)\\
&\geqslant \delta \sum_{A \text{ not biased}} \mu_0(A)u_A. & (3.11)
\end{aligned}$$

Combining this with (3.1), we obtain

$$\begin{aligned}
(1 - \alpha_0)\mu_0 - \left(\frac{\beta_0}{\delta}\right)\mu_1 &\leqslant \sum_A u_A\mu_0(A) - \sum_{A \text{ unbiased}} u_A\mu_0(A) & (3.12)\\
&\leqslant \sum_{A \text{ biased}} u_A\mu_0(A). & (3.13)
\end{aligned}$$

Since the left hand side is positive, the sum on the right cannot be empty. The claim follows from this. □

Claim 3.7. *Let $\delta > 0$. Fix a product distribution μ, and let A be a subcube such that $\mu_1(A) \leqslant \delta\mu_0(A)$. Suppose $(w_R : R \text{ a subcube})$ satisfies (3.5) and (3.6). Let $\mathcal{B} = \{B : \mathsf{support}(B) \cap \mathsf{support}(A) = \varnothing\}$. Then,*

$$\sum_{B \in \mathcal{B}} \mu_1(B) w_B \leqslant \beta_1 + \delta,$$

Proof.

$$\sum_{B \in \mathcal{B}} \mu_1(B) w_B \leqslant \sum_{B \in \mathcal{B}} \mu(B) w_B \tag{3.14}$$

$$\leqslant \frac{1}{\mu(A)} \sum_{B \in \mathcal{B}} \mu(A \cap B) w_B \tag{3.15}$$

$$\leqslant \frac{1}{\mu(A)} \sum_{B \in \mathcal{B}} \mu_0(A \cap B) w_B + \frac{1}{\mu(A)} \sum_{B \in \mathcal{B}} \mu_1(A \cap B) w_B. \tag{3.16}$$

Let us bound the two terms on the right separately. For the first term, by our assumption (3.6), we have

$$\sum_{B \in \mathcal{B}} w_B \mu_0(A \cap B) \leqslant \beta_1 \mu_0(A) \leqslant \beta_1 \mu(A).$$

For second term, by assumption (3.5), we have

$$\sum_{B \in \mathcal{B}} \mu_1(A \cap B) w_B \leqslant \mu_1(A) \leqslant \delta\mu_0(A) \leqslant \delta\mu(A).$$

By using these bounds in (3.16), we establish our claim. □

Proof of Lemma 3.5. We will prove the theorem by induction on b. We assume that

$$\mu_0, \mu_1 \geqslant \frac{1}{4},$$

for otherwise, we can reliably guess the answer without making any query. Similarly, we assume that (3.8) holds, for otherwise, we have

$$\mu_0 \leqslant \frac{\beta_0 \mu_1}{(1 - \alpha_0)\delta},$$

and we may answer 1 without making any query, and still keep the error within bounds.

Base case ($b = 0$): The only subcube R that may appear in the inequalities (3.1)–(3.6) is the one with empty support (R contains all inputs). Since $\mu_0, \mu_1 \geqslant \frac{1}{4}$; we then obtain from (3.4) and (3.6) that $1 - \alpha_1 \leqslant w_R \leqslant \beta$ or $\alpha_1 + \beta_1 \geqslant 1$; so the claim holds trivially.

Induction step ($b \geqslant 1$): Using (3.1) and (3.2) and Claim 3.6, we conculde that there is a rectangle A_0 with bias δ. We start by querying the bits in the support of A_0. For each result $\sigma \in \{0,1\}^a$, we are left with a new product distribution μ^σ on the inputs for which we need to eventually guess the answer.

We proceed as follows. By Claim 3.7, we have, as B ranges over subcubes whose supports are disjoint from A_0's, that

$$\sum_B \mu_1(B)w_B \leqslant \beta_1 + \delta.$$

It follows from (3.4) that (now summing over all R whose supports intersect A_0's)

$$\begin{aligned}\sum_R \mu_1(R)w_R &\geqslant (1-\alpha_1)\mu_1 - \beta_1 - \delta \\ &\geqslant (1-(\alpha_1 + 4\beta_1 + 4\delta))\mu_1. \end{aligned} \tag{3.17}$$

For each outcome σ for the bits queried, let μ^σ be the resulting conditional distribution on inputs, where variables in support of A_0 are fixed at σ in μ^σ. Set $w_R = 0$ for R whose support is disjoint from A_0's, and define α_1^σ by

$$\sum_R \mu_1^\sigma(R)w_R := (1-\alpha_1^\sigma)\mu_1^\sigma.$$

Then, μ^σ is an $(\alpha_0, \beta_0, \alpha_1^\sigma, \beta_1, a, b-1)$-product distribution for f (recall our convention that we do not include the index of a fixed bit in the support of our subcubes). Furthermore, by (3.17)

$$\mathbb{E}[\alpha_1^\sigma] \leqslant \alpha_1 + 4\beta_1 + 4\delta. \tag{3.18}$$

By induction, μ^σ it admits a decision tree of depth at most $a(b-1)$ that errs with probability at most

$$\varepsilon^\sigma \leqslant \frac{1}{4} + \alpha_1^\sigma + \beta_1 + 4(b-1)(\beta_1+\delta) + \frac{\beta_0}{(1-\alpha_0)\delta},$$

when inputs are drawn according to μ^σ. It follows from (3.18), the overall error is

$$\begin{aligned}\mathbb{E}_\sigma[\varepsilon^\sigma] &\leqslant \mathbb{E}_\sigma\left[\frac{1}{4} + \alpha_1^\sigma + \beta_1 + 4(b-1)(\beta_1+\delta) + \frac{\beta_0}{(1-\alpha_0)\delta}\right] \\ &\leqslant \frac{1}{4} + \alpha_1 + \beta_1 + 4b(\beta_1+\delta) + \frac{\beta_0}{(1-\alpha_0)\delta}.\end{aligned}$$

□

References

[1] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM J. Comput.*, 42(3):1327–1363, 2013. (Preliminary Version in *42nd STOC*, 2010). doi:10.1137/100811969. 1

[2] Lila Fontes, Rahul Jain, Iordanis Kerenidis, Sophie Laplante, Mathieu Laurière, and Jérémie Roland. Relative discrepancy does not separate information and communication complexity. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Proc. 42nd International Colloq. of Automata, Languages and Programming (ICALP), Part I*, volume 9134 of *LNCS*, pages 506–516. Springer, 2015. doi:10.1007/978-3-662-47672-7_41. 1

[3] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *Proc. 55th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 176–185, 2014. doi:10.1109/FOCS.2014.27. 1

[4] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of communication and external information. Technical Report TR15-088, Elect. Colloq. on Comput. Complexity (ECCC), 2015. URL: http://eccc.hpi-web.de/report/2015/088/. 1

[5] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. In *Proc. 47th ACM Symp. on Theory of Computing (STOC)*, pages 557–566, 2015. doi:10.1145/2746539.2746572. 1

[6] Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proc. 25th IEEE Conf. on Computational Complexity*, pages 247–258, 2010. arXiv:0910.4266, doi:10.1109/CCC.2010.31. 1, 2, 15, 16, 17

[7] Rahul Jain and Penghui Yao. A strong direct product theorem in terms of the smooth rectangle bound. 2012. arXiv:1209.0263. 2

[8] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Proc. 53th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 500–509, 2012. arXiv:1204.1505, doi:10.1109/FOCS.2012.68. 0, 1, 17

[9] Gillat Kol. Interactive compression for product distributions. Technical Report TR15-168, Elect. Colloq. on Comput. Complexity (ECCC), 2015. URL: http://eccc.hpi-web.de/report/2015/168/. 0, 1

[10] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997. doi:10.1017/CBO9780511574948. 2, 4, 7

[11] Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995. (Preliminary version in *35th FOCS*, 1994). doi:10.1007/BF01192527. 4

[12] Clifford D. Smyth. Reimer's inequality and Tardos' conjecture. In *Proc. 34th ACM Symp. on Theory of Computing (STOC)*, pages 218–221, 2002. doi:10.1145/509907.509942. 2

[13] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proc. 11th ACM Symp. on Theory of Computing (STOC)*, pages 209–213, 1979. doi:10.1145/800135.804414. 2

[14] Andrew Chi-Chih Yao. Lower bounds by probabilistic arguments (extended abstract). In *Proc. 24th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 420–428, 1983. doi:10.1109/SFCS.1983.30. 2

A Query Complexity LP bounds

Definition A.1 (Partition bound [6])**.** *The ε-query partition bound of f, denoted $\mathsf{qprt}_\varepsilon(f)$, is given by the optimal value of the following linear program.*

Primal

$$\min \quad \sum_z \sum_A w_{z,A} \cdot 2^{|A|}$$

s.t.

$$\forall x : \quad \sum_{A: x \in A} w_{f(x),A} \geqslant 1 - \varepsilon$$

$$\forall x : \quad \sum_{A: x \in A} \sum_z w_{z,A} = 1$$

$$\forall (z, A) : \quad w_{z,A} \geqslant 0$$

Dual

$$\max \quad (1-\varepsilon) \sum_x \mu_x + \sum_x \varphi_x$$

s.t.

$$\forall (z, A) : \quad \sum_{x \in A \cap f^{-1}(z)} \mu_x + \sum_{x \in A} \varphi_x \leqslant 2^{|A|}$$

$$\forall x : \quad \mu_x \geqslant 0, \varphi_x \in \mathbb{R}$$

The success for relaxed partition bound can be boosted as in the following claim.

Claim A.2. *Let $\varepsilon > \delta > 0$. Then* $\log \mathsf{rqprt}_\delta(f) \leqslant \left(\frac{2}{(0.5-\varepsilon)^2} \log \frac{1}{\delta}\right) \log \mathsf{rqprt}_\varepsilon(f)$.

Proof. Let t be an odd number. Let $\{w_{z,A}\}$ be an optimal solution for the primal of $\mathsf{qprt}_\varepsilon(f)$. Define, G_z be the set of all t-tuples $(z_1, \ldots, z_t)$ such that z is the majority of $(z_1, \ldots, z_t)$. Let G_A be the set of all t-tuples $(A_1, \ldots, A_t)$ such that A is the intersection of $(A_1, \ldots, A_t)$. Define

$$v_{z,A} := \sum_{(z_1,\ldots,z_t)\in G_z,(A_1,\ldots,A_t)\in G_A} \left(\Pi_{i=1}^t w_{z_i,A_i}\right).$$

Fix x. Consider,

$$\begin{aligned}
\sum_z \sum_{A\ni x} v_{z,A} &= \sum_z \sum_{A\ni x} \sum_{(z_1,\ldots,z_t)\in G_z,(A_1,\ldots,A_t)\in G_A} \left(\Pi_{i=1}^t w_{z_i,A_i}\right) \\
&= \sum_{(z_1,\ldots,z_t),(A_1\ni x,\ldots,A_t\ni x)} \left(\Pi_{i=1}^t w_{z_i,A_i}\right) \\
&= \Pi_{i=1}^t \left(\sum_{z_i} \sum_{A_i\ni x} w_{z_i,A_i}\right) \\
&\leqslant 1.
\end{aligned}$$

$$\begin{aligned}
\sum_{R\ni x} v_{f(x),A} &= \sum_{A\ni x} \sum_{(z_1,\ldots,z_t)\in G_{f(x)},(A_1,\ldots,A_t)\in G_A} \left(\Pi_{i=1}^t w_{z_i,A_i}\right) \\
&= \sum_{(z_1,\ldots,z_t)\in G_{f(x)},(A_1\ni x,\ldots,A_t\ni x)} \left(\Pi_{i=1}^t w_{z_i,A_i}\right) \\
&\geqslant 1 - \exp(-(0.5-\varepsilon)^2 t/2). \quad \text{(using Chernoff bounds)}
\end{aligned}$$

$$\begin{aligned}
\sum_z \sum_A v_{z,A} &= \sum_z \sum_A \sum_{(z_1,\ldots,z_t)\in G_z,(A_1,\ldots,A_t)\in G_A} \left(\Pi_{i=1}^t w_{z_i,A_i}\right) \\
&= \sum_{(z_1,\ldots,z_t),(A_1,\ldots,A_t)} \left(\Pi_{i=1}^t w_{z_i,A_i}\right) \\
&= \Pi_{i=1}^t \left(\sum_{z_i} \sum_{A_i} w_{z_i,A_i}\right).
\end{aligned}$$

Hence the claim follows. □

B Communication Complexity LP bounds

The partition bound: The partition bound was introduced by Jain and Klauck [6] as a linear programming bound to lower bound the public-coin randomized communication complexity.

Definition B.1 (Partition bound [6]). *For a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and $\varepsilon \in (0,1)$, the ε-partition bound of f, denoted $\mathsf{qprt}_\varepsilon(f)$, is defined to be the optimal value of the following linear program. Below R represents a rectangle in $\mathcal{X} \times \mathcal{Y}$ and $(x,y,z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$.*

<u>*Primal*</u>

$$\begin{aligned}
\min \quad & \sum_z \sum_R w_{z,R} \\
\sum_{R:(x,y)\in R} w_{f(x,y),R} &\geqslant 1-\varepsilon, && \forall (x,y) \\
\sum_{R:(x,y)\in R} \sum_z w_{z,R} &= 1, && \forall (x,y) \\
w_{z,R} &\geqslant 0, && \forall (z,R) \ .
\end{aligned}$$

<u>*Dual*</u>

$$\begin{aligned}
\max \quad (1-\varepsilon) \sum_{(x,y)} \mu_{x,y} + \sum_{(x,y)} \varphi_{x,y} \\
\sum_{(x,y)\in R\cap f^{-1}(z)} \mu_{x,y} + \sum_{(x,y)\in R} \varphi_{x,y} &\leqslant 1, && \forall (z,R) \\
\mu_{x,y} &\geqslant 0, && \forall (x,y) \\
\varphi_{x,y} &\in \mathbb{R}, && \forall (x,y) \ .
\end{aligned}$$

It easily follows from the definition that $\mathsf{CC}^{\mathrm{pub}}_\varepsilon(f) \geqslant \log(\mathsf{prt}_\varepsilon(f))$.

Relaxed partition bound: Kerenidis *et al.* [8] defined a relaxation of the partition bound by relaxing the equality constraint in the primal.

Definition B.2 (Relaxed partition bound [8]). *For a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and $\varepsilon \in (0,1)$, the ε-relaxed partition bound of f, denoted $\mathsf{rprt}_\varepsilon(f)$, is given by the optimal value of the following linear program.*

<u>*Primal*</u>

$$\begin{aligned}
\min \quad & \sum_z \sum_R w_{z,R} \\
\sum_{R:(x,y)\in R} w_{f(x,y),R} &\geqslant 1-\varepsilon, && \forall (x,y) \\
\sum_{R:(x,y)\in R} \sum_z w_{z,R} &\leqslant 1, && \forall (x,y) \\
w_{z,R} &\geqslant 0, && \forall (z,R) \ .
\end{aligned}$$

<u>*Dual*</u>

$$\begin{aligned}
\max \quad (1-\varepsilon) \sum_{(x,y)} \mu_{x,y} - \sum_{(x,y)} \varphi_{x,y} \\
\sum_{(x,y)\in R\cap f^{-1}(z)} \mu_{x,y} - \sum_{(x,y)\in R} \varphi_{x,y} &\leqslant 1, && \forall (z,R) \\
\mu_{x,y} &\geqslant 0, && \forall (x,y) \\
\varphi_{x,y} &\geqslant 0, && \forall (x,y) \ .
\end{aligned}$$

Clearly, $\mathsf{CC}^{\mathrm{pub}}_\varepsilon(f) \geqslant \log(\mathsf{prt}_\varepsilon(f)) \geqslant \log(\mathsf{rprt}_\varepsilon(f)) \log \mathsf{srec}_{\varepsilon,\varepsilon}(f)$.

Proposition 2.5 follows from the above observation and the following boosting claim for rprt.

Claim B.3. *Let $\varepsilon > \delta > 0$. Then $\mathsf{rprt}_\delta(f) \leqslant O(\log \frac{1}{\delta}) \cdot \mathsf{rprt}_\varepsilon(f)$.*

ECCC ISSN 1433-8092

http://eccc.hpi-web.de