Direct Product Theorems for Classical Communication Complexity via Subdistribution Bounds

Rahul Jain [*]
U. Waterloo

Hartmut Klauck [†]
Goethe-Universität Frankfurt

Ashwin Nayak [‡]
U. Waterloo & Perimeter

December 17, 2007

Abstract

A basic question in complexity theory is whether the computational resources required for solving k independent instances of the same problem scale as k times the resources required for one instance. We investigate this question in various models of classical communication complexity.

We introduce a new measure, the *subdistribution bound*, which is a relaxation of the well-studied rectangle or corruption bound in communication complexity. We nonetheless show that for the communication complexity of Boolean functions with constant error, the subdistribution bound is the same as the latter measure, up to a constant factor. We prove that the one-way version of this bound tightly captures the one-way public-coin randomized communication complexity of any relation, and the two-way version bounds the two-way public-coin randomized communication complexity from below. More importantly, we show that the bound satisfies the strong direct product property under product distributions for both one- and two-way protocols, and the weak direct product property under arbitrary distributions for two-way protocols. These results subsume and strengthen, in a unified manner, several recent results on the direct product question.

The simplicity and broad applicability of our technique is perhaps an indication of its potential to solve yet more challenging questions regarding the direct product problem.

[*] School of Computer Science, and Institute for Quantum Computing, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L 3G1, Canada. Email: rjain@cs.uwaterloo.ca. Research supported in part by ARO/NSA USA.

[†] Institut für Informatik, Goethe-Universität Frankfurt, 60054 Frankfurt am Main, Germany. E-mail: hklauck@gmail.com. Supported by DFG grant KL 1470/1.

[‡] Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L 3G1, Canada. E-mail: anayak@math.uwaterloo.ca. Research supported in part by NSERC Canada, CIFAR, MITACS, CFI, OIT, and an ERA from the Province of Ontario. A.N. is also Associate Member, Perimeter Institute for Theoretical Physics, Waterloo, Canada. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.

1 Introduction

Consider two parties, Alice and Bob, who wish to communicate (classically) to solve several instances of the *same* computational problem. The problem is modeled as a relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, and Alice receives an input $x \in \mathcal{X}$, and Bob an input $y \in \mathcal{Y}$. The goal is to find an element $z \in \mathcal{Z}$ that satisfies the relation, i.e., to find a z such that $(x, y, z) \in f$. Given a communication protocol to solve f, a straightforward method for solving k instances of f is to run the protocol independently on each problem instance. This method has complexity that scales as k times the complexity of the original protocol. Moreover, when the protocol is randomized, and is guaranteed to succeed with probability at least $2/3$, then the probability of simultaneously succeeding on all k instances is only guaranteed to be at least $(2/3)^k$. A basic question in complexity theory is whether this method of solution is essentially optimal. A proof of its optimality is called a *strong direct product* theorem.

Direct product results and their variants appear in many different areas of complexity theory, ranging from hardness amplification in the theory of pseudo-randomness (see, e.g., [GNW95]), to parallel repetition in interactive proof systems (see, e.g., [Raz98, CSUU07]), to time-space trade-offs in concrete models of computation (for some recent examples, see [Aar04, KŠdW04, AŠdW06]). We concentrate on the setting of classical communication complexity. Forms of the direct product property for communication have repercussions for other areas of computational complexity. Karchmer, Raz, and Wigderson [KRW95] showed that a *direct sum* result for certain relations would imply $\mathsf{NC}^1 \neq \mathsf{NC}^2$. Bar-Yossef, Jayram, Kumar, and Sivakumar [BYJKS04] use direct sum results to place space lower bounds in the datastream model [BYJKS04]. Pătraşcu and Thorup [PT06] used direct sum type results to prove stronger lower bounds for approximate near-neighbour (ANN) search in the cell probe model. Work on the direct sum property has also inspired earlier lower bounds for ANN due to Chakrabarti and Regev [CR04].

Although they seem highly plausible, it is well-known that strong direct product results fail to hold for several modes of communication and computation. For example, testing the equality of $k = \log n$ pairs of n-bit strings with a constant-error private-coin communication protocol has complexity $\mathrm{O}(k \log k + \log n) = \mathrm{O}(\log n \log\log n)$ (see, e.g., [KN97, Example 4.3, page 43]), where we might expect a complexity of $\Omega(k \log n) = \Omega(\log^2 n)$. Similarly, Shaltiel [Sha03] gives an example for which a strong direct product result fails to hold for average case (i.e., distributional) communication complexity.

Notwithstanding the above mentioned counterexamples, various forms of direct product result have been discovered in special cases. Early attempts at the question can be found in [IRW94], and the references therein. Parnafes, Raz, and Wigderson [PRW97] prove a direct product result for "collections" of protocols. In their result the bound on the success probability, however, is only shown to behave like $2^{-\Omega(k/c)}$ for the communication complexity c of the problem at hand. Shaltiel [Sha03] proves a strong direct product property in cases where the discrepancy method is used under the uniform distribution; Klauck, Špalek, and de Wolf [KŠdW04] prove it for the quantum communication complexity of Set Disjointness; Beame, Pitassi, Segerlind, and Wigderson [BPSW07] prove it in cases where the *rectangle* or *corruption bound* is tight under product distributions; and Gavinsky [Gav06] proves it for the one-way complexity of a certain class of relational problems. The result by Beame *et al.* for instance allows the conclusion that solving k instances of Set Disjointness with communication complexity $\mathrm{o}(k\sqrt{n})$ has success probability at most $2^{-\Omega(k)}$. De Wolf [dW05] proves a strong direct product theorem for the one-way public-coin randomized communication complexity of the Index function. In more recent work (albeit subsequent to ours), Ben-Aroya, Regev, and de Wolf [BARdW07] derive a similar direct product theorem for the one-way *quantum* communication complexity of Index. Since Index captures the notion of VC-dimension, similar results follow for the one-way distributional (classical and quantum) communication complexity of any Boolean function under the worst case *product* distribution.

Whether the strong direct product theorem holds in general for public-coin randomized protocols remains a frustrating open question in communication complexity theory. Research on weaker types of property, namely the *direct sum* or the *weak direct product* property, has met with better success. A direct sum theorem states that solving k instances with constant probability of success incurs at least k times the cost of solving 1 instance. (A strong direct product theorem would show that even with probability of success that is exponentially small in k, the cost would be k times the cost of solving one instance.)

For deterministic protocols it is known that k times the square root of the deterministic complexity of a function f is needed to compute k instances of f (see, e.g., [KN97, Exercise 4.11, page 46]). It is also straightforward to show that the deterministic *one-way* communication complexity of every function f has the direct sum property. For randomized protocols, Chakrabarti, Shi, Wirth, and Yao [CSWY01] give a lower bound for the direct sum problem in the simultaneous message passing (SMP) model in terms of “information cost”. This has also been extended to two-way classical and quantum communication [BYJKS04, JRS03b].

Jain, Radhakrishnan, and Sen [JRS05b] show a tight direct sum theorem for the one-way and SMP models for both quantum and randomized classical communication, along with a weak direct sum result for two-way communication. In other work, Jain, Radhakrishnan, and Sen [JRS03a] give a direct sum type lower bound for bounded round private-coin protocols in terms of the average case communication complexity under product distributions. Harsha, Jain, McAllester, and Radhakrishnan [HJMR07] have strengthened the latter lower bound by reducing to a large extent its dependence on the number of rounds. (As mentioned above, these have influenced the work of Chakrabarti and Regev [CR04] on the approximate nearest neighbour problem in the cell probe model. Pătraşcu and Thorup [PT06] use direct sum type results to prove bigger lower bounds for this problem.)

In a weak direct product theorem, one shows that the success probability of solving k instances of a problem with the resources needed to solve one instance (with probability 2/3) goes down exponentially with k. Klauck [Kla04] shows such a result for the rectangle/corruption bound under arbitrary distributions, leading to the conclusion that solving k instances of Set Disjointness with communication complexity $\mathrm{o}(n)$ is possible only with success probability $2^{-\Omega(k)}$.

In this article, we develop a new information-theoretic framework for proving results of a direct-product flavour. We begin by introducing *subdistribution bounds*, measures of hardness of computing a function (more generally, a relation) through various kinds of two-party communication protocol. Subdistribution bounds are based on the notion of *relative co-min-entropy* of two distributions (see Section 2.2 for a formal definition) which in turn is closely connected to *relative entropy*, a well-studied notion in information theory. This allows us to draw on a rich, burgeoning body of techniques from information theory for its analysis. Subdistribution bounds are in fact a relaxation of rectangle/corruption bounds. We show that these quantities are nonetheless within a constant factor of each other for Boolean functions. (Lemma 3.2 contains the precise statement for relations.) We are therefore also able to draw on existing work on the rectangle bound. In particular, we can infer estimates for subdistribution bounds for explicit relations from estimates on rectangle bounds.

In the setting of public-coin randomized *one-way* communication, we show that the one-way variant of subdistribution complexity equals, essentially up to a constant factor, the communication complexity of any relation. The public-coin randomized two-way communication complexity of any relation is bounded from below by its (two-way) subdistribution complexity. More importantly, we show that both the one- and two-way subdistribution bounds satisfy the strong direct product property under any *product* distribution. This way we get strong direct product lower bounds for both kinds of randomized protocol. In particular, we establish strong direct product theorems for problems whose one- or two-way complexity is achieved by the rectangle/corruption bound under product distributions. Finally, we prove that the two-way subdistribution bound satisfies the weak direct product property under *arbitrary* distributions. As a consequence, we get weak direct product theorems for problems whose two-way complexity is achieved by the rectangle/corruption bound.

The proofs we present for the direct product properties of the subdistribution bound belong to a line of work based on the powerful *substate theorem* due to Jain, Radhakrishnan, and Sen [JRS02]. The substate theorem establishes an approximate equivalence between relative co-min-entropy and relative entropy. This equivalence, and the super-additivity of relative entropy enable us to give a simple information-theoretic explanation of why the direct product properties hold.

We point out that the our approach provides a unified view of several recent works on the topic, simultaneously generalizing and strengthening them. These works include the strong direct product property for the rectangle/corruption bound for Boolean functions due to Beame *et al.* [BPSW07] (and its consequence for the two-way classical communication complexity of Set Disjointness, which was also independently shown by Klauck *et al.* [KŠdW04, Theorem 20]), a direct product property for the one-way classical communication complexity of certain relations due to Gavinsky [Gav06], the direct product theorem for the one-way classical communication complexity of Index due to de Wolf [dW05] (which also follows from the subsequent work of Ben-Aroya *et al.* [BARdW07]), and the weak direct product property for the rectangle/corruption bound for Boolean functions due to Klauck [Kla04].

The subdistribution approach is strictly more powerful than those followed in all of the above mentioned works, except possibly that of Beame *et al.* For example, the direct product theorem we derive for the one-way randomized communication complexity of the Hidden Matching relation is optimal, and better by a logarithmic factor both in the communication and in the error-exponent than the one due to [Gav06]. The methods in [dW05] and [BARdW07] are specialized for Boolean functions, and do not apply to relations. Specifically, it is unclear how these methods could yield lower bounds for Hidden Matching. Of course, the results of [BARdW07] also pertain to quantum communication; these do not follow from our work. (We leave the extension of the subdistribution framework to quantum communication to future work.) Because of the intimate connection between the subdistribution and the rectangle/corruption bounds, it is plausible that the method followed by Beame *et al.* be extensible, with additional arguments, to the case of one-way communication and general relations.

The subdistribution technique makes an important link between rectangle/corruption methods and methods from information theory, two of the most successful approaches to proving lower bounds for communication complexity. We believe that the results described above signify its potential to solve yet more challenging questions regarding the direct product problem, and communication complexity in general.

Organization of the article

We follow standard terminology and notation in communication complexity, as in the text [KN97]. For completeness, this is summarized in Section 2.1. Section 2.2 contains the notation, and the definition and properties of information theoretic concepts that we use. We introduce subdistribution bounds in Section 3 and relate them to rectangle/corruption bounds. We characterize one-way communication complexity in terms of the one-way subdistribution bound in Section 4. In Section 5.1 we present strong direct product results in the setting of two-way communication. These are extended to one-way communication in Section 5.2. The weak direct product theorem for the subdistribution bound is derived in Section 6. We describe the immediate consequences of our direct product theorems, in particular how they imply results in prior works on the topic, in the sections in which the theorems occur. The remaining consequences are described in Section 7. Several proofs are deferred to Appendix B.

2 Preliminaries

2.1 Communication complexity

In this section we briefly review the model of communication complexity. For a comprehensive introduction to the subject we refer the reader to the text by Kushilevitz and Nisan [KN97].

We consider the two-party model of communication. Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets, and let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. In a two-party communication protocol the parties, say Alice and Bob, get inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively. They alternately send messages to each other with the goal of determining an element $z \in \mathcal{Z}$ such that $(x, y, z) \in f$. We assume that for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ given as input, there is at least one $z \in \mathcal{Z}$ such that $(x, y, z) \in f$.

One-way communication

We first consider the *one-way* model of communication, in which there is a single message, from Alice to Bob at the end of which Bob determines the answer z from the single message from Alice, and his input y. (In the one-way protocols we consider, the single message is always from Alice to Bob.) Let $0 \leq \epsilon < 1/3$, and let μ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. We let $\mathsf{D}^{1,\mu}_{\epsilon}(f)$ represent the *distributional one-way communication complexity* of f under μ with expected error ϵ, i.e., the communication of the best private-coin one-way protocol for f, with *distributional error* (average error over the coins and the inputs) at most ϵ under μ. We note that $\mathsf{D}^{1,\mu}_{\epsilon}(f)$ is achieved by a deterministic one-way protocol, and will henceforth restrict ourselves to deterministic protocols in the context of distributional communication complexity. We let $\mathsf{R}^{1,\mathsf{pub}}_{\epsilon}(f)$ represent the public-coin *randomized one-way communication complexity* of f with worst case error ϵ, i.e., the communication of the best public-coin randomized one-way protocol for f with error for each input (x, y) being at most ϵ. The analogous quantity for private coin randomized protocols is denoted by $\mathsf{R}^{1}_{\epsilon}(f)$. The following is a consequence of the *min-max* theorem in game theory [KN97, Theorem 3.20, page 36].

Lemma 2.1 (Yao principle) $\mathsf{R}^{1,\mathsf{pub}}_{\epsilon}(f) = \max_{\mu} \mathsf{D}^{1,\mu}_{\epsilon}(f)$.

The communication complexity of a relation may reduce significantly when μ is restricted to product distributions over $\mathcal{X} \times \mathcal{Y}$. We define $\mathsf{R}^{1,\square}_{\epsilon}(f) \overset{\Delta}{=} \max_{\mu \text{ product}} \mathsf{D}^{1,\mu}_{\epsilon}(f)$.

The VC-dimension of a Boolean function f is an important combinatorial concept and has close connections with the one-way communication complexity of f.

Definition 2.1 (Vapnik-Chervonenkis (VC) dimension) *A set S is said to be shattered by a set $\mathcal{G}$ of Boolean functions from S to $\{0,1\}$, if $\forall R \subseteq S, \exists g_R \in \mathcal{G}$ such that $\forall s \in S, (s \in R) \iff (g_R(s) = 1)$.*

Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function. For all $x \in \mathcal{X}$ let $f_x : \mathcal{Y} \to \{0,1\}$ be defined as $f_x(y) \overset{\Delta}{=} f(x,y), \forall y \in \mathcal{Y}$. Let $\mathcal{F} \overset{\Delta}{=} \{f_x : x \in \mathcal{X}\}$. Then the Vapnik-Chervonenkis dimension of f is defined as $\mathsf{VC}(f) \overset{\Delta}{=} \max_{S \subseteq \mathcal{Y}} \{|S| : S$ is shattered by $\mathcal{F}\}$.

Kremer, Nisan, and Ron [KNR99, Theorem 3.2] relate VC-dimension to communication complexity. The tighter lower bound (stated below) on the communication complexity in terms of the error ϵ in the communication protocol appears in Ambainis, Nayak, Ta-Shma, and Vazirani [ANTSV99, Theorem 1.1].

Theorem 2.2 ([KNR99, ANTSV99]) *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function, and let $\epsilon \in (0, 1/2)$. Then there is a universal constant κ_0 such that*

$$(1 - \mathsf{H}_2(\epsilon)) \cdot \mathsf{VC}(f) \quad \leq \quad \mathsf{R}^{1,\square}_{\epsilon}(f) \quad \leq \quad \kappa_0 \cdot \frac{1}{\epsilon} \log \frac{1}{\epsilon} \cdot \mathsf{VC}(f),$$

where $\mathsf{H}_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$ is the binary entropy function defined on $[0,1]$.

Two-way communication

Next we consider two-way protocols, which are defined analogously. These allow communication between Alice and Bob over multiple rounds at the end of which both parties output the *same* element $z \in \mathcal{Z}$ *that depends upon the transcript of the protocol alone*. Following Kushilevitz and Nisan [KN97], we assume Alice and Bob disregard their inputs when they determine their output. This is unlike in one-way protocols, where we (necessarily have to) allow Bob to determine his output from Alice's message *and his input*. The relevant complexity measures for this model are denoted $\mathsf{D}^{\mu}_{\epsilon}(f)$, $\mathsf{R}^{\square}_{\epsilon}(f)$, $\mathsf{R}^{\mathsf{pub}}_{\epsilon}(f)$, $\mathsf{R}_{\epsilon}(f)$ etc. (without the superscript '1'). Lemma 2.1 holds for two-way protocols *mutatis mutandis*.

An arbitrary two-way communication protocol (in which the two parties consider their inputs for the computation of their respective outputs) may be converted into the form above. One party may send an additional message consisting of his/her output. The consequent increase in communication complexity is at most $\log |\mathcal{Z}|$.

2.2 Information theory

In this section we present some information theoretic notation, definitions and facts that we use in our proofs. For an introduction to information theory, we refer the reader to the text by Cover and Thomas [CT91]. Most of the facts stated in this section without proof may be found in this book.

All logarithms in the article are taken with base 2. For an integer $t \geq 1$, $[t]$ represents the set $\{1, \ldots, t\}$. For square matrices P, Q, by $Q \geq P$ we mean that $Q - P$ is positive semi-definite. For a matrix A, $\|A\|_1$ denotes its ℓ_1 norm.

Specializing from the quantum case, we view a discrete probability distribution P as a positive semi-definite trace one diagonal matrix indexed by its (finite) sample space. For a distribution P with support on set $\mathcal{X}$, and $x \in \mathcal{X}$, $P(x)$ denotes the (x, x) diagonal entry of P, and $P(\mathcal{E}) = \sum_{x \in \mathcal{E}} P(x)$ denotes the probability of the event $\mathcal{E} \subseteq \mathcal{X}$. For a random variable X, we sometimes also let X represent its distribution.

Let $\mathcal{X}, \mathcal{Y}$ be sets and let P be a distribution with support on $\mathcal{X} \times \mathcal{Y}$. For $x \in \mathcal{X}$, we define $P_{\mathcal{X}}(x) = \sum_{y \in \mathcal{Y}} P(x, y)$, the probability of x in the marginal distribution on $\mathcal{X}$; $P_{\mathcal{Y}}(y)$ is similarly defined for $y \in \mathcal{Y}$. Further, if $y \in \mathcal{Y}$ occurs with probability $P_{\mathcal{Y}}(y) > 0$, we define $P_{\mathcal{X}}(x|y) = \frac{P(x,y)}{P_{\mathcal{Y}}(y)}$, the conditional probability given the event $\mathcal{X} \times \{y\}$. The distribution P is said to be a *product* distribution if there are distributions Q, R on $\mathcal{X}, \mathcal{Y}$ respectively such that $P = Q \otimes R$, where $\otimes$ denotes the tensor product operation. Equivalently, for a product distribution, $P(x, y) = P_{\mathcal{X}}(x) \cdot P_{\mathcal{Y}}(y)$.

The *relative entropy* or the *Kullback-Leibler divergence* of the distribution P with respect to the distribution Q is defined as $\mathsf{S}(P \,\|\, Q) \stackrel{\Delta}{=} \mathrm{Tr}(P \log P - P \log Q)$. Relative entropy is jointly convex in its arguments.

Lemma 2.3 *Let P_1, P_2, Q_1, Q_2 be probability distributions. Then for $r \in [0, 1]$,*

$$\mathsf{S}(rP_1 + (1-r)P_2 \,\|\, rQ_1 + (1-r)Q_2) \quad \leq \quad r\mathsf{S}(P_1 \,\|\, Q_1) + (1-r)\mathsf{S}(P_2 \,\|\, Q_2).$$

Relative entropy satisfies the following *chain rule*:

Lemma 2.4 (Chain rule for relative entropy) *Let $M_1, \ldots, M_k$ and $N_1, \ldots, N_k$ be jointly distributed random variables. For $1 \leq i \leq k$, let $\tilde{M}_i$ represent the random variable $M_1 \ldots M_{i-1}$. Similarly define $\tilde{N}_i$. Then*

$$\mathsf{S}(M_1 \ldots M_k \,\|\, N_1 \ldots N_k) \quad = \quad \sum_{i=1}^{k} \mathbb{E}_{m \sim \tilde{M}_i}[\mathsf{S}(M_i|\tilde{M}_i = m \,\|\, N_i|\tilde{N}_i = m)].$$

Lemma 2.5 *Let $M_1 M_2$ be random variables and let $N_1 N_2$ be mutually independent random variables. Then*

$$\mathsf{S}(M_1 M_2 \,\|\, N_1 N_2) \quad \geq \quad \mathsf{S}(M_1 \,\|\, N_1) + \mathsf{S}(M_2 \,\|\, N_2).$$

Proof: Using the chain rule (Lemma 2.4), the independence of N_1 and N_2, and finally convexity (Lemma 2.3), we have

$$\begin{aligned} \mathsf{S}(M_1 M_2 \,\|\, N_1 N_2) &= \mathsf{S}(M_1 \,\|\, N_1) + \mathbb{E}_{m \sim M_1}[\mathsf{S}(M_2|M_1 = m \,\|\, N_2|N_1 = m)] \\ &= \mathsf{S}(M_1 \,\|\, N_1) + \mathbb{E}_{m \sim M_1}[\mathsf{S}(M_2|M_1 = m \,\|\, N_2)] \\ &\geq \mathsf{S}(M_1 \,\|\, N_1) + \mathsf{S}(M_2 \,\|\, N_2), \end{aligned}$$

as claimed. ■

For distributions P, Q, with support on set $\mathcal{X}$, we define

$$\mathsf{S}_\infty(P \,\|\, Q) \quad \stackrel{\Delta}{=} \quad \inf\{c : Q \geq P/2^c\},$$

as the *relative co-min-entropy* of P with respect to Q. This quantity measures the least scaling (i.e., shrinking) of the distribution P with which it "sits inside" the distribution Q. For a finite sample space $\mathcal{X}$, we may equivalently define relative co-min-entropy as

$$\mathsf{S}_\infty(P \,\|\, Q) \quad \stackrel{\Delta}{=} \quad \max_{x \in \mathcal{X}} \ \log_2 \frac{P(x)}{Q(x)}.$$

Note that the relative co-min-entropy of P with respect to the uniform distribution on $\mathcal{X}$ is precisely $\log|\mathcal{X}| - \mathsf{H}_\infty(P)$, where $\mathsf{H}_\infty(P) = \min_x \log \frac{1}{P(x)}$ is the *min-entropy* of P.

The notion of relative co-min-entropy is closely connected to the notion of relative entropy. The monotonicity of the logarithm function implies that:

Lemma 2.6 *Let P, Q be distributions. Then $\mathsf{S}(P \,\|\, Q) \ \leq \ \mathsf{S}_\infty(P \,\|\, Q)$.*

This fact is a special case of Theorem 1(7) in [JRS05a]. The converse of this statement also holds in an approximate sense, and gives us a powerful operational characterization of relative entropy. This fact has come to be known as the *substate theorem* [JRS02, Proposition 1].

Lemma 2.7 (Substate theorem) *Let P, Q be probability distributions over the same finite sample space such that $\mathsf{S}(P \,\|\, Q) \ \leq \ c$. Then for all $r > 1$, there exist distributions P_r such that $\|P - P_r\|_1 \leq \frac{2}{r}$ and*

$$(1 - \tfrac{1}{r})\tfrac{P_r}{2^{r(c+1)}} \quad \leq \quad Q. \tag{1}$$

The condition in Eq. (1) is equivalent to

$$\mathsf{S}_\infty(P_r \,\|\, Q) \quad \leq \quad r(c+1) + \log \tfrac{r}{r-1}.$$

The following fact is readily verified:

Lemma 2.8 *If P, Q are distributions on the same sample space such that $\|P - Q\|_1 \leq \epsilon$, then for any event $\mathcal{E}$, we have $|P(\mathcal{E}) - Q(\mathcal{E})| \leq \epsilon/2$.*

The following fact may be verified from the definition of relative co-min-entropy.

Lemma 2.9 *Let X_1, X_2, Y_1, Y_2 be random variables. Then $\mathsf{S}_\infty(X_1 \,\|\, Y_1) \leq \mathsf{S}_\infty(X_1 X_2 \,\|\, Y_1 Y_2)$.*

Random variables X, Y, Z form a *Markov chain*, represented as $X \to Y \to Z$, iff for all x, y, the conditional random variable $Z|(XY = xy)$ is equal to $Z|(Y = y)$. The following lemma may be verified readily from this definition.

Lemma 2.10 *If $X \to Y \to Z$ is a Markov chain, then so is $Z \to Y \to X$.*

We use various forms of the *Markov inequality* from probability theory [CT91] in our arguments without proof.

3 Subdistribution bounds

Here we introduce,*subdistribution bounds*, new measures of communication complexity. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 \leq \epsilon \leq 1/3$.

Definition 3.1 *Let λ, μ be distributions on $\mathcal{X} \times \mathcal{Y}$.*

1. **ϵ-monochromatic:** *We say that the distribution λ is (ϵ, z)-*monochromatic *for f if the probability* $\Pr_{XY \sim \lambda}[(X, Y, z) \in f] \geq 1 - \epsilon$*. We say that λ is ϵ-*monochromatic *for f if it is (ϵ, z)-monochromatic for f for some $z \in \mathcal{Z}$.*
2. **one-way ϵ-monochromatic:** *We call λ* one-way ϵ-monochromatic *for f if there is a function $g : \mathcal{Y} \rightarrow \mathcal{Z}$ such that* $\Pr_{XY \sim \lambda}[(X, Y, g(Y)) \in f] \geq 1 - \epsilon$*.*
3. **one-message-like:** *We call λ* one-message-like *for μ with respect to $\mathcal{X}$ if for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, whenever $\lambda_{\mathcal{X}}(x) > 0$, we have $\mu_{\mathcal{X}}(x) > 0$ and $\lambda_{\mathcal{Y}}(y|x) = \mu_{\mathcal{Y}}(y|x)$. We similarly define the notion of "one-message-like with respect to $\mathcal{Y}$". In the context of one-way protocols, we use the term "one-message-like" to mean "one-message-like with respect to $\mathcal{X}$".*
4. **SM-like:** *We call λ* SM-like (simultaneous-message-like) *for μ, if there is a distribution θ on $\mathcal{X} \times \mathcal{Y}$ such that θ is one-message-like for μ with respect to $\mathcal{X}$ and λ is one-message-like for θ with respect to $\mathcal{Y}$.*

These definitions are motivated by properties of distributions that arise in one-way, two-way, and simultaneous-message communication protocols. For instance, the distribution λ is one-way ϵ-monochromatic precisely when there is a one-way communication protocol for f with *zero* communication cost and distributional error at most ϵ under λ. Furthermore, suppose $\mathcal{P}$ is a deterministic one-way protocol for f, with a single message from Alice to Bob. Let X, Y denote random variables with joint distribution μ, corresponding to Alice and Bob's inputs respectively. For any message string m in $\mathcal{P}$, we may readily verify that the conditional distribution $XY|(M = m)$ is one-message-like for μ (with respect to $\mathcal{X}$). Finally, suppose the distributional error made by $\mathcal{P}$ is at most ϵ. Then, for any $\delta \in (0, 1]$, the distribution of $XY|(M = m)$ is one-way $\frac{\epsilon}{\delta}$-monochromatic for f with probability at least $1 - \delta$ over the messages m.

We begin by defining subdistribution bounds on one-way communication complexity.

Definition 3.2 (One-way subdistribution bounds) *For a distribution μ over $\mathcal{X} \times \mathcal{Y}$, define $\mathsf{sub}^1_{\mathcal{Y}}(f, \epsilon, \mu) \stackrel{\Delta}{=} \min_\lambda \mathsf{S}_\infty(\lambda \,\|\, \mu)$, where λ ranges over all distributions which are both one-message-like for μ (with respect to $\mathcal{X}$) and one-way ϵ-monochromatic for f. We define the* one-way subdistribution bound *as $\mathsf{sub}^1_{\mathcal{Y}}(f, \epsilon) \stackrel{\Delta}{=} \max_\mu \mathsf{sub}^1_{\mathcal{Y}}(f, \epsilon, \mu)$, where μ ranges over all distributions on $\mathcal{X} \times \mathcal{Y}$. When the maximization is restricted to product distributions μ, we refer to the quantity as the* one-way product subdistribution bound $\mathsf{sub}^{1,[]}_{\mathcal{Y}}(f, \epsilon)$*.*

Remark: First, in the above definition, the subscript $\mathcal{Y}$ is used to emphasize the fact that in the definition of one-way ϵ-monochromatic, we allow for different values of output depending upon Bob's input $y \in \mathcal{Y}$ in a zero-communication protocol for f under distribution λ. Second, note that a distribution λ which is one-message-like for a product distribution μ is itself a product distribution. Third, when we take the distribution λ to range over μ conditioned upon rectangles of the form $S \times \mathcal{Y}$, where $S \subseteq \mathcal{X}$, we get a one-way variant of the rectangle or corruption bound in communication complexity as introduced by Yao, and applied by Razborov and others. We elaborate on the precise connection between the rectangle and subdistribution bounds below.

We define two-way subdistribution bounds in a manner analogous to the one-way bounds. We also consider variants of such bounds corresponding to fixed outputs, since these variants better capture communication complexity in some important cases.

Definition 3.3 (Two-way subdistribution bounds) *For a distribution μ over $\mathcal{X} \times \mathcal{Y}$, define $\mathsf{sub}(f, \epsilon, \mu) \overset{\Delta}{=} \min_{\lambda} \mathsf{S}_{\infty}(\lambda \,\|\, \mu)$, where λ ranges over all distributions which are both SM-like for μ and ϵ-monochromatic for f. When we restrict the minimization to distributions that are SM-like and (ϵ, z)-monochromatic, for some fixed value $z \in \mathcal{Z}$, we denote the resulting quantity $\mathsf{sub}(f, \epsilon, z, \mu)$. We define the* two-way subdistribution bound *as $\mathsf{sub}(f, \epsilon) \overset{\Delta}{=} \max_{\mu} \mathsf{sub}(f, \epsilon, \mu)$, where μ ranges over all distributions on $\mathcal{X} \times \mathcal{Y}$. Similarly, $\mathsf{sub}(f, \epsilon, z) \overset{\Delta}{=} \max_{\mu} \mathsf{sub}(f, \epsilon, z, \mu)$. When the maximization is restricted to product distributions μ, we refer to the quantities as* two-way product subdistribution bounds $\mathsf{sub}^{\square}(f, \epsilon)$ *and* $\mathsf{sub}^{\square}(f, \epsilon, z)$.

Remark: First, suppose m is a possible message transcript in a deterministic two-way protocol $\mathcal{P}$ for f when run on inputs X, Y distributed according to μ. Let M denote the random variable corresponding to the transcript of $\mathcal{P}$. We may readily verify that the conditional distribution $XY|(M = m)$ is SM-like for μ. Second, the distributions that are SM-like for a *product* distribution μ are precisely the set of all product distributions over $\mathcal{X} \times \mathcal{Y}$. Third, it is important to restrict λ appropriately in the definition of $\mathsf{sub}(f, \epsilon, \mu)$ to get non-trivial bounds. For instance, if we do not restrict λ to product distributions in the definition of $\mathsf{sub}^{\square}(f, \epsilon, \mu)$ (with μ being a product distribution), the quantity is at most 1 for Boolean functions: consider the possibly non-product distribution λ that results from conditioning upon $f^{-1}(1)$ or $f^{-1}(0)$, whichever has higher probability under μ. The distribution λ is 0-monochromatic for f, and "sits well inside μ" (i.e., $\mathsf{S}_{\infty}(\lambda \| \mu) \leq 1$), since the event on which we condition has probability $\geq \frac{1}{2}$.

To state the precise connection between the subdistribution bound and the rectangle/corruption bound from communication complexity, we define the latter bound precisely. A rectangle in $\mathcal{X} \times \mathcal{Y}$ is a subset of the form $S \times T$, where $S \subseteq \mathcal{X}, T \subseteq \mathcal{Y}$. For a distribution μ, and an event $R \subseteq \mathcal{X} \times \mathcal{Y}$, let μ_R denote the conditional distribution of μ given the event R.

Definition 3.4 (Rectangle/corruption bounds) *For a possibly non-product distribution μ, define $\mathsf{rec}(f, \epsilon, \mu) \overset{\Delta}{=} \min_R \mathsf{S}_{\infty}(\mu_R \,\|\, \mu)$, where R ranges over all rectangles in $\mathcal{X} \times \mathcal{Y}$ such that μ_R is ϵ-monochromatic for f. Define $\mathsf{rec}(f, \epsilon) \overset{\Delta}{=} \max_{\mu} \mathsf{rec}(f, \epsilon, \mu)$. When the maximization is restricted to product distributions μ, we get the two-way* product *rectangle bound $\mathsf{rec}^{\square}(f, \epsilon)$. The quantities $\mathsf{rec}(f, \epsilon, z)$ and $\mathsf{rec}^{\square}(f, \epsilon, z)$ for a fixed value $z \in \mathcal{Z}$ are defined in a manner analogous to the two-way subdistribution bounds.*

As is evident from the above definitions, subdistribution bounds are a relaxation of the corresponding rectangle bounds, and are therefore always dominated by the latter. Nevertheless, we show that they are approximately equal to each other.

We begin with this connection for the *product* two-way bound.

Lemma 3.1 *Let μ be a product distribution on $\mathcal{X} \times \mathcal{Y}$ and let $\delta \in (0, 1)$. Then*

$$\mathsf{rec}(f, \epsilon, \mu) \quad \geq \quad \mathsf{sub}^{\square}(f, \epsilon, \mu) \quad \geq \quad \mathsf{rec}\Big(f, \frac{\epsilon}{\delta^2}, \mu\Big) - \log \tfrac{1}{(1-\delta)^2} \ .$$

The same inequalities hold between $\mathsf{rec}(f, \epsilon, z, \mu)$ and $\mathsf{sub}^{\square}(f, \epsilon, z, \mu)$ mutatis mutandis *for any $z \in \mathcal{Z}$.*

Proof: Let $\mu = \mu_A \otimes \mu_B$ be a product distribution on $\mathcal{X} \times \mathcal{Y}$. By definition, $\mathsf{rec}(f, \epsilon, \mu) \geq \mathsf{sub}^{\square}(f, \epsilon, \mu)$. For the second inequality we argue as follows. Consider any ϵ-monochromatic product distribution $\lambda = \lambda_A \otimes \lambda_B$ along with an output $z \in \mathcal{Z}$ which makes it ϵ-monochromatic. View λ as a convex combination of distributions λ_x on $\{x\} \times \mathcal{Y}$. Note that the marginal distribution of λ_x on $\mathcal{Y}$ is λ_B for all x. Using the Markov Inequality, we get a subset $S \subseteq \mathcal{X}$ such that $\lambda_A(S) \geq 1 - \delta$ and for each $x \in S$, the distribution λ_x is (ϵ/δ)-monochromatic for the same output z. Therefore, the distribution $\pi = \mu_{A,S} \otimes \lambda_B$, where $\mu_{A,S}$ is the distribution on $\mathcal{X}$ conditioned upon event S, is also (ϵ/δ)-monochromatic for f with output z. Similarly we identify a subset $T \subseteq \mathcal{Y}$ such that $\lambda_B(T) \geq 1 - \delta$, and each distribution π_y on $\mathcal{X} \times \{y\}$ with marginal $\mu_{A,S}$ on $\mathcal{X}$ is (ϵ/δ^2)-monochromatic for every $y \in T$.

Thus, we get a rectangle $R = S \times T$ with probability $\lambda(R) \geq (1-\delta)^2$ such that the distribution μ_R, the distribution μ conditioned on R, is (ϵ/δ^2)-monochromatic. Moreover, if $\mathsf{S}_\infty(\lambda \| \mu) = c$, then $\mu(R) \geq \lambda(R) \cdot 2^{-c} \geq (1-\delta)^2 \cdot 2^{-c}$. In other words, $\mathsf{rec}(f, \epsilon/\delta^2, \mu) \leq c + \log \frac{1}{(1-\delta)^2}$. Minimizing over all such λ, we see that $\mathsf{rec}(f, \epsilon/\delta^2, \mu) \leq \mathsf{sub}^{\square}(f, \epsilon, \mu) + \log \frac{1}{(1-\delta)^2}$. ■

The case of non-product distributions is more technical, and is deferred to Appendix A. We simply state the connection here.

Lemma 3.2 *Let μ be any distribution on $\mathcal{X} \times \mathcal{Y}$. There exist universal positive constants $\kappa_1, \kappa_2, \kappa_3$ such that*

$$\begin{aligned} \mathsf{rec}(f, \epsilon, \mu) &\geq \mathsf{sub}(f, \epsilon, \mu) \\ &\geq \kappa_1 \cdot \min\left\{ \mathsf{rec}(f, \kappa_2\epsilon, \mu), \quad \mathsf{rec}(f, 0, \mu) - \log_2 |\mathcal{Z}| - \log_2 \frac{1}{\epsilon} \right\} - \kappa_3, \end{aligned}$$

Similar inequalities hold between $\mathsf{rec}(f, \epsilon, z, \mu)$ and $\mathsf{sub}(f, \epsilon, z, \mu)$ for any $z \in \mathcal{Z}$.

Thus, in spite of being a relaxation of rectangle bounds, subdistribution bounds give us lower bounds that are as strong as the ones obtained from the former. Accurate estimates for the subdistribution bounds for explicit functions may be obtained from the corresponding rectangle bounds. Furthermore, subdistribution bounds have the distinct advantage of being readily amenable to analysis with tools from information theory, as is illustrated by the proof of Theorem 5.1 (which states that they satisfy the direct product property).

The rectangle bound $\mathsf{rec}(f, \epsilon)$ is well-known to be a lower bound for two-way randomized communication complexity. We refer the reader to [BPSW07, Section 3] for a precise formulation of this bound, and state a consequence of this connection for Boolean functions.

Theorem 3.3 *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, $\epsilon \in (0,1]$, μ be any probability distribution on $\mathcal{X} \times \mathcal{Y}$. For $z \in \{0,1\}$, let $p_z = \mu(f^{-1}(z))$, and $\delta \in [0, \epsilon p_z)$. Then*

$$\mathsf{R}^{\mathsf{pub}}_{\delta}(f) \quad \geq \quad \max_z \left[\mathsf{rec}(f, \epsilon, z) - \log_2 \frac{1}{p_z - \delta/\epsilon} \right].$$

The product rectangle bound $\mathsf{rec}^{\square}(f, \epsilon)$ may be arbitrarily smaller than the (unrestricted) rectangle bound [She07], but is still known to give strong bounds for important functions. For example, when f is the $\mathsf{Set\ Disjointness}$ problem on n-bit inputs, $\mathsf{R}^{\mathsf{pub}}_{1/3}(f) = \Theta(n) = \mathsf{rec}(f, 1/3, 0)$ [KN97, Section 4.6, Lemma 4.49]. On the other hand, $\mathsf{rec}^{\square}(f, 1/3, 0)$ and $\mathsf{R}^{\square}_{1/3}(f)$ are both $\mathrm{O}(\sqrt{n} \log n)$ and at least $\Omega(\sqrt{n})$ [BFS86].

4 A characterization of one-way communication complexity

In this section we present a new characterization of randomized one-way communication complexity in terms of the one-way subdistribution bound. Throughout this section, we use the term "one-message-like" to mean "one-message-like with respect to $\mathcal{X}$".

We begin by showing that the one-way communication complexity of a relation is always larger than the subdistribution bound.

Lemma 4.1 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 \leq \epsilon \leq 1/3$ and $k > 0$ be non-negative real numbers. Then*

$$\mathsf{R}^{1,\mathsf{pub}}_{\epsilon(1-2^{-k})}(f) \quad \geq \quad \mathsf{sub}^{1}_{\mathcal{Y}}(f,\epsilon) - k.$$

Proof: For any distribution μ on $\mathcal{X} \times \mathcal{Y}$, we show

$$\mathsf{D}^{1,\mu}_{\epsilon(1-2^{-k})}(f) \quad \geq \quad \mathsf{sub}^{1}_{\mathcal{Y}}(f,\epsilon,\mu) - k. \tag{2}$$

Maximizing over μ, noting that distributional complexity is bounded by worst-case complexity (see the Yao min-max principle, Lemma 2.1), and the definition of $\mathsf{sub}^{1}_{\mathcal{Y}}(f,\epsilon)$ we get our bound.

Let $c \stackrel{\Delta}{=} \mathsf{sub}^{1}_{\mathcal{Y}}(f,\epsilon,\mu)$. If $\lfloor c-k \rfloor \leq 0$, Eq. (2) holds vacuously. Otherwise, let $\mathcal{P}$ be a deterministic one-way protocol with communication at most $\lfloor c-k \rfloor$. Let the random variables X, Y with joint distribution μ represent the inputs of Alice and Bob respectively. Let M represent the correlated random variable corresponding to Alice's message. For a message string m with $p_m \stackrel{\Delta}{=} \Pr[M=m] > 0$ let ϵ_m denote the probability of error of $\mathcal{P}$ conditional on $M=m$. Let $\mathcal{M}$ be the set of messages m such that $p_m > 2^{-c}$. Since there are at most 2^{c-k} messages, we get that $\sum_{m \notin \mathcal{M}} p_m \leq 2^{-k}$. Let λ_m be the distribution of $XY|(M=m)$. Note that $\mathsf{S}_\infty(\lambda_m \,\|\, \mu) = -\log_2 p_m < c$ for for $m \in \mathcal{M}$. Since λ_m is one-message-like for μ, from the definition of $\mathsf{sub}^{1}_{\mathcal{Y}}(f,\epsilon,\mu)$ we have $\epsilon_m > \epsilon$. Hence the overall error of the protocol $\mathcal{P}$ is $> \epsilon(1-2^{-k})$. Therefore, by its definition $\mathsf{D}^{1,\mu}_{\epsilon(1-2^{-k})}(f) > \lfloor c-k \rfloor$, which is the communication in $\mathcal{P}$. ■

For the other direction, we first show that for a relation f with low subdistribution complexity, any distribution μ may be decomposed into a small number of one-message-like distributions that are one-way ϵ-monochromatic for f.

Lemma 4.2 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $0 \leq \epsilon < 1$, and $c \stackrel{\Delta}{=} \mathsf{sub}^{1}_{\mathcal{Y}}(f,\epsilon)$. For any distribution μ on $\mathcal{X} \times \mathcal{Y}$, and $\delta \in (0,1]$, there exists an integer $r \geq 1$, distributions $\{\lambda_j, j \in [r+1]\}$ on $\mathcal{X} \times \mathcal{Y}$ and numbers $\{p_j, j \in [r+1], 0 \leq p_j \leq 1\}$ such that:*

1. *$\forall j \in [r+1], \lambda_j$ is one-message-like for μ and one-way ϵ-monochromatic for f,*
2. *$\mu = \sum_{j=1}^{r+1} p_j \lambda_j$,*
3. *$p_{r+1} \leq \delta$, and*
4. *For $j \in [r], p_j > 2^{-c}\delta$ and $r < \frac{2^c}{\delta}$.*

Proof: By hypothesis, $c = \mathsf{sub}^{1}_{\mathcal{Y}}(f,\epsilon) = \max_\nu \mathsf{sub}^{1}_{\mathcal{Y}}(f,\epsilon,\nu)$. This means for every distribution ν there exists a distribution θ_ν with the properties that θ_ν is one-message-like for ν, one-way ϵ-monochromatic for f, and $\mathsf{S}_\infty(\theta_\nu \,\|\, \nu) \leq c$.

We obtain the distributions $\lambda_1, \ldots, \lambda_{r+1}$ in the decomposition of μ inductively:

- Let $\mu_1 \stackrel{\Delta}{=} \mu, \lambda_1 \stackrel{\Delta}{=} \theta_{\mu_1}$ and $p_1 \stackrel{\Delta}{=} 2^{-\mathsf{S}_\infty(\lambda_1 \,\|\, \mu)}$. By definition of relative co-min-entropy, we have $p_1\lambda_1 \leq \mu$.
- Suppose for some $j \geq 1$, distributions $\lambda_1, \ldots, \lambda_j$ and numbers $p_1, \ldots, p_j$ have been obtained such that $\sum_{k=1}^{j} p_k\lambda_k \leq \mu$.

  Let $q_{j+1} \stackrel{\Delta}{=} \left\| \mu - \sum_{k=1}^{j} p_k\lambda_k \right\|_1$, and $\mu_{j+1} \stackrel{\Delta}{=} \frac{1}{q_{j+1}} \left(\mu - \sum_{k=1}^{j} p_k\lambda_k \right)$. By construction, μ_{j+1} is a probability distribution on $\mathcal{X} \times \mathcal{Y}$.

  In case $q_{j+1} > \delta$, we let $\lambda_{j+1} \stackrel{\Delta}{=} \theta_{\mu_{j+1}}$ and $p_{j+1} \stackrel{\Delta}{=} q_{j+1}2^{-\mathsf{S}_\infty(\lambda_{j+1} \,\|\, \mu_{j+1})}$ and move to $j+2$. Note that by definition of relative co-min-entropy, we have $p_{j+1}\lambda_{j+1} \leq q_{j+1}\mu_{j+1} \leq \mu - \sum_{k=1}^{j} p_k\lambda_k$, i.e., $\sum_{k=1}^{j+1} p_k\lambda_k \leq \mu$.

  In case $q_{j+1} \leq \delta$ we stop the process and let $\lambda_{j+1} \stackrel{\Delta}{=} \mu_{j+1}, p_{j+1} \stackrel{\Delta}{=} q_{j+1}$ and $r \stackrel{\Delta}{=} j$.

Part 1 of the lemma is immediate from our construction and the following properties of the 'one-message-like' relation. Let ν, σ, τ be distributions over $\mathcal{X} \times \mathcal{Y}$.

- If σ is one-message-like for ν, and $p \geq 0$ is such that $p\sigma \leq \nu$, the distribution $\frac{\nu - p\sigma}{\|\nu - p\sigma\|_1}$ is also one-message-like for ν.
- The distributions that are one-message-like for a fixed distribution ν form a convex set. I.e., if σ, τ are one-message-like for ν, the distribution $p\sigma + (1-p)\tau$ is also one-message-like for ν for any $0 \leq p \leq 1$.
- The 'one-message-like' relation is transitive. I.e., if σ is one-message-like for τ, and τ is one-message-like for ν, then σ is one-message-like for ν.

Parts 2 and 3 of the lemma may be verified from our construction. For Part 4 we note that for any $1 \leq j \leq r$,

$$p_j \quad = \quad q_j \, 2^{-\mathsf{S}_\infty(\lambda_j \,\|\, \mu_j)} \quad > \quad \delta 2^{-c}.$$

Since $\sum_{j=1}^{r} p_j \leq 1$, we get $r < 2^c/\delta$. ■

Using the above decomposition of distributions, we can design efficient protocols for relations with small subdistribution complexity.

Lemma 4.3 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, and $0 \leq \epsilon \leq 1/6$ and $0 < \delta \leq 1/6$. Then,*

$$\mathsf{R}^{1,\mathsf{pub}}_{\epsilon+\delta}(f) \quad \leq \quad \mathsf{sub}^1_{\mathcal{Y}}(f, \epsilon) + \log\frac{1}{\delta} + 2.$$

Proof: We show that for every distribution μ on $\mathcal{X} \times \mathcal{Y}$,

$$\mathsf{D}^{1,\mu}_{\epsilon+\delta}(f) \quad \leq \quad \mathsf{sub}^1_{\mathcal{Y}}(f, \epsilon) + \log\frac{1}{\delta} + 2. \tag{3}$$

The result then follows from the Yao min-max principle (Lemma 2.1).

We exhibit a private coin protocol $\mathcal{P}$ for f whose distributional error under μ is at most $\epsilon + \delta$ and communication is at most $c + \log(1/\delta) + 2$, where $c \stackrel{\Delta}{=} \mathsf{sub}^1_{\mathcal{Y}}(f, \epsilon)$. From $\mathcal{P}$ we also get a deterministic protocol with the same communication and distributional error. This implies Eq. (3).

In the protocol $\mathcal{P}$, Alice and Bob start with their inputs XY in distribution μ. Using the decomposition of μ as given by Lemma 4.2, we define a random variable M that is correlated with XY.

We then argue that M may be produced from the knowledge of X alone, and therefore be used as a message to derive a protocol with small distributional error.

Let $\mu = \sum_{j \in [r+1]} p_j \lambda_j$ with p_j, λ_j and r as given by Lemma 4.2 for δ as in the statement of this lemma. The random variable M has support in $[r+1]$. The joint distribution of XYM is defined by

$$\Pr[XYM = (x,y,j)] \quad = \quad p_j \, \lambda_j(x,y),$$

for $(x,y,j) \in \mathcal{X} \times \mathcal{Y} \times [r+1]$. Note that $\Pr[M = j] = p_j$ and the distribution of $XY|(M = j)$ is λ_j. Since for all j, the distribution λ_j is one-message-like for μ, we have $Y|(X = x, M = m) = Y|(X = x)$ for all x, m. Hence $M \to X \to Y$ is a Markov chain. From Lemma 2.10, $Y \to X \to M$ is also a Markov chain. Therefore, the random variable M is a function of X alone, and Alice can generate it using private coins.

To summarize the protocol $\mathcal{P}$, on input x, Alice generates message M as above using private coins, and sends it to Bob. From the construction of XYM, on receiving message j, Bob knows that the conditional distribution on XY is λ_j. On each λ_j with $j \in [r]$ we can ensure that the error of $\mathcal{P}$ is at most ϵ since λ_j is one-way ϵ-monochromatic. On message $r+1$, which occurs with probability at most δ, the error may be as large as 1. Therefore $\mathcal{P}$ has distributional error at most $\epsilon + \delta$ on μ. The communication in $\mathcal{P}$ is bounded by $\lceil \log(r+1) \rceil \le c + \log(1/\delta) + 2$. ■

Combining the bounds in Lemmata 4.1 and 4.3 with standard probability amplification techniques, we get our characterization of one-way communication complexity in terms of the subdistribution bound.

Theorem 4.4 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $0 \le \epsilon \le 1/6$. There are universal constants κ_1, κ_2 such that*

$$\mathsf{sub}^1_{\mathcal{Y}}(f,\epsilon) - 1 \quad \le \quad \kappa_1 \cdot \mathsf{R}^{1,\mathsf{pub}}_\epsilon(f) \quad \le \quad \kappa_2 \left[\mathsf{sub}^1_{\mathcal{Y}}(f,\epsilon) + \log \frac{1}{\epsilon} + 2 \right].$$

Remark: From proofs of Lemma 4.3 and Lemma 4.1, we also conclude that for a distribution μ such that $\mathsf{sub}^1_{\mathcal{Y}}(f,\epsilon) = \mathsf{sub}^1_{\mathcal{Y}}(f,\epsilon,\mu)$ we have $\mathsf{D}^{1,\mu}_\epsilon(f) = \Theta(\mathsf{sub}^1_{\mathcal{Y}}(f,\epsilon,\mu))$ (for a constant ϵ). However for other distributions, $\mathsf{sub}^1_{\mathcal{Y}}(f,\epsilon,\mu)$ may be much smaller than $\mathsf{D}^{1,\mu}_\epsilon(f)$. As an example consider the function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined as $f(x,y) \overset{\Delta}{=} x_1 \vee \bigoplus_{i=2}^n x_i \wedge y_i$. While the one-way communication required for computing this function with distributional error at most 1/5 under the uniform distribution U is $\Omega(n)$, we have $\mathsf{sub}^1_{\mathcal{Y}}(f,0,\mathsf{U}) \le 1$. This is because the distribution with $x_1 = 1$ and remaining bits uniform has 0 error and sits inside U with a scaling of 1/2.

The proof of Theorem 4.4 readily adapts to give a similar relationship between $\mathsf{R}^{1,\Box}_\epsilon(f)$, which is the maximum distributional communication complexity of f under product distributions, and $\mathsf{sub}^{1,\Box}_{\mathcal{Y}}(f,\epsilon)$.

Theorem 4.5 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $0 \le \epsilon \le 1/6$. There are universal constants κ_1, κ_2 such that*

$$\mathsf{sub}^{1,\Box}_{\mathcal{Y}}(f,\epsilon) - 1 \quad \le \quad \kappa_1 \cdot \mathsf{R}^{1,\Box}_\epsilon(f) \quad \le \quad \kappa_2 \left[\mathsf{sub}^{1,\Box}_{\mathcal{Y}}(f,\epsilon) + \log \frac{1}{\epsilon} + 2 \right].$$

Since the one-way distributional communication complexity under product distributions of a Boolean function is captured by its VC-dimension (Theorem 2.2) both quantities in the above theorem are of the same order as the VC-dimension of f (for constant ϵ). The precise dependence on ϵ (when it is not set to a constant) may be inferred from the preceding theorems.

Corollary 4.6 *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a Boolean function. Let $0 \leq \epsilon \leq 1/6$ be a constant. Then $\mathsf{R}_{\epsilon}^{1,\square}(f) = \Theta(\mathsf{sub}_{\mathcal{Y}}^{1,\square}(f, \epsilon)) = \Theta(\mathsf{VC}(f))$.*

5 Direct product theorems for communication complexity

5.1 Two-way protocols

Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. We define the *k-fold product* of f, $f^{\otimes k} \subseteq \mathcal{X}^k \times \mathcal{Y}^k \times \mathcal{Z}^k$ as $f^{\otimes k} \stackrel{\Delta}{=} \{(x_1, \ldots x_k, y_1, \ldots, y_k, z_1, \ldots, z_k) : \forall i \in [k], (x_i, y_i, z_i) \in f\}$. This relation captures k independent instances of the relation f. We show that the two-way product subdistribution bound satisfies the direct product property by considering f and its k-fold product.

Theorem 5.1 *Let $\epsilon, \delta \in (0, 1/6)$, k be a positive integer, and let $q \stackrel{\Delta}{=} (1 - \epsilon/2)^{(1-\delta)k}$. Let $\mu \stackrel{\Delta}{=} \mu_A \otimes \mu_B$ be any product distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathsf{sub}^{\square}(f, \epsilon, \mu) > \frac{48}{\delta\epsilon}$. Then,*

$$\mathsf{sub}^{\square}(f^{\otimes k}, 1 - 2q, \mu^{\otimes k}) \quad > \quad \frac{\delta\epsilon}{16} \cdot k \cdot \mathsf{sub}^{\square}(f, \epsilon, \mu).$$

The same relation holds mutatis mutandis between $\mathsf{sub}^{\square}(f^{\otimes k}, 1 - 2q, z, \mu^{\otimes k})$ and $\mathsf{sub}^{\square}(f, \epsilon, u, \mu)$, where $z = u^k$, and $u \in \mathcal{Z}$ is any fixed output.

Proof: Let $c \stackrel{\Delta}{=} \mathsf{sub}^{\square}(f, \epsilon, \mu)$ and $l \stackrel{\Delta}{=} \frac{\delta\epsilon}{16} \cdot k \cdot c$. Let $\lambda \stackrel{\Delta}{=} \lambda_A \otimes \lambda_B$ be a product distribution on $\mathcal{X}^k \times \mathcal{Y}^k$, such that $\mathsf{S}_{\infty}(\lambda \parallel \mu^{\otimes k}) \leq l$. Let XY be joint random variables distributed according to λ. For $i \in [k]$, let X_i, Y_i represent the components of X, Y respectively in the ith coordinate. The symbol $\mathbf{1}$ denotes a sequence of appropriate length of ones (that is implied by the context).

We show that for any output string $z = z_1 \ldots z_k \in \mathcal{Z}^k$, the distributional error under λ is greater than $1 - 2q$. Formally, define Boolean random variables S_i such that $S_i = 1$ iff the output in the ith coordinate is correct, i.e., $(X_i, Y_i, z_i) \in f$; $S_i = 0$ otherwise. We show the following.

Lemma 5.2 $\Pr_{XY \sim \lambda}[S_1 \ldots S_k = \mathbf{1}] \;\leq\; 2q.$

This lemma directly implies our theorem. ■

Proof of Lemma 5.2: Let $t \stackrel{\Delta}{=} \lceil (1 - \delta)k \rceil$. Our goal is to identify t indices $i_1, \ldots, i_t \in [k]$ such that for each successive index i_j in this sequence, the probability, conditioned upon success on the previous $j - 1$ coordinates, that the protocol succeeds with output z_{i_j} for the coordinate i_j is bounded by $1 - \frac{\epsilon}{2}$. (This implies our lemma.) We do this by choosing the coordinate i_j such that the marginal distribution of XY in that coordinate "sits well" inside μ, and is a product distribution. We ensure that this property holds even when we condition on success in the previous coordinates. Ensuring a product distribution involves conditioning on the inputs to one party (say, Bob) in the previous coordinates. As a consequence, we only identify the required t coordinates for all but a small fraction of "atypical" values for the conditioned input variables. We elaborate on this below.

For a string $y \in \mathcal{Y}^k$ and $i \in [k]$, let y_i denote the sub-string in the ith coordinate of y. We extend this notation to a subset of coordinates $I = \{i_1, \ldots, i_j\} \subseteq [k]$ as $y_I = y_{i_1} \ldots y_{i_j}$ (where the coordinates in the subset are always taken in a canonical order). Similarly for $x \in \mathcal{X}^k$.

In the interest of readability, we sometimes use non-standard notation in our arguments below. For a subset $I \subseteq [k]$, we abbreviate $X_I Y_I$ as XY_I. Similarly, we write XY_i for $X_i Y_i$. The subscript (I, w), where $I \subseteq [k]$ and $w \in \mathcal{Y}^{|I|}$, indicates conditioning on the event $Y_I = w$. For example, $XY_{i,(I,w)}$ is the random variable $X_i Y_i$ conditioned upon the event $Y_I = w$.

Let $X'Y'$ be distributed according to $\mu^{\otimes k}$. We identify a set $\mathcal{B}_I \subseteq \mathcal{Y}^{|I|}$ of "atypical" inputs sub-strings for Bob for each subset I. Let $w \in \mathcal{B}_I \subseteq \mathcal{Y}^{|I|}$, iff

$$\mathsf{S}_\infty(XY_{(I,w)} \,\|\, X'Y'_{(I,w)}) \quad > \quad l + 2k.$$

In Appendix B we bound the probability that Bob's input has an atypical sub-string.

Lemma 5.3 $\Pr_{XY \sim \lambda}[\, (\exists I \subset [k])\; Y_I \in \mathcal{B}_I] \;<\; 2^{-k}$.

Inputs with sub-strings in a set $\mathcal{B}_I$ are precisely the ones for which we are not able to carry out the line of argument outlined above.

We also identify a set $\mathcal{L}_I \subseteq \mathcal{Y}^{|I|}$ of "lucky" input sub-strings for Bob, for each $I \subseteq [k]$ of size less than t. Let $w \in \mathcal{L}_I$ iff $\Pr[S_I = \mathbf{1} | Y_I = w] < 2^{-k}$. Since $2^{-k} \leq q$, for such lucky sub-strings we already have $\Pr[S_1 \ldots S_k = \mathbf{1} | Y_I = w] < q$.

The following lemma captures the main step in our proof.

Lemma 5.4 *Let $I \subseteq [k]$ be of size less than t,and let $w \in \mathcal{Y}^{|I|}$. Then, either*

1. *The sub-string $w \in \mathcal{B}_I$, i.e., $\mathsf{S}_\infty(XY_{(I,w)} \,\|\, X'Y'_{(I,w)}) > l + 2k$, or*
2. *The sub-string $w \in \mathcal{L}_I$, i.e., $\Pr[S_I = \mathbf{1} \mid Y_I = w] < 2^{-k}$, or*
3. *There exists an $i \in [k] - I$, such that $\Pr[S_i = 1 \mid S_I = \mathbf{1}, Y_I = w] < 1 - \frac{\epsilon}{2}$.*

Below we sketch how this implies Lemma 5.2; the technical details are deferred to Appendix B. Lemma 5.4 allows us to select t indices on which the success probability of the protocol is bounded appropriately, so long as parts 1 and 2 are not satisfied. Part 1 is satisfied only for a 2^{-k} fraction of inputs, and we ignore these. As we successively add indices to $I = \{j_1, j_2, \ldots, j_m\}$, if for any value of $m \leq t$, part 2 of the Lemma 5.4 holds, then, in that "branch of conditioning" on the value of Y_I, the probability of success on all k coordinates is bounded by 2^{-k}. If part 2 does not hold for any $m \leq t - 1$, then we keep choosing the indices as given by part 3. As long as Bob's input Y does not contain an atypical sub-string, either part 2 or 3 hold. Therefore we get that the probability of success on all k instances is at most $q + 2^{-k}$. Along with Lemma 5.3 this implies that $\Pr[S_1 \cdots S_k = \mathbf{1}] < 2q$. ■

For the final piece of the argument we prove a key property of sub-distributions.

Lemma 5.5 *Let $0 < \eta < 1/2$ and $\zeta \leq 1$. Let $\mu \overset{\Delta}{=} \mu_A \otimes \mu_B$ and $\omega \overset{\Delta}{=} \omega_A \otimes \omega_B$ be product distributions on $\mathcal{X} \times \mathcal{Y}$. If $\mathsf{S}(\omega \,\|\, \mu) < \eta \cdot \mathsf{sub}^{\square}(f, \zeta, \mu)$, and $\mathsf{sub}^{\square}(f, \zeta, \mu) > \frac{9}{\eta}$, then any zero-communication protocol for f with output $u \in \mathcal{Z}$ has error at least $\zeta - 4\eta$ under ω, i.e., $\Pr_{XY \sim \omega}[(X, Y, u) \notin f] \geq \zeta - 4\eta$.*

Proof: Suppose $\mathsf{sub}^{\square}(f,\zeta,\mu) = d$, $\mathsf{S}(\omega_A \,\|\, \mu_A) = s_A$ and $\mathsf{S}(\omega_B \,\|\, \mu_B) = s_B$. Note that $\mathsf{S}(\omega \,\|\, \mu) = s_A + s_B < \eta d$. Let $r \overset{\Delta}{=} 1/2\eta$. Applying Lemma 2.7 to ω_A and ω_B separately, we get a distribution $\omega' = \omega'_A \otimes \omega'_B$ with $\|\omega - \omega'\|_1 \leq 4/r$ and $\mathsf{S}_\infty(\omega' \,\|\, \mu) \leq r(s_A + s_B + 2) + 2\log\frac{r}{r-1} < d$. This implies, from definition of $\mathsf{sub}^{\square}(f,\zeta,\mu) = d$, that any zero-communication protocol with output $u \in \mathcal{Z}$ has error $> \zeta$ under ω'. Since $\|\omega - \omega'\|_1 \leq 4/r = 8\eta$, Lemma 2.8 tells us that the protocol has error at least $\zeta - 4\eta$ under ω. ■

Proof of Lemma 5.4: We follow the previously described non-standard notation for conditional random variables. In addition, a superscript '**1**' indicates conditioning on the event $S_I = \mathbf{1}$, with I and S_I as in the statement of the lemma.

To prove the lemma, we show that when parts 1 and 2 are false, part 3 holds. By hypothesis, we have

$$\begin{array}{rrcll}
 & \mathsf{S}_\infty(XY_{(I,w)} \,\|\, X'Y'_{(I,w)}) & \leq & l + 2k & \\
\Rightarrow & \mathsf{S}_\infty(XY^{\mathbf{1}}_{(I,w)} \,\|\, X'Y'_{(I,w)}) & \leq & l + 3k, & \text{since } \Pr[S_{I,(I,w)} = \mathbf{1}] \geq 2^{-k}; \\
\Rightarrow & \mathsf{S}_\infty(XY^{\mathbf{1}}_{(I,w),[k]-I} \,\|\, X'Y'_{[k]-I}) & \leq & l + 3k, & \text{from Lemma 2.9;} \\
\Rightarrow & \mathsf{S}(XY^{\mathbf{1}}_{(I,w),[k]-I} \,\|\, X'Y'_{[k]-I}) & \leq & l + 3k, & \text{from Lemma 2.6;} \\
\Rightarrow & \sum_{i \in [k]-I} \mathsf{S}(XY^{\mathbf{1}}_{i,(I,w)} \,\|\, X'Y'_i) & \leq & l + 3k, & \text{from Lemma 2.5;} \\
\Rightarrow & \exists (i \in [k] - I) \quad \mathsf{S}(XY^{\mathbf{1}}_{i,(I,w)} \,\|\, X'Y'_i) & \leq & \frac{l+3k}{k-(1-\delta)k} < \frac{\epsilon c}{8} &
\end{array} \tag{4}$$

In the third inequality, we also used the independence of $X'Y'_I$ and $X'Y'_{[k]-I}$. The last inequality follows from $l = \frac{\delta\epsilon}{16}kc$ and the assumption that $\mathsf{sub}^{\square}(f,\epsilon,\mu) = c > \frac{48}{\delta\epsilon}$.

We show in Appendix B that:

Lemma 5.6 *The distribution of the random variables* $XY^{\mathbf{1}}_{i,(I,w)}$ *is product on* $\mathcal{X} \times \mathcal{Y}$.

Lemma 5.5 tells us that the error in the ith coordinate is therefore at least $\epsilon - \frac{\epsilon}{2} \geq \frac{\epsilon}{2}$. This implies part 3 of the lemma. ■

The direct product property of the subdistribution bound translates to a similar result for the communication complexity of two-way protocols. Its proof appears in Appendix B.

Theorem 5.7 *Let* $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ *be a relation. Let* $\epsilon, \delta \in (0, 1/6)$ *and* k *be a positive integer. Let* $q \overset{\Delta}{=} (1-\epsilon/2)^{(1-\delta)k}$. *Suppose* $\mathsf{sub}^{\square}(f,\epsilon) > \frac{48}{\delta\epsilon}$. *Then,*

$$\mathsf{R}^{\mathsf{pub}}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{R}^{\square}_{1-3q}(f^{\otimes k}) \quad > \quad k \cdot \left[\frac{\delta\epsilon}{16} \cdot \mathsf{sub}^{\square}(f,\epsilon) - 1\right].$$

The same lower bound holds with $\mathsf{sub}^{\square}(f,\epsilon,u)$ *substituted for* $\mathsf{sub}^{\square}(f,\epsilon)$, *for any fixed output* $u \in \mathcal{Z}$.

Theorem 5.1 along with Lemma 3.1 implies the direct product for the rectangle/corruption bound due to Beame *et al.* [BPSW07, Theorem 4.2] (with different parameters). The (two-way) product rectangle bound $\mathsf{rec}^{\square}(f, 1/3, 0)$ for the Set Disjointness function f is $\Omega(\sqrt{n})$ [BFS86]. As a consequence (see [BPSW07, Theorem 4.8]), there is a constant κ such that any two-way protocol for its k-fold product with communication at most $\kappa k\sqrt{n}$ has success probability at most $2^{-\Omega(k)}$.

5.2 One-way protocols

We now explain how the same ideas as in the two-way case lead to a direct product result for one-way communication. The primary difference in this case is that the output of the protocol cannot in general be inferred from the single message sent by Alice. To handle this, we define a variant of the product subdistribution bound which is symmetric with respect to Alice and Bob.

Definition 5.1 (One-way symmetric product subdistribution bound) *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $\mu \stackrel{\Delta}{=} \mu_A \otimes \mu_B$ be a product distribution on $\mathcal{X} \times \mathcal{Y}$. Let $\mathsf{sub}^{1,\square}(f, \epsilon, \mu) \stackrel{\Delta}{=} \min_\lambda \mathsf{S}_\infty(\lambda \,\|\, \mu)$, where λ ranges over all (product) distributions that are* one-message-like *for μ and ϵ-monochromatic for f. We define the* one-way symmetric product subdistribution bound *as $\mathsf{sub}^{1,\square}(f, \epsilon) \stackrel{\Delta}{=} \max_\mu \mathsf{sub}^{1,\square}(f, \epsilon, \mu)$, where μ ranges over all* product distributions *on $\mathcal{X} \times \mathcal{Y}$.*

Note that a distribution that is one-message-like for a product distribution is itself a product distribution.

The following relationships between the one-way symmetric product subdistribution bound and the one-way product subdistribution bound are straightforward and we state them without proof.

Lemma 5.8 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $\mu \stackrel{\Delta}{=} \mu_A \otimes \mu_B$ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Then*

1. $\mathsf{sub}^{1,\square}(f, \epsilon, \mu) \geq \mathsf{sub}^{1,\square}_{\mathcal{Y}}(f, \epsilon, \mu), \quad \mathsf{sub}^{1,\square}(f, \epsilon) \geq \mathsf{sub}^{1,\square}_{\mathcal{Y}}(f, \epsilon).$
2. $\mathsf{sub}^{1,\square}_{\mathcal{Y}}(f, \epsilon, \mu) + \log|\mathcal{Z}| \geq \mathsf{sub}^{1,\square}(f, \epsilon, \mu), \quad \mathsf{sub}^{1,\square}_{\mathcal{Y}}(f, \epsilon) + \log|\mathcal{Z}| \geq \mathsf{sub}^{1,\square}(f, \epsilon).$

We arrive at the following direct product result for one-way symmetric product subdistribution bound along the lines of Theorem 5.1.

Theorem 5.9 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 < \epsilon, \delta < 1/6$ and k be a positive integer. Let $q \stackrel{\Delta}{=} (1-\epsilon/2)^{(1-\delta)k}$. Let $\mu \stackrel{\Delta}{=} \mu_A \otimes \mu_B$ be any product distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathsf{sub}^{1,\square}(f, \epsilon, \mu) > \frac{48}{\delta\epsilon}$. Then*

$$\mathsf{sub}^{1,\square}(f^{\otimes k}, 1 - 2q, \mu^{\otimes k}) \quad > \quad \frac{\delta\epsilon}{16} \cdot k \cdot \mathsf{sub}^{1,\square}(f, \epsilon, \mu).$$

This implies the following direct product result for one-way communication. Its proof is presented in Appendix B.

Corollary 5.10 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 < \epsilon, \delta < 1/6$ and k be a positive integer. Let $q \stackrel{\Delta}{=} (1 - \epsilon/2)^{(1-\delta)k}$. Suppose $\mathsf{sub}^{\square}(f, \epsilon) > \frac{48}{\delta\epsilon}$. Then,*

$$\mathsf{R}^{1,\mathsf{pub}}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{R}^{1,\square}_{1-3q}(f^{\otimes k}) \quad > \quad k \cdot \left[\frac{\delta\epsilon}{16} \cdot \mathsf{sub}^{1,\square}_{\mathcal{Y}}(f, \epsilon) - \log|\mathcal{Z}| - 1\right].$$

This combined with Theorem 4.6 subsumes the strong direct product result due to de Wolf [dW05] for the one-way randomized communication complexity of Index. As has been previously noted, similar results immediately follow for other functions like Set Disjointness and Inner product, whose one-way communication complexity is captured by their VC-dimension.

In fact for a total function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$, we can avoid the loss of the $\log|\mathcal{Z}|$ term.

Theorem 5.11 *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a total function. Let $0 < \epsilon, \delta < 1/6$ and k be a positive integer. Let $q \overset{\Delta}{=} (1-\epsilon/2)^{(1-\delta)k}$. There are universal constants $\gamma_0, \gamma_1 > 0$ such that if $\mathsf{sub}^{\Box}(f, \epsilon) > \frac{\gamma_0}{\delta\epsilon}$. Then,*

$$\mathsf{R}^{1,pub}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{R}^{1,\Box}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{sub}^{1,\Box}_{\mathcal{Y}}(f^{\otimes k}, 1-2q) - k \quad > \quad k \cdot \left[\frac{\delta\epsilon}{\gamma_1} \cdot \mathsf{sub}^{1,\Box}_{\mathcal{Y}}(f, \epsilon) - 1 \right].$$

The proof will be included in the full version of this article.

6 The weak direct product property

Here we derive, in our framework, the weak direct product property of the subdistribution bound. In combination with Lemma 3.2, this subsumes the same result due to Klauck [Kla04] for the rectangle bound in the case of Boolean functions.

Theorem 6.1 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $\epsilon \in (0, 1/4)$, k be a positive integer, and let $q \overset{\Delta}{=} (1-\epsilon/2)^k$. Let μ be any distribution on $\mathcal{X} \times \mathcal{Y}$ such that $k < \mathsf{rec}(f, \epsilon, \mu)/4$. Then,*

$$\mathsf{rec}(f^{\otimes k}, 1-q, \mu^{\otimes k}) \quad > \quad \frac{1}{4} \cdot \mathsf{rec}(f, \epsilon, \mu).$$

The same inequality holds between $\mathsf{rec}(f^{\otimes k}, 1-q, z, \mu^{\otimes k})$ and $\mathsf{rec}(f, \epsilon, u, \mu)$ mutatis mutandis *for $z = u^k$ and any $u \in \mathcal{Z}$.*

Proof: Let $c \overset{\Delta}{=} \mathsf{rec}(f, \epsilon, \mu)$. Let R be a rectangle on $\mathcal{X}^k \times \mathcal{Y}^k$, such that $\mathsf{S}_\infty(\mu_R^{\otimes k} \parallel \mu^{\otimes k}) \leq c/4$. Let XY be joint random variables distributed according to $\mu_R^{\otimes k}$. We use the same notation as in the proof of the strong direct product theorem.

We show that for any output string $z = z_1 \ldots z_k \in \mathcal{Z}^k$, the distributional error under $\mu_R^{\otimes k}$ is greater than $1 - q$. Formally, define boolean random variables S_i such that $S_i = 1$ iff the output in the ith coordinate is correct, i.e., $(X_i, Y_i, z_i) \in f$; $S_i = 0$ otherwise. We show the following.

Lemma 6.2 $\Pr_{XY \sim \mu_R^{\otimes k}}[S_1 \ldots S_k = \mathbf{1}] \; < \; q.$

This lemma directly implies our theorem. ■

Proof of Lemma 6.2: For $i \in [k]$, let $\tilde{i}$ denote the set $[k] - \{i\}$. Let $X'Y'$ be distributed according to $\mu^{\otimes k}$. Now,

$$\begin{aligned} c/4 &\geq \mathsf{S}_\infty(XY \parallel X'Y') \\ &\geq \mathsf{S}(XY \parallel X'Y') \\ &\geq \mathbb{E}_{w \sim XY_{\tilde{1}}}[\mathsf{S}(XY_{1,(\tilde{1},w)} \parallel X'Y'_{1,(\tilde{1},w)})] \\ &= \mathbb{E}_{w \sim XY_{\tilde{1}}}[\mathsf{S}_\infty(XY_{1,(\tilde{1},w)} \parallel X'Y'_{1,(\tilde{1},w)})] \end{aligned}$$

The third inequality above follows from the Chain rule for relative entropy (Lemma 2.4). For the last inequality note that for every w, the distribution of $XY_{1,(\tilde{1},w)}$ corresponds to a rectangle and for a rectangle $T \subseteq X \times Y$, $\mathsf{S}(\mu_T \parallel \mu) = \mathsf{S}_\infty(\mu_T \parallel \mu)$.

Now from Markov's inequality, $\Pr_{w \sim XY_{\tilde{1}}}[\mathsf{S}_\infty(XY_{1,(\tilde{1},w)} \parallel X'Y'_{1,(\tilde{1},w)}) \leq c/2] > 1/2$. For w, such that $\mathsf{S}_\infty(XY_{1,(\tilde{1},w)} \parallel X'Y'_{1,(\tilde{1},w)}) \leq c/2$, from definition of $\mathsf{rec}(f, \epsilon, \mu)$, we get that $\Pr[S_1 = 1 | XY_{\tilde{1}} = w] \leq 1 - \epsilon$. Hence overall $\Pr[S_1 = 1] < 1 - \epsilon/2$.

Let us now condition on $S_1 = 1$. The superscript $\mathbf{1}$ on a random variable indicates conditioning on $S_1 = 1$. If $\Pr[S_1 = 1] \leq 2^{-k}$ then we are done (since $2^{-k} \leq q$). Hence lets assume that $\Pr[S_1 = 1] > 2^{-k}$. Now $\mathsf{S}_\infty(XY^{\mathbf{1}} \parallel X'Y') < k + c/4 < c/2$. Hence as before we would get $\Pr[S_2 = 1 | S_1 = 1] < 1 - \epsilon/2$. Proceeding this way we would finally obtain $\Pr[S_1 S_2 \ldots S_k = \mathbf{1}] < q$.

■

The same proof shows that (two-way) subdistribution bound satisfies the weak direct product property; the only difference is that we reason about a distribution λ that is SM-like for μ. Any such distribution remains SM-like in any one coordinate when conditioned on the remaining inputs and on success in any of the remaining coordinates.

7 Consequences

7.1 Entanglement versus communication

Some of the most important questions in quantum communication concern the power of entanglement. Here we consider quantum communication complexity, as introduced by Yao [Yao93], and investigated extensively thereafter. For definitions concerning quantum computing we refer the reader to Nielsen and Chuang's monograph [NC00].

There are several models of quantum communication complexity: with entanglement and quantum communication, with entanglement and classical communication, and without entanglement but with quantum communication. Due to the phenomenon of quantum teleportation [BBC+93], any protocol with shared entanglement and c qubits of *quantum* communication may be converted to a protocol with an additional c shared EPR-pairs, and a total of $2c$ classical communication.

We are interested in the question whether the quantum communication complexity can be reduced drastically by allowing prior entanglement. So far only small savings in the quantum communication complexity are known when entanglement is allowed. Basically, superdense coding [BW92] allows us to compress classical messages by a factor of 2 when entanglement is available, hence saving a factor of 2 in the quantum communication complexity for the model with entanglement. Also, entanglement can be used like public randomness, leading to additive $\Theta(\log n)$ savings for some functions, e.g., Equality. This gives rise to the question as to how much entanglement is actually necessary to compute a function optimally.

In the analogous situation for public randomness, Newman [New91] shows that $O(\log n)$ public random bits are enough to compute any function with optimal communication complexity. His proof is a black box simulation, in the sense that is does not change the protocol, but rather chooses uniformly at random from a polynomial-size set of strings and runs the protocol with this randomness. Can the amount of entanglement be reduced in the same way for quantum protocols? Jain *et al.* [JRS05b] showed that in fact such a black box approach does not work. Recently, Gavinsky [Gav06] showed that there is a relation that can be computed with $\mathrm{O}(k \log n)$ communication and entanglement in a simultaneous message passing protocol, while every one-way protocol with $\mathrm{o}(k/\log n)$ entanglement and only classical messages needs communication $\Omega(k\sqrt{n}/\log n)$. Hence trying to work with less entanglement increases the communication complexity, or requires drastic changes to the protocol, e.g., going from classical to quantum messages.

Gavinsky derives his result using a direct product theorem for the one-way communication complexity of a certain class of relations. Here we follow the same approach, but use the direct product theorem we prove for one-way communication complexity to get stronger trade-offs.

We begin by defining the relation used in the result. Recall that a perfect matching is an undirected graph in which there is exactly one edge incident on each vertex.

Definition 7.1 (Hidden Matching Relation) *In the* hidden matching *problem* HM_n*,* Alice *gets a string* $x \in \{0,1\}^{2n}$*, and* Bob *gets a perfect matching* M *on* $2n$ *vertices.* Bob *is required to output an edge* $\{j,k\}$ *in* M *along with the bit* $x_j \oplus x_k$*.*

Bar-Yossef, Jayram, and Kerenidis [BYJK04] show that there is a large gap between the classical and quantum one-way complexity of the relation HM_n.

Theorem 7.1 ([BYJK04]) *The one-way quantum communication complexity (with no error and with no prior shared entanglement) of the hidden matching relation* HM_n *is* $\mathrm{O}(\log n)$*. Moreover,* $\mathsf{R}^{1,\Box}_{1/3}(\mathsf{HM}_n) = \Omega(\sqrt{n})$*.*

As mentioned above, with quantum teleportation we can implement the quantum protocol for HM_n as a one-way protocol with $\mathrm{O}(\log n)$ shared EPR-pairs and $\mathrm{O}(\log n)$ classical communication.

Like Gavinsky, we show that a certain amount of entanglement is necessary to preserve the optimal communication complexity of the k-fold product of Hidden Matching.

Theorem 7.2 *The relation* $\mathsf{HM}_n^{\otimes k}$*, with input length* $\Theta(kn\log n)$*, can be computed exactly (with no error) by a one-way quantum protocol with prior entanglement in the form of* $\mathrm{O}(k\log n)$ *shared EPR-pairs, and (classical) communication* $\mathrm{O}(k\log n)$*. There is a constant* $\gamma > 0$ *such that any one-way quantum protocol which uses an entangled state on* γk *qubits needs classical communication* $\Omega(k\sqrt{n})$*.*

Proof: By Theorem 7.1 and the remark following it, $\mathsf{HM}_n^{\otimes k}$ can be computed by a one-way protocol with no error with $\mathrm{O}(k\log n)$ EPR-pairs and using $\mathrm{O}(k\log n)$ bits of classical communication.

From the direct product theorem we prove for one-way classical protocols, Theorem 5.10, there is a constant $d > 0$ such that

$$\mathsf{R}^{1,\mathsf{pub}}_{1-2^{-dk}}(\mathsf{HM}_n^{\otimes k}) \quad \geq \quad \mathsf{R}^{1,\Box}_{1-2^{-dk}}(\mathsf{HM}_n^{\otimes k}) \quad > \quad \Omega(k(\sqrt{n} - 2\log n - 2)) \quad = \quad \Omega(k\sqrt{n}). \qquad (5)$$

Suppose we are given a one-way protocol for $\mathsf{HM}_n^{\otimes k}$ with entanglement ρ over $dk/2$ qubits, classical communication c, and error at most $1/3$. The initial state of the protocol is the entangled state given to Alice and Bob, in tensor product with their inputs. The entire computation of the protocol (unitary operations and measurements) followed by the acceptance criterion is captured by a POVM element E that depends upon the input alone, and acts on the entangled state. The probability of acceptance is then $\mathrm{Tr}(E\rho)$. We replace the entangled state by the *maximally mixed state* over $dk/2$ qubits. This decreases the success probability of the protocol to no worse than $(2/3)\cdot 2^{-dk/2} > 2^{-dk}$. This holds because any quantum state ρ on l qubits (formally a positive semi-definite $2^l \times 2^l$ matrix with trace 1) "sits inside" the maximally mixed state U_l with probability at least 2^{-l}, i.e., $\mathsf{U}_l - 2^{-l}\rho \geq 0$.

An l-qubit maximally mixed state is physically and computationally equivalent to the uniform distribution on l-bit strings. This is a product distribution. As a result, we are left with a private-coin randomized protocol for $\mathsf{HM}_n^{\otimes k}$ with classical communication c and success probability $> 2^{-dk}$. From Eq. (5) we conclude that $c = \Omega(k\sqrt{n})$. ■

If we choose $k = n^p$ for some constant $p > 0$, we get a polynomial gap between the two bounds in the theorem, when the entanglement used is reduced only slightly (from $\Theta(n^p \log n)$ to $\mathrm{O}(n^p)$).

7.2 One-way direct product bound due to Gavinsky

In this section we show that a direct product lower bound shown by Gavinsky [Gav06] follows directly from the one-way direct product theorem we prove.

For a set S, let U_S denote the uniform distribution over the set.

Theorem 7.3 (Gavinsky) *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, where $\mathcal{X} = \{0,1\}^m$. Let $\sigma \in [\log m, m]$, and $\log(1/\delta) \geq 4 + 6(\log|\mathcal{Z}|)/\log m$. Assume that for any random variable X taking values in $\mathcal{X}$ with $\mathsf{H}(X) \geq m - \sigma$, we have*

$$\Pr_{(Y,Z)\sim\mathsf{U}_{\mathcal{Y}\times\mathcal{Z}}}\left[\Pr[(X,Y,Z) \in f] \geq \frac{2}{3}\right] \quad \leq \quad \frac{\delta}{|\mathcal{Z}|}. \tag{6}$$

Then, for $m \geq 64$ and $k \geq \log m$, for any set $B \subseteq \mathcal{X}^k$ of size at least $2^{km - k\sigma/logm}$ the following holds:

$$\Pr_{Y\sim\mathsf{U}_{\mathcal{Y}^k}}\left[\exists z \in \mathcal{Z}^k \ : \ |B_{y,z}| \geq (2/3)^{k/\log m}\,|B|\right] \quad \leq \quad 2^{-k},$$

where $B_{y,z} = \left\{x \in B \ : \ (x,y,z) \in f^{\otimes k}\right\}$.

Proof: Consider any random variable X with min-entropy at least $m - \sigma$ (and therefore Shannon entropy also at least as much), and $Y \sim \mathsf{U}_{\mathcal{Y}}$. The distribution of XY is one-way for $\mathsf{U}_{\mathcal{X}\times\mathcal{Y}}$ and has relative co-min-entropy at most σ. The hypothesis Eq. (6) implies that the probability that any $z \in \mathcal{Z}$ satisfies the relation is at most δ. Let $\tilde{X} \sim \mathsf{U}_B, \tilde{Y} \sim \mathsf{U}_{\mathcal{Y}^k}$. The random variables $\tilde{X}\tilde{Y}$ are one-message-like for the k-fold product of the uniform distribution, and have relative co-min-entropy at most $k\sigma/\log m$. By our direct product theorem, therefore, every zero communication protocol for $f^{\otimes k}$ under this distribution succeeds with probability at most $2^{-\Omega(k)}$. This is the essence of the conclusion of the theorem. ■

Acknowledgements

We thank Ronald de Wolf for sharing with us his write-up [dW05] on the strong direct product theorem for Index and for many useful discussions. We also thank Paul Beame for clarifying the results of [BPSW07] to us.

References

[Aar04] Scott Aaronson. Limitations of quantum advice and one-way communication. In *Proceedings of the 19th Annual IEEE Conference on Computational Complexity*, pages 320–332, 2004.

[ANTSV99] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pages 376–383. ACM Press, 1999.

[AŠdW06] Andris Ambainis, Robert Špalek, and Ronald de Wolf. A new quantum lower bound method, with applications to direct product theorems and time-space tradeoffs. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, pages 618–633. ACM Press, May21–23 2006.

[BARdW07] Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing. Technical Report arXiv:0705.3806v1, ArXiv.org Preprint Archive, `http://arxiv.org/`, May 2007.

[BBC$^+$93] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895–1899, 1993.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science*, pages 337–347, 1986.

[BPSW07] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A direct sum theorem for corruption and a lower bound for the multiparty communication complexity of Set Disjointness. *Computational Complexity*, 2007. To appear.

[BW92] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69(20):2881–2884, 1992.

[BYJK04] Ziv Bar-Yossef, T. S. Jayram, and Iordanis Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 128–137, 2004.

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi 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. Special issue on FOCS 2002.

[CR04] Amit Chakrabarti and Oded Regev. An optimal randomised cell probe lower bound for approximate nearest neighbour searching. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 473–482, 2004.

[CSUU07] Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. In *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity*, 2007. To appear.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew C.-C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications. John Wiley & Sons, New York, NY, USA, 1991.

[dW05] Ronald de Wolf. Random access codes, direct product theorems, and multiparty communication complexity. Unpublished manuscript, incorporated into [BARdW07], 2005.

[Gav06] Dmitry Gavinsky. On the role of shared entanglement. Technical Report quant-ph/0604052, ArXiv.org Preprint Archive, `http://www.arxiv.org/abs/quant-ph/`, April 2006.

[GNW95] Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR-lemma. Technical Report TR95-050, Electronic Colloquium on Computational Complexity, `http://http://eccc.hpi-web.de/eccc/`, 1995. Revision 1, January 1999.

[HJMR07] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity*, 2007. To appear.

[IRW94] Russell Impagliazzo, Ran Raz, and Avi Wigderson. A direct product theorem. In *Proceedings of the Ninth Annual IEEE Structure in Complexity Theory Conference*, pages 88–96, 1994.

[JRS02] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 429–438, 2002.

[JRS03a] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Proceedings of the Thirtieth International Colloquium on Automata Languages and Programming*, volume 2719 of *Lecture notes in Computer Science*, pages 300–315. Springer, Berlin/Heidelberg, 2003.

[JRS03b] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of Set Disjointness. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 220–229, 2003.

[JRS05a] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. On divergence, relative entropy and the substate property. Technical Report quant-ph/0506210, ArXiv.org Preprint Archive, `http://www.arxiv.org/abs/quant-ph/`, June 2005.

[JRS05b] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 285–296, 2005.

[Kla04] Hartmut Klauck. Quantum and classical communication-space tradeoffs from rectangle bounds. In *Proceedings of the 24th Annual IARCS International Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 3328 of *Lecture notes in Computer Science*, pages 384–395. Springer, Berlin/Heidelberg, 2004.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, UK, 1997.

[KNR99] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999. Corrected version available at `http://www.eng.tau.ac.il/ danar/Public-pdf/KNR-fix.pdf`.

[KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-logarithmic depth lower bounds via direct sum in communication complexity. *Computational Complexity*, 5:191–204, 1995.

[KŠdW04] Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 12–21, 2004.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, UK, 2000.

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, 1991.

[PRW97] Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 363–372, 1997.

[PT06] Mihai Pătraşcu and Mikkel Thorup. Higher lower bounds for near-neighbor and further rich problems. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 646–654. IEEE Computer Society Press, Los Alamitos, CA, USA, 2006.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[Sch86] Alexander Schrijver. *Theory of Linear and Integer Programming.* John Wiley and Sons, 1986.

[Sha03] Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1–2):1–22, 2003.

[She07] Alexander A. Sherstov. Communication complexity under product and nonproduct distributions. Technical Report TR07-072, Electronic Colloquium on Computational Complexity, `http://eccc.hpi-web.de/eccc/`, August 2007.

[Yao93] Andrew Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361, 1993.

Appendix

A Subdistribution versus the rectangle bound

In this section, we show that the two-way subdistribution bounds under an arbitrary (possibly non-product) distributions are within a constant factor of the corresponding rectangle bounds. We begin with some observations.

The set of all distributions T such that $S_\infty(T \,\|\, Q) \le k$ is a convex polytope. We call a distribution P a *k-restriction* of Q iff P is the distribution of Q conditioned on an event $\mathcal{E}$ which is the support of some extreme point of this polytope.

Lemma A.1 *Any k-restriction P of Q has relative co-min entropy $S_\infty(P \,\|\, Q) \le k$ and is within $\eta = 2^{-(H_\infty(Q)-k-1)}$ of the corresponding extreme point (in ℓ_1 distance).*

Proof: Let $\mathcal{S}$ be the sample space on which Q is defined. Let P be a k-restriction of Q derived from the distribution $\tilde{P}$ which is an extreme point of the polytope of distributions T such that $\mathsf{S}_\infty(T \,\|\, Q) \le k$.

Let $\mathcal{E} = \operatorname{supp}(\tilde{P})$, so that $P = \tilde{P}|\mathcal{E}$. For each $i \in \mathcal{S}$, we have $0 \le \tilde{P}(i) \le 2^k Q(i)$. Since $\tilde{P}$ is an extreme point, except possibly for one sample point, we have $\tilde{P}(i) = 0$, or $\tilde{P}(i) = 2^k Q(i)$. If this holds for all i, we have $Q(\mathcal{E}) = 2^{-k}$, and $P = \tilde{P}$. Evidently, $S_\infty(P \,\|\, Q) = k$ and $\left\|P - \tilde{P}\right\|_1 = 0 \le \eta$.

Suppose there is a unique $i_0 \in \mathcal{E}$ such that $0 < \tilde{P}(i_0) < 2^k Q(i_0)$. Then

$$\begin{aligned} Q(\mathcal{E}) &= Q(i_0) + \sum_{i \in \mathcal{E}, i \ne i_0} Q(i) \\ &= Q(i_0) + \frac{1 - \tilde{P}(i_0)}{2^k} \qquad (7) \\ &> \frac{\tilde{P}(i_0)}{2^k} + \frac{1 - \tilde{P}(i_0)}{2^k} \quad = \quad \frac{1}{2^k}. \end{aligned}$$

So $\mathsf{S}_\infty(P \,\|\, Q) = -\log Q(\mathcal{E}) < k$. Furthermore, using Eq. (7) we get

$$\begin{aligned} \left\|P - \tilde{P}\right\|_1 &= \sum_{i \in \mathcal{E}} \left|P(i) - \tilde{P}(i)\right| \\ &= \left|\frac{Q(i_0)}{Q(\mathcal{E})} - 1 + 2^k(Q(\mathcal{E}) - Q(i_0))\right| + \sum_{i \in \mathcal{E}, i \ne i_0} \left|\frac{Q(i)}{Q(\mathcal{E})} - 2^k Q(i)\right| \\ &= 2\left(1 - \frac{Q(i_0)}{Q(\mathcal{E})}\right)(2^k Q(\mathcal{E}) - 1) \\ &\le 2(2^k Q(\mathcal{E}) - 1) \\ &= 2(2^k Q(i_0) - \tilde{P}(i_0)). \end{aligned}$$

This is at most η. ■

Since every point in a convex polytope is a convex combination of its extreme points [Sch86], we have

Corollary A.2 *Every distribution P such that $S_\infty(P \,\|\, Q) \le k$, is a convex combination of distributions that within η (in ℓ_1 distance) of k-restrictions of Q, with $\eta = 2^{-(H_\infty(Q) - k - 1)}$.*

Therefore, when Q has sufficiently high min-entropy, i.e., $H_\infty(Q) \gg k$, then we may identify its k-restrictions with the corresponding extreme points.

Definition A.1 (Sampling matrix) *A* sampling matrix *M is any positive semi-definite diagonal matrix such that $I - M$ is also positive semi-definite, where I is the identity matrix of the same dimension.*

In the following, we identify a sampling matrix M with its diagonal, and abbreviate the entry $M(x, x)$ as $M(x)$. The following is immediate.

Lemma A.3 *Let λ, μ be two distributions on $\mathcal{X} \times \mathcal{Y}$.*

1. *The distribution λ is one-message-like for μ with respect to $\mathcal{X}$ if and only if there exists a non-zero sampling matrix M such that $\lambda = \frac{(M\otimes I)\mu}{\mathrm{Tr}(M\otimes I)\mu}$.*
2. *If λ is one-message-like for μ with respect to $\mathcal{X}$ and M is the corresponding sampling matrix, then $\mathsf{S}_\infty(\lambda\|\mu) = -\log \mathrm{Tr}((M\otimes I)\mu) + \max_{x\in\mathcal{X}} \log M(x)$.*
3. *The distribution λ is SM-like for μ if and only if there exist non-zero sampling matrices M, N such that $\lambda = \frac{(M\otimes N)\mu}{\mathrm{Tr}(M\otimes N)\mu}$.*
4. *If λ is SM-like for μ and M, N are the corresponding sampling matrices, then $\mathsf{S}_\infty(\lambda\|\mu) = -\log \mathrm{Tr}((M\otimes N)\mu) + \max_{x\in\mathcal{X}, y\in\mathcal{Y}} \log(M(x)\cdot N(y))$.*

Proof: Consider a distribution λ that is one-message-like for μ with respect to $\mathcal{X}$, and let $k = \mathsf{S}_\infty(\lambda\|\mu) = \mathsf{S}_\infty(\lambda_\mathcal{X}\|\mu_\mathcal{X})$. Define $M(x) = \lambda_\mathcal{X}(x)/2^k\mu_\mathcal{X}(x)$. The diagonal matrix M is a sampling matrix, and $\lambda = (M\otimes I)\mu/\mathrm{Tr}((M\otimes I)\mu)$. Conversely, any distribution of the latter form is one-message-like for μ with respect to $\mathcal{X}$. This proves part 1. Part 3 follows along the same lines and its proof is omitted.

For part 2, we have

$$\begin{aligned}
\mathsf{S}_\infty(\lambda\|\mu) &= \max_{x\in\mathcal{X}, y\in\mathcal{Y}} \log \frac{\lambda(x,y)}{\mu(x,y)} \\
&= \max_{x\in\mathcal{X}, y\in\mathcal{Y}} \log \frac{M(x)\mu(x,y)}{\mu(x,y)\cdot \mathrm{Tr}((M\otimes I)\mu)} \\
&= -\log \mathrm{Tr}((M\otimes I)\mu) + \max_{x\in\mathcal{X}} \log M(x).
\end{aligned}$$

Part 4 follows along the same lines and its proof is omitted. ■

The following lemma is useful in the approximation of successive k-restrictions of a distribution over a product space.

Lemma A.4 *Let μ, ν, ω be probability distributions on $\mathcal{X}\times\mathcal{Y}$, such that ν is one-message-like for μ with respect to $\mathcal{X}$, and ω is one-message-like for ν with respect to $\mathcal{Y}$. Let $S \subseteq \mathcal{X}$ and $T \subseteq \mathcal{Y}$ be such that*

$$\begin{aligned}
\left\|\nu - \mu_{S\times\mathcal{Y}}\right\|_1 &\leq \epsilon_1, \quad \textit{and} \\
\left\|\omega - \nu_{\mathcal{X}\times T}\right\|_1 &\leq \epsilon_2.
\end{aligned}$$

Then

$$\left\|\omega - \mu_{S\times T}\right\|_1 \leq \frac{1.5\,\epsilon_1}{\nu(\mathcal{X}\times T)} + \epsilon_2.$$

Proof: We have

$$\begin{aligned}
\left\|\omega - \mu_{S\times T}\right\|_1 &\leq \left\|\omega - \nu_{\mathcal{X}\times T}\right\|_1 + \left\|\nu_{\mathcal{X}\times T} - \mu_{S\times T}\right\|_1 \\
&\leq \epsilon_2 + \left\|\nu_{\mathcal{X}\times T} - \mu_{S\times T}\right\|_1. \qquad (8)
\end{aligned}$$

We bound the second term in Eq. (8) above as

$$\begin{aligned}
&\left\|\nu_{\mathcal{X}\times T} - \mu_{S\times T}\right\|_1 \\
&\quad= \sum_{x\in\mathcal{X}, y\in T} \left| \frac{\nu(x,y)}{\nu(\mathcal{X}\times T)} - \frac{\mu_{S\times\mathcal{Y}}(x,y)}{\mu_{S\times\mathcal{Y}}(S\times T)} \right| \\
&\quad\leq \sum_{x\in\mathcal{X}, y\in T} \left| \frac{\nu(x,y)}{\nu(\mathcal{X}\times T)} - \frac{\mu_{S\times\mathcal{Y}}(x,y)}{\nu(\mathcal{X}\times T)} \right| + \sum_{x\in\mathcal{X}, y\in T} \left| \frac{\mu_{S\times\mathcal{Y}}(x,y)}{\nu(\mathcal{X}\times T)} - \frac{\mu_{S\times\mathcal{Y}}(x,y)}{\mu_{S\times\mathcal{Y}}(S\times T)} \right|. \qquad (9)
\end{aligned}$$

Further, the first term in Eq. (9) above is bounded as

$$\frac{1}{\nu(\mathcal{X}\times T)}\sum_{x\in\mathcal{X},y\in T}|\nu(x,y)-\mu_{S\times\mathcal{Y}}(x,y)| \quad\leq\quad \frac{\epsilon_1}{\nu(\mathcal{X}\times T)}.$$

Similarly, the second term in Eq. (9) is bounded as

$$\begin{aligned}\sum_{x\in\mathcal{X},y\in T}\left|\frac{\mu_{S\times\mathcal{Y}}(x,y)}{\nu(\mathcal{X}\times T)}-\frac{\mu_{S\times\mathcal{Y}}(x,y)}{\mu_{S\times\mathcal{Y}}(S\times T)}\right| \quad&=\quad \mu_{S\times\mathcal{Y}}(S\times T)\cdot\left|\frac{1}{\nu(\mathcal{X}\times T)}-\frac{1}{\mu_{S\times\mathcal{Y}}(S\times T)}\right|\\ &\leq\quad \frac{\epsilon_1}{2\,\nu(\mathcal{X}\times T)}.\end{aligned}$$

This gives us the claimed bound. ■

We now move to the connection between the subdistribution and the rectangle bound.

Lemma A.5 *Let $f\subseteq\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ be a relation. Let $\epsilon,\delta\in(0,1)$. Let μ be a distribution on $\mathcal{X}\times\mathcal{Y}$, $a=\max_x\mu_{\mathcal{X}}(x)=2^{-\mathsf{H}_\infty(\mu_{\mathcal{X}})}$, $b=\max_y\mu_{\mathcal{Y}}(y)=2^{-\mathsf{H}_\infty(\mu_{\mathcal{Y}})}$, and $c=\mathsf{sub}(f,\epsilon,\mu)$. If $(a+b)2^{(c+1)/(1-\delta)+2}\leq\epsilon$, then*

$$\mathsf{rec}(f,\epsilon,\mu)\quad\geq\quad\mathsf{sub}(f,\epsilon,\mu)\quad\geq\quad(1-\delta)\cdot\mathsf{rec}\bigg(f,\left(1+\frac{1}{\delta}\right)\epsilon,\mu\bigg)-1.$$

The same inequalities hold between $\mathsf{rec}(f,\epsilon,z,\mu)$ and $\mathsf{sub}(f,\epsilon,z,\mu)$ mutatis mutandis *for any fixed $z\in\mathcal{Z}$.*

Proof: The first inequality follows from definitions. For the second inequality let

$$\lambda\quad=\quad\frac{(M\otimes N)\mu}{\mathrm{Tr}((M\otimes N)\mu)}$$

be the SM-like distribution that is ϵ-monochromatic for f and achieves $\mathsf{S}_\infty(\lambda\|\mu)=c=\mathsf{sub}(f,\epsilon,\mu)$. Let $\theta\overset{\Delta}{=}\frac{(M\otimes I)\mu}{\mathrm{Tr}(M\otimes I)\mu}$, where I is the identity matrix, $c_0=\mathsf{S}_\infty(\theta\|\mu)$ and $c_1=\mathsf{S}_\infty(\lambda\|\theta)$. It follows from parts 2 and 4 of Lemma A.3 that

$$c_0+c_1\quad=\quad\mathsf{S}_\infty(\theta\|\mu)+\mathsf{S}_\infty(\lambda\|\theta)\quad=\quad\mathsf{S}_\infty(\lambda\|\mu)\quad=\quad c.$$

By applying Lemma A.2 to the pair $\theta_{\mathcal{X}},\mu_{\mathcal{X}}$, we see that $\theta=\sum_i p_i\theta_i$ with $\mathsf{S}_\infty(\theta_i\|\mu)\leq c_0$ and $\sum_i p_i=1$. Furthermore, for all i, there is a rectangle $R_i=S_i\times\mathcal{Y}$ in $\mathcal{X}\times\mathcal{Y}$ such that

$$\|\theta_i-\mu_{R_i}\|_1\quad\leq\quad 2^{c_0+1}a, \tag{10}$$

and $\mathsf{S}_\infty(\mu_{R_i}\|\mu)\leq c_0$. (Recall that the notation μ_{R_i} refers to the distribution μ conditioned upon the event R_i.)

For all i, let $\lambda_i\overset{\Delta}{=}\frac{(I\otimes N)\theta_i}{\mathrm{Tr}(I\otimes N)\theta_i}$. We have,

$$\lambda\quad=\quad\frac{(I\otimes N)\theta}{\mathrm{Tr}\ (I\otimes N)\theta}\quad=\quad\sum_i p_i\frac{(I\otimes N)\theta_i}{\mathrm{Tr}\ (I\otimes N)\theta}\quad=\quad\sum_i\frac{p_i\ \mathrm{Tr}\ (I\otimes N)\theta_i}{\mathrm{Tr}(I\otimes N)\theta}\ \lambda_i.$$

Let $q_i \stackrel{\Delta}{=} p_i \frac{\mathrm{Tr}(I\otimes N)\theta_i}{\mathrm{Tr}(I\otimes N)\theta}$, and $c_{1i} = \mathsf{S}_\infty(\lambda_i \| \theta_i)$. Note that $\sum_i q_i = 1$ and $\lambda = \sum_i q_i \lambda_i$. Using Lemma A.3, part 2,

$$c_{1i} - c_1 \quad = \quad \mathsf{S}_\infty(\lambda_i \| \theta_i) - \mathsf{S}_\infty(\lambda \| \theta) \quad = \quad \log \frac{\mathrm{Tr}\ (I\otimes N)\theta}{\mathrm{Tr}\ (I\otimes N)\theta_i}.$$

Therefore,

$$\begin{aligned}
\sum_i q_i \, (c_{1i} - c_1) \quad &= \quad \sum_i q_i \log \frac{\mathrm{Tr}\ (I\otimes N)\theta}{\mathrm{Tr}\ (I\otimes N)\theta_i} \\
&= \quad \sum_i p_i \frac{\mathrm{Tr}\ (I\otimes N)\theta_i}{\mathrm{Tr}\ (I\otimes N)\theta} \ \log \frac{\mathrm{Tr}\ (I\otimes N)\theta}{\mathrm{Tr}\ (I\otimes N)\theta_i} \\
&\leq \quad \sum_i p_i \quad = \quad 1. \qquad (11)
\end{aligned}$$

The last inequality above follows from the fact that $x \log(1/x) \leq 1$ for $x > 0$.

Applying Lemma A.2 to the marginal distributions $\lambda_{i,\mathcal{Y}}$ and $\theta_{i,\mathcal{Y}}$, we again express $\lambda_i = \sum_j r_{ij} \gamma_{ij}$, for all i. Here $\mathsf{S}_\infty(\gamma_{ij} \| \theta_i) \leq c_{1i}$ for all i, j, and $\sum_j r_{ij} = 1$. Furthermore, for each i, j there is a rectangle $R_{ij} = \mathcal{X} \times T_{ij} \subseteq \mathcal{X} \times \mathcal{Y}$ such that

$$\left\| \gamma_{ij} - \theta_{ij} \right\|_1 \quad \leq \quad 2^{c_{1i}+1} b_i \qquad (12)$$

and $\mathsf{S}_\infty(\theta_{ij} \| \theta_i) \leq c_{1i}$, where θ_{ij} is the distribution θ_i conditioned upon the event R_{ij}, and $b_i = \max_y \theta_{i,\mathcal{Y}}(y) = 2^{-\mathsf{H}_\infty(\theta_{i,\mathcal{Y}})}$.

Since

$$\begin{aligned}
\mathsf{S}_\infty(\theta_{ij} \| \mu) \quad &\leq \quad \mathsf{S}_\infty(\theta_{ij} \| \theta_i) + \mathsf{S}_\infty(\theta_i \| \mu) \\
&\leq \quad c_{1i} + c_0,
\end{aligned}$$

we have

$$\begin{aligned}
\sum_{i,j} q_i r_{ij} \, \mathsf{S}_\infty(\theta_{ij} \| \mu) \quad &\leq \quad \sum_{i,j} q_i r_{ij} \, (c_{1i} + c_0) \\
&\leq \quad \sum_i q_i \, c_{1i} + c_0 \\
&\leq \quad c_1 + 1 + c_0 \quad = \quad c + 1.
\end{aligned}$$

Let $z \in \mathcal{Z}$ be an output such that λ is (ϵ, z)-monochromatic for f. Let $\epsilon_{ij} \geq 0$ be such that γ_{ij} is (ϵ_{ij}, z)-monochromatic for f. Then $\sum_{i,j} q_i \, r_{ij} \, \epsilon_{ij} \leq \epsilon$. By the Markov Inequality, there is a set $\mathcal{I}$ of pairs ij such that $\epsilon_{ij} \leq \epsilon/\delta$, and $\sum_{i,j \in \mathcal{I}} q_i \, r_{ij} \geq 1 - \delta$. Thus, the expectation of $c_{1i} + c_0$ conditioned on $i, j \in \mathcal{I}$ is at most $\frac{1}{1-\delta}(c+1)$. So there exist a pair $i_0, j_0 \in \mathcal{I}$, such that $\mathsf{S}_\infty(\theta_{i_0 j_0} \| \mu) \leq c_{1i_0} + c_0 \leq \frac{1}{1-\delta}(c+1)$. By the construction of $\mathcal{I}$, we have $\epsilon_{i_0 j_0} \leq \epsilon/\delta$.

Let R be the rectangle $S_{i_0} \times T_{i_0 j_0}$. We claim that

$$\begin{aligned}
\mathsf{S}_\infty(\mu_R \| \mu) \quad &\leq \quad \frac{1}{1-\delta}(c+1), \quad \text{and} \qquad (13) \\
\left\| \gamma_{i_0 j_0} - \mu_R \right\|_1 \quad &\leq \quad \epsilon. \qquad (14)
\end{aligned}$$

The first property says that R is a rectangle with probability at least $2^{-(c+1)/(1-\delta)}$ under μ, and the second implies that the rectangle is $((1 + 1/\delta)\epsilon, z)$-monochromatic. This in turn implies the statement of the lemma.

For Eq. (13), observe that for all $(x, y) \in R_{ij}$, we have

$$\theta_{ij}(x, y) \quad \leq \quad 2^{c_{1i}}\theta_i(x, y) \quad \leq \quad 2^{c_{1i}+c_0}\mu(x, y).$$

Applying this to $i = i_0, j = j_0$, and summing up over $(x, y) \in R = R_{i_0 j_0}$, we get $\mu(R) \geq 2^{-(c_{1i_0}+c_0)} \geq 2^{\frac{-(c+1)}{1-\delta}}$. This is quivalent to the first claim.

For Eq. (14), we invoke the bounds in Eqs. (10) and (12), along with Lemma A.4, as applied to $\mu, \theta_i, \gamma_{ij}$. We get

$$\left\|\gamma_{ij} - \mu_{R_{ij}}\right\|_1 \quad \leq \quad \frac{1.5}{\theta_i(\mathcal{X} \times T_{ij})} 2^{c_0+1}a + 2^{c_{1i}+1}b_i.$$

Since $\theta_i \leq 2^{c_0}\mu$, we have $b_i \leq 2^{c_0}b$. Moreover, $\theta_i(\mathcal{X} \times T_{ij}) \geq 2^{-c_{1i}}$, as $\theta_{ij} \leq 2^{c_{1i}}\theta_i$. Taking $i = i_0, j = j_0$, therefore,

$$\begin{aligned}\left\|\gamma_{i_0 j_0} - \mu_R\right\|_1 &\leq 2^{c_0+c_{1i_0}+2}a + 2^{c_{1i_0}+c_0+1}b \\ &\leq (a + b)2^{(c+1)/(1-\delta)+2} \\ &\leq \epsilon,\end{aligned}$$

by hypothesis. This completes the proof of the lemma. ■

The relationship between the subdistribution bound and the rectangle bound stated in Section 3 now follows.

Proof of Lemma 3.2: Let $a = 2^{-\mathsf{H}_\infty(\mu_\mathcal{X})}$ and $b = 2^{-\mathsf{H}_\infty(\mu_\mathcal{Y})}$. Now either,

$$(a + b)2^{(c+1)/(1-\delta)+2} \quad \leq \quad \epsilon$$

and the hypothesis of Lemma A.5 is satisfied and Lemma 3.2 follows. Otherwise assume W.l.o.g that (if instead it is b then a similar argument follows),

$$a \quad \geq \quad \epsilon \cdot 2^{-(c+1)/(1-\delta)-3} \tag{15}$$

For any particular $x \in \mathcal{X}$, we can find a subset $T \subset \mathcal{Y}$ and an output $z \in \mathcal{Z}$ such that $(x, T, z) \in f$ and $\mu(\{x\} \times T) \geq \mu_\mathcal{X}(x)/|\mathcal{Z}|$. So $\mathsf{rec}(f, 0, \mu)$ is upper bounded by $\mathsf{H}_\infty(\mu_\mathcal{X}) + \log|\mathcal{Z}|$. Now from Eq. (15) we have,

$$\begin{aligned}\mathsf{sub}(f, \epsilon, \mu) &\geq (1 - \delta)(\mathsf{H}_\infty(\mu_\mathcal{X}) - \log\frac{1}{\epsilon} - 3) - 1 \\ &\geq (1 - \delta)(\mathsf{rec}(f, 0, \mu) - \log|\mathcal{Z}| - \log\frac{1}{\epsilon} - 3) - 1,\end{aligned}$$

and Lemma 3.2 follows. ■

B Proofs of some lemmas and theorems

Proof of Lemma 5.3: Let $w \in \mathcal{B}_I$ for some $I \subseteq [k]$. Since

$$\mathsf{S}_\infty(XY_{(I,w)} \parallel X'Y'_{(I,w)}) \quad > \quad l + 2k,$$

there exist $x, y \in \mathcal{X}^k \times \mathcal{Y}^k$ with $y_I = w$ such that

$$\frac{1}{2^{l+2k}} \cdot \Pr[X = x, Y = y \mid Y_I = w] \quad > \quad \Pr[X' = x, Y' = y \mid Y'_I = w].$$

Since $\mathsf{S}_\infty(\lambda \parallel \mu^{\otimes k}) \leq l$ we have, $\Pr[X = x, Y = y] \leq 2^l \cdot \Pr[X' = x, Y' = y]$. Combining these, we get

$$\Pr[Y_I = w] \quad < \quad 2^{-2k} \cdot \Pr[Y'_I = w].$$

Summing up over all possibilities for w, we get

$$\Pr[Y_I \in \mathcal{B}_I] \quad < \quad 2^{-2k}.$$

Therefore, by the union bound over subsets I,

$$\Pr[\,(\exists I \subseteq [k])\; Y_I \in \mathcal{B}_I] \quad < \quad \sum_{I \subseteq [k]} 2^{-2k} \quad = \quad 2^{-k}.$$

■

Proof of Lemma 5.2: Here we rigorously complete the proof of this lemma following the informal sketch in Section 5.1.

In order to bound $\Pr[S_1 \dots S_k = \mathbf{1}]$, we recursively define a subset $J = \{j_1, \dots, j_t\} \subseteq [k]$ of size t for every $y \in \mathcal{Y}^k$. The set J depends upon Bob's input y, and therefore is a random variable correlated with XY. For the purposes of analysis, we also introduce Boolean random variables A_m, L_m, for $m \in [t]$.

Since $\mathsf{S}_\infty(\lambda \| \mu) \leq l$, parts 1 and 2 of Lemma 5.4 are false (with the $I = \emptyset$ and w set to the null string). Let j_1 be the smallest index given by part 3 of the lemma. We set $J = \{j_1\}$, $A_1 = 0 = L_1$.

Suppose indices $I = \{j_1, \dots, j_m\}$ have been defined for input y for some $m \in [t]$. If $y_I \in \mathcal{B}_I \cup \mathcal{L}_I$, i.e., part 1 or 2 of Lemma 5.4 is satisfied with $w = y_I$, then we extend I arbitrarily to a subset J of size t containing I. If part 1 is satisfied we define $A_p = 1, L_p = 0$ for all $p > m$. If part 2 is, then we set $L_p = 1, A_p = 0$ for all $p > m$. Otherwise, we let j_{m+1} be the smallest index i given by part 3 of Lemma 5.4 for I as above and $w = y_I$, and set $A_{m+1} = 0 = L_{m+1}$. Thus, the random variables A_p, L_p are monotonically non-decreasing functions that indicate if parts 1 or 2 were satisfied at any point in the recursive definition of J. In particular, they indicate if the input y is atypical or is lucky.

Lemma 5.3 tells us that $\Pr[A_t = 1] \leq \Pr[\,(\exists I \subseteq [k])\; Y_I \in \mathcal{B}_I] < 2^{-k}$. Since

$$\begin{aligned}
&\Pr[S_1 \dots S_k = \mathbf{1}] \\
&\quad = \quad \Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 1] + \Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 0] \\
&\quad < \quad 2^{-k} + \Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 0],
\end{aligned}$$

if we show that

$$\Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 0] \quad < \quad q + 2^{-k}, \tag{16}$$

we would get a bound of $q + 2^{-k+1} \leq 2q$ as required to prove our lemma.

Now,

$$\begin{aligned}
&\Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 0] \\
&\quad = \quad \Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 0,\ L_t = 1] + \Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 0,\ L_t = 0] \\
&\quad < \quad 2^{-k} + \Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 0,\ L_t = 0], \qquad (17)
\end{aligned}$$

since $L_t = 1$ implies that there is a subset J as defined above such that

$$\Pr[S_1 \dots S_k = \mathbf{1},\ A_t = 0,\ L_t = 1] \quad \leq \quad \mathbb{E}_{Y_J} \Pr[S_J = \mathbf{1} | Y_J] \quad < \quad 2^{-k}.$$

We bound the second term in Eq. (17) by an inductive argument. We show that for all $m \in [t]$,

$$\Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_m = 0,\ L_m = 0] \quad < \quad (1 - \epsilon/2)^m. \tag{18}$$

This is true for $m = 1$ by virtue of Lemma 5.4. Assume that Eq. 18 holds for some $m \geq 1$. Then,

$$\begin{aligned}
&\Pr[S_{j_1} \dots S_{j_{m+1}} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0] \\
&= \sum_{w \in \mathcal{Y}^m} \Pr[S_{j_{m+1}} = 1 \mid S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0,\ Y_{j_1} \cdots Y_{j_m} = w] \\
&\quad \times \Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0,\ Y_{j_1} \cdots Y_{j_m} = w] \\
&< (1 - \epsilon/2) \cdot \sum_{w \in \mathcal{Y}^m} \Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0,\ Y_{j_1} \cdots Y_{j_m} = w] \\
&= (1 - \epsilon/2) \cdot \Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0] \\
&\leq (1 - \epsilon/2) \cdot \Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_m = 0,\ L_m = 0] \\
&< (1 - \epsilon/2)^{m+1}.
\end{aligned}$$

Here, we invoked part 3 of Lemma 5.4 in the first inequality, the monotone non-decreasing property of A_p, L_p in the penultimate step, and the induction hypothesis in the final step. This proves that the second term in Eq. (17) is bounded by q, and therefore Eq. (16) holds. ■

Proof of Lemma 5.6: Recall that $XY \sim \lambda = \lambda_A \otimes \lambda_B$, and therefore are in a product distribution. Therefore, $XY_{(I,w)} = XY|(Y_I = w) = X \otimes (Y|(Y_I = w))$ are in a product distribution. Also $S_I|(Y_I = w) = \mathbf{1}$ is the event $(X_I, w, z_I) \in f^{\otimes|I|}$. So $XY^{\mathbf{1}}_{(I,w)} = XY|(Y_I = w, S_I = \mathbf{1})$ are also in a product distribution. Consequently, the marginal of these random variables on the ith coordinate is also in a product distribution. ■

Proof of Theorem 5.7: The first inequality follows from the definitions. For the second inequality consider a product distribution μ such that $\mathsf{sub}^{\square}(f, \epsilon) = \mathsf{sub}^{\square}(f, \epsilon, \mu)$. Arguing as in the proof of Lemma 4.1, and noting that the the conditional distribution of the inputs given any message is still a product distribution, we get

$$\mathsf{D}^{\mu^{\otimes k}}_{1-2q-2^{-k}}(f^{\otimes k}) \quad > \quad \mathsf{sub}^{\square}(f^{\otimes k}, 1 - 2q, \mu^{\otimes k}) - k.$$

Since $q \geq 2^{-k}$, we get:

$$\begin{aligned}
\mathsf{R}^{\square}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{D}^{\mu^{\otimes k}}_{1-3q}(f^{\otimes k}) \quad &\geq \quad \mathsf{D}^{\mu^{\otimes k}}_{1-2q-2^{-k}}(f^{\otimes k}) \\
&> \quad \mathsf{sub}^{\square}(f^{\otimes k}, 1 - 2q, \mu^{\otimes k}) - k \\
&> \quad \frac{\delta\epsilon}{16} \cdot k \cdot \mathsf{sub}^{\square}(f, \epsilon) - k.
\end{aligned}$$

The last inequality above follows from Theorem 5.1. ■

Proof of Theorem 5.10: The first inequality follows from the definitions. For the second inequality consider a product distribution μ such that $\mathsf{sub}^{1,\square}(f, \epsilon) = \mathsf{sub}^{1,\square}(f, \epsilon, \mu)$. Arguing as in the proof of Lemma 4.1, we get

$$\mathsf{D}^{1,\mu^{\otimes k}}_{1-2q-2^{-k}}(f^{\otimes k}) \quad > \quad \mathsf{sub}^{1,\square}_{\mathcal{Y}}(f^{\otimes k}, 1 - 2q, \mu^{\otimes k}) - k.$$

Now since $q \geq 2^{-k}$, we get

$$\begin{aligned}
\mathsf{R}^{1,[]}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{D}^{1,\mu^{\otimes k}}_{1-3q}(f^{\otimes k}) &\geq \mathsf{D}^{1,\mu^{\otimes k}}_{1-2q-2^{-k}}(f^{\otimes k}) \\
&> \mathsf{sub}^{1,[]}_{\mathcal{Y}}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) - k \\
&\geq \mathsf{sub}^{1,[]}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) - k \log|\mathcal{Z}| - k \qquad (19) \\
&> \frac{\delta\epsilon}{16} \cdot k \cdot \mathsf{sub}^{1,[]}(f, \epsilon, \mu) - k \log|\mathcal{Z}| - k \qquad (20) \\
&= \frac{\delta\epsilon}{8} \cdot k \cdot \mathsf{sub}^{1,[]}(f, \epsilon) - k \log|\mathcal{Z}| - k \\
&\geq k \cdot \left[\frac{\delta\epsilon}{8} \cdot \mathsf{sub}^{1,[]}_{\mathcal{Y}}(f, \epsilon) - \log|\mathcal{Z}| - 1\right]. \qquad (21)
\end{aligned}$$

The Eq. (20) follows from Theorem 5.9. Eq. (19) and (21) follow from Lemma 5.8. ■