Lower Bounds for Elimination via Weak Regularity*

Arkadev Chattopadhyay[1], Pavel Dvořák[2], Michal Koucký[2], Bruno Loff[2], and Sagnik Mukhopadhyay[5]

1 **Tata Institute of Fundamental Research, Mumbai, India**
arkadev.c@tifr.res.in
2 **Charles University, Prague, Czech Republic**
koblich@iuuk.mff.cuni.cz
3 **Charles University, Prague, Czech Republic**
koucky@iuuk.mff.cuni.cz
4 **Charles University, Prague, Czech Republic**
bruno@iuuk.mff.cuni.cz
5 **Tata Institute of Fundamental Research, Mumbai, India**
sagnik@tifr.res.in

Abstract

We consider the problem of elimination in communication complexity, that was first raised by Ambainis et al. [1] and later studied by Beimel et al. [4] for its connection to the famous direct sum question. In this problem, let $f : \{0,1\}^{2n} \to \{0,1\}$ be any boolean function. Alice and Bob get k inputs $x_1, \ldots, x_k$ and $y_1, \ldots, y_k$ respectively, with $x_i, y_i \in \{0,1\}^n$. They want to output a k-bit vector v, such that there exists one index i for which $v_i \neq f(x_i, y_i)$. We prove a general result lower bounding the randomized communication complexity of the elimination problem for f using its discrepancy. Consequently, we obtain strong lower bounds for the functions Inner-Product and Greater-Than, that work for exponentially larger values of k than the best previous bounds.

To prove our result, we use a pseudo-random notion called regularity that was first used by Raz and Wigderson [19]. We show that functions with small discrepancy are regular. We also observe that a weaker notion, that we call weak-regularity, already implies hardness of elimination. Finally, we give a different proof, borrowing ideas from Viola [23], to show that Greater-Than is weakly regular.

1998 ACM Subject Classification F.1.1 Models of computation, F.2.2. Analysis of algorithms and Problem Complexity

Keywords and phrases communication complexity, elimination, discrepancy, regularity, greater-than

Digital Object Identifier 10.4230/LIPIcs.STACS.2017.21

1 Introduction

There is a great interest in computational complexity in so called direct sum questions which consider the complexity of solving k independent instances of a problem at once. In

* The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 616787. The first author is partially supported by a Ramanujan Fellowship of the DST, India and the last author is partially supported by a TCS fellowship.

most reasonable models of computation one would expect the cost of solving k independent instances at once to be k-times the cost of solving a single instance. This is indeed true in some models of computation such as decision trees [18, 10]. However, in some computational models we know that one can achieve savings when solving k-instances simultaneously: in randomized communication complexity [12], in distributional complexity [21] and in zero-error average complexity [15]. For some other models of computations such as boolean circuits we do not know whether some savings are possible. However, in all these cases there is a great interest in the direct sum question as its understanding would shed light on various aspects of complexity. A direct sum theorem for communication complexity would imply the separation of NC^1 from NC^2 [13]. The direct sum for information complexity is used to prove lower bounds in communication complexity (see for example [2, 11, 3, 5]). This motivates the study of the direct sum question.

There are several other problems that are closely related to the direct sum question. In a problem introduced by Beimel et al. [4], one gets k independent instances of some computational problem and has to decide the correct answer for one of the instances of his choice. Provided that the instances are independent there is a hope of picking some easy instance among the k instances. Another problem studied in the literature is the problem of selection, where one should select a positive instance among a k independent instances. This problem was studied exhaustively in the context of structural complexity theory (starting with [20]). Another problem is the problem of distinguishing k positive instances from k negative instances [4].

The least difficult among all these problems is the problem of *elimination*. If $f : \{0,1\}^{2n} \to \{0,1\}$ is a two-player boolean function, the elimination problem $\mathit{elim} \circ f^k$ gives Alice and Bob k n-bit strings each, and asks them to find a k-bit vector z of answers to the k instances which differs from the correct answer for at least one of the instances (i.e. $z_i \neq f(x_i, y_i)$ for some i). Hence, one eliminates a single incorrect vector of answers. In the context of communication complexity this problem was posed by Ambainis et al. [1] and further studied by Beimel et al. [4]. To solve the elimination problem one merely has to solve one of the k instances and negate its answer. Both papers [1, 4] provide examples where one can achieve some savings, typically $\ll k$ bits of communication, and they also provide lower bounds for particular functions.

Ambainis et al. established $\Omega(n/\log n \log\log n)$ lower bound on randomized communication complexity of elimination for the Inner-Product – $\mathit{elim} \circ \mathrm{IP}^k$ – and Disjointness – $\mathit{elim} \circ \textsc{Disj}^k$ – for constant k. Their result can be extended to slightly growing k, but for $k \geq \Omega(\log n)$ the lower bound becomes trivial. Beimel et al. [4] establish a general relationship, showing that the (public-coin) randomized communication complexity of elimination for f^k is lower-bounded by the (public-coin) randomized complexity of f with error roughly $\frac{1}{2} - 2^{-k}$. Due to this large allowed error, the lower bound also becomes trivial for large k.

In this work we also consider the problem of elimination in the setting of (public-coin) randomized communication complexity. We identify two properties of boolean functions that are closely related to the randomized communication complexity of elimination: the *regularity* and *weak-regularity*. Regularity can be thought of as a generalization of discrepancy for functions with non-boolean output. The notion was used by Raz and Wigderson [19] to prove lower bounds in communication complexity. We establish a close relationship between regularity of f^k and the discrepancy of f for any function f. We then relax the notion of regularity to weak-regularity and show that weak-regularity implies lower bounds for the elimination problem. The two results together allow us to lower-bound the randomized communication-complexity of the elimination problem by the inverse-log of discrepancy:

▶ **Theorem 1.** *For any boolean function f and a distribution μ on inputs of f,*

$$\mathcal{R}^{\varepsilon}(elim \circ f^k) \geq \log \frac{1}{Disc(f)} - \log k + \log(1 - \varepsilon \cdot 2^k) - O(1).$$

One corollary of this theorem is a lower-bound of $\Omega(n)$ for elimination of IP^k for k as large as exponential in n. The best known result before our work, due to Beimel et al., does not give any non-trivial lower bound for $k \geq n$. Similarly, we show a bound of $\Omega(\log n)$ for the elimination of GT^k, for $k < n^{1/4}$, where GT is the Greater-Than function. Previous results yielded interesting bounds only for $k \leq \log n$.

Our discrepancy to regularity reduction relies on a sophisticated result of Lee et al. [17] which establishes a direct product theorem for discrepancy. For the Greater-Than function GT we also provide an elementary proof of weak-regularity. This gives a hope of establishing stronger lower bounds for the elimination of functions that have large discrepancy, such as set disjointness. Our proof of weak-regularity designs a hard distribution for GT. Our distribution borrows ideas from the work of Viola [23], but extends them in a novel way. While Viola's distribution is easy for distinguishing k negative instances from k positive instances for $k = O(n)$, our distribution seems hard even for distinguishing exponentially many positive or negative instances.

2 The elimination problem

For a boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, its corresponding *elimination problem*, denoted $elim \circ f^k$ for $k \geq 1$, is defined as follows: Alice and Bob are given k strings $\bar{x} = x_1, \ldots, x_k$, and $\bar{y} = y_1, \ldots, y_k$ respectively, of n-bits each; they must then communicate in order to agree on an *output* string $\mathsf{out} \in \{0,1\}^k$, such that $\mathsf{out} \neq f^k(\bar{x}, \bar{y})$ – i.e., they must *eliminate* a single possibility for the value of $f(x_1, y_1) \ldots f(x_k, y_k)$, from among all possible 2^k values it could take.

The elimination problem was first studied in the context of communication complexity in [1], and later in [4]. Other problems of a similar flavor are studied in those works – enumeration and selection in [1], choice and agreement in [4] – but, in fact, any lower-bounds for elimination will imply the same lower-bounds for these problems. Since proving lower-bounds for elimination will give the same lower-bounds for these other problems, we will not describe the other problems in further detail.

Instead, we will focus on proving lower-bounds against *randomized* protocols for the elimination problem. In this setting, we say that Alice and Bob *succeed* when they output a string $\mathsf{out} \in \{0,1\}^k$ other than $f^k(\bar{x}, \bar{y})$, as above; otherwise we say that they *made an error*.

▶ **Definition 2.** The *randomized communication complexity* of $elim \circ f^k$, denoted $\mathcal{R}^{\varepsilon}(elim \circ f^k)$, is the maximum length of the smallest randomized protocol with shared randomness, which on every input $\bar{x}, \bar{y}$ will succeed except with *error probability* $\leq \varepsilon$.

2.1 Basic observations

The first thing to realize is that if both players compute f for a single instance, then they can output any vector that negates f at that coordinate. So clearly $\mathcal{R}^{\varepsilon}(elim \circ f^k) \leq \mathcal{R}^{\varepsilon}(f) \leq \mathcal{D}(f)$, where $\mathcal{R}^{\varepsilon}(f)$ and $\mathcal{D}(f)$ denote the randomized and deterministic communication complexity of f, respectively.

The next thing to notice is that the randomized task becomes trivial for $\varepsilon \geq 2^{-k}$. Indeed, both players can use their shared randomness to choose a uniformly random $\mathsf{out} \in \{0,1\}^k$,

and outputting this out causes them to succeed with probability $1 - 2^{-k}$. Hence we will assume from this point onward that $\varepsilon \leq 2^{-k}$.

As usual, we will make use of Yao's principle to prove our lower-bounds. So let $\rho \sim \{0,1\}^{2nk}$ be a distribution, and $f : \{0,1\}^{2n} \to \{0,1\}$ a boolean function as before.

▶ **Definition 3.** The *communication complexity* of $elim \circ f^k$ over ρ, denoted $\mathcal{D}^{\varepsilon}_{\rho}(elim \circ f^k)$, is the maximum length of the smallest deterministic protocol which succeeds with error probability $\leq \varepsilon$ on inputs drawn from ρ.

Yao's principle then tells us that

$$\mathcal{R}^{\varepsilon}(elim \circ f^k) = \sup_{\rho} \mathcal{D}^{\varepsilon}_{\rho}(elim \circ f^k).$$

We will use the easy direction ($\geq$) to prove lower-bounds, by presenting a *hard distribution* ρ for which $\mathcal{D}^{\varepsilon}_{\rho}(elim \circ f^k)$ is high.

In preparation to this, let us think what it means when we say $\mathcal{D}^{\varepsilon}_{\rho}(elim \circ f^k) \leq C$. It means that we can partition the space $\{0,1\}^{nk} \times \{0,1\}^{nk}$ into $\leq 2^C$ combinatorial rectangles $R = A \times B$, and in each rectangle we will output a single $\mathsf{out} = \mathsf{out}(R) \in \{0,1\}^k$. Now, each rectangle R is itself naturally partitioned into 2^k *pieces*, one piece for each $z \in \{0,1\}^k$:

$$R^z = \{(\bar{x}, \bar{y}) \in R \mid f^k(x,y) = z\}.$$

(These pieces are possibly empty and non-rectangular.) Then if the success probability is high, it must happen that on the ρ-large rectangles, R^{out} has very little mass. Indeed, the error probability is the sum of every $\rho(R^{\mathsf{out}(R)})$ over the various rectangles R induced by the protocol.

This (vague) condition is both necessary and sufficient, because if we do have a (protocol-induced) partition of $\{0,1\}^{nk} \times \{0,1\}^{nk}$ into rectangles, and on every such rectangle R there is a piece R^z with very little mass (less than $2^{-k}\rho(R)$, say), we may simply output $\mathsf{out} = z$ on this rectangle, and we will have a non-trivial protocol for elimination.

2.2 Regularity

So a natural way of proving a lower-bound would be to show that, under some carefully chosen hard distribution ρ, every rectangle R gets split into pieces R^z all of which are *non-vanishing*. We may eventually come to the following natural definition:

▶ **Definition 4.** Let $n, k \geq 1$ be natural numbers, and $\delta \in [0,1]$; let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a boolean function, and ρ be a distribution over $\{0,1\}^{nk} \times \{0,1\}^{nk}$.

Then f is said to be *δ-weakly-regular* with respect to ρ if for every R and z,

$$\rho(R^z) \geq 2^{-k}(\rho(R) - \delta),$$

where R ranges over combinatorial rectangles in $\{0,1\}^{nk} \times \{0,1\}^{nk}$ and z ranges over $\{0,1\}^k$.

In this definition, if $\rho(R) < \delta$ then the condition is satisfied trivially so δ bounds from below the mass of rectangles for which each R^z should have non-trivial mass.

Now take any protocol π communicating less than C bits, and let R range over the monochromatic rectangles induced by π. Then if f is δ-weakly-regular w.r.t. ρ, π's error probability is

$$\varepsilon = \sum_R \rho(R^{\mathsf{out}(R)}) \geq \sum_R 2^{-k}\big(\rho(R) - \delta\big) \geq 2^{-k}\big(1 - \delta \cdot 2^C\big).$$

This proves that:

▶ **Theorem 5.** *If f is δ-weakly-regular with respect to ρ, then*

$$\mathcal{D}^{\varepsilon}_{\rho}(elim \circ f^k) \geq \log \frac{1-\varepsilon \cdot 2^k}{\delta}.$$

The notion of *weak* regularity naturally arises from first considerations of the elimination problem. Interestingly, it is the weak version of a stronger notion, which we define below, variants of which first appeared in a paper of Raz and Wigderson [19][1] on randomized communication complexity of Karchmer–Wigderson games [14], and also later – in disguise – in related work on simulation theorems [9][2].

The stronger notion says that f is *regular* with respect to ρ if every large rectangle R is partitioned into rectangles R^z of roughly equal mass (i.e., each R^z comprises approximately 2^{-k} fraction of all the ρ-mass of R). Formally:

▶ **Definition 6.** Let n, k, δ, f, and ρ be as in Definition 4. Then f is said to be δ-*regular* with respect to ρ if for every R and z

$$2^{-k}\big(\rho(R)-\delta\big) \leq \rho(R^z) \leq 2^{-k}\big(\rho(R)+\delta\big),$$

where R ranges over combinatorial rectangles in $\{0,1\}^{nk} \times \{0,1\}^{nk}$ and z ranges over $\{0,1\}^k$.

It is obvious that regularity implies weak regularity. We conjecture that the opposite is not true, i.e., that the two notions can be separated for some function f. A formal statement of our conjecture appears in the conclusion (§5). In Section 3 we will establish a strong two-way connection between regularity and matrix discrepancy.

The reader may wonder why the notion is called *regularity*. Indeed, if we think of a bipartite graph G_z with $\{0,1\}^{nk}$ on each side, and put an edge between $\bar{x}$ and $\bar{y}$ if $f^k(\bar{x},\bar{y}) = z$, then the regularity property implies that the edge-density between any two large sets A and B will be roughly what one would expect if the edges of G_z had been chosen at random. This is regularity in the sense of Szemerédi [22].

Regularity is a form of pseudo-randomness

We now give an alternative definition for regularity, which we found curious and worth noting. Say that a set $\mathcal{G} \subset \{0,1\}^{nk} \times \{0,1\}^{nk}$ is δ-*pseudorandom against combinatorial rectangles under* ρ if for any such rectangle $R \subset \{0,1\}^{nk} \times \{0,1\}^{nk}$ we have $\rho(R) - \delta \leq \frac{\rho(R \cap \mathcal{G})}{\rho(\mathcal{G})} \leq \rho(R) + \delta$. Call $\mathcal{G}$ a δ-*hitting set for combinatorial rectangles under* ρ if $\frac{\rho(R \cap \mathcal{G})}{\rho(\mathcal{G})} \geq \rho(R) - \delta$.

Now suppose that $\rho = \mu^k$, where μ is a balanced distribution over $\{0,1\}^n$. It then follows that f is δ-regular with respect to ρ if and only if, for every $z \in \{0,1\}^k$, $(f^k)^{-1}(z)$ is δ-pseudorandom against combinatorial rectangles under ρ. In the same way, f is weakly regular if every inverse image $(f^k)^{-1}(z)$ is a δ-hitting set

It should also be clear that, if δ is small enough and π is a sufficiently-short protocol, or part of a protocol, then whatever output or transcript distribution π may have on ρ, the corresponding distribution on the conditional $\rho \mid (f^k)^{-1}(z)$ will be close, so, in some sense, π cannot distinguish the two distributions. We found this to be a good intuition, although in our results it is used only implicitly.

[1] See the comments on p. 10, before Lemma 2.2, of the full (preprint) version.
[2] When the authors show that block-wise density implies uniformity, they are essentially showing that large R are broken into equally sized R^z.

2.3 The contents of our paper

The next two sections form the core of the paper. In Section 3 we will explain the relationship between discrepancy and regularity, and this will give us the strongest possible randomized lower-bounds for the elimination problem of any function with sufficiently high discrepancy. This result improves the bound from [4] (we will see how in Section 3), and establishes that regularity holds for a large class of functions, a result which is interesting in its own right.

We then present, in Section 4, a proof of weak regularity of the Greater-Than function. Braverman and Weinstein [6] have proven that the Greater-Than function has small discrepancy ($O(\frac{1}{\sqrt{n}})$), and this, together with the previous Section 3, is enough to prove a comparable lower-bound. However, we feel that this proof is interesting for two reasons. First of all, it is a proof from first principles, whereas the proof that discrepancy implies regularity uses the XOR lemma for discrepancy of Lee et al. [17], which can be considered heavy machinery. Second, perhaps more importantly, it is noteworthy that we are still unable to prove elimination lower-bounds for functions with large discrepancy, such as disjointness. Because regularity is a stronger property than discrepancy, there is little hope to use the regularity property to prove such lower-bounds. But because our second proof is based on weak regularity, we can hope that the techniques therein will one day allow us to understand the elimination problem for high-discrepancy functions.

3 Discrepancy and regularity

For the remainder of this section, let $n, k \geq 1$ be natural numbers; let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a boolean function, and μ be a distribution over $\{0,1\}^n \times \{0,1\}^n$; let the letter R always denote a combinatorial rectangle in $\{0,1\}^{nk} \times \{0,1\}^{nk}$ and let z always denote an element of $\{0,1\}^k$. For a given R, let R^z be as defined in Section 2:

$$R^z = \{(\bar{x}, \bar{y}) \in R \mid \forall i \; f(x_i, y_i) = z_i\}.$$

Begin by recalling the well-known notion of matrix discrepancy [16]:

▶ **Definition 7.** The *discrepancy* of f with respect to μ equals

$$\mathrm{Disc}_\mu(f) = \max_{A \subset \{0,1\}^n, B \subset \{0,1\}^n} \left| \sum_{a \in A, b \in B} \mu(a,b)(-1)^{f(a,b)} \right|.$$

We will now show that discrepancy and regularity (Definition 6) are closely related. On one hand, it is easy to see that regularity for $k = 1$ is equivalent to discrepancy. To see this, notice that f is δ-regular with respect to μ if and only if $\mu(R^0)$ and $\mu(R^1)$ are both within $\frac{\mu(R)}{2} - \frac{\delta}{2}$ and $\frac{\mu(R)}{2} + \frac{\delta}{2}$ for all R, which again happens if and only if $\mathrm{Disc}_\mu(f) \leq \delta$.

▶ **Lemma 8.** *Let $k = 1$. Then f is δ-regular with respect to μ if and only if $Disc_\mu(f) \leq \delta$.*

It is somewhat harder to show the relation for $k > 1$. Our proof makes use of the following remarkable result, which was proven in [17, Theorem 19]:

▶ **Lemma 9** (XOR-lemma for discrepancy)**.** *Let μ^t be the t-fold product of μ, and $\oplus_t f$ be the t-fold XOR of f. Then*

$$Disc_{\mu^t}(\oplus_t f) \leq 64^t \cdot Disc_\mu(f)^t$$

Our main result in this section is the following:

▶ **Lemma 10.** *Every f is $O(k \cdot Disc_\mu(f))$-regular with respect to μ^k, whenever $k \leq \frac{1}{64 Disc_\mu(f)}$.*

The constant hidden in the O-notation is $64e$ (e is Euler's number).

Proof. Abbreviate $\rho \equiv \mu^k$. For any given rectangle $R = A \times B \subset \{0,1\}^{nk} \times \{0,1\}^{nk}$, we will compute $\rho(R^z)$ directly. Let a range over A and b over B, and set $f_j = f$ when $z_j = 1$ and $f_j = 1 - f$ when $z_j = 0$; then

$$\rho(R^z) = \sum_{a,b} \rho(a,b) \prod_{j \in [k]} \frac{1 + (-1)^{f_j(a_j,b_j)}}{2};$$

Expanding the product and separating out the resulting "1" term:

$$\rho(R^z) = 2^{-k} \left(\sum_{a,b} \rho(a,b) + \sum_{\varnothing \neq T \subset [k]} \sigma_T \right),$$

$$\sigma_T = \sum_{a,b} \rho(a,b) \prod_{j \in T} (-1)^{f_j(a_j,b_j)}$$

The left term is simply $\rho(R)$, so we now bound $|\sigma_T|$. Say $|T| = t$; let $a' \in \{0,1\}^{n(k-t)}$ range over the projection[3] $A_{[k]\setminus T}$ and let a'' range over the elements of A_T such that $a'a'' \in A$; similarly for b' and b'' with respect to B; then

$$|\sigma_T| \leq \sum_{a',b'} \mu^{k-t}(a',b') \cdot \left| \sum_{a'',b''} \mu^t(a'',b'') \prod_{j \in T} (-1)^{f(a''_j,b''_j)} \right|.$$

This inequality follows from the triangle inequality, since $\rho = \mu^k$. Notice that f_j may be replaced by f, since any resulting multiplication with -1 gets absorbed by taking the absolute value. Now the innermost sum is upper-bounded by $\mathrm{Disc}_{\mu^t}(\oplus_t f)$. Let $D = 64 \cdot \mathrm{Disc}_\mu(f)$. Now Lemma 9 allows us to simplify:

$$|\sigma_T| \leq \sum_{a',b'} \mu^{k-t}(a',b') \cdot D^t \leq D^t.$$

But then

$$\sum_{\varnothing \neq T \subset [k]} |\sigma_T| \leq \sum_{t=1}^{k} \binom{k}{t} D^t = (1+D)^k - 1.$$

By taking the derivative with respect to D, we find that $(1+D)^k - 1 \leq ekD$ whenever $k \leq \frac{1}{D}$. So we conclude that

$$\rho(R) \cdot 2^{-k} - ekD \cdot 2^{-k} \leq \rho(R^z) \leq \rho(R) \cdot 2^{-k} + ekD \cdot 2^{-k}.$$ ◀

[3] A_S for $S \subset [k]$ is the projection of A into the coordinates in S, i.e., those settings a' of the coordinates in S which can be completed with some a'' in the coordinates in $[p] \setminus S$ to get a string in A.

Lower-bounds from discrepancy

The connection between regularity and discrepancy results in a general lower-bound for the elimination problem of small-discrepancy functions:

▶ **Corollary 11.** *For any boolean function $f : \{0,1\}^{2n} \to \{0,1\}$, any distribution μ over $\{0,1\}^{2n}$, and any $k \leq \frac{1}{64 Disc_\mu(f)}$,*

$$\mathcal{D}^{\varepsilon}_{\mu^k}(elim \circ f^k) \geq \log \frac{1}{Disc_\mu(f)} - \log k + \log(1 - \varepsilon \cdot 2^k) - O(1).$$

From Corollary 11, Yao's principle now gives us Theorem 1 mentioned in the introduction. This result should be compared with a similar result, which is implicit in [4]:

▶ **Proposition 12** ([4])**.** *For any f and μ,*

$$\mathcal{D}^{\varepsilon}_{\mu^k}(elim \circ f^k) \geq \log \frac{1}{Disc_\mu(f)} - k + \log(1 - \varepsilon \cdot 2^k) - O(1).$$

As the reader may see, the gain that we achieve is to show the theorem for much larger k. Our result will still hold for an exponentially larger k than what was previously allowed.

Recall the *Greater-Than* function GT_n – $\mathsf{GT}_n(x,y) = 1$ if and only if $x \geq y$, where x and y are two n-bit numbers written in base 2. Braverman and Weinstein [6] have provided a distribution μ on which $\mathrm{Disc}_\mu(\mathsf{GT}_n) = \Omega(\frac{1}{\sqrt{n}})$, and so as a corollary we obtain the following lower-bound:

▶ **Corollary 13.** $\mathcal{R}^{\varepsilon}(elim \circ \mathsf{GT}_n^k) \geq \frac{1}{2}\log n - \log k + \log(1 - \varepsilon \cdot 2^k)$.

When IP_n equals the Inner-Product mod-2 function, we get:

▶ **Corollary 14.** $\mathcal{R}^{\varepsilon}(elim \circ \mathrm{IP}_n^k) \geq \frac{n}{2} - \log k + \log(1 - \varepsilon \cdot 2^k)$.

While our improvement may seem minor at first, we believe it is actually very significant. Notice that, remarkably, the Inner-Product lower-bound is linear even for $k = 2^{\Omega(n)}$. For the Greater-Than function, the previously known lower-bounds would be meaningless for any $k \geq \log n$, whereas our lower-bounds can go as far as $k = \Omega(n)$. We conjecture that these lower-bounds are optimal when the allowed error is $\Omega(2^{-k})$; we will have more to say in the conclusion.

4 Lower-bound for *elim* ∘ GT from first principles

Our "hard" distribution μ on $\{0,1\}^n \times \{0,1\}^n$ is as follows. Let m and ℓ be integers such that $n = m\ell$. (Think of $m = \sqrt{n}$.) We split each n-bit output into m blocks of ℓ bits each. We set $X = X_1 \dots X_m$, where each block X_i is uniformly and independently selected from the set $\mathcal{B}_\ell = \{0,1\}^\ell - \{0^\ell, 1^\ell\}$ (i.e. we forbid the all-0s and all-1s strings). Then we pick a uniformly-random block-index $J \in [m]$, and a uniformly-random bit $Z \in \{-1, 1\}$, and set $Y = X_1 \dots X_{J-1}(X_J + Z)0 \dots 0$, where X_J is interpreted as an integer. Let $\bar{\mu}$ denote the distribution of (X, Y, Z, J) generated by this process; then μ is the projection of $\bar{\mu}$ onto (X, Y).

Then the main theorem of this section is:

▶ **Theorem 15.** GT_n *is $n^{-1/17}$-weakly-regular with respect to μ^k, provided $k \leq n^{1/4}$.*

Our proof is inspired by a paper of Viola [23] who proved lower bound on the randomized communication complexity of GT_n. To prove Theorem 15 we will review some basic lemmas and definitions.

If A and B are two random variables over the same universe, we will use $\Delta(A;B)$ to denote the *statistical distance* (or *total variation distance*) of their distributions; we use $H(A)$ to denote the *entropy* of A's distribution. Definitions of these concepts may be found on [7], or via a simple internet search.

▶ **Lemma 16** (Pinsker's inequality). *Let V be a random variable taking values in a set S, and let U be a uniform variable over S. Then $\Delta(V;U) \leq \sqrt{\log|S| - H(V)}$.*

See [8, p. 44] for a proof of the above.

▶ **Lemma 17.** *For $x \geq 2$, it holds $\log(2^x - 2) \geq x - \frac{1}{2^{x-2}}$.*

Proof. We will prove an equivalent inequality

$$\log(2^x - 2) - \log(2^x) \geq -\frac{1}{2^{x-2}}.$$

By convexity of the exponential function we have $1 - y \geq 2^{-2y}$ for $y \in [0; \frac{1}{2}]$. Then

$$\log(2^x - 2) - \log(2^x) = \log(1 - 2^{1-x}) \geq \log(2^{-2^{2-x}}) = -\frac{1}{2^{x-2}}. \qquad \blacktriangleleft$$

Proof of Theorem 15. Let the random variables $X = X^1 \dots X^k$, $Y = Y^1 \dots Y^k$, $J = J^1 \dots J^k$ and $Z = Z^1 \dots Z^k$ be drawn according to the distribution $\bar{\mu}^k$, by the process given above, so that (X, Y) is distributed according to μ^k.

Fix a large rectangle $R = R_1 \times R_2$ – i.e., a rectangle such that $\mu^k(R) \geq \frac{1}{n}$. Let $\mathcal{X} = (\mathcal{B}_\ell)^{mk}$ be the support of X. Since μ^k has zero mass outside of $\mathcal{X} \times \{0,1\}^{kn}$, assume without loss of generality that $R_1 \subset \mathcal{X}$.

Let W denote a random variable distributed as Y conditioned on $X \in R_1$, and W_z be distributed as Y conditioned on $X \in R_1$ and $Z = z$. We will prove that, for any $z \in \{0,1\}^k$,

$$\Delta(W; W_z) \leq \frac{1}{n^{1/17}}. \tag{I}$$

This implies Theorem 15, because[4]

$$\begin{aligned}
2^k \mu^k(R^z) &= 2^k \cdot \Pr_{\mu^k}[X \times Y \in R^z] \\
&= 2^k \cdot \Pr_{\mu^k}[X \in R_1, Y \in R_2, Z = z] \\
&= \Pr_{\mu^k}[X \in R_1] \cdot \Pr_{\mu^k}[Y \in R_2 \mid X \in R_1, Z = z] \\
&\geq \Pr_{\mu^k}[X \in R_1] \cdot \left(\Pr_{\mu^k}[Y \in R_2 \mid X \in R_1] - \frac{1}{n^{1/17}} \right) \\
&\geq \Pr_{\mu^k}[X \in R_1] \cdot \Pr_{\mu^k}[Y \in R_2 \mid X \in R_1] - \frac{1}{n^{1/17}} \\
&= \mu^k(R) - \frac{1}{n^{1/17}}
\end{aligned}$$

[4] We should note that the same property will trivially hold if $\mu^k(R) < \frac{1}{n}$.

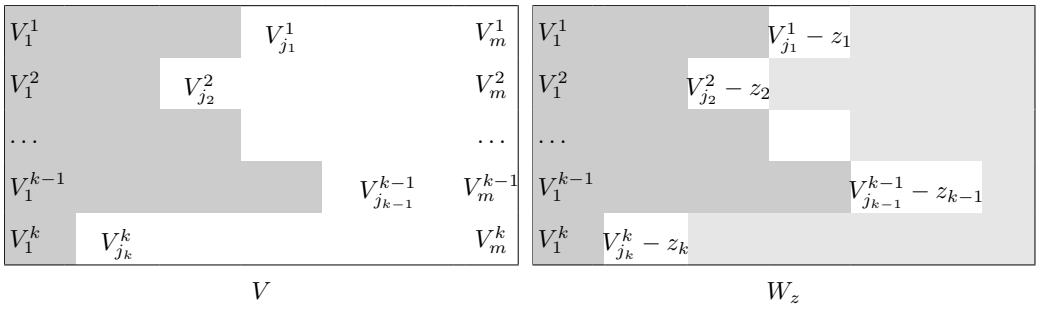

Figure 1 V and W_z. $V_{<j}$ appears in dark gray; light gray marks zeros.

To prove (I), it suffices bounding $\Delta(W_b; W_a) \leq \frac{1}{n^{1/17}}$ for arbitrary $a, b \in \{1, -1\}^k$, because:

$$\begin{aligned}\Delta(W; W_a) &= \frac{1}{2}\sum_{\bar{y}}\big|\mathsf{Pr}[W = \bar{y}] - \mathsf{Pr}[W_a = \bar{y}]\big| \\ &= \frac{1}{2}\sum_{\bar{y}}\Big|\sum_{b\in\{1,-1\}^k}\mathsf{Pr}[Z = b]\cdot\mathsf{Pr}[W_b = \bar{y}] - \mathsf{Pr}[W_a = \bar{y}]\Big| \\ &\leq \sum_{b\in\{1,-1\}^k}\frac{1}{2^k}\cdot\frac{1}{2}\sum_{s}\big|\mathsf{Pr}[W_b = \bar{y}] - \mathsf{Pr}[W_a = \bar{y}]\big| = \sum_{b\in\{1,-1\}^k}\frac{1}{2^k}\Delta(W_b; W_a)\end{aligned}$$

Then let V denote a random variable distributed as X conditioned on $X \in R_1$; since $R_1 \subset \mathcal{X}$ and X is uniform over $\mathcal{X}$, V itself is drawn uniformly from R_1. First we prove that

$$H(V) \geq nk - o(1) - \log n \tag{II}$$

Since V is uniform on R_1, $H(V) = \log|R_1|$. As we assumed R was large,

$$\frac{|R_1|}{|\mathcal{X}|} = Pr_{(X,Y)}[X \in R_1] \geq \frac{1}{n}.$$

Thus by Lemma 17:

$$H(V) \;\geq\; \log\frac{|\mathcal{X}|}{n} = \log\big[(2^{n/m}-2)^{mk}\big] - \log n \;\geq\; mk\Big(\frac{n}{m} - \frac{1}{2^{\frac{n}{m}}-2}\Big) - \log n \;=\; nk - o(1) - \log n.$$

The variable V can be divided into same blocks as the variable X. Thus, $V = V^1, \dots, V^k$ where each V^i is an n-bit number. Each V^i is $V_1^i \dots V_m^i$ where each V_j^i is an ℓ-bit number. Let j be a vector in $[m]^k$. By $V_{<j}$ we denote all blocks V_s^i for $i \in [k]$ and $s < j_i$. The universe of $V_{<j}$ is denoted by $\mathcal{V}_{<j}$. For $j \in [m]^k$, $v \in \mathcal{V}_{<j}$ and $z \in \{-1, 1\}^k$, let $V_{j,v,z}$ be the random variable $V_{j_1}^1 - z_1, \dots, V_{j_k}^k - z_k$, conditioned on $V_{<j} = v$. This is the same distribution as the projection of Y, conditioned on $J = j, Z = z$ and $V_{<j} = v$, onto blocks given by j. Figure 1 illustrates the notation.

Then from the triangle inequality we get

$$\Delta(W_b; W_a) \leq \sum_{j\in[m]^k}\sum_{v\in\mathcal{V}_{<j}}\mathsf{Pr}_{J,V}[J = j, V_{<j} = v]\;\;\Delta(V_{j,v,b}; V_{j,v,a}).$$

Now notice that $V_{j,v,a} = \pi(V_{j,v,b}) = \pi'(V_j)$ for some permutations π and π' of the domain $\{0,1\}^{\frac{nk}{m}}$; then for U uniform over $\{0,1\}^{\frac{nk}{m}}$,

$$\Delta(V_{j,v,b}; V_{j,v,a}) \leq \Delta(V_{j,v,b}; U) + \Delta(U; V_{j,v,a}) = 2\Delta(U; V_{j_1}^1, \dots, V_{j_k}^k \mid V_{<j} = v).$$

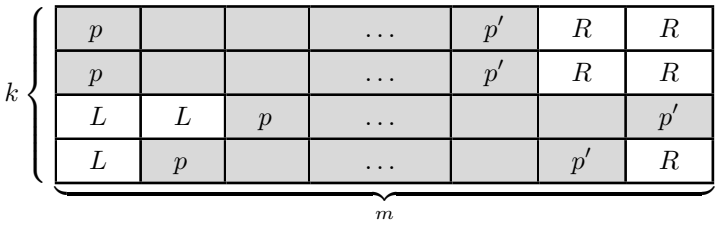

Figure 2 An example how to cover all blocks in the sum M (gray blocks) and the variable L and R. Each rectangle represents a block of $\frac{n}{m}$ bits.

Pinsker's inequality then tells us that $\Delta(V_j;U) \leq \sqrt{\frac{nk}{m} - H(V^1_{j_1},\ldots,V^k_{j_k} \mid V_{<j}=v)}$. We may thus bound

$$\begin{aligned}
\Delta(W_b;W_a) &\leq \frac{1}{m^k}\sum_{jv} \Pr_V[V_{<j}=v]\sqrt{\frac{nk}{m} - H(V^1_{j_1}\ldots V^k_{j_k}|V_{<j}=v)} \\
&\leq \sqrt{\frac{1}{m^k}\sum_{jv} Pr_V[V_{<j}=v]\Big(\frac{nk}{m} - H(V^1_{j_1}\ldots V^k_{j_k}|V_{<j}=v)\Big)} \quad \text{(by concavity of } \sqrt{\cdot}\text{)} \\
&= \sqrt{\frac{nk}{m} - \frac{1}{m^k}\sum_{j} H(V^1_{j_1}\ldots V^k_{j_k}|V_{<j})} \qquad \text{(by definition of conditional entropy)}
\end{aligned}$$

Now we need to bound the sum $\sum_j H(V^1_{j_1}\ldots V^k_{j_k}|V_{<j})$. The sum is over all vectors j in $[m]^k$. We will divide the summands into parts that allow us to use the chain rule. We call a vector $p \in [m]^k$ a *pattern* if p contains 1 in some coordinate. We denote the set of all patterns by $\mathcal{P}$. For a pattern $p \in \mathcal{P}$ we define a width $w(p)$ of pattern p as the maximum of entries of p:

$$w(p) = \max_{i\in[k]} p_i.$$

Denote the set of all patterns of width w by $\mathcal{P}_w$. For an integer s, $(p+s) = (p_1+s, p_2+s,\ldots,p_k+s)$. In this way, we can rewrite the sum of entropies:

$$\sum_{j\in[m]^k} H(V^1_{j_1}\ldots V^k_{j_k}|V_{<j}) = \sum_{w\in[m]}\sum_{p\in\mathcal{P}_w}\sum_{s=0}^{m-w} H(V^1_{p_1+s}\ldots V^k_{p_k+s}|V_{<(p+s)}). \tag{III}$$

Let fix some pattern $p \in \mathcal{P}_w$ and bound the last sum $M = \sum_{s=0}^{m-w} H(V^1_{p_1+s}\ldots V^k_{p_k+s}|V_{<(p+s)})$. Let p' be a vector such that we add $m-w$ to every entry of p, i.e., $p' = (p+m-w)$. Let L be blocks of V "to the left" of p and R be blocks "to the right" of p'. Formally,

$$L = V^1_1\ldots V^1_{p_1-1} \;\ldots\; V^k_1\ldots V^k_{p_k-1} \qquad R = V^1_{p'_1+1}\ldots V^1_m \;\ldots\; V^k_{p'_k+1}\ldots V^k_m$$

The variables L and R are chosen in a way that they, together with the blocks used in the sum M, "cover" all blocks of V. For a better understanding see Figure 2.

Note that variables L and R contains together $(w-1)k$ blocks (i.e., $\frac{n}{m}(w-1)k$ bits) independently of the choice of $p \in \mathcal{P}_w$. The chain rule then says that $H(V) = H(L) + M + H(R|V_{\leq p'})$, and so

$$M = H(V) - H(L) - H(R|V_{\leq p'}) \geq H(V) - \frac{n}{m}(w-1)k.$$

We are ready to bound the sums from Equation (III).

$$\begin{aligned}
\sum_{w\in[m]}\sum_{p\in\mathcal{P}_w}\sum_{s=0}^{m-w} H(V^1_{p_1+s}\dots V^k_{p_k+s}|V_{<(p+s)}) &\geq \sum_{w\in[m]}\sum_{p\in\mathcal{P}_w} H(V) - \frac{n}{m}(w-1)k \\
&\geq \sum_{w\in[m]}\sum_{p\in\mathcal{P}_w} nk - o(1) - \log n - \frac{n}{m}(w-1)k && \text{(by II)} \\
&= \frac{nk}{m}\Big(\sum_{w\in[m]}\sum_{p\in\mathcal{P}_w} m-w+1\Big) - |\mathcal{P}|\Big(o(1)+\log n\Big) \\
&= m^k\frac{nk}{m} - |\mathcal{P}|\Big(o(1)+\log n\Big)
\end{aligned}$$

Bounding the number of all patterns by $|\mathcal{P}| \leq km^{k-1}$, we can also bound

$$\begin{aligned}
\Delta(W_a;W_b) &\leq \sqrt{\frac{nk}{m} - \frac{1}{m^k}\sum_j H(V^1_{j_1}\dots V^k_{j_k}|V_{<j})} \\
&\leq \sqrt{\frac{nk}{m} - \frac{1}{m^k}\Big(m^k\frac{nk}{m} - |\mathcal{P}|\Big(o(1)+\log n\Big)\Big)} \\
&= \sqrt{\frac{km^{k-1}}{m^k}\big(\log n + o(1)\big)} && \text{(for } m=\sqrt{n}, k\leq n^{1/4}\text{)} \\
&= \sqrt{\frac{1}{n^{\frac{1}{4}}}\big(\log n + o(1)\big)} \leq \frac{1}{n^{\frac{1}{17}}} && \blacktriangleleft
\end{aligned}$$

5 Conclusion and open problems

We have given strong lower bounds on the elimination problem with an exponentially improved dependence on k for functions with small discrepancy. We have singled out two measures of complexity, regularity and weak regularity, which appeared implicitly in previous works on communication complexity. We have found regularity to be a natural property to keep in mind, whenever Alice and Bob are given multiple instances $x_1, y_1, \dots x_k, y_k$, and must solve some problem that depends on the pointwise application $f^k(\bar{x}, \bar{y})$.

The first question that is not at all clear to us is: how do these notions relate to each other? Is there a function f for which the elimination problem is hard, and which is still not weakly-regular? Is there a function f which is weakly-regular but not regular?

Open problem. Can we separate the notion of elimination from weak regularity, and weak regularity from regularity?

The second problem we would like to understand is the following. As far as we are able to tell, it could be that elimination is simply as hard as communication of a single instance. We would like to see settled the following conjecture:

Conjecture (Elimination is as hard as communication)

$$\mathcal{R}^{\varepsilon}(\mathrm{elim}\circ f^k) \geq \Omega(\mathcal{R}^{\delta}(f)) \qquad \text{(for } \delta=\Omega(1),\ \varepsilon=\Omega(2^{-k}) \text{ and } k\leq 2^{\Omega(\mathcal{R}^{\delta}(f))}\text{)}$$

Third, and finally, we would like to know if our lower-bounds are tight with respect to the parameter k. We do not know whether this is the case. For example, it could be that if

k is as large as $10n$, say, then it would be possible to solve $elim \circ \mathsf{GT}^k$ with $o(\log n)$ bits; or if $k \geq 2^{10n}$, say, solving $elim \circ \mathrm{IP}_n^k$ would be possible with $o(n)$ bits of communication. Or it could be that stronger lower-bounds can be proven, with much larger k. This is an open question.

Acknowledgements. Part of the research for this work was done at the Institut Henri Poincaré, as part of the *Nexus of Information and Computation Theories* workshop.

References

1. Andris Ambainis, Harry Buhrman, William Gasarch, Bala Kalyanasundaram, and Leen Torenvliet. The communication complexity of enumeration, elimination, and selection. *Journal of Computer and System Sciences*, 63(2):148–185, 2001.
2. Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proceedings of the 43rd Symposium on Foundations of Computer Science (FOCS)*, pages 209–218, 2002.
3. Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd Symposium on Theory of Computing (STOC)*, pages 67–76, 2010.
4. Amos Beimel, Sebastian Ben Daniel, Eyal Kushilevitz, and Enav Weinreb. Choosing, agreeing, and eliminating in communication complexity. *Computational Complexity*, 23(1):1–42, 2014.
5. Mark Braverman and Anup Rao. Information equals amortized communication. In *Proceedings of the 52nd Symposium on Foundations of Computer Science (FOCS)*, pages 748–757, 2011.
6. Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 459–470. Springer, 2012.
7. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.
8. Imre Csiszar and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
9. Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proceedings of the 47th Symposium on Theory of Computing (STOC)*, pages 257–266. ACM, 2015. 0.
10. R. Jain, H. Klauck, and M. Santha. Optimal direct sum results for deterministic and randomized decision tree complexity. *Information Processing Letters*, 110(20):893–897, 2010.
11. T. S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In *Proceedings of the 35th Symposium on Theory of Computing (STOC)*, pages 673–682, 2003.
12. M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. *SIAM Journal on Discrete Mathematics*, 8(1):76–92, 1995.
13. M. Karchmer, R. Raz, and A. Wigderson. Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Computational Complexity*, 5(3/4):191–204, 1995.
14. Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM Journal on Discrete Mathematics*, 3(2):255–265, 1990.
15. Gilat Kol, Shay Moran, Amir Shpilka, and Amir Yehudayoff. Direct sum fails for zero-error average communication. In *Proceedings of the 5th Innovations in Theoretical Computer Science (ITCS)*, pages 517–522, 2014.

16 Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, New York, NY, USA, 1997. 0.

17 Troy Lee, Adi Shraibman, and Robert Špalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd Conference on Computational Complexity (CCC)*, pages 71–80. IEEE, 2008.

18 Noam Nisan, Steven Rudich, and Mike Saks. Products and help bits in decision trees. *SIAM Journal on Computing*, 28(3):1035–1050, 1999.

19 Ran Raz and Avi Wigderson. Probabilistic communication complexity of boolean relations. In *Proceedings of the 30th Symposium on Foundations of Computer Science (FOCS)*, pages 562–567, 1989.

20 Alan L. Selman. P-selective sets, tally languages, and the behavior of polynomial time reducibilities on NP. In *Proceedings of the 6th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 546–555, 1979.

21 Ronen Shaltiel. Towards proving strong direct product theorems. In *Proceedings of the 16th Conference on Computational Complexity (CCC)*, pages 107–119, 2001.

22 Endre Szemerédi. Regular partitions of graphs. Technical report, DTIC Document, 1975.

23 Emanuele Viola. The communication complexity of addition. *Combinatorica*, 35(6):703–747, 2015.