Semi-Direct Sum Theorem and Nearest Neighbor under ℓ_∞

Mark Braverman[1]

Department of Computer Science, Princeton University, 35 Olden St. Princeton NJ 08540, USA
mbraverm@cs.princeton.edu

Young Kun Ko

Department of Computer Science, Princeton University, 35 Olden St. Princeton NJ 08540, USA
yko@cs.princeton.edu

Abstract

We introduce semi-direct sum theorem as a framework for proving asymmetric communication lower bounds for the functions of the form $\bigvee_{i=1}^{n} f(x, y_i)$. Utilizing tools developed in proving direct sum theorem for information complexity, we show that if the function is of the form $\bigvee_{i=1}^{n} f(x, y_i)$ where Alice is given x and Bob is given y_i's, it suffices to prove a lower bound for a single $f(x, y_i)$. This opens a new avenue of attack other than the conventional combinatorial technique (i.e. "richness lemma" from [12]) for proving randomized lower bounds for asymmetric communication for functions of such form.

As the main technical result and an application of semi-direct sum framework, we prove an information lower bound on c-approximate Nearest Neighbor (ANN) under ℓ_∞ which implies that the algorithm of [9] for c-approximate Nearest Neighbor under ℓ_∞ is optimal even under randomization for both decision tree and cell probe data structure model (under certain parameter assumption for the latter). In particular, this shows that randomization cannot improve [9] under decision tree model. Previously only a deterministic lower bound was known by [1] and randomized lower bound for cell probe model by [10]. We suspect further applications of our framework in exhibiting randomized asymmetric communication lower bounds for big data applications.

2012 ACM Subject Classification Theory of computation → Communication complexity

Keywords and phrases Asymmetric Communication Lower Bound, Data Structure Lower Bound, Nearest Neighbor Search

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2018.6

1 Introduction

Direct Sum Theorem in communication and information complexity is a key technique in lower bounding the communication and information complexity of computing functions of the following form

$$F(\vec{x}, \vec{y}) = \bigvee_{i=1}^{n} f(x_i, y_i)$$

where Alice is given a length n string $\vec{x}$ and Bob is given another length n string $\vec{y}$ according to some distribution. Many fundamental functions, namely Disjointness and Equality (under

[1] Research supported in part by an NSF Awards, DMS-1128155, CCF- 1525342, and CCF-1149888, a Packard Fellowship in Science and Engineering, and the Simons Collaboration on Algorithms and Geometry. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018).
Editors: Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer; Article No. 6; pp. 6:1–6:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

de Morgan's Law), are of the above form. Direct Sum Theorem allows us to reduce computing f to computing F. This implies that the lower bound for f can be translated to a lower bound for F. Since f is a function over a smaller number of bits (1-bit AND in the case of Disjointness) providing lower bounds for f is much easier than for F from the technical perspective.

Asymmetric communication complexity addresses a different setting where the bit-size of Bob's input is much larger than Alice's input. One example of such setting is determining whether $S \cap T = \emptyset$ when $|S| \ll |T|$ (Lopsided Disjointness). In this setting, it is more meaningful to lower bound the length of message sent by Alice and Bob separately, instead of bounding the total length of the transcript as in the symmetric setting. In asymmetric setting, the trivial protocol where Alice sends her whole input is usually the most efficient protocol in terms of total number of bits communicated between Alice and Bob.

Lower bounds in asymmetric communication are not only interesting from pure communication complexity theory perspective but also from its applications to lower bounds in data structures as seen in [11, 12, 14]. It is now a well-taught fact in graduate communication complexity classes that asymmetric communication lower bounds translate to data structure (cell probe and decision tree models) lower bounds. We explain this connection formally in the appendix Section B.

The key technical tool that was used for showing asymmetric communication lower bounds is **"Richness Lemma"** from [12]. This lemma is an extension of monochromatic rectangle lower bound for symmetric communication complexity. The main observation is that lower bounds on the height and width of any monochromatic rectangle translates to lower bounds on asymmetric communication complexity. But similar to symmetric communication setting, rectangle lower bounds usually require complicated combinatorial lemmas. Further adding to this technical complication, it is also not a great tool for analyzing the performance of randomized protocols. For randomized protocols, rectangles are no longer monochromatic but are allowed to be "roughly" monochromatic, which makes it harder to argue that such rectangle does not exist.

To avoid technical complications from asymmetric communication complexity, [13, 10] introduces a notion of "robust expansion," to lower bound the number of cells required in cell-probe data structure model even under randomization. However, this method does not imply lower bound in decision tree model as shown in [1] for the deterministic case.

Instead of using "Richness Lemma," we take an information theoretic approach in establishing asymmetric communication lower bound, similar to the one observed in [8] for lopsided disjointness. But we emphasize that unlike in [8] or lopsided disjointness, the functions in consideration are of the form

$$F(x, \vec{y}) = \bigvee_{i=1}^{n} f(x, y_i).$$

These functions are especially interesting for big data applications. In particular, it provides a lower bound for the following set of queries:

> Alice (user) with input x queries Bob (database) with $y_1, \dots, y_n$ whether there exists a data point that satisfies certain condition (i.e. $f(x, y_i)$).

More importantly, unlike in [8], Alice's input is recycled over the singleton function f, while Bob's inputs are all distinct. Therefore, simple application of Direct Sum Theorem does not suffice for this application.

As in Direct Sum Theorem, we reduce the task of computing f to computing F. Indeed, as in symmetric communication setting, direct sum does not necessarily hold if one considers the total length of the transcript that is the number of communicated bits. Instead, we introduce "asymmetric information cost," and analogous "asymmetric information complexity," as introduced in [8, 15] as the key complexity measure in asymmetric communication. We then show that asymmetric information cost lower bounds asymmetric communication cost with an analogous argument from symmetric communication setting.

1.1 Technical Results

Semi-Direct Sum

With a properly defined notion on asymmetric information cost, it is rather straightforward to prove the following semi-direct sum theorem.

▶ **Theorem 1** (Semi-Direct Sum (informal)). *Consider a protocol* Π *for computing* F *with the following guarantee :*

- *For all* $(x, \vec{y}) \in F^{-1}(1)$, $\Pr_\pi[\pi(x, \vec{y}) \neq 1] < \varepsilon_1$
- *For all* $(x, \vec{y}) \in F^{-1}(0)$, $\Pr_\pi[\pi(x, \vec{y}) \neq 0] < \varepsilon_0$

that is with a prior free error guarantee. Suppose μ *is a distribution such that* $X, Y_1, \dots, Y_n$ *are i.i.d. Furthermore, suppose* $f(X, Y_i) = 1$ *with probability at most* $o(1/n)$ *under* μ*. Then there exists a protocol* Π' *(with input* X *and* Y_i*) for computing a singleton* f *with the following guarantee :*

- *For all* $(x, y_i) \in f^{-1}(1)$, $\Pr_{\pi'}[\pi'(x, y_i) \neq 1] < \varepsilon_1 + 0.01$
- *For all* $(x, y_i) \in f^{-1}(0)$, $\Pr_{\pi'}[\pi'(x, y_i) \neq 0] < \varepsilon_0 + 0.01$

with $n \cdot I_{(x,y_i)\sim\mu}(\Pi'; Y_i|X) \leq I_{(x,\vec{y})\sim\mu}(\Pi; Y_1, \dots Y_n|X)$ *and* $|\Pi'_a| \leq |\Pi_a|$*, that is the total number of bits sent by Alice.*

This allows us to translate a lower bound on a singleton function f, to a lower bound on the whole F if the distribution for measuring the information cost is i.i.d. In particular, if one shows that for any Π' with the prior free error guarantee require $I_{(x,y_i)\sim\mu}(\Pi'; Y_i|X) > \beta$, this implies that Π requires $I_{(x,\vec{y})\sim\mu}(\Pi; \vec{Y}|X) > \beta n$.

We emphasize that the error guarantee is "prior free" in a sense that the protocol must be correct (up to some error parameters) for all inputs, instead of being correct on average under some distribution. In particular, the probability of error is only over the protocol, not the input. Under the guarantee, we can convert the prior free lower bound to a lower bound for a fixed prior using minimax argument as in [3]. More formally, we have the following theorem.

▶ **Theorem 2** (Minimax Theorem). *Fix some* $0 < \alpha < 1$*. Consider a protocol* Π *for computing* F *such that for any* $(x, \vec{y})$, $\Pr_\pi[\pi(x, \vec{y}) \neq F(x, \vec{y})] < \varepsilon/\alpha$*. Then there exists a "hard" distribution* μ *on* X *and* $\vec{Y}$ *such that for any protocol* Π' *such that* $\Pr_{(x,\vec{y})\sim\mu}[\pi'(x, \vec{y}) \neq F(x, \vec{y})] < \varepsilon$,

$$\frac{I_{(x,\vec{y})\sim\mu}(\Pi'; \vec{Y}|X)}{1-\alpha} \geq \sup_\nu I_{(x,\vec{y})\sim\nu}(\Pi; \vec{Y}|X).$$

Intuitively, this implies that the asymmetric information cost for protocols with distributional error guarantee and prior free error guarantee are at most constant factor within each other. In particular, setting $\alpha = 1/2$, and exhibiting a lower bound on $I_{(x,\vec{y})\sim\nu}(\Pi; \vec{Y}|X)$ for a particular ν (which is a product distribution from semi-direct sum theorem), we exhibit an existence of a hard distribution μ such that any protocol that errors at most ε on average must reveal a large amount of information about $\vec{Y}$.

The main technical disadvantage of the rectangle bound is that it is extremely challenging to give any bounds on non-product distribution over Y_i's as mentioned in [1]. We avoid this complication by a standard technique in information complexity. We first provide a lower bounds on prior-free protocols. Then using minimax argument, we can argue that a hard distribution exists **without having to explicitly construct a hard distribution**.

Nearest Neighbor Lower Bound

As the main application of our approach, we revisit c-approximate Nearest Neighbor Search under ℓ_∞-norm, improving the corresponding asymmetric communication lower bound to asymmetric information lower bound. This **strengthens the deterministic asymmetric communication lower bound given in [1] to a randomized asymmetric communication lower bound**, and obtain two corollaries for decision tree model and cell-probe model respectively. (1) For cell-probe model, this **reproves and simplifies the proof of [10]**; (2) For decision tree model, this shows that **[9] is tight for decision tree model even under randomization**, substantially improving upon [1] which only proved tightness for deterministic decision tree model. Therefore, this **closes the remaining gap for c-approximate Nearest Neighbor Search under ℓ_∞-norm under (randomized) decision tree model**.

Formally, consider the partial function

$$F(x,\vec{y}) = \begin{cases} 1 & \text{if } \exists y_i \ \|x - y_i\|_\infty \le 1 \\ 0 & \text{if } \forall y_i \ \|x - y_i\|_\infty \ge c. \end{cases}$$

for $c > 3$ with $x, y_1, \ldots, y_n \in \{0, \ldots, M\}^d$ where Alice (user) is given x and Bob (database) is given $y_1, \ldots, y_n$. The goal of the database is to compute F. [9] showed an unorthodox yet efficient, deterministic data structure which achieves $O(\log_\rho \log d)$ approximation using $O(dn^\rho \text{poly} \log n)$ space. Surprisingly, [1] showed the optimality of [9] under any deterministic data structure (under decision tree model) while [10] showed that this is optimal for $n^{o(1)}$-word size (randomized) cell probe data structure. Whether randomization can improve [9] under decision tree model remained an open problem for over a decade. We answer this negatively by using semi-direct sum theorem. In particular, we prove the following asymmetric information lower bound.

▶ **Theorem 3.** *Set $\rho = (\tau \log d)^{1/c}$, and $d \ge (2 \cdot \log n)^{\frac{1}{1-\tau}}$ for some constant $1 > \tau > 0$. Let Π be a protocol (with both private and public randomness) that computes F with the following guarantee*

- *For all $(x,\vec{y}) \in F^{-1}(1)$, $\Pr[\pi(x,y) \ne 1] < 0.05$*
- *For all $(x,\vec{y}) \in F^{-1}(0)$, $\Pr[\pi(x,y) \ne 0] < 0.05$*

There exists a distribution u such that for any such Π and for any sufficiently small constant $\delta > 0$, if $|\Pi_a| \le \delta\rho \log n$ i.e. the number of bits sent by Alice, then $I_b = I_{(x,\vec{y}) \sim u^{\otimes n+1}}(\Pi; \vec{Y}|X) \ge n^{1-O(\delta)}$.

This is an analogous theorem to asymmetric communication lower bound provided in [1]. But we point out that the error guarantee required is prior free and this is an information lower bound which is stronger than communication lower bound. This theorem, together with standard techniques in translating asymmetric communication lower bound to decision tree lower bound, shows that [9] is optimal under randomized decision tree model with prior free error guarantee.

[1] raised an interesting technical bottleneck that the hard distribution for randomized asymmetric communication (with distributional error guarantee) must be a *non-product distribution* over the inputs. We avoid this technical bottleneck by applying the minimax

theorem, thereby exhibiting a lower bound with distributional error guarantee over the inputs without explicitly constructing a hard (non-product) distribution. Furthermore, using the standard technique for translating asymmetric communication lower bound to cell-probe lower bound, this reproves [10].

2 Preliminary

2.1 Asymmetric Information Cost

In this section, we define analogous quantities for asymmetric communication. We refer the reader to Section A for the definitions in symmetric settings. First, recall the following natural definition for defining asymmetric communication cost.

▶ **Definition 4** (Asymmetric Communication Cost). Protocol Π has asymmetric communication cost of (a, b) if $|\Pi_a| := \sum_{i:\text{odd}} |\Pi_i| \leq a$ and $|\Pi_b| := \sum_{i:\text{even}} |\Pi_i| = b$.

Since we have two parameters, naive notion of lower bound on one quantity no longer applies. Here, the lower bound will be of the form "if $|\Pi_a| < a$, then $|\Pi_b| > b(a)$."

We introduce analogous definitions for information cost, as first defined in [15], extending the intuition explained in Section A.

▶ **Definition 5** (Asymmetric Information Cost). Protocol Π has asymmetric information cost of $[I_a, I_b]$ under μ if $I_a^\mu(\Pi) := I_\mu(\Pi; X|Y) \leq I_a$ and $I_b^\mu(\Pi) := I_\mu(\Pi; Y|X) \leq I_b$.

Similar to information cost of a protocol being a lower bound for communication cost of a protocol, it is straightforward prove an analogous lemma for asymmetric setting.

▶ **Lemma 6** (Asymmetric Information Cost lower bounds Communication Cost). *For any protocol Π and any distribution μ, $I_a^\mu(\Pi) \leq |\Pi_a|$ and $I_b^\mu(\Pi) \leq |\Pi_b|$.*

Proof. Suppose at round r, Alice sends a_r bits. Now we can write the information cost incurred at round r, that is how much additional information Bob gains, as $I_\mu(M_{r+1}; X|Y, \Pi_r)$, where M_{r+1} is the message sent by Alice. Now we can write

$$I_\mu(M_{r+1}; X|Y, \Pi_r) \leq H_\mu(M_{r+1}|Y, \Pi_r) \leq a_r$$

where the last inequality holds since M_{r+1} is of length a_r and thus can have entropy of at most a_r. Applying Fact 26,

$$I_\mu(\Pi; X|Y) = \sum_r I_\mu(M_{r+1}; X|Y, M_1, \ldots, M_r) \leq \sum_r a_r = |\Pi_a|$$

Similarly we also get $I_\mu(\Pi; Y|X) \leq |\Pi_b|$. ◀

Indeed, it is no longer meaningful to argue the infimum of one quantity. Instead, we impose condition on $|\Pi_a|$ or $I_a^\mu(\Pi)$ then argue about the infimum of $I_b^\mu(\Pi)$. Then similar to distributional information complexity and prior-free information complexity as defined in [3], we can define an analogous notion for asymmetric setting conditioned on $|\Pi_a| \leq a$.

▶ **Definition 7** (Distributional Asymmetric Information Complexity). Distributional asymmetric information complexity for Bob of f under μ with error ε and subject to $|\Pi_a| < a$ is defined as

$$\mathsf{IC}_\mu^{<a}(f, \varepsilon) = \inf_\Pi I_b^\mu(\Pi)$$

where the infimum is taken over the set of protocols that achieve $\Pr_{(x,y)\sim\mu,\pi\sim\Pi}[\pi(x, y) \neq f(x, y)] < \varepsilon$.

▶ **Definition 8** (Prior Free Asymmetric Information Complexity)**.** Prior free asymmetric information complexity for Bob of f with error ε and $|\Pi_a| < a$ is defined as

$$\mathsf{IC}^{<a}(f,\varepsilon) = \inf_{\Pi} \max_{\mu} I_b^{\mu}(\Pi)$$

where the infimum is taken over the set of protocols that achieve $\Pr[\pi(x,y) \neq f(x,y)] < \varepsilon$ for all (x,y).

Indeed, one can also similarly define it with an upper bound on the information revealed by Alice say I_a^{μ}, rather than $|\Pi_a|$. But for our application (to data structure lower bound), the above definition suffices. Then using a standard Minimax argument, we can also that prior free asymmetric information complexity lower bounds distributional asymmetric information complexity up to some constant factor in the error parameter.

▶ **Theorem 9** (Theorem 2 rephrased)**.** *For any f, $\varepsilon \geq 0$, and $\alpha \in (0,1)$, there exists μ on (x,y) such that*

$$\mathsf{IC}^{<a}(f,\varepsilon/\alpha) \leq \frac{\mathsf{IC}_{\mu}^{<a}(f,\varepsilon)}{1-\alpha}$$

Proof. The proof follows from a standard minimax technique (by Theorem 3.5 of [3]) for translating prior free lower bound to distribution lower bound. We define the following two-player zero-sum game, where Player 1 comes up with a protocol Π for f conditioned on $|\Pi_a| < a$ (note that such set of protocols is closed under convex combination) and Player 2 comes up with a distribution μ on (x,y)'s with the following payoff :

$$P(\Pi,\mu) := (1-\alpha) \cdot \frac{I_b^{\mu}(\Pi)}{I} + \alpha \cdot \frac{\Pr_{(x,y)\sim\mu}[\pi(x,y) \neq f(x,y)]}{\varepsilon}$$

where $I := \max_{\mu} \mathsf{IC}_{\mu}^{<a}(f,\varepsilon)$. Then the rest of the argument follows from [3]. ◀

We remark that our hard distribution for distributional error guarantee therefore is not explicitly defined since the proof is non-constructive and follows from bounding $\mathsf{IC}^{<a}(f,\varepsilon/\alpha)$. Since $\mathsf{IC}^{<a}(f,\varepsilon/\alpha)$ is maximum over the distributions, it suffices to exhibit a lower bound on a particular (in our case a product distribution) distribution. But *this does not imply that μ is a product distribution as well*, since a convex combination of product distributions is not necessarily a product distribution as well.

3 Semi-Direct Sum Theorem

In this section, we prove semi-direct sum theorem for prior free information cost, that is where the protocol is guaranteed to be correct with good probability on all inputs when the underlying distribution is a product distribution.

Let $F(x,y)$ be of form $\bigvee_{i=1}^{n} f(x,y_i)$, that is OR of functions on singletons. Recall that in symmetric setting (refer to [4]) we prove direct sum by extracting a strategy for a single copy (x_i,y_i) from a protocol that solves for $(\vec{x},\vec{y})$. Similarly, we can prove semi-direct sum theorem for those set of functions, by extracting a strategy for $f(x,y_i)$ from a protocol for $F(x,\vec{y})$. More precisely, we prove the following theorem.

▶ **Theorem 10** (Semi Direct Sum Theorem)**.** *Consider a protocol Π for computing F with the following guarantee :*

- *For all $(x,\vec{y}) \in F^{-1}(1)$, $\Pr_{\pi}[\pi(x,\vec{y}) \neq 1] < \varepsilon_1$*

- *For all* $(x, \vec{y}) \in F^{-1}(0)$, $\Pr_\pi[\pi(x, \vec{y}) \neq 0] < \varepsilon_0$
- $|\Pi_a| \leq a$ *and* $I_b^\mu(\Pi) \leq b$

where μ *is a distribution such that* $Y_1, \ldots, Y_n$ *are i.i.d. with* X *distributed independently as well. Furthermore, suppose* $f(X, Y_i) = 1$ *with probability at most* $o(1/n)$ *under* μ*. Then there exists a protocol* Π' *(with input* X *and* Y_i*) for computing a singleton* f *with the following guarantee :*

- *For all* $(x, y_i) \in f^{-1}(1)$, $\Pr_{\pi'}[\pi'(x, y_i) \neq 1] < \varepsilon_1 + 0.01$
- *For all* $(x, y_i) \in f^{-1}(0)$, $\Pr_{\pi'}[\pi'(x, y_i) \neq 0] < \varepsilon_0 + 0.01$
- $|\Pi'_a| \leq a$ *and* $I_b^\mu(\Pi') \leq b/n$.

and $|\Pi'_a| \leq |\Pi_a|$*, that is the total number of bits sent by Alice.*

Before proving the theorem, we prove necessary facts in information theory.

▶ **Proposition 11.** *Suppose* $Y_1, \ldots Y_n$ *are all independent given* X*. Then*

$$I(\Pi; Y_i | X, Y_1, \ldots Y_{i-1}) \geq I(\Pi; Y_i | X)$$

Proof. Via our independence assumption, $I(Y_i; Y_1, \ldots, Y_{i-1}|X) = 0$. Then applying Fact 26, we get the desired inequality. ◀

▶ **Lemma 12** (Semi Direct Sum). *Suppose given* X, Y_i*'s are i.i.d. Then*

$$\frac{1}{n} \cdot I(\Pi; Y|X) \geq I(\Pi; Y_i|X)$$

Proof. First we show that $I(\Pi; Y|X) \geq \sum_i I(\Pi; Y_i|X)$. By Fact 27 we have

$$I(\Pi; Y_1, \ldots, Y_n|X) = I(\Pi; Y_1|X) + I(\Pi; Y_2|Y_1, X) + \ldots + I(\Pi; Y_n|Y_1, \ldots, Y_{n-1}, X)$$

Now for each term $I(\Pi; Y_i|Y_1, \ldots, Y_{i-1}, X)$, we can lower bound it with $I(\Pi; Y_i|X)$ from Proposition 11 due to our assumption on independence. Applying the lower bound term by term we have $I(\Pi; Y|X) \geq \sum_i I(\Pi; Y_i|X)$. By our assumption on the distribution, $I(\Pi; Y_i|X) = I(\Pi; Y_j|X)$ for all $i \neq j$. Thus we have the desired inequality. ◀

Proof of Theorem 10.

Protocol 1 Protocol Π'

1. Alice and Bob jointly and publicly samples $J \in [n]$.
2. Bob privately samples $Y_1 \ldots Y_{J-1}, Y_{J+1} \ldots Y_n$.
3. Alice and Bob set $X = x$ and $Y_J = y$ then run protocol Π on $(x, \vec{y})$

Consider Protocol 1. First observe that $F(X, Y) = f(X, Y_J)$ with high probability, since by our assumption on the density of $f^{-1}(0)$ under μ,

$$\Pr\left[\bigvee_{\substack{i=1\\ i\neq J}}^{n} f(X, Y_i) = 0\right] \leq (1 - o(1/n))^{n-1} = o(1).$$

Therefore, Π' satisfies the claimed guarantee if Π has the guarantee. Also $|\Pi'_a| = |\Pi_a|$ by design. The bound on $I_b^\mu(\Pi')$ follows from Proposition 11 and Lemma 12 via following observation

$$\begin{aligned} I(\Pi'; Y|X) &= I(\Pi; Y|X) \leq I(\Pi, J; Y|X) = I(J; Y|X) + I(\Pi; Y_J|J, X) \\ &= I(\Pi; Y_J|J, X) = \frac{1}{n} \cdot \sum_{i=1}^{n} I(\Pi; Y_i|X) \leq \frac{I(\Pi; Y_1 \ldots Y_n|X)}{n}. \end{aligned}$$

◀

As a contrapositive of Theorem 10, we obtain the following corollary which will be the main component in establishing asymmetric information lower bound.

▶ **Corollary 13.** *Suppose there exists a product distribution μ on X and Y with $\Pr_\mu[f(x,y) = 1] = o(1/n)$ such that for any protocol Π that computes f with $\Pr[f(x,y) \neq \pi(x,y)] < \varepsilon$ for all (x,y) with $|\Pi_a| < a$, $I_\mu(\Pi; Y|X) \geq b$. Then there exists a (product distribution) μ_n such that for any protocol Π^n that computes F with $\Pr[F(x,\vec{y}) \neq \pi_n(x,\vec{y})] < \varepsilon + 0.01$ for all $(x,\vec{y})$ with $|\Pi^n_a| < a$, $I_{\mu_n}(\Pi^n; \vec{Y}|X) \geq n \cdot b$. In other words, $\mathsf{IC}^{<a}(f,\varepsilon) \geq n \cdot b$.*

4 Nearest Neighbor in ℓ_∞ Lower Bound

In this section, we prove Theorem 3. To utilize Theorem 10, we focus on prior free information cost lower bound for the following function for $x, y \in \{0, \dots, M\}^d$.

$$f(x,y) := \begin{cases} 1 & \text{if } \|x-y\|_\infty \leq 1 \\ 0 & \text{if } \|x-y\|_\infty \geq c \end{cases}$$

We focus on lower bounding $\mathsf{IC}^{<a}(f,\varepsilon)$ for some sufficiently small constant ε. The main idea of the proof is that any protocol that achieves the desired error guarantee must be distinguishing between $f^{-1}(1)$ and $f^{-1}(0)$ for any given x. In other words, whether y is in the neighborhood of x or not.

To make the proof simpler, we prove the following Compression Lemma whose proof we attach in Section C which allows us to assume without loss of generality (with some minor costs) that the protocol is one round where Bob's reply is a single bit.

▶ **Lemma 14** (Compression for Bob). *Consider Π such that $I_{u^{\otimes 2}}(\Pi; X|Y) < I_a$ and $I_{u^{\otimes 2}}(\Pi; Y|X) < I_b$ that computes f. Further, assume that $I_a \cdot I_b = o(1)$. Then Π can be compressed to a one round protocol $\tau \sim \Pi'$ with $I_{u^{\otimes 2}}(\Pi'; X|Y) < O(I_a)$ and $I_{u^{\otimes 2}}(\Pi'; Y|X) < O\left(H(\sqrt{I_a I_b})\right)$ with the following guarantee:*

- *For $(x,y) \in f^{-1}(1)$, $\Pr[\tau(x,y) \neq 1] < \varepsilon_1 + 0.05$*
- *There exists $Z \subset f^{-1}(0)$, with $\mu(Z) = \mu(f^{-1}(0)) - 0.01$ such that for all $(x,y) \in S$, $\Pr[\tau(x,y) \neq 0] < \varepsilon_0 + 0.05$*

where $\varepsilon_0 = \max_{(x,y)\in f^{-1}(0)} \Pr[\pi(x,y) \neq f(x,y)]$ and similarly $\varepsilon_1 = \max_{(x,y)\in f^{-1}(1)} \Pr[\pi(x,y) \neq f(x,y)]$.

We also prove the following simple observation which allows us to assume without loss of generality that only Bob has access to private randomness for a single round protocol.

▶ **Lemma 15.** *Suppose there is a single round protocol Π where*

- *Alice and Bob both have access to both private and public randomness.*
- *Alice sends at most a-bits $|\Pi_a| \leq a$*
- *Bob sends at most b-bits of information $I(\Pi; Y|X) \leq b$.*

Then there is a protocol Π' where Alice does not have access to private randomness but $|\Pi'_a| \leq a$ and $I(\Pi'; Y|X) \leq b$

Proof. Consider the following simple modification to Π' from Π. Alice additionally samples from public randomness R^a_{pub} (thereby revealing the randomness of Bob) instead of using private randomness and follows Π. Bob ignores R^a_{pub} and follows Π. By design $|\Pi'_a| \leq a$. Let M_b denote the reply by Bob and M_a the message by Alice. Then

$$I(\Pi'; Y|X) = I(M_b; Y|X, R^a_{pub}, M_a) \leq I(M_b; Y|X, M_a) = I(\Pi; Y|X)$$

since $I(Y; R^a_{pub}|M_a, X) = 0$. ◀

Lemma 15 implies that it suffices to lower bound the case where Alice does not have access to private randomness. Then we set

$$u(x) := \begin{cases} 2^{-2\rho^x} & \text{if } x \in [M] \\ 1 - \sum_{i\in[M]} 2^{-2\rho^i} & \text{if } x = 0. \end{cases}$$

with $u(\vec{x})$ with $\vec{x}$ as a d-dimensional vector defined as a product of $u(x_i)$'s. Then we prove the main technical theorem whose proof is attached in Section D.

▶ **Theorem 16** (Single Function Lower Bound). *Let Π be a one round protocol where Alice does not have access to private randomness, Bob replies with one bit and computes f with the following guarantee*

- *For $(x, y) \in f^{-1}(1)$, $\Pr[\pi(x, y) \neq 1] < 0.1$*
- *There exists $S \subset f^{-1}(0)$, with $u(S) \geq u(f^{-1}(0)) - 0.01$ such that for all $(x, y) \in S$, $\Pr[\pi(x, y) \neq 0] < 0.1$*

For such Π, for any sufficiently small constant $\delta > 0$, if $I_a = I_{(x,y)\sim u^{\otimes 2}}(\Pi; X|Y) \leq \delta\rho \log n$, then $I_b = I_{(x,y)\sim u^{\otimes 2}}(\Pi; Y|X) \geq n^{-O(\delta)}$.

The main intuition of the proof follows from bounding ℓ_1-norm between transcripts from $(x, y) \in f^{-1}(1)$ and $(x, y) \in Z$. If the information sent by Bob is lower than the claimed bound, then the protocol cannot distinguish between $f^{-1}(1)$ and $f^{-1}(0)$ (more technically close in terms of ℓ_1-norm) which is a contradiction.

Then combining Theorem 16 with Lemma 14 and Lemma 15, we get the following prior free bound as a corollary conditioned on $|\Pi_a| \leq \delta\rho \log n$.

▶ **Corollary 17.** *Let Π be any protocol that computes f with the following guarantee*

- *For all $(x, y) \in f^{-1}(1)$, $\Pr[\pi(x, y) \neq 1] < 0.05$*
- *For all $(x, y) \in f^{-1}(0)$, $\Pr[\pi(x, y) \neq 0] < 0.05$*

For such Π, for any sufficiently small constant $\delta > 0$, if $I_a = I_{(x,y)\sim u\times u}(\Pi; X|Y) \leq \delta\rho \log n$, then $I_b = I_{(x,y)\sim u\times u}(\Pi; Y|X) \geq n^{-O(\delta)}$.

Proof. Suppose we have a protocol that computes f with the guarantee, with $I_a \leq \delta\rho \log n$ and $I_b = I(\Pi; Y|X) < n^{-\omega(\delta)}$. Then by Lemma 14, we can compress the protocol to a one-round protocol with 1-bit response from Bob with information cost $O(\delta\rho \log n)$ and $n^{-\omega(\delta)}$ respectively for Alice and Bob, with the following error guarantee.

- For all $(x, y) \in f^{-1}(1)$, $\Pr[\tau(x, y) \neq 1] < 0.1$
- There exists $S \subset f^{-1}(0)$, with $u(S) \geq u(f^{-1}(0)) - 0.01$ such that for all $(x, y) \in S$, $\Pr[\tau(x, y) \neq 0] < 0.1$

This is indeed a contradiction to Theorem 16. ◀

Then finally combining Corollary 17 with Corollary 13, and with a guarantee that $f(x, y) = 1$ with probability $o(1/n)$, the proof of which we append in Claim 40, we get a lower bound for any protocol that computes $F(x, \vec{y})$ with prior free error guarantee, when the information cost is measured over the distribution where all X and Y's are distributed according to u.

▶ **Theorem 18.** *Let Π be any protocol that computes F with the following guarantee*

- *For all $(x, \vec{y}) \in F^{-1}(1)$, $\Pr[\pi(x, y) \neq 1] < 0.04$*
- *For all $(x, \vec{y}) \in F^{-1}(0)$, $\Pr[\pi(x, y) \neq 0] < 0.04$*

For such Π, for any sufficiently small constant $\delta > 0$, if $|\Pi_a| \leq \delta\rho \log n$, then $I_b = I(\Pi; Y|X) \geq n^{1-O(\delta)}$.

Via the minimax argument, we get the following for distributional error setting (from applying Theorem 2)

▶ **Corollary 19.** *For any sufficiently small constant $\delta > 0$, there exists μ such that for any protocol Π that computes F with $\Pr_{(x,\vec{y})\sim\mu}[\pi(x,\vec{y}) \neq F(x,\vec{y})] < 0.01$, if $|\Pi_a| \leq \delta\rho\log n$, then $I_b = I(\Pi; Y|X) \geq n^{1-O(\delta)}$.*

This yields the desired data structure lower bounds from the translations in Section B. For the definition of data structures, we refer the reader to Section B as well.

▶ **Corollary 20** (Cell-Probe Lower Bound)**.** *Consider any randomized cell-probe data structure solving d-dimensional near-neighbor search under ℓ_∞ with approximation factor $c = O(\log_\rho \log d)$, with the guarantee that if there exists $y_i \in \vec{y}$ that is close to x, the querier outputs 1 with > 0.95 probability. If the word size is $w = n^{1-\delta}$ for some $\delta > 0$, the data structure requires space $n^{\Omega(\rho/t)}$ for query time t.*

▶ **Corollary 21** (Decision Tree Lower Bound)**.** *Let $\delta > 0$ be arbitrary constant. A decision tree of depth $r = n^{1-2\delta}$ and node size $w = n^\delta$ that solves d-dimensional near-neighbor search under ℓ_∞ with approximation $c = O(\log_\rho \log d)$, must have size $s = n^{\Omega(\rho)}$.*

References

1 Alexandr Andoni, Dorian Croitoru, and Mihai Patrascu. Hardness of nearest neighbor under l-infinity. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 424–433. IEEE, 2008.

2 Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

3 Mark Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015. doi:10.1137/130938517.

4 Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 151–160. ACM, 2013.

5 Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, 2014.

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

7 Thomas M Cover and Joy A Thomas. *Elements of information theory.* John Wiley &amp; Sons, 2012.

8 Anirban Dasgupta, Ravi Kumar, and D. Sivakumar. Sparse and lopsided set disjointness via information theory. In Anupam Gupta, Klaus Jansen, José Rolim, and Rocco Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 517–528, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

9 Piotr Indyk. On approximate nearest neighbors under l∞ norm. *Journal of Computer and System Sciences*, 63(4):627–638, 2001.

10 Michael Kapralov and Rina Panigrahy. Nns lower bounds via metric expansion for l∞ and emd. In *International Colloquium on Automata, Languages, and Programming*, pages 545–556. Springer, 2012.

11 Peter Bro Miltersen. Lower bounds for union-split-find related problems on random access machines. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 625–634. ACM, 1994.

12 Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. In *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pages 103–111. ACM, 1995.

13 Rina Panigrahy, Kunal Talwar, and Udi Wieder. Lower bounds on near neighbor search via metric expansion. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 805–814. IEEE, 2010.

14 Mihai Patrascu. *Lower Bound Techniques for Data Structures*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2008. AAI0821553.

15 Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. How to compress asymmetric communication. In *Proceedings of the 30th Conference on Computational Complexity*, pages 102–123. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015.

A Information Theory

In this section, we provide necessary backgrounds on information theory and information complexity that are used in this paper. For further reference, we refer the reader to [7] and [6, 3, 4, 2, 5].

▶ **Definition 22** (Entropy). The entropy of a random variable X is defined as

$$H(X) := \sum_x \Pr[X = x] \log \frac{1}{\Pr[X = x]}.$$

Similarly, the conditional entropy is defined as

$$H(X|Y) := \mathop{\mathbb{E}}_{Y}\left[\sum_x \Pr[X = x|Y = y] \log \frac{1}{\Pr[X = x|Y = y]}\right].$$

As an abuse of notation, we also denote binary entropy function $H : [0, 1] \to [0, 1]$ as

$$H(p) := p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}$$

▶ **Definition 23** (Mutual Information). Mutual information between X and Y (conditioned on Z) is defined as

$$I(X; Y|Z) := H(X|Z) - H(X|Y Z).$$

▶ **Definition 24** (KL-Divergence). KL-Divergence between two distributions μ and ν is defined as

$$D(\mu||\nu) := \sum_x \mu(x) \log \frac{\mu(x)}{\nu(x)}.$$

In order to bound mutual information, it suffices to bound KL-divergence, due to following fact.

▶ **Fact 25** (KL-Divergence and Mutual Information). *The following equality between mutual information and KL-Divergence holds*

$$I(A; B|C) = \mathop{\mathbb{E}}_{B,C}\left[D(A|_{B=b,C=c}||A|_{C=c})\right].$$

▶ **Fact 26.** *Let A, B, C, D be random variables such that $I(A; D|C) = 0$. Then*

$$I(A; B|C) \leq I(A; B|C, D).$$

▶ **Fact 27** (Mutual Information Chain Rule).

$$I(A; B_1, \dots B_n | C) = \sum_{i=1}^{n} I(A; B_i | C, B_1, \dots B_{i-1})$$

▶ **Fact 28.** *Let $p(x,y)$ and $q(x,y)$ be joint distributions of two random variables X and Y. Then*

$$D(p(x,y) || q(x,y)) = D(p(x) || q(x)) + \mathop{\mathbb{E}}_{x \sim p} D(p(y|x) || q(y|x))$$

To provide context to asymmetric information costs that will be introduced in the next section, we first introduce usual definition of information cost of f.

▶ **Definition 29** (Information Cost). The information cost of Protocol Π with input X for Alice Y for Bob under distribution μ is defined as

$$\mathsf{IC}_\mu(\Pi) = I_\mu(\Pi; X|Y) + I_\mu(\Pi; Y|X)$$

where $I_\mu(\Pi; X|Y)$ denote mutual information where X and Y are distributed according to μ, similarly for $I_\mu(\Pi; Y|X)$.

Intuitively, $I(\Pi; X|Y)$-term captures how much information Alice must inject to the transcript to compute F. Similarly $I(\Pi; Y|X)$-term captures how much information Bob must inject to the transcript. This intuition comes useful in defining analogous quantities for the asymmetric communication setting.

B Data Structure and Asymmetric Communication Lower Bound

In this section, we introduce relevant models in data structure for which an asymmetric communication lower bound translates to their lower bounds.

Cell Probe Data Structure

Cell probe data structure, which is similar to Random Access machine, can be formally defined as following.

▶ **Definition 30** (Cell-Probe). Cell Probe data structure is defined as the following model for computing F. Alice (user) sends a cell-address in each round. Bob (database) answers with the contents in the queried cell. Word size w is the maximum number of bits in the cell.

The following classic “translation” theorem holds.

▶ **Theorem 31** (Lemma 1 of [12]). *If there exists a (randomized) cell-probe database that solves F with cell size s, word size w and query time t, then there exists $(2t \log s, 2tw)$-(randomized) communication protocol for F.*

In other words, asymmetric communication lower bounds translates to cell-probe database lower bounds. This is the key observation in [1].

Decision Tree Data Structure

Decision tree data structure model is formally defined as following.

▶ **Definition 32** (Decision Tree). Depending on $\vec{y}$ and the randomness, Bob constructs a decision tree $T_{\vec{y}}$, a complete binary tree, in which:

- Each node v contains a predicate function $f_v : \mathcal{X} \to \{0,1\}$ where $f_v \in \mathcal{F}$, the set of allowed predicates.
- Each edge is labeled 0 or 1, which denotes the answer to the parent's f_v.
- Each leaf is labeled 0 or 1, which denotes the final output to F.

Size s of the tree is the number of nodes. Depth r is the depth of the tree. Predicate size w is $\log_2 |\mathcal{F}|$.

Under this model, a more efficient translation theorem holds which results in better lower bound under decision tree model.

▶ **Theorem 33** ([1]). *If there exists a (randomized) decision tree that solves F with size s, depth d, and node size w, then there exists $(O(\log s), O(dw \log s))$-(randomized) communication protocol for F.*

As a corollary of Lemma 14 and Theorem 3, we get the following corollaries in data structure lower bound. Combining Theorem 31 with Theorem 18 (or Corollary 19 depending on the error guarantee), we get the following corollary.

▶ **Corollary 20.** *Consider any randomized cell-probe data structure solving d-dimensional near-neighbor search under ℓ_∞ with approximation factor $c = O(\log_\rho \log d)$, with the guarantee that if there exists $y_i \in \vec{y}$ that is close to x, the querier outputs 1 with > 0.95 probability. If the word size is $w = n^{1-\delta}$ for some $\delta > 0$, the data structure requires space $n^{\Omega(\rho/t)}$ for query time t.*

Combining Theorem 33 with Theorem 18 (or Corollary 19 depending on the error guarantee), we get the following corollary.

▶ **Corollary 21.** *Let $\delta > 0$ be arbitrary constant. A decision tree of depth $r = n^{1-2\delta}$ and node size $w = n^{\delta}$ that solves d-dimensional near-neighbor search under ℓ_∞ with approximation $c = O(\log_\rho \log d)$, must have size $s = n^{\Omega(\rho)}$.*

C Proof of Compression Lemma

In this section, we show that any protocol such that Alice sends at most I_a bits, and Bob sends at most I_b bits of information with $I_b \ll 1$ and $I_a \cdot I_b = o(1)$ can be compressed to a one-round protocol such that Alice sends at most $O(I_a)$ bits and Bob sends at most 1 bits and $H(\sqrt{I_a I_b})$-bits of information.

First we use the following compression theorem to compress the size of the transcript where τ is the resulting compressed protocol:

▶ **Theorem 34** (Theorem 1.2 in [15]). *Suppose $I_a = \omega(1)$. Then any protocol Π such that Alice sends at most I_a-bits of information and Bob sends at most I_b-bits of information can be simulated with $O(I_a 2^{O(I_b)})$ bits of communication with the following guarantee :*

- *For $(x,y) \in F^{-1}(1)$, $\Pr[\pi(x,y) \neq \tau(x,y)] < \varepsilon_1 + 0.01$*
- *There exists $\Gamma \subset F^{-1}(0)$, with $\mu(\Gamma) = \mu(F^{-1}(0)) - 0.01$ such that for all $(x,y) \in \Gamma$, $\Pr[\pi(x,y) \neq \tau(x,y)] < \varepsilon_0 + 0.01$*

where ε_1 and ε_0 refers to the respective guarantee for the original protocol Π.

▶ Remark. Theorem 1.2 in [15] is not stated as the above form. But setting the output as 1 on high divergence (x, y) pairs and adjusting the constants give the above guarantee. Also note that $2^{O(I_b)} \leq 2$ in our setting where $I_b \ll 1$.

Theorem 34 allows us to assume the length of the protocol $|\Pi| = |\Pi_a| + |\Pi_b| \leq O(I_a)$ for our setting. Now we introduce the following round compression protocol and its guarantee as Lemma 14.

Protocol 2 Compression Protocol

- Alice samples $\pi \sim \Pi_x$, the distribution of protocol conditioned on Alice's input x.
- Alice sends π to Bob
- Bob answers 1 if π agrees with his input y and private randomness r_b. Rejects otherwise.

▶ **Lemma 14.** *Consider* Π *such that* $I(\Pi; X|Y) < I_a$ *and* $I(\Pi; Y|X) < I_b$ *that computes* f *under the distribution* μ*. Further, assume that* $I_a \cdot I_b = o(1)$*. Then* Π *can be compressed to a one round protocol* $\tau \sim \Pi'$ *with* $I(\Pi'; X|Y) < O(I_a)$ *and* $I(\Pi'; Y|X) < O\left(H(\sqrt{I_a I_b})\right)$ *with the following guarantee:*

- *For* $(x, y) \in f^{-1}(1)$, $\Pr[\tau(x, y) \neq 1] < \varepsilon_1 + 0.05$
- *There exists* $Z \subset f^{-1}(0)$*, with* $\mu(Z) = \mu(f^{-1}(0)) - 0.01$ *such that for all* $(x, y) \in S$, $\Pr[\tau(x, y) \neq 0] < \varepsilon_0 + 0.05$

where $\varepsilon_0 = \max_{(x,y)\in f^{-1}(0)} \Pr[\pi(x, y) \neq f(x, y)]$ *and similarly* $\varepsilon_1 = \max_{(x,y)\in f^{-1}(1)} \Pr[\pi(x, y) \neq f(x, y)]$.

Towards proving Lemma 14, we first prove two facts on non-negative numbers.

▶ **Claim 35.** *Let* $\Pi_{x,y}$ *denote the distribution of transcript conditioned on* $X = x$ *and* $Y = y$*. Similarly let* Π_x *denote the distribution conditioned on* $X = x$*. Similarly let* $M_i|_{M_{<i},x,y}$ *and* $M_i|_{M_{<i},x}$ *denote the distribution of message at* i*-th round conditioned on all previous messages and* x, y *or just* x*. Then*

$$D(\Pi_{x,y}||\Pi_x) = \sum_{i=1}^{R} \mathop{\mathbb{E}}_{M_{<i}|x,y} D(M_i|_{M_{<i},x,y}||M_i|_{M_{<i},x})$$

Proof. Follows from Fact 28. ◀

▶ **Claim 36.** *Consider* R *non-negative numbers* $I_1, \ldots I_R$*. Then*

$$\sum_{r=1}^{R} \sqrt{I_r} \leq \sqrt{R \cdot \sum_{r=1}^{R} I_r}$$

Proof. Follows from Cauchy-Schwarz. ◀

Proof of Lemma 14. Note that without loss of generality, the total amount of bits communicated in Protocol 2 is $|\Pi| + 1 = O(I_a)$ since one can compress the transcript using Theorem 34. This immediately implies that $I(\Pi'; X|Y) < O(I_a)$, since the number of bits sent by Alice is at most $O(I_a)$. Similarly, it also implies that the number of round (R) is at most $O(I_a)$.

Now, to prove the full lemma, it suffices to bound the probability of Bob sending 0. We bound by computing the probability of Bob rejecting the transcript at each round $i + 1$, (i.e. the odd round messages) then summing up these probabilities.

Let M_i denote the message sent in i-th round of the protocol. If $i+1$ is odd (the round where Bob sends message), then we have

$$M_{i+1}|M_{\le i}, X, Y = M_{i+1}|M_{\le i}, Y$$

since it is a protocol and Bob's response only depends on Y and $M_{\le i}$. Then by Pinsker's inequality, we bound the error of sampling. The probability of making error at $i+1$-th round can be bounded using Pinsker's inequality as

$$\begin{aligned} &\mathop{\mathbb{E}}_{M_{\le i},X,Y}[\|M_{i+1}|M_{\le i},X,Y - M_{i+1}|M_{\le i},X\|_1] \\ &\le O\left(\mathop{\mathbb{E}}_{M_{\le i},X,Y}\left[\sqrt{D(M_{i+1}|M_{\le i},X,Y||M_{i+1}|M_{\le i},X)}\right]\right) \\ &\le O\left(\sqrt{\mathop{\mathbb{E}}_{M_{\le i},X,Y}[D(M_{i+1}|M_{\le i},X,Y||M_{i+1}|M_{\le i},X)]}\right) \qquad (1) \end{aligned}$$

where the last inequality is using the concavity of $\sqrt{x}$. Then we can sum and bound the probability of error as

$$\begin{aligned} &\sum_{\substack{i=0\\ i:odd}}^{R-1}\mathop{\mathbb{E}}_{M_{\le i},X,Y}[\|M_{i+1}|M_{\le i},X,Y - M_{i+1}|M_{\le i},X\|_1] \\ &\le O\left(\sum_{\substack{i=0\\ i:odd}}^{R-1}\sqrt{\mathop{\mathbb{E}}_{M_{\le i},X,Y}[D(M_{i+1}|M_{\le i},X,Y||M_{i+1}|M_{\le i},X)]}\right) \\ &\le O\left(\sqrt{R\cdot\sum_{\substack{i=0\\ i:odd}}^{R-1}\mathop{\mathbb{E}}_{M_{\le i},X,Y}[D(M_{i+1}|M_{\le i},X,Y||M_{i+1}|M_{\le i},X)]}\right) \\ &\le O\left(\sqrt{R\cdot\mathop{\mathbb{E}}_{X,Y}[D(\Pi_{X,Y}||\Pi_X)]}\right) = O\left(\sqrt{R\cdot I_b}\right) \qquad (2) \end{aligned}$$

where the second inequality follows from Claim 36 and the third inequality follows from Claim 35. Now $R < O(I_a)$ from the total number of bits communicated by Alice, which upper bounds the number of rounds, is $O(I_a)$ from Theorem 34. This ℓ_1-norm bound implies that Bob answers 0 with at most $O(\sqrt{I_aI_b})$-probability in expectation over (x,y). Then the amount of information that Alice learns is at most $O\left(H(\sqrt{I_aI_b})\right)$.

It remains to show that the error rate guarantee is preserved. We divide into two cases.

- If $f(x,y)=1$, by design there is no guarantee lost on $(x,y)\in f^{-1}(1)$, except the loss from Theorem 34.
- Now suppose $f(x,y)=0$. Recall that the compression fails with probability at most $O(\sqrt{I_aI_b})$ in expectation. Let $Z^0 := \{(x,y)|\|\Pi|X=x,Y=y-\Pi|X=x\|_1 < (I_aI_b)^{2/3}\}$. By Markov's inequality, $\mu(Z^0)\ge 1-o(1)$. Set $Z = Z^0\cap\Gamma\cap f^{-1}(0)$, where Γ is the set from Theorem 34. It is easy to check that the probability of error only increases by $o(1)$ for any $(x,y)\in Z$ since $(I_aI_b)^{2/3}=o(1)$ by assumption while being in Γ only increases the error rate by 0.01. Thus Z indeed satisfies the guarantee. ◀

D Omitted Proof from Section 4

Recall that

$$u(x) := \begin{cases} 2^{-2\rho^x} & \text{if } x \in [M] \\ 1 - \sum_{i \in [M]} 2^{-2\rho^i} & \text{if } x = 0. \end{cases}$$

with $u(x)$ defined as the product of $u(x_i)$'s. Under such u, we make use of the following lemma from [1].

▶ **Lemma 37** (Isoperimetric Inequality (Lemma 9 of [1])). *Consider any set $S \subseteq \{0, \dots M\}^d$. Let $N(S)$ be the set of points at distance at most 1 from S under ℓ_∞. Then $u(N(S)) \geq u(S)^{1/\rho}$.*

Then we get the following corollary in terms of KL-Divergence.

▶ **Corollary 38.** *Let $u|_S$ denote u restricted on a subset $S \subseteq \{0, \dots M\}^d$. Then for any S*

$$u(N(S)) \geq 2^{-D(u|_S || u)/\rho}$$

Proof. Observe that $D(u|_S||u) = -\log u(S)$ or $u(S) = 2^{-D(u|_S||u)}$. Then Lemma 37 implies the desired inequality. ◀

We also state the following simple fact about norm.

▶ **Fact 39** (ℓ_1-norm convex). *Consider a family of distributions μ and $\{\nu_i\}_i$. Then*

$$\|\mu - \mathop{\mathbb{E}}_i[\nu_i]\|_1 \leq \mathop{\mathbb{E}}_i [\|\mu - \nu_i\|_1]$$

Proof. Follows from the convexity of ℓ_1-norm. ◀

Now we are ready to prove Theorem 16.

Proof of Theorem 16. Let $a := \delta\rho \log n$ and $b := 2^{-2000a/\rho}$. Recall that we assume that the protocol is one round, therefore the distribution on the protocol Π is on Alice's message π_a and Bob's message π_b. Also let $\Pi_b|\pi_a$ be the distribution on Bob's message given Alice's message as π_a. If $\pi_b \sim \Pi_b|\pi_a$, note that each Bob's message π_b conditioned on π_a induces a prior on Bob's input y, μ_{π_b}. Suppose further $\mathbb{E}_{\pi_b \sim \Pi_b|\pi_a} D(\mu_{\pi_b}||u) < 10b$. Note that such message π_a must exist by Markov argument.

Since Alice does not have access to private randomness, prior on X conditioned on π_a is exactly a subset on $\{0, \dots M\}^d$, which we denote as S_{π_a}.

With π_a fixed, with an abuse of notation let $\Pi_{x,y}$ denote the distribution on the transcript (the remaining message, Π_b) conditioned on input being (x, y) and Alice's message. And let μ denote the distribution on x conditioned on the message π_a, that is $u|_{S_{\pi_a}}$. Note that Π_{x,y_1} and Π_{x,y_2} are close in ℓ_1 norm if $f(x, y_1) = f(x, y_2) = 1$ or $(x, y_1), (x, y_2) \in Z$, since the protocol is a one round protocol and Alice's message is fixed. In particular,

$$\|\Pi_{x,y_1} - \Pi_{x,y_2}\|_1 < 0.2$$

since both Π_{x,y_1} and Π_{x,y_2} outputs $f(x, y_1) = f(x, y_2)$ with at least 0.9 probability for $f^{-1}(1)$ and Z. Similarly, if $(x, y_1) \in Z$ and $f(x, y_2) = 1$, $\|\Pi_{x,y_1} - \Pi_{x,y_2}\|_1 > 1.8$. Now consider

$$\nu_{f^{-1}(0)} := \mathop{\mathbb{E}}_{x \sim \mu} \mathop{\mathbb{E}}_{\substack{y \sim u \\ (x,y) \in Z}} [\Pi_{x,y}]$$

$$\nu_{f^{-1}(1)} := \mathop{\mathbb{E}}_{x \sim \mu} \mathop{\mathbb{E}}_{\substack{y \sim u \\ f(x,y) = 1}} [\Pi_{x,y}]$$

First observe that from Fact 39 if $(x, y) \in Z$,

$$\|\Pi_{x,y} - \nu_{f^{-1}(0)}\|_1 \leq \mathop{\mathbb{E}}_{x\sim\mu} \mathop{\mathbb{E}}_{\substack{y'\sim u \\ (x,y')\in Z}} [\|\Pi_{x,y} - \Pi_{x,y'}\|_1] < 0.2 \tag{3}$$

Similarly, if $f(x, y) = 1$,

$$\|\Pi_{x,y} - \nu_{f^{-1}(1)}\|_1 \leq \mathop{\mathbb{E}}_{x\sim\mu} \mathop{\mathbb{E}}_{\substack{y'\sim u \\ f(x,y')=1}} [\|\Pi_{x,y} - \Pi_{x,y'}\|_1] < 0.2. \tag{4}$$

Then by triangular inequality if $(x, y_1) \in Z$ and $f(x, y_2) = 1$,

$$1.8 < \|\Pi_{x,y_1} - \Pi_{x,y_2}\|_1 \leq \|\Pi_{x,y_1} - \nu_{f^{-1}(0)}\|_1 + \|\nu_{f^{-1}(0)} - \nu_{f^{-1}(1)}\|_1 + \|\Pi_{x,y_2} - \nu_{f^{-1}(1)}\|_1 \tag{5}$$

Combining (3), (4) and (5) we get

$$\|\nu_{f^{-1}(0)} - \nu_{f^{-1}(1)}\|_1 > 1.4.$$

Now to derive the contradiction, we define ν and ν_0 as following.

$$\nu := \Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in Z] \cdot \nu_{f^{-1}(0)} + \Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in f^{-1}(1)] \cdot \nu_{f^{-1}(1)}$$
$$+ \Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \notin f^{-1}(1),\ (x, y) \notin Z] \cdot \nu_{f^{-1}(*)}$$
$$\nu_0 := \nu_{f^{-1}(0)}$$

where $\nu_{f^{-1}(*)}$ denotes expected prior of transcripts from inputs that are neither in Z nor $f^{-1}(1)$. Then we have

$$\|\nu_0 - \nu\|_1 \geq \|\Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in f^{-1}(1)] \cdot \left(\nu_{f^{-1}(0)} - \nu_{f^{-1}(1)}\right)\|_1$$
$$= \Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in f^{-1}(1)] \cdot \|\nu_{f^{-1}(0)} - \nu_{f^{-1}(1)}\|_1 \geq \Omega\left(\Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in f^{-1}(1)]\right)$$

since the difference is minimized when $\nu_{f^{-1}(*)} = \nu_{f^{-1}(0)}$. This via Pinsker's inequality implies

$$D(\nu_0||\nu) \geq \Omega\left(\Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in f^{-1}(1)]^2\right).$$

while by our assumption, it must be the case $D(\nu_0||\nu) < 20b$ since $\Pr[(x, y) \in Z] > 1/2$.

To derive a contradiction, now we lower bound $\Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in f^{-1}(1)]$. By Corollary 38,

$$\Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in f^{-1}(1)]^2 = u(N(\mathsf{Supp}(\mu)))^2 \geq 2^{-2D(\mu||u)/\rho}$$

since $\mu = u|_{S_{\pi_a}}$. This implies that $\Pr_{\substack{x\sim\mu \\ y\sim u}}[(x, y) \in f^{-1}(1)]^2 \geq 2^{-2a/\rho}$ which is in contradiction to our setting of b. ◀

▶ **Claim 40** (Density). *Set $\rho = (\varepsilon \log d)^{1/c}$, and $d \geq (2 \cdot \log n)^{\frac{1}{1-\varepsilon}}$. Then*

$$\Pr_{(x,y)\sim u\times u}[f(x, y) = 0] \geq 1 - o(1/n)$$

Proof. We prove by bounding $\Pr_{(x,y)\sim u\times u}[\|x - y\|_\infty \leq c]$.

$$\Pr_{(x,y)\sim u\times u}[\|x - y\|_\infty < c] \leq \Pr_{x\sim u}[\|x\|_\infty < c] < \left(1 - 2^{-\rho^c}\right)^d = \left(1 - \frac{1}{d^\varepsilon}\right)^d \leq 2^{-d^{1-\varepsilon}} = o(1/n).$$

where the first inequality holds since $\max_{x\in\mathcal{X}} u(x) = u((0, \ldots, 0))$ by the property of u. ◀