Space lower bounds for distance approximation in the data stream model

Michael Saks*
Department of Mathematics
Rutgers University
New Brunswick, NJ
saks@math.rutgers.edu

Xiaodong Sun*
Department of Mathematics
Rutgers University
New Brunswick, NJ
sunxd@math.rutgers.edu

ABSTRACT

We consider the problem of approximating the distance of two d-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$ in the data stream model. In this model, the $2d$ coordinates are presented as a "stream" of data in some arbitrary order, where each data item includes the index and value of some coordinate and a bit that identifies the vector ($\mathbf{x}$ or $\mathbf{y}$) to which it belongs. The goal is to minimize the amount of memory needed to approximate the distance. For the case of L^p-distance with $p \in [1, 2]$, there are good approximation algorithms that run in polylogarithmic space in d (here we assume that each coordinate is an integer with $O(\log d)$ bits). Here we prove that they do not exist for $p > 2$. In particular, we prove an optimal approximation-space tradeoff of approximating L^∞ distance of two vectors. We show that any randomized algorithm that approximates L^∞ distance of two length d vectors within factor of d^δ requires $\Omega(d^{1-4\delta})$ space. As a consequence we show that for $p > 2/(1-4\delta)$, any randomized algorithm that approximate L^p distance of two length d vectors within a factor d^δ requires $\Omega(d^{1-\frac{2}{p}-4\delta})$ space.

The lower bound follows from a lower bound on the two-party one-round communication complexity of this problem. This lower bound is proved using a combination of information theory and Fourier analysis.

1. INTRODUCTION

Many applications in science and commerce require the processing of *massive data sets*, sets whose size alone imposes significant limitations on the way the data can be stored and manipulated. The need to process such sets effectively gives rise to a variety of fundamental problems, and several related theoretical models have been proposed to capture these problems. Two of these models are the *data stream model* and the *sketch model*.

In the *data stream model* we are trying to compute some function $f(x_1, \dots, x_m)$. In this case, the data is the set of m pairs, (i, x_i) and the data arrives in some arbitrary order. We assume that m is much larger than the memory available, so we can not store all of the data as it arrives. The problem is to minimize the amount of space needed to compute f.

In the *sketch model* we are trying to compute some function $f(\mathbf{x}, \mathbf{y})$ of two vectors $\mathbf{x}, \mathbf{y}$ stored at different sites. The vectors are so long that it would be expensive to transmit the whole vector. The problem is to find a sketch function g and another function h s.t. $h(g(\mathbf{x}), g(\mathbf{y}))$ is a good approximation of $f(\mathbf{x}, \mathbf{y})$ and the size of the sketches $g(\mathbf{x})$, $g(\mathbf{y})$ is significantly smaller.

Recently, various researchers have considered the problem of estimating the distance between two vectors in these models, where the distance measure is the L^p-distance for some $p \geq 1$, i.e., $\rho_p(\mathbf{x}, \mathbf{y}) = \left(\sum_i (x_i - y_i)^p\right)^{\frac{1}{p}}$. Results of Alon, Matias and Szegedy [1], Feigenbaum, Kannan, Strauss, Viswanathan [6], Fong, Strauss [7] and Indyk [8] show that for $p \in [1, 2]$, there are algorithms (in both the data stream and sketch models) which give approximation factor arbitrarily close to 1 that run in space polylogarithmic in d. To our knowledge, before the present paper nothing was known for this problem for $p > 2$.

As observed by Alon, Matias and Szegedy [1], for any function $f(\mathbf{x}, \mathbf{y})$ of two vectors, any protocol for f in either the data stream or sketch models using space at most S gives rise to a one round communication protocol using at most S bits of communication.

Our results

We prove that any randomized one round communication protocol that approximates L^∞ distance of two length d vectors within factor of d^δ requires $\Omega(d^{1-4\delta})$ communication. As a consequence, we get that any randomized one round communication protocol that approximates L^p distance for $p > 2/(1-4\delta)$ within d^δ requires $\Omega(d^{1-\frac{2}{p}-4\delta})$ communication. By the above observation of Alon, et al., these communication bounds translate into space bounds in the data stream and sketch models. For $p = \infty$, this tradeoff is essentially optimal, i.e., one can get a d^δ approximation with communication $\tilde{\Omega}|(d^{1-\frac{2}{p}-4\delta})$ communication. To do this, divide $\mathbf{x}$ and $\mathbf{y}$ into $t = d^{1-4\delta}$ vectors $\mathbf{x}^j, \mathbf{y}^j$ $(1 \leq j \leq t)$ of length $d^{4\delta}$ and use the L_2 algorithm to estimate the L_2 distance between each $\mathbf{x}^j, \mathbf{y}^j$. Since $a_j = d^\delta \rho_2(\mathbf{x}^j, \mathbf{y}^j)$ is within factor of d^δ of $\rho_\infty(\mathbf{x}^j, \mathbf{y}^j)$, we can take our approximation to be $\max_j a_j$.

*Research supported by NSF grant CCR-9700239, CCR-9988526 and by DIMACS

STOC'02, May 19-21, 2002, Montreal, Quebec, Canada.
Copyright 2002 ACM 1-58113-495-9/02/0005 ...$5.00.

Our proof proceeds as follows. We do a few transformations of the problem in order to get it in a more convenient form. First, we consider a decision version of the problem where the problem is to distinguish between instances where the L^∞ distance is less than d or greater than $d^{1+2\delta}$ (and we don't care about instances whose distance is in between). A communication lower bound on this decision problem carries over to the approximation problem. We consider this problem in the distributional model: we select a probability distribution over inputs and prove a communication lower bound on any deterministic algorithm that solves the problem on most instances. By Yao's lemma, this implies the same lower bound on randomized complexity. We also transform the domain of the problem from $\mathbb{Z}^d$ to the d-dimensional torus $\mathbb{Z}_n^d$ (for some appropriate n).

Observe that the (partial) decision function we are investigating can be written in the form $F = \vee_{i=1}^d g_i$, where g_i is the corresponding one-dimensional function on coordinate i. We give a new approach to proving distributional communication lower bounds for (partial) functions of this form. We select a distribution μ on $(\mathbf{x}, \mathbf{y})$ by selecting a distribution ν on pairs of integers (x, y) and taking μ to be the product of d copies of ν. We show that for this distribution, if Π is a communication protocol that computes F with small error, then for most i, Π also computes g_i in the following relaxed sense: when $g_i = 1$ the protocol makes very small error, and when $g_i = 0$, the protocol is correct with non-negligible fraction of the time. On the other hand, we show that if the total communication is small, then for most i the amount of information transmitted about the pair (x_i, y_i) is so small that even such a relaxed requirement can not be met. Since coordinates are chosen independently, this statement has a strong intuitive appeal. However, this intuition is misleading. Indeed, there is a subtle but significant difficulty introduced by the (unavoidable) fact that for each i, x_i and y_i are not chosen independently. To overcome this difficulty requires a rather involved argument combining information theory and Fourier analysis. At this point, our proof only works for the case of one round communication protocols, which is enough for the data stream model lower bounds. However, our approach is in principle applicable to general communication lower bounds.

Related work

For the frequency moment problem in the data stream model, Alon, Matias and Szegedy [1] obtained results similar to ours. The frequency moment problem is essentially equivalent to the following problem: given k integer vectors $\mathbf{y}^1$, $\ldots$, $\mathbf{y}^k$ each of length d, estimate the L^p norm of their sum. Here the coordinates of the vectors are assumed to be of size at most polynomial in d. Although no explicit approximation-space tradeoff was given in [1], analyzing the argument in the paper gives that any algorithm that estimates L^∞ norm of vector sum within a factor better than d^δ requires space $\Omega(d^{1-10\delta})$. Their results are proved by reducing the problem to a k-party communication problem.

While the form of their bound is similar to ours, the results are incomparable. Their bounds hold even in the case that the vectors are restricted to be nonnegative. Note that in this special case, there is an efficient $\sqrt{k}$ approximation algorithm, since the maximum entry of all of the vectors multiplied by $\sqrt{k}$ is within a $\sqrt{k}$ factor of the maximum entry of the sum, In particular when $k = 2$ there is a $\sqrt{2}$ factor approximation requiring only logarithmic space.

In this framework, our result says that if we consider the case $k = 2$, where the first vector is nonnegative and the second is nonpositive then it is provably much harder to estimate the maximum entry of the vector sum.

The lower bound in [1] is obtained from a lower bound on a version of set-disjointness in the k-party communication model. An easy reduction shows that this lower bound carries over to a space lower bound in data stream model for the frequency moment problem.

Independently of our work, Bar-Yossef, Jayram, Kumar and Sivakumar [3] also proposed to use information theory to study communication complexity problems in the one-way and simultaneous communication models. In particular, in the simultaneous communication model, they obtained optimum lower bound for the multi-party set-disjointness problem in [1] mentioned above. As this paper was going to press, Bar-Yossef, Jayram, Kumar and Sivakumar [4] reported that, after seeing a preliminary version of our paper, they obtained optimum lower bound for the distance approximation problem in the general communication complexity model.

Our paper is organized as follows. In section 2, we review the model, give a precise formulation of the problem, and give some mathematical tools. In section 3, we present the general framework of our lower bound technique and use it to give a non-trivial lower bound for set-disjointness problem. In section 4, we present the main result. In section 5, we prove the main technical lemma used in section 4. In section 6, we present a reduction from the lower bound for toroidal L^∞ distance to usual L^∞ distance.

2. PRELIMINARIES

2.1 Communication complexity

We briefly review the two party communication model, and refer the reader to [9] for details. Two parties, referred to as Alice and Bob, each begin with an input; Alice has $x \in S_1$ and Bob has $y \in S_2$. They alternately send messages to each other about their inputs. A k-round deterministic communication protocol Π specifies a function from the pair (x, y) to a sequence $\Pi(x, y) = (a_1, b_1, a_2, b_2, \ldots, a_k, b_k)$. Each a_i and b_i is a binary string called a message, and $a_1, a_2, \ldots, a_k$ are the messages sent by Alice and $b_1, b_2, \ldots, b_k$ are the messages sent by Bob. Each successive message depends on the input of the sender and the previous messages. The sequence $\Pi(x, y)$ is called the *transcript* of Π on input (x, y). For $j \leq 2k$ we write $\Pi_j(x, y)$ for the subsequence of $\Pi(x, y)$ consisting of the first j messages; such a subsequence is called a *partial transcript of length j*. $Trans(\Pi)$ denotes the set of all transcripts, and $Trans^*(\Pi)$ is the set of partial transcripts of all lengths. If σ, τ are partial transcripts we write $\sigma \prec \tau$ if σ is a prefix of τ.

The last message b_k of the transcript τ is regarded as the output of the protocol and is denoted $\mathtt{OUT}(\tau)$. Thus the output of the protocol on input x, y is obtained by applying Π followed by $\mathtt{OUT}$; we denote this composition by $\mathtt{OUT}[\Pi]$. The function $\mathtt{OUT}[\Pi]$ on domain $S_1 \times S_2$ is the *function computed by* Π.

For a partial transcript τ, $\Pi^{-1}(\tau)$ denotes $\{(x, y) : \tau \prec \Pi(x, y)\}$. We have the following fundamental fact (See, for example, Lemma 1.16 of [9]):

LEMMA 2.1. *For any $\tau \in Trans^*(\Pi)$, $\Pi^{-1}(\tau)$ is a product set in $S_1 \times S_2$.*

We may therefore define $\Pi_A^{-1}(\tau) \subset S_1$ and $\Pi_B^{-1}(\tau) \subseteq S_2$ so that $\Pi^{-1}(\tau) = \Pi_A^{-1}(\tau) \times \Pi_B^{-1}(\tau)$.

In a randomized protocol, Alice (resp. Bob) generates an auxiliary string of r_A (resp. r_B) of random bits and Alice's (resp. Bob's) messages may depend on r_A (resp. r_B). The transcript $\Pi(x, y)$ is then a random variable and the output $\texttt{OUT}[\Pi]$ is a random function which maps $S_1 \times S_2$ to a distribution over output values.

The *cost* of a deterministic (resp. randomized) protocol Π on input (x, y) is the number of bits (resp., maximum number of bits) in the transcript $\Pi(x, y)$. The *complexity* of Π is the maximum over inputs (x, y) of the cost of Π on (x, y).

A *problem specification* with output domain T is a function f that maps each $(x, y) \in S_1 \times S_2$ to a nonempty subset of T. $f(x, y)$ is the set of *acceptable outputs* on input (x, y). In the case that $T = \{0, 1\}$, we say that f defines a *decision problem.* For decision problems, we view f as a (partial) function from $S_1 \times S_2$ to $\{0, 1, *\}$ instead of $\{\{0\}, \{1\}, \{0, 1\}\}$. We say that a randomized protocol ϵ-computes f if on every input (x, y) the probability that $\texttt{OUT}[\Pi](x, y) \notin f(x, y)$ is at most ϵ. $RCC_\epsilon(f)$ denotes the minimum complexity of any randomized protocol that ϵ-computes f. Define $ROCC_\epsilon(f)$ to be the minimum complexity of any randomized one-round protocol that ϵ-computes f.

Much of this paper is focused on *distributional communication complexity.* Let μ be a probability distribution on $S_1 \times S_2$. We write $\mu(x, y)$ for the probability that μ assigns to the pair (x, y). We denote random variables by capital letters, e.g. (X, Y) denotes a random input pair chosen according to μ. Associated with a k-round protocol are random variables $A_1, B_1, \dots, A_k, B_k$ where A_i and B_i are the messages sent by Alice and Bob in the round i.

Let Π be a deterministic communication protocol. Then μ induces a probability distribution on the transcript $\Pi(X, Y)$ as well as on the output $\texttt{OUT}[\Pi](X, Y)$. We say that Π ϵ-computes f relative to μ if for (X, Y) selected according to μ the probability that $\texttt{OUT}[\Pi](x, y) \notin f(x, y)$ is at most ϵ. The *distributional complexity of f with respect to* μ, $DCC_\epsilon^\mu(f)$, is the minimum communication complexity of any deterministic protocol that ϵ-computes f relative to μ. We also define $DOCC_\epsilon^\mu(f)$, to be the minimum communication complexity of any deterministic one-round protocol that ϵ-computes f relative to μ. The following well known lemma of Yao[13] connects distributional complexity and randomized complexity:

LEMMA 2.2. *For any probability distribution μ on $S_1 \times S_2$, $DCC_\epsilon^\mu(f) \le RCC_\epsilon(f)$ and $DOCC_\epsilon^\mu(f) \le ROCC_\epsilon(f)$.*

In this paper we will prove lower bounds on distributional communication complexity, and the above lemma shows that the same bounds apply to randomized communication complexity.

In the case that the output set of f is $\{0, 1\}$, we need a more refined measure of the quality of a deterministic protocol Π relative to distribution μ. We say that Π (ϵ_0, ϵ_1)-*computes* f relative to μ if $\Pr_\mu[\Pi(X, Y) = 1 | f(X, Y) = 0] \le \epsilon_0$ and $\Pr_\mu[\Pi(X, Y) = 0 | f(X, Y) = 1] \le \epsilon_1$. Trivially, we have:

PROPOSITION 2.3. *If a deterministic protocol Π ϵ-computes a boolean function f relative to μ, then Π (ϵ_0, ϵ_1)-computes f where $\epsilon_0 = \epsilon / \Pr_\mu[f(X, Y) = 0]$ and $\epsilon_1 = \epsilon / \Pr_\mu[f(X, Y) = 1]$.*

For $\tau \in Trans^*(\Pi)$, we need to understand how conditioning on the event $\tau \prec \Pi(X, Y)$ changes the distribution of (X, Y). Let $\alpha^\tau \in \{0, 1\}^{S_1}$ be the characteristic vector of the set $\Pi_A^{-1}(\tau)$, and $\beta^\tau \in \{0, 1\}^{S_2}$ be the characteristic vector of $\Pi_B^{-1}(\tau)$. Applying the definition of conditional probability and Lemma 2.1 immediately gives:

LEMMA 2.4. *Let μ be a distribution on $S_1 \times S_2$ and Π a communication protocol and let $\tau \in Trans^*(\Pi)$, and let μ' be the distribution μ conditioned on the event $\tau \prec \Pi(X, Y)$. Then (1) for $(x, y) \in S_1 \times S_2$, $\mu'(x, y) = \alpha^\tau(x)\mu(x, y)$ $\beta^\tau(y)/\mu(\Pi^{-1}(\tau))$, and (2) if μ is a product distribution (so that X, Y are independent) then so is μ'.*

2.2 Distance problems

Throughout this paper, d and n are positive integers. If S is a set, we denote elements of S^d in bold: $\mathbf{x} = (x_1, \dots, x_d)$. For $i \in [d]$, $S^{d \setminus i}$ denotes the set of partial vectors that are undefined in position i. An element of $S^{d \setminus i}$ is denoted by superscripting with i, e.g., $\mathbf{y}^i$. If $\mathbf{x} \in S^d$ and $i \in [d]$ then, $\mathbf{x}^i \in S^{d \setminus i}$ is obtained by restricting x in the obvious way.

For $p > 0$, the L^p distance between $\mathbf{x}, \mathbf{y} \in [n]^d$ is defined as $\rho_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^d |x_i - y_i|^p\right)^{\frac{1}{p}}$. The L^∞ distance between $\mathbf{x}, \mathbf{y}$ is defined as $\rho_\infty(\mathbf{x}, \mathbf{y}) = \max_{1 \le i \le d} |x_i - y_i|$ and the toroidal L^∞ distance is $\rho_\odot(\mathbf{x}, \mathbf{y}) = \max_{1 \le i \le d} ||x_i - y_i||_n$ where $||z||_n = \min(|z|, n - |z|)$ for $z \in [-n, n]$.

We will prove lower bounds on the one-round communication complexity of the following problems whose input set is $S^d \times S^d$. Let n, d be positive integer and ρ be a metric on $[n]^d$ and let $K \ge 1$ and $\theta_L \le \theta_U$ be positive constants.

The distance estimation problem(DEP) for (n, d, ρ, K) is to estimate $\rho(\mathbf{x}, \mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in [n]^d$ within a factor K. Thus, the acceptable output for the protocol is a number z such that $\rho(\mathbf{x}, \mathbf{y})/K \le z \le K\rho(\mathbf{x}, \mathbf{y})$.

The distance threshold decision problem(DTDP) for $(n, d, \rho, \theta_L, \theta_U)$ is to output 0 if $\rho(\mathbf{x}, \mathbf{y}) \le \theta_L$, output 1 if $\rho(\mathbf{x}, \mathbf{y}) \ge \theta_U$. For $\theta_L < \rho(\mathbf{x}, \mathbf{y}) < \theta_U$ either 0 or 1 is acceptable.

Suppose Π is any randomized communication protocol with domain $S_1 \times S_2$ that outputs a real number, and w is any real number, we define $\Pi[w]$ to be the protocol that runs Π and outputs 0 if the output of Π is less than w and 1 if the output of Π is greater than w.

PROPOSITION 2.5. *If Π solves DEP(n, d, ρ, K) with error probability at most ϵ and $K < \sqrt{\theta_U/\theta_L}$ then $\Pi[\sqrt{\theta_L \theta_U}]$ solves DTDP$(n, d, \rho, \theta_L, \theta_U)$ with error probability at most ϵ.*

Combining this proposition and Lemma 2.2 we conclude that to prove a lower bound on the ϵ-error randomized complexity of DEP(n, d, ρ, K) it is enough to prove a lower bound on the ϵ-error distributional complexity of DTDP$(n, d, \rho, \theta, K^2\theta)$ relative to any distribution μ of our choice, and for any θ of our choice.

2.3 Technical preliminaries

Information theory

We review some elementary concepts from information theory (see e.g., [5]). The setting for our discussion is that we have a probability space and random variables on the space, each of which takes values from a finite set. If X is such a random variable taking values from S then its distribution function p is a stochastic function on S. The *entropy* $H(X)$ of X is defined to be $h(p)$. If A is an event, the conditional entropy of X given A, $H(X|A)$, is $h(q)$ where q is the conditional distribution function for X. For random variables X, Y we define $H(X|Y) = H(X,Y) - H(X)$; this is equivalent to $H(X|Y) = \sum_{t \in T} H(X|Y=t)Prob(Y=t)$ where T is the set of possible values of Y. The *mutual information* between X, Y is defined as:

$$\begin{aligned} I(X:Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) = H(Y) - H(Y|X). \end{aligned}$$

For random variables X, Y, Z, the *conditional mutual information* $I(X:Y|Z)$ is defined as:

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

The main technical fact about entropy is its subadditivity:

Lemma 2.6. *For any random variables $X_1, \dots, X_n$, $H(X_1, \dots, X_n) \le \sum_{i=1}^{n} H(X_i)$.*

This implies, in particular that $H(X|Y) \le H(X)$ for any random variables X and Y.

The following simple facts are easily derived from the definitions or from subadditivity.

Lemma 2.7. *Given four random variables X, Y, Z, W, we have $I(X:YZ|W) = I(X:Z|W) + I(X:Y|ZW)$.*

Lemma 2.8. *Given three random variables X, Y, Z, we have $I(X:Y|Z) = I(XZ:YZ) - H(Z)$.*

Lemma 2.9. *Given three random variables X, Y, Z, we have $I(X:Y|Z) \le H(X)$.*

Lemma 2.10. *Given three random variables X, Y, Z and a function f, we have $I(X:f(Y)|Z) \le I(X:Y|Z)$.*

Using these facts, it is easy to deduce:

Lemma 2.11. *Let $\mathbf{X} = (X_1, \dots, X_d), Y, Z$ be random variables with $X_1, \dots, X_n$ mutually independent conditioned on Z. Then $I(Y:\mathbf{X}|Z) \ge \sum_{i=1}^{d} I(Y:X_i|Z)$.*

Some inequalities

This section contains some elementary inequalities. We omit the easy and routine proofs for lack of space.

If p is a nonnegative real valued function on the finite set S, we write $\bar{p}$ for the average of p, and $h(p) = \sum_{s \in S} p(s)\log(1/p(s))$ ($\log x$ always denotes the logarithm base 2). Also, if $T \subseteq S$ we write $p(T)$ for $\sum_{s \in T} p(s)$. Note that here we do not require that $p(S) = 1$; if $p(S) = 1$ we say that p is a *stochastic function.*

The convexity of the function $(1+x)\log(1+x)$ implies:

Lemma 2.12. *Let p be a nonnegative valued function on the set S. Then:*

$$h(p) \le |S|\bar{p}(\log 1/\bar{p})$$

In the case that p is a probability distribution, the right hand side is just the entropy of the uniform distribution on S. The quantity $|S|\bar{p}(\log 1/\bar{p}) - h(p)$ is the *entropy deficiency* of p and is denoted $h^-(p)$. (Note that the definition of $h^-(p)$ requires that the set S be clear.) We will derive some upper and lower bounds on $h^-(p)$ in terms of p.

We have the following routine estimates of $(1+x)\log(1+x)$. For $x \ge -1$,

$$\frac{(1+x)x}{\ln 2} \ge (1+x)\log(1+x) \ge \frac{x}{\ln 2} \tag{1}$$

If $\delta \in [0, 1/2]$, then for $x \ge -1$ and $|x| \ge \delta$:

$$(1+x)\log(1+x) \ge \frac{x}{\ln 2} + \frac{\delta^2}{4} \tag{2}$$

If $x \ge 1$ then:

$$(1+x)\log(1+x) \ge \frac{x}{\ln 2} + \frac{x}{4} \tag{3}$$

From the upper bound in (1) we get:

Lemma 2.13. *For any nonnegative valued function p on set S,*

$$h^-(p) \le \frac{1}{p \ln 2} \sum_{s \in S} (p(s) - \bar{p})^2$$

Using the lower bounds on $(1+x)\log(1+x)$ in (1), (2) and (3) one can show:

Lemma 2.14. *Let p be a nonnegative valued function on set S. Let $\delta \in [0, 1/2]$. Suppose that $T \subseteq S$ satisfies $|\frac{p(T)}{|T|} - \bar{p}| \ge \delta\bar{p}$. Then $h^-(p) \ge \delta^2|T|/(4|S|)$.*

Lemma 2.15. *Let p be a nonnegative valued function on set S. Suppose that $T \subseteq S$ satisfies $\frac{p(T)}{|T|} \ge 2\bar{p}$. Then $h^-(p) \ge p(T)/8$.*

Corollary 2.16. *Let p be a stochastic function on set S and let $T \subseteq S$. (1) If $\frac{p(T)}{|T|} \le \frac{1}{2|S|}$, then $h^-(p) \ge \frac{|T|}{16|S|}$. (2) If $\frac{p(T)}{|T|} \ge \frac{2}{|S|}$, then $h^-(p) \ge \frac{p(T)}{8}$.*

We also need another technical fact concerning the convexity of certain functions on $\mathbb{R}^d$.

Lemma 2.17. *Let g, h be linear functions mapping $\mathbb{R}^d$ to $\mathbb{R}$ and let W be the subset of the domain where h is positive. Then $f = \frac{g^2}{h}$ is a convex function on W.*

3. A NEW APPROACH TO COMMUNICATION COMPLEXITY LOWER BOUNDS

3.1 The framework

Let S be a set and $\mathbf{x} = (x_1, x_2 \dots, x_d)$, $\mathbf{y} = (y_1, y_2 \dots, y_d)$ vectors in S^d. Let $g : S \times S \longrightarrow \{0, 1, *\}$ be a partial function. We define $g^{\vee d}(\mathbf{x}, \mathbf{y}) = \vee_{i=1}^{d} g_i(\mathbf{x}, \mathbf{y})$ where $g_i(\mathbf{x}, \mathbf{y}) = g(x_i, y_i)$. Here $\vee_{i=1}^{d} z_i = 1$ if $z_i = 1$ for some $i \in [d]$, $\vee_{i=1}^{d} z_i = 0$ if $z_i = 0$ for all $i \in [d]$ and $\vee_{i=1}^{d} z_i = *$ otherwise. In this section we present a framework for proving lower bounds on communication complexity for boolean functions of the form $g^{\vee d}$.

We begin by choosing the distribution μ on $S^d \times S^d$. We have two requirements for the distribution: (i) if we write $W_i = (X_i, Y_i)$ then the distributions of $W_1, \dots, W_d$ should be mutually independent, and (ii) the probabilities that $g^{\vee d} = 1$ and $g^{\vee d} = 0$ should be bounded away from 0 independent of d. To accomplish this we choose a distribution ν on $S \times S$ so that $\Pr_\nu[g(X,Y) = 1] = \Theta(1/d)$ and $\Pr_\nu[g(X,Y) = 0] = 1 - \Theta(1/d)$; the second condition does not follow from the first since we have to consider $*$ values for g. (There are other considerations in the choice of ν which we will deal with later.) The product distribution $\mu = \nu^d$ on $(S \times S)^d$ then satisfies (i) and (ii).

We now observe that if a deterministic protocol computes f with small error, then for most i, it must output 1 on almost all inputs for which $g_i = 1$, and must output 0 on a nontrivial fraction of inputs for which $g_i = 0$.

Lemma 3.1. *Let Π be a deterministic protocol that (ϵ, ϵ)-computes $g^{\vee d} = \vee_{i=1}^d g_i$ relative to μ. There exists some $I \subseteq [d]$ s.t. $|I| > (1 - 2\sqrt{\epsilon})d$ and Π $(\frac{3}{4}, 2\sqrt{\epsilon})$-computes g_i relative to μ for all $i \in I$.*

Proof. Define the random variables $T = \Pi(\mathbf{X}, \mathbf{Y})$, $G_i = g_i(\mathbf{X}, \mathbf{Y})$ and $G = g^{\vee d}(\mathbf{X}, \mathbf{Y})$. First, we have

$$\begin{aligned} \Pr[\mathtt{OUT}(T) = 0 | G_i = 0] &\geq \Pr[\mathtt{OUT}(T) = 0 \wedge G_i = 0] \\ &\geq \Pr[\mathtt{OUT}(T) = 0 \wedge G = 0] \\ &= \Pr[\mathtt{OUT}(T) = 0 | G = 0] \Pr[G = 0] \\ &\geq (1 - \epsilon)\tfrac{1}{e} > \tfrac{1}{4} \end{aligned}$$

Let $I' = \{i \in [d] : \Pr[\mathtt{OUT}(T) = 0 | G_i = 1] \geq 2\sqrt{\epsilon}\}$. We want to show $|I'| < 2\sqrt{\epsilon}d$. Suppose not, then we pick $J \subseteq I'$ s.t. $|J| = 2\sqrt{\epsilon}d$ and use the inclusion-exclusion inequality to obtain

$$\begin{aligned} &\Pr[\mathtt{OUT}(T) = 0 \wedge G = 1] \\ &\quad \geq \Pr[\mathtt{OUT}(T) = 0 \wedge \vee_{i \in J}(G_i = 1)] \\ &\quad = \Pr[\vee_{i \in J}(\mathtt{OUT}(T) = 0 \wedge G_i = 1)] \\ &\quad \geq \textstyle\sum_{i \in J} \Pr[\mathtt{OUT}(T) = 0 \wedge G_i = 1] \\ &\qquad - \textstyle\sum_{i,j \in J, i \neq j} \Pr[\mathtt{OUT}(T) = 0 \wedge (G_i = G_j = 1)] \\ &\quad \geq \textstyle\sum_{i \in J} \Pr[\mathtt{OUT}(T) = 0 \wedge G_i = 1] \\ &\qquad - \textstyle\sum_{i,j \in J, i \neq j} \Pr[(G_i = G_j = 1)] \\ &\quad \geq |J| \tfrac{2\sqrt{\epsilon}}{d} - \tbinom{|J|}{2} \tfrac{1}{d^2} \geq 2\epsilon \end{aligned}$$

Therefore, we obtain the following contradiction

$$\Pr[\mathtt{OUT}(T) = 0 | G = 1] \geq 2\epsilon/(1 - 1/e) > \epsilon.$$

□

The main part of the argument is based on the information theoretic intuition that if the communication complexity is small then on average the communication does not reveal too much information about X_i, Y_i therefore, for most i, the algorithm will not be able to approximate g_i even in the weak sense given by the above lemma.

3.2 Lower bound for set-disjointness

As a warmup, let us use our framework to give a nontrivial communication complexity lower bound for set-disjointness problem. In the set-disjointness problem, we are given two boolean vectors of length n, and we wish to determine whether there is a coordinate where they are both 1. This problem was first studied by Babai, Frankl and Simon [2] and they obtained a lower bound $\Omega(\sqrt{n})$. Their result was later improved to $\Omega(n)$ by Kalyanasundaram and Schnitger [10] and a simplified proof was presented by Razborov [12]. Here we illustrate our framework by proving an $\Omega(\sqrt{n})$ lower bound.

Partition the n-coordinates into $d = \sqrt{n}$ blocks of $\sqrt{n}$ bits each, and restrict attention to boolean vectors that have exactly one 1 in each block. We can represent such a vector $\mathbf{z}$ by a vector $\mathbf{x} \in [d]^d$ where x_i indicates the position of the 1 within the i-th block of $\mathbf{z}$. With this restriction the set-disjointness problem becomes: evaluate $f(\mathbf{x}, \mathbf{y}) = g^{\vee d}(\mathbf{x}, \mathbf{y})$ where for $x, y \in [d]$, $g(x, y) = 1$ if $x = y$ and 0 otherwise. We will prove a lower bound on the distributional complexity of this problem, for the distribution $\mu = \nu^d$, where ν is the uniform distribution $[d] \times [d]$. Throughout this section, let $S = [d]$ and $f = \vee_{i=1}^d g_i$ where $g_i(\mathbf{x}, \mathbf{y}) = g(x_i, y_i)$.

Let $\mathbf{X}, \mathbf{Y}$ be random variables chosen according to μ. Fix a two-party protocol Π that takes input from $[d]^d \times [d]^d$. Let T denote the (random) transcript $\Pi(\mathbf{X}, \mathbf{Y})$. Clearly the entropy $H(T)$ is a lower bound on the communication complexity of Π. Using Lemmas 2.9 and 2.11:

Lemma 3.2. *$H(T) \geq I(T : \mathbf{X}, \mathbf{Y}) \geq \sum_{i=1}^d I(T : X_i, Y_i)$.*

Our goal now is to show that if Π ϵ-computes f (on the given distribution, for suitably small ϵ) then for most indices i, $I(T : X_i, Y_i)$ is bounded below by a constant (which will thus give an $\Omega(d)$ communication lower bound). By Lemma 3.1 for most indices i, the protocol Π $(\frac{3}{4}, 2\sqrt{\epsilon})$ computes g_i. Therefore it is enough to give a lower bound on $I(T : X_i, Y_i)$ for each such i. This is stated as the following lemma.

Lemma 3.3. *Let $d > 5$, $i \in [d]$ and $\delta > 0$. Suppose that Π $(\frac{3}{4}, \delta)$-computes g_i. If $\delta \leq \frac{1}{80}$, then $I(T : X_i, Y_i) \geq \frac{1}{640}$.*

Proof. Let G_i be the random variable $g_i(X, Y)$. Assume $\Pr[\mathtt{OUT}(T) = 1 | G_i = 0] \leq \frac{3}{4}$ and $\Pr[\mathtt{OUT}(T) = 0 | G_i = 1] \leq \delta$.

For $\lambda > 0$, define the set

$$W_\lambda = \{\tau \in \mathit{Trans}(\Pi) : H(X_i, Y_i) - H(X_i, Y_i | T = \tau) \geq \lambda\}.$$

Notice that $H(X_i, Y_i) - H(X_i, Y_i | T = \tau)$ is nonnegative for any τ since X_i and Y_i are uniformly distributed. We have

$$\begin{aligned} & I(T : X_i, Y_i) \\ = \; & \textstyle\sum_{\tau \in \mathit{Trans}(\Pi)} (H(X_i, Y_i) - H(X_i, Y_i | T = \tau)) \Pr[T = \tau] \\ \geq \; & \textstyle\sum_{\tau \in W_\lambda} (H(X_i, Y_i) - H(X_i, Y_i | T = \tau)) \Pr[T = \tau] \\ \geq \; & \lambda \Pr[T \in W_\lambda] \end{aligned}$$

Our goal is to lower bound $\Pr[T \in W_\lambda]$ for a suitable constant λ. We start with:

Claim 3.4. *For any τ, if $\Pr[X_i = Y_i | T = \tau] < \frac{1}{8d}$, $\tau \in W_{\frac{1}{64}}$.*

Proof of Claim 3.4. By the second part of Lemma 2.4, X_i, Y_i conditioned on $T = \tau$ are independent. Thus, $H(X_i, Y_i) - H(X_i, Y_i | T = \tau) = (\log d - H(X_i | T = \tau)) + (\log d - H(Y_i | T = \tau)$, where the two terms are nonnegative, so it suffices to show that one of them is at least 1/64.

Let U_X be the set of $j \in [d]$ such that $\Pr[X_i = j | T = \tau] \leq 1/2d$, and define U_Y analogously. If both $|U_X|$ and $|U_Y|$ are smaller than $d/4$, then there are $d/2$ indices j for which $\Pr[X_i = Y_i = j | T = \tau] \geq \frac{1}{4d^2}$, which contradicts the hypothesis that $\Pr[X_i = Y_i | T = \tau] < \frac{1}{8d}$. So without loss of generality $|U_X| \geq d/4$. For $j \in [d]$, define $p_j = \Pr[X_i = j|T = \tau]$. Applying Lemma 2.14 with $\delta = \frac{1}{2}$ gives $\log d - H(X_i|T = \tau) \geq \frac{1}{64}$, to prove the claim.

For $\gamma > 0$, let $B_\gamma = \{\tau \in \mathtt{OUT}^{-1}(0) : \Pr[X_i = Y_i|T = \tau] > \gamma\}$. By the claim, $W_{\frac{1}{64}} \subseteq \mathtt{OUT}^{-1}(0) - B_{\frac{1}{8d}}$, so it suffices to lower bound $\Pr[\mathtt{OUT}(T) = 0] - \Pr[T \in B_{\frac{1}{8d}}]$. We have:

$$\begin{aligned}\Pr[\mathtt{OUT}(T) = 0] \geq\ & \Pr[\mathtt{OUT}(T) = 0|G_i = 0]\Pr[G_i = 0]\\ \geq\ & (1 - \tfrac{3}{4})\tfrac{d-1}{d} > \tfrac{1}{5},\end{aligned}$$

We upper bound $\Pr[T \in B_\gamma]$ as follows.

$$\begin{aligned}\delta\ & \geq \Pr[\Pi(T) = 0|G_i = 1]\\ & = d\Pr[\mathtt{OUT}(T) = 0 \wedge G_i = 1]\\ & \geq d\textstyle\sum_{\tau \in B_\gamma} \Pr[T = \tau]\Pr[X_i = Y_i|T = \tau]\\ & \geq d\gamma\Pr[T \in B_\gamma].\end{aligned}$$

Thus $\Pr[T \in B_{\frac{1}{8d}}] \leq 8\delta$, and so $\Pr[T \in W_{\frac{1}{64}}] \geq \frac{1}{5} - 8\delta$. Choosing $\delta = \frac{1}{80}$ we obtain:

$$I(T : X_i, Y_i) \geq \frac{1}{64}\Pr[T \in W_{\frac{1}{64}}] \geq \frac{1}{640}.$$

□

PROPOSITION 3.5. *Any deterministic protocol that $(\frac{1}{160^2}, \frac{1}{160^2})$-computes the set-disjointness problem must use $\Omega(\sqrt{n})$ bits.*

PROOF. Assume that Π $(\frac{1}{160^2}, \frac{1}{160^2})$-computes the set-disjointness problem. By Lemma 3.1, Π $(\frac{3}{4}, \frac{1}{80})$-computes g_i for at least $\frac{79}{80}d$ indices $i \in [d]$. By Lemma 3.3, $I(\Pi(\mathbf{X}, \mathbf{Y}) : X_i, Y_i) \geq \frac{1}{640}$ for all such i. By Lemma 3.2, $H(\Pi(\mathbf{X}, \mathbf{Y})) \geq \frac{79}{80}d\frac{1}{640} = \Omega(\sqrt{n})$. □

4. SPACE LOWER BOUND FOR THE L^∞ DTDP

For the rest of the paper, d and n are integers with n prime and $n \geq 2d^{1+\delta}$, where δ is a small positive constant to be chosen. In this section and the next, we prove a lower bound on the one-way communication complexity of the DTDP (distance threshold decision problem) for vectors in $\mathbb{Z}_n{}^d$ under distance measure $\rho_\odot$ (recall that $\rho_\odot(\mathbf{x}, \mathbf{y}) = \max_{1 \leq i \leq d} ||x_i - y_i||_n$ where $||z||_n = \min(|z|, n - |z|)$) with lower threshold d and upper threshold $d^{1+\delta}$. In section 6, we use this lower bound to get a similar bound for the case of ρ_∞ distance.

We now recast this decision problem in the framework of section 3.1. Let $g : \mathbb{Z}_n \times \mathbb{Z}_n \longrightarrow \{0, 1, *\}$ s.t. $g(x, y) = 0$ whenever $||x - y||_n \leq d$, $g(x, y) = 1$ whenever $||x - y||_n \geq d^{1+\delta}$ and $g(x, y) = *$ otherwise. Let $f = g^{\vee d} = \vee_{i=1}^d g_i$ where $g_i(\mathbf{x}, \mathbf{y}) = g(x_i, y_i)$. We seek a lower bound for the one-round communication complexity of f for some distribution μ. Following the outline in section 3.1, we want a product distribution $\mu = \nu^d$ where ν is a distribution on $\mathbb{Z}_n \times \mathbb{Z}_n$ that maps to pairs within distance $d - 1$ with probability $1 - 1/d$ and to pairs of distance at least $d^{1+\delta}$ with probability $1/d$. A natural choice for such distribution is the uniform distribution over the set $P = \{(x, x + z) : x \in \mathbb{Z}_n, z \in [d - 1] \cup \{d^{1+\delta}\}\}$. If we select (X_i, Y_i) with this distribution then X_i and Y_i are each uniform on $\mathbb{Z}_n$ but are not independent. This seems unavoidable and complicates the proof. Indeed, this complication will force us to restrict attention to one-round protocols (although, we believe that it will eventually be possible to remove this restriction.) Note that we do have that if we define $\Delta_i = Y_i - X_i$ then X_i and Δ_i are independent and also Y_i and Δ_i are independent.

Let us see what goes wrong if we try to mimic the proof of the set-disjointness lower bound to obtain a lower bound on the communication complexity of f. Let Π be a protocol and let $T = \Pi(\mathbf{X}, \mathbf{Y})$ be the transcript of the protocol. As in the previous proof, we seek to lower bound $H(T)$, and we write it as a sum $\sum_{i=1}^d I(T : X_i, Y_i)$. We want a counterpart to Lemma 3.3, that says that if Π $(\frac{3}{4}, \delta)$-computes g_i (for some appropriate δ), then $I(T : X_i, Y_i)$ can not be too small.

As in the proof of Lemma 3.3, we can write $I(T : X_i, Y_i)$ as a sum over transcripts τ of $(H(X_i, Y_i) - H(X_i, Y_i|T = \tau))\Pr[T = \tau]$. Even though X_i, Y_i are not independent it is still true that the pair (X_i, Y_i) is uniformly distributed over its support and therefore $(H(X_i, Y_i) - H(X_i, Y_i|T = \tau))$ is always nonnegative. Defining W_λ as before, $I(T : X_i, Y_i) \geq \lambda\Pr[T \in W_\lambda]$. Again, we would like to choose λ so that for "most" i, $\lambda\Pr[T \in W_\lambda]$ is not too small (note that W_λ depends implicitly on i). In the previous proof, we defined $B_\gamma = \{\tau \in \Pi^{-1}(0) : \Pr[g_i = 1|T = \tau] > \gamma\}$ and showed that for $\gamma = 1/8d$, and $\lambda = 1/64$ (1) $\Pr[\tau \in \mathtt{OUT}^{-1}(0) - B_\gamma] > 1/10$ and (2) $\mathtt{OUT}^{-1}(0) - B_\gamma \subseteq W_\lambda$. Condition (1) is still true. However, the proof of the second condition, which relied heavily on the independence of X_i and Y_i, falls apart. Specifically, independence was needed for the claim in Lemma 3.3 that says that if τ is a transcript and the "entropy loss" $H(X_i, Y_i) - H(X_i, Y_i|T = \tau)$ is small then $\Pr[g_i(X_i, Y_i) = 1|T = \tau]$ can not be much smaller than $\Pr[g_i(X_i, Y_i) = 1] = 1/d$. For this claim, we needed not only that X_i and Y_i are independent, but that X_i, Y_i conditioned on $T = \tau$ are still independent, and this fact followed from the second part of Lemma 2.4.

Lacking independence we try to modify the claim by showing: (i) For any transcript τ, the distribution of X_i, Y_i conditioned on $T = \tau$ is "nice" in some sense, and (ii) something like the claim holds if we replace independence by "niceness". This line of thought led us to consider conditioning not just on the value of T, but also on the value of $\mathbf{Y}^i$ (Bob's input apart from Y_i). We can then prove that the distribution of X_i, Y_i conditioned on $(T = \tau, \mathbf{Y}^i = \mathbf{y}^i)$ is nice in some sense that enables us to prove something like the claim. This enables us to prove an analog of Lemma 3.3 and thereby show that for most indices $i \in [d]$, $I(T, \mathbf{Y}^i|X_i, Y_i)$ is bounded below by some constant. However, this does not finish the proof because, while $H(T) \geq \sum_i I(T|X_i, Y_i)$, it is not the case that $H(T) \geq \sum_i I(T, \mathbf{Y}^i|X_i, Y_i)$. So we need to lower bound $H(T)$ in terms of $\sum_i I(T, \mathbf{Y}^i|X_i, Y_i)$. Intuitively, it seems reasonable that $I(T, \mathbf{Y}^i|X_i, Y_i)$ should not be much different than $I(T|X_i, Y_i)$ since $I(T, \mathbf{Y}^i|X_i, Y_i)$ is the amount of information that $T, \mathbf{Y}^i$ reveal about X_i, Y_i, and since $\mathbf{Y}^i$ is independent of (X_i, Y_i) it should not affect things. However, this intuition is wrong. Suppose that $T = \sum_i Y_i \bmod n$. Then $(T, \mathbf{Y}^i)$ determines Y_i, so $I(T, \mathbf{Y}^i|X_i, Y_i) = \log n$, while $I(T|X_i, Y_i) = 0$. Note also here that $H(T) = \log n$, which is factor $1/n$ of $\sum_i I(T, \mathbf{Y}^i|X_i, Y_i) = n \log n$. This would seem to kill this approach, but it turns out that if the transcript T only depends on X (i.e., consists only of a message from Alice) then we can show that $H(T)$ can't be much smaller than $\sum_i I(T, \mathbf{Y}^i | X_i, Y_i)$. This argument (which is the main technical result of the paper) is given in Section 5.

This argument appears like it should give a lower bound for 1-round protocols, but there is one remaining difficulty in the above sketch. For a one-round protocol Π it is not true that the transcript T depends only on X since it is of the form (A, B) where A is the message from Alice and B is the output declared by Bob. So we must further modify our counterpart to Lemma 3.3 so that we condition only on the values of A and $\mathbf{Y}^i$ rather than T and $\mathbf{Y}^i$. More precisely, we prove:

LEMMA 4.1. *Let $d > 5$, $i \in [d]$ and $\delta > 0$. Suppose that Π is a one-round protocol that $(\frac{3}{4}, \delta)$-computes g_i. If $\delta \leq \frac{1}{800}$*

$$I(A, \mathbf{Y}^i : X_i, Y_i) \geq \frac{1}{6400},$$

PROOF. The transcript of the one-round protocol Π is (A, B) where Bob's message B is the output of the protocol. Let G_i be the random variable $g_i(X, Y)$. Assume $\Pr[B = 1 | G_i = 0] \leq \frac{3}{4}$ and $\Pr[B = 0 | G_i = 1] = \delta$. We will show that if $\delta \leq 1/800$ then $I(A, \mathbf{Y}^i : X_i, Y_i) \geq 1/6400$.

For a a possible message of Alice and $\mathbf{y}^i \in [n]^{d \setminus i}$, let $E(a, \mathbf{y}^i)$ denote the event that $A = a$ and $\mathbf{Y}^i = \mathbf{y}^i$. For $\lambda > 0$, define the set

$$W_\lambda \;=\; \{(a, \mathbf{y}^i) : H(X_i, Y_i) - H(X_i, Y_i | E(a, \mathbf{y}^i)) \geq \lambda\}$$

By a computation analogous to that in the proof of Lemma 3.3:

$$I(A, \mathbf{Y}^i : X_i, Y_i) \geq \lambda \Pr[(A, \mathbf{Y}^i) \in W_\lambda]. \tag{4}$$

For $\gamma, \gamma' > 0$, define

$$\begin{aligned} V_\gamma &= \{(a, \mathbf{y}^i) : \Pr[B = 0 | E(a, \mathbf{y}^i)] \geq \gamma\} \\ U_{\gamma'} &= \{(a, \mathbf{y}^i) : \Pr[B = 0 | (G_i = 1) \wedge E(a, \mathbf{y}^i)]\} \leq \gamma'. \end{aligned}$$

For suitable parameters γ, γ', and λ, we will show $V_\gamma \cap U_{\gamma'} \subseteq W_\lambda$ and give a lower bound $\Pr[(A, \mathbf{Y}^i) \in V_\gamma \cap U_{\gamma'}]$. Together with (4) this will give a lower bound on $I(A, \mathbf{Y}^i : X_i, Y_i)$. We proceed by a sequence of claims.

CLAIM 4.2. $\Pr[(A, \mathbf{Y}^i) \in V_{1/10}] \geq \frac{1}{10}$.

Proof of Claim 4.2. We have $\Pr[B = 0] \geq 1/5$ by the same derivation as in the proof of Lemma 3.3 (with B replacing $\mathtt{OUT}(T)$.) Now

$$\begin{aligned} \Pr[B = 0] \leq\;& \Pr[(B = 0) \wedge ((A, \mathbf{Y}^i) \in V_\gamma)] \\ & + \Pr[(B = 0) \wedge ((A, \mathbf{Y}^i) \notin V_\gamma)] \\ \leq\;& \Pr[(A, \mathbf{Y}^i) \in V_\gamma] \\ & + \Pr[B = 0 | (A, \mathbf{Y}^i) \notin V_\gamma] \\ \leq\;& \Pr[(A, \mathbf{Y}^i) \in V_\gamma] + \gamma, \end{aligned}$$

from which the claim follows immediately.

For the next two claims, we fix a pair $(a, \mathbf{y}^i)$. By definition, (X_i, Y_i) has uniform distribution ν over the nd element set $P = \{(x, y) \in {\mathbb{Z}_n}^d : y - x \in [d-1] \cup \{d^{1+\delta}\}\}$. Let $\nu'(x, y) = \Pr[(X_i, Y_i) = (x, y) | E(a, \mathbf{y}^i)]$. We can view ν' as a distribution on P. The condition $(a, \mathbf{y}^i) \in W_\lambda$ says that the entropy deficiency $h^-(\nu') \geq \lambda$.

CLAIM 4.3. *There is a stochastic function q on $[n]$ such that $\nu'(x, y) = n\nu(x, y)q(x)$.*

Proof of Claim 4.3. For $x \in [n]$, let $R(a, x)$ be the set of $\mathbf{x}^i \in [n]^{d \setminus i}$ such that on input $(\mathbf{X}^i, X_i) = (\mathbf{x}^i, x)$, Alice sends a. Then:

$$\begin{aligned} \nu'(x, y) &= \tfrac{\Pr[((X_i, Y_i) = (x, y)) \wedge (A = a) \wedge (\mathbf{Y}^i = \mathbf{y}^i)]}{\Pr[(A = a) \wedge (\mathbf{Y}^i = \mathbf{y}^i)]} \\ &= \tfrac{\Pr[((X_i, Y_i) = (x, y)) \wedge (\mathbf{X}^i \in R(a, x)) \wedge (\mathbf{Y}^i = \mathbf{y}^i)]}{\Pr[(A = a) \wedge (\mathbf{Y}^i = \mathbf{y}^i)]} \\ &= \tfrac{\Pr[(X_i, Y_i) = (x, y)] \Pr[(\mathbf{X}^i \in R(a, x)) \wedge (\mathbf{Y}^i = \mathbf{y}^i)]}{\Pr[(A = a) \wedge (\mathbf{Y}^i = \mathbf{y}^i)]} \\ &= \nu(x, y) n q(x), \end{aligned}$$

where $q(x) = \frac{\Pr[(\mathbf{X}^i \in R(a, x)) \wedge (\mathbf{Y}^i = \mathbf{y}^i)]}{n \Pr[(A = a) \wedge (\mathbf{Y}^i = \mathbf{y}^i)]}$ depends on x, a and $\mathbf{y}^i$ but not on y. The crucial line in the above derivation is the third, where we use the independence of (X_i, Y_i) with respect to (X^i, Y^i). Finally, ν being uniform on P and ν' being stochastic implies that q is stochastic.

It follows from this claim, and the definition of $h^-(\cdot)$ that $h^-(\nu') = h^-(q)$.

CLAIM 4.4. *If $(a, \mathbf{y}^i) \in V_{\frac{1}{10}} \cap U_{\frac{1}{40}}$ then $(a, \mathbf{y}^i) \in W_{\frac{1}{320}}$.*

Proof of Claim 4.4. Assume that $\Pr[B = 0 | E(a, \mathbf{y}^i)] \geq 1/10$ and $\Pr[B = 0 | E(a, \mathbf{y}^i) \wedge G_i + 1] \leq \frac{1}{40}$. We need to show $h^-(\nu') \geq 1/320$.

When conditioned on $E(a, \mathbf{y}^i)$, B only depends on Y_i. Let $L = L(a, \mathbf{y}^i)$ be the set of $y \in [n]$ which cause $B = 0$ under this conditioning, and let $P(L) = \{(x, y) \in P : y \in L\}$, so $|P(L)| = d|L|$. Then $\Pr[B = 0 | E(a, \mathbf{y}^i)] = \nu'(P(L))$. If $|L| \leq n/20$, then $\nu'(P(L))/|P(L)| \geq 2/|P|$ and the second part of Corollary 2.16 implies $h^-(\nu') \geq 1/80$.

So assume $|L| > n/20$. Let $L' = \{x : \exists y \in L, \text{ s.t.} y - x = d^{1+\delta}\}$. Notice that

$$\begin{aligned} & \Pr[B = 0 \wedge G_i = 1 | E(a, \mathbf{y}^i)] \\ =\;& \nu'(\{(x, y) \in P : y \in L, y - x = d^{1+\delta}\}) = \tfrac{q(L')}{d}. \end{aligned}$$

Then:

$$\begin{aligned} & \Pr[B = 0 | E(a, \mathbf{y}^i) \wedge (G_i = 1)] \\ =\;& \tfrac{\Pr[B = 0 \wedge (G_i = 1) | E(a, \mathbf{y}^i)]}{\Pr[G_i = 1 | E(a, \mathbf{y}^i)]} \\ =\;& \tfrac{q(L')/d}{1/d} = q(L') \end{aligned}$$

which means that $q(L') \leq 1/40$. Since $|L'| = |L| > n/20$, the first part of corollary 2.16, implies that $h^-(\nu') = h^-(q) > 1/320$, to complete the proof of the claim.

CLAIM 4.5. *For $\gamma' > 0$,* $\Pr[(A, \mathbf{Y}^i) \notin U_{\gamma'}] \leq \frac{\Pr[B = 0 | G_i = 1]}{\gamma'}$.

Proof of Claim 4.5.

$$\begin{aligned} & \Pr[B = 0 | G_i = 1] \\ \geq\;& \Pr[(A, \mathbf{Y}^i) \notin V_{\gamma'} | G_i = 1] \\ & \cdot \Pr[B = 0 | (G_i = 1) \wedge ((A, \mathbf{Y}^i) \notin V_{\gamma'})] \\ \geq\;& \Pr[(A, \mathbf{Y}^i) \notin V_{\gamma'}] \gamma', \end{aligned}$$

where the last inequality uses the independence of $(A, \mathbf{Y}^i)$ and G_i, and the definition of $V_{\gamma'}$. This proves the claim.

Taking $\gamma' = \frac{1}{40}$ in the claim, we have that if $\Pr[B = 0 | G_i = 1] \leq \frac{1}{800}$, then $\Pr[(A, \mathbf{Y}^i) \notin U_{\frac{1}{40}}] \leq \frac{1}{20}$. By Claim 4.2 $\Pr[(A, \mathbf{Y}^i) \in V_{1/10} \cap U_{1/40}] \geq 1/20$. By Claim 4.4 this is also a lower bound on $\Pr[(A, \mathbf{Y}^i) \in W_{1/320}]$. By (4), $I(A, \mathbf{Y}^i : X_i, Y_i) \geq \frac{1}{6400}$. □

Next, we want a counterpart to Lemma 3.2, which will lower bound $H(C)$ in terms of $\sum_i I(A, \mathbf{Y}^i : X_i, Y_i)$. As indicated earlier, unlike Lemma 3.2, this does not seem to follow easily from elementary properties of entropy.

LEMMA 4.6 (MAIN LEMMA). *Let A be first round communication that only depends on $\mathbf{X}$ and μ is the distribution over $(\mathbf{X}, \mathbf{Y})$ as defined earlier in this section. Then*

$$\frac{1}{d}\sum_{i=1}^{d} I(A, \mathbf{Y}^i : X_i, Y_i) = O(\frac{n^2}{d^3}(H(A) + \log d + \log n))$$

This is the key technical result of the paper. We give the proof in the next section.

Comparing the upper bound and lower bound implied by Lemma 4.1 and Lemma 4.6, We obtain

THEOREM 4.7. *Let $\delta < \frac{1}{2}$, $n \geq 100d^{1+\delta}$, $\frac{d^3}{n^2} \gg \log d$. Any one-round protocol Π that $(\frac{1}{1600^2}, \frac{1}{1600^2})$-computes toroidal L^∞ distance threshold decision problem of two length d vectors with $\theta_U/\theta_L \leq d^\delta$ must use $\Omega(\frac{d^3}{n^2})$ bits.*

PROOF. Assume that Π $(\frac{1}{1600^2}, \frac{1}{1600^2})$-computes toroidal L^∞ distance threshold decision problem with threshold d^δ. By Lemma 3.1, Π $(\frac{3}{4}, \frac{1}{800})$-computes g_i for at least $\frac{799}{800}d$ indices $i \in [d]$. By Lemma 4.1, $I(A, \mathbf{Y}^i : X_i, Y_i) \geq \frac{1}{6400}$ for all such i. By Lemma 4.6,

$$H(A) \geq \Omega(\frac{d^2}{n^2}\frac{799}{800}d\frac{1}{6400}) - \log d - \log n = \Omega(\frac{d^3}{n^2}).$$

□

5. THE PROOF OF THE LEMMA 4.6.

The goal is give a good upper bound on $\sum_i I(A, \mathbf{Y}^i : X_i, Y_i)$. We begin with:

LEMMA 5.1.

$$\sum_i I(A, \mathbf{Y}^i : X_i, Y_i) \leq \sum_i I(A : Y_i|\mathbf{Y}^i) + H(A).$$

PROOF. By Lemma 2.7, $I(A, \mathbf{Y}^i : X_i, Y_i) = I(\mathbf{Y}^i : X_i, Y_i) + I(A : X_i, Y_i|\mathbf{Y}^i) = I(A : X_i, Y_i|\mathbf{Y}^i)$ since $\mathbf{Y}^i$ is independent of X_i, Y_i. Recall that $\Delta_i = Y_i - X_i$. Then $I(A : X_i, Y_i|\mathbf{Y}^i) = I(A : \Delta_i, Y_i|\mathbf{Y}^i)$ by Lemma 2.10. Again by Lemma 2.7:

$$\sum_i I(A : \Delta_i, Y_i|\mathbf{Y}^i) = \sum_i I(A : Y_i|\mathbf{Y}^i) + \sum_i I(A : \Delta_i|\mathbf{Y})$$

Apply Lemma 2.11 to the second sum with $X_i \leftarrow \Delta_i$, $Y \leftarrow A$ and $Z \leftarrow \mathbf{Y}$, to get:

$$\sum_i I(A : \Delta_i|\mathbf{Y}) \leq I(A : \Delta|\mathbf{Y}) \leq H(A).$$

□

So it remains to upper bound $\sum_i I(A : Y_i|\mathbf{Y}^i)$ in terms of $H(A)$. Here we need to have some additional definitions. If $\phi : S^d \longrightarrow \mathbb{C}$, we define $\phi^i : S^{d\setminus i} \longrightarrow \mathbb{C}$ by $\phi^i(\mathbf{y}^i) = \frac{1}{|S|}\sum_{y_i \in S}\phi(\mathbf{y})$. For $\mathbf{x}^i \in [n]^{d\setminus i}$, the set $Line(\mathbf{x}^i) = \{\mathbf{z} \in [n]^d : \mathbf{z}^i = \mathbf{x}^i\}$ is called the *line determined by* $\mathbf{x}^i$. For a a possible value of A, let g_a be the function on $[n]^d$ defined by $g_a(\mathbf{y}) = \Pr[A = a|\mathbf{Y} = \mathbf{y}]$. Note that $\Pr[A = a|\mathbf{Y}^i = \mathbf{y}^i] = g_a^i(\mathbf{y}^i)$. By definition of mutual information

$$\begin{aligned} & \textstyle\sum_i I(A : Y_i|\mathbf{Y}^i) = \sum_i H(A|\mathbf{Y}^i) - H(A|\mathbf{Y}) \\ = & \textstyle\sum_a \sum_i \sum_{\mathbf{y}} \Pr[\mathbf{Y} = \mathbf{y}] g_a(\mathbf{y}) \log g_a(\mathbf{y}) - g_a^i(\mathbf{y}^i) \log g_a^i(\mathbf{y}^i) \end{aligned}$$

Our goal is to get the following upper bound of the inner double sum of the above for each value of a.

$$\begin{aligned} & \textstyle\sum_i \sum_{\mathbf{y}} \Pr[\mathbf{Y} = \mathbf{y}][g_a(\mathbf{y}) \log g_a(\mathbf{y}) - g_a^i(\mathbf{y}^i) \log g_a^i(\mathbf{y}^i)] \\ = & O\left(\tfrac{n^2}{d^2}\Pr[A = a] \log \Pr[A = a]\right) \end{aligned}$$

Define $Q(a) = \{\mathbf{x} \in [n]^d : \Pi_1(\mathbf{x}, \mathbf{y}) = a, \forall \mathbf{y} \in [n]^d\}$.

LEMMA 5.2. *Let $p = \Pr[A = a]$. If $\max_{i,\mathbf{x} \in [n]^d} |Q(a) \cap Line(\mathbf{x}^i)| \leq K$ where $K \geq 1$, then*

$$\begin{aligned} & \textstyle\sum_{i=1}^{d} \sum_{\mathbf{y}} \Pr[\mathbf{Y} = \mathbf{y}][g_a(\mathbf{y}) \log g_a(\mathbf{y}) - g_a^i(\mathbf{y}^i) \log g_a^i(\mathbf{y}^i)] \\ = & O(\tfrac{Kn^2}{d^2} p(\log(1/p) + \log d)). \end{aligned}$$

PROOF. Fix a and we simply write g for g_a in this proof. The proof of this lemma involves bounding the left hand side by a quadratic function of $g(\mathbf{y})$, which is then bounded above using Fourier analysis. Fix i,

$$\begin{aligned} & \sum_{\mathbf{y}} \Pr[\mathbf{Y} = \mathbf{y}][g(\mathbf{y}) \log g(\mathbf{y}) - g^i(\mathbf{y}^i) \log g^i(\mathbf{y}^i)] \\ = & \frac{1}{n^{d-1}} \sum_{\mathbf{y}^i} \frac{1}{n} \sum_{y_i} [g(\mathbf{y}) \log g(\mathbf{y}) - g^i(\mathbf{y}^i) \log g^i(\mathbf{y}^i)] \\ = & \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) < p/d} \sum_{y_i} [g(\mathbf{y}) \log g(\mathbf{y}) - g^i(\mathbf{y}^i) \log g^i(\mathbf{y}^i)] \\ & + \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) \geq p/d} \sum_{y_i} [g(\mathbf{y}) \log g(\mathbf{y}) - g^i(\mathbf{y}^i) \log g^i(\mathbf{y}^i)] \end{aligned}$$

By Lemma 2.12, the first sum is bounded by

$$\begin{aligned} & \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) < p/d} \sum_{y_i} [g(\mathbf{y}) \log g(\mathbf{y}) - g^i(\mathbf{y}^i) \log g^i(\mathbf{y}^i)] \\ \leq & \frac{1}{n^{d-1}} \sum_{\mathbf{y}^i, g^i(\mathbf{y}) < p/d} g^i(\mathbf{y}^i) \log(1/g^i(\mathbf{y}^i)) \leq \frac{p}{d} \log \frac{d}{p} \end{aligned}$$

We will use Fourier analysis to upper bound the second sum. The appendix contains needed definitions and a lemma. Let the Fourier expansion of g be $g(\mathbf{x}) = \sum_\alpha \widehat{g}(\alpha)\omega^{\alpha \cdot \mathbf{x}}$. We split the above sum into two parts $g(\mathbf{x}) = g_1(\mathbf{x}) + g_2(\mathbf{x})$, where $g_1(\mathbf{x})$ contains all the terms where weight $|\alpha|$ is less than $(2\log(1/p) + \log d)\frac{2n^2}{d^2}$ and $g_2(\mathbf{x})$ contains all the high weight terms.

Using Lemma 2.13, the second sum is bounded by

$$\begin{aligned} & \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) \geq p/d} \sum_{y_i} [g(\mathbf{y}) \log g(\mathbf{y}) - g^i(\mathbf{y}^i) \log g^i(\mathbf{y}^i)] \\ \leq & \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) \geq p/d} \frac{2}{g^i(\mathbf{y}^i)} \sum_{y_i} (g(\mathbf{y}) - g^i(\mathbf{y}^i))^2 \\ \leq & \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}) \geq p/d} \frac{2}{g^i(\mathbf{y}^i)} \sum_{y_i} 2(g_1(\mathbf{y}) - g_1^i(\mathbf{y}^i))^2 \\ & + \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) \geq p/d} \frac{2}{g^i(\mathbf{y}^i)} \sum_{y_i} 2(g_2(\mathbf{y}) - g_2^i(\mathbf{y}^i))^2 \end{aligned}$$

Let f be the characteristic function of $Q(a)$. Clearly, one may observe that g is a convolution of f with averaging function Λ_D defined in the appendix with $D = \{-1, \ldots, -(d-1), -d^{1+\delta}\}$. For the heavy weight part corresponding to $g_2(\mathbf{x})$, we use the fact $\widehat{g}(\alpha) = n^d \widehat{f}(\alpha)\widehat{\Lambda}(\alpha)$ and the bound over $n^d|\widehat{\Lambda}(\alpha)|$ given by Lemma A.1 with $s = d$ to obtain

$$\begin{array}{ll} & \sum_i \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) \geq p/d} \frac{2}{g^i(\mathbf{y}^i)} \sum_{\mathbf{y}_i} 2(g_2(\mathbf{y}) - g_2^i(\mathbf{y}^i))^2 \\ \leq & \sum_i \frac{2}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) \geq p/d} \frac{d}{p} \sum_{y_i} 2(g_2(\mathbf{y}) - g_2^i(\mathbf{y}^i))^2 \\ \leq & \sum_i \frac{4}{n^d} \sum_{\mathbf{y}^i} \frac{d}{p} \sum_{y_i} (g_2(\mathbf{y}) - g_2^i(\mathbf{y}^i))^2 \\ = & \sum_i \frac{4}{n^d} \sum_{\mathbf{y}^i} \frac{d}{p} \sum_{y_i} (g_2(\mathbf{y}))^2 - (g_2^i(\mathbf{y}^i))^2 \\ = & \frac{4d}{p} \sum_{\alpha, |\alpha| \geq (2 \log 1/p + \log d) \frac{2n^2}{d^2}} |\alpha||\widehat{g}(\alpha)|^2 \\ \leq & \frac{4d}{p} \sum_{\alpha, |\alpha| \geq (2 \log 1/p + \log d) \frac{2n^2}{d^2}} |\alpha||\widehat{f}(\alpha)|^2 e^{-\frac{|\alpha| d^2}{2n^2}} \\ \leq & \frac{4d}{p}(2 \log 1/p + \log d) \frac{2n^2}{d^2} \frac{p^2}{d} \sum_\alpha |\widehat{f}(\alpha)|^2 \\ = & \frac{8n^2}{d^2} p^2 (2 \log 1/p + \log d) \end{array}$$

Before doing analysis on low weight part, we argue that most of L^2 norm of f is in the high weight terms.

According to assumption, $f^i(\mathbf{x}^i) \leq K/n$ whenever $f^i(\mathbf{x}^i) > 0$, we have

$$||f^i||_2^2 \leq \frac{K}{n} ||f||_2^2$$

for each i. If we sum over all i, we obtain

$$\sum_\alpha (d - |\alpha|) |\widehat{f}(\alpha)|^2 = \sum_{i=1}^d ||f^i||_2^2 = \frac{dK}{n} ||f||_2^2$$

This implies

$$||f_1(X)||_2^2 \leq \sum_{\alpha \leq d/2} |\widehat{f}(\alpha)|^2 \leq \frac{2K}{n} ||f||_2^2$$

i.e. L^2 norm of $f_1(\mathbf{x})$ is small.

To get a bound for the low weight part, we use Lemma 2.17 and the fact g_1 is also a convolution of functions f_1 and Λ_D. We have

$$\begin{array}{ll} & \sum_{i=1}^d \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) \geq p/d} \frac{2}{g^i(\mathbf{y}^i)} \sum_{y_i} 2(g_1(\mathbf{y}) - g_1^i(\mathbf{y}^i))^2 \\ \leq & \sum_{i=1}^d \frac{1}{n^d} \sum_{\mathbf{y}^i, g^i(\mathbf{y}^i) \neq 0} \frac{2}{g^i(\mathbf{y}^i)} \sum_{y_i} 2(g_1(\mathbf{y}) - g_1^i(\mathbf{y}^i))^2 \\ \leq & \sum_{i=1}^d \frac{1}{n^d} \sum_{\mathbf{x}^i, f^i(\mathbf{x}^i) \neq 0} \frac{2}{f^i(\mathbf{x}^i)} \sum_{x_i} 2(f_1(\mathbf{x}) - f_1^i(\mathbf{x}^i))^2 \\ \leq & \sum_{i=1}^d \frac{1}{n^d} \sum_{\mathbf{x}^i} 2n \sum_{x_i} 2(f_1(\mathbf{x}) - f_1^i(\mathbf{x}^i))^2 \\ = & 4n \sum_{i=1}^d \frac{1}{n^d} \sum_{\mathbf{x}} (f_1(\mathbf{x}))^2 - (f_1^i(\mathbf{x}^i))^2 \\ = & 4n \sum_{\alpha, |\alpha| \leq (2 \log(1/p) + \log d) \frac{2n^2}{d^2}} |\alpha||\widehat{f}(\alpha)|^2 \\ = & (4n)(2 \log(1/p) + \log d) \frac{2n^2}{d^2} ||f_1||_2^2 \\ \leq & (4n)(2 \log(1/p) + \log d) \frac{2n^2}{d^2} \frac{2K}{n} ||f||_2^2 \\ \leq & O(\frac{Kn^2}{d^2} p(\log(1/p) + \log d) \end{array}$$

Summarizing the bounds we have obtains, we have:

- The first sum is bounded by $\frac{p}{d} \log \frac{d}{p} = \frac{p}{d}(\log \frac{1}{p} + \log d)$
- The heavy weight part of the second sum is bounded by $\frac{8n^2}{d^2} p^2 (2 \log 1/p + \log d)$
- The light weight part of the second sum is bounded by $O(\frac{Kn^2}{d^2} p(\log(1/p) + \log d))$.

Therefore, the whole thing is bounded by $O(\frac{Kn^2}{d^2} p(\log(1/p) + \log d))$. □

Suppose now that K is an upper bound on $\max_a |Q(a) \cap Line(\mathbf{x}^i)|$. Then using Lemma 5.2:

$$\begin{array}{l} \sum_{i=1}^d I(A : Y_i | \mathbf{Y}^i) \\ \quad = \sum_a \sum_i \sum_{\mathbf{y}} \Pr[\mathbf{Y} = \mathbf{y}] \\ \qquad \cdot (g_a(\mathbf{y}) \log g_a(\mathbf{y}) - g_a^i(\mathbf{y}^i) \log g_a^i(\mathbf{y}^i)) \\ \quad = \sum_a O(\frac{n^2}{d^2} \Pr[A = a](\log(1/\Pr[A = a]) + \log d)) \\ \quad = O(\frac{Kn^2}{d^2}(H(A) + \log d)). \end{array}$$

If K is a constant, the lemma is proved. Otherwise, we can apply the above with A replaced by $A' = (A, \sum_{i=1}^d X_i \pmod n)$, which has $K = 1$. Since $H(A') = H(A) + \log n$, this completes the proof of Lemma 4.6.

6. FROM THE TORUS TO THE INTEGER LATTICE

We have now proved the lower bound we wanted, but for the distance $\rho_\odot$ rather than ρ_∞. We now prove a lower bound for latter distance by reducing the approximation problem for the former to it.

Let $T = {\mathbb{Z}_n}^d$ be d-dimensional torus. Let $i = 1, \ldots, d$, $n = \gamma d^{1+\delta}$ with γ to be determined later. Let the distribution $\mu = \nu^d$ on $[n]^d \times [n]^d$ be defined as in Section 4. Let $||x||_n = \min(|x|, n - |x|)$ for $x \in [-n, n]$. Let $\ell = \lceil \frac{n}{2} \rceil$. We define a map P from the torus $T = {\mathbb{Z}_n}^d$ to integer lattice $L = [\ell]^d$ s.t. $P(x_1, x_2, \ldots, x_d) = (||x_1||_n, ||x_2||_n, \ldots, ||x_d||_n)$.

Let $g : [n] \times [n] \longrightarrow \{0, 1, *\}$ and $\widetilde{g} : [\ell] \times [\ell] \longrightarrow \{0, 1, *\}$ s.t.

$$g(x, y) = \begin{cases} 1 & \text{if } ||x - y||_n \geq d^{1+\delta} \\ 0 & \text{if } ||x - y||_n \leq d \\ * & \text{otherwise} \end{cases}$$

and

$$\widetilde{g}(x, y) = \begin{cases} 1 & \text{if } |x - y| \geq d^{1+\delta} \\ 0 & \text{if } |x - y| \leq d \\ * & \text{otherwise} \end{cases}$$

Let $f = \vee_{i=1}^n g_i : [n]^d \times [n]^d \longrightarrow \{0, 1, *\}$ where $g_i(\mathbf{x}, \mathbf{y}) = g(x_i, y_i)$. Let $\widetilde{f} = \vee_{i=1}^n \widetilde{g}_i : [\ell]^d \times [\ell]^d \longrightarrow \{0, 1, *\}$ where $\widetilde{g}_i(\mathbf{x}, \mathbf{y}) = \widetilde{g}(x_i, y_i)$.

Let $\widetilde{\nu}$ be the distribution on $[\ell] \times [\ell]$ induced by ν and map $|| \cdot ||_n$, and let $\widetilde{\mu} = \widetilde{\nu}^d$. We want to show that if a one-round protocol $\widetilde{\Pi}$ does well on $\widetilde{f}$, it induces a one-round protocol that does well on f. we first show that under distribution μ, $\widetilde{f}(P(\cdot), P(\cdot))$ is a good approximation of $f(\cdot, \cdot)$.

Lemma 6.1. *Let $f, \widetilde{f}, P$ be defined as above and $n = \gamma d^{1+\delta}$ Then* $\Pr_\mu[\widetilde{f}(P(\mathbf{X}), P(\mathbf{Y})) \neq 0 \wedge f(\mathbf{X}, \mathbf{Y}) = 0] = 0$ *and* $\Pr_\mu[\widetilde{f}(P(\mathbf{X}), P(\mathbf{Y})) \neq 1 \wedge f(\mathbf{X}, \mathbf{Y}) = 1] \leq \frac{2}{\gamma}$.

Proof. Clearly $||\cdot||_n$ is a metric and satisfies the triangular inequality $|||x||_n - ||y||_n| \leq ||x - y||_n$. Thus $f(\mathbf{X}, \mathbf{Y}) = 0$ implies

$$|||X_i||_n - ||Y_i||_n| \leq ||X_i - Y_i||_n \leq d, \qquad \forall i,$$

i.e. $\widetilde{f}(P(\mathbf{X}), P(\mathbf{Y})) = 0$. Therefore $\Pr_\mu[\widetilde{f}(P(\mathbf{X}), P(\mathbf{Y})) \neq 0 \wedge f(\mathbf{X}, \mathbf{Y}) = 0] = 0$.

Now we want to analyze the case when $f(\mathbf{X}, \mathbf{Y}) = 1$ but $\widetilde{f}(P(\mathbf{X}), P(\mathbf{Y})) \neq 1$. In this case, there must be some $i \in [d]$ s.t $||X_i - Y_i||_n = d^{1+\delta}$ but $|||X_i||_n - ||Y_i||_n| < d^{1+\delta}$. On the other hand, $||X_i - Y_i||_n \neq |||X_i||_n - ||Y_i||_n|$ implies $||X_i||_n + ||Y_i||_n = ||X_i - Y_i||_n = d^{1+\delta}$. This can only happen when $||X_i||_n \leq d^{1+\delta} + d$.

Let B_i denote the event that $g_i(X_i, Y_i) = 1 \wedge \widetilde{g}(X_i, Y_i) \neq 1$. Certainly,

$$\Pr_\mu[B_i] \leq \Pr_\mu[Y_i - X_i \equiv d^{1+\delta} (\text{mod } n) \wedge ||X_i||_n \leq d^{1+\delta}] \leq \frac{2}{\gamma d}.$$

Now we have

$$\Pr{}_{\mu}(\widetilde{f}(P(\mathbf{X}),P(\mathbf{Y}))\neq 1 \wedge f(\mathbf{X},\mathbf{Y})=1)\leq\sum_{i=1}^{d}\Pr{}_{\mu}(B_i)\leq\frac{2}{\gamma}.$$

□

Lemma 6.1 combined with Theorem 4.7 implies

Theorem 6.2. *Let $\delta<\frac{1}{2}$. Any one-round protocol that $(\frac{1}{2\times 1600^2},\frac{1}{2\times 1600^2})$-computes $\widetilde{f}$ with respect to $\widetilde{\mu}$ must use $\Omega(d^{1-2\delta})$ space.*

The straightforward proof is omitted for lack of space.

This theorem also implies space lower bounds for approximating L^p distance for $p>2$.

Corollary 6.3. *Any one-round protocol that approximates L^∞ distance of two length d vectors within factor of d^δ requires $\Omega(d^{1-4\delta})$ space. For $p>2$, Any one-round protocol that approximates L^p distance of two length d vectors within factor of d^δ requires $\Omega(d^{1-\frac{2}{p}-4\delta})$ space.*

7. REFERENCES

[1] Noga Alon, Yossi Matias, and Mario Szegedy. The Space Complexity of Approximating the Frequency Moments. *STOC 1996* 20–29.

[2] L. Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). *FOCS 1986* 337–347.

[3] Ziv Bar-Yossef, T.S. Jayram, Ravi Kumar, and D. Sivakumar. Information Theory Methods in Communication Complexity. to appear in *Proceedings of IEEE Conference on Computational Complexity 2002*.

[4] Ziv Bar-Yossef, T.S. Jayram, Ravi Kumar, and D. Sivakumar. Personal communication.

[5] T.M. Cover and J.A. Thomas. Elements of Information Theory. John Wiley & Sons, Inc., 1991.

[6] Joan Feigenbaum, Sampath Kannan, Martin Strauss, and Mahesh Viswanathan. An Approximate L^1-Difference Algorithm for Massive Data Streams. *FOCS 1999* 501–511.

[7] Jessica H. Fong and Martin Strauss. An Approximate L^p-Difference Algorithm for Massive Data Streams. *STACS 2000* 193–204.

[8] Piotr Indyk. Stable Distributions, Pseudorandom Generators, Embeddings and Data Stream Computation. *FOCS 2000* 189–197.

[9] Eyal Kushilevitz and Noam Nisan. Communication complexity. Cambridge University Press, 1997.

[10] Balasubramanian Kalyanasundaram and Georg Schnitger. The Probabilistic Communication Complexity of Set Intersection (preliminary Version) *Proc. of 2nd Structure in Complexity Theory 1987* 41–49.

[11] W.B. Johnson and J. Lindenstrauss. Extensions of Lipshitz mapping into Hilbert space. *Contemporary Mathematics*, **26**(1984) 189–206.

[12] A. A. Razborov. On the distributed complexity of disjointness *Theoretical Computer Science* **106**(2) (1992) 385–390.

[13] A. C. Yao. Lower bounds by probabilistic arguments. *FOCS 1983* 420–428.

APPENDIX

A. FOURIER ANALYSIS

Let f be a function defined on $[n]^d$ where m is identified with $m\,(\text{mod } n)$ and n a prime. The Fourier expansion of f is defined as $f(\mathbf{x})=\sum_{\alpha\in\mathbb{Z}_n{}^d}\widehat{f}(\alpha)\omega^{\alpha\cdot\mathbf{x}}$ where $\omega=e^{\frac{2\pi i}{n}}$ and $\widehat{f}(\alpha)=\frac{1}{n^d}\sum_{\mathbf{x}\in[n]^d}f(\mathbf{x})\omega^{\alpha\cdot\mathbf{x}}$.

Let f^i be the average of f over the ith coordinate, we have

$$f^i(\mathbf{x}^i)=\frac{1}{n}\sum_{\mathbf{y},\mathbf{y}^i=\mathbf{x}^i}f(\mathbf{y})=\sum_{\alpha\in\mathbb{Z}_n{}^d,\alpha_i=0}\widehat{f}(\alpha)\omega^{\alpha\cdot\mathbf{x}}.$$

Therefore,

$$\widehat{f^i}(\alpha)=\begin{cases}\widehat{f}(\alpha) & \text{if } \alpha_i=0\\ 0 & \text{otherwise}\end{cases}$$

Let $h=f*g$ the convolution of f and g s.t. $h(\mathbf{x})=\sum_{\mathbf{y}\in[n]^d}f(\mathbf{x}-\mathbf{y})g(\mathbf{y})$. Then $\widehat{h}(\alpha)=n^d\widehat{f}(\alpha)\widehat{g}(\alpha)$.

The Fourier transform satisfies the *Parseval identity*

$$\sum_{\alpha\in\mathbb{Z}_n{}^d}\left|\widehat{f}(\alpha)\right|^2=n^d\sum_{\mathbf{x}\in[n]^d}|f(\mathbf{x})|^2$$

Let S be a subset of $[n]$ of size s and Λ_S the function that defines averaging over $S^d\in[n]^d$ as

$$\Lambda_S(\mathbf{x})=\begin{cases}\frac{1}{s^d} & \text{if } \mathbf{x}\in S^d\\ 0 & \text{otherwise}\end{cases}$$

We have the following estimate of $\widehat{\Lambda}_S(\alpha)$.

Lemma A.1. *Let Λ_S be the function defined above and $10<s<n/6$. We have*

$$n^d|\widehat{\Lambda}_S(\alpha)|\leq e^{-\frac{|\alpha|s^2}{2n^2}}$$

where $|\alpha|$ is the weight of α.

Proof. Let $\omega=e^{\frac{2\pi i}{n}}$ First, we observe that the maximum of

$$\left|\frac{1}{s}\sum_{x\in T\subset[n]}\omega^x\right|,$$

where T is any subset of $[n]$ of size s, is achieved when T is a set of s consecutive numbers. In that case, the value is

$$\left|\frac{1}{s}\sum_{0\leq x<s}\omega^x\right|=\frac{\sin(\pi s/n)}{s\sin(\pi/n)}\leq e^{-\frac{s^2}{2n^2}}.$$

The Fourier transformation gives

$$\begin{aligned}n^d\widehat{\Lambda}_S(\alpha)\quad &=\textstyle\sum_{\mathbf{x}\in[n]^d}\Lambda_S(\mathbf{x})\omega^{\alpha\cdot\mathbf{x}}\\ &=\textstyle\prod_{i=1}^{d}\frac{1}{s}\sum_{x\in S}\omega^{\alpha_i x}\\ &=\textstyle\prod_{i=1,\alpha_i\neq 0}^{d}\frac{1}{s}\sum_{y\in S(\alpha_i)}\omega^{y}\end{aligned}$$

where $S(\alpha_i)=\{\alpha_i x\,(\text{mod } n): x\in S\}$. Therefore,

$$n^d|\widehat{\Lambda}_S(\alpha)|=\prod_{\substack{i=1\\ \alpha_i\neq 0}}^{d}\left|\frac{1}{s}\sum_{y\in S(\alpha_i)}\omega^y\right|\leq\prod_{\substack{i=1\\ \alpha_i\neq 0}}^{d}e^{-\frac{s^2}{2n^2}}=e^{-\frac{|\alpha|s^2}{2n^2}}$$

□