Lower bounds for Edit Distance and Product Metrics via Poincaré-Type Inequalities

Alexandr Andoni[*]
Princeton U./CCI
andoni@mit.edu

T.S. Jayram
IBM Almaden
jayram@almaden.ibm.com

Mihai Pătraşcu[†]
AT&T Labs
mip@alum.mit.edu

Abstract

We prove that any sketching protocol for edit distance achieving a constant approximation requires nearly logarithmic (in the strings' length) communication complexity. This is an exponential improvement over the previous, doubly-logarithmic, lower bound of [Andoni-Krauthgamer, FOCS'07]. Our lower bound also applies to the Ulam distance (edit distance over non-repetitive strings). In this special case, it is polynomially related to the recent upper bound of [Andoni-Indyk-Krauthgamer, SODA'09].

From a technical perspective, we prove a direct-sum theorem for sketching product metrics that is of independent interest. We show that, for any metric X that requires sketch size which is a sufficiently large constant, sketching the max-product metric $\ell_\infty^d(X)$ requires $\Omega(d)$ bits. The conclusion, in fact, also holds for arbitrary two-way communication. The proof uses a novel technique for information complexity based on Poincaré inequalities and suggests an intimate connection between non-embeddability, sketching and communication complexity.

1 Introduction

The edit distance, as the most natural similarity metric between two strings, shows up in algorithmic questions of many different flavors:

computation: How fast can we estimate the edit distance between two large strings?

nearest neighbor: Can we preprocess a set of strings using little space, such that database elements close to a query string can be retrieved efficiently?

communication: If two parties have similar versions of a document, how little can they communicate to estimate the difference between their versions?

Variations on these question are ubiquitous. Applications range from computational biology, to allowing programmers to synchronize and archive code changes, to helping users who cannot spel.

Communication complexity. The main result in this paper is an improved lower bound for the communication complexity of edit distance. Assume that the two strings come from $\{0,1\}^d$. In FOCS'07, Andoni and Krauthgamer [AK07] showed that, to approximate edit distance within any constant factor, the two parties need to communicate $\Omega(\log\log d)$ bits. All throughout the paper, by "approximating edit distance" we mean the decision version of the problem: where two players are to decide whether the strings are at edit distance at most R or at least αR for some threshold R and approximation α.

Here, we exponentially improve their result to show that a constant factor approximation requires $\Omega(\frac{\log d}{\log\log d})$ bits of communication. In general, we obtain that with c bits of communication, the two parties cannot approximate edit distance up to a factor better than $\Omega(\frac{\log d}{\log\log d}/c)$.

For the general edit distance, there seems to be no consensus on how much communication should be necessary. The current state of the upper bounds is certainly dismal: there is no sublinear protocol achieving constant approximation.

[*] Work done while at MIT.

[†] Work done while at IBM Almaden.

However, much better results are known for a restricted subset of the edit distance, called the Ulam distance. Formally, the Ulam distance is defined similarly to the edit distance, except that the two strings are requires to be permutations on $[d]$. This is meant to capture the (arguably practical) scenario of nonrepetitive edit distance: each long enough block of characters appears uniquely in each string.

Our lower bound, as well as the previous result of [AK07], holds even in the restricted case of Ulam distance. On the upper bound front, a recent paper of Andoni, Indyk, and Krauthgamer [AIK09] from SODA'09 gave a protocol with $O(\log^6 d)$ bits of communication, which approximates the edit distance to within some fixed constant. Thus, our lower bound is polynomially close to the best known upper bound. In fact, we conjecture that our lower bound is *tight* for the Ulam distance, up to doubly-logarithmic factors.

To prove this result, we design a new communication complexity technique which is geared towards products spaces. Using the powerful information complexity paradigm for communication complexity [CSWY01, BJKS04], we reduce this problem to a direct sum question for communication complexity. We introduce a novel technique for proving information complexity lower bounds based on *Poincaré-type* inequalities. The latter are an indispensable tool in obtaining non-embeddability results [Mat02] and our result demonstrates that they are also intimately connected with communication complexity.

In some sense, our lower bound is the best possible result without exhibiting a separation between the edit distance and its special case, the Ulam distance. Such a separation appears like a significant milestone lying ahead.

Metric embeddings. Though our results focus on the communication problem, they are significant in the broader context of edit distance questions.

The most promising current attack on the edit distance is through embedding it into simpler metrics. An embedding is a mapping from strings to some normed space X. The embedding is said to have distortion α if, for any strings x, y, $\mathrm{ed}(x, y) \le \|f(x) - f(y)\|_X \le \alpha \cdot \mathrm{ed}(x, y)$. Then, if the target metric admits a fast (approximate) nearest neighbor solution, one can immediately obtain a nearest neighbor solution for edit distance, where the approximation is further multiplied by α. A similar statement holds for communication protocols as well.

To demonstrate the power of this idea, one only needs to mention that the state of the art on all fronts comes from metric embeddings. In STOC'05, Ostrovski and Rabani [OR05] described an embedding of edit distance into the space ℓ_1 with distortion $2^{O(\sqrt{\log d \log\log d})}$. This is currently the best approximation for both a nearest neighbor data structure of polynomial space as well as for estimating the edit distance in $d^{1+o(1)}$ time. The latter result was achieved by Andoni and Onak [AO09] only recently and requires additional ideas since it is unknown whether the embedding can be implemented in sub-quadratic time. The embedding also yields the best known communication protocol with, say, $\mathrm{polylog}(d)$ communication.

Given this success, proving *non-embeddability* results became an important direction. The question of (non-)embeddability of edit distance into ℓ_1 appears on the Matoušek's list of open problems [Mat07], as well as in Indyk's survey [Ind01]. From the first non-embeddability bound of $3/2$ of [ADG+03], the bound has been improved to $\Omega(\log^{0.5-o(1)} d)$ by Khot and Naor [KN06], and later to the state-of-the-art $\Omega(\log d)$ bound of Krauthgamer and Rabani [KR06]. Later, Andoni and Krauthgamer [AK07] prove an $\Omega(\frac{\log d}{\log\log d})$ lower bound for embedding into more general classes of spaces, which includes ℓ_1.

Recent evidence, however, shows these lower bounds are unsatisfactory from a qualitative perspective. Traditionally, researchers have searched for embeddings into classic spaces from real analysis, such as the Manhattan norm ℓ_1, the Euclidean norm ℓ_2, or perhaps ℓ_∞. However, there seems to be no inherent reason to restrict ourselves to such mathematically "nice" spaces[1]. Indeed, one can consider other "target spaces", with the only restriction that the target metric is still *computationally* nice, in the sense of having efficient nearest neighbor data structures, or fast communication protocols.

The first compelling examples of this direction are given by Andoni, Indyk, and Krauthgamer [AIK09] in SODA'09. They show that the Ulam metric can be embedded into the rather unusual metric, $\ell_1\left(\ell_\infty\left((\ell_1)^2\right)\right)$, with constant distortion. As a consequence, they obtain state-of-the-art results regarding the Ulam metric:

- a nearest-neighbor data structure of $O(n^{1+\varepsilon})$ space with $\mathrm{poly}(d, \log n)$ query time and $O(\log\log d)$ approximation.
- a communication protocol for estimating the Ulam norm with $O(\log^6 d)$ bits of communication and an $O(1)$ approximation guarantee.

Let us look closer at their target space: $\ell_1\left(\ell_\infty\left((\ell_1)^2\right)\right)$. This distance can be computed

[1] At least for our applications at hand. We note that, in other applications, such as sparsest cut problem, there is a general interest of embedding finite metric spaces into, say, ℓ_1.

using a combination of the standard ℓ_1 and ℓ_∞ norms. The inner term, $(\ell_1)^2$, is the square of the Manhattan norm (which, technically speaking, is not itself a metric). To define the distance in the space $\ell_\infty\left((\ell_1)^2\right)$, imagine the two points as two-dimensional matrices and compute the difference matrix. On each row, the $(\ell_1)^2$ norm is applied reducing the matrix to a vector. On the resulting vector, we apply the ℓ_∞ norm, yielding the $\ell_\infty\left((\ell_1)^2\right)$ norm. The final distance is obtained by iterating this again, on three-dimensional arrays, with ℓ_1 on the outside. Metrics obtained through this composition process are called *product metrics.* We note that the dimension of the ℓ_∞ component is only $O(\log d)$, which is an important feature as ℓ_∞ is metrically the hardest and it governs the performance of the nearest neighbor search and communication protocol for Ulam distance.

Note however, the success of product metrics casts doubt on the relevance of the statements on non-embeddability results into classic spaces such as ℓ_1 or $(\ell_1)^2$. On the other hand, proving lower bounds for embedding into metrics such as $\ell_1\left(\ell_\infty\left((\ell_1)^2\right)\right)$ seems like a fool's game, given the large number of possible variations.

The proper attack, we believe, is to switch from inherently geometric statements of non-embeddability, and replace them with an information theoretic approach, of a more computer-science flavor. The metrics that we may want to embed into have, almost by definition, low communication protocols for distance estimation (since we only care to embed into "computationally efficient" metrics). Thus, a communication lower bound immediately implies non-embeddability into a large class of metrics of interest.

For example, we obtain that Ulam metric does not embed with constant distortion into the spaces $\ell_\infty(M)$, where M can be any of ℓ_1, $(\ell_1)^2$, ℓ_2, or $(\ell_2)^2$, and ℓ_∞ has dimension $k = o\left(\frac{\log d}{\log\log^{O(1)} d}\right)$. This follows from the fact that these metrics have communication complexity of $O(k \log k)$ for constant approximation (via the standard sketches for ℓ_1 and ℓ_2 [KOR00]).

Technical contribution. Our technical contribution is a new direct sum result in communication complexity, geared towards metrics.

Recall that a *distance* (or *dissimilarity*) function [DD06] on a space $\mathcal{X}$ is a non-negative function d on $\mathcal{X}^2$ that is both symmetric ($d(x,y) = d(y,x)$) and reflexive ($d(x,x) = 0$). We consider "distance" functions g that are also decision problems, i.e. range$(g) = \{0,1\}$. In the usual application, g corresponds to a *distance threshold estimation problem (DTEP)* of distinguishing instances of distance at most R or at least αR for some threshold R and approximation α. Note that g is a partial function, i.e. dom$(g) \subseteq \mathcal{X}^2$.

Suppose the sketch complexity of g is a sufficiently large constant. Let the function $f(\mathbf{x},\mathbf{y}) = \bigvee_{i=1}^{n} g(x_i, y_i)$. We show that the communication complexity of f is $\Omega(n)$. Such a result is somewhat easy to show if g were defined on all of $\mathcal{X}^2$. This is because it is possible to identify 2-point sets $A, B \subseteq \mathcal{X}$ such that the restriction g to $A \times B$ is isomorphic to the AND function on two bits. (Here is a proof sketch: in the distance matrix of g, the diagonal entries are all equal to 0. Since g has large communication complexity, $g(x_1, y_1) = 1$ for some $x_1 \neq y_1$. Moreover, if $g(x,y) = 1$ for all $x \neq y$, then g is the non-equality function whose communication complexity is $O(1)$. Therefore $g(x_0, y_0) = 0$ for some $x_0 \neq y_0$. Set $A = \{x_0, x_1\}$ and $B = \{y_0, y_1\}$. The argument can also be extended to arbitrary total Boolean functions.) Then, f has a copy of the disjointness function embedded inside it for which the $\Omega(n)$ bound is a classical result in communication complexity [KS92, Raz92]. Things are quite different when g is a partial function. For example, let $g(x,y) = 0$ if $|x-y| \leq 1$ and $|x-y| > 3$, where $0 \leq x, y \leq 4$. By triangle inequality, any 2×2 sub-matrix in the distance matrix of g in which all 4 points are legal inputs for g cannot be isomorphic to AND.

We tackle this problem by resorting to information complexity which, informally speaking, characterizes the minimum amount of information about the inputs that must be revealed by the players in a valid protocol. Introduced as a formal measure first by Chakrabarti, Shi, Wirth, and Yao [CSWY01] for two-party simultaneous protocols, this was later extended to handle non-product distributions for general protocols by Bar-Yossef, Jayram, Ravi Kumar and Sivakumar [BJKS04]. Appropriately, both these papers used this measure to prove direct sum theorems. In order to apply this methodology to our setting (in particular, the Bar-Yossef *et al.* approach), we are faced with two issues: (1) how to define the hard distribution and (2) how to prove an information complexity lower bound. For the former, the sketch complexity of g suggests using the distribution given by Yao's Lemma. But is not clear how to use it to prove an information complexity bound.

We introduce a new technique for proving information complexity bounds based on certain type of inequalities that arise in functional analysis called *Poincaré-type* inequalities. In metric embeddings, such inequalities have been an indispensable tool in obtaining non-embeddability results [Mat02] and in some cases are equivalent to non-embeddability in $(\ell_2)^2$ [LLR95]. In our case, we consider Poincaré-type inequalities for distance threshold estimation. Our main technical result shows how such inequalities can be used to obtain strong information complexity lower bounds. A special case of this argument was considered by Bar-Yossef *et al.* [BJKS04] for $\mathbb{R}$ (under ℓ_1), and by Jayram and Woodruff [JW09] for the Hamming cube.

To complete the argument, we need to prove appropriate Poincaré-type inequalities for distance threshold estimation. Indeed, such inequalities are known for special cases, e.g., the above results for $\mathbb{R}$ and the Hamming cube are based on such inequalities, but there is no general characterization. We give a characterization of a class of Poincaré-type inequalities for any DTEP g: in fact it is equivalent to the fact that g has sketch complexity at least some large constant! (Recall that this was the assumption with which we started.) Neither direction is technically hard and in fact, one of them is implied by an earlier argument in [AK07].

Organization. We start the presentation by reviewing communication complexity and the notion of *information complexity* of a communication protocol, as developed in [BJKS04]. This will be presented in Section 2. Next, in Section 3, we show how, using the direct sum on information cost theorem, one can obtain communication complexity lower bounds from certain Poincaré-type inequality on a metric. Further, in Section 4, we show that we may obtain this Poincaré-type inequality from a more standard lower bound on constant-sized protocols. Combining these two steps, together with the lower bound of [AK07] for constant-size protocols, we will obtain our main result: improved communication lower bounds for edit and Ulam metrics.

2 Preliminaries

We consider the two-party communication model where Alice gets an input in $\mathcal{X}$ and Bob gets an input in $\mathcal{Y}$. Their goal is to solve some communication problem f, defined on a *legal* subset $\mathcal{L} \subseteq \mathcal{X} \times \mathcal{Y}$, by sending messages to each other. In other words, f is a partial function. We adopt the standard *blackboard* model where the messages are all written on a shared medium. A *protocol* $\mathcal{P}$ specifies the rules for Alice and Bob to send messages to each other for all inputs in $\mathcal{X} \times \mathcal{Y}$. The protocol is said to be *simultaneous* or *sketch*-based if Alice and Bob each write just a single message based only on their respective inputs. The sequence of messages written on the blackboard is called the *transcript*. The maximum length of the transcript (in bits) over all inputs is the *communication cost* of the protocol $\mathcal{P}$. The output of the protocol (which need not be part of the transcript) is given by a referee looking only at the transcript and not the inputs. The protocol is allowed to be randomized in which each player, as well as the referee, has *private* access to an unlimited supply of random coins. The protocol solves the communication problem f if the answer on any input $(x, y) \in \mathcal{L}$ equals $f(x, y)$ with probability at least $1 - \delta$. Unless mentioned explicitly, δ will be a small constant and such protocols will be called as *correct* protocols. Note that the protocol itself is legally defined for all inputs in $\mathcal{X} \times \mathcal{Y}$ although no restriction is placed on the answer of the protocol outside the legal set $\mathcal{L}$. The *communication complexity* of f, denoted by $\mathrm{CC}(f)$, is the minimum communication cost of a correct protocol for f.

In the next two sections, we will review the *information complexity* paradigm for proving communication lower bounds via *direct sum* arguments, as developed in [BJKS04].[2]

2.1 Information Complexity

Notation. Random variables will be denoted by upper case Roman or Greek letters, and the values they take by (typically corresponding) lower case letters. Probability distributions will be denoted by lower case Greek letters. A random variable X with distribution μ is denoted by $X \sim \mu$. If μ is the uniform distribution over a set $\mathcal{W}$, then this is also denoted as $X \in_R \mathcal{W}$. Vectors will be denoted in bold case.

Definition 2.1. *A distribution μ over $\mathcal{X} \times \mathcal{Y}$ is* partitioned *by η if there exists a joint probability space (X, Y, F) such that $(X, Y) \sim \mu$, $F \sim \eta$, and (X, Y) are jointly independent conditioned on F i.e.* $\Pr(X, Y \mid F) = \Pr(X \mid F) \cdot \Pr(Y \mid F)$. □

Definition 2.2. *Let $\mathcal{P}$ be a randomized private-coin protocol on the input domain $\mathcal{X} \times \mathcal{Y}$ and let its random coins be denoted by the random variable R. Suppose μ is a distribution over $\mathcal{X} \times \mathcal{Y}$ partitioned by η in some joint probability space (X, Y, F) where $(X, Y) \sim \mu$ and $F \sim \eta$. Extend this to a joint probability space over (X, Y, F, R) such that (X, Y, F) is independent of R. Now, let $\Pi = \Pi(X, Y, R)$ be the random variable denoting the transcript of the protocol, where the randomness is* both *over the input distribution and the random coins of the protocol $\mathcal{P}$. The* (conditional) information cost *of $\mathcal{P}$ under (μ, η) is defined to be $\mathrm{I}(X, Y : \Pi \mid F)$, i.e., the (Shannon) conditional mutual information between (X, Y) and Π conditioned on F.*

The information complexity *of a problem f under (μ, η), denoted by $\mathrm{IC}_\mu(f \mid \eta)$, is defined to be the minimum information cost of a correct protocol for f under (μ, η).* □

[2] The methodology also applies to general multi-party number-in-hand communication which is not needed here.

Since $\mathrm{I}(X, Y : \Pi \mid D) \le H(\Pi) \le |\Pi|$, it follows that $\mathrm{CC}(f) \ge \mathrm{IC}_\mu(f \mid \eta)$.

2.2 Direct Sum

Suppose f is a 0-1 decision problem that can be expressed in terms of a simpler problem g:

$$f(\mathbf{x}, \mathbf{y}) \triangleq \bigvee_{j=1}^{n} g(x_i, y_i).$$

Let g be defined on a set $\mathcal{L} \subseteq \mathcal{X} \times \mathcal{Y}$. The legal inputs for f are pairs $(\mathbf{x}, \mathbf{y})$ such that $(x_i, y_i) \in \mathcal{L}$ for all i. Identify $\mathcal{X}^n \times \mathcal{Y}^n$ with $(\mathcal{X} \times \mathcal{Y})^n$ so that the set of legal pairs equals $\mathcal{L}^n$. Any communication protocol for f is well-defined for all inputs in $\mathcal{X}^n \times \mathcal{Y}^n$.

To relate the information complexity of f to that of g, we proceed as follows. Suppose ν is a distribution over $\mathcal{L}$ partitioned by ζ. We say that ν is *collapsing* if its support is contained in $g^{-1}(0)$. Define the distributions $\mu \triangleq \nu^n$ and $\eta \triangleq \zeta^n$ by taking the n-fold product of ν and ζ, respectively. μ is partitioned by η in some joint probability space of $(\mathbf{X}, \mathbf{Y}, \mathbf{F})$ where:

- for every i, $(X_i, Y_i) \sim \nu$ and $F_i \sim \zeta$;
- the triples over all i of (X_i, Y_i, F_i) are jointly independent of each other.

Proposition 2.1. (Direct Sum [BJKS04]) *Let $\mathcal{L} \subseteq \mathcal{X} \times \mathcal{Y}$ be the domain of a decision function g. Define $f(\mathbf{x}, \mathbf{y}) \triangleq \bigvee_{j=1}^{n} g(x_i, y_i)$. Let ν be a collapsible distribution over $\mathcal{L}$ partitioned by η. Then,*

$$\mathrm{CC}(f) \ge \mathrm{IC}_{\nu^n}(f \mid \zeta^n) \ge n \cdot \mathrm{IC}_\nu(g \mid \zeta).$$

Consequently, the goal will be to prove a lower bound on the information complexity of g. For the applications considered in this paper, the information complexity of g will be an $O(1)$ quantity. Here, it will be fruitful to transition from information measures to statistical divergences, which is the subject of the next section.

2.3 Hellinger Distance

Notation. Let $\|\cdot\|$ denote the standard ℓ_2 norm.

Fix a protocol $\mathcal{P}$, and let $\pi(u, v)$ denote the probability distribution over transcripts induced by $\mathcal{P}$ on input (u, v), where the randomness is over the private coins of $\mathcal{P}$. Let $\pi(u, v)_\tau$ denote the probability that the transcript equals τ. Viewing $\pi(u, v)$ as an element of ℓ_1, note that it belongs to the unit simplex since $\sum_\tau \pi(u, v)_\tau = 1$.

Let $\psi(u, v) \in \ell_2$ be obtained via the square-root map $\pi(u, v) \mapsto \psi(u, v) = \sqrt{\pi(u, v)}$. This means $\psi(u, v)_\tau = \sqrt{\pi(u, v)_\tau}$ for all τ. Now, $\|\psi(u, v)\| = \sum_\tau \pi(u, v)_\tau = 1$, and so $\psi(u, v) \in \mathbb{S}_+$, where $\mathbb{S}_+$ denotes the unit sphere in ℓ_2 restricted to the non-negative orthant. Following [Jay09], $\psi(u, v)$ is called the *transcript wave function* of (u, v) in $\mathcal{P}$.

Definition 2.3. (Hellinger Distance) *The* Hellinger distance *between $\psi_1, \psi_2 \in \mathbb{S}_+$ is a scaled Euclidean distance defined as*

$$h(\psi_1, \psi_2) \triangleq \tfrac{1}{\sqrt{2}} \|\psi_1 - \psi_2\| \qquad \square$$

The scaling ensures that Hellinger distance is always between 0 and 1. In this paper, we will mostly be dealing with the *square* of the Hellinger distance, for which the following notation is not only convenient but also emphasizes the geometric nature of Hellinger distance.

Notation. Let $\hat{\|}\psi\hat{\|} \triangleq \frac{1}{2}\|\psi\|^2$ for $\psi \in \ell_2$ so that $h^2(\psi_1, \psi_2) = \hat{\|}\psi_1 - \psi_2\hat{\|}$. □

We summarize the relevant properties of Hellinger distance that are needed in this paper in Appendix A.

3 Information Complexity via Poincaré-type Inequalities

In this section we present a new technique for proving information complexity lower bounds. Fix a decision problem $g : \mathcal{L} \to \{0, 1\}$, where $\mathcal{L} \subseteq \mathcal{X} \times \mathcal{X}$, that is also a distance function on $\mathcal{L}$. Formally, g is symmetric—$(x, y) \in \mathcal{L} \iff (y, x) \in \mathcal{L}$ and $g(x, y) = g(y, x)$ for all $(x, y) \in \mathcal{L}$—and reflexive—$(x, x) \in \mathcal{L}$ for all $x \in \mathcal{X}$ and $g(x, x) = 0$ for all $x \in \mathcal{X}$.

Suppose that there are two distributions η_0 on $g^{-1}(0)$ and η_1 on $g^{-1}(1)$ with the following property. For some fixed $\alpha > 0$ and $\beta \ge 0$, the following inequality holds that for all vector-valued functions $\rho : \mathcal{X} \to \mathbb{S}_+$:

$$\mathbb{E}_{(x,y)\sim\eta_0} \hat{\|}\rho(x) - \rho(y)\hat{\|} \ge \alpha \cdot \mathbb{E}_{(x,y)\sim\eta_1} \hat{\|}\rho(x) - \rho(y)\hat{\|} - \beta. \tag{3.1}$$

Call the above an (α, β)-Poincaré inequality for g with respect to η_0 and η_1.

Theorem 3.1. *Let $g : \mathcal{L} \to \{0, 1\}$ be a distance function for some $\mathcal{L} \subseteq \mathcal{X}^2$ that satisfies an (α, β)-Poincaré inequality with respect to distributions η_0 on $g^{-1}(0)$ and η_1 on $g^{-1}(1)$. Then, there exists a collapsible distribution ν partitioned by some distribution ζ such that*

$$\mathrm{IC}_\nu(g \mid \zeta) \ge \frac{\alpha(1 - 2\sqrt{\delta}) - \beta}{4}$$

Proof. Let the random variables (U, V, S, T) be defined jointly as follows:

- $S \in_R \{\textsc{a}, \textsc{b}\}$ and $T \sim \eta_0$.

- Suppose $T = (u, v) \in \mathcal{X}^2$. Then we have two cases. If $S = \textsc{a}$, then $U \in_R \{u, v\}$ and $V = v$. Otherwise $S = \textsc{b}$, and here $U = u$ and $V \in_R \{u, v\}$.

We let ν be the distribution of (U, V) and ζ be the distribution of (S, T). It follows that ν is partitioned by ζ. Since $(x, x) \in g^{-1}(0)$ for all x, the support of ν is contained in $g^{-1}(0)$, so ν is collapsible.

Let Π denote the transcript random variable in a correct protocol for g. We bound the information cost of this protocol as follows. Let $Q(s, u, v)$ denote the event "$S = s \wedge T = (u, v)$" for $s \in \{\textsc{a}, \textsc{b}\}$ and $(u, v) \in \mathcal{X}^2$.

$$\begin{aligned}
&\mathrm{I}(U, V : \Pi \mid S, T) \\
&= \sum_{\substack{s \in \{\textsc{a},\textsc{b}\} \\ (u,v) \in \mathcal{X}^2}} \Pr[Q(s, u, v)] \cdot \mathrm{I}(U, V : \Pi \mid Q(s, u, v)) \\
&= \tfrac{1}{2} \cdot \mathbb{E}_{(u,v)\sim\eta_0} \mathrm{I}(U, V : \Pi \mid Q(\textsc{a}, u, v)) + \\
&\qquad\qquad \mathrm{I}(U, V : \Pi \mid Q(\textsc{b}, u, v)) \\
&\geq \tfrac{1}{2} \cdot \mathbb{E}_{(u,v)\sim\eta_0} \hat{\|}\psi(u, u) - \psi(u, v)\hat{\|} + \hat{\|}\psi(u, v) - \psi(v, v)\hat{\|}
\end{aligned}$$

where the last inequality follows by applying the Mutual-information-to-Hellinger-distance property of Proposition A.1. Since $\hat{\|}\cdot\hat{\|}$ is the square of a metric, applying Cauchy-Schwarz followed by the triangle inequality yields:

$$\mathrm{I}(U, V : \Pi \mid S, T) \geq \tfrac{1}{4} \cdot \mathbb{E}_{(u,v)\sim\eta_0} \hat{\|}\psi(u, u) - \psi(v, v)\hat{\|}$$

Now, we apply the Poincaré-type inequality satisfied by g (Equation (3.1)) by setting $\rho(x) = \psi(x, x)$ for all x. We obtain:

$$\begin{aligned}
&\mathrm{I}(U, V : \Pi \mid S, T) \\
&\geq \tfrac{1}{4} \cdot \left(\alpha \cdot \mathbb{E}_{(u,v)\sim\eta_1} \hat{\|}\psi(u, u) - \psi(v, v)\hat{\|} - \beta\right)
\end{aligned} \tag{3.2}$$

For the expression within the expectation in the RHS, fix an (u, v) in the support of η_1. By the Pythagorean property of Proposition A.1,

$$\begin{aligned}
&\hat{\|}\psi(u, u) - \psi(v, v)\hat{\|} \\
&\geq \tfrac{1}{2} \cdot \left(\hat{\|}\psi(u, u) - \psi(u, v)\hat{\|} + \hat{\|}\psi(v, u) - \psi(v, v)\hat{\|}\right)
\end{aligned}$$

Since $g(u, v) = g(v, u) = 1$ for (u, v) in the support of η_1, and $g(u, u) = g(v, v) = 0$, we can apply the Soundness property of Proposition A.1 in the above inequality to get:

$$\hat{\|}\psi(u, u) - \psi(v, v)\hat{\|} \geq 1 - 2\sqrt{\delta}$$

Substituting this bound in (3.2), we get

$$\mathrm{I}(U, V : \Pi \mid S, T) \geq \frac{\alpha(1 - 2\sqrt{\delta}) - \beta}{4}$$

Combining the above main theorem and the direct sum theorem, Theorem 2.1, we obtain the following:

Corollary 3.1. *Let g be a 0-1 distance function that satisfies an (α, β)-Poincaré inequality. Let $f(\mathbf{x}, \mathbf{y}) = \bigvee_{i=1}^{n} g(x_i, y_i)$. Then, $\mathrm{CC}(f) \geq cn/4$ where $c = \alpha(1 - 2\sqrt{\delta}) - \beta$.*

Example 3.1. *In [BJKS04], the authors prove a communication lower bound for estimating ℓ_∞ via an information complexity and direct sum paradigm. The function g that they consider is defined as follows. Let $u, v \in [0, m]$; $g(u, v) = 0$ if $|u - v| \leq 1$ and $g(u, v) = 1$ if $|u - v| = m$. The authors show an $\Omega(1/m^2)$ information complexity lower bound for this problem. We can obtain the same bound via Corollary 3.1.*

Consider any mapping $\rho : [0, m] \to \mathbb{S}_+$. By Cauchy-Schwarz and triangle inequality,

$$\mathbb{E}_{u \in_R [0..m-1]} \hat{\|}\rho(u) - \rho(u + 1)\hat{\|} \geq \frac{1}{m^2} \cdot \hat{\|}\rho(0) - \rho(m)\hat{\|},$$

which is just a $(1/m^2, 0)$-Poincaré inequality. By Corollary 3.1, we obtain an $\Omega(1/m^2)$ information complexity bound.

Example 3.2. *Consider the Hamming cube $H = \{0, 1\}^d$ and its associated metric $|\cdot|$. In [JW09], the authors define a function g using H as follows. Let $x, y \in \{0, 1\}^d$; $g(x, y) = 0$ if $|x - y| \leq 1$ and $g(x, y) = 1$ if $|x - y| = d$. The authors show an $\Omega(1/d)$ information complexity lower bound for this problem and use it to derive space lower bounds for estimating cascaded norms in a data stream.*

Consider any mapping $\rho : H \to \mathbb{S}_+$. Let η_0 denote the uniform distribution on the edges of H, i.e., pairs (u, v) such that $|u - v| = 1$. Let η_1 denote the distribution on the diagonals of H, i.e., pairs $(u, \overline{u})$ where $\overline{u}$ denotes the bit-wise complement of u. The well-known "short-diagonals" property [Mat02] of the Hamming cube states that

$$\mathbb{E}_{(u,v)\sim\eta_0} \hat{\|}\rho(u) - \rho(v)\hat{\|} \geq \frac{1}{d} \cdot \mathbb{E}_{(u,v)\sim\eta_1} \hat{\|}\rho(u) - \rho(v)\hat{\|}.$$

This is a $(1/d, 0)$-Poincaré inequality, which by Corollary 3.1 yields an $\Omega(1/d)$ information complexity bound.

4 Poincaré-type Inequalities via Hardness of Sketching

Suppose g is a 0-1 distance function whose sketch complexity is at least some large constant C for protocols with error probability at most 1/3. We show that this implies a Poincaré-type inequality for g under a suitable distribution derived from the hardness of g via

Yao's lemma. This result can be interpreted as a converse to a result in [AK07], where the authors show that a Poincaré-type inequality implies a sketching lower bound. Together with the results of the previous section, this will enable us to derive new communication complexity lower bounds.

Let $\varepsilon = .1$, and suppose $C = \Omega(1/\varepsilon^4 \cdot \log^2 1/\varepsilon)$. First we note that any protocol for g with success probability $\geq \frac{1}{2} + \varepsilon/3$ has size at least $C' = \Omega(C \cdot \varepsilon^2)$.

By Yao's principle, there exists a hard distribution ψ for protocols of size $< C'$. We decompose the distribution ψ into two distributions with distinct support: for $i \in \{0, 1\}$, we define distribution $(x, y) \sim \eta_i$ to be the distribution ψ conditioned on $g(x, y) = i$. Let $p_i = \Pr_\psi[g(x, y) = i]$ for $i \in \{0, 1\}$.

Claim 4.1. *For any vector-valued function $\rho : \mathcal{X} \to \mathbb{S}_+$, we have that*

$$|\mathbb{E}_{(x,y)\sim\eta_1} \|\rho(x) - \rho(y)\|^2 - \mathbb{E}_{(x,y)\sim\eta_0} \|\rho(x) - \rho(y)\|^2| < \varepsilon. \tag{4.3}$$

Proof. Note that $p_0, p_1 \geq \frac{1}{2} - \varepsilon/3$ (otherwise, there exists a trivial 1-bit protocol with success probability at least $\frac{1}{2} + \varepsilon/3$).

For the sake of contradiction assume Equation (4.3) does not hold, and, w.l.o.g.,

$$\mathbb{E}_{(x,y)\sim\eta_1} \|\rho(x) - \rho(y)\|^2 - \mathbb{E}_{(x,y)\sim\eta_0} \|\rho(x) - \rho(y)\|^2 \geq \varepsilon.$$

Then, we show how to design a simultaneous-message protocol of size $O(1/\varepsilon^2 \cdot \log^2 1/\varepsilon) < C'$ that has success probability $\geq \frac{1}{2} + \varepsilon/3$.

Namely, we take a randomized protocol that estimates the quantity $\|\rho(x) - \rho(y)\|^2$ up to additive $\varepsilon/10$ term, with probability $1 - \varepsilon/10$, using the ℓ_2 estimation algorithm. Specifically, since $\|\rho(x) - \rho(y)\|^2 \leq 4$, we can just use a $(1 + \varepsilon/40)$-multiplicative ℓ_2 estimation protocol (e.g., via embedding ℓ_2 into the Hamming space and then using the [KOR00] sketch). Note that the protocol has size $O(1/\varepsilon^2)$ (for [KOR00] sketch), times $O(\log 1/\varepsilon)$ (to boost the success probability to $\geq 1 - \varepsilon/10$), times another $O(\log 1/\varepsilon)$ (to guess the right scale); in other words, the size of the protocol is less than C'.

Let z_{xy} be the estimate given by the ℓ_2 estimation protocol on input (x, y). The protocol accepts with probability exactly z_{xy}. The resulting success probability is at least:

$$\begin{aligned}
&p_1 \cdot \mathbb{E}_{\eta_1}(1 - \varepsilon/10) z_{xy} + p_0 \cdot \mathbb{E}_{\eta_0}(1 - \varepsilon/10)(1 - z_{xy}) \\
&\geq 1 - \tfrac{\varepsilon}{3} + \mathbb{E}_{\eta_1} \hat{\|}\rho(x) - \rho(y)\hat{\|} - \mathbb{E}_{\eta_0} \hat{\|}\rho(x) - \rho(y)\hat{\|} - \tfrac{3\varepsilon}{10} \\
&\geq \tfrac{1}{2} + \tfrac{\varepsilon}{3}.
\end{aligned}$$

This is a contradiction. The claim follows.

Combining the above claim with Corollary 3.1, we get:

Corollary 4.1. *Let g be a 0-1 distance function whose simultaneous-message communication complexity is at least C, for some large absolute constant C, with error probability at most 1/3. Then, the general communication complexity of the problem $f(\mathbf{x}, \mathbf{y}) = \bigvee_{i=1}^n g(x_i, y_i)$ is $\Omega(n)$.*

4.1 Applications to Product Spaces and Edit Distance

We first state our general corollaries, which hold for *product spaces*. We then show how they imply our lower bound on Ulam and edit distances.

We define two types of product spaces. Let (X, d) be a metric space. A *max-product* of $k \geq 1$ copies of X is the metric (X^k, d_∞), denoted $\ell_\infty(X)$ or $\bigoplus_{\ell_\infty}^k X$, where the distance between $x = (x_1, \ldots x_k), y = (y_1, \ldots y_k) \in X^k$ is $d_\infty(x, y) = \max_{i\in[k]} d(x_i, y_i)$. Similarly, we define the *sum-product*, which is the metric (X^k, d_1), denoted $\ell_1(X)$ or $\bigoplus_{\ell_1}^k X$, where the distance between $x, y \in X^k$ is $d_1(x, y) = \sum_{i\in[k]} d(x_i, y_i)$.

We now define the *distance threshold estimation problem (DTEP)* for a given metric X, approximation factor $\alpha \geq 1$, and a threshold $R > 0$. The problem is defined on pairs of points $x, y \in X$ as follows. The No instances are those where $d(x, y) \leq R$. The Yes instances are those where $d(x, y) > \alpha R$. We denote this problem as $\mathrm{DTEP}(X, \alpha, R)$.

We are now ready to state the corollaries of the direct sum theorem.

Corollary 4.2. (Max-product) *There is an absolute constant $C > 1$ such that the following holds. Fix some metric X, threshold $R > 0$, and approximation $\alpha \geq 1$. Suppose* $\mathrm{DTEP}(X, \alpha, R)$ *has communication complexity at least C.*

Then, for any $k \geq 1$, $\mathrm{DTEP}(\bigoplus_{\ell_\infty}^k X, \alpha, R)$*, defined by the max-product of k copies of X, has communication complexity of $\Omega(k)$.*

Proof. Let $g : X^2 \to \{0, 1\}$ be the function corresponding to $\mathrm{DTEP}(X, \alpha, R)$. Note that $g(x, x) = 0$ (No) as $d(x, x) = 0 \leq R$ by the definition of the metric. Then, for any $k \geq 1$ $\mathrm{DTEP}(\bigoplus_{\ell_\infty}^k X, \alpha, R)$ corresponds to the function $\bigvee_{i=1}^k g_i$ where each $g_i = g$ for $i \in [n]$. The result then follows from Theorem 4.1.

Corollary 4.3. (Sum-product) *There are an absolute constants $C > 1$ and $c > 0$ such that the following holds. Fix some metric X, threshold $R > 0$, and approximation $\alpha \geq 1$. Suppose* $\mathrm{DTEP}(X, \alpha, R)$ *has simultaneous communication complexity at least C.*

For any $1 \leq k \leq c\alpha$ the following holds. Consider the space $\bigoplus_{\ell_1}^k X$ whose metric is given by the sum-

product of k copies of X. Then $\mathrm{DTEP}(\bigoplus_{\ell_1}^k X, \alpha/k, kR)$ *has communication complexity* $\Omega(k)$.

Proof. We reduce the DTEP for the max-product space of X to the DTEP for the sum-product space of X via the identity mapping. This is because for any $x, y \in X^k$ such that $d_\infty(x, y) \leq R$, we have that $d_1(x, y) \leq kR$ (i.e., when we view the points x, y in the metric of sum-product of X). Similarly, when $d_\infty(x, y) > \alpha R$, then $d_1(x, y) > \alpha R = \frac{\alpha}{k} \cdot kR$. The result then follows using the previous corollary.

We are now ready to prove our main result for the Ulam and edit distance.

Theorem 4.1. *Let d be the length of strings. There exists some threshold $R > 1$ such that for constant approximation, the DTEP for Ulam distance requires $\Omega\left(\frac{\log d}{\log\log d}\right)$ communication. More generally, for any approximation $\alpha \leq O\left(\frac{\log d}{\log\log d}\right)$, the DTEP for Ulam distance has communication complexity of $\Omega\left(\frac{\log d}{\alpha \cdot \log\log d}\right)$.*

Same lower bound holds for edit distance over binary strings as well.

Proof. We use the following result of [AK07, Theorem 1.1].

Theorem 4.2. ([AK07]) *There exists some absolute constant c' and threshold $d^{0.1} \leq R \leq d^{0.49}$, such that, for any approximation at most $\phi(d) = c' \frac{\log d}{\log\log d}$, the DTEP for Ulam distance has communication complexity more than C.*

Let $k = \phi(d^{0.99})/\alpha$. Let us denote the Ulam distance on strings of length l by Ulam_l. Then, consider the DTEP for the sum-product of k copies of $\mathrm{Ulam}_{d^{0.99}}$. The above theorem, in conjunction with Corollary 4.3, implies that, for approximation α, the DTEP for $\bigoplus_{\ell_1}^k \mathrm{Ulam}_{d^{0.99}}$ has communication complexity at least $\Omega(k) = \Omega\left(\frac{\log d}{\alpha \cdot \log\log d}\right)$.

It remains to show that we can reduce the DTEP for $\bigoplus_{\ell_1}^k \mathrm{Ulam}_{d^{0.99}}$ to the DTEP for Ulam distance for strings of length d. Indeed, we can map the metric $\bigoplus_{\ell_1}^k \mathrm{Ulam}_{d^{0.99}}$ into Ulam_d preserving all distances. For $x = (x_1, \ldots x_k) \in \bigoplus_{\ell_1}^k \mathrm{Ulam}_{d^{0.99}}$, just construct $\zeta(x) \in \mathrm{Ulam}_d$ by concatenating $x_1 \circ x_2 \circ \ldots x_k$ using a new alphabet for each coordinate $i \in [k]$, and appending $d - kd^{0.99}$ more copies of a symbol $\perp$ that does not appear in any of the other alphabets. It is immediate to check that, for any $x, y \in \bigoplus_{\ell_1}^k \mathrm{Ulam}_{d^{0.99}}$, we have that $\mathrm{ed}(\zeta(x), \zeta(y)) = \sum_i \mathrm{ed}(x_i, y_i)$.

The result for edit distance on binary strings follows from the result on Ulam metric together with Theorem 1.2 from [AK07], which shows a reduction from the latter to the former.

References

[ADG+03] Alexandr Andoni, Michel Deza, Anupam Gupta, Piotr Indyk, and Sofya Raskhodnikova. Lower bounds for embedding edit distance into normed spaces. In *SODA*, pages 523–526, 2003.

[AIK09] Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer. Overcoming the ℓ_1 non-embeddability barrier: Algorithms for product metrics. In *SODA*, pages 865–874, 2009.

[AK07] Alexandr Andoni and Robert Krauthgamer. The computational hardness of estimating edit distance. In *FOCS*, pages 724–734, 2007. Accepted to *SIAM Journal on Computing* (FOCS'07 special issue).

[AO09] Alexandr Andoni and Krzysztof Onak. Approximating edit distance in near-linear time. In *STOC*, pages 199–204, 2009.

[BJKS04] Ziv Bar-Yossef, T.S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

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

[DD06] Michel Deza and Elena Deza. *Dictionary of Distances.* Elsevier Science, 2006.

[Ind01] P. Indyk. Tutorial: Algorithmic applications of low-distortion geometric embeddings. In *FOCS*, pages 10–33, 2001.

[Jay09] T.S. Jayram. Hellinger strikes back: A note on the multi-party information complexity of AND. In *RANDOM*, 2009. To Appear.

[JW09] T.S. Jayram and David Woodruff. The data stream space complexity of cascaded norms. In *FOCS*, 2009. To appear.

[KN06] Subhash Khot and Assaf Naor. Nonembeddability theorems via fourier analysis. *Math. Ann.*, 334(4):821–852, 2006. Preliminary version appeared in FOCS'05.

[KOR00] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Comput.*, 30(2):457–474, 2000. Preliminary version appeared in STOC'98.

[KR06] Robert Krauthgamer and Yuval Rabani. Improved lower bounds for embeddings into l_1. In *SODA*, pages 1010–1017, 2006.

[KS92] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[Mat02] Jiri Matousek. *Lectures on Discrete Geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

[Mat07] J. Matoušek. Collection of open problems on low-distortion embeddings of finite metric spaces, March 2007. Available online. Last access in August, 2007.

[OR05] R. Ostrovsky and Y. Rabani. Low distortion embeddings for edit distance. In *Proceedings of the 37th ACM Symposium on Theory of Computing*, pages 218–224, 2005.

[Raz92] A. A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

A Hellinger distance and Communication Protocols

Proposition A.1. ([BJKS04]) *Let $\mathcal{P}$ be a randomized private-coin protocol on $\mathcal{X} \times \mathcal{Y}$. Let $(u_1, v_1), (u_2, v_2) \in \mathcal{X} \times \mathcal{Y}$ be two distinct inputs whose transcript wave functions in $\mathcal{P}$ are denoted by $\psi(u_1, v_1)$ and $\psi(u_2, v_2)$, respectively.*

Mutual information to Hellinger distance:
Suppose $(U, V) \in_R \{(u_1, v_1), (u_2, v_2)\}$. If Π denotes the transcript random variable, then

$$\mathrm{I}(U, V : \Pi) \geq \hat{\|}\psi(u_1, v_1) - \psi(u_2, v_2)\hat{\|}.$$

Soundness:
Suppose $\mathcal{P}$ is a correct protocol for a decision problem g defined on $\mathcal{L} \subseteq \mathcal{X} \times \mathcal{Y}$. Suppose $(u_1, v_1), (u_2, v_2) \in \mathcal{L}$ such that $g(u_1, v_1) \neq g(u_2, v_2)$. Then,

$$\hat{\|}\psi(u_1, v_1) - \psi(u_2, v_2)\hat{\|} \geq 1 - 2\sqrt{\delta}.$$

Pythagorean property:
Consider the combinatorial rectangle of 4 inputs $\{u_1, u_2\} \times \{v_1, v_2\}$ and label them as $A = (u_1, v_1)$, $B = (u_1, v_2)$, $C = (u_2, v_1)$ and $D = (u_2, v_2)$. Then,

$$\begin{aligned} &\hat{\|}\psi(A) - \psi(D)\hat{\|} \\ &\geq \begin{cases} \frac{1}{2} \cdot (\hat{\|}\psi(A) - \psi(B)\hat{\|} + \hat{\|}\psi(C) - \psi(D\hat{\|}) \\ \frac{1}{2} \cdot (\hat{\|}\psi(A) - \psi(C)\hat{\|} + \hat{\|}\psi(B) - \psi(D)\hat{\|}) \end{cases} \end{aligned}$$

□

The first property in the above proposition is just a restatement of the fact that the Jensen-Shannon distance between $\psi(u)$ and $\psi(v)$ is bounded from below by their Hellinger distance. The next property follows by relating Hellinger to variational distance and then invoking the correctness of the protocol. The last property relies on the structure of *deterministic* communication protocols, namely, that the transcripts partition the space of inputs into *combinatorial rectangles*. The property itself can be seen as one generalization to randomized protocols. (In [BJKS04], another property is shown which generalizes the cut-and-paste property of deterministic communication protocols. This is not needed for our results.)