Distance Estimation Protocols for General Metrics

Alexandr Andoni*
MIT
andoni@mit.edu

Robert Krauthgamer
Weizmann Inst. and IBM Almaden
robert.krauthgamer@weizmann.ac.il

July 3, 2008

Abstract

Many algorithms involving metric spaces rely on a basic primitive that estimates the distance between points x, y in a metric space M. A common framework for this task is provided by communication complexity: Alice and Bob receive points x and y respectively as private inputs, and they wish to decide whether the distance $d(x, y)$ is "small" or "large" using only a small amount of communication. The precise model may depend on the intended application, e.g. sketching and one-way protocols pose further limits on the number of rounds, but are very useful for near neighbor searching and for streaming algorithms.

So far, the extensive work devoted to distance estimation protocols focused on particular metrics, such as L_p or edit distance. Seeking a broader perspective, the following meta-question has recently emerged (cf. [McG06, Question #5]): **"Characterize which metric spaces admit efficient communication protocols?"**

We address this meta-question by studying *general* metric spaces, seeking to understand the extreme cases and how they are affected by metric (or topological) structure. At one extreme, we show that expanders have the largest possible communication complexity. This is proved for every approximation factor, and thus provides a strict generalization of known non-embeddability results for expanders. At the other extreme we have various metrics that admit very efficient protocols (even more efficient than say L_1) achieving arbitrarily good approximation. Our results determine, for example, whether $\mathbb{R}^3$ is easier or harder than a tree metric.

We also compare the power of sketching protocols and that of embeddings, offering a partial converse to the well-known fact that sketching generalizes the "embeddings approach".

*Most of the work was done while the author was visiting IBM Almaden Research Center.

1 Introduction

Many algorithms involving metric spaces rely on a basic primitive that estimates the distance between points x, y in a metric space $\mathcal{M}$. A common framework for this task is provided by communication complexity: Alice and Bob receive points x and y respectively as private inputs, and they wish to decide whether the distance $d_{\mathcal{M}}(x, y)$ is $\leq R$ or $> \alpha R$, for predetermined threshold $R > 0$ (this is merely the decision version of approximation within factor $\alpha \geq 1$). The precise model may depend on the intended application, although we always allow the use of public randomness. In fact, we shall concentrate on sketching and one-way protocols, which pose further limits on the number of rounds (see precise definitions below), but are extremely useful e.g. for near neighbor searching and for streaming algorithms.

The recent years have seen increased motivation for studying the communication complexity of metric spaces — see Section 1.2 for known motivations and new potential applications. So far, the extensive work devoted to this focused on a few metric spaces of key importance, such as ℓ_p spaces, for which we now know near-complete answers for all $p \in [1, \infty]$ [IM98, KOR00, SS02, BJKS04, Woo04], and more recently Earthmover distance [AIK08] and edit distance [BYJKK04, AK07]. It is thus timely to seek a broader perspective, and the following meta-question emerges quite naturally:

Meta-question: *Characterize which metric spaces admit efficient communication protocols.*

This question was raised, in a slightly different form, in the open problems collection [McG06, Question #5]. We are aware of only one result in this direction, due to [GIM07], which addresses a family that includes all Bregman divergences between distributions.

The present work is geared towards the above meta-question. First, we seek to understand whether the extreme cases correspond to metric (or topological) structure. We thus study metrics with the largest possible communication complexity, finding along the way a strict generalization of known non-embeddability results. At the other extreme, we study metrics with very efficient protocols, even lower complexity than L_1. We also compare the power of communication protocols and that of embeddings, offering a partial converse to the well-known fact that sketching generalizes the "embeddings approach". Before going into more details, we set up terminology and review basic facts regarding these concepts.

A Hierarchy of Protocols. We use standard communication complexity models, which date back to [Yao82]. Restrictions on the order of messages give rise to different types of protocols:

Two-way: Alice and Bob exchange messages in any order.

One-way: Alice sends just one message to Bob, who is to decide on the output.

Simultaneous (aka sketching): Alice and Bob each sends a message to a third party, the referee, who is to decide on the output, based only on the two messages and the public randomness.

In all three cases, the *communication complexity* (or *size*) is defined to be the minimum number of communication bits needed by a protocol that (on every input x, y) succeeds with probability at least $2/3$. The three protocol types are progressively more restricted, meaning that a two-way protocols can simulate one-way protocol with no additional communication, and similarly one-way protocols can simulate simultaneous protocols.

The Embeddings Prism. Communication protocols for distance estimation are often related to the following notion. An *embedding* of a metric $\mathcal{M}$ into a *host* metric, such as ℓ_1, is a map $\varphi : \mathcal{M} \mapsto \ell_1$ with the property that for all $x, y \in \mathcal{M}$, the quantity $\|\varphi(x) - \varphi(y)\|_1$ approximates

$d_{\mathcal{M}}(x, y)$ within approximation α, which is often called the *distortion* of φ. (See e.g. [IM03] and the references therein.) The embedding can be viewed as a communication protocol (even a sketching one), where Alice and Bob send $\varphi(x)$ and $\varphi(y)$, respectively, and their message determine the estimate $\|\varphi(x) - \varphi(y)\|_1$. As stated, this model is incomparable to the sketching model; on the one hand, a full representation of $\varphi(x)$ might be rather lengthy to communicate; on the other hand the protocol's structure is very restricted, e.g. to using only an ℓ_1 difference. Nevertheless, sketching protocols do generalize the embeddings model, if we accept a small increase in the approximation by a factor of $1 + \varepsilon$. This claim follows from [KOR00], who prove that ℓ_1 metrics admit a sketching protocol achieving $1 + \varepsilon$ approximation with $O(1/\varepsilon^2)$ communication.

Obtaining tight (non)embeddability results for some notorious metrics is a key open problem, see e.g. the list [Mat07, Problems 2.18 and 2.20]. Recalling that sketching generalizes embeddings, we know by the contrapositive that lower bounds for sketching imply lower bounds for embedding. Indeed this indirect approach was recently employed for the edit distance and Ulam metric [AK07], yielding non-embeddability results that were not known earlier (i.e. before studying sketching).

A Computational Perspective. We advocate viewing the communication complexity of (estimating distances in) a metric space $\mathcal{M}$ as a "measure" of complexity for $\mathcal{M}$, especially with regard to different computational tasks involving this metric. Such "measures" of complexity are instructive in explaining or predicting the tractability of different computational problems. For example, the doubling dimension was suggested in [GKL03] as a proxy for a metric's low complexity, and a sequence of subsequent papers showed that this notion is indeed a useful parameter for a diverse set of problems, such as Nearest Neighbor Search (NNS), the Traveling Salesman Problem (TSP), and low-stretch routing. This proxy is often useful when the metric $\mathcal{M}$ can only be accessed in a black-box manner by querying the distance function. Another proxy is provided by embedding into ℓ_1, which is indeed more useful in several cases, but we feel that communication complexity offers an even better alternative, and may generalize or outperform these other two in many contexts.

1.1 Our Results

First, we study the two ends of the spectrum of communicaton complexity, namely metrics of very large and very small communication complexity. For every n-point metric, a sketch of size $O(\log n)$ clearly suffices for exact distance estimation ($\alpha = 1$). Two fundamental questions that immediately arise are (i) is this upper bound tight? and (ii) does approximation help, and if so, by how much? To answer question (ii), we show that for every desired approximation $1 \le \alpha \le O(\log n)$ there is a one-way protocol of complexity $O(\frac{\log n}{\alpha})$. This bound offers a smooth tradeoff between approximation and communication; the only values previously known were $\alpha = 1$ and $\alpha = O(\log n)$ (the latter follows by, e.g., embedding into ℓ_1 [Bou85] then using protocol for ℓ_1 [KOR00]). Furthermore, we prove that this communication upper bound is tight, for all α, thus answering question (i) more generally (not only for $\alpha = 1$).[1] Our matching lower bound is proved for expander graphs, thus offering a generalization to the known non-embeddability results for expanders; indeed, the case $\alpha = O(\log n)$ in our lower bound implies (but does not follow from) non-embeddability of expanders [LLR95, LM00]. These results appear in Section 2.

[1]A simple counting argument over unweighted graphs is inapplicable because the protocol may depend on the metric, but here is a simple alternative: every symmetric function $f : [n] \times [n] \to \{0, 1\}$ corresponds to an unweighted graph, in which we can take the shortest-path metric, hence the distance $d(x, y)$ is 1 whenever $f(x, y) = 1$ (and ≥ 2 otherwise). Known functions f (e.g. a random one) require $\Omega(\log n)$ communication, implying the same communication lower bound for factor 2 distance estimation.

At the other extreme we focus on metrics with very efficient sketching protocols — even lower communication complexity than L_1, which admits sketching with $1+\varepsilon$ approximation using sketch-size $O(1/\varepsilon^2)$ (for any desired $\varepsilon > 0$). Specifically, we give near-tight bounds for paths, trees, and low-dimensional $\mathbb{R}^k$ under ℓ_1 norm (and easily extends to ℓ_p for every fixed $p \geq 1$). We give a fuller account in Table 1. It is interesting to note that there is an exponential gap between, say, $\mathbb{R}^3$ and tree metrics (even though both embed in ℓ_1 *isometrically*, i.e. with distortion 1). Previously, no bounds were known for all these metrics. Our bound for $\mathbb{R}^k$ follows from a more general result for *sum-product spaces.* For a metric $\mathcal{M}$ with distance $d_{\mathcal{M}}$, define the *sum-product metric* $\bigoplus^k \mathcal{M}$ to be the set of all tuples $\langle x_1, \ldots, x_k\rangle \in \mathcal{M}^k$ endowed with distance function $d_{\ell_1,\mathcal{M}}(\langle x_1, \ldots, x_k\rangle, \langle y_1, \ldots y_k\rangle) = \sum_{i=1}^k d_{\mathcal{M}}(x_i, y_i)$. For example, the k-dimensional ℓ_1 space is simply $\bigoplus^k \mathbb{R}$, and the real hyperbolic space $\mathbb{H}^k$ embeds (roughly) into the product of k trees [BS05, BDS07]. We note that many other metrics, such as edit distance, Earthmover distance, and squared-ℓ_2, are closed under the sum-product operator. These results appear in Section 3.

	Metric	Approx.	Communication UB/LB		Protocol Types/Qualifications
here	arbitrary	α	$O(\frac{\log n}{\alpha})$	$\Omega(\frac{\log n}{\alpha})$	1-way. LB: expanders
	ℓ_1 on $\mathbb{R}^k$	$1+\varepsilon$	$k \cdot \tilde{O}(\log \frac{1}{\varepsilon})$	$k \cdot \tilde{\Omega}(\log(2 + \frac{1}{k\varepsilon}))$	UB: all protocols. LB: 1-way. $k \leq \frac{1}{\varepsilon^2}$
	$\bigoplus^k$ trees	$1+\varepsilon$	$\tilde{O}(k/\varepsilon)$	$\tilde{\Omega}(1/\varepsilon)$	UB: all protocols. LB: 1-way.
known	L_1, L_2 (any dim.)	$1+\varepsilon$	$O(1/\varepsilon^2)$	$\Omega(1/\varepsilon^2)$	UB: all protocols. LB: 1-way. [KOR00, Woo04]
	ℓ_p on $[m]^k$, $p \in (2, \infty]$	α	$\tilde{O}(k^{1-2/p}/\alpha^2)$	$\Omega(k^{1-2/p}/\alpha^2)$	for all protocols [SS02, BJKS04]

Table 1: Our results and some known bounds. For each metric, we present approximation factor, communication upper bound (UB) and lower bound (LB).

A natural approach to answer our meta-question is to relate the communication complexity of a metric $\mathcal{M}$, say for sketching protocols, to embedding of $\mathcal{M}$ into, say, squared-ℓ_2. As mentioned previously, sketching generalizes the "embeddings approach". An immediate question is whether there is a converse statement, i.e., *does efficient sketching imply low-distortion embeddings?* After delineating the precise statement based on a few examples, we offer such a partial converse that requires some technical conditions, most notably that $\mathcal{M}$ has low aspect ratio. See Section 4 for details.

We conclude with a discussion of some concrete problems that emerge from our results and may bring us closer to a resolution of the meta-question. Due to lack of space, this part is relegated to Appendix A.

1.2 Applications of Distance Estimation Protocols

To put the communication complexity approach in a larger context, we discuss some known and potential scenarios where distance estimation protocols play a key role in either designing algorithms, or proving tight lower bounds for such algorithms. We remind that one such strong connection — to embeddings — was discussed above.

Linear-time Distance Estimation. In its simplest form, the *distance estimation (DE)* problem just asks to compute (exactly or approximately) the distance $d_{\mathcal{M}}(x, y)$ between inputs $x, y \in \mathcal{M}$, say on a RAM machine. Sometimes, it turns out that casting the DE problem in a restricted setting, such as a communication protocol, helps in designing new algorithms. Attacking a problem by restricting the computational model might seem counter-intuitive, but this approach actually helped for obtaining near-linear time DE algorithms for several metrics such as variants of edit distance and Earthmover distance, e.g. by embedding them into L_1 [CPSV00, MS00, CM02, Cha02, IT03, AIK08].

Nearest Neighbor Search (NNS). This problem asks to construct a data structure on a dataset of n points from a metric $\mathcal{M}$, so as to support queries of the form: given as query a point $q \in \mathcal{M}$, report the data point closest to q. This is a central problem in many areas and thus was studied extensively; see [Ind03] and references therein for details.

It can be seen, using ideas from previous work, that low-complexity distance estimation protocols imply a host of NNS algorithms. For example, a one-way protocol of size s can be used to design an NNS scheme with polynomial space and fast queries ($n^{O(s)}$ space and $O(s \log n)$ query time). Alternatively, such a protocol implies an NNS scheme with near-linear space and sublinear queries (roughly, $n^{1+O(\varepsilon)}$ space and $O(n^{1-\varepsilon/s})$ query time for any desired $\varepsilon > 0$). We provide the details in Appendix D.1, for completeness, since we are not aware of an explicit proofs of these statements in the literature. The two schemes mentioned above are proved via a unified approach, which is a generalization of the hashing techniques that were devised in [KOR00, IM98] for NNS in ℓ_1. Additionally, a sketching protocol of size s can be used in NNS algorithms based on a linear-scan, as a fast filtering mechanism. While these algorithms take a time linear in n, the running time per point may be only $O(s \log n)$, independent of the size (dimension) of a point. (In particular, this algorithm may significantly reduce the number of exact distance evaluations.)

Distance labeling. Several concepts were proposed in the literature to address the broad goal of representing a metric space in a more succinct yet easily accessible method than simply storing the entire matrix of distances. Depending on the precise formulation of efficiency, these concepts including spanners [PS89], distance labeling schemes [Pel00], distance oracles [TZ04], and network triangulation [KSW04]. Among these, the closest to distance estimation is distance labeling, though the two are generally incomparable because (a) distance labeling applies to a family of finite metrics, while distance estimation operates on one fixed (possibly infinite) metric; and (b) distance labeling is deterministic and solves the "search" version, while distance estimation is randomized and solves the decision version. In many scenarios (e.g. networking) a sketching algorithm can be more advantageous, potentially leading to smaller storage per point, due to the above differences. Another potential advantage is that the sketch (label) of each point is often computed locally and independently of the number of relevant points (e.g. if the metric is infinite). However, a protocol may sometimes depend on the metric space in an undesirable way (e.g. scan all n points in search of some event).

We observe (Appendix D.2) that a low-complexity one-way protocol for an infinite metric $\mathcal{M}$ implies an approximate distance labeling scheme with small labels for the family of all finite submetrics of $\mathcal{M}$. To see one example, observe that an infinite n-ary tree contains as a submetric every tree of size n with integer edge weights. This example demonstrates also the aforementioned differences in efficiency, e.g. sketching offers some guarantees even with only $O(1)$ bits per point.

Streaming algorithms. In the data-stream model, the input is a sequence of elements arriving in a sequential fashion, without providing access to earlier elements. The main complexity measure of an algorithm in this model is its use of space (storage). Quite often, although sometimes implicitly, the data stream contains a description of x followed by that of y, and we wish to estimate some distance $d_{\mathcal{M}}(x, y)$. As observed in [AMS96], a lower bound on the one-way communication complexity of $\mathcal{M}$ immediately implies a lower bound on the space requirement of a streaming algorithm for this purpose. This method of proving space lower bounds for streaming algorithms was used in [SS02, BJKS04] to obtain a streaming lower bound for ℓ_∞. In particular, their main technical result is a tight communication lower bound for one-way protocols for distance estimation of L_p's for $p > 2$.

1.3 Preliminaries

Fix a metric space $\mathcal{M}$, and $r > 0, \alpha \geq 1$. Let $\mathrm{DE}_{\mathcal{M},r,\alpha}$ (stands for Distance Estimation) denote the promise problem of deciding, upon input $(x, y) \in \mathcal{M} \times \mathcal{M}$ whether $d_{\mathcal{M}}(x, y) \leq r$ or $d_{\mathcal{M}}(x, y) > \alpha r$.

2 Metrics with Large Communication Complexity

We start our investigation with the regime of high communication complexity, proving that expander graphs, when equipped with the shortest distance metric, are the "hardest" metrics for one-way and for two-way protocols. Namely, we show that for every approximation $\alpha \geq 1$, *all n-point metrics* admit communication upper bound of $O(\frac{\log n}{\alpha})$, and that this bound is tight for expander graphs.

Theorem 2.1. *Consider any metric $(\mathcal{M}, d_{\mathcal{M}})$ with n points. Then, for every $r > 0$ and $1 \leq \alpha \leq \log n$, the one-way communication complexity of $\mathrm{DE}_{\mathcal{M},r,\alpha}$ is $O(\frac{\log n}{\alpha})$.*

Theorem 2.2. *Consider an (expander) graph G of size n, which is d-regular and its adjacency matrix has second eigenvalue at most $\lambda_2 d < d$. Then, for every $1 \leq \alpha \leq O(\log n)$, the two-way communication complexity of $\mathrm{DE}_{G,r,\alpha}$ is $\Omega_{\lambda_2,d}(\frac{\log n}{\alpha})$.*

These results generalize the statement that expanders are the hardest metrics for embedding into normed spaces such as ℓ_1 or ℓ_2. The equivalent of Theorem 2.1 for embeddings is Bourgain's theorem [Bou85] that every n-point metric embeds into ℓ_1 with $O(\log n)$ distortion. The equivalent of Theorem 2.2 for embeddings says that expanders require distortion $\Omega(\log n)$ in any embedding into ℓ_1 and $(\ell_2)^p$ for constant p [AR98, LLR95, LM00, Mat02]. We proceed to proving Theorem 2.1 and Theorem 2.2.

Proof of Theorem 2.1. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a finite metric space. Henceforth, let $B(x, t)$ denote the (closed) ball of radius t around x. For a partition P of $\mathcal{M}$, we shall call the elements $C \in P$ clusters. The partition is called Δ-bounded if its clusters all have diameter at most Δ. For a point $x \in X$, we let $P(x)$ denote the cluster $C \in P$ containing x. We shall need the following lemma due to Mendel and Naor [MN07, Lemma 3.1].

Lemma 2.3 ([MN07])**.** *Let $(\mathcal{M}, d_{\mathcal{M}})$ be a finite metric space. Then for every $\Delta > 0$ there exists a probability distribution over Δ-bounded partitions P of $\mathcal{M}$ such that for every $0 < t < \Delta/8$ and every $x \in \mathcal{M}$,*

$$\Pr_P[B(x,t) \subseteq P(x)] \geq \left(\frac{|B(x,\Delta/8)|}{|B(x,\Delta)|}\right)^{16t/\Delta}.$$

Fix $r > 0$ and $\alpha \geq \Omega(1)$. Let $c > 0$ be a sufficiently large constant, e.g. $c = 3$ suffices. Alice and Bob use their shared randomness to compute $T = cn^{16/\alpha}$ decompositions of the shortest-path metric of G, each drawn independently from the distribution in Lemma 2.3, using parameters $\Delta = \alpha r$ and $t = r$. Denote these randomly chosen partitions by $P_1, \ldots, P_T$. Call $v \in V$ padded in decomposition P_i if $B(v, r) \subseteq P_i(x)$. By the lemma, every $v \in V$ is padded in every partition P_i with probability $\geq (\frac{1}{n})^{16/\alpha}$.

The one-way protocol proceeds as follows. Alice find a partition P_i in which its input x is padded, and sends to Bob the value of $i \in [T]$ and $h(P_i(x))$ where $h : 2^V \mapsto \{0,1\}^c$ is a universal hash function, chosen using the shared randomness. If there is no such P_i, then Alice sends an error message, and we shall just assume the protocol fails. Bob computes for its input y the value $h(P_i(y))$ using the same hash function $h(\cdot)$ as Alice and the i he received, and accepts (declares that $d(x,y) \leq r$) if and only if the hash value he computed is equal to the one he received from Alice, namely $h(P_i(y)) = h(P_i(x))$.

Clearly, the communication complexity is $O(\log T) + c = O(\frac{\log n}{\alpha})$. It thus remains to bound the error probability. One obvious term is the probability the protocol fails because x is not padded in all the T partitions, which occurs with probability $\leq (1 - (\frac{1}{n})^{16/\alpha})^T < e^{-3}$. Assume now the protocol did not fail, i.e. x is padded in P_i. If $d(x,y) \leq r$ then $y \in B(x,r) \subseteq P_i(x)$, hence $P_i(x) = P_i(y)$ and Bob must accept. If $d(x,y) > \alpha r = \Delta$ then necessarily $P_i(x) \neq P_i(y)$, and their hash values must collide for Bob to accept, which occurs with probability $\leq 2^{-c}$. We conclude that in either case, the protocol errs with probability at most $e^{-c} + 2^{-c} \leq 1/4$.

For $\alpha = O(1)$, Alice can just send a hash value of its point, which takes $O(\log n) = O(\frac{\log n}{\alpha})$ communication. $\square$

Proof of Theorem 2.2. Fix $r = \Theta(\frac{\log_d n}{\alpha})$ and consider the following two distributions. The "far distribution", μ_0, is uniformly random over pairs (x,y). The "close distribution", μ_1, chooses a random x and then generates y by taking a random walk from x of r steps. Technically, the "far distribution" may generate pairs of points that are not at distance $\geq \alpha r = \Omega(\log n)$, but this happens with probability $2^{-\Omega(\log n)}$, which we will ignore for the sake of clarity (see an example of the formal treatment of this issue in [AIK08] or [AK07]).

Suppose l is the complexity of a two-way protocol of $\mathrm{DE}_{G,r,\alpha}$. Then, Lemma 3.1 from [AK07] implies that there exist boolean function $\mathcal{H}^A, \mathcal{H}^B : V \to \{0,1\}$ such that

$$\Pr_{\mu_0}[\mathcal{H}^A(x) \neq \mathcal{H}^B(y)] - \Pr_{\mu_1}[\mathcal{H}^A(x) \neq \mathcal{H}^B(y)] \geq 2^{-O(l)}.$$

Let $A = (\mathcal{H}^A)^{-1}(1)$ and $B = (\mathcal{H}^B)^{-1}(1)$, and let $E^r(A,B)$ denote the set of length r walks starting in A and ending in B. Then

$$\Pr_{\mu_0}[\mathcal{H}^A(x) \neq \mathcal{H}^B(y)] - \Pr_{\mu_1}[\mathcal{H}^A(x) \neq \mathcal{H}^B(y)] = 2\Pr_{\mu_1}[\mathcal{H}^A(x)\mathcal{H}^B(y)] - 2\Pr_{\mu_0}[\mathcal{H}^A(x)\mathcal{H}^B(y)] = 2\frac{|E^r(A,B)|}{d^r|V|} - 2\frac{|A|\cdot|B|}{|V|^2}.$$

Applying the expander mixing lemma to the graph G^r defined by these walks,

$$\left| |E^r(A,B)| - \frac{d^r|A|\cdot|B|}{|V|} \right| \leq (\lambda_2 d)^r \sqrt{|A|\cdot|B|} \leq (\lambda_2 d)^r |V|.$$

Altogether, we conclude that $2^{-O(l)} \leq \lambda_2^r$, hence $l \geq \Omega(r \log \frac{1}{\lambda_2}) = \Omega(\frac{\log n}{\alpha} \cdot \log_d(1/\lambda_2))$. $\square$

3 Metrics with Very Efficient Protocols

At the other end of the spectrum we have metrics admitting very low space communication protocols. We focus on path metrics, trees, and products of these spaces, which embed isometrically into ℓ_1, and thus admit sketching with approximation $1+\varepsilon$ and sketch-size $O(1/\varepsilon^2)$ by [KOR00]. We answer the basic question of whether better dependence on ε is possible for these specific families.

First we establish near-tight bounds on the communication complexity of a simple path (equivalently, the real line $\mathbb{R}$) and of a complete binary tree (which contains every finite tree as a submetric, up to scaling). It is interesting to note that the complexity of these two metrics is different, in fact there is an exponential gap between them (in terms of ε).

Using these two metrics as a baseline, we then establish a form of a hierarchy of simple metrics. This hierarchy is obtained by applying the sum-product operation. For example, a sum-product of k path metrics (real lines) yields a k-dimensional ℓ_1. The sum-product operation is very natural, and many common metrics are closed under this operation.

Technically, we design "generic" protocols for $\oplus_k \mathcal{M}$, the sum-product of $\mathcal{M}$ with itself k times, by assuming $\mathcal{M}$ admits an efficient protocol and using it as a black-box. Applying these results to the two basic metrics (the path and the tree), we get protocols for the sum-product of paths (i.e. $\mathbb{R}^k$) and for the sum-product of trees (which effectively include hyperbolic spaces). We accompany these protocols with a lower bound for the sum-product of paths, which nearly matches our protocol for all $k \leq 1/2\varepsilon$; beyond this value, the complexity stabilizes to $\Theta(1/\varepsilon^2)$, the complexity of full-dimensional ℓ_1. An immediate consequence is that our generic protocol is nearly the best possible.

3.1 Path and Tree Metrics

We shall use the term path metric P to denote the metric on $\mathbb{R}$ with $d_P(x,y) = |x-y|$.

Theorem 3.1. *Let P be a path metric. Then there is a sketching algorithm for the distance estimation problem $\mathrm{DE}_{P,r,1+\varepsilon}$ that uses sketches of size $O(\log \frac{1}{\varepsilon})$. Furthermore, every one-way protocol for this problem must have size $\Omega(\frac{\log(1/\varepsilon)}{\log\log(1/\varepsilon)}) = \tilde{\Omega}(\log \frac{1}{\varepsilon})$.*

Deferring this proof to Appendix B, we now state and prove the result on tree metrics.

Theorem 3.2. *Let T be a (possibly infinite) tree metric. Then there is a sketching algorithm for the distance estimation problem $\mathrm{DE}_{T,r,1+\varepsilon}$ that uses sketches of size $O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$. Furthermore, every one-way protocols for this problem, even for complete binary trees, must use size $\Omega(\frac{1}{\varepsilon})$.*

We remark that the protocol in this theorem is randomized with one-sided error. The lower bound holds even for randomized, two-sided error protocols.

Proof of Theorem 3.2. We first prove the upper bound on the sketch size. We shall fix a point in T as the root and assume that $\varepsilon < 1/2$. We may assume without loss of generality (by subdividing and scaling) that all tree edges have unit length, and that r and $\varepsilon r/3$ are integral. We may further assume that all the points are at distance at least r from the root, by adding to T a path of length r hanging off the root of T, making the other endpoint of this path the new root of T.

Sketching protocol. We now describe Alice's algorithm for her input $x \in T$; Bob's algorithm for his input $y \in T$ is identical. Given her input x, Alice first finds all the strict ancestors of x (in the tree T) whose distance from x is at most $r(1+\varepsilon/3)$ and whose depth (distance from the root) is a multiple of $\varepsilon r/3$. Denote this sequence of points as S_x (ordered by their distance from x). Next, Alice hashes the name (unique identifier) of each point in this set S_x into $O(\log(1/\varepsilon))$ bits using the shared randomness, letting $\sigma(x)$ be the resulting sequence of hash values. Bob uses the same algorithm to compute $\sigma(y)$. It is easy to verify that the claimed bound on communication since $|S_x| \leq \frac{3}{\varepsilon}+1$, and each of these points is hashed into $O(\log\frac{1}{\varepsilon})$ bits.

The referee's algorithm is as follows. Given two sketches $a = \sigma(x)$ and $b = \sigma(y)$, he checks whether there are i, j such that the ith hash value in a equals the jth hash value in b and such that $i + j \leq \frac{3}{\varepsilon}+2$. If so, he decides that $d_T(x,y) \leq r$; otherwise, he decides that $d_T(x,y) > r(1+\varepsilon)$.

We now show that the referee's decision is correct with high probability. Fix $x, y \in T$ such that $d_T(x,y) \leq r$. Let the point $z \in T$ be their least common ancestor. Let $z' \in T$ be the ancestor of z

closest to z (i.e. the first point of the path from z to the root) whose depth is a multiple of $\varepsilon r/3$. By the triangle inequality $d(x, z') \le r(1 + \varepsilon/3)$, and thus $z' \in S_x$. A similar argument shows $z' \in S_y$, and their hash values in $s(x)$ and $s(y)$ are clearly equal. Furthermore, the position i of z' in S_x and the position j of z' in S_y satisfy $i + j \le \frac{d(x,z)+d(z,z')}{\varepsilon r/3} + \frac{d(y,z)+d(z,z')}{\varepsilon r/3} = \frac{d(x,y)+2d(z,z')}{\varepsilon r/3} \le \frac{3}{\varepsilon} + 2$, hence the referee will accept (with probability 1).

Next, fix $x, y \in T$ such that $d_T(x, y) > r(1 + \varepsilon)$. Consider some i and j such that $i + j \le \frac{3}{\varepsilon} + 2$. We claim that the ith point in S_x and the jth point in S_y must be distinct points in T. Indeed, if they were the same point $z \in T$, then by the triangle inequality $d(x, y) \le d(x, z) + d(z, y) \le (i+j)\varepsilon r/3 \le r(1+2\varepsilon/3)$, which contradicts our assumption. Using the claim and since hash values have $O(\log \frac{1}{\varepsilon})$ bits, we get that the two corresponding hash values in $s(x)$ and in $s(y)$ will be equal with probability at most $\varepsilon^3/32$. The number of such pairs i, j is at most $(\frac{3}{\varepsilon} + 1)^2 \le 16/\varepsilon^2$, and by a union bound we get that the referee will reject with high probability (at least $1 - \varepsilon/2 \ge 3/4$).

Lower bound. We proceed to proving the lower bound on communication. The proof is by reduction from a modified version of the indexing problem. We first recall the definition of this problem, and then describe our reduction from this problem to distance estimation.

Modified Indexing. A classical hard problem for one-way communication complexity is the indexing problem IND_n: Alice is given a string $\mathbf{v} \in \{0, 1\}^n$ and Bob is given an index $i \in \{1, \ldots, n\}$, and based on a single message from Alice, Bob has to output $\mathbf{v}_i$. It is well known that in any one-way protocol solving this problem, even a randomized one, Alice's message has to be of length $\Omega(n)$. We will need a variant of this problem (in which Bob has more information), denoted $\mathrm{IND}_{n,k}$: Alice gets a string $\mathbf{v} \in \{0, 1\}^n$; Bob gets an index $i \ge k$ and also the $k - 1$ bits in $\mathbf{v}$ that precede position i; the goal is again to output $\mathbf{v}_i$. Lemma 2 from [BYJKK04] shows that the modified indexing problem requires $\Omega(n - k)$ communication from Alice (for every n and k).

The reduction. Fix $0 < \varepsilon < 1/3$ and let $n = 3/\varepsilon$ and $k = 2n/3$. We now describe the reduction from $\mathrm{IND}_{n,k}$ to distance estimation in a complete binary tree T of depth n. The root of the tree, denoted root, is at depth 0. Let $\mathbf{v} \in \{0, 1\}^n$ be Alice's input, and let $(i, \mathbf{v}_{i-(k-1)}, \ldots, \mathbf{v}_{i-1})$ be Bob's input. Notice that the tree T is predetermined and thus requires no communication, and that every point in T corresponds to a unique $0-1$ string of length at most n. Alice lets the string $x \in \{0, 1\}^{2n/3}$ consist of the last $2n/3$ bits in her input $\mathbf{v}$, i.e. $x = \mathbf{v}_{n/3+1} \ldots \mathbf{v}_n$. Bob lets the string $y \in \{0, 1\}^{2i-n}$ consist of the last $i - n/3 - 1$ bits his input contains from $\mathbf{v}$, concatenated with $i - 2n/3 + 1$ zeros, i.e., $y = \mathbf{v}_{n/3+1} \ldots \mathbf{v}_{i-1} 0^{i-2n/3+1}$. Next, Alice and Bob consider their strings x, y as points in T, and apply the assumed one-way protocol for $\mathrm{DE}_{T,r,1+\varepsilon}$, for $r = n/3$; notice that $\varepsilon = 3/n = 1/r$. Finally, if the distance estimation protocol decides that $d_T(x, y) \le r$, then Bob outputs 0; otherwise he outputs 1.

The key observation is that for every $x, y \in T$, if j_{xy} is the length of the longest common prefix of x and y (i.e. the first j bits in x and y are equal), then $d_T(x, y) = d_T(x, \mathrm{root}) + d_T(y, \mathrm{root}) - 2j_{xy}$. It follows that if $\mathbf{v}_i = 0$, then $j_{xy} \ge i - n/3$; thus $d_T(x, y) \le 2n/3 + (2i - n) - 2(i - n/3) = n/3 = r$, and with high probability the distance estimation protocol will report this. Similarly, if $\mathbf{v}_i = 1$, then $j_{xy} \le i - n/3 - 1$; thus $d_T(x, y) \ge 2n/3 + (2i - n) - 2(i - n/3 - 1) = n/3 + 2 > r(1 + \varepsilon)$, and with high probability the distance estimation protocol will report this.

Now if there exists a one-way protocol for $\mathrm{DE}_{T,r,1+\varepsilon}$ that uses at most $s = s(\varepsilon)$ bits of communication, then we obtain a one-way protocol for $\mathrm{IND}_{n,2n/3}$ with $n = 1/\varepsilon$ that uses at most s bits of communication. But as mentioned above, we know that the latter must be $\Omega(n)$, and we conclude that $s \ge \Omega(n) = \Omega(\frac{1}{\varepsilon})$. □

3.2 Products of Metrics

We now consider the complexity of product metrics. We prove two results of the following type: if the problem $\mathrm{DE}_{N,r,\alpha}$ has a protocol with communication s_α (for all $r > 0$), then $\bigoplus^k N$ has communication protocol of communication linear in $k \cdot s_\alpha$, potentially for a slightly bigger approximation. One result, Proposition B.1, gives a protocol with $\tilde{O}(\frac{\alpha k}{\varepsilon}) \cdot s_\alpha$ communication for $(1+\varepsilon)\alpha$ approximation for $\mathrm{DE}_{\bigoplus^k N,r,(1+\varepsilon)\alpha}$. A second result, Proposition B.1, achieves a smaller communication, $\tilde{O}(k \log \frac{1}{\varepsilon}) \cdot s_\alpha$, but requires a technical condition on N. We defer the statements and their proofs to Appendix B. We note that both proofs proceed (roughly) by suitably choosing a random r_i for each metric N_i and sketching each N_i for the respective r_i.

Applying the results for sum-product metrics, we obtain upper bounds for two specific families: (i) sum-product of paths, and (ii) sum-product of trees. For both families, we obtain a sketching protocol whose communication grows almost linearly with k, and for the former family we also provide a nearly matching lower bound. Note that both families embed isometrically into ℓ_1, hence $\mathrm{DE}_{M,r,1+\varepsilon}$ can be solved using sketch-size $O(1/\varepsilon^2)$ [KOR00], and if k (the dimension) is unbounded (or even about $1/\varepsilon^2$), this sketch-size is known to be optimal [Woo04]. But for relatively small k, e.g. $k \ll 1/\varepsilon$, the bounds we provide here are better. We prove the following theorem in Appendix B.

Theorem 3.3. *Let $P^k = P_1 \times \cdots \times P_k$ be the product metric of k (possibly infinite) path metrics, and let $T^k = T_1 \times \cdots \times T_k$ be a product metric of k (possibly infinite) tree metrics. Then for every $r > 0$ and $\varepsilon \le 1$, there is a sketching algorithm for the following distance estimation problems:*

- $\mathrm{DE}_{P^k,r,1+\varepsilon}$ *has sketching complexity $O(k \log \frac{k}{\varepsilon})$. Furthermore, every one-way algorithm must use $\Omega(k \cdot \frac{\log(1/\varepsilon k)}{\log\log(1/\varepsilon k)}) = \tilde{\Omega}(k \log \frac{1}{\varepsilon k})$ communication if $2k \le 1/\varepsilon$, or $\Omega(k)$ if $k \le 1/\varepsilon^2$.*
- $\mathrm{DE}_{T^k,r,1+\varepsilon}$ *has sketching complexity $O(\frac{k}{\varepsilon} \log^2(\frac{k}{\varepsilon})) = \tilde{O}(\frac{k}{\varepsilon})$.*

4 Embeddings into Normed Spaces

We now compare the power of sketching protocols and that of embeddings, offering a partial converse to the well-known fact that sketching generalizes the "embeddings approach". The motivation here is to characterize sketchable metrics by their embeddability into normed spaces (see also open problem 2 in Appendix A).

It is natural to compare sketching to the normed spaces such as $(\ell_2)^p$, for $p \in \mathbb{N}$, the p^{th} power of ℓ_2. The metric $(\ell_2)^p$ is the space on $\mathbb{R}^d$, endowed with the semimetric distance $d(x,y) = \|x-y\|_2^p$. We note that $(\ell_2)^p$ is isometrically embeddable into $(\ell_1)^p$ (defined similarly), which, in turn, is isometrically embeddable into $(\ell_2)^{2p}$.

The feature of $(\ell_2)^p$ is that sketching $(\ell_2)^p$ takes $O(p^2/\varepsilon^2)$ communication for approximation $1+\varepsilon$. Furthermore, the spaces $(\ell_2)^p$ are known to be progressively richer spaces as p increases. In fact, Deza and Maehara [DM90] proved that every metric on n points is isometrically embeddable into space $(\ell_2)^p$ for $p = O(n)$.

We prove that small communication sketching implies embedding of the respective metric into $(\ell_2)^p$, for relatively small p's, in particular much smaller than given by Deza-Maehara. Contrary to Deza-Maehara (which concerned isometric embeddings), our result concerns approximations to true distances.

Our resulting embedding is *asymmetric*. Namely, we construct *two* functions ψ_1 and ψ_2 both taking $\mathcal{M}$ into $(\ell_2)^p$, and we measure distance between x and y as $\|\psi_1(x) - \psi_2(y)\|_2^p$. Later, we explain why the asymmetry is necessary, at least if we want a small p. The formal statement of the result follows. The aspect ratio Δ of a metric is the ratio of the maximum distance over the minimum (non-zero) distance.

Theorem 4.1. *Let $\mathcal{M}$ be a metric of aspect ratio $\Delta > 1$, and fix approximation factor $\alpha \geq 1$. Suppose that for each $r \in [1, \Delta]$, $DE_{\mathcal{M},r,\alpha}$ admits a sketching protocol of size s. Then, for every $\varepsilon > 0$ there are $\psi_1, \psi_2 : \mathcal{M} \to (\ell_2)^p$ for $p = \left(\frac{\log \Delta}{\varepsilon}\right)^{O(s)}$, such that for all $x \neq y \in \mathcal{M}$,*

$$d_{\mathcal{M}}(x, y) \leq \|\psi_1(x) - \psi_2(y)\|_2^p \leq (1 + \varepsilon)\alpha \cdot d_{\mathcal{M}}(x, y).$$

Proof of Theorem 4.1. First boost the error probability of the sketching to be at most δ (small, to be fixed later). This increases the sketch size by a factor of $O(\log 1/\delta)$.

The high-level proof is as follows. Fix $I = \lceil \log_{1+\varepsilon} \Delta \rceil + 1$. In step 1, we construct embedding $\phi_{j,i}$ for all $j \in \{1, 2\}$ and $i \in [I]$, from the sketching algorithm for the radii $r_i = (1 + \varepsilon)^{i-1}$. In step 2, we construct $\phi'_{j,i}$ from $\phi_{j,i}$. In the third final step, we construct each of ψ_j, for $j \in \{1, 2\}$, by concatenating $\phi_{j,i}$ for $i \in [I]$.

Step 1. Fix $i \in [I]$ and radius $r_i = (1 + \varepsilon)^{i-1}$. We construct functions $\phi_{j,i}$, viewed as an embedding into L_1, which will satisfy

$$1 - 1/S \leq \|\phi_{1,i}(x) - \phi_{2,i}(y)\|_1 \leq 1$$

as well as the following guarantees. If $d(x, y) \leq r_i$ then $\|\phi_{j,i}(x) - \phi_{j,i}(y)\|_1 \leq 1 - (1 - \delta)/S$. If $d(x, y) > \alpha r_i$ then $\|\phi_{j,i}(x) - \phi_{j,i}(y)\|_1 \geq 1 - \delta/S$.

We obtain $\phi_{j,i}$ from the assumed sketching algorithm. Fix $S \triangleq 2^{sO(\log 1/\delta)} = (1/\delta)^{O(s)}$ to be the number of different messages Alice or Bob can transmit. Let R be the random string that is used by Alice, Bob, and the referee. We will set $\phi_{j,i} = (\frac{\Pr[R]}{2S} \cdot \phi^R_{j,i})_R$, where each $\phi^R_{j,i}(x) \in \{-1, 0, 1\}^{S^2}$. Fixing R, all algorithms are deterministic. Let A^R be the binary matrix of size $S \times S$ that describes the output of the referee: $A^R_{i,j} = 1$ if the referee outputs "close" on receiving message i from Alice and j from Bob; $A^R_{i,j} = 0$ otherwise. For each possible value of R, we generate $S \times S$ coordinates, a coordinate corresponding to a tuple of message from Alice (treated as a number from 0 to $S - 1$) and a message from Bob. For $\phi^R_{1,i}(x)$, we set precisely one entire row (corresponding to Alice's message) to 1 and the rest to 0. For $\phi^R_{2,i}(x)$, we set the column indexed by Bob's message to be equal to a vector $\in \{\pm 1\}^S$; in particular, a coordinate of this column is 1 if the corresponding entry of A^R is 1 and it is -1 otherwise.

Step 2. For some $k > 0$, we construct $\phi'_{j,i}$ such that $\|\phi'_{1,i}(x) - \phi'_{2,i}(y)\|_2^2 = 1 - e^{-k \cdot \|\phi_{1,i}(x) - \phi_{2,i}(y)\|_1}$. This is possible due to Schoenberg's theorem [Sch38] (constructively it can be obtained as in [RR07]).

We use $k = 2 \ln SI$. Let $t = e^{-k}$ and assume $t < .5$.

Step 3. Construct ψ_j by concatenating scaled $\phi_{j,i}$: $\psi_j = \exp\left[\frac{1}{2} t^{1-1/S}\right] \frac{1}{I} (\phi'_{j,i})_{i \in [I]}$. We set $p = \frac{I \ln(1+\varepsilon)}{t^{1-(1-\delta)/S} - t}$ and $\delta = 1/I$. Then we have $S = I^{O(s)}$ and then $p = I \cdot \ln(1+\varepsilon) \cdot e^k \cdot S/k = (I/\varepsilon)^{O(s)}$. We defer the analysis of the distortion of the resulting embedding to the appendix C. $\square$

We note that the embedding necessarily has to be *asymmetric* if we target to obtain the value of p as in the theorem. Namely, there exists a metric that has $O(1)$ sketch for exact DE but embedding (symmetrically) it into $(\ell_2)^p$ requires either $2 - o(1)$ distortion or $p = \Omega(n)$. We note that Deza-Maehara [DM90] proved that there exists an n-point metric that requires power $p = \Omega(n)$ to be embedded isometrically into $(\ell_2)^p$. In contrast, we prove that the same statement holds even if we allow a distortion of nearly 2. We prove the statement for the graph $K_{n,n}$. The proof of the following observation is in Appendix C.

Observation 4.2. *Let $p(n, D)$ be the minimum p such that $K_{n,n}$ embeds into $(\ell_2)^p$ with distortion $\leq D$. Then, for all $\varepsilon > 0$, $p(n, 2 - \varepsilon) \geq \Omega_\varepsilon(n)$.*

Acknowledgments

We would like to thank Piotr Indyk and T.S. Jayram for helpful discussions.

References

[AIK08] Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer. Earth mover distance over high-dimensional spaces. January 2008.

[AK07] A. Andoni and R. Krauthgamer. The computational hardness of estimating edit distance. In *48th Annual Symposium on Foundations of Computer Science*, 2007.

[AMS96] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Proceedings of the Symposium on Theory of Computing*, pages 20–29, 1996.

[AR98] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301, 1998.

[BDS07] S. Buyalo, A. Dranishnikov, and V. Schroeder. Embedding of hyperbolic groups into products of binary trees. *Inventiones Mathematicae*, 169(1), 2007.

[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.

[Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[BS05] S. Buyalo and V. Schroeder. Embedding of hyperbolic spaces in the product of trees. *Geometriae Dedicata*, 113:75–93, 2005.

[BYJKK04] Z. Bar-Yossef, T. S. Jayram, Robert Krauthgamer, and Ravi Kumar. Approximating edit distance efficiently. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 550–559. IEEE, October 2004.

[Cha02] M. Charikar. Similarity estimation techniques from rounding. *Proceedings of the Symposium on Theory of Computing*, 2002.

[CM02] Graham Cormode and S. Muthukrishnan. The string edit distance matching problem with moves. In *Proceedings of the 13th annual ACM-SIAM symposium on Discrete algorithms*, pages 667–676. SIAM, 2002.

[CPSV00] Graham Cormode, Mike Paterson, Suleyman Cenk Sahinalp, and Uzi Vishkin. Communication complexity of document exchange. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 197–206, 2000.

[DM90] M. Deza and H. Maehara. Metric transforms and Euclidean embeddings. *Trans. Amer. Math. Soc.*, 317(2):661–671, 1990.

[GIM07] Sudipto Guha, Piotr Indyk, and Andrew McGregor. Sketching information divergences. In *Proceedings of the Annual Conference on Learning Theory*, pages 424–438, 2007.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science*, pages 534–543, 2003.

[GP03] C. Gavoille and C. Paul. Optimal distance labeling for interval and circular-arc graphs. In *11th Annual European Symposium*, volume 2832 of *Lecture Notes in Computer Science*, pages 254–265. Springer, 2003.

[IM98] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *30th Annual ACM Symposium on Theory of Computing*, pages 604–613, May 1998.

[IM03] P. Indyk and J. Matoušek. Low distortion embeddings of finite metric spaces. *CRC Handbook of Discrete and Computational Geometry*, 2003.

[Ind03] P. Indyk. Nearest neighbors in high-dimensional spaces. *CRC Handbook of Discrete and Computational Geometry*, 2003.

[IT03] P. Indyk and N. Thaper. Fast color image retrieval via embeddings. *Workshop on Statistical and Computational Theories of Vision (at ICCV)*, 2003.

[KNR99] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

[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 (electronic), 2000.

[KSW04] J. M. Kleinberg, A. Slivkins, and T. Wexler. Triangulation and embedding using small sets of beacons. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 444–453, 2004.

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

[LM00] N. Linial and A. Magen. Least-distortion Euclidean embeddings of graphs: products of cycles and expanders. *J. Combin. Theory Ser. B*, 79(2):157–171, 2000.

[Mat02] J. Matousek. *Lectures on Discrete Geometry*. Springer, 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.

[McG06] Andrew McGregor. Open problems in data streams and related topics. *IITK Workshop on Algorithms For Data Streams*, 2006. Available at `http://www.cse.iitk.ac.in/users/sganguly/workshop.html`.

[MN07] M. Mendel and A. Naor. Ramsey partitions and proximity data structures. *J. Eur. Math. Soc.*, 9(2):253–275, 2007.

[MS00] S. Muthukrishnan and S. C. Sahinalp. Approximate nearest neighbors and sequence comparisons with block operations. In *32nd Annual ACM Symposium on Theory of Computing*, pages 416–424, 2000.

[NS06] A. Naor and G. Schechtman. Planar earthmover is not in l_1. *Proceedings of the Symposium on Foundations of Computer Science*, 2006.

[Pel00] D. Peleg. Proximity-preserving labeling schemes. *Journal of Graph Theory*, 33(3):167–176, 2000.

[PS89] D. Peleg and A. A. Schäffer. Graph spanners. *J. Graph Theory*, 13(1):99–116, 1989.

[RR07] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, 2007.

[Sch38] I. J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536, November 1938.

[SS02] Michael Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the Symposium on Theory of Computing*, pages 360–369, 2002.

[TZ04] M. Thorup and Y. Zhang. Tabulation based 4-universal hashing with applications to second moment estimation. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2004.

[Woo04] D. Woodruff. Optimal space lower bounds for all frequency moments. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 167–175, 2004.

[Yao82] A. C. C. Yao. On constructing minimum spanning trees in k-dimensional spaces and related problems. *SIAM Journal of Computing*, 22:721–736, 1982.

A Open Problems and Discussion

We mention some concrete problems derived from the meta-question.

1. *Is there a family of n-point metrics and approximation $\alpha(n)$ such that the sketching complexity differs its one-way complexity by more than a constant factor? Same question but for sketching vs. two-way complexity?*

   At the moment we do not have any separation between sketching and one-way communication complexity of distance estimation. Note that, for general functions (not distance estimation), such separations are known [BJKS04].

2. *Is there a sketching algorithm with $O(\frac{\log n}{\alpha})$ communication for a general metric?*

   A positive answer would improve Theorem 2.1 from one-way to sketching protocols, and obtain tight results for the latter. Currently the two extreme cases, $\alpha = 1$ and $\alpha = \log n$, are easily achieved. A negative answer would resolve also Problem 1.

3. *Fix a metric $\mathcal{M}$ and suppose that $\bigoplus_{\mathcal{M}}^k$ is sketchable with approximation α using $O(s)$ space, for sufficiently large k. Can the distortion of an (asymmetric) embedding of $\mathcal{M}$ into squared-ℓ_2 be upper bounded by $f(s) \cdot \alpha$?*

   A positive answer would mean that non-embeddability of a metric $\mathcal{M}$ implies lower bounds for sketching $\bigoplus \mathcal{M}$. For example, it would (roughly) imply that planar Earth-Mover Distance (EMD) requires $\omega(1)$ sketch complexity for $O(1)$ approximation (using the non-embeddability result of [NS06] and the fact that EMD is closed under taking sum-products).

   Put differently, we ask whether a solution for a specific problem in a specific model (embedding $\mathcal{M}$ into squared-ℓ_2) can always be obtained from a solution for a *harder* problem ($\bigoplus \mathcal{M}$) in a *more general* model (sketching). A more straightforward question is whether sketching $\mathcal{M}$ (as opposed to $\bigoplus \mathcal{M}$) implies embeddability of $\mathcal{M}$ into squared-ℓ_2. We suspect the answer to the latter question is negative, say there might be a metric in $(\ell_2)^{10}$ (thus admits efficient sketching) that cannot be embedded with $O(1)$ distortion into squared-ℓ_2.

4. *If the sketching complexity of an n-point metric $\mathcal{M}$ is $\Omega(\frac{\log n}{\alpha})$ for all $1 \leq \alpha \leq \log n$, is it necessarily true that embedding $\mathcal{M}$ into squared-ℓ_2 requires $\tilde{\Omega}(\log n)$ distortion?*

   A similar question regarding embedding into ℓ_1 is open as well.

B Proofs from Section 3

Proof of Theorem 3.1. We prove here only the upper bound on the sketch size. The lower bound is proved (as a special case) in Theorem 3.3. Without loss of generality, we shall assume that εr is an integer and that $\varepsilon \leq 1$.

The Sketching Algorithm. We now describe Alice's algorithm for her input $x \in P$; Bob's algorithm for his input $y \in P$ is identical. Alice first uses the shared randomness to choose uniformly at random an "offset" value $v \in [8r]$, and rounds down $x+v$ to its nearest multiple of $8r$, denoting this last value by x'. She then uses the shared randomness to choose a random function h (for better efficiency, she may choose h as a hash function from an approximate pairwise-independent hash family) with range $[\frac{20}{\varepsilon}]$. Next, Alice computes her sketch to be $\sigma(x) = \lfloor \frac{x+v-x'}{\varepsilon r/4} + h(x') \rfloor$, and sends this value to the referee. Note that $0 \le x+v-x' < 8r$ and $0 \le \sigma(x) < \frac{8r}{\varepsilon r/4} + 1 + \frac{20}{\varepsilon} \le O(\frac{1}{\varepsilon})$, and hence the sketch size is $|\sigma(x)| \le O(\log \frac{1}{\varepsilon})$.

The referee's algorithm is very simple. Given the two sketches $a = \sigma(x)$ and $b = \sigma(y)$, he decides that $d_P(x,y) \le r$ if and only if $|a-b| \le \frac{4}{\varepsilon} + 1$.

Correctness. It remains to prove that the referee's decision is correct with high probability. Fix first $x, y \in P$ with $d_P(x,y) \le r$. Then clearly $\Pr_v[x' \neq y'] = \frac{d_P(x,y)}{8r} \le 1/8$. In other words, with probability at least $7/8$, v is chosen such that $x' = y'$. Assuming this is the case, we have $h(x') = h(y')$ and $|(x+v-x')-(y+v-y')| = |x-y| \le r$, implying that $|\sigma(x)-\sigma(y)| \le \frac{r}{\varepsilon r/4}+1 = \frac{4}{\varepsilon}+1$, and thus the referee decides correctly that $d_P(x,y) \le r$.

Fix next $x, y \in P$ with $d_P(x,y) > (1+\varepsilon)r$. We now have two cases: (i) $x' = y'$, and thus $|\sigma(x)-\sigma(y)| > \frac{(1+\varepsilon)r}{\varepsilon r/4} - 1 = \frac{4}{\varepsilon}+3$, implying that the referee correctly decides that $d_P(x,y) > (1+\varepsilon)r$. (ii) $x' \neq y'$ and then using the independence between $h(x')$ and $h(y')$ we get that $|\sigma(x)-\sigma(y)| \le \frac{4}{\varepsilon}+1$ occurs with probability at most $(\frac{4}{\varepsilon}+1)/(\frac{20}{\varepsilon}) \le 1/4$, i.e. the referee decides correctly with probability at least $3/4$. □

Proposition B.1. *Let $\mathcal{M} = N_1 \times \cdots \times N_k$ be the sum-product of k metrics. Suppose there exist $\alpha \ge 1$ and s_α such that for every $r > 0$ and every metric N_i, the distance estimation problem $\mathrm{DE}_{N_i,r,\alpha}$ admits a sketching protocol with communication s_α. Then for every $r > 0$ and $\varepsilon \le 1$, there is a sketching protocol for the problem $\mathrm{DE}_{M,r,(1+\varepsilon)\alpha}$ that has $\tilde{O}(\frac{\alpha k}{\varepsilon}) \cdot s_\alpha$ communication.*

Proof of Proposition B.1. Let x, y denote the inputs to Alice and Bob, respectively. For each metric N_i, Alice, Bob, and the referee execute the assumed sketching algorithm applied to x_i, y_i with several possible parameter settings; specifically, for $\varepsilon' = \varepsilon/3$ and every value $r' = r, r/(1+\varepsilon'), r/(1+\varepsilon)^2, \ldots, \frac{\varepsilon' r}{\alpha k}$, they execute this algorithm with parameters r' and ε'. For each metric N_i we get $\log_{1+\varepsilon'}(\alpha k/\varepsilon')$ executions, which adds up to $t = O(\frac{\alpha k}{\varepsilon'} \log \frac{\alpha k}{\varepsilon'})$ executions. We shall assume in the sequel that the error probability of each of these algorithms is at most $1/4t$ (which is easily achieved by repeating each such algorithm $O(\log t)$ times). Alice and Bob run all these sketching algorithms simultaneously, and concatenate the resulting sketches, which yields a total sketch-size $s_\alpha \cdot O(t \log t) = s_\alpha \cdot O(\frac{\alpha k}{\varepsilon} \log^2 \frac{\alpha k}{\varepsilon})$.

After completing all these executions, the referee performs the following steps. (1) If there is a metric N_i where the protocol applied with $r' = r$ reported that $d_{N_i}(x_i, y_i) > \alpha r'$, then the referee decides that $d_M(x,y) > \alpha r$ (and halts). (2) The referee identifies all metrics N_i for which the protocol applied with $r' = \frac{\varepsilon' r}{\alpha k}$ reported that $d_{N_i}(x_i, y_i) \le r'$, sets his estimate for each of these distances to be $\rho_i = \alpha r' = \frac{\varepsilon' r}{k}$; clearly, ρ_i is an estimate from above for $d_{N_i}(x_i, y_i)$. (3) In all other metrics N_i, the referee may assume that $\frac{\varepsilon' r}{\alpha k} < d_{N_i}(x_i, y_i) \le \alpha r$.[2] Under this last assumption, the referee can determine an upper estimate $\rho_i$for $d_{N_i}(x_i, y_i)$ that has multiplicative accuracy $\alpha(1+\varepsilon')$. Indeed, the referee finds the smallest value r' in which the algorithm reported that $d_{N_i}(x_i, y_i) \le r'$, and sets his estimate to be $\rho_i = \alpha r'$; by the outcome for this value of r', it cannot be the case that

[2] This is true with high probability, because a too large value would have been noticed in step 1, and a too small value would have been noticed in step 2.

$d_{N_i}(x_i, y_i) > \alpha r' = \rho_i$, and by the outcome for the "next smaller" value $r'/(1+\varepsilon')$, it cannot be the case that $d_{N_i}(x_i, y_i) \le r'/(1+\varepsilon)' = \frac{\rho_i}{(1+\varepsilon')\alpha}$. (4) Finally, the referee decides that $d_M(x,y) \le r$ if and only if his estimates from steps 2 and 3 satisfy $\sum_{i=1}^k \rho_i \le (1+2\varepsilon')\alpha r$.

It remains to prove that with high probability the referee's output is correct. Each of the t algorithms is correct with probability at least $1 - 1/4t$; applying a union bound, with probability at least 3/4 they are all correct, and in the sequel, we shall assume that this is indeed the case. Observe that if the referee halts in step 1, then the protocol succeeds because necessarily $d_{N_i}(x_i, y_i) > r$. Assume now that the referee does not halt in step 1. Then his estimates ρ_i in step 2 have total additive error at most $k \cdot \frac{\varepsilon' r}{k} \le \varepsilon' r$, an the estimates in step 3 have (separately and thus in total) a multiplicative error of at most $(1+\varepsilon')\alpha$. It follows that whenever $d_M(x,y) \le r$, the referee's sum in step 4 will be

$$\sum_i \rho_i \le \varepsilon' r + (1+\varepsilon')\alpha d_M(x,y) \le (1+2\varepsilon')r.$$

And whenever $d_N(x,y) > (1+\varepsilon)\alpha r$, the referee's sum in step 4 will be $\sum_i \rho_i \ge \sum_i d_{N_i}(x_i, y_i) > (1+\varepsilon)\alpha r$. In both cases, the protocol succeeds. We have thus shown a sketching algorithm that uses sketch size $O(\frac{k}{\varepsilon} \log k \log \frac{1}{\varepsilon}) = \tilde{O}(\frac{k}{\varepsilon})$ and succeeds with probability at least 3/4. □

Proposition B.2. *Let $\mathcal{M} = N_1 \times \cdots \times N_k$ be the sum-product of k metrics. Suppose there exist $1 \le \alpha \le 2$ and s_α such that for every $r > 0$ and every metric N_i, there is a sketching protocol that, using s_α communication, finds $\rho \in [\alpha r/2, \alpha r]$ such that*

$$\Pr[\min\{d_{N_i}(x_i, y_i), \alpha r\} \le \rho \le \alpha \cdot \max\{d_{N_i}(x_i, y_i), r/2\}] \ge 3/4.$$

Then for every $r > 0$ and $\varepsilon \le 1$, there is a sketching protocol for the problem $\mathrm{DE}_{M,r,(1+\varepsilon)\alpha}$ that has $\tilde{O}(k \log \frac{1}{\varepsilon}) \cdot s_\alpha$ communication.

Proof of Proposition B.2. The proof is similar to that of Proposition B.1, and we only describe the differences. For each metric N_i, Alice, Bob and the referee use fewer values of r', namely, $r' = r, r/2, \ldots, \varepsilon' r/k$, hence $t = O(k \cdot \log \frac{k}{\varepsilon'})$. Again, we assume that each of these has success probability at least $1 - 1/3t$ by employing $O(\log t)$ repetitions. It follows that with high probability all these t executions succeed, and we assume in the sequel that this last event indeed happens. Finally, observe that the assumed sketching algorithm (with various values r') maintains the three properties we used in the aforementioned proof:

1. If $r'/2 \le d_{N_i}(x_i, y_i) \le \alpha r'$, the algorithm's output ρ provides an upper estimate on $d_{N_i}(x_i, y_i)$ within multiplicative accuracy α.
2. If $d_{N_i}(x_i, y_i) \le r'/2$, the algorithm reports $\rho = \alpha r'/2$.
3. If $d_{N_i}(x_i, y_i) \ge \alpha r'$, the algorithm reports $\rho = \alpha r'$.

The total sketch size is $O(t \log t) \cdot s_\alpha = O(k \cdot \log \frac{k}{\varepsilon} \cdot (\log k + \log \log \frac{1}{\varepsilon})) \cdot s_\alpha$. □

Proof of Theorem 3.3. We prove the theorem for product of paths and product of trees separately.

Product of paths. The upper bound follows from Proposition B.2, by plugging in the sketching algorithm of Theorem 3.1. We choose $\alpha = 1 + \varepsilon$, and obtain a sketching algorithm for $\mathrm{DE}_{M,r,(1+\varepsilon)^2}$. An inspection of the proof of the latter shows its sketching algorithm not only solves a relaxed (gap) decision problem, but rather provides the stronger guarantee required by the proposition (e.g. whenever the distance is in the range $[r/2, r]$, the algorithm can provide an additive $O(\varepsilon r)$ approximation). Since $s_\alpha = O(\log \frac{1}{\varepsilon})$, this immediately yields sketch size $O(k \log^2(\frac{k}{\varepsilon}) \cdot s_\alpha) \le O(k \log^3(\frac{k}{\varepsilon}))$.

This can be further improved as follows. It is easy to verify that the $O(\log t)$ term in the proof of Proposition B.2 is not really needed, because by changing some constants in the proof of Theorem 3.1, one can reduce the error probability to be at most ε^2. Furthermore, a $1+\varepsilon$ multiplicative accuracy in the entire range $[\varepsilon r/k, r]$ can be similarly obtained with error probability ε^2/k using sketch-size $O(\log \frac{k}{\varepsilon})$. This would lead to a total sketch size $O(k \cdot \log \frac{k}{\varepsilon})$.

We proceed to proving the lower bound. Notice that $M = \ell_1^k$, i.e. k-dimensional ℓ_1. The bound of $\Omega(k)$ for $k \leq 1/\varepsilon^2$ follows from the lower bound for ℓ_1 [Woo04]. Now assume that $2k \leq 1/\varepsilon$.

Assume there exists a one-way protocol that estimates the distance within factor $1+\varepsilon$ using sketch size $s = s(k, \varepsilon)$. We will use it to construct a protocol for indexing on $L = k \log \frac{1}{\varepsilon k}$ bits. In the indexing problem, Alice gets a bit-string x' of length L, and Bob gets an index $j \in [L]$. Let Alice "break" her input into k equal-size blocks, and consider each block as a binary representation of an integer between 1 and $v = \frac{1}{\varepsilon k}$. This gives rise to a point $x \in M$. Alice could then execute the assumed protocol for distance estimation in M and send a message to Bob.

The idea is that Bob can pretend to have any (fixed) point $y \in M$, or even t points in M if we were to increase the success probability to $1 - 1/3t$ (by $O(\log t)$ repetitions, which increases the sketch size accordingly). In particular, Bob can use the point y that is 1 in all but the jth coordinate, and try different values for the jth coordinate. Notice that the value for Bob's jth coordinate that gives the smallest ℓ_1 distance between x and y is exactly the jth coordinate of x. This value has $\log \frac{1}{\varepsilon k}$ bits, and Bob can find this value using $\log \frac{1}{\varepsilon k}$ iterations, revealing one bit at a time. (Note: This requires decreasing the error probability by using repetitions, i.e. increasing the sketch size s by an $O(\log \log \frac{1}{\varepsilon k})$ factor.)

Strictly speaking, Bob cannot evaluate the distance $d_M(x, y)$, but only apply a approximate decision version with respect to a fixed value r. But we can get around this shortcoming as follows. Let us actually use $2k$ dimensions, i.e. $M = \ell_1^{2k}$, where the first k dimensions are as before (i.e., Alice reads them off her input x' and Bob sets all of them but the jth coordinate to 1.) Alice will then let dimensions $k+1, ..., 2k$ be the same as dimensions $1, ..., k$, respectively, and Bob will set his coordinates $k+1, ..., 2k$ to be $1+v$ (which is about the largest possible value for Alice). Notice if Bob were to set his jth coordinate to 1, i.e. $y_j = 1$, then regardless of x we would get $d_M(x, y) = k \cdot v = 1/\varepsilon$ (because coordinates 1 and $k+1$ contribute together $|x_1 - 1| + |1 + v - x_1| = v$, and so forth). We will set $r = 1/\varepsilon$, thus $(1+\varepsilon)r = r + 1$, implying that a $1+\varepsilon$ approximation can distinguish an additive increase of 1 in the distance $d_M(x, y)$. Bob can now figure out the jth coordinate of x (with high probability) iteratively bit by bit (recall it is represented by $\log \frac{1}{\varepsilon k}$ bits). The precise details are somewhat complicated to describe (see also the proof in [KNR99] for the Greater-Than problem), and we will only show as an example how Bob can reveal the most significant bit (MSB) of x_j. Let Bob could plug in y_j the value $1+v$ instead of 1; now, if $d_M(x, y) \leq r$ then x_j is closer to $1+v$ than to 1, hence the MSB of x_j is 1. (We may need to assume here that $1+v$ is a power of 2.) To apply such an argument in subsequent iterations, we might need an additional coordinate $2k+1$ that will be used to "offset" the target distance to be exactly r.

We thus obtain a one-way communication protocol that uses $O(s \cdot \log \log \frac{1}{\varepsilon k})$ bits to solve the indexing problem. But since there is a known $\Omega(L)$ lower bound for this problem, we conclude that $s \geq \Omega(\frac{k \log(1/\varepsilon k)}{\log \log(1/\varepsilon k)})$.

Product of Paths. The upper bound follows from Proposition B.2, by plugging in the sketching algorithm of Theorem 3.2. We choose $\alpha = 1+\varepsilon$, and obtain a sketching algorithm for $\mathrm{DE}_{M,r,(1+\varepsilon)^2}$. An inspection of the proof of the latter shows its sketching algorithm not only solves a relaxed (gap) decision problem, but rather provides the stronger guarantee required by the proposition (e.g. whenever the distance is in the range $[r/2, r]$, the algorithm can provide an additive $O(\varepsilon r)$ approximation). Since $s_\alpha = O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$, this immediately yields sketch size $O(k\log^2(\frac{k}{\varepsilon})\cdot s_\alpha) \leq O(\frac{k}{\varepsilon}\log^3(\frac{k}{\varepsilon}))$.

This can be further improved as follows. It is easy to verify that the $O(\log t)$ term in the proof of Proposition B.2 is not really needed, because by changing some constants in the proof of Theorem 3.2, one can reduce the error probability to be at most ε^2. This would lead to a total sketch size $O(k\cdot\log\frac{k}{\varepsilon}\cdot s_\alpha) \leq O(\frac{k}{\varepsilon}\log^2(\frac{k}{\varepsilon}))$. □

C Proofs from Section 4

Completion of the proof of Theorem 4.1. It is actually more convenient for the analysis to first define ψ_j's as

$$\psi_j = \tfrac{1}{I}(\phi'_{j,i})_{i\in[I]}.$$

Define $u = \lfloor \log_{1+\varepsilon} d(x,y)\rfloor$, and $a = \lfloor \log_{1+\varepsilon}\alpha\rfloor$, and let $m = u + a + 2$. Then, ψ_j's satisfy:

$$\|\psi_1(x)-\psi_2(y)\|_2^2 \leq \tfrac{1}{I}\Big((I-m)\cdot(1-t^{1-(1-\delta)/S}) + m\cdot(1-t)\Big) = 1 - t^{1-(1-\delta)/S} + \tfrac{m}{I}(t^{1-(1-\delta)/S}-t)$$

$$\|\psi_1(x)-\psi_2(y)\|_2^2 \geq \tfrac{1}{I}\Big((I-u)\cdot(1-t^{1-1/S}) + u\cdot(1-t^{1-\delta/S})\Big) = 1 - t^{1-1/S} + \tfrac{u}{I}t^{-\delta/S}(t^{1-(1-\delta)/S}-t).$$

Raising $\|\psi_1(x)-\psi_2(y)\|_2^2$ to power $p = \frac{I\ln(1+\varepsilon)}{t^{1-(1-\delta)/S}-t}$, we get the following upper bound

$$\|\psi_1(x)-\psi_2(y)\|_2^{2p} \leq \exp\left[-pt^{1-(1-\delta)/S} + p\tfrac{m}{I}(t^{1-(1-\delta)/S}-t)\right] \approx \exp\left[-pt^{1-(1-\delta)/S}\right]\alpha d(x,y). \quad (1)$$

where $\approx$ means up to approximation $1+O(\varepsilon)$. For the lower bound, we use the approximation $1-\xi \geq e^{-\xi-2\xi^2}$ for $\xi \leq .5$, to get

$$\|\psi_1(x)-\psi_2(y)\|_2^{2p} \geq \exp\left[-pt^{1-1/S}\right](d(x,y))^{t^{-\delta/S}}\exp\left[-2p(t^{1-1/S})^2\right]. \quad (2)$$

Note that $d(x,y)^{t^{-\delta/S}} \geq d(x,y)$. We approximate $t^{1-f/S}$, for $f\in[0,1]$, as $t(1+kf/S) \leq t^{1-f/S} = te^{kf/S} \leq t(1+kf/S+2k^2/S^2)$. Then, the approximation of the resulting embedding is the ratio between RHS of Eqn (1) and RHS of Eqn. (2):

$$\beta \leq \alpha\exp\left[\ln(1+\varepsilon)I\frac{-t^{1-(1-\delta)/S}+t^{1-1/S}}{t^{1-(1-\delta)/S}-t}\right]\cdot\exp\left[-2p(t^{1-1/S})^2\right]$$

$$\implies \beta \leq \alpha\exp\left[\ln(1+\varepsilon)I\cdot\frac{k/S-(1-\delta)k/S+\frac{2k^2}{S^2}}{1+(1-\delta)k/S-1-\frac{2k^2}{S^2}}\right](1+\varepsilon)^{4I\sqrt{t}/(k/S)}\beta \leq \alpha(1+\varepsilon)^{O(\delta I)}\cdot(1+\varepsilon)^{4IS\sqrt{t}/k}.$$

We set $\delta = 1/I$. Also note that $4IS\sqrt{t}/k = O(e^{-k/2}\cdot IS/k) < 1$. Concluding, we have $S = I^{O(s)}$ and then $p = I\cdot\ln(1+\varepsilon)\cdot e^k\cdot S/k = (I/\varepsilon)^{O(s)}$. To obtain the stated results we rescale the embeddings ψ_j's by $\exp\left[\frac{1}{2}t^{1-1/S}\right]$, as well as the values of p and ε. □

Proof of Observation 4.2. Suppose there exists an embedding $\phi : K_{n,n} \to (\ell_2)^p$ of distortion $D < 2$. Let the vertices of $K_{n,n}$ be A_i and B_i, $i \in [n]$. Then there exists an embedding ψ of $(K_{n,n})^{2/p}$ into $(\ell_2)^2$ such that, for $u, v \in K_{n,n}$,

$$(d_{K_{n,n}}(u,v))^{2/p} \leq \|\psi(u) - \psi(v)\|_2^2 \leq (D \cdot d_{K_{n,n}}(u,v))^{2/p}.$$

We can construct a new embedding ψ' such that $\|\psi'(u) - \psi'(v)\|_2^2$, for $v \neq u$, takes only two values: "small distance" and "large distance". ψ' is obtained by concatenating $\psi \circ \mathcal{I}$ for all $\mathcal{I} : K_{n,n} \to K_{n,n}$ that are isometric automorphisms. Then, for $i \neq j$,

$$\|\psi'(A_i) - \psi'(A_j)\|_2^2 = \|\psi'(B_i) - \psi'(B_j)\|_2^2 = \sum_{i \neq j} (\|\psi(A_i) - \psi(A_j)\|_2^2 + \|\psi(B_i) - \psi(B_j)\|_2^2) \cdot n!(n-2)! \geq 2(n!)^2 2^{2/p},$$

and also, for all i, j,

$$\|\psi'(A_i) - \psi'(B_j)\|_2^2 = \sum_{i,j} \|\psi(A_i) - \psi(B_j)\|_2^2 \cdot 2(n-1)!(n-1)! \leq 2(n!)^2 D^{2/p},$$

But negative type inequality then says that for $b_{A_i} = +1$ and $b_{B_j} = -1$, we must have:

$$\sum_{i<j} b_{A_i} b_{A_j} 2^{2/p} + \sum_{i<j} b_{B_i} b_{B_j} 2^{2/p} + \sum_{i,j} b_{A_i} b_{B_j} D^{2/p} \leq 0,$$

or $2^{2/p} \cdot n(n-1) - D^{2/p} \cdot n^2 \leq 0$. This implies $D \geq 2(1 - 1/n)^{p/2}$. □

D Applications of Distance Estimation Protocols

D.1 Nearest Neighbor Search via Communication Protocols

We show that one can obtain algorithms for nearest neighbor search (NNS) problem from small complexity communication protocols of the distance estimation problem. In particular, below we show that upper bounds on one-way protocol for distance estimation implies NNS algorithm. The algorithm exhibits a smooth tradeoff between the space requirement and query time.

Lemma D.1. *Consider metric $\mathcal{M}$ for which distance estimation problem $\mathrm{DE}_{\mathcal{M},r,\alpha}$ admits one-way protocol of size s. Suppose Alice's computation time is τ_A, and time for distance computation in $\mathcal{M}$ is τ_d. Then, for any δ, with $0 < \delta \leq s$, there exists a NNS for $\mathcal{M}$ achieving approximation α with space $n^{1+O(\delta)}$ and query time $O(n^{1-\delta/s} \cdot \tau_d + \tau_A \log n)$. Preprocessing time is $O(n^{1+O(\delta)} \cdot \tau_B)$, where τ_B is Bob's computation time per message.*

Proof. Let $t = \log_2 n^{\delta/s}$. Boost the success probability of the one-way protocol by repeating it $\Theta(t) = \Theta(\frac{\delta}{s} \cdot \log n)$ times, decreasing the error probability to $2^{-t} = n^{\delta/s}$. A "hash table" is an index of all possible strings of length st (messages of the boosted protocol): for each possible string store all dataset points for which Bob would answer "close". The query algorithm just looks at the corresponding bucket and computes the distance all points in the bucket.

The space requirement is $2^{s\Theta(t)} \cdot n = n^{O(\delta)} \cdot n$ ($2^{s\Theta(t)}$ is the number of possible buckets). Expected query time is $n^{1-\delta/s}$ because, in expectation, a fraction of 2^{-t} of "far" points may be confused as "close" points, thus colliding with the query point. Thus, there are $(n-1) \cdot 2^{-t} \leq n^{1-\delta/s}$ points in the bucket in expectation. The true "close" point will be in the bucket with probability $\geq 1 - 2^{-t}$. □

D.2 Distance labeling via Communication Complexity

Let $\mathcal{F}$ be a family of finite metric spaces. A k-bit approximate distance labeling scheme for $\mathcal{F}$ provides (i) for every $\mathcal{M} \in \mathcal{F}$, a mapping $f_{\mathcal{M}} : \mathcal{M} \mapsto \{0,1\}^k$, and (ii) a decoder $g : \{0,1\}^k \times \{0,1\}^k \mapsto \mathbb{R}$, such that for all $\mathcal{M} \in \mathcal{F}$ and $x, y \in \mathcal{M}$, the distance $d_{\mathcal{M}}(x,y)$ is approximated by $g(f_{\mathcal{M}}(x), f_{\mathcal{M}}(y))$.

It is easy to see that distance labeling immediately implies a deterministic sketching protocol for the same metric, with the same approximation factor and the same size. Indeed, we simply let the sketch of a point be its label in the distance labeling algorithm, and let the referee perform the decoder's algorithm and decide accordingly. But this is generally not useful for finite metrics, because such a metric $\mathcal{M}$ always admits a trivial $O(\log|\mathcal{M}|)$-size sketching protocol.

Conversely, a one-way distance estimation protocol for an infinite metric $\mathcal{M}$ implies an approximate distance labeling scheme for the family of all finite submetrics of $\mathcal{M}$.

The next proposition shows that a one-way distance estimation protocol for an infinite metric $\mathcal{M}$ implies an approximate distance labeling scheme for the family of all finite submetrics of $\mathcal{M}$. Examples for infinite metric spaces that are relevant here include trees (binary or of unbounded degree), doubling metrics, the real hyperbolic space $\mathcal{H}^k$, and so forth.

Proposition D.2. *Let $\mathcal{M}$ be an infinite metric space and let $\alpha, s \geq 1$. Suppose that for every $r > 0$ there is a sketching protocol of size s for $\mathrm{DE}_{\mathcal{M},r,\alpha}$. Then for every $0 < \varepsilon \leq 1$, the family of finite submetrics of $\mathcal{M}$ admits a $(1+\varepsilon)\alpha$-approximate distance labeling scheme with $O(\frac{s}{\varepsilon} \cdot \log n \cdot \log(n\Delta))$-bit labels, where n is the size of the finite submetric and Δ is its diameter.*

Proof. Fix $\mathcal{M}$ and let $\mathcal{M}'$ be a finite submetric of it of size n. We show a probabilistic construction that succeeds with high probability. The label of each $x \in \mathcal{M}'$ contains, for each $r = 1, 1+\varepsilon, (1+\varepsilon)^2, \ldots, \Delta$ a concatenation of $O(\log(n\Delta))$ independent executions of the sketching protocol applied to x. (It Alice's and Bob's algorithms are different, then it includes both.) It is easy to see that the $O(\log(n\Delta))$ repetitions can be used (by a majority rule) to increase the sketching algorithm's success probability to $1 - 1/4n^2\Delta$. The distance labeling decoding function iterators over these values of r, reporting the smallest $r > 0$ for which the sketching algorithm determined that $d_{\mathcal{M}}(x,y) > \alpha r$. Using a union bound, it is easily verified that probability at least $3/4$, for all $x, y \in \mathcal{M}'$ the decoder's output is larger than $d(x,y)$ by at most a $(1+\varepsilon)\alpha$ factor. $\square$