Sketching and Embedding are Equivalent for Norms*

Alexandr Andoni†
Simons Institute, UC Berkeley
Berkeley, CA, USA
andoni@mit.edu

Robert Krauthgamer‡
Weizmann Institute
Rehovot, Israel
robert.krauthgamer@weizmann.ac.il

Ilya Razenshteyn
CSAIL, MIT
Cambridge, MA, USA
ilyaraz@mit.edu

ABSTRACT

An outstanding open question [51, Question #5] asks to characterize metric spaces in which distances can be estimated using *efficient sketches*. Specifically, we say that a sketching algorithm is efficient if it achieves constant approximation using constant sketch size. A well-known result of Indyk (J. ACM, 2006) implies that a metric that admits a constant-distortion embedding into ℓ_p for $p \in (0, 2]$ also admits an efficient sketching scheme. But is the converse true, i.e., is embedding into ℓ_p the only way to achieve efficient sketching?

We address these questions for the important special case of *normed spaces*, by providing an almost complete characterization of sketching in terms of embeddings. In particular, we prove that a finite-dimensional normed space allows efficient sketches if and only if it embeds (linearly) into $\ell_{1-\varepsilon}$ with constant distortion. We further prove that for norms that are closed under sum-product, efficient sketching is equivalent to embedding into ℓ_1 with constant distortion. Examples of such norms include the *Earth Mover's Distance* (specifically its norm variant, called Kantorovich-Rubinstein norm), and the *trace norm* (a.k.a. Schatten 1-norm or the nuclear norm). Using known non-embeddability theorems for these norms by Naor and Schechtman (SICOMP, 2007) and by Pisier (Compositio. Math., 1978), we then conclude that these spaces do not admit efficient sketches either, making progress towards answering another open question [51, Question #7].

Finally, we observe that resolving whether "sketching is equivalent to embedding into ℓ_1 for general norms" (i.e., without the above restriction) is *equivalent* to resolving a well-known open problem in Functional Analysis posed by Kwapien in 1969.

*See `arXiv:1411.2577` for the full version

†Work was done in part while all authors were at Microsoft Research Silicon Valley

‡Work supported in part by a US-Israel BSF grant #2010418 and an Israel Science Foundation grant #897/13

STOC'15, June 14–17, 2015, Portland, Oregon, USA.

ACM 978-1-4503-3536-2/15/06 ...$15.00.
DOI: http://dx.doi.org/10.1145/2746539.2746552.

1. INTRODUCTION

One of the most exciting notions in the modern algorithm design is that of *sketching*, where an input is summarized into a small data structure. Perhaps the most prominent use of sketching is to estimate *distances between points*, one of the workhorses of similarity search. For example, some early uses of sketches have been designed for detecting duplicates and estimating resemblance between documents [17, 18, 22]. Another example is Nearest Neighbor Search, where many algorithms rely heavily on sketches, under the labels of dimension reduction (like the Johnson-Lindenstrauss Lemma [37, 30]) or Locality-Sensitive Hashing (see e.g. [33, 45, 4]). Sketches see widespread use in streaming algorithms, for instance when the input implicitly defines a high-dimensional vector (via say frequencies of items in the stream), and a sketch is used to estimate the vector's ℓ_p norm. The situation is similar in compressive sensing, where acquisition of a signal can be viewed as sketching. Sketching—especially of distances such as ℓ_p norms—was even used to achieve improvements for *classical* computational tasks: see e.g. recent progress on numerical linear algebra algorithms [68], or dynamic graph algorithms [2, 40]. Since sketching is a crucial primitive that can lead to many algorithmic advances, it is important to understand its power and limitations.

A primary use of sketches is for *distance estimation* between points in a metric space (X, d_X), such as the Hamming space. The basic setup here asks to design a *sketching* function $\mathbf{sk} : X \to \{0, 1\}^s$, so that the distance $d_X(x, y)$ can be estimated given only the sketches $\mathbf{sk}(x), \mathbf{sk}(y)$. In the decision version of this problem, the goal is to determine whether the inputs x and y are "close" or "far", as formalized by the *Distance Threshold Estimation Problem* [61], denoted $\mathrm{DTEP}_r(X, D)$, where, for a threshold $r > 0$ and approximation $D \geq 1$ given as parameters in advance, the goal is to decide whether $d_X(x, y) \leq r$ or $d_X(x, y) > Dr$. Throughout, it will be convenient to omit r from the subscript.[1] Efficient sketches $\mathbf{sk}$ almost always need to be randomized, and hence we allow randomization, requiring (say) 90% success probability.

The diversity of applications gives rise to a variety of natural and important metrics M for which we want to solve DTEP: Hamming space, Euclidean space, other ℓ_p norms, the Earth Mover's Distance, edit distance, and so forth. Sketches for Hamming and Euclidean distances are now classic and well-understood [33, 45]. In particular,

[1]When X is a normed space it suffices to consider $r = 1$ by simply scaling the inputs x, y.

both are "efficiently sketchable": one can achieve approximation $D = O(1)$ using sketch size $s = O(1)$ (most importantly, independent of the dimension of X). Indyk [32] extended these results to efficient sketches for every ℓ_p norm for $p \in (0, 2]$. In contrast, for ℓ_p-spaces with $p > 2$, efficient sketching (constant D and s) was proved impossible using information-theoretic arguments [61, 12]. Extensive subsequent work investigated sketching of other important metric spaces,[2] or refined bounds (like a trade-off between D and s) for "known" spaces.[3]

These efforts provided beautiful results and techniques for many specific settings. Seeking a broader perspective, a foundational question has emerged [51, Question #5]:

QUESTION 1. *Characterize metric spaces which admit efficient sketching.*

To focus the question, efficient sketching will mean constant D and s for us. Since its formulation circa 2006, progress on this question has been limited. The only known characterization is by [31] for distances that are decomposable by coordinates, i.e., $d_X(x, y) = \sum_i \varphi(x_i, y_i)$ for some φ.

1.1 The embedding approach

To address DTEP in various metric spaces more systematically, researchers have undertaken the approach of metric embeddings. A *metric embedding* of X is a map $f : X \to Y$ into another metric space (Y, d_Y). The *distortion* of f is the smallest $D' \geq 1$ for which there exists a scaling factor $t > 0$ such that for every $x, y \in X$

$$d_Y(f(x), f(y)) \leq t \cdot d_X(x, y) \leq D' \cdot d_Y(f(x), f(y)).$$

If the target metric Y admits sketching with parameters D and s, then X admits sketching with parameters DD' and s, by the simple composition $\mathbf{sk}' : x \mapsto \mathbf{sk}(f(x))$. This approach of "reducing" sketching to embedding has been very successful, including for variants of the Earth Mover's Distance [22, 34, 24, 54, 5], and for variants of edit distance [13, 58, 23, 6, 29, 53, 28, 27]. The approach is obviously most useful when Y itself is efficiently sketchable, which holds for all $Y = \ell_p$, $p \in (0, 2]$ [32], and in fact the embeddings mentioned above are all into ℓ_1, except for [6] which employs a more complicated target space. We remark that in many cases the distortion D' achieved in the current literature is not constant and depends on the "dimension" of X.

Extensive research on embeddability into ℓ_1 has resulted in several important distortion lower bounds. Some address the aforementioned metrics [41, 54, 44, 8], while others deal with metric spaces arising in rather different contexts such as Functional Analysis [59, 25, 26], or Approximation Algorithms [49, 9, 43, 42]. Nevertheless, obtaining (optimal) distortion bounds for ℓ_1-embeddability of several metric spaces of interest, are still well-known open questions [50].

Yet sketching is a more general notion, and one may hope to achieve better approximation by bypassing embeddings into ℓ_1. As mentioned above, some limited success in bypassing an ℓ_1-embedding has been obtained for a variant of edit distance [6], albeit with a sketch size depending mildly on the dimension of X. Our results disparage these hopes, at least for the case of *normed spaces.*

1.2 Our results

Our main contribution is to show that efficient sketchability of norms is *equivalent* to embeddability into $\ell_{1-\varepsilon}$ with constant distortion. Below we only assert the "sketching $\implies$ embedding" direction, as the reverse direction follows from [32], as discussed above.

THEOREM 1. *Let X be a finite-dimensional normed space, and suppose that $0 < \varepsilon < 1/3$. If X admits a sketching algorithm for* $\mathrm{DTEP}(X, D)$ *for approximation $D > 1$ with sketch size s, then X linearly embeds into $\ell_{1-\varepsilon}$ with distortion $D' = O(sD/\varepsilon)$.*

One can ask whether it is possible to improve Theorem 1 by showing that X, in fact, embeds into ℓ_1. Since many non-embeddability theorems are proved for ℓ_1, such a statement would lift such results to lower bounds for sketches. Indeed, we show results in this direction too. First of all, the above theorem also yields the following statement.

THEOREM 2. *Under the conditions of Theorem 1, X linearly embeds into ℓ_1 with distortion $O(sD \cdot \log(\dim X))$.*

We would like however a stronger statement: efficient sketchability for norms is equivalent to embeddability into ℓ_1 with constant distortion (i.e., independent of the dimension of X as above). Such a stronger statement in fact requires the resolution of an open problem posed by Kwapien in 1969 (see [39, 14]). To be precise, Kwapien asks whether every finite-dimensional normed space X that embeds into $\ell_{1-\varepsilon}$ for $0 < \varepsilon < 1$ with distortion $D_0 \geq 1$ must also embed into ℓ_1 with distortion D_1 that depends only on D_0 and ε but not on the dimension of X (this is a reformulation of the finite-dimensional version of the original Kwapien's question). In fact, by Theorem 1, the "efficient sketching $\implies$ embedding into ℓ_1 with constant distortion" statement is *equivalent* to a positive resolution of the Kwapien's problem. Indeed, for the other direction, observe that a potential counter-example to the Kwapien's problem must admit efficient sketches by [32] but is not embeddable into ℓ_1.

To bypass the resolution of the Kwapien's problem, we prove the following variant of the theorem using a result of Kalton [39]: efficient sketchability is equivalent to ℓ_1-embeddability with constant distortion for norms that are *"closed" under sum-products.* A sum-product of two normed spaces X and Y, denoted $X \oplus_{\ell_1} Y$, is a normed space derived from $X \times Y$ by setting $\|(x, y)\| = \|x\| + \|y\|$. It is easy to verify that ℓ_1, the Earth Mover's Distance, and the trace norm are all closed under taking sum-products (potentially with an increase in the dimension). Again, we only need to show the "sketching $\implies$ embedding" direction, as the reverse direction follows from [32]. We discuss the application of this theorem to the Earth Mover's Distance in Section 1.3.

THEOREM 3. *Let $(X_n)_{n=1}^{\infty}$ be a sequence of finite-dimensional normed spaces. Suppose that for every $i_1, i_2 \geq 1$ there exists $m = m(i_1, i_2) \geq 1$ such that $X_{i_1} \oplus_{\ell_1} X_{i_2}$ embeds isometrically into X_m. Assume that every X_n admits a sketching*

---

[2] Other metric spaces include edit distance [10, 13, 58, 8] and its variants [29, 53, 28, 27, 23, 6], the Earth Mover's Distance in the plane or in hypercubes [22, 34, 24, 41, 5, 3], cascaded norms of matrices [36], and the trace norm of matrices [47].

[3] These refinements include the Gap-Hamming-Distance problem [67, 35, 19, 20, 21, 63, 65], and LSH in ℓ_1 and ℓ_2 spaces [52, 57].

algorithm for DTEP(X_n, D) *for fixed approximation* $D > 1$ *with fixed sketch size* s *(both independent of* n*). Then, every* X_n *linearly embeds into* ℓ_1 *with bounded distortion (independent of* n*).*

Overall, we almost completely characterize the norms that are efficiently sketchable, thereby making a significant progress on Question 1. In particular, our results suggest that the embedding approach (embed into ℓ_p for some $p \in (0, 2]$, and use the sketch from [32]) is essentially unavoidable for norms. It is interesting to note that for general metrics (not norms) the implication "efficient sketching $\Longrightarrow$ embedding into ℓ_1 with constant distortion" is false: for example the Heisenberg group embeds into ℓ_2-squared (with bounded distortion) and hence is efficiently sketchable, but it is not embeddable into ℓ_1 [46, 25, 26] (another example of this sort is provided by Khot and Vishnoi [43]). At the same time, we are not aware of any counter-example to the generalization of Theorem 1 to general metrics.

1.3 Applications

To demonstrate the applicability of our results to concrete questions of interest, we consider two well-known families of normed spaces, for which we obtain the first non-trivial lower bounds on the sketching complexity.

Trace norm. Let $\mathcal{T}_n$ be the vector space $\mathbb{R}^{n\times n}$ (all real square $n \times n$ matrices) equipped with the trace norm (also known as the nuclear norm and Schatten 1-norm), which is defined to be the sum of singular values. It is well-known that $\mathcal{T}_n$ embeds into ℓ_2 (and thus also into ℓ_1) with distortion $\sqrt{n}$ (observe that the trace norm is within $\sqrt{n}$ from the Frobenius norm, which embeds isometrically into ℓ_2). Pisier [59] proved a matching lower bound of $\Omega(\sqrt{n})$ for distortion of any embedding of $\mathcal{T}_n$ into ℓ_1.

This non-embeddability result, combined with our Theorem 2, implies a sketching lower bound for the trace norm. Before, only lower bounds for specific types of sketches (linear and bilinear) were known [47].

Corollary 1. *For any sketching algorithm for* DTEP$(\mathcal{T}_n, D)$ *with sketch size* s *the following bound must hold:*

$$sD = \Omega\left(\frac{\sqrt{n}}{\log n}\right).$$

The Earth Mover's Distance. The (planar) Earth Mover's Distance (also known as the transportation distance, Wasserstein-1 distance, and Monge-Kantorovich distance) is the vector space $\mathrm{EMD}_n = \{p \in \mathbb{R}^{[n]^2} : \sum_i p_i = 0\}$ endowed with the norm $\|p\|_{\mathrm{EMD}}$ defined as the minimum cost needed to transport the "positive part" of p to the "negative part" of p, where the transportation cost per unit between two points in the grid $[n]^2$ is the ℓ_1-distance between them (for a formal definition see [54]). It is known that this norm embeds into ℓ_1 with distortion $O(\log n)$ [34, 22, 54], and that any ℓ_1-embedding requires distortion $\Omega(\sqrt{\log n})$ [54].

We obtain the first sketching lower bound for EMD_n, which in particular addresses a well-known open question [51, Question #7]. Its proof is a direct application of Theorem 3 (which we *can* apply, since EMD_n is obviously closed under taking sum-products), to essentially "upgrade" the known non-embeddability into ℓ_1 [54] to non-sketchability. Strictly speaking, EMD_n is a generalization of the version of EMD metric commonly used in computer science applications: given two weighted sets $A, B \subset [n]^2$ of the same total weight, their EMD distance is the min-cost matching between A and B. Nevertheless we show in the full version that efficient sketching of EMD on weighted sets implies efficient sketching of the EMD norm. Hence, the non-sketchability of EMD_n norm applies to EMD on weighted sets as well.

Corollary 2. *No sketching algorithm for* DTEP(EMD_n, D) *can achieve approximation* D *and sketch size* s *that are constant (independent of* n*).*

The reason we can not apply Theorem 2 and get a clean quantitative lower bound for sketches of EMD_n is the factor $\log(\dim X)$ in the statement of Theorem 2. Indeed, the lower bound on the distortion of an embedding of EMD_n into ℓ_1 proved in [54] is $\Omega(\sqrt{\log n})$, which is smaller than $\log(\dim X) = \Theta(\log n)$.

1.4 Other related work

Another direction for "characterization" is one for streaming algorithms, where we are given a vector $x \in \mathbb{R}^n$ under updates of the form (i, δ), with the semantics that the coordinate i has to be increased by $\delta \in \mathbb{R}$.

There are two known results in this vein. First, [16] characterized the streaming complexity of computing the sum $\sum_i \varphi(x_i)$, for some fixed φ (e.g., $\varphi(x) = x^2$ for ℓ_2 norm), when the updates are positive. They gave a precise property of φ that determines whether the complexity of the problem is small. Second, [48] showed that, in certain settings, streaming algorithms may as well be *linear*, i.e., the sketch $f(x) = Ax$ for a matrix A. The size of the sketch is increased by a factor logarithmic in the dimension of x.

1.5 Proof overview

Following common practice, we think of sketching as a communication protocol. In fact, our results hold for protocols with an *arbitrary* number of rounds (and access to public randomness).

Our proof of Theorem 1 can be divided into two parts: *information-theoretic* and *analytic*. First, we use information-theoretic tools to convert an efficient *protocol* for DTEP(X, D) into a so-called *threshold map* from X to a Hilbert space. Our notion of a threshold map can be viewed as a very weak definition of embeddability (see Definition 4 for details). Second, we use techniques from nonlinear functional analysis to convert a threshold map to a *linear map* into $\ell_{1-\varepsilon}$.

Information-theoretic part. To get a threshold map from a protocol for DTEP(X, D), we proceed in three steps. First, using the fact that X is a *normed space*, we are able to give a good protocol for DTEP$(\ell_\infty^k(X), Dk)$ (Lemma 1). The space $\ell_\infty^k(X)$ is a product of k copies of X with the norm $\|(x_1, \ldots, x_k)\| = \max_i \|x_i\|$. Then, invoking the main result from [7], we conclude non-existence of certain Poincaré-type inequalities for X (Theorem 6, in the contrapositive).

Finally, we use convex duality together with a compactness argument to conclude the existence of a desired threshold map from X to a Hilbert space (Lemma 2, again in the contrapositive).

Analytic part. We proceed from a threshold map by upgrading it to a *uniform embedding* (see Definition 1) of X

into a Hilbert space (Theorem 7). For this we adapt arguments from [38, 60]. We use two tools from nonlinear functional analysis: an extension theorem for 1/2-Hölder maps from a (general) metric space to a Hilbert space [66] (Theorem 8), and a symmetrization lemma for maps from metric abelian groups to Hilbert spaces [1] (Lemma 4).

Then we convert a uniform embedding of X into a Hilbert space to a *linear* embedding into $\ell_{1-\varepsilon}$ by applying the result of Aharoni, Maurey and Mityagin [1] together with the result of Nikishin [56].

To prove a quantitative version of this step, we "open the black boxes" of [1] and [56], and thus obtain explicit bounds on the distortion of the resulting map. We accomplish this in the full version.

Embeddings into ℓ_1. To prove Theorem 2 (which has dependence on the dimension of X), we note it is a simple corollary of Theorem 1 and a result of Zvavitch [69], which gives a dimension reduction for subspaces of $\ell_{1-\varepsilon}$.

Norms closed under sum-product. Finally, we prove Theorem 3 — embeddability into ℓ_1 for norms closed under sum-product — by proving and using a finitary version of the theorem of Kalton [39] (Lemma 5), instead of invoking Nikishin's theorem as above. We prove the finitary version by reducing it to the original statement of Kalton's theorem via a compactness argument.

Let us point out that Naor and Schechtman [54] showed how to use (the original) Kalton's theorem to upgrade a uniform embedding of EMD_n into a Hilbert space to a linear embedding into ℓ_1 (they used this reduction to show uniform non-embeddability of EMD_n). Their proof used certain specifics of EMD. In contrast, to get Theorem 3 for general norms, we seem to need a finitary version of Kalton's theorem.

We also note that in Theorem 1, Theorem 2 and Theorem 3, we can conclude embeddability into $\ell_{1-\varepsilon}^d$ and ℓ_1^d respectively, where d is *near-linear* in the dimension of the original space. This conclusion uses the known dimension reduction theorems for subspaces from [64, 69].

2. PRELIMINARIES

We remind a few definitions and standard facts from functional analysis that will be useful for our proofs. A central notion in our proofs is the notion of *uniform embeddings*, which is a weaker version of embeddability.

Definition 1. *For two* metric spaces X *and* Y *we say that a map* $f\colon X \to Y$ *is a* uniform embedding*, if there exist two non-decreasing functions* $L, U\colon \mathbb{R}_+ \to \mathbb{R}_+$ *such that for every* $x_1, x_2 \in X$ *one has* $L(d_X(x_1, x_2)) \le d_Y(f(x_1), f(x_2)) \le U(d_X(x_1, x_2))$*,* $U(t) \to 0$ *as* $t \to 0$ *and* $L(t) > 0$ *for every* $t > 0$*. The functions* $L(\cdot)$ *and* $U(\cdot)$ *are called* moduli *of the embedding.*

Definition 2. *An* inner product space *is a real vector space* X *together with an* inner product $\langle \cdot, \cdot \rangle\colon X\colon X \to \mathbb{R}$*, which is a symmetric positive-definite bilinear form.*

Any inner product space is a normed space: we can set $\|x\| = \sqrt{\langle x, x \rangle}$. For a normed space X we denote B_X its closed unit ball.

Definition 3. *A* Hilbert space X *is an inner product space that is* complete *as a metric space.*

The main example of a Hilbert space is ℓ_2: the space of all real sequences $\{x_n\}$ with $\sum_i x_i^2 < \infty$, where the dot product is defined as

$$\langle x, y \rangle = \sum_i x_i y_i.$$

Finally, we denote $\dim X$ the dimension of a finite-dimension vector space X.

3. FROM SKETCHES TO UNIFORM EMBEDDINGS

Our main technical result shows that, for a finite-dimensional normed space X, good sketches for $\mathrm{DTEP}(X, D)$ imply a good uniform embedding of X into a Hilbert space (Definition 1). Below is the formal statement.

Theorem 4. *Suppose a finite-dimensional normed space* X *admits a public-coin randomized communication protocol for* $\mathrm{DTEP}(X, D)$ *of size* s *for approximation* $D > 1$*. Then, there exists a map* $f\colon X \to H$ *to a Hilbert space such that for all* $x_1, x_2 \in X$*,*

$$\min\left\{1, \frac{\|x_1 - x_2\|_X}{s \cdot D}\right\} \le \|f(x_1) - f(x_2)\|_H \le K \cdot \|x_1 - x_2\|_X^{1/2},$$

where $K > 1$ *is an absolute constant.*

Theorem 4 implies a *qualitative* version of Theorem 1 using the results of Aharoni, Maurey, and Mityagin [1] and Nikishin [56] (see Theorem 5).

Theorem 5 ([1, 56]). *For every fixed* $0 < \varepsilon < 1$*, any finite-dimensional normed space* X *that is uniformly embeddable into a Hilbert space is linearly embeddable into* $\ell_{1-\varepsilon}$ *with a distortion that depends only on* ε *and the moduli of the assumed uniform embedding.*

To prove the full (quantitative) versions of Theorems 1 and 2, we "open the black boxes" of [1, 56] in the full version.

In the rest of this section, we prove Theorem 4 according to the outline in Section 1.5, putting the pieces together in Section 3.4.

3.1 Sketching implies the absence of Poincaré inequalities

Sketching is often viewed from the perspective of a two-party communication complexity. Alice receives input x, Bob receives y, and they need to communicate to solve the DTEP problem. In particular, a sketch of size s implies a communication protocol that transmits s bits: Alice just sends her sketch $\mathbf{sk}(x)$ to Bob, who computes the output of DTEP (based on that message and his sketch $\mathbf{sk}(y)$). We assume here a public-coins model, i.e., Alice and Bob have access to a common (public) random string that determines the sketch function $\mathbf{sk}$.

To characterize sketching protocols, we build on results of Andoni, Jayram and Pătraşcu [7, Sections 3 and 4]. This works in two steps: first, we show that a protocol for $\mathrm{DTEP}(X, D)$ implies a sketching algorithm for $\mathrm{DTEP}(\ell_\infty^k(X), kD)$, with a loss of factor k in approximation (Lemma 1, see the proof in the end of the section). As usual, $\ell_\infty^k(X)$ is a normed space derived from X, by taking the vector space X^k and letting the norm of a vector $(x_1, \ldots x_k) \in X^k$ be the maximum of the norms of its k components. The second step is to apply a result from [7] (Theorem 6), which asserts that sketching for $\ell_\infty^k(X)$ precludes certain Poincaré inequalities for the space X.

LEMMA 1. *Let X be a finite-dimensional normed space that for some $D \geq 1$ admits a communication protocol for* $\mathrm{DTEP}(X, D)$ *of size s. Then for every integer k, the space $\ell_\infty^k(X)$ admits sketching with approximation kD and sketch size $s' = O(s)$.*

PROOF. Fix a threshold $t > 0$, and recall that we defined the success probability of sketching to be 0.9. By our assumption, there is a sketching function $\mathbf{sk}$ for X that achieves approximation D and sketch size s for threshold kt. Now define a "sketching" function $\mathbf{sk}'$ for $\ell_\infty^k(X)$ by choosing random signs $\varepsilon_1, \ldots, \varepsilon_k \in \{\pm 1\}$, letting $\mathbf{sk}' : x \mapsto \mathbf{sk}(\sum_{i=1}^k \varepsilon_i x_i)$, and using the same decision procedure used by $\mathbf{sk}$ (for X).

Now to examine the performance of $\mathbf{sk}'$, consider $x, y \in \ell_\infty^k(X)$. If their distance is at most t, then we always have that $\|\sum_{i=1}^k \varepsilon_i x_i - \sum_{i=1}^k \varepsilon_i y_i\| \leq \sum_{i=1}^k \|x_i - y_i\| \leq kt$ (i.e., for every realization of the random signs). Thus with probability at least 0.9 the sketch will declare that x, y are "close".

If the distance between x, y is greater than $kD \cdot t$, then for some coordinate, say $i = 1$, we have $\|x_1 - y_1\| > kD \cdot t$. Letting $z = \sum_{i \geq 2} \varepsilon_i (x_i - y_i)$, we can write $\|\sum_{i=1}^k \varepsilon_i x_i - \sum_{i=1}^k \varepsilon_i y_i\| = \|\varepsilon_1 (x_1 - y_1) + z\| = \|(x_1 - y_1) + \varepsilon_1 z\|$. The last term must be at least $\|x_1 - y_1\|$ under at least one of the two possible realizations of ε_1, because by the triangle inequality $2\|x_1 - y_1\| \leq \|(x_1 - y_1) + z\| + \|(x_1 - y_1) - z\|$. We see that with probability $1/2$ we have $\|\sum_{i=1}^k \varepsilon_i x_i - \sum_{i=1}^k \varepsilon_i y_i\| \geq \|x_1 - y_1\| > D \cdot kt$, and thus with probability at least $1/2 \cdot 0.9 = 0.45$ the sketch will declare that x, y are "far". This last guarantee is not sufficient for $\mathbf{sk}'$ to be called a sketch, but it can easily be amplified.

The final sketch $\mathbf{sk}''$ for $\ell_\infty^k(X)$ is obtained by $O(1)$ independent repetitions of $\mathbf{sk}'$, and returning "far" if at least 0.3-fraction of the repetitions come up with this decision. These repetitions amplify the success probability to 0.9, while increasing the sketch size to $O(s)$. $\square$

We now state the theorem of [7] that we use (in the contrapositive).

THEOREM 6 ([7]). *Let X be a metric space, and fix $r > 0$, $D \geq 1$. Suppose there are $\alpha > 0$, $\beta \geq 0$, and two symmetric probability measures μ_1, μ_2 on $X \times X$ such that*

- *The support of μ_1 is finite and is only on pairs with distance at most r;*
- *The support of μ_2 is finite and is only on pairs with distance greater than Dr; and*
- *For every $f : X \to B_{\ell_2}$ (where B_{ℓ_2} is the unit ball of ℓ_2),*

$$\mathop{\mathbb{E}}_{(x,y)\sim\mu_1} \|f(x) - f(y)\|^2 \geq \alpha \cdot \mathop{\mathbb{E}}_{(x,y)\sim\mu_2} \|f(x) - f(y)\|^2 - \beta.$$

Then for every integer k, the communication complexity of $\mathrm{DTEP}(\ell_\infty^k(X), D)$ *with probability of error $\delta_0 > 0$ is at least $\Omega(k) \cdot \big(\alpha(1 - 2\sqrt{\delta_0}) - \beta\big)$.*

We remark that [7] does not explicitly discuss protocols with public randomness, but rather private-coin protocols. While one can often use Newman's theorem [55] to extend such lower bounds to public coin protocols, we cannot afford to apply it here. Nonetheless, communication bounds that are based on information complexity (as in [7] or [11]) extend "black box" to public-coin protocols, see e.g. the argument in [15]. For completeness, we describe the entire reduction for our setting in the full version of the paper.

3.2 The absence of Poincaré inequalities implies threshold maps

We now prove that non-existence of Poincaré inequalities implies the existence a "threshold map", as formalized in Lemma 2 below. First we define the notion of threshold maps.

DEFINITION 4. *A map $f : X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is called an $(s_1, s_2, \tau_1, \tau_2, \tau_3)$-*threshold map *for $0 < s_1 < s_2$, $0 < \tau_1 < \tau_2 < \tau_3$, if for all $x_1, x_2 \in X$:*

- *if $d_X(x_1, x_2) \leq s_1$, then $d_Y(f(x_1), f(x_2)) \leq \tau_1$;*
- *if $d_X(x_1, x_2) \geq s_2$, then $d_Y(f(x_1), f(x_2)) \geq \tau_2$; and*
- *$d_Y(f(x_1), f(x_2)) \leq \tau_3$.*

Again, it is more convenient to prove the contrapositive statement:

LEMMA 2. *Suppose X is a metric space that does not allow an $(s_1, s_2, \tau_1, \tau_2, +\infty)$-threshold map to a Hilbert space. Then, for every $\delta > 0$ there exist two symmetric probability measures μ_1, μ_2 on $X \times X$ such that*

- *The support of μ_1 is finite and is only on pairs with distance at most s_1;*
- *The support of μ_2 is finite and is only on pairs with distance at least s_2; and*
- *For every $f : X \to B_{\ell_2}$,*

$$\mathop{\mathbb{E}}_{(x,y)\sim\mu_1} \|f(x) - f(y)\|^2 \geq \left(\frac{\tau_1}{\tau_2}\right)^2 \cdot \mathop{\mathbb{E}}_{(x,y)\sim\mu_2} \|f(x) - f(y)\|^2 - \delta.$$

During the course of the proof, we denote $\binom{X}{2}$ the set of all *unordered* pairs $\{x, y\}$ with $x, y \in X$, $x \neq y$. We prove Lemma 2 via the following three claims. The first one uses standard arguments about embeddability of finite subsets (see, e.g., Proposition 8.12 in [14]). For the proof see the full version of the paper.

CLAIM 1. *For every metric space X and every $0 < s_1 < s_2$, $0 < \tau_1 < \tau_2 < \tau_3$ there exists an $(s_1, s_2, \tau_1, \tau_2, \tau_3)$-threshold map of X to a Hilbert space iff the same is true for every finite subset of X.*

CLAIM 2. *Suppose that (X, d_X) is a* finite *metric space and $0 < s_1 < s_2$, $0 < \tau_1 < \tau_2 < \tau_3$. Assume that there is* no *$(s_1, s_2, \tau_1, \tau_2, \tau_3)$-threshold map of X to ℓ_2. Then, there exist two symmetric probability measures μ_1, μ_2 on $X \times X$ such that*

- *μ_1 is supported only on pairs with distance at most s_1, while μ_2 is supported only on pairs with distance at least s_2; and*

- *for every* $f\colon X \to \ell_2$,

$$\mathop{\mathbb{E}}_{(x,y)\sim\mu_1} \|f(x)-f(y)\|^2 \geq \left(\frac{\tau_1}{\tau_2}\right)^2 \cdot \mathop{\mathbb{E}}_{(x,y)\sim\mu_2} \|f(x)-f(y)\|^2 - \left(\frac{2\tau_1}{\tau_3}\right)^2 \cdot \sup_{x\in X} \|f(x)\|^2.$$

Proof. This Claim can be proved using convex duality. For the details see the full version. □

We are now ready to prove Lemma 2.

Proof Proof of Lemma 2. Let $\tau_3 > \tau_2$ be sufficiently large so that $(2\tau_1/\tau_3)^2 < \delta$. Then X has no $(s_1, s_2, \tau_1, \tau_2, \tau_3)$-threshold map to a Hilbert space, and by Claim 1 there exists a finite subset $X' \subset X$ that has no $(s_1, s_2, \tau_1, \tau_2, \tau_3)$-threshold map to a Hilbert space (which without loss of generality can be chosen to be ℓ_2). Now using Claim 2 we obtain measures μ_1 and μ_2 as required. □

3.3 Threshold maps imply uniform embeddings

We now prove that threshold embeddings imply uniform embeddings, formalized as follows.

Theorem 7. *Suppose that X is a finite-dimensional normed space that admits a $(1, D, \tau_1, \tau_2, +\infty)$-threshold map to a Hilbert space for some $D > 1$ and for some $0 < \tau_1 < \tau_2$ with $\tau_2 > 8\tau_1$. Then there exists a map h of X into a Hilbert space such that for every $x_1, x_2 \in X$,*

$$(\tau_2^{1/2} - (8\tau_1)^{1/2}) \cdot \min\left\{1, \frac{\|x_1 - x_2\|}{2D+4}\right\} \leq \|h(x_1) - h(x_2)\| \leq (2\tau_1\|x_1 - x_2\|)^{1/2}. \quad (1)$$

In particular, h is a uniform embedding of X into a Hilbert space with moduli that depend only on τ_1, τ_2 and D.

Let us point out that in [38, 60], Johnson and Randrianarivony prove that for a Banach space coarse embeddability into a Hilbert space is equivalent to uniform embeddability. Our definition of a threshold map is weaker than that of a coarse embedding (for the latter see [38] say), but we show that we can adapt the proof of [38, 60] to our setting as well (at least whenever the gap between τ_1 and τ_2 is large enough). Since we only need one direction of the equivalence, we present a part of the argument from [38] with one (seemingly new) addition: Claim 5. The resulting proof is arguably simpler than the combination of [38] and [60], and yields a clean quantitative bound (1).

Intuition. Let us provide some very high-level intuition of the proof of Theorem 7. We start with a threshold map f from X to a Hilbert space. First, we show that f is Lipschitz on pairs of points that are sufficiently far. In particular, f, restricted on a sufficiently crude net N of X, is Lipschitz. This allows us to use a certain extension theorem to extend the restriction of f on N to a Lipschitz function on the whole X, while preserving the property that f does not contract too much distances that are sufficiently large. Then, we get a required uniform embedding by performing a certain symmetrization step.

The real proof is different in the number of details: in particular, instead of being Lipschitz the real property we will be trying to preserve is different.

Useful facts. To prove Theorem 7, we need the following three results.

Lemma 3 ([62]). *For a set S and a map f from S to a Hilbert space, there exists a map g from S to a Hilbert space such that $\|g(x_1) - g(x_2)\| = \|f(x_1) - f(x_2)\|^{1/2}$ for every $x_1, x_2 \in S$.*

Lemma 4 (essentially [1]). *Suppose that f is a map from an abelian group G to a Hilbert space such that for every $g \in G$ we have $\sup_{g_1 - g_2 = g} \|f(g_1) - f(g_2)\| < +\infty$. Then, there exists a map f' from G to a Hilbert space such that $\|f'(g_1) - f'(g_2)\|$ depends only on $g_1 - g_2$ and for every $g_1, g_2 \in G$ we have*

$$\inf_{g_1' - g_2' = g_1 - g_2} \|f(g_1') - f(g_2')\| \leq \|f'(g_1) - f'(g_2)\| \leq \sup_{g_1' - g_2' = g_1 - g_2} \|f(g_1') - f(g_2')\|.$$

Definition 5. *We say that a map $f\colon X \to Y$ between metric spaces is $1/2$-Hölder with constant C, if for every $x_1, x_2 \in X$ one has $d_Y(f(x_1), f(x_2)) \leq C \cdot d_X(x_1, x_2)^{1/2}$.*

Theorem 8 (Theorem 19.1 in [66]). *Let (X, d_X) be a metric space and let H be a Hilbert space. Suppose that $f\colon S \to H$, where $S \subset X$, is a $1/2$-Hölder map with a constant $C > 0$. Then there exists a map $g\colon X \to H$ that coincides with f on S and is $1/2$-Hölder with the constant C.*

We are now ready to prove Theorem 7.

Proof Proof of Theorem 7. We prove the theorem via the following sequence of claims. Suppose that X is a finite-dimensional normed space. Let f be a $(1, D, \tau_1, \tau_2, +\infty)$-threshold map to a Hilbert space.

The first claim is well-known and is a variant of Proposition 1.11 from [14].

Claim 3. *For every $x_1, x_2 \in X$ we have $\|f(x_1) - f(x_2)\| \leq \max\{1, 2 \cdot \|x_1 - x_2\|\} \cdot \tau_1$.*

Proof. If $\|x_1 - x_2\| \leq 1$, then $\|f(x_1) - f(x_2)\| \leq \tau_1$, and we are done. Otherwise, let us take $y_0, y_1, \ldots, y_l \in X$ such that $y_0 = x_1$, $y_l = x_2$, $\|y_i - y_{i+1}\| \leq 1$ for every i, and $l = \lceil \|x_1 - x_2\| \rceil$. We have

$$\|f(x_1) - f(x_2)\| \leq \sum_{i=0}^{l-1} \|f(y_i) - f(y_{i+1})\| \leq l\tau_1 = \lceil \|x_1 - x_2\| \rceil \cdot \tau_1 \leq 2\|x_1 - x_2\| \cdot \tau_1,$$

where the first step is by the triangle inequality, the second step follows from $\|y_i - y_{i+1}\| \leq 1$, and the last step follows from $\|x_1 - x_2\| \geq 1$. □

The proof of the next claim essentially appears in [38].

Claim 4. *There exists a map g from X to a Hilbert space such that for every $x_1, x_2 \in X$,*

- $\|g(x_1) - g(x_2)\| \leq (2\tau_1 \cdot \|x_1 - x_2\|)^{1/2}$;
- *if* $\|x_1 - x_2\| \geq D + 2$, *then* $\|g(x_1) - g(x_2)\| \geq \tau_2^{1/2} - (8\tau_1)^{1/2}$;

Proof. From Claim 3 and Lemma 3 we can get a map g' from X to a Hilbert space such that for every $x_1, x_2 \in X$

- $\|g'(x_1) - g'(x_2)\| \le \max\left\{1, (2\|x_1 - x_2\|)^{1/2}\right\} \cdot \tau_1^{1/2}$;
- if $\|x_1 - x_2\| \ge D$, then $\|g'(x_1) - g'(x_2)\| \ge \tau_2^{1/2}$.

Let $N \subset X$ be a 1-net of X such that all the pairwise distances between points in N are more than 1. The map g' is 1/2-Hölder on N with a constant $(2\tau_1)^{1/2}$, so we can apply Theorem 8 and get a map g that coincides with g' on N and is 1/2-Hölder on the whole X with a constant $(2\tau_1)^{1/2}$. That is, for every $x_1, x_2 \in X$ we have

- $\|g(x_1) - g(x_2)\| \le (2\tau_1 \cdot \|x_1 - x_2\|)^{1/2}$;
- if $x_1 \in N$, $x_2 \in N$ and $\|x_1 - x_2\| \ge D$, then $\|g(x_1) - g(x_2)\| \ge \tau_2^{1/2}$.

To conclude that g is as required, let us lower bound $\|g(x_1) - g(x_2)\|$ whenever $\|x_1 - x_2\| \ge D + 2$. Suppose that $x_1, x_2 \in X$ are such that $\|x_1 - x_2\| \ge D + 2$. Let u_1 be the closest to x_1 point from N and, similarly, let $u_2 \in N$ be the closest net point to x_2. Observe that

$$\begin{aligned}\|u_1 - u_2\| &\ge \|x_1 - x_2\| - \|x_1 - u_1\| - \|x_2 - u_2\| \\ &\ge (D + 2) - 1 - 1 = D.\end{aligned}$$

We have

$$\begin{aligned}&\|g(x_1) - g(x_2)\| \\ &\ge \|g(u_1) - g(u_2)\| - \|g(u_1) - g(x_1)\| - \|g(u_2) - g(x_2)\| \\ &\ge \tau_2^{1/2} - 2(2\tau_1)^{1/2},\end{aligned}$$

as required, where the second step follows from the inequality $\|g(u_1) - g(u_2)\| \ge \tau_2^{1/2}$, which is true, since $u_1, u_2 \in N$, and that g is 1/2-Hölder with a constant $(2\tau_1)^{1/2}$. □

The following claim completes the proof of Theorem 7.

Claim 5. *There exists a map h from X to a Hilbert space such that for every $x_1, x_2 \in X$:*

- $\|h(x_1) - h(x_2)\| \le (2\tau_1 \cdot \|x_1 - x_2\|)^{1/2}$;
- *one has*

$$\begin{aligned}&\|h(x_1) - h(x_2)\| \\ &\ge (\tau_2^{1/2} - (8\tau_1)^{1/2}) \cdot \min\{1, \|x_1 - x_2\|/(2D + 4)\}.\end{aligned}$$

Proof. We take the map g from Claim 4 and apply Lemma 4 to it. Let us call the resulting map h. The first desired condition for h follows from a similar condition for g and Lemma 4. Let us prove the second one.

If $x_1 = x_2$, then there is nothing to prove. If $\|x_1 - x_2\| \ge D + 2$, then by Claim 4 and Lemma 4, $\|h(x_1) - h(x_2)\| \ge \tau_2^{1/2} - (8\tau_1)^{1/2}$, and we are done. Otherwise, let us consider points $y_0, y_1, \ldots, y_l \in X$ such that $y_0 = 0$, $y_i - y_{i-1} = x_1 - x_2$ for every i, and $l = \left\lceil \frac{D+2}{\|x_1 - x_2\|} \right\rceil$. Since $\|y_l - y_0\| = \|l(x_1 - x_2)\| = l\|x_1 - x_2\| \ge D + 2$, we have

$$\begin{aligned}\tau_2^{1/2} - (8\tau_1)^{1/2} &\le \|h(y_l) - h(y_0)\| \le \sum_{i=1}^{l} \|h(y_i) - h(y_{i-1})\| \\ &= l \cdot \|h(x_1) - h(x_2)\| \le \frac{2D + 4}{\|x_1 - x_2\|} \cdot \|h(x_1) - h(x_2)\|,\end{aligned}$$

where the equality follows from the conclusion of Lemma 4. □

Finally, observe that Theorem 7 is merely a reformulation of Claim 5.

3.4 Putting it all together

We now show that Theorem 4 follows by applying Lemma 1, Theorem 6, Lemma 2, and Theorem 7, in this order, with an appropriate choice of parameters.

Proof Proof of Theorem 4. Suppose $\mathrm{DTEP}(X, D)$ admits a protocol of size s. By setting $k = Cs$ in Lemma 1 (C is a large absolute constant, to be chosen later), we conclude that $\mathrm{DTEP}(\ell_\infty^{Cs}(X), CsD)$ admits a protocol of size $s' = O(s)$.

Now choosing C large enough and applying Theorem 6 (in contrapositive), we conclude that X has no Poincaré inequalities for distance scales 1 and CsD, $\alpha = 0.01$ and $\beta = 0.001$.

Applying Lemma 2 (in contrapositive) we conclude that X allows a $(1, CsD, 1, 10, +\infty)$-threshold map to a Hilbert space.

Using Theorem 7 it follows that there is a map h from X to a Hilbert space, such that for all $x_1, x_2 \in X$,

$$\min\left\{1, \frac{\|x_1 - x_2\|}{s \cdot D}\right\} \le \|h(x_1) - h(x_2)\| \le K \cdot \|x_1 - x_2\|^{1/2},$$

where $K > 1$ is an absolute constant, and this proves the theorem. □

4. EMBEDDING INTO ℓ_1 VIA SUM-PRODUCTS

Finally, we prove Theorem 3: good sketches for norms closed under the sum-product imply embeddings into ℓ_1 with constant distortion. First we invoke Theorem 4 and get a sequence of good uniform embeddings into a Hilbert space, whose moduli depend only on the sketch size and the approximation. Then, we use the main result of this section: Lemma 5. Before stating the lemma, let us remind a few notions. For a metric space X, recall that the metric space $\ell_1^k(X) = \bigoplus_{\ell_1}^k X_n$ is the direct sum of k copies of X, with the associated distance defined as a sum-product (ℓ_1-product) over the k copies. We define $\ell_1(X)$ similarly. We also denote $X \oplus_{\ell_1} Y$ the sum-product of X and Y.

Lemma 5. *Let $(X_n)_{n=1}^{\infty}$ be a sequence of finite-dimensional normed spaces. Suppose that for every $i_1, i_2 \ge 1$ there exists $m = m(i_1, i_2) \ge 1$ such that $X_{i_1} \oplus_{\ell_1} X_{i_2}$ is isometrically embeddable into X_m. If every X_n admits a uniform embedding into a Hilbert space with moduli independent of n, then every X_n is linearly embeddable into ℓ_1 with distortion independent of n.*

Note that Theorem 3 just follows from combining Lemma 5 with Theorem 4.

Before proving Lemma 5, we state the following two useful theorems. The first one (Theorem 9) follows from the fact that uniform embeddability into a Hilbert space is determined by embeddability of finite subsets [14]. The second one (Theorem 10) follows by composing results of Aharoni, Maurey, and Mityagin [1] and Kalton [39].

Theorem 9 (Proposition 8.12 from [14]). *Let $A_1 \subset A_2 \subset \ldots$ be metric spaces and let $A = \bigcup_i A_i$. If every A_n is uniformly embeddable into a Hilbert space with moduli independent of n, then the whole A is uniformly embeddable into a Hilbert space.*

Theorem 10 ([1, 39]). *A Banach space X is linearly embeddable into L_1 iff $\ell_1(X)$ is uniformly embeddable into a Hilbert space.*

We are now ready to proceed with the proof of Lemma 5.

Proof Proof of Lemma 5. Let $X = X_1 \oplus_{\ell_1} X_2 \oplus_{\ell_1} \ldots$. More formally,

$$X = \Big\{(x_1, x_2, \ldots) : x_i \in X_i, \sum_i \|x_i\| < \infty\Big\},$$

where the norm is set as follows:

$$\big\|(x_1, x_2, \ldots)\big\| = \sum_i \|x_i\|.$$

We claim that the space $\ell_1(X)$ embeds uniformly into a Hilbert space. To see this, consider $U_p = \ell_1^p(X_1 \oplus_{\ell_1} X_2 \oplus_{\ell_1} \ldots \oplus_{\ell_1} X_p)$, which can be naturally seen as a subspace of $\ell_1(X)$. Then, $U_1 \subset U_2 \subset \ldots \subset U_p \subset \ldots \subset \ell_1(X)$ and $\bigcup_p U_p$ is dense in $\ell_1(X)$. By the assumption of the lemma, U_p is isometrically embeddable into X_m for some m, thus, U_p is uniformly embeddable into a Hilbert space with moduli independent of p. Now, by Theorem 9, $\bigcup_p U_p$ is uniformly embeddable into a Hilbert space. Since $\bigcup_p U_p$ is dense in $\ell_1(X)$, the same holds also for the whole $\ell_1(X)$, as claimed.

Finally, since $\ell_1(X)$ embeds uniformly into a Hilbert space, we can apply Theorem 10 and conclude that X is linearly embeddable into L_1. The lemma follows since X contains every X_i as a subspace. $\square$

5. ACKNOWLEDGMENTS

We are grateful to Assaf Naor for pointing us to [1, 39], as well as for numerous very enlightening discussions throughout this project. We also thank Gideon Schechtman for useful discussions and explaining some of the literature. We thank Piotr Indyk for fruitful discussions and for encouraging us to work on this project.

6. REFERENCES

[1] I. Aharoni, B. Maurey, and B. S. Mityagin. Uniform embeddings of metric spaces and of Banach spaces into Hilbert spaces. *Israel J. Math.*, 52(3):251–265, 1985.

[2] K. Ahn, S. Guha, and A. McGregor. Analyzing graph structure via linear measurements. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 459–467, 2012.

[3] A. Andoni, K. Do Ba, P. Indyk, and D. Woodruff. Efficient sketches for Earth-Mover Distance, with applications. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, 2009.

[4] A. Andoni and P. Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Communications of the ACM*, 51(1):117–122, 2008.

[5] A. Andoni, P. Indyk, and R. Krauthgamer. Earth mover distance over high-dimensional spaces. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 343–352, 2008.

[6] A. Andoni, P. Indyk, and R. Krauthgamer. Overcoming the ℓ_1 non-embeddability barrier: Algorithms for product metrics. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 865–874, 2009.

[7] A. Andoni, T. Jayram, and M. Pătraşcu. Lower bounds for edit distance and product metrics via Poincaré-type inequalities. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010.

[8] A. Andoni and R. Krauthgamer. The computational hardness of estimating edit distance. *SIAM Journal on Computing*, 39(6):2398–2429, 2010. Previously appeared in FOCS'07.

[9] 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 (electronic), 1998.

[10] Z. Bar-Yossef, T. S. Jayram, R. Krauthgamer, and R. Kumar. Approximating edit distance efficiently. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 550–559, 2004.

[11] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *CCC '02: Proceedings of the 17th IEEE Annual Conference on Computational Complexity*, page 93, Washington, DC, USA, 2002. IEEE Computer Society.

[12] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. Previously in FOCS'02.

[13] T. Batu, F. Ergün, and C. Sahinalp. Oblivious string embeddings and edit distance approximations. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 792–801, 2006.

[14] Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

[15] M. Braverman and A. Garg. Public vs private coin in bounded-round information. In *Proceedings of International Colloquium on Automata, Languages and Programming (ICALP)*, pages 502–513, 2014.

[16] V. Braverman and R. Ostrovsky. Zero-one frequency laws. In *Proceedings of the Symposium on Theory of Computing (STOC)*, 2010.

[17] A. Broder. On the resemblance and containment of documents. *Proceedings of Compression and Complexity of Sequences*, pages 21–29, 1997.

[18] A. Broder, S. Glassman, M. Manasse, and G. Zweig. Syntactic clustering of the web. *Proceedings of the Sixth International World Wide Web Conference*, pages 391–404, 1997.

[19] J. Brody and A. Chakrabarti. A multi-round communication lower bound for gap Hamming and some consequences. In *Proceedings of the IEEE Conference on Computational Complexity*, 2009.

[20] J. Brody, A. Chakrabarti, O. Regev, T. Vidick, and R. De Wolf. Better gap-hamming lower bounds via better round elimination. In *Proceedings of the 13th International Conference on Approximation, and 14 the International Conference on Randomization, and Combinatorial Optimization: Algorithms and Techniques*, APPROX/RANDOM'10, pages 476–489. Springer-Verlag, 2010.

[21] A. Chakrabarti and O. Regev. An optimal lower bound on the communication complexity of

gap-hamming-distance. *SIAM J. Comput.*, 41(5):1299–1317, 2012. Previously in STOC'11.

[22] M. Charikar. Similarity estimation techniques from rounding. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 380–388, 2002.

[23] M. Charikar and R. Krauthgamer. Embedding the Ulam metric into ℓ_1. *Theory of Computing*, 2(11):207–224, 2006.

[24] M. Charikar, Q. Lv, and K. Li. Image similarity search with compact data structures. In *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 2004.

[25] J. Cheeger and B. Kleiner. Differentiating maps into L^1, and the geometry of BV functions. *Ann. of Math. (2)*, 171(2):1347–1385, 2010.

[26] J. Cheeger, B. Kleiner, and A. Naor. Compression bounds for Lipschitz maps from the Heisenberg group to L_1. *Acta mathematica*, 207(2):291–373, 2011.

[27] G. Cormode and S. Muthukrishnan. The string edit distance matching problem with moves. *ACM Transactions on Algorithms*, 3(1), 2007. Previously in SODA'02.

[28] G. Cormode, S. Muthukrishnan, and S. C. Sahinalp. Permutation editing and matching via embeddings. In *28th International Colloquium on Automata, Languages and Programming*, volume 2076 of *Lecture Notes in Computer Science*, pages 481–492. Springer, 2001.

[29] G. Cormode, M. Paterson, S. C. Sahinalp, and U. Vishkin. Communication complexity of document exchange. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 197–206, 2000.

[30] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, 2003.

[31] S. Guha, P. Indyk, and A. McGregor. Sketching information divergences. *Journal of Machine Learning*, 72(1–2):5–19, 2008. Previously appeared in COLT'07.

[32] P. Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. *J. ACM*, 53(3):307–323, 2006. Previously appeared in FOCS'00.

[33] P. Indyk and R. Motwani. Approximate nearest neighbor: towards removing the curse of dimensionality. *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 604–613, 1998.

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

[35] T. Jayram, R. Kumar, and D. Sivakumar. The one-way communication complexity of hamming distance. *Theory of Computing*, 4(1):129–135, 2008.

[36] T. Jayram and D. Woodruff. The data stream space complexity of cascaded norms. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, 2009.

[37] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Connecticut, 1982)*, volume 26 of *Contemporary Mathematics*, pages 189–206. 1984.

[38] W. B. Johnson and N. L. Randrianarivony. l_p $(p > 2)$ does not coarsely embed into a Hilbert space. *Proc. Amer. Math. Soc.*, 134(4):1045–1050 (electronic), 2006.

[39] N. J. Kalton. Banach spaces embedding into L_0. *Israel J. Math.*, 52(4):305–319, 1985.

[40] B. Kapron, V. King, and B. Mountjoy. Dynamic graph connectivity in polylogarithmic worst case time. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1131–1142, 2013.

[41] S. Khot and A. Naor. Nonembeddability theorems via Fourier analysis. *Math. Ann.*, 334(4):821–852, 2006. Preliminary version appeared in FOCS'05.

[42] S. Khot and R. Saket. Sdp integrality gaps with local ℓ_1-embeddability. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 565–574. IEEE Computer Society, 2009.

[43] S. A. Khot and N. K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 53–62, 2005.

[44] R. Krauthgamer and Y. Rabani. Improved lower bounds for embeddings into L_1. *SIAM J. Comput.*, 38(6):2487–2498, 2009.

[45] 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.

[46] J. R. Lee and A. Naor. Lp metrics on the heisenberg group and the goemans-linial conjecture. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 99–108. IEEE, 2006.

[47] Y. Li, H. L. Nguyên, and D. P. Woodruff. On sketching matrix norms and the top singular vector. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.

[48] Y. Li, H. L. Nguyên, and D. P. Woodruff. Turnstile streaming algorithms might as well be linear sketches. In *Proceedings of the Symposium on Theory of Computing (STOC)*, 2014.

[49] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 577–591, 1994.

[50] J. Matoušek and A. Naor. Open problems on embeddings of finite metric spaces. August 2011. Available online at `http://kam.mff.cuni.cz/~matousek/metrop.ps.gz`. Last accessed in November, 2014.

[51] A. McGregor. Open problems in data streams and related topics. *IITK Workshop on Algorithms For Data Streams*, 2006. Available at `http://sublinear.info`.

[52] R. Motwani, A. Naor, and R. Panigrahy. Lower bounds on locality sensitive hashing. In *Proceedings of the ACM Symposium on Computational Geometry (SoCG)*, pages 253–262, 2006.

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

[54] A. Naor and G. Schechtman. Planar earthmover is not in L_1. *SIAM Journal on Computing*, 37(3):804–826, 2007. An extended abstract appeared in FOCS'06.

[55] I. Newman. Private vs. common random bits in communication complexity. *Inf. Process. Lett.*, 39(2):67–71, July 1991.

[56] E. Nikišin. A resonance theorem and series in eigenfunctions of the laplacian. *Mathematics of the USSR-Izvestiya*, 6(4):788–806, 1972.

[57] R. O'Donnell, Y. Wu, and Y. Zhou. Optimal lower bounds for locality-sensitive hashing (except when q is tiny). *ACM Trans. Comput. Theory*, 6(1):5:1–5:13, Mar. 2014.

[58] R. Ostrovsky and Y. Rabani. Low distortion embedding for edit distance. *J. ACM*, 54(5), 2007. Preliminary version appeared in STOC'05.

[59] G. Pisier. Some results on Banach spaces without local unconditional structure. *Compositio Math.*, 37(1):3–19, 1978.

[60] N. L. Randrianarivony. Characterization of quasi-Banach spaces which coarsely embed into a Hilbert space. *Proc. Amer. Math. Soc.*, 134(5):1315–1317 (electronic), 2006.

[61] M. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 360–369, 2002.

[62] I. J. Schoenberg. On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Ann. of Math. (2)*, 38(4):787–793, 1937.

[63] A. A. Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(8):197–208, 2012.

[64] M. Talagrand. Embedding subspaces of L_1 into l_1^N. *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.

[65] T. Vidick. A concentration inequality for the overlap of a vector on a large set, with application to the communication complexity of the gap-Hamming-Distance problem. *Chicago Journal of Theoretical Computer Science*, 2012(1), July 2012.

[66] J. H. Wells and L. R. Williams. *Embeddings and extensions in analysis*. Springer-Verlag, New York-Heidelberg, 1975.

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

[68] D. P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10:1–157, 2014.

[69] A. Zvavitch. More on embedding subspaces of L_p into l_p^N, $0 < p < 1$. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 269–280. Springer, Berlin, 2000.