# Constant-Distortion Embeddings of Hausdorff Metrics into Constant-Dimensional $\ell_p$ Spaces

**Artūrs Bačkurs*[1] and Anastasios Sidiropoulos†[2]**

1 **MIT, Cambridge MA, USA**
backurs@mit.edu
2 **The Ohio State University, Columbus OH, USA**
sidiropoulos.1@osu.edu

## Abstract

We show that the Hausdorff metric over constant-size pointsets in constant-dimensional Euclidean space admits an embedding into constant-dimensional $\ell_\infty$ space with constant distortion. More specifically for any $s, d \geq 1$, we obtain an embedding of the Hausdorff metric over pointsets of size $s$ in $d$-dimensional Euclidean space, into $\ell_\infty^{s^{O(s+d)}}$ with distortion $s^{O(s+d)}$. We remark that any metric space $M$ admits an isometric embedding into $\ell_\infty$ with dimension proportional to the size of $M$. In contrast, we obtain an embedding of a space of infinite size into constant-dimensional $\ell_\infty$.

We further improve the distortion and dimension trade-offs by considering probabilistic embeddings of the snowflake version of the Hausdorff metric. For the case of pointsets of size $s$ in the real line of bounded resolution, we obtain a probabilistic embedding into $\ell_1^{O(s \log s)}$ with distortion $O(s)$.

**1998 ACM Subject Classification** F.2.2 Nonnumerical Algorithms and Problems

**Keywords and phrases** metric embeddings, Hausdorff metric, distortion, dimension

**Digital Object Identifier** 10.4230/LIPIcs.APPROX-RANDOM.2016.1

## 1 Introduction

Low-distortion embeddings between metric spaces have given rise to a plethora of tools in computer science and mathematics [10, 15, 2, 5, 17]. The most well-studied case is embedding into $\ell_p^d$, that is $\mathbb{R}^d$ endowed with the $\ell_p$ distance. In this case the most important parameters are the distortion of the embedding and the dimension of the target space; the former quantifies the extent to which the geometry of the input space is preserved, while the latter affects the complexity of various algorithmic methods performed on the target space.

In most embeddings of finite metric spaces both of these parameters depend on the size of the input space. Prototypical such examples are Bourgain's Theorem [4] which asserts that any $n$-point metric admits an embedding into $\ell_2$ with distortion $O(\log n)$, and the seminal result of Johnson and Lindenstrauss [14] asserting that any $n$-point subset of $\ell_2$ admits an embedding into $\ell_2^{O(\varepsilon^{-2} \log n)}$ with distortion $1 + \varepsilon$.

However, in several applications one seeks an embedding of some input space of infinite size. One such application is in algorithms and data structures (e.g. nearest neighbor data structures) with approximation guarantee independent of the input size. Another application

* Artūrs Backurs was supported by NSF and Simons Foundation.
† Anastasios Sidiropoulos is supported by NSF under grant CCF-1423230 and award CAREER-1453472.

is when designing an oblivious or streaming algorithm that requires an embedding of the input space that can be computed independently at each point without having access to the rest of the input (e.g. [12, 13]).

A classical example of an embedding of an infinite metric is Dvoretzky's Theorem [6] which asserts that for any $k \geq 1$, there exists $d \geq 1$ such that $\ell_2^k$ admits an embedding into any $d$-dimensional normed space with distortion $1+\varepsilon$.

Interestingly, the case of input spaces that are not normed, is much less understood. One important such space is given by the Hausdorff metric which is used to measure the dissimilarity between two pointsets. Given two finite pointsets $A = \{a_1, \ldots, a_s\}$ and $B = \{b_1, \ldots, b_s\}$, the Hausdorff distance is defined as

$$H_s(A, B) = \max(h(A, B), h(B, A)),$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} m(a, b)$$

and $m(\cdot, \cdot)$ is underlying metric on the points of $A$ and $B$. We use the notation $H_{s,d}$ to denote the Hausdorff distance with underlying metric $\ell_2^d$. We will omit subscripts if the cardinality of the pointsets or the underlying metric is clear from the context, or the statement is valid independent from the cardinality or the underlying space.

We study embeddings of the Hausdorff metric over finite subsets of Euclidean space. This is an infinite space since there are infinitely many possible subsets even in the real line. Therefore known results for embedding finite metrics into $\ell_p$ space are not directly applicable in this case.

## 1.1 Our results and techniques

### 1.1.1 Embedding for Hausdorff metric over pointsets in $\mathbb{R}^1$

We show that there exists an embedding of $H_{s,1}$ into $\ell_\infty^{s^{O(s)}}$ with distortion $s^{O(s)}$. Let $\mathcal{M}$ be a collection of metric spaces on the same pointset $X$. We say that a metric is a $\ell_\infty$*-metric over* $\mathcal{M}$ if for any pair of points in $X$ the distance is given by the maximum over all distances in $\mathcal{M}$. Our result is obtained via iteratively embedding $H_s$ into an $\ell_\infty$-metric over $H_{s-1}$ metrics. The key property in this mapping is that it preserves all distances in the infinite space $H_s$. Repeating this process we inductively obtain an embedding of $H_s$ into an $\ell_\infty$-metric over $H_1$ metrics. Since $H_1 = \mathbb{R}^1$, the resulting embedding is into $\ell_\infty$.

### 1.1.2 Embedding for Hausdorff metric over pointsets in $\mathbb{R}^d$

We extend the above approach to $H_{s,d}$. This is done by embedding $H_{s,d}$ into an $\ell_\infty$-metric over $H_{s,d-1}$ metrics. By repeating this embedding we obtain an embedding of $H_{s,d}$ into an $\ell_\infty$ metric over $H_{s,1}$ metrics. Combining with the above embedding we obtain the desired embedding of $H_{s,d}$ into $\ell_\infty$.

### 1.1.3 Probabilistic embeddings

The above embeddings obtain distortion and dimension that depend only on $s$. We show how to exponentially improve the dependence of both parameters on $s$ by considering probabilistic embeddings of the snowflake version of $H_s$ into $\ell_1$.

**Table 1** Summary of our and previous results on embedding Hausdorff distance into $\ell_p$ spaces. Column "dimension" specifies the dimension of the target $\ell_p$ space. We consider Hausdorff distance over pointsets of size $s$ coming from the underlying space. Here, $\varepsilon > 0$ is a small constant.

| | Underlying space | To | Dimension | Distortion | Comments |
|---|---|---|---|---|---|
| Theorem 10 | $\ell_2^d$ | $\ell_\infty$ | $s^{O(s+d)}$ | $s^{O(s+d)}$ | |
| Theorem 23 | $\ell_p^1$ | $\ell_1$ | $O(s \log s)$ | $O(s/\alpha)$ | Snowflaked embedding with parameter $\alpha$ |
| [11] | $\{0, \ldots, \Delta\}_\infty^d$ | $\ell_\infty$ | $s^2 \cdot e^{O(d)} \cdot \log^2 \Delta$ | $1+\varepsilon$ | Threshold embedding |
| [16] | $\ell_2^d$ | $\ell_\infty$ | $e^{O(ds)}$ | $1+\varepsilon$ | Snowflaking |

### 1.1.4 Embedding into high-dimensional $\ell_1$ space

To improve the distortion of the embedding, we relax the requirements of the embedding. First, we embed a snowflake version of the Hausdorff distance into $\ell_1$. This means that we embed the distance $H_s^{1-\alpha}$ for some $\alpha > 0$ into $\ell_1$. Second, we allow that the expansion property holds in expectation (see Section 2 for a formal definition). This allows us to achieve distortion $O(s/\alpha)$, which is an exponential improvement over the deterministic embedding. The embedding uses ideas that were previously used to construct embeddings for earth-mover distance [9, 1]. In particular, we recursively subdivide the underlying metric space into cells and designate a coordinate in the target space for every cell. Instead of counting the number of points that fall into each cell (as was done in the case of embeddings of earth-mover distance), we instead detect whether at least one point falls into the cell. To achieve distortion that does not depend on the size of the underlying metric, we use ideas developed in [1], embedding a snowflake version of the Hausdorff distance.

### 1.1.5 Embedding into low-dimensional $\ell_1$ space

To improve the dimension of the target $\ell_1$ space, we further relax the requirements of the embedding. We allow that the embedding contracts with probability bounded by some small constant. This allows us to reduce the dimension exponentially. The dimension of the target $\ell_1$ space becomes $O(s \log s)$. This improvement is obtained by observing that a vector resulting from the embedding into high dimensional space, is essentially sparse; that is, the main contribution to the $\ell_1$ norm comes from few non-zero entries. This suggests that we can use dimensionality reduction techniques for $\ell_1$ space for sparse vectors. To this end we use a construction from [3]. We remark that a similar dimensionality reduction idea was used in [1]. We note that we can decrease the probability of contraction to an arbitrary $\delta > 0$ by combining $O(\log(1/\delta))$ independent copies of the embedding.

In Table 1 we summarize our results and highlight the previous work on embedding Hausdorff distance into simpler spaces.

## 1.2 Related work

Farach-Colton and Indyk [7] have studied the problem of embedding the Hausdorff metric over finite pointsets into $\ell_\infty$. However, they only obtain embeddings that approximately preserve distances that are within a fixed range $[r, R]$, for some $0 < r < R$. This weaker guarantee is sufficient for designing an approximate nearest neighbor data structure. However, in order to obtain an embedding that preserves all distances up to some small distortion, one has to concatenate $O(\log \Delta)$ such embeddings, where $\Delta$ is the spread of the metric. Since $\Delta$ is in general unbounded, this leads to a host space of arbitrarily large dimension.

Indyk [11] studied *threshold embeddings* for Hausdorff distance. In this setting the goal is to obtain an embedding so that the following two conditions hold: First, the embedding is a contraction. Second, if the distance between two points in the original space is at least $r$, then their distance in the target space is at least $r'$, for some $r \geq r' > 0$. The distortion of a threshold embedding is defined to be the ratio $r/r'$. The dimension of the target space in [11] depends on the size of the underlying metric, which can be unbounded.

Previous works [8, 16] studied embeddings of snowflake metrics. They showed that, if the *doubling dimension* of a metric is $t$, then it is possible to embed such a metric into $\ell_\infty^{e^{O(t)}}$ with distortion $1+\epsilon$, for any constant $\epsilon > 0$. We will not define doubling dimension here but we note that for the case of Hausdorff metric over $\ell_p^d$, it is bounded by $O(ds)$.

## 2 Preliminaries

▶ **Definition 1.** Consider the Hausdorff distance over pointsets in some underlying space. Let $f$ be a function that maps pointsets to vectors in some $\ell_\infty$-space. We say that $f$ is an *embedding* if there exist $L \geq l > 0$ such that

$$l \cdot \|f(A) - f(B)\|_\infty \leq H(A, B) \leq L \cdot \|f(A) - f(B)\|_\infty \tag{1}$$

for all pointsets $A$ and $B$. The quantity $L/l$ is the *distortion* of the embedding.

▶ **Definition 2.** Let $\mathcal{D}$ be a probability distribution over functions that map pointsets of some space into some $\ell_\infty$-space. We say that a function $f$ chosen from $\mathcal{D}$ is a *probabilistic embedding*. Moreover, if there exist $L \geq l > 0$ such that for all sets $A, B$, we have

$$l \cdot E_f[\|f(A) - f(B)\|_\infty] \leq H(A, B)$$

and

$$\Pr_f [H(A, B) \leq L \cdot \|f(A) - f(B)\|_\infty] \geq 2/3,$$

then the *distortion* of $f$ is defined to be $L/l$. Note that the choice of $2/3$ is arbitrary; we can amplify it by sampling independent copies of the function $f$ and concatenating the resulting embeddings.

▶ **Definition 3.** A probabilistic embedding $f$ is called a *snowflaked embedding* with parameter $\alpha > 0$ if it satisfies the following properties: There exist $L \geq l > 0$ such that for all sets $A, B$, we have

$$l \cdot E_f[\|f(A) - f(B)\|_\infty] \leq H^{1-\alpha}(A, B)$$

and

$$\Pr_f \left[H^{1-\alpha}(A, B) \leq L \cdot \|f(A) - f(B)\|_\infty\right] \geq 2/3.$$

The *distortion* of $f$ is defined to be $L/l$.

## 2.1 Notation

Given two vectors $A \in \mathbb{R}^x$, $A = (a_1, \ldots a_x)^T$ and $B \in \mathbb{R}^y$, $B = (b_1, \ldots, b_y)^T$, we denote concatenation of $A$ and $B$ by

$$A \oplus B := (a_1, \ldots, a_x, b_1, \ldots, b_y)^T.$$

For an integer $n$, we denote the set $\{1, 2, \ldots, n\}$ by $[n]$. For any $x, y \in \mathbb{R}$, we denote the set $\{x \leq z \leq y \mid z \in \mathbb{R}\}$ by $[x, y]$. For any $X \subseteq \mathbb{R}$ and $y \in \mathbb{R}$, we denote the set $\{x \cdot y \mid x \in X\}$ by $X \cdot y$. Similarly, we denote the set $\{x - y \mid x \in X\}$ by $X - y$. For any function $g : \mathbb{R} \to \mathbb{R}$ and $X \subseteq \mathbb{R}$, we denote the set $\{g(x) \mid x \in X\}$ by $g(X)$.

# 3 Embedding for Hausdorff metric over pointsets in $\mathbb{R}^1$

Below we will work with $H_{s,1}(A, B)$ and we will write $H_s(A, B)$ instead of $H_{s,1}(A, B)$.

We define $10s^2$ functions $f^i : [0, 1] \to [0, 1]$, one for each $i \in \{1, \ldots, 10s^2\}$, as follows.

$$f^i(x) := \begin{cases} \frac{x}{y_i} & \text{if } x \leq y_i; \\ \frac{1-x}{1-y_i} & \text{otherwise,} \end{cases}$$

where $y_i = \frac{1}{3} + \frac{i}{3(10s^2+1)}$. Notice that the function $f^i$ satisfies the following four properties:

1. $f^i$ achieves the maximum value 1 at $y_i$;
2. $\frac{1}{3} < y_i < \frac{2}{3}$;
3. $f^i(0) = f^i(1) = 0$;
4. $f^i$ grows linearly in the interval $[0, y_i]$ and decreases linearly in the interval $[y_i, 1]$.

To prove our results, we need the following lemma.

▶ **Lemma 4.** *Let $A = \{a_1, \ldots, a_s\} \subseteq \mathbb{R}$ and $B = \{b_1, \ldots, b_s\} \subseteq \mathbb{R}$ with $0 = a_1 \leq a_2 \leq \ldots \leq a_s = 1$ and $0 = b_1 \leq b_2 \leq \ldots \leq b_s = 1$. We have*

$$H_s(A, B) \in [1/10, 1000s^2] \cdot \max_i H_{s-1}\left(f^i(A), f^i(B)\right).$$

**Proof.** Notice that, since $f^i(0) = f^i(1) = 0$, we have $|f^i(A)|, |f^i(B)| = s - 1$. This is why we have $H_{s-1}$ in the right side of the equation in the statement of the lemma. From now on we will use $H$ instead of $H_s$ or $H_{s-1}$.

Now we will establish $H(A, B) \geq \frac{1}{10} \max_i(H(f^i(A), f^i(B)))$. It is sufficient to show that $H(A, B) \geq \frac{1}{10} H(f^i(A), f^i(B))$ for all $i \in [10s^2]$. Fix $i \in [10s^2]$. We can check that for all $x, y \in [0, 1]$, $|f^i(x) - f^i(y)| \leq 10|x - y|$ ($f^i$ is a piece-wise linear function with the derivative bounded by 3 in absolute value in every piece). That is, $f^i$ is a Lipschitz function with constant 10 in the interval $[0, 1]$. That means that $f^i$ can increase distance between any two points by a factor of at most 10. Therefore, inequality $H(A, B) \geq \frac{1}{10} H(f^i(A), f^i(B))$ follows.

It remains to show that $H(A, B) \leq 1000s^2 \max_i H_{s-1}\left(f^i(A), f^i(B)\right)$. The remainder of the proof is devoted to show this inequality. We need to show that there exists $i \in [10s^2]$ such that $H(f^i(A), f^i(B)) \geq \frac{1}{1000s^2} H(A, B)$. We will show that there exists $i$ such that

$$\forall a \in A, \quad d(f^i(a), f^i(B)) \geq \frac{1}{1000s^2} d(a, B) \tag{2}$$

and

$$\forall b \in B, \quad d(f^i(b), f^i(A)) \geq \frac{1}{1000s^2} d(b, A), \tag{3}$$

where function $d$ is defined as follows. For point $y$ and finite pointset $X$,

$$d(y, X) := \min_{x \in X} \|x - y\|.$$

The first inequality shows that for every point from $A$, the distance to the closest point from $B$ decreases by a factor of at most $1000s^2$ if we apply map $f^i$ to the point and to the set $B$. Similarly, the second inequality shows that for every point from $B$, the distance to the closest point from $A$ decreases by a factor of at most $1000s^2$ if we apply map $f^i$ to the point and to the set $A$. By the definition of Hausdorff distance, it is sufficient to show these two inequalities to establish what we need.

To prove (2) and (3), we will use the following proposition.

▶ **Proposition 5.** *For all $i_1 \neq i_2$ and $x, y \in [0, 1]$ with $x \leq y$,*

$$\frac{1}{1000s^2} \min(x, y - x) \leq \max\left(|f^{i_1}(x) - f^{i_1}(y)|, |f^{i_2}(x) - f^{i_2}(y)|\right) \tag{4}$$

*and*

$$\frac{1}{1000s^2} \min(1 - y, y - x) \leq \max\left(|f^{i_1}(x) - f^{i_1}(y)|, |f^{i_2}(x) - f^{i_2}(y)|\right). \tag{5}$$

**Proof.** We will show (4). The proof of (5) is analogous. W.l.o.g., $i_1 < i_2$. We have that $y_{i_1} < y_{i_2}$ (see the definition of function $f^i$). If $x \geq y_{i_1}$, we have that $|f^{i_1}(x) - f^{i_1}(y)| \geq y - x$ by the definition of the function $f^{i_1}$. Similarly, if $y \leq y_{i_2}$, we have that $|f^{i_2}(x) - f^{i_2}(y)| \geq y - x$ by the definition of function $f^{i_2}$. Therefore, if $x \geq y_{i_1}$ or $y \leq y_{i_2}$,

$$y - x \leq \max\left(|f^{i_1}(x) - f^{i_1}(y)|, |f^{i_2}(x) - f^{i_2}(y)|\right) \tag{6}$$

and we are done proving (4) and (5).

Now we consider the complement case: $x \leq y_{i_1}$ and $y \geq y_{i_2}$. We will show inequality

$$\frac{1}{1000s^2} x \leq \max\left(|f^{i_1}(x) - f^{i_1}(y)|, |f^{i_2}(x) - f^{i_2}(y)|\right). \tag{7}$$

Notice that, by combining (6) and (7), we get (4). Suppose that

$$|f^{i_1}(x) - f^{i_1}(y)| < \frac{1}{1000s^2} x \tag{8}$$

since otherwise we have established (7). By the definition of $f^{i_1}$ and $f^{i_2}$,

$$f^{i_1}(x) - f^{i_2}(x) = 3x \cdot \left( \frac{1}{1 + \frac{i_1}{10s^2+1}} - \frac{1}{1 + \frac{i_2}{10s^2+1}} \right) \geq \frac{x}{20s^2}, \tag{9}$$

where we use the fact that $i_2 - i_1 \geq 1$. Using inequalities $f^{i_2}(y) \geq f^{i_1} y$, (8) and (9), we get

$$\begin{aligned} |f^{i_2}(x) - f^{i_2}(y)| &\geq \left(f^{i_2}(y) - f^{i_1}(y)\right) + \left(f^{i_1}(x) - f^{i_2}(x)\right) - |f^{i_1}(y) - f^{i_1}(x)| \\ &\geq \frac{x}{20s^2} - \frac{x}{1000s^2} \geq \frac{x}{1000s^2}. \end{aligned}$$

This establishes (7). ◀

Now we continue the proof of Lemma 4. We will use several times the fact that $\{0, 1\} \subseteq A, B$. Consider $a \in A$ and $b \in B$ with $a \leq b$. By (4), inequality

$$\frac{1}{1000s^2} d(a, B) \leq \frac{1}{1000s^2} \min(a, b - a) \leq |f^i(a) - f^i(b)| \tag{10}$$

holds for all indices $i \in [10s^2]$ except at most one. Consider $a \in A$ and $b \in B$ with $a > b$. By (5), inequality

$$\frac{1}{1000s^2} d(a, B) \leq \frac{1}{1000s^2} \min(1 - a, a - b) \leq |f^i(a) - f^i(b)| \tag{11}$$

holds for all indices $i \in [10s^2]$ except at most one. By fixing $a \in A$ and considering all $b \in B$, from (10) and (11) we have that

$$\frac{1}{1000s^2} d(a, B) \leq d(f^i(a), f^i(B)) \tag{12}$$

holds for all indices $i \in [10s^2]$ except for at most $s$ indices.

Analogously we get that for any fixed $b \in B$,

$$\frac{1}{1000s^2} d(b, A) \leq d(f^i(b), f^i(A)) \tag{13}$$

holds for all indices $i \in [10s^2]$ except for at most $s$ indices.

From (12) we get that (2) holds for all but $s^2$ indices. From (13) we get that (3) holds for all but $s^2$ indices. By definition of $f^i$, we consider $10s^2$ indices. We conclude that there must be at least $10s^2 - 2s^2 \geq 1$ index that satisfy both (2) and (3). This concludes the proof of the lemma. ◀

We define function $g^i_t(x) : \mathbb{R} \to \mathbb{R}$, parameterized by $t \in \mathbb{R}$ and $i \in [10s^2]$, as follows:

$$g^i_t(x) := t \cdot f^i\left(\frac{x}{t}\right).$$

▶ **Lemma 6.** *Let $A = \{a_1, \ldots, a_s\} \subseteq \mathbb{R}$ and $B = \{b_1, \ldots, b_s\} \subseteq \mathbb{R}$ with $a_1 \leq \ldots \leq a_s$ and $b_1 \leq \ldots \leq b_s$, and $a_1 = b_1$, and $a_s = b_s$. We have*

$$H_s(A, B) \in [1/10, 1000s^2] \cdot \max_i H_{s-1}\left(g^i_{a_s - a_1}(A - a_1), g^i_{b_s - b_1}(B - b_1)\right).$$

**Proof.** Hausdorff distance is shift invariant, that is, for any $x \in \mathbb{R}$, $H(A, B) = H(A-x, B-x)$. Because of this and $a_1 = b_1$, we can assume that $a_1 = b_1 = 0$. Then the inequality we want to prove simplifies to

$$H_s(A, B) \in [1/10, 1000s^2] \cdot \max_i H_{s-1}\left(g^i_{a_s}(A), g^i_{b_s}(B)\right). \tag{14}$$

By the definition of Hausdorff distance, for any positive $y \in \mathbb{R}$, $H(A, B) = \frac{H(x \cdot A, x \cdot B)}{x}$. Because of this equality, expression (14) follows from Lemma 4 and the definition of $g^i_t$ and $f^i$. ◀

▶ **Lemma 7.** *Let $A = \{a_1, \ldots, a_s\} \subseteq \mathbb{R}$ and $B = \{b_1, \ldots, b_s\} \subseteq \mathbb{R}$ with $a_1 \leq \ldots \leq a_s$ and $b_1 \leq \ldots \leq b_s$. We have*

$$\begin{aligned} H_s(A, B) \in &[1/1000, 10^6 s^2] \\ &\cdot \max\left(|a_1 - b_1|, |a_s - b_s|, \max_i H_{s-1}\left(g^i_{a_s - a_1}(A - a_1), g^i_{b_s - b_1}(B - b_1)\right)\right). \end{aligned}$$

**Proof.** We define pointsets $A' = \{a'_1, \ldots, a'_s\}$ and $B' = \{b'_1, \ldots, b'_s\}$ from $A$ and $B$ in the following way.

1. We set $a'_i = a_i$ and $b'_i = b_i$ for all $i \in [s]$;
2. if $a'_1 < b'_1$, we set $b'_1$ to be equal to $a'_1$;

3. if $a'_1 > b'_1$, we set $a'_1$ to be equal to $b'_1$;
4. if $a'_s < b'_s$, we set $a'_s$ to be equal to $b'_s$;
5. if $a'_s > b'_s$, we set $b'_s$ to be equal to $a'_s$.

We define $M := \max(|a_1 - b_1|, |a_s - b_s|)$.

▶ **Proposition 8.**

$$\left|H\left(g^i_{a_s-a_1}(A-a_1), g^i_{b_s-b_1}(B-b_1)\right) - H\left(g^i_{a'_s-a'_1}(A'-a'_1), g^i_{b'_s-b'_1}(B'-b'_1)\right)\right| \le 100M.$$

**Proof.** Notice that for every $x$, such that $0 \le x \le a_s - a_1$,

$$|g^i_{a_s-a_1}(x) - g^i_{a'_s-a'_1}(x)| \le 50M. \tag{15}$$

This is true because $g^i$ is a Lipshitz function with Lipshitz constant at most 10 and $|a_s - a'_s|, |a_1 - a'_1| \le M$. Similarly, for every $x$, $0 \le x \le b_s - b_1$,

$$|g^i_{b_s-b_1}(x) - g^i_{b'_s-b'_1}(x)| \le 50M. \tag{16}$$

(15) and (16) mean that, as we apply function $g^j$ to set $A'$ instead of $A$ and to $B'$ instead of $B$, every point in the resulting sets (after application of $g^j$) changes its position by at most $50M$, and the assertion follows. ◀

By Lemma 6,

$$H(A', B') \in [1/10, 1000s^2] \cdot \max_i H\left(g^i_{a'_s-a'_1}(A'-a'_1), g^i_{b'_s-b'_1}(B'-b'_1)\right).$$

We get

$$\begin{aligned}
&\max\left(|a_1-b_1|, |a_s-b_s|, \max_i H\left(g^i_{a_s-a_1}(A-a_1), g^i_{b_s-b_1}(B-b_1)\right)\right)\\
\le& M + \max_i H\left(g^i_{a_s-a_1}(A-a_1), g^i_{b_s-b_1}(B-b_1)\right)\\
\le& M + 100M + \max_i H\left(g^i_{a'_s-a'_1}(A'-a'_1), g^i_{b'_s-b'_1}(B'-b'_1)\right)\\
\le& 101M + 10H(A', B')\\
\le& 111M + 10H(A, B)\\
\le& 200H(A, B).
\end{aligned}$$

In the second inequality we use Proposition 8. In the third inequality we use the result of Lemma 6. In the second to last inequality we use $H(A', B') \le M + H(A, B)$, which follows from the definition of $A'$ and $B'$. In the last inequality we use $H(A, B) \ge M$, which follows from the definition of $M$. This shows the lower bound in the statement of the lemma. We prove the upper bound now.

We have

$$\begin{aligned}
H(A, B) &\le M + H(A', B')\\
&\le M + 1000s^2 \max_i H\left(g^i_{a'_s-a'_1}(A'-a'_1), g^i_{b'_s-b'_1}(B'-b'_1)\right)\\
&\le M + 1000s^2\left(\max_i H\left(g^i_{a_s-a_1}(A-a_1), g^i_{b_s-b_1}(B-b_1)\right) + 100M\right)\\
&\le 10^6 s^2 \cdot \max\left(M, H\left(g^i_{a_s-a_1}(A-a_1), g^i_{b_s-b_1}(B-b_1)\right)\right).
\end{aligned}$$

The second inequality follows from Lemma 6. The third inequality follows from Proposition 8. We established the upper bound of the lemma and the proof of Lemma 7 is complete. ◀

▶ **Theorem 9.** *There exists an embedding of $H_{s,1}$ into $\ell_\infty^{s^{O(s)}}$ with distortion $s^{O(s)}$.*

**Proof.** We will construct embedding $f$ of $H_s$ into $\ell_\infty^{s^{O(s)}}$ with distortion $s^{O(s)}$. Let $A = \{a_1, \ldots, a_s\} \subseteq \mathbb{R}$ and $B = \{b_1, \ldots, b_s\} \subseteq \mathbb{R}$ with $a_1 \leq \ldots \leq a_s$ and $b_1 \leq \ldots \leq b_s$. By Lemma 7, we can bound $H(A, B)$ in terms of the maximum of $|a_1 - b_1|$, $|a_s - b_s|$ and

$$H_{s-1}\left(g^i_{a_s-a_1}(A - a_1), g^i_{b_s-b_1}(B - b_1)\right)$$

over all $i \in [10s^2]$. By Lemma 7, we lose a factor $O(s^2)$ in the distortion. Notice that pointsets $g^i_{a_s-a_1}(A - a_1)$ and $g^i_{b_s-b_1}(B - b_1)$ are of size $s - 1$. That is, we decreased the number of points in the sets by 1. Also notice that the functions $g^i_{a_s-a_1}$ and $g^i_{b_s-b_1}$ depend only on sets $A$ and $B$, respectively. The idea now is to apply this expression recursively until we arrive at pointsets of size 1, which we can embed into $l_\infty$ trivially. More precisely, we define the following recursive embedding $h^s$ of pointset $A$ of size $s$. If $s \geq 2$,

$$h^s(A) := \left(a_1, a_s, h^{s-1}(g^1_{a_s-a_1}(A - a_1)), \ldots, h^{s-1}(g^{10s^2}_{a_s-a_1}(A - a_1))\right).$$

If $s = 1$, then $h^1(A) = (a_1)$. $h^s(A)$ is concatenation of values $a_1$, $a_s$ and $10s^2$ vector defined recursively by $h^{s-1}$. We define $f(A) := h^{|A|}(A)$. We call the recursive embedding at most $s$ times, each time increases number of dimensions by a factor of $O(s^2)$ and the distortion by a factor of $O(s^2)$. This means that the final distortion is $\leq [O(s^2)]^s \leq s^{O(s)}$ and the dimension is $\leq [O(s^2)]^s \leq s^{O(s)}$. ◀

## 4 Embedding for Hausdorff metric over pointsets in $\mathbb{R}^d$

▶ **Theorem 10.** *There exists an embedding of $H_{s,d}$ into $\ell_\infty^{s^{O(s+d)}}$ with distortion $s^{O(s+d)}$ for an arbitrary integer $d \geq 1$.*

**Proof.** It suffices to consider the case $d > 1$, since the case $d = 1$ has been handled in the previous Section. Given sets $A, B \subseteq \ell_2^d$ of size $|A| = |B| = s$, we show how to produce sets $A_1, \ldots, A_{2s^2+1}$ and $B_1, \ldots, B_{2s^2+1}$ with the following properties.

1. Each $A_i$ depends on $A$ only. Each $B_i$ depends on $B$ only.
2. For every $i$, $A_i, B_i \subseteq \ell_2^{d-1}$ and $|A_i| = |B_i| = s$.
3. For every $i$, $H_{s,d-1}(A_i, B_i) \leq H_{s,d}(A, B)$.
4. There exists $i$ such that $H_{s,d-1}(A_i, B_i) \geq \frac{1}{Cs^2} H_{s,d}(A, B)$ for sufficiently large constant $C$.

From the properties we see that

$$H_{s,d}(A, B) \geq \max_{i=1,\ldots,2s^2+1} H_{s,d-1}(A_i, B_i) \geq \frac{1}{Cs^2} H_{s,d}(A, B)$$

where $A$ and $B$ are any two subsets of $d$ dimensional space and $A_i, B_i$ are subsets of $d - 1$ dimensional space. If we repeat the construction $d - 1$ times in total, we get embedding that satisfies inequality

$$H_{s,d}(A, B) \geq \max_{j=1,\ldots,(2s^2+1)^{d-1}} H_{s,1}(A'_j, B'_j) \geq \frac{1}{(Cs^2)^{d-1}} H_{s,d}(A, B).$$

Now we apply Theorem 9 to embed sets $A'_j, B'_j$ into $l_\infty^{s^{O(s)}}$ with distortion $s^{O(s)}$. The final dimension of the embedding is $(2s^2+1)^{d-1} \cdot s^{O(s)} = s^{O(s+d)}$ as promised. The final distortion of the embedding is $(Cs^2)^{d-1} \cdot s^{O(s)} = s^{O(s+d)}$ as promised.

In the remainder of the proof we show how to construct the embedding with the four properties stated at the beginning of the proof. Consider the first two vectors of the standard basis of $\ell_2^d$. These two vectors span a plane. Choose $2s^2+1$ unit vectors $v_1, \dots, v_{2s^2+1}$ in this plane so that angle between vectors $v_i$, $v_{i+1}$ is $2\pi/(2s^2+1)$ for all $i = 1, \dots, 2s^2+1$ and $v_1$ is the first standard basis vector of $\ell_2^d$. We define $v_{2s^2+2} := v_1$. We build $A_i$ ($B_i$, resp.) by projecting $A$ ($B$, resp.) on the hyperplane perpendicular to $v_i$ for all $i = 1, \dots 2s^2+1$. The first property follows from the definition of $A_i$ and $B_i$. The second property follows because every hyperplane of $\ell_2^d$ span $\ell_2^{d-1}$. The third property follows because projection on hyperplane can only decrease interpoint distances. It remains to show the fourth property. Consider any pair of points $a \in A$ and $b \in B$. There can be at most two values $i$ such that

$$\|\Pi_i(a) - \Pi_i(b)\|_2 < \frac{1}{10000s^2}\|a-b\|_2, \tag{17}$$

where $\Pi_i$ denotes projection on the hyperplane defined by vector $v_i$. This is true because of the following considerations. Consider $i$ such that inequality (17) does not hold. Then we must have

$$|v_i \cdot (a-b)| > \left(1 - \frac{1}{10000s^2}\right) \cdot \|v_i\|_2 \cdot \|a-b\|_2.$$

However, this can happen to at most two vectors $v_i$ by the construction of $v_i$. Because there are $2s^2+1$ vectors $v_i$ and at most $s^2$ pairs $(a,b)$, $a \in A$, $b \in B$ determine distance $H_s(A,B)$, the fourth property follows. ◀

## 5 Probabilistic embedding

▶ **Theorem 11.** *For any $\alpha \in (0, 1/2)$ and integer $\Delta > 0$, there exists a probabilistic embedding $f$ of $H_s$ over subsets of $[\Delta]$ into $8\Delta$-dimensional $\ell_1$ space $\ell_1^{8\Delta}$ that satisfies the following properties. For any two pointsets $A, B \subseteq [\Delta]$ with $|A| = |B| = s$,*

1. $\frac{1}{10} H_s^{1-\alpha}(A,B) \le \|f(A) - f(B)\|_1$;
2. $E[\|f(A) - f(B)\|_1] \le 100s/\alpha \cdot H_s^{1-\alpha}(A,B)$,

*where the expectation in the second property is over the randomness of the embedding.*

**Proof.** W.l.o.g. we assume that $\log_2 \Delta$ is positive integer. If this is not so, we increase $\Delta$ to $2^{\lceil \log_2 \Delta \rceil}$. For integer $y$ and finite set $X \subseteq \mathbb{R}$, we define

$$y + X := \{x + y \mid x \in X\}.$$

For integer $i$, $0 \le i \le \log_2(2\Delta)$, and finite set $X \subseteq \mathbb{R}$, we define vector $f_i(X) \in \mathbb{R}^{\frac{2\Delta}{2^i}}$ as follows. For $l = 1, \dots, \frac{2\Delta}{2^i}$,

$$(f_i(X))_l = \begin{cases} 1 & \exists x \in X \text{ such that } (l-1)2^i < x \le l \cdot 2^i; \\ 0 & \text{otherwise.} \end{cases}$$

For integer $v$ and finite set $X \subseteq \mathbb{R}$, we define embedding $f_v(X)$:

$$f_v(X) := \bigoplus_{i=0}^{\log_2(2\Delta)} 2^{i(1-\alpha)} f_i(v + X).$$

Embedding $f$ is defined by choosing $v \in \{1, \dots, \Delta\}$ uniformly at random and setting $f(X) = f_v(X)$. Now we will show that the embedding satisfy the stated properties. Let $h = H_s(A,B)$.

The following lemma establishes the first inequality.

▶ **Lemma 12.** *For every $v$,*

$$\|f_v(A) - f_v(B)\|_1 \geq \frac{1}{10} h^{1-\alpha}.$$

**Proof.** We assume that $h \geq 2$. If $h = 1$, then $\|f_v(A) - f_v(B)\|_1 \geq 1 \geq \frac{1}{10}$. If $h = 2$, there is nothing to prove. W.l.o.g., let $a \in A$ be such that $d(a, B) = h$. Let $i = (\log_2 h) - 1 \geq 0$ and $l = \left\lceil \frac{a+v}{2^i} \right\rceil$. Because $a \in A$ and by the choice of $l$, we have

$$(f_i(v + A))_l = 1.$$

Because $d(a, B) = h$ and by the choice of $i$ and $l$, we have

$$(f_i(v + B))_l = 0.$$

Therefore, we conclude:

$$\begin{aligned}\|f_v(A) - f_v(B)\|_1 &\geq 2^{i(1-\alpha)} \|f_i(v + A) - f_i(v + B)\|_1 \\ &\geq \left(\frac{h}{2}\right)^{1-\alpha} |\,(f_i(v + A))_l - (f_i(v + B))_l\,| \geq \frac{1}{10} h^{1-\alpha}.\end{aligned}$$ ◀

The following lemma establishes the second inequality.

▶ **Lemma 13.** *$E[\|f(A) - f(B)\|_1] \leq 100s/\alpha \cdot h^{1-\alpha}$.*

**Proof.**

▶ **Proposition 14.** *For every $i \in \{0, \ldots, \log_2(2\Delta)\}$,*

$$E_v[\|f_i(v + A) - f_i(v + B)\|_1] \leq 2s \min(1, 2h/2^i).$$

**Proof.** We define an undirected bipartite graph $G = (A, B, E)$ as follows. For every $a \in A$, we add edge $(a, b)$, $b \in B$ such that $d(a, B) = |a - b|$. If there are multiple possibilities for $b$, we choose one $b$ arbitrarily. For every $b \in B$, we add edge $(a, b)$, $a \in A$ such that $d(b, A) = |a - b|$. If there are multiple possibilities for $a$, we choose one $a$ arbitrarily.

By the definition of Hausdorff distance and $f_i$, we get

$$E_v[\|f_i(v + A) - f_i(v + B)\|_1] \leq \sum_{(a,b) \in E} \Pr_{v \in [\Delta]} \left[ \left\lceil \frac{a+v}{2^i} \right\rceil \neq \left\lceil \frac{b+v}{2^i} \right\rceil \right].$$

We can upper bound every probability in the latter quantity by $\min(1, 2h/2^i)$ because for every $(a, b) \in E$, $|a - b| \leq h$. We get the bound stated in the proposition because $|E| \leq 2s$ by the definition of graph $G$. ◀

Using this proposition, we get

$$\begin{aligned}E[\|f(A) - f(B)\|_1] &\leq \sum_{i=0}^{\log_2(2\Delta)} 2^{i(1-\alpha)} E_v[\|f_i(v + A) - f_i(v + B)\|_1] \\ &\leq 2s \sum_{i=0}^{\log_2(2\Delta)} 2^{i(1-\alpha)} \min(1, 2h/2^i) \\ &\leq 2s \sum_{i=1}^{1+\log_2 h} \left(2^{i(1-\alpha)}\right) + 4sh \sum_{i=2+\log_2 h}^{\log_2(2\Delta)} 2^{-i\alpha} \\ &\leq 20sh^{1-\alpha} + 4sh \left(2^{-\alpha}\right)^{2+\log_2 h} \sum_{i=0}^{\infty} \left(2^{-\alpha}\right)^i\end{aligned}$$

$$\leq 20sh^{1-\alpha} + \frac{20sh^{1-\alpha}}{\alpha}$$
$$\leq 100s/\alpha \cdot h^{1-\alpha},$$

which is what we needed. ◀

◀

▶ **Lemma 15.** *Let $U := \log_2 s + \log_2 h + 10$. Then*

$$\Pr_v[\forall i \geq U\ ,\ f_i(v+A) = f_i(v+B)] \geq 0.9.$$

**Proof.** Since $f_U(v+A) = f_U(v+B)$ implies that $f_i(v+A) = f_i(v+B)$ for all $i \geq U$, it is sufficient to prove that $\Pr_v[f_U(v+A) = f_U(v+B)] \geq 0.9$.

Let $G = (A, B, E)$ be the bipartite graph defined in Proposition 14. Since for all $(a, b) \in E$, $|a - b| \leq h$, we have

$$\begin{aligned}
\Pr_v[f_U(v+A) = f_U(v+B)] &\geq \Pr_v\left[\forall (a,b) \in E\ ,\ \left\lceil \frac{a+v}{2^i} \right\rceil = \left\lceil \frac{b+v}{2^i} \right\rceil\right] \\
&\geq 1 - \sum_{(a,b)\in E} \Pr_v\left[\left\lceil \frac{a+v}{2^i} \right\rceil \neq \left\lceil \frac{b+v}{2^i} \right\rceil\right] \\
&\geq 1 - |E| \cdot \frac{2h}{2^U} \geq 1 - 2s \cdot \frac{2h}{2^U} \\
&\geq 0.9.
\end{aligned}$$

◀

▶ **Lemma 16.** *Let $L := \log_2 h - \frac{1}{1-\alpha} \log s - 20$. For every $v \in \{1, \ldots, \Delta\}$,*

$$\left\| \bigoplus_{i=0}^{L} 2^{i(1-\alpha)}(f_i(v+A) - f_i(v+B)) \right\|_1 \leq \frac{h^{1-\alpha}}{1000}.$$

**Proof.** We use the definition of $\oplus$ and $L$:

$$\sum_{i=0}^{L} \left\| 2^{i(1-\alpha)}(f_i(v+A) - f_i(v+B)) \right\| \leq 2s \cdot 5 \cdot 2^{L(1-\alpha)} \leq \frac{h^{1-\alpha}}{1000}.$$

◀

From Lemmas 15 and 16, we get that, with probability $\geq 0.9$ the following happens. Almost all $\ell_1$ mass of $\|f(A) - f(B)\|_1$ comes from $U - L - 1 \leq 100 \log_2 s$ vectors $f_i(v+A) - f_i(v+B)$. $f_i(v+A) - f_i(v+B)$ that correspond to $i \geq U$ or $i < L$ contribute at most $h^{1-\alpha}/1000$ to the $\ell_1$ mass. Also notice that, by Theorem 11, the $\ell_1$ mass of $f(A) - f(B)$ is at least $h^{1-\alpha}/10$. We get that we lose relatively small amount of $\ell_1$ mass by discarding of many vectors $f^i$. We will use these observations in Theorem 23 below to reduce the dimensionality of the target $\ell_1$ space in Theorem 11.

▶ **Definition 17.** Let $G = (A, B, E)$ be a bipartite graph. We call it $r$-regular if, for every vertex $a \in A$, the degree of $a$ is equal to $r$.

▶ **Definition 18.** Graph $G = (A, B, E)$ is called random $r$-regular bipartite graph if it comes from a distribution defined by the following process. Initially, $E = \emptyset$. For every $a \in A$ we choose a subset of $r$ distinct vertices of $B$ uniformly at random and connect $a$ to the all chosen vertices.

▶ **Definition 19.** Let $G = ([n],[m],E)$ be $r$-regular bipartite graph for some integers $r, n, m \geq 1$. We define matrix $\Phi_G \in \mathbb{R}^{m\times n}$ as follows.

$$(\Phi_G)_{i,j} = \begin{cases} \frac{1}{r} & \text{if } (i,j) \in E; \\ 0 & \text{otherwise.} \end{cases}$$

for all $i \in [m]$ and $j \in [n]$.

The following lemma can be shown using the probabilistic method.

▶ **Lemma 20.** *Let $G = ([n],[O(n/\delta^2)],E)$ be random $r$-regular bipartite graph for $r = O(1/\delta)$. For any subset $X$ of vertices, let $N(X)$ denote the set of neighbors of vertices in $X$. Then we have*

$$\Pr_G[\forall X \subseteq A \ : \ |N(X)| \geq (1-\delta)r|X|] \geq 0.99.$$

The following result was shown in [3].

▶ **Theorem 21.** *Let $G = ([n],[m],E)$ be some $r$-regular bipartite graph with the property that*

$$\forall X \subseteq A \ : \ |N(X)| \geq (1-\delta)r|X|.$$

*Let $\Phi_G$ be the matrix according as in Definition 19. Then we have that for all $x \in \mathbb{R}^n$,*

$$(1 - O(\delta))\|x\|_1 \leq \|\Phi_G x\|_1 \leq \|x\|_1.$$

Below we will need the following lemma.

▶ **Lemma 22.** *Let $G = ([n'],O(n/\delta^2),E)$ be random $r$-regular bipartite graph for $r = O(1/\delta)$. Then for every $x \in \mathbb{R}^{n'}$ with $\|x\|_0 \leq n$ (number of non-zero entries of $x$ is at most $n$),*

$$\Pr_G[(1 - O(\delta))\|x\|_1 \leq \|\Phi_G x\|_1 \leq \|x\|_1] \geq 0.99.$$

**Proof.** Consider matrix $\Phi_G$ restricted to the columns corresponding to the non-zero entries of $x$. This matrix correspond to random $r$-regular bipartite graph with at most $n$ vertices on the left side. By Lemma 20, this matrix will satisfy the requirement for Theorem 21 with probability at least 0.99, concluding the proof. ◀

▶ **Theorem 23.** *For any $\alpha \in (0, 1/2)$ and integer $\Delta > 0$, there exists a probabilistic embedding $f'$ of $H_s$ into $\ell_1^{O(s \log s)}$ that satisfies the following properties. For any two pointsets $A, B \subseteq [\Delta]$ with $|A| = |B| = s$,*
1. $\frac{1}{100} H_s^{1-\alpha}(A,B) \leq \|f'(A) - f'(B)\|_1$ *with probability* $\geq 2/3$;
2. $E[\|f'(A) - f'(B)\|_1] \leq 100s/\alpha \cdot H_s^{1-\alpha}(A,B)$.

**Proof.** Let $C_1, C_2 > 0$ be large constants and $\delta > 0$ be a small enough constant that we will set later.

Let $G = ([8\Delta], \frac{C_1 \cdot 200 s \log_2 s}{\delta^2}, E)$ be random $\frac{C_2}{\delta}$-regular bipartite graph. By Lemma 22, for all $x \in \mathbb{R}^{8\Delta}$ with $\|x\|_0 \leq 200 s \log_2 s$,

$$\Pr_G[0.9 \cdot \|x\|_1 \leq \|\Phi_G x\|_1 \leq \|x\|_1] \geq 0.99, \tag{18}$$

where we choose $C_1$ and $C_2$ be large enough constants and $\delta > 0$ to be a small enough constant so that $1 - O(\delta) \geq 0.9$.

Let $f_v(X)$ be the embedding as in Theorem 11. For graph $G$ and integer $v$, we define embedding $f'_{G,v}(X) := \Phi_G \cdot f_v(X)$. Embedding $f'(X)$ is defined by choosing uniformly random $v \in [\Delta]$ and $G$. Property 2 in the theorem follows since, for some $G$ and $v$,

$$\|f'(A) - f'(B)\|_1 = \|\Phi_G \cdot (f_v(A) - f_v(B))\|_1 \leq \|f_v(A) - f_v(B)\|_1$$

and property 2 of Theorem 11. We used the fact that matrix $\Phi_G$ is a left stochastic matrix (that is, all columns of it sum up to 1) to conclude the inequality.

The remainder of the proof is devoted to show the first property. Consider entries of $f_v(A) - f_v(B)$ corresponding to embedding $f_i$ for $i = U, \ldots, \log_2(2\Delta)$. By Lemma 15, with probability $\geq 0.9$, all these entries are 0. We assume that this happens from now on. Consider entries of $f_v(A) - f_v(B)$ that correspond to embeddings $f_i$ for $i = 0, \ldots, L$. By Lemma 16, the total sum of absolute values of these entries is upper bounded by $h^{1-\alpha}/1000$. We set all these entries (corresponding to $f_0, \ldots, f_L$) to 0. Because $\Phi_G$ is left stochastic, we change the value of $\|f'(A) - f'(B)\|_1$ by at most $h^{1-\alpha}/1000$. Now the only entries that are nonzero in $f_v(A) - f_v(B)$ correspond to $f_{L+1}, \ldots, f_{U-1}$. The total number of nonzero entries is at most

$$(|A| + |B|) \cdot (U - L - 1) \leq 2s \cdot 100 \log_2 s \leq 200 s \log_2 s.$$

By (18), we have that with probability $\geq 0.99$,

$$\begin{aligned}
\|f'(A) - f'(B)\|_1 &\geq 0.9 \cdot \|f_v(A) - f_v(B)\|_1 - h^{1-\alpha}/1000 \\
&\geq 0.9 \cdot \left(\frac{1}{10} h^{1-\alpha} - h^{1-\alpha}/1000\right) - h^{1-\alpha}/1000 \\
&\geq \frac{1}{100} h^{1-\alpha},
\end{aligned}$$

which is what we need. In the second inequality we used the first property from Theorem 11 and the fact that we did set all entries, corresponding to $f_i$ with $i \leq L$, to 0. The lower bound holds with probability $1 - 0.1 - 0.01 \geq 2/3$ by using the union bound. ◀

**Acknowledgments.** We thank Piotr Indyk for many helpful discussions.

## References


1. Artūrs Bačkurs and Piotr Indyk. Better embeddings for planar earth-mover distance over sparse sets. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, SOCG'14, pages 280:280–280:289, New York, NY, USA, 2014. ACM. doi:10.1145/2582112.2582120.
2. Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.
3. Radu Berinde, Anna C. Gilbert, Piotr Indyk, H. Karloff, and Martin J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 798–805. IEEE, 2008.
4. J. Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Mathematics*, 52(1-2):46–52, 1985. doi:10.1007/BF02776078.
5. Manuel Costa, Miguel Castro, Antony Rowstron, and Peter Key. Pic: Practical internet coordinates for distance estimation. In *Distributed Computing Systems, 2004. Proceedings. 24th International Conference on*, pages 178–187. IEEE, 2004.
6. Aryeh Dvoretzky. *Some results on convex bodies and Banach spaces.* Hebrew University, 1960.

**7** Martin Farach-Colton and Piotr Indyk. Approximate nearest neighbor algorithms for hausdorff metrics via embeddings. In *Foundations of Computer Science, 1999. 40th Annual Symposium on*, pages 171–179. IEEE, 1999.

**8** Sariel Har-Peled and Manor Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM Journal on Computing*, 35(5):1148–1184, 2006.

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

**10** Piotr Indyk. Algorithmic applications of low-distortion geometric embeddings. In *focs*, page 10. IEEE, 2001.

**11** Piotr Indyk. *High-dimensional Computational Geometry*. PhD thesis, Stanford University, 2001.

**12** Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM (JACM)*, 53(3):307–323, 2006.

**13** Piotr Indyk, Avner Magen, Anastasios Sidiropoulos, and Anastasios Zouzias. On-line embeddings. In *Proc. of APPROX*, 2010.

**14** William B Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.

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

**16** Ofer Neiman. Low dimensional embeddings of doubling metrics. In Christos Kaklamanis and Kirk Pruhs, editors, *Approximation and Online Algorithms*, volume 8447 of *Lecture Notes in Computer Science*, pages 12–23. Springer International Publishing, 2014.

**17** Santosh S Vempala. *The random projection method*, volume 65. American Mathematical Soc., 2005.