

Dimension Reduction Techniques for ℓ_p ($1 \leq p \leq 2$), with Applications

Yair Bartal¹ and Lee-Ad Gottlieb²

1 Hebrew University, Jerusalem, Israel

yair@cs.huji.ac.il

2 Ariel University, Ariel, Israel

leead@ariel.ac.il

Abstract

For Euclidean space (ℓ_2), there exists the powerful dimension reduction transform of Johnson and Lindenstrauss [26], with a host of known applications. Here, we consider the problem of dimension reduction for all ℓ_p spaces $1 \leq p \leq 2$. Although strong lower bounds are known for dimension reduction in ℓ_1 , Ostrovsky and Rabani [40] successfully circumvented these by presenting an ℓ_1 embedding that maintains fidelity in only a bounded distance range, with applications to clustering and nearest neighbor search. However, their embedding techniques are specific to ℓ_1 and do not naturally extend to other norms.

In this paper, we apply a range of advanced techniques and produce bounded range dimension reduction embeddings for all of $1 \leq p \leq 2$, thereby demonstrating that the approach initiated by Ostrovsky and Rabani for ℓ_1 can be extended to a much more general framework. We also obtain improved bounds in terms of the intrinsic dimensionality. As a result we achieve improved bounds for proximity problems including snowflake embeddings and clustering.

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Dimension reduction, embeddings

Digital Object Identifier 10.4230/LIPIcs.SoCG.2016.16

1 Introduction

Dimension reduction for normed space is a fundamental tool for algorithms and related fields. A much celebrated result for dimension reduction is the well-known l_2 flattening lemma of Johnson and Lindenstrauss [26]: For every n -point subset of l_2 and every $0 < \varepsilon < 1$, there is a mapping into l_2^k that preserves all interpoint distances in the set within factor $1 + \varepsilon$, with target dimension $k = O(\varepsilon^{-2} \log n)$. The dimension reducing guarantee of the Johnson-Lindenstrauss (JL) transform is remarkably strong, and has the potential to make algorithms with a steep dependence on the dimension tractable. It can be implemented as a simple linear transform, and has proven to be a very popular tool in practice, even spawning a stream of literature devoted to its analysis, implementation and extensions (see for example [5, 4, 2]).

Given the utility and impact of the JL transform, it is natural to ask whether these strong dimension reduction guarantees may hold for other ℓ_p spaces as well (see [23] for further motivation). This is a fundamental open problem in embeddings, and has attracted significant research. Yet the dimension reduction bounds known for ℓ_p norms other than ℓ_2 are much weaker than those given by the JL transform, and it is in fact known that a linear dimension reduction mapping in the style the JL transform is quite unique to the ℓ_2 -norm [27]. Further, strong lower bounds on dimension reduction are known for ℓ_1 [7] and for ℓ_∞



© Yair Bartal and Lee-Ad Gottlieb;

licensed under Creative Commons License CC-BY

32nd International Symposium on Computational Geometry (SoCG 2016).

Editors: Sándor Fekete and Anna Lubiw; Article No. 16; pp. 16:1–16:15

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

[34], and it is a plausible conjecture that ℓ_2 is the only ℓ_p space which admits the strong distortion and dimension properties of the JL transform.

Ostrovsky and Rabani [30, 39, 40] successfully circumvented the negative results for ℓ_1 , and presented a dimension reduction type embedding for the Hamming cube – and by extension, all of l_1 – which preserves fidelity only in a bounded range of distances. They further demonstrated that their embedding finds use in algorithms for nearest neighbor search (NNS) and clustering, as these can be reduced to subproblems where all relevant distance are found in a bounded range. In fact, a number of other proximity problems can also be reduced to bounded range subproblems, including the construction of distance oracles and labels [22, 10], snowflake embeddings [19, 11], and ℓ_p -difference of a pair of data streams. Hence, we view a bounded range embedding as a framework for the solution of multiple important problems. Note also that for spaces with fixed aspect ratio (a fixed ratio between the largest and smallest distances in the set – a common scenario in the literature, see [25]), a bounded range mapping is in effect a complete dimension reduction embedding.

The dimension reduction embedding of Ostrovsky and Rabani is very specific to ℓ_1 , and does not naturally extend to other norms. The central contribution of the paper is to bring advanced techniques to bear on this problem, thereby extending this framework to all $1 \leq p \leq 2$.

1.1 Our contribution

We first present a basic embedding in Section 3, which shows that we can reduce dimension while realizing a certain distance transform with low distortion. Using this result, we are able to derive two separate dimension-reducing embeddings:

- Range embedding. In Theorem 4, we present an embedding which preserves distances in a given range with $(1 + \varepsilon)$ distortion. The target dimension is $O(\log n)$, with dependence on the range parameter and ε . This generalizes the approach of Ostrovsky and Rabani [39, 40] to all $1 \leq p \leq 2$. This embedding can be applied in the streaming setting as well, and it can also be modified to achieve target dimension polynomial in the *doubling dimension* of the set (Lemma 5).¹
- Snowflake embedding. An α -snowflake embedding is one in which each inter-point distance t in the origin space is replaced by distance t^α in the embedded space (for $0 < \alpha < 1$). It was observed in [19, 11] that the snowflake of a finite metric space in l_2 may be embedded in dimension which is close to the intrinsic dimension of the space (measured by its doubling dimension), and this may be independent of n . In [19] the case of l_1 was considered as well, however the resulting dimension had doubly exponential dependence on the doubling dimension. We demonstrate that the basic embedding can be used to build a snowflake for ℓ_p for all $1 \leq p \leq 2$ with dimension polynomial in the doubling dimension; this is found in Lemma 7. For ℓ_1 this provides a *doubly exponential* improvement over

¹ Our range embedding has the additional property that it can be used to embed ℓ_p^m into $\ell_q^{O(\log n)}$ (that is, m -dimensional ℓ_p into $O(\log n)$ -dimensional ℓ_q , for $1 \leq q < p$) with $(1 + \varepsilon)$ -distortion in time $O(m \log n)$, with dependence on the range parameter and ε . This is a fast version of the embedding of [28], a common tool for embedding ℓ_p into more malleable spaces such as l_1 . (The embedding of [28] is particularly useful for nearest neighbor search, see [30, 21].) Note that [28] features a large overhead cost $O(m^2)$, and since m can be as large as $\Theta(n)$, this overhead can be the most expensive step in algorithms for ℓ_p . We also note that the embedding of Theorem 4 can be used to produce efficient algorithms for approximate nearest neighbor search and ℓ_p difference, although other efficient techniques have already been developed for these specific problems (see for example [8, 37] for NNS, and [24, 32] for ℓ_p difference).

the previously known dimension bound [19], while generalizing the $p \in \{1, 2\}$ results of [19, 11] to all $1 \leq p \leq 2$.

To highlight the utility of our embeddings, we demonstrate (in Section 5) their applicability in deriving better runtime bounds for clustering problems:

We first consider the k -center problem, and show that our snowflake embedding can be used to provide an efficient algorithm. For ℓ_p -spaces of low doubling dimension and fixed p , $1 \leq p \leq 2$, we apply our snowflake embedding in conjunction with the clustering algorithm of Agarwal and Procopiu [3], and obtain a $(1 + \varepsilon)$ -approximation to the k -center problem in time $O(n(2^{\tilde{O}(\text{ddim}(S))} + \log k)) + (k \cdot \varepsilon^{-\text{ddim}(S)/\varepsilon^2})^{k^{1 - (\varepsilon/\text{ddim}(S))^{O(1)}}}$.

We then consider the min-sum clustering problem, and show that our snowflake embedding can be used to provide an efficient algorithm for this problem. For ℓ_p -spaces of low doubling dimension ($1 \leq p \leq 2$), we apply our snowflake embedding in conjunction with the clustering algorithm of Schulman [44], and obtain a $(1 + \varepsilon)$ -approximation to the min-sum clustering problem in randomized time $n^{O(1)} + 2^{2^{O(d')}} (\varepsilon \log n)^{O(d')}$ where $d' = (\text{ddim}/\varepsilon)^{O(1)}$.

1.2 Related work

For results on dimension reduction for ℓ_p spaces, see [43] for $p < 2$, and also [45] for $p = 1$, [46] for $1 < p < 2$, and [34] for $p = \infty$. Other related notions of dimension reduction have been suggested in the literature. Indyk [24] devised an analogue to the JL-Lemma which uses p -stable distributions to produce estimates of interpoint distances; strictly speaking, this is not an embedding into ℓ_p (e.g. it uses median over the coordinates). Motivated by the nearest neighbor search problem, Indyk and Naor [25] proposed a weaker form of dimension reduction, and showed that every doubling subset $S \subset \ell_2$ admits this type of dimension reduction into ℓ_2 of dimension $O(\text{ddim}(S))$. Roughly speaking, this notion is weaker in that distances in the target space are allowed to err in one direction (be too large) for all but one pair of points. Similarly, dimension reduction into ordinal embeddings (where only relative distance is approximately preserved) was considered in [6, 14]. Bartal, Recht and Schulman [11] developed a variant of the JL-Lemma that is local – it preserves the distance between every point and the \hat{k} points closest to it. Assuming $S \subset \ell_2$ satisfies a certain growth rate condition, they achieve, for any desired \hat{k} and $\varepsilon > 0$, an embedding of this type with distortion $1 + \varepsilon$ and dimension $O(\varepsilon^{-2} \log \hat{k})$. The same paper built on a result of [19] to show that for ℓ_2 a $(1 + \varepsilon)$ -approximate α -snowflake embeds into $\tilde{O}(\varepsilon^{-3} \text{ddim}(S))$ dimensions (for fixed α). For ℓ_1 , [19] showed that a $(1 + \varepsilon)$ -approximate α -snowflake embeds into $2^{\varepsilon^{-\tilde{O}(\text{ddim}(S))}}$ dimensions.

2 Preliminaries

2.1 Embeddings and metric transforms

Following [13], we define an *oblivious* embedding to be an embedding which can be computed for any point of a database set X or query set Y , without knowledge of any other point in X or Y . (This differs slightly from the definition put forth by Indyk and Naor [25].) Familiar oblivious embeddings include standard implementations of the JL-Lemma for ℓ_2 , the dimension reduction mapping of Ostrovsky and Rabani [40] for the Hamming cube, and the embedding of Johnson and Schechtman [28] for ℓ_p , $p \leq 2$. A *transform* is a function mapping from the positive reals to the positive reals, and a *metric transform* maps a metric distance function to another metric distance function on the same set of points. An embedding is

transform preserving with respect to a transform if it achieves the distances defined by that transform.

Let $(X, d_X), (Y, d_Y)$ be metric spaces. For distance scales $0 \leq a \leq b \leq \infty$, an $[a, b]$ -embedding of X into Y with distortion D is a mapping $f : X \rightarrow Y$ such that for all $x, y \in X$ such that $d_X(x, y) \in [a, b]$: $1 \leq c \cdot \frac{d_Y(f(x), f(y))}{d_X(x, y)} \leq D$. (Here, c is any scaling constant.) Then f is a *range embedding* with range $[a, b]$. If f has the additional property that for all x, y such that $d(x, y) < a$ we have $c \cdot \frac{d_Y(f(x), f(y))}{a} \leq D$, then we say that f is range preserving from below. Similarly, if for all x, y such that $d(x, y) > b$ we have $c \cdot \frac{d_Y(f(x), f(y))}{b} \geq 1$, then we say that f is range preserving from above. And if f is range preserving from above and below, then we say that f is a *strong* range embedding.

Let $R > 1$ be a parameter. We say that X admits an R -range embedding into Y with distortion D if for every $u > 0$ there exists an $[u, uR]$ embedding of X into Y with distortion D . As above, an R -range embedding may be range preserving from above or below. We will usually take $u = 1$.

2.2 Doubling dimension, nets and hierarchies

For a metric (\mathcal{X}, ρ) , let $\lambda > 0$ be the smallest value such that every ball in \mathcal{X} can be covered by λ balls of half the radius. The *doubling dimension* of \mathcal{X} is $\text{ddim}(\mathcal{X}) = \log_2 \lambda$. Note that $\text{ddim}(\mathcal{X}) \leq \log n$. The following packing property can be shown (see, for example [29]): Suppose that $S \subset \mathcal{X}$ has a minimum interpoint distance of at least α . Then $|S| \leq \left(\frac{2 \text{diam}(S)}{\alpha} \right)^{\text{ddim}(\mathcal{X})}$.

Given a metric space S , $S' \subset S$ is a γ -net of S if the minimum interpoint distance in S' is at least γ , while the distance from every point of S to its nearest neighbor in S' is less than γ . Let S have minimum inter-point distance 1. A *hierarchy* is a series of $\lceil \log \Delta \rceil$ nets (Δ being the aspect ratio of S), where each net S_i is a 2^i -net of the previous net S_{i-1} . The first (or bottom) net is $S_0 = S$, and the last (or top) net S_t contains a single point called the *root*. For two points $u \in S_i$ and $v \in S_{i-1}$, if $d(u, v) < 2^i$ then we say that u covers v , and this definition allows v to have multiple covering points in S_i . The closest covering point of v is its *parent*. The distance from a point in S_i to its ancestor in S_j is at most $\sum_{k=i+1}^j 2^k = 2 \cdot (2^j - 2^{i+1}) < 2 \cdot 2^j$.

Given S , a hierarchy for S can be built in time $2^{O(\text{ddim}(S))}n$, and this term bounds the size of the hierarchy as well [22, 16]. The height of the hierarchy is $O(\min\{n, \log \Delta\})$.

2.3 Probabilistic partitions

Probabilistic partitions are a common tool used in embeddings. Let (X, d) be a finite metric space. A partition P of X is a collection of non-empty pairwise disjoint clusters $P = \{C_1, C_2, \dots, C_t\}$ such that $X = \cup_j C_j$. For $x \in X$ we denote by $P(x)$ the cluster containing x . We will need the following decomposition lemma due to Gupta, Krauthgamer and Lee [20], Abraham, Bartal and Neiman [1], and Chan, Gupta and Talwar [15].² Let ball $B(x, r) = \{y \mid \|x - y\| \leq r\}$.

► **Theorem 1** (Padded Decomposition of doubling metrics [20, 1, 15]). *There exists a constant $c_0 > 1$, such that for every metric space (X, d) , every $\varepsilon \in (0, 1)$, and every $\delta > 0$, there is a*

² [20] provided slightly different quantitative bounds than in Theorem 1. The two enumerated properties follow, for example, from Lemma 2.7 in [1], and the bound on support-size m follows by an application of the Lovász Local Lemma sketched therein.

multi-set $\mathcal{D} = [P_1, \dots, P_m]$ of partitions of X , with $m \leq c_0 \varepsilon^{-1} \dim(X) \log \dim(X)$, such that (i) Bounded radius: $\text{diam}(C) \leq \delta$ for all clusters $C \in \bigcup_{i=1}^m P_i$; and (ii) Ball padding: If P is chosen uniformly from \mathcal{D} , then for all $x \in X$, $\Pr_{P \in \mathcal{D}} [B(x, \frac{\delta}{c_0 \dim(X)}) \subseteq P(x)] \geq 1 - \varepsilon$.

2.4 Stable distributions

The density of symmetric p -stable random variables ($0 < p \leq 2$) is $h(x) = \frac{1}{\pi} \int_0^\infty \cos(tx) e^{-t^p} dt$ [47, Section 2.2]. The density function $h(x)$ is unimodal and bell-shaped. It is well known that $\frac{c_p}{1+x^{p+1}} \leq h(x) \leq \frac{c'_p}{1+x^{p+1}}$ for constants c_p, c'_p that depend only on p [38] (and we may ignore the dependence on p for the purposes of this paper). Using this approximation for $h(x)$, it is easy to see that for $0 < q < p$ and p -stable random variable g , $\mathbb{E}[g^q] = \int_0^\infty x^q h(x) dx \approx \int_0^1 x^q dx + \int_1^\infty x^{q-(p+1)} dx \approx \frac{1}{p-q}$. (Here we use the notation \approx in the same sense as $\Theta(\cdot)$.) Also, for $b > 1$, $\int_0^b x^p h(x) dx \approx \int_0^1 x^p dx + \int_1^b \frac{1}{x} dx \approx 1 + \ln b$. Let $g_1, g_2, \dots, g_m \in G$ be a set of i.i.d. symmetric p -stable random variables, and let v be a real m -length vector. A central property of p -stable random variables is that $\sum_{j=1}^m g_j v_j$ is distributed as $g(\sum_{j=1}^m v_j^p)^{1/p} = g \|v\|_p$, for all $g \in G$. When all $g_i \in G$ are normalized as $\mathbb{E}[|g_i|^q] = 1$, we have that $\mathbb{E} \left[\left| \sum_{j=1}^m g_j v_j \right|^q \right] = \mathbb{E}[|g|^q] \|v\|_p^q = \|v\|_p^q$.

3 Basic transform-preserving embedding

In Theorem 3 below, we present an embedding which realizes a certain distance transform with low distortion, while also reducing dimension to $O(\log n)$ (with dependence on the range and the desired distortion). We then demonstrate that the transform itself has several desirable properties. In the next section, we will use this transform-preserving embedding to obtain a range embedding and a snowflake embedding.

3.1 Embedding

We first present a randomized embedding into a single coordinate, and show that it is transform-preserving in expectation only (Lemma 2). We then show that a concatenation of many single-coordinate embeddings yields a single embedding which (approximately) preserves the bounded transform with high probability, and this gives Theorem 3.

The following single-coordinate embedding is inspired by the Nash device of Bartal *et al.* [11] (see also [42]), and is related to the spherical threshold function of Mendel and Naor [35, Lemma 5.9]. Broadly speaking, our embedding uses the sine function as a dampening tool, which serves to mitigate undesirable properties of p -stable distributions (i.e., their heavy tails). Our central contribution is to give a tight analysis of the transform preserved by our embedding.³

Let $S \subset \ell_p$ be a set of m -dimensional vectors for $1 \leq p \leq 2$. Let $g_j \in G$ be k i.i.d. symmetric p -stable random variables, and fix parameters s (the *threshold*) and $0 \leq \phi \leq 2\pi$. Further define the fixed constant $P_q = \mathbb{E}[|\cos \theta|^q] = \frac{1}{2\pi} \int_0^{2\pi} |\cos \theta|^q d\theta$ for $1 \leq q \leq p$ (and note that $\frac{1}{2} = \frac{1}{2\pi} \int_0^{2\pi} \cos^2 \theta d\theta \leq P_q \leq \frac{1}{2\pi} \int_0^{2\pi} |\cos \theta| d\theta = \frac{2}{\pi}$). Then the embedding $F_{\phi,s} : S \rightarrow L_q^1$

³ It may be possible to replace our embedding with that of Mendel and Naor [35], but one would still require a tighter analysis, and the final dimension would likely increase. Note also that our embedding is into the reals, while [35] embed into the complex numbers; as ℓ_p^d over \mathbb{C} embeds into $\ell_p^{O(\varepsilon^{-2} \sqrt{p} 2^{p/2} d)}$ over \mathbb{R} with distortion $1 + \varepsilon$ (a consequence of Dvoretzky's theorem [36]), the embedding of [35] can in fact be used to achieve an embedding into the reals with increased dimension.

(i.e., into a single coordinate of L_q) for vector $v \in S$ is defined by $F_s(v) = F_{\phi,s}(v) = \frac{s}{2P_q^{1/q}} \sin(\phi + \frac{2}{s} \sum_{i=1}^m g_i v_i)$. Note that $0 \leq F_s(v) < s$. For vectors $v, w \in S$ we have that $|F_s(v) - F_s(w)|^q = \frac{s^q}{2^q P_q} \left| \sin(\phi + \frac{2}{s} \sum_{i=1}^m g_i v_i) - \sin(\phi + \frac{2}{s} \sum_{i=1}^m g_i w_i) \right|^q$
 $= \frac{s^q}{P_q} \left| \sin(\frac{1}{s} \sum_{i=1}^m g_i (v_i - w_i)) \cos(\phi + \frac{1}{s} \sum_{i=1}^m g_i (v_i + w_i)) \right|^q$. Now, when ϕ is a random variable chosen uniformly from the range $[0, 2\pi]$ and independently of set G , we have that $\mathbb{E}[|F_s(v) - F_s(w)|^q] = \frac{s^q}{P_q} \mathbb{E} \left[\left| \sin(\frac{1}{s} \sum_{i=1}^m g_i (v_i - w_i)) \cos(\phi + \frac{1}{s} \sum_{i=1}^m g_i (v_i + w_i)) \right|^q \right] = \frac{s^q}{P_q} \mathbb{E} \left[\left| \sin(\frac{1}{s} \sum_{i=1}^m g_i (v_i - w_i)) \cos(\phi) \right|^q \right] = \frac{s^q}{P_q} \mathbb{E} \left[\left| \sin(\frac{1}{s} \sum_{i=1}^m g_i (v_i - w_i)) \right|^q \right] \cdot \mathbb{E} \left[|\cos(\phi)|^q \right] = s^q \mathbb{E} \left[\left| \sin(\frac{1}{s} \sum_{i=1}^m g_i (v_i - w_i)) \right|^q \right]$, where the second equality follows from the periodicity of the cosine function, and the independence of ϕ and G .

This is our single-coordinate embedding. In Lemma 2 below, we will describe its behavior on interpoint distances – that is, we derive useful bounds on $\mathbb{E}[|F_s(v) - F_s(w)|^q]$. Now $\frac{1}{s} \sum_{i=1}^m g_i (v_i - w_i)$ is distributed as $g \frac{\|v-w\|_p}{s}$ for random variable $g \in G$, so we will set $a = \frac{\|v-w\|_p}{s}$ and will derive bounds for $H(a) = \mathbb{E}[|\sin(ag)|^q] = s^{-q} \mathbb{E}[|F_s(v) - F_s(w)|^q]$. But first we introduce the full embedding f , which is a scaled concatenation of k single-coordinate embeddings: Independently for each coordinate i , create a function $F_{\phi_i} = F_{\phi_i,s}$ by fixing a random angle $0 \leq \phi_i \leq 2\pi$ and a family of k i.i.d. symmetric p -stables. Then coordinate $f(v)_i$ is defined by $f(v)_i = k^{-1/p} F_{\phi_i,s}(v)$ which can be constructed in $O(m)$ time per coordinate. Note that for $t = \|v - w\|_p$, $\mathbb{E}[|f(v)_i - f(w)_i|^q] = \frac{1}{k} \mathbb{E}[|F_{\phi_i}(v) - F_{\phi_i}(w)|^q] = \frac{s^q}{k} H(t/s)$, and so $\mathbb{E}[|f(v) - f(w)|_p^q] = s^q H(t/s)$. Below, we will show that with high probability f is transform-preserving with respect to H with low distortion.

3.2 Analysis of basic embedding

We now show that both the single-coordinate embedding and the full embedding have desirable properties. Recall that $h(u)$ is the density function of p -stable random variables. Set $Q = 2 \int_0^\infty u^q h(u) du \approx \frac{1}{p-q}$. Also, for $a < 1$ and some fixed $a^2 < \varepsilon < 1$, set $Q_a = \frac{1}{2} \int_0^{\sqrt{\varepsilon}/a} u^p h(u) du = \frac{1}{2} \left[\int_0^1 u^p h(u) du + \Theta(\ln(\sqrt{\varepsilon}/a)) \right] \approx 1 + \ln(\sqrt{\varepsilon}/a)$.

► **Lemma 2.** *Let g be a symmetric p -stable random variable. For $1 \leq q \leq p \leq 2$ and any fixed $0 < \varepsilon < \frac{1}{2}$, $H(a) = \mathbb{E}[|\sin(ag)|^q]$ obeys the following:*

- (a) *Threshold: $H(a) \leq 1$.*
- (b) *Bi-Lipschitz for small scales: When $q < p$ and $a \leq \min\{\varepsilon^{\frac{1}{2} + \frac{1}{p-q}}, \sqrt{\varepsilon}(1 + (p-q)\varepsilon^{-(\frac{q}{2}+1)})^{\frac{1}{p-q}}\}$, we have $1 - O(\varepsilon) \leq \frac{H(a)}{a^q Q} \leq 1 + O(\varepsilon)$.
When $q = p$ and $a \leq \sqrt{\varepsilon} e^{-\varepsilon^{-(\frac{q}{2}+1)}}$, we have $1 - O(\varepsilon) \leq \frac{H(a)}{a^q Q_a} \leq 1 + O(\varepsilon)$.*
- (c) *Bi-Lipschitz for intermediate scales: When $q < p$ and $a < 1$, let $\delta = 1 - a^{p-q}$, and we have $H(a) = \Theta(1 + \delta Q) a^q$. When $q = p$ we have $H(a) = \Theta(1 + \ln(1/a)) a^q$.*
- (d) *Lower bound for large scales: When $a \geq 1$, $H(a) > \frac{1}{8}$.*
- (e) *Smoothness: When $a \leq 1$, $\frac{|H((1+\varepsilon)a) - H(a)|}{H(a)} = O(\varepsilon)$.*

It follows that the distance transform implied by our single-coordinate embedding F_s achieves the bounds of Lemma 2 scaled by s^q , if only in expectation. Item 2 implies that for very small a (i.e., when the inter-point distance under consideration is sufficiently small with respect to the parameter s) the embedding has very small expected distortion (at least when $q < p$). Weaker distortion bounds hold for distances smaller than s (item 3). For distances greater than s , these are contracted to the threshold (item 1) or slightly smaller (item 4). The smoothness property will be useful for constructing the snowflake in Section 4.3. We proceed to consider the full embedding:

► **Theorem 3.** *Let $1 \leq q \leq p \leq 2$ and $0 < \varepsilon < \frac{1}{2}$, and consider an n -point set $S \in l_p^m$. Set threshold $s > 1$. Then with constant probability the oblivious embedding $f : S \rightarrow l_q^k$ for $k = O\left(\frac{\log n}{\varepsilon^2} \cdot \min\left\{s^{2q}, \max\left\{\frac{s^{2q-p}}{2q-p}, \varepsilon s^q\right\}\right\}\right)$, satisfies the following for each point pair $v, w \in S$, where $t = \|v - w\|_p$:*

1. *Threshold: $\|f(v) - f(w)\|_q^q < s^q$.*
2. *Bi-Lipschitz for large scales: When $t \geq 1$, we have $(1 - \varepsilon)s^q H(t/s) \leq \|f(v) - f(w)\|_q^q \leq (1 + \varepsilon)s^q H(t/s)$.*
3. *Bounded expansion for small scales: When $t < 1$, we have $\|f(v) - f(w)\|_q^q \leq s^q H(1/s) + \varepsilon$. The embedding can be constructed in $O(mk)$ time per point.*

Theorem 3 demonstrates that in a certain range, there exists a transform-preserving embedding with high probability. (We note that when $2q$ is close to p , better dimension bounds can be obtained by embedding into an interim value $q + c$ for some c and then embedding into q .)

Proof of Lemma 2.

Item 1. This follows trivially from the fact that $|\sin(x)| \leq 1$.

Item 2. Note that since the density function h is symmetric about 0,

$$H(a) = 2 \int_0^{\sqrt{\varepsilon}/a} |\sin(au)|^q h(u) du + 2 \int_{\sqrt{\varepsilon}/a}^\infty |\sin(au)|^q h(u) du. \tag{1}$$

We show that under the conditions of the item, the second term in Equation (1) is dominated by the first. Considering the second term, we have that $2 \int_{\sqrt{\varepsilon}/a}^\infty |\sin(au)|^q h(u) du < 2 \int_{\sqrt{\varepsilon}/a}^\infty h(u) du < 2c'_p \int_{\sqrt{\varepsilon}/a}^\infty \frac{1}{u^{p+1}} du = \frac{2c'_p}{p} \left(\frac{a}{\sqrt{\varepsilon}}\right)^p$.

Turning to the first term, recall the Taylor series expansion $\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$; so when $x < \sqrt{\varepsilon}$ we have that $x(1 - \frac{\varepsilon}{6}) < \sin(x) < x$. Also note that the conditions of the item give that $a < \sqrt{\varepsilon}$, and so $\frac{\sqrt{\varepsilon}}{a} > 1$. Hence, when $q < p$ we have $2 \int_0^{\sqrt{\varepsilon}/a} |\sin(au)|^q h(u) du > 2(1 - \varepsilon/6)^q a^q \int_0^{\sqrt{\varepsilon}/a} u^q h(u) du > 2(1 - \varepsilon/3)a^q c_p \left[\int_0^1 \frac{u^q}{1+u^{p+1}} du + \int_1^{\sqrt{\varepsilon}/a} \frac{u^q}{1+u^{p+1}} du\right] > 2(1 - \varepsilon/3)a^q c_p \left[\int_0^1 \frac{u^q}{2} du + \int_1^{\sqrt{\varepsilon}/a} \frac{u^{q-p-1}}{2} du\right] = (1 - \varepsilon/3)a^q c_p \left[\frac{1}{q+1} + \frac{1-(a/\sqrt{\varepsilon})^{p-q}}{p-q}\right]$. When $q = p$, the same analysis gives $2 \int_0^{\sqrt{\varepsilon}/a} |\sin(au)|^q h(u) du > (1 - \varepsilon/3)a^q c_p \left[\frac{1}{q+1} + \ln(\sqrt{\varepsilon}/a)\right]$

We proceed to show that the first term of Equation (1) exceeds the second by a factor of $\Omega(\varepsilon^{-1})$. For $q = p$ this holds trivially when $a^q \ln(\sqrt{\varepsilon}/a) \geq \frac{1}{\varepsilon} \left(\frac{a}{\sqrt{\varepsilon}}\right)^p$ – or equivalently, when $\ln(\sqrt{\varepsilon}/a) \geq \varepsilon^{-(\frac{q}{2}+1)}$ – which in turn holds exactly when $a \leq \sqrt{\varepsilon} e^{-\varepsilon^{-(\frac{q}{2}+1)}}$. For $q < p$, the condition is fulfilled when $a^q \geq \frac{1}{\varepsilon} \left(\frac{a}{\sqrt{\varepsilon}}\right)^p$, which holds exactly when $a \leq \varepsilon^{\frac{p/2+1}{p-q}}$. Better, the condition is also fulfilled when $a^q \left[\frac{1-(a/\sqrt{\varepsilon})^{p-q}}{p-q}\right] \geq \frac{1}{\varepsilon} \left(\frac{a}{\sqrt{\varepsilon}}\right)^p$, and this is equivalent to satisfying $a^{p-q} \left[\frac{p-q}{\varepsilon^{\frac{p}{2}+1}} + \frac{1}{\varepsilon^{(p-q)/2}}\right] \leq 1$; this holds exactly when $a \leq \sqrt{\varepsilon}(1 + (p - q)\varepsilon^{-(q/2+1)})^{-1/(p-q)}$.

As the first term of Equation (1) dominates the second, it follows that when $q < p$ then for some constant c $H(a) = 2 \int_0^{\sqrt{\varepsilon}/a} |\sin(au)|^q h(u) du + 2 \int_{\sqrt{\varepsilon}/a}^\infty |\sin(au)|^q h(u) du \leq 2(1 + c\varepsilon) \int_0^{\sqrt{\varepsilon}/a} |\sin(au)|^q h(u) du < 2(1 + c\varepsilon)a^q \int_0^\infty u^q h(u) du = (1 + c\varepsilon)a^q Q$. Further, we have that $H(a) = 2 \int_0^{\sqrt{\varepsilon}/a} |\sin(au)|^q h(u) du + 2 \int_{\sqrt{\varepsilon}/a}^\infty |\sin(au)|^q h(u) du > 2 \int_0^{\sqrt{\varepsilon}/a} |\sin(au)|^q h(u) du > 2(1 - \varepsilon/3)a^q \int_0^{\sqrt{\varepsilon}/a} u^q h(u) du > (1 - \varepsilon/3)(1 - c'\varepsilon)a^q Q$, where the final inequality follows

from noting that since $a \leq \varepsilon^{\frac{1}{2} + \frac{1}{p-q}}$, $\int_{\sqrt{\varepsilon}/a}^{\infty} u^q h(u) du \leq \int_{\varepsilon^{-1/(p-q)}}^{\infty} u^q h(u) du \approx \frac{\varepsilon}{p-q} \approx \varepsilon Q$, and so $\int_0^{\sqrt{\varepsilon}/a} u^q h(u) du = \int_0^{\infty} u^q h(u) du - \int_{\sqrt{\varepsilon}/a}^{\infty} u^q h(u) du > (1 - c'\varepsilon)Q$ for some c' . This completes the analysis for $q < p$. The same analysis gives that when $q = p$, $H(a) < 2(1 + c\varepsilon)a^q \int_0^{\sqrt{\varepsilon}/a} u^q h(u) du = (1 + c\varepsilon)a^q Q_a$, and $H(a) > 2(1 - \varepsilon/3)a^q \int_0^{\sqrt{\varepsilon}/a} u^q h(u) du = (1 - \varepsilon/3)a^q Q_a$.

Item 3. The analysis is similar to that presented in the proof of Item 2. Noting that when $0 \leq x \leq 1$, $\sin(x) \approx x$, and recalling that under the conditions of the item $a \leq 1$, we have for $q < p$ that $H(a) = 2 \int_0^{1/a} |\sin(au)|^q h(u) du + 2 \int_{1/a}^{\infty} |\sin(au)|^q h(u) du = \Theta \left(a^q \int_0^{1/a} u^q h(u) du \right) + O \left(\int_{1/a}^{\infty} h(u) du \right) = \Theta \left(\left(1 + \frac{1-a^{p-q}}{p-q} \right) a^q \right) + O(a^p) = \Theta \left(\left(1 + \frac{\delta}{p-q} \right) a^q \right) + O(a^p) = \Theta \left(\left(1 + \frac{\delta}{p-q} \right) a^q \right)$. Similarly, for $q = p$ we have $H(a) = \Theta \left(a^q \int_0^{1/a} u^p h(u) du \right) + O \left(\int_{1/a}^{\infty} h(u) du \right) = \Theta \left(a^q \int_0^{1/a} u^p h(u) du \right) \approx (1 + \ln(1/a)) a^p$.

Item 4. First note that when $p \geq 1$, $h(x) = \frac{1}{\pi} \int_0^{\infty} \cos(tx) e^{-t^p} dt < \frac{1}{\pi} \int_0^{\infty} e^{-t^p} dt < \frac{1}{\pi}$, so $\int_0^b h(u) du < \frac{b}{\pi}$. Since $h(x)$ is a symmetric density function, we have $\int_0^{\infty} h(u) du = \frac{1}{2}$, and so $\int_b^{\infty} h(u) du > \frac{1}{2} - \frac{b}{\pi}$. We have for any $0 < \theta < \frac{\pi}{2}$,

$$\begin{aligned} H(a) &\geq 2 \int_{\theta/a}^{\infty} |\sin(au)|^q h(u) du > 2 \sum_{i=0}^{\infty} \int_{\frac{i\pi+\theta}{a}}^{\frac{(i+1)\pi-\theta}{a}} |\sin(au)|^q h(u) du \\ &> 2 |\sin(\theta)|^q \sum_{i=0}^{\infty} \int_{\frac{i\pi+\theta}{a}}^{\frac{(i+1)\pi-\theta}{a}} h(u) du > 2 |\sin(\theta)|^q \left[1 - \frac{2\theta}{\pi} \right] \sum_{i=0}^{\infty} \int_{\frac{i\pi+\theta}{a}}^{\frac{(i+1)\pi+\theta}{a}} h(u) du \\ &= 2 |\sin(\theta)|^q \left[1 - \frac{2\theta}{\pi} \right] \int_{\theta/a}^{\infty} h(u) du > |\sin(\theta)|^q \left[1 - \frac{2\theta}{\pi} \right] \left[1 - \frac{2\theta}{a\pi} \right], \end{aligned}$$

where the fourth inequality follows from the fact that $h(x)$ is monotone decreasing for $x \geq 0$. The claimed result follows by taking $a = 1$ (the maximum value of a) and $\theta = \frac{\pi}{4}$, and recalling that $q \leq 2$.

Item 5. As in the proof of Item 3 above, we have that for $q \leq p$ and $a \leq 1$, $H(a) \approx \int_0^{1/a} |\sin(au)|^q h(u) du + \int_{1/a}^{\infty} h(u) du = \int_0^{1/a} |\sin(au)|^q h(u) du + O(a^p) \approx \int_0^{1/a} |\sin(au)|^q h(u) du$. Now recall the Taylor series expansion $\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots > 1 - \frac{x^2}{2}$ (when $0 \leq x \leq 1$), and note that as a consequence of the Mean Value Theorem, $\|A\|^q - \|B\|^q = qC^{q-1} \|A - B\|$ for some $|A| \geq C \geq |B|$. Further noting that when $u \leq \frac{1}{a}$ we have $au \leq 1$ and that when $u \leq \frac{1}{\varepsilon a}$ we have $\varepsilon au \leq 1$, we conclude that

$$\begin{aligned} H(a(1 + \varepsilon)) - H(a) &= 2 \int_0^{\infty} [|\sin(a(1 + \varepsilon)u)|^q - |\sin(au)|^q] h(u) du \\ &\leq 2 \int_0^{\infty} q [\max\{|\sin(a(1 + \varepsilon)u)|, |\sin(au)|\}]^{q-1} \\ &\quad [|\sin(a(1 + \varepsilon)u)| - |\sin(au)|] h(u) du \\ &\leq 2 \int_0^{\infty} q [\max\{|\sin(au) \cos(\varepsilon au)| + |\sin(\varepsilon au) \cos(au)|, |\sin(au)|\}]^{q-1} \\ &\quad [|\sin(au) \cos(\varepsilon au)| + |\sin(\varepsilon au) \cos(au)| - |\sin(au)|] h(u) du \end{aligned}$$

$$\begin{aligned}
 &\leq 2 \int_0^\infty q[|\sin(au)|^{q-1} + |\sin(\varepsilon au)|^{q-1}][|\sin(\varepsilon au)|]h(u)du \\
 &= O\left(\int_0^{1/a} \varepsilon|\sin(au)|^qh(u)du + \varepsilon a \int_{1/a}^{1/(\varepsilon a)} uh(u)du + \int_{1/(\varepsilon a)}^\infty h(u)du\right) \\
 &= O\left(\int_0^{1/a} \varepsilon|\sin(au)|^qh(u)du + \varepsilon a^p + \varepsilon^p a^p\right) \\
 &= O(\varepsilon H(a)). \quad \blacktriangleleft
 \end{aligned}$$

Proof of Theorem 3. The first claim of the theorem follows from the fact that for all i , $0 \leq F_{\phi_i,s}(v) < s$. To prove the rest of the theorem, we may make use of Hoeffding’s inequality. When $1 \leq t \leq s$, we have by Lemma 2234 that $s^q H(t/s) = s^q \Omega(1) = \Omega(1)$. Then Claim 10 implies that for some $k = O\left(\frac{s^{2q}}{\varepsilon^2} \log n\right)$, $\Pr\left[\|f(v) - f(w)\|_q^q - s^q H(t/s) > \varepsilon s^q H(t/s)\right] = \Pr\left[\|f(v) - f(w)\|_q^q - s^q H(t/s) > \varepsilon \Omega(1)\right] \leq \frac{1}{n^2}$, so this distortion guarantee can hold simultaneously for all point pairs. When $t < 1$, first note that when $H(t/s) = \Theta(1)$, we can use the same proof as for $1 \leq t \leq s$ above. If $H(t/s) = o(1)$, $\Pr\left[\|f(v) - f(w)\|_q^q > s^q H(1/s) + \varepsilon\right] = \Pr\left[\|f(v) - f(w)\|_q^q - s^q H(t/s) > s^q H(1/s) - s^q H(t/s) + \varepsilon\right] < \Pr\left[\|f(v) - f(w)\|_q^q - s^q H(t/s) > \varepsilon\right] \leq \frac{1}{n^2}$.

An alternate bound can be derived using Bennett’s inequality (Claim 11). For this, it suffices to take $k = O\left(\frac{s^{2q} \log n}{\sigma^2 V\left(\frac{s^q \varepsilon [s^q H(t/s)]}{\sigma^2}\right)}\right)$, with the variance term $\sigma^2 = \Theta\left(s^{2q} \mathbb{E}[|\sin(ag)|^{2q}]\right)$.

We will prove the case $t \geq 1$, and the case $t < 1$ follows as above. Set $r = \frac{s^q \varepsilon [s^q H(t/s)]}{\sigma^2}$. If $r \geq 1$, then we have $V(r) = \Omega(r)$, and so $k = O\left(\frac{s^q \log n}{\varepsilon [s^q H(t/s)]}\right) = O\left(\frac{s^q \log n}{\varepsilon}\right)$. Otherwise $r < 1$, and we have $V(r) = \Theta(r^2)$, from which we derive $k = O\left(\frac{\sigma^2 \log n}{\varepsilon^2 [s^q H(t/s)]^2}\right)$. Now, if $t \geq s$, we recall by 24 that $H(t/s) = \Theta(1)$, and noting that $\sigma^2 \approx s^{2q} \mathbb{E}[|\sin(ag)|^{2q}] = O(s^{2q})$ we obtain $k = O\left(\frac{\log n}{\varepsilon^2}\right)$. When $1 \leq t < s$, we have by 223 that $H(t/s) = \Omega((t/s)^q)$, and so $k = O\left(\frac{\sigma^2 \log n}{\varepsilon^2 t^{2q}}\right)$. In this case we require a better bound on σ^2 : Setting $a = \frac{t}{s} < 1$ we have (by analysis similar to the proof of Lemma 2) $\sigma^2 \approx s^{2q} \mathbb{E}[|\sin(ag)|^{2q}] \approx s^{2q} \left[a^{2q} \int_0^1 u^{2q} du + a^{2q} \int_1^{1/a} u^{2q-p-1} du + \int_{1/a}^\infty u^{-p-1} du \right] \approx s^{2q} \left[\frac{a^{2q}}{2q+1} + \frac{a^p}{2q-p} + \frac{a^p}{p} \right] \approx s^{2q} \frac{a^p}{2q-p}$. It follows that $k = O\left(\frac{a^{p-2q} \log n}{(2q-p)\varepsilon^2}\right) = O\left(\frac{s^{2q-p} \log n}{(2q-p)\varepsilon^2}\right)$. \blacktriangleleft

4 Range and snowflake embeddings

In Section 3 above, we presented our basic embedding. Here, we show how to use the basic embedding to derive a dimension-reducing embedding that preserves distances in a fixed range (Section 4.1). We also show that the basic embedding can be used to derive a dimension-reducing snowflake embedding (Section 4.3). As a precursor to the snowflake embedding, we show that the basic and range embeddings can be improved to embed into the doubling dimension of the space (Section 4.2).

4.1 Range embedding

By combining Theorem 3 and Lemma 2 and choosing an appropriate parameter s , we can achieve a dimension-reducing oblivious strong range embedding for ℓ_p . This is the central contribution of our paper:

► **Theorem 4.** Let $1 \leq q \leq p \leq 2$ and $0 < \varepsilon < \frac{1}{2}$, and consider an n -point set $S \in \mathcal{L}_p^m$. Fix range $R > 1$ and set threshold s as follows: When $q < p$, $s \approx R\varepsilon^{-1/2}(1 + (p - q)\varepsilon^{-(q/2+1)})^{1/(p-q)}$, and when $q = p$, $s \approx \max\{R^{1/\varepsilon}, R\varepsilon^{-1/2}e^{\varepsilon^{-(\frac{3}{2}+1)}}\}$. Then there exists an oblivious embedding $f : S \rightarrow \ell_q^k$ for $k = O\left(\frac{\log n}{\varepsilon^2} \cdot \min\left\{s^{2q}, \max\left\{\frac{s^{2q-p}}{2^{q-p}}, \varepsilon s^q\right\}\right\}\right)$, which satisfies the following for each point pair $v, w \in S$, where $t = \|v - w\|_p$:

(a) *Threshold:* When $q < p$ we have $\|f(v) - f(w)\|_q^q \leq \frac{s^q}{Q}$.

When $q = p$ we have $\|f(v) - f(w)\|_q^q \leq \frac{s^q}{Q_{R/s}}$.

(b) *Bounded expansion and contraction for large scales:* When $t > R$, we have $\|f(v) - f(w)\|_q^q = O(t^q)$, and $\|f(v) - f(w)\|_q^q \geq R^q + \varepsilon$.

(c) *Bi-Lipschitz for intermediate scales:* When $1 \leq t \leq R$, we have $(1 - \varepsilon)t^q \leq \|f(v) - f(w)\|_q^q \leq (1 + \varepsilon)t^q$.

(d) *Bounded expansion for small scales:* When $t < 1$, we have $\|f(v) - f(w)\|_q^q \leq 1 + \varepsilon$.

The embedding can be constructed in $O(mk)$ time per point.

Proof. We use the construction of Theorem 3 with the stated value of s , and then scale down by a factor of $Q^{1/q}$ or $Q_{R/s}^{1/q}$. (Note that $Q, Q_{R/s} = \Omega(1)$.) Then the threshold guarantee follows immediately from Theorem 31 and the scaling step.

We will now prove the rest of the theorem for the case $p < q$. First, the bi-Lipschitz claim for values $1 \leq t \leq R$ follows immediately from Theorem 32 and Lemma 22, when noting that for an appropriate choice of s , $\frac{t}{s} \leq \varepsilon^{1/2}(1 + (p - q)\varepsilon^{-(q/2+1)})^{-1/(p-q)}$. Similarly, the bounded expansion claim for $t < 1$ follows from Theorem 33 and the aforementioned bi-Lipschitz guarantee at $t = 1$. Finally, the bounded expansion for $t > R$ follows from Theorem 32, and the bounded contraction follows from Theorem 32 combined with fact that when $a > R/s$, $H(a) > H(R/s)$ (as a consequence of Lemma 234 for an appropriate choice of s).

For $p = q$, we have essentially the same proof, only noting that the function $Q_a = Q_{t/s}$ is monotone decreasing in t , and since $s = O(R^{1/\varepsilon})$ we have for all $1 \leq t \leq R$ that $\frac{Q_{t/s}}{Q_{R/s}} \leq \frac{Q_{1/s}}{Q_{R/s}} = O\left(\frac{1 + \varepsilon^{-1} \log R}{1 + (\varepsilon^{-1} - 1) \log R}\right) = O(1 + \varepsilon)$. ◀

4.2 Intrinsic dimensionality reduction

Here we show that the guarantees of Theorem 3 and Theorem 4 can be achieved by (non-oblivious) embeddings whose target dimension is independent of n , and depends only on the doubling dimension of the space. The following lemma is derived by applying the framework of [19] to Theorem 3.

► **Lemma 5.** Let $1 \leq q \leq p \leq 2$ and $0 < \varepsilon < \frac{1}{2}$, and consider an n -point set $S \in \mathcal{L}_p^m$. Set threshold $s > 1$. There exists an embedding $f : S \rightarrow \ell_q^k$ for $k = \tilde{O}\left(\frac{\text{ddim}^2(S)}{\varepsilon^3} \cdot s \cdot \min\left\{s'^{2q}, \max\left\{\frac{s'^{2q-p}}{2^{q-p}}, \varepsilon s'^q\right\}\right\}\right)$ for $s' = \frac{s \text{ddim}(S)}{\varepsilon}$, which satisfies the following for each pair $v, w \in S$, where $t = \|v - w\|_p$:

(a) *Threshold:* $\|f(v) - f(w)\|_q^q \leq s^q$.

(b) *Bi-Lipschitz for intermediate scales:* When $1 \leq t \leq s$, we have $(1 - \varepsilon)s^q H(t/s) \leq \|f(v) - f(w)\|_q^q \leq (1 + \varepsilon)s^q H(t/s)$.

(c) *Strong bounded expansion for small scales:* When $t < 1$, for some constant c $\|f(v) - f(w)\|_q^q \leq \min\{(1 + \varepsilon)s^q H(1/s), c \text{ddim}(S)t\}$.

Given a point hierarchy for S , the embedding can be constructed in $2^{\tilde{O} \text{ddim}(S)} + O(mk)$ time per point.

Proof. Similar to what was done in [19], we compute for S a padded decomposition with padding s . This is a multiset $[P_1, \dots, P_m]$ where each partition P_i is a set of clusters, and

every point is s -padded in a $(1 - \frac{\varepsilon}{s})$ fraction of the partitions. Each cluster has diameter bounded by $O(s \text{ddim}(S))$, and the support is $m = \tilde{O}(s\varepsilon^{-1} \text{ddim}(S))$. Using the hierarchy, this can be done in time $2^{\tilde{O}(\text{ddim}(S))}$ per point [10].

We embed each partition P_i separately as follows: For each cluster $C \in P_i$, we extract from C an $\frac{\varepsilon}{\text{ddim}(S)}$ -net $N \subset C$. Now each cluster net has aspect ratio $\frac{s \text{ddim}^2(S)}{\varepsilon}$, and so $|N| = (\frac{s}{\varepsilon})^{\tilde{O}(\text{ddim}(S))}$. We then scale N up by a factor of $\frac{\text{ddim}(S)}{\varepsilon}$ so that the minimum distance is 1, invoke the embedding of Theorem 3 with parameter s' on N , and scale back down. This procedure has a runtime cost of $O(mk)$ per point, thresholds all distances at s , and reduces dimension to $\tilde{O}\left(\frac{\text{ddim}(S)}{\varepsilon^2} \cdot \min\left\{s'^{2q}, \max\left\{\frac{s'^{2q-p}}{2q-p}, \varepsilon s'^q\right\}\right\}\right)$. We then concatenate the m cluster partitions together and scale down by $m^{1/q}$, achieving an embedding of dimension k for the net points. We then extend this embedding to all points using the $c \text{ddim}(S)$ -factor Lipschitz extension of Lee and Naor [31] for metric space.

For the net points, item 1 holds immediately. For item 2, we note that when $1 \leq t \leq s$, the points fall in the same cluster in a fraction $(1 - \frac{\varepsilon}{s})$ of the partitions, and these partitions contribute the correct amount to the interpoint distance. However, in a fraction $\frac{\varepsilon}{s}$ of the partition the points are not found in the same cluster, and in these cases the contribution may be as large as s . These partitions account for an additive value of at most $\frac{\varepsilon}{s} \cdot s = \varepsilon \leq \varepsilon t$. Item 3 follows in an identical manner.

For the non-net points, item 1 holds since we may assume that all interpoint distance are at most s , since the non-net points outside the convex hull of the net-points can all be projected onto the hull, and this can only improve the quality of the extension. Item 23 follow from the embedding of the $\frac{\varepsilon}{\text{ddim}(S)}$ -net: By the guarantees of the extension, an embedded non-net point may be at distance at most $O(\varepsilon)$ from its closest net-point, and then the items follow by an appropriate scaling down of ε . ◀

Similarly, the exact guarantees of Theorem 4 can be achieved by a non-oblivious embedding with dimension independent of n .

► **Corollary 6.** *Let $1 \leq q \leq p \leq 2$ and consider an n -point set $S \in l_p^m$. Set threshold $s > 1$. Fix range $R > 1$ and set threshold s as follows: When $q < p$, $s \approx R\varepsilon^{-1/2}(1 + (p - q)\varepsilon^{-(q/2+1)})^{1/(p-q)}$, and when $q = p$, $s \approx \max\{R^{1/\varepsilon}, R\varepsilon^{-1/2}e^{\varepsilon^{-(\frac{q}{2}+1)}\}$. Then there exists an embedding $f : S \rightarrow l_q^k$ satisfying items 134 of Theorem 4. The target dimension is $k = \tilde{O}\left(\frac{\text{ddim}^2(S)}{\varepsilon^3} \cdot s \cdot \min\left\{s'^{2q}, \max\left\{\frac{s'^{2q-p}}{2q-p}, \varepsilon s'^q\right\}\right\}\right)$ for $s' = \frac{s \text{ddim}(S)}{\varepsilon}$. Given a point hierarchy for S , the embedding can be constructed in $2^{\tilde{O}(\text{ddim}(S))} + O(mk)$ time per point.*

Comment

We conjecture that the $\frac{\text{ddim}^2(S)}{\varepsilon^3}$ terms can be reduced to $\frac{\text{ddim}(S)}{\varepsilon^2}$ by combining the randomness used separately for the construction of the padded decomposition, threshold embedding f , and the Lipschitz extension (as was done in [11] for l_2).

4.3 Snowflake embedding

The embedding of Lemma 5 implies a global dimension-reduction snowflake embedding for l_p . The proof uses the framework presented in [19], and appears in the full version of this paper.

► **Lemma 7.** *Let $0 < \varepsilon < 1/4$, $0 < \alpha < 1$, and $\tilde{\alpha} = \min\{\alpha, 1 - \alpha\}$. Every finite subset $S \subset l_p$ admits an embedding $\Phi : S \rightarrow l_q^k$ ($1 \leq q \leq p \leq 2$) for $k = \tilde{O}\left(\frac{\text{ddim}^6(S)}{\tilde{\alpha}^2 \varepsilon^8}\right)$.*

$\min \left\{ s'^{2q}, \max \left\{ \frac{s'^{2q-p}}{2q-p}, \varepsilon s'^q \right\} \right\}$ where $s' = \left(\frac{\text{ddim}(S)}{\varepsilon} \right)^4$ and $1 \leq \frac{\|\Phi(x) - \Phi(y)\|_q}{\|x - y\|_p} \leq 1 + \varepsilon$, for all $x, y \in S$. Given a point hierarchy for S , the embedding can be constructed in $2^{\tilde{O}(\text{ddim}(S))} + O(mk)$ time per point.

5 Clustering

Here we show that our snowflake embedding can be used to produce faster algorithms for the k -center and min-sum clustering problems. In both cases, we obtain improvements whenever $(\text{ddim}/\varepsilon)^{\Theta(1)}$ is smaller than the ambient dimension.

5.1 k -center clustering

In the k -center clustering problem, the goal is to partition the input set into k clusters, where the objective function to be minimized is the maximum radius among the clusters. Agarwal and Procopiuc [3] considered this problem for d -dimensional set $S \subset \ell_p$, and gave an exact algorithm that runs in time $n^{O(k^{1-1/d})}$, and used this to derive a $(1 + \varepsilon)$ -approximation algorithm that runs in $O(nd \log k) + (k\varepsilon^{-d})^{O(k^{1-1/d})}$. Here, the cluster centers are points of the ambient space \mathbb{R}^d in which S resides, chosen to minimize the maximum distance from points in S to their nearest center. The authors claim that the algorithm can be applied to the *discrete* problem as well, where all centers are chosen from S , and in fact the algorithm applies to the more general problem where the centers are chosen from a set S' satisfying $S \subset S' \subset \mathbb{R}^d$.⁴ Clearly, the runtime of the algorithm can be improved if the dimension is lowered.

► Theorem 8. *Given d -dimensional point set $S \subset \ell_p$ for constant p , $1 < p \leq 2$, a $(1 + \varepsilon)$ -approximate solution to the k -center problem on set S can be computed in time $O(nd(2^{\tilde{O}(\text{ddim}(S))} + \log k)) + (k \cdot \varepsilon^{-\text{ddim}(S)} \log(1/\varepsilon)/\varepsilon^2)^{k^{1-(\varepsilon/\text{ddim}(S))^{O(1)}}$. This holds for the discrete case as well.*

Proof. We first consider the discrete case. Let r^* be the optimal radius. As in [3], we run the algorithm of Feder and Greene [18] in time $O(n \log k)$ to obtain a value \tilde{r} satisfying $r^* \leq \tilde{r} < 2r^*$. We then build a hierarchy and extract a $\frac{\varepsilon}{2}\tilde{r}$ -net $V \subset S$ in time $2^{\tilde{O}(\text{ddim}(S))}nd$ (assuming the word-RAM model [12]). Since all points of S are contained in k balls of radius r , we have $|V| = k\varepsilon^{-O(\text{ddim}(S))}$. Further, a k -clustering for V is a $(1 + O(\varepsilon))$ -approximate k -clustering for S . We then apply the snowflake embedding of Lemma 7 to embed V into $(\text{ddim}(S)/\varepsilon)^{O(1)}$ -dimensional ℓ_p , and run the exact algorithm of [3] on V in the embedded space. Since a snowflake preserves the ordering of distances, the returned solution for the embedded space is a valid solution in the origin space as well.

Turning to the general (non-discrete) case, the above approach is problematic in that the embedding makes no guarantees on embedded points not in V . To address this, we construct a set W of candidate center points in the ambient space \mathbb{R}^d : Recall that the problem of finding the minimum enclosing ℓ_p -ball admits a core-set of size $O(\varepsilon^{-2})$ [9]. (That is, for any discrete point set there exists a subset of size $O(\varepsilon^{-2})$ with the property that the center of the subset is also the center of a $(1 + \varepsilon)$ -approximation to the smallest ball covering the original set.) We take all distinct subsets $C \subset V$ of size $|C| = O(\varepsilon^{-2})$ and radius at most \tilde{r} , compute

⁴ The Euclidean core-set algorithm of [9] runs in time $k^{O(k/\varepsilon^2)} \cdot nd$, and can readily be seen to apply to all ℓ_p for constant p , $1 < p \leq 2$ (see [41] for a simple approach). The algorithm of [3] compares favorably to the core-set algorithm when d is small.

the center point of each subset (see [41]), and add its candidate center to W . It follows that $|W| = k\varepsilon^{-O(\text{ddim}(S)/\varepsilon^2)}$ and $\text{ddim}(W) = \log \varepsilon^{-O(\text{ddim}(S)/\varepsilon^2)} = O(\text{ddim}(S) \log(1/\varepsilon)/\varepsilon^2)$. As above, we use the snowflake embedding of Lemma 7 to embed $V \cup W$ into $(\text{ddim}(S)/\varepsilon)^{O(1)}$ -dimensional ℓ_p and run the exact algorithm of [3] on the embedded space, covering the points of V with candidate centers from W . The returned solution is a valid solution in the origin space as well. \blacktriangleleft

5.2 Min-sum clustering

In the min-sum clustering problem, the goal is to partition the points of an input set into k clusters, where the objective function is the sum of distances between each intra-cluster point pair. Schulman [44] designed algorithms for min-sum clustering under ℓ_1, ℓ_2 and ℓ_2 -squared costs, and their runtimes depend exponentially on the dimension. We will obtain faster runtimes for min-sum clustering for ℓ_p ($1 \leq p \leq 2$) by using our snowflake embedding as a preprocessing step to reduce dimension. Ultimately, we will embed the space into ℓ_2 using a $\frac{1}{2}$ -snowflake, and then solve min-sum clustering with ℓ_2 -squared costs in the embedded space; this is equivalent to solving the original ℓ_p problem. We will prove the case of ℓ_1 , and the other cases are simpler.

We are given an input set $S \in \ell_1$, and set $c = 2 - \frac{1}{d'}$ for some value $d' = \left(\frac{\text{ddim}(S)}{\varepsilon}\right)^{O(1)}$. We note that the $\frac{1}{c}$ -snowflake of ℓ_1 is itself in ℓ_c [17], and also that this embedding into ℓ_c can be computed in polynomial time (with arbitrarily small distortion) using semi-definite programming [19], although the target dimension may be large. We compute this embedding, and then reduce dimension by invoking our snowflake embedding (Lemma 7) to compute a $(\frac{c}{2} = 1 - \frac{1}{2d'})$ -snowflake in ℓ_c , with dimension d' . We then consider the vectors to be in ℓ_2 instead of ℓ_c , which induces a distortion of $1 + O(\varepsilon)$. Finally, we run the algorithms of Schulman on the final Euclidean space. As we have replaced the original ℓ_1 distances with their $(\frac{1}{c} \cdot \frac{c}{2} = \frac{1}{2})$ -snowflake and embedded into ℓ_2 , solving min-sum clustering with ℓ_2 -squared costs on the embedded space solves the original ℓ_1 problem with distortion $1 + O(\varepsilon)$.

The following lemma follows from the embedding detailed above, in conjunction with [44, Propositions 14,28, full version]. For ease of presentation, we will assume that $k = O(1)$.

► **Lemma 9.** *Given a set of n points $S \in \mathbb{R}^d$, set $d' = (\text{ddim}(S)/\varepsilon)^{O(1)}$. a $(1 + O(\varepsilon))$ -approximation to the ℓ_p min-sum k -clustering for S , for $k = O(1)$, can be computed*

- in deterministic time $n^{O(d')} 2^{2^{O(d')}}$.
- in randomized time $n^{O(1)} + 2^{2^{O(d')}} (\varepsilon \log n / \delta)^{O(d')}$, with probability $1 - \delta$.

Acknowledgements. We thank Piotr Indyk, Robi Krauthgamer and Assaf Naor for helpful conversations.

References

- 1 I. Abraham, Y. Bartal, and O. Neiman. Embedding metric spaces in their intrinsic dimension. In *SODA'08*, pages 363–372. SIAM, 2008.
- 2 D. Achlioptas. Database-friendly random projections. In *PODS'01*, pages 274–281, New York, NY, USA, 2001. ACM.
- 3 P. K. Agarwal and C. M. Procopiuc. Exact and approximation algorithms for clustering. *Algorithmica*, 33(2):201–226, 2002.
- 4 N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *STOC'06*, pages 557–563. ACM, 2006.

- 5 N. Alon. Problems and results in extremal combinatorics, part I. *Disc. Math.*, 273, 2003.
- 6 N. Alon, M. Bădoiu, E.D. Demaine, M. Farach-Colton, M. Hajiaghayi, and A. Sidiropoulos. Ordinal embeddings of minimum relaxation: General properties, trees, and ultrametrics. *ACM Transactions on Algorithms*, 4(4), 2008.
- 7 A. Andoni, M. Charikar, O. Neiman, and H.L. Nguyen. Near linear lower bounds for dimension reduction in ℓ_1 . In *FOCS'11*. IEEE Computer Society, 2011.
- 8 Alexandr Andoni. *Nearest Neighbor Search: the Old, the New, and the Impossible*. PhD thesis, MIT, 2009.
- 9 M. Bădoiu, S. Har-Peled, and P. Indyk. Approximate clustering via core-sets. In *STOC'02*, pages 250–257, New York, NY, USA, 2002. ACM.
- 10 Y. Bartal, L. Gottlieb, T. Kopelowitz, M. Lewenstein, and L. Roditty. Fast, precise and dynamic distance queries. In *SODA'11*, pages 840–853. SIAM, 2011.
- 11 Y. Bartal, B. Recht, and L. Schulman. Dimensionality reduction: beyond the Johnson-Lindenstrauss bound. In *SODA'11*, pages 868–887. SIAM, 2011.
- 12 Yair Bartal and Lee-Ad Gottlieb. A linear time approximation scheme for euclidean tsp. In *FOCS'13*, pages 698–706, Washington, DC, USA, 2013. IEEE Computer Society.
- 13 T. Batu, F. Ergun, and C. Sahinalp. Oblivious string embeddings and edit distance approximations. In *SODA'06*, pages 792–801, 2006.
- 14 M. Bădoiu, E. Demaine, M. Hajiaghayi, A. Sidiropoulos, and M. Zadimoghaddam. Ordinal embedding: Approximation algorithms and dimensionality reduction. In *APPROX'08*, pages 21–34. 2008.
- 15 H. Chan, A. Gupta, and K. Talwar. Ultra-low-dimensional embeddings for doubling metrics. *J. ACM*, 57(4):1–26, 2010. doi:10.1145/1734213.1734215.
- 16 R. Cole and L. Gottlieb. Searching dynamic point sets in spaces with bounded doubling dimension. In *38th annual ACM symposium on Theory of computing*, pages 574–583, 2006.
- 17 M. M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Springer-Verlag, Berlin, 1997.
- 18 Tomás Feder and Daniel Greene. Optimal algorithms for approximate clustering. In *STOC'88*, pages 434–444, New York, NY, USA, 1988. ACM.
- 19 L. Gottlieb and R. Krauthgamer. A nonlinear approach to dimension reduction. In *SODA'11*, pages 888–899. SIAM, 2011.
- 20 A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *FOCS'03*, pages 534–543, 2003.
- 21 S. Har-Peled, P. Indyk, and R. Motwani. Approximate nearest neighbor: Towards removing the curse of dimensionality. *Theory of Computing*, 8(14):321–350, 2012.
- 22 S. Har-Peled and M. Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM Journal on Computing*, 35(5):1148–1184, 2006.
- 23 P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *FOCS'01*, pages 10–33, 2001.
- 24 P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006.
- 25 P. Indyk and A. Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):31, 2007.
- 26 W.B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conf. in modern analysis and probability*, pages 189–206. AMS, 1984.
- 27 W.B. Johnson and A. Naor. The Johnson-Lindenstrauss lemma almost characterizes Hilbert space, but not quite. In *SODA'09*, pages 885–891. SIAM, 2009.
- 28 W.B. Johnson and G. Schechtman. Embedding l_p^m into l_1^m . *Acta Math.*, 149(1-2):71–85, 1982.
- 29 R. Krauthgamer and J.R. Lee. Navigating nets: Simple algorithms for proximity search. In *SODA'04*, pages 791–801, January 2004.

- 30 E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *STOC'98*, pages 614–623. ACM, 1998.
- 31 J.R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Inventiones Mathematicae*, 160:59–95, 2005.
- 32 Ping Li. Estimators and tail bounds for dimension reduction in ℓ_α ($0 < \alpha \leq 2$) using stable random projections. In *SODA*, pages 10–19, 2008.
- 33 G. Lugosi. Concentration-of-measure inequalities. Manuscript, available at <http://www.econ.upf.edu/~lugosi/anu.ps>, 2004.
- 34 J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.
- 35 M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Advances in Mathematics*, 189(2):451–494, 2004.
- 36 V. D. Milman and G. Schechtman. *Asymptotic Theory of Finite-dimensional Normed Spaces*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.
- 37 H. Nguyen. Approximate nearest neighbor search in ℓ_p . Manuscript, available at <http://arxiv.org/pdf/1306.3601v1>, 2013.
- 38 J.P. Nolan. *Stable Distributions – Models for Heavy Tailed Data*. Birkhauser, Boston, 2012. In progress, Chapter 1 online at academic2.american.edu/~jpnolan.
- 39 R. Ostrovsky and Y. Rabani. Polynomial time approximation schemes for geometric k-clustering. In *FOCS'00*, pages 349–358. IEEE Computer Society, 2000.
- 40 R. Ostrovsky and Y. Rabani. Polynomial-time approximation schemes for geometric min-sum median clustering. *J. ACM*, 49(2):139–156, 2002.
- 41 R. Panigrahy. Minimum enclosing polytope in high dimensions. *CoRR*, 2004.
- 42 A. Rahimi and B. Recht. Random features for large-scale kernel machines. In *NIPS*, 2007.
- 43 G. Schechtman. Dimension reduction in ℓ_p , $0 < p < 2$. Manuscript, available at <http://arxiv.org/pdf/1110.2148v1.pdf>, 2011.
- 44 L. J. Schulman. Clustering for edge-cost minimization (extended abstract). In *STOC*, pages 547–555. ACM, 2000. Full version available as ECCO report TR99-035. doi:10.1145/335305.335373.
- 45 M. Talagrand. Embedding subspaces of L_1 into l_1^N . *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.
- 46 M. Talagrand. Embedding subspaces of L_p in l_p^N . In *Geometric aspects of functional analysis*, volume 77 of *Oper. Theory Adv. Appl.*, pages 311–325. Birkhäuser, Basel, 1995.
- 47 V.M. Zolotarev. *One-dimensional Stable Distributions*. AMS, 1986.

A Probability theory

The following is a simplified version of Hoeffding's inequality [33]:

► **Claim 10.** Let X_1, \dots, X_k be independent real-valued random variables, and assume $|X_i| \leq s$ with probability one. Let $\bar{X} = \frac{1}{k} \sum_{i=1}^k X_i$. Then for any $z > 0$, $\Pr [|\bar{X} - \mathbb{E}[\bar{X}]| \geq z] \leq 2 \exp\left(-\frac{2kz^2}{s^2}\right)$

The following is a restatement of Bennett's inequality [33]:

► **Claim 11.** Let X_1, \dots, X_k be independent real-valued random variables, and assume $|X_i| \leq s$ with probability one. Let $\bar{X} = \frac{1}{k} \sum_{i=1}^k X_i$ and set $\sigma^2 = \frac{1}{k} \sum_{i=1}^k \text{Var}\{X_i\}$. Then for any $z > 0$, $\Pr [|\bar{X} - \mathbb{E}[\bar{X}]| \geq z] \leq 2 \exp\left(-\frac{k\sigma^2}{s^2} V\left(\frac{sz}{\sigma^2}\right)\right)$ where $V(u) = (1+u) \ln(1+u) - u$ for $u \geq 0$.

Note that for $u \geq 1$, we have $V(u) = \Omega(u \log u)$, while for $0 \leq u < 1$, $V(u) = \Theta(u^2)$.