

A Nonlinear Approach to Dimension Reduction

Lee-Ad Gottlieb* Robert Krauthgamer*
Weizmann Institute of Science

Abstract

The ℓ_2 flattening lemma of Johnson and Lindenstrauss [JL84] is a powerful tool for dimension reduction. It has been conjectured that the target dimension bounds can be refined and bounded in terms of the intrinsic dimensionality of the data set (for example, the doubling dimension). One such problem was proposed by Lang and Plaut [LP01] (see also [GKL03, Mat02, ABN08, CGT10]), and is still open. We prove another result in this line of work:

The snowflake metric $d^{1/2}$ of a doubling set $S \subset \ell_2$ can be embedded with arbitrarily low distortion into ℓ_2^D , for dimension D that depends solely on the doubling constant of the metric.

In fact, the target dimension is polylogarithmic in the doubling constant. Our techniques are robust and extend to the more difficult spaces ℓ_1 and ℓ_∞ , although the dimension bounds here are quantitatively inferior than those for ℓ_2 .

1 Introduction

Dimension reduction, in which high-dimensional data is faithfully represented in a low-dimensional space, is a key tool in several fields. Probably the most prevalent mathematical formulation of this problem considers the data to be a set $S \subset \ell_2$, and the goal is to map the points in S into a low-dimensional ℓ_2^k . (Here and throughout, ℓ_p^k denotes the space \mathbb{R}^k endowed with the ℓ_p -norm; ℓ_p is the infinite-dimensional counterpart of all sequences that are p -th power summable.) A celebrated result in this area is the so-called JL-Lemma [JL84]:

THEOREM 1.1. (JOHNSON AND LINDENSTRAUSS) *For every n -point subset $S \subset \ell_2$ and every $0 < \varepsilon < 1$, there is a mapping $\Psi_{\text{JL}} : S \rightarrow \ell_2^k$ that preserves all interpoint distances in S within factor $1 + \varepsilon$, and has target dimension $k = O(\varepsilon^{-2} \log n)$.*

*This work was supported in part by The Israel Science Foundation (grant #452/08), and by a Minerva grant. Weizmann Institute of Science, Rehovot, Israel. Email: {lee-ad.gottlieb, robert.krauthgamer}@weizmann.ac.il

This positive result is remarkably strong; in fact the map Ψ_{JL} is an easy to describe linear transformation. It has found many applications, and has become a basic tool. It is natural to seek the optimal (minimum) target dimension k possible in this theorem. The logarithmic dependence on $n = |S|$ is necessary, as can be easily seen by volume arguments, and Alon [Alo03] further proved that the JL-Lemma is optimal up to a factor of $O(\log \frac{1}{\varepsilon})$. These lower bounds are existential, meaning that there are sets S for which the result of the JL-Lemma cannot be improved. However, it may still be possible to significantly reduce the dimension for sets S that are “intrinsically” low-dimensional. This raises the interesting and fundamental question of bounding k in terms of parameters other than n , which we formalize next.

We recall some basic terminology involving metric spaces. The *doubling constant* of a metric (M, d_M) , denoted $\lambda(M)$, is the smallest $\lambda \geq 1$ such that every (metric) ball in M can be covered by at most λ balls of half its radius. We say that M is *doubling* if its doubling constant $\lambda(M)$ is bounded independently of $|M|$. It is sometimes more convenient to refer to $\dim(M) \stackrel{\text{def}}{=} \log_2 \lambda(M)$, which is known as the *doubling dimension* of M [GKL03]. An *embedding* of one metric space (M, d_M) into another (N, d_N) is a map $\Psi : M \rightarrow N$. We say that Ψ attains *distortion* $D' \geq 1$ if Ψ preserves every pairwise distance within factor D' , namely, there is a scaling factor $s > 0$ such that

$$1 \leq \frac{d_N(\Psi(x), \Psi(y))}{s \cdot d_M(x, y)} \leq D', \quad \forall x, y \in M.$$

The following problem was posed independently by [LP01] and [GKL03] (see also [Mat02, ABN08, CGT10]):

QUESTION 1. *Does every doubling subset $S \subset \ell_2$ embed with distortion D' into ℓ_2^D for D, D' that depend only on $\lambda(S)$?*

This question is still open and seems very challenging. Resolving it in the affirmative seems to require completely different techniques than the JL-Lemma, since such an embedding cannot be achieved by a linear

map [IN07, Remark 4.1]. For algorithmic applications, it would be ideal to resolve positively an even stronger variant of Question 1, where the target distortion D' is an absolute constant independent of $\lambda(S)$, or even $1 + \varepsilon$ as in the JL-Lemma. This stronger version has not been excluded, and is still open as well.

1.1 Results and Techniques We present dimension reduction results for doubling subsets in Euclidean spaces. In fact, we devise a robust framework that extends even to the spaces ℓ_1 and ℓ_∞ . Our results incur constant or $1 + \varepsilon$ distortion, with target dimension that depends not on $|S|$ but rather on $\dim(S)$ (and this dependence is unavoidable due to volume arguments). We remark that such guarantees – very low distortion and dimension – are highly sought-after in metric embeddings, but rarely achieved. We state our results in the context of finite metrics (subsets of ℓ_p); they extend to infinite subsets of L_p via standard arguments.

Snowflake Embedding. Our primary embedding achieves distortion $1 + \varepsilon$ for the *snowflake metric* d^α of an input metric d (i.e. the snowflake metric is obtained by raising every pairwise distance to power $0 < \alpha < 1$). It is instructive to view α as a fixed constant, say $\alpha = 1/2$. We prove the following in Section 3.3. Throughout, we use $\tilde{O}(f)$ to denote $f \cdot (\log f)^{O(1)}$.

THEOREM 1.2. *Let $0 < \varepsilon < 1/4$, $0 < \alpha < 1$, and $\tilde{\alpha} = \min\{\alpha, 1 - \alpha\}$. Every finite subset $S \subset \ell_2$ admits an embedding $\Phi : S \rightarrow \ell_2^k$ for $k = \tilde{O}(\varepsilon^{-4} \tilde{\alpha}^{-1} (\dim S)^2)$, such that*

$$1 \leq \frac{\|\Phi(x) - \Phi(y)\|_2}{\|x - y\|_2^\alpha} \leq 1 + \varepsilon, \quad \forall x, y \in S.$$

Notice the difference between our theorem and Question 1: Our embedding achieves better distortion $1 + \varepsilon$, but it applies to the (often easier) snowflake metric d^α . Our result is also related to the following theorem of Assouad [Ass83]: *For every doubling metric (M, d) and every $0 < \alpha < 1$, the snowflake metric d^α embeds into ℓ_2^D with distortion D' , where D, D' depend only on $\lambda(M)$ and α .* Note the theorem’s vast generality – the only requirements are the doubling property (which by volume arguments is an obvious necessity) and that the data be a metric – at the nontrivial price that the distortion achieved depends on $\lambda(M)$. Compared to Assouad’s theorem, our embedding achieves a much stronger distortion $1 + \varepsilon$, but requires the additional assumption that the input metric is Euclidean.

Previously, Theorem 1.2 was only known to hold in the special case where $S = \mathbb{R}$ (the real line). For this case, Kahane [Kah81] and Talagrand [Tal92] exhibit a $1 + \varepsilon$ distortion embedding of the snowflake

metric $|x - y|^\alpha$ into ℓ_2^k . Kahane’s [Kah81] shows an embedding of $|x - y|^{1/2}$ (also known as Wilson’s helix) into dimension $k = O(1/\varepsilon)$, while Talagrand [Tal92] shows how to embed every snowflake metric $|x - y|^\alpha$, $\alpha \in (0, 1)$, with dimension $k = O(K(\alpha)/\varepsilon^2)$ (which is larger). Thus, our theorem can be viewed as a generalization of [Kah81, Tal92] to arbitrary doubling subset of ℓ_2 (or other ℓ_p), albeit with a somewhat worse dependence on ε .

Embedding for a Single Scale. Most of our technical work is devoted to designing an embedding that preserves distances at a single scale $r > 0$, while still maintaining a one-sided Lipschitz condition for all scales. We now state our most basic new result, which achieves only a constant distortion (for the desired scale).

THEOREM 1.3. *For every scale $r > 0$ and every $0 < \delta < 1/4$, every finite set $S \subset \ell_2$ admits an embedding $\varphi : S \rightarrow \ell_2^k$ for $k = \tilde{O}(\log \frac{1}{\delta} \cdot (\dim S)^2)$, satisfying:*

- (a). *Lipschitz:* $\|\varphi(x) - \varphi(y)\|_2 \leq \|x - y\|_2$ for all $x, y \in S$;
- (b). *Bi-Lipschitz at scale r :* $\|\varphi(x) - \varphi(y)\|_2 = \Omega(\|x - y\|_2)$ whenever $\|x - y\|_2 \in [\delta r, r]$; and
- (c). *Boundedness:* $\|\varphi(x)\|_2 \leq r$ for all $x \in S$.

The constant factor accuracy achieved by this theorem is too weak to achieve the $1 + \varepsilon$ distortion asserted in Theorem 1.2. While we cannot improve condition (b) to a factor of $1 + \varepsilon$, we are able to refine it in a useful way. Roughly speaking, we introduce a “correction” function $\tilde{G} : \mathbb{R} \rightarrow \mathbb{R}$, such that whenever $\|x - y\|_2 \in [\delta r, r]$,

$$\frac{\|\varphi(x) - \varphi(y)\|_2}{\|x - y\|_2} = (1 \pm \varepsilon) \tilde{G}\left(\frac{\|x - y\|_2}{r}\right). \quad (1.1)$$

This function \tilde{G} does not depend on r and equals $\Theta(1)$ in the appropriate range. Using the correction function, we obtain very accurate bounds on distances in the target space, at the price of increasing the dimension by a factor of $\tilde{O}(1/\varepsilon^3)$. This high-level idea is implemented in Theorem 3.1, which immediately implies Theorem 1.3, although the precise bound therein slightly differs from Equation (1.1).

Embeddings for a single scale are commonly used in the embeddings literature, though not in the context of dimension reduction. It is plausible that in some applications, a single-scale embedding may suffice, or even provide better bounds than our snowflake embedding (or Question 1).

Other ℓ_p Spaces. Our dimension reduction framework extends to ℓ_p (i.e. $S \subset \ell_p$ and $\Phi : S \rightarrow \ell_p^k$) for both $p = 1$ and $p = \infty$, as discussed in Section

4. The bounds we obtain therein are worse than in the ℓ_2 case, namely the dimension k is at least exponential in $\dim(S)$, which is to be expected because of strong lower bounds known in terms of $n = |S|$ (see [BC05, LN04] for ℓ_1 and [Mat96] for ℓ_∞). We remark that previous work on dimension reduction in ℓ_p spaces [JL84, Sch87, Tal90, Bal90, Tal95, Mat96] did not establish any dimension bound in terms of $\lambda(S)$; these bounds are all expressed in terms of $n = |S|$, or of the dimension of S as a linear subspace.

For ultrametrics, our framework provides even stronger bounds, which resolve Question 1 in the affirmative, as follows. Ultrametrics embed isometrically (i.e. with distortion 1) into ℓ_2 , hence Theorem 1.2 is immediately applicable. We can then eliminate the snowflake operator (i.e. achieve $\alpha = 1$) by the observation that (M, d) is an ultrametric if and only if (M, d^2) is an ultrametric, and thus Theorem 1.2 is applicable to the ultrametric d^2 with $\alpha = 1/2$. Moreover, the dimension bound can be improved by replacing some steps with more specialized machinery. However, in retrospect a near-optimal bound may be obtained by minor refinements of [ABN09, Lemma 12].

Technical contribution. The main technical challenge is to keep both distortion and dimension under tight control. We use a relatively large number of the tools developed recently in the metric embeddings literature, combining them in a technically non-trivial manner that yields a rather strong outcome ($1 + \varepsilon$ distortion). Several of the tools we use are nonlinear, hence our approach can potentially be used to circumvent the limitation on linear embeddings observed by [IN07].

Our results may also be viewed as partial progress towards Question 1: Observe that Theorem 1.2 answers that question positively in the special case where also the square of the given Euclidean metric is known to be Euclidean (e.g. for all ultrametrics). Further, Theorem 1.3 achieves bounds that relax those required by Question 1. Moreover, if the answer to Question 1 is negative (which is not unlikely), then our results may be essentially the closest alternative.

1.2 Related work A summary of some related work on embeddings, meant to put our results in context, is found in Table 1. Very recently, and subsequent to the public posting of this paper, the authors of [BRS11] concluded that an extension of their work, coupled with the framework presented here, yields a stronger version of Theorem 1.2, where the target dimension is improved to $\tilde{O}(\varepsilon^{-3}\tilde{\alpha}^{-1} \dim S)$. This additional result has been appended to the most recent version of [BRS11].

1.3 Applications In many settings, data is provided as points in ℓ_p , and it is extremely advantageous to represent the data using a low-dimensional space. For instance, the cost of many data processing tasks (in terms of runtime, storage or accuracy) grows exponentially in the embedding dimension. In many such cases, our machinery can reduce the embedding dimension to close to the data’s doubling dimension, leading to significant performance improvement. This approach suitable for problems (i) that depend on pairwise distances but can tolerate small distortion; and (ii) whose algorithms depend heavily on the embedding dimension, so that the improved performance given by the lower dimension outweighs the overhead cost of computing the dimensionality reduction.

We provide in Section 5 two examples where our dimensionality reduction results have immediate algorithmic applications. The first one is an approximate Distance Labeling Scheme, where the main complexity measure is the storage required at each network node. The second example is approximation algorithms for clustering algorithms, where running time is typically exponential in the dimension. In both cases, the final approximation obtained is $1 + \varepsilon$.

On a more conceptual level, our embeddings may explain a common empirical phenomenon regarding low-dimensional data: Many heuristics that represent (non-Euclidean) input data as points in Euclidean space find that low-dimensional Euclidean space is sufficient to yield a fair representation, see e.g. [NZ02] for networking and [TdSL00, RS00] for machine learning. Our results can be interpreted as conveying the following principle: Intrinsically low-dimensional data that admits a meaningful representation in ℓ_2 , can actually be represented in *low-dimensional* ℓ_2 .

Implementation. All our embedding results are algorithmic — they are constructive and can be computed in polynomial time. The details are mostly straightforward, and we do not address this issue explicitly. It is possible that the running time may be improved further, and perhaps even be brought close to linear. (For example, the Gaussian transform is computed quickly via the Gram matrix.) Two nontrivial steps in this direction are the implementation of Kirschbraun’s Theorem, which is usually solved as a semidefinite program, and our use of padded partitions, which require an application of the Lovász Local Lemma.

2 Preliminaries and tools

Doubling dimension. For a metric (X, d) , let λ be the infimum value such that every ball in X can be covered by λ balls of half the radius. The *doubling dimension* of X is $\dim(X) = \log_2 \lambda$. A metric is *doubling* when

Reference	Origin	Target space	Distortion	Dimension	Snowflake α
[JL84]	ℓ_2	ℓ_2	$1 + \varepsilon$	$O(\varepsilon^{-2} \log n)$	none ($\alpha = 1$)
[Ass83]	doubling	ℓ_2	$2^{O(\dim S)}$	$2^{O(\dim S)}$	fixed $\alpha < 1$
[GKL03]	doubling	ℓ_2	$\tilde{O}(\dim S)$	$\tilde{O}(\dim S)$	fixed $\alpha < 1$
[HM06]	doubling	ℓ_∞	$1 + \varepsilon$	$\varepsilon^{-O(\dim S)}$	fixed $\alpha < 1$
[ABN08]	doubling	$\ell_p, p \geq 1$	$O(\log^{1+\varepsilon} n)$	$O(\varepsilon^{-1} \dim S)$	none
Theorem 1.2	ℓ_2	ℓ_2	$1 + \varepsilon$	$\tilde{O}(\varepsilon^{-4} \dim^2 S)$	fixed $\alpha < 1$

Table 1: A sampling of related work; holds for arbitrary $\varepsilon \in (0, 1)$

its doubling dimension is finite. The following property can be demonstrated via a repetitive application of the doubling property.

PROPERTY 2.1. *For set S with doubling dimension $\log \lambda$, if the minimum interpoint distance in S is at least α , and the diameter of S is at most β , then $|S| \leq \lambda^{O(\log(\beta/\alpha))}$.*

ϵ -nets. For a point set S , an ϵ -net of S is a subset $T \subset S$ with the following properties: (i) Packing: For every pair $u, v \in T$, $d(u, v) \geq \epsilon$. (ii) Covering: Every point $u \in S$ is strictly within distance ϵ of some point $v \in T$: $d(y, x) < \epsilon$.

Lipschitz norm. Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \rightarrow Y$ is said to be K -Lipschitz (for $K > 0$) if for all $x, x' \in X$ we have $d_Y(f(x), f(x')) \leq K \cdot d_X(x, x')$. The Lipschitz constant (or Lipschitz norm) of f , denoted $\|f\|_{\text{Lip}}$, is the infimum over $K > 0$ satisfying the above. A 1-Lipschitz function is called in short Lipschitz. We recall the following basic property of Lipschitz functions: Let $f : X \rightarrow \ell_2^k$ and $g : X \rightarrow \mathbb{R}$. Then their product $fg : x \rightarrow g(x) \cdot f(x)$ has Lipschitz norm

$$\|fg\|_{\text{Lip}} \leq \|f\|_{\text{Lip}} \cdot \max_x |g(x)| + \|g\|_{\text{Lip}} \cdot \max_x \|f(x)\|.$$

Extension Theorem. The Kirszbraun Theorem [Kir34] states that if S and X are Euclidean spaces, $T \subset S$, and there exists a Lipschitz function $f : T \rightarrow X$; then there exists a function $\tilde{f} : S \rightarrow X$ that has the same Lipschitz constant as f and also *extends* f , i.e. $\tilde{f}|_T = f$, meaning that the restriction of \tilde{f} to T is identical to f .

Bounded distances and the Gaussian Transform. A *metric transform* maps a distance function to another distance function on the same set of points (e.g. maps (X, d) to $(X, d^{1/2})$). We say that a metric transform is *bounded* (by $T > 0$) if it always results with a distance function where all interpoint distances are bounded (by $T > 0$). The *Gaussian transform* is a metric transform that maps value t to $G_r(t) = r(1 - e^{-t^2/r^2})^{1/2}$, where $r > 0$ is a parameter. Schoenberg [Sch38, DL97] showed that the Gaussian transform maps Euclidean spaces to

Euclidean spaces. That is, for every $r > 0$ and $X \subset L_2$ there is an embedding $g : X \rightarrow L_2$ such that for all $x, y \in X$ we have $\|g(x) - g(y)\|_2 = G_r(\|x - y\|_2)$. It is easily verified that

$$G_r(t) \leq t, \quad \forall t \geq 0, \quad (2.2)$$

thus $\|g\|_{\text{Lip}} \leq 1$. In addition, $G_r(t) \leq r$ for all t , hence the Gaussian transform is bounded. The idea of using bounded transforms for embeddings is due to [BRS11].

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 [GKL03], Abraham, Bartal and Neiman [ABN08], and Chan, Gupta and Talwar [CGT10]. Let $B(x, r) = \{y \mid \|x - y\| \leq r\}$.

THEOREM 2.1. (PADDED DECOMPOSITION) *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 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*

1. *Bounded radius: $\text{diam}(C) \leq \Delta$ for all clusters $C \in \bigcup_{i=1}^m P_i$.*
2. *Padding: If P is chosen uniformly from \mathcal{D} , then for all $x \in X$,*

$$\Pr_{P \in \mathcal{D}} [B(x, \frac{\Delta}{c_0 \dim(X)}) \subseteq P(x)] \geq 1 - \varepsilon.$$

Remark: [GKL03] provided slightly different quantitative bounds than in Theorem 2.1. The two enumerated properties follow, for example, from Lemma 2.7 in [ABN08], and the bound on support-size m follows by an application of the Lovász Local Lemma sketched therein.

3 Dimension Reduction for ℓ_2

In this section we first design a single scale embedding that achieves distortion $1 + \varepsilon$ after including a correction

function. This result is stated in Theorem 3.1 below, which is a refined version of Theorem 1.3. We then use this single scale embedding to prove Theorem 1.2 in Section 3.3. Throughout this section, the norm notation $\|\cdot\|$ denotes ℓ_2 -norms. We make no attempt to optimize constants. Following Section 2, define $G : \mathbb{R} \rightarrow \mathbb{R}$ by $G(t) = (1 - e^{-t^2})^{1/2}$, and let

$$G_r(t) = r \cdot G(t/r) = r(1 - e^{-t^2/r^2})^{1/2}.$$

THEOREM 3.1. *For every scale $r > 0$ and every $0 < \delta, \varepsilon < 1/4$, every finite set $S \subset \ell_2$ admits an embedding $\varphi : S \rightarrow \ell_2^k$ for $k = \tilde{O}(\varepsilon^{-3} \log \frac{1}{\delta} \cdot (\dim S)^2)$, satisfying:*

- (a). *Lipschitz:* $\|\varphi(x) - \varphi(y)\| \leq \|x - y\|$ for all $x, y \in S$.
(b). $1 + \varepsilon$ *distortion to the Gaussian (at scales near r):*
For all $x, y \in S$ with $\delta r \leq \|x - y\| \leq \frac{r}{\delta}$,

$$\frac{1}{1 + \varepsilon} \leq \frac{\|\varphi(x) - \varphi(y)\|}{G_r(\|x - y\|)} \leq 1;$$

- (c). *Boundedness:* $\|\varphi(x)\| \leq r$ for all $x \in S$.

In the sequel, we shall prove bounds that are slightly weaker than those stated above but only by a constant $C > 1$, e.g. $\|\varphi\|_{\text{Lip}} \leq 1 + C\varepsilon$. The actual theorem follows immediately from these bounds by scaling of φ by $\frac{1}{1+C\varepsilon}$ and scaling of ε by $1/C$.

3.1 Embedding for a single scale Our construction of the embedding φ for Theorem 3.1 proceeds in seven steps, as described below. Let $\lambda = \lambda(S)$. All the hidden constants are absolute, i.e. independent of $\lambda, \varepsilon, \delta$ and r . It is plausible that the dependence of target dimension on $\log \lambda$ can be improved to be near-linear, by carefully combining some of these steps.

Step 1 (Net Extraction): Let $N \subseteq S$ be an $(\varepsilon\delta r)$ -net in S .

Step 2 (Padded Decomposition): Compute for N a padded decomposition with padding $\frac{3r}{\delta}$. More specifically, by Theorem 2.1, there is a multiset $[P_1, \dots, P_m]$ of partitions of N , where every point is $\frac{3r}{\delta}$ -padded in $1 - \varepsilon$ fraction of the partitions, all clusters have diameter bounded by $\Delta = O(\frac{r}{\delta} \log \lambda)$, and $m = O(\varepsilon^{-1} \log \lambda \log \log \lambda)$.

Step 3 (Bounding Distances): In each partition P_i and each cluster $C \in P_i$, bound the interpoint distances in C at maximum value r . Specifically, using a Gaussian transform as per Section 2, obtain a map $g_C : C \rightarrow \ell_2$ such that

$$\|g_C(x) - g_C(y)\|_2^2 = G_r(\|x - y\|)^2 = r^2(1 - e^{-\|x - y\|_2^2/r^2}), \quad \forall x, y \in C.$$

Step 4 (Dimension Reduction): For each partition P_i and each cluster $C \in P_i$, the point set $g_C(C) \subset \ell_2$ admits a dimension reduction, with distortion $1 + \varepsilon$. Specifically, by the JL-Lemma there is a map $\Psi_{\text{JL}} : g_C(C) \rightarrow \ell_2^{k'}$ such that

$$\frac{\|t - t'\|}{1 + \varepsilon} \leq \|\Psi_{\text{JL}}(t) - \Psi_{\text{JL}}(t')\| \leq \|t - t'\|, \quad \forall t, t' \in g_C(C), \quad (3.3)$$

and the target dimension is (using Property 2.1)

$$\begin{aligned} k' &= O(\varepsilon^{-2} \log |C|) \\ &= O(\varepsilon^{-2} \log(\lambda^{O(\log(\Delta/\varepsilon\delta r))})) \\ &= O(\varepsilon^{-2} \log \frac{1}{\varepsilon\delta} \cdot \log \lambda \log \log \lambda). \end{aligned}$$

Composing the last two steps, define $f_C = \Psi_{\text{JL}} \circ g_C$ mapping $C \rightarrow \ell_2^{k'}$.

Step 5 (Gluing Clusters): For each partition P_i , “glue” the cluster embeddings f_C by smoothing them near the boundary. Specifically, for each cluster $C \in P_i$, assume by translation that f_C attains the origin, i.e. there exists $z_C \in C$ such that $\|f_C(z_C)\| = 0$. Define $h_C : C \rightarrow \mathbb{R}$ by $h_C(x) = \min_{y \in N \setminus C} \|x - y\|$, as a proxy for x 's distance to the boundary of its cluster. Now define $\varphi_i : N \rightarrow \ell_2^{k'}$ by

$$\varphi_i(x) = f_{P_i(x)}(x) \cdot \min\{1, \frac{\delta}{r} h_{P_i(x)}(x)\};$$

recall that $P_i(x)$ is the unique cluster $C \in P_i$ containing x .

Step 6 (Gluing Partitions): Combine the maps obtained in the previous step via direct sum and scaling. Specifically, define $\varphi : N \rightarrow \ell_2^{mk'}$ by $\varphi = m^{-1/2} \bigoplus_{i=1}^m \varphi_i$.

Step 7 (Extension beyond the Net): Use the Kirszbraun theorem to extend the map φ to all of S , without increasing the Lipschitz constant.

3.2 Proof of Theorem 3.1 Let us show the embedding φ constructed above indeed satisfies the conclusion of Theorem 3.1. By construction, the target dimension is $mk' = O(\varepsilon^{-3} \log \frac{1}{\varepsilon\delta} (\log \lambda \log \log \lambda)^2)$.

We first focus on points in the net N , and later extend the analysis to all points in S . Let us start with a few immediate observations.

LEMMA 3.1. *For every $x, y \in N$ and every $i \in \{1, \dots, m\}$,*

- (i). $\|f_{P_i(x)}(x)\| \leq r$.
(ii). If $P_i(x) = P_i(y) = C$ then $\|f_C(x) - f_C(y)\| \leq G_r(\|x - y\|) \leq \|x - y\|$.

(iii). If $P_i(x) \neq P_i(y)$ then $h_{P_i(x)}(x) \leq \|x - y\|$.

Proof. [Proof of Lemma 3.1] For assertion (i), recall that by the translation, every cluster, and in particular $C = P_i(x)$, contains a point $z_C \in N$ such that $f_C(z_C) = 0$. Thus, using Equation (3.3) we have

$$\|f_C(x) - f_C(z_C)\| \leq \|g_C(x) - g_C(z_C)\| = G_r(\|x - z_C\|) \leq r.$$

To prove the assertion (ii), use Equations (2.2) and (3.3), to get

$$\begin{aligned} \|f_C(x) - f_C(y)\| &\leq \|\Psi_{\text{JL}}\|_{\text{Lip}} \cdot \|g_C(x) - g_C(y)\| \\ &\leq G_r(\|x - y\|) \\ &\leq \|x - y\|. \end{aligned}$$

For assertion (iii), since $C = P_i(x) \neq P_i(y)$ we have that $y \in N \setminus C$, and so $h_C(x) = \min_{z \in N \setminus C} \|x - z\| \leq \|x - y\|$. \blacksquare

Analysis for the net N . We now prove assertions (a)-(c) of Theorem 3.1 for (only) net points. (We shall need this later to complete the proof of the theorem.) To this end, fix $x, y \in N$.

(a) Lipschitz: If $\|x - y\| > \frac{r}{\delta}$, we use the boundedness condition and the fact that $\delta \leq \frac{1}{4}$ to get

$$\|\varphi(x) - \varphi(y)\| \leq \|\varphi(x)\| + \|\varphi(y)\| \leq 2r < \frac{r}{\delta} \leq \|x - y\|.$$

Assume now that $\|x - y\| \leq \frac{r}{\delta}$. Then by Step 6

$$\|\varphi(x) - \varphi(y)\|^2 = \frac{1}{m} \sum_{i=1}^m \|\varphi_i(x) - \varphi_i(y)\|^2. \quad (3.4)$$

To bound the righthand side, fix $i \in \{1, \dots, m\}$ and consider separately the following three cases.

Case 1: x is padded. The padding is $\frac{3r}{\delta}$, hence x and y belong to the same cluster $C = P_i(x) = P_i(y)$. Furthermore, $h_C(x) \geq \frac{3r}{\delta}$, thus $\varphi_i(x) = f_C(x)$; similarly $h_C(y) \geq h_C(x) - \|x - y\| \geq \frac{2r}{\delta}$, and thus $\varphi_i(y) = f_C(y)$. Using Lemma 3.1(ii)

$$\|\varphi_i(x) - \varphi_i(y)\| = \|f_C(x) - f_C(y)\| \leq \|x - y\|.$$

Case 2: x is not padded and $P_i(x) \neq P_i(y)$. By Lemma 3.1(iii),

$$\begin{aligned} \|\varphi_i(x)\| &\leq \|f_{P_i(x)}(x)\| \cdot \frac{\delta}{r} h_{P_i(x)}(x) \\ &\leq \delta h_{P_i(x)}(x) \\ &\leq \delta \|x - y\|. \end{aligned}$$

Using a similar bound for $\varphi_i(y)$, we obtain

$$\begin{aligned} \|\varphi_i(x) - \varphi_i(y)\| &\leq \|\varphi_i(x)\| + \|\varphi_i(y)\| \\ &\leq 2\delta \|x - y\| \\ &\leq \|x - y\|. \end{aligned}$$

Case 3: x is not padded and x, y belong to the same cluster $P_i(x) = P_i(y) = C$. Restrict φ_i to C and write it as the product of the two functions $z \mapsto f_C(z)$ and $\tilde{h}_C : z \mapsto \min\{1, \frac{\delta}{r} h_C(z)\}$. It follows that

$$\begin{aligned} \frac{\|\varphi_i(x) - \varphi_i(y)\|}{\|x - y\|} &\leq \|f_C\|_{\text{Lip}} \cdot \max_{x \in C} |\tilde{h}_C(x)| \\ &\quad + \|\tilde{h}_C\|_{\text{Lip}} \cdot \max_{x \in C} \|f_C(x)\| \end{aligned}$$

It easy to verify that $\max_{z \in C} |\tilde{h}_C(z)| \leq 1$ and $\|\tilde{h}_C\|_{\text{Lip}} \leq \frac{\delta}{r} \cdot \|h_C\|_{\text{Lip}} \leq \frac{\delta}{r}$. Plugging in these estimates and bounds on f_C obtained from Lemma 3.1, we have

$$\frac{\|\varphi_i(x) - \varphi_i(y)\|}{\|x - y\|} \leq 1 \cdot 1 + \frac{\delta}{r} \cdot r = 1 + \delta.$$

Now combine these three cases by plugging into Equation (3.4). Since x is padded in at least $1 - \varepsilon$ fraction of the partitions P_i , and for the remaining partitions we can use the worst bound among the three cases, we get

$$\begin{aligned} \|\varphi(x) - \varphi(y)\|^2 &\leq (1 - \varepsilon) \|x - y\|^2 + \varepsilon(1 + \delta) \|x - y\|^2 \\ &= (1 + \varepsilon\delta) \|x - y\|^2. \end{aligned}$$

(b) Distortion to the Gaussian: We assume henceforth a slightly extended range $\frac{1}{2}\delta r \leq \|x - y\| \leq \frac{2r}{\delta}$. Observe that $\frac{G_r(t)}{t}$ is monotonically decreasing in t , hence

$$\frac{G_r(\|x - y\|)}{\|x - y\|} \geq \frac{G_r(2r/\delta)}{(2r/\delta)} > \frac{G_r(8r)}{(2r/\delta)} > \frac{\delta}{3}, \quad (3.5)$$

We proceed by considering the exact same three cases as above.

Case 1': x is padded. By the analogous case above, $P_i(x) = P_i(y) = C$ and

$$\|\varphi_i(x) - \varphi_i(y)\| = \|f_C(x) - f_C(y)\|.$$

By (3.3) we have $1 - \varepsilon \leq \frac{\|f_C(x) - f_C(y)\|}{\|g_C(x) - g_C(y)\|} \leq 1$, where, by construction, the denominator equals $G_r(\|x - y\|)$. Altogether, we get

$$1 - \varepsilon \leq \frac{\|\varphi_i(x) - \varphi_i(y)\|}{G_r(\|x - y\|)} \leq 1.$$

Case 2': x is not padded and $P_i(x) \neq P_i(y)$. Combining the analogous case above and Equation (3.5), we have

$$\|\varphi_i(x) - \varphi_i(y)\| \leq 2\delta \|x - y\| < 6G_r(\|x - y\|).$$

Case 3': x is not padded and x, y belong to the same cluster $P_i(x) = P_i(y) = C$. Refining the analysis in the analogous case above and using Equation (3.5), we have

$$\begin{aligned} \|\varphi_i(x) - \varphi_i(y)\| &= \|f_C(x)\tilde{h}_C(x) - f_C(y)\tilde{h}_C(y)\| \\ &\leq \|f_C(x)\tilde{h}_C(x) - f_C(x)\tilde{h}_C(y)\| \\ &\quad + \|f_C(x)\tilde{h}_C(y) - f_C(y)\tilde{h}_C(y)\| \\ &\leq \|f_C(x)\| \cdot |\tilde{h}_C(x) - \tilde{h}_C(y)| \\ &\quad + \|f_C(x) - f_C(y)\| \cdot |\tilde{h}_C(y)| \\ &\leq r \cdot \frac{\delta}{r} \|x - y\| \\ &\quad + G_r(\|x - y\|) \cdot 1 \\ &\leq 4G_r(\|x - y\|). \end{aligned}$$

Again combine these three cases by plugging into Equation (3.4) and recalling that x is padded in at least $1 - \varepsilon$ fraction of partitions; we thus get

$$(1 - \varepsilon)^2 \leq \frac{\|\varphi(x) - \varphi(y)\|^2}{G_r(\|x - y\|)^2} \leq (1 - \varepsilon) + \varepsilon \cdot 36 = 1 + 35\varepsilon. \quad (3.6)$$

For later use, let us record that

$$\begin{aligned} \|\varphi(x) - \varphi(y)\| &\leq G_r(\|x - y\|)(1 + 35\varepsilon)^{1/2} \\ &\leq G_r(\|x - y\|)(1 + 18\varepsilon) \\ &\leq \frac{11}{2}G_r(\|x - y\|). \end{aligned} \quad (3.7)$$

(c). Boundedness: By the fact $0 \leq h_{P_i(x)}(x) \leq 1$ and Lemma 3.1(i),

$$\|\varphi(x)\|^2 \leq \frac{1}{m} \sum_{i=1}^m \|\varphi_i(x)\|^2 \leq \frac{1}{m} \sum_{i=1}^m \|f_{P_i(x)}(x)\|^2 \leq r^2.$$

This completes the analysis for net points $x, y \in N$.

Analysis for entire S . To complete the proof of Theorem 3.1, we need to prove assertions (a)-(c) for all points in S . This follows via standard arguments using the above proof for net points and appealing to the Kirschbraun theorem to argue that φ is 1-Lipschitz on the entire S . For technical reasons, it also requires some careful estimates on $G_r(t)$. This part of the proof is deferred to the full paper.

3.3 Snowflake Embedding We now use Theorem 3.1 (the single scale embedding) to prove Theorem 1.2 (embedding for d^α). For simplicity, we will first prove the theorem for $\alpha = 1/2$, and then extend the proof to arbitrary $0 < \alpha < 1$.

Fix a finite set $S \subset \ell_2$ and $0 < \varepsilon < 1/4$. Assume without loss of generality that the minimum

interpoint distance in S is 1. Define $p = 3\lceil \log_{1+\varepsilon}(\frac{1}{\varepsilon}) \rceil = O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$, and the set $I = \{i \in \mathbb{Z} : \varepsilon^5 \leq (1 + \varepsilon)^i \leq \varepsilon^{-5} \text{diam}(S)\}$. For each $i \in I$, let $\varphi_i : S \rightarrow \ell_2^k$ be the embedding that achieves the bounds of Theorem 1.3 for S and ε with respect to parameters $r = (1 + \varepsilon)^i$ and $\delta = (1 + \varepsilon)^{-p} = \Theta(\varepsilon^3)$. Notice that each φ_i has target dimension $k = \tilde{O}(\varepsilon^{-3} \log^2 \lambda)$.

We shall now use the following technique due to Assouad [Ass83]. First, each φ_i is scaled by $1/\sqrt{r} = (1 + \varepsilon)^{-i/2}$. They are then grouped in a round robin fashion into $2p$ groups, and the embeddings in each group are summed up. This yields $2p$ embeddings, each into ℓ_2^k , which are combined using a direct-sum, resulting in one map Φ into ℓ_2^{2pk} .

Formally, let $i \equiv_{2p} j$ denote that two integers i, j are equal modulo $2p$. Define $\Phi : S \rightarrow \ell_2^{2pk}$ using the direct sum $\Phi = \bigoplus_{j \in [2p]} \Phi_j$, where each $\Phi_j : S \rightarrow \ell_2^k$ is given by

$$\Phi_j = \sum_{i \in I: i \equiv_{2p} j} \frac{\varphi_i}{(1 + \varepsilon)^{i/2}}.$$

For $M = M(\varepsilon) > 0$ that will be defined later, our final embedding is $\Phi/\sqrt{M} : S \rightarrow \ell_2^{2pk}$, which has target dimension $pk \leq \tilde{O}(\varepsilon^{-4} \log^2 \lambda)$, as required (for $\alpha = 1/2$). It thus remains to prove the distortion bound. We will need the following lemma.

LEMMA 3.2. *Let $\Phi : S \rightarrow \ell_2^{2pk}$ be as above, let $x, y \in S$, and define $B_i = \frac{\|\varphi_i(x) - \varphi_i(y)\|}{(1 + \varepsilon)^{i/2}}$. Then for every interval $A \subset I$ of size $2p$ (namely $A = \{a, a+1, \dots, a+2p-1\}$),*

$$\begin{aligned} \|\Phi(x) - \Phi(y)\|^2 &\leq \sum_{i \in A} \left(B_i + \sum_{i' \in I \setminus A: i' \equiv_{2p} i} B_{i'} \right)^2 \\ &\leq \sum_{a \leq i < a+p} \left(B_i^2 + B_{i+p}^2 + 2(B_i + B_{i+p}) \sum_{i' \in I \setminus A: i' \equiv_p i} B_{i'} \right) \\ &\quad + \left(\sum_{i' \in I \setminus A: i' \equiv_p i} B_{i'} \right)^2 \end{aligned}$$

and also

$$\begin{aligned} \|\Phi(x) - \Phi(y)\|^2 &\geq \sum_{i \in A} \left(\max \left\{ 0, B_i - \sum_{i' \in I \setminus A: i' \equiv_{2p} i} B_{i'} \right\} \right)^2 \\ &\geq \sum_{a \leq i < a+p} \left(B_i^2 + B_{i+p}^2 - 2(B_i + B_{i+p}) \sum_{i' \in I \setminus A: i' \equiv_p i} B_{i'} \right) \end{aligned}$$

Proof. By construction,

$$\begin{aligned} \|\Phi(x) - \Phi(y)\|^2 &= \sum_{j \in [2p]} \|\Phi_j(x) - \Phi_j(y)\|^2 \\ &= \sum_{i \in A} \left\| \sum_{i' \in I: i' \equiv_p i} \frac{\varphi_{i'}(x) - \varphi_{i'}(y)}{(1+\varepsilon)^{i/2}} \right\|^2. \end{aligned}$$

Fix $i \in A$ and let us bound the term corresponding to i . The first required inequality now follows by separating (among all $i' \in I$ with $i' \equiv_p i$) the term for $i' = i$ from the rest, and applying the triangle inequality for vectors $v_1, \dots, v_s \in \ell_2^k$, namely, $\|\sum_l v_l\| \leq \sum_l \|v_l\|$. We then bound the sum of the terms for indices i and $i+p$. The second inequality follows similarly by separating the term for $i' = i$ from the rest, and applying the following triangle inequality for vectors $u, v_1, \dots, v_s \in \ell_2^k$, namely, $\|u + \sum_l v_l\| \geq \max\{0, \|u\| - \sum_l \|v_l\|\}$. We then bound the sum of the terms for indices i and $i+p$. (Note that $(\max\{0, b-c\})^2 > b^2 - 2bc$). ■

The proof of Theorem 1.2 proceeds by demonstrating that, for an appropriate choice of A (meaning p and a), the leading terms in the above summations (B_i^2 and B_{i+p}^2 for $a \leq i < a+p$) dominate the sum of all other terms of the summations. Fix $x, y \in S$, and let $i^* \in I$ be such that $(1+\varepsilon)^{i^*} \leq \|x-y\| \leq (1+\varepsilon)^{i^*+1}$. We wish to apply Lemma 3.2. To this end, let $A = \{i^* - p, \dots, i^* + p - 1\}$ and consider $i \in A$. We have the following lemma:

LEMMA 3.3. *The following hold for all $i^* - p \leq i < i^*$:*

- (a). $2\varepsilon^3 \cdot B_i \geq \sum_{i' \in I \setminus A: i' \equiv_p i, i' < i} B_{i'}$
- (b). $2\varepsilon^3(1+\varepsilon) \cdot B_{i+p} \geq \sum_{i' \in I \setminus A: i' \equiv_p i, i' > i} B_{i'}$
- (c). $2\varepsilon^3(1+\varepsilon) \cdot (B_i + B_{i+p}) \geq \sum_{i' \in I \setminus A: i' \equiv_p i} B_{i'}$

The proof of Lemma 3.3 is deferred to the full version.

We also have the following lemma, whose proof is again deferred to the full version of this paper.

LEMMA 3.4. *Let $0 < \eta < 1/3$ and suppose $0 < t' \leq (1+\eta)t$. Then $\frac{G_r(t')}{G_r(t)} \leq 1 + 3\eta$.*

Finally observe that for $i^* - p \leq j < i^* + p$

$$\delta \leq \frac{\|x-y\|}{(1+\varepsilon)^j} \leq \frac{1}{\delta},$$

hence we can apply Theorem 3.1(b) to obtain

$$\frac{1}{1+\varepsilon} \leq \frac{\|\varphi_j(x) - \varphi_j(y)\|}{G_{(1+\varepsilon)^j}(\|x-y\|)} \leq 1.$$

Now, plugging (3.8) and Lemma 3.3 into Lemma 3.2, we obtain

$$\begin{aligned} \|\Phi(x) - \Phi(y)\|^2 &\geq \sum_{i \in A: i < i^*} \left(B_i^2 + B_{i+p}^2 - 2\varepsilon^3(B_i + B_{i+p})^2 \right) \\ &\geq (1-4\varepsilon^3) \sum_{i \in A: i < i^*} \left(B_i^2 + B_{i+p}^2 \right) \\ &= (1-4\varepsilon^3(1+\varepsilon)) \sum_{i \in A} B_i^2 \\ &\geq \frac{1-4\varepsilon^3(1+\varepsilon)}{(1+\varepsilon)^2} \sum_{i \in A} \left(\frac{G_{(1+\varepsilon)^i}(\|x-y\|)}{(1+\varepsilon)^{i/2}} \right)^2 \\ &\geq \frac{1-4\varepsilon^3(1+\varepsilon)}{(1+\varepsilon)^2} \\ &\quad \sum_{i^*-p \leq i < i^*+p} \left((1+\varepsilon)^{i/2} \cdot G((1+\varepsilon)^{i^*-i}) \right)^2 \\ &\geq (1-4\varepsilon^3(1+\varepsilon))(1+\varepsilon)^{i^*-2} \\ &\quad \sum_{b: -p \leq b < p} \left((1+\varepsilon)^{b/2} \cdot G((1+\varepsilon)^{-b}) \right)^2 \\ &\geq (1-4\varepsilon^3(1+\varepsilon))(1+\varepsilon)^{-1} \|x-y\| \\ &\quad \sum_{b: -p \leq b < p} \left((1+\varepsilon)^{b/2} \cdot G((1+\varepsilon)^{-b}) \right)^2. \end{aligned}$$

and similarly, using also Lemma 3.4

$$\begin{aligned} \|\Phi(x) - \Phi(y)\|^2 &\leq (1+8\varepsilon^3(1+\varepsilon)) \sum_{i \in A: i < i^*} \left(B_i^2 + B_{i+p}^2 \right) \\ &= (1+8\varepsilon^3(1+\varepsilon)) \sum_{i \in A} B_i^2 \\ &\leq (1+8\varepsilon^3(1+\varepsilon)) \sum_{i \in A} \left(\frac{G_{(1+\varepsilon)^i}(\|x-y\|)}{(1+\varepsilon)^{i/2}} \right)^2 \\ &\leq (1+8\varepsilon^3(1+\varepsilon)) \\ &\quad \sum_{i^*-p \leq i < i^*+p} \left((1+\varepsilon)^{i/2} \cdot G((1+\varepsilon)^{i^*-i+1}) \right)^2 \\ &\leq (1+8\varepsilon^3(1+\varepsilon))(1+\varepsilon)^{i^*} (1+3\varepsilon)^2 \\ &\quad \sum_{b: -p \leq b < p} \left((1+\varepsilon)^{b/2} \cdot G((1+\varepsilon)^{-b}) \right)^2 \\ &\leq (1+8\varepsilon^3(1+\varepsilon))(1+3\varepsilon)^2 \|x-y\| \\ &\quad \sum_{b: -p \leq b < p} \left((1+\varepsilon)^{b/2} \cdot G((1+\varepsilon)^{-b}) \right)^2. \end{aligned}$$

$$(3.8) \quad \text{Setting } M = \sum_{b: -p \leq b < p} \left((1+\varepsilon)^{b/2} \cdot G((1+\varepsilon)^{-b}) \right)^2, \text{ which clearly depends only on } \varepsilon \text{ (and is in particular}$$

independent of x, y), we combine the last two estimates to obtain

$$(1+\varepsilon)^{-1} - 4\varepsilon^3 \leq \frac{\|\Phi(x) - \Phi(y)\|^2}{M \|x - y\|} \leq (1+8\varepsilon^3(1+\varepsilon))(1+3\varepsilon)^2$$

We conclude that the final embedding Φ/\sqrt{M} achieves distortion $1 + O(\varepsilon)$ for $\alpha = 1/2$.

Arbitrary $0 < \alpha < 1$. Turning to proving the theorem for arbitrary values of $0 < \alpha < 1$, we repeat the previous construction and proof with $p = \frac{3}{\alpha} \lceil \log_{1+\varepsilon}(\frac{1}{\varepsilon}) \rceil$ and $\delta = (1 + \varepsilon)^{-p\tilde{\alpha}} = \Theta(\varepsilon^3)$, where $\tilde{\alpha} = \min\{\alpha, 1 - \alpha\}$. As before, $\varphi_i : S \rightarrow \ell_2^k$ is the embedding that achieves the bounds of Theorem 1.3 for S , and $\Phi : S \rightarrow \ell_2^{pk}$ is defined by the direct sum $\Phi = \bigoplus_{j \in [2p]} \Phi_j$, where each $\Phi_j : S \rightarrow \ell_2^k$ is given by

$$\Phi_j = \sum_{i \in I: i \equiv j} \frac{\varphi_i}{(1 + \varepsilon)^{i(1-\alpha)}}.$$

The final embedding is $\Phi/\sqrt{M} : S \rightarrow \ell_2^{pk}$ (for the same M as above), which has target dimension $pk \leq \tilde{O}(\varepsilon^{-4} \tilde{\alpha}^{-1} \log^2 \lambda)$, as required.

We need to make only small changes to the preceding proof of distortion: In the statement and proof of Lemma 3.2 and elsewhere, the dividing term $(1 + \varepsilon)^{i/2}$ is replaced by $(1 + \varepsilon)^{i(1-\alpha)}$. Note that the increase in value of p (and the introduction of the term $\tilde{\alpha}$), is necessary for Lemma 3.3 to hold in this setting. No other changes to the proof are necessary, and this completes the proof of Theorem 1.2.

4 Extension to Other ℓ_p Spaces

We explain how our results and techniques can be extended to other ℓ_p spaces. For concreteness, we consider only two important spaces, namely ℓ_1 and ℓ_∞ . A number of key tools used in our previous embeddings are specific to ℓ_2 , for example the JL-Lemma, the Gaussian transform, and the Kirschbraun theorem, and we must therefore find suitable replacements for these tools. Note however that there is no Lipschitz extension theorem for ℓ_1 .

The primary result of this section is a variant of our snowflake embedding, Theorem 1.2. We note that the snowflake operator is necessary in this theorem, as for $\alpha = 1$ (and either $p = 1$ or $p = \infty$) Lee, Mendel and Naor [LMN05, Theorem 1.3] have shown that the target dimension cannot be bounded as a function of $\lambda(S)$, independently of $|S|$.

THEOREM 4.1. *Let $0 < \varepsilon < 1/4$, $0 < \alpha < 1$ and $p = \{1, \infty\}$. Every finite subset $S \subset \ell_p$ with $\lambda = \lambda(S)$ admits an embedding $\Phi : S \rightarrow \ell_p^k$ satisfying*

$$1 \leq \frac{\|\Phi(x) - \Phi(y)\|_p}{\|x - y\|_p^\alpha} \leq 1 + \varepsilon, \quad \forall x, y \in S;$$

with $k = (1 - \alpha)^{-1} \exp\{\lambda^{O(\log(1/\varepsilon) + \log \log \lambda)}\}$ for $p = 1$, and $k = (1 - \alpha)^{-1} \lambda^{O(\log(1/\varepsilon) + \log \log \lambda)}$ for $p = \infty$.

Recall that our (refined) single scale embedding for ℓ_2 (Theorem 3.1), coupled with an application of Assouad's technique, were sufficient to prove Theorem 1.2. Similarly, single scale embeddings for ℓ_1 and ℓ_∞ , coupled with a standard application of Assouad's technique, are sufficient to prove Theorem 4.1. We present single scale embeddings for ℓ_1 and ℓ_∞ below, and Theorem 4.1 then follows easily.

4.1 Single Scale Embedding for ℓ_1 We can extend Theorem 3.1 to ℓ_1 spaces as follows. For $r > 0$ define $L_r : \mathbb{R} \rightarrow \mathbb{R}$, called the Laplace distance transform, by $L_r(t) = r(1 - e^{-t/r})$. Observe that $L_r(t) = r \cdot G(\sqrt{t/r})^2$.

THEOREM 4.2. *For every scale $r > 0$ and every $0 < \delta, \varepsilon < 1/4$, every finite set $S \subset \ell_1$ admits an embedding $\varphi : S \rightarrow \ell_1^k$ for $k = \exp\{\lambda^{O(\log(1/\varepsilon\delta) + \log \log \lambda)}\}$. satisfying:*

- (a). *Lipschitz:* $\|\varphi(x) - \varphi(y)\|_1 \leq \|x - y\|_1$ for all $x, y \in S$.
- (b). *$1 + \varepsilon$ distortion to the Laplace transform (at scales near r):* For all $x, y \in S$ with $\delta r \leq \|x - y\|_1 \leq \frac{r}{\delta}$,

$$\frac{1}{1 + \varepsilon} \leq \frac{\|\varphi(x) - \varphi(y)\|_1}{L_r(\|x - y\|_1)} \leq 1;$$

- (c). *Boundedness:* $\|\varphi(x)\|_1 \leq r$ for all $x \in S$.

Proof. [Proof Sketch] We would like to utilize the framework designed for ℓ_2 in Section 3.1. However, a few problems arise. Let us point them out explain how to solve them.

- Step 7: This step is not possible for ℓ_1 norm, since there is no ℓ_1 -analogue of the Kirschbraun theorem. Instead, we modify the entire construction (specifically, steps 2-6) so that they work with the entire data set S , not only with the net N . The effect of this will be seen shortly. (In ℓ_2 , the same approach of discarding step 7 can be achieved by applying the Kirschbraun Theorem separately in every cluster in step 4, but this approach does not seem to have any advantages.)
- Step 2: We apply a padded decomposition to the entire set S (and not only to the net N) with essentially same parameters and bounds. Thus, from now on each cluster $C \in P_i$ is a subset of S (but not necessarily of N).
- Step 3: Instead of the Gaussian transform, we apply the Laplace transform L_r , i.e. g_C now

satisfies $\|g_C(x) - g_C(y)\|_1 = L_r(\|x - y\|)$ for all $x, y \in C$. Such an embedding $g_C : C \rightarrow \ell_1$ is known to exist, see [DL97, Corollary 9.1.3]. The effect is clearly quite similar to that of the Gaussian transform. The fact that C is not a subset of N is not an issue.

- Step 4: We need to find a weak analogue to the JL lemma, but there is an additional complication of having to deal with points not in the net N . Specifically, we need a map $\Psi : g_C(C) \rightarrow \ell_1^{k'}$ which satisfies: (i) Ψ is 1-Lipschitz on the entire cluster $g_C(C)$; and (ii) Ψ achieves $1 + \varepsilon$ distortion on the cluster net points $g_C(C \cap N)$. Observe that the former requirement is non-standard and does not follow from ‘standard’ dimension reduction theorems for a finite subsets of ℓ_1 . However, we describe below one simple dimension reduction for ℓ_1 which does extend to our setting. (This construction was observed jointly with Gideon Schechtman.) We suspect that the dimension can be further reduced, since the current construction is an isometry on $g_C(C \cap N)$, and does not exploit the $1 + \varepsilon$ distortion allowed by requirement (ii). However, an improved map Ψ cannot be linear, since in the worst case such a linear map requires dimension $k = 2^{\Omega(|C \cap N|)}$ [FJS91, Corollary 12.A].

Construct Ψ as follows. Since the metric $g_C(C) \in \ell_1$, it can be written as a conic combination of cut metrics, i.e. there are $\gamma_A \geq 0$ for $A \subset C$ such that

$$\|g_C(x) - g_C(y)\|_1 = \sum_A \gamma_A |1_A(x) - 1_A(y)|, \quad \forall x, y \in C,$$

where $1_A(x) = 1$ if $x \in A$ and 0 otherwise. In other words, $g_C(x) \rightarrow \sum_A \alpha_A 1_A(x)$ is an isometric embedding of $g_C(C)$ into ℓ_1 . Let Ψ have one coordinate for every subset $B \subset C \cap N$; this coordinate is given by $g_C(x) \rightarrow \sum_{A: A \cap (C \cap N) = B} \gamma_A 1_A(x)$. In words, we add together coordinates whenever they correspond to different A but have the same $A \cap (C \cap N)$. Observe that Ψ is 1-Lipschitz for all $g_C(x)$, $x \in C$, simply because adding two coordinates together can only decrease distances, and that it is isometric for $g_C(x)$, $x \in C \cap N$, because now if coordinates corresponding to A and A' are added together then necessarily $1_A(x) = 1_{A'}(x)$ and similarly $1_A(y) = 1_{A'}(y)$. Observe that $k' = 2^{|C \cap N|}$ where $|C \cap N| \leq \lambda^{O(\Delta/\varepsilon\delta r)}$.

- Step 5: There is only a minor change; since we do not restrict attention to net points, we now define $h_C(x) = \min_{y \in S \setminus C} \|x - y\|_1$.
- Step 6: There is only a minor change to the scaling factor, namely $\varphi = m^{-1} \bigoplus_{i=1}^m \varphi_i$.

The rest of the proof is quite similar to the one presented for ℓ_2 , and the final dimension obtained is $mk' = O(\varepsilon^{-1} \log \lambda \log \log \lambda) \cdot \exp(\lambda^{O(\log(\varepsilon^{-1} \delta^{-2} \log \lambda))}) = \exp\{\lambda^{O(\log(1/\varepsilon\delta) + \log \log \lambda)}\}$. ■

4.2 Single Scale Embedding for ℓ_∞ We can also extend Theorem 3.1 to ℓ_∞ spaces as follows. For $r > 0$ define $T_r : \mathbb{R} \rightarrow \mathbb{R}$, called the threshold transform, by $T_r(s) = \min\{s, r\}$.

THEOREM 4.3. *For every scale $r > 0$ and every $0 < \delta, \varepsilon < 1/4$, every finite set $S \subset \ell_\infty$ admits an embedding $\varphi : S \rightarrow \ell_\infty^{k'}$ for $k' = \lambda^{O(\log(1/\varepsilon\delta) + \log \log \lambda)}$ satisfying:*

- (a). *Lipschitz:* $\|\varphi(x) - \varphi(y)\|_\infty \leq \|x - y\|_\infty$ for all $x, y \in S$.
- (b). *$1 + \varepsilon$ distortion to the threshold transform (at scales near r):* For all $x, y \in S$ with $\delta r \leq \|x - y\|_\infty \leq \frac{r}{\sqrt{\delta}}$,

$$\frac{T_r(\|x - y\|_\infty)}{1 + \varepsilon} \leq \|\varphi(x) - \varphi(y)\|_\infty \leq T_r(\|x - y\|_\infty)$$

- (c). *Boundedness:* $\|\varphi(x)\|_\infty \leq r$ for all $x \in S$.

Proof. [Proof Sketch] The proof is quite similar to that of Theorem 4.2, except for a few changes in some of the arguments.

- Step 3: The thresholding of distances is easily achieved by a simple variant of the well-known Frechet embedding. Formally, g_C has $|C|$ coordinates, one for every point $z \in C$, and that coordinate is given by $x \mapsto \min\{\|z - x\|_\infty, r\}$. It is easily verified that $\|g_C(x) - g_C(y)\|_\infty = T_r(\|x - y\|_\infty)$ for all $x, y \in C$.
- Step 4: The required bound is again obtained by a simple variant of the well-known Frechet embedding. Formally, $\Psi : g_C(C) \rightarrow \ell_\infty^{k'}$ has one coordinate for every point $z \in g_C(C \cap N)$, and that coordinate is given by $t \mapsto \|t - z\|_\infty$. Thus, $k' = |C \cap N|$. It is easily verified that this map Ψ is 1-Lipschitz on the entire cluster $g_C(C)$ and also isometric on $g_C(C \cap N)$.
- Step 6: The scaling factor is different and now $\varphi = \frac{1}{1+2\sqrt{\delta}} \bigoplus_{i=1}^m \varphi_i$. The resulting embedding is 1-Lipschitz: We consider the worst partition P_i (without averaging the partitions). Case 3 in the Lipschitz analysis for ℓ_2 yields a Lipschitz constant of $1 + \delta$, and $\frac{1+\delta}{1+2\sqrt{\delta}} < 1$.

We remark that it suffices to use a padded decomposition with padding probability $1/2$ (instead of $1 - \varepsilon$), but asymptotically this change does not improve the dimension. The final dimension obtained

is $mk' = O(\varepsilon^{-1} \log \lambda \log \log \lambda) \cdot (\lambda^{O(\log(\varepsilon^{-1} \delta^{-2} \log \lambda))}) = \lambda^{O(\log(1/\varepsilon\delta) + \log \log \lambda)}$.

The lower bound on $\|\varphi(x) - \varphi(y)\|_\infty$ for pair x, y follows when x is padded (case 1' in the distortion to the Gaussian analysis for ℓ_2), where we have $\|\varphi(x) - \varphi(y)\|_\infty = T_r(\|x - y\|_\infty)$; we further stipulate without loss of generality that $\delta \leq \frac{\varepsilon^2}{4}$, so that the scaling factor is $1 + 2\sqrt{\delta} \leq 1 + \varepsilon$.

For the upper bound we have (from cases 2' and 3') $\|\varphi(x) - \varphi(y)\|_\infty \leq \max\{2\delta\|x - y\|_\infty, \delta\|x - y\|_\infty + T_r(\|x - y\|_\infty)\} \leq 2\delta\|x - y\|_\infty + T_r(\|x - y\|_\infty)$. We consider two possibilities:

1. $\delta r \leq \|x - y\|_\infty \leq r$. Then $T_r(\|x - y\|_\infty) = \|x - y\|_\infty$ and $2\delta\|x - y\|_\infty + T_r(\|x - y\|_\infty) = (1 + 2\delta)T_r(\|x - y\|_\infty)$.
2. $r < \|x - y\|_\infty \leq \frac{r}{\sqrt{\delta}}$. Then $T_r(\|x - y\|_\infty) = r$ and $2\delta\|x - y\|_\infty + T_r(\|x - y\|_\infty) \leq (1 + 2\sqrt{\delta})r = (1 + 2\sqrt{\delta})T_r(\|x - y\|_\infty)$.

The final result follows from the scaling factor in Step 6. \blacksquare

5 Algorithmic applications

Here we illustrate the effectiveness and potential of our results for various algorithmic tasks by describing two immediate (theoretical) applications.

Distance Labeling Scheme (DLS). Consider this problem for the family of n -point ℓ_2 metrics with a given bound on the doubling dimension. As usual, we assume the interpoint distances are in the range $[1, R]$. Our snowflake embedding into ℓ_2^k (Theorem 1.2 for $\alpha = \frac{1}{2}$) immediately provides a DLS with approximation $(1 + \varepsilon)^2 \leq 1 + 3\varepsilon$, simply by rounding each coordinate to a multiple of $\varepsilon/2k$. We have:

LEMMA 5.1. *Every finite subset ℓ_2 with $\lambda = \lambda(S)$ possesses a $(1 + \varepsilon)$ -approximate distance labeling scheme with label size*

$$k \cdot \log \frac{R}{\varepsilon/2k} = \tilde{O}(\varepsilon^{-4}(\dim S)^2) \log R.$$

Notice that, apart from the $\log R$ term, this bound is independent of n . The published bounds of this form (see [HM06] and references therein) apply to the more general family of all doubling metrics (not necessarily Euclidean) but require exponentially larger label size, roughly $(1/\varepsilon)^{O(\dim S)}$.

Approximation algorithms for clustering. Clustering problems are often defined as an optimization problem whose objective function is expressed in terms of distances between data points. For example, in the k -center problem one is given a metric (S, d) and

is asked to identify a subset of centers $C \subset S$ that minimizes the objective $\max_{x \in S} d(x, C)$. When the data set S is Euclidean (and the centers are discrete, i.e. from S), one can apply our snowflake embedding (Theorem 1.2) and solve the problem in the target space, which has low dimension k . Indeed, it is easy to see how to map solutions from the original space to the target space and vice versa, with a loss of at most a $(1 + \varepsilon)^2 \leq 1 + 3\varepsilon$ factor in the objective.

For other clustering problems, like k -median or min-sum clustering, the objective function is the *sum* of certain distances. The argument above applies, except that now in the target space we need an algorithm that solves the problem with ℓ_2 -squared costs. For instance, to solve the k -median problem in the original space, we can use an algorithm for k -means in the target space. Schulman [Sch00] has designed algorithms for min-sum clustering under both ℓ_2 and ℓ_2 -squared costs, and their run time depend exponentially on the dimension. The following lemma follows from our snowflake embedding and [Sch00, Propositions 14,28]. For simplicity, we will assume that $k = O(1)$.

LEMMA 5.2. *Given a set of n points $S \in \mathbb{R}^d$, a $(1 + \varepsilon)$ -approximation to the ℓ_2 min-sum k -clustering for S , for $k = O(1)$, can be computed*

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

where $d' = \min\{d, \tilde{O}(\varepsilon^{-4} \dim^2 S)\}$.

Acknowledgments The authors thank Assaf Naor and Gideon Schechtman for useful discussions and references, and Yair Bartal for helpful comments on an earlier version of this paper.

References

- [ABN08] I. Abraham, Y. Bartal, and O. Neiman. Embedding metric spaces in their intrinsic dimension. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 363–372. SIAM, 2008.
- [ABN09] I. Abraham, Y. Bartal, and O. Neiman. On low dimensional local embeddings. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 875–884. SIAM, 2009.
- [Alon03] N. Alon. Problems and results in extremal combinatorics. I. *Discrete Math.*, 273(1-3):31–53, 2003.
- [Ass83] P. Assouad. Plongements lipschitziens dans \mathbf{R}^n . *Bull. Soc. Math. France*, 111(4):429–448, 1983.
- [Bal90] K. Ball. Isometric embedding in l_p -spaces. *European J. Combin.*, 11(4):305–311, 1990.

- [BC05] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1 . *J. ACM*, 52(5):766–788, 2005.
- [BRS11] Y. Bartal, B. Recht, and L. Schulman. A Nash-type dimensionality reduction for discrete subsets of L_2 . Unpublished manuscript, 2007. A revised version, entitled “Dimensionality reduction: beyond the Johnson-Lindenstrauss bound” appears in Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms, 2011.
- [CGT10] H. Chan, A. Gupta, and K. Talwar. Ultra-low-dimensional embeddings for doubling metrics. *J. ACM*, 57(4):1–26, 2010.
- [DL97] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*. Springer-Verlag, Berlin, 1997.
- [FJS91] T. Figiel, W. B. Johnson, and G. Schechtman. Factorizations of natural embeddings of l_p^n into l_r . II. *Pacific J. Math.*, 150(2):261–277, 1991.
- [GKL03] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Annual IEEE Symposium on Foundations of Computer Science*, pages 534–543, October 2003.
- [HM06] 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.
- [IN07] P. Indyk and A. Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):31, 2007.
- [JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.
- [Kah81] J.-P. Kahane. Hélices et quasi-hélices. In *Mathematical analysis and applications, Part B*, volume 7 of *Adv. in Math. Suppl. Stud.*, pages 417–433. Academic Press, New York, 1981.
- [Kir34] M.D. Kirszbraun. Über die zusammenziehenden und lipschitzchen transformationen. *Fundam. Math.*, pages 77–108, 1934.
- [LMN05] J. R. Lee, M. Mendel, and A. Naor. Metric structures in L_1 : dimension, snowflakes, and average distortion. *European J. Combin.*, 26(8):1180–1190, 2005.
- [LN04] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1 . *Geom. Funct. Anal.*, 14(4):745–747, 2004.
- [LP01] U. Lang and C. Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.
- [Mat96] J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.
- [Mat02] J. Matoušek. Open problems on low-distortion embeddings of finite metric spaces. Available at <http://kam.mff.cuni.cz/~matousek/metrop.ps>, 2002. Revised March 2007.
- [NZ02] T. S. E. Ng and H. Zhang. Predicting internet network distance with coordinates-based approaches. In *INFOCOM*, volume 1, pages 170–179, 2002.
- [RS00] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [Sch38] I. J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536, November 1938.
- [Sch87] G. Schechtman. More on embedding subspaces of L_p in l_p^n . *Compositio Math.*, 61(2):159–169, 1987.
- [Sch00] L. J. Schulman. Clustering for edge-cost minimization (extended abstract). In *32nd Annual ACM Symposium on Theory of Computing*, pages 547–555. ACM, 2000. Full version available as ECCC report TR99-035.
- [Tal90] M. Talagrand. Embedding subspaces of L_1 into l_1^N . *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.
- [Tal92] M. Talagrand. Approximating a helix in finitely many dimensions. *Ann. Inst. H. Poincaré Probab. Statist.*, 28(3):355–363, 1992.
- [Tal95] M. Talagrand. Embedding subspaces of L_p in l_p^N . In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 311–325. Birkhäuser, Basel, 1995.
- [TdSL00] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.