A Nonlinear Approach to Dimension Reduction

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

April 13, 2010

Abstract

The $\ell_2$ 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]), 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 $\ell_2^D$, 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 $\ell_1$ and $\ell_\infty$, although the dimension bounds here are quantitatively inferior than those for $\ell_2$.

*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
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 $\ell_k$. (Here and throughout, $\ell_k$ denotes the space $\mathbb{R}^k$ endowed with the $\ell_p$-norm; $\ell_p$ 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:

**Theorem 1.1** (Johnson and Lindenstrauss [JL84]). For every $n$-point subset $S \subset \ell_2$ and every $0 < \varepsilon < 1$, there is a mapping $\Psi_{JL} : S \to \ell_2$ that preserves all interpoint distances in $S$ within factor $1 + \varepsilon$, and has target dimension $k = O(\varepsilon^{-2} \log n)$.

This positive result is remarkably strong; in fact the map $\Psi_{JL}$ 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 $\lambda$ 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 $\dim(M) \overset{\text{def}}{=} \log_2 \lambda(M)$, 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 $\Psi : M \to N$. We say that $\Psi$ attains distortion $D' \geq 1$ if $\Psi$ 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]):

**Question 1.** Does every doubling subset $S \subset \ell_2$ embed with distortion $D'$ into $\ell_D$ 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 $\ell_1$ and $\ell_\infty$. 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 $\ell_p$); they extend to infinite subsets of $L_p$ via standard arguments.

**Snowflake Embedding.** Our primary embedding achieves distortion $1 + \varepsilon$ for the snowflake metric $d^\alpha$ 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 $\alpha$ 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 $0 < \varepsilon < 1/4$ and $0 < \alpha < 1$. Every finite subset $S \subset \ell_2$ admits an embedding $\Phi : S \to \ell^k_2$ for $k = \tilde{O}(\varepsilon^{-4}(1-\alpha)^{-1}(\dim S)^2)$, 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 $d^\alpha$. 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 $d^\alpha$ embeds into $\ell^D_2$ with distortion $D'$, where $D, D'$ depend only on $\lambda(M)$ and $\alpha$. 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 $\ell_2^k$. 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 $\ell_2$ (or other $\ell_p$), albeit with a somewhat worse dependence on $\varepsilon$.

**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 $\varphi : S \to \ell_2^k$ 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 $x \in S$.  

2
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 $\tilde{G} : \mathbb{R} \to \mathbb{R}$, such that whenever $\|x-y\|_2 \in [\delta r, r]$,

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

(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).

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 $\ell_p$ Spaces. Our dimension reduction framework extends to $\ell_p$ (i.e. $S \subset \ell_p$ and $\Phi : S \to \ell^k_p$) for both $p = 1$ and $p = \infty$, as discussed in Section 4. The bounds we obtain therein are worse than in the $\ell_2$ 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 $\ell_1$ and [Mat96] for $\ell_\infty$). We remark that previous work on dimension reduction in $\ell_p$ 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 $\ell_2$, 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 [N07].

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 [BRS07]
A sampling of related work; holds for arbitrary $\varepsilon \in (0, 1)$

| Reference | Origin | Target space | Distortion | Dimension | Snowflake $\alpha$ |
|-----------|--------|--------------|------------|-----------|-------------------|
| [JL84]    | $\ell_2$ | $\ell_2$ | $1 + \varepsilon$ | $O(\varepsilon^{-2} \log n)$ | none ($\alpha = 1$) |
| [Ass83]   | doubling | $\ell_2$ | $2^{O(\dim S)}$ | $2^{O(\dim S)}$ | fixed $\alpha < 1$ |
| [GKL03]   | doubling | $\ell_2$ | $\tilde{O}(\dim S)$ | $\tilde{O}(\dim S)$ | fixed $\alpha < 1$ |
| [HM06]    | doubling | $\ell_\infty$ | $1 + \varepsilon$ | $\varepsilon^{-O(\dim S)}$ | fixed $\alpha < 1$ |
| [ABN08]   | doubling | $\ell_p$, $p \geq 1$ | $O((\log^{1+\varepsilon} n)$ | $O(\varepsilon^{-1} \dim S)$ | none |
| Theorem 1.2 | $\ell_2$ | $\ell_2$ | $1 + \varepsilon$ | $\tilde{O}(\varepsilon^{-3} \dim^2 S)$ | fixed $\alpha < 1$ |

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

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 $O(\varepsilon^{-3} \dim S)$. This additional result has been appended to the most recent version of [BRS07].

### 1.3 Applications

In many settings, data is provided as points in $\ell_p$, 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 3 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 [TDLSL00, RS00] for machine learning. Our results can be interpreted as conveying the following principle: Intrinsically low-dimensional data that admits a meaningful representation in $\ell_2$, can actually be represented in low-dimensional $\ell_2$.

**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 Kirszbraun’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 \(\lambda\) be the infimum value such that every ball in \(X\) can be covered by \(\lambda\) balls of half the radius. The doubling dimension of \(X\) is \(\dim(X) = \log_2 \lambda\). A metric is doubling when 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 \(\alpha\), and the diameter of \(S\) is at most \(\beta\), then \(|S| \leq \lambda^{O(\log(\beta/\alpha))}\).

\(\epsilon\)-nets. For a point set \(S\), an \(\epsilon\)-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 \(\epsilon\) 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 \(f : X \to Y\) 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 \(f : X \to \ell^2_p\) and \(g : X \to \mathbb{R}\). Then their product \(fg : x \to g(x) \cdot f(x)\) 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 \cite{Kir34} states that if \(S\) and \(X\) are Euclidean spaces, \(T \subset S\), and there exists a Lipschitz function \(f : T \to X\); then there exists a function \(\hat{f} : S \to X\) that has the same Lipschitz constant as \(f\) and also extends \(f\), i.e. \(\hat{f}|_T = f\), meaning that the restriction of \(\hat{f}\) 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 \cite{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 \(g : X \to L_2\) 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

\[
G_r(t) \leq t, \quad \forall t \geq 0,
\]

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 \cite{BRS07}.

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 \(P = \{C_1, C_2, \ldots, C_t\}\) 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 \cite{GKL03} and Abraham, Bartal and Neiman \cite{ABN08}. Let \(B(x, r) = \{y | ||x - y|| \leq r\}\).
Theorem 2.1 (Padded Decomposition of doubling metrics [GKL03, ABN08]). There exists a constant $c_0 > 1$, such that for every metric space $(X, d)$ and every $\Delta > 0$, there is a multi-set $D = [P_1, \ldots, P_m]$ 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 $D$, then for all $x \in X$,
   \[ \Pr_{P \in D} \left[ B(x, \frac{\Delta}{c_0 \dim(X)}) \subseteq P(x) \right] \geq 1 - \varepsilon. \]

Remark: [GKL03] provided slightly different quantitative bounds than in Theorem 2.1. The two enumerated properties follow 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 $\ell_2$

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 $\ell_2$-norms. We make no attempt to optimize constants.

Following Section 2, define $G : \mathbb{R} \to \mathbb{R}$ by
   \[ G(t) = \left(1 - e^{-t^2}\right)^{1/2}, \]
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 $\varphi : S \to \ell_2^k$ for $k = \tilde{O}(\varepsilon^{-3} \log \frac{1}{r} \cdot (\dim(S))^2)$, 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 $\varphi$ by $\frac{1}{1 + C \varepsilon}$ and scaling of $\varepsilon$ by $1/C$.

3.1 Embedding for a single scale

Our construction of the embedding $\varphi$ 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 $\frac{3\varepsilon}{r}$. More specifically, by Theorem 2.1, there is a multiset $[P_1, \ldots, P_m]$ of partitions of $N$, where every point is $\frac{3\varepsilon}{r}$-padded in $1 - \varepsilon$ fraction of the partitions, all clusters have diameter bounded by $\Delta = O(\frac{\lambda}{\varepsilon} \log \lambda)$, 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 inter-point distances in $C$ at maximum value $r$. Specifically, using a Gaussian transform as per Section 2, obtain a map $g_C : C \to \ell_2$ such that
\[
\|g_C(x) - g_C(y)\|_2^2 = G_r(\|x - y\|)^2 = r^2(1 - e^{-\|x - y\|^2/r^2}), \quad \forall x, y \in C.
\]

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 $\Psi_{JL} : g_C(C) \to \ell_2''$ such that
\[
\frac{\|t - t'\|}{1 + \varepsilon} \leq \|\Psi_{JL}(t) - \Psi_{JL}(t')\| \leq \|t - t'\|, \quad \forall t, t' \in g_C(C),
\]
and the target dimension is (using Property 2.1)
\[
k' = O(\varepsilon^{-2} \log |C|) = O(\varepsilon^{-2} \log(\lambda^O(\log(\Delta/e\delta r)))) = O(\varepsilon^{-2} \log \frac{1}{\varepsilon^2} \cdot \log \lambda \log \log \lambda).
\]
Composing the last two steps, define $f_C = \Psi_{JL} \circ g_C$ mapping $C \to \ell_2''$.

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 $h_C : C \to \mathbb{R}$ 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 $\varphi_i : N \to \ell_2''$ by
\[
\varphi_i(x) = f_{P_i}(x) \cdot \min\{1, \frac{\delta}{r} h_{P_i}(x)\};
\]
recalling 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 $\varphi : N \to \ell_2''$ 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 $\varphi$ to all of $S$, without increasing the Lipschitz constant.

3.2 Proof of Theorem 3.1

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

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 3.2. For every $x, y \in N$ and every $i \in \{1, \ldots, m\}$,

(i). $\|f_{P_i}(x)\| \leq r$.

(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 $h_{P_i}(x) \leq \|x - y\|$.
Proof of Lemma 3.2. 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) 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) and (3), to get

$$
\|f_C(x) - f_C(y)\| \leq \|\Psi_{Lr}\| \|g_C(x) - g_C(y)\| \leq G_r(\|x - y\|) \leq \|x - y\|.
$$

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

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 $\|x - y\| > \frac{r}{5}$, we use the boundedness condition and the fact that $\delta \leq \frac{1}{2}$ to get

$$
\|\varphi(x) - \varphi(y)\| \leq \|\varphi(x)\| + \|\varphi(y)\| \leq 2r < \frac{r}{5} \leq \|x - y\|.
$$

Assume now that $\|x - y\| \leq \frac{r}{5}$. Then by Step 6

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

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

Case 1: $x$ is padded. The padding is $\frac{2r}{5}$, hence $x$ and $y$ belong to the same cluster $C = P_i(x) = P_i(y)$. Furthermore, $h_C(x) \geq \frac{2r}{5}$, thus $\varphi_i(x) = f_C(x)$; similarly $h_C(y) \geq h_C(x) - \|x - y\| \geq \frac{2r}{5}$, and thus $\varphi_i(y) = f_C(y)$. Using Lemma 3.2(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 3.2(iii),

$$
\|\varphi_i(x)\| \leq \|f_{P_i(x)}(x)\| \cdot \frac{2}{r} h_{P_i(x)}(x) \leq \delta h_{P_i(x)}(x) \leq \delta \|x - y\|.
$$

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

$$
\|\varphi_i(x) - \varphi_i(y)\| \leq \|\varphi_i(x)\| + \|\varphi_i(y)\| \leq 2\delta \|x - y\| \leq \|x - y\|.
$$

Case 3: $x$ is not padded and $x, y$ belong to the same cluster $P_i(x) = P_i(y) = C$. Restrict $\varphi_i$ to $C$ and write it as the product of the two functions $z \mapsto f_C(z)$ and $h_C : z \mapsto \min\{1, \frac{2}{r} h_C(z)\}$. It follows that

$$
\|\varphi_i(x) - \varphi_i(y)\| \leq \|f_C\|_{\text{Lip}} \cdot \max_{x \in C} |\tilde{h}_C(x)| + |\tilde{h}_C|_{\text{Lip}} \cdot \max_{x \in C} \|f_C(x)\|.
$$

It easy to verify that $\max_{x \in C} |\tilde{h}_C(z)| \leq 1$ and $|\tilde{h}_C|_{\text{Lip}} \leq \frac{2}{r} \cdot \|h_C\|_{\text{Lip}} \leq \frac{2}{r}$. Plugging in these estimates and bounds on $f_C$ obtained from Lemma 3.2, we have

$$
\frac{\|\varphi_i(x) - \varphi_i(y)\|}{\|x - y\|} \leq 1 \cdot 1 + \frac{2}{r} \cdot r = 1 + \frac{2}{r}.
$$
Now combine these three cases by plugging into Equation (4). Since $x$ is padded in at least $1 - \epsilon$ fraction of the partitions $P_i$, and for the remaining partitions we can use the worst bound among the three cases, we get

$$\|\varphi(x) - \varphi(y)\|^2 \leq (1 - \epsilon)\|x - y\|^2 + \epsilon(1 + \delta)\|x - y\|^2 = (1 + \epsilon\delta)\|x - y\|^2.$$ 

**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 $G_r(t)$ is monotonically decreasing in $t$, hence

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

(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) we have $1 - \epsilon \leq \|f_C(x) - f_C(y)\| \leq 1$, where, by construction, the denominator equals $G_r(\|x - y\|)$. Altogether, we get

$$1 - \epsilon \leq \frac{\|f_C(x) - f_C(y)\|}{G_r(\|x - y\|)} \leq 1.$$ 

**Case 2':** $x$ is not padded and $P_i(x) \neq P_i(y)$. Combining the analogous case above and Equation (3), we have

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

**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), we have

$$\|\varphi_i(x) - \varphi_i(y)\| = \|f_C(x)\bar{h}_C(x) - f_C(y)\bar{h}_C(y)\|$$ 

$$\leq \|f_C(x)\bar{h}_C(x) - f_C(x)\bar{h}_C(y)\| + \|f_C(x)\bar{h}_C(y) - f_C(y)\bar{h}_C(y)\|$$ 

$$\leq \|f_C(x)\| \cdot |\bar{h}_C(x) - \bar{h}_C(y)| + \|f_C(x) - f_C(y)\| \cdot |\bar{h}_C(y)|$$ 

$$\leq r \cdot \frac{\delta}{3}\|x - y\| + G_r(\|x - y\|) \cdot 1$$ 

$$\leq 4G_r(\|x - y\|).$$

Again combine these three cases by plugging into Equation (3) and recalling that $x$ is padded in at least $1 - \epsilon$ fraction of partitions; we thus get

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

(6)

For later use, let us record that

$$\|\varphi(x) - \varphi(y)\| \leq G_r(\|x - y\|)(1 + 35\epsilon)^{1/2} \leq G_r(\|x - y\|)(1 + 18\epsilon) \leq \frac{11}{2}G_r(\|x - y\|).$$ 

(7)

**Boundedness:** By the fact $0 \leq h_{P_i(x)}(x) \leq 1$ and Lemma 3.2(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 $x, y \in N$. 

9
**Analysis for entire $S$.** We extend the previous analysis to all points in $S$. Fix $x, y \in S$, and let $x', y' \in N$ be the net points closest to $x$ and $y$, respectively. Recalling that $N$ is an $\varepsilon\delta r$-net, we have $\|x - x'|, \|y - y'| \leq \varepsilon\delta r$. To prove the Lipschitz requirement, recall that Step 8 extends $\varphi$ from the net $N$ to the entire $S$ using the Kirszbraun theorem, i.e. without increasing its Lipschitz norm, hence

$$\|\varphi(x) - \varphi(y)\| \leq \|x - y\|.$$  

Using this Lipschitz condition and the triangle inequality, we immediately obtain the boundedness requirement:

$$\|\varphi(x)\| \leq \|\varphi(x')\| + \|\varphi\|_{\text{Lip}}\|x - x'\| \leq (1 + \varepsilon\delta) r.$$

To prove the requirement of distortion to the Gaussian (which is slightly more involved) assume further that $\delta r \leq \|x - y\| \leq \frac{\delta}{3}$. By the triangle inequality,

$$\left|\|x - y\| - \|x' - y'\|\right| \leq \|x - x'\| + \|y - y'\| \leq 2\varepsilon\delta r. \tag{8}$$

We conclude that $(1 - 2\varepsilon)\delta r \leq \|x' - y'\| \leq (\frac{1}{3} + 2\varepsilon\delta) r$, and hence $x', y' \in N$ possess the bound for distortion to the Gaussian. It also follows that $2\varepsilon\delta r \leq 4(1 - 2\varepsilon)\varepsilon\delta r \leq 4\varepsilon\|x' - y'\|$. Using the Lipschitz condition on $\varphi$, and the above distortion to the Gaussian for net points (Equation (8)), we similarly derive

$$\left|\|\varphi(x) - \varphi(y)\| - \|\varphi(x') - \varphi(y')\|\right| \leq \|x - x'\| + \|y - y'\| \leq 2\varepsilon\delta r \leq 4\varepsilon\|x' - y'\| < 22\varepsilon G_r(\|x' - y'\|). \tag{9}$$

We shall need the following bound on the behavior of $G_r(t)$.

**Lemma 3.3.** Let $0 < \eta < 1/3$ and suppose $0 < t' \leq (1 + \eta)t$. Then $\frac{G_r(t')}{G_r(t)} \leq 1 + 3\eta$.

**Proof of Lemma 3.3.** Observe that $G_r(t)$ is monotonically increasing (in $t$), and thus

$$\frac{G_r(t')}{G_r(t)} \leq \frac{G_r((1 + \eta)t)}{G_r(t)} \leq \frac{G((1 + \eta)t/r)}{G(t/r)}.$$  

Letting $s = t/r$, we have

$$\frac{G((1 + \eta)s)^2}{G(s)^2} - 1 = \frac{G((1 + \eta)s)^2 - G(s)^2}{G(s)^2} = \frac{e^{-s^2} - e^{-(1+\eta)^2s^2}}{1 - e^{-s^2}} \leq \frac{e^{-s^2}(1 - e^{-3\eta s^2})}{1 - e^{-s^2}}. \tag{10}$$

Recall that for all $0 \leq z \leq 1$ we have $1 - z \leq e^{-z} \leq 1 - z + z^2/2 \leq 1 - z/2$. Using this estimate, we now have three cases:

- When $s^2 \leq 1$, the righthand side of (10) is at most $\frac{13\eta s^2}{s^2} \leq 6\eta$.
- When $1 \leq s^2 \leq 1/3\eta$, the righthand side of (10) is at most $\frac{e^{-s^2 - 3\eta s^2}}{1 - 1/e} \leq 6\eta s^2 e^{-s^2} \leq 6\eta/e$, where the last inequality follows from the observation that $z \mapsto ze^{-z}$ is monotonically decreasing for all $z \geq 1$.  
- When $s^2 \geq 1/3\eta$, the righthand side of (10) is at most $\frac{e^{-s^2}}{1 - 1/e} \leq \frac{e^{-s^2 - 3\eta s^2}}{1 - 1/e} \leq 6\eta/e$, where the last inequality follows similarly to the previous case.
Altogether, we conclude that \( \frac{G_{r}(t')}{t_{r}(t')} \leq \frac{G((1+\eta)s)}{t_{r}(s)} \leq \sqrt{1+6\eta} \leq 1 + 3\eta. \)

We are now ready to complete the proof of distortion to the Gaussian (for the entire set \( S \)). Similar to the derivation of Equation (8), we derive \( \|x' - y'\| \leq (1 + 2\varepsilon)\|x - y\| \) and by Lemma 3.3 we get \( G_{r}(\|x - y\|) \leq (1 + 6\varepsilon)G_{r}(\|x - y\|) \). Together with Equation (8) and the upper bound for net points (Equation (7)), we obtain

\[
\|\varphi(x) - \varphi(y)\| \leq \|\varphi(x') - \varphi(y')\| + 22\varepsilon G_{r}(\|x' - y'\|) \\
\leq (1 + 18\varepsilon + 22\varepsilon)G_{r}(\|x' - y'\|) \\
\leq (1 + 40\varepsilon)(1 + 6\varepsilon)G_{r}(\|x - y\|).
\]

The other direction is analogous. By (8) we have \( \|x - y\| \leq (1 + 4\varepsilon)\|x' - y'\| \) and by Lemma 3.3 we get \( G_{r}(\|x - y\|) \leq (1 + 12\varepsilon)G_{r}(\|x' - y'\|) \). Together with (8) and the lower bound for net points (Equation (7)), we obtain

\[
\|\varphi(x) - \varphi(y)\| \geq \|\varphi(x') - \varphi(y')\| - 22\varepsilon G_{r}(\|x' - y'\|) \\
\geq (1 - \varepsilon - 22\varepsilon)G_{r}(\|x' - y'\|) \\
\geq (1 - 23\varepsilon)(1 - 12\varepsilon)G_{r}(\|x - y\|).
\]

### 3.3 Snowflake Embedding

We now use Theorem 3.1 (the single scale embedding) to prove Theorem 1.2 (embedding for \( d^{\alpha} \)). 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 \( p = \lfloor 6\log_{1+\varepsilon}\left(\frac{1}{\varepsilon}\right) \rfloor = O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}) \), 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 \( \varphi_{i} : S \to \ell_{2}^{k} \) be the embedding that achieves the bounds of Theorem 3.3 for \( S \) and \( \varepsilon \) with respect to parameters \( r = (1 + \varepsilon)^{i} \) and \( \delta = (1 + \varepsilon)^{-p/2} = \Theta(\varepsilon^{3}) \). Notice that each \( \varphi_{i} \) has target dimension \( k = O(\varepsilon^{-3} \log^{2} \lambda) \).

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

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

\[
\Phi_{j} = \sum_{i \in I: \ i \equiv_{p} j} \frac{\varphi_{i}}{(1 + \varepsilon)^{i/2}}.
\]

For \( M = M(\varepsilon) > 0 \) that will be defined later, our final embedding is \( \Phi/\sqrt{M} : S \to \ell_{2}^{pk} \), which has target dimension \( pk \leq O(\varepsilon^{-4} \log^{2} \lambda) \), as required (for \( \alpha = 1/2 \)). It thus remains to prove the distortion bound. The key idea is that in each \( \Phi_{j} \), most of the contribution to \( \|\Phi_{j}(x) - \Phi_{j}(y)\| \) comes from a single \( \varphi_{i} \), and we can further estimate that contribution quite accurately: It behaves roughly like \( (1 + \varepsilon)^{i/2}\|x - y\| \). We need the following lemma.
Lemma 3.4. Let $\Phi : S \to \ell^p_2$ 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 $p$ (namely $A = \{a, a+1, \ldots, a+p-1\}$),

$$\|\Phi(x) - \Phi(y)\|^2 \leq \sum_{i \in A} \left( B_i + \sum_{i' \in I \setminus A : i' \equiv_p i} B_{i'} \right)^2$$

$$\|\Phi(x) - \Phi(y)\|^2 \geq \sum_{i \in A} \left( \max \left\{ 0, B_i - \sum_{i' \in I \setminus A : i' \equiv_p i} B_{i'} \right\} \right)^2$$

Proof. By construction,

$$\|\Phi(x) - \Phi(y)\|^2 = \sum_{j \in [p]} \|\Phi_j(x) - \Phi_j(y)\|^2 = \sum_{i \in A} \sum_{i' \equiv_p i} \varphi_{i'}(x) - \varphi_{i'}(y)^2 \|1 + \varepsilon\|^2.$$

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 $v_1, \ldots, v_s \in \ell^p_2$, namely, $\|\sum_i v_i\| \leq \sum_i |v_i|$. The second inequality follows similarly by separating the term for $i' = i$ from the rest, and applying the following triangle inequality for vectors $u, v_1, \ldots, v_s \in \ell^p_2$, namely, $\|u + \sum_i v_i\| \geq \max\{0, \|u\| - \sum_i |v_i|\}$. □

The proof of Theorem 1.2 proceeds by demonstrating that, for an appropriate choice of $A$ (meaning $p$ and $a$), the leading term in the above summation ($B_i$ for $i \in A$) dominates the sum of the other terms of the summation (terms $B_i$ for $i' \in I \setminus A : i' \equiv_p i$). 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 3.4. To this end, let $A = \{i^* - p/2 + 1, \ldots, i^* + p/2\}$ and consider $i \in A$. Observe that

$$\delta \leq (1 + \varepsilon)^{-p/2} \leq (1 + \varepsilon)^{i^* - i} \leq \frac{\|x - y\|(1 + \varepsilon)^i}{(1 + \varepsilon)^i} \leq (1 + \varepsilon)^{i^* + 1 - i} \leq (1 + \varepsilon)^{p/2} \leq \frac{1}{\delta},$$

hence we can apply Theorem 3.1(b) to obtain

$$\frac{1}{1 + \varepsilon} \leq \frac{\|\varphi_i(x) - \varphi_i(y)\|}{G_{\varepsilon}(1 + \varepsilon)^{i/2} \|x - y\|} \leq 1. \quad (11)$$

Combining this with the monotonicity of $G_r$ and Lemma 3.3, and noting that $G_r(1) > \frac{1 + \varepsilon}{2}$ when $\varepsilon < \frac{1}{4}$, we further obtain

$$\frac{\|\varphi_i(x) - \varphi_i(y)\|}{(1 + \varepsilon)^{i/2}} \geq \frac{(1 + \varepsilon)^{i-1}G(1)}{(1 + \varepsilon)^{i/2}} \geq \frac{1}{2}(1 + \varepsilon)^{i/2}. \quad (12)$$

By Theorem 3.1(a) and (c), for all $i' \in I$,

$$\|\varphi_{i'}(x) - \varphi_{i'}(y)\| \leq \min\{\|x - y\|, (1 + \varepsilon)^{i'}\}.$$

and thus

$$\sum_{i' \in I \setminus A : i' \equiv_p i} \frac{\|\varphi_{i'}(x) - \varphi_{i'}(y)\|}{(1 + \varepsilon)^{i'/2}} \leq \sum_{i' < i : i' \equiv_p i} \frac{\|\varphi_{i'}(x) - \varphi_{i'}(y)\|}{(1 + \varepsilon)^{i'/2}} + \sum_{i' > i : i' \equiv_p i} \frac{\|\varphi_{i'}(x) - \varphi_{i'}(y)\|}{(1 + \varepsilon)^{i'/2}} \leq \sum_{i' < i : i' \equiv_p i} (1 + \varepsilon)^{i'/2} + \sum_{i' > i : i' \equiv p} \frac{\|x - y\|}{(1 + \varepsilon)^{i'/2}}.$$
Recalling that a geometric series with ratio less than $\frac{1}{2}$ sums to less than twice the largest term,

$$\sum_{i' \in I \setminus A; i' \equiv_p i} \frac{\|\varphi_{i'}(x) - \varphi_{i'}(y)\|}{(1 + \varepsilon)^{p/2}} \leq 2(1 + \varepsilon)^{(i - p)/2} + 2\|x - y\|(1 + \varepsilon)^{(i + p)/2}$$

$$\leq 2(1 + \varepsilon)^{i/2}[(1 + \varepsilon)^{-p/2} + \frac{1}{2}(1 + \varepsilon)^{1-p/2}]$$

$$\leq \varepsilon(1 + \varepsilon)^{i/2}.$$  

Observe that the last bound is at most $2\varepsilon$ times $\left[\frac{1}{2}\right]$. Using this information and plugging $(11)$ into Lemma $3.3$, we obtain

$$\|\Phi(x) - \Phi(y)\|^2 \geq \sum_{i \in A} \left(1 + \frac{2\varepsilon}{1 + \varepsilon} \cdot G((1 + \varepsilon)^i \cdot \|x - y\|)\right)^2$$

$$\geq \sum_{-p/2 < i - i^* \leq p/2} \left(1 + \frac{2\varepsilon}{1 + \varepsilon} \cdot (1 + \varepsilon)^i \cdot G((1 + \varepsilon)^{i^* - i})\right)^2$$

$$\geq (1 - 2\varepsilon)^2(1 + \varepsilon)^{2i^* - 2} \sum_{b: -p/2 < b \leq p/2} \left((1 + \varepsilon)^b \cdot G((1 + \varepsilon)^{-b})\right)^2,$$

and similarly, using also Lemma $3.3$

$$\|\Phi(x) - \Phi(y)\|^2 \leq \sum_{i \in A} \left(1 + \frac{2\varepsilon}{1 + \varepsilon} \cdot G((1 + \varepsilon)^i \cdot \|x - y\|)\right)^2$$

$$\leq \sum_{-p/2 < i - i^* \leq p/2} \left((1 + 2\varepsilon)(1 + 4\varepsilon)(1 + \varepsilon)^i \cdot G((1 + \varepsilon)^{i^* - i})\right)^2$$

$$\leq (1 + 4\varepsilon)^4(1 + \varepsilon)^{2i^*} \sum_{b: -p/2 < b \leq p/2} \left((1 + \varepsilon)^b \cdot G((1 + \varepsilon)^{-b})\right)^2.$$

Setting $M = \sum_{b: -p/2 < b \leq p/2} \left((1 + \varepsilon)^b \cdot G((1 + \varepsilon)^{-b})\right)^2$, which clearly depends only on $\varepsilon$ (and is in particular independent of $x, y$), we combine the last two estimates to obtain

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

We conclude that the final embedding $\Phi/\sqrt{M}$ 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 $p = \frac{3}{1-\alpha} \left[\log_1 + \varepsilon \left(\frac{1}{2}\right)\right]$ and $\delta = (1 + \varepsilon)^{-p(1-\alpha)} = \Theta(\varepsilon^3)$. As before, $\varphi_i : S \to \ell_2^{pk}$ is the embedding that achieves the bounds of Theorem $1.3$ for $S$, and $\Phi : S \to \ell_2^{pk}$ is defined by the direct sum $\Phi = \bigoplus_{j \in \mathbb{N}_p} \Phi_j$, where each $\Phi_j : S \to \ell_2$ is given by

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

The final embedding is $\Phi/\sqrt{M} : S \to \ell_2^{pk}$ (for the same $M$ as above), which has target dimension $pk \leq \tilde{O}(\varepsilon^{-4}(1 - \alpha)^{-1} \log^2 \lambda)$, as required.
We need to make only small changes to the preceding proof of distortion: In the statement and proof of Lemma 3.4, the dividing terms \((1 + \varepsilon)^{1/2}\) and \((1 + \varepsilon)^{\theta/2}\) are replaced by \((1 + \varepsilon)^{(1-\alpha)}\) and \((1 + \varepsilon)^i\), respectively. The same substitution is made to modify Equation 12 and obtain \(\frac{\|\phi(x) - \phi(y)\|}{(1+\varepsilon)^{(1-\alpha)}} \geq \frac{1}{2} (1 + \varepsilon)^{(1-\alpha)}\), and to the subsequent geometric series, from which we derive \(\sum_{i \in \ell^1 \setminus \{0\}} \|\phi_i(x) - \phi_i(y)\| \leq \varepsilon (1 + \varepsilon)^{(1-\alpha)}\). (Note that the geometric series still has a ratio of less than 1/2, due to the increase in value of \(p\).) No other changes to the proof are necessary, and this completes the proof of Theorem 1.2.

4 Extension to Other \(\ell_p\) Spaces

We briefly explain how our results and techniques can be extended to other \(\ell_p\) spaces. For concreteness, we consider only two important spaces, namely \(\ell_1\) and \(\ell_\infty\). A number of key tools used in our previous embeddings are specific to \(\ell_2\), for example the JL-Lemma, the Gaussian transform, and the Kirszbraun theorem, and we must therefore find suitable replacements for these tools. Note however that there is no Lipschitz extension theorem for \(\ell_1\).

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 \(\Phi : S \to \ell_p^k\) satisfying
\[
1 \leq \frac{\|\Phi(x) - \Phi(y)\|_p}{\|x - y\|_\alpha^p} \leq 1 + \varepsilon, \quad \forall x, y \in S;
\]
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 \(\ell_2\) (Theorem 3.1), coupled with an application of Assouad’s technique, were sufficient to prove Theorem 1.2. Similarly, single scale embeddings for \(\ell_1\) and \(\ell_\infty\), coupled with a standard application of Assouad’s technique, are sufficient to prove Theorem 4.1. We present single scale embeddings for \(\ell_1\) and \(\ell_\infty\) below, and Theorem 4.1 then follows easily.

4.1 Single Scale Embedding for \(\ell_1\)

We can extend Theorem 3.1 to \(\ell_1\) spaces as follows. For \(r > 0\) define \(L_r : \mathbb{R} \to \mathbb{R}\), 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 \(\varphi : S \to \ell_1^k\) for \(k = \exp\{\lambda^{O(\log(1/\varepsilon) + \log \log \lambda)}\}\) 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 Sketch. We would like to utilize the framework designed for \( \ell_2 \) 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 \( \ell_1 \) norm, since there is no \( \ell_1 \)-analogue of the Kirszbraun 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 \( \ell_2 \), the same approach of discarding step 7 can be achieved by applying the Kirszbraun 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 \( C \in P_l \) 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 \( g_C : C \to \ell_1 \) 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 \( \Psi : g_C(C) \to \ell_{1}' \) which satisfies: (i) \( \Psi \) is 1-Lipschitz on the entire cluster \( g_C(C) \); and (ii) \( Psi \) 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 \( \ell_1 \). However, we describe below one simple dimension reduction for \( \ell_1 \) 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 \( \Psi \) cannot be linear, since in the worst case such a linear map requires dimension \( k = \Omega(|C \cap N|) \) [FJS97, Corollary 12.A].

Construct \( \Psi \) 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, \( g_C(x) \to \sum_A \alpha_A 1_A(x) \) is an isometric embedding of \( g_C(C) \) into \( \ell_1 \). Let \( \Psi \) have one coordinate for every subset \( B \subset C \cap N \); this coordinate is given by \( g_C(x) \to \sum_A \alpha_A 1_A(x) \). In words, we add together coordinates whenever they correspond to different \( A \) but have the same \( A \cap (C \cap N) \). Observe that \( \Psi \) 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 \( k' = 2^{\Omega(|C \cap N|)} \) 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 \( h_C(x) = \min_{y \in S \cap C} \| x - y \|_1 \).
- Step 6: There is only a minor change to the scaling factor, namely \( \varphi = m^{-1} \sum_{r=1}^n \varphi_i \).
The rest of the proof is quite similar to the one presented for $\ell_2$, and the final dimension obtained is $mk' = O(\varepsilon^{-1} \log \lambda \log \log \lambda) \cdot \exp(\lambda^{O(\log(\varepsilon^{-1}\delta^{-2}\log \log \lambda))) = \exp\{\lambda^{O(\log(1/\delta)+\log \log \lambda)}\}. \square$

4.2 Single Scale Embedding for $\ell_\infty$

We can also extend Theorem 4.1 to $\ell_\infty$ spaces as follows. For $r > 0$ define $T_r : \mathbb{R} \to \mathbb{R}$, 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 $\varphi : S \to \ell_k^{\infty}$ for $k = \lambda^{O(\log(1/\varepsilon\delta)+\log \log \lambda)}$, satisfying:

(a). Lipschitz: $\|\psi(x) - \psi(y)\|_\infty \leq \|x - y\|_\infty$ 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 $\delta r \leq \|x - y\|_{\infty}$:

$$T_r(\|x - y\|_\infty) \leq \|\psi(x) - \psi(y)\|_\infty \leq T_r(\|x - y\|_\infty)$$

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

**Proof Sketch.** The proof is quite similar to that of Theorem 4.1, 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, $\Psi : g_C(C) \to \ell_{k'}^{\infty}$ 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 $\Psi$ 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 $\ell_2$ 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 $mk' = O(\varepsilon^{-1} \log \lambda \log \log \lambda) \cdot \max\{\lambda^{O(\log(1/\varepsilon\delta)+\log \log \lambda)}\} = \max\{\lambda^{O(\log(1/\delta)+\log \log \lambda)}\}$.

The lower bound on $\|\psi(x) - \psi(y)\|_\infty$ for pair $x, y$ follows when $x$ is padded (case 1 in the distortion to the Gaussian analysis for $\ell_2$), where we have $\|\psi(x) - \psi(y)\|_\infty = T_r(\|x - y\|_{\infty})$; 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’), $\|\psi(x) - \psi(y)\|_\infty \leq \max\{2\delta \|x - y\|_{\infty}, \delta \|x - y\|_{\infty}, T_r(\|x - y\|_{\infty}) + T_r(\|x - y\|_{\infty})\} \leq 2\delta \|x - y\|_{\infty} + T_r(\|x - y\|_{\infty})$. 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. $r < \|x - y\|_{\infty} \leq \frac{r}{2\sqrt{\delta}}$. 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. \square
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 \( \ell_2 \) 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 \( \ell^k_2 \) (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 \( \varepsilon/k \). We have:

**Lemma 5.1.** Every finite subset \( \ell_2 \) 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 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 \( \ell_2 \)-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 \( \ell_2 \) and \( \ell_2 \)-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 \( \ell_2 \) min-sum \( k \)-clustering for \( S \), for \( k = O(1) \), can be computed

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

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

Acknowledgments

The authors thank Assaf Naor and Gideon Schechtman for useful discussions and references.
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] Ittai Abraham, Yair Bartal, and Ofer Neiman. On low dimensional local embeddings. In SODA ’09: Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 875–884, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

[Alo03] 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 $l_r$-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.

[BRS07] Y. Bartal, B. Recht, and L. Schulman. A Nash-type dimensionality reduction for discrete subsets of $L_2$. Manuscript, available at http://www.ist.caltech.edu/~brecht/publications.html, 2007.

[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.

[JM84] 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 $l^p_r$. 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, New York, NY, USA, 2000. ACM.

[Tal90] M. Talagrand. Embedding subspaces of $L_1$ into $l^N_1$. 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 $l^N_p$. 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.