Volume distortion for subsets of Euclidean spaces

James R. Lee*
University of Washington

Abstract

In [Rao 1999], it is shown that every n-point Euclidean metric with polynomial aspect ratio admits a Euclidean embedding with k-dimensional distortion bounded by $O(\sqrt{\log n \log k})$, a result which is tight for constant values of k. We show that this holds without any assumption on the aspect ratio, and give an improved bound of $O(\sqrt{\log n}(\log k)^{1/4})$. Our main result is an upper bound of $O(\sqrt{\log n}\log\log n)$ independent of the value of k, nearly resolving the main open questions of [Dunagan-Vempala 2001] and [Krauthgamer-Linial-Magen 2004]. The best previous bound was $O(\log n)$, and our bound is nearly tight, as even the 2-dimensional volume distortion of an n-vertex path is $\Omega(\sqrt{\log n})$.

1 Introduction

In the geometry of finite metric spaces, bi-Lipschitz mappings between pairs of metric spaces play a central role. Given metric spaces $(X, d_X), (Y, d_Y)$, and a map $f : X \to Y$, one defines the *Lipschitz norm* of f by

$$\|f\|_{\mathrm{Lip}} = \sup_{x \neq y \in X} \frac{d_Y(f(x), f(y))}{d_X(x, y)},$$

i.e. the maximum amount by which distances are expanded under f. If f is injective, we define the *distortion* by $\mathsf{Dst}(f) = \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$. If $\|f\|_{\mathrm{Lip}} \leq 1$, we say that f *is non-expansive*. In the present paper, we will be concerned primarily with the case when $Y = L_2$ and X is finite. In this case, one defines $c_2(X) = \inf_{f:X \hookrightarrow L_2} \mathsf{Dst}(f)$, where the infimum is over all injective maps from X into a Hilbert space. This quantity is referred to as the *Euclidean distortion of* X. In his study of the graph bandwidth problem, Feige [Fei00] introduced a higher-dimensional analogue of distortion for maps into Euclidean spaces which is useful for controlling the behavior of finite subsets under random projections (for a nice discussion of this and its application to bandwidth, see [Mat02, Ch. 15]).

*This research was done while the author was a postdoctoral fellow at the Institute for Advanced Study, Princeton, NJ.

For a k-point subset $T \subseteq \mathbb{R}^{k-1}$, let $\mathsf{conv}(T)$ denote the convex hull of T, and define $\mathsf{Evol}(T) = \mathrm{vol}_{k-1}(\mathsf{conv}(T))$, where vol_{k-1} denotes the $(k-1)$-dimensional Lebesgue measure The *volume* of an arbitrary k-point metric space S, denoted $\mathsf{Vol}(S)$, is then defined as the supremum of $\mathsf{Evol}(\Psi(S))$ over all non-expansive maps $\Psi : S \to \mathbb{R}^{k-1}$. Given a non-expansive map $f : X \to L_2$, we define the $(k-1)$-*dimensional distortion of* f by

$$\mathsf{Dst}^{k-1}(f) = \sup_{S \subseteq X : |S|=k} \left[\frac{\mathsf{Vol}(S)}{\mathsf{Evol}(f(S))} \right]^{\frac{1}{k-1}}.$$

In words, we measure how well f achieves the maximal Euclidean volume *simultaneously for all subsets of size* k. Observe that the 1-dimensional distortion corresponds with the standard notion, i.e. $\mathsf{Dst}^1(f) = \mathsf{Dst}(f)$, which considers only pairs of points. For larger values of k, Dst^k measures the distortion of higher-order structures in X. We define $c_2^k(X) = \inf_{f:X \to L_2} \mathsf{Dst}^k(f)$, where the infimum is over all injective, non-expansive maps f. This quantity is called the k-*dimensional volume distortion of* X.

For a non-expansive map $f : X \to L_2$, we define the *rigidity* of f, written $\mathsf{rigidity}(f)$ as the minimum value R such that the following holds: For every $x \in X$, $Y \subseteq X$, we have

$$\mathsf{dist}_2\left(f(x), \mathrm{span}\{f(y)\}_{y \in Y}\right) \geq \frac{d(x,Y)}{R}.$$

We call such a map R-*rigid*. Finally, define $r_2(X) = \inf\{\mathsf{rigidity}(f) \mid f : X \hookrightarrow L_2\}$. It is an easy (but non-trivial) observation that $c_2^k(X) \leq r_2(X)$ for any $k \leq |X|$ (see, e.g. [KLM04, Sec. 2]).

Previous work. The asymptotics of $c_2^k(X)$ and $r_2(X)$ as a function of $n = |X|$ are well-studied, because of their intrinsic geometric appeal, and the application of bounds on $c_2^k(X)$ to graph-theoretic layout problems [Fei00, Vem98, DV01]. The first bounds, given by Feige [Fei00], were based on a new analysis of Bourgain's embedding [Bou85] and showed that $c_2^k(X) \leq O(\log n + \sqrt{k \log k \log n})$ for any n-point metric space X. Later, Rao [Rao99] showed that $c_2^k(X) \leq O(\log n)^{3/2}$, for any $1 \leq k \leq n$. (Rao's paper does not contain this result, but as observed by A. Gupta, it follows from his work in combination with known metric partitioning techniques [LS93, Bar96]). In fact, using Rao's technique one obtains the stronger bound $r_2(X) \leq O(\log n)^{3/2}$. Finally, Krauthgamer, et. al. [KLMN05] gave the optimal bound $r_2(X) \leq O(\log n)$. (We remark that this bound is a special case of our analysis for the Euclidean subsets, see Corollary 3.6.) The matching lower bound (based on expander graphs) is proved in [KLM04, Sec. 3.5].

The Euclidean case. One natural case, which arises in the analysis of a semi-definite program for bandwidth [DV01], occurs when X is an n-point subset of some Euclidean space. The rounding algorithm of [DV01] proceeds in three steps (following Feige's original algorithm [Fei00]):

1. Solve an SDP for the graph bandwidth problem, applied to a graph on n vertices. This yields an n-point subset $S \subseteq \mathbb{R}^n$.

2. Embed the subset S back into $\mathbb{R}^n$ using an embedding f with small volume distortion.

3. Project the subset $f(S) \subseteq \mathbb{R}^n$ onto a random line, and output the induced linear ordering.

Step (2) is a pre-processing step used to ensure that the set of points behaves well under random projection. We refer to [Vem98, Fei00, DV01, Mat02] for details about graph bandwidth, and how volume distortion relates to random projections.

For the Euclidean case, Rao [Rao99] exhibited a bound of $c_2^k(X) \leq O(\sqrt{\log n \log k})$, with the caveat that the ratio of the maximum to minimum pairwise distance in X must be bounded by some polynomial in n. Furthermore, this bound is essentially tight for constant values of k, as exhibited independently by Dunagan and Vempala [DV01], and Krauthgamer, Linial, and Magen [KLM04]: If P_n is the path metric on n-points, then $c_2^2(P_n) = \Omega(\sqrt{\log n})$. In those papers, it is asked whether $c_2^k(X) \leq O(\sqrt{\log n})$ for every n-point subset of some Euclidean space and every value $1 \leq k \leq n$. In the present work, we nearly resolve this open problem.

Theorem 1.1. *For any n-point subset $X \subseteq \mathbb{R}^n$, we have $r_2(X) = O(\sqrt{\log n} \log\log n)$. In particular, $c_2^k(X) = O(\sqrt{\log n} \log\log n)$ for any $1 \leq k \leq n$.*

Furthermore, for constant values of k, we achieve the optimal distortion without requiring bounds on the ratio of maximum to minimum distance in X.

Theorem 1.2. *For any n-point subset $X \subseteq \mathbb{R}^n$, and $1 \leq k \leq n$, $c_2^k(X) = O(\sqrt{\log n}\,(\log k)^{\frac{1}{4}})$.*

Clearly this result is dominated by the preceding theorem for $k \gg 2^{(\log\log n)^4}$. These theorems are proved in sections 4.1 and 4.2, respectively. We remark that every step of the proofs can be made algorithmic (i.e. can be carried out in time polynomial in n) in a straightforward way. Using the latter theorem in the algorithm of [DV01] yields a marginal improvement of $O(\log\log n)^{1/4}$ to the best-known approximation ratio for graph bandwidth. We obtain only this small improvement because in their analysis, one takes $k = \Theta(\log n)$.

Our approach makes a connection between the value of $r_2(X)$ and a seemingly simpler parameter which we now define.

Definition 1.3. *For a number $d \in \mathbb{N}$, let $\mathsf{h}(d)$ be the smallest value such that there exists a non-expansive map $F_d : \mathbb{R}^d \to L_2$ satisfying the following conditions for every $x \in \mathbb{R}^d$.*

1. $\|F_d(x)\|_2 \leq 30/\mathsf{h}(d)$.

2. *If $B(x,1)$ is the ball of radius 1 around x, then*

$$\mathsf{dist}_2\left(F_d(x), \mathrm{span}\{F_d(y)\}_{y \in \mathbb{R}^d \setminus B(x,1)}\right) \geq \frac{1}{\mathsf{h}(d)},$$

where $\mathsf{dist}_2(x,S) = \inf_{y \in S} \|x-y\|_2$ for a subset $S \subseteq L_2$.

The value 30 is somewhat arbitrary (as any large enough constant would suffice).

Observe that $\mathsf{h}(\cdot)$ is monotone in the sense that $\mathsf{h}(d+1) \geq \mathsf{h}(d)$ for all $d \geq 1$, since $\mathbb{R}^d \subseteq \mathbb{R}^{d+1}$ with the canonical identification. The connection between $\mathsf{h}(d)$ and $r_2(X)$ is contained in the following theorem.

Theorem 1.4. *If $\mathsf{h}(d) \leq O(d^{\varepsilon})$ for some $\varepsilon \geq \frac{1}{2}$, then*

$$\sup\left\{r_2(X) : X \subseteq L_2, |X| = n\right\} \leq O(\log n)^{\varepsilon} \log\log n.$$

At the highest level, the proof of Theorem 1.4 constructs a rigid embedding for $X \subseteq \mathbb{R}^n$ by decomposing it into various subsets (these subsets are formed from a combination of random partitioning and variable-rate random sampling), projecting such a subset into a low-dimensional subspace, and then applying a variant of an appropriate map $F_d : \mathbb{R}^d \to L_2$ from the family defined above. The different embeddings are then glued together using smooth partitions of unity; see Section 1.2 for a more detailed proof overview.

Our second contribution, which completes the proof, is a bound on the value of $\mathsf{h}(d)$.

Theorem 1.5. $\mathsf{h}(d) = O(\sqrt{d})$.

Theorem 1.5 is proved in Section 2, while the transference argument of Theorem 1.4 combines results from Sections 2 and 3, and is completed in Section 4. After introducing some preliminaries, we present an overview of the proof in Section 1.2. Finally, in Section 4.3, we outline an approach which might achieve the optimal bound of $O(\sqrt{\log n})$.

1.1 Preliminaries

For a metric space (X, d) and a subset $S \subseteq X$, we write $N_\delta(S) = \{x \in X : d(x, S) \leq \delta\}$. We write $B(x, r) = \{x \in X : d(x, y) \leq r\}$ for the closed ball of radius R about x, and $A(x, r_1, r_2) = B(x, r_2) \setminus B(x, r_1)$.

Hilbert spaces and random mappings. Throughout the paper, L_2 represents a separable, infinite-dimensional Hilbert space. Given a Hilbert space Z, and two maps $f, g : X \to Z$, we define the map $f \oplus g : X \to (Z \oplus Z)$ by $(f \oplus g)(x) = (f(x), g(x))$. We extend this definition to more than two maps (and even countably infinite sums) in the obvious way. If $Z = L_2$, we will routinely view $f \oplus g$ as a function taking values in L_2 (under some canonical isomorphism). Of course if X is finite (as will usually be the case), one can assume that $Z = \mathbb{R}^{|X|-1}$.

Often, it will be useful to construct embeddings into Hilbert spaces of random variables. Given a probability space $(\Omega, \Pr)$, we let $L_2(\Omega, \Pr)$ denote the space of all L_2-valued random variables defined (and measurable) with respect to $(\Omega, \Pr)$. Given such an $A \in L_2(\Omega, \Pr)$, one has $\|A\|_{L_2(\Omega,\Pr)} = \sqrt{\mathbb{E}_\Omega \|A\|_2^2}$. If X is finite, then one can often convert a mapping $f : X \to L_2(\Omega, \Pr)$ to a map $f' : X \to \mathbb{R}^d$ by randomly sampling coordinates from the distribution of the embedding. In all of our constructions, $\mathrm{poly}(k, \log|X|)$ random samples suffice when trying to preserve the k-dimensional volume distortion achieved by f.

Decomposability. We now recall the notion of *padded decomposability*. Given a partition P of X and $x \in X$ we denote by $P(x) \in P$ the unique element of P to which x belongs.

In what follows we sometimes refer to $P(x)$ as the *cluster* of x. Following [KLMN05] we define the *modulus of padded decomposability* of X, denoted α_X, as the least constant $\alpha > 0$ such that for every $\tau > 0$ there is a distribution ν over partitions of X with the following properties.

1. For all $P \in \mathrm{supp}(\nu)$ and all $C \in P$ we have that $\mathrm{diam}(C) < \tau$.
2. For every $x \in X$ we have that

$$\nu\{P : \ B(x, \tau/\alpha) \subseteq P(x)\} \geq \tfrac{1}{2}.$$

The results of [LS93, Bar96] imply that $\alpha_X = O(\log |X|)$, and this will be used in our proof.

1.2 Proof overview

We recall that our goal is to prove that for every n-point subset $X \subseteq \mathbb{R}^n$, we have $r_2(X) \leq O(\sqrt{\log n} \log\log n)$. Our approach breaks into three steps.

1. Handling a single scale.
We show that there exists a constant $C > 0$ such that for every finite subset $S \subseteq L_2$, and every $\tau \geq 0$, there exists a non-expansive map $f_{S,\tau} : S \to L_2$ such that for every $x \in S$,

$$\mathsf{dist}_2\left(f_{S,\tau}(x), \mathrm{span}\left\{f_{S,\tau}(y)\right\}_{y \in S \setminus B(x,\tau)}\right) \geq \frac{\tau}{C\sqrt{\log |S|}}. \tag{1}$$

This is done by first obtaining a bi-Lipschitz projection of S into a $k = O(\log |S|)$-dimensional Euclidean space [JL84]. Once in $\mathbb{R}^k$, we ignore the set S, and concentrate on proving (1) for all $x \in \mathbb{R}^k$. This is crucial since our bound on the Lipschitz constant of $f_{S,\tau}$ will depend on the fact that $\mathbb{R}^k$ is a geodesic space (or at least "coarsely" geodesic at scales smaller than τ/k.).

Our construction then proceeds using a method of "local random projections" introduced by Rao [Rao99]. Essentially, using Rao's method, we are able to enjoy the benefits of random projection for close pairs of points, while maintaining complete independence for pairs $x, y \in \mathbb{R}^k$ with $\|x - y\|_2 \gg \tau$. This independence is necessary to achieve the strong lower bound required by (1). We remark that Rao's analysis is only able to obtain the bound (1) with $\sqrt{\log |S|}$ replaced by $(\log |S|)^{\frac{3}{4}}$.

2. Passing to a dependence on the local growth ratio.
The next goal is to obtain, for every $\tau \geq 0$, a non-expansive map $\varphi_\tau : X \to L_2$ which satisfies

$$\mathsf{dist}_2\left(\varphi_\tau(x), \mathrm{span}\left\{\varphi_\tau(y)\right\}_{y \in X \setminus B(x,\tau)}\right) \geq \frac{\tau}{C\sqrt{\log \frac{|B(x,\tau)|}{|B(x,\tau/4)|}}}. \tag{2}$$

The actual lower bound we obtain is slightly weaker (this is one source of the extra $O(\log\log n)$ term in our result).

This is done by extending the framework of [ALN08] for smoothly piecing together global single-scale maps from maps defined only on small subsets (these subsets are formed out of a combination of random partitioning and random sampling). There are a number of difficulties involved in extending this to the domain of volume distortion (as opposed to one-dimensional distortion). In applying the method of [ALN08], we are confronted with the problem of extending non-expansive maps $f : S \to L_2$ from a subset $S \subseteq X$ to maps $\tilde{f} : X \to L_2$ which are non-expansive on the whole space. One difference from [ALN08] is our use of Kirszbraun's extension theorem [Kir34] for extending Lipschitz maps between Hilbert spaces. This is necessary in our setting because the maps produced by step (1) above are not of Fréchet type, and thus extension is a non-trivial issue.

A more serious difficulty arises in the process of extension: We must not only maintain a Lipschitz bound (i.e. an upper bound on $\|\varphi_t\|_{\mathrm{Lip}}$), but we must also extend the lower bound of (1) to apply to the span of larger sets of points. (Observe that in (2) we consider $y \in X \setminus B(x, \tau)$, while in (1), we have only a guarantee for $y \in S \setminus B(x, \tau)$.) Here, we make use of the power of rigid embeddings: they behave particularly nicely under partitions of unity. By a partition of unity on a metric space (X, d), we mean a family of maps $\{\rho_t : X \to [0, 1]\}$ such that for every $x \in X$, $\sum_t \rho_t(x) = 1$. Such a family is distinguished from an arbitrary set of weight functions on X by the fact that we usually require some smoothness condition from each $\rho_t : X \to [0, 1]$. In our case, all partitions of unity will be Lipschitz, and we will care greatly about the norms $\|\rho_t\|_{\mathrm{Lip}}$.

In order to apply the techniques of [Lee05, ALN08], given a function $\phi : X \to L_2$, we are often confronted with the problem of analyzing the product function $g(x) = \rho_t(x) \cdot \phi(x)$ (this is the map ϕ "localized" under the partition of unity ρ_t). Bounds on $\|g\|_{\mathrm{Lip}}$ are controlled in the usual way (the chain rule) via the quantities $\sup_{x \in X} \|\phi(x)\|_2$, $\|\phi\|_{\mathrm{Lip}}$, and $\|\rho_t\|_{\mathrm{Lip}}$. However, providing good control on the lower bounds becomes more delicate.

Fortunately, the simple inequality (where $\{c_y\} \subseteq \mathbb{R}$ are real constants),

$$\left\| \phi(x) - \sum_{y \in Y} c_y g(y) \right\|_2 = \left\| \phi(x) - \sum_{y \in Y} c_y \rho_t(y) \phi(y) \right\|_2 \geq \mathsf{dist}_2 \left(\phi(x), \mathrm{span}\{\phi(y)\}_{y \in Y} \right) \qquad (3)$$

allows us to freely use well-behaved partitions of unity since the $\{\rho_t(y)\}$ multipliers are absorbed into the span.

In particular, this allows us to "dampen" the map φ_τ away from various subsets $S \subseteq X$, while absorbing this dampening into the span (see Claim 3.7).

3. Gluing for volume distortion. The last step in the proof is to establish an analogue of the scale gluing methodology of [Lee05, KLMN05] for embeddings with small volume distortion, as opposed to bi-Lipschitz embeddings. This is taken up in Section 3. Since our embeddings are not of Fréchet type, our starting point is the author's work [Lee05], based on combining single-scale embeddings under partitions of unity. We are able to adapt those techniques to our setting by again using the observation (3) above (see Theorem 3.5).

This allows us to transform the ensemble of maps $\{\varphi_\tau\}_{\tau \geq 0}$ from (2) into a genuine rigid embedding that simultaneously handles all the scales.

2 Local random projections

This section contains most of our results specifically about the geometry of finite-dimensional Euclidean spaces. First, we prove Theorem 1.5, yielding the estimate $\mathsf{h}(d) = O(\sqrt{d})$. In fact, for simplicity, we will only prove it for compact subsets $Z \subseteq \mathbb{R}^d$. The general case follows by a standard argument which we omit for the sake of simplicity. We remark that the compact case is all that is required for applications throughout the paper.

Lemma 2.1. *There exists a constant $\beta \in (0,1)$ such that the following holds. Let $\tau \geq 0, d \in \mathbb{N}$ be given, and let $Z \subseteq \mathbb{R}^d$ be a compact subset. Then there exists a map $F : \mathbb{R}^d \to L_2$ such that*

1. $||F||_{\mathrm{Lip}} \leq 1$,
2. $\|F(x)\|_2 \leq 28\beta\tau/\sqrt{d}$ *for all* $x \in Z$,
3. *For all $x \in Z$, denoting $C(x) = Z \setminus B(x,\tau)$,*

$$\mathsf{dist}_2\left(F(x), \mathrm{span}\{F(y)\}_{y \in C(x)}\right) \geq \frac{\beta\tau}{\sqrt{d}}.$$

Proof. Without loss of generality, we assume that $d \geq 3$. We may assume that Z is convex by simply replacing Z with the closure of its convex hull. Clearly we may also assume that $N_\tau(Z) \subseteq \mathcal{B}$ for some closed ball $\mathcal{B}$ of radius R, for some sufficiently large value $R > 0$. Let vol_d denote the d-dimensional Lebesgue measure, normalized so that $\mathrm{vol}_d(\mathcal{B}) = 1$. Let T be a uniformly random m-point subset of $\mathcal{B}$, where $m = \mathrm{vol}_d(\mathcal{B})/\mathrm{vol}_d(B(x,\tau/4))$ (we may assume that m is an integer by enlarging $\mathcal{B}$). Thus for every $z \in Z$, we have $\mathbb{E}\,|T \cap B(z,\tau/4)| = 1$.

We now define the random mapping $\gamma : Z \to \mathbb{R}$ by

$$\gamma(x) = \min\left(\max\left\{0, \mathsf{dist}_2(x,T) - \frac{\tau}{4}\left(1 - \frac{2}{d}\right)\right\}, \frac{\tau}{2d}\right) \tag{4}$$

By standard volume arguments, we have

$$\mathrm{vol}_d\left(B\left(x, \tfrac{\tau}{4}\right) \setminus B\left(x, \left(1 - \tfrac{1}{d}\right)\tfrac{\tau}{4}\right)\right) \approx \mathrm{vol}_d\left(B\left(x, \left(1 - \tfrac{2}{d}\right)\tfrac{\tau}{4}\right)\right),$$

up to constant factors (independent of the dimension d). So for $x \in Z$, we expect $\gamma(x)$ to take each of the values 0 and $\frac{\tau}{4d}$, with constant probability (corresponding to the events $\mathsf{dist}_2(x,T) \leq (1-\frac{2}{d})\frac{\tau}{4}$ and $\mathsf{dist}_2(x,T) \geq (1-\frac{1}{d})\frac{\tau}{4}$, respectively). Furthermore, if $\|x-y\|_2 \geq \tau$, we expect that $\gamma(x)$ and $\gamma(y)$ behave essentially independently. The truncation of $\gamma(x)$ at $\frac{\tau}{2d}$ ensures that, conditioned on the value $|B(x,\tau/4) \cap T|$, $\gamma(x)$ and $\gamma(y)$ are indeed independent. We will use this independence between $\gamma(x)$ and $\{\gamma(y)\}_{y \in Z \setminus B(x,\tau)}$ to achieve a good embedding.

Define $\Gamma : Z \to L_2(\mathcal{B}^m)$ by $\Gamma(x) = \gamma(x)$, so that

$$||\Gamma(x)||_{L_2(\mathcal{B}^m)} = \sqrt{\mathbb{E}\,|\gamma(x)|^2}.$$

Fix some $x \in Z$ and values $\{c_y\}_{y \in C(x)} \subseteq \mathbb{R}$. Let $\mathcal{E}_{\text{close}}$ be the event that $|B(x, \tau/4) \cap T| = 1$, and note that $\Pr(\mathcal{E}_{\text{close}}) \geq \frac{1}{5}$. Now observe that after conditioning on $\mathcal{E}_{\text{close}}$, the values $\{\gamma(y)\}_{y \in C(x)}$ are mutually independent of $\gamma(x)$. This follows because if $\mathsf{dist}_2(y, T) \geq \tau/2$, then $\gamma(y) = \tau/(2d)$. Since $\|x - y\|_2 \geq \tau$ for every $y \in C(x)$, the value of $\mathsf{dist}_2(x, T)$ cannot affect $\gamma(y)$, conditioned on $\mathcal{E}_{\text{close}}$. Thus we will be able to establish a lower bound if we can exhibit some variation in the value of $\gamma(x)$, conditioned on $\mathcal{E}_{\text{close}}$. Define $\mathcal{E}_{\text{in}} = \{\mathsf{dist}_2(x, T) \leq \left(1 - \frac{2}{d}\right) \frac{\tau}{4}\}$ and $\mathcal{E}_{\text{out}} = \{\mathsf{dist}_2(x, T) \geq \left(1 - \frac{1}{d}\right) \frac{\tau}{4}\}$. Simple volume computations show that $\Pr[\mathcal{E}_{\text{in}} \,|\, \mathcal{E}_{\text{close}}], \Pr[\mathcal{E}_{\text{out}} \,|\, \mathcal{E}_{\text{close}}] \geq \frac{1}{8}$. Finally, note that if $\mathcal{E}_{\text{in}}$ occurs, then $\gamma(x) = 0$ and if $\mathcal{E}_{\text{out}}$ occurs, then $\gamma(x) \geq \tau/(4d)$.

It follows that

$$\begin{aligned}
\left\| \Gamma(x) - \textstyle\sum_{y \in Y} c_y \Gamma(y) \right\|^2_{L_2(\mathcal{B}^m)} &= \mathbb{E} \left| \gamma(x) - \textstyle\sum_{y \in Y} c_y \gamma(y) \right|^2 \\
&\geq \Pr[\mathcal{E}_{\text{close}}] \cdot \mathbb{E} \left[\left| \gamma(x) - \textstyle\sum_{y \in Y} c_y \gamma(y) \right|^2 \,\middle|\, \mathcal{E}_{\text{close}} \right] \\
&\geq \Pr[\mathcal{E}_{\text{close}}] \cdot \min \left\{ \Pr[\mathcal{E}_{\text{in}} \,|\, \mathcal{E}_{\text{close}}], \Pr[\mathcal{E}_{\text{out}} \,|\, \mathcal{E}_{\text{close}}] \right\} \cdot \left(\frac{1}{2} \cdot \frac{\tau}{4d} \right)^2 \\
&\geq \left(\frac{\tau}{56\, d} \right)^2,
\end{aligned}$$

where the penultimate inequality follows from the fact that $\sum_{y \in Y} c_y \gamma(y)$ is independent of $\gamma(x)$ when conditioned on $\mathcal{E}_{\text{close}}$. We conclude that

$$\mathsf{dist}_2 \left(\Gamma(x), \text{span}\{\Gamma(y)\}_{y \in C(x)} \right) \geq \frac{\tau}{56\, d}. \tag{5}$$

For every $x \in Z$, we have

$$\|\Gamma(x)\|_{L_2(\mathcal{B}^m)} \leq \frac{\tau}{2d}, \tag{6}$$

since $\gamma(x) \leq \frac{\tau}{2d}$ with probability 1. We will now establish the following claim.

Claim 2.2. *For some constant $C \geq 1$, $\|\Gamma\|_{\text{Lip}} \leq C/\sqrt{d}$.*

Assuming the claim, we finish the proof using the following classical theorem of Kirszbraun [Kir34].

Theorem 2.3. *Let H, H' be Hilbert spaces, $S \subseteq H$, and $f : S \to H'$ a Lipschitz map. Then there exists an extension $\tilde{f} : H \to H'$ such that $\tilde{f}|_S = f$, and $\|\tilde{f}\|_{\text{Lip}} \leq \|f\|_{\text{Lip}}$.*

Using Theorem 2.3, we obtain a map $\tilde{\Gamma} : \mathbb{R}^d \to L_2$ such that $\|\tilde{\Gamma}\|_{\text{Lip}} = \|\Gamma\|_{\text{Lip}}$ and $\tilde{\Gamma}|_Z = \Gamma$, Finally, we set $F = (\sqrt{d}/C) \cdot \tilde{\Gamma}$. Observe that this rescaling satisfies $\|F\|_{\text{Lip}} \leq 1$, and improves the lower bound (5) to satisfy condition (3) of Lemma 2.1 with $\beta = \frac{1}{56 \cdot C}$. We now move onto the proof of the claim.

Proof of Claim 2.2. In what follows, we use $\|\cdot\| = \|\cdot\|_2$ to denote the 2-norm on $\mathbb{R}^d$. Recall that, for $x, y \in Z$, we wish to prove that $\|\Gamma(x) - \Gamma(y)\|_{L_2(\mathcal{B}^m)} \leq \frac{C}{\sqrt{d}} \|x - y\|$ for some fixed constant $C > 0$. The idea is that we can think of the map $x \mapsto \mathsf{dist}_2(x, T)$ like a "local"

projection onto a randomly oriented direction (indeed, the closest point to x from T has a spherically symmetric distribution about x, conditioned on $\mathsf{dist}_2(x,T) \le \tau$). If $x, y \in Z$ share the same closest point, then they experience the same "projection." The only caveat is that, conditioned on x and y sharing the same closest point, the distribution of that point is no longer spherically symmetric with respect to x or y.

In proving the claim, we may assume that $||x-y|| \le \delta$ for any $\delta > 0$. If the Lipschitz condition holds for such x, y, then it holds for all pairs as follows. Let $x', y' \in Z$ be arbitrary, and let ℓ be the line connecting x', y'. Since Z is convex, we have $\ell \subseteq Z$. Let $x' = x_1, x_2, \ldots, x_k = y'$ be a subdivision of ℓ such that $\|x_i - x_{i+1}\| \le \delta$ for every $1 \le i \le k-1$. Observe that under our assumptions,

$$\|\Gamma(x') - \Gamma(y')\|_{L_2(\mathcal{B}^m)} \le \sum_{i=1}^{k-1} \|\Gamma(x_i) - \Gamma(x_{i+1})\|_{L_2(\mathcal{B}^m)} \le \frac{C}{\sqrt{d}} \sum_{i=1}^{k-1} \|x_i - x_{i+1}\| = \frac{C}{\sqrt{d}} \|x' - y'\|.$$

Before proceeding, we need to apply the following standard volume estimate.

Lemma 2.4. *If $u, v \in \mathbb{R}^d$, and $s > 0$, then*

$$\frac{\mathrm{vol}_d\left(B(u,s) \cap B(v,s)\right)}{\mathrm{vol}_d(B(u,s))} \ge 1 - \frac{4\sqrt{d}\|u-v\|}{s}.$$

Let $\mathcal{E}_{\mathrm{share}}$ be the event that x and y share the same closest point in T, i.e. there exists $z \in T$ such that $||x-z|| = \mathsf{dist}_2(x,T)$ and $||y-z|| = \mathsf{dist}_2(y,T)$ (having a non-unique closest point is an event of measure zero, which we ignore). Using Lemma 2.4, there exists some $\delta > 0$ such that if $\|x-y\| \le \delta$, then $\Pr(\mathcal{E}_{\mathrm{share}}) \ge 1 - \frac{1}{d}$. To see this, observe that for any $R > 0$, if $\mathsf{dist}_2(x,T) = R$, then $\mathcal{E}_{\mathrm{share}}$ happens unless there is a sample point in $B(y,R) \setminus B(x,R)$, and the probability of this can be bounded using Lemma 2.4.

Proposition 2.5. *For $\delta > 0$ small enough, for every $x, y \in Z$ with $\|x-y\| \le \delta$, we have*

$$\mathbb{E}\left[|\gamma(x) - \gamma(y)|^2 \,\middle|\, \mathcal{E}_{\mathrm{share}}, \mathsf{dist}_2(x,T), \mathsf{dist}_2(y,T) \le \tau\right] \le \frac{\|x-y\|^2}{\Omega(d)}.$$

Proof. Let $z \in T$ be the closest point to x from T, and assume without loss that $z \notin \{x, y\}$. Let $u \in S^{d-1}$ be a uniformly chosen unit vector. By standard arguments, we have the bound

$$\Pr\left[|\langle x-y, u\rangle|^2 > t\frac{\|x-y\|^2}{d}\right] \le 2\,e^{-t/2},$$

see e.g. [Mat02, Ch. 14]. Observe that when $z \in B(x,\tau)$, the random vector $\frac{x-z}{\|x-z\|}$ is distributed identically to u, so using the fact that $\Pr(\mathcal{E}_{\mathrm{share}}) \ge \frac{1}{2}$ (assuming δ small enough),

$$\Pr\left[\left|\left\langle x-y, \frac{x-z}{\|x-z\|}\right\rangle\right|^2 > t\frac{\|x-y\|^2}{d} \,\middle|\, \mathcal{E}_{\mathrm{share}}, \mathsf{dist}_2(x,T) \le \tau\right] \le 4\,e^{-t/2}. \tag{7}$$

Indeed, the same bound holds with the role of x and y exchanged. Finally, we observe that if $\mathcal{E}_{\text{share}}$ occurs, then

$$\begin{aligned} |\gamma(x) - \gamma(y)| \leq |\mathsf{dist}_2(x,T) - \mathsf{dist}_2(y,T)| &= |\,\|x-z\| - \|y-z\|\,| \\ &\leq \left|\left\langle x-y, \frac{x-z}{\|x-z\|}\right\rangle\right| + \left|\left\langle x-y, \frac{y-z}{\|y-z\|}\right\rangle\right| \end{aligned} \quad (8)$$

The last inequality may be easily checked for $z = (0,0), x, y \in \mathbb{R}^2$. Combining (7) and (8), and integrating over t easily yields the desired expectation bound.

□

Let $\mathcal{E} = \{\mathcal{E}_{\text{share}}, \mathsf{dist}_2(x,T), \mathsf{dist}_2(y,T) \leq \tau\}$, and note that $\Pr(\mathcal{E}) \geq 1 - \frac{1}{2d}$ using the fact that $\Pr(B(x,\tau) \cap T = \emptyset) \leq 4^{-d}$ (and a similar bound holds for y, even conditioned on $\mathsf{dist}_2(x,T) \leq \tau$). Now, observe that

$$|\gamma(x) - \gamma(y)| \leq |\mathsf{dist}_2(x,T) - \mathsf{dist}_2(y,T)| \leq \|x-y\|,$$

hence using Proposition 2.5,

$$\begin{aligned} \|\Gamma(x) - \Gamma(y)\|^2_{L_2(\mathcal{B}^m)} &= \mathbb{E}\,|\gamma(x) - \gamma(y)|^2 \\ &\leq \Pr(\mathcal{E}) \cdot \mathbb{E}\left[|\gamma(x) - \gamma(y)|^2 \,\middle|\, \mathcal{E}\right] + (1 - \Pr(\mathcal{E})) \cdot \|x-y\|^2 \\ &\leq \frac{\|x-y\|^2}{\Omega(d)}. \end{aligned}$$

It follows that $\|\Gamma\|_{\text{Lip}} \leq 1/\Omega(\sqrt{d})$. □

Our bound on $\|\Gamma\|_{\text{Lip}}$ completes the proof. □

3 Gluing for volume-preserving embeddings

This sections concerns the following two theorems. First, we need a definition. Given a metric space (X,d) and a 1-Lipschitz map $\varphi : X \to L_2$, we define $\mathsf{rigidity}_{\leq k}(\varphi)$ to be the smallest value $R \geq 0$ such that for every $S \subseteq X$ with $|S| \leq k$, we have

$$\mathsf{dist}_2\left(\varphi(x), \mathrm{span}\{\varphi(y)\}_{y \in S}\right) \geq \frac{d(x,S)}{R}.$$

For a space X, we define $\mathsf{rigidity}_{\leq k}(X)$ to be the infimal value of $\mathsf{rigidity}_{\leq k}(\varphi)$ over all 1-Lipschitz maps $\varphi : X \to L_2$. We recall that $c_2^k(X) \leq \mathsf{rigidity}_{\leq k}(X)$ (see [Fei00, KLM04]).

Theorem 3.1. *Let (X,d) be an n-point metric space, let $\mathcal{A}$ be a collection of subsets of X, and let $R \geq 1$. Suppose that for every $\tau \geq 0$, there exists a 1-Lipschitz map $\psi_\tau : X \to L_2$ which satisfies the following.*

1. *For every $x \in X$, $\|\psi_\tau(x)\|_2 \leq \tau$.*

2. *For every $x \in X$, and every subset $S \in \mathcal{A}$,*

$$\mathsf{dist}_2\left(\psi_\tau(x), \operatorname{span}\{\psi_\tau(y)\}_{y \in S \setminus B(x,\tau)}\right) \geq \tau/R.$$

Then there exists a 1-Lipschitz map $\varphi : X \to L_2$ such that for every $x \in X$, and every $S \in \mathcal{A}$,

$$\mathsf{dist}_2\left(\varphi(x), \operatorname{span}\{\varphi(y)\}_{y \in S}\right) \geq \frac{d(x,S)}{O(\sqrt{R \log n})}.$$

In particular, if $\mathcal{A} = \{S \subseteq X : |S| \leq k\}$, then $\mathsf{rigidity}_{\leq k}(\varphi) = O(\sqrt{R \log n})$, hence also $c_2^k(X) \leq O(\sqrt{R \log n})$.

The next theorem is more technical, so we begin with an informal description. We assume the existence of maps $\psi_{S,\tau} : X \to L_2$ which are "volume-preserving" for points in $S \subseteq X$ at scale τ, where the volume distortion depends only on $|S|$. The theorem glues these maps together to obtain a map which is volume-preserving for all points at all scales.

Theorem 3.2. *Let X be an n-point metric space. Let $L > 0$, $\varepsilon \in [\frac{1}{2}, 1], \beta \leq \frac{1}{2}$ be constants. Suppose that for every $\tau \geq 0$, and every subset $S \subseteq X$ there is a 1-Lipschitz map $\psi_{S,\tau} : X \to L_2$ satisfying the following. Let $\delta_S = \beta(\log |S|)^{-\varepsilon}$.*

1. *For every $x \in X$, $||\psi_{S,\tau}(x)||_2 \leq L\delta_S\tau$.*
2. *For every $x \in S$,*

$$\mathsf{dist}_2\left(\psi_{S,\tau}(x), \operatorname{span}\{\psi_{S,\tau}(y) : y \in N_{L\delta_S\tau}(S) \setminus B(x,\tau)\}\right) \geq \delta_S\tau.$$

Then there exists a map $\varphi : X \to L_2$ with $\mathsf{rigidity}(\varphi) = O\left((\log n)^\varepsilon \log\log n\right)$.

3.1 Proof of Theorem 3.1

3.1.1 The quotient-decomposition technique

In this section, we handle a base case. First, we recall a result that follows from the techniques of [Rao99], together with the decomposition theorem of [CKR01, FRT03], and which first appeared in [KLMN05]. We say that a measure μ on a finite space X is non-degenerate if $\mu(x) > 0$ for every $x \in X$.

Theorem 3.3. *Let X be an finite metric space, let μ be a non-degenerate measure on X, and let $\tau > 0$. Then there exists a 1-Lipschitz map $\gamma_\tau : X \to L_2$ such that for every $x \in X$, $\|\gamma_\tau(x)\|_2 \leq \tau$, and*

$$\mathsf{dist}_2\left(\gamma_\tau(x), \operatorname{span}\{\gamma_\tau(y)\}_{y \in X \setminus B(x,\tau)}\right) \geq \frac{\tau}{O\left(1 + \log \frac{\mu(B(x,\tau))}{\mu(B(x,\tau/4))}\right)}.$$

For a metric space (X, d), we define the *ε-path quotient of X* to be the metric space $(X_\varepsilon, d_\varepsilon)$ which is defined as follows. X_ε is the set of equivalence classes of X under the transitive closure of the relation defined by $x \sim_\varepsilon y \iff d(x, y) \le \varepsilon$, while d_ε is the path metric in X_ε (with respect to d): For $x \in X$, let $\bar{x} \in X_\varepsilon$ represent the equivalence class of x, then for $x, y \in X$, we define

$$d_\varepsilon(\bar{x}, \bar{y}) = \inf \left\{ \sum_{i=0}^{k-1} d(x_i, x_{i+1}) : k \in \mathbb{N}, x_0 \in \bar{x}, x_k \in \bar{y}, x_j \in X, 0 \le j \le k \right\},$$

where as usual for $A, B \subseteq X$, $d(A, B) = \inf_{a \in A, b \in B} d(a, b)$. Also, starting from a measure μ on X, there is natural induced measure μ_ε on X_ε defined by $\mu_\varepsilon(\bar{x}) = \sum_{x \in \bar{x}} \mu(x)$.

We now prove the main theorem of this section. The result is essentially known, but has not appeared anywhere previously. The proof relies on quotients to retain a bound on the Lipschitz constant of the embedding (see e.g., [Bar96, Mat02]), and the observation from [GKL03] that "volume growth" at one scale is roughly maintained under path-quotients.

Theorem 3.4. *Let (X, d) be an n-point metric space. Then there exists a map $\Phi : X \to L_2$ satisfying $\|\Phi\|_{\mathrm{Lip}} \le O(\sqrt{\log n})$, and for every $x \in X$, for every $k \in \mathbb{Z}$,*

$$\mathsf{dist}_2\left(\Phi(x), \mathrm{span}\{\Phi(y)\}_{y \in X \setminus B(x, 2^k)}\right) \ge \frac{2^k}{O\left(1 + \log\left[\frac{|B(x,2^k)|}{|B(x,2^{k-3})|}\right]\right)}.$$

Proof. For each $k \in \mathbb{Z}$, define $\varepsilon_k = 2^{k-1}/n$. Letting $B_\varepsilon(\cdot, \cdot)$ represent balls in $(X_\varepsilon, d_\varepsilon)$, we have the following two sets of inequalities.

1. For every $x \in X$, $\mu_{\varepsilon_k}(B_{\varepsilon_k}(\bar{x}, 2^{k-1})) \le \mu(B(x, 2^k))$.

This follows from the fact that for $x, y \in X$, one has

$$d_{\varepsilon_k}(\bar{x}, \bar{y}) \ge d(x, y) - (n-1) \cdot \varepsilon_k \ge d(x, y) - 2^{k-1}, \tag{9}$$

because every shortest path in X has at most $n-1$ steps (thus at most $(n-1) \cdot \varepsilon_k$ distance is contracted in the quotient). The second family of inequalities follows trivially:

2. For every $x \in X$, $\mu_{\varepsilon_k}(B_{\varepsilon_k}(\bar{x}, R)) \ge \mu(B(x, R))$ for every $R \ge 0$.

In particular, combining (1) and (2), we see that for $x \in X$,

$$\log\left[\frac{\mu_{\varepsilon_k}(B_{\varepsilon_k}(\bar{x}, 2^{k-1}))}{\mu_{\varepsilon_k}(B_{\varepsilon_k}(\bar{x}, 2^{k-3}))}\right] \le \log\left[\frac{\mu(B(x, 2^k))}{\mu(B(x, 2^{k-3}))}\right].$$

Now, let $\mu(\cdot) = |\cdot|$ be the counting measure on X. Applying Theorem 3.3 to $(X_{\varepsilon_k}, d_{\varepsilon_k}), \mu_{\varepsilon_k}$ with $\tau = 2^{k-1}$, we obtain a 1-Lipschitz map $\gamma_k : X_{\varepsilon_k} \to L_2$. We may clearly think of γ_k as a map also on X by defining $\gamma_k(x) = \gamma_k(\bar{x})$. Finally, we define $\Phi : X \to \ell_2(L_2)$ by

$$\Phi(x) = \bigoplus_{k \in \mathbb{Z}} [\gamma_k(x) - \gamma_k(x_0)],$$

where $x_0 \in X$ is some fixed basepoint.

To see that Φ is well-defined (i.e. that the 2-norm of $\Phi(x)$ is bounded for every $x \in X$), let $x, y \in X$ be fixed, and let $k_0 \in \mathbb{Z}$ be such that $d(x,y) \in [2^{k_0}, 2^{k_0+1}]$, then

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} \|\gamma_k(x) - \gamma_k(y)\|_2^2 &= \sum_{k \le \lceil \log_2 n\rceil + k_0 + 2} \|\gamma_k(x) - \gamma_k(y)\|_2^2 && (10)\\
&\le \sum_{k<k_0} 2^{2k} + \sum_{k=k_0}^{\lceil \log_2 n\rceil + k_0 + 2} d(x,y)^2 && (11)\\
&\le O(\log n)\cdot d(x,y)^2, && (12)
\end{aligned}$$

where in (10) we have used the fact that for $k \ge \lceil \log_2 n\rceil + k_0 + 2$, we have $\varepsilon_k \ge 2^{k_0+1} \ge d(x,y)$, hence $\gamma_k(x) = \gamma_k(y)$ because x and y belong to the same equivalence class of X_{ε_k}, in (11) we have used $\|\gamma_k(x)\|_2 \le 2^{k-1}$ and $\|\gamma_k\|_{\mathrm{Lip}} \le 1$ from Theorem 3.3, and in (12) we evaluated a geometric sum. It follows that $\|\Phi\|_{\mathrm{Lip}} \le O(\sqrt{\log n})$.

Now, fix $x \in X$, and $k \in \mathbb{Z}$. Observe that by (9), we have $d(x,y) \ge 2^k \implies d_{\varepsilon_k}(\bar{x}, \bar{y}) \ge 2^{k-1}$. It follows that

$$\begin{aligned}
\mathsf{dist}_2\left(\Phi(x), \mathrm{span}\{\Phi(y)\}_{y\in X\setminus B(x,2^k)}\right) &\ge \mathsf{dist}_2\left(\gamma_{k-1}(x), \mathrm{span}\{\gamma_{k-1}(y)\}_{y\in X\setminus B(x,2^k)}\right)\\
&\ge \mathsf{dist}_2\left(\gamma_{k-1}(\bar{x}), \mathrm{span}\{\gamma_{k-1}(\bar{y})\}_{\bar{y}\in X_{\varepsilon_k}\setminus B_{\varepsilon_k}(\bar{x},2^{k-1})}\right)\\
&\ge \frac{2^k}{O\left(1+\log\left[\frac{\mu_{\varepsilon_k}(B_{\varepsilon_k}(\bar{x},2^{k-1}))}{\mu_{\varepsilon_k}(B_{\varepsilon_k}(\bar{x},2^{k-3}))}\right]\right)}\\
&\ge \frac{2^k}{O\left(1+\log\left[\frac{|B(x,2^k)|}{|B(x,2^{k-3})|}\right]\right)}.
\end{aligned}$$

□

3.1.2 Multi-scale gluing

In this section, we prove some multi-scale gluing lemmas for volume-preserving embeddings, and finish the proof of Theorem 3.1. The following theorem adapts a construction of the author [Lee05] to the case of volume-preserving maps.

Theorem 3.5 (Gluing for rigidity)**.** *Let (X,d) be an n-point metric space and $A, B \ge 1$, $\eta > 0$, and for every $m \in \mathbb{Z}$, let $\phi_m : X \to L_2$ be a 1-Lipschitz map such that $||\phi_m(x)||_2 \le \eta \cdot 2^m/B$ for every $x \in X$. Then there is a map $\varphi : X \to L_2$ satisfying*

1. *$||\varphi||_{\mathrm{Lip}} \le O(\eta\sqrt{\log n \log(AB)})$.*
2. *For every $x \in X$, $m \in \mathbb{Z}$, and $Y \subseteq X$,*

$$\mathsf{dist}_2\big(\varphi(x), \mathrm{span}\,\{\varphi(y)\}_{y\in Y}\big) \ge \sqrt{\left\lfloor \log \frac{|B(x,2^{m+1}A)|}{|B(x,2^m/B)|}\right\rfloor}\cdot \mathsf{dist}_2\big(\phi_m(x), \mathrm{span}\,\{\phi_m(y)\}_{y\in Y}\big).$$

Proof. First, we restate the construction of [Lee05] in a slightly modified form. Let $\rho : X \to \mathbb{R}_{\geq 0}$ be any $2B$-Lipschitz map with $\rho \equiv 1$ on $[1/B, 2A]$, and $\rho \equiv 0$ outside $[1/2B, 4A]$. For $x \in X$ and $t \geq 0$, define

$$R(x,t) = \sup\{R : |B(x,R)| \leq 2^t\},$$

and observe that $R(\cdot, t)$ is 1-Lipschitz for every value of t. And for each $m \in \mathbb{Z}$, define

$$\rho_{m,t}(x) = \rho\left(\frac{R(x,t)}{2^m}\right).$$

Now, for each $t \in \{1, 2, \ldots, \lceil \log_2 n \rceil\}$, define $\psi_t : X \to \ell_2(L_2)$,

$$\psi_t(x) = \bigoplus_{m \in \mathbb{Z}} \rho_{m,t}(x) \cdot \phi_m(x).$$

Finally, let $\varphi = \psi_1 \oplus \psi_2 \oplus \cdots \oplus \psi_{\lceil \log_2 n \rceil}$.

First, we bound $||\psi_t||_{\text{Lip}}$ as follows.

$$||\psi_t(x) - \psi_t(y)||_2^2 = \sum_{\substack{m \in \mathbb{Z} \\ \rho_{m,t}(x) + \rho_{m,t}(y) > 0}} ||\rho_{m,t}(x)\widehat{\phi}_m(x) - \rho_{m,t}(y)\widehat{\phi}_m(y)||_2^2.$$

The number of non-zero summands above is at most $O(\log A + \log B)$. Furthermore, each summand can be bounded as follows.

$$\begin{aligned} ||\rho_{m,t}(x)\widehat{\phi}_m(x) - \rho_{m,t}(y)\widehat{\phi}_m(y)||_2 &\leq ||\widehat{\phi}_m(x)||_2 |\rho_{m,t}(x) - \rho_{m,t}(y)| + ||\widehat{\phi}_m(x) - \widehat{\phi}_m(y)||_2 |\rho_{m,t}(y)| \\ &\leq \left(||\rho_{m,t}||_{\text{Lip}} \frac{\eta \cdot 2^m}{B} + ||\phi_m||_{\text{Lip}}\right) d(x,y) \\ &\leq (2\eta + 1)\, d(x,y), \end{aligned}$$

where we have used $\|\rho_{m,t}\|_{\text{Lip}} \leq 2^{-m}\|\rho\|_{\text{Lip}} \leq 2^{-m+1}B$. Thus $||\psi_t||_{\text{Lip}} \leq O(\eta\sqrt{\log(AB)})$. It follows that $||\varphi||_{\text{Lip}} \leq O(\eta\sqrt{\log n \log(AB)})$, as claimed.

To verify (2), fix $x \in X$, $m \in \mathbb{Z}$, and real constants c_y for $y \in Y$. Then,

$$\begin{aligned} \left\| \varphi(x) - \sum_{y \in Y} c_y \varphi(y) \right\|_2^2 &= \sum_{t=1}^{\lceil \log n \rceil} \left\| \psi_t(x) - \sum_{y \in Y} c_y \psi_t(y) \right\|_2^2 \\ &\geq \sum_{t=1}^{\lceil \log n \rceil} \left\| \rho_{m,t}(x)\phi_m(x) - \sum_{y \in Y} c_y \rho_{m,t}(y)\phi_m(y) \right\|_2^2. \end{aligned}$$

Now observe that when $\rho_{m,t}(x) = 1$, we have

$$\left\| \rho_{m,t}(x)\phi_m(x) - \sum_{y \in Y} c_y \rho_{m,t}(y)\phi_m(y) \right\|_2 \geq \mathsf{dist}_2\big(\phi_m(x), \operatorname{span}\{\phi_m(y)\}_{y \in Y}\big). \tag{13}$$

Hence it suffices to count the number of values of t for which $\rho_{m,t}(x) = 1$. By our definitions we have that

$$\rho_{m,t}(x) = 1 \iff \frac{2^m}{B} \le R(x,t) \le 2^{m+1}A \iff t \in [\log|B(x,2^m/B)|, \log|B(x,2^{m+1}A)|].$$

This completes the proof since the lower bound (13) holds for $\left\lfloor \log \frac{|B(x,2^{m+1}A)|}{|B(x,2^m/B)|} \right\rfloor$ values of t. □

Now we complete the proof of Theorem 3.1, along the lines of [KLMN05].

Proof of Theorem 3.1. In the applications of Theorem 3.5 that follow, we set $A = 2, B = 8$, and $\eta = 16$. Let $\{\psi_\tau\}_{\tau \ge 0}$ be as in the statement of the theorem, and let $\Psi : X \to L_2$ be the map resulting from the application of Theorem 3.5 to the ensemble $\{\psi_{2^m}\}_{m\in\mathbb{Z}}$. Let $\gamma_\tau : X \to L_2$ be the maps from Theorem 3.3 applied to X, and let $\Gamma : X \to L_2$ be the result of applying Theorem 3.5 to $\{\gamma_{2^m}\}_{m\in\mathbb{Z}}$. Finally, let $\Phi : X \to L_2$ be the map from Theorem 3.4. Consider $\varphi = \Psi \oplus \Gamma \oplus \Phi$.

First, we have $\|\varphi\|_{\mathrm{Lip}} \le \|\Psi\|_{\mathrm{Lip}} + \|\Gamma\|_{\mathrm{Lip}} + \|\Phi\|_{\mathrm{Lip}} \le O(\sqrt{\log n})$. Now fix $x \in X$, and $Y \in \mathcal{A}$. Let $m \in \mathbb{Z}$ be such that $d(x,Y) \in (2^m, 2^{m+1}]$. Observe that if $\log \frac{|B(x,2^{m+3})|}{|B(x,2^{m-3})|} \le 1$, then

$$\mathrm{dist}_2\left(\varphi(x), \mathrm{span}\{\varphi(y)\}_{y\in Y}\right) \ge \mathrm{dist}_2\left(\Phi(x), \mathrm{span}\{\Phi(y)\}_{y\in Y}\right) \ge \Omega(1)\cdot 2^m = \Omega(1)\cdot d(x,Y).$$

So we may assume that $\log \frac{|B(x,2^{m+3})|}{|B(x,2^{m-3})|} \ge 1$ in the arguments that follow. In this case,

$$\begin{aligned}
&\mathrm{dist}_2\left(\varphi(x), \mathrm{span}\{\varphi(y)\}_{y\in Y}\right)^2 \\
&\ge \mathrm{dist}_2\left(\Psi(x), \mathrm{span}\{\Psi(y)\}_{y\in Y}\right)^2 + \mathrm{dist}_2\left(\Gamma(x), \mathrm{span}\{\Gamma(y)\}_{y\in Y}\right)^2 \\
&\ge \left\lfloor \log \frac{|B(x,2^{m+3})|}{|B(x,2^{m-3})|} \right\rfloor \left[\mathrm{dist}_2\left(\psi_{2^m}(x), \mathrm{span}\{\psi_{2^m}(y)\}_{y\in Y}\right)^2 + \mathrm{dist}_2\left(\gamma_{2^m}(x), \mathrm{span}\{\gamma_{2^m}(y)\}_{y\in Y}\right)^2\right] \\
&\ge \left\lfloor \log \frac{|B(x,2^{m+3})|}{|B(x,2^{m-3})|} \right\rfloor \left[\left(\frac{2^m}{R}\right)^2 + \left(\frac{2^m}{O\left(1+\log\frac{|B(x,2^{m+3})|}{|B(x,2^{m-3})|}\right)}\right)^2\right] \\
&\ge \Omega(1)\cdot d(x,Y)^2 \left[\frac{\log\frac{|B(x,2^{m+3})|}{|B(x,2^{m-3})|}}{R^2} + \frac{1}{\log\frac{|B(x,2^{m+3})|}{|B(x,2^{m-3})|}}\right] \ge \Omega(1)\cdot\frac{d(x,Y)^2}{R},
\end{aligned}$$

where the last line follows from the AM-GM inequality. If we now replace φ by $\frac{1}{\|\varphi\|_{\mathrm{Lip}}}\varphi$, where $\|\varphi\|_{\mathrm{Lip}} \le O(\sqrt{\log n})$, then φ becomes 1-Lipschitz, and the proof is complete. □

One corollary of Theorem 3.1 is the optimal bound for general n-point metric spaces.

Corollary 3.6. *For any n-point metric space (X,d), one has $r_2(X) \le O(\log n)$.*

Proof. Apply Theorem 3.1 with $\mathcal{A} = 2^X$ to the ensemble of maps $\{\gamma_\tau\}$ from Lemma 3.3 (with the counting measure $\mu(\cdot) = |\cdot|$). □

3.2 Proof of Theorem 3.2

The proof of Theorem 3.2 requires most of the results of the previous section, along with a number of other ideas. Our approach follows [ALN08], but with new machinery to deal with volume-preserving embeddings. The following claim gives a way of extending our embeddings to larger distances by dampening out the effects of distant points.

Claim 3.7. *Let (X, d) be a metric space, and suppose that for some number $\epsilon > 0$, we have a map $\psi : X \to L_2$ such that $\|\psi\|_{\text{Lip}} \leq 1$, and $\|\psi(x)\|_2 \leq \epsilon L$ for all $x \in X$. Let $S, U \subseteq X$, and suppose that for $x \in S$, we have*

$$\mathsf{dist}_2\left(\psi(x), \text{span}\{\psi(y)\}_{y\in U}\right) \geq \epsilon.$$

Define $U' = U \cup (X \setminus N_{\epsilon L}(S))$. Then there exists a map $\varphi : X \to L_2$ which satisfies $||\varphi||_{\text{Lip}} \leq 1$, $\|\varphi(x)\|_2 \leq \epsilon L$ for every $x \in X$, and for any $x \in N_{\epsilon/4}(S)$,

$$\mathsf{dist}_2\left(\varphi(x), \text{span}\{\varphi(y)\}_{y\in U'}\right) \geq \epsilon/4.$$

Proof. Let $\rho(x) = \max(0, 1 - d(x, S)/(\epsilon L))$ so that $\|\rho\|_{\text{Lip}} \leq 1/\epsilon L$, and define $\varphi(x) = \frac{1}{2}\rho(x)\psi(x)$. Then for every $x, y \in X$,

$$||\varphi(x) - \varphi(y)||_2 \leq \tfrac{1}{2}\left(|\rho(x)| \cdot ||\psi||_{\text{Lip}} + ||\psi(y)||_2 \cdot ||\rho||_{\text{Lip}}\right) d(x, y) \leq d(x, y).$$

Now, suppose that $x \in N_{\epsilon/4}(S)$, and $\{c_y\}_{y\in U'} \subseteq \mathbb{R}$. Let $x' \in S$ be such that $d(x, x') \leq \epsilon/4$, and recall that $y \notin N_{\epsilon L}(S)$ implies $\rho(y) = 0$, hence

$$\begin{aligned}
\left\|\varphi(x) - \textstyle\sum_{y\in U'} c_y\varphi(y)\right\|_2 &\geq \left\|\varphi(x') - \textstyle\sum_{y\in U'} c_y\varphi(y)\right\|_2 - ||\varphi(x) - \varphi(x')||_2 \\
&\geq \tfrac{1}{2}\left\|\psi(x') - \textstyle\sum_{y\in U} c_y\rho(y)\psi(y)\right\|_2 - \epsilon/4 \\
&\geq \tfrac{1}{2}\,\mathsf{dist}_2\left(\psi(x'), \text{span}\{\psi(y)\}_{y\in U}\right) - \epsilon/4 \\
&\geq \epsilon/2 - \epsilon/4 = \epsilon/4.
\end{aligned}$$

□

The first step is to reduce the number of points we need to embed by randomly sampling a reasonably dense subset of our space. We will need to use different sampling rates in different regions of the space (depending upon the local volume growth), and thus we arrive at different guarantees corresponding to the various subsets $T_\tau(A; k)$ defined below.

Lemma 3.8. *Assume that X satisfies the conditions of Theorem 3.2, and suppose that $A \subseteq X$ and $k \geq 2$. Define*

$$T_\tau(A; k) = \left\{x \in A : |A| \leq k \;\middle|\; B\left(x, \frac{\beta\tau}{4(\log k)^\varepsilon}\right)\right\}. \tag{14}$$

Then there exists a 1-Lipschitz map $\lambda_{A,k} : X \to L_2$ such that

$$\operatorname{dist}_2\left(\lambda_{A,k}(x), \operatorname{span}\{\lambda_{A,k}(y)\}_{y\in X\setminus B(x,\tau)}\right) \geq \frac{\beta\tau}{8(\log k)^{\varepsilon}},$$

whenever $x \in T_\tau(A;k)$. Furthermore, for all $x \in X$, $\|\lambda_{A,k}(x)\|_2 \leq L\delta_S\tau$.

Proof. Let S be a uniformly random subset $S \subseteq A$ with $|S| = \min\{|A|, k\}$. Let $h_S : X \to L_2$ be the map defined by applying Claim 3.7 to the map $\psi_{S,\tau} : X \to L_2$. Let μ be the distribution over which the random subsets $S \subseteq A$ are defined, and let $\lambda_{A,k} : X \to L_2(\mu)$ be given by $\lambda_{A,k}(x) = h_S(x)$. Observe that $\|\lambda_{A,k}\|_{\mathrm{Lip}} \leq 1$ because this is true of each $\psi_{S,\tau}$, and hence by Claim 3.7, it is also true of each h_S.

Fix $x \in T_\tau(A;k)$. Then by the definition of $T_\tau(A;k)$, with probability at least $1/e$, we have

$$S \cap B\left(x, \frac{\beta\tau}{4(\log k)^{\varepsilon}}\right) \neq \emptyset. \tag{15}$$

Letting $\epsilon = \delta_S\tau$, we see that conditioned on (15), $x \in N_{\epsilon/4}(S)$. Setting $U = N_{\epsilon L}(S)\setminus B(x,\tau)$, we have in this case (by assumption (2) on the map $\psi_{S,\tau}$ in Theorem 3.2),

$$\mathsf{dist}_2\left(\psi_{S,\tau}(x), \operatorname{span}\left\{\psi_{S,\tau}(y)\right\}_{y\in U}\right) \geq \delta_S\tau = \epsilon.$$

It follows that in the statement of Claim 3.7, $U' = U \cup (X \setminus N_{\epsilon L}(S)) = X \setminus B(x,\tau)$. So that with probability $1/e$, we have

$$\operatorname{dist}_2\left(h_S(x), \operatorname{span}\{h_S(y)\}_{y\in X\setminus B(x,\tau)}\right) \geq \frac{\beta\tau}{4(\log k)^{\varepsilon}}.$$

Hence for any $x \in T_\tau(A;k)$,

$$\operatorname{dist}_2\left(\lambda_{A,k}(x), \operatorname{span}\{\lambda_{A,k}(y)\}_{y\in X\setminus B(x,\tau)}\right) \geq \frac{1}{\sqrt{e}}\cdot\frac{\beta\tau}{4(\log k)^{\varepsilon}} \geq \frac{\beta\tau}{8(\log k)^{\varepsilon}}.$$

□

Our next step is to construct embeddings separately for different localities of the space. These local embeddings are stitched together in a smooth way using partitions of unity derived from random partitions of the space.

Fix a finite metric space (X,d), and for $K \geq 1$, $\tau \geq 0$, define

$$S_\tau(K) = \left\{x \in X : |B(x, 8\tau\alpha_X)| \leq K \,\middle|\, B\left(x, \frac{\beta\tau}{4(\log K)^{\varepsilon}}\right)\right\},$$

where we recall that α_X is the modulus of decomposability of X.

Lemma 3.9 (Localization)**.** *Assume that X satisfies the conditions of Theorem 3.2. Then for every $\tau \geq 0, k \geq 1$, there exists a 1-Lipschitz map $\Lambda_{\tau,k} : X \to L_2$ such that for every $x \in S_\tau(k)$,*

$$\operatorname{dist}_2\left(\Lambda_{\tau,k}(x), \operatorname{span}\{\Lambda_{\tau,k}(y)\}_{y \in X \setminus B(x,\tau)}\right) \geq \frac{\beta\tau}{24(\log k)^\varepsilon},$$

and $\|\Lambda_{\tau,k}(x)\|_2 \leq L\delta_S\tau$ for all $x \in X$.

Proof. Let $D = 4\tau\alpha_X$ and take P_D to be a random partition from the α_X-padded bundle for X with diameter bound D. Define a random mapping $\rho : X \to \mathbb{R}$ by

$$\rho(z) = \min\left\{1, \frac{d(z, X \setminus P_D(z))}{\tau}\right\}.$$

Clearly $\|\rho\|_{\mathrm{Lip}} \leq 1/\tau$. For each $U \in P_D$, let $\lambda_{U,k} : X \to L_2$ be the corresponding map from Lemma 3.8. Also, for every such U, let σ_U be a $\{0,1\}$-valued Bernoulli random variable independent of all other variables in the proof. Finally, define a random map $\Lambda_{\tau,k} : X \to L_2$ by

$$\Lambda_{\tau,k}(z) = \tfrac{1}{2}\rho(z) \cdot \sigma_{P_D(z)} \cdot \lambda_{P_D(z),k}(z).$$

Clearly $\|\Lambda_{\tau,k}(x)\|_2 \leq \|\lambda_{P_D(x),k}(x)\|_2 \leq L\delta_S\tau$ for every $x \in X$. Moreover, we claim that $\|\Lambda_{\tau,k}\|_{\mathrm{Lip}} \leq 1$. Indeed, fix $u, v \in X$. If $P_D(u) = P_D(v) = U$ then

$$\begin{aligned} ||\Lambda_{\tau,k}(u) - \Lambda_{\tau,k}(v)||_2 &\leq \tfrac{1}{2}|\rho(u) - \rho(v)| \cdot ||\lambda_{U,k}(u)||_2 + \tfrac{1}{2}||\lambda_{U,k}(u) - \lambda_{U,k}(v)||_2 \cdot |\rho(v)| \\ &\leq \tfrac{1}{2}(\tau||\rho||_{\mathrm{Lip}} + ||\lambda_{U,k}||_{\mathrm{Lip}})\, d(u,v) \\ &\leq d(u,v). \end{aligned}$$

Otherwise, assume that $P_D(u) \neq P_D(v)$. In particular,

$$d(u,v) \geq \max\{d(u, X \setminus P_D(u)), d(v, X \setminus P_D(v))\}.$$

It follows that

$$\begin{aligned} ||\Lambda_{\tau,k}(u) - \Lambda_{\tau,k}(v)||_2 &\leq ||\Lambda_{\tau,k}(u)||_2 + ||\Lambda_{\tau,k}(v)||_2 \\ &\leq \frac{d(u, X \setminus P_D(u))}{2\tau} \cdot \tau + \frac{d(v, X \setminus P_D(v))}{2\tau} \cdot \tau \\ &\leq d(u,v). \end{aligned}$$

Now suppose that $x \in S_\tau(k)$. Observe that since $\operatorname{diam}(P_D(x)) \leq D$, we have $P_D(x) \subseteq B(x, 2D)$. It follows that since $x \in S_\tau(k)$, we have $x \in T_\tau(P_D(x); k)$ (recall equation (14)). Moreover, using the defining property of the α_X-padded bundle, with probability at least $\frac{1}{2}$,

we have $d(x, X \setminus P_D(x)) \geq \tau$, which implies $\rho(x) = 1$. If $\{c_y\}_{y \in X \setminus B(x,\tau)} \subseteq \mathbb{R}$, then

$$\begin{aligned}
&\mathbb{E} \left\| \Lambda_{\tau,k}(x) - \sum_y c_y \Lambda_{\tau,k}(y) \right\|_2^2 \\
&\geq \tfrac{1}{2}\,\mathbb{E}\left[\left\| \Lambda_{\tau,k}(x) - \sum_y c_y \Lambda_{\tau,k}(y) \right\|_2^2 \,\Big|\, \rho(x) = 1 \right] \\
&= \tfrac{1}{2}\,\mathbb{E}'\left[\left\| \sigma_{P_D(x)} \left(\lambda_{P_D(x),k}(x) - \sum_{y \in P_D(x)} c_y \rho(y) \lambda_{P_D(x),k}(y) \right) - \sum_{y \notin P_D(x)} \sigma_{P_D(y)}\, c_y \rho(y) \lambda_{P_D(y),k}(y) \right\|_2^2 \right],
\end{aligned}$$

where $\mathbb{E}'[\cdot] = \mathbb{E}[\cdot \,|\, \rho(x) = 1]$, and we recall that $y \in P_D(x) \implies P_D(y) = P_D(x)$. Now, observe that the values $\left\{\sigma_{P_D(y)}, \rho(y), \lambda_{P_D(y),k}(y)\right\}_{y \notin P_D(x)}$ are independent of the random variable $\sigma_{P_D(x)}$. We use the following simple fact: If A, B are (possibly dependent) real-valued random variables and σ is an independent $\{0,1\}$-valued Bernoulli random variable, then $\mathbb{E}|\sigma A - B|^2 \geq \frac{1}{4}\mathbb{E}|A|^2$ (in particular, by integrating, the same holds if A, B are random elements in some Hilbert space). It follows that the last line of the above expression is at least

$$\begin{aligned}
&\tfrac{1}{8}\mathbb{E}' \left\| \lambda_{P_D(x),k}(x) - \sum_{y \in P_D(x)} c_y \rho(y) \lambda_{P_D(x),k}(y) \right\|_2^2 \\
&\qquad \geq \tfrac{1}{8}\mathbb{E}'\left[\mathrm{dist}_2\left(\lambda_{P_D(x),k}(x), \mathrm{span}\{\lambda_{P_D(x),k}(y)\}_{y \in X \setminus B(x,\tau)} \right)^2 \right] \geq \frac{1}{8}\left(\frac{\beta\tau}{8(\log k)^\varepsilon} \right)^2,
\end{aligned}$$

where the final line follows from Lemma 3.8.

Denoting by (Ω, μ) the probability space on which $\Lambda_{\tau,k}$ is defined, we can think of $\Lambda_{\tau,k}$ as a mapping of X into the Hilbert space $L_2(L_2, \mu)$ for which we have just argued that

$$\mathrm{dist}_2\left(\Lambda_{\tau,k}(x), \mathrm{span}\{\Lambda_{\tau,k}(y)\}_{y \in X \setminus B(x,\tau)} \right) \geq \frac{\beta\tau}{8\sqrt{8}(\log k)^\varepsilon} \geq \frac{\beta\tau}{24(\log k)^\varepsilon}.$$

□

Now we complete the proof of Theorem 3.2 along the lines of [ALN08].

Proof of Theorem 3.2. We claim that for every $K \in [2, n]$ there exists a map $f_K : X \to L_2$ which satisfies

1. $||f_K||_{\mathrm{Lip}} \leq O(L\sqrt{\log n \cdot \log\log n})$.
2. For every $m \in \mathbb{Z}$ and $x \in S_{2^m}(K)$, we have

$$\mathsf{dist}_2\left(f_K(x), \mathrm{span}\left\{f_K(y)\right\}_{y \in X \setminus B(x, 2^m)} \right) \geq \sqrt{\left\lfloor \log \frac{|B(x, 2^{m+3}\alpha_X)|}{|B(x, \beta 2^{m-2}/(\log K)^\varepsilon)|} \right\rfloor} \cdot \frac{\beta 2^m}{24(\log K)^\varepsilon}.$$

Indeed, f_K is obtained from an application of Theorem 3.5 to the mappings $\{\Lambda_{2^m,K}\}_{m\in\mathbb{Z}}$ from Lemma 3.9 with $A = 4\alpha_X$ and $B = 4(\log K)^\varepsilon/\beta$ (and using the fact that $\alpha_X = O(\log n)$).

Observe that for every $m \in \mathbb{Z}$, $S_{2^m}(n) = X$. Hence, defining $K_0 = n$ and $K_{j+1} = \sqrt{K_j}$, as long as $K_j \geq 4$, we obtain mappings $f_0, \ldots, f_j : X \to L_2$ satisfying

1. $\|f_j\|_{\mathrm{Lip}} \leq O(L\sqrt{\log n \cdot \log\log n})$.
2. For all $x \in S_{2^m}(K_j) \setminus S_{2^m}(K_{j+1})$, we have

$$\begin{aligned}
\mathsf{dist}_2\left(f_j(x), \operatorname{span}\{f_j(y)\}_{y\in X\setminus B(x,2^m)}\right) &\geq \sqrt{\left\lfloor \log \frac{|B(x,2^{m+3}\alpha_X)|}{|B(x,\beta 2^{m-2}/(\log K_j)^\varepsilon)|}\right\rfloor} \cdot \frac{\beta 2^m}{24(\log K_j)^\varepsilon} \\
&\geq \sqrt{\lfloor \log K_{j+1}\rfloor} \cdot \frac{\beta 2^m}{24(\log K_j)^\varepsilon} \qquad (16)\\
&\geq \frac{\beta 2^m}{48(\log K_j)^{\varepsilon-\frac{1}{2}}}, \qquad (17)
\end{aligned}$$

where in (16) we used the fact that $x \notin S_{2^m}(K_{j+1})$, and in (17) we used the fact that $K_{j+1} = \sqrt{K_j} \geq 2$.

This procedure ends after N steps, where $N \leq O(\log\log n)$. Every $x \in S_{2^m}(K_N)$ satisfies

$$|B(x, 2^{m+3}\alpha_X)| \leq 4|B(x, \beta 2^m/[4(\log K)^\varepsilon])|.$$

In particular, $|B(x,2^m)| \leq 4|B(x,2^{m-3})|$. By Theorem 3.4, there is a mapping $f_{N+1} : X \to L_2$ with $\|f_{N+1}\|_{\mathrm{Lip}} \leq O(\sqrt{\log n})$ and such that for $x \in S_{2^m}(K_N)$, we have

$$\mathsf{dist}_2\left(f_{N+1}(x), \operatorname{span}\{f_{N+1}(y)\}_{y\in X\setminus B(x,2^m)}\right) \geq 2^m/O(1).$$

(This follows because, for such x, we have $\log \frac{|B(x,2^m)|}{|B(x,2^{m-3})|} = O(1)$.)

Let $M = \max_{j=1,\ldots,N+1} \|f_j\|_{\mathrm{Lip}}$. Consider the map $\Phi = \frac{1}{M\sqrt{N+1}} \bigoplus_{j=0}^{N+1} f_j$, which has $\|\Phi\|_{\mathrm{Lip}} \leq 1$. Recall that $M\sqrt{N+1} \leq O(\sqrt{\log n}\log\log n)$. Let $x \in X$, $Y \subseteq X$ be arbitrary, and choose $m \in \mathbb{Z}$ such that $d(x,Y) \in (2^m, 2^{m+1}]$. If $x \in S_{2^m}(K_N)$, then

$$\begin{aligned}
\mathsf{dist}_2\left(\Phi(x), \operatorname{span}\{\Phi(y)\}_{y\in Y}\right) &\geq \mathsf{dist}_2\left(\Phi(x), \operatorname{span}\{\Phi(y)\}_{y\in X\setminus B(x,2^m)}\right) \\
&\geq \tfrac{1}{M\sqrt{N+1}}\, \mathsf{dist}_2\left(f_{N+1}(x), \operatorname{span}\{f_{N+1}(y)\}_{y\in X\setminus B(x,2^m)}\right) \\
&\geq \frac{2^m}{O(1)M\sqrt{N+1}} \geq \frac{d(x,Y)}{O(\sqrt{\log n}\log\log n)}.
\end{aligned}$$

Otherwise, without loss of generality there is $j \in \{0,\ldots,N-1\}$ such that $x \in S_{2^m}(K_j) \setminus S_{2^m}(K_{j+1})$, in which case by (17)

$$\begin{aligned}
\mathsf{dist}_2\left(\Phi(x), \operatorname{span}\{\Phi(y)\}_{y\in Y}\right) &\geq \mathsf{dist}_2\left(\Phi(x), \operatorname{span}\{\Phi(y)\}_{y\in X\setminus B(x,2^m)}\right) \\
&\geq \tfrac{1}{M\sqrt{N+1}}\, \mathsf{dist}_2\left(f_j(x), \operatorname{span}\{f_j(y)\}_{y\in X\setminus B(x,2^m)}\right) \\
&\geq \frac{d(x,Y)}{O(\log n)^\varepsilon \log\log n}
\end{aligned}$$

It follows that $\mathsf{rigidity}(\Phi) \leq O(\log n)^{\varepsilon} \log\log n$. $\square$

4 Upper bounds for rigid embeddings

4.1 The transference theorem

Now we finish our proof of the transference theorem (Theorem 1.4). For the sake of simplicity, we may assume that $\mathsf{h}(d) \geq 50$ for all $d \geq 1$.

Lemma 4.1. *There exists a constant $C \geq 1$ such that the following holds. Let H be a Hilbert space, and suppose that $S \subseteq H$ with $|S| = k$. Let $\tau \geq 0$ be given and let $\delta = \tau/(4 \cdot \mathsf{h}(C \log k))$, where $\mathsf{h}(\cdot)$ is from Definition 1.3. Then there exists a map $\varphi : H \to L_2$ such that $||\varphi||_{\mathrm{Lip}} \leq 1$ and $\|\varphi(x)\|_2 \leq 30\delta$ for all $x \in H$, and such that for all $x \in S$, we have*

$$\mathsf{dist}_2\left(\varphi(x), \mathrm{span}\left\{\varphi(y) : y \in N_{30\delta}(S) \setminus B(x,\tau)\right\}\right) \geq \delta.$$

Proof. Using the Johnson-Lindenstrauss flattening lemma [JL84], there exists a map $g : S \to \mathbb{R}^d$ with $d = C \log k$ such that $||g||_{\mathrm{Lip}} \leq 1$ and $||g^{-1}||_{\mathrm{Lip}} \leq 2$, where $C \geq 1$ is a universal constant. By the Kirszbraun extension theorem [Kir34], there exists $\tilde{g} : H \to \mathbb{R}^d$ with $||\tilde{g}||_{\mathrm{Lip}} \leq 1$ and $\tilde{g}(x) = g(x)$ for $x \in S$.

Now, let $Z \subseteq \mathbb{R}^d$ be a compact set which contains $N_\tau(\mathrm{Im}(g))$ (this set is bounded because S, and hence $\mathrm{Im}(g)$, is finite). Let $F : \mathbb{R}^d \to L_2$ be a 1-Lipschitz map satisfying the following.

1. For every $x \in \mathbb{R}^d$, $\|F(x)\|_2 \leq \frac{30\tau}{4\mathsf{h}(d)} = 30\delta$.
2. For every $x \in Z$,
$$\mathsf{dist}_2\left(F(x), \mathrm{span}\{F(y)\}_{y \in Z \setminus B(x,\tau/4)}\right) \geq \frac{\tau}{4\mathsf{h}(d)} = \delta. \tag{18}$$

Such a map follows immediately from the definition of $\mathsf{h}(\cdot)$ after scaling by $\tau/4$. Finally, define $\varphi : H \to L_2$ by $\varphi(x) = F(\tilde{g}(x))$. We now make the following observations.

1. $||\varphi||_{\mathrm{Lip}} \leq ||F||_{\mathrm{Lip}} \cdot ||\tilde{g}||_{\mathrm{Lip}} \leq 1$.
2. For every $x \in H$, $||\varphi(x)||_2 = ||F(\tilde{g}(x))||_2 \leq 30\delta$.
3. For every $x \in S$,
$$\mathrm{dist}_2\left(\varphi(x), \mathrm{span}\{\varphi(y) : y \in N_{30\delta}(S) \setminus B(x,\tau)\}\right) \geq \delta$$
To see this, fix $x \in S$, and suppose that $y \in N_{30\delta}(S) \setminus B(x,\tau)$. Let $y' \in S$ satisfy $\|y - y'\|_H \leq 30\delta$. Then clearly $\|\tilde{g}(y) - \tilde{g}(y')\|_2 \leq \|y - y'\|_H \leq \tau$, implying that

$\tilde{g}(y) \in N_\tau(\mathrm{Im}(g)) \subseteq Z$. Furthermore,

$$\begin{aligned} \|\tilde{g}(x) - \tilde{g}(y)\|_2 &\geq \|\tilde{g}(x) - \tilde{g}(y')\|_2 - \|\tilde{g}(y) - \tilde{g}(y')\|_2 \\ &\geq \tfrac{1}{2}\|x - y'\|_H - 30\delta \qquad (19) \\ &\geq \tfrac{1}{2}\left(\|x - y\|_H - \|y - y'\|_H\right) - 30\delta \\ &\geq \tau/2 - 45\delta > \tau/4, \qquad (20) \end{aligned}$$

where in (19), we use the fact that $\|g^{-1}\|_{\mathrm{Lip}} \leq 2$, and in (20), we use the assumption that $\mathsf{h}(d) \geq 50$.

To finish, write

$$\begin{aligned} &\mathrm{dist}_2\left(\varphi(x), \mathrm{span}\{\varphi(y) : y \in N_{30\delta}(S) \setminus B(x,\tau)\}\right) \\ &\quad= \mathrm{dist}_2\left(F(\tilde{g}(x)), \mathrm{span}\{F(\tilde{g}(y)) : y \in N_{30\delta}(S) \setminus B(x,\tau)\}\right) \\ &\quad\geq \mathrm{dist}_2\left(F(\tilde{g}(x)), \mathrm{span}\{F(y) : y \in Z \setminus B(\tilde{g}(x), \tau/4)\}\right) \geq \delta, \end{aligned}$$

where in the final line, we use the fact that for every $y \in N_{30\delta}(S) \setminus B(x,\tau)$, we have shown above that $\tilde{g}(y) \in Z \setminus B(\tilde{g}(x), \tau/4)$, and to finish we have employed estimate (18).

□

Now we can finish the proof of Theorem 1.4.

Proof of Theorem 1.4. Suppose that $\mathsf{h}(d) \leq C \cdot d^\varepsilon$ for some constant $C \geq 1$ and $\varepsilon \geq \frac{1}{2}$, and let $X \subseteq L_2$ with $|X| = n$. In this case, applying Lemma 4.1 to a subset $S \subseteq X$ and a value $\tau \geq 0$ yields a 1-Lipschitz map $\psi_{S,\tau} : X \to L_2$ satisfying $\|\psi_{S,\tau}(x)\|_2 \leq 30\delta_S\tau$ and

$$\mathsf{dist}_2\left(\psi_{S,\tau}(x), \mathrm{span}\left\{\psi(y) : y \in N_{30\delta_S\tau}(S) \setminus B(x,\tau)\right\}\right) \geq \delta_S\tau$$

for $x \in S$, and some $\delta_S \geq 1/O(\log|S|)^\varepsilon$. Applying Theorem 3.2 to this ensemble of maps shows that $\mathsf{rigidity}(X) \leq O(\log n)^\varepsilon \log\log n$. □

Applying Theorem 1.4 with the result of Theorem 1.5 immediately yields our desired bound.

Corollary 4.2. *For any n-point subset $X \subseteq L_2$, $\mathsf{rigidity}(X) \leq O(\sqrt{\log n}\log\log n)$.*

4.2 Optimal bounds for constant values of k

Now we prove Theorem 1.2. In fact, we prove a slightly stronger estimate (see the beginning of Section 3 for the definition of $\mathsf{rigidity}_{\leq k}$). We recall that for $k = O(1)$, the theorem is optimal up to a universal constant since $\mathsf{rigidity}_{\leq 2}(P_n) \geq c_2^2(P_n) \geq \Omega(\sqrt{\log n})$ where P_n is the path metric on an n-point path [DV01, KLM04], and clearly $\mathsf{rigidity}_{\leq 1}(P_n) = c_2(P_n) = 1$. (Map the kth point of P_n to $(k, \ell) \in \mathbb{R}^2$ and let $\ell \to \infty$. The rigidity goes to 1.)

Theorem 4.3. *For any n-point subset $X \subseteq L_2$, we have* $\mathsf{rigidity}_{\leq k}(X) \leq O(\sqrt{\log n}(\log k)^{1/4})$.

Proof. We may clearly assume that $X \subseteq \mathbb{R}^n$. Let $\tilde{P}_{n,d} : \mathbb{R}^n \to \mathbb{R}^d$ be a projection onto a random $d = O(\log k)$-dimensional subspace, let $P_{n,d} = \sqrt{\frac{n}{d}} \cdot \tilde{P}_{n,d}$, and let $Z_{n,d} \subseteq \mathbb{R}^d$ a compact set such that $P_{n,d}(X) \subseteq Z_{n,d}$. Let $F_{d,\tau} : \mathbb{R}^d \to L_2$ be a random 1-Lipschitz map satisfying the following conditions for some $\frac{1}{30} \geq \delta \geq 1/O(\sqrt{d})$.

1. For every $x \in \mathbb{R}^d$, $\|F_{d,\tau}(x)\|_2 \leq 30\delta\tau \leq \tau$.
2. For every $x \in Z_{n,d}$,

$$\mathsf{dist}_2\left(F_{d,\tau}(x), \mathrm{span}\{F_{d,\tau}(y)\}_{y \in Z_{n,d} \setminus B(x,\tau/4)}\right) \geq \delta\tau.$$

The existence of such a map follows from Theorem 1.5. (For those worried about measurability, note that since X is finite, we may actually choose a finite collection of projections $P^{(1)}_{n,k}, \ldots, P^{(N)}_{n,k}$ in the arguments that follow. In fact, $N = O(\log n)$ suffices; see [JL84, MS86]. In this case, choosing a random projection consists of choosing some $P^{(i)}_{n,k}$ uniformly at random.)

Observe that for every $x \in X$, $F_{d,\tau}(P_{n,d}(x))$ is an L_2-valued random variable. Let H be the Hilbert space of L_2-valued random variables over the probability space on which $\{F_{d,\tau}(P_{n,d}(x))\}_{x \in X}$ is defined, equipped with norm $\|X\|_H = \sqrt{\mathbb{E}\,\|X\|_2^2}$. We define $\psi_\tau : X \to H$ by $\psi_\tau(x) = F_{d,\tau}(P_{n,d}(x))$. First, we have, for every $x \in X$, $\|\psi_\tau(x)\|_H \leq \tau$ since $\|F_{d,\tau}(y)\|_2 \leq \tau$ for every $y \in \mathbb{R}^d$. Secondly, for $x, y \in X$,

$$\begin{aligned}
\|\psi_\tau(x) - \psi_\tau(y)\|_H^2 &= \mathbb{E}\,\|F_{d,\tau}(P_{n,d}(x)) - F_{d,\tau}(P_{n,d}(y))\|_2^2 \\
&\leq \mathbb{E}\,\|P_{n,d}(x) - P_{n,d}(y)\|_2^2 \\
&= \|x - y\|_2^2,
\end{aligned}$$

implying that $\|\psi_\tau\|_{\mathrm{Lip}} \leq 1$.

Now, fix $x \in X$, and consider any subset $S \subseteq X$ with $|S| = k$. From [JL84], by choosing $d = O(\log k)$ large enough, we know that with probability at least $\frac{1}{2}$ over the choice of $P_{n,d} : \mathbb{R}^n \to \mathbb{R}^d$, we have, for every $y \in S$, $\|P_{n,d}(x) - P_{n,d}(y)\|_2 \geq \frac{1}{2}\|x - y\|_2$. Call this event $\mathcal{E}_{x,S}$. It follows that

$$\begin{aligned}
&\mathsf{dist}_2\left(\psi_\tau(x), \mathrm{span}\{\psi_\tau(y)\}_{y \in S \setminus B(x,\tau)}\right)^2 \\
&\quad= \mathbb{E}\left[\mathsf{dist}_2\left(F_{\tau,d}(P_{n,d}(x)), \mathrm{span}\{F_{\tau,d}(P_{n,d}(y))\}_{y \in S \setminus B(x,\tau)}\right)^2\right] \\
&\quad\geq \tfrac{1}{2}\mathbb{E}\left[\mathsf{dist}_2\left(F_{\tau,d}(P_{n,d}(x)), \mathrm{span}\{F_{\tau,d}(P_{n,d}(y))\}_{y \in S \setminus B(x,\tau)}\right)^2 \,\middle|\, \mathcal{E}_{x,S}\right] \\
&\quad\geq \tfrac{1}{2}\mathbb{E}\left[\mathsf{dist}_2\left(F_{\tau,d}(P_{n,d}(x)), \mathrm{span}\{F_{\tau,d}(z)\}_{z \in Z_{n,d} \setminus B(P_{n,d}(x),\tau/4)}\right)^2 \,\middle|\, \mathcal{E}_{x,S}\right] \\
&\quad\geq \tfrac{1}{2}\delta^2\tau^2,
\end{aligned}$$

where the penultimate line follows from the fact that, conditioned on $\mathcal{E}_{x,S}$, $y \in S \setminus B(x,\tau) \implies P_{n,d}(y) \in Z \setminus B(P_{n,d}(x), \tau/4)$. We conclude that

$$\mathsf{dist}_2\left(\psi_\tau(x), \mathrm{span}\{\psi_\tau(y)\}_{y \in S \setminus B(x,\tau)}\right) \geq \frac{\tau}{O(\sqrt{\log k})}.$$

Applying Theorem 3.1 to the ensemble $\{\psi_\tau : X \to L_2\}_{\tau \geq 0}$, with $\mathcal{A} = \{S \subseteq X : |S| \leq k\}$, we conclude the existence of a map $\varphi : X \to L_2$ with $\mathsf{rigidity}_{\leq k}(\varphi) \leq O(\sqrt{\log n}(\log k)^{1/4})$. □

4.3 Discussion

We remark that to prove $r_2(X) \leq O(\sqrt{\log |X|})$ for any finite subset $X \subseteq L_2$, it suffices to show the existence of the following ensemble of maps. For every $\tau \geq 0$, there should exist a 1-Lipschitz map $\psi_\tau : X \to L_2$ satisfying both $\|\psi_\tau(x)\|_2 \leq O(1)\tau$ for all $x \in X$, and

$$\mathsf{dist}_2\left(\psi_\tau(x), \mathrm{span}\{\psi_\tau(y)\}_{y \in S \setminus B(x,\tau)}\right) \geq \Omega(1) \cdot \frac{\tau}{\sqrt{1 + \log \frac{|B(x,\alpha\,\tau)|}{|B(x,\beta\,\tau)|}}},$$

for every $x \in X$, and with $\alpha > \beta > 0$ two fixed constants. One may now apply Theorem 3.5 to finish.

Acknowledgements. We thank Sanjeev Arora and Assaf Naor for relevant discussions during the work of [ALN08], and Satish Rao for related conversations over the years.

References

[ALN08] Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21(1):1–21, 2008.

[Bar96] Yair Bartal. Probabilistic approximations of metric space and its algorithmic application. In *37th Annual Symposium on Foundations of Computer Science*, pages 183–193, October 1996.

[Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[CKR01] Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 8–16, Philadelphia, PA, 2001.

[DV01] J. Dunagan and S. Vempala. On Euclidean embeddings and bandwidth minimization. In *Randomization, approximation, and combinatorial optimization*, pages 229–240. Springer, 2001.

[Fei00] Uriel Feige. Approximating the bandwidth via volume respecting embeddings. *J. Comput. System Sci.*, 60(3):510–539, 2000.

[FRT03] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 448–455, 2003.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science*, pages 534–543, 2003.

[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.

[Kir34] M. D. Kirszbraun. Über die zusammenziehenden und Lipschitzchen Transformationen. *Fund. Math.*, (22):77–108, 1934.

[KLM04] Robert Krauthgamer, Nathan Linial, and Avner Magen. Metric embeddings–beyond one-dimensional distortion. *Discrete Comput. Geom.*, 31(3):339–356, 2004.

[KLMN05] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: A new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005.

[Lee05] James R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 92–101, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

[LS93] Nathan Linial and Michael Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993.

[Mat02] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[MS86] Vitali D. Milman and Gideon Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[Rao99] Satish Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999.

[Vem98] Santosh Vempala. Random projection: A new approach to VLSI layout. In *39th Annual Symposium on Foundations of Computer Science*, 1998.