Euclidean distortion and the Sparsest Cut

[Extended abstract]

Sanjeev Arora*
Princeton University
arora@cs.princeton.edu

James R. Lee†
U.C. Berkeley
jrl@cs.berkeley.edu

Assaf Naor
Microsoft Research
anaor@microsoft.com

ABSTRACT

We prove that every n-point metric space of negative type (in particular, every n-point subset of L_1) embeds into a Euclidean space with distortion $O(\sqrt{\log n}\log\log n)$, a result which is tight up to the $O(\log\log n)$ factor. As a consequence, we obtain the best known polynomial-time approximation algorithm for the Sparsest Cut problem with general demands. If the demand is supported on a subset of size k, we achieve an approximation ratio of $O(\sqrt{\log k}\log\log k)$.

Categories and Subject Descriptors

F.2.0 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity—*General*

General Terms

Algorithms, Theory

Keywords

Approximation Algorithms, Sparsest Cut, Semidefinite Programming, Metric Embeddings

1. INTRODUCTION

Geometric embeddings of finite metric spaces, a topic originally studied in geometric analysis, became an integral part of theoretical computer science following work of Linial, London, and Rabinovich [20]. They gave an algorithmic version of a result of Bourgain [6] which shows that every n-point metric space embeds into L_2 with distortion $O(\log n)$. This geometric viewpoint offers a way to understand the approximation ratios achieved by linear programming (LP) and semidefinite programming (SDP) relaxations for cut problems [20, 4]. It soon became apparent that further progress in understanding SDP relaxations would involve improving Bourgain's general bound of $O(\log n)$ for n-point metric spaces of negative type. For instance, the approximation ratio achieved by a well-known SDP relaxation for the general Sparsest Cut problem is known to coincide exactly with the best-possible distortion bound achievable for the embedding of n-point metrics of negative type into L_1—a striking connection between pure mathematics and algorithm design.

Further progress on these problems required new insights into the structure of metric spaces of negative type, and the design of more sophisticated and flexible embedding methods for finite metrics. Coincidentally, significant progress was made recently on both these fronts. Arora, Rao and Vazirani [3] proved a new structural theorem about metric spaces of negative type and used it to design an $O(\sqrt{\log n})$-approximation algorithm for *uniform case* of the Sparsest Cut Problem. Krauthgamer, Lee, Mendel and Naor [16] introduced a new embedding method called *measured descent* which unified and strengthened many existing embedding techniques, and they used it to solve a number of open problems in the field.

These breakthroughs indeed resulted in improved embeddings for negative type metrics; Chawla, Gupta, and Räcke [8] used the structural theorem of [3] (specifically, its stronger form in Lee [17]), in conjunction with measured descent to show that every n-point metric of negative type embeds into L_2 with distortion $O(\log n)^{3/4}$. In the present work, we show how one can achieve distortion $O(\sqrt{\log n}\log\log n)$. This almost matches the 35-year-old lower bound of $\sqrt{\log n}$ from Enflo [10]. Our methods use the results of [3, 17, 8] essentially as a "black box," together with a delicate enhancement of the measured descent technique.

Recall that a metric space (X, d) is said to be of *negative type* if $(X, \sqrt{d})$ is isometric to a subset of Euclidean space. In particular, it is well known that L_1 is of negative type. (We also remind the reader that L_2 is isometrically equivalent to a subset of L_1.) The parameter $c_2(X)$, known as the *Euclidean distortion* of X, is the least distortion with which X embeds into Hilbert space, i.e. it is the minimum of $\mathrm{dist}(f) = \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$ over all bijections $f : X \hookrightarrow L_2$. The mathematical investigation of the problem we study dates back to the work of Enflo [10], who showed that the Euclidean distortion of the Hamming cube $\Omega_d = \{0,1\}^d$ equals $\sqrt{d} = \sqrt{\log_2 |\Omega_d|}$.

*Supported by David and Lucile Packard Fellowship and NSF grant CCR-0205594.

†Supported by NSF grant CCR-0121555 and an NSF Graduate Research Fellowship. Portions of this work were done while the author was an intern in the Microsoft Research Theory Group.

The following natural question is folklore in geometric and functional analysis:

"Is the discrete d-dimensional hypercube the most non-Euclidean 2^d-point subset of L_1?"

A positive answer to this question would imply that any n-point subset of L_1 embeds in L_2 with distortion $O(\sqrt{\log n})$. In fact, motivated by F. John's theorem in convex geometry (see [24]), Johnson and Lindenstrauss [14] asked in 1983 whether every n-point metric space embeds into L_2 with distortion $O(\sqrt{\log n})$. Here, the analogy between finite dimensional normed spaces and finite metric spaces is not complete: Bourgain [6] has shown that for any n-point metric space X, $c_2(X) = O(\log n)$, and this result is existentially optimal [20, 4]. By now we understand that finite metric spaces (namely expander graphs) can exhibit an isoperimetric profile which no normed space can achieve, and this is the reason for the discrepancy with John's theorem. However, it is known (see [16]) that several natural restricted classes of metrics do adhere to the $O(\sqrt{\log n})$ Euclidean distortion suggested by John's theorem. (Additional remarks on relationships to Banach spaces can be found in Section 5.) Arguably, for applications in theoretical computer science, the most important restricted class of metrics are those of negative type, yet improvements over Bourgain's theorem for such metrics have long resisted the attempts of mathematicians and computer scientists.

The present paper is devoted to proving that up to double logarithmic factors, the answer to the above question is positive. This yields a general tool for the rounding of certain classes of semi-definite programs. As a result, we obtain the best-known polynomial time algorithm for the approximation of the Sparsest Cut problem with general demands, improving over the previous bounds due to [8] and the preceding works [20] and [4] (which yield an $O(\log n)$ approximation). This problem is described in Section 4. We now state our main result.

THEOREM 1.1. *Let (X, d) be an n-point metric space of negative type. Then*

$$c_2(X) = O\left(\sqrt{\log n} \cdot \log\log n\right).$$

In Section 2.2, we present a high-level overview of the proof.

Related work. Until recently, there was no real evidence to the conjecture that any n-point subset of L_1 embeds in Hilbert space with distortion $O(\sqrt{\log n})$. In the paper [18], Lee, Mendel and Naor show that any n-point subset of L_1 embeds into Hilbert space with *average distortion* $O(\sqrt{\log n})$. Arora, Rao, and Vazirani [3] have shown that $O(\sqrt{\log n})$ distortion is achievable using a different notion of average distortion, which turns out to be more relevant for bounding the actual distortion. As described above, combining their result with the measured descent technique of Krauthgamer, Lee, Mendel and Naor [16], Chawla, Gupta, and Räcke [8] have recently proved that for any n-point metric space X of negative type, $c_2(X) = O(\log n)^{3/4}$. It was conjectured [22, pg. 379] that n-point metrics of negative type embed into L_1 with distortion $O(1)$. Recently, Khot and Vishnoi [15] have obtained a lower bound of $\Omega(\log\log n)^{\delta}$, for some constant $\delta > 0$.

Our results also suggest that the dimension reduction lower bound of Brinkman and Charikar [7] (see also [19]) is tight for certain distortions. They show that embedding certain n-point subsets of L_1 into ℓ_1^d with distortion D requires that $d \geq n^{\Omega(1/D^2)}$. Theorem 1.1, together with theorems of Johnson and Lindenstrauss [14] and Figiel, Lindenstrauss, and Milman [12], yields an embedding of every n-point subset of L_1 into $\ell_1^{O(\log n)}$ with distortion $O(\sqrt{\log n}\log\log n)$.

2. PRELIMINARIES

In this section we present Theorem 2.1, which is one of the main tools used in the proof of Theorem 1.1. It is a concatenation of the result of Arora, Rao, and Vazirani [3], its strengthening by Lee [17], and the "reweighting" method of Chawla, Gupta, and Räcke [8], who use it in conjunction with [16] to achieve distortion $O(\log^{\frac{3}{4}} n)$. We now present a high-level sketch of the proof. Complete details can be found in the full version of [17], where a more general result is proved; the statement actually holds for metric spaces which are quasisymmetrically equivalent to subsets of Hilbert space, and not only for those of negative type. (See [13] for the definition of quasi-symmetry; the relevance of such maps to the techniques of [3] was first pointed out in [26]).

THEOREM 2.1. *There exist constants $C \geq 1$ and $0 < p < \frac{1}{2}$ such that for every n-point metric space (Y, d) of negative type and every $\Delta > 0$, the following holds. There exists a distribution μ over subsets $U \subseteq Y$ such that for every $x, y \in Y$ with $d(x, y) \geq \frac{\Delta}{16}$,*

$$\mu\left(U : y \in U \text{ and } d(x, U) \geq \frac{\Delta}{C\sqrt{\log n}}\right) \geq p.$$

PROOF (HIGH-LEVEL SKETCH). Let $g : Y \to \ell_2$ be such that

$$d(x, y) = ||g(x) - g(y)||_2^2$$

for all $x, y \in Y$. By [23], there exists a map $T : \ell_2 \to \ell_2$ such that $||T(z)||_2 \leq \sqrt{\Delta}$ for all $z \in \ell_2$ and

$$\frac{1}{2} \leq \frac{||T(z) - T(z')||_2}{\min\{\sqrt{\Delta}, ||z - z'||_2\}} \leq 1.$$

for all $z, z' \in \ell_2$. As in [17], we let $f : Y \to \mathbb{R}^n$ be the map given by $f = T \circ g$ (we remark that this map can be computed efficiently). Then f is a bi-Lipschitz embedding (with distortion 2) of the metric space $(Y, \sqrt{\min\{\Delta, d\}})$ into the Euclidean ball of radius Δ.

Let $0 < \sigma < 1$ be some constant. The basic idea is to choose a random $u \in S^{n-1}$ and define

$$\begin{aligned} L_u &= \{x \in Y : \langle x, u\rangle \leq \tfrac{-\sigma\sqrt{\Delta}}{\sqrt{n}}\} \\ R_u &= \{x \in Y : \langle x, u\rangle \geq \tfrac{\sigma\sqrt{\Delta}}{\sqrt{n}}\}. \end{aligned}$$

One then prunes the sets by iteratively removing any pairs of nodes $x \in L_u$, $y \in R_u$ with $d(x, y) \leq \Delta/\sqrt{\log n}$. At the end one is left with two sets L'_u, R'_u. The main result of [3, 17] is that with high probability (over the choice of u), the number of pairs pruned from $L_u \times R_u$ is not too large.

Let $S_\Delta = \{(x, y) \in Y \times Y : d(x, y) \geq \frac{\Delta}{16}\}$. The reweighting idea of [8] is to apply the above procedure to a weighted version of the point set as follows. Let $w : Y \times Y \to \mathbb{Z}^+$ be an integer-valued weight function on pairs, with $w(x, y) =$

$w(y,x)$, $w(x,x) = 0$, and $w(x,y) > 0 \iff (x,y) \in S_\Delta$. This weight function can be viewed as yielding a new set of points where each point x is replaced by $\sum_{y\in Y} w(x,y)$ copies, with $w(x,y)$ of them corresponding to the pair (x,y). One could think of running the above procedure on this new point set; note that the pruning procedure above may remove some or all copies of x. Then, as observed in [8], the theorems of [3, 17] imply that with high probability, after the pruning, we still have

$$\sum_{x\in L'_u, y\in R'_u} w(x,y) \ge \tfrac{1}{8} \sum_{x,y} w(x,y).$$

The distribution μ mentioned in the statement of the theorem is defined using a family of $O(\log n)$ weight functions described below. Sampling from μ consists of picking a weight function from this family and a random direction $u \in S^{n-1}$, and then forming sets L'_u, R'_u as above using the weight function. One then outputs the set U of all points x for which any "copy" falls into L'_u.

Now we define the family of weight functions. The first one has $w(x,y) = n^4$ for all $(x,y) \in S_\Delta$. Given any weight function, obtain another one by doing the thought experiment of picking a random u and constructing L'_u, R'_u for the point set described by this weight function. For every pair $(x,y) \in S_\Delta$ that is left unpruned with probability $\ge \frac{1}{10}$ (i.e., is in $L'_u \times R'_u$ for $\ge 1/10$ of the $u \in S^{n-1}$), we lower the weight of $w(x,y)$ by a factor 2. A simple argument (presented in [8]) shows that by repeating this $O(\log n)$ times we obtain $O(\log n)$ weight functions such that for every pair $(x,y) \in S_\Delta$ the following is true: If one picks a random weight function and a random direction $u \in S^{n-1}$ and forms the set L'_u, R'_u according to the weight, then with constant probability (over the choice of weight function and $u \in S^{n-1}$), we have $(x,y) \in L'_u \times R'_u$. $\square$

2.1 Padded decomposability and random zero sets

Theorem 2.1 is the only way the negative type property will be used in what follows. It is therefore helpful to introduce it as an abstract property of metric spaces. Let (X,d) be an n-point metric space.

Definition 2.2 (Random zero-sets). *Given Δ, $\zeta > 0$, and $p \in (0,1)$ we say that X admits a random zero set at scale Δ which is ζ-spreading with probability p if there is a distribution μ over subsets $Z \subseteq X$ such that for every $x, y \in X$ with $d(x,y) \ge \Delta$,*

$$\mu \quad Z \subseteq X : \; y \in Z \text{ and } d(x,Z) \ge \frac{\Delta}{\zeta} \quad \ge p.$$

We denote by $\zeta(X;p)$ the least $\zeta > 0$ such that for every $\Delta > 0$, X admits a random zero set at scale Δ which is ζ-spreading with probability p. Finally, given $k \le n$ we define

$$\zeta_k(X;p) = \max_{\substack{Y\subseteq X\\ |Y|\le k}} \zeta(Y;p).$$

With this definition, Theorem 2.1 implies that there exists a universal constant $p \in (0,1)$ such that for every n-point metric space (X,d) of negative type, $\zeta(X;p) = O(\sqrt{\log n})$.

We now recall the related 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 [16] we define the *modulus of padded decomposability* of X, denoted α_X, as the least constant $\alpha > 0$ such that for every $\Delta > 0$ 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 $\mathrm{diam}(C) < \Delta$.

2. For every $x \in X$ we have that

$$\nu\{P : \; B(x, \Delta/\alpha) \subseteq P(x)\} \ge \frac{1}{2}.$$

As observed in [16], the results of [21, 5] imply that $\alpha_X = O(\log|X|)$, and this will be used in our proof.

The following useful fact relates the notions of padded decomposability and random zero sets. Its proof is motivated by an argument of Rao [28].

Fact 2.3. $\zeta(X;1/8) \le \alpha_X$.

Proof. Fix $\Delta > 0$ and let P be a partition of X into subsets of diameter less than Δ. Given $x \in X$ we denote by $\pi_P(x)$ the largest radius r for which $B(x,r) \subseteq P(x)$. Let $\{\varepsilon_C\}_{C\in P}$ be i.i.d. symmetric $\{0,1\}$-valued Bernoulli random variables. Let Z_P be a random subset of X given by

$$Z_P = \bigcup_{C\in P:\; \varepsilon_C = 0} C \,.$$

If $x, y \in X$ satisfy $d(x,y) \ge \Delta$ then $P(x) \ne P(y)$. It follows that

$$\Pr[y \in Z_P \wedge d(x, Z_P) \ge \pi_P(x)] \ge \frac{1}{4}.$$

By the definition of α_X, there exists a distribution over partitions P of X into subsets of diameter less than Δ such that for every $x \in X$ with probability at least $1/2$, $\pi_P(x) \ge \Delta/\alpha_X$. The required result now follows by considering the random zero set Z_P. $\square$

2.2 Proof overview and connection to past work

Apart from Theorem 2.1, our presentation is self-contained. In the informal description that follows, we omit unimportant constants, floors, ceilings, etc. in order to focus on the essential ideas.

Let (X,d) be an arbitrary n-point metric space. First, we recall that using $\alpha_X = O(\log n)$ it is easy to show that $c_2(X) = O(\log^{\frac{3}{2}} n)$ for all finite metric spaces via the "trivial" concatenation technique, where one uses a new set of coordinates for each of the $O(\log \Phi)$ relevant scales $\Delta = 2^k$. A single scale is handled by forming the map $f_k : X \to L_2$ given by $f_k(x) = d(x, Z_k)$, where Z_k is a random zero-set as in Definition 2.2. Using a standard contraction trick (see Matousek's survey chapter [22]), the dependence on $O(\log \Phi)$ is reduced to a dependence on $O(\log n)$.

To obtain Bourgain's stronger bound $c_2(X) = O(\log n)$, Krauthgamer et al. [16] introduce a nontrivial way to glue together the distributions arising from various scales. Let $\rho(x,R) = \log \frac{|B(x,R)|}{|B(x,R/4)|}$ be the "local volume growth" at x. In essence, the method of measured descent [16] shows that, given ζ-spreading zero-sets for each scale 2^k, it is possible to construct a map $\varphi : X \to L_2$ which is $O(\sqrt{\log n})$-Lipschitz and satisfies the following. For every $k \in \mathbb{Z}$ and every $x, y \in X$ with $d(x, y) \approx 2^k$,

$$||\varphi(x) - \varphi(y)||_2 \geq \sqrt{\rho(x, 2^k)} \cdot \frac{2^k}{\zeta}$$

Using $\alpha_X = O(\log n)$, one can derive $\zeta \approx O(\log n)$, which yields distortion $O(\log^{\frac{3}{2}} n)/\sqrt{\rho(x, 2^k)}$; this is again $\Omega(\log^{\frac{3}{2}} n)$ in the worst case.

Using the decomposition lemma of [11], it is possible to obtain $\zeta \approx \rho(x, 2^k)$. The resulting distortion for the pair x, y is $O(\sqrt{(\log n)\rho(x, 2^k)}) = O(\log n)$, recovering Bourgain's bound. Combining this gluing technique with the improved zero-sets available for metrics of negative type (Theorem 2.1), it is possible to achieve distortion $O(\log^{\frac{3}{4}} n)$ [8]. To see this, note that when $\rho(x, 2^k) \leq \sqrt{\log n}$, the above bound is $O(\log^{\frac{3}{4}} n)$. On the other hand, when $\rho(x, 2^k) \geq \sqrt{\log n}$, one uses the negative type assumption to achieve $\zeta \approx \sqrt{\log n}$, and the distortion is again $O(\log^{\frac{3}{4}} n)$. In order to do better, we must dispense with the auxiliary embedding corresponding to $\zeta \approx \rho(x, 2^k)$, and instead employ a more delicate technique.

If we could somehow achieve $\zeta \approx \sqrt{\rho(x, 2^k)}$, then clearly we would obtain distortion $O(\sqrt{\log n})$ as the factors of $\rho(x, 2^k)$ would cancel. It is currently unknown whether such distributions exist. The obstacle lies in the intrinsically "non-local" structure of the Arora-Rao-Vazirani chaining argument [3]. Instead, we try to simulate a contribution of $2^k/\sqrt{\rho(x, 2^k)}$ by applying Theorem 2.1 to localized random samples of the space whose size n' satisfies $n' \ll n$. Ideally the samples relevant to x would have $n' \approx \exp(\rho(x, 2^k))$ points so that $\sqrt{\log n'} \approx \sqrt{\rho(x, 2^k)}$. On the other hand, the samples must be dense enough so that the locally constructed map admits a useful extension to the entire scale-2^k neighborhood of x. Making matters more difficult, the localization and sampling processes must vary smoothly across the entire space (to maintain the Lipschitz property), and must be intimately intertwined with the descent construction across all scales. To facilitate this requires a more delicate gluing procedure, which is carried out in Section 3.1.

3. PROOF OF THEOREM 1.1

The main technical result of this paper is contained in the following lemma.

LEMMA 3.1 (ENHANCED DESCENT). *Let (X, d) be an n-point metric space and fix $p \leq 1/8, K \geq 2$ and $\zeta \geq \zeta_K(X; p)$. For every $m \in \mathbb{Z}$, let $S_m(K)$ be the set*

$$x \in X : \; B \;\; x, 2^{m+5}\alpha_X \;\; \leq \frac{K}{16} \cdot \; B \;\; x, \frac{2^{m-9}}{\zeta} \qquad .$$

Then there exists a mapping $\phi : X \to L_2$ such that

1. *$||\phi(x) - \phi(y)||_2^2 \leq O(\log n \log \alpha_X)\, d(x, y)^2$ for all $x, y \in X$,*
2. *For all $m \in \mathbb{Z}$, $x \in S_m(K)$ and $y \in X$ such that $d(x, y) \in [2^{m-1}, 2^m]$,*

$$||\phi(x) - \phi(y)||_2^2 \;\; \geq \;\; \frac{p}{64}^{\;5} \cdot \frac{d(x, y)^2}{\zeta^2} \cdot \log \frac{|B(x, 2^{m+5}\alpha_X)|}{|B(x, 2^{m+3}/\zeta)|} \qquad (1)$$

Before proving Lemma 3.1 we show how it implies Theorem 1.1. The argument below actually yields more general results. For example if we assume that $\zeta_k(X; p) = O(\log k)^\theta$ for some $p \in (0, 1/8)$, $\theta \geq \frac{1}{2}$ and all $k \leq n$ then

$$\begin{aligned} c_2(X) &= O_p((\log n)^\theta \sqrt{\log \alpha_X \log \log n}) \\ &= O_p((\log n)^\theta \log \log n), \end{aligned}$$

where $O_p(\cdot)$ may contain an implicit constant which depends only on p.

PROOF OF LEMMA 3.1 $\Longrightarrow$ THEOREM 1.1. Combining Theorem 2.1, Lemma 3.1 and the fact that $\alpha_X = O(\log n)$ we obtain the following statement. There exists a constant $A > 0$ such that for every $K \geq 2$ there is a mapping $\phi : X \to L_2$ satisfying the following conditions.

1. $\|\phi(x) - \phi(y)\|_2^2 \leq A \log n \cdot \log \log n \cdot d(x, y)^2$ for all $x, y \in X$.
2. Define $S'_m(K)$ to be the set

$$u \in X : \; B \;\; u, 2^{m+5}\alpha_X \;\; \leq K \cdot \; B \;\; u, \frac{2^m}{A\sqrt{\log K}} \qquad .$$

Then for all $m \in \mathbb{Z}$, $x \in S'_m(K)$ and $y \in X$ such that $d(x, y) \in [2^{m-1}, 2^m]$,

$$||\phi(x) - \phi(y)||_2^2 \geq \frac{d(x, y)^2}{A \log K} \cdot \log \frac{|B(x, 2^{m+5}\alpha_X)|}{|B(x, A2^m/\sqrt{\log K})|}.$$

Observe that for every $m \in \mathbb{Z}$, $S'_m(n) = X$. Hence, defining $K_0 = n$ and $K_{j+1} = K_j^{1/A^4}$, as long as $K_j \geq 2$, we obtain mappings $\phi_0, \ldots, \phi_j : X \to L_2$ satisfying

1. $||\phi_j(x) - \phi_j(y)||_2^2 \leq A \log n \cdot \log \log n \cdot d(x, y)^2$ for all $x, y \in X$.
2. For all $x \in S_m(K_j) \setminus S_m(K_{j+1})$ and $y \in X$ such that $d(x, y) \in [2^{m-1}, 2^m]$:

$$\begin{aligned} &||\phi_j(x) - \phi_j(y)||_2^2 \\ &\geq \frac{d(x, y)^2}{A \log K_j} \cdot \log \frac{|B(x, 2^{m+5}\alpha_X)|}{|B(x, A2^m/\sqrt{\log K_j})|} \\ &= \frac{d(x, y)^2}{A \log K_j} \cdot \log \frac{|B(x, 2^{m+5}\alpha_X)|}{B \;\; x, A2^m/\sqrt{\log K_{j+1}^{A^4}}} \\ &= \frac{d(x, y)^2}{A \log K_j} \cdot \log \frac{|B(x, 2^{m+5}\alpha_X)|}{|B(x, 2^m/(A\sqrt{\log K_{j+1}}))|} \\ &\geq \frac{d(x, y)^2}{A \log K_j} \cdot \log K_{j+1} = \frac{d(x, y)^2}{A^5}. \end{aligned}$$

This procedure ends after N steps, where $N \leq \frac{\log \log n}{\log A}$. Every $x \in S_m(K_N)$ satisfies

$$|B(x, 2^{m+5}\alpha_X)| \leq e^{A^4} |B(x, 2^{m+1}/A)|.$$

By the result of [16] there is a mapping $\phi_{N+1} : X \to L_2$ which is Lipschitz with constant $O(\sqrt{\log n})$ and for every $x, y \in S_m(K_N)$, $\|\phi_{N+1}(x) - \phi_{N+1}(y)\|_2 \geq \Omega(1) \cdot d(x, y)$.

Consider the map $\Phi = \bigoplus_{j=0}^{N+1} \phi_j$, which is Lipschitz with constant $O \;\; \sqrt{\log n} \cdot \log \log n$. For every $x, y \in X$ choose $m \in \mathbb{Z}$ such that $d(x, y) \in [2^{m-1}, 2^m]$. If $x, y \in S_m(K_N)$ then

$$\|\Phi(x) - \Phi(y)\|_2 \geq \|\phi_{N+1}(x) - \phi_{N+1}(y)\|_2 \geq \Omega\,(1) \cdot d(x, y).$$

Otherwise, there is $j \in \{0, \dots, N-1\}$ such that $x \in S_m(K_j) \setminus S_m(K_{j+1})$, in which case

$$\|\Phi(x) - \Phi(y)\|_2 \geq \|\phi_{j+1}(x) - \phi_{j+1}(y)\|_2 \geq \Omega(1) \cdot d(x, y).$$

□

3.1 Proof of Lemma 3.1: Enhanced descent

We begin with a simple definition.

DEFINITION 3.2. *For every $x \in X$ and $t > 0$ define*

$$\kappa(x, t) = \max\{\kappa \in \mathbb{Z} : |B(x, 2^\kappa)| < 2^t\}. \tag{2}$$

The following simple lemma says that the values of $\kappa(\cdot, \cdot)$ cannot change too rapidly when moving between nearby points. This fact will be used several times in the ensuing arguments, and played a similar role in [16]. We defer the proof to the Appendix.

LEMMA 3.3 (SMOOTHNESS). *For $x \in X$, let $i \in \mathbb{Z}$ and $m, t \in \mathbb{Z}^+$ be such that $B(x, 2^{i+m-1}) \leq 2^t \leq B(x, 2^{i+m})$. Then every $z \in X$ for which $d(x, z) \leq 2^{\min\{m, m+i-2\}}$ satisfies:*

$$\kappa(z, t) \in \{m + i - 3, m + i - 2, m + i - 1, m + i, m + i + 1\}.$$

Notation. We introduce some notation which will be used in the forthcoming proofs. Write $\alpha = \alpha_X$, and for every $j \in \mathbb{Z}$ let P_j denote a random partition of X satisfying the following.

1. For all $C \in P_j$ we have that $\operatorname{diam}(C) \leq 2^{j+4}\alpha$.
2. For every $x \in X$ we have that $\nu\{P : B(x, 2^{j+4}) \subseteq P_j(x)\} \geq \frac{1}{2}$.

We also fix $p \in (0, 1/8)$ and for every $k \leq n$ let $\zeta_k = \zeta_k(X; p)$. For $S \subseteq X$ let $\Psi_j(S)$ denote a random zero set of S at scale 2^{j-3} which is $\zeta_{|S|}$-spreading with probability p. For each $C \subseteq X$ let $\widetilde{C}$ be a uniformly random subset of C of size $\min\{K, |C|\}$.

The distribution on Fréchet-type embeddings. The embeddings we produce will be of Fréchet-type, i.e. every coordinate $f_i : X \to \mathbb{R}$ will be of the form $f_i(x) = d(x, U)$ for some $U \subseteq X$. Let $I = [-\log_2 \zeta_K + 3, \log_2 \alpha + 6] \cap \mathbb{Z}$ and $J = \{0, 1, \dots, \log_2 n\}$. For each $i \in I$ and $t \in J$, we describe a distribution W_t^i on sets. Our random embedding consists of mapping x to $f(x) = (d(x, W_t^i) : i \in I, t \in J)$. Such a mapping is clearly Lipschitz with constant $\sqrt{|I| \cdot |J|} = O(\sqrt{\log n \cdot \log \alpha})$ (here we use Fact 2.3, i.e. $\zeta_K \leq \alpha$).

Let $\{\sigma_m\}_{m \in \mathbb{Z}}$ be a sequence of random variables taking each of the values $\{0, 1, 2\}$ with probability $\frac{1}{3}$, which are independent of all the other random variables appearing in this proof. (In general, the reader should assume that samplings from various distributions are independent of one another.) Then the random subset W_t^i is defined as

$$\begin{aligned} W_t^i = \Big\{ & x \in X : \sigma_{\kappa(x,t)-i} = 0 \quad \text{or} \\ & x \in \Psi_{\kappa(x,t)-i} \quad \widetilde{P_{\kappa(x,t)-i}}(x) \quad \text{and } \sigma_{\kappa(x,t)-i} = 1 \quad \text{or} \\ & x \in \widetilde{P_{\kappa(x,t)-i}}(x) \text{ and } \sigma_{\kappa(x,t)-i} = 2 \quad \Big\}. \end{aligned}$$

For the rest of the proof, let m be any integer, fix $x, y \in X$ such that $d(x, y) \in [2^{m-1}, 2^m]$, and assume that $x \in S_m(K)$. Let $s_i = \log_2 |B(x, 2^{i+m})|$, with $s_{\min I}$ and $s_{\max I}$ corresponding to the smallest and largest $i \in I$. The next claim follows from the "smoothness" of Lemma 3.3.

CLAIM 3.4. *For $i \in I$ and all $t \in \mathbb{Z} \cap [s_{i-1}, s_i]$, every $w \in B(x, 2^m/\zeta_K)$ satisfies*

$$m - 3 \leq \kappa(w, t) - i \leq m + 1.$$

Now we define

$$\begin{aligned} N(x) &= \#\{(i, t) : i \in I, t \in [s_{i-1}, s_i] \cap \mathbb{Z}\} \\ &= \#\{t : t \in [s_{\min I}, s_{\max I}] \cap \mathbb{Z}\}. \end{aligned}$$

Observe that

$$N(x) \geq \log_2 \frac{|B(x, 2^{m+5}\alpha)|}{|B(x, 2^{m+3}/\zeta_K)|}.$$

We are going to get a contribution to $||f(x) - f(y)||_2^2$ from the sets $\{W_t^i\}$ where $t \in \mathbb{Z} \cap [s_{i-1}, s_i]$ for some $i \in I$. The number of such pairs is $N(x)$. Thus clearly we get the desired lower bound (1) if we can prove that for these values of i and t, we have

$$\mathbb{E}\, |d(x, W_t^i) - d(y, W_t^i)|^2 \geq \frac{p}{64}^5 \cdot \frac{2^{2m}}{\zeta_K}. \tag{3}$$

To prove (3) we fix $i \in I, t \in [s_{i-1}, s_i] \cap \mathbb{Z}$ and let

$$M = \{m - 3, m - 2, m - 1, m, m + 1\}$$

be the range of values from Claim 3.4.

3.1.1 Partitions and padding

For any $j \in M$ we have that $\operatorname{diam}(P_j(x)) \leq 2^{j+4}\alpha \leq 2^{m+5}\alpha$, so $B(x, 2^{m+5}\alpha) \supseteq P_j(x)$. Since $x \in S_m(K)$, it follows that $|P_j(x)| \leq \frac{K}{16}|B(x, 2^{m-9}/\zeta_K)|$. Recall that for $j \in M$ the random partition P_j satisfies

$$\begin{aligned} \Pr[d(x, X \setminus P_j(x)) \geq 2^{m+1}] &\geq \Pr[d(x, X \setminus P_j(x) \geq 2^{j+4}] \\ &\geq 1/2. \end{aligned}$$

Define the event

$$\mathcal{E}_{\text{pad}}^j = \{d(x, X \setminus P_j(x)) \geq 2^{m+1}\},$$

and let

$$\mathcal{E}_{\text{pad}} = \bigcap_{j \in M} \mathcal{E}_{\text{pad}}^j.$$

Note that, by independence, we have $\Pr[\mathcal{E}_{\text{pad}}] \geq 2^{-5}$.

Suppose that $\mathcal{E}_{\text{pad}}^j$ occurs, then since $d(x, y) \leq 2^m$, we have $y \in P_j(x)$, implying $P_j(x) = P_y(x)$. Furthermore, since

$$B(x, 2^{m-9}/\zeta_K) \subseteq P_j(x),$$

when we sample down $P_j(x)$ to $\widetilde{P_j(x)}$, a set of size at most K, Lemma A.1, part (1), ensures that

$$\Pr[\widetilde{P_j(x)} \cap B(x, 2^{m-9}/\zeta_K) = \emptyset] \leq e^{-15}.$$

To this end, we denote

$$\mathcal{E}_{\text{hit}}^j = \left\{\widetilde{P_j(x)} \cap B(x, 2^{m-9}/\zeta_K) \neq \emptyset\right\},$$

and we define $\mathcal{E}_{\text{hit}} = \bigcap_{j \in M} \mathcal{E}_{\text{hit}}^j$. Since the events $\{\mathcal{E}_{\text{hit}}^j\}_{j \in M}$ are independent even after conditioning on $\mathcal{E}_{\text{pad}}$, the preceding discussion yields the following lemma.

LEMMA 3.5. $\Pr(\mathcal{E}_{\mathrm{hit}} \cap \mathcal{E}_{\mathrm{pad}}) \geq 2^{-5}(1-5e^{-15}) > 2^{-6}$.

3.1.2 Obtaining a separation

We introduce the following events which mark different "phases" of the embedding. For $\ell \in \{1,2\}$, let

$$\mathcal{E}_\ell^\sigma = \{\sigma_j = \ell \text{ for all } j \in M\}\,.$$

Note that $\Pr[\mathcal{E}_\ell^\sigma] \geq 3^{-5}$ for each $\ell \in \{1,2\}$. Now we study the distance from x to W_t^i in phase 1.

CLAIM 3.6. *If $\mathcal{E}_1^\sigma \cap \mathcal{E}_{\mathrm{pad}}$ occurs, then*

$$d(x, W_t^i) \geq \min\left(\frac{2^m}{\zeta_K}, \min_{j\in M}\left\{d(x, \Psi_j(\widetilde{P_j(x)}))\right\}\right). \quad (4)$$

PROOF. Fix a point $w \in B(x, 2^m/\zeta_K)$, and let $j = \kappa(w,t)-i$. By Claim 3.4, $j \in M$, hence $\mathcal{E}_{\mathrm{pad}}$ implies that $w \in P_j(x)$. Since $\mathcal{E}_1^\sigma$ occurs, we have $w \in W_t^i$ if and only if $w \in \Psi_j(\widetilde{P_j(x)})$. $\square$

If $\mathcal{E}_{\mathrm{pad}} \cap \mathcal{E}_{\mathrm{hit}}$ occurs, then for each $j \in M$, there exists a point $w_j \in \widetilde{P_j(x)}$ such that $d(x, w_j) \leq 2^{m-9}/\zeta_K$. So to get a lower bound on $d(x, W_t^i)$, we can restrict our attention to $\{w_j\}_{j\in M}$.

CLAIM 3.7. *If $\mathcal{E}_{\mathrm{hit}} \cap \mathcal{E}_{\mathrm{pad}} \cap \mathcal{E}_1^\sigma$ occurs and*

$$d\left(w_j, \Psi_j\left(\widetilde{P_j(x)}\right)\right) \geq \varepsilon$$

for every $j \in M$, then

$$d(x, W_i^t) \geq \min\left(\frac{2^m}{\zeta_K}, \varepsilon - \frac{2^{m-9}}{\zeta_K}\right).$$

PROOF. For every $j \in M$,

$$\begin{aligned} d\left(x, \Psi_j\left(\widetilde{P_j(x)}\right)\right) &\geq d\left(w_j, \Psi_j\left(\widetilde{P_j(x)}\right)\right) - d(x, w_j) \\ &\geq \varepsilon - \frac{2^{m-9}}{\zeta_K}. \end{aligned}$$

Now apply Claim 3.6. $\square$

There are two types of points $y \in X$ which occur in the argument that follows. As a warmup, we first dispense with the easy type.

Type I: There exists $z \in B(y, 2^{m-7}/\zeta_K)$ for which $\kappa(z,t) - i \notin M$.

Fix this z and let $j' = \kappa(z,t)-i$. Assume that $\mathcal{E}_{\mathrm{hit}} \cap \mathcal{E}_{\mathrm{pad}} \cap \mathcal{E}_1^\sigma$ occurs, as well as the independent event $\sigma_{j'} = 0$. Note that using Lemma 3.5 along with independence, the probability of this event is at least $q = 2^{-6} \cdot 3^{-5} \cdot (1/3)$.

Now, applying the definition of ζ_K to the sets $\widetilde{P_j(x)} = \widetilde{P_j(w_j)}$ for $j \in M$, we conclude that there is an event $\mathcal{E}_{\mathrm{zero}}$ which occurs with probability at least p^5, and such that for every $j \in M$,

$$d\left(w_j, \Psi_j\left(\widetilde{P_j(x)}\right)\right) \geq \frac{2^{j-3}}{\zeta_K} \geq \frac{2^{m-6}}{\zeta_K}.$$

Applying Claim 3.7 with $\varepsilon = 2^{m-6}/\zeta_K$, we conclude that, in this case,

$$d(x, W_t^i) \geq \frac{5 \cdot 2^{m-9}}{\zeta_K}.$$

Since $\sigma_{j'} = 0$, we have $z \in W_t^i$, hence with probability at least $q \cdot p^5 \geq (p/16)^5$, we have

$$|d(x, W_t^i) - d(y, W_t^i)| \geq \frac{5\cdot 2^{m-9}}{\zeta_K} - d(y,z) \geq \frac{2^{m-9}}{\zeta_K}.$$

This completes the analysis of Type I points.

3.1.3 A case analysis on the fate of y

We now analyze the complement of the set of Type I points.

Type II: For all $z \in B(y, 2^{m-7}/\zeta_K)$, $\kappa(z,t) - i \in M$.

First, we define the following key event.

$$\mathcal{E}_{\mathrm{close}} = \left\{\exists j \in M,\ \exists z \in \widetilde{P_j(y)} \text{ such that } d(y,z) \leq \frac{2^{m-7}}{\zeta_K} \text{ and } \kappa(z,t) - i = j\right\}. \quad (5)$$

Also, let $\mathcal{E}_{\mathrm{far}} = \neg\mathcal{E}_{\mathrm{close}}$.

These two events concern the distance of y to the various sample sets. Since we do not make the assumption that $y \in S_m(K)$, we cannot argue that a random sample point lands near y with non-negligible probability, thus we must handle both possibilities $\mathcal{E}_{\mathrm{close}}$ and $\mathcal{E}_{\mathrm{far}}$. This is the main purpose of the two phases, i.e. the events $\mathcal{E}_\ell^\sigma$ for $\ell \in \{1,2\}$. Thus the proof now breaks down into two sub-cases corresponding to the occurrences of $\mathcal{E}_{\mathrm{close}}$ and $\mathcal{E}_{\mathrm{far}}$, respectively.

CLAIM 3.8 (THE CLOSE CASE). *Conditioned on the event $\mathcal{E}_{\mathrm{hit}} \cap \mathcal{E}_{\mathrm{pad}} \cap \mathcal{E}_1^\sigma \cap \mathcal{E}_{\mathrm{close}}$ occurring, with probability at least p^5,*

$$|d(x, W_t^i) - d(y, W_t^i)| \geq \frac{2^{m-9}}{\zeta_K}.$$

PROOF. If the event $\mathcal{E}_{\mathrm{close}} \cap \mathcal{E}_{\mathrm{pad}}$ occurs, then there exists some $j_0 \in M$ and $z \in \widetilde{P_{j_0}(x)} = \widetilde{P_{j_0}(y)}$ such that $d(y,z) \leq 2^{m-7}/\zeta_K$ and $\kappa(z,t) - 1 = j_0$. Additionally, if $\mathcal{E}_{\mathrm{pad}} \cap \mathcal{E}_{\mathrm{hit}}$ occurs, then for every $j \in M$ there is $w_j \in \widetilde{P_j(x)}$ with $d(w_j, x) \leq 2^{m-9}/\zeta_K$. It follows that, for every $j \in M$,

$$\begin{aligned} d(w_j, z) &\geq d(x,y) - d(x,w_j) - d(y,z) \\ &\geq 2^{m-1} - \frac{5\cdot 2^{m-9}}{\zeta_K} \\ &\geq 2^{m-2} \\ &\geq 2^{j-3}. \end{aligned}$$

Hence applying the definition of ζ_K to the sets $\widetilde{P_j(x)}$ for $j \in M$, we conclude that there is an event $\mathcal{E}_{\mathrm{zero}}$ with probability at least p^5, independent of $\mathcal{E}_{\mathrm{hit}}, \mathcal{E}_{\mathrm{pad}}, \mathcal{E}_{\mathrm{close}}$ and $\mathcal{E}_1^\sigma$, such that if the event $\mathcal{E}_{\mathrm{hit}} \cap \mathcal{E}_{\mathrm{pad}} \cap \mathcal{E}_{\mathrm{close}} \cap \mathcal{E}_1^\sigma \cap \mathcal{E}_{\mathrm{zero}}$ occurs then $z \in \Psi_{j_0}\left(\widetilde{P_{j_0}(y)}\right)$, and for every $j \in M$,

$$d\left(w_j, \Psi_j\left(\widetilde{P_j(x)}\right)\right) \geq \frac{2^{j-3}}{\zeta_K} \geq \frac{2^{m-6}}{\zeta_K}.$$

Applying Claim 3.7, it follows that $d(x, W_t^i) \geq \frac{5\cdot 2^{m-9}}{\zeta_K}$. Furthermore, since $z \in \Psi_{j_0}\left(\widetilde{P_{j_0}(y)}\right)$ and $\mathcal{E}_1^\sigma$ occurs, we have $\sigma_{j_0} = 1$, hence $z \in W_t^i$. It follows that

$$|d(x, W_t^i) - d(y, W_t^i)| \geq \frac{5\cdot 2^{m-9}}{\zeta_K} - d(y,z) \geq \frac{2^{m-9}}{\zeta_K},$$

completing the proof. $\square$

We now analyze the probability of the previous event.

LEMMA 3.9. $\Pr[\mathcal{E}_{\text{hit}} \cap \mathcal{E}_{\text{pad}} \cap \mathcal{E}_1^\sigma \mid \mathcal{E}_{\text{close}}] \geq 2^{-6} \cdot 3^{-5}$.

PROOF. Since

$$\begin{aligned}\Pr[\mathcal{E}_{\text{hit}} \cap \mathcal{E}_{\text{pad}} \cap \mathcal{E}_1^\sigma \mid \mathcal{E}_{\text{close}}] &= 3^{-5} \cdot \Pr[\mathcal{E}_{\text{hit}} \cap \mathcal{E}_{\text{pad}} \mid \mathcal{E}_{\text{close}}] \\ &= 2^{-5} \cdot 3^{-5} \cdot \Pr[\mathcal{E}_{\text{hit}} \mid \mathcal{E}_{\text{pad}}, \mathcal{E}_{\text{close}}],\end{aligned}$$

we need only argue that $\Pr[\mathcal{E}_{\text{hit}} \mid \mathcal{E}_{\text{pad}}, \mathcal{E}_{\text{close}}] \geq \frac{1}{2}$. But this follows by applying Lemma A.1, part 2(a), to the sets $X = P_j(x)$, $A = B(x, 2^{m-9}/\zeta_K)$, $B = B(y, 2^{m-7}/\zeta_K)$, and concluding that

$$\Pr[\neg\mathcal{E}_{\text{hit}}^j \mid \mathcal{E}_{\text{pad}}, \mathcal{E}_{\text{close}}] \leq e^{(1-15K)/K} \leq e^{-14}.$$

Thus $\Pr[\mathcal{E}_{\text{hit}} \mid \mathcal{E}_{\text{pad}}, \mathcal{E}_{\text{close}}] \geq 1 - 5e^{-14} \geq \frac{1}{2}$. □

Now we proceed to analyze the case when $\mathcal{E}_{\text{far}}$ occurs. By Claim 3.4, every $w \in B(x, 2^{m-9}/\zeta_K)$ satisfies $\kappa(w,t) - i \in M$. By the pigeonhole principle, some $j^* \in M$ must occur as the value of $\kappa(w,t) - i$ in at least a $\frac{1}{5}$th of them. Together with the growth condition implied by $x \in S_m(K)$, we conclude that

$$\begin{aligned}\{w \in B(x, 2^{m-9}/\zeta_K) \ :\ & \kappa(w,t) - i = j^*\} \\ &\geq \frac{16}{5K}|B(x, 2^{m+5}\alpha)| \\ &\geq \frac{3}{K}|B(x, 2^{m+5}\alpha)|.\end{aligned}$$

Define the event $\mathcal{E}_{\text{hit}}^*$ to be

$$\{\exists w \in \widetilde{P_{j^*}(x)} \cap B(x, 2^{m-9}/\zeta_K) \text{ with } \kappa(w,t) - i = j^*\},$$

and observe that by Lemma A.1, part (1), $\Pr[\mathcal{E}_{\text{hit}}^*] \geq 1 - e^{-3}$.

CLAIM 3.10 (THE FAR CASE). *If $\mathcal{E}_{\text{pad}} \cap \mathcal{E}_{\text{far}} \cap \mathcal{E}_2^\sigma \cap \mathcal{E}_{\text{hit}}^*$ occurs, then*

$$|d(x, W_t^i) - d(y, W_t^i)| \geq \frac{2^{m-9}}{\zeta_K}.$$

PROOF. Assume that the event $\mathcal{E}_{\text{pad}} \cap \mathcal{E}_{\text{far}} \cap \mathcal{E}_2^\sigma \cap \mathcal{E}_{\text{hit}}^*$ occurs, and let $w \in \widetilde{P_{j^*}(x)} \cap B(x, 2^{m-9}/\zeta_K)$ be the point guaranteed by $\mathcal{E}_{\text{hit}}^*$. Since $\sigma_j = 2$, we have $w \in W_t^i$, so that $d(x, W_t^i) \leq 2^{m-9}/\zeta_K$).

On the other hand, we claim that $d(y, W_t^i) \geq 2^{m-7}/\zeta_K$. Indeed, first note that $\mathcal{E}_{\text{pad}}$ implies that for all $j \in M$, $d(y, X \setminus P_j(y)) \geq 2^m$. Suppose that $z \in W_t^i$ and $d(y,z) \leq 2^{m-7}/\zeta_K$. Let $j' = \kappa(z,t) - i$. In this case, we have $j' \in M$, hence $\sigma_{j'} = 2$, and this implies that $z \in \widetilde{P_{j'}(z)} = \widetilde{P_{j'}(y)}$. But in this case, $\mathcal{E}_{\text{far}}$ implies that $d(y,z) > 2^{m-7}/\zeta_K$, yielding a contradiction. It follows that

$$|d(x, W_t^i) - d(y, W_t^i)| \geq \frac{2^{m-7}}{\zeta_K} - \frac{2^{m-9}}{\zeta_K} \geq \frac{2^{m-9}}{\zeta_K}.$$

□

LEMMA 3.11. $\Pr[\mathcal{E}_{\text{pad}} \cap \mathcal{E}_2^\sigma \cap \mathcal{E}_{\text{hit}}^* \mid \mathcal{E}_{\text{far}}] \geq 2^{-5} \cdot 3^{-5} \cdot (1 - e^{-3})$.

PROOF.

$$\begin{aligned}\Pr[\mathcal{E}_{\text{pad}} \cap \mathcal{E}_2^\sigma \cap \mathcal{E}_{\text{hit}}^* \mid \mathcal{E}_{\text{far}}] &= 3^{-5} \cdot 2^{-5} \cdot \Pr[\mathcal{E}_{\text{hit}}^* \mid \mathcal{E}_{\text{far}}, \mathcal{E}_{\text{pad}}] \\ &\geq \Pr[\mathcal{E}_{\text{hit}}^* \mid \mathcal{E}_{\text{pad}}] \\ &\geq 3^{-5} \cdot 2^{-5} \cdot (1 - e^{-3}).\end{aligned}$$

The penultimate inequality follows from the fact that conditioning on $\mathcal{E}_{\text{far}}$ cannot decrease the probability of $\mathcal{E}_{\text{hit}}^*$, as in Lemma A.1, part 2(b). □

To finish with the analysis of the Type II points, we apply Claim 3.8 together with Lemma 3.9 and Claim 3.10 with Lemma 3.11 to conclude that

$$\begin{aligned}\mathbb{E}|d(x, W_t^i) - d(y, W_t^i)|^2 \geq\ & (\Pr[\mathcal{E}_{\text{good}} \cap \mathcal{E}_{\text{close}} \cap \mathcal{E}_1^\sigma \cap \mathcal{E}_{\text{zero}}] + \\ & \Pr[\mathcal{E}_{\text{pad}} \cap \mathcal{E}_{\text{far}} \cap \mathcal{E}_2^\sigma \cap \mathcal{E}_{\text{hit}}^*])\ \frac{2^{2m-18}}{\zeta_K^2} \\ =\ & (\Pr[\mathcal{E}_{\text{close}}] \Pr[\mathcal{E}_{\text{good}} \cap \mathcal{E}_1^\sigma \cap \mathcal{E}_{\text{zero}} \mid \mathcal{E}_{\text{close}}] + \\ & \Pr[\mathcal{E}_{\text{far}}] \Pr[\mathcal{E}_{\text{pad}} \cap \mathcal{E}_2^\sigma \cap \mathcal{E}_{\text{hit}}^* \mid \mathcal{E}_{\text{far}}])\ \frac{2^{2m-18}}{\zeta_K^2} \\ \geq\ & \tfrac{1}{2} \min\{2^{-6} \cdot 3^{-5} \cdot p^5, 2^{-5} \cdot 3^{-5} \cdot (1 - e^{-3})\} \frac{2^{2m-18}}{\zeta_K^2} \\ \geq\ & \frac{p}{64}\ ^5\ \frac{2^{2m}}{\zeta_K^2}.\end{aligned}$$

Since we have proved that (3) holds for both Type I points and Type II points, the proof is complete. □

4. THE SPARSEST CUT PROBLEM WITH NON-UNIFORM DEMANDS

4.1 Computing the Euclidean distortion

In this section, we remark that the maps used to prove Theorem 1.1 have a certain “auto-extendability” property which will be used in the next section. We also recall that it is possible to find near-optimal Euclidean embeddings using semi-definite optimization.

THEOREM 4.1. *Let (Y,d) be an arbitrary metric space, and fix a k-point subset $X \subseteq Y$. If the space (X,d) is a metric of negative type, then there exists a probability space (Ω,μ), and a map $f : Y \to L_2(\Omega,\mu)$ such that*

1. *For every $\omega \in \Omega$, the map $x \mapsto f(x)(\omega)$ is 1-Lipschitz.*
2. *For every $x, y \in X$,*

$$||f(x) - f(y)||_2 \geq \frac{d(x,y)}{C\sqrt{\log k}\log\log k},$$

where $C > 0$ is some universal constant.

PROOF. We observe that the map used to prove Theorem 1.1 is a Fréchet embedding (note that the map from [16] employed in the proof of Theorem 1.1 for the case $x, y \in S_m(K_n)$ is also Fréchet). In other words, there is a probability space (Ω,μ) over subsets $A_\omega \subseteq X$ for $\omega \in \Omega$, and we obtain a map $g : X \to L_2(\mu)$ given by $g(x)(\omega) = d(x, A_\omega)$. We can then define the extension $f : Y \to L_2(\mu)$ by

$$f(y)(\omega) = d(y, A_\omega).$$

This ensures that the map $x \mapsto f(x)(\omega)$ is 1-Lipschitz on Y for every $\omega \in \Omega$. □

COROLLARY 4.2. *Let (Y,d) be an arbitrary metric space, and fix a k-point subset $X \subseteq Y$. If the space (X,d) is a metric of negative type, then there exists a 1-Lipschitz map $f : Y \to L_2$ such that the map $f|_X : X \to L_2$ has distortion $O(\sqrt{\log k}\log\log k)$.*

Now we suppose that (Y,d) is an n-point metric space and $X \subseteq Y$ is a k-point subset.

CLAIM 4.3. *There exists a polynomial-time algorithm (in terms of n) which, given X and Y, computes a map $f : Y \to L_2$ such that $f|_X$ has minimal distortion among all 1-Lipschitz maps f.*

PROOF. We give a semi-definite program computing the optimal f.

SDP (5.1)

max	ε	
s.t.	$x_u \in \mathbb{R}^n$	$\forall u \in Y$
	$(x_u - x_v)^2 \le d(u,v)^2$	$\forall u, v \in Y$
	$(x_u - x_v)^2 \ge \varepsilon\, d(u,v)^2$	$\forall u, v \in X$

□

4.2 The Sparsest Cut

Let V be an n-point set with two symmetric weights on pairs $w_N, w_D : V \times V \to \mathbb{R}_+$ (i.e. $w_N(x,y) = w_N(y,x)$ and $w_D(x,y) = w_D(y,x)$). For a subset $S \subseteq V$, we define the *sparsity of* S by

$$\Phi_{w_N,w_D}(S) = \frac{\sum_{u\in S, v\in \bar{S}} w_N(u,v)}{\sum_{u\in S, v\in \bar{S}} w_D(u,v)},$$

and we let $\Phi^*(V) = \min_{S\subseteq V} \Phi_{w_N,w_D}(S)$. (The set V is usually thought of as the vertex set of a graph with $w_N(u,v)$ supported only on edges (u,v), but this is unnecessary since we allow arbitrary weight functions.)

Computing the value of $\Phi^*(V)$ is NP-hard. The following semi-definite program is a relaxation of $\Phi^*(V)$.

SDP (5.2)

min	$\sum_{u,v\in V} w_N(u,v)\,(x_u - x_v)^2$	
s.t.	$x_u \in \mathbb{R}^n$	$\forall u \in V$
	$\sum_{u,v\in V} w_D(u,v)\,(x_u - x_v)^2 = 1$	
	$(x_u - x_v)^2 \le (x_u - x_w)^2 + (x_w - x_v)^2$	$\forall u, v, w \in V$

Furthermore, an optimal solution to this SDP can be computed in polynomial time.

The algorithm. We now give our algorithm for rounding SDP (5.2). Suppose that the weight function w_D is supported only on pairs u, v for which $u, v \in U \subseteq V$, and let $k = |U|$. Let $M = 20 \log n$.

1. Solve SDP (5.2), yielding a solution $\{x_u\}_{u\in V}$.
2. Consider the metric space (V, d) given by $d(u,v) = (x_u - x_v)^2$.
3. Applying SDP (5.1) to U and (V, d) (where $Y = V$ and $X = U$), compute the optimal map $f : V \to \mathbb{R}^n$.
4. Choose $\beta_1, \ldots, \beta_M \in \{-1, +1\}^n$ independently and uniformly at random.
5. For each $1 \le i \le M$, arrange the points of V as $v_1^i, \ldots, v_n^i$ so that
$$\langle \beta_i, f(v_j^i)\rangle \le \langle \beta_i, f(v_{j+1}^i)\rangle \text{ for each } 1 \le j \le n-1.$$
6. Output the sparsest of the Mn cuts
$$(\{v_1^i, \ldots, v_m^i\}, \{v_{m+1}^i, \ldots, v_n^i\}),\quad 1 \le m \le n-1, 1 \le i \le M.$$

CLAIM 4.4. *With constant probability over the choice of $\beta_1, \ldots, \beta_M$, the cut $(S, \bar{S})$ returned by the algorithm has*

$$\Phi(S) \le O(\sqrt{\log k}\log\log k)\,\Phi^*(V). \tag{6}$$

PROOF. Let $S \subseteq \mathbb{R}^n$ be the image of V under the map f. Consider the map $g : S \to \ell_1^M$ given by $g(x) = (\langle\beta_1, x\rangle, \ldots, \langle\beta_M, x\rangle)$. It is well-known (see, e.g. [1, 24]) that, with constant probability over the choice of $\{\beta_i\}_{i=1}^M \subseteq S^{n-1}$, g has distortion $O(1)$ (where S is equipped with the Euclidean metric). In this case, we claim that (6) holds.

To see this, let $S_1, S_2, \ldots, S_{Mn} \subseteq V$ be the Mn cuts which are tested in line (6). It is a standard fact that there exist constants $\alpha_1, \alpha_2, \ldots, \alpha_{Mn}$ such that for every $x, y \in V$,

$$||g(f(x)) - g(f(y))||_1 = \sum_{i=1}^{Mn} \alpha_i \rho_{S_i}(x,y),$$

where $\rho_{S_i}(x,y) = 1$ if x and y are on opposite sides of the cut $(S_i, \bar{S}_i)$ and $\rho_{S_i}(x,y) = 0$ otherwise.

Assume (by scaling) that $g \circ f : Y \to \ell_1^M$ is 1-Lipschitz. Let Λ be the distortion of $g \circ f$. By Corollary 4.2, $\Lambda = O(\sqrt{\log k}\log\log k)$. Recalling that $w_D(u,v) > 0$ only when $u, v \in U$,

$$\begin{aligned}
\Phi^*(V) &\ge \frac{\sum_{u,v\in V} w_N(u,v)(x_u - x_v)^2}{\sum_{u,v\in U} w_D(u,v)(x_u - x_v)^2} \\
&\ge \frac{1}{\Lambda}\frac{\sum_{u,v\in V} w_N(u,v)\,||g(f(u)) - g(f(v))||_1}{\sum_{u,v\in U} w_D(u,v)\,||g(f(u)) - g(f(v))||_1} \\
&= \frac{1}{\Lambda}\frac{\sum_{i=1}^{Mn}\alpha_i \sum_{u,v\in V} w_N(u,v)\rho_{S_i}(u,v)}{\sum_{i=1}^{Mn}\alpha_i \sum_{u,v\in U} w_D(u,v)\rho_{S_i}(u,v)} \\
&\ge \frac{1}{\Lambda}\min_i \frac{\sum_{u,v\in V} w_N(u,v)\rho_{S_i}(u,v)}{\sum_{u,v\in U} w_D(u,v)\rho_{S_i}(u,v)} \\
&= \frac{\Phi(S)}{\Lambda}.
\end{aligned}$$

This completes the proof. □

5. CONCLUDING REMARKS

- There are two factors of $O(\sqrt{\log\log n})$ which keep our bound from being optimal up to a constant factor. One factor of $\sqrt{\log\log n}$ arises because the ratio
$$\frac{|B(x, 2^{m+5}\alpha_X)|}{|B(x, 2^{m+3}/\zeta_K(X;p))|}$$
in Lemma 3.1 involves a pair of radii $R_1 = 2^{m+5}\alpha_X$ and $R_2 = 2^{m+3}/\zeta_K(X;p)$ for which $R_1/R_2 = \Omega(\log n)$. This arises out of a certain non-locality property which seems inherent to the method of proof in [3]. Note that even improving the method to $R_1 \approx 2^m$ is insufficient, as the ratio may still be quite large. The other factor arises because, in proving Theorem 1.1, we invoke Lemma 3.1 with $O(\log\log n)$ different values of the parameter K. It is likely removable by a more technical induction, but we chose to present the simpler proof.

- It is an interesting open problem to understand the exact distortion required to embed n-point negative type metrics into L_1. As mentioned, the best lower-bound is $\Omega(\log \log n)^\delta$ [15]. We also note that assuming an appropriate form of the *Unique Games Conjecture* is true, the general Sparsest Cut problem is hard to approximate within a factor of $\Omega(\log \log n)^\delta$ for some $\delta > 0$; this was recently shown independently in [15] and [9].

- For the uniform case of Sparsest Cut, it is possible to achieve a $O(\sqrt{\log n})$ approximation in quadratic time without solving the SDP [2]. Whether such an algorithm exists for the general case is an open problem.

- There is no asymptotic advantage in embedding n-point negative type metrics into L_p for some $p \in (1, \infty)$, $p \neq 2$ (observe that since L_2 is isometric to a subset of L_p for all $p \geq 1$, our embedding into Hilbert space is automatically also an embedding into L_p). Indeed, for $1 < p < 2$ it is shown in [19] that there are arbitrarily large n-point subsets of L_1 that require distortion $\Omega(\sqrt{(p-1)\log n})$ in any embedding into L_p. For $2 < p < \infty$ it follows from [27, 25] that there are arbitrarily large n-point subsets of L_1 whose minimal distortion into L_p is $1 + \Theta \sqrt{\frac{\log n}{p}}$ (the dependence on n follows from [27], and the optimal dependence on p follows from the results of [25]). Thus, up to multiplicative constants depending on p (and the double logarithmic factor in Theorem 1.1), our result is optimal for all $p \in (1, \infty)$.

- Let $(X, d_X), (Y, d_Y)$ be metric spaces and $\eta : [0, \infty) \to [0, \infty)$ a strictly increasing function. A one to one mapping $f : X \hookrightarrow Y$ is called a *quasisymmetric embedding with modulus* η if for every $x, a, b \in X$ such that $x \neq b$,

$$\frac{d_Y(f(x), f(a))}{d_Y(f(x), f(b))} \leq \eta \quad \frac{d_X(x,a)}{d_X(x,b)} \quad .$$

  We refer to [13] for an account of the theory of quasisymmetric embeddings. Observe that metrics of negative type embed quasisymmetrically into Hilbert space. It turns out that our embedding result generalizes to any n point metric space which embeds quasisymmetrically into Hilbert space. Indeed, if (X, d) embeds quasisymmetrically into L_2 with modulus η then, as shown in the full version of [17], there exists constants $p = p(\eta)$ and $C = C(\eta)$, depending only on η, such that $\zeta(X; p) \leq C\sqrt{\log n}$.

Acknowledgements

We thank Satish Rao for useful discussions.

6. REFERENCES

[1] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. System Sci.*, 66(4):671–687, 2003. Special issue on PODS 2001 (Santa Barbara, CA).

[2] S. Arora, E. Hazan, and S. Kale. $O(\sqrt{\log n})$ approximation to SPARSEST CUT in $\tilde{O}(n^2)$ time. In *45th Annual Syposium on Foundations of Computer Science*, pages 238–247. IEEE Computer Society, 2004.

[3] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *36th Annual Symposium on the Theory of Computing*, pages 222–231, 2004.

[4] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301 (electronic), 1998.

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

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

[7] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in ℓ_1. In *Proceedings of the 44th Annual IEEE Conference on Foundations of Computer Science*. ACM, 2003.

[8] S. Chawla, A. Gupta, and H. Räcke. An improved approximation to sparsest cut. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, Vancouver, 2005. ACM.

[9] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On embeddability of negative type metrics into ℓ_1. Manuscript, 2004.

[10] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105, 1969.

[11] J. Fakcharoenphol, S. Rao, and K. 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.

[12] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.

[13] J. Heinonen. *Lectures on analysis on metric spaces.* Universitext. Springer-Verlag, New York, 2001.

[14] 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.

[15] S. Khot and N. Vishnoi. On embeddability of negative type metrics into ℓ_1. Manuscript, 2004.

[16] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: A new embedding method for finite metrics. *Geom. Funct. Anal.*, 2004. To appear.

[17] J. R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, Vancouver, 2005. ACM.

[18] J. R. Lee, M. Mendel, and A. Naor. Metric structures in L_1: Dimension, snowflakes, and average distortion. *European J. Combin.*, 2004. To appear.

[19] 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.

[20] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic

applications. *Combinatorica*, 15(2):215–245, 1995.
[21] N. Linial and M. Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993.
[22] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.
[23] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Adv. Math.*, 189(2):451–494, 2004.
[24] V. D. Milman and G. 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.
[25] A. Naor, Y. Peres, O. Schramm, and S. Sheffield. Markov chains in smooth normed spaces and Gromov hyperbolic metric spaces. Preprint, 2004.
[26] A. Naor, Y. Rabani, and A. Sinclair. Quasisymmetric embeddings, the observable diameter, and expansion properties of graphs. Preprint, 2004.
[27] A. Naor and G. Schechtman. Remarks on non linear type and Pisier's inequality. *J. Reine Angew. Math.*, 552:213–236, 2002.
[28] S. 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. ACM.

APPENDIX

A. MISC. PROOFS

LEMMA A.1 (SAMPLING LEMMA). *Suppose that $k, n \in \mathbb{N}$, $1 \le k \le n$. Let X be an n-point set and let $\widetilde{X}_k$ be chosen uniformly at random from all k-point subsets of X. Then*

1. *For every $A \subseteq X$, $\Pr[\widetilde{X}_k \cap A = \emptyset] \le e^{-k|A|/n}$.*
2. *For every $A, B \subseteq X$ such that $A \cap B = \emptyset$,*
   (a) $\Pr[\widetilde{X}_k \cap A = \emptyset \mid \widetilde{X}_k \cap B \neq \emptyset] \le \Pr[\widetilde{X}_{k-1} \cap A = \emptyset] \le e^{(1-k)|A|/n}$
   (b) $\Pr[\widetilde{X}_k \cap A = \emptyset \mid \widetilde{X}_k \cap B = \emptyset] \le e^{-k|A|/n}$.

PROOF. The proof of (1) is an easy calculation. To prove 2(a), note that choosing $\widetilde{X}_k$ uniformly subject to $\widetilde{X}_k \cap B \neq \emptyset$ is the same as first choosing $z \in B$ uniformly at random, then choosing a uniform $k-1$ point subset $S \subseteq X \setminus \{z\}$ and returning $S \cup \{z\}$. $\square$

PROOF OF LEMMA 3.3. By definition,

$$|B(z, 2^{\kappa(z,t)})| < 2^t \le |B(x, 2^{\kappa(z,t)+1})|.$$

For the upper bound, we have

$$|B(x, 2^{\kappa(z,t)} - 2^{m-s})| \le |B(z, 2^{\kappa(z,t)})| < 2^t \le |B(x, 2^{i+m})|,$$

implying that $2^{\kappa(z,t)} - 2^{m-s} < 2^{i+m}$, which yields $2^{\kappa(z,t)} < 2^{m+1+i}$. For the lower bound, we have

$$\begin{aligned} |B(x, 2^{\kappa(z,t)+1} + 2^{m-s})| &\ge |B(z, 2^{\kappa(z,t)+1})| \\ &\ge 2^t \ge |B(x, 2^{m+i-1})|. \end{aligned}$$

We conclude that $2^{\kappa(z,t)+1} + 2^{m-s} \ge 2^{m+i-1}$, which implies that $2^{\kappa(z,t)+1} \ge 2^{m+i-2}$. $\square$