Euclidean distortion and the Sparsest Cut

Sanjeev Arora* James R. Lee† Assaf Naor ‡

Abstract

We prove that every n-point metric space of negative type (and, in particular, every n-point subset of L_1) embeds into a Euclidean space with distortion $O(\sqrt{\log n} \cdot \log \log n)$, a result which is tight up to the iterated logarithm factor. As a consequence, we obtain the best known polynomial-time approximation algorithm 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 $O(\sqrt{\log k} \cdot \log \log k)$.

1 Introduction

Bi-Lipschitz embeddings of finite metric spaces, a topic originally studied in geometric analysis and Banach space theory, became an integral part of theoretical computer science following work of Linial, London, and Rabinovich [27]. They presented an algorithmic version of a result of Bourgain [7] 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 [27, 5]. 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 [4] 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 [21] 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 [10] used the structural theorem of [4] (specifically, its stronger form due to

*Supported by David and Lucile Packard Fellowship and NSF grant CCR-0205594. Princeton University. arora@cs.princeton.edu

†Supported by NSF grant CCR-0121555 and an NSF Graduate Research Fellowship. U. C. Berkeley. jrl@cs.berkeley.edu

‡Microsoft Research. anaor@microsoft.com

Lee [22]), 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 $O(\sqrt{\log n} \cdot \log \log n)$. This almost matches the 35-year-old lower bound of $\sqrt{\log n}$ from Enflo [13]. Our methods use the results of [4, 22, 10] essentially as a "black box," together with an 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 $\mathsf{distortion}(f) = \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$ over all bijections $f : X \hookrightarrow L_2$. The mathematical investigation of the problem we study here goes back to the work of Enflo [13], who showed that the Euclidean distortion of the Hamming cube $\Omega_d = \{0,1\}^d$ equals $\sqrt{d} = \sqrt{\log_2 |\Omega_d|}$. 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 [32]), Johnson and Lindenstrauss [19] 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 [7] has shown that for any n-point metric space X, $c_2(X) = O(\log n)$, and this result is existentially optimal [27, 5]. 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 [21]) that several natural restricted classes of metrics do adhere to the $O(\sqrt{\log n})$ Euclidean distortion suggested by John's theorem. 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 iterated 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 [10] and the preceding works [27] and [5] (which yield an $O(\log n)$ approximation). This problem is described in Section 1.1. We now state our main result. In the case of metrics of negative type (and not just L_1 metrics), it answers positively (up to iterated logarithms) a well known conjecture in theoretical computer science and metric geometry (stated explicitly by Goemans in [16]).

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

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

Related work. Until recently, there was little solid evidence behind the conjecture that any n-point subset of L_1 embeds in Hilbert space with distortion $O(\sqrt{\log n})$. In the paper [23], 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 [4] 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 [21], Chawla, Gupta, and Räcke [10] 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 [29, pg. 379] that n-point metrics of negative type embed into L_1 with distortion $O(1)$. Recently, Khot and Vishnoi [20] 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 [8] (see also [24]) 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 [19] and Figiel, Lindenstrauss, and Milman [15], yields an embedding of every n-point subset of L_1 into $\ell_1^{O(\log n)}$ with distortion $O(\sqrt{\log n}\cdot\log\log n)$.

1.1 Algorithmic application: The Sparsest Cut problem with general demands

In this section, we briefly describe an application of Theorem 1.1 to the Sparsest Cut problem with general demands (and its relation to the *multi-commodity flow* problem). This is a fundamental NP-hard combinatorial optimization problem—we refer the interested reader to the articles [26, 2, 27, 5], the survey [37], and Chapter 21 of the book [38] for additional information on Sparsest Cut, and its applications to the design of approximation algorithms.

Let $G = (V, E)$ be a graph (network), with a *capacity* $C(e) \geq 0$ associated to every edge $e \in E$. Assume that we are given k pairs of vertices $(s_1, t_1), ..., (s_k, t_k) \in V \times V$ and $D_1, \dots, D_k \geq 1$. We think of the s_i as *sources*, the t_i as *targets*, and the value D_i as the *demand* of the *terminal pair* (s_i, t_i) for some *commodity* κ_i. The problem is said to have *uniform demands* if *every* pair $u, v \in V$ occurs as some (s_i, t_i) pair with $D_i = 1$.

In the *MaxFlow* problem the objective is to maximize the *fraction* λ of the demand that can be shipped simultaneously for all the commodities, subject to the capacity constraints. Denote this maximum by λ^*. A trivial upper bound on λ^* is the *cut ratio*. Given any subset $S \subseteq V$, we write

$$\Phi(S) = \frac{\sum_{uv\in E} C(uv)\cdot|\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\sum_{i=1}^k D_i\cdot|\mathbf{1}_S(s_i) - \mathbf{1}_S(t_i)|},$$

where $\mathbf{1}_S$ is the characteristic function of S. The value $\Phi^* = \min_{S\subseteq V}\Phi(S)$ is the minimum over all cuts (partitions) of V, of the ratio between the total capacity crossing the cut and the total demand crossing the cut. In the case of a single commodity (i.e. $k = 1$) the classical MaxFlow-MinCut theorem states that $\lambda^* = \Phi^*$, but in general this is no longer the case. It is known [26, 27, 5] that $\Phi^* = O(\log k)\lambda^*$. This result is perhaps the first striking application of metric embeddings in combinatorial optimization (specifically, it uses Bourgain's embedding theorem [7]).

Computing Φ^* is NP-hard [30]. Moreover, finding a cut for which Φ^* is (approximately) attained is a basic step in approximation algorithms for several NP-hard problems [26, 2, 37]. The best known algorithm for computing Φ^* in the case of uniform demands is due to [4], where an approximation ratio of $O(\sqrt{\log n})$ is achieved. In the case of non-uniform demands, an approximation ratio of $O(\log k)^{3/4}$ is obtained in [10]. Here, as an application of Theorem 1.1, we prove the following theorem:

Theorem 1.2. *Using the above notation, there exists a polynomial-time algorithm which produces a subset $S \subseteq V$ for which*

$$\Phi(S) = O\left(\sqrt{\log k}\cdot\log\log k\right)\Phi^*.$$

Structure of the paper: This paper is organized as follows. In Section 2 we present an informal overview of the ideas involved in the proof of Theorem 1.1. Section 3 is devoted to various preliminaries on the geometry of metrics of negative type. Theorem 1.1 is proved in Section 4, and the algorithm of Theorem 1.2 is described and analyzed in Section 5. We end with Section 6, which contains additional remarks and open problems.

2 Overview of the proof of Theorem 1.1

Remarks on notation. When we write $E \gtrsim F$ for two expressions E and F, we intend this to mean that there exists some $\epsilon > 0$ such that $E \geq \epsilon F$, where ϵ is intended to be a universal constant, independent of the variables or parameters on which E and F depend.

We will often work with Hilbert spaces of the following form: If H is a Hilbert space, and (Ω, μ) is a probability space, we use $L_2(H, \Omega, \mu)$ to denote the Hilbert space of H-valued random variables Z with norm $||Z||_{L_2(H,\Omega,\mu)} = \sqrt{\mathbb{E}\,||Z||_H^2}$. When H, Ω are clear from context, we simply write $L_2(\mu)$ and denote $||\cdot||_H$ by $||\cdot||_2$.

Our proof of Theorem 1.1 has little to do with metrics of negative type; the connection to such spaces comes through the techniques of [4, 22, 10] and is laid out in Section 3. Instead, we present a general theorem about gluing together various maps from finite metric spaces into Hilbert spaces (and, more generally, L_p spaces for $p \in [1, \infty)$). Our starting point is the following type of ensemble.

Let (X, d) be an n-point metric space. Suppose that for every $\tau \geq 0$, and every subset $S \subseteq X$, there exists a 1-Lipschitz map $\varphi_{S,\tau} : X \to L_2$ with

$$||\varphi_{S,\tau}(x) - \varphi_{S,\tau}(y)||_2 \geq \frac{\tau}{\sqrt{\log |S|}} \tag{1}$$

whenever $x, y \in S$ and $d(x, y) \in [\tau, 2\tau]$. In general, $\sqrt{\log |S|}$ could be a different function of $|S|$, but we restrict ourselves here for simplicity. Additionally, let us temporarily define $\varphi_\tau = \varphi_{X,\tau}$ for every $\tau \geq 0$ so that for the maps $\{\varphi_\tau\}$, condition (1) holds for all $x, y \in X$ and $|S| = n$. The problem we are now confronted with is how to combine the ensemble of maps $\{\varphi_{S,\tau}\}$ together to obtain a genuinely bi-Lipschitz map.

There is an obvious approach which comes to mind: Let $R \subseteq \mathbb{Z}$ be such that for all $x, y \in X$, there exists $k \in R$ such that $d(x, y) \in [2^k, 2^{k+1}]$. Now define the map $\varphi : X \to L_2$ by $\varphi = \bigoplus_{k \in R} \varphi_{2^k}$. Clearly we have both $||\varphi||_{\mathrm{Lip}} \leq \sqrt{R}$ and, for all $x, y \in X$,

$$||\varphi(x) - \varphi(y)||_2 \geq \frac{d(x, y)}{2\sqrt{\log n}},$$

hence $\mathsf{distortion}(\varphi) = O(\sqrt{R \log n})$. Trivially, we can choose R so that $|R| \leq n^2$. A slightly more delicate argument yields such an R with $|R| \leq O(n)$. Unfortunately, we are searching for a bound of the form $\mathsf{distortion}(\varphi) \approx \sqrt{\log n}$, making this construction useless.

Nevertheless, the key to a better gluing of the given ensemble does lie in the delicate interplay between the distributions of distances in X and the number of points in various regions of the space. The technique of *measured descent* from [21] relies essentially on two facts about finite metric spaces. First, the identity

$$\sum_{k \in \mathbb{Z}} \log \frac{|B(x, \alpha \cdot 2^k)|}{|B(x, 2^k)|} = O(\log n \log \alpha) \tag{2}$$

for any number $\alpha \geq 2$. (In [21], a fixed constant value of α was used, but for us the quantitative dependence is crucial, as we will have α depending on n.) This gives a simple bound on the rate that a finite metric space can expand over all its scales, and is implicitly used in earlier works under the name of "region growing" [26, 14].

For the purposes of this description, we will state the second fact less concretely. Basically, in certain settings, one can think of the ratio $|B(x, \alpha\tau)|/|B(x, \tau)|$ as the "local cardinality of the space" around x at scale τ. As an example, if $X = \mathbb{R}^d$, $B(x, \cdot)$ represents a Euclidean ball, and $|\cdot|$ is the Lebesgue measure, then this ratio approximates the number of τ-net points that can be packed inside a ball of radius $\alpha\tau$. Later, it will become necessary to randomly partition X into pieces of diameter at most 2τ while ensuring that pairs $x, y \in X$ with $d(x, y) \ll \tau$ are usually in the same component of the partition (see Section 3.1 on padded decomposability). It is known [9, 14] that the properties of such partitions near x depends on the local value $\log \frac{|B(x,2\tau)|}{|B(x,\tau)|}$.

Following [21], this relationship is used in [22] to prove (roughly) that, given the maps $\{\varphi_\tau\}_{\tau \geq 0}$ defined above, there exists a map $\varphi : X \to L_2$ such that $||\varphi||_{\text{Lip}} \leq O(\sqrt{\log n})$ and, for $x, y \in X$ with $d(x, y) \in [2^k, 2^{k+1}]$,

$$||\varphi(x) - \varphi(y)||_2 \gtrsim \sqrt{\log \frac{|B(x, 2^{k+1})|}{|B(x, 2^k)|}} \left(\underbrace{||\varphi_{2^k}(x) - \varphi_{2^k}(y)||}_{\text{(I)}} + \underbrace{\frac{d(x,y)}{\log \frac{|B(x,2^{k+1})|}{|B(x,2^k)|}}}_{\text{(II)}} \right) \tag{3}$$

The contribution (II) comes from random partitioning and Rao's technique [36], and is valid for any metric space X. Observing that (I) $\gtrsim d(x, y)/\sqrt{\log n}$, and using AM-GM in (3), one arrives at the lower bound

$$||\varphi(x) - \varphi(y)||_2 \gtrsim \frac{d(x,y)}{(\log n)^{\frac{1}{4}}},$$

hence $\mathsf{distortion}(\varphi) \leq O(\log n)^{\frac{3}{4}}$. While not obvious at present, the identity (2) is what allows [22] to get the leading $\sqrt{\cdot}$ factor in (3) while keeping $||\varphi||_{\text{Lip}}$ small (see Theorem 4.5).

In order to get the distortion near $O(\sqrt{\log n})$, we have to dispense with the contribution (II) which is not derived from the ensemble $\{\varphi_{S,\tau}\}$. Instead, we would like to pass from the ensemble $\{\varphi_{S,\tau}\}$ to a family of maps $\{\tilde{\varphi}_\tau : X \to L_2\}$ for which the contribution of (I) in (3) is replaced by

$$||\tilde{\varphi}_{2^k}(x) - \tilde{\varphi}_{2^k}(y)||_2 \gtrsim \frac{d(x,y)}{\sqrt{\log \frac{|B(x,2^{k+1})|}{|B(x,2^k)|}}}. \tag{4}$$

Clearly this would finish the proof. Roughly, the construction of $\tilde{\varphi}_{2^k}$ proceeds as follows. We first randomly partition X into components of diameter about $2^k\alpha$ for some appropriately chosen $\alpha = \alpha(n)$. Writing the random partition as $X = C_1 \cup C_2 \cup \cdots \cup C_m$, we then derive subsets $\tilde{C}_i \subseteq C_i$ by randomly sampling points from each C_i. Then, we use an appropriately constructed (random) partition of unity to glue the collection of maps $\{\varphi_{\tilde{C}_i,2^k}\}_{i=1}^m$ together. To ensure that the resulting map still has $||\tilde{\varphi}_{2^k}||_{\text{Lip}} \leq O(1)$, the partition of unity is constructed carefully using properties of the random partition (this bears some resemblance to the technique of [25] for extending Lipschitz functions).

The key to the proof is the way in which the random samples $\tilde{C}_i$ are chosen. We have to maintain the property that $\tilde{C}_i$ is a "good representative" of C_i at scale 2^k (i.e. we need that, on average, $\tilde{C}_i$ is a 2^k-dense in C_i). On the other hand, we need to maintain the invariant that if $x \in C_i$, then

$$\log|\tilde{C}_i| \approx \log \frac{|B(x, \alpha 2^k)|}{|B(x, 2^k)|},$$

so that we can achieve a bound similar to (4) (recall that the quality of the map $\varphi_{S,\tau}$ depends on $|S| = |\tilde{C}_i|$). Unfortunately, this is impossible since for distinct $x, x' \in C_i$, the above ratios can be quite different. Instead, we have a number of phases, one for each estimate of the possible ratio (see the proof of Theorem 1.1). For this to work, we have to give up on achieving (4) exactly, and instead we weave together the *inter-scale* (Lemma 4.4) gluing of (3) with the *intra-scale* (Theorem 4.5) gluing of (4) to obtain a nearly-tight bound of $O(\sqrt{\log n} \cdot \log\log n)$.

3 Single scale embeddings

In this section we present Theorem 3.1, and derive from it Lemma 3.5 which is one of the main tools used in the proof of the Main Theorem (1.1). It is a concatenation of the result of Arora, Rao, and Vazirani [4], its strengthening by Lee [22], and the "reweighting" method of Chawla, Gupta, and Räcke [10], who use it in conjunction with [21] to achieve distortion $O(\log n)^{\frac{3}{4}}$. For the sake of completeness, we present below a sketch of the proof of Theorem 3.1. Complete details can be found in the full version of [22], 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 [18] for the definition of quasisymmetry; the relevance of such maps to the techniques of [4] was first pointed out in [34]).

Theorem 3.1. *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 (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 [31], 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 [22], 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 [4, 22] 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 [10] 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 $w(x,y) > 0$ only if $(x,y) \in S_\Delta$. 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 applying 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 [10], the theorems of [4, 22] imply that with high probability, after the pruning, we still have

$$\sum_{x \in L'_u, y \in R'_u} w(x,y) \gtrsim \sum_{x,y} w(x,y).$$

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. Let us call these sets $L'_u(w), R'_u(w)$. One then outputs the set U of all points x for which any "copy" falls into $L'_u(w)$.

Now we define the family of weight functions. The initial weight function has $w_0(x,y) = n^4$ for all $(x,y) \in S_\Delta$. Given w_k, obtain w_{k+1} as follows. If

$$\mu\left\{u \in S^{n-1} : (x,y) \in L'_u(w_k) \times R'_u(w_k)\right\} \geq 0.1,$$

we set $w_{k+1}(x,y) = \frac{1}{2} w_k(x,y)$. Otherwise, we set $w_{k+1}(x,y) = w_k(x,y)$. A simple argument (presented in [10]) 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 w and a random direction $u \in S^{n-1}$, then with constant probability we have $(x,y) \in L'_u(w) \times R'_u(w)$. □

3.1 Padded decomposability and random zero sets

Theorem 3.1 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 3.2 (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 $d(x,y) \geq \Delta$,*

$$\mu\left\{Z \subseteq X : \ y \in Z \text{ and } d(x,Z) \geq \frac{\Delta}{\zeta}\right\} \geq p.$$

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 $k \leq n$ we define

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

With this definition, Theorem 3.1 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.

Definition 3.3 (Decomposition bundle, modulus of padded decomposability)**.** *Following [21] we say that $\{P_\Delta\}_{\Delta>0}$ is an α-padded decomposition bundle of a metric space X if for every $\Delta > 0$, P_Δ is a random partition of X (whose distribution we denote by ν) 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*

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

The modulus of padded decomposability of X, denoted α_X, is defined as the largest constant $\alpha > 0$ such that X admits an α-padded decomposition bundle.

As observed in [21], the results of [28, 6] imply that $\alpha_X = O(\log |X|)$, and this will be used in the ensuing arguments. The following useful fact relates the notions of padded decomposability and random zero sets. Its proof is motivated by an argument of Rao [36].

Fact 3.4. *$\zeta(X; 1/8) \leq \alpha_X$.*

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 $d(x,y) \geq \Delta$ then $P(x) \neq P(y)$. It follows that

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

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 $1/2$, $\pi_P(x) \geq \Delta/\alpha_X$. The required result now follows by considering the random zero set Z_P. $\square$

We end this section with the following simple lemma, which shows that the existence of random zero sets implies the existence of embeddings into L_2 which are bi-Lipschitz on a fixed distance scale.

Lemma 3.5 (Random zero sets yield single scale embeddings)**.** *For every finite metric space X, every $S \subseteq X$ every $p \in (0,1)$ and every $\tau > 0$, there exists a 1-Lipschitz mapping $\varphi : X \to L_2$ such that for every $x, y \in S$ with $d(x,y) \geq \tau$,*

$$\|\varphi(x) - \varphi(y)\|_2 \geq \frac{\tau\sqrt{p}}{\zeta(S;p)}.$$

Proof. By the definition of $\zeta(S,p)$ there exists a distribution μ over subsets $Z \subseteq S$ such that for every $x, y \in S$ with $d(x,y) \geq \tau$,

$$\mu\left\{Z \subseteq S:\ y \in Z \text{ and } d(x,Z) \geq \frac{\tau}{\zeta(S;p)}\right\} \geq p.$$

Define $\varphi : X \to L_2(\mu)$ by $\varphi(x) = d(x,Z)$. Clearly φ is 1-Lipschitz. Moreover, for every $x, y \in S$ with $d(x,y) \geq \tau$,

$$\|\varphi(x) - \varphi(y)\|_{L_2(\mu)}^2 = \mathbb{E}_\mu\left[d(x,Z) - d(y,Z)\right]^2 \geq p \cdot \left(\frac{\tau}{\zeta(S;p)}\right)^2.$$

□

4 Proof of Theorem 1.1

The primary result of this section is the following theorem.

Theorem 4.1. *Let (X,d) be an n-point metric space. Suppose there exist constants $C > 0$ and $\frac{1}{2} \leq \varepsilon \leq 1$, such that for every $\tau \geq 0$, and every subset $S \subseteq X$, there exists a 1-Lipschitz map $\varphi_{S,\tau} : X \to L_2$ with*

$$||\varphi_{S,\tau}(x) - \varphi_{S,\tau}(y)||_2 \geq \frac{\tau}{C(\log|S|)^\varepsilon}$$

whenever $x, y \in S$ and $d(x,y) \in [\tau, 6\tau]$. Then $c_2(X) \leq O(1) \cdot C(\log n)^\varepsilon \log\log n$.

Theorem 4.1 implies Theorem 1.1. Indeed, if X is an n point metric space such that for some $p \in (0,1)$, $\varepsilon \in [1/2, 1]$, and $C > 0$, we have for every $k \leq n$, $\zeta_k(X;p) \leq C(\log k)^\varepsilon$, then Theorem 4.1 together with Lemma 3.5 imply that

$$c_2(X) = O\left(\frac{C(\log n)^\varepsilon \log\log n}{\sqrt{p}}\right).$$

Theorem 1.1 follows since by Theorem 3.1 we know that for some universal constant $p \in (0,1)$, if X is a metric space of negative type then for all k, $\zeta_k(X;p) = O\left(\sqrt{\log n}\right)$.

The proof of Theorem 4.1 will be broken down into several steps. In what follows we fix a finite metric space X, and for $K \geq 1$, $\tau \geq 0$, define

$$S_\tau(K) = \left\{x \in X : |B(x, 8\tau\alpha_X)| \leq K \left|B\left(x, \frac{\tau}{12C(\log K)^\varepsilon}\right)\right|\right\}.$$

Lemma 4.2 (Embedding neighborhoods)**.** *Let $S \subseteq X$, $\tau \geq 0$, and assume that there exists a 1-Lipschitz map $\varphi : X \to L_2$ satisfying*

$$||\varphi(x) - \varphi(y)||_2 \geq \frac{\tau}{L}$$

for $x, y \in S$, $d(x,y) \in [\tau/2, 3\tau]$ and some $L \geq 2$. Then there is a 1-Lipschitz map $h : X \to L_2$ with

$$||h(x) - h(y)||_2 \geq \frac{\tau}{9L}$$

whenever $d(x,S) \leq \frac{\tau}{6L}$, $y \in X$, and $d(x,y) \in [\tau, 2\tau]$.

Proof. Define $g : X \to \mathbb{R}$ by $g(x) = d(x, S)$, and set $h = \frac{1}{\sqrt{2}}(\varphi \oplus g)$. If $d(y, S) > \frac{\tau}{3L}$, then

$$||h(x) - h(y)||_2 \geq \frac{1}{\sqrt{2}}||g(x) - g(y)||_2 \geq \frac{1}{\sqrt{2}}(d(y, S) - d(x, S)) \geq \frac{1}{\sqrt{2}} \cdot \frac{\tau}{6L}.$$

Otherwise, let $x', y' \in S$ be such that $d(x, x') \leq \frac{\tau}{6L}, d(y, y') \leq \frac{\tau}{3L}$, and observe that

$$d(x', y') \in \left[d(x, y) - \frac{\tau}{6L} - \frac{\tau}{3L}, d(x, y) + \frac{\tau}{6L} + \frac{\tau}{3L}\right] \subseteq \left[\frac{\tau}{2}, 3\tau\right].$$

Using our assumptions on φ, we have

$$||\varphi(x) - \varphi(y)||_2 \geq ||\varphi(x') - \varphi(y')||_2 - ||\varphi||_{\text{Lip}} \left(\frac{\tau}{6L} + \frac{\tau}{3L}\right) \geq \frac{\tau}{2L},$$

hence $||h(x) - h(y)||_2 \geq \frac{1}{\sqrt{2}} \cdot \frac{\tau}{2L}$. □

Lemma 4.3 (Random subsets)**.** *Assume that X satisfies the conditions of Theorem 4.1, and suppose that $U \subseteq X$ and $k \geq 2$. Define*

$$T_\tau(U; k) = \left\{x \in U : |U| \leq k \left|B\left(x, \frac{\tau}{12C(\log k)^\varepsilon}\right)\right|\right\}. \tag{5}$$

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

$$||\gamma_{U,k}(x) - \gamma_{U,k}(y)||_2 \geq \frac{\tau}{30C(\log k)^\varepsilon}$$

whenever $x \in T_\tau(U; k), y \in X$ and $d(x, y) \in [\tau, 2\tau]$.

Proof. Let S be a uniformly random subset $S \subseteq U$ with $|S| = \min\{|U|, k\}$. Let $h_S : X \to L_2$ be the map defined by $h_S = \frac{1}{\sqrt{2}}(\varphi_{S,\tau/2} \oplus g)$ where $g(x) = d(x, S)$. Define $\gamma_{U,k} : X \to L_2(L_2, \mu)$, where μ is the distribution of the random subset S, by $\gamma_{U,k}(x) = h_S(x)$ (recall that $h_S(x)$ is a *random* element of L_2). Note that $\gamma_{U,k}$ is 1-Lipschitz because the same is true for each h_S.

Let $L = 2C(\log |S|)^\varepsilon$. Observe that, by the definition of $T_\tau(U; k)$, with probability at least $1/e$, we have

$$S \cap B\left(x, \frac{\tau}{6L}\right) = S \cap B\left(x, \frac{\tau}{12C(\log k)^\varepsilon}\right) \neq \emptyset.$$

Assuming this holds, we see that $d(x, S) \leq \frac{\tau}{6L}$. Thus by Lemma 4.2, $||h_S(x) - h_S(y)||_2 \geq \frac{\tau}{9L}$. It follows that

$$||\gamma_{U,k}(x) - \gamma_{U,k}(y)||_2 \geq \frac{1}{\sqrt{e}} \cdot \frac{\tau}{9L} \geq \frac{\tau}{30C(\log k)^\varepsilon}.$$

□

In what follows we shall use the fact that for every $\tau > 0$ there exists a mapping $G_\tau : L_2 \to L_2$ such that for every $x, y \in L_2$,

$$\|G_\tau(x)\|_2 = \|G_\tau(y)\|_2 = \tau \quad \text{and} \quad \tfrac{1}{2}\min\{\tau, \|x - y\|_2\} \leq \|G_\tau(x) - G_\tau(y)\|_2 \leq \min\{\tau, \|x - y\|_2\}. \tag{6}$$

The existence of G_τ is precisely Lemma 5.2 in [31]. As in [22], we will use the map G_τ to control the Lipschitz constant of various functions under partitions of unity.

Lemma 4.4 (Localization)**.** *Assume that X satisfies the conditions of Theorem 4.1. 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 $x \in S_\tau(k), y \in X$ with $d(x,y) \in [\tau, 3\tau]$,*

$$||\Lambda_{\tau,k}(x) - \Lambda_{\tau,k}(y)||_2 \geq \frac{\tau}{240C(\log k)^\varepsilon}.$$

Proof. Let $D = 4\tau\alpha_X$ and take P_D to be a random partition from the α_X-padded bundle ensured by Definition 3.3. 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 $\gamma_{U,k}$ be the corresponding map from Lemma 4.3. Finally, define a random map $\Lambda_{\tau,k} : X \to L_2$ by

$$\Lambda_{\tau,k}(z) = \tfrac{1}{2}\rho(z) \cdot \widehat{\gamma}_{P_D(z),k}(z),$$

where for $f : X \to L_2$ we write $\widehat{f} = G_\tau \circ f$, where G_τ is as in (6).

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

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

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 $x \in S_\tau(k), y \in X$, and $d(x,y) \in [\tau, 3\tau]$. Observe that since $\mathrm{diam}(P_D(x)) \leq D$, 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 (5)). Moreover, using the defining property of the α_X-padded bundle, with probability at least $\frac{1}{2}$, we have $d(x, X \setminus P_D(x)) \geq 5\tau$. Since we are assuming that $d(x,y) \leq 3\tau$, this implies that $\rho(x) = \rho(y) = 1$. It follows that

$$\begin{aligned}
\mathbb{E}\,||\Lambda_{\tau,k}(x) - \Lambda_{\tau,k}(y)||_2 &\geq \tfrac{1}{2} \cdot \tfrac{1}{2}\mathbb{E}\,||\widehat{\gamma}_{P_D(x),k}(x) - \widehat{\gamma}_{P_D(x),k}(y)||_2 \\
&\geq \tfrac{1}{8}\mathbb{E}\left(\min\left\{||\gamma_{P_D(x),k}(x) - \gamma_{P_D(x),k}(y)||_2, \tau\right\}\right) \\
&\geq \frac{\tau}{240C(\log k)^\varepsilon}.
\end{aligned}$$

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 $L_2(L_2, \mu)$ which has the required properties. □

The following theorem is a generalization of the Gluing Lemma in [22]. In particular, it is important for us that part (2) treats x and y symmetrically, unlike in [22].

Theorem 4.5 (Inter-scale gluing)**.** *Given any n-point metric space (X, d) and constants $A, B \geq 1$, and for every $m \in \mathbb{Z}$, a 1-Lipschitz map $\phi_m : X \to L_2$, there exists a map $\varphi : X \to L_2$ which satisfies*

1. $||\varphi||_{\mathrm{Lip}} \leq O(\sqrt{\log n \log(AB)})$.
2. *For every $x, y \in X$ we have*

$$||\varphi(x) - \varphi(y)||_2 \geq \max_{m \in \mathbb{Z}} \left(\sqrt{\left\lfloor \log \frac{|B(x, 2^{m+1}A)|}{|B(x, 2^m/B)|} \right\rfloor} \cdot \min \left\{ \frac{2^m}{B}, ||\phi_m(x) - \phi_m(y)||_2 \right\} \right).$$

Proof. Let $\rho : X \to \mathbb{R}_+$ 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

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

Write $\widehat{\phi}_m = G_{2^m/B} \circ \phi_m$, where $G_{2^m/B}$ is as in (6). Now, for each $t \in \{1, 2, \ldots, \lceil \log_2 n \rceil\}$, define $\psi_t : X \to \ell_2(L_2)$,

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

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

First, we bound $||\psi_t||_{\mathrm{Lip}}$ 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.

$$\begin{aligned} ||\rho_{m,t}(x)\widehat{\phi}_m(x) - \rho_{m,t}(y)\widehat{\phi}_m(y)||_2 &\leq ||\widehat{\phi}_m(x)||_2|\rho_{m,t}(x) - \rho_{m,t}(y)| + ||\widehat{\phi}_m(x) - \widehat{\phi}_m(y)||_2|\rho_{m,t}(y)| \\ &\leq \left(||\rho_{m,t}||_{\mathrm{Lip}} \cdot \frac{2^m}{B} + ||\phi_m||_{\mathrm{Lip}} \right) d(x, y) \\ &\leq 4\, d(x, y). \end{aligned}$$

Thus $||\psi_t||_{\mathrm{Lip}} \leq O(\sqrt{\log(AB)})$. It follows that $||\varphi||_{\mathrm{Lip}} \leq O(\sqrt{\log n \log(AB)})$, as claimed.

It remains to prove the lower bound. To this end, fix $m \in \mathbb{Z}$, $x, y \in X$ and observe that if $\rho_{m,t}(x) = 1$, then

$$\begin{aligned} ||\psi_t(x) - \psi_t(y)||_2 &\geq ||\widehat{\phi}_m(x) - \widehat{\phi}_m(y)||_2 - (1 - \rho_{m,t}(y)) \cdot ||\widehat{\phi}_m(y)||_2 \\ &\geq \frac{1}{2} \min \left\{ \frac{2^m}{B}, ||\phi_m(x) - \phi_m(y)||_2 \right\} - \frac{2^m}{B} \cdot (1 - \rho_{m,t}(y)). \end{aligned} \tag{7}$$

On the other hand

$$\begin{aligned}||\psi_t(x)-\psi_t(y)||_2 &\geq \|\widehat{\phi}_m(x)-\rho_{m,t}(y)\widehat{\phi}_m(y)\|_2\\ &\geq \|\widehat{\phi}_m(x)\|_2-\rho_{m,t}(y)\|\widehat{\phi}_m(y)\|_2=\frac{2^m}{B}\cdot(1-\rho_{m,t}(y)).\end{aligned}\tag{8}$$

Averaging (7) and (8) we get that

$$||\psi_t(x)-\psi_t(y)||_2 \geq \frac{1}{4}\min\left\{\frac{2^m}{B}, ||\phi_m(x)-\phi_m(y)||_2\right\}.\tag{9}$$

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

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

This completes the proof since the lower bound (9) holds for $\left\lfloor \log\frac{|B(x,2^{m+1}A)|}{|B(x,2^m/B)|}\right\rfloor$ values of t. □

We also present the following base case which is a variant of Bourgain's argument [7].

Claim 4.6 (Small ratios)**.** *Let X be an n-point metric space, and $\lambda\geq 1$. For each $\tau>0$, define the subset*

$$J_\lambda(\tau)=\{x\in X:|B(x,\tau/2)|\leq\lambda|B(x,\tau/4)|\}.$$

Then there exists a map $F:X\to L_2$ with $\|F\|_{\mathrm{Lip}}\leq O\left(\sqrt{\log n}\right)$, and such that for every $x,y\in X$ with $x\in J_\lambda(\tau)$ and $d(x,y)\geq\tau$, we have

$$||F(x)-F(y)||_2\geq\epsilon(\lambda)\tau,$$

where $\epsilon(\lambda)>0$ is a constant depending only on λ.

Proof. For each $t\in\{1,2,\ldots,\lceil\log n\rceil\}$, let $W_t\subseteq X$ be a random subset which contains each point of X independently with probability 2^{-t}. Let $g_t(x)=d(x,W_t)$, and consider the *random* map $f=g_1\oplus\cdots\oplus g_{\lceil\log n\rceil}\in\ell_2^n$. Finally, we define $F:X\to L_2(\ell_2^n,\mu)$ by $F(x)=f(x)$, where μ is the distribution over which the random subsets $\{W_t\}$ are defined. Observe that $\|F\|_{\mathrm{Lip}}\leq O\left(\sqrt{\log n}\right)$.

Fix $x,y\in X$ such that $d(x,y)\geq\tau$ and $x\in J_\lambda(\tau)$. Let $t\in\mathbb{N}$ be such that $2^t\leq|B(x,\tau/2)|\leq 2^{t+1}$. Let $\mathcal{E}_{\mathrm{far}}$ be the event $\{d(x,W_t)\geq\tau/2\}$ and let $\mathcal{E}_{\mathrm{close}}$ be the event $\{d(x,W_t)\leq\tau/4\}$. Also, define the event $\mathcal{E}^y_{\mathrm{close}}=\{d(y,W_t)<\tau/2\}$. Observe that each of the events $\mathcal{E}_{\mathrm{close}},\mathcal{E}_{\mathrm{far}}$ are independent of $\mathcal{E}^y_{\mathrm{close}}$ since the former events depend only on $W_t\cap B^\circ(x,\tau/2)$, and the latter on $W_t\cap B^\circ(y,\tau/2)$, where $B^\circ(\cdot,\cdot)$ denotes an open ball in X. It follows that

$$\begin{aligned}&\|F(x)-F(y)\|^2_{L_2(\mu)}=\mathbb{E}_\mu\|f(x)-f(y)\|_2^2\\ &\geq \mathbb{E}_\mu|g_t(x)-g_t(y)|_2^2\\ &\geq \Pr(\mathcal{E}^y_{\mathrm{close}})\cdot\min\left\{\Pr(\mathcal{E}_{\mathrm{far}}),\Pr(\mathcal{E}_{\mathrm{close}})\right\}\cdot\left(\frac{1}{2}\cdot\frac{\tau}{4}\right)^2+\Pr(\neg\mathcal{E}^y_{\mathrm{close}})\cdot\Pr(\mathcal{E}_{\mathrm{close}})\cdot\left(\frac{\tau}{4}\right)^2\\ &\geq \epsilon(\lambda)\tau^2.\end{aligned}$$

The final inequality holds true because $x\in J_\lambda(\tau)$ implies that each of $\Pr(\mathcal{E}_{\mathrm{far}})$ and $\Pr(\mathcal{E}_{\mathrm{close}})$ can be bounded from below by some $\epsilon'(\lambda)>0$. □

We are now in position to conclude the proof of Theorem 4.1

Proof of Theorem 4.1. We claim that for every $K \in [2, n]$ there exists a map $f_K : X \to L_2$ which satisfies

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

$$||f_K(x) - f_K(y)||_2^2 \gtrsim \left\lfloor \log \frac{|B(x, 2^{m+3}\alpha_X)|}{|B(x, 2^m/[12C(\log K)^\varepsilon])|} \right\rfloor \cdot \frac{2^{2m}}{C^2(\log K)^{2\varepsilon}}.$$

Indeed, f_K is obtained from an application of Theorem 4.5 to the mappings $\{\Lambda_{2^m,K}\}_{m\in\mathbb{Z}}$ from Lemma 4.4 with $A = 4\alpha_X$ and $B = 12C(\log K)^\varepsilon$ (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 $K_j \ge 4$, we obtain mappings $f_0, \dots, f_j : X \to L_2$ satisfying

1. $\|f_j\|_{\mathrm{Lip}} \le O(\sqrt{\log n \cdot \log\log n})$.
2. For all $x \in S_{2^m}(K_j) \setminus S_{2^m}(K_{j+1})$ and $y \in X$ such that $d(x, y) \in [2^m, 2^{m+1}]$ and we have

$$\begin{aligned} ||f_j(x) - f_j(y)||_2^2 &\gtrsim \left\lfloor \log \frac{|B(x, 2^{m+3}\alpha_X)|}{|B(x, 2^m/[12C(\log K_j)^\varepsilon])|} \right\rfloor \cdot \frac{2^{2m}}{C^2(\log K_j)^{2\varepsilon}} \\ &\gtrsim \left\lfloor \log \frac{|B(x, 2^{m+3}\alpha_X)|}{|B(x, 2^m/[12C(\log K_{j+1})^\varepsilon])|} \right\rfloor \cdot \frac{d(x,y)^2}{C^2(\log K_j)^{2\varepsilon}} && (10) \\ &\gtrsim \lfloor \log K_{j+1} \rfloor \cdot \frac{d(x,y)^2}{C^2(\log K_j)^{2\varepsilon}} && (11) \\ &\gtrsim \frac{d(x,y)^2}{C^2(\log K_j)^{2\varepsilon-1}}, && (12) \end{aligned}$$

where in (10) we used the fact that $K_{j+1} \le K_j$ and $d(x, y) \le 2^{m+1}$, in (11) we used the fact that $x \notin S_{2^m}(K_{j+1})$, and in (12) we used the fact that $K_{j+1} = \sqrt{K_j} \ge 2$.

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

$$|B(x, 2^{m+3}\alpha_X)| \le 4|B(x, 2^m/[12C])|.$$

In particular, using the notation of Claim 4.6, $x \in J_4(2^m)$. By Claim 4.6, there is a mapping $f_{N+1} : X \to L_2$ which is Lipschitz with constant $O(\sqrt{\log n})$ and for every $x, y \in S_{2^m}(K_N) \subseteq J_4(2^m)$, $\|f_{N+1}(x) - f_{N+1}(y)\|_2 \gtrsim d(x, y)$.

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

$$\|\Phi(x) - \Phi(y)\|_2 \ge \|f_{N+1}(x) - f_{N+1}(y)\|_2 \gtrsim d(x, y).$$

Otherwise, without loss of generality there is $j \in \{0, \dots, N-1\}$ such that $x \in S_{2^m}(K_j) \setminus S_{2^m}(K_{j+1})$, in which case by (12)

$$\|\Phi(x) - \Phi(y)\|_2 \ge \|f_{j+1}(x) - f_{j+1}(y)\|_2 \gtrsim \frac{d(x,y)}{C(\log K_j)^{\varepsilon-\frac{1}{2}}} \ge \frac{d(x,y)}{C(\log n)^{\varepsilon-\frac{1}{2}}}.$$

□

5 The sparsest cut problem with general demands

This section is devoted to the proof of Theorem 1.2. Our argument follows the well known approach for deducing the algorithmic Theorem 1.2 from the embedding result contained in Theorem 1.1 (see e.g. [27, 5, 16]).

5.1 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 programming [27].

Corollary 5.1. *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 1-Lipschitz map $f : Y \to L_2$ such that the map $f|_X : X \to L_2$ has distortion $O\left(\sqrt{\log k} \cdot \log\log k\right)$.*

Proof. We observe that the maps used to prove Theorem 1.1, i.e. those produced in Lemma 3.5 and Claim 4.6, are of Fréchet-type. In other words, there is a probability space (Ω, μ) over subsets $A_\omega \subseteq X$ for $\omega \in \Omega$, and we obtain a maps $\varphi_{S,\tau} : X \to L_2(\mu)$ given by $\varphi_{S,\tau}(x)(\omega) = d(x, A_\omega)$. We can then define the extension $\varphi_{S,\tau} : Y \to L_2(\mu)$ by

$$\varphi_{S,\tau}(y)(\omega) = d(y, A_\omega).$$

Thus by extending the ensemble of maps $\{\varphi_{S,\tau}\}$ to the larger space Y before the application of Theorem 4.1, we can ensure that the final embedding is 1-Lipschitz on Y. □

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

Claim 5.2. *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, which can be solved in polynomial time using the methods of [17].

SDP (5.1)

$$\begin{array}{lll} \max & \varepsilon & \\ \text{s.t.} & x_u \in \mathbb{R}^n & \forall u \in Y \\ & \|x_u - x_v\|_2^2 \le d(u,v)^2 & \forall u, v \in Y \\ & \|x_u - x_v\|_2^2 \ge \varepsilon\, d(u,v)^2 & \forall u, v \in X \end{array}$$

□

5.2 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

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

and we let $\Phi^*(V, w_N, w_D) = \min_{S \subseteq V} \Phi_{w_N,w_D}(S)$. (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 $\Phi^*(V, w_N, w_D)$ is NP-hard [30]. The following semi-definite program is well known to be a relaxation of $\Phi^*(V, w_N, w_D)$ (see e.g. [16]).

SDP (5.2)

$$
\begin{array}{rl}
\min & \sum_{u,v \in V} w_N(u,v) \, \|x_u - x_v\|_2^2 \\
\text{s.t.} & x_u \in \mathbb{R}^n \qquad\qquad \forall u \in V \\
& \sum_{u,v \in V} w_D(u,v) \, \|x_u - x_v\|_2^2 = 1 \\
& \|x_u - x_v\|_2^2 \le \|x_u - x_w\|_2^2 + \|x_w - x_v\|_2^2 \\
& \qquad\qquad\qquad\quad \forall u, v, w \in V
\end{array}
$$

Furthermore, an optimal solution to this SDP can be computed in polynomial time [17, 16].

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 $k = |U|$. Denote $M = 20 \log n$.

1. Solve SDP (5.2), yielding a solution $\{x_u\}_{u \in V} \subseteq \mathbb{R}^n$.
2. Consider the metric space (V, d) given by $d(u,v) = \|x_u - x_v\|_2^2$.
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
$$(\{v_1^i, \ldots, v_m^i\}, \{v_{m+1}^i, \ldots, v_n^i\}), \qquad 1 \le m \le n-1, \ 1 \le i \le M.$$

Claim 5.3. *With constant probability over the choice of $\beta_1, \ldots, \beta_M$, the cut $(S, V \setminus S)$ returned by the algorithm has*
$$\Phi_{w_N,w_D}(S) \le O\left(\sqrt{\log k} \log\log k\right) \Phi^*(V, w_N, w_D). \tag{13}$$

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, 32]) 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 (13) 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 [27, 12] that there exist constants $\alpha_1, \alpha_2, \ldots, \alpha_{Mn} \ge 0$ 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 $(S_i, V \setminus S_i)$ 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 5.1, $\Lambda = O\left(\sqrt{\log k}\log\log k\right)$. Recalling that $w_D(u,v) > 0$ only when $u, v \in U$,

$$\begin{aligned}
\Phi^*(V, w_N, w_D) &\geq \frac{\sum_{u,v\in V} w_N(u,v)\|x_u - x_v\|_2^2}{\sum_{u,v\in U} w_D(u,v)\|x_u - x_v\|_2^2} \\
&\geq \frac{1}{\Lambda}\frac{\sum_{u,v\in V} w_N(u,v)\,||g(f(u)) - g(f(v))||_1}{\sum_{u,v\in U} w_D(u,v)\,||g(f(u)) - g(f(v))||_1} \\
&= \frac{1}{\Lambda}\frac{\sum_{i=1}^{Mn}\alpha_i\sum_{u,v\in V} w_N(u,v)\rho_{S_i}(u,v)}{\sum_{i=1}^{Mn}\alpha_i\sum_{u,v\in U} w_D(u,v)\rho_{S_i}(u,v)} \\
&\geq \frac{1}{\Lambda}\min_i \frac{\sum_{u,v\in V} w_N(u,v)\rho_{S_i}(u,v)}{\sum_{u,v\in U} w_D(u,v)\rho_{S_i}(u,v)} \\
&= \frac{\Phi_{w_N,w_D}(S)}{\Lambda}.
\end{aligned}$$

This completes the proof. □

6 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 Theorem 4.5 is applied with $A, B \sim \mathrm{polylog}(n)$. The need for such values arises out of a certain non-locality property which seems inherent to the method of proof in [4]. We remark that achieving $A = O(1)$ is probably possible, and it seems that B is the difficult factor.

  The other factor arises because, in proving Theorem 1.1, we invoke Theorem 4.5 for $O(\log\log n)$ 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-known lowerbound is $\Omega(\log\log n)^\delta$ [20]. We also note that assuming a strong form of the *Unique Games Conjecture* is true, the general Sparsest Cut problem is hard to approximate within a factor of $\Omega(\log\log n)$ [11, 20].

- For the uniform case of Sparsest Cut, it is possible to achieve a $O(\sqrt{\log n})$ approximation in quadratic time without solving an SDP [3]. 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 [24] 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 [35, 33] that there are arbitrarily large n-point subsets of L_1 whose minimal distortion into L_p is $1+\Theta\left(\sqrt{\frac{\log n}{p}}\right)$ (the dependence on n follows from [35], and the optimal dependence on p follows from the results of [33]). 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 $x \neq b$,

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

We refer to [18] 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 [22], 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.

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] A. Agrawal, P. Klein, R. Ravi, and S. Rao. Approximation through multicommodity flow. In *31st Annual Symposium on Foundations of Computer Science*, pages 726–737. IEEE Computer Soc., Los Alamitos, CA, 1990.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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*, pages 514–523, 2003.

[9] G. Calinescu, H. Karloff, and Y. 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.

[10] S. Chawla, A. Gupta, and H. Räcke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 102–111, Vancouver, 2005.

[11] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On embeddability of negative type metrics into ℓ_1. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 144–153, 2005.

[12] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997.

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

[14] 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.

[15] 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.

[16] M. X. Goemans. Semidefinite programming and combinatorial optimization. *Mathematical Programming*, 49:143–161, 1997.

[17] M. Grötschel, L. L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1:169–197, 1981.

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

[19] 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.

[20] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th Annual IEEE Conference on Foundations of Computer Science*, 2005. To appear.

[21] 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.

[22] 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*, pages 92–101, Vancouver, 2005.

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

[24] 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.

[25] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[26] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999.

[27] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[28] N. Linial and M. Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993.

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

[30] D. W. Matula and F. Shahrokhi. The maximum concurrent flow problem. *J. ACM*, 37(2):318–334, 1990.

[31] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Adv. Math.*, 189(2):451–494, 2004.

[32] 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.

[33] A. Naor, Y. Peres, O. Schramm, and S. Sheffield. Markov chains in smooth normed spaces and Gromov hyperbolic metric spaces. Preprint, 2004.

[34] A. Naor, Y. Rabani, and A. Sinclair. Quasisymmetric embeddings, the observable diameter, and expansion properties of graphs. *J. Funct. Anal.* To appear.

[35] A. Naor and G. Schechtman. Remarks on non linear type and Pisier's inequality. *J. Reine Angew. Math.*, 552:213–236, 2002.

[36] 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.

[37] D. B. Shmoys. Cut problems and their application to divide-and-conquer. In *Approximation Algorithms for NP-hard Problems, (D.S. Hochbaum, ed.)*, pages 192–235. PWS, 1997.

[38] V. V. Vazirani. *Approximation algorithms*. Springer-Verlag, Berlin, 2001.