Distance scales, embeddings, and metrics of negative type*

[preliminary draft]

James R. Lee †
University of California, Berkeley

February 2, 2006

Abstract

We introduce a new number of new techniques for the construction of low-distortion embeddings of a finite metric space. These include a generic Gluing Lemma which avoids the overhead typically incurred from the naïve concatenation of maps for different scales of a space. We also give a significantly improved and quantitatively optimal version of the main structural theorem of Arora, Rao, and Vazirani on separated sets in metrics of negative type. The latter result offers a simple hyperplane rounding algorithm for the computation of an $O(\sqrt{\log n})$-approximation to the Sparsest Cut problem with uniform demands, and has a number of other applications to embeddings and approximation algorithms.

1 Introduction

Low distortion embeddings of finite metric spaces into normed spaces (in particular L_1 and L_2) have provided an essential tool in the construction and analysis of approximate algorithms for a variety of fundamental problems (see, e.g. [LLR95, AR98, Rao99, Fei00, ARV04], the surveys [Ind01, Lin02], and the book chapter [Mat02, Ch. 15]).

A seminal result in this field is the optimal $O(\log k)$-approximate max-flow/min-cut theorem for multiflow instances with k commodities, discovered independently by Linial, London, and Rabinovich [LLR95] and Aumann and Rabani [AR98]. Their arguments are based on a theorem of Bourgain stating that every n-point metric space embeds into a Euclidean space with distortion at most $O(\log n)$ [Bou85]. These same techniques also yield an $O(\log n)$-approximation algorithm for the Sparsest Cut problem on graphs with n nodes.

One potential path to better approximations is a well-known semi-definite program (SDP) relaxation for the Sparsest Cut problem (see, e.g. [Goe97, Lin02, ARV04]). It is known that the integrality gap of this SDP is equal to the worst distortion required to embed any n-point metric of negative type into L_1 (see [Mat02, Ch. 15]). The primary goal of the present work is to study the construction of such embeddings. A metric space (X, d) is said to be of *negative type*[1] if there

---

*A preliminary version of this paper appeared in the 16th annual Symposium on Discrete Algorithms (SODA 2005).

†Computer Science Division, University of California, Berkeley, CA 94720. Supported by NSF grant CCR-0121555 and an NSF Graduate Research Fellowship. Email: `jrl@cs.berkeley.edu`

[1]The name comes from the fact that if (X, d) is a metric space of negative type, then for every $\{x_1, x_2, \ldots, x_k\} \subseteq X$, the matrix $\{d(x_i, x_j)\}_{i,j}$ is negative definite.

exists a mapping $f : X \to L_2$ such that $d(x,y) = \|f(x) - f(y)\|_2^2$ for every $x, y \in X$ (equivalently, the metric space $(X, \sqrt{d})$ embeds isometrically into L_2).

Until recently, relevant properties of negative-type metrics were poorly understood. In an important breakthrough, Arora, Rao, and Vazirani [ARV04] obtained deep results about the structure of such spaces. In particular, they showed that if (X, d) is a negative-type metric on n-points such that $\operatorname{diam}(X) \leq 1$ and $\frac{1}{n^2}\sum_{x,y \in X} d(x,y) = \Omega(1)$, then there exist two "well-separated" subsets $A, B \subseteq X$ with $|A|, |B| = \Omega(n)$ and such that $d(A,B) = \min_{a \in A, b \in B} d(a,b) \geq 1/O(\sqrt{\log n})$. In particular, this allows one to obtain an $O(\sqrt{\log n})$-approximation for the Sparsest Cut problem with uniform demands (i.e. the edge expansion problem in graphs).

Quantitatively, this theorem is tight, but for constructing low-distortion embeddings (and to handle general demands) it is essential to have more than a single pair of separated sets (A, B). We show that it is possible to choose these pairs of separated sets in a somewhat "random" manner. This requires significant changes to the [ARV04] induction argument, which we outline in the next section.

Another key ingredient in the construction of many embeddings for finite spaces is the ability to deal with distinct "scales" of a space separately. Recently, Krauthgamer, Lee, Mendel, and Naor introduced a technique called *measured descent* [KLMN05] which allows one to do this in a non-trivial way for special kinds of maps (those produced via Fréchet embeddings). In this work, we show how appropriately constructed partitions of unity can be used to glue together *arbitrary* embeddings. We give an application to non-Fréchet Euclidean embeddings for finite subsets of L_p with $1 < p < 2$, which yields the first non-trivial Euclidean embeddings of such spaces.

Subsequently, many of the questions addressed by the initial version of this paper have been partially answered, and in some cases rely heavily on the theorems and techniques developed here. For that reason, we also spend time developing the various corollaries required for these other results.

Subsequent work. Chawla, Gupta, and Räcke [CGR05] have broken the $O(\log n)$ barrier for embedding n-point negative-type metrics into L_2. They achieve a distortion bound of $O(\log n)^{3/4}$ by combining the measured descent technique of [KLMN05] with the strong form of the [ARV04] geometric structure theorem provided in the present work (Theorem 4.7). Arora, Lee, and Naor subsequently obtained a near-optimal bound of $O(\sqrt{\log n}\log\log n)$. Their approach also requires Theorem 4.7, as well as a variant of our Gluing Lemma 2.1.

There are a number of other recently developed approximation algorithms that require our strong version of the [ARV04] techniques in order to obtain better analyses. In these cases, the approximation ratio obtained using [ARV04] is $O(\log n)^{2/3}$, while $O(\sqrt{\log n})$ is achievable using Theorem 4.7 and its consequences. These include the vanishing term for vertex cover [Kar01], Min 2CNF Deletion and directed cut problems [ACMM05], and directed vertex ordering problems [CHKR06]. Finally, in [Lee06], the author constructs embeddings of Euclidean metrics with near-optimal volume distortion (in the sense of Feige [Fei00]) which crucially require the gluing techniques introduced in this paper.

1.1 Preliminaries

All logarithms are base 2 unless otherwise specified.

Distortion. A map $f : X \to Y$ between two metric spaces (X, d_X) and (Y, d_Y) is said to be

L-Lipschitz (or *Lipschitz* with constant L) if

$$d_Y(f(x), f(y)) \leq L \cdot d_X(x, y)$$

for all $x, y \in X$. One defines

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

as the smallest L for which f is L-Lipschitz. If, in addition, f is injective and $\|f^{-1}\|_{\mathrm{Lip}}$ is finite, we say that f is *bi-Lipschitz* and denote the distortion of f by $\mathrm{distortion}(f) = \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$. If f has distortion D, we refer to it as a D-embedding of X into Y. The least distortion with which X may be embedded in Y is denoted $c_Y(X)$. When $Y = L_p$ we use the notation $c_Y(\cdot) = c_p(\cdot)$. Finally, the parameter $c_2(X)$ is called the Euclidean distortion of X.

Concatenation, extension, restriction. Given two maps $f_1 : X \to Y_1$ and $f_2 : X \to Y_2$, we define their *direct sum* $f_1 \oplus f_2$ as the map $f : X \to Y_1 \times Y_2$ given by $f(x) = (f_1(x), f_2(x))$. This extends naturally to a direct sum of more than two functions, which we sometimes denote by $\bigoplus_{\alpha \in I} f_\alpha$ where I is some index family. For a subset $S \subseteq X$, and a map $f : S \to Y$, we refer to the map $\tilde{f} : X \to Y$ as an *extension of* f if $\tilde{f}(x) = f(x)$ for all $x \in S$. In general, we denote the restriction of $\tilde{f}$ to S by $\tilde{f}|_S$.

Subsets. We denote by $B(x, r) = \{y \in X : d_X(x, y) \leq r\}$ and $B^\circ(x, r) = \{y \in X : d_X(x, y) < r\}$ the closed ball and open ball of radius r about x, respectively. For a subset $S \subseteq X$, we write $d_X(x, S) = \inf_{y \in S} d_X(x, y)$, and define $\mathrm{diam}(S) = \sup_{x,y \in S} d_X(x, y)$.

Finally, we say that a subset N of X is an *ϵ-net* if it satisfies (1) For every $x, y \in N, d(x, y) \geq \epsilon$ and (2) $X \subseteq \bigcup_{y \in N} B(y, \epsilon)$. Such nets always exist for any $\epsilon > 0$. For finite metrics, they can be constructed greedily. For arbitrary metrics, proof of their existence is an easy application of Zorn's lemma.

1.2 Results and techniques

We now discuss briefly the main results of the paper, and the techniques involved in their proof. For two metric spaces X and Y, we define $D_n(X, Y) = \sup\{c_Y(A) : A \subseteq X, |A| = n\}$. In words, $D_n(X, Y)$ describes the distortion required to embed every n-point subspace of X into Y. Additionally, for a family of metric spaces $\mathcal{F}$, we define $D_n(\mathcal{F}, Y) = \sup\{D_n(X, Y) : X \in \mathcal{F}\}$. If we denote by NEG the family of all negative-type metrics, then our main object of study becomes $D_n(\mathsf{NEG}, L_1)$. We recall that since L_2 embeds isometrically into L_1, we have $D_n(X, L_1) \leq D_n(X, L_2)$ for any metric space X (see, e.g. [BL00]).

Gluing single-scale embeddings. In Section 2, we reduce the problem of constructing a non-trivial embedding of a finite metric space to that of "handling a single scale." Let (X, d_X) and (Y, d_Y) be metric spaces, and let $\tau \in \mathbb{R}_+$ be a positive real number. If $f : X \to Y$ is a 1-Lipschitz map such that for every $x, y \in X$ with $d(x, y) \in [\tau, 2\tau]$, we have

$$d_Y(f(x), f(y)) \geq \frac{\tau}{K},$$

then we will call f a *scale-τ* embedding with deficiency K. Analogous to $D_n(\cdot, \cdot)$, we define $K_n^\tau(X, Y)$ to be the infimal value K such that for every n-point subset $A \subseteq X$, there exists a scale-τ map from A into Y with deficiency at most K. Finally, we define $K_n(X, Y) = \sup_{\tau > 0} K_n^\tau(X, Y)$.

Certainly one has $D_n(X,Y) \geq K_n(X,Y)$ for any pair of spaces X, Y, but a priori it is not clear whether any meaningful relationship holds in the other direction, even for $Y = L_2$. We exhibit such a relationship.

Lemma 1.1 (Gluing Lemma)**.** *For any metric space X, one has*

$$D_n(X, L_2) \leq O(\sqrt{K_n(X, L_2) \cdot \log n}).$$

Related theorems hold when L_2 is replaced by L_p for any $p \in [1, \infty)$. By Bourgain's embedding theorem, we know that $D_n(X, L_2) = O(\log n)$ for any X, and the above theorem characterizes spaces which admit better embeddings than the general case: $D_n(X, L_2) = o(\log n)$ if and only if $K_n(X, L_2) = o(\log n)$. We remark that bounding $K_n(X, L_2)$ is often much easier than bounding $D_n(X, L_2)$, and we will see examples of this in Section 3.

In some sense, Lemma 1.1 is tight. If planar denotes the family of all planar graph metrics, then $K_n(\mathsf{planar}, L_2) = O(1)$ by a result of Rao [Rao99] (see also [KPR93]). On the other hand, $D_n(\mathsf{planar}, L_2) = \Theta(\sqrt{\log n})$ (the upper bound is from [Rao99], and the lower bound is from Newman and Rabinovich [NR03]). Additionally, if $\mathcal{F}$ denotes a family of shortest-path metrics on $O(1)$-degree expander graphs, then one has both $K_n(\mathcal{F}, L_2) = \Theta(\log n)$ and $D_n(\mathcal{F}, L_2) = \Theta(\log n)$ (see [LLR95, AR98] and also [Mat02, Ch. 15]). On the other hand, for values of $K_n(\cdot, L_2)$ strictly between $O(1)$ and $\Theta(\log n)$, the correct asymptotic dependence in Lemma 1.1 is unknown. We conjecture the optimal bound:

Conjecture 1. *For any metric space X, $D_n(X, L_2) = O\left(K_n(X, L_2) + \sqrt{\log n}\right)$.*

We remark that if we replace L_2 by L_1, then much less is known. Our results imply that $D_n(X, L_1) = O(\sqrt{K_n(X, L_1) \cdot \log n})$, but there is no lower bound ruling out a statement such as $D_n(X, L_1) = O(K_n(X, L_1))$. If such a result were true, it would imply, e.g. that every planar graph metric embeds into L_1 with $O(1)$-distortion, resolving a long-standing open problem [GNRS99].

The main technical contribution which allows us to concentrate on single scale embeddings throughout is the Gluing Lemma proved in Section 2. Our approach to proving Lemma 1.1 is based on the ideology of measured descent, but is technically quite different. In particular, the authors of [KLMN05] proved a version of the Lemma 1.1 when the single scale maps are of a special form known as Fréchet embeddings.

For general embeddings, we use a standard analytical tool called a *partition of unity.* If X is a metric space, then a family of maps $\{\rho_t : X \to [0, 1]\}_{t \in T}$ (for some countable index set T) is called a partition of unity if $\sum_{t \in T} \rho_t(x) = 1$ for every $x \in X$. Such families are distinguished from arbitrary probability measures on X because we will often require the constituent functions $\rho_t : X \to [0, 1]$ to be smooth in some sense. In this present context, we will be chiefly concerned with quantitative bounds on $\|\rho_t\|_{\text{Lip}}$. Given a family of embeddings, e.g. $\{f_t : X \to L_2\}_{t \in T}$, we can glue them together in two different ways, by defining

$$F_+(x) = \sum_{t \in T} \rho_t(x) f_t(x) \qquad \text{or} \qquad F_\oplus(x) = \bigoplus_{t \in T} \rho_t(x) f_t(x),$$

depending on the intended use.

The Euclidean distortion of L_p spaces, $1 < p < 2$**.** One of the most natural questions about finite metrics that arises when one searches for analogues with the local theory of Banach spaces involves the distortion required to embed n-point subsets of L_p into L_q for $p \neq q$. In other words, given n-points represented in some L_p norm, what is the distortion required to represent them under the L_q norm? (By changing their coordinate representations, of course.) Previous knowledge can be summarized by the following results. We discuss first the known upper bounds (the notations $O(\cdot)$ and $\Omega(\cdot)$ may hide a multiplicative factor which depends on p). The last three hold for all $p \in [1, \infty]$.

1. $D_n(L_p, L_q) = 1$ for $1 \leq q \leq p \leq 2$.
2. $D_n(L_p, L_\infty) = 1$.
3. $D_n(L_2, L_p) = 1$.
4. $D_n(L_p, L_2) = O(\log n)$.

(2) follows because $D_n(X, L_\infty) = 1$ for any metric space X, using Fréchet's embedding (see, e.g. [Mat02]); (1) and (3) follow from the existence of symmetric p-stables (see, e.g. [BL00]); and (4) follows because $D_n(X, L_2) = O(\log n)$ for any metric space X by Bourgain's theorem [Bou85]. Thus the only previous upper bounds either hold for general metric spaces (and thus don't depend on the properties of n-point subsets of L_p), or involve isometric embeddability.

In Section 3, we prove that, for $1 \leq p \leq 2$,

$$D_n(L_p, L_2) \leq (\log n)^{\frac{1}{2}+\frac{1}{p}(\frac{1}{p}-\frac{1}{2})}.$$

Note that this bound is better than $O(\log n)$ whenever $p > 1$.

Our embedding constructs a scale-1 map, but unlike many embedding theorems (e.g. [Bou85, Rao99, KLMN05]), is not based on piece-wise line embeddings. To counter this, a sophisticated extension lemma of Marcus and Pisier [MP84] is used. We proceed by first embedding the $p/2$ power of the L_p metric into L_2. (It is well known that the metric $\|x - y\|_p^{p/2}$ is isometric to a subset of L_2). Then, we use a lemma of [MN04] to "truncate" all the distances in the image to be at most 1, incurring only bounded distortion. After restricting the embedding to an ε-net, the distortion is $\varepsilon^{p/2-1}$. The Marcus-Pisier lemma is used to extend the map to the rest of the space, and one can prove that all pairs with $\|x - y\|_p \in [1, 2)$ are mapped far apart as long as ε is sufficiently small. After calculating the optimal value of ε, we obtain a scale-1 map with deficiency $K = O(\log n)^{\frac{2}{p}(\frac{1}{p}-\frac{1}{2})}$, and applying the Gluing Lemma completes the proof (by scale invariance of L_p spaces, we need only generate a map for some fixed scale).

Other bounds from recent work. In an earlier version of this paper, we showed the following: For every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$D_n(L_1, L_{2+\epsilon}) \leq O(\log n)^{1-\delta}.$$

This proof was based on the Gluing Lemma, along with our improved structure theorem for NEG-metrics, and a rather inefficient induction argument. Subsequently, the papers [CGR05, ALN05] have replaced the induction argument with a far more sophisticated approach. Arora, Lee, and Naor achieve an upper bound of $D_n(\mathsf{NEG}, L_2) \leq O(\sqrt{\log n} \log\log n)$. Since $L_1 \in \mathsf{NEG}$, this implies that

$$\sqrt{\log n} \leq D_n(L_1, L_2) \leq O(\sqrt{\log n} \log\log n),$$

where the lower bound is due to Enflo [Enf69]. We remark that our argument for $L_p, 1 < p < 2$ yields the bound $K_n(L_p, L_2) \leq O(\log n)^{\frac{2}{p}(\frac{1}{p}-\frac{1}{2})}$, which remains the best-known for certain values of

p (i.e. the exponent approaches 0 as $p \to 2$). For completeness, we summarize the known bounds when the target space is L_2. For $1 \leq p \leq 2$,

$$(\log n)^{\frac{1}{p}-\frac{1}{2}} \leq K_n(L_p, L_2) \leq D_n(L_p, L_2) \leq O(\sqrt{\log n} \log\log n),$$

where the lower bound is from [Enf69]. And for $p \geq 2$,

$$\Omega\left(\frac{\log n}{\log\log n}\right)^{\frac{1}{2}-\frac{1}{p}} \leq K_n(L_p, L_2) \leq D_n(L_p, L_2) \leq O(\log n),$$

where the lower bound follows from differentiation and cotype considerations (see, e.g. [BL00] or [MN06] for a "non-linear" proof based on metric cotype).

We also mention one case where one has nearly tight bounds. For any $1 \leq p < \infty$, we have

$$\Omega(\sqrt{\log n}) \leq K_n(L_1, L_p) \leq D_n(L_1, L_p) \leq D_n(L_1, L_2) \leq O(\sqrt{\log n} \log\log n).$$

The lower bound is due to [LN04] for $1 \leq p < 2$ and to [NS02] for $2 < p < \infty$.

Random projections and metrics of negative type. Our final ingredient is an improved structure theorem for metrics of negative type. In fact, the technical core of the argument contains a deep fact about configurations of finite sets of points in Euclidean spaces.

In Section 4, we obtain the "Big Core" Theorem, which is a quantitatively optimal version of the main geometric lemma of [ARV04] (it shows that a certain geometric object we call a *core* must be composed of many points). The beautiful induction employed in [ARV04] degrades with each step, and this bounds the number of iterations for which the induction can be carried out. Here, we show how to avoid this degradation by performing the time-honored technique of *region growing* as a phase of every induction step.

One benefit of the improved core theorem is that it allows one to significantly simplify the $O(\sqrt{\log n})$-approximation for the Sparsest Cut problem obtained in [ARV04]. In that paper, the authors present a simple hyperplane rounding algorithm which they show achieves an approximation ratio of $O(\log^{2/3} n)$. A more complicated algorithm is used to obtain the $O(\sqrt{\log n})$-approximation. The Big Core Theorem (Theorem 4.7) shows that the [ARV04] hyperplane rounding algorithm already achieves $O(\sqrt{\log n})$. This is discussed in Section 4.3, along with corollaries required for other applications that cannot rely on [ARV04].

2 Single-scale embeddings and the Gluing Lemma

Let (X, d) be an n-point metric space. The goal of this section is to prove the following lemma which shows that the ideology of [KLMN05] can be extended beyond Fréchet embeddings, to any collection of single-scale embeddings. The lemma is stated for the case of Euclidean embeddings, but after the proof we discuss the extension to any $p \in [1, \infty)$.

Lemma 2.1 (Gluing Lemma)**.** *Suppose that for each $m \in \mathbb{Z}$, there exists a scale-2^m embedding $\phi_m : X \to L_2$ with deficiency K. Then $c_2(X) \leq O(\sqrt{K \log n})$.*

For applications to L_1 (and hence NEG), using the Gluing Lemma as a black box can never yield a result stronger than $D_n(L_1, L_2) = O(\log n)^{3/4}$. It suffices to show that for $\tau = k$, any scale-τ map for the k-dimensional cube requires deficiency $\Omega(\sqrt{k})$. To see this, let $Q_k = (\{0,1\}^k, d)$ be the cube with the L_1 metric, and suppose that $f : Q_k \to L_2$ is a scale-k map with deficiency K, then using the standard Poincaré inequality for the cube (see, e.g. [Mat02]),

$$\frac{2^{k-1} \cdot k^2}{K^2} \leq \sum_{x,y \in Q_k : d(x,y)=k} \|f(x) - f(y)\|_2^2 \leq \sum_{x,y \in Q_k : d(x,y)=1} \|f(x) - f(y)\|_2^2 \leq k \cdot 2^{k-1},$$

so that $K \geq \sqrt{k}$. We now proceed with the proof of Lemma 2.1.

Proof of Lemma 2.1. Let $\|\cdot\| = \|\cdot\|_2$, and for $x, y \in X$, define

$$\rho_m(x,y) = \begin{cases} x & \text{if } |B(x, 2^m)| \geq |B(y, 2^m)| \\ y & \text{otherwise.} \end{cases}$$

The following lemma forms the technical core of our argument.

Lemma 2.2. *Given, for every $m \in \mathbb{Z}$, a 1-Lipschitz map $h_m : X \to L_2$, there exists a map $H : X \to L_2$ which satisfies*

1. $\|H\|_{\mathrm{Lip}} \leq O(\sqrt{\log n})$.
2. *For every $m \in \mathbb{Z}$, and every $x, y \in X$ with $d(x,y) \in [2^m, 2^{m+1})$, we have*

$$\|H(x) - H(y)\|_2 \geq \sqrt{\left\lfloor \log \frac{|B(\rho_{m-3}(x,y), 2^{m+1})|}{|B(\rho_{m-3}(x,y), 2^{m-3})|} \right\rfloor} \cdot \|h_m(x) - h_m(y)\|_2.$$

Before we prove the preceding lemma, let us see that it suffices to complete the proof of Lemma 2.1. First, we need the following two results. The first theorem uses Rao's technique [Rao99] applied to the decomposition theorem of [FRT03] to construct a map for each scale, and is a standard construction; see, e.g. [KLMN05]. Our intention is to apply Lemma 2.2 to this ensemble of maps.

Theorem 2.3. *For every $m \in \mathbb{Z}$, there exists a map $f_m : X \to L_2$ with $\|f_m\|_{\mathrm{Lip}} \leq 1$ and such that for all $x, y \in X$ with $d(x,y) \in [2^m, 2^{m+1}]$,*

$$\|f_m(x) - f_m(y)\| \geq \frac{d(x,y)}{1 + O\left(\log \frac{|B(x,2^{m+1})|}{|B(x,2^{m-3})|}\right)}$$

The next map handles the case where

$$\log \frac{|B(x, 2^{m+1})|}{|B(x, 2^{m-3})|} < 1.$$

We use a simple variant of Bourgain's argument [Bou85].

Proposition 2.4. *There exists a map $M : X \to L_2$ such that*

1. $\|M\|_{\mathrm{Lip}} = O(\sqrt{\log n})$.
2. *For all $m \in \mathbb{Z}$, for all $x, y \in X$ satisfying $d(x,y) \in [2^m, 2^{m+1}]$ and $\log \frac{|B(x,2^{m-1})|}{|B(x,2^{m-2})|} < 1$,*

$$\|M(x) - M(y)\| \geq \Omega(1)\, d(x,y).$$

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 $f = g_1 \oplus \cdots \oplus g_{\lceil \log n \rceil}$. 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 $\|F\|_{\text{Lip}} \leq O\left(\sqrt{\log n}\right)$.

Fix $x, y \in X$ such that $d(x, y) \in [2^m, 2^{m+1}]$ and with $\log \frac{|B(x,2^{m-1})|}{|B(x,2^{m-2})|} < 1$. Let $t \in \mathbb{N}$ be such that $2^t \leq |B(x, 2^{m-1})| \leq 2^{t+1}$. Let $\mathcal{E}_{\text{far}}$ be the event $\left\{d(x, W_t) \geq 2^{m-1}\right\}$ and let $\mathcal{E}_{\text{close}}$ be the event $\left\{d(x, W_t) \leq 2^{m-2}\right\}$. Also, define the event $\mathcal{E}^y_{\text{close}} = \{d(y, W_t) < 2^{m-1}\}$. Observe that each of the events $\mathcal{E}_{\text{close}}, \mathcal{E}_{\text{far}}$ are independent of $\mathcal{E}^y_{\text{close}}$ since the former events depend only on $W_t \cap B^\circ(x, 2^{m-1})$, and the latter on $W_t \cap B^\circ(y, 2^{m-1})$. It follows that

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

The final inequality holds true because $\log \frac{|B(x,2^{m-1})|}{|B(x,2^{m-2})|} < 1$ implies that $\Pr(\mathcal{E}_{\text{far}}), \Pr(\mathcal{E}_{\text{close}}) = \Omega(1)$. □

Let $F : X \to L_2$ and $\Phi : X \to L_2$ be the maps obtained by applying Lemma 2.2 to the collections $\{f_m\}$ and $\{\phi_m\}$, respectively. Let $M : X \to L_2$ be the map from Proposition 2.4. Let $\Psi = F \oplus \Phi \oplus M$ be our final embedding.

Using property (1) of Lemma 2.2 and property (1) of Proposition 2.4, we see that

$$\|\Psi\|_{\text{Lip}} \leq \|F\|_{\text{Lip}} + \|\Phi\|_{\text{Lip}} + \|M\|_{\text{Lip}} = O(\sqrt{\log n}).$$

For the lower bound, fix $x, y \in X$. Assume without loss that $x = \rho_{m-3}(x, y)$, and set $A = \log \frac{|B(x,2^{m+1})|}{|B(x,2^{m-3})|}$. In this case, we have

$$\begin{aligned}
\|\Psi(x) - \Psi(y)\|^2 = \;& \|F(x) - F(y)\|^2 + \|\Phi(x) - \Phi(y)\|^2 + \|M(x) - M(y)\|^2 \\
&\geq \Omega(1)\, d(x, y)^2 \left(\frac{\lfloor A \rfloor}{(1 + A)^2} + \frac{\lfloor A \rfloor}{K^2} + \mathbf{1}_{\{A<1\}}\right) \\
&\geq \Omega(1)\, \frac{d(x, y)^2}{K},
\end{aligned}$$

where we observe that $\lfloor A \rfloor = 0$ implies $\mathbf{1}_{\{A<1\}} = 1$. We conclude that $\text{distortion}(\Psi) = O(\sqrt{K \log n})$.

Now we move onto the proof of Lemma 2.2.

Proof of Lemma 2.2. For every $t \in \{1, 2, \ldots, \lceil \log_2 n \rceil\}$ we produce a map $\psi_t : X \to L_2$. Our final map will be

$$H = \psi_1 \oplus \cdots \oplus \psi_{\lceil \log_2 n \rceil}.$$

We now show how to construct the map ψ_t.

First, we need to "truncate" the maps $\{h_m\}$ so that they glue properly. This is done using the following observation from [MN04]. Let $\ell_2^{\leq D}$ be the metric space $(\ell_2, \hat{d})$ where

$$\hat{d}(x, y) = \min\{\|x - y\|_2, D\}.$$

Lemma 2.5 ([MN04]). *There exists a map $G : \ell_2^{\leq D} \to \ell_2$ with distortion 2 and such that for every $x \in \ell_2$, $\|G(x)\|_2 \leq 2D$. In particular,*

$$\tfrac{1}{2}\min\{D, \|x-y\|_2\} \leq \|G(x) - G(y)\|_2 \leq \min\{D, \|x-y\|_2\}. \tag{1}$$

Proof (sketch). It suffices to prove the lemma for $D = 1$. See [BL00] for the appropriate background on positive and negative definite kernels and their relation to isometric embeddings in a Hilbert space. We assume the terminology used there. Observe that $(x, y) \mapsto \|x - y\|_2^2$ is a negative definite kernel, hence $e^{-\|x-y\|_2^2}$ is positive definite and thus $1 - e^{-\|x-y\|_2^2}$ is negative definite. It follows that there exists a map $G : \ell_2 \to \ell_2$ such that

$$\|G(x) - G(y)\|_2 = \sqrt{1 - e^{-\|x-y\|_2^2}}.$$

Now (1) follows by observing that for $a < 1$, $\sqrt{1 - e^{-a^2}} \approx a$. Clearly we may translate G so that $\|G(x)\| \leq 2$ for every $x \in \ell_2$. ☐

Using the above lemma, for every map $h_m : X \to L_2$, we may pass to a map $\hat{h}_m : X \to L_2$ which satisfies

$$\tfrac{1}{2}\min\{2^m, \|h_m(x) - h_m(y)\|\} \leq \|\hat{h}_m(x) - \hat{h}_m(y)\| \leq \|h_m(x) - h_m(y)\|,$$

and $\|\hat{h}_m(x)\| \leq 2^{m+1}$ for every $x \in X$. Now we continue with the construction of the map ψ_t.

Constructing the partition of unity. Define $R(x, t) = \sup\{R : |B(x, R)| \leq 2^t\}$. Let $\rho : \mathbb{R} \to \mathbb{R}_+$ be any $O(1)$-Lipschitz map with $\mathrm{supp}(\rho) \subset [2^{-4}, 2^4]$, $\rho \equiv 1$ on $[2^{-3}, 2^3]$, and $\rho \leq 1$ everywhere. Define

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

and set

$$\psi_t(x) = \bigoplus_{m \in \mathbb{Z}} \rho_{m,t}(x)\hat{h}_m(x).$$

Bounding the Lipschitz constant. Observe that for every t, the map $x \mapsto R(x, t)$ is 1-Lipschitz. It follows from the above lemma and the definition of ρ that

$$|\rho_{m,t}(x) - \rho_{m,t}(y)| \leq \tfrac{O(1)}{2^m}|R(x,t) - R(y,t)| \leq O(1)\frac{d(x,y)}{2^m}.$$

Now we write

$$\|\psi_t(x) - \psi_t(y)\|^2 = \sum_{m \in \mathbb{Z}} \|\rho_{m,t}(x)\hat{h}_m(x) - \rho_{m,t}(y)\hat{h}_m(y)\|^2.$$

Notice that there are only $O(1)$ non-zero terms in this sum because we have $\rho_{m,t}(x) \neq 0$ or $\rho_{m,t}(y) \neq 0$ only for $O(1)$ values of $m \in \mathbb{Z}$. To bound a non-zero summand, we use the chain rule:

$$\begin{aligned}\|\rho_{m,t}(x)\hat{h}_m(x) - \rho_{m,t}(y)\hat{h}_m(y)\| &\leq \|\hat{h}_m(x)\| \cdot |\rho_{m,t}(x) - \rho_{m,t}(y)| + |\rho_{m,t}(y)| \cdot \|\hat{h}_m(x) - \hat{h}_m(y)\| \\ &\leq 2^{m+1} \cdot O(1)\frac{d(x,y)}{2^m} + d(x,y) \leq O(1) \cdot d(x,y),\end{aligned}$$

where in the last line we have use the fact that h_m, and hence $\hat{h}_m$ is 1-Lipschitz for every $m \in \mathbb{Z}$. The upper bound of $\|H\|_{\text{Lip}} \leq O(\sqrt{\log n})$ follows immediately.

Obtaining a lower bound. To prove the lower bound, we fix $x, y \in X$ with $d(x,y) \in [2^m, 2^{m+1})$. Notice that if $\rho_{m,t}(x) = \rho_{m,t}(y) = 1$, then we have

$$\|\psi_t(x) - \psi_t(y)\| \geq \|\hat{h}_m(x) - \hat{h}_m(y)\| \geq \tfrac{1}{2}\|h_m(x) - h_m(y)\|.$$

Now we are left to count the number of values $t \in \{1, 2, \ldots, \lceil \log_2 n \rceil\}$ for which $\rho_{m,t}(x) = 1 = \rho_{m,t}(y)$.

Notice that $\rho_{m,t}(x) = 1$ if and only if $R(x,t) \in [2^{m-3}, 2^{m+3}]$, which happens if and only if

$$t \in \left[\log |B(x, 2^{m-3})|, \log |B(x, 2^{m+3})|\right].$$

Similarly, $\rho_{m,t}(y) = 1$ if and only if

$$t \in \left[\log |B(y, 2^{m-3})|, \log |B(y, 2^{m+3})|\right].$$

Assume without loss of generality that $x = \rho_{m-3}(x,y)$ so that $|B(x, 2^{m-3})| \geq |B(y, 2^{m-3})|$, then

$$t \in [\log |B(x, 2^{m-3})|, \log |B(x, 2^{m+1})|]$$

implies $\rho_{m,t}(x) = \rho_{m,t}(y) = 1$ (recalling that $R(x,t) \leq 2^{m+1}$ implies $R(y,t) \leq 2^{m+3}$ since $R(\cdot, t)$ is 1-Lipschitz and $d(x,y) \leq 2^{m+1}$). Hence the number of values of t for which $\rho_{m,t}(x) = \rho_{m,t}(y) = 1$ is at least

$$\left\lfloor \log \frac{|B(x, 2^{m+1})|}{|B(x, 2^{m-3})|} \right\rfloor.$$

We conclude that

$$\|H(x) - H(y)\|^2 \geq \frac{1}{2} \left\lfloor \log \frac{|B(\rho_{m-3}(x,y), 2^{m+1})|}{|B(\rho_{m-3}(x,y), 2^{m-3})|} \right\rfloor \cdot \|h_m(x) - h_m(y)\|^2,$$

completing the proof. □

□

Remark 2.1. *A simple modification of the proof shows that for any $p \in [1, \infty)$, a similar lemma holds with the conclusion that*

$$c_p(X) \leq O(K^{1-1/q} \log^{1/q} n)$$

where $q = \max\{2, p\}$, though there is one caveat: For $p \neq 2$, one must also assume that $\|\phi_m(x)\|_p \leq O(2^m)$ for every $x \in X$, since an analogue of Lemma 2.5 does not hold.

3 An application to $D_n(L_p, L_2)$

In this section, we apply the Gluing Lemma to obtain improved Euclidean embeddings of n-point subsets of L_p for $p \in (1, 2)$. The main theorem follows.

Theorem 3.1 (Embedding L_p spaces)**.** *For $1 < p \leq 2$,*

$$D_n(L_p, L_2) \leq O(\log n)^{\frac{1}{2} + \frac{1}{p}(\frac{1}{p} - \frac{1}{2})}.$$

Proof. Let $X \subseteq L_p$ be an n-point subset. It suffices to construct a scale-1 map $f : X \to L_2$ (because L_p is scale invariant) with deficiency $K = O(\log n)^{\frac{2}{p}(\frac{1}{p}-\frac{1}{2})}$, and then apply the Gluing Lemma. It is well-known that there exists a map $T : X \to L_2$ with $\|T(x) - T(y)\|_2 = \|x - y\|_p^{p/2}$. Applying Lemma 2.5, we also obtain a map $G : L_2^{\leq 1} \to L_2$.

Now we define $\hat{T} : X \to L_2$ by $\hat{T}(x) = G(T(x))$, so that

$$\frac{1}{2} \min\{1, \|x - y\|_p^{p/2}\} \leq \|\hat{T}(x) - \hat{T}(y)\|_2 \leq \min\{1, \|x - y\|_p^{p/2}\}.$$

Observe that we can truncate the image like this because we are only concerned with constructing a scale-1 embedding. Such an approach would fail if we were also required to preserve larger distances.

In order that $\|x - y\|_p^{p/2}$ not be too much larger than $\|x - y\|_p$, we need to ensure that $\|x - y\|_p$ is sufficiently large. To do this, we take an ε-net N in X, where $\varepsilon < \frac{1}{4}$ will be determined shortly. Observe that the function $\hat{T}|_N$ is Lipschitz with constant $\varepsilon^{p/2-1}$. In order to extend $\hat{T}|_N$ to all of X, we will need the following lemma.

Lemma 3.2 (Marcus-Pisier)**.** *For any $1 < p \leq 2$, let $X \subseteq L_p$ be an n-point subset, and let $f : X \to L_2$ be any map, then there exists an extension $\tilde{f} : L_p \to L_2$ with*

$$\|\tilde{f}\|_{\mathrm{Lip}} \leq C(p)(\log n)^{\frac{1}{p}-\frac{1}{2}} \|f\|_{\mathrm{Lip}},$$

where $C(p)$ is a constant depending only on p.

It follows that we can extend $\hat{T}|_N$ to all of X, obtaining an extension $\tilde{T}$ with

$$\|\tilde{T}\|_{\mathrm{Lip}} \leq C(p)(\log n)^{\frac{1}{p}-\frac{1}{2}} \varepsilon^{p/2-1}.$$

Now, fix a pair $x, y \in X$ with $\|x-y\|_p \in [1, 2]$, and let $x', y' \in N$ be such that $\|x-x'\|_p, \|y-y'\|_p \leq \varepsilon$, then

$$\begin{aligned} \|\tilde{T}(x) - \tilde{T}(y)\|_2 &\geq \|\hat{T}(x') - \hat{T}(y')\|_2 - \|\tilde{T}\|_{\mathrm{Lip}} \cdot (\|x - x'\|_p + \|y - y'\|_p) \\ &\geq \frac{1}{4} - 2\varepsilon\|\tilde{T}\|_{\mathrm{Lip}}. \end{aligned}$$

Now choosing

$$\varepsilon = \left[\frac{1}{16C(p)} (\log n)^{\frac{1}{2}-\frac{1}{p}} \right]^{2/p}$$

so that $\varepsilon \leq 1/(16\|\tilde{T}\|_{\mathrm{Lip}})$ implies $\|\tilde{T}(x) - \tilde{T}(y)\| \geq \frac{1}{16}$. A calculation shows that

$$\|\tilde{T}\|_{\mathrm{Lip}} \leq O(\log n)^{\frac{2}{p}(\frac{1}{p}-\frac{1}{2})}.$$

After scaling $\tilde{T}$ to be 1-Lipschitz, we obtain a scale-1 embedding with deficiency $O(\|\tilde{T}\|_{\mathrm{Lip}})$, completing the proof. $\square$

4 The Big Core Theorem

In this section, we study various properties of "covers" and random projections in Euclidean spaces. Our eventual goal is to provide an optimal version of the [ARV04] chaining argument, which is embodied in the "Big Core" Theorem of Section 4.2. In Section 4.3, we present an application to rounding the SDP for Sparsest Cut with uniform demands.

4.1 Covers, matchings, and cores

We consider the space $\mathbb{R}^d$ equipped with the 2-norm $\|\cdot\| = \|\cdot\|_2$. For a subset $S \subseteq \mathbb{R}^d$, we sometimes write $\mathsf{dist}_2(x, S) = \inf_{y \in S} \|x - y\|$. Let S^{d-1} denote the $(d-1)$-dimensional unit sphere. When we talk about a random vector $u \in S^{d-1}$, we are referring to the (unique) uniform surface measure. We will require a large deviation bound for random projections (see, e.g. [Led96]).

Lemma 4.1. *If $z \in \mathbb{R}^d$, then*

$$\Pr_{u \in S^{d-1}} \left[\langle z, u \rangle \geq \frac{\sigma}{\sqrt{d}} \right] \leq \exp\left(\frac{-\sigma^2}{2\|z\|^2} \right).$$

Euclidean covers. We will say that a point $x \in \mathbb{R}^d$ is (σ, δ, ℓ)*-covered by a set* $C \subseteq \mathbb{R}^d$ if the following two conditions are satisfied.

1. For every $y \in C$, $\|x - y\| \leq \ell$.
2. $\Pr_{u \in S^{d-1}} \left[\exists\, y \in C : \langle x - y, u \rangle \geq \frac{\sigma}{\sqrt{d}} \right] \geq \delta$.

We also say that a set of points $S \subseteq \mathbb{R}^d$ is (σ, δ, ℓ)-covered by a set $C \subseteq \mathbb{R}^d$ if every $x \in S$ is (σ, δ, ℓ)-covered by $B_C(x, \ell) = B(x, \ell) \cap C$.

We now give three lemmas dealing with the structural properties of covers (each of these claims appears, in some form, in [ARV04]). First, we have the following simple estimate on the size of covers.

Lemma 4.2. *If x is (σ, δ, ℓ)-covered by a set C, then*

$$|C| \geq \delta \cdot \exp\left(\frac{\sigma^2}{2\,\ell^2} \right).$$

Proof. For any $y \in C$, apply Lemma 4.1 with $z = x - y$. Since $\|x - y\| \leq \ell$, y can "cover" x in at most an $\exp(-\sigma^2/(2\ell^2))$ fraction of directions. □

The next lemma shows that if x is covered by C, then the cover can be extended to a nearby point y with only a small loss in the parameters.

Lemma 4.3. *Suppose that x is (σ, δ, ℓ)-covered by C, and $z \in \mathbb{R}^d$. Then for every $t \geq 0$, z is $\big(\sigma - t \cdot \|x - z\|, \delta - \exp(-t^2/2), \ell + \|x - z\|\big)$-covered by C.*

Proof. In order to have $\langle x - y, u \rangle \geq \sigma/\sqrt{d}$ for some $y \in C$, but $\langle z - y, u \rangle < [\sigma - t \cdot \|x - y\|]\,/\sqrt{d}$, it must be the case that $\langle x - z, u \rangle \geq t \cdot \|x - z\|/\sqrt{d}$. But by Lemma 4.1, the probability of this (over a random choice of $u \in S^{d-1}$) is at most $\exp(-t^2/2)$. In addition, clearly $\|y - z\| \leq \ell + \|x - z\|$ for every $y \in C$. □

Corollary 4.4. *If a subset $S \subseteq \mathbb{R}^d$ is (σ, δ, ℓ)-covered by C, then for every $\epsilon, t \geq 0$, the neighborhood $N_\epsilon(S) = \{z \in \mathbb{R}^d : \mathsf{dist}_2(x, S) \leq \epsilon\}$ is $(\sigma - \epsilon t, \delta - \exp(-t^2/2), \ell + \epsilon)$-covered by C.*

Our final lemma concerns the tradeoff between the different parameters of a cover. In particular, by paying slightly in σ (the length of the projection), we can boost δ very close to 1, as long as ℓ is small enough. The lemma follows immediately from Levy's isoperimetric inequality for S^{d-1}. Our lemma is just a restatement of Lemma 9 in [ARV04], so we omit its proof.

Lemma 4.5. *Suppose that x is (σ, δ, ℓ)-covered by C, then for any $\gamma > \sqrt{2\log(2/\delta)} + t$, x is also $(\sigma - 2\ell\gamma, 1 - \exp(-t^2/2), \ell)$-covered by C.*

Matching covers and cores. For a finite set X, let $\mathcal{M}(X)$ denote the set of partial matchings on X (i.e. the set of partial matchings in the complete graph on X). Given a subset $Y \subseteq X$, we say that Y is (σ, δ, ℓ)-*matching covered* by X if there exists a map

$$M : S^{d-1} \to \mathcal{M}(X)$$

such that the following conditions hold.

1. For every $u \in S^{d-1}$ and $(x, y) \in M(u)$, we have $\langle x - y, u \rangle \geq \frac{\sigma}{\sqrt{d}}$, and $\|x - y\| \leq \ell$.
2. For every $y \in Y$,
$$\Pr_{u \in S^{d-1}} [\exists\, x \in X : (x, y) \in M(u)] \geq \delta.$$

We refer to M as the *matching cover* of Y (where the set X here is implicit). If Y is (σ, δ, ℓ)-matching covered by itself, then we call Y a *(σ, δ, ℓ)-core.*

In the next section, we will use a modification of the [ARV04] chaining argument to give optimal lower bounds on the size of a core in terms of the structural properties of the set Y. First, we show how to do one step of the "chaining." This is where our approach departs from that of [ARV04]. From now on, all covers, matching covers, and cores are assumed to be finite subsets of $\mathbb{R}^d$.

For subsets $S \subseteq Y \subseteq \mathbb{R}^d$, define

$$\Gamma_Y(S, r) = \{y \in Y : \mathsf{dist}_2(y, S) \leq r\}.$$

Additionally, for $k \in \mathbb{N}$, define $\Gamma_Y^k(S, r)$ inductively by

$$\Gamma_Y^k(S, r) = \Gamma_Y(\Gamma_Y^{k-1}(S, r), r),$$

with $\Gamma_Y^1(S, r) = \Gamma_Y(S, r)$. Note that $\Gamma_Y^k(S, r)$ is *not* necessarily the $(k \cdot r)$-neighborhood of S in Y. This will become very important in the next section.

Proposition 4.6 (One step)**.** *Suppose that $\mathcal{C} \subseteq \mathbb{R}^d$ is a $(\sigma_0, \delta_0, \ell_0)$-core. Additionally, suppose that $S \subseteq \mathcal{C}$ is $(\sigma, 1 - \frac{\delta_0}{2}, \ell)$-covered by $\mathcal{C}$. Let $\beta = |S|/|\Gamma_{\mathcal{C}}(S, \ell_0)|$. Then there exists a subset $S' \subseteq \Gamma_{\mathcal{C}}(S, \ell_0)$ with the following properties.*

1. $|S'| \geq \frac{\delta_0}{4}|S|$.
2. *S' is $\left(\sigma + \sigma_0, \frac{\delta_0 \beta}{4}, \ell + \ell_0\right)$-covered by $\mathcal{C}$.*

Proof. Let $M : S^{d-1} \to \mathcal{M}(\mathcal{C})$ be the matching cover of $\mathcal{C}$ by itself. Consider a point $x \in S$. Since S is $(\sigma, 1 - \frac{\delta_0}{2}, \ell)$-covered by $\mathcal{C}$, for a $1 - \frac{\delta_0}{2}$ fraction of directions $u \in S^{d-1}$, there exists some $y_u \in B_{\mathcal{C}}(x, \ell)$ such that $\langle x - y_u, u \rangle \geq \frac{\sigma}{\sqrt{d}}$. In addition (since $\mathcal{C}$ is a core), for a δ_0 fraction of $u \in S^{d-1}$, there exists a point z_u such that $(z_u, x) \in M(u)$, which implies that $\langle z_u - x, u \rangle \geq \frac{\sigma_0}{\sqrt{d}}$ and $z \in B_{\mathcal{C}}(x, \ell_0)$ (in particular, $z \in \Gamma_{\mathcal{C}}(S, \ell_0)$).

By a trivial intersection bound, for a $\frac{\delta_0}{2}$ fraction of $u \in S^{d-1}$, both events happen simultaneously, and we have $\langle z_u - y_u, u \rangle \geq \frac{\sigma + \sigma_0}{\sqrt{d}}$. In this case, we define $A(z_u, u) = y_u$. Observe that this is well-defined; since $M(u)$ is a matching, $A(z_u, u)$ is assigned at most once. Doing this for every $x \in S, u \in S^{d-1}$ defines a partial assignment $A : \mathcal{C} \times S^{d-1} \to \mathcal{C}$.

Define a measure μ_A on $\mathcal{C}$ by

$$\mu_A(z) = \Pr_{u \in S^{d-1}} [A(z, u) \text{ is defined}] .$$

First, we have $\mu_A(\mathcal{C}) \geq \frac{\delta_0}{2} |S|$ by construction. Secondly, we have $\mu_A(z) > 0$ only if $z \in \Gamma_{\mathcal{C}}(S, \ell_0)$, and trivially $\mu_A(z) \leq 1$ for every $z \in \mathcal{C}$. Define

$$S' = \left\{ z \in \mathcal{C} : \mu_A(z) \geq \frac{\delta_0 \beta}{4} \right\} ,$$

and observe that

$$\frac{\delta_0}{2} |S| = \mu_A(\mathcal{C}) \leq |\Gamma_{\mathcal{C}}(S, \ell_0)| \cdot \frac{\delta_0 \beta}{4} + |S'| = \frac{\delta_0}{4} |S| + |S'|.$$

We conclude that $|S'| \geq \frac{\delta_0}{4} |S|$. Additionally, every $z \in \mathcal{C}$ is $(\sigma + \sigma_0, \mu_A(z), \ell + \ell_0)$-covered by the set $\{A(z, u) : A(z, u) \text{ is defined}\}$, so S' is itself $(\sigma + \sigma_0, \frac{\delta_0 \beta}{4}, \ell + \ell_0)$-covered by $\mathcal{C}$. □

4.2 NEG metrics, chaining, and the size of a core

In this section, we seek to lower bound the size of a core $\mathcal{C} \subseteq \mathbb{R}^d$ in terms of its structural properties. First of all, if $\mathcal{C}$ is a (σ, δ, ℓ)-core, then one can apply the (trivial) bound of Lemma 4.2 to conclude that (for σ, δ fixed constants), $|\mathcal{C}| \geq \exp(\Omega(1/\ell^2))$. And unless one knows more about the structure of $\mathcal{C}$, this bound is asymptotically the best possible.

Recall the definition of an NEG metric: This is a metric space (X, d_X) for which their exists a map $f : X \to L_2$ with $d_X(x, y) = \|f(x) - f(y)\|^2$. We will study cores which are the images of such embeddings. In other words, subsets $S \subseteq \mathbb{R}^d$ for which the distance function $(x, y) \mapsto \|x - y\|^2$ is a metric on S. Such a set $S \subseteq \mathbb{R}^d$ must have a very restricted structure. For instance, it is an easy observation that every triple in S subtends an angle of at most 90 degrees. Our main theorem is that, for sets of this type, the size of a core must be significantly larger.

Theorem 4.7 (Big Core Theorem)**.** *Suppose that $\mathcal{C} \subseteq \mathbb{R}^d$ is a (σ, δ, ℓ)-core for some $\sigma, \delta \in (0, \frac{1}{2}]$. Suppose furthermore that the distance function $d_{\mathcal{C}}(x, y) = \|x - y\|^2$ is a metric on $\mathcal{C}$. Then*

$$|\mathcal{C}| \geq \exp\left(\Omega\left(\frac{\sigma^6}{\ell^4 \log^2(1/\delta)}\right)\right) .$$

We remark that for $\sigma, \delta > 0$ fixed, the authors of [ARV04] obtain a weaker core theorem: $|\mathcal{C}| \geq \exp(\Omega(1/\ell^3))$. Theorem 4.7 is optimal in terms of the asymptotic dependence on ℓ. To see this, observe that the set $\mathcal{C} = \{-1, 1\}^d$ is an $\left(\Omega(1), \Omega(1), d^{-1/4}\right)$ core. (We leave this as an exercise to the reader. It follows easily from the consequences of Theorem 4.7 given in Section 4.3.)

Proof of Theorem 4.7. Let

$$R = \left\lfloor \frac{\sigma^2}{2^{11} \cdot \ell^2 \log(8/\delta^2)} \right\rfloor .$$

We claim the following.

Claim 4.8. *There exists a subset $S_R \subseteq \mathcal{C}$ such that S_R is $(\frac{\sigma}{4}R, 1-\frac{\delta}{2}, 1)$-covered by $\mathcal{C}$.*

To see that this claim finishes the proof, observe that Lemma 4.2 implies that

$$|\mathcal{C}| \geq \exp(\Omega(\sigma R)^2) \geq \exp(\Omega(\sigma^6/[\ell^4 \log^2(1/\delta)])).$$

We move on to a proof of the claim.

Proof of Claim 4.8. The proof is by induction on r.

Inductive assumption. For $0 \leq r \leq R$, there exists a subset $S_r \subseteq \mathcal{C}$ satisfying the following conditions.

1. S_r is $(\frac{\sigma}{4}r, 1-\frac{\delta}{2}, 2\ell\sqrt{r})$-covered by $\mathcal{C}$.
2. $|S_r| \geq \left(\frac{\delta}{4}\right)^r |\mathcal{C}|$
3. $|S_r| \geq \delta |\Gamma_{\mathcal{C}}(S_r, \ell)|$.

The base case. Let $S_0 = Y$. Then since S_0 is trivially $(0, 1, 0)$-matching covered (every point covers itself in all directions), the inductive assumption is satisfied.

Now assume that S_{r-1} satisfies the inductive assumption and that $r \leq R$. The construction of S_r proceeds in three steps.

(S1) **Use the core to extend the set S_{r-1} to $S'_r \subseteq \Gamma_{\mathcal{C}}(S_{r-1}, \ell)$.**

We apply Proposition 4.6 to the set S_{r-1} and the core $\mathcal{C}$ to obtain S'_r. Observe that by property (3) of the inductive assumption, the value of β in Proposition 4.6 is at least δ. It follows that S'_r is $(\frac{\sigma}{4}(r-1)+\sigma, \frac{\delta^2}{4}, \ell')$-covered by $\mathcal{C}$ for some ℓ' (the value of which we address in step (S3)). Additionally, using property (1) of Proposition 4.6, $|S'_r| \geq \left(\frac{\delta}{4}\right)|S_{r-1}| \geq \left(\frac{\delta}{4}\right)^r |\mathcal{C}|$.

(S2) **Grow S'_r until it stops expanding.**

Let $k \geq 0$ be the first value for which $|\Gamma^k_{\mathcal{C}}(S'_r, \ell)| \geq \delta |\Gamma^{k+1}_{\mathcal{C}}(S'_r, \ell)|$. Let $S_r = \Gamma^k_{\mathcal{C}}(S'_r, \ell)$. Notice that the neighborhood condition (3) is satisfied by construction, i.e. $|S_r| \geq \delta |\Gamma_{\mathcal{C}}(S_r, \ell)|$. Condition (2) is satisfied since $S_r \supseteq S'_r$.

We claim that S_r is $(\frac{\sigma}{4}(r-1)+\frac{\sigma}{2}, \frac{\delta^2}{8}, \ell'')$ covered by $\mathcal{C}$ (for some ℓ'' addressed in (S3)). First, since we had $|S'_r| \geq \left(\frac{\delta}{4}\right)^r |\mathcal{C}|$, it follows that

$$k \leq \log_{1/\delta} \left(\frac{4}{\delta^2}\right)^r \leq 3r. \tag{4}$$

It follows that for every $a \in S_r$, there exists a $b \in S'_r$ and a sequence $a = a_0, a_1, \ldots, a_k = b$ of points in $\mathcal{C}$ such that $\|a_i - a_{i+1}\| \leq \ell$ for $i = 0, 1, \ldots, k-1$. Now use the fact that $d_{\mathcal{C}}(x, y) = \|x - y\|^2$ is a metric on $\mathcal{C}$ to conclude that

$$\|x - y\|^2 \leq \sum_{i=0}^{k-1} \|a_i - a_{i+1}\|^2 \leq 3r\ell^2, \tag{5}$$

i.e. $\|x - y\| \leq \ell\sqrt{3r}$, i.e. $S_r \subseteq N_\epsilon(S'_r)$ for $\epsilon = \ell\sqrt{3r}$. Applying Corollary 4.4 to S'_r with $t = \sigma/(2\epsilon)$, we conclude that S_r is $\left(\frac{\sigma}{4}(r-1) + \frac{\sigma}{2}, \frac{\delta^2}{4} - \exp(-t^2/2), \ell''\right)$-covered by $\mathcal{C}$. This yields our desired conclusion as long as $\exp(-t^2/2) \leq \frac{\delta^2}{8}$. This is true as long as

$$r \leq \frac{\sigma^2}{24\,\ell^2 \log(8/\delta^2)}, \tag{6}$$

which holds since $r \leq R$.

(S3) **Bounding ℓ'' and boosting the cover to $1 - \frac{\delta}{2}$.**

First we consider the size of ℓ''. Observe that in (S1), in augmenting our cover with Proposition 4.6, we go at most "one step" (along some "edge" of the matching cover) when passing from S_{r-1} to S'_r (this corresponds to the fact that in property (2) of Proposition 4.6, the set S' is covered by vectors of length at most $\ell + \ell_0$, where ℓ_0 is the length of a vector in the matching cover). Additionally, using the bound (4), we see that the *total* number of steps taken by (S2) is at most $3r$. Using a similar calculation to the one in line (5), we conclude that $\ell'' \leq 2\ell\sqrt{r}$. It follows that S_r is $\left(\frac{\sigma}{4}(r-1) + \frac{\sigma}{2}, \frac{\delta^2}{8}, 2\ell\sqrt{r}\right)$-covered by $\mathcal{C}$. Now we would like to apply Lemma 4.5 with $\gamma = \sigma/(8\ell'')$ and $t = \sqrt{2\log(2/\delta)}$ to conclude that S_r is also $\left(\frac{\sigma}{4}r, 1 - \frac{\delta}{2}, 2\ell\sqrt{r}\right)$-covered by $\mathcal{C}$. This is possible as long as

$$\gamma > \sqrt{2\log(2/\delta)} + t = 2\sqrt{2\log(2/\delta)},$$

which holds whenever

$$r < \frac{\sigma^2}{2^{11} \cdot \ell^2 \log(2/\delta)}, \tag{7}$$

which is true since $r \leq R$.

This completes the induction. □

4.3 Applications

In this section, we present some applications of Theorem 4.7.

4.3.1 Finding separated sets

We prove the following theorem from [ARV04] about separated sets. Our proof yields a very simple algorithm to find such sets. Afterward, we mention some extensions which are useful for various other approximation algorithms, and which require Theorem 4.7.

Theorem 4.9 (Separated sets, [ARV04]). *Let (X, d) be an n-point metric of negative type with $\mathrm{diam}(X) \le 1$ and $\frac{1}{n^2}\sum_{x,y\in X} d(x,y) \ge \alpha > 0$. Then there exist subsets $A, B \subseteq X$ with $|A|, |B| = \Omega(\alpha n)$ and $d(A,B) \ge 1/O(\sqrt{\log n})$, where the $O(\cdot)$ notations hides some dependence on α.*

Proof. Since (X, d) is of negative type, there exists a map $f : X \to \mathbb{R}^n$ such that $d(x,y) = \|f(x) - f(y)\|^2$. We will associate X with the image of f. Since $\mathrm{diam}(X) \le 1$, we may assume that $X \subseteq B(0,1) \subseteq \mathbb{R}^n$. Let $\sigma = \sigma(\alpha), \ell = \ell(\alpha)$ be constants to be chosen later. For $u \in S^{n-1}$, we define $L_u = \{x \in X : \langle x, u\rangle \le \frac{-\sigma}{\sqrt{n}}\}$ and $R_u = \{x \in X : \langle x, u\rangle \ge \frac{\sigma}{\sqrt{n}}\}$. Using standard estimates, with probability $p = \Omega(\alpha)$, one has $|L_u|, |R_u| \ge \frac{\alpha}{16} n$ for some choice of $\sigma = \Theta(\alpha)$ (see, e.g. [ARV04, Lemma 4]).

Define a bipartite graph $G = (L_u \times R_u, E)$ with $(x,y) \in E$ if $x \in L_u, y \in R_u$ and $\|x - y\| \le \ell$. Let $M(u)$ be a maximal matching in G, and let L'_u, R'_u be the sets L_u, R_u with the endpoints of $M(u)$ removed. Then by construction, $d(L'_u, R'_u) \ge \ell^2$ (since $\|x - y\| \ge \ell$ for all $x \in L'_u, y \in R'_u$).

Let

$$q = \Pr_{u\in S^{d-1}}\left[|M(u)| \ge \frac{\alpha}{32} n \,\middle|\, |L_u|, |R_u| \ge \frac{\alpha}{16} n\right].$$

Under this definition, with probability $p(1-q)$, we have $|L'_u|, |R'_u| \ge \frac{\alpha}{32} n$, hence we can take $A = L'_u$ and $B = R'_u$. So we are left to show that we can simultaneously choose $\ell \ge O(\log n)^{-1/4}$ and $q < 1$. In fact, we will argue that from the matchings $\{M(u) : u \in S^{d-1}\}$, we can construct a core $\mathcal{C} \subseteq X$. Applying Theorem 4.7, we will conclude that $|\mathcal{C}| \gg n$ unless ℓ is large enough and q is small enough, yielding a contradiction.

The idea is to define an (infinite) graph $H = (X \times S^{d-1}, E_H)$ with an edge $\{(x,u), (y,u)\}$ whenever $(x,y) \in M(u)$. We define the "degree" of a point $x \in X$ by

$$d_H(x) = \Pr_{u\in S^{d-1}}[\exists \text{ and edge in } E_H \text{ with endpoint } (x,u)\ ].$$

Observe that since $M(u)$ is a matching, there is at most one edge containing (x,u) as an endpoint. It follows by construction that $\sum_{x\in X} d_H(x) \ge \frac{pq\alpha}{32} n$.

We now iteratively remove any node $x \in X$ (thus removing all nodes (x,u) from H for $u \in S^{d-1}$ and all edges adjacent to these nodes) that has $d_H(x) \le \frac{pq\alpha}{64}$. Let H' be the graph remaining after all such nodes have been removed, and let $\mathcal{C}$ be the set of points not removed from H. Then $d_{H'}(x) \ge \frac{pq\alpha}{64}$ for every $x \in \mathcal{C}$, and clearly H' is non-empty since the total degree removed is at most $\frac{pq\alpha}{64} n < \frac{pq\alpha}{32} n$.

It follows that $\mathcal{C}$ is a $(2\sigma, pq\alpha/64, \ell)$-core, hence by Theorem 4.7, we have

$$|\mathcal{C}| \ge \exp\left(\Omega\left(\frac{\sigma^6}{\ell^4 \log^2(\frac{1}{pq\alpha})}\right)\right) \ge \exp\left(\Omega\left(\frac{\alpha^6}{\ell^4 \log^2(\frac{1}{q\alpha^2})}\right)\right).$$

This yields a contradiction when the latter bound exceeds n. Thus for every fixed choice of q, there exists an $\ell \ge O(\log n)^{-1/4}$ such that with probability $p(1-q)$, we have $d(L'_u, R'_u) \ge \ell^2 \ge O(\log n)^{-1/2}$, completing the proof. □

In fact, the proof of Theorem 4.9 yields more information which does not follow from [ARV04]. Variants of the following corollary are used in the analysis of SDPs for vertex cover [Kar01] and the directed sparsest cut problem [ACMM05].

Corollary 4.10. *Let $X \subseteq \mathbb{R}^n$ be an $2n$-point subset such that the distance function $d(x,y) = \|x-y\|^2$ is a metric on X. Suppose furthermore that for every $x \in X$, $\|x\| = 1$ and if $x \in X$ then also $-x \in X$ (in [ACMM05], this is referred to as a* symmetric unit-ℓ_2^2 representation*). Then there exist subsets $A, B \subseteq X$ for which $A = -B$ and $d(A,B) \geq 1/O(\sqrt{\log n})$.*

Proof. It is easy to check that since X is symmetric and all the points of X lie on the unit sphere, one has $\frac{1}{n^2}\sum_{x,y\in X} d(x,y) \geq \frac{1}{2}$, thus we are in position to apply Theorem 4.9. To ensure that $A = -B$, observe that if $x \in L_u$, then $-x \in R_u$. Since we have freedom in choosing the maximal matching $M(u)$, whenever we add (x,y) to $M(u)$, we can also add $(-y,-x)$. This ensures that $L'_u = -R'_u$ after removing the endpoints of $M(u)$. $\square$

The next corollary is useful for the Sparsest Cut problem with general demands, as shown in [CGR05]. We explain the application to embeddings in Section 4.3.3.

Corollary 4.11. *Let (X,d) be an n-point metric space of negative type with* $\mathrm{diam}(X) \leq 1$*, and let $\omega : X \times X \to \mathbb{Z}_+$ be a symmetric weight function for which $0 \leq w(a,b) \leq \mathrm{poly}(n)$ for all $a, b \in X$, and such that $w(a,b) > 0$ only if $d(a,b) \geq \alpha$. Then there exist subsets $A, B \subseteq X$ such that $\omega(A,B) \geq \Omega(1)\omega(X,X)$ and $d(A,B) \geq 1/O(\sqrt{\log n})$, where we have extended $\omega(\cdot,\cdot)$ to subsets in the obvious way, and the $\Omega(\cdot), O(\cdot)$ notations hides some dependence on α.*

Proof. The idea of [CGR05] is to define $\pi : X \to \mathbb{N}$ by $\pi(x) = \sum_{y\in X}\omega(x,y)$, then to replace every copy of $x \in X$ by $\pi(x)$ copies. Let n' be the new number of points, and observe that $n' = \mathrm{poly}(n)$ by assumption. (We allow each copy of x to participate in matchings $M(u)$ as distinct points.) It is not difficult to see that by choosing parameters appropriately in the proof of Theorem 4.7, for a random $u \in S^{d-1}$, with constant probability p we have $\omega(L_u, R_u) \geq \left(\frac{\alpha}{16}\right)^2 w(X,X)$. Furthermore, we can ensure that with probability p conditioned on this, we have $|M(u)| \leq \left(\frac{\alpha}{16}\right)\frac{n'}{4}$. (Note that in forming L'_u, R'_u from L_u, R_u, we have $x \in (L_u \setminus L'_u) \cup (R_u \setminus R'_u)$ only if *all copies of* x participate in $M(u)$.)

But then pruning out $M(u)$ from $L_u \times R_u$ can remove at most a $\frac{1}{2}\left(\frac{\alpha}{16}\right)$ fraction of the weight, hence

$$\omega(L'_u, R'_u) \geq \left[\left(\frac{\alpha}{16}\right)^2 - \frac{1}{2}\left(\frac{\alpha}{16}\right)^2\right] w(X,X) \geq \frac{1}{2}\left(\frac{\alpha}{16}\right)^2 w(X,X).$$

Furthermore, $d(L'_u, R'_u) \geq 1/O(\sqrt{\log n'}) \geq 1/O(\sqrt{\log n})$. $\square$

4.3.2 Approximating edge expansion in graphs

We now give a simple rounding algorithm for obtaining an $O(\sqrt{\log n})$-approximation for the Sparsest Cut problem with uniform demands. The same algorithm was analyzed in [ARV04] to give a guarantee of $O(\log n)^{2/3}$.

Let $G = (V,E)$ be an n-vertex graph. In the *Sparsest Cut* problem (with uniform demands), one wishes to find the cut with the smallest *edge expansion*. This is the cut $S \subseteq V$ which minimizes $|E(S,\bar{S})|/(|S|\cdot|\bar{S}|)$, where $E(S,\bar{S})$ is the set of edges crossing the cut $(S,\bar{S})$. The edge expansion of G is defined as

$$\Phi(G) = \min_{S\subseteq V}\frac{|E(S,\bar{S})|}{|S|\cdot|\bar{S}|}.$$

Algorithm ROUNDSDP

0. If for some $i \in V$, $|B(i, \frac{1}{4})| \geq \frac{n}{4}$, set $L = B(i, \frac{1}{4})$ and go to (5).
1. Let $i_0 \in V$ be the vertex for which $|B(i_0, 2)|$ is maximized.
2. Choose $u \in S^{n-1}$ uniformly at random (according to the Haar measure).
3. Let $L = \{j \in V : \langle u, v_j - v_{i_0} \rangle \leq \frac{-\sigma_0}{\sqrt{n}}\}$ and $R = \{j \in V : \langle u, v_j - v_{i_0} \rangle \geq \frac{\sigma_0}{\sqrt{n}}\}$.
4. As long as there exists $(i, j) \in L \times R$ with $d(i, j) \leq C_0/\sqrt{\log n}$,
   set $L \leftarrow L - \{i\}, R \leftarrow R - \{j\}$.
5. Sort the points $i \in V$ according to the value $d(i, L)$: $\{p_1, p_2, \ldots, p_n\} = V$.
6. Output the sparsest of the $n-1$ cuts $(\{p_1, \ldots, p_k\}, \{p_{k+1}, \ldots, p_n\})$ for $1 \leq k \leq n-1$.

Figure 1: Hyperplane rounding algorithm

Consider the following SDP. It is well known that if sdp is the optimal value of this program, then $n^2\Phi(G) \geq \mathsf{sdp}$. (see, e.g. [Goe97]).

$$
\begin{aligned}
\min \quad & \sum_{ij \in E} \|v_i - v_j\|^2 \\
\text{s.t.} \quad & v_i \in \mathbb{R}^n, \qquad \forall i \in V, \\
& \|v_i - v_j\|^2 \leq \|v_i - v_k\|^2 + \|v_k - v_j\|^2, \forall i, j, k \in V, \\
& \frac{1}{n^2} \sum_{i,j \in V} \|v_i - v_j\|^2 = 1.
\end{aligned}
$$

Here, we show that the simple rounding procedure of Figure 1 yields an $O(\sqrt{\log n})$-approximation to the Sparsest Cut. For simplicity, let $C_0, \sigma_0 > 0$ be some fixed constants, and let $d(i, j) = \|v_i - v_j\|^2$. We write $B(i, r) = \{j : d(i, j) \leq r\}$, and for $L \subseteq V$, $d(i, L) = \min_{j \in L} d(i, j)$.

Lemma 4.12. *For appropriate choices of $C_0, \sigma_0 > 0$, if line (4) is reached, then with constant probability over the choice of $u \in S^{n-1}$, $|L|, |R| = \Omega(n)$ after line (4).*

Proof. Since we have reached line (4), every ball $B(i, \frac{1}{4})$ contains less than $n/4$ points. Also, we have $|B(i_0, 2)| \geq n/2$. Otherwise,

$$n^2 = \sum_{i,j} d(i, j) = \sum_i \left(\sum_j d(i, j) \right) > n \cdot \frac{n}{2} \cdot 2$$

yielding a contradiction. It follows that the set $\mathcal{B} = B(i_0, 2)$ contains at least $n/2$ points. Furthermore, since every $i \in \mathcal{B}$ has $|B(i, \frac{1}{4})| \leq n/4$, the average distance in $\mathcal{B}$ is at least $1/8$. It now follows from applying the proof of Theorem 4.9 to $\mathcal{B}$ that, for appropriate choices of σ_0, C_0, we have $|L|, |R| = \Omega(n)$ with $\Omega(1)$ probability after the execution of line (4). (To see that we are simulating the proof algorithmically, just rescale so that $\text{diam}(\mathcal{B}) \leq 1$, and observe that the average distance in $\mathcal{B}$ is still $\Omega(1)$. Notice that the pruning of line (4) is simply removing a maximal matching $M(u)$ from L_u and R_u in the proof of Theorem 4.9.) $\square$

Now we analyze lines (5) and (6).

Lemma 4.13. *Upon reaching line (5), we have*

$$\frac{1}{n^2}\sum_{i,j}|d(i,L)-d(j,L)| \geq \frac{1}{O(\sqrt{\log n})}.$$

Proof. If line (4) was executed, then using Lemma 4.12, we see that

$$\frac{1}{n^2}\sum_{i,j}|d(i,L)-d(j,L)| \geq \frac{1}{n^2}\sum_{i\in L, j\in R}|d(i,L)-d(j,L)| \geq \frac{|L|}{n^2}\sum_{j\in R}|d(j,L)| \geq \frac{1}{O(\sqrt{\log n})}.$$

Otherwise, there exists an index i' for which $|B(i',\frac{1}{4})| \geq n/4$, and $L = B(i',\frac{1}{4})$. In this case,

$$n^2 = \sum_{i,j} d(i,j) \leq \sum_{i,j}\left(d(i,i')+d(j,i')\right) = 2n\sum_i d(i,i') \leq 2n\sum_i\left(d(i,L)+\tfrac{1}{4}\right),$$

hence $\frac{1}{n}\sum_i d(i,L) \geq \frac{1}{4}$. Therefore,

$$\frac{1}{n^2}\sum_{i,j}|d(i,L)-d(j,L)| \geq \frac{1}{n^2}\sum_{j\in L, i\notin L}|d(j,L)| \geq \frac{1}{4n}\sum_{i\notin L} d(i,L) \geq \frac{1}{16}.$$

□

To finish, let $S_k = \{p_1, p_2, \ldots, p_k\}$ (from line (6)) for each $1 \leq k \leq n-1$, and observe that for all $i,j \in V$,

$$|d(i,L)-d(j,L)| = \sum_{k=1}^{n-1}|\mathbf{1}_{S_k}(i)-\mathbf{1}_{S_k}(j)|\cdot|d(p_i,L)-d(p_{i+1},L)|,$$

therefore

$$\begin{aligned}
\min_k \frac{|E(S_k,\bar{S}_k)|}{|S_k|\cdot|\bar{S}_k|} &= \min_k \frac{\sum_{ij\in E}|\mathbf{1}_{S_k}(i)-\mathbf{1}_{S_k}(j)|}{\sum_{i,j\in V}|\mathbf{1}_{S_k}(i)-\mathbf{1}_{S_k}(j)|} \\
&\geq \frac{\sum_{k=1}^{n-1}\left(\sum_{ij\in E}|\mathbf{1}_{S_k}(i)-\mathbf{1}_{S_k}(j)|\cdot|d(p_i,L)-d(p_{i+1},L)|\right)}{\sum_{k=1}^{n-1}\left(\sum_{i,j\in V}|\mathbf{1}_{S_k}(i)-\mathbf{1}_{S_k}(j)|\cdot|d(p_i,L)-d(p_{i+1},L)|\right)} \\
&= \frac{\sum_{ij\in E}|d(i,L)-d(j,L)|}{\sum_{i,j\in V}|d(i,L)-d(j,L)|} \\
&\leq \frac{\sum_{ij\in E}d(i,j)}{n^2/O(\sqrt{\log n})} \leq O(\sqrt{\log n})\cdot\Phi(G),
\end{aligned}$$

where in the final line we employed Lemma 4.13 and $\mathsf{sdp}/n^2 \leq \Phi(G)$.

4.3.3 Embedding NEG metrics into L_1

In this section, we give a self-contained variant of the [CGR05] proof that $D_n(\mathsf{NEG}, L_2) \leq O(\log n)^{3/4}$. The following simple combinatorial lemma is implicit in [CGR05].

Lemma 4.14. *Let S be any set of size n, let $\varepsilon > 0$ be given, and let $\mathcal{F}$ be a family of subsets of S. Let $\omega : S \to \mathbb{Z}_+$ be a weight function which satisfies $1 \le \omega(v) \le n^{2/\varepsilon}$ for every $v \in S$. If for every such weight function, there exists a subset $T(\omega) \in \mathcal{F}$ such that $\omega(T) \ge \varepsilon \cdot \omega(S)$, then there exists a family of $k = O(\log n/\varepsilon)$ subsets $T_1, T_2, \ldots, T_k \in \mathcal{F}$ such that for every $v \in S$,*

$$|\{1 \le i \le k : v \in T_i\}| \ge \frac{\varepsilon}{2}k.$$

Proof. We define a family of weight functions inductively. Initially, $\omega_0(v) = 1$ for every $v \in S$. Given $\omega_{k-1} : S \to \mathbb{Z}_+$, we define $\omega_k(v) = 2\omega_{k-1}(v)$ if $v \notin T(\omega_{k-1})$ and $\omega_k(v) = \omega_{k-1}(v)$ otherwise.

First, observe that $\omega_k(S) \le [2(1-\varepsilon)]^k \omega_0(S) = [2(1-\varepsilon)]^k n$. Furthermore, if some element $v \in S$ has occurred in at most t of the sets $\{T(\omega_i) : 1 \le i \le k\}$, then $\omega_k(v) \ge 2^{k-t}$. For $t = \log_2 n$ and $k = 2\log_2 n/\varepsilon$, this is a contradiction because

$$2^{k-t} > [2(1-\varepsilon)]^k n.$$

Finally, observe that the maximum weight of an element in ω_k is at most $2^k \le n^{2/\varepsilon}$. □

We now prove the main theorem of this section.

Theorem 4.15. *Let (X, d) be an n-point metric of negative type. Then for every $\tau \ge 0$, there exists a family of $K = O(\log n)$ pairs of disjoint subsets $(A_1, B_1), \ldots, (A_K, B_K)$ with $A_i, B_i \subseteq X$ such that $d(A_i, B_i) \ge \tau/O(\sqrt{\log n})$ for every $i \in [K]$, and such if $x, y \in X$ satisfy $d(x, y) \in [\tau, 2\tau]$, then*

$$|\{1 \le i \le K : x \in A_i, y \in B_i\}| = \Omega(K).$$

Proof. First, assume that $\mathrm{diam}(X) \le \tau$. We will show how to remove this assumption at the end of the proof. Let $S = \{(x, y) : d(x, y) \le \tau\}$. For a subset $T \subseteq S$, let $T^1, T^2 \subseteq X$ be the subsets of points which occur as a first or second coordinate of some pair in T, respectively. We define $\mathcal{F} = \{T \subseteq S : d(T^1, T^2) \ge \tau/O(\sqrt{\log n})\}$. By Corollary 4.11, for every $\omega : S \to \mathbb{Z}_+$, there exists a pair $(A, B) \in \mathcal{F}$ for which $\omega(A \times B) \ge \Omega(1)\omega(S)$ (there is some abuse of notation here, but ω extends to $X \times X$ because $\omega(a, b) = 0$ unless $(a, b) \in S$, so there is no confusion). Thus by using Lemma 4.14, the claim follows. (Each set T_i from Lemma 4.14 yields two pairs $(A_i, B_i), (A'_i, B'_i)$ with $A_i = T_i^1, B_i = T_i^2$, and $A'_i = T_i^2, B'_i = T_i^1$.)

Now we address the assumption that $\mathrm{diam}(X) \le \tau$. There are two ways to handle this. One is carried out in [CGR05], and requires a redefinition of the sets L_u, R_u in the proof of Theorem 4.9. We take an alternate path here. We may assume that $X \subseteq \mathbb{R}^n$ and $d(x, y) = \|x - y\|^2$. Using Lemma 2.5, there exists a map $g : X \to \mathbb{R}^n$ such that

$$\tfrac{1}{2}\min\{\tau, \|x - y\|\} \le \|g(x) - g(y)\| \le \min\{\tau, \|x - y\|\}.$$

We would like to simply apply the above argument to the set $g(X) \subseteq \mathbb{R}^n$ with τ replaced by $\tau/2$ since we now have $\mathrm{diam}(g(X)) \le \tau$. The problem is that the map $(x, y) \mapsto \|g(x) - g(y)\|^2$ may no longer form a metric. However, for any set of points $x_1, x_2, \ldots, x_k \in g(X)$, we still have the inequality $\|x_1 - x_k\|^2 \le 2\sum_{i=1}^{k-1} \|x_i - x_{i+1}\|^2$ (where 2 is a universal constant independent of k), which is the only property needed in the proof of Theorem 4.7 (see equation 5 in that proof). Further weakenings of the assumption in Theorem 4.7 are discussed in the next section. □

If we combine the preceding theorem with the Gluing Lemma, we arrive at an embedding of NEG metrics into L_2.

Theorem 4.16 ([CGR05]). *Every n-point metric of negative type embeds in L_2 with distortion $O(\log n)^{3/4}$.*

Proof. Using the Gluing Lemma, we need only show that for any $\tau \geq 0$, there exists a 1-Lipschitz map $f_\tau : X \to L_2$ such that for every $x, y \in X$ with $d(x, y) \in [\tau, 2\tau]$, we have $\|f_\tau(x) - f_\tau(y)\| \geq \tau/O(\sqrt{\log n})$. Let $(A_1, B_1), \ldots, (A_K, B_K)$ be the subsets yielded by applying Theorem 4.15 to X and τ. Consider the embedding $f_\tau : X \to \mathbb{R}^K$ given by

$$f_\tau(x) = \tfrac{1}{\sqrt{K}} \left(d(x, A_1), d(x, A_2), \ldots, d(x, A_K) \right).$$

Clearly f_τ is 1-Lipschitz. Furthermore, for any $x, y \in X$ with $d(x, y) \in [\tau, 2\tau]$, we have

$$\|f_\tau(x) - f_\tau(y)\|^2 = \frac{1}{K} \sum_{i=1}^{K} |d(x, A_i) - d(y, A_i)|^2 \geq \frac{1}{K} \cdot \Omega(K) \cdot d(A_i, B_i)^2 \geq \frac{\tau}{O(\sqrt{\log n})}.$$

□

Acknowledgements

The author is grateful to Sanjeev Arora, Assaf Naor, and Satish Rao for useful discussions throughout the course of this work.

References

[ACMM05] A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. $O(\sqrt{\log n})$ approximation algorithms for Min UnCut, Min 2CNF Deletion, and directed cut problems. In *STOC*. ACM, 2005.

[ALN05] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the Sparsest Cut. In *37th Annual Symposium on the Theory of Computing*, pages 553–562, 2005. Full version available at `http://www.cs.berkeley.edu/~jrl`.

[AR98] Yonatan Aumann and Yuval Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301 (electronic), 1998.

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *36th Annual Symposium on the Theory of Computing*, pages 222–231, 2004.

[BL00] Yoav Benyamini and Joram Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

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

[CGR05] 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, 2005.

[CHKR06] M. Charikar, M. T. Hajiaghayi, H. Karloff, and S. B. Rao. ℓ_2^2 spreading metrics for ordering problems. To appear, SODA, 2006.

[Enf69] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105, 1969.

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

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

[GNRS99] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and l_1-embeddings of graphs. In *40th Annual Symposium on Foundations of Computer Science*, pages 399–408. IEEE Computer Soc., Los Alamitos, CA, 1999.

[Goe97] Michel X. Goemans. Semidefinite programming in combinatorial optimization. *Math. Programming*, 79(1-3, Ser. B):143–161, 1997. Lectures on mathematical programming (ismp97) (Lausanne, 1997).

[Ind01] Piotr Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd Annual Symposium on Foundations of Computer Science*, pages 10–33. IEEE Computer Society, 2001.

[Kar01] G. Karakostas. A better approximation ratio for the vertex cover problem. In *32nd International Colloquium on Automata, Languages and Programming*, volume 3580 of *Lecture Notes in Computer Science*, pages 1043–1050. Springer, 2001.

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

[KPR93] Philip N. Klein, Serge A. Plotkin, and Satish Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 682–690, 1993.

[Led96] Michel Ledoux. Isoperimetry and Gaussian analysis. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, volume 1648 of *Lecture Notes in Math.*, pages 165–294. Springer, Berlin, 1996.

[Lee06] J. R. Lee. Volume distortion for subsets of Euclidean spaces. Available at `http://www.cs.berkeley.edu/~jrl`. Submitted, 2006.

[Lin02] Nathan Linial. Finite metric-spaces—combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pages 573–586, Beijing, 2002. Higher Ed. Press.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[LN04] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geom. Funct. Anal.*, 14(4):745–747, 2004.

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

[MN04] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. To appear, *Advances in Mathematics*, 2004.

[MN06] M. Mendel and A. Naor. Metric cotype. Ann. Math. To appear, 2006.

[MP84] M. B. Marcus and G. Pisier. Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. *Acta Math.*, 152(3-4):245–301, 1984.

[NR03] Ilan Newman and Yuri Rabinovich. A lower bound on the distortion of embedding planar metrics into Euclidean space. *Discrete Comput. Geom.*, 29(1):77–81, 2003.

[NS02] Assaf Naor and Gideon Schechtman. Remarks on non linear type and Pisier's inequality. *J. Reine Angew. Math.*, 552:213–236, 2002.

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