

# Embeddings of Negative-type Metrics and An Improved Approximation to Generalized Sparsest Cut

SHUCHI CHAWLA

University of Wisconsin – Madison

and

ANUPAM GUPTA

Carnegie Mellon University

and

HARALD RÄCKE

Toyota Technological Institute

---

In this paper, we study metrics of *negative type*, which are metrics  $(V, d)$  such that  $\sqrt{d}$  is an Euclidean metric; these metrics are thus also known as “ $\ell_2$ -squared” metrics. We show how to embed  $n$ -point negative-type metrics into Euclidean space  $\ell_2$  with distortion  $D = O(\log^{3/4} n)$ . This embedding result, in turn, implies an  $O(\log^{3/4} k)$ -approximation algorithm for the Sparsest Cut problem with non-uniform demands. Another corollary we obtain is that  $n$ -point subsets of  $\ell_1$  embed into  $\ell_2$  with distortion  $O(\log^{3/4} n)$ .

Categories and Subject Descriptors: F.2.0 [Analysis of Algorithms and Problem Complexity]: General

General Terms: Algorithms, Theory

Additional Key Words and Phrases: approximation algorithm, embedding, metrics, negative-type metric, sparsest cut

---

## 1. INTRODUCTION

The area of finite metric spaces and their embeddings into “simpler” spaces lies in the intersection of the areas of mathematical analysis, computer science and discrete geometry. Over the past decade, this area has seen hectic activity, partly due to the fact that it has proved invaluable in many algorithmic applications. Many examples

---

This research was performed while the first author was a graduate student and the third author was a postdoctoral researcher at Carnegie Mellon University. The first and third authors were supported in part by NSF ITR award CCR-0122581 (The ALADDIN project), and the second author was supported by an NSF CAREER award CCF-0448095, and by an Alfred P. Sloan Fellowship.

Corresponding author’s address: Anupam Gupta, Computer Science Department, Carnegie Mellon University, Pittsburgh PA 15213.

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 2008 ACM 1529-3785/2008/0700-0001 \$5.00

can be found in the surveys by Indyk [2001] and Linal [2002], or in the chapter by Matoušek [2002].

One of the first major applications of metric embeddings in Computer Science was an  $O(\log k)$  approximation to the Sparsest Cut problem with non-uniform demands (henceforth called the *Generalized Sparsest Cut* problem) [Linal et al. 1995; Aumann and Rabani 1998]. This result was based on a fundamental theorem of Bourgain [1985] in the local theory of Banach spaces, which showed that any finite  $n$ -point metric could be embedded into  $\ell_1$  space (and indeed, into any of the  $\ell_p$  spaces) with distortion  $O(\log n)$ . The connection between these results uses the fact that the Generalized Sparsest Cut problem seeks to minimize a linear function over all cuts of the graph, which is equivalent to optimizing over all  $n$ -point  $\ell_1$  metrics. Since this problem is NP-hard, we can optimize over all  $n$ -point metrics instead, and then use an algorithmic version of Bourgain’s embedding to embed into  $\ell_1$  with only an  $O(\log n)$  loss in performance.

A natural extension of this idea is to optimize over a smaller class of metrics that contains  $\ell_1$ ; a natural candidate for this class is NEG, the class of  $n$ -point *metrics of negative type*<sup>1</sup>. These are just the metrics obtained by squaring an Euclidean metric, and hence are often called “ $\ell_2$ -squared” metrics. It is known that the following relationships hold:

$$\ell_2 \text{ metrics} \subseteq \ell_1 \text{ metrics} \subseteq \text{NEG metrics.} \quad (1)$$

Since it is possible to optimize over NEG via semidefinite programming, this gives us a semidefinite relaxation for the Generalized Sparsest Cut problem [Goemans 1997]. Now if we could prove that  $n$ -point metrics in NEG embed into  $\ell_1$  with distortion  $D$  and this embedding can be found in polynomial time, we would get a  $D$ -approximation for Sparsest Cut; while this  $D$  has been conjectured to be  $O(\sqrt{\log n})$  or even  $O(1)$ , no bounds better than the  $O(\log n)$  were known prior to this work. (See the Section 1.3 for subsequent progress towards the resolution of this conjecture.)

In a recent breakthrough, Arora, Rao, and Vazirani [2004] showed that every  $n$ -point metric in NEG has a contracting embedding into  $\ell_1$  such that the sum of the distances decreases by only  $O(\sqrt{\log n})$ . Formally, they showed that the SDP relaxation had an integrality gap of  $O(\sqrt{\log n})$  for the case of uniform demand Sparsest Cut; however, this is equivalent to the above statement by the results of Rabinovich [2003].

We extend the techniques of Arora, Rao, and Vazirani to give embeddings for  $n$ -point metrics in NEG into  $\ell_2$  with distortion  $O(\log^{3/4} n)$ . More generally, we obtain the following theorem.

**THEOREM 1.1.** *Given  $(V, d)$ , a negative-type metric, and a set of terminal-pairs  $\mathbb{D} \subseteq V \times V$  with  $|\mathbb{D}| = k$ , there is a contracting embedding  $\varphi : V \rightarrow \ell_2$  such that for all pairs  $(x, y) \in \mathbb{D}$ ,*

$$\|\varphi(x) - \varphi(y)\|_2 \geq \frac{1}{O(\log^{3/4} k)} d(x, y).$$

<sup>1</sup>Note that NEG usually refers to all *distances* of negative-type, even those that do not obey the triangle inequality. In this paper, we will use NEG only to refer to negative-type *metrics*.

Note that the above theorem requires the embedding to be contracting for all node pairs, but the resulting contraction needs to be small only for the terminal pairs. In particular, when  $\mathbb{D} = V \times V$ , the embedding is an  $O(\log^{3/4} n)$ -distortion embedding into  $\ell_2$ . Though we also give a randomized polynomial-time algorithm to find this embedding, let us point out that optimal embeddings into  $\ell_2$  can be found using semidefinite programming [Linial et al. 1995, Thm. 3.2(2)].

Finally, let us note some simple corollaries.

**THEOREM 1.2.** *Every  $n$ -point metric in NEG embeds into  $\ell_1$  with  $O(\log^{3/4} n)$  distortion, and every  $n$ -point metric in  $\ell_1$  embeds into Euclidean space  $\ell_2$  with  $O(\log^{3/4} n)$  distortion. These embeddings can be found in polynomial time.*

The existence of both embeddings follows immediately from (1). To find the map  $\text{NEG} \rightarrow \ell_1$  in polynomial time, we can use the fact that every finite  $\ell_2$  metric can be embedded into  $\ell_1$  isometrically; if we so prefer, we can find a distortion- $\sqrt{3}$  embedding into  $\ell_1$  in deterministic polynomial time using families of 4-wise independent random variables [Linial et al. 1995, Lemma 3.3].

**THEOREM 1.3.** *There is a randomized polynomial-time  $O(\log^{3/4} k)$ -approximation algorithm for the Sparsest Cut problem with non-uniform demands.*

Theorem 1.3 thus extends the results of Arora et al. [2004] to the case of non-uniform demands, albeit it proves a weaker result than the  $O(\sqrt{\log k})$  approximation that they achieve for uniform demands.

The proof of Theorem 1.3 follows from the fact that the existence of distortion- $D$  embeddings of negative-type metrics into  $\ell_1$  implies an integrality gap of at most  $D$  for the semidefinite programming relaxation of the Sparsest Cut problem. Furthermore the embedding can be used to find such a cut as well. (For more details about this connection of embeddings to the Sparsest Cut problem, see the survey by Shmoys [1997, Sec. 5.3]; the semidefinite programming relaxation can be found in the survey by Goemans [1997, Sec. 6]).

## 1.1 Our Techniques

The proof of the Main Theorem 1.1 proceeds thus: we first classify the terminal pairs in  $\mathbb{D}$  by distance scales. We define the scale- $i$  set  $\mathbb{D}_i$  to be the set of all pairs  $(x, y) \in \mathbb{D}$  with  $d(x, y) \approx 2^i$ . For each scale  $i$ , we find a partition of  $V$  into components such that for a constant fraction of the terminal pairs  $(x, y) \in \mathbb{D}_i$ , the following two “good” events happens: (1)  $x$  and  $y$  lie in different components of the partition, and (2) the distance from  $x$  to any component other than its own is at least  $\eta 2^i$ , and the same for  $y$ . Here  $\eta = 1/O(\sqrt{\log k})$ . Informally, both  $x$  and  $y$  lie deep within their *distinct* components, and this happens for a constant fraction of the pairs  $(x, y) \in \mathbb{D}_i$ . This partition defines a contracting embedding of the points into a one-dimensional  $\ell_1$  metric (a line) such that every pair  $(x, y)$ , for which the above “good” events happen, has low distortion. (The details of this process are given in Section 3; the proofs use ideas from the paper by Arora, Rao, and Vazirani [2004] and the subsequent improvements by Lee [2005].)

Note that the good event happens for only a constant fraction of the pairs in  $\mathbb{D}_i$ , and we have little control over which of the pairs will be the lucky ones. However, to obtain low distortion for every terminal pair, we want a partitioning

scheme that separates a random constant fraction of the pairs in  $\mathbb{D}_i$ . To this end, we employ a simple reweighting scheme (reminiscent of the Weighted Majority algorithm [Littlestone and Warmuth 1994] and many other applications). We just duplicate each unlucky pair and repeat the above process  $O(\log k)$  times. Since each pair that is unlucky gets a higher weight in the subsequent runs, a simple argument given in Section 4 shows that each pair in  $\mathbb{D}_i$  will be separated in at least  $\log k$  of these  $O(\log k)$  partitions. (Picking one of these partitions uniformly at random would now ensure that each pair is separated with constant probability.)

We therefore obtain a good partition for each distance scale individually. We could now use these  $O(\log k)$  partitions naïvely, by concatenating the corresponding “line-embeddings”, to construct an embedding where the contraction for the pairs in  $\mathbb{D}$  would be bounded by  $\sqrt{\log k}/\eta = O(\log k)$ . However, this would be no better than the previous bounds, and hence we have to be more careful. We slightly adapt the *measured descent* embeddings of Krauthgamer et al. [2005] to combine the  $O(\log k)$  partitions for the various distance scales to get a distortion- $O(\sqrt{\log k}/\eta) = O(\log^{3/4} k)$  embedding. The details of the embedding are given in Section 5.

## 1.2 Related Work

This work adopts and adapts techniques of Arora, Rao and Vazirani [2004], who gave an  $O(\sqrt{\log n})$ -approximation for the uniform demand case of Sparsest Cut. In fact, using their results about the behavior of projections of negative-type metrics almost as a black-box, we obtain an  $O(\log^{5/6} k)$ -approximation for Generalized Sparsest Cut. Our approximation factor is further improved to  $O(\log^{3/4} k)$  by the results of Lee [2005] showing that the hyperplane separator algorithm of Arora et al. [2004, Section 3] itself gives an  $O(\sqrt{\log n})$ -approximation for the uniform demand case.

As mentioned above, there has been a large body of work on low-distortion embeddings of finite metrics; see, e.g., [Bartal 1998; Bourgain 1985; Chekuri et al. 2006; Fakcharoenphol et al. 2003; Gupta et al. 2003; Gupta et al. 2004; Krauthgamer et al. 2005; Linial et al. 1995; Matoušek 1996; Matoušek 1999; Rao 1999], and our work stems in spirit from many of these papers. However, it draws most directly on the technique of *measured descent* developed by Krauthgamer et al. [2005].

Independently of our work, Lee [2005] has used so-called “scale-based” embeddings to give low-distortion embeddings from  $\ell_p$  ( $1 < p < 2$ ) into  $\ell_2$ . The paper gives a “Gluing Lemma” of the following form: if for every distance scale  $i$ , we are given a contracting embedding  $\phi_i$  such that each pair  $x, y$  with  $d(x, y) \in [2^i, 2^{i+1})$  has  $\|\phi_i(x) - \phi_i(y)\| \geq \frac{d(x, y)}{K}$ , one can glue them together to get an embedding  $\phi : \mathbb{D} \rightarrow \ell_2$  with distortion  $O(\sqrt{K \log n})$ . His result is a generalization of [Krauthgamer et al. 2005], and of our Lemma 5.2; using this gluing lemma, one can derive an  $\ell_2$  embedding from the decomposition bundles of Theorem 4.5 without using any of the ideas in Section 5.

## 1.3 Subsequent Work

Following the initial publication of this work, Arora, Lee, and Naor [2005] built upon our techniques to obtain an embedding from negative-type metrics into  $\ell_2$  with distortion  $O(\sqrt{\log n} \log \log n)$ , implying an approximation to Generalized Sparsest

Cut with the same factor of approximation. Their improvement lies in a stronger gluing lemma. This result is essentially tight, as it is known that embedding negative-type metrics into  $\ell_2$  requires  $\Omega(\sqrt{\log n})$  distortion in the worst case [Enflo 1969].

This improvement was coupled with considerable progress on lower bounds for embeddability into  $\ell_1$ —Khot and Vishnoi [2005] showed the existence of a negative-type metric that must incur a distortion of  $\Omega(\log \log n)^{1/6-\epsilon}$  when embedded into  $\ell_1$ . This was subsequently improved to a lower bound of  $\Omega(\log \log n)$  on the distortion by Krauthgamer and Rabani [2006]. In related work, Chawla et al. [2006] showed that it is NP-hard to approximate the Sparsest Cut problem within a factor of  $o(\sqrt{\log \log n})$ , assuming an appropriate version of the Unique Games Conjecture of Khot [2002]. (Khot and Vishnoi [2005] independently show a weaker hardness of  $\Omega(\log \log n)^{1/6-\epsilon}$ , under the same assumption.)

## 2. NOTATION AND DEFINITIONS

### 2.1 Sparsest Cut

In the *Generalized Sparsest Cut* problem, we are given an undirected graph  $G = (V, E)$  with edge capacities  $c_e$ , and  $k$  source-sink (terminal) pairs  $\{s_i, t_i\}$  with each pair having an associated demand  $D_i$ . For any subset  $S \subseteq V$  of the nodes of the graph, let  $D(S, \bar{S})$  be the net demand going from the terminals in  $S$  to those outside  $S$ , and  $C(S, \bar{S})$  the total capacity of edges exiting  $S$ . Now the *generalized sparsest cut* is defined as follows (if there is unit demand between all pairs of vertices, then the problem is just called the Sparsest Cut problem).

$$\Phi = \min_{\substack{S \subseteq V \\ D(S, \bar{S}) \neq 0}} \frac{C(S, \bar{S})}{D(S, \bar{S})} = \min_{\text{cut metric } \delta_S} \frac{\sum_{(u,v) \in E} c_{uv} \delta_S(u, v)}{\sum_{ij} D_i \delta_S(s_i, t_i)} = \min_{d \in \ell_1} \frac{\sum_{(u,v) \in E} c_{uv} d(u, v)}{\sum_{ij} D_i d(s_i, t_i)}$$

Here a cut metric  $\delta_S$  is defined as  $\delta_S(x, y) = 1$  if exactly one of  $x$  and  $y$  is in the set  $S$ , and 0 otherwise. So the second equality just follows from the definition. The third equality is less trivial; see, e.g., [Aumann and Rabani 1998] for a proof.

This problem is NP-hard [Shahrokhi and Matula 1990], as is optimizing linear functions over the cone of  $\ell_1$ -metrics [Karzanov 1985]. There is much work on the sparsest cut problem (see, e.g., [Leighton and Rao 1988; Shmoys 1997]), and  $O(\log k)$  approximations were previously known [Linial et al. 1995; Aumann and Rabani 1998]. These algorithms proceeded by relaxing the problem and optimizing over *all* metrics instead of over  $\ell_1$ -metrics in the above equation, and then rounding the resulting fractional solution. A potentially stronger relaxation is obtained by optimizing only over metrics  $d \in \text{NEG}$  instead of over all metrics:

$$\Phi_{\text{NEG}} = \min_{d \in \text{NEG}} \frac{\sum_{(u,v) \in E} c_{uv} d(u, v)}{\sum_i D_i d(s_i, t_i)}$$

This quantity is the value of the semidefinite relaxation of the problem, and can be approximated well in polynomial time (see, e.g., [Goemans 1997]). Since  $\ell_1 \subseteq \text{NEG}$ , it follows that  $\Phi_{\text{NEG}} \leq \Phi$ . On the other hand, if we can embed  $n$ -point metrics in NEG into  $\ell_1$  with distortion at most  $D$  in polynomial time, we can obtain a solution of value at most  $D \times \Phi_{\text{NEG}}$ . It follows that  $\Phi \leq D \times \Phi_{\text{NEG}}$ , and the solution is a  $D$  approximation to Generalized Sparsest Cut.

## 2.2 Metrics.

The input to our embedding procedure is a *negative-type metric*  $(V, d)$  with  $|V| = n$ . We can, and indeed will, use the following standard correspondence between finite metrics and graphs: we set  $V$  to the node set of the graph  $G = (V, E = V \times V)$ , where the length of an edge  $(x, y)$  is set to  $d(x, y)$ . This correspondence allows us to perform operations like deleting edges to partition the graph. By scaling, we can assume that the smallest distance in  $(V, d)$  is 1, and the maximum distance is some value  $\Delta(d)$ , the *diameter* of the graph.

It is well-known that any *negative-type* distance space admits a geometric representation as the *square of a Euclidean metric*; i.e., there is a map  $\psi : V \rightarrow \mathbb{R}^n$  such that  $\|\psi(x) - \psi(y)\|_2^2 = d(x, y)$  for every  $x, y \in V$  [Deza and Laurent 1997, Thm. 6.2.2]. Furthermore, the fact that  $d$  is a metric implies that the angle subtended by any two points at a third point is non-obtuse. Since this map can be found in polynomial time using semidefinite programming, we will assume that we are also given such a map  $\psi$ . For any node  $x \in V$ , we use  $\vec{x}$  to denote the point  $\psi(x) \in \mathbb{R}^n$ .

## 2.3 Terminal Pairs.

We are also given a set of *terminal pairs*  $\mathbb{D} \subseteq V \times V$ ; these are the pairs of nodes for which we need to ensure a small contraction. In the sequel, we will assume that each node in  $V$  takes part in *at most one* terminal-pair in  $\mathbb{D}$ . This is without loss of generality: if a node  $x$  belongs to several terminal pairs, we add new vertices  $x_i$  to the graph at distance 0 from  $x$ , and replace  $x$  in the  $i$ -th terminal pair with  $x_i$ . (Since this transformation adds at most  $O(|\mathbb{D}|)$  nodes, it does not asymptotically affect our results.) Note that a result of this is that  $\mathbb{D}$  may have two terminal pairs  $(x, y)$  and  $(x', y')$  such that  $d(x, x') = d(y, y') = 0$ .

A node  $x \in V$  is a *terminal* if there is a (unique)  $y$  such that  $(x, y) \in \mathbb{D}$ ; call this node  $y$  the *partner* of  $x$ . Define  $\mathbb{D}_i$  to be the set of node-pairs whose distance according to  $d$  is approximately  $2^i$ .

$$\mathbb{D}_i = \{(x, y) \in \mathbb{D} \mid 2^i \leq d(x, y) < 2^{i+1}\} \quad (2)$$

We use the phrase *scale- $i$*  to denote the distances in the interval  $[2^i, 2^{i+1})$ , and hence  $\mathbb{D}_i$  is merely the set of terminal pairs that are at distance *scale- $i$* . If  $(x, y) \in \mathbb{D}_i$ , then  $x$  and  $y$  are called *scale- $i$  terminals*. Let  $\mathcal{D}$  be the set of all terminal nodes, and  $\mathcal{D}_i$  be the set of *scale- $i$  terminals*.

The *radius  $r$  ball* around  $x \in V$  is naturally defined to be  $\mathbf{B}(x, r) = \{z \in V \mid d(x, z) \leq r\}$ . Given a set  $S \subseteq V$ , the ball  $\mathbf{B}(S, r) = \cup_{x \in S} \mathbf{B}(x, r)$ .

## 2.4 Metric Decompositions: Suites and Bundles

Much of the paper will deal with finding decompositions of metrics (and of the underlying graph) with specific properties; let us define these here. Given a distance scale  $i$  and a partition  $P_i$  of the graph, let  $C_i(v)$  denote the component containing a vertex  $v \in V$ . We say that a pair  $(x, y) \in \mathbb{D}_i$  is  *$\delta$ -separated* by the partition  $P_i$  if

- the vertices  $x$  and  $y$  lie in different components; i.e.,  $C_i(x) \neq C_i(y)$ , and
- both*  $x$  and  $y$  are “far from the boundary of their components”, i.e.,  $d(x, V \setminus C_i(x)) \geq \delta d(x, y)$  and  $d(y, V \setminus C_i(y)) \geq \delta d(x, y)$ .

A *decomposition suite*  $\Pi$  is a collection  $\{P_i\}$  of partitions, one for each distance scale  $i$  between 1 and  $\lceil \log \Delta(d) \rceil$ . Given a *separation function*  $\delta(x, y) : V \times V \rightarrow [0, 1]$ , the decomposition suite  $\Pi$  is said to  $\delta(x, y)$ -*separate*  $(x, y) \in \mathbb{D}$  if for the distance scale  $i$  such that  $(x, y) \in \mathbb{D}_i$ ,  $(x, y)$  is  $\delta(x, y)$ -separated by the corresponding partition  $P_i \in \Pi$ .

Finally, a  $\delta(x, y)$ -*decomposition bundle* is a collection  $\{\Pi_j\}$  of decomposition suites such that for each  $(x, y) \in \mathbb{D}$ , at least a constant fraction of the  $\Pi_j$   $\delta(x, y)$ -separate<sup>2</sup> the pair  $(x, y)$ .

In Section 3, we show how to create a decomposition suite that  $\Omega(1/\sqrt{\log k})$ -separates a constant fraction of the pairs  $(x, y) \in \mathbb{D}_i$ , for all distance scales  $i$ . Using this procedure and a simple reweighting argument, we construct a  $\Omega(1/\sqrt{\log k})$ -decomposition bundle with  $O(\log k)$  suites. Finally, in Section 5, we show how decomposition bundles give us embeddings of the metric  $d$  into  $\ell_2$ .

### 3. CREATING DECOMPOSITION SUITES

In this section, we will give the procedure **Project-&-Prune** that takes a distance scale  $i$ , and constructs a partition  $P_i$  of  $V$  that  $\eta$ -separates at least a constant fraction of the pairs in  $\mathbb{D}_i$ . Here we use  $\eta = \frac{1}{4c\sqrt{\log k}}$ , where  $c$  is a constant to be defined later; let us also define  $f = \frac{1}{4\eta} = c\sqrt{\log k}$ .

**Procedure Project-&-Prune:**

**Input:** The metric  $(V, d)$ , its graph representation  $G$ , its geometric representation where  $x \in V$  is mapped to  $\vec{x} \in \mathbb{R}^n$ , and, a distance scale  $i$ . We assume that terminal pairs  $(x, y) \in \mathbb{D}_i$  are disjoint.

- (1) **Project.** In this step, we pick a random direction and project the points in  $V$  on the line in this direction. Formally, we pick a random unit vector  $u$ . Let  $p_x = \sqrt{n} \langle \vec{x}, u \rangle$  be the *normalized projection* of the point  $\vec{x}$  on  $u$ .
- (2) **Bucket.** Let  $\ell = 2^{i/2}$ , and set  $\beta = \ell/6$ . Informally, we will form *buckets* by dividing the line into intervals of length  $\beta$ . We then group the terminals in  $\mathcal{D}_i$  according to which interval (mod 4) they lie in. (See Figure 1.) Formally, for each  $a = 0, 1, 2, 3$ , define

$$A_a = \{ x \in \mathcal{D}_i \mid p_x \in \cup_{m \in \mathbb{Z}} ((4m + a)\beta, (4m + 1 + a)\beta) \}$$

A terminal pair  $(x, y) \in \mathbb{D}_i$  is *split* by  $A_a$  if  $x \in A_a$  and  $y \in A_{(a+2) \bmod 4}$ . If the pair  $(x, y)$  is not split by any  $A_a$ , we remove both  $x$  and  $y$  from the sets  $A_a$ . For  $a \in \{0, 1\}$ , let  $\mathbb{B}_a \subseteq \mathbb{D}_i$  be the set of terminal pairs split by  $A_a$  or  $A_{a+2}$ .

- (3) **Prune.** If there exist terminals  $x \in A_a$  and  $y \in A_{(a+2) \bmod 4}$  for some  $a \in \{0, 1\}$  (not necessarily belonging to the same terminal pair) with  $d(x, y) < \ell^2/f$ , we remove  $x$  and  $y$  and their partners from the sets  $\{A_a\}$ . We repeat until no such pairs remain.
- (4) **Cleanup.** For each  $a$ , if  $(x, y) \in \mathbb{B}_a$  and the above pruning step has removed either of  $x$  or  $y$  (recall that we remove the other one as well), then we remove

<sup>2</sup>Although for the decomposition bundle that we construct, the value of  $\delta(x, y)$  is independent of the points  $x$  and  $y$ , in constructing an embedding using the decomposition bundle we give a more general theorem that works also when  $\delta(x, y)$  is a function of  $x$  and  $y$ .

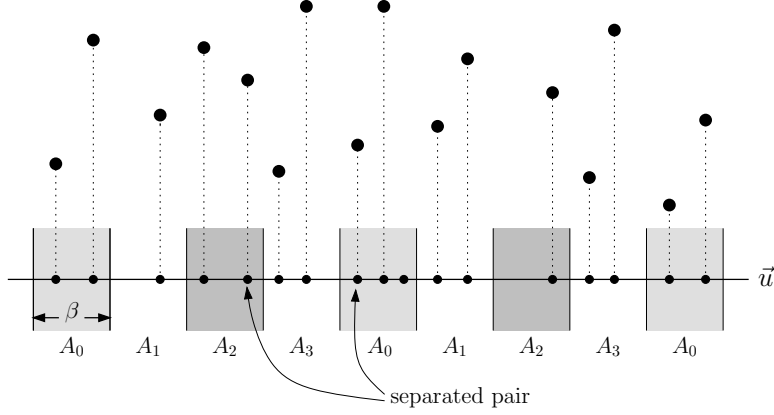


Fig. 1. Projection and Bucketing

$(x, y)$  from  $\mathbb{B}_a$ . Once this is done,  $\mathbb{B}_a = \mathbb{D}_i \cap (A_a \times A_{(a+2) \bmod 4})$  is once again the set of terminal pairs split by  $A_a$  or  $A_{a+2}$ .

- (5) If  $\max\{|\mathbb{B}_0|, |\mathbb{B}_1|\} \leq \frac{1}{64} |\mathbb{D}_i|$ , go back to Step 1, else go to Step 6.
- (6) Say the set  $\mathbb{B}_a$  has more pairs than  $\mathbb{B}_{(1-a) \bmod 2}$ . Define the partition  $P_i$  by deleting all the edges in  $G$  at distance  $\ell^2/2f$  from the set  $A_a$ . More formally, let  $C = \mathbf{B}(A_a, \ell^2/2f)$ , and define the partition  $P_i$  to be  $G[C]$  and  $G[V \setminus C]$ , the components induced by  $C$  and  $V \setminus C$ .

Note the procedure above ensures that for *any pair of terminals*  $(x, y) \in A_a \times A_{(a+2) \bmod 4}$ , the distance  $d(x, y)$  is at least  $\ell^2/f = 2^i/f$ , even if  $(x, y) \notin \mathbb{D}_i$ . Why do we care about these pairs? It is because the separation of  $\ell^2/f$  between the sets  $A_a$  and  $A_{(a+2) \bmod 4}$  ensures that the balls of radius  $\frac{\ell^2}{2f}$  around these sets are disjoint.

This in turn implies that terminal pairs  $(x, y) \in \mathbb{D}_i \cap (A_a \times A_{(a+2) \bmod 4})$  are  $\eta$ -separated upon deleting the edges in Step 6 from the graph  $G$ . Indeed, for such a pair  $(x, y)$ , the components  $C_i(x)$  and  $C_i(y)$ , obtained upon deleting the edges at distance  $\frac{\ell^2}{2f}$  from the set  $A_a$ , are distinct, and both  $d(x, V \setminus C_i(x))$  and  $d(y, V \setminus C_i(y))$  are at least  $\frac{\ell^2}{2f} \geq \frac{d(x, y)}{4f}$ .

The following theorem now shows that the procedure Project-&-Prune terminates quickly.

**THEOREM 3.1.** *For any distance scale  $i$ , the procedure Project-&-Prune terminates in a constant number of iterations. This gives us a random polynomial-time algorithm that outputs a partition  $P_i$  which  $\eta$ -separates at least  $\frac{1}{64} |\mathbb{D}_i|$  pairs of  $\mathbb{D}_i$ .*

The proof of this theorem has two parts, which we will prove in the next two subsections. We first show that the sets  $\mathbb{B}_0$  and  $\mathbb{B}_1$  contain most of  $\mathbb{D}_i$  before the pruning step (with a high probability over the random direction  $u$ ). We then show that the pruning procedure removes only a constant fraction of the pairs from these sets  $\mathbb{B}_0$  and  $\mathbb{B}_1$  with a constant probability. In fact, the size of  $\mathbb{B}_0 \cup \mathbb{B}_1$  remains at least  $|\mathbb{D}_i|/32$  even after the pruning, and then it follows that the larger of these sets

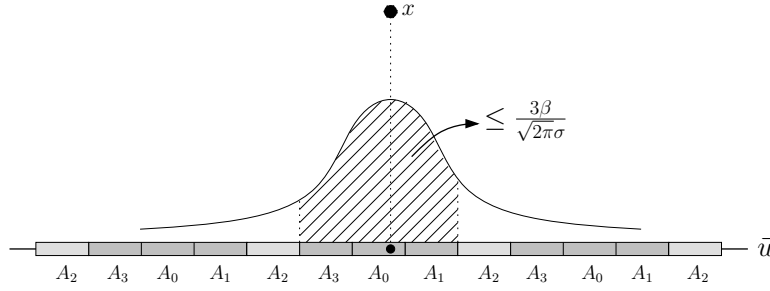


Fig. 2. The distribution of projected edge lengths in the proof of Lemma 3.2. If  $y$  falls into a light-shaded interval, the pair  $(x, y)$  is split.

must have half of the terminal pairs, proving the theorem.

### 3.1 The Projection Step

LEMMA 3.2. *Fix a distance scale  $i$ . At the end of the bucketing stage, the set  $\mathbb{B}_0 \cup \mathbb{B}_1$  contains at least  $\frac{1}{16}|\mathbb{D}_i|$  terminal pairs w.p.  $\frac{1}{15}$ .*

PROOF. Recall that a terminal pair  $(x, y) \in \mathbb{D}_i$  is split if  $x$  lies in the set  $A_a$  and  $y$  lies in  $A_{(a+2) \bmod 4}$  for some  $a \in \{0, 1, 2, 3\}$ . Also, we defined  $\ell^2 = 2^i$ , and hence  $(x, y) \in \mathbb{D}_i$  implies that  $\|\vec{x} - \vec{y}\|^2 = d(x, y) \in [\ell^2, 2\ell^2)$ . Consider the normalized projections  $p_x$  and  $p_y$  of the vectors  $\vec{x}, \vec{y} \in \mathbb{R}^n$  on the random direction  $u$ , and note that  $p_y - p_x$  is distributed (nearly) as a Gaussian random variable  $Z_u \sim N(0, \sigma^2)$  with a standard deviation  $\sigma \in [\ell, \sqrt{2}\ell)$  (see Figure 2.)

Now consider the bucket of width  $\beta$  in which  $p_x$  lies. The pair  $(x, y)$  will not be separated if  $p_y$  lies in either the same bucket, or in either of the adjoining buckets. (The probability of each of these three events is at most  $\frac{1}{\sqrt{2\pi}\sigma} \times \beta$ .) Also, at least  $\frac{1}{4}$  of the remainder of the distribution causes  $(x, y)$  to be split, since each good interval is followed by three bad intervals with less measure.

Putting this together gives us that the probability of  $(x, y)$  being split is at least

$$\frac{1}{4} \left( 1 - 3\beta \frac{1}{\sqrt{2\pi}\sigma} \right) \geq \frac{1}{4} \left( 1 - \frac{3(\ell/6)}{\sqrt{2\pi}\ell} \right) \geq \frac{1}{8}$$

Since each pair  $(x, y) \in \mathbb{D}_i$  is separated with probability  $1/8$ , the linearity of expectations and inverse Markov's inequality implies that at least one-sixteenth of  $\mathbb{D}_i$  must be split at the end of the bucketing stage with probability  $\frac{1}{15}$ .  $\square$

### 3.2 The Pruning Step

We now show that a constant fraction of the terminal pairs in  $\mathbb{D}_i$  also survive the pruning phase. This is proved by contradiction, and follows the argument of Arora et al. [2004].

Assume that, with a large probability (over the choice of the random direction  $u$ ), a large fraction of the terminal pairs in  $\mathbb{D}_i$  (say  $\frac{63}{64}|\mathbb{D}_i|$ ) get removed in the pruning phase. By the definition of the pruning step, the projection of  $\vec{x} - \vec{y}$  on  $u$  must have been large for such a removed pair  $(x, y)$ . (Recall again that the removed pair  $(x, y)$  is not necessarily a terminal pair.) In our algorithm, this happens when

$d(x, y) < \ell^2/f$ , or equivalently when  $\beta > \sqrt{d(x, y)} \times \sqrt{f}/6$ . Since the expected value of  $|p_x - p_y|$  is exactly  $\sqrt{d(x, y)}$ , while  $p_x$  and  $p_y$  are separated by at least one bucket of width  $\beta$ , this implies that the expectation is exceeded by a factor of at least  $\sqrt{f}/6 = \Omega(\log^{1/4} k)$ . Setting  $t = \sqrt{f}/6$ , we can say that such a pair  $(x, y)$  is “stretched by a factor  $t$  in the direction  $u$ ”. For any given direction  $u$ , the stretched pairs removed in the pruning step are disjoint, and hence form a large matching  $M_u$ .

Arora et al. showed the following geometric property—for a given set  $W$  and some constant  $C$ , the number of disjoint  $t$ -stretched pairs in  $W \times W$  cannot be more than  $C|W|$  with constant probability (over the choice of  $u$ ); however, their proof only proved this for stretch  $t = \Omega(\log^{1/3} |W|)$ . The dependence on  $t$  was improved subsequently by Lee [2005] to  $t = \Omega(\log^{1/4} |W|)$ .

In order to make the above discussion more precise, let us first recall the definition of a stretched set of points.

**DEFINITION 3.3** [ARORA ET AL. 2004], DEFN. 4. *A set of  $n$  points  $\vec{x}_1, \dots, \vec{x}_n$  in  $\mathbb{R}^n$  is said to be  $(t, \gamma, \beta)$ -stretched at scale  $l$ , if for at least a  $\gamma$  fraction of the directions  $u$ , there is a partial matching  $M_u = \{(x_i, y_i)\}_i$ , with  $|M_u| \geq \beta n$ , such that for all  $(x, y) \in M_u$ ,  $d(x, y) \leq l^2$  and  $\langle u, \vec{x} - \vec{y} \rangle \geq tl/\sqrt{n}$ . That is, the pair  $(x, y)$  is stretched by a factor of  $t$  in direction  $u$ .*

**THEOREM 3.4** [ARORA ET AL. 2004], THM. 5. *For any  $\gamma, \beta > 0$ , there is a  $C = C(\gamma, \beta)$  such that if  $t > C \log^{1/3} n$ , then no set of  $n$  points in  $\mathbb{R}^n$  can be  $(t, \gamma, \beta)$ -stretched for any scale  $l$ .*

The above theorem has been subsequently improved by Lee to the following (as implied by [Lee 2005, Thm. 4.1]).

**THEOREM 3.5.** *For any  $\gamma, \beta > 0$ , there is a  $C = C(\gamma, \beta)$  such that if  $t > C \log^{1/4} n$ , then no set of  $n$  points in  $\mathbb{R}^n$  can be  $(t, \gamma, \beta)$ -stretched for any scale  $l$ .*

Summarizing the implication of Theorem 3.5 in our setting, we get the following corollary.

**COROLLARY 3.6.** *Let  $W$  be a set of vectors corresponding to some subset of terminals satisfying the following property: with probability  $\Theta(1)$  over the choice of a random unit vector  $u$ , there exist subsets  $S_u, T_u \subseteq W$  and a constant  $\rho$  such that  $|S_u| \geq \rho|W|$  and  $|T_u| \geq \rho|W|$ , and the length of the projection  $|\langle u, \vec{x} - \vec{y} \rangle| \geq \ell/(6\sqrt{n})$  for all  $\vec{x} \in S_u$  and  $\vec{y} \in T_u$ . Then with probability  $\Theta(1)$  over the choice of  $u$ , the pruning procedure applied to sets  $S_u$  and  $T_u$  returns sets  $S'_u$  and  $T'_u$  with  $|S'_u| \geq \frac{3}{4}|S_u|$  and  $|T'_u| \geq \frac{3}{4}|T_u|$ , such that for all  $\vec{x} \in S'_u$  and  $\vec{y} \in T'_u$ ,  $d(x, y) \geq \ell^2/f$ .*

**PROOF.** For a unit vector  $u$ , let  $M(u)$  denote the matching obtained by taking the pairs  $(x, y)$  of terminals that are deleted by the pruning procedure when given the vector  $u$ . Note that pairs  $(x, y) \in M(u)$  have the property that  $d(x, y) < \ell^2/f$  and  $|p_x - p_y| > \ell/6$ . For the sake of contradiction, suppose there is a constant  $\gamma$  such that the matchings  $M(u)$  are larger than  $\rho/4|W|$  with probability at least  $1 - \gamma$  over the choice of  $u$ .

Using Definition 3.3 above, we get that the vectors in  $W$  form an  $(6\sqrt{f}, \gamma, \rho/4)$ -stretched set at scale  $\ell/\sqrt{f}$ . Theorem 3.5 now implies that  $6\sqrt{f} = 6\sqrt{c}(\log k)^{1/4}$

must be at most  $C \log^{1/4} |W|$ . However, since  $|W| \leq 2k$ , setting the parameter  $c$  suitably large compared to  $C$  would give us the contradiction.  $\square$

Finally, we are in a position to prove Theorem 3.1 using Lemma 3.2 and Corollary 3.6.

**PROOF OF THEOREM 3.1.** Define  $W$  to be  $\mathcal{D}_i$ , the set of all terminals that belong to some terminal pair in  $\mathbb{D}_i$ . Let  $a$  be the index corresponding to the larger of  $\mathbb{B}_0$  and  $\mathbb{B}_1$  before the pruning step, and set  $S_u = A_a$  and  $T_u = A_{(a+2) \bmod 4}$  for this value of  $a$ . Lemma 3.2 assures us that  $|S_u| = |T_u| \geq \frac{1}{32} |\mathbb{D}_i| = \frac{1}{16} |W|$  with probability  $\frac{1}{15}$  (over the random choice of the vector  $u \in \mathbb{R}^n$ ). Furthermore, for each  $\vec{x} \in S_u$  and  $\vec{y} \in T_u$ , the fact that  $|p_x - p_y| \geq \beta$  translates to the statement that  $\langle \vec{x} - \vec{y}, u \rangle \geq \ell / (6\sqrt{n})$ .

These vectors satisfy the conditions of Corollary 3.6, and hence we can infer that with a constant probability, the pruning procedure removes at most  $\frac{1}{4} |S_u|$  and  $\frac{1}{4} |T_u|$  vertices from  $S_u$  and  $T_u$  respectively. Their partners may be pruned in the cleanup step as well, and hence the total number of terminal pairs pruned is at most  $\frac{1}{2} |S_u|$ . Thus the number of terminal pairs remaining in  $\mathbb{D}_i \cap (S'_u \times T'_u)$  is at least  $\frac{1}{2} |S_u| \geq \frac{1}{64} |\mathbb{D}_i|$  pairs. Recall that for each pair  $(x, y) \in \mathbb{D}_i \cap (S'_u \times T'_u)$ , this pair is  $\eta$ -separated: both  $d(x, V \setminus C_i(x))$  and  $d(y, V \setminus C_i(y))$  are at least  $\eta d(x, y)$ .

Since this happens with a constant probability, we will need to repeat Steps 1-3 of the procedure (each time with a new unit vector  $u$ ) only a constant number of times until we find a partition that  $\eta$ -separates at least  $\frac{1}{64} |\mathbb{D}_i|$  of the terminal pairs; this proves the result.  $\square$

Running the procedure **Project-&-Prune** for each distance scale  $i$  between 1 and  $\lfloor \log \Delta(d) \rfloor$ , we can get the following result with  $\gamma = \frac{1}{64}$ .

**THEOREM 3.7.** *Given a negative-type metric  $d$ , we can find in randomized polynomial time a decomposition suite  $\Pi = \{P_i\}$  that  $\eta$ -separates a constant fraction  $\gamma$  of the terminal pairs at each distance scale  $i$ .*

In the next section, we will extend this result to get a set of  $O(\log k)$  decomposition suites  $\{\Pi_j\}$  so that each terminal pair  $(x, y) \in \mathbb{D}$  is separated in a constant fraction of the  $\Pi_j$ 's.

#### 4. OBTAINING DECOMPOSITION BUNDLES: WEIGHTING AND WATCHING

To start off, let us observe that the result in Theorem 3.7 can be generalized to the case where terminal pairs have an associated weight  $w_{xy} \in \{0, 1, 2, \dots, k\}$ .

**LEMMA 4.1.** *Given terminal pairs  $(x, y) \in \mathbb{D}$  with weights  $w_{xy} \in \{0, 1, \dots, k\}$ , there is a randomized polynomial time algorithm that outputs a decomposition suite  $\Pi$  which, for each distance scale  $i$ ,  $\Omega(1/\sqrt{\log k})$ -separates terminals with total weight at least  $\gamma \sum_{(x,y) \in \mathbb{D}_i} w_{xy}$  with  $\gamma = \frac{1}{64}$ .*

**PROOF.** The proof is almost immediate: we replace each terminal pair  $(x, y) \in \mathbb{D}_i$  having weight  $w_{xy} > 0$  with  $w_{xy}$  new terminal pairs  $(x^j, y^j)$ , where the points  $\{x^j\}$  and  $\{y^j\}$  are placed at distance 0 to  $x$  and  $y$  respectively. Doing this reduction for all weighted pairs gives us an unweighted instance with a set  $\mathbb{D}'_i$  of terminal pairs with  $|\mathbb{D}'_i| \leq k |\mathbb{D}_i|$ . Now Theorem 3.7 gives us a decomposition suite

$\eta$ -separating at least  $\frac{1}{64} |\mathbb{D}'_i|$  of the new terminal pairs at distance scale  $i$ , where  $\eta = 1/O(\sqrt{\log \mathbb{D}'_i}) = 1/O(\sqrt{\log k})$ . Finally, observing that the separated terminal pairs at scale  $i$  contribute at least  $\frac{1}{64} \sum_{(x,y) \in \mathbb{D}_i} w_{xy}$  completes the claim.  $\square$

In the sequel, we will associate weights with the terminal pairs in  $\mathbb{D}$  and run the procedure from Lemma 4.1 repeatedly. The weights start off at  $k$ , and the weight of a pair that is separated in some iteration is halved in the subsequent iteration; this reweighting ensures that all pairs are separated in significantly many rounds. (Note: this weighting argument is fairly standard and has been used, e.g., in geometric algorithms [Clarkson 1995], machine learning [Littlestone and Warmuth 1994], and many other areas; see Welzl [1996] for a survey.)

**The Algorithm:**

- (1) Initialize  $w^{(0)}(x, y) = 2^{\lceil \log k \rceil}$  for all terminal pairs  $(x, y) \in \mathbb{D}$ . Set  $j = 0$ .
- (2) Use the algorithm from Lemma 4.1 to obtain a decomposition suite  $\Pi_j$ . Let  $T_j$  be the set of terminal pairs  $\eta$ -separated by this decomposition.
- (3) For all  $(x, y) \in T_j$ , set  $w^{(j+1)}(x, y) \leftarrow w^{(j)}(x, y)/2$ ; set  $w^{(j+1)}(x, y) \leftarrow w^{(j)}(x, y)$  for the others. If  $w^{(j+1)}(x, y) < 1$  then  $w^{(j+1)}(x, y) \leftarrow 0$ .
- (4) Increment  $j \leftarrow j + 1$ . If  $\sum_{(x,y) \in \mathbb{D}_i} w^{(j)}(x, y) \geq 1$  for some  $i$ , go to step 2, else halt.

Note that the distance function  $d$  in each iteration of the algorithm remains the same.

LEMMA 4.2. *In each iteration  $j$  of the above algorithm, for all scales  $i$ , the following holds*

$$\sum_{(x,y) \in \mathbb{D}_i} w^{(j+1)}(x, y) \leq \left(1 - \frac{\gamma}{2}\right) \sum_{(x,y) \in \mathbb{D}_i} w^{(j)}(x, y)$$

PROOF. In each iteration, the algorithm of Lemma 4.1 separates at least a  $\gamma$  fraction of the weight  $\sum_{(x,y) \in \mathbb{D}_i} w^j(x, y)$  for all  $i$ , and hence the total weight in the next round drops by at least half this amount.  $\square$

Noting that initially we have  $\sum_{(x,y) \in \mathbb{D}_i} w^{(0)}(x, y) \leq 2k^2$ , one derives the following simple corollary:

COROLLARY 4.3. *The above algorithm has at most  $\frac{4}{\gamma} \log k$  iterations.*

LEMMA 4.4. *For all distance scales  $i$ , every pair  $(x, y) \in \mathbb{D}_i$  is  $\eta$ -separated in at least  $\log k$  iterations.*

PROOF. Since we start off with  $w^{(0)}(x, y) \geq k$  and end with  $w^{(j)}(x, y) < 1$ , the weight  $w^{(j)}(x, y)$  must have been decremented at least  $\log k$  times. Each such reduction corresponds to a round  $j$  in which  $(x, y)$  was  $\eta$ -separated by  $\Pi_j$ .  $\square$

THEOREM 4.5. *The above procedure outputs an  $\eta$ -decomposition bundle with at most  $\frac{4}{\gamma} \log k$  decomposition suites, such that each terminal pair  $(x, y)$  is  $\eta$ -separated in at least  $\log k$  of these suites.*

## 5. EMBEDDING VIA DECOMPOSITION BUNDLES

In the previous sections we have constructed a decomposition bundle with a large separation between terminal pairs. Now, we show how to obtain a small distortion  $\ell_2$ -embedding from this. The proof mainly follows the lines of the results in Krauthgamer et al. [2005].

**THEOREM 5.1.** *Given an  $\alpha(x, y)$ -decomposition bundle for the metric  $d$  and a set  $\mathbb{D}$ , there exists a randomized contracting embedding  $\varphi : V \rightarrow \ell_2$ , such that for each pair  $(x, y) \in \mathbb{D}$ ,*

$$\|\varphi(x) - \varphi(y)\|_2 \geq \Omega\left(\sqrt{\frac{\alpha(x, y)}{\log k}}\right) \cdot d(x, y)$$

Note that for  $\alpha(x, y) = \Omega(1/\sqrt{\log k})$  this theorem implies Theorem 1.1.

Along the lines of the reasoning in [Krauthgamer et al. 2005], we define a measure of “local expansion”. Let

$$V(x, y) = \max \left\{ \log \frac{|B(x, 2d(x, y))|}{|B(x, d(x, y)/8)|}, \log \frac{|B(y, 2d(x, y))|}{|B(y, d(x, y)/8)|} \right\}$$

where  $B(x, r)$  denotes the set of *terminal nodes* within the ball of radius  $r$  around  $x$ . We derive Theorem 5.1 from the following lemma.

**LEMMA 5.2.** *Given an  $\alpha(x, y)$ -decomposition bundle, there is a randomized contracting embedding  $\varphi : V \rightarrow \ell_2$  such that for every pair  $(x, y)$  with constant probability*

$$\|\varphi(x) - \varphi(y)\|_2 \geq \Omega\left(\sqrt{\frac{V(x, y)}{\log k}} \cdot \alpha(x, y)\right) \cdot d(x, y) .$$

By repeatedly applying Lemma 5.2, we obtain the following guarantee:

**COROLLARY 5.3.** *Given an  $\alpha(x, y)$ -decomposition bundle, there is a randomized contracting embedding  $\varphi : V \rightarrow \ell_2$  such that for every pair  $(x, y)$ ,*

$$\|\varphi(x) - \varphi(y)\|_2 \geq \Omega\left(\sqrt{\frac{V(x, y)}{\log k}} \cdot \alpha(x, y)\right) \cdot d(x, y) .$$

**PROOF.** The corollary follows by applying Lemma 5.2 repeatedly and independently for each decomposition suite several times. Then concatenating and rescaling the resulting maps gives with high probability an embedding that fulfills the corollary. In passing, we note that this algorithm (using independent repetitions) may result in an embedding with a large number of dimensions, which may not be algorithmically desirable. However, it shows the *existence* of such an embedding, and we can then use semidefinite programming followed by random projections to obtain a nearly-optimal embedding of the metric into  $\ell_2$  with  $O(\log n)$  dimensions in randomized polynomial time.  $\square$

To see that the above corollary implies Theorem 5.1, we use a decomposition due to Calinescu et al. [2004] and Fakcharoenphol et al. [2003] (and its extension to general measures, as observed by Lee and Naor [2005] and by Krauthgamer et al. [2005]) that has the property that with probability at least  $1/2$ , a pair  $(x, y)$

is  $\Omega(1/V(x, y))$ -separated in this decomposition. Applying the corollary to this decomposition bundle, we get an embedding  $\varphi_1$ , such that

$$\|\varphi_1(x) - \varphi_1(y)\|_2 \geq \Omega\left(\frac{1}{\sqrt{V(x, y) \cdot \log k}}\right) \cdot d(x, y) .$$

Applying the corollary to the decomposition bundle assumed by the theorem gives an embedding  $\varphi_2$  with

$$\|\varphi_2(x) - \varphi_2(y)\|_2 \geq \Omega\left(\sqrt{\frac{V(x, y)}{\log k}} \cdot \alpha(x, y)\right) \cdot d(x, y) .$$

Concatenating the two mappings and rescaling, we get a contracting embedding  $\varphi = \frac{1}{\sqrt{2}}(\varphi_1 \otimes \varphi_2)$ , with

$$\begin{aligned} \|\varphi(x) - \varphi(y)\|_2 &\geq \Omega\left(\frac{1}{\sqrt{\log k}} \cdot \left(\frac{1}{V(x, y)^{\frac{1}{2}}} + V(x, y)^{\frac{1}{2}} \alpha(x, y)\right)\right) \cdot d(x, y) \\ &\geq \Omega\left(\sqrt{\frac{\alpha(x, y)}{\log k}}\right) \cdot d(x, y) \end{aligned}$$

as desired. Now it remains to prove Lemma 5.2.

### 5.1 The embedding

Let  $T = \{1, \dots, \log k\}$  and  $Q = \{0, \dots, m - 1\}$ , for some suitably chosen constant  $m$ . In the following we define an embedding into  $|T| \cdot |Q|$  dimensions. For  $t \in T$ , let  $r_t(x)$  denote the minimum radius  $r$  such that the ball  $B(x, r)$  contains at least  $2^t$  terminal nodes. We call  $r_t(x)$  the  $t$ -radius of  $x$ . Further, let  $\ell_t(x) \in \mathbb{N}$  denote the distance class this radius belongs to (i.e.,  $2^{\ell_t(x)-1} \leq r_t(x) \leq 2^{\ell_t(x)}$ ).

Pick a decomposition suite  $\Pi = \{P_s\}$  from the decomposition bundle at random. In the following  $\delta(x, y)$  denotes the separation-factor between  $x$  and  $y$  in this suite, i.e.,  $\delta(x, y) = \frac{1}{d(x, y)} \min\{d(x, V \setminus C_s(x)), d(y, V \setminus C_s(y))\}$  if  $C_s(y) \neq C_s(x)$  and 0, otherwise. Observe that with constant probability we have  $\delta(x, y) \geq \alpha(x, y)$ .

The standard way to obtain an embedding from a decomposition suite is to create a coordinate for every distance scale and embed points in this coordinate with respect to the partitioning for this scale. For example, one could assign a random color, 0 or 1, to each cluster  $C \in P_i$ . Let  $W_i$  denote the set of nodes contained in clusters with color 0 in partitioning  $P_i$ . By setting the  $i$ -th coordinate of the image  $\varphi(x)$  of a point  $x$  to  $d(x, W_i)$ , a pair  $(x, y)$  gets a distance  $\Omega(\delta(x, y)d(x, y))$  with probability 1/2, because this is the probability that the clusters  $C_i(x)$  and  $C_i(y)$  get different colors (in this case the distance is  $\Omega(\delta(x, y)d(x, y))$  since both nodes are at least that far away from the boundary of their cluster). Overall this approach gives an embedding into  $\ell_2$  with contraction  $O(\sqrt{\log k}/\delta(x, y))$ , and has e.g. been used by Rao [1999] for getting a  $\sqrt{\log n}$  embedding of planar metrics into  $\ell_2$ .

In order to improve this, along the lines of the recent measured descent idea from [Krauthgamer et al. 2005], the goal is to construct an embedding in which the distance between  $(\varphi(x), \varphi(y))$  increases as the local expansion  $V(x, y)$  increases. This can be achieved by constructing a coordinate for every  $t \in T$  and then embed points in this coordinate according to the partitioning for the corresponding distance

scale  $\ell_t(x)$  (i.e., different points use different distance scales depending on their local expansion). Thereby, for a pair with a high  $V(x, y)$ -value the nodes will often ( $\approx V(x, y)$  times) be embedded according to the partitioning for distance scale  $i = \lfloor \log d(x, y) \rfloor$  that corresponds to  $d(x, y)$ . Therefore, the pair  $(x, y)$  gets a larger distance (by a factor of roughly  $\sqrt{V(x, y)}$ ) in this embedding than in the standard approach.

However, transferring the rest of the standard analysis to this new idea has some difficulties. If we define the set  $W_t$  as the nodes  $x$  that are colored 0 in the partitioning for scale  $\ell_t(x)$  we cannot argue that for a pair  $(x, y)$  either  $d(x, W_t)$  or  $d(y, W_t)$  is large, because nodes  $u$  very close to  $x$  or  $y$  may have distance scales  $\ell_t(u)$  that are different from  $\ell_t(x)$  or  $\ell_t(y)$ . In order to ensure local consistency such that all nodes close to  $x$  obtain their color from the same partitioning, we construct several coordinates in the embedding for every  $t$ , such that for each distance scale  $\ell_t(x)$  there is a coordinate in which all nodes close to  $x$  derive their color from the partitioning for scale  $\ell_t(x)$ . The details are as follows.

Let  $Q = \{0, \dots, m-1\}$  denote the set of indices of coordinates corresponding to each value of  $t$ . For each  $q \in Q$ , we partition the distance scales into *groups*  $g_q$  of size  $m$  each, and let the median scale in each group represent that group for the coordinate corresponding to  $q$ . In the  $(q, t)^{\text{th}}$  coordinate, the color of a node is chosen according to the median distance scale in the group  $g_q$  to which  $\ell_t(x)$  belongs.

In particular, let  $g_q(\ell) := \lceil \frac{\ell-q}{m} \rceil$ . Note that each distance group contains (at most)  $m$  consecutive distance classes which means that distances within a group differ at most by a constant factor – all distances in group  $g$  are in  $\Theta(2^{m \cdot g})$ . We define a mapping  $\pi_q$  between distance classes that maps all classes of a group to the median distance class in this group (the value of  $\pi_q$  for the first and last distance group is rounded off appropriately; we omit a precise definition for the sake of clarity).

$$\pi_q(\ell) := i + m \cdot g_q(\ell) - \lfloor \frac{m}{2} \rfloor$$

Observe that this partitioning satisfies the key property that for each distance class  $i$ , there exists a  $q$  such that  $\pi_q(i) = i$ .

Based on this mapping we define a set  $W_t^q$  for each choice of  $t \in T$  and  $q \in Q$  by  $W_t^q = \{x \in V : \text{color}_{\pi_q(\ell_t(x))}(x) = 0\}$ , where  $\text{color}_i(x)$  denotes the color of the cluster that contains  $x$  in partitioning  $P_i$ . Note that all nodes whose  $t$ -radii fall into the same distance group (w.r.t. parameter  $q$ ) derive their color (and hence whether they belong to  $W_t^q$ ) from the same partitioning.

Based on the sets  $W_t^q$  we define an embedding  $\varphi_{t,q} : V \rightarrow \mathbb{R}$  for each coordinate  $(t, q) — \varphi_{t,q}(x) = d(x, W_t^q)$ . The embedding  $\varphi : V \rightarrow \mathbb{R}^{|T| \cdot |Q|}$  is defined by

$$\varphi(x) := \otimes_{t,q} \varphi_{t,q}(x). \quad (3)$$

In the next section, we analyse the distortion of the map  $\varphi$ .

## 5.2 The Analysis

Since each coordinate of  $\varphi$  maps point  $x$  to its distance from some subset of points, it follows that each coordinate of this embedding is contracting. Therefore, we have

for all  $x, y \in V$

$$\begin{aligned} \|\varphi(x) - \varphi(y)\|_2 &\leq \sqrt{|T| \cdot |Q| \cdot d(x, y)^2} \\ &\leq O(\sqrt{\log k}) \cdot d(x, y) \end{aligned}$$

Now, we show that for a pair  $x, y$  that is  $\delta(x, y)$ -separated in the partitioning corresponding to its distance scale  $\lfloor \log(d(x, y)) \rfloor$ , with a constant probability, we get

$$\|\varphi(x) - \varphi(y)\|_2 \geq \Omega(\delta(x, y) \cdot d(x, y)) \cdot \sqrt{V(x, y)} \quad (4)$$

This gives Lemma 5.2 since  $\delta(x, y) > \alpha(x, y)$  with constant probability.

Fix a pair  $(x, y)$  that is  $\delta(x, y)$ -separated in the partitioning for distance scale  $\lfloor \log(d(x, y)) \rfloor$ . Without loss of generality assume that the maximum in the definition of  $V(x, y)$  is attained by the first term, i.e.  $\frac{|B(x, 2d(x, y))|}{|B(x, d(x, y)/8)|} \geq \frac{|B(y, 2d(x, y))|}{|B(y, d(x, y)/8)|}$ . We show that for any  $t$  with  $|B(x, d(x, y)/8)| \leq 2^t \leq |B(x, 2d(x, y))|$ , there is a  $q \in Q$  such that the coordinate  $(t, q)$  gives a large contribution, i.e.,  $|\varphi_{t, q}(x) - \varphi_{t, q}(y)| \geq \Omega(\delta(x, y) \cdot d(x, y))$ . Equation 4 then follows.

We fix an integer  $t$  with  $\log(|B(x, d(x, y)/8)|) \leq t \leq \log(|B(x, 2d(x, y))|)$ , and we use  $i = \lfloor \log d(x, y) \rfloor$  to denote the distance class of  $d(x, y)$ . Clearly, the distance class  $\ell_t(x)$  of the  $t$ -radius of  $x$  is in  $\{i - 4, \dots, i + 2\}$ , because  $d(x, y)/8 \leq r_t(x) \leq 2d(x, y)$ . The following claim gives a similar bound on the  $t$ -radius for nodes that are close to  $x$ .

**CLAIM 5.4.** *Let  $z \in B(x, \frac{1}{16}d(x, y))$ . Then  $\ell_t(z) \in \{i - 5, i + 3\}$ .*

**PROOF.** For the  $t$ -radius  $r_t(z)$  around  $z$  we have  $r_t(x) - d(x, y)/16 \leq r_t(z) \leq r_t(x) + d(x, y)/16$ . Since  $d(x, y)/8 \leq r_t(x) \leq 2d(x, y)$  we get  $\frac{1}{16}d(x, y) \leq r_t(z) \leq \frac{33}{16}d(x, y)$ , which yields the claim.  $\square$

In the following we choose  $m$  (the number of distances classes within a group) as 10, and  $q$  such that  $\pi_q(i) = i$ , i.e.,  $i$  is the median of its distance group. Then the above claim ensures that for all nodes  $z \in B(x, \frac{1}{16}d(x, y))$ , the distance class  $\ell_t(z)$  is in the same distance group as  $i$ . Furthermore, these nodes choose their color (that decides whether they belong to  $W_t^q$ ) according to the partitioning for distance scale  $i$ . Recall that  $x$  is  $\delta(x, y)$ -separated in this partitioning. Therefore, we can make the following claim.

**CLAIM 5.5.** *If  $x$  does not belong to the set  $W_t^q$ , then,*

$$d(x, W_t^q) \geq \min\{\frac{1}{16}, \delta(x, y)\} d(x, y) \geq \frac{1}{16} \delta(x, y) d(x, y)$$

Now, we consider the following events concerning the distances of  $x$  and  $y$  from  $W_t^q$ , respectively.

- $X_0 = \{d(x, W_t^q) = 0\}$ , i.e.,  $x \in W_t^q$
- $X_{\text{far}} = \{d(x, W_t^q) > \frac{1}{16}\delta(x, y)d(x, y)\}$
- $Y_{\text{close}} = \{d(y, W_t^q) \leq \frac{1}{32}\delta(x, y)d(x, y)\}$
- $Y_{\text{far}} = \{d(y, W_t^q) > \frac{1}{32}\delta(x, y)d(x, y)\}$

These events only depend on the random colorings chosen for the partitionings in different distance classes. First we claim that the events  $X_0$  and  $X_{\text{far}}$  are independent

of events  $Y_{\text{close}}$  and  $Y_{\text{far}}$ . To see this, note that  $X_0$  and  $X_{\text{far}}$  only depend on colors chosen for nodes in  $B(x, \frac{1}{16}\delta(x, y)d(x, y))$ . Our choice of  $q$  ensures that these colors are derived from the partitioning for distance class  $i$ , and Claim 5.4 implies that all nodes in  $B(x, \frac{1}{16}\delta(x, y)d(x, y))$  get the color assigned to the cluster  $C_i(x)$ .

The events  $Y_{\text{close}}$  and  $Y_{\text{far}}$ , however, depend on colors chosen for nodes that lie in the ball  $B(y, \delta(x, y) \frac{1}{32}d(x, y))$ . Such a color is either derived from a partitioning for a distance class different from  $i$  (in this case independence is immediate), or it is equal to the color assigned to the cluster  $C_i(y)$ , using the fact that  $d(y, V \setminus C_i(y)) \geq \delta(x, y)d(x, y)$ . In the latter case the independence follows, since  $x$  and  $y$  lie in different clusters in this partitioning as they are separated by it.

If  $X_0 \cap Y_{\text{far}}$  or  $X_{\text{far}} \cap Y_{\text{close}}$  happens, then the dimension  $(t, q)$  gives a contribution of  $\Omega(\delta(x, y)d(x, y))$ . This happens with probability

$$\begin{aligned} & \Pr[X_0 \cap Y_{\text{far}} \uplus X_{\text{far}} \cap Y_{\text{close}}] \\ &= \Pr[X_0 \cap Y_{\text{far}}] + \Pr[X_{\text{far}} \cap Y_{\text{close}}] \\ &= \Pr[X_0] \cdot \Pr[Y_{\text{far}}] + \Pr[X_{\text{far}}] \cdot \Pr[Y_{\text{close}}] \\ &= \Pr[X_0] \cdot \Pr[Y_{\text{far}}] + \Pr[X_{\text{far}}] \cdot (1 - \Pr[Y_{\text{far}}]) \\ &= 1/2 . \end{aligned}$$

Here we used the fact that  $\Pr[X_0] = \Pr[X_{\text{far}}] = 1/2$  which holds due to Claim 5.5. This completes the proof of Lemma 5.2.

#### ACKNOWLEDGMENTS

We would like to thank Sanjeev Arora, Avrim Blum, Hubert Chan, Vineet Goyal, James R. Lee, Satish Rao, R. Ravi, and Mohit Singh for useful discussions, and Jiri Matoušek for pointing us to [Welzl 1996]. Many thanks to James R. Lee for sending us a manuscript of [Krauthgamer et al. 2005].

#### REFERENCES

- ARORA, S., LEE, J., AND NAOR, A. 2005. Euclidean distortion and the sparsest cut. In *Proceedings of the 37th ACM Symposium on Theory of Computing (STOC)*. ACM Press, New York, NY, USA, 553–562.
- ARORA, S., RAO, S., AND VAZIRANI, U. 2004. Expander flows, geometric embeddings, and graph partitionings. In *Proceedings of the 36th ACM Symposium on Theory of Computing (STOC)*. ACM Press, New York, NY, USA, 222–231.
- AUMANN, Y. AND RABANI, Y. 1998. An  $O(\log k)$  approximate min-cut max-flow theorem and approximation algorithm. *SIAM Journal on Computing* 27, 1, 291–301.
- BARTAL, Y. 1998. On approximating arbitrary metrics by tree metrics. In *Proceedings of the 30th ACM Symposium on Theory of Computing (STOC)*. ACM Press, New York, NY, USA, 161–168.
- BOURGAIN, J. 1985. On Lipschitz embeddings of finite metric spaces in Hilbert space. *Israel Journal of Mathematics* 52, 1-2, 46–52.
- CĂLINESCU, G., KARLOFF, H. J., AND RABANI, Y. 2004. Approximation algorithms for the 0-extension problem. *SIAM Journal on Computing* 34, 2, 358–372. Also in *Proc. 12th SODA*, 2001, pp. 8–16.
- CHAWLA, S., KRAUTHGAMER, R., KUMAR, R., RABANI, Y., AND SIVAKUMAR, D. 2006. On the hardness of approximating sparsest cut and multicut. *Computational Complexity* 15, 2, 94–114. Also in *Proc. 20th CCC*, 2005, pp. 144–153.

- CHEKURI, C., GUPTA, A., NEWMAN, I., RABINOVICH, Y., AND SINCLAIR, A. 2006. Embedding  $k$ -outerplanar graphs into  $\ell_1$ . *SIAM Journal on Discrete Mathematics* 20, 1, 119–136.
- CLARKSON, K. L. 1995. Las Vegas algorithms for linear and integer programming when the dimension is small. *Journal of the ACM* 42, 2, 488–499.
- DEZA, M. M. AND LAURENT, M. 1997. *Geometry of Cuts and Metrics*. Algorithms and Combinatorics, vol. 15. Springer-Verlag, Berlin, Germany.
- ENFLO, P. 1969. On the nonexistence of uniform homeomorphisms between  $L_p$ -spaces. *Arkiv För Matematik* 8, 103–105.
- FAKCHAROENPHOL, J., RAO, S. B., AND TALWAR, K. 2003. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35th ACM Symposium on Theory of Computing (STOC)*. ACM Press, New York, NY, USA, 448–455.
- GOEMANS, M. X. 1997. Semidefinite programming and combinatorial optimization. *Mathematical Programming* 79, 1–3, 143–161.
- GUPTA, A., KRAUTHGAMER, R., AND LEE, J. R. 2003. Bounded geometries, fractals, and low-distortion embeddings. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, Washington, DC, USA, 534–543.
- GUPTA, A., NEWMAN, I., RABINOVICH, Y., AND SINCLAIR, A. 2004. Cuts, trees and  $\ell_1$ -embeddings of graphs. *Combinatorica* 24, 2, 233–269. Also in *Proc. 35th FOCS*, 1999, pp. 399–409.
- INDYK, P. 2001. Algorithmic applications of low-distortion geometric embeddings. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, Washington, DC, USA, 10–33.
- KARZANOV, A. V. 1985. Metrics and undirected cuts. *Mathematical Programming* 32, 2, 183–198.
- KHOT, S. 2002. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th ACM Symposium on Theory of Computing (STOC)*. ACM Press, New York, NY, USA, 767–775.
- KHOT, S. AND VISHNOI, N. 2005. The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into  $\ell_1$ . In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, Washington, DC, USA, 53–62.
- KRAUTHGAMER, R., LEE, J., MENDEL, M., AND NAOR, A. 2005. Measured descent: A new embedding method for finite metrics. *Geometric and Functional Analysis* 15, 4, 839–858. Also in *Proc. 45th FOCS*, 2004, pp. 434–443.
- KRAUTHGAMER, R. AND RABANI, Y. 2006. Improved lower bounds for embeddings into  $\ell_1$ . In *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1010–1017.
- LEE, J. R. 2005. On distance scales, embeddings, and efficient relaxations of the cut cone. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 92–101.
- LEE, J. R. AND NAOR, A. 2005. Extending Lipschitz functions via random metric partitions. *Inventiones Mathematicae* 160, 1, 59–95.
- LEIGHTON, F. T. AND RAO, S. B. 1988. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, Washington, DC, USA, 422–431.
- LINIAL, N. 2002. Finite metric spaces – combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. III*. Higher Education Press, Beijing, China, 573–586.
- LINIAL, N., LONDON, E., AND RABINOVICH, Y. 1995. The geometry of graphs and some of its algorithmic applications. *Combinatorica* 15, 2, 215–245. Also in *Proc. 35th FOCS*, 1994, pp. 577–591.
- LITTLESTONE, N. AND WARMUTH, M. K. 1994. The weighted majority algorithm. *Information and Computation* 108, 2, 212–261.
- MATOUŠEK, J. 1999. On embedding trees into uniformly convex Banach spaces. *Israel Journal of Mathematics* 114, 221–237. (Czech version in: *Lipschitz distance of metric spaces*, C. Sc. degree thesis, Charles University, 1990).

- MATOUŠEK, J. 2002. *Lectures on Discrete Geometry*. Graduate Texts in Mathematics, vol. 212. Springer, New York, NY, USA.
- MATOUŠEK, J. 1996. On the distortion required for embedding finite metric spaces into normed spaces. *Israel Journal of Mathematics* 93, 333–344.
- RABINOVICH, Y. 2003. On average distortion of embedding metrics into the line and into  $\ell_1$ . In *Proceedings of the 35th ACM Symposium on Theory of Computing (STOC)*. ACM Press, New York, NY, USA, 456–462.
- RAO, S. B. 1999. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th ACM Symposium on Computational Geometry*. ACM Press, New York, NY, USA, 300–306.
- SHAHROKHI, F. AND MATULA, D. W. 1990. The maximum concurrent flow problem. *Journal of the ACM* 37, 2, 318–334.
- SHMOYS, D. B. 1997. Cut problems and their application to divide-and-conquer. In *Approximation Algorithms for NP-hard Problems*, D. S. Hochbaum, Ed. PWS Publishing, Boston, MA, USA, 192–235.
- WELZL, E. 1996. Suchen und Konstruieren durch Verdoppeln. In *Highlights der Informatik*, I. Wegener, Ed. Springer, Berlin, Germany, 221–228.