

Near Linear Lower Bound for Dimension Reduction in ℓ_1

Alexandr Andoni
Microsoft Research SVC

Mountain View, CA, USA
Email: andoni@mit.edu

Moses S. Charikar, Ofer Neiman, Huy L. Nguyen
Department of Computer Science
Center for Computational Intractability
Princeton University, Princeton, NJ, USA
Email: {moses, oneiman, hlnghuyen}@cs.princeton.edu

Abstract— Given a set of n points in ℓ_1 , how many dimensions are needed to represent all pairwise distances within a specific distortion? This dimension-distortion tradeoff question is well understood for the ℓ_2 norm, where $O((\log n)/\epsilon^2)$ dimensions suffice to achieve $1 + \epsilon$ distortion. In sharp contrast, there is a significant gap between upper and lower bounds for dimension reduction in ℓ_1 . A recent result shows that distortion $1 + \epsilon$ can be achieved with n/ϵ^2 dimensions. On the other hand, the only lower bounds known are that distortion δ requires $n^{\Omega(1/\delta^2)}$ dimensions and that distortion $1 + \epsilon$ requires $n^{1/2 - O(\epsilon \log(1/\epsilon))}$ dimensions. In this work, we show the first near linear lower bounds for dimension reduction in ℓ_1 . In particular, we show that $1 + \epsilon$ distortion requires at least $n^{1 - O(1/\log(1/\epsilon))}$ dimensions.

Our proofs are combinatorial, but inspired by linear programming. In fact, our techniques lead to a simple combinatorial argument that is equivalent to the LP based proof of Brinkman-Charikar for lower bounds on dimension reduction in ℓ_1 .

Keywords— dimension reduction, metric embedding

1. INTRODUCTION

In this paper, we study dimension reduction questions of the following form: Given a set X of n points in \mathbb{R}^d with distances measured by the ℓ_p norm, the objective is to find an embedding $f : X \rightarrow \mathbb{R}^{d'}$ with $d' \ll d$ that roughly preserves pairwise distances. We say that the *distortion* of the embedding is δ if there exists a constant c such that for all $x, y \in X$:

$$c\|x - y\|_p \leq \|f(x) - f(y)\|_p \leq c \cdot \delta \|x - y\|_p$$

Dimension reduction has received a lot of attention in recent years, due to its numerous applications in computer science (see [8], [9], [11], [12], [7] and the references therein). Furthermore, it is a topic of interest from a purely mathematical point of view and has been well-studied in the area of functional analysis.

For the $\|\cdot\|_2$ norm, the famous Johnson-Lindenstrauss Lemma [10] states that for any $\epsilon > 0$ one can linearly embed an n point subset of ℓ_2 into $d' = O((\log n)/\epsilon^2)$ dimensions with distortion at most $1 + \epsilon$. This is essentially tight, as shown by Alon [1]. Unfortunately, the situation for spaces other than Euclidean is far from understood. In this work, we focus on the ℓ_1 norm for which there are huge gaps between currently known upper and lower bounds.

Ball [2] showed that in order to embed some n point subset of ℓ_1 isometrically (with distortion 1), $\binom{n}{2}$ dimensions are necessary (and always suffice). A strong lower bound for dimension reduction in ℓ_1 was shown by Brinkman and Charikar [4], with a simpler geometric proof by Lee and Naor [13]. They exhibit an n point subset of ℓ_1 requiring dimension at least $n^{\Omega(1/D^2)}$ for distortion D . They also showed that for distortion $1 + \epsilon$ the dimension required is at least $n^{1/2 - \tilde{O}(\epsilon)}$. There has been some progress on upper bounds: Indyk [8] showed a weak form of dimension reduction in ℓ_1 , which is already sufficient for certain applications such as norm estimation in streaming, and approximate nearest neighbor search (see, for example, Datar et al. [6]). For dimension reduction with $1 + \epsilon$ distortion, Schechtman [15], and Talagrand [16] showed that $O((n \log n)/\epsilon^2)$ dimensions suffice. This was recently improved by Newman and Rabinovich [14] to $O(n/\epsilon^2)$ based on the sparsification techniques of Batson, Spielman, and Srivastava [3].

In the $1 + \epsilon$ distortion regime, the gap between the best known lower and upper bounds is $\Omega(n^{1/2})$. We narrow this gap substantially, and show that when the distortion is close to 1, the required number of dimensions is near linear.

Theorem 1. *For any $\epsilon > 0$ and integer $n > 0$ there exists an n point subset X of ℓ_1 , such that any embedding of X into ℓ_1 with distortion $1 + \epsilon$ requires dimension at least $n^{1 - O(1/\log(1/\epsilon))}$.*

1.1. Techniques

The lower bound of [4] and [13] was obtained by using the so-called diamond graph, which is defined recursively: G_0 is a single edge, and G_i is obtained from G_{i-1} by replacing every edge with a 4-cycle where two antipodal vertices of the cycle are identified with the original end points of the edge. [4] established the bound by exhibiting a dual solution of a certain linear program that captures the constraints imposed on pairwise distances by a line embedding. The alternate, simpler proof of [13] first showed a lower bound on embedding the diamond graph in ℓ_p for p close to 1, and then used the relation between ℓ_p and ℓ_1 norms in order to achieve the dimension lower bound for embedding into ℓ_1 . The metric based on the diamond graph in these lower

bounds is isometrically embeddable in \sqrt{n} dimensions – hence it cannot establish a lower bound better than \sqrt{n} .

The basis for our near linear lower bound is a series-parallel graph that generalizes the diamond graph: in the recursive construction, edges are replaced by cycles of length $2/\epsilon$ – we call this the recursive cycle graph. Similarly to [5], [4], we use the notion of *stretch* (defined in the sequel) to argue about the number of dimensions needed for a low distortion ℓ_1 embedding. Informally, a stretch s embedding is a distribution over line embeddings where no pairwise distance is increased by a factor of more than s . We show that a Poincaré inequality – a linear inequality (involving distances for antipodal pairs of the cycles in the graph and edges) — must hold for all embeddings of limited stretch but cannot hold for any low distortion embedding. This implies that any low distortion embedding must have high stretch, and hence, high dimension.

While the use of *stretch* for proving dimension lower bounds has been established in previous work, using this proof technique involves the following two steps: (1) identifying an appropriate Poincaré inequality, and (2) proving that it holds for any line embedding with low stretch. Both steps are challenging in general and this is our main technical contribution. We made crucial use of linear programming in executing these two steps.

We briefly elaborate on this methodology for proving dimension lower bounds. In order to capture the limitations of line embeddings with low stretch, we write (1) linear constraints on pairwise distances (viewed as variables) using the fact that they arise from a line embedding and (2) constraints imposed by the stretch $\leq s$ condition. In general, there is no principled way to write down the appropriate constraints (1) for any line embedding. A key idea we use in expressing these constraints is to incorporate information about the line embedding by using a convenient renaming of vertices which is a function of the line embedding. Next, we write (3) a set of constraints on average distances on certain groups of pairwise distances implied by the distortion $\leq \delta$ condition, using the intuition about the particular metric under consideration, that low stretch embeddings tend to expand certain groups of distances, and contract other groups of distances. For the lower bound example in this paper, these groups are the set of edges and for every level, the set of antipodal pairs at that level. The infeasibility of this system of linear inequalities leads to a proof that no stretch s embedding can have distortion $\leq \delta$. In fact the Poincaré inequality that establishes this is obtained from a linear combination of the constraints (3) where the multipliers are the values of the dual variables for these constraints.

This is in fact how we derived the inequality we use to establish our dimension lower bound. Although the proof we present here makes no reference to linear programming, it is in fact a combinatorial interpretation of the LP based proof (and more insightful than an LP duality based proof, we

believe). Our proof uses a certain charging argument and inspired by this, we give a simple combinatorial proof of the original result of [4] that is equivalent to the original LP based proof. The Poincaré-type inequality and the distances involved are similar to the case of the recursive cycle graph but with a slight twist: some of the distances are directed distances i.e. they could take positive and negative values.

1.2. Organization

In Section 2 we recall the notion of stretch and its connection to distortion and dimension. In Section 3.2 we demonstrate our techniques in a relatively simple setting, and prove Theorem 2, a lower bound for C_n , the cycle graph on n vertices. Then in Section 3.3 we extend the methods to the recursive cycle graph, and prove Theorem 1. Finally in Section 4 we show the versatility of our technique, and reprove [4] lower bound by a somewhat simpler argument.

2. PRELIMINARIES

Definition 1. An embedding of a metric (X, d) to ℓ_1 with stretch s and distortion D is a distribution over mappings $f : X \rightarrow \mathbb{R}$ that satisfies the following conditions for all $x, y \in X$,

- 1) **stretch constraint:** $|f(x) - f(y)| \leq s \cdot d(x, y)$ for all f in the support of the distribution.
- 2) **distortion constraint:** $d(x, y)/D \leq \mathbb{E}_f[|f(x) - f(y)|] \leq d(x, y)$

The following claim relating stretch to dimension was shown in [5]

Claim 1. *If a metric space (X, d) has an embedding into ℓ_1^s with distortion α , then it also has embedding with distortion α and stretch at most s .*

In order to show the tradeoff between stretch and distortion, we will prove certain Poincaré-type inequalities and apply the following lemma, which essentially was proven in [5]. A proof is provided below for completeness.

Lemma 2. *Consider a metric (X, d) . If there are positive numbers $\alpha_{x,y}, \beta_{x,y}$ and γ such that for any line embedding $f : X \rightarrow \mathbb{R}$ of stretch at most s ,*

$$\sum_{x,y \in X} \beta_{x,y} |f(x) - f(y)| - \sum_{x,y \in X} \alpha_{x,y} |f(x) - f(y)| \leq \gamma \quad (1)$$

then any embedding to ℓ_1 of stretch at most s has distortion at least $\frac{\sum_{x,y} \beta_{x,y} d(x,y)}{\sum_{x,y} \alpha_{x,y} d(x,y) + \gamma}$.

Proof: Consider an embedding, i.e. a distribution over line embeddings f , with stretch s and distortion D . By definition,

$$\sum_{x,y \in X} \alpha_{x,y} \mathbb{E}_f[|f(x) - f(y)|] + \gamma \leq \sum_{x,y \in X} \alpha_{x,y} d(x, y) + \gamma$$

and

$$\sum_{x,y \in X} \beta_{x,y} \mathbb{E}_f[|f(x) - f(y)|] \geq \sum_{x,y \in X} \beta_{x,y} d(x,y) / D$$

Since inequality (1) holds for all line embeddings, we have

$$\sum_{x,y \in X} \beta_{x,y} \mathbb{E}_f[|f(x) - f(y)|] \leq \sum_{x,y \in X} \alpha_{x,y} \mathbb{E}_f[|f(x) - f(y)|] + \gamma$$

Combining the above inequalities, we get $\frac{\sum_{x,y} \beta_{x,y} d(x,y)}{D} \leq \sum_{x,y} \alpha_{x,y} d(x,y) + \gamma$ i.e. $D \geq \frac{\sum_{x,y} \beta_{x,y} d(x,y)}{\sum_{x,y} \alpha_{x,y} d(x,y) + \gamma}$. ■

We note that, like in the setting of embeddability into ℓ_1 (without dimension restriction), if a metric X does not admit an embedding of stretch s for a certain distortion, then there exists a proof of non-embeddability of the above form (that is, there exist α and β satisfying the condition of Lemma 2).

Notation 1. Let $x \circ y$ denote the concatenation of two strings x and y .

3. THE LOWER BOUND

3.1. Proof Overview

Our lower bound is based on the shortest path metric of a recursively constructed graph. The basic ingredient of the construction is the cycle C_{2k} and we first establish a dimension-distortion tradeoff for this simple graph. The intuition for the lower bound comes from the basic case of the Borsuk-Ulam theorem: in any continuous embedding of the cycle into \mathbb{R} , an antipodal pair is mapped to the same point. This intuition is made precise in a discrete setting by the notion of *flipped* antipodals (see Definition 2). We establish an inequality relating antipodal distances that “tend to contract” to edge distances that “tend to expand” in a line embedding. The proof then uses a charging argument, based on the flipped antipodals and the triangle inequality, to bound the lengths of the antipodals by the lengths of the edges. Finally the stretch bound is used on some edges that were charged “often”.

C_{2k} has an isometric embedding with stretch k . Our proof shows that any distortion $1 + \epsilon$ embedding of C_{2k} must have stretch at least $\Omega(1/\epsilon)$. This lower bound on stretch is tight up to constants, since C_{2k} has a $1 + \epsilon$ distortion embedding with stretch $O(1/\epsilon)$.

Next we consider a certain product construction of the cycle, that we call the recursive cycle graph G_t – this has $(2k)^t$ edges and $\Theta((2k)^t)$ vertices. For this product construction, we show (roughly speaking) that the stretch bound is raised to the power t . In particular, we show that any $1 + 1/k$ distortion embedding of G_t must have stretch $(\Omega(k))^t$ by reasoning about the lengths of edges and antipodal pairs. This gives our main result showing a near linear lower bound for $1 + \epsilon$ distortion embeddings of ℓ_1 . The lower bound we show for G_t is almost optimal in the following sense: G_t has an embedding with stretch k^t that is isometric for all edges and antipodal pairs. Also, there is

a stretch $(O(1/\epsilon))^t$ embedding of G_t that has distortion at most $1 + \epsilon$ for edges and antipodal pairs.

3.2. Simple Case

We begin by proving the following theorem for the shortest path metric of the cycle C_{2k} . In this process, we will establish some machinery that will be useful for the general case later.

Theorem 2. For any integer k and $\epsilon \geq 1/k$, any embedding of C_{2k} into ℓ_1 with distortion at most $1 + \epsilon$ requires dimension at least $\Omega(1/\epsilon)$.

Assume w.l.o.g that k is even, and label the vertices of C_{2k} by $a_1, a_2, \dots, a_k, a_{-1}, a_{-2}, \dots, a_{-k}$, so that for every $1 \leq i \leq k$, (a_i, a_{-i}) is an antipodal pair denoted by d_i . In what follows we identify indices $k+i$ with $-i$ and $-(k+i)$ with i for any $1 \leq i \leq k$. Fix some $f : C_{2k} \rightarrow \mathbb{R}$, and let $L = \sum_{i=1}^k |f(a_i) - f(a_{-i})|$ and $E = \sum_{i=1}^k |f(a_i) - f(a_{i+1})| + |f(a_{-i}) - f(a_{-(i+1)})|$.

Lemma 3. Let $f : C_{2k} \rightarrow \mathbb{R}$ be a line embedding with stretch at most s , then the following Poincaré-type inequality holds

$$2L - k(1 - 2\epsilon)E \leq s \cdot (2\epsilon k)^2. \quad (2)$$

Before proving this lemma, let us show that it implies Theorem 2. Apply Lemma 2 with the appropriate coefficients $\alpha_{a_i, a_{i+1}} = \alpha_{a_{-i}, a_{-(i+1)}} = k(1 - 2\epsilon)$, $\beta_{a_i, a_{-i}} = 2$ for all $1 \leq i \leq k$ (set α and β to zero everywhere else) and $\gamma = s \cdot (2\epsilon k)^2$, to conclude that any embedding $F : C_{2k} \rightarrow \ell_1$ with stretch at most s must have distortion at least

$$\frac{\sum_{x,y \in V(C_{2k})} \beta_{x,y} d(x,y)}{\sum_{x,y \in V(C_{2k})} \alpha_{x,y} d(x,y) + \gamma} = \frac{2k^2}{2k^2(1 - 2\epsilon) + s \cdot (2\epsilon k)^2} = \frac{1}{1 - 2\epsilon + 2s\epsilon^2},$$

where it was used that the total antipodal length in C_{2k} is k^2 and the total edge length is $2k$. Finally set $s = 1/(2\epsilon)$ and obtain distortion at least $\frac{1}{1 - 2\epsilon + \epsilon} > 1 + \epsilon$. By Claim 1 this suggests that an embedding into fewer than $s = 1/(2\epsilon)$ dimensions cannot have distortion bounded by $1 + \epsilon$, so it concludes the proof of Theorem 2. We now turn to prove the main lemma.

Proof of Lemma 3:

The notion of *flipped* antipodal pairs is key to showing that some antipodal distances tend to be short relative to distances of edges in any line embedding.

Definition 2. Let $f : C_{2k} \rightarrow \mathbb{R}$ and fix $1 \leq i \leq k$. Two antipodal pairs d_i, d_{i+1} are called *flipped* under f if $\text{sign}(f(a_i) - f(a_{-i})) \neq \text{sign}(f(a_{i+1}) - f(a_{-(i+1)}))$ (or if $\text{sign}(f(a_i) - f(a_{-i})) = 0$).

Claim 4. In any map $f : C_{2k} \rightarrow \mathbb{R}$ there exist flipped antipodal pairs d_i, d_{i+1} , which satisfy:

$$\begin{aligned} & |f(a_i) - f(a_{i+1})| + |f(a_{-i}) - f(a_{-(i+1)})| \\ & \geq |f(a_i) - f(a_{-i})| + |f(a_{i+1}) - f(a_{-(i+1)})| \end{aligned} \quad (3)$$

Proof: To see the existence of the flipped antipodal pairs, consider the values $f(a_1) - f(a_{-1}), f(a_2) - f(a_{-2}), \dots, f(a_k) - f(a_{-k}), f(a_{-1}) - f(a_1)$. Since the first is the opposite of the last, the sign must change at some point, and there will be the flipped pair.

To prove (3), assume w.l.o.g that $f(a_i) \geq f(a_{-i})$ and $f(a_{i+1}) \leq f(a_{-(i+1)})$. Then

$$\begin{aligned} & (f(a_i) - f(a_{-i})) + (f(a_{-(i+1)}) - f(a_{i+1})) \\ & = (f(a_i) - f(a_{i+1})) + (f(a_{-(i+1)}) - f(a_{-i})) \\ & \leq |f(a_i) - f(a_{i+1})| + |f(a_{-(i+1)}) - f(a_{-i})|. \end{aligned}$$

■

Labeling Cycle Edges: Consider the two flipped antipodal pairs d_i, d_{i+1} guaranteed to exist by Claim 4. Let e_0 be the length under f of the edge connecting a_i to a_{i+1} , i.e. $e_0 = |f(a_i) - f(a_{i+1})|$, and $e_{0'} = |f(a_{-i}) - f(a_{-(i+1)})|$. Similarly, let $e_j, e_{-j}, e_{j'}$ and $e_{-j'}$ be the length under f of the four edges of distance j from $e_0, e_{0'}$ respectively, for $j = 1, \dots, k/2$, see Figure 1 (note that the last two edges have multiple names under this labeling, but will not be considered in our proof). Furthermore, let $E_j = e_j + e_{-j} + e_{j'} + e_{-j'}$ (for ease of notation we define $e_{-0} = e_{-0'} = 0$). We also abuse notation and write $d_j = |f(a_j) - f(a_{-j})|$.

By (3) we have that

$$d_i + d_{i+1} \leq e_0 + e_{0'} = E_0.$$

Consider now the two adjacent antipodals d_{i-1} and d_{i+2} . By the triangle inequality

$$d_{i-1} + d_{i+2} \leq (e_{-1} + d_i + e_{1'}) + (e_1 + d_{i+1} + e_{-1'}) \leq E_0 + E_1.$$

In a similar manner for any integer $0 \leq j < k/2$ we bound

$$\begin{aligned} d_{i-j} + d_{i+1+j} & \leq d_i + d_{i+1} + \sum_{h=1}^j (e_h + e_{h'} + e_{-h} + e_{-h'}) \\ & \leq E_0 + E_1 + \dots + E_j. \end{aligned}$$

So we have the following *antipodal charging inequality* (see Figure 2):

$$2L \leq 2 \sum_{j=0}^{k/2-1} (E_0 + \dots + E_j) = \sum_{j=0}^{k/2-1} (k-2j)E_j. \quad (4)$$

For simplicity assume that $\epsilon = q/k$ for some integer $q \geq 1$, then

$$\begin{aligned} 2L - k(1-2\epsilon)E & \leq \sum_{j=0}^{k/2-1} (k-2j)E_j - \sum_{j=0}^{k/2-1} (k-2q)E_j \\ & \leq 2 \sum_{j=0}^{q-1} (q-j)E_j. \end{aligned}$$

Since f has stretch at most s , $E_0 \leq 2s$ and $E_j \leq 4s$ for $j \geq 1$. Hence,

$$\begin{aligned} 2L - k(1-2\epsilon)E & \leq 4sq + 8s \sum_{j=1}^{q-1} (q-j) \\ & = 4sq + 8s(q-1)q/2 \\ & = s \cdot (2\epsilon k)^2, \end{aligned}$$

which concludes the proof of the lemma. ■

3.3. Recursive Cycle Graph

In this section we extend our techniques to prove Theorem 1. The bulk of the proof is showing that the recursive cycle graph requires high dimension for small distortion, then in Section 3.4 we exhibit a subset of n points in ℓ_1 that has the same dimension-distortion tradeoff. Let $k = 1/\epsilon$, and w.l.o.g assume that k is an even integer. The graph is defined recursively as follows: G_0 is a single edge, and for $i > 0$, G_i is obtained from G_{i-1} by replacing every edge $(u, v) \in E(G_{i-1})$ with a disjoint copy of C_{2k} , where some antipodal pair, say a_1 and a_{-1} , are associated with u, v respectively, see Figure 3. The edge (u, v) is called the *parent edge* for the cycle it induces. We call G_t the level t graph, note that the number of edges in G_t is $(2k)^t$, the number of vertices is $n = \Theta((2k)^t)$ and the diameter is $k^t \approx n^{1-1/\log k}$.

Fix some $f : G_t \rightarrow \mathbb{R}$. In what follows we define a labeling scheme for the edges and antipodals of G_t (note that this labeling depends on f), for ease of presentation, the label of a pair will also indicate its length under f . For $1 \leq i \leq t$, let \mathcal{C}_i be the collection of level i cycles (note that $|\mathcal{C}_i| = (2k)^{i-1}$). Let $A = \{-k/2, \dots, k/2, (-k/2)', \dots, (k/2)'\}$ be a set of indices. The single edge of G_0 is labeled e_\emptyset , and the rest of the edges and antipodals are labeled recursively. For level $i > 0$, fix some $C \in \mathcal{C}_i$ whose parent is an edge labeled e_x for $x \in A^{i-1}$. Recall that by Claim 4 C will have a pair of flipped antipodals under f . An edge in C is labeled by $e_{x \circ z}$ where $z \in A$ chosen by the same labeling for the single cycle case, which is determined by the flipped antipodals, and is depicted in Figure 1. The flipped antipodals will be named $d_{x \circ 0}, d_{x \circ 0'}$, and the two antipodals of distance j from the flipped ones will be named $d_{x \circ j}, d_{x \circ j'}$ (for notational convenience we let $d_{x \circ -j} = d_{x \circ -j'} = 0$ and ignore the multiple labeled edges, as these are never charged).

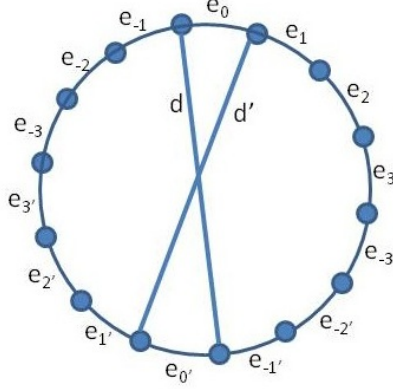


Figure 1. Notation for cycle edges, d and d' are the flipped antipodals

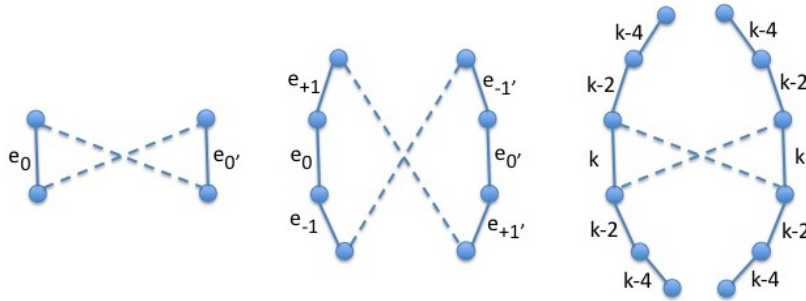


Figure 2. **Antipodal charging scheme:** in two left images the antipodals are charged to the depicted edges, right image shows the total edge charge (the dashed lines are the flipped antipodals)

Let $L_i = \sum_{x \in A^i} d_x$ be the total antipodal length in level i , and $E = \sum_{x \in A^t} e_x$ the total (level t) edge length.

Lemma 5. *Let $f : G_t \rightarrow \mathbb{R}$ be a line embedding with stretch at most s , then the following Poincaré-type inequality holds*

$$\sum_{i=1}^t 2^{t+1-i} L_i - (tk - 2t)E \leq s \cdot t(32)^t. \quad (5)$$

Once again, before attempting to prove this lemma, we show that it implies Theorem 1. Apply Lemma 2 with the appropriate coefficients: for $x \in A^t$ let $\alpha_{e_x} = tk - 2t$, and for all $1 \leq i \leq t$ and $x \in A^i$ let $\beta_{d_x} = 2^{t+1-i}$ (set α and β to zero everywhere else) and $\gamma = s \cdot t(32)^t$. Note that there are $(2k)^{i-1} \cdot k$ antipodals in level i , each of length k^{t+1-i} , so the total antipodal length in level i is $2^{i-1}k^{t+1}$, and the total edge length is $(2k)^t$. Now, any embedding $F : G_t \rightarrow \ell_1$ with stretch at most s must have distortion at least

$$\frac{\sum_{x,y \in V(G_t)} \beta_{x,y} d(x,y)}{\sum_{x,y \in V(G_t)} \alpha_{x,y} d(x,y) + \gamma} = \frac{\sum_{i=1}^t 2^{t+1-i} \cdot 2^{i-1} k^{t+1}}{(tk - 2t)(2k)^t + s \cdot t(32)^t}$$

$$= \frac{t2^t k^{t+1}}{t2^t k^{t+1} - 2t(2k)^t + s \cdot t(32)^t}$$

Finally set $s = (k/16)^t$ and obtain distortion at least

$$\frac{t2^t k^{t+1}}{t2^t k^{t+1} - 2t(2k)^t + t(2k)^t} = \frac{1}{1 - 1/k} > 1 + 1/k.$$

By Claim 1 this suggests that an embedding into fewer than $(\Omega(k))^t = n^{1-O(1/\log k)}$ dimensions cannot have distortion bounded by $1 + 1/k$ (recall that $\epsilon = 1/k$). It remains to prove the main lemma.

Proof of Lemma 5:

As the inequality (5) suggests, we will charge antipodal lengths from all levels to the level t edges. The following claim will be useful.

Claim 6. *Let e_x be the length of a level i edge in G_t for some $1 \leq i \leq t$ and $x \in A^i$. Let $E(e_x) = \sum_{y \in A^{t-i}} e_{xoy}$ (this is the total length of the level t edges whose ancestor is e_x), then*

$$e_x \leq \frac{E(e_x)}{2^{t-i}}.$$

Proof: By inverse induction on i . When $i = t$ this is immediate as $E(e_x) = e_x$. Assume it holds for $i + 1$, and note that the triangle inequality suggests that $e_x \leq \frac{1}{2} \sum_{z \in A} e_{xoz}$ (in words, the length of an antipodal is at most half of the sum of all edges in a cycle). Now apply the induction hypothesis on each e_{xoz} , and note that $E(e_x) = \sum_{z \in A} E(e_{xoz})$ to obtain

$$e_x \leq \frac{1}{2} \sum_{z \in A} e_{xoz} \leq \frac{1}{2} \sum_{z \in A} \frac{E(e_{xoz})}{2^{t-(i+1)}} = \frac{E(e_x)}{2^{t-i}}.$$

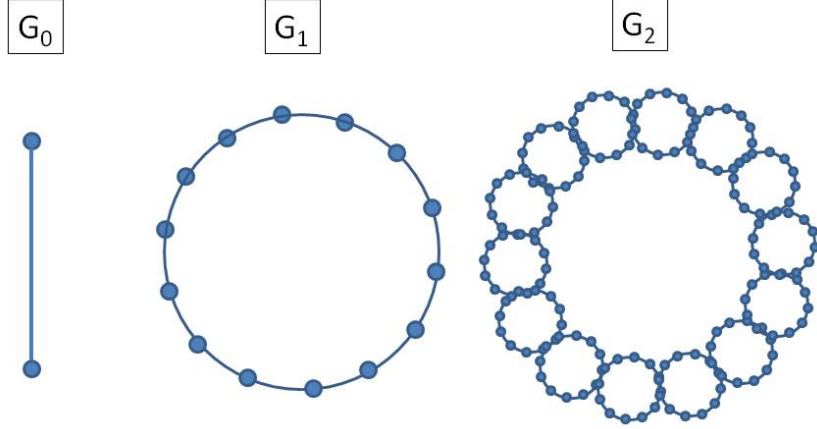


Figure 3. Recursive cycle graph, the first three levels

Fix some $i > 0$. By the antipodal charging inequality (4), for any $x \in A^{i-1}$

$$\begin{aligned}
& 2^{t+1-i} \sum_{z \in A} d_{xoz} \\
& \leq 2^{t-i} \sum_{j=0}^{k/2-1} (k-2j)(e_{xoj} + e_{xoj'} + e_{xo-j} + e_{xo-j'}) \\
& \leq \sum_{y \in A^{t-i}} \sum_{j=0}^{k/2-1} (k-2j)(e_{xojoy} + e_{xoj'o_y} + e_{xo-joy} + e_{xo-j'o_y}),
\end{aligned}$$

where the second inequality is by Claim 6 (for notational convenience we let $e_{x_0-0oy} = e_{x_0-0'o_y} = 0$ for any x, y). Now,

$$\begin{aligned}
2^{t+1-i} L_i &= 2^{t+1-i} \sum_{x \in A^{i-1}} \sum_{z \in A} d_{xoz} \\
&\leq \sum_{x \in A^{i-1}} \sum_{y \in A^{t-i}} \sum_{j=0}^{k/2-1} (k-2j)(e_{xojoy} + e_{xoj'o_y} + e_{xo-joy} + e_{xo-j'o_y}).
\end{aligned} \tag{6}$$

Next we attempt to bound $\sum_{i=1}^t 2^{t+1-i} L_i$. Consider $x \in A^t$ and an edge labeled e_x . The number of times that this edge length was charged by L_i in (6) depends only on the symbol x_i . For $z \in A$ let $|z|$ be the value of z , which is simply the number (without the minus or prime signs), and for a vector $x \in A^t$, $|x| = \sum_{i=1}^t |x_i|$. Now (6) suggests that if $|x_i| = j$ this adds $k-2j$ to the charge of e_x . We conclude that the total number of times e_x appears is exactly $\sum_{i=1}^t (k-2|x_i|) = tk - 2|x|$, in other words,

$$\sum_{i=1}^t 2^{t+1-i} L_i \leq \sum_{x \in A^t} (tk - 2|x|) e_x = tkE - 2 \sum_{x \in A^t} |x| e_x.$$

Plugging this into (5) it remains to show that

$$2 \sum_{x \in A^t} (t - |x|) e_x \leq s \cdot t(32)^t. \tag{7}$$

Notice that $0 \leq |x| \leq tk/2$, so the only edges that have a positive sign on the LHS of the inequality are those edges for which $|x| < t$, and these are the edge on which we will apply the stretch assumption, that for any edge $e_x \leq s$. Write

$$\begin{aligned}
2 \sum_{x \in A^t} (t - |x|) e_x &\leq 2 \sum_{j=0}^{t-1} \sum_{x \in A^t: |x|=j} (t-j) e_x \\
&\leq 2ts \sum_{j=0}^{t-1} \sum_{x \in A^t: |x|=j} 1,
\end{aligned}$$

and now we are left with a counting task. For each $1 \leq i \leq t$ fix some value $v_i \in \{0, 1, \dots, k/2\}$, and notice that there are at most 4^t vectors $x \in A^t$ for which $|x_i| = v_i$ for all $1 \leq i \leq t$. The number of choices for the v_i that sum exactly to j , is at most the number of ways of writing j as the sum of t non-negative integers, which is bounded by 2^{j+t} (because there are only t possible terms). It follows that

$$\sum_{j=0}^{t-1} \sum_{x \in A^t: |x|=j} 1 \leq t(4^t)^2 = t16^t,$$

and finally that

$$2 \sum_{x \in A^t} (t - |x|) e_x \leq 2ts \cdot t16^t \leq s \cdot t(32)^t,$$

where we used that $2t \leq 2^t$. This concludes the proof of (7), and thus the proof of the lemma. \blacksquare

3.4. Canonical ℓ_1 Embedding of Recursive Cycle Graph

Unfortunately, the metric induced by the recursive cycle graph is not necessarily an ℓ_1 metric. Since we are interested

in the low distortion regime, a bi-Lipschitz map of G_t into ℓ_1 will not suffice, so we would like to argue that there exists an ℓ_1 metric with the same "edges" and "antipodals" as the recursive cycle graph. This will suffice to prove the lower bound, because these are the only distances that appeared in the Poincaré-type inequality.

Definition 3. For an integer $t \geq 1$, an embedding $g_t : G_t \rightarrow \ell_1$ is called *canonical* if it satisfies the following:

- g_t is an isometry on edges and antipodals.
- For any edge $(u, v) \in E(G_t)$, there exists a single coordinate i such that $|(g_t(u) - g_t(v))_i| = 1$.

Claim 7. For any integer $t \geq 1$ there is a canonical embedding of G_t to ℓ_1 .

Proof: By induction on t , the base case for $t = 0$ is trivial. Assume we have a canonical embedding $g_{t-1} : G_{t-1} \rightarrow \ell_1$, and define g_t as follows: for every $x \in V(G_{t-1})$ let $g_t(x) = \bigoplus_{j=1}^k g_{t-1}(x)$. It remains to extend g_t to the new points. Let $C = C_{2k}$ be some level t cycle, and note that the antipodal pair a_1, a_{-1} is already fixed by g_{t-1} . Since $(a_1, a_{-1}) \in E(G_{t-1})$ and by the induction hypothesis, there is a single dimension i in which $|(g_{t-1}(a_1) - g_{t-1}(a_{-1}))_i| = 1$, so there are k dimensions I in which $|(g_t(a_1) - g_t(a_{-1}))_i| = 1$ for $i \in I$. Denote by e_1, \dots, e_k the standard basis unit vectors spanning I , and define for each $1 \leq j \leq k-1$, $g(a_{j+1}) = g(a_j) + e_j$ and $g(a_{-(j+1)}) = g(a_{-j}) - e_j$.

Note that by the construction of the recursive cycle graph, for any pair $x, y \in V(G_{t-1})$, $d_{G_t}(x, y) = k \cdot d_{G_{t-1}}(x, y)$, so the edges and antipodals that were preserved by g_{t-1} are also preserved by g_t . Furthermore, it can be checked that g_t is isometry on every level t cycle, and that edges are mapped to one dimension. ■

4. A COMBINATORIAL VIEW OF THE BRINKMAN-CHARIKAR PROOF

In this section we give an alternate view of the Linear Programming proof by Brinkman-Charikar [4] for dimension reduction in ℓ_1 in the large distortion regime using the diamond graphs.

The diamond graphs are constructed as follows. G_0 consists of a single edge of length 1. One end point of the edge is designated as the head and the other end point is designated as the tail. G_i is constructed from G_{i-1} as follows. For each edge (s, t) in G_{i-1} , where s is the head and t is the tail, we replace it with a diamond s, u, t, v with edge length 2^{-i} . s is designated as the head of the edges (s, u) and (s, v) . t is designated as the tail of the edges (u, t) and (v, t) . The other end points are designated as heads and tails accordingly. The edge (s, t) is called an edge of level $i-1$ and (v, u) is called a diagonal of level i .

Now we define some notations with respect to a line embedding $f : G_t \rightarrow \mathbb{R}$. Label the original edge in G_0 the

empty string. Consider the diamond in G_i corresponding to the edge (s, t) in G_{i-1} with label $x \in \{0, 1\}^{i-1}$. Of the two new vertices introduced by the replacement of (s, t) with a diamond, let u be the vertex with smaller f value and v be the vertex with larger f value i.e. $f(u) \leq f(v)$. The difference $f(s) - f(t)$ is called the directed length of the edge (s, t) with head s and tail t . The absolute value of the directed length of (s, t) is the length of (s, t) . The absolute difference $|f(u) - f(v)|$ is called the diagonal length of the diamond s, u, t, v . Label the edges from s to u and from v to t with $x \circ 0$ and label the edges from s to v and from u to t with $x \circ 1$, as shown in Figure 4. Note that for any $x \in \{0, 1\}^i$, there are 2^i edges in level i with label x . Let E_i be the sum of the directed lengths of level i edges, D_i be the sum of the lengths of level i diagonals, $E_{i,x}$ be the sum of the directed lengths of level i edges whose labels end with x , and F_i be the sum of the lengths of the level i edges. Below are some properties relating E_i, D_i , and $E_{i,x}$.

Property 1. $E_i + D_{i+1} = E_{i+1,0}$

Proof: Consider an arbitrary diamond in G_{i+1} corresponding to an edge (s, t) in G_i . Of the two new vertices introduced by the replacement of (s, t) with a diamond, let u be the vertex with smaller f value and v be the vertex with larger f value i.e. $f(u) \leq f(v)$. We have $(f(s) - f(t)) + (f(v) - f(u)) = (f(s) - f(u)) + (f(v) - f(t))$. Summing over all diamonds of G_{i+1} , we get the property. ■

Property 2. $E_{i,x} = \frac{1}{2}(E_{i+1,x \circ 0} + E_{i+1,x \circ 1})$. In particular, $E_i = \frac{1}{2}(E_{i+1,0} + E_{i+1,1}) = \frac{1}{2}E_{i+1}$.

Property 3. For any binary string x of length less than i , we have $E_{i,x} = E_{i,0 \circ x} + E_{i,1 \circ x}$

Theorem 3. Any embedding of G_t into ℓ_1 with distortion at most d requires $2^{\Omega(t/d^2)}/d^2$ dimensions.

Note that the number of vertices in G_t is $n = \Theta(4^t)$ so the theorem states that any embedding of G_t into ℓ_1 with distortion at most d requires $\frac{1}{d^2} n^{\Omega(1/d^2)}$ dimensions. Now we get to the proof.

Proof: In order to establish the lower bound, we will prove the following Poincaré-type inequality for any line embedding $f : G_t \rightarrow \mathbb{R}$ with stretch at most $s \leq 2^{\Theta(t/d^2)}/d^2$,

$$d \sum_{i=0}^{t-1} 2^{t-i} D_{i+1} - (t-2d)F_t \leq (d+1) \cdot 2^t \quad (8)$$

First, we show the inequality implies the theorem. Apply Lemma 2 with $\beta_{u,v} = d \cdot 2^{t-i}$ for all diagonals (u, v) of level $i+1$, $\alpha_{s,t} = (t-2d)$ for all edges (s, t) of level t (set β, α zero everywhere else), and $\gamma = (d+1)2^t$. Any embedding with stretch at most $2^{\Theta(t/d^2)}/d^2$ has distortion at least

$$\begin{aligned} & \frac{\sum_{x,y \in V(G_t)} \beta_{x,y} d(x,y)}{\sum_{x,y \in V(G_t)} \alpha_{x,y} d(x,y) + \gamma} \\ &= \frac{d \sum_{i=0}^{t-1} 2^{t-i} \cdot 4^i \cdot 2^{-i}}{(t-2d) \cdot 4^t \cdot 2^{-t} + (d+1) \cdot 2^t} = \frac{dt \cdot 2^t}{(t+1-d)2^t} \geq d \end{aligned}$$

By Claim 1, the theorem immediately follows.

Now we proceed to prove (8). By Property 1,

$$\sum_{i=0}^{t-1} 2^{t-i} D_{i+1} = \sum_{i=0}^{t-1} 2^{t-i} E_{i+1,0} - \sum_{i=0}^{t-1} 2^{t-i} E_i$$

By Property 2, $E_{i,x} = \frac{1}{2^{i-1}} \sum_{y \in \{0,1\}^{t-i}} E_{t,x \circ y}$. Apply this identity and Property 3 to the RHS and simplify, we get a sum $\sum_{x \in \{0,1\}^t} c_x E_{t,x}$ where c_x are coefficients depending on x . The contribution to c_x from $\sum_{i=0}^{t-1} 2^{t-i} E_{i+1,0}$ is $2(t - |x|_1)$ and from $-\sum_{i=0}^{t-1} 2^{t-i} E_i$ is $-t$. Thus, $c_x = t - 2|x|_1$. Let $W = \sum_{x \in \{0,1\}^t} m_x E_{t,x}$ where

$$m_x = \begin{cases} t - 2d & \text{if } |x|_1 \leq t/2 - t/(2d) \\ 2d - t & \text{if } |x|_1 \geq t/2 + t/(2d) \\ d(t - 2|x|_1) & \text{otherwise} \end{cases}$$

We have

$$\begin{aligned} d \sum_{i=0}^{t-1} 2^{t-i} D_{i+1} - W &= \sum_{x \in \{0,1\}^t} E_{t,x} (dc_x - m_x) \\ &= \sum_{x \in \{0,1\}^t: |\frac{t}{2} - |x|_1| \geq \frac{t}{2d}} E_{t,x} (d(t - 2|x|_1) - m_x) \end{aligned}$$

For any $x \in \{0,1\}^t$ and $m \in \mathbb{R}$, by the definition of stretch, $mE_{t,x} \leq |m| \cdot |E_{t,x}| \leq |m|s$, so

$$\begin{aligned} & \sum_{x \in \{0,1\}^t: |\frac{t}{2} - |x|_1| \geq \frac{t}{2d}} E_{t,x} (d(t - 2|x|_1) - m_x) \\ & \leq 2s \sum_{x \in \{0,1\}^t: |x|_1 \leq \frac{t}{2} - \frac{t}{2d}} |d(t - 2|x|_1) - (t - 2d)| \\ & \leq 2s \sum_{i=0}^{\frac{t}{2} - \frac{t}{2d}} \binom{t}{i} |(d(t - 2i) - t + 2d)| \leq sd^2(1+d)2^{t-\Theta(t/d^2)} \end{aligned}$$

The last inequality follows from a series of calculations in [4]. Notice that $|m_x| \leq t - 2d \forall x$ so $W \leq (t - 2d)F_t$. Thus, when the stretch is bounded by $s \leq 2^{\Theta(t/d^2)}/d^2$, we have

$$d \sum_{i=0}^{t-1} 2^{t-i} D_{i+1} - (t-2d)F_t \leq d \sum_{i=0}^{t-1} 2^{t-i} D_{i+1} - W \leq (d+1)2^t$$

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under CCF 0832797 and AF 0916218. Part of this work was done while A. A. was a postdoc at the Center for Computational Intractability at Princeton University. H. N. was supported by a Gordon Wu fellowship. ■

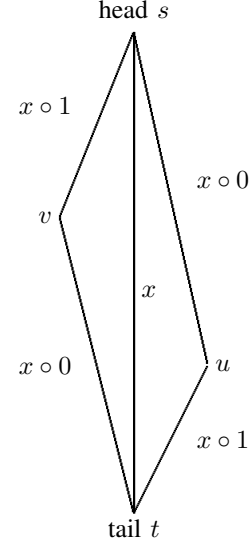


Figure 4. Labels of edges of a diamond s, u, t, v with respect to a line embedding f with $f(u) \leq f(v)$.

REFERENCES

- [1] N. Alon, “Problems and results in extremal combinatorics,” *I, Discrete Math.*, vol. 273, pp. 31–53, 2003.
- [2] K. Ball, “Markov chains, Riesz transforms and Lipschitz maps,” *Geom. Funct. Anal.*, vol. 2, no. 2, pp. 137–172, 1992.
- [3] J. D. Batson, D. A. Spielman, and N. Srivastava, “Twice-ramanujan sparsifiers,” in *STOC ’09: Proceedings of the 41st annual ACM symposium on Theory of computing*. New York, NY, USA: ACM, 2009, pp. 255–262.
- [4] B. Brinkman and M. Charikar, “On the impossibility of dimension reduction in l_1 ,” *J. ACM*, vol. 52, pp. 766–788, September 2005. [Online]. Available: <http://doi.acm.org/10.1145/1089023.1089026>
- [5] M. Charikar and A. Sahai, “Dimension reduction in the l_1 norm,” in *FOCS ’02: Proceedings of the 43rd Symposium on Foundations of Computer Science*. Washington, DC, USA: IEEE Computer Society, 2002, pp. 551–560.
- [6] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, “Locality-sensitive hashing scheme based on p -stable distributions,” in *SCG ’04: Proceedings of the twentieth annual symposium on Computational geometry*. New York, NY, USA: ACM, 2004, pp. 253–262.
- [7] P. Indyk, “Algorithmic applications of low-distortion geometric embeddings,” in *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, ser. FOCS ’01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 10–. [Online]. Available: <http://portal.acm.org/citation.cfm?id=874063.875596>
- [8] —, “Stable distributions, pseudorandom generators, embeddings, and data stream computation,” *J. ACM*, vol. 53, no. 3, pp. 307–323, 2006.
- [9] P. Indyk and R. Motwani, “Approximate nearest neighbors: towards removing the curse of dimensionality,” in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, ser. STOC ’98. New York,

- NY, USA: ACM, 1998, pp. 604–613. [Online]. Available: <http://doi.acm.org/10.1145/276698.276876>
- [10] W. B. Johnson and J. Lindenstrauss, “Extensions of Lipschitz mappings into a Hilbert space,” in *Conference in modern analysis and probability (New Haven, Conn., 1982)*. Providence, RI: Amer. Math. Soc., 1984, pp. 189–206.
- [11] J. M. Kleinberg, “Two algorithms for nearest-neighbor search in high dimensions,” in *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, ser. STOC '97. New York, NY, USA: ACM, 1997, pp. 599–608. [Online]. Available: <http://doi.acm.org/10.1145/258533.258653>
- [12] E. Kushilevitz, R. Ostrovsky, and Y. Rabani, “Efficient search for approximate nearest neighbor in high dimensional spaces,” *SIAM Journal on Computing*, vol. 30, no. 2, pp. 457–474, 2000. [Online]. Available: <http://link.aip.org/link/?SMJ/30/457/1>
- [13] A. Naor and J. R. Lee, “Embedding the diamond graph in l_p and dimension reduction in l_1 ,” *Geometric and Functional Analysis (GAFA)* 14(4), pp. 745–747, 2004.
- [14] I. Newman and Y. Rabinovich, “On cut dimension of l_1 metrics and volumes, and related sparsification techniques,” in *arxiv*. arxiv.org/abs/1002.3541, 2010.
- [15] G. Schechtman, “More on embedding subspaces of l_p in l_r ,” *Compositio Math.*, vol. 61(2), pp. 159–169, 1987.
- [16] M. Talagrand, “Embedding subspaces of l_1 into l_1^n ,” *Proceedings of the American Mathematical Society*, vol. 108, no. 2, pp. pp. 363–369, 1990. [Online]. Available: <http://www.jstor.org/stable/2048283>