A LOWER BOUND ON DIMENSION REDUCTION FOR TREES IN $\ell_1$

JAMES R. LEE AND MOHAMMAD MOHARRAMI

Abstract. There is a constant $c > 0$ such that for every $\varepsilon \in (0, 1)$ and $n \geq 1/\varepsilon^2$, the following holds. Any mapping from the $n$-point star metric into $\ell_1^d$ with bi-Lipschitz distortion $1 + \varepsilon$ requires dimension

$$d \geq \frac{c \log n}{\varepsilon^2 \log(1/\varepsilon)}.$$

1. Introduction

Consider an integer $n \geq 1$. The $n$-node star is the simple, undirected graph $G_n = (V_n, E_n)$ with $|V_n| = n$, where one node has degree $n - 1$ and all other nodes have degree one. We write $\rho_n$ for the shortest-path metric on $G_n$ where each edge is equipped with a unit weight. We use $\ell_1^d$ to denote the space $\mathbb{R}^d$ equipped with the $\ell_1$ norm. Our main theorem follows.

**Theorem 1.** There is a constant $c > 0$ such that the following holds. Consider any $\varepsilon \in (0, 1/16)$ and $n \geq 1/\varepsilon^2$. Suppose there exists a 1-Lipschitz mapping $f : V_n \to \ell_1^d$ such that $\|f(x) - f(y)\|_1 \geq (1 - \varepsilon)\rho_n(x,y)$ for all $x, y \in V_n$. Then,

$$d \geq \frac{c \log n}{\varepsilon^2 \log(1/\varepsilon)}.$$

One can achieve such a mapping with $d \leq O\left(\frac{\log n}{\varepsilon^2}\right)$, thus the theorem is tight up to the factor of $c/\log(1/\varepsilon)$. In general, de Mesmay and the authors [11] proved that every $n$-point tree metric admits a distortion $1 + \varepsilon$ embedding into $\ell_1^{C(\varepsilon)\log n}$ where $C(\varepsilon) \leq O((\frac{1}{\varepsilon})^4 \log \frac{1}{\varepsilon})$. For the special case of complete trees where all internal nodes have the same degree (such as the $n$-star), they achieve $C(\varepsilon) \leq O(\frac{1}{\varepsilon^2})$.

We recall that given two metric spaces $(X, d_X)$ and $(Y, d_Y)$ and a map $f : X \to Y$, one defines the Lipschitz constant of $f$ by

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

The bi-Lipschitz distortion of $f$ is the quantity $\text{dist}(f) = \|f\|_{\text{Lip}} : \|f^{-1}\|_{\text{Lip}}$, which is taken as infinite when $f$ is not one-to-one. If there exists such a map $f$ with distortion $D$, we say that $X$ $D$-embeds into $Y$. 

1
A finite tree metric is a finite, graph-theoretic tree $T = (V, E)$, where every edge is equipped with a positive length. The metric $d_T$ on $V$ is given by taking shortest paths. Since every finite tree metric embeds isometrically into $\ell_1$, one can view the preceding statements as quantitative bounds on the dimension required to achieve such an embedding with small distortion (instead of isometrically).

Such questions have a rich history. Perhaps most famously, if $X$ is an $n$-point subset of $\ell_2$, then a result of Johnson and Lindenstrauss [9] states that $X$ admits a $(1 + \varepsilon)$-embedding into $\ell^n_2$ where $d = O\left(\frac{\log n}{\varepsilon^2}\right)$. Alon [2] proved that this is tight up to a $\log(1/\varepsilon)$ factor: If $X \subseteq \ell^n_2$ is an orthonormal basis, then any $D$-embedding of $X$ into $\ell_1^d$ requires $d \geq \Omega(\log n)$.

The situation for finite subsets of $\ell_1$ is quite a bit more delicate. Talagrand [18], following earlier results of Bourgain-Lindenstrauss-Milman [5] and Schechtman [17], showed that every $n$-dimensional subspace $X$ of $\ell_1$ (and, in particular, every $n$-point subset) admits a $(1 + \varepsilon)$-embedding into $\ell_1^d$, with $d \leq O\left(\frac{n \log n}{\varepsilon^2}\right)$. For $n$-point subsets, this was improved to $d \leq O\left(\frac{n}{\varepsilon^2}\right)$ by Newman and Rabinovich [15], using the spectral sparsification techniques of Batson, Spielman, and Srivastava [4].

In contrast, Brinkman and Charikar [6] proved that there exist $n$-point subsets $X \subseteq \ell_1$ such that any $D$-embedding of $X$ into $\ell_1^d$ requires $d \geq n^{\Omega(1/D^2)}$ (see also [12] for a simpler argument). Thus the exponential dimension reduction achievable in the $\ell_2$ case cannot be matched for the $\ell_1$ norm. More recently, it has been show by Andoni, Charikar, Neiman, and Nguyen [3] that there exist $n$-point subsets such that any $(1 + \varepsilon)$-embedding requires dimension at least $n^{1-O(1/\log(1/\varepsilon))}$. Regev [16] has given an elegant proof of both these lower bounds based on information theoretic arguments. Our proof takes some inspiration from Regev’s approach.

We note that Theorem 1 has an analog in coding theory. Let $U_n = \{e_1, e_2, \ldots, e_n\} \subseteq \ell_1$. Then any $(1 + \varepsilon)$-embedding of $U_n$ into the Hamming cube $\{0, 1\}^d$ requires $d \geq \Omega(\log n)/\varepsilon^2\log(1/\varepsilon)$. This was proved in 1977 by McEliece, Rodemich, Rumsey, and Welch [14] using the Delsarte linear programming bound [8]. The corresponding coding question concerns the maximum number of points $x_1, x_2, \ldots \in \{0, 1\}^d$ which satisfy $(1 - \varepsilon)d/2 \leq \|x_i - x_j\|_1 \leq (1 + \varepsilon)d/2$ for $i \neq j$. Alon’s result for $\ell_2$ [2] yields this bound as a special case since $\|x - y\|_2 = \sqrt{\|x - y\|_1}$ when $x, y \in \{0, 1\}^d$.

On the one hand, the lower bound of Theorem 1 is stronger since it applies to the target space $\ell_1^d$ and not simply $\{0, 1\}^d$. On the other hand, it is somewhat weaker since embedding $U_n$ corresponds to embedding only the leaves of the star graph $G_n$, while our lower bound requires an embedding of the internal vertex as well. In fact, this is used in a fundamental and crucial way in our proof. Still, in Section 3, we prove the following somewhat weaker lower bound using simply the set $U_n$. 


Theorem 2. There is a constant $c > 0$ such that for every $\varepsilon \in (0, 1)$, for all $n$ sufficiently large, any $(1 + \varepsilon)$-embedding of $U_n \subseteq \ell_1^n$ into $\ell_1^d$ requires
$$d \geq \frac{c \log n}{\varepsilon \log \frac{1}{\varepsilon}}.$$

For the case of isometric embeddings (i.e., $\varepsilon = 0$), Alon and Pudlák [1] showed that if $U_n$ embeds isometrically in $\ell_1^n$, then $d \geq \Omega(\frac{n}{\log n})$. Our proof of Theorem 2 bears some similarity to their approach.

Finally, we mention that if $B_h$ denotes the height-$h$ complete binary tree (which has $2^{h+1} - 1$ nodes), then it was proved by Charikar and Sahai [7] that for every $h \geq 1$ and $\varepsilon > 0$, $B_h$ admits a $(1 + \varepsilon)$-embedding into $\ell_1^d$ with $d \leq O(h^2/\varepsilon^2)$. It was asked in [13] whether one could achieve $d \leq O(h/\varepsilon^2)$ and this was resolved positively in [11]. From Theorem 1, one can deduce that this upper bound is asymptotically tight up to the familiar factor of $\log(1/\varepsilon)$. This corollary is proved in Section 4.

Corollary 3. For any $\varepsilon > 0$ and $k \geq 2$, the following holds. For $h$ sufficiently large, any $(1 + \varepsilon)$-embedding of the complete $k$-ary, height-$h$ tree into $\ell_1^d$ requires
$$d \geq \frac{\Omega(h \log k)}{\varepsilon^2 \log(1/\varepsilon)}.$$

2. Proof of Theorem 1

We will first bound the number of “almost disjoint” probability measures that can be put on a finite set. Then we will translate this to a lower bound for the dimension required for embedding the $n$-star into $\ell_1^d$ with distortion $1 + \varepsilon$.

Let $X$ be a finite ground set, and let $S$ be a set of measures $X$. We say that $S$ is $\varepsilon$-unrelated if, for all distinct elements $\mu, \nu \in S$,
$$\|\mu - \nu\|_{TV} \geq \frac{1}{2}(\mu(X) + \nu(X)) - \varepsilon,$$
where $\|\cdot\|_{TV}$ denotes the total variation distance. The following lemma is an easy corollary of a fact from [16]. We include the proof here for completeness.

Lemma 4. For every $\varepsilon \in (0, 1)$ and $k \in \mathbb{N}$, if there exists a map $f : (V_n, \rho_n) \to \ell_1^k$ with distortion $1 + \varepsilon$, then there exists an $\varepsilon$-unrelated set of probability measures on $\{1, \ldots, 2k + 1\}$ of size $n - 1$.

Proof. Let $r \in V_n$ denote the the vertex of degree $n - 1$. By translation and scaling, we may assume that $f(r) = 0$ and $f$ is 1-Lipschitz. Thus for all vertices $v \in V_n$, we have $\|f(v)\|_1 \leq 1$. For each vertex $v \in V_n \setminus \{r\}$ define the measure $\mu_v$ as follows
$$\mu_v(\{i\}) = \begin{cases} \max(0, f(v)_i) & 1 \leq i \leq k \\ \max(0, -f(v)_i) & k + 1 \leq i \leq 2k \\ 1 - \|f(v)\|_1 & i = 2k + 1, \end{cases}$$
where we use $f(v)_i$ to denote the $i$th coordinate of $f(v)$.

Note that for all $u, v \in V_n \setminus \{r\}$ we have

$$\|\mu_u - \mu_v\|_{TV} = \frac{1}{2} \left( \|f(u) - f(v)\|_1 + \left| (1 - \|f(u)\|_1) - (1 - \|f(u)\|_1) \right| \right) \geq \|f(u) - f(v)\|_1.$$  

Since $f$ has distortion $1 + \varepsilon$, for any two distinct vertices $u, v \in V_n$, we have

$$\|f(u) - f(v)\|_1 \geq \left( \frac{2}{1 + \varepsilon} \right) \geq 2(1 - \varepsilon).$$

Therefore the collection $\{\mu_v : v \in V_n \setminus \{r\}\}$ satisfies the conditions of the lemma.  

The next lemma is the final ingredient that we need to prove Theorem 1. Let $M_k$ be the set of all measures $\{1, 2, \ldots, k\}$, and let $P_k$ be the set of all probability measures on $\{1, 2, \ldots, k\}$.

**Lemma 5.** There exists a universal constant $C \geq 1$ such that for $\varepsilon \leq 1/16$, the following holds. If there is an $\varepsilon$-unrelated set $S \subseteq P_k$, then there exists a $\frac{1}{2}$-unrelated set $T \subseteq P_k$ of size at least $\frac{|S|}{14}$ such that for all $\mu \in T$, we have $|\text{supp}(\mu)| \leq \lceil C\varepsilon(\varepsilon + \frac{1}{n}) \rceil$.

Before proving the lemma, we use it to finish the proof of Theorem 1.

**Proof of Theorem 1.** Suppose that there is a map from the $n$-star to $\ell_1^d$ with distortion $1 + \varepsilon$. Then by Lemma 4, there exists an $\varepsilon$-unrelated set of probability measures on $\{2d + 1\}$ of size $n - 1$. Thus by Lemma 5, there must exist a $\frac{1}{2}$-unrelated set $S$ of probability measures on $\{1, \ldots, 2d + 1\}$ of size $\Omega(n)$ such that every measure in $S$ has support size at most

$$\left\lfloor C \cdot \varepsilon \cdot \left( \varepsilon + \frac{1}{n - 1} \right) \cdot (2d + 1) \right\rfloor,$$

for some universal constant $C \geq 1$.

We now divide the problem into two cases. In the case that $C\varepsilon(\varepsilon + \frac{1}{|S|})(2d + 1) < 1$, every measure in $S$ is supported on exactly one element, therefore $|S| \leq 2d + 1$. Hence,

$$d \geq \Omega(|S|) \geq \Omega(n) \geq \Omega\left(\frac{\log n}{\varepsilon^2 \log(1/\varepsilon)}\right),$$

where we have used the assumption that $n \geq 1/\varepsilon^2$.

In the second case, we have $C\varepsilon(\varepsilon + \frac{1}{|S|})(2d + 1) \geq 1$. Since $\frac{1}{|S|} = O(\varepsilon)$, each element $\mu \in S$ has $|\text{supp}(\mu)| \leq O(\varepsilon^2 d)$. Thus for some constant $c > 0$, there are at most $\left(\frac{2d + 1}{\varepsilon^2 d}\right) \leq \exp\left(\frac{O(\varepsilon^2 d \log(1/\varepsilon) d)}{O(\varepsilon^2 d)}\right)$ different supports of size $O(\varepsilon^2 d)$ for the measures in $S$. 


A LOWER BOUND ON DIMENSION REDUCTION FOR TREES IN $\ell_1$

Since $S$ is a $\frac{1}{2}$-unrelated set of probability measures, for any $\mu, \nu \in S$, we have

$$\|\mu - \nu\|_{TV} \geq \frac{1}{2}.$$  

In particular, if we fix a set $Q \subseteq X$, then by a simple $|Q|$-dimensional volume argument,

$$|\mu \in S : \text{supp}(\mu) \subseteq Q| \leq 3^{|Q|}.$$  

All together, we have

$$|S| \leq 3^{O(e^2d)} \cdot e^{O(e^2d \log(1/\varepsilon))} \leq e^{O(e^2d \log(1/\varepsilon))}.$$  

Hence, $d \geq \Omega\left(\frac{\log |S|}{\varepsilon^2 \log(1/\varepsilon)}\right)$, completing the proof.  

**Remark 6.** We note that there is a straightforward volume lower bound for large distortions $D \geq 1$: Any $D$-embedding of the $n$-star into $\ell_1^d$ requires $d \geq \Omega\left(\frac{\log n}{\log D}\right)$. This is simply because the maximal number of disjoint $\ell_1$ balls of radius $1/D$ that can be packed in an $\ell_1$ ball of radius 2 is $(2D)^d$ in $d$ dimensions.

We are left to prove Lemma 5. We start by recalling some simple properties of the total variation distance. For a finite set $S$ and measures $\mu, \nu : 2^S \rightarrow [0, \infty)$, we define

$$\min(\mu, \nu)(T) = \sum_{x \in T} \min\{\mu(\{x\}), \nu(\{x\})\}.$$  

For $k \in \mathbb{N}$, and measures $\mu, \nu \in M_k$, we have

$$\|\mu - \nu\|_{TV} = \frac{1}{2}\left(\mu([k]) + \nu([k])\right) - \min(\mu, \nu)([k]),$$  

where we use the notation $[k] = \{1, 2, \ldots, k\}$. We also use the following partial order on measures on the set $S$: $\mu \preceq \nu$, if and only if for all $T \subseteq S$, $\mu(T) \leq \nu(T)$. The following observation is immediate from (1).

**Observation 7.** Fix $k \in \mathbb{N}$, $\varepsilon > 0$, and measures $\mu, \nu, \mu', \nu' \in M_k$, such that $\mu' \preceq \mu$ and $\nu' \preceq \nu$. If

$$\|\mu - \nu\|_{TV} \geq \frac{1}{2}(\mu([k]) + \nu([k])) - \varepsilon,$$

then

$$\|\mu' - \nu'\|_{TV} \geq \frac{1}{2}(\mu'([k]) + \nu'([k])) - \varepsilon.$$  

We will require the following fact in the proof of Lemma 5.

**Lemma 8.** Consider $\delta \in (0, 1)$ and a finite subset $S \subseteq [0, \infty)$ such that

$$\delta \cdot (|S| - 1) \cdot \sum_{x \in S} x \geq \sum_{x,y \in S, x \neq y} \min(x, y).$$  

Then there exists a set $T \subseteq S$, such that $\sum_{x \in T} x \geq \frac{1}{2} \sum_{x \in S} x$ and $|T| \leq \lceil \delta(|S| - 1) \rceil$.  

Proof. Let \( n = |S| \), and let \( a_1 \geq \cdots \geq a_n \geq 0 \) be the elements of \( S \) in decreasing order. Then,

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} \min(a_i, a_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{\max(i,j)} = \sum_{i=1}^{n} 2(i-1) a_i.
\]

Letting \( k = \lceil \delta(|S| - 1) \rceil \), we have

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} \min(a_i, a_j) \geq \sum_{i=k+1}^{n} 2(i-1) a_i \geq 2k \sum_{i=k+1}^{n} a_i \geq 2\delta(|S| - 1) \sum_{i=k+1}^{n} a_i.
\]

Combining this inequality and (2) implies that \( \sum_{i=k+1}^{n} a_i \leq \frac{1}{2} \sum_{x \in S} x \), therefore \( \sum_{i=1}^{k} a_i \geq \frac{1}{2} \sum_{x \in S} x \). Hence the set \( T = \{a_1, \ldots, a_k\} \) satisfies both conditions of the lemma.

Proof of Lemma. We will show that each of the following statements implies the next one.

I) There exists an \( \varepsilon \)-unrelated set \( S \subseteq P_k \) of size \( n \).

II) There exists an \( \varepsilon \)-unrelated set \( S \subseteq M_k \) of size \( n \) such that
(a) for all \( \mu \in S \), \( \mu([k]) \leq 1 \);
(b) \( \sum_{\mu \in S} \mu([k]) \geq n/4 \);
(c) \( \sum_{\mu \in S} |\text{supp}(\mu)| < (2\varepsilon n + 1)k \);

III) There exists an \( \varepsilon \)-unrelated set \( S \subseteq M_k \) of size at least \( n/14 \) such that
(a) for all \( \mu \in S \), \( |\text{supp}(\mu)| < 14 \left( 2\varepsilon + \frac{1}{n} \right) k \);
(b) for all \( \mu \in S \), we have \( \mu([k]) \geq 1/8 \);

IV) There exists a set satisfying all the conditions of the lemma.

For ease of notation, given a subset \( S \subseteq M_k \), we define,

\[
\Delta_S = \sum_{\mu, \nu \in S, \mu \neq \nu} \min(\mu, \nu).
\]

Note that, if for some \( \varepsilon \in [0, 1] \), \( S \subseteq P_k \) is \( \varepsilon \)-unrelated, then (II) implies that

\[
\Delta_S([k]) \leq \sum_{\mu, \nu \in S, \mu \neq \nu} \frac{1}{2} (\mu([k]) + \nu([k])) - \|\mu - \nu\|_TV \\
\leq \sum_{\mu, \nu \in S, \mu \neq \nu} (1 - (1 - \varepsilon)) \\
= \varepsilon |S| \cdot (|S| - 1).
\]
I ⇒ II: Suppose that $S_i \subseteq \mathcal{P}_k$ is $\varepsilon$-unrelated, and let $X$ be a random variable with state space $\{1, \ldots, k\}$ such that

$$\mathbb{P}(X = i) = \frac{\sum_{\mu \in S_i} \mu\{i\}}{|S_i|}.$$  

We have

$$\mathbb{E} \left[ \frac{\Delta_{S_i}(X)}{\sum_{\mu \in S_i} \mu(X)} \right] = \frac{1}{|S_i|} \sum_{i=1}^{k} \Delta_{S_i}(\{i\}) = \frac{1}{|S_i|} \Delta_{S_i}([k]) \leq \varepsilon(|S_i| - 1),$$

Markov’s inequality implies that

$$\mathbb{P} \left( \frac{\Delta_{S_i}(X)}{\sum_{\mu \in S_i} \mu(X)} \leq 2\varepsilon(|S_i| - 1) \right) \geq \frac{1}{2}.$$  

So if we let

$$A = \left\{ i : \frac{\Delta_{S_i}(\{i\})}{\sum_{\mu \in S_i} \mu(\{i\})} \leq 2\varepsilon(|S_i| - 1) \right\},$$

then we have

$$\frac{1}{|S_i|} \sum_{\mu \in S_i} \mu(A) = \sum_{\mu \in S_i} \sum_{i \in A} \mathbb{P}(X = i) \geq \frac{1}{2}. \quad (4)$$

By Lemma 8, for all $i \in A$ there exists a set $W_i \subseteq S_i$ such that $|W_i| \leq \lfloor 2\varepsilon(|S_i| - 1) \rfloor$, and

$$\sum_{\mu \in W_i} \mu(\{i\}) \geq \frac{1}{2} \sum_{\mu \in S_i} \mu(\{i\}). \quad (5)$$

For $\mu \in S_i$, let $Y_{\mu} = \{ i : \mu \in W_i \}$. For any $Y \subseteq S$, define $R_Y : 2^S \rightarrow 2^S$ by $R_Y(T) = T \cap Y$ for elements of $Y$ and zero elsewhere.

Let $S_{II} = \{ \mu \circ R_{Y_{\mu}} \}_{\mu \in S_i}$. Since $\mu \circ R_{Y_{\mu}} \preceq \mu$, $S_{II}$ satisfies II(a). Furthermore, Observation 7 implies that the collection $\{ \mu \circ R_{Y_{\mu}} \}_{\mu \in S_i}$ is $\varepsilon$-unrelated. Furthermore, $S_{II}$ satisfies II(b) because

$$\sum_{\mu \in S_i} \mu(Y_{\mu}) \geq \frac{1}{2} \sum_{i \in A} \sum_{\mu \in S_i} \mu(\{i\}) = \frac{1}{2} \sum_{\mu \in S_i} \mu(A) \geq \frac{1}{4} |S_i|. $$

Finally, condition II(c) holds because

$$\sum_{\mu \in S_i} |\text{supp}(\mu \circ R_{Y_{\mu}})| \leq \sum_{i \in A} |W_i| < 2\varepsilon(|S_i| - 1)|A| + |A| \leq (2\varepsilon|S_i| + 1)k.$$  

II ⇒ III: Suppose that $S_{II} \subseteq \mathcal{M}_k$ is an $\varepsilon$-unrelated collection of cardinality $n$ satisfying all the conditions of II. We have $\max\{\mu([k])\}_{\mu \in S_{II}} \leq 1$ and $\sum_{\mu \in S_{II}} \mu([k]) \geq |S_{II}|/4$. Therefore, there exists a subcollection $S' \subseteq S_{II}$ such that for all $\mu \in S'$, we have $\mu([k]) \geq 1/8$, and

$$|S'| \geq \left( \frac{1/4 - 1/8}{1 - 1/8} \right) |S_{II}| \geq \frac{n}{7}. $$


By Markov’s inequality, there exists a collection of measures $S_{III}$ such that $|S_{III}| \geq \frac{1}{2}|S'| \geq \frac{1}{4}|S|$, where for all $\mu \in S_{III}$,

$$\text{supp}(\mu) \leq 2 \frac{\sum_{\mu \in S'} |\text{supp}(\mu)|}{|S'|} \leq 2 \frac{\sum_{\mu \in S_{III}} |\text{supp}(\mu)|}{|S'|} \leq 2 \frac{\sum_{\mu \in S_{III}} |\text{supp}(\mu)|}{|S_{III}|/7} \leq 14k \left(2\varepsilon + \frac{1}{n}\right).$$

The set $S_{III}$ has size at least $\frac{n}{14}$ and by construction satisfies conditions (a) and (b) of III.

**III $\Rightarrow$ IV:** Suppose $S_{III} \subseteq M_k$ is an $\varepsilon$-unrelated collection of cardinality at least $n/14$. For each measure $\mu \in S_{III}$, let $Z_\mu \subseteq \{1, \ldots, k\}$ be the set of $\lceil 16 \cdot (4k(2\varepsilon + \frac{1}{n})) \\rceil$ elements of $\{1, \ldots, k\}$ that has the largest measures with respect to $\mu$ (breaking ties arbitrarily). Since $\varepsilon \leq \frac{1}{8}$, for all $\mu \in S_{III}$ we have

$$\mu(Z_\mu) \geq \frac{1}{8} \left(\frac{16 \cdot (4k(2\varepsilon + \frac{1}{n}))}{14k(2\varepsilon + \frac{1}{n})}\right) = 2\varepsilon. \quad (6)$$

Let $S_{IV} = \{\frac{\mu R_{Z_\mu}}{\mu(Z_\mu)} : \mu \in S_{III}\}$. Clearly $S_{IV} \subseteq P_k$, and $|S_{IV}| \geq \frac{n}{14}$. Moreover, by our construction for all $\bar{\mu} \in S_{IV}$, $|\text{supp}(\bar{\mu})| \leq \lceil 224\varepsilon (2\varepsilon + \frac{1}{n})k \rceil$. To complete the proof we need to show $S_{IV}$ is $\frac{1}{2}$-unrelated. Note that if $\mu, \nu \in S_{III}$, then Observation 7 implies that

$$\min(\nu \circ R_{Z_\nu}, \mu \circ R_{Z_\mu})([k]) \leq \frac{\mu(Z_\mu) + \nu(Z_\mu)}{2} - \|\mu \circ R_{Z_\nu} - \mu \circ R_{Z_\mu}\|_{TV} \leq \varepsilon.$$

Therefore,

$$\frac{\|\mu \circ R_{Z_\mu} - \nu \circ R_{Z_\nu}\|_{TV}}{\mu(Z_\mu) - \nu(Z_\nu)} \leq 1 - \min \left(\frac{\mu \circ R_{Z_\mu}}{\mu(Z_\mu)}, \frac{\nu \circ R_{Z_\nu}}{\nu(Z_\nu)}\right)([k]) \leq 1 - \min \left(\frac{\mu \circ R_{Z_\mu}}{2\varepsilon}, \frac{\nu \circ R_{Z_\nu}}{2\varepsilon}\right)([k]) \geq \frac{1}{2},$$

completing the proof. \qed

3. Nearly Equilateral Sets in $\ell_1^d$

We will need the following result of Kahane [10].

**Theorem 9.** For every $\varepsilon \in (0, 1)$, there exists a mapping $K_\varepsilon : \mathbb{R} \to \ell_2^d$ such that $d \leq O(1/\varepsilon)$ and the following holds: For every $x, y \in \mathbb{R}$,

$$(1 - \varepsilon)\sqrt{|x - y|} \leq \|K_\varepsilon(x) - K_\varepsilon(y)\|_2 \leq \sqrt{|x - y|}.$$

**Proof of Theorem** Suppose that $f : U_n \to \ell_1^d$ is a $(1 + \varepsilon)$-embedding scaled so that $f$ is 1-Lipschitz. Consider the mapping $g : U_n \to \ell_2^{O(d/\varepsilon)}$ given by

$$g(x) = \left(K_\varepsilon(f(x)_1), K_\varepsilon(f(x)_2), \ldots, K_\varepsilon(f(x)_d)\right),$$
where $f(x)_i$ denotes the $i$th coordinate of $f(x)$. By Theorem 9, for any $x, y \in U_n$, we have $\|g(x) - g(y)\|_2^2 \leq \|f(x) - f(y)\|_1$. On the other hand,

$$\|g(x) - g(y)\|_2^2 \geq (1 - \varepsilon)^2 \|f(x) - f(y)\|_1 \geq \frac{(1 - \varepsilon)^2}{1 + \varepsilon},$$

implying that $\|g(x) - g(y)\|_2 \geq (1 - \varepsilon)/\sqrt{1 + \varepsilon} \geq 1 - 2\varepsilon$. Thus $g$ is a $(1 + 2\varepsilon)$-embedding of $U_n$ into $\ell_2^{O(d/\varepsilon)}$. But now by 2, for $n$ sufficiently large, we have

$$d \geq \frac{\Omega(\log n)}{\varepsilon \log \frac{1}{\varepsilon}}.$$

$\square$

4. Extension to $k$-ary trees

We now prove Corollary 3. Combining the next lemma with Theorem 1 yields the desired result.

**Lemma 10.** For $h, k \geq 2$, let $B_{h,k}$ be a complete $k$-ary tree of height $h$. If $B_{h,k}$ admits a $(1 + \varepsilon)$-embedding into $\ell_1^d$ for some $0 \leq \varepsilon \leq \frac{1}{8}$, then the $(1 + k^{[h/2]})$-star admits a $(1 + 4\varepsilon)$-embedding into $\ell_1^d$.

**Proof.** Suppose that $f : B_{h,k} \to \ell_1^d$ is a $(1 + \varepsilon)$-embedding. We may assume, without loss, that $f$ is 1-Lipschitz. Letting $n = (1 + k^{[h/2]})$, we construct an embedding $g : V_n \to \ell_1^d$ of the $n$-star as follows.

Let $r \in V_n$ denote the vertex of degree $n - 1$. We put $g(r) = 0$. Let $S$ be the set of vertices in $B_{h,k}$ at height $[h/2]$ (we use the convention that root has height zero). For any vertex $v \in S$, pick an arbitrary leaf $x_v$ in the subtree rooted at $v$. Associate to every vertex $w \in V_n \setminus \{r\}$ a distinct element $\tilde{w} \in S$ and put

$$g(w) = \frac{f(x_{\tilde{w}}) - f(\tilde{w})}{h - [h/2]}.$$

Since $f$ is 1-Lipschitz, the same holds for $g$. Moreover for any two distinct elements $u, v \in S$ we have

$$2(h - [h/2]) + d_{B_{h,k}}(u, v) = d_{B_{h,k}}(u, v) + d_{B_{h,k}}(x_v, v) + d_{B_{h,k}}(x_u, u)$$

$$\leq (1 + \varepsilon)\|f(x_u) - f(x_v)\|_1$$

$$\leq (1 + \varepsilon)\|(f(x_u) - f(u)) - (f(x_v) - f(v))\|_1$$

$$+ (1 + \varepsilon)\|f(u) - f(v)\|_1$$

$$\leq (1 + \varepsilon)\|(f(x_u) - f(u)) - (f(x_v) - f(v))\|_1$$

$$+ (1 + \varepsilon) d_{B_{h,k}}(u, v).$$
Therefore,

\[ (1 + \varepsilon)\| (f(x_u) - f(u)) - (f(x_v) - f(v)) \|_1 \geq 2(h - \lceil h/2 \rceil) - \varepsilon d_{B_{h,k}}(u, v) \]

\[ \geq 2(h - \lceil h/2 \rceil) - 2\varepsilon \lceil h/2 \rceil \]

\[ \geq 2(h - \lceil h/2 \rceil) - 4\varepsilon (h - \lceil h/2 \rceil) \]

\[ \geq (2 - 4\varepsilon)(h - \lceil h/2 \rceil). \]

Since \( \varepsilon \leq 1/8 \), the preceding inequality bounds the distortion of \( g \) by \( \frac{1+\varepsilon}{1-2\varepsilon} \leq 1 + 4\varepsilon \), completing the proof. \( \square \)

Acknowledgments. We are grateful to Jiri Matoušek for suggesting the approach of Section 3. This research was partially supported by NSF grant CCF-1217256.

References

[1] N. Alon and P. Pudlák. Equilateral sets in \( l^n_p \). Geom. Funct. Anal., 13(3):467–482, 2003.

[2] Noga Alon. Problems and results in extremal combinatorics. I. Discrete Math., 273(1-3):31–53, 2003. EuroComb’01 (Barcelona).

[3] A. Andoni, M. Charikar, O. Neiman, and H. L. Nguyen. Near linear lower bound for dimension reduction in \( L_1 \). To appear, Proceedings of the 52nd Annual IEEE Conference on Foundations of Computer Science, 2011.

[4] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. In Proceedings of the 41st Annual ACM Symposium on Theory of Computing, pages 255–262, 2009.

[5] J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. Acta Math., 162(1-2):73–141, 1989.

[6] Bo Brinkman and Moses Charikar. On the impossibility of dimension reduction in \( \ell_1 \). J. ACM, 52(5):766–788, 2005.

[7] Moses Charikar and Amit Sahai. Dimension reduction in the \( \ell_1 \) norm. In 39rd Annual Symposium on Foundations of Computer Science, 2002.

[8] P. Delsarte. An algebraic approach to the association schemes of coding theory. Philips Res. Rep. Suppl., (10):vi+i+97, 1973.

[9] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In Conference in modern analysis and probability (New Haven, Conn., 1982), volume 26 of Contemp. Math., pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[10] Jean-Pierre Kahane. Hélices et quasi-hélices. In Mathematical analysis and applications, Part B, volume 7 of Adv. in Math. Suppl. Stud., pages 417–433. Academic Press, New York, 1981.

[11] J. R. Lee, A. de Mesmay, and M. Moharrami. Dimension reduction for finite trees in \( \ell_1 \). Prelim. version in SODA 2012. Preprint: arXiv:1108.2290, 2011.

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

[13] J. Matoušek. Open problems on low-distortion embeddings of finite metric spaces. Online: http://kam.mff.cuni.cz/~matousek/metrop.ps.

[14] Robert J. McEliece, Eugene R. Rodemich, Howard Rumsey, Jr., and Lloyd R. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. IEEE Trans. Information Theory, IT-23(2):157–166, 1977.
[15] Ilan Newman and Yuri Rabinovich. On cut dimension of $\ell_1$ metrics and volumes, and related sparsification techniques. CoRR, abs/1002.3541, 2010. Prelim. version in SODA 2012.

[16] Oded Regev. Entropy-based bounds on dimension reduction in $L_1$. arXiv:1108.1283, 2011.

[17] Gideon Schechtman. More on embedding subspaces of $L_p$ in $l_1^n$. Compositio Math., 61(2):159–169, 1987.

[18] Michel Talagrand. Embedding subspaces of $L_1$ into $l_1^n$. Proc. Amer. Math. Soc., 108(2):363–369, 1990.