Trees and Markov convexity

[Extended Abstract]

James R. Lee Assaf Naor Yuval Peres

Abstract

We give combinatorial, geometric, and probabilistic characterizations of the distortion of tree metrics into L_p spaces. This requires the development of new embedding techniques, as well as a method for proving distortion lower bounds which is based on the wandering of Markov chains in Banach spaces, and a new metric invariant we call *Markov convexity.* Trees are thus the first non-trivial class of metric spaces for which one can give a simple and complete characterization of their distortion into a Hilbert space, up to universal constants. Our results also yield an efficient algorithm for constructing such embeddings.

1 Introduction

Geometric embeddings of discrete metric spaces, a topic originally studied in geometric analysis, became an integral part of theoretical computer science following the pioneering work of Linial, London, and Rabinovich [LLR95]. The use of metric embeddings and high-dimensional convex geometry in approximation algorithms, especially for partitioning problems like cuts and clustering, has grown significantly in recent years. We mention, for instance, the recent works [ARV04, AN04, KLMN04, ALN05, CGR05, FHL05, ACMM05, AMMN05]. Furthermore, the study of various bi-Lipschitz invariants has frequently sparked exciting algorithmic progress.

In the present work, we focus on the problem of embedding trees into L_p spaces for $p > 1$ (it is easy to see that every tree embeds isometrically into L_1). This problem is well-studied (and well-understood) if one is concerned with the worst-case distortion of trees in terms of their cardinality. For ease of exposition, we focus on the case $p = 2$ for the present time. In [Bou86], Bourgain shows that the complete binary tree of height k (on $n = 2^k$ points) incurs distortion $\Theta(\sqrt{\log \log n})$ for any embedding into L_2 (see also the accounts in [LS03, Mat99] for alternate proofs). Other papers [LMS98, Mat99] have studied upper bounds on the distortion in terms of n. In particular, Matousek [Mat99] shows that Bourgain's lower bound is the best-possible: Every n-point tree metric embeds into L_2 with distortion $O(\sqrt{\log \log n})$. These works also exhibit similar tight bounds for all $p \geq 1$. We also mention a result of Gupta, Krauthgamer, and Lee [GKL03] which shows that every tree which is doubling embeds into L_2 with distortion depending only upon the doubling constant.

But even the simple and natural question of which tree metrics embed into L_2 with $O(1)$ distortion has been left unanswered. In this paper, we give combinatorial, geometric, and probabilistic characterizations of the distortion of tree metrics into L_p spaces. Trees are thus the first non-trivial class of metric spaces for which one can give a simple and complete characterization of their distortion into a Hilbert space, up to universal constants. See Section 2 for a detailed discussion of our results, as well as an overview of the proof techniques.

1.1 Preliminaries

We recall that for a metric space (X, d), the parameter $c_p(X)$ represents the least distortion with which X embeds into an L_p space, i.e. it is the minimum of $\mathrm{dist}(f) = ||f||_{\mathrm{Lip}} \cdot ||f^{-1}||_{\mathrm{Lip}}$ over all bijections $f : X \to L_p$. If X embeds into a space Y with distortion D, we will say that X *D-embeds into* Y, and sometimes write $X \overset{D}{\hookrightarrow} Y$ to denote this. For the purposes of the present abstract, we will consider every tree metric T to have an associated graph $G(T) = (V, E)$ with an assignment of lengths to edges $\ell : E \to \mathbb{R}_+$. We will require that $|T| = |V|$, but this is without loss of generality [Gup01], as will be discussed in the full version. For $x, y \in V$, we let $d_T(x, y)$ be the length of the unique path from x to y in T. Finally, we sometimes write $A \lesssim B$ to mean $A \leq O(B)$.

2 Results and techniques

Recall that our goal is to understand the distortion of tree metrics into Banach spaces; in particular, we would like to understand what combinatorial and geometric properties control the distortion of trees into Hilbert space and other L_p spaces. To state our first result, we introduce the notion of *Markov convexity.*

Markov convexity, binary trees, and distortion.

Let $\{X_t\}_{t=0}^{\infty}$ be a Markov chain on a state space Ω. Given an integer $k \geq 0$ we denote by $\{\widetilde{X}_t(k)\}_{t=0}^{\infty}$ the process which equals X_t for time $t \leq k$, and evolves independently (with respect to the same transition probabilities) for time $t > k$. Observe that for $k < 0$, $\widetilde{X}_t(k) = X_t$ for all $t \geq 0$.

DEFINITION 2.1. *Let (X, d) be a metric space and $p > 0$. We shall say that (X, d) is* Markov p-convex *with constant Π if for every Markov chain $\{X_t\}_{t=0}^{\infty}$ on a state space Ω, and every $f : \Omega \to X$, we have for every $m \in \mathbb{N}$,*

$$\sum_{k=0}^{m} \sum_{t=1}^{2^m} \frac{\mathbb{E}\left[d\left(f\left(X_t\right), f(\widetilde{X}_t(t-2^k))\right)^p\right]}{2^{kp}} \leq \Pi^p \sum_{t=1}^{2^m} \mathbb{E}[d(f(X_t), f(X_{t-1}))^p].$$

The lease constant Π above is called the Markov p-convexity constant of X, and is denoted $\Pi_p(X)$. We shall say that X is Markov p-convex if $\Pi_p(X) < \infty$.

To understand this notion, recall that the chains X_t and $\widetilde{X}_t(t-2^k)$ run together for the first $t - 2^k$ steps, and then evolve independently for the remaining 2^k steps. Thus the left hand side is measuring the sum, over many "scales" $k = 0, 1, 2, \ldots$ of the average of the pth power of the normalized "drift" of the chain in X with respect to scale 2^k. We will say that X has *non-trivial Markov convexity* if X is Markov p-convex for some $p < \infty$. We note that L_2 is Markov 2-convex (see Section 4). More generally, the name comes in part from the fact that if X is a p-convex Banach space, then X is also Markov p-convex. In Bourgain's paper [Bou86], there is an implicit "non-linear" notion of uniform convexity related to the presence of complete binary trees. For the results in this paper, we require a more powerful "Markov variant," analogous to Ball's notion of Markov type [Bal92]. We defer a more detailed discussion to the full version.

For two metric spaces X and Y, we say that X *is finitely-representable in* Y if there exists a constant C such that every finite subset of X C-embeds into Y. Let B_∞ be the infinite complete binary tree. For simplicity, we first state our results only for L_2. Our first theorem follows.

THEOREM 2.1. *If T is a tree metric, then the following conditions are equivalent.*

1. *T admits a bi-Lipschitz embedding into a Hilbert space.*
2. *B_∞ is not finitely-representable in T.*
3. *T is Markov 2-convex.*
4. *T has non-trivial Markov convexity.*

We conjecture that (2) and (4) are equivalent for *any* metric space X. To understand what this theorem says for finite tree metrics, let us discuss a quantitative version (which will imply the above theorem by standard compactness arguments).

In what follows we denote the complete unweighted binary tree of height k by B_k. Given a metric space (X, d), $k \in \mathbb{N}$ and $c > 1$, we denote

$$\mathscr{B}_X(c) = \max\left\{k \in \mathbb{N} : \; B_k \overset{c}{\hookrightarrow} X\right\}.$$

THEOREM 2.2. *Let T be a finite weighted tree. Then for every $c > 1$,*

$$\frac{1}{c} \left(\log \mathscr{B}_T(c)\right)^{\frac{1}{2}} \lesssim c_2(T) \lesssim \left(\frac{c}{c-1} \cdot \mathscr{B}_T(c)\right)^{\frac{1}{2}}.$$

Thus the ability of a tree T to embed into Hilbert space is controlled by the height of binary trees which embed into T. We note that an appropriate version of the above theorem holds for embeddings into L_p for any $1 < p < \infty$. The lower bound follows immediately from Bourgain's lower bound for binary trees [Bou86], so our contribution is contained in the upper bound, whose proof is a mix of combinatorial and analytic techniques. In Section 3.1, we discuss *monotone colorings* of trees. Given a graph-theoretic tree $T = (V, E)$, a monotone coloring is a map $\chi : E \to \mathbb{N}$ such that every color class $\chi^{-1}(c)$, for $c \in \mathbb{N}$, is a monotone path in T (i.e. a subset of some root-leaf path). These types of colorings have been used previously for embeddings of trees [LMS98, Mat99, GKL03], but in our case the construction must be significantly more delicate. For instance, in [LMS98, Mat99] the authors were only interested in colorings which minimize the number of colors used (in particular, such colorings depend only on the topology of T, and do not take into account edge lengths).

Strong colorings and weak prototypes.

We say that a monotone coloring is δ-strong if, for every $x, y \in T$, half the distance from x to y is colored by classes of length at least $\delta\, d_T(x, y)$. Our proof proceeds in the following manner. In Section 3.1, we give some procedure for coloring the edges of T. If the coloring fails to be δ-strong, we recursively construct larger and larger binary trees inside T until eventually we find an embedded binary tree of height $\Omega(\log(1/\delta))$. The coloring uses a"scale selector" function, and a family $\{\mu_j\}_{j \in \mathbb{Z}}$ of weight functions (one for each scale) on connected subtrees of T. There are two key difficulties. The first involves choosing the right weight functions which allow the recursive reconstruction to take place. The second involves handling all of the scales simultaneously. (In order to construct a good embedding, one has to use the information that there are no large embedded binary trees *at every scale* of the space.)

To finish the proof of the upper bound, we use the δ-strong coloring to guide us in laying out the tree in L_2. In particular, in Section 3.1, we prove the following theorem, which is based on a variation of Matousek's argument [Mat99].

THEOREM 2.3. *If T admits a δ-strong coloring, then* $c_2(T) \le O\left(\sqrt{\log(1/\delta)}\right)$.

One might wonder whether the exponential quantitative gap in Theorem 2.2 is necessary. Certainly the lower bound cannot be improved. Somewhat surprisingly, the upper bound is also best-possible as we show in Section 4.2.

THEOREM 2.4. *There exists an infinite family of tree metrics $\{C_n\}_{n\in\mathbb{N}}$ with $\mathscr{B}_{C_n}(c) \to \infty$ as $n \to \infty$ for every $c > 1$ and such that*

$$c_2(C_n) = \Omega\left(\frac{c}{c-1}\mathscr{B}_{C_n}(c)\right)^{\frac{1}{2}} = \Omega\left(\sqrt{\log\log|C_n|}\right).$$

In other words, there are families of trees whose Euclidean distortion (in terms of their size) is the worst possible, but which are very far from complete binary trees. The height of the largest embedded binary tree is only $O(\log\log|C_n|)$, so using Bourgain's lower bound as a black box would only yield $\Omega(\sqrt{\log\log\log|C_n|})$. We refer to the family $\{C_n\}$ as "Cantor trees," since every root-leaf path in C_n resembles a (finitary version of) the "middle-thirds" Cantor set. This new lower bound is based on Markov convexity. In particular, since $\Pi_2(L_2) \le 2\sqrt{2}$ (see Section 4), we have the following.

THEOREM 2.5. *For any metric space X, $c_2(X) \ge \Omega(\Pi_2(X))$.*

In the full version we discuss applications of Markov convexity to proving lower bounds for balls in the Cayley graphs of certain groups. In a number of cases, e.g. the lamplighter group over the cycle, this provides the first known non-trivial lower bound. More relevant to the present discussion, we prove lower bounds on the distortion of certain families of trees we call *weak prototypes*. We say that a tree metric T is an (ε,δ)-*weak prototype with height ratio* R if the following conditions are satisfied.

1. Every non-leaf vertex of T has exactly two children.
2. Every root-leaf path of T is (ε,δ)-weak.
3. If h is the length of the shortest root-leaf path in T and h' is the length of the longest, then $h'/h \le R$.

Here, a path metric P is (ε,δ)-weak if at least an ε-fraction of the length of P is composed of edges of length at most $\delta\ell(P)$, where $\ell(P)$ is the length of P. We then have the following theorem whose proof is deferred to the full version. See Section 4.2 for a warmup.

THEOREM 2.6. *If T is an (ε,δ)-weak prototype with height ratio R, then*

$$\Pi_2(T) \ge \Omega\left(\sqrt{\frac{\varepsilon}{R}\log(1/\delta)}\right).$$

This theorem is proved by building an appropriate Markov chain on any given (ε,δ)-weak tree.

Tight characterizations of distortion

All the notions we have discussed so far are intimately inter-related for trees. Let $\delta^*(T) > 0$ be the largest δ for which T admits a δ-strong coloring. Combined with the previous discussion, our most precise embedding theorem establishes the following result. The proof of this result is deferred to the full version due to lack of space.

THEOREM 2.7. *Let T be a tree metric, then*

$$c_2(T) = \Theta(\Pi_2(T)) = \Theta\left(\sqrt{\log(1/\delta^*(T))}\right).$$

In other words, up to universal constants, δ-strong colorings, combined with the embedding of Theorem 2.3 are the optimal way to embed trees into L_2, and furthermore, Markov 2-convexity inequalities provide the optimal lower bounds. We state also the version for L_p.

THEOREM 2.8. *For any tree metric T,*

$$c_p(T) = \Theta(\Pi_q(T)) = \Theta\left(\log^{1/q}(1/\delta^*(T))\right),$$

where $q = 2$ for $1 < p \le 2$, and $q = p$ for $p \ge 2$.

The proof of Theorem 2.7 is similar in spirit to that of Theorem 2.2. Using a more sophisticated family of weight functions $\{\rho_j\}_{j\in\mathbb{Z}}$, we construct a coloring of T. If the coloring fails to be δ-strong, then we find an embedded copy of some $(\Omega(1), O(\delta))$-weak prototype tree, and apply Theorem 2.6. The problem is that, unlike complete binary trees, weak prototypes can have a highly non-recursive structure (in particular, subtrees of weak prototypes are not themselves weak prototypes). This makes "recursive reconstruction" a more delicate process. Recall that the weight functions $\{\rho_j\}$ are defined on subtrees H of T. To handle the non-recursive nature of weak prototypes, the new weight functions look not only "down" the subtree but also use information "from above," so to speak.

3 Binary trees and distortion

Let $T = (V, E)$ be a *finite* graph-theoretic tree with positive edge lengths $\ell : E \to (0,\infty)$, and let d_T be the induced path metric on V. We also fix some arbitrary root $r \in T$. We now proceed to the proof of Theorem 2.2.

3.1 Coloring based upper bounds A *monotone path* in T is a connected subset (also called a *segment* in what follows) of some root-leaf path. By an *edge-coloring of* T, we mean a map $\chi : E \to \mathbb{Z}$. We say that the coloring is *monotone* if every color class $\chi^{-1}(m)$, for $m \in \mathbb{Z}$, is a monotone path. For $u, v \in V$ we let $P(u,v) \subseteq E$ denote the unique path from u to v, and set $P(v) = P(v,r)$. Given an edge coloring $\chi : E \to \mathbb{Z}$, $k \in \chi(E)$, and $u, v \in V$, we write

$$\ell_k^\chi(u,v) := \sum_{\substack{e \in P(u,v)\\ \chi(e)=k}} \ell(e).$$

We also set $\ell_k^\chi(v) = \ell_k^\chi(v,r)$. Finally, given $u, v \in V$ we let $\mathbf{lca}(u,v)$ denote the *least common ancestor* of u and v in T.

DEFINITION 3.1. (ε-GOOD COLORING) *We say that a coloring $\chi : E \to \mathbb{Z}$ is ε-good if it is monotone, and for every $u, v \in T$, the unique path from u to v contains a monochromatic segment of length at least $\varepsilon d_T(u,v)$. We define $\varepsilon^*(T)$ to be the largest ε for which T admits an ε-good coloring.*

To get tighter control on the Euclidean distortion of trees we also introduce the notion of δ-strong colorings.

DEFINITION 3.2. (δ-STRONG COLORING) *We say that a coloring $\chi : E \to \mathbb{Z}$ is δ-strong if it is monotone, and for every $u, v \in T$*

$$\sum_{k \in \mathbb{Z}} \ell_k^\chi(u,v) \cdot \mathbf{1}_{\{\ell_k^\chi(u,v) \ge \delta d_T(u,v)\}} \ge \tfrac{1}{2} d_T(u,v).$$

In words, we demand that at least half of the shortest path joining u and v is covered by color classes of length at least $\delta d_T(u,v)$. As before, we define $\delta^*(T)$ to be the largest δ for which T admits an δ-strong coloring.

The notions of δ-strong colorings and ε-good colorings are related via the following simple lemma whose proof we defer to the full version.

LEMMA 3.1. *Every weighted tree T satisfies $\delta^*(T) \ge 2^{-3/\varepsilon^*(T)}$.*

The relation of strong colorings to bounds on distortion is contained in the following theorem.

THEOREM 3.1. *For every weighted tree $T = (V,E)$ and $p \ge 1$,*

$$c_p(T) \lesssim \left[\log\left(\frac{1}{\delta^*(T)}\right)\right]^{\min\left\{\frac{1}{p},\frac{1}{2}\right\}}.$$

Proof. We may assume that $p \in [2,\infty)$, since if $p \in [1,2)$ the required result follows by embedding T into $\ell_2 \subseteq L_p$. Fix $\delta < \min\{\delta^*(T), 1/2\}$ and let $\chi : E \to \mathbb{Z}$ be a δ-strong coloring. Let $\{e_k\}_{k\in\mathbb{Z}}$ be a system of orthonormal vectors. For $v \in V$ we denote by $(k_1(v), \ldots, k_{m_v}(v))$ the sequence of color classes encountered on the path from the root to v. We shall denote by $d_j(v)$ the distance that the color class $k_j(v)$ contributes to the path from the root to v, i.e.

$$d_j(v) = \sum_{\substack{e \in P(v)\\ \chi(e)=k_j(v)}} \ell(e).$$

Now let

$$s_i(v) = \sum_{j=i}^{m_v} \max\left\{0, d_j(v) - \frac{\delta}{2}\sum_{h=i}^{j} d_h(v)\right\},$$

and define $f : V \to \ell_p(\mathbb{Z})$ by

$$f(v) = \sum_{i=1}^{m_v} [d_i(v)]^{1/p}[s_i(v)]^{(p-1)/p} e_{k_i(v)}.$$

We will break the proof into several claims.

CLAIM 3.1. *For all $v \in V$ and $j \in \{1, \ldots, m_v\}$,*

$$s_i(v) \ge \frac{1}{4}\sum_{j=i}^{m_v} d_j(v).$$

Proof. This is where the fact that χ is a δ-strong coloring comes in. Indeed,

$$\begin{aligned} s_i(v) &= \sum_{j=i}^{m_v} \max\left\{0, d_j(v) - \frac{\delta}{2}\sum_{h=i}^{j} d_h(v)\right\} \\ &\ge \sum_{\substack{j\in\{i,\ldots,m_v\}\\ d_j(v) \ge \delta \sum_{h=i}^{m_v} d_h(v)}} \frac{d_j(v)}{2} \ge \frac{1}{4}\sum_{j=i}^{m_v} d_j(v). \end{aligned}$$

CLAIM 3.2. $\|f\|_{\mathrm{Lip}} \le [5\log(3/\delta)]^{1/p}$.

Proof. We need to show that for every edge $(u,v) \in E$, $\|f(u) - f(v)\|_p \le 10[\log(1/\delta)]^{1/p}$. Assume that v is further than u from the root of T. In this case $k_1(u) = k_1(v), \ldots, k_{m_u}(u) = k_{m_u}(v)$ and $m_v \in \{m_u, m_u+1\}$. If $m_v = m_u + 1$ then we denote for the sake of simplicity $d_{m_v}(u) = s_{m_v}(u) = 0$. With this notation we have that

$$\begin{aligned} &\|f(u) - f(v)\|_p^p = \\ &\sum_{i=1}^{m_v}\left|[d_i(u)]^{1/p}[s_i(u)]^{(p-1)/p} - [d_i(v)]^{1/p}[s_i(v)]^{(p-1)/p}\right|^p = \\ &\left(\sum_{i=1}^{m_v-1} d_i(v)\left|[s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p}\right|^p\right) + \\ &\left|[d_{m_v}(u)]^{1/p}[s_{m_v}(u)]^{(p-1)/p} - [d_{m_v}(v)]^{1/p}[s_{m_v}(v)]^{(p-1)/p}\right|^p. \end{aligned}$$

Note that by our definitions, $s_{m_v}(u) = d_{m_v}(u)$ and $s_{m_v}(v) = d_{m_v}(v)$. Letting

$$C_i(u,v) = \left|[s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p}\right|^p,$$

we have

$$\begin{aligned} &\|f(u) - f(v)\|_p^p = \\ &\quad\sum_{i=1}^{m_v-1} d_i(v)C_i(u,v) + |d_{m_v}(u) - d_{m_v}(v)|^p \\ &\le \sum_{i=1}^{m_v-1} d_i(v)C_i(u,v) + [d_T(u,v)]^p. \end{aligned} \tag{3.1}$$

Observe that for all $i \in \{1, \ldots, m_v - 1\}$, $s_i(v) \geq s_i(u)$. Thus

$$(3.2)\quad C_i(u,v) = \left|[s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p}\right| \leq \frac{|s_i(u) - s_i(v)|}{[s_i(v)]^{1/p}},$$

where we used the elementary inequality $y^\alpha - x^\alpha \leq \frac{y-x}{y^{1-\alpha}}$, which is valid for all $y \geq x > 0$ and $\alpha \in (0,1)$.

Observe that for every $i \leq m_v - 1$,

$$\begin{aligned} s_i(v) - s_i(u) &= \max\left\{0, d_{m_v}(v) - \frac{\delta}{2}\sum_{h=i}^{m_v} d_h(v)\right\} \\ &- \max\left\{0, d_{m_v}(u) - \frac{\delta}{2}\sum_{h=i}^{m_v} d_h(u)\right\} \\ (3.3)\quad &\leq d_T(u,v). \end{aligned}$$

Thus, combining (3.2) and (3.3) we see that

$$\begin{aligned} (3.4)\quad & \sum_{i=1}^{m_v-1} d_i(v) \left|[s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p}\right|^p \\ \leq & \sum_{i=1}^{m_v-1} d_i(v) \cdot \frac{|s_i(u) - s_i(v)|^p}{s_i(v)} \\ \leq & [d_T(u,v)]^p \cdot \sum_{\substack{i\in\{1,\ldots,m_v-1\}\\ s_i(u)\neq s_i(v)}} \frac{d_i(v)}{s_i(v)} \\ \leq & 4[d_T(u,v)]^p \cdot \sum_{\substack{i\in\{1,\ldots,m_v-1\}\\ s_i(u)\neq s_i(v)}} \frac{d_i(v)}{\sum_{j=i}^{m_v} d_j(v)}, \end{aligned}$$

where in the last line we used Claim 3.1.

Observe that for every $x_1, \ldots, x_k > 0$, $\sum_{i=1}^{k} \frac{x_i}{x_i + x_{i+1} + \cdots + x_k + 1} \leq \sum_{i=k}^{1} \int_{x_{i+1}+\cdots+x_k}^{x_i+\cdots+x_k} \frac{dt}{t+1} = \int_0^{x_1+\cdots+x_k} \frac{dt}{t+1} = \log(x_1 + \cdots + x_k + 1)$. Thus

$$\begin{aligned} & \sum_{\substack{i\in\{1,\ldots,m_v-1\}\\ s_i(u)\neq s_i(v)}} \frac{d_i(v)}{\sum_{j=i}^{m_v} d_j(v)} \\ = & \sum_{\substack{i\in\{1,\ldots,m_v-1\}\\ s_i(u)\neq s_i(v)}} \frac{d_i(v)/d_{m_v}(v)}{\left(\sum_{j=i}^{m_v-1} d_j(v)/d_{m_v}(v)\right) + 1} \\ (3.5)\quad \leq & \log\left(1 + \frac{1}{d_{m_v}(v)} \sum_{\substack{i\in\{1,\ldots,m_v-1\}\\ s_i(u)\neq s_i(v)}} d_i(v)\right). \end{aligned}$$

Let i be the smallest integer in $\{1, \ldots, m_v - 1\}$ such that $s_i(u) \neq s_i(v)$. Then by the definition of $s_i(\cdot)$,

$$d_{m_v}(v) > \frac{\delta}{2}\sum_{h=i}^{m_v} d_h(v).$$

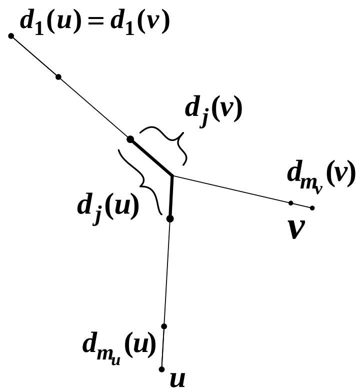

Figure 1: A schematic description of the location of u and v in the tree T. The bold segment corresponds to the color class $k_j(u) = k_j(v)$.

It follows that

$$\log\left(1 + \frac{1}{d_{m_v}(v)} \sum_{\substack{i\in\{1,\ldots,m_v-1\}\\ s_i(u)\neq s_i(v)}} d_i(v)\right) \leq \log\left(1 + \frac{\sum_{h=i}^{m_v} d_h(v)}{d_{m_v}(v)}\right) \leq \log\left(1 + \frac{2}{\delta}\right).$$

Plugging this bound into (3.5), and using (3.4) and (3.1), we get that

$$\begin{aligned} \|f(u) - f(v)\|_p &\leq \left[4\log\left(1 + \frac{2}{\delta}\right) + 1\right]^{1/p} d_T(u,v) \\ &\leq [5\log(3/\delta)]^{1/p} \cdot d_T(u,v). \end{aligned}$$

Our final claim bounds $\|f^{-1}\|_{\mathrm{Lip}}$.

CLAIM 3.3. *The embedding f is invertible, and $\|f^{-1}\|_{\mathrm{Lip}} \leq 48$.*

Proof. Fix $u, v \in V$, $u \neq v$, and let j be the integer satisfying $k_1(u) = k_1(v), \ldots, k_j(u) = k_j(v)$ and $k_{j+1}(u) \neq k_{j+1}(v)$. It follows that $d_1(u) = d_1(v), \ldots, d_{j-1}(u) = d_{j-1}(v)$, and we may assume without loss of generality that $d_j(u) \geq d_j(v)$. With this notation (see Figure 3.1 below),

$$d_T(u,v) = d_j(u) - d_j(v) + \sum_{i=j+1}^{m_u} d_i(u) + \sum_{i=j+1}^{m_v} d_i(v). \quad (3.6)$$

On the other hand,

$$\|f(u)-f(v)\|_p^p \geq \left|[d_j(u)]^{1/p}[s_j(u)]^{(p-1)/p}-[d_j(v)]^{1/p}[s_j(v)]^{(p-1)/p}\right|^p + \sum_{i=j+1}^{m_u} d_i(u)[s_i(u)]^{p-1} + \sum_{i=j+1}^{m_v} d_i(v)[s_i(v)]^{p-1}. \tag{3.7}$$

Using Claim 3.1 we see that

$$\begin{aligned} \sum_{i=j+1}^{m_u} d_i(u)[s_i(u)]^{p-1} &\geq \frac{1}{4^{p-1}}\sum_{i=j+1}^{m_u} d_i(u)\left(\sum_{h=i}^{m_u} d_h(u)\right)^{p-1} \\ &\geq \frac{1}{4^{p-1}}\sum_{i=j+1}^{m_u}\int_{d_{i+1}(u)+\cdots d_{m_u}(u)}^{d_i(u)+\cdots d_{m_u}(u)} t^{p-1}dt \\ &= \frac{1}{4^{p-1}}\int_0^{d_{j+1}(u)+\cdots d_{m_u}(u)} t^{p-1}dt \\ &= \frac{1}{p4^{p-1}}\cdot\left(\sum_{i=j+1}^{m_u} d_i(u)\right)^p. \end{aligned} \tag{3.8}$$

Similarly,

$$\sum_{i=j+1}^{m_v} d_i(v)[s_i(v)]^{p-1} \geq \frac{\left(\sum_{i=j+1}^{m_v} d_i(v)\right)^p}{p\,4^{p-1}} \tag{3.9}$$

We now consider two cases.

Case 1. $\frac{d_j(u)-d_j(v)}{2} \leq \sum_{i=j+1}^{m_v} d_i(v)$. In this case, using (3.6) we see that

$$\begin{aligned} [d_T(u,v)]^p &\leq 3^p\left(\sum_{i=j+1}^{m_u} d_i(u)+\sum_{i=j+1}^{m_v} d_i(v)\right)^p \\ &\leq 3^p\cdot 2^{p-1}\left[\left(\sum_{i=j+1}^{m_u} d_i(u)\right)^p+\left(\sum_{i=j+1}^{m_v} d_i(v)\right)^p\right] \\ &\leq p4^{p-1}\cdot 3^p\cdot 2^{p-1}\|f(u)-f(v)\|_p^p, \end{aligned}$$

where in the last inequality we plugged the bounds (3.8) and (3.9) into (3.7). Thus we get that

$$\|f(u)-f(v)\|_p \geq \frac{d_T(u,v)}{48},$$

as required.

Case 2. $\frac{d_j(u)-d_j(v)}{2} > \sum_{i=j+1}^{m_v} d_i(v)$. In this case we observe that

$$\begin{aligned} s_j(u) &= \sum_{i=j}^{m_u}\max\left\{0, d_i(u)-\frac{\delta}{2}\sum_{h=j}^{i} d_h(u)\right\} \\ &\geq \left(1-\frac{\delta}{2}\right)d_j(u), \end{aligned}$$

and similarly

$$\begin{aligned} s_j(v) &\leq \left(1-\frac{\delta}{2}\right)d_j(v)+\sum_{i=j+1}^{m_v} d_i(v) \\ &\leq \left(1-\frac{\delta}{2}\right)d_j(v)+\frac{d_j(u)-d_j(v)}{2}. \end{aligned}$$

Thus letting $C_0 = \left(1-\frac{\delta}{2}\right)^{(p-1)/p}$,

$$\begin{aligned} &[d_j(u)]^{1/p}[s_j(u)]^{(p-1)/p}-[d_j(v)]^{1/p}[s_j(v)]^{(p-1)/p} \\ &\geq C_0\left[d_j(u)-d_j(v)\cdot\left(1+\frac{d_j(u)-d_j(v)}{(2-\delta)d_j(v)}\right)^{(p-1)/p}\right] \\ &\geq C_0\left[d_j(u)-d_j(v)\cdot\left(1+\frac{d_j(u)-d_j(v)}{(2-\delta)d_j(v)}\right)\right] \\ &= \left(1-\frac{\delta}{2}\right)^{(p-1)/p}\cdot\frac{1-\delta}{2-\delta}\cdot[d_j(u)-d_j(v)] \\ &\geq \frac{d_j(u)-d_j(v)}{4}, \end{aligned}$$

where we used the fact that $\delta \leq \frac{1}{2}$. Using (3.7) and the bounds (3.8) and (3.9), it follows that

$$\begin{aligned} &\|f(u)-f(v)\|_p^p \\ &\geq \frac{[d_j(u)-d_j(v)]^p}{4^p}+\frac{\left(\sum_{i=j+1}^{m_u} d_i(u)\right)^p+\left(\sum_{i=j+1}^{m_v} d_i(v)\right)^p}{p4^{p-1}} \\ &\geq \frac{\left(d_j(u)-d_j(v)+\sum_{i=j+1}^{m_u} d_i(u)+\sum_{i=j+1}^{m_v} d_i(v)\right)^p}{p4^p\cdot 3^{p-1}} \\ &\geq \frac{1}{24^p}\cdot[d_T(u,v)]^p. \end{aligned}$$

Claim 3.3, together with Claim 3.2, concludes the proof of Theorem 3.1.

3.2 Relating coloring bounds to the containment of large binary trees

The following theorem, in conjunction with Theorem 3.1 and lemma 3.1, implies Theorem 2.2. We recall that $\varepsilon^*(T)$ is the largest value $\varepsilon > 0$ such that T admits an ε-good coloring.

THEOREM 3.2. *For every weighted tree $T = (V,E)$ and every $c > 1$,*

$$\mathscr{B}_T(c) \geq \frac{c-1}{250c}\cdot\frac{1}{\varepsilon^*(T)}.$$

Proof. We start by introducing some notation. For a vertex $v \in V$ we denote by $\mathscr{C}(v)$ the set of all children of v in T. Given $u \in \mathscr{C}(v)$ we denote by $T_u = (V_u, E_u)$ the subtree rooted at u. We also let F_u denote the tree $F_u = (V_u \cup \{v\}, E_u \cup \{(v,u)\})$, i.e. F_u is T_u plus the "incoming" edge (v,u).

Recall that $B_k = (V_k, E_k)$ is the complete binary tree of height k. Let r_k be the root of B_k, and define an auxiliary tree M_k by $M_k = (V_k \cup \{m_k\}, E_k \cup \{(m_k, r_k)\})$ (i.e. M_k is B_k with an extra incoming edge). Given a connected subtree H of T rooted at r_H, we shall say that H *admits a copy of* M_k *at scale* j if there exits a one to one mapping $f : M_k \to H$ such that

1. $f(m_k) = r_H$
2. $\|f\|_{\mathrm{Lip}} \le \frac{9c}{c-1} \cdot 4^j$ and $\|f^{-1}\|_{\mathrm{Lip}} \le \frac{c-1}{9 \cdot 4^j}$ (thus in particular $\mathsf{dist}(f) \le c$).

We define $\mu_j(H)$ to be the largest value $k \in \mathbb{N}$ for which H admits a copy of M_k at scale j, or $\mu_j(H) = -1$ if no such k exists.

We shall now define a function $g : V \to \mathbb{Z} \cup \{\infty\}$ and a coloring $\chi : E \to \mathbb{Z}$. These mappings will be constructed by induction as follows. We start by setting $g(r) = \infty$. Assume inductively that the construction is done so that whenever $v \in V$ is such that $g(v)$ is defined, if u is a vertex on the path $P(v)$ then $g(u)$ has already been defined, and for every edge $e \in E$ incident with v, $\chi(e)$ is defined.

Let $v \in V$ be a vertex closest to the root r for which $g(v)$ hasn't yet been defined. Then, by our assumption, for every $e \in P(v)$, $\chi(e)$ has been defined, and for every vertex u other than v lying on the path $P(v)$, $g(u)$ has been defined. Let $\beta_\chi(v) \subseteq V$ denote the set of *breakpoints* of χ in $P(v)$, i.e. the set of vertices on $P(v)$ for which the incoming and outgoing edges have distinct colors (for convenience, in what follows we shall also consider the root r as a breakpoint of χ). We define

$$g(v) = \max\left\{ j \in \mathbb{Z} : \forall u \in \beta_\chi(v), d_T(u,v) \ge 4^{\min\{g(u), j\}} \right\}$$

Having defined $g(v)$ we choose one of its children $w \in \mathscr{C}(v)$ for which

$$\mu_{g(v)}(F_w) = \max_{z \in \mathscr{C}(v)} \mu_{g(v)}(F_z).$$

Letting u be the father of v on the path $P(v)$, we set $\chi(v,w) = \chi(u,v)$, and we assign arbitrary new (i.e. which haven't been used before) distinct colors to each of the edges $\{(v,z)\}_{z \in \mathscr{C}(v) \setminus \{w\}}$. In other words, given the "scale" $j = g(v)$ we order the children of v according to the size of the copy of M_k which they admit beneath them at scale j. We then continue coloring with the color $\chi(u,v)$ the path $P(v)$ along the edge joining v and its child which admits the largest M_k at scale j, and color the remaining edges incident with v by arbitrary distinct new colors.

This definition clearly results in a monotone coloring χ. To motivate this somewhat complicated construction, we shall now prove some of the crucial properties of χ and g which will be used later.

CLAIM 3.4. *Let P be any monotone path in T, and let $(b_1, b_2, \dots, b_m)$ be the sequence of breakpoints along P ordered down the tree (i.e. in increasing distance from the root). Assume that $j \in \mathbb{Z}$ satisfies for every $i \in \{2, \dots, m\}$, $d_T(b_i, b_{i-1}) \le 4^j$, and assume also that $d_T(b_1, b_m) \ge \frac{30c}{c-1} \cdot 4^j$. Then there exists a subsequence of the indices $1 \le i_1 < i_2 < \dots < i_k \le m$ such that*

1. *$k \ge \frac{c-1}{20c \cdot 4^j} \cdot d_T(b_1, b_m)$.*
2. *For every $s \in \{1, \dots, k\}$, $g(b_{i_s}) = j$.*
3. *For every $s \in \{1, \dots, k-1\}$ we have $\frac{9}{c-1} \cdot 4^j \le d_T(b_{i_{s+1}}, b_{i_s}) \le \frac{9c}{c-1} \cdot 4^j$.*

Proof. We shall show that if $i \in \{1, \dots, m\}$ is such that $d_T(b_i, b_m) > \frac{4^{j+1}}{3}$ then there exists an index $t \in \{1, \dots, m\}$ with $g(b_t) = j$ and $d_T(b_t, b_i) \le 4^{j+1}$. Assuming this fact for the moment, we conclude the proof as follows. Let i_1 be the smallest integer in $\{2, \dots, m\}$ such that $g(b_{i_1}) = j$. Then $d_T(b_{i_1}, b_1) \le 4^{j+1}$. Assuming we defined $i_1 < i_2 < \dots < i_s$, if $d_T(b_{i_s}, b_m) \le \frac{9c}{c-1} \cdot 4^j$ we stop the construction, and otherwise we let t be the smallest integer bigger than i_s such that $d_T(b_t, b_{i_s}) \ge \frac{4c+5}{c-1} \cdot 4^j$. Since $d_T(b_{t-1}, b_{i_s}) < \frac{4c+5}{c-1} \cdot 4^j$, we know that $d_T(b_t, b_{i_s}) < \frac{4c+5}{c-1} \cdot 4^j + 4^j$. Thus $d_T(b_t, b_m) > d_T(b_{i_s}, b_m) - \frac{4c+5}{c-1} \cdot 4^j - 4^j > \frac{4^{j+1}}{3}$ (because we are assuming that $d_T(b_{i_s}, b_m) < \frac{9c}{c-1} \cdot 4^j$). So, there exists $i_{s+1} \in \{1, \dots, m\}$ such that $g(b_{i_{s+1}}) = j$ and $d_T(b_{i_{s+1}}, b_t) \le 4^{j+1}$. Since by construction $d_T(b_t, b_{i_s}) \ge \frac{4c+5}{c-1} \cdot 4^j > 4^{j+1}$ we deduce that $i_{s+1} > i_s$ and

$$\begin{aligned} \frac{9}{c-1} \cdot 4^j &\le d_T(b_t, b_{i_s}) - d_T(b_{i_{s+1}}, b_t) \le d_T(b_{i_{s+1}}, b_{i_s}) \\ &\le d_T(b_{i_{s+1}}, b_t) + d_T(b_t, b_{i_s}) \le \frac{9c}{c-1} \cdot 4^j. \end{aligned}$$

This construction terminates after k steps, in which case we have that

$$\begin{aligned} d_T(b_1, b_m) &= d_T(b_1, b_{i_1}) + \sum_{s=1}^{k-1} d_T(b_{i_s}, b_{i_{s+1}}) + d_T(b_{i_k}, b_m) \\ &\le 4^{j+1} + (k-1) \cdot \frac{9c}{c-1} \cdot 4^j + \frac{9c}{c-1} \cdot 4^j. \end{aligned}$$

Since $d_T(b_1, b_m) \ge \frac{30c}{c-1} \cdot 4^j$, this implies the required result.

It remains to show that if $i \in \{1, \dots, m\}$ is such that $d_T(b_i, b_m) > \frac{4^{j+1}}{3}$ then there exists $t \in \{1, \dots, m\}$ with

$g(b_t) = j$ and $d_T(b_t, b_i) \le 4^{j+1}$. We first claim that for every $i \in \{1, \dots, m\}$ there is a breakpoint $w \in \beta_\chi(b_i)$ with $g(w) \ge j$ and $d_T(w, b_i) < \frac{4^{j+1}}{3}$. Indeed, if $g(b_i) \ge j$ then there is nothing to prove, so assume that $g(b_i) < j$. By the definition of g there exists a breakpoint $w_1 \in \beta_\chi(b_i)$ such that

$$4^{\min\{g(w_1), g(b_i)\}} \le d_T(w_1, b_i) < 4^{\min\{g(w_1), g(b_i)+1\}}.$$

Thus necessarily $g(w_1) \ge g(b_i) + 1$ and $d_T(w_1, b_i) < 4^{g(b_i)+1} < 4^j$. If $g(b_i) + 1 \ge j$ then we are done by taking $w = w_1$. Otherwise, continuing in this manner we find a breakpoint $w_2 \in \beta_\chi(w_1) \subseteq \beta_\chi(b_i)$ with $g(w_2) \ge g(w_1) + 1 \ge g(b_i) + 2$ and $d_T(w_2, w_1) < 4^{g(w_1)+1}$. This procedure terminates when we find a sequence $b_i = w_0, w_1, w_2, \dots, w_t$ with $g(w_t) \ge j$, $g(w_{t-1}) < j$, and for every $0 \le s \le t-1$, $g(w_{s+1}) \ge g(w_s) + 1$ and $d_T(w_{s+1}, w_s) < 4^{g(w_s)+1}$. Thus

$$\begin{aligned} d_T(b_i, w_t) &= \sum_{s=0}^{t-1} d_T(w_{s+1}, w_s) \\ &< \sum_{s=0}^{t-1} 4^{g(w_s)+1} \sum_{s=-\infty}^{j} 4^s = \frac{4^{j+1}}{3}. \end{aligned}$$

Now, assume that $d_T(b_i, b_m) > \frac{4^{j+1}}{3}$. Let s be the largest integer in $\{i+1, \dots, m\}$ such that $d_T(b_s, b_i) \le \frac{4^{j+1}}{3}$ (such an integer s exists since $d_T(b_i, b_{i+1}) \le 4^j$). Then $\frac{4^{j+1}}{3} < d_T(b_{s+1}, b_i) \le \frac{4^{j+1}}{3} + 4^j$. By the previous argument there is a break point $w \in \beta_\chi(b_{s+1})$ with $g(w) \ge j$ and $d_T(w, b_{s+1}) < \frac{4^{j+1}}{3}$. This implies that $w = b_t$ for some $t \in \{i+1, \dots, s+1\}$, and $d_T(b_i, b_t) \le \frac{4^{j+1}}{3} + 4^j$.

We proved that as long as b_i satisfies $d_T(b_i, b_m) > \frac{4^{j+1}}{3}$, there are $1 \le t \le i \le s \le m$ such that $g(b_s) \ge j$, $g(b_t) \ge j$, and $d_T(b_t, b_i) \le \frac{4^{j+1}}{3}$, $d_T(b_s, b_i) \le \frac{4^{j+1}}{3} + 4^j$. Thus, by the definition of g,

$$4^{\min\{g(b_s), g(b_t)\}} \le d_T(b_s, b_t) \le \frac{2 \cdot 4^{j+1}}{3} + 4^j < 4^{j+1}.$$

It follows that either $g(b_s) = j$ or $g(b_t) = j$, as required.

To conclude the proof of Theorem 3.2 we may assume that $\varepsilon^*(T) < \frac{c-1}{240c}$, since otherwise the assertion of Theorem 3.2 is trivial. Fix $\frac{c-1}{240c} > \varepsilon > \varepsilon^*(T)$. By the definition of $\varepsilon^*(T)$, the coloring χ constructed above is not ε-good. Thus, there exist two vertices $u, v \in V$ such that the path $P(u, v)$ does not contain a monochromatic segment of length at least $\varepsilon d_T(u, v)$. We may assume without loss of generality that u is an ancestor of v, and let $(b_1, b_2, \dots, b_m)$ be the sequence of breakpoints along this path, enumerated down the tree (i.e. from u to v, not necessarily including u or v). Denoting $D = d_T(u, v)$ we have that $d_T(u, b_1), d_T(v, b_m), d_T(b_i, b_{i+1}) \le \varepsilon D$ for all $i \in \{1, \dots, m-1\}$. Fix $j \in \mathbb{Z}$ such that

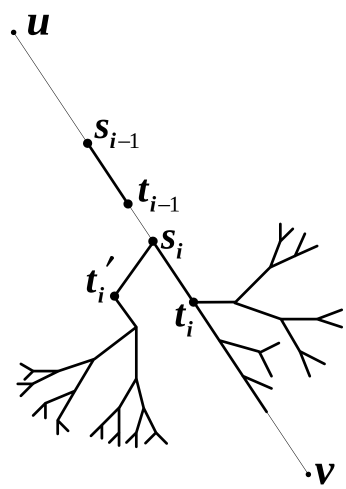

Figure 2: A schematic description of the gluing procedure in the inductive step. Because s_i was a breakpoint it must have two copies of M_{k-i-1} at scale j below it.

$\varepsilon D \le 4^j \le 4\varepsilon D$. This choice implies that $d_T(b_i, b_{i+1}) \le 4^j$ and $d_T(b_1, b_m) \ge (1 - 2\varepsilon) D \ge \frac{1-2\varepsilon}{4\varepsilon} \cdot 4^j \ge \frac{30c}{c-1} \cdot 4^j$. By Claim 3.4 there is an integer $k \ge \frac{(c-1)(1-2\varepsilon)D}{20c \cdot 4^j} \ge \frac{c-1}{250c} \cdot \frac{1}{\varepsilon} + 2$ (using the upper bound on ε) and a sequence of breakpoints $s_1, \dots, s_k$ on the path $P(u, v)$ (ordered down the tree) such that $g(s_1) = \dots = g(s_k) = j$ and for $i \in \{1, \dots, k-1\}$, $\frac{9}{c-1} \cdot 4^j \le d_T(s_i, s_{i+1}) \le \frac{9c}{c-1} \cdot 4^j$.

The proof of Theorem 3.2 will be complete once we show that $\mathscr{B}_T(c) \ge k - 2$. For $i \in \{1, \dots, k\}$ let t_i be the child of s_i along the path $P(u, v)$. We will prove by reverse induction on $i \in \{1, \dots, k-1\}$ that $\mu_j(F_{t_i}) \ge k - i - 1$, implying the required result. The base case is true, i.e. $\mu_j(F(t_{k-1}) \ge 0$, since the pair (s_{k-1}, s_k) constitutes a copy of M_0 at scale j.

Assuming that $\mu_j(F_{t_i}) \ge k - i - 1$ we shall prove that $\mu_j(F_{t_{i-1}}) \ge k - i$. Since s_i was a breakpoint, the construction of χ implies that there must be a child t_i' of s_i, *other than* t_i, for which $\mu_j(F_{t_i'}) \ge \mu_j(F_{t_i}) \ge k - i - 1$. Thus, there exist one to one mappings $f, f' : M_{k-i-1} \to T$ such that $f(m_{k-i-1}) = f'(m_{k-i-1}) = s_i$, $f(M_{k-i-1}) \subseteq F_{t_i}$, $f'(M_{k-i-1}) \subseteq F_{t_i'}$, $\|f\|_{\mathrm{Lip}}, \|f'\|_{\mathrm{Lip}} \le \frac{9c}{c-1} \cdot 4^j$, and $\|f^{-1}\|_{\mathrm{Lip}}, \|(f')^{-1}\|_{\mathrm{Lip}} \le \frac{c-1}{9 \cdot 4^j}$. Thinking of M_{k-i} as two disjoint copies of M_{k-i-1}, joined at the root m_{k-i}, we may glue f and f' to an embedding $\overline{f}$ of M_{k-i} by setting $\overline{f}(m_{k-i}) = s_{i-1}$. Since $\frac{9}{c-1} \cdot 4^j \le d_T(s_i, s_{i+1}) \le \frac{9c}{c-1} \cdot 4^j$, this results in an embedding at scale j of M_{k-i} into $F_{t_{i-1}}$, as required (see Figure 3.2).

4 Markov convexity and distortion

We start by showing that Hilbert space is Markov 2-convex. This has essentially been proved by Bourgain in [Bou86], and the following proof is a rephrasing of Bourgain's argument.

LEMMA 4.1. *For every $m \geq 0$ and every collection of points $y_0, y_1, \ldots, y_{2^m} \in L_2$,*

$$\sum_{i=1}^{2^m} \|y_i - y_{i-1}\|_2^2 = \frac{\|y_{2^m} - y_0\|_2^2}{2^m} + \sum_{k=1}^{m} \frac{1}{2^k} \sum_{j=1}^{2^{m-k}} \|y_{j2^k} - 2y_{(2j-1)2^{k-1}} + y_{(j-1)2^k}\|_2^2$$

Proof. Clearly since all the distances are squared, we may assume that $y_0, y_1, \ldots, y_{2^m} \in \mathbb{R}$. This can be proved by induction on m, however, we will prove this using Parseval's identity. Let $M = 2^m$. Consider the Haar orthonormal basis of $\mathbb{R}^M$ which is defined by the following vectors: For any $1 \leq k \leq m$ and any $1 \leq j \leq 2^{m-k}$ let $I(k\,;\,j)$ denote the set of indices $\{(j-1)2^k, \ldots, j2^k\}$ and define

$$\psi_{I(k\,;\,j)}(i) = \begin{cases} \frac{1}{2^{k/2}}, & (j-1)2^k < i \leq (2j-1)2^{k-1}\,; \\ -\frac{1}{2^{k/2}}, & (2j-1)2^{k-1} < i \leq j2^k. \end{cases}$$

Together with the vector $\psi_1 = \frac{1}{\sqrt{M}}(1, \ldots, 1)$ this gives 2^M orthonormal vectors in $\mathbb{R}^M$. Now let $z \in \mathbb{R}^M$ by $z_i = y_i - y_{i-1}$, so LHS of the lemma becomes $\sum_{i=1}^M z_i^2$ which by Parseval's identity is

$$\langle z, z\rangle = \langle z, \psi_1\rangle^2 + \sum_{k=1}^{m} \sum_{j=1}^{2^{m-k}} \langle z, \psi_{I(k\,;\,j)}\rangle^2,$$

which can easily be seen to be the RHS of the lemma.

THEOREM 4.1. *Hilbert space is Markov 2-convex. In fact, $\Pi_2(L_2) \leq 2\sqrt{2}$.*

Proof. First, we can obviously apply Lemma 4.1 to the sets of vectors $y_j, \ldots, y_{2^m}$ for $j = 0, 1, 2, \ldots, 2^m$. Summing these inequalities yields

$$\sum_{k=1}^{m} 2^{-2k} \sum_{t=2^k}^{2^m} \|y_t - 2y_{t-2^{k-1}} + y_{t-2^k}\|_2^2 \leq 4 \sum_{i=1}^{2^m} \|y_i - y_{i-1}\|_2^2.$$

Let $\{X_t\}_{t=0}^{\infty}$ be a Markov chain on a state space Ω, and take $f : \Omega \to L_2$. By the above inequality

$$4 \sum_{t=2^k}^{2^m} \mathbb{E}\|f(X_t) - f(X_{t-1})\|_2^2 \geq \sum_{k=1}^{m} 2^{-2k} \sum_{t=1}^{2^m} \mathbb{E}\|f(X_t) - 2f(X_{t-2^{k-1}}) + f(X_{t-2^k})\|_2^2$$

Observe that for every two i.i.d. random vectors $Z, Z' \in L_2$, and every constant $a \in L_2$, $\mathbb{E}\|Z - Z'\|_2^2 \leq 2\mathbb{E}\|Z - a\|_2^2$. Thus, using the fact that conditioned on $\mathcal{X} = (X_0, \ldots, X_{t-2^{k-1}})$ the random vectors $f(X_t)$ and $f(\widetilde{X}_t(t - 2^{k-1}))$ are i.i.d., we see that

$$\begin{aligned} &\mathbb{E}\|f(X_t) - f(\widetilde{X}_t(t-2^{k-1}))\|_2^2 \\ &\quad = \mathbb{E}\left(\mathbb{E}\left(\|f(X_t) - f(\widetilde{X}_t(t-2^{k-1}))\|_2^2 \middle| \mathcal{X}\right)\right) \\ &\quad \leq 2\,\mathbb{E}\|f(X_t) - 2f(X_{t-2^{k-1}}) + f(X_{(t-2^k)})\|_2^2. \end{aligned}$$

The proof is complete.

4.1 Lower bounds First, we show how Markov convexity can be used to prove Bourgain's theorem for binary trees.

LEMMA 4.2. *Let (X, d) be a Markov p-convex metric space. For every $k \in \mathbb{N}$, denote by $c_X(B_m)$ the distortion required to embed B_m into X, then*

$$c_X(B_m) \geq \frac{m^{1/p}}{\Pi_p(X)}.$$

Proof. Let $\{X_t\}_{t=0}^{\infty}$ be the forward random walk on B_m (which goes left/right each with probability 1/2), starting from the root. And assume that $f : B_m \to X$ is bi-Lipschitz. Then

$$\sum_{i=1}^{2^m} \mathbb{E}[d(f(X_i), f(X_{i-1}))]^p \leq 2^m \|f\|_{\mathrm{Lip}}^p. \tag{4.10}$$

Moreover, in the forward random walk, after splitting at time r with probability at least $\frac{1}{2}$ two independent walks will accumulate distance which is at least twice the number of steps. Thus

$$\begin{aligned} &\sum_{k=1}^{m} \sum_{t=1}^{2^m} \frac{\mathbb{E}\left[d\left(f\left(X_t\right), f\left(\widetilde{X}_t\left(t-2^k\right)\right)\right)^p\right]}{2^{kp}} \\ &\qquad \geq \frac{1}{\|f^{-1}\|_{\mathrm{Lip}}^p} \sum_{k=1}^{m} \sum_{t=1}^{2^m} \frac{1}{2^{kp}} \cdot \frac{1}{2} \cdot 2^{(k+1)p} \\ &\qquad \geq \frac{m2^m}{\|f^{-1}\|_{\mathrm{Lip}}^p}. \end{aligned} \tag{4.11}$$

Combining (4.10) and (4.11) with the definition of Markov p-convexity yields the required result.

4.2 The Cantor trees By an *unweighted spherically symmetric tree* (SST), we mean a finite graph theoretic tree T with root $r \in T$ which satisfies the following property: Let $v \in T$ be any internal node with children x, y and let T_x, T_y be the subtrees rooted at x and y; then there is a graph isomorphism from T_x to T_y which maps x to y. In words, all subtrees under a common parent node are isomorphic. We will consider only SSTs where every internal node has one or two children. In this case, the tree T is completely specified by (1) the number of nodes on a root leaf path, and (2) the subset of those nodes which have two children (i.e. the places along the path where the tree branches).

We first define a sequence of (graph-theoretic) paths P_i for $i = 0, 1, 2, \ldots$. We will set $\ell_i = \ell(P_i)$ to be the length of path P_i, and we will have for each P_i a *branching subset* $S_i \subseteq P_i$. We define P_0 as a single node and $S_0 = P_0$. Inductively, define P_{i+1} as follows: We glue end-to-end a copy of P_i, then a path of length 2^i, then another copy of P_i, where gluing two paths P and Q together means identifying the last node of P with the first node of Q. Thus $\ell_i = 2\ell_{i-1} + 2^i = i \cdot 2^i$. Furthermore, we define S_{i+1} as the set of all nodes in P_{i+1} which came from some set S_i (i.e. those nodes which are in one of the two copies of P_i and correspond to a node of S_i). Thus S_i is precisely the set of nodes which originate from P_0, and $|S_i| = 2^i$. Having defined the pairs (S_i, P_i), we now pass to the induced SST C_i (as discussed earlier). We leave the following claim as an exercise to the interested reader. Details will appear in the full version.

CLAIM 4.1. $\mathscr{B}_{C_i}(c) \le O(\frac{c-1}{c} i)$.

THEOREM 4.2. $\Pi_2(C_i) \ge \Omega(\sqrt{i})$, *hence*

$$\begin{aligned} c_2(C_i) &= \Omega(\sqrt{i}) \\ &= \Omega\left(\frac{c}{c-1}\mathscr{B}_{C_i}(c)\right)^{\frac{1}{2}} = \Omega\left(\sqrt{\log\log|C_i|}\right). \end{aligned}$$

Proof. (sketch) Recall that $\ell_i = i \cdot 2^i$ is the height of C_i, and define $\{X_t\}_{t=0}^{\infty}$ to be the forward random walk starting from the root (which walks steadily down the tree, and upon reaching a vertex with two children decides to go left/right each with probability half), with the leaves as absorbing states. Let $m = \lfloor \log_2 \ell_i \rfloor$, and note that $m = \Theta(i)$. In order to estimate $\Pi_2(C_k)$ from below, we need a bound on sums of the form $\sum_{t=1}^{2^m} 2^{-2k}\, \mathbb{E}\left[d\left(X_t, \widetilde{X}_t(t-2^k)\right)^2\right]$.

Observe that if X_t encounters a branch point between time $t - 2^{k-1}$ and time $t - 2^k$, then the chains X_t and $\widetilde{X}_t(t-2^k)$ will, with probability at least $\frac{1}{2}$, drift apart for 2^{k-1} steps, accruing distance at least 2^k. We claim that this occurs for at least an $\Omega(k/i)$ fraction of the times $t \in \{1, 2, \ldots, 2^m\}$. Assuming this is true, the above sum is at least $\Omega(2^m k/i)$. Since $\sum_{k=1}^{m} \Omega(2^m k/i) \ge \Omega(i \cdot 2^m)$, and $\sum_{t=1}^{2^m} \mathbb{E}[d(X_t, X_{t-1})^2] \le 2^m$, we conclude that $\Pi_2(C_i) \ge \Omega(\sqrt{i})$.

To verify the claim, it suffices to compute the amount of length of P_i taken up by segments of length at least 2^k which contain no branch points (i.e. points of S_i). The total length is $2^i + 2 \cdot 2^{i-1} + 4 \cdot 2^{i-2} + \ldots + 2^{i-k}2^k = (k-i) \cdot 2^i = \frac{k-i}{i}\ell(P_i)$. It follows that an $\Omega(k/i)$ fraction of P_i is composed of nodes for which there is a branch point within distance 2^{k-1}.

References

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

[ALN05] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the Sparsest Cut. To appear, STOC, 2005.

[AMMN05] N. Alon, K. Makarychev, Y. Makarychev, and A. Naor. Quadratic forms on graphs. To appear, STOC, 2005.

[AN04] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. In *STOC*. ACM, 2004.

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

[Bal92] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[Bou86] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986.

[CGR05] S. Chawla, A. Gupta, and H. Räcke. An improved approximation to sparsest cut. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, Vancouver, 2005. ACM. To appear.

[FHL05] U. Feige, M. T. Hajiaghayi, and J. R. Lee. Improved approximation algorithms for minimum-weight vertex separators. To appear, STOC, 2005.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science*, pages 534–543, 2003.

[Gup01] Anupam Gupta. Steiner points in tree metrics don't (really) help. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms (Washington, DC, 2001)*, pages 220–227, Philadelphia, PA, 2001. SIAM.

[KLMN04] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: A new embedding method for finite metrics. *Geom. Funct. Anal.*, 2004. To appear.

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

[LMS98] Nathan Linial, Avner Magen, and Michael E. Saks. Low distortion Euclidean embeddings of trees. *Israel J. Math.*, 106:339–348, 1998.

[LS03] Nathan Linial and Michael Saks. The Euclidean distortion of complete binary trees. *Discrete Comput. Geom.*, 29(1):19–21, 2003.

[Mat99] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel J. Math.*, 114:221–237, 1999.