Markov Convexity and Local Rigidity of Distorted Metrics

[Extended Abstract]*

Manor Mendel†
The Open University of Israel
mendelma@gmail.com

Assaf Naor‡
Courant Institute
naor@cims.nyu.edu

ABSTRACT

The geometry of discrete tree metrics is studied from the following perspectives:

(1) Markov p-convexity, which was shown by Lee, Naor, and Peres to be a property of p-convex Banach space, is shown here to be equivalent to p-convexity of Banach spaces.

(2) On the other hand, there exists an example of a metric space which is not Markov p-convex for any $p < \infty$, but does not uniformly contain complete binary trees. Note that the previous item implies that Banach spaces contain complete binary trees uniformly if and only if they are not Markov p-convex for any $p < \infty$.

(3) For every $B > 4$, a metric space X is constructed such that all tree metrics can be embedded in X with distortion at most B, but when large complete binary trees are embedded in X, the distortion tends to B. Therefore the class of finite tree metrics do exhibit a dichotomy in the distortions achievable when embedding them in other metric spaces. This is in contrast to the dichotomy exhibited by the class of finite subsets of L_1, and the class of all finite metric spaces.

Categories and Subject Descriptors

G.2.m [**Mathematics of Computing**]: Discrete Mathematics

*Full version will appear in `http://www.arxiv.org`

†Partially supported by an Israel Science Foundation (ISF) grant no. 221/07, and a US-Israel Bi-national Science Foundation (BSF) grant no. 2006009. Part of the research was done while visiting Microsoft Research.

‡partially supported by NSF grants CCF-0635078 and DMS-0528387 and by US-Israel Bi-national Science Foundation (BSF) grant no. 2006009. Part of the research was done while being affiliated with Microsoft Research.

SCG'08, June 9–11, 2008, College Park, Maryland, USA.

General Terms

Theory

Keywords

Uniform convexity, p-convexity, tree metrics, Markov convexity, metric dichotomy, BD-Ramsey

1. INTRODUCTION

A Banach space $(X, \|\cdot\|_X)$ is said to be finitely representable in a Banach space $(Y, \|\cdot\|_Y)$ if there exists a constant $D < \infty$ such that for every finite dimensional linear subspace $F \subseteq X$ there is a linear operator $T : F \to Y$ satisfying $\|x\|_X \le \|Tx\|_Y \le D\|x\|_X$ for all $x \in F$. In 1976 Ribe [31] proved that if two Banach spaces X and Y are uniformly homeomorphic, i.e. there is a bijection $f : X \to Y$ such that f, f^{-1} are uniformly continuous, then X is finitely representable in Y and vice versa. This remarkable theorem motivated what is known today as the "Ribe program": the search for purely metric reformulations of basic linear concepts and invariants from the local theory of Banach spaces. This program has been put forth by Bourgain in 1986 [8] (see also the discussion in [25, 26]). Since its inception the Ribe program attracted the work of many mathematicians, and led to the development of several satisfactory metric theories which extend important concepts and results of Banach space theory—we refer to [25] for a historical discussion of these topics. So far, progress on Ribe's program has come hand-in-hand with striking applications to metric geometry, group theory, functional analysis, and computer science. The present paper contains further progress in this direction—we obtain a metric characterization of p-convexity in Banach spaces, discuss some of its metric consequences, and construct unexpected counter-examples which indicate that further progress on Ribe's program can uncover non-linear phenomena which do not have analogues in Banach space theory. In doing so, we answer questions posed by Lee-Naor-Peres and Fefferman, and improve a theorem of Bates, Johnson, Lindenstrauss, Preiss and Schechtman—these results will be explained in detail below.

A Banach space $(X, \|\cdot\|_X)$ is said to be p-convex if there exists a norm $|||\cdot|||$ which is equivalent to $\|\cdot\|_X$ (i.e. for some $a, b > 0$, $a\|x\|_X \le |||x||| \le b\|x\|_X$ for all $x \in X$), and a constant $K > 0$ satisfying:

$$|||x||| = |||y||| = 1 \implies \left|\left|\left|\frac{x+y}{2}\right|\right|\right| \le 1 - K|||x-y|||^p.$$

X is called superreflexive if it is p convex for some $p < \infty$ (historically, this is not the original definition of superreflexivity[1], but it is equivalent to it due to a deep theorem of Pisier [29], which builds on important work of James [14] and Enflo [10]). Ribe's theorem implies that p-convexity, and hence also superreflexivity, is preserved under uniform homeomorphisms. The first major success in Ribe's program is a famous theorem of Bourgain [8] which obtains a metrical characterization of superreflexivity as follows. Let B_n be the complete unweighted binary tree of depth n, equipped with the natural graph-theoretical metric. Then a Banach space X is superreflexive if and only if

$$\lim_{n\to\infty} c_X(B_n) = \infty.$$

Here, and in what follows, given two metric spaces $(\mathcal{V}, d_{\mathcal{V}})$, $(\mathcal{W}, d_{\mathcal{W}})$, the parameter $c_{\mathcal{W}}(\mathcal{V})$ denotes smallest distortion with which $\mathcal{V}$ embeds into $\mathcal{W}$, i.e. the infimum over all $D > 0$ such that there exists a scaling factor $r > 0$ and a mapping $f : \mathcal{V} \to \mathcal{W}$ satisfying $rd_{\mathcal{V}}(x, y) \le d_{\mathcal{W}}(x, y) \le rDd_{\mathcal{V}}(x, y)$ for all $x, y \in \mathcal{V}$.

Bourgain's theorem characterizes superreflexivity of Banach spaces in terms of their metric structure, but it leaves open the characterization of p-convexity. The notion of p-convexity is crucial for many applications in Banach space theory and metric geometry, and it turns out that the completion of Ribe's program for p-convexity requires additional work beyond Bourgain's superreflexivity theorem. In [19] Lee, Naor and Peres defined a bi-Lipschitz invariant of metric spaces called *Markov convexity*, which is motivated by Ball's notion of Markov type [3] and Bourgain's argument in [8].

Definition 1. Let $\{X_t\}_{t=0}^{\infty}$ be a Markov chain on a state space Ω. Given an integer $k \ge 0$, we denote by $\{\widetilde{X}_t(k)\}_{t=0}^{\infty}$ the process which equals X_t for time $t \le k$, and evolves independently (with respect to the same transition probabilities) for time $t > k$. Fix $p > 0$. A metric space (X, d_X) is called *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 $m \in \mathbb{N}$,

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

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

To obtain some intuition about Definition 1, consider the standard downward random walk on the directed rooted infinite binary tree, and consider an arbitrary embedding of the states in a Markov p-convex space. By the triangle inequality, for every fixed k, the summand in the left-hand side of (1) is at most the right-hand side of (1) (with $\Pi = 1$). Equality (or almost equality) occurs, for example, when the underlying metric coincides with the tree metric of the infinite binary tree. In Markov p-convex spaces, on the other hand, the sum over k of the left-hand side of (1) is *uniformly bounded* by the right-hand side of (1), and therefore Markov p-convex spaces cannot contain the complete infinite binary tree (or even uniformly all the finite complete binary trees), similarly to p-convex Banach spaces (as was discussed above).

We refer to [19] for more information on the notion of Markov p-convexity. In particular, it is shown in [19] that the Markov 2-convexity constant of an arbitrary weighted tree T is, up to constant factors, the Euclidean distortion of T—we refer to [19] for L_p versions of this statement and their algorithmic applications. Moreover, it was shown in [19], via a modification of an argument of Bourgain [8], that if a Banach space X is p-convex then it is also Markov p-convex. It was asked in [19] if the converse is also true. Here we answer this question positively:

THEOREM 1. *A Banach space is p-convex if and only if it is Markov p-convex.*

Thus Markov p-convexity is equivalent to p-convexity in Banach spaces, completing Ribe's program in this case. Our proof of Theorem 1 is based on a renorming method of Pisier [29]. It can be viewed as a non-linear variant of Pisier's method, and it requires several subtle changes in Pisier argument in order to adapt it to a non-linear inequality such as (1).

Results similar to Theorem 1 have been obtained for the notions of type and cotype of Banach spaces (see [9, 30, 3, 27, 25, 24]), and have been used to transfer some of the linear theory to the setting of general metric spaces. This led to several applications to problems in metric geometry. Apart from the applications of Markov p-convexity that were obtained in [19], here we show that this invariant is preserved under Lipschitz quotients. The notion of Lipschitz quotient was introduced by Gromov (see section 1.25 in [12]). Given two metric spaces (X, d_X) and (Y, d_Y), a surjective mapping $f : X \to Y$ is called a Lipschitz quotient if it is Lipschitz, and it is also "Lipschitzly open" in the sense that there exists a constant $c > 0$ such that for every $x \in X$ and $r > 0$, $f(B_X(x, r)) \supseteq B_Y(f(x), r/c)$. Here we show that if (X, d_X) is Markov p-convex and (Y, d_Y) is a Lipschitz quotient of X then Y is also Markov p-convex. In [6] Bates, Johnson, Lindenstrauss, Preiss and Schechtman investigated in detail Lipschitz quotients of Banach spaces. Their results imply that if $2 \le p < q$ then L_q is not a Lipschitz quotient of L_p. Since L_p is p-convex, it is also Markov p-convex. Hence also all of its subsets are Markov p convex. But, L_q is not p-convex, so we deduce that L_q is not a Lipschitz quotient of any subset of L_p. Thus this new "invariant approach" to the result in [6] significantly extends it. Note that the method of [6] is based on a differentiation argument, and hence it crucially relies on the fact that the Lipschitz quotient mapping is defined on all of L_p and not just on an arbitrary subset of L_p.

In light of Theorem 1 it is natural to ask whether Bourgain's characterization of superreflexivity holds for general metric spaces. Namely, is it true that for any metric space X, if $\lim_{n\to\infty} c_X(B_n) = \infty$ then X is Markov p-convex for some $p < \infty$? This question was asked in [19]. Here we show that the answer is negative: the Laakso graphs (defined in Section 3.1) are not Markov p-convex for any

[1] Superreflexivity is equivalent to having an equivalent norm $|||\cdot|||$ which is uniformly convex, i.e. for every $\varepsilon \in (0, 1)$ there exists $\delta > 0$ such that if $|||x||| = |||y||| = 1$ and $|||x-y||| = \varepsilon$ then $|||x+y||| \le 2-\delta$. For finite dimensional spaces uniform convexity is equivalent to strict convexity, but this is no longer the case for infinite dimensional spaces, or when we wish to obtain dimension-free quantitative bounds.

$p < \infty$, even though they do not contain B_n with distortion uniformly bounded in n (the last assertion follows immediately from the fact that they are doubling with constant 16—see [16]). This is in sharp contrast to the situation in other metric characterizations of linear notions such as metric type and metric cotype, where it is shown in [9] that any metric space with no non-trivial metric type must contain the Hamming cubes $(\{0,1\}^n, \|\cdot\|_1)$ with distortion independent of n. An analogous result is obtained in [25] for metric spaces with no non-trivial metric cotype. Our result also implies that for $p > 2$ the n-point Laakso graph incurs distortion $\Omega\big((\log n)^{1/p}\big)$ in any embedding into L_p. The case of L_p embeddings of the Laakso graphs (which are series parallel and doubling) when $1 < p \le 2$ was already solved in [28, 15, 18, 17] using the uniform 2-convexity property of L_p, but these proofs do not extend to the case $p > 2$.

1.1 There is no dichotomy for tree metrics

Bourgain's proof of his superreflexivity characterization [8] implies that for any Banach space X either $c_X(B_n) = 1$ for all n or there exists $\alpha > 0$ such that $c_X(B_n) \ge (\log n)^\alpha$ for large enough n. Similar dichotomic results are known to hold for arbitrary metric spaces X when the binary trees B_n are replaced by the Hamming cubes $(\{0,1\}^n, \|\cdot\|_1)$ or the integer grids $(\{0,\ldots,m\}^n, \|\cdot\|_\infty)$. Namely, Bourgain, Milman and Wolfson [9] used non-linear type to prove that for any metric space X, either $c_X(\{0,1\}^n, \|\cdot\|_1) = 1$ for all n or there exists $\alpha > 0$ such that $c_X(\{0,1\}^n, \|\cdot\|_1) \ge n^\alpha$ for large enough n. A similar result for the ℓ_∞ grids $(\{0,...,m\}^n, \|\cdot\|_\infty)$ was proved by Mendel and Naor in [25] using the theory of metric cotype. These results can be viewed alternatively as follows: the Hamming metric on $\{0,1\}^n$ is *locally rigid* in the following sense. For every $D, \varepsilon > 0$ and $d \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that if ρ is *any* metric on $\{0,1\}^n$ which is D-equivalent to the ℓ_1 (Hamming) metric, then $(\{0,1\}^d, \|\cdot\|_1)$ embeds with distortion $1+\varepsilon$ into $(\{0,1\}^n, \rho)$. An analogous statement holds true for the ℓ_∞ grids. In words, any distorted metric on a large enough Hamming cube must contain a large subset which is an almost undistorted Hamming cube (and similarly for the ℓ_∞ grids).

Motivated by these results, C. Fefferman studied the case of large binary trees, and asked us whether they also possess the local rigidity property. Namely, is it true that if ρ is a metric on the infinite binary tree B_∞ which is D-equivalent to the usual graph-theoretical metric on B_∞, then for every n and $\varepsilon > 0$, B_n embeds into (B_∞, ρ) with distortion $1+\varepsilon$? Surprisingly, despite the analogies with the Banach space case, it turns out that the answer to this question is negative!

THEOREM 2. *For every $D > 4$ there exists a metric ρ on B_∞ which is D-equivalent to the usual graph-theoretical metric, yet for every $\delta > 0$ there exists $n(\delta) \in \mathbb{N}$ such that if B_n embeds into (B_∞, ρ) with distortion $D-\delta$ then $n \le n(\delta)$.*

Moreover, Bourgain's dichotomic result for embeddings of B_n into Banach spaces fails for general metric spaces:

THEOREM 3. *Let $\eta \in (0, 0.5)$, and let $s(n)$ be a non-decreasing sequence such that $s(n)/n$ is non increasing, and $4 < s(n) \le c\eta \frac{\log n}{\log\log n}$ (for a universal constant $c > 0$). Then there exists a metric space X for which*

$$(1-2\eta)s(n/40s(n)) \le c_X(B_n) \le s(n).$$

Thus, unlike the case of Banach spaces, we can get an intermediate behavior of the growth rate of $\{c_X(B_n)\}_{n=1}^\infty$ when X is a metric space. For example, there exists a metric space X for which $c_X(B_n)$ is, say, of order $\log^* n$. Our constructions in Theorem 2 and Theorem 3 are a new family of metric spaces called H-trees (H can stand of "horizontally distorted" or "Heisenberg", since they exhibit a "preferred direction" which is loosely similar to the geometry of the Heisenberg group). The proofs of these theorems are quite delicate, as they rely on structural results for H-trees, as well as quantitative versions of metric differentiation similar to Matoušek's arguments in [21].

Another corollary of the theory of metric cotype in [25] is the following dichotomy: for any family of metric spaces $\mathcal{F}$, either for every finite metric space M and every $\varepsilon > 0$ there exists $X \in \mathcal{F}$ such that $c_X(M) \le 1+\varepsilon$, or there exists $\alpha = \alpha(\mathcal{F}) > 0$ and n-point metric spaces M_n such that for all $X \in F$ we have $c_X(M_n) = \Omega\left((\log n)^\alpha\right)$. More generally, a class of finite metric spaces $\mathcal{F}$ is said to have the *dichotomy property* if for every host space H, either $\sup_{X\in\mathcal{F}} c_H(X) = 1$, or otherwise $D_N(H,\mathcal{F}) = \sup\{c_H(X) : X \in \mathcal{F},\ |X| \le N\}$, is unbounded (or increases rapidly to infinity, depending on the strength of the dichotomy). Matoušek [21] showed that several classes of finite metric spaces have the dichotomy property. Among them, finite subsets of L_p, for fixed $p \in [1,\infty]$. In particular, the case $p = \infty$ means that the class of all finite metric spaces has the dichotomy property (since L_∞ contains all the finite metric spaces isometrically). The relevance of these types of dichotomies to computer science was observed in [2], and has been expanded upon in [23]. Our results here show that there is no metric dichotomy for trees.

1.2 Reflections

Discrete trees have appeared in numerous algorithmic contexts, mainly due to their relatively simple structure on the one hand, and their expressibility on the other hand. It is natural that the geometry of their distances is also extensively studied. Relevant literature includes [8, 22, 13, 19]. The results of this paper are a continuation of these investigations, and we believe that some of them uncover unexpected phenomena.

We have demonstrated that tree metrics behave markedly differently when embedded in Banach spaces compared to embedding in general metric spaces, which is in sharp contrast to similar questions regarding Hamming cubes, and L_∞ grids.

The use of Markov convexity to prove nonexistence of Lipschitz quotient maps may also apply in spirit to prove nonexistence of weaker notions of quotient maps. In particular, we hope that this approach will eventually lead to a resolution of a question left open in [5] about the possibility of obtaining almost tight approximation algorithms for the *group Steiner tree* problem [11] on arbitrary finite metric spaces using the so called *path-distortion of multi embedding*.

An interesting question, which is raised by Theorem 1 and is also a part of the Ribe program, is finding a metric characterization of q-smoothness. A Banach space $(X, \|\cdot\|_X)$ is called q-smooth if it admits an equivalent norm $|||\cdot|||$ such that there is a constant $S > 0$ satisfying:

$$|||x||| = 1 \ \wedge\ y \in X \implies \frac{|||x+y||| + |||x-y|||}{2} \le 1 + S|||y|||^q.$$

A Banach space X is p-convex if and only if its dual space X^* is q-smooth, where $\frac{1}{p} + \frac{1}{q} = 1$ [20]. It is known that a

Banach space X is p-convex for some $p < \infty$ (i.e. superreflexive) if and only if it is q-smooth for some $q > 1$ (this follows from [14, 29]). Hence Bourgain's metric characterization of superreflexivity can be viewed as a statement about uniform smoothness as well. However, we still lack a metric characterization of the more useful notion of q-smoothness. From past experience, it is reasonable to believe that such a characterization will be interesting and useful.

2. P-CONVEXITY AND MARKOV CONVEXITY COINCIDE

In this section we prove Theorem 1, i.e. that for Banach spaces p-convexity and Markov p-convexity are the same properties. We first show that p-convexity implies Markov p-convexity, and in fact it implies a stronger inequality that is stated in Proposition 2.1 below. The slightly weaker assertion that p-convexity implies Markov p-convexity was first proved in [19], based on an argument from [8]. Our argument here is simpler and more general.

It was proved in [4] that a Banach space X is p-convex if and only if it admits an equivalent norm $\|\cdot\|$ for which there exists $K > 0$ such that for every $a, b \in X$

$$2\|a\|^p + \frac{2}{K^p}\|b\|^p \leq \|a+b\|^p + \|a-b\|^p. \tag{2}$$

Proposition 2.1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a discrete time stochastic process on a state space Ω. Denote by $\{\widetilde{X}_t(s)\}_{t\in\mathbb{Z}}$ the process such that $(\ldots, X_{s-1}, X_s) = (\ldots, \widetilde{X}_{s-1}, \widetilde{X}_s)$ while the variables $(X_{s+1}, X_{s+2}, \ldots)$ and $(\widetilde{X}_{s+1}, \widetilde{X}_{s+2}, \ldots)$ are identically distributed (but not necessarily independent). Let $(X, \|\cdot\|)$ be a Banach space whose norm satisfies (2) and fix $f : \Omega \to X$. Then,*

$$\sum_{k=0}^{m-1}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[\left\|f(X_t) - f\left(\widetilde{X}_t(t-2^k)\right)\right\|^p\right]}{2^{kp}} \leq (4K)^p \sum_{t\in\mathbb{Z}} \mathbb{E}\|f(X_t) - f(X_{t-1})\|^p.$$

Proof. The proof is based on a geometric argument which essentially iterates (2), and for lack of space is deferred to the full version of this paper. □

We next prove the more interesting direction: a Markov p-convex Banach space is also p-convex.

Theorem 4. *Let X be a Banach space which is Markov p-convex with constant Π. Then for every $\varepsilon \in (0,1)$ there exists a norm $|||\cdot|||$ on X such that for all $x, y \in X$,*

$$(1-\varepsilon)\|x\| \leq |||x||| \leq \|x\|,$$

and

$$\left|\left|\left|\frac{x+y}{2}\right|\right|\right|^p \leq \frac{|||x|||^p + |||y|||^p}{2} - \frac{1-(1-\varepsilon)^p}{4\Pi^p(p+1)} \cdot \left|\left|\left|\frac{x-y}{2}\right|\right|\right|^p.$$

Thus the norm $|||\cdot|||$ satisfies (2) with constant $K = O\left(\frac{\Pi}{\varepsilon^{1/p}}\right)$.

Proof. Recall that the fact that X is Markov p-convex with constant Π implies that for every Markov chain $\{X_t\}_{t\in\mathbb{Z}}$ with values in X we have

$$\sum_{k=0}^{m}\sum_{t=1}^{2^m} \frac{\mathbb{E}\left\|X_t - \widetilde{X}_t(t-2^k)\right\|^p}{2^{kp}} \leq \Pi^p \sum_{t=1}^{2^m} \mathbb{E}\|X_t - X_{t-1}\|^p. \tag{3}$$

For $x \in X$ we shall say that a Markov chain $\{X_t\}_{t=-\infty}^{2^m}$ is an m-admissible representation of x if $X_t = 0$ for $t \leq 0$ and $\mathbb{E}X_t = tx$ for $t \in \{1, \ldots, 2^m\}$. Fix $\varepsilon \in (0,1)$, and denote $\eta = 1-(1-\varepsilon)^p$. For every $m \in \mathbb{N}$ define

$$|||x|||_m = \inf\left\{\left(\frac{1}{2^m}\sum_{t=1}^{2^m}\mathbb{E}\|X_t - X_{t-1}\|^p - \frac{\eta}{\Pi^p}\cdot\frac{1}{2^m}\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\mathbb{E}\left\|X_t - \widetilde{X}_t(t-2^k)\right\|^p}{2^{kp}}\right)^{1/p}\right\}, \tag{4}$$

where the infimum in (4) is taken over all m-admissible representations of x. Note that such a representation of x always exists, since we can define $X_t = 0$ for $t \leq 0$ and $X_t = tx$ for $t \in \{1, \ldots, 2^m\}$. This example shows that $|||x|||_m \leq \|x\|$. On the other hand if $\{X_t\}_{t=-\infty}^{2^m}$ is an m-admissible representation of x then

$$\begin{aligned}
&\sum_{t=1}^{2^m}\mathbb{E}\|X_t - X_{t-1}\|^p - \frac{\eta}{\Pi^p}\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\mathbb{E}\left\|X_t - \widetilde{X}_t(t-2^k)\right\|^p}{2^{kp}} \\
&\geq (1-\eta)\sum_{t=1}^{2^m}\mathbb{E}\|X_t - X_{t-1}\|^p \qquad (5)\\
&\geq (1-\varepsilon)^p\sum_{t=1}^{2^m}\|\mathbb{E}X_t - \mathbb{E}X_{t-1}\|^p \qquad (6)\\
&= (1-\varepsilon)^p\sum_{t=1}^{2^m}\|tx - (t-1)x\|^p = 2^m(1-\varepsilon)^p\|x\|^p,
\end{aligned}$$

where in (5) we used (3), and in (6) we used the convexity of the function $z \mapsto \|z\|^p$. In conclusion we see that for all $x \in X$,

$$(1-\varepsilon)\|x\| \leq |||x|||_m \leq \|x\|. \tag{7}$$

Now take $x, y \in X$ and fix $\delta \in (0,1)$. Let $\{X_t\}_{t=-\infty}^{2^m}$ be an admissible representation on x and $\{Y_t\}_{t=-\infty}^{2^m}$ be an admissible representation of y which is stochastically independent of $\{X_t\}_{t\in\mathbb{Z}}$, such that

$$\sum_{t=1}^{2^m}\mathbb{E}\|X_t - X_{t-1}\|^p - \frac{\eta}{\Pi^p}\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\mathbb{E}\left\|X_t - \widetilde{X}_t(t-2^k)\right\|^p}{2^{kp}} \leq 2^m\left(|||x|||_m^p + \delta\right), \tag{8}$$

and

$$\sum_{t=1}^{2^m}\mathbb{E}\|Y_t - Y_{t-1}\|^p - \frac{\eta}{\Pi^p}\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\mathbb{E}\left\|Y_t - \widetilde{Y}_t(t-2^k)\right\|^p}{2^{kp}} \leq 2^m\left(|||y|||_m^p + \delta\right). \tag{9}$$

Define a Markov chain $\{Z_t\}_{t=-\infty}^{2^{m+1}} \subseteq X$ as follows. For $t \leq -2^m$ set $Z_t = 0$ while with probability $\frac{1}{2}$ we let $(Z_{-2^m+1}, Z_{-2^m+2}, \ldots, Z_{2^{m+1}})$ equal

$$\Big(\underbrace{0, \ldots, 0}_{2^m \text{ times}}, X_1, X_2, \ldots, X_{2^m}, X_{2^m}+Y_1, X_{2^m}+Y_2, \ldots, X_{2^m}+Y_{2^m}\Big),$$

and with probability $\frac{1}{2}$ we let $(Z_{-2^m+1}, Z_{-2^m}, \ldots, Z_{2^{m+1}})$ equal

$$\Big(\underbrace{0, \ldots, 0}_{2^m \text{ times}}, Y_1, Y_2, \ldots, Y_{2^m}, X_1+Y_{2^m}, X_2+Y_{2^m}, \ldots, X_{2^m}+Y_{2^m}\Big).$$

For brevity, $\{Z_t\}$ was defined somewhat informally. It can be realized by a Markov chain on the state space $\Omega \times \Omega' \times \{0,1\} \times \mathbb{Z}$ where Ω is the state space of $\{X_t\}$, Ω' is the state space of $\{Y_t\}$, such that at time -2^m the Markov chain of $\{Z_t\}$ flips a fair coin and decides on sub-state '0' or sub-state '1'.

Hence, $Z_t = 0$ for $t \le 0$, for $t \in \{1,\dots,2^m\}$ we have $\mathbb{E}Z_t = \frac{\mathbb{E}X_t + \mathbb{E}Y_t}{2} = t \cdot \frac{x+y}{2}$ and for $t \in \{2^m+1,\dots,2^{m+1}\}$ we have

$$\mathbb{E}Z_t = \frac{1}{2}\mathbb{E}(X_{2^m} + Y_{t-2^m}) + \frac{1}{2}\mathbb{E}(X_{t-2^m} + Y_{2^m}) = \frac{2^m x + (t-2^m)y}{2} + \frac{(t-2^m)x + 2^m y}{2} = t \cdot \frac{x+y}{2}.$$

Thus $\{Z_t\}_{t=-\infty}^{2^{m+1}}$ is a $(m+1)$-admissible representation of $\frac{x+y}{2}$. The definition (4) implies that

$$2^{m+1}\left|\left|\left|\frac{x+y}{2}\right|\right|\right|_{m+1}^p \le \sum_{t=1}^{2^{m+1}} \mathbb{E}\left\|Z_t - Z_{t-1}\right\|^p - \frac{\eta}{\Pi^p}\sum_{k=0}^{m+1}\sum_{t=1}^{2^{m+1}} \frac{\mathbb{E}\left\|Z_t - \widetilde{Z}_t(t-2^k)\right\|^p}{2^{kp}}. \quad (10)$$

Note that

$$\sum_{t=1}^{2^{m+1}} \mathbb{E}\left\|Z_t - Z_{t-1}\right\|^p = \sum_{t=1}^{2^m} \mathbb{E}\left\|X_t - X_{t-1}\right\|^p + \sum_{t=1}^{2^m} \mathbb{E}\left\|Y_t - Y_{t-1}\right\|^p. \quad (11)$$

Moreover

$$\sum_{k=0}^{m+1}\sum_{t=1}^{2^{m+1}} \frac{\mathbb{E}\left\|Z_t - \widetilde{Z}_t(t-2^k)\right\|^p}{2^{kp}} = \frac{1}{2^{(m+1)p}}\sum_{t=1}^{2^{m+1}} \mathbb{E}\left\|Z_t - \widetilde{Z}_t(t-2^{m+1})\right\|^p + \sum_{k=0}^{m}\sum_{t=1}^{2^{m+1}} \frac{\mathbb{E}\left\|Z_t - \widetilde{Z}_t(t-2^k)\right\|^p}{2^{kp}}. \quad (12)$$

We bound each of the terms in (12) separately. Note that by construction we have for every $t \in \{1,\dots,2^m\}$

$$Z_t - \widetilde{Z}_t\left(t-2^{m+1}\right) = Z_t - \widetilde{Z}_t\left(1-2^{m+1}\right) = \begin{cases} X_t - Y_t & \text{w/ prob. } 1/4, \\ Y_t - X_t & \text{w/ prob. } 1/4, \\ X_t - \widetilde{X}_t(1) & \text{w/ prob. } 1/4, \\ Y_t - \widetilde{Y}_t(1) & \text{w/ prob. } 1/4. \end{cases}$$

Thus the first term in (12) can be bounded from below as follows.

$$\sum_{t=1}^{2^{m+1}} \mathbb{E}\left\|Z_t - \widetilde{Z}_t(t-2^{m+1})\right\|^p \ge \frac{1}{2}\sum_{t=1}^{2^m}\mathbb{E}\|X_t - Y_t\|^p \ge \frac{1}{2}\sum_{t=1}^{2^m}\|\mathbb{E}X_t - \mathbb{E}Y_t\|^p = \frac{\|x-y\|^p}{2}\sum_{t=1}^{2^m} t^p \ge \frac{2^{(m+1)(p+1)}}{2p+2}\left(\frac{1}{2} - \frac{1}{2^{m+1}}\right)^{p+1}\|x-y\|^p. \quad (13)$$

We now bound the second term in (12) as follows.

$$\begin{aligned}\sum_{k=0}^{m}\sum_{t=1}^{2^{m+1}} \frac{\mathbb{E}\left\|Z_t - \widetilde{Z}_t(t-2^k)\right\|^p}{2^{kp}} &\ge \sum_{k=0}^{m}\sum_{t=1}^{2^m} \frac{\frac{1}{2}\mathbb{E}\left\|X_t - \widetilde{X}_t(t-2^k)\right\|^p + \frac{1}{2}\mathbb{E}\left\|Y_t - \widetilde{Y}_t(t-2^k)\right\|^p}{2^{kp}} \\ &\quad + \sum_{k=0}^{m}\sum_{t=2^m+1}^{2^{m+1}} \frac{\frac{1}{2}\mathbb{E}\left\|X_{t-2^m} - \widetilde{X}_{t-2^m}(t-2^m-2^k)\right\|^p}{2^{kp}} \quad (14) \\ &\quad + \frac{\frac{1}{2}\mathbb{E}\left\|Y_{t-2^m} - \widetilde{Y}_{t-2^m}(t-2^m-2^k)\right\|^p}{2^{kp}} \quad (15) \\ &= \sum_{k=0}^{m}\sum_{t=1}^{2^m} \frac{\mathbb{E}\left\|X_t - \widetilde{X}_t(t-2^k)\right\|^p}{2^{kp}} + \sum_{k=0}^{m}\sum_{t=1}^{2^m} \frac{\mathbb{E}\left\|Y_t - \widetilde{Y}_t(t-2^k)\right\|^p}{2^{kp}}. \quad (16)\end{aligned}$$

To derive (14), note that for $2^m + 2^k \ge t > 2^m$, we have

$$Z_t - \widetilde{Z}_t(t-2^k) = X_{t-2^m} - \widetilde{X}_{t-2^m}(t-2^m-2^k) + Y_{2^m} - \widetilde{Y}_{2^m}(t-2^k).$$

Using Jensen inequality, and the independence of $\{X_t\}$ and $\{Y_t\}$, we conclude that

$$\mathbb{E}\left\|X_{t-2^m} - \widetilde{X}_{t-2^m}(t-2^m-2^k) + Y_{2^m} - \widetilde{Y}_{2^m}(t-2^k)\right\|^p \ge \mathbb{E}\left\|X_{t-2^m} - \widetilde{X}_{t-2^m}(t-2^m-2^k)\right\|^p.$$

The term (15) is derived similarly.

Combining (8), (9), (10), (11), (12), (13) and (16), and letting δ tend to 0, we see that

$$\left|\left|\left|\frac{x+y}{2}\right|\right|\right|_{m+1}^p \le \frac{|||x|||_m^p + |||y|||_m^p}{2} - \frac{\eta}{\Pi^p}\cdot\frac{1}{2p+2}\left(\frac{1}{2} - \frac{1}{2^{m+1}}\right)^{p+1}\|x-y\|^p. \quad (17)$$

Define for $w \in X$,

$$|||w||| = \limsup_{m\to\infty} |||w|||_m.$$

Then a combination of (7) and (17) yields that

$$(1-\varepsilon)\|x\| \le |||x||| \le \|x\|,$$

and

$$\left|\left|\left|\frac{x+y}{2}\right|\right|\right|^p \le \frac{|||x|||^p + |||y|||^p}{2} - \frac{\eta}{\Pi^p(p+1)2^{p+2}}\cdot\|x-y\|^p \le \frac{|||x|||^p + |||y|||^p}{2} - \frac{\eta}{4\Pi^p(p+1)}\cdot\left|\left|\left|\frac{x-y}{2}\right|\right|\right|^p. \quad (18)$$

Note that (18) implies that the set $\{x \in X : |||x||| \le 1\}$ is convex, so that $|||\cdot|||$ is a norm on X. This concludes the proof of Theorem 4. $\square$

3. OBSERVATIONS ON MARKOV CONVEXITY

3.1 A doubling space which is not Markov convex

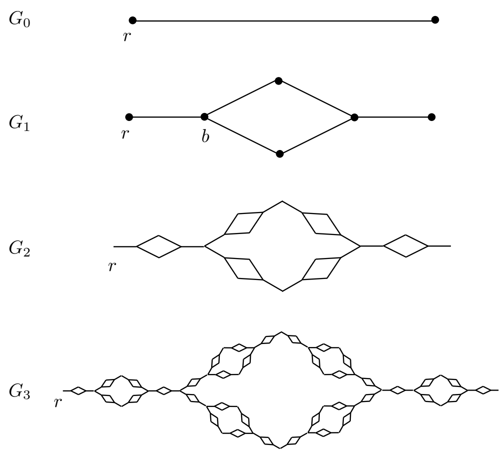

Figure 1: The Laakso graphs.

Consider the Laakso graphs, $\{G_i\}_{i=0}^{\infty}$, which are defined as follows. G_0 is the graph on two vertices with one edge. To construct G_i, take six copies of G_{i-1} and scale their metric by a factor of $\frac{1}{4}$. We glue four of them cyclicly by identifying pairs of endpoints, and attach at two opposite gluing points the remaining two copies. See Figure 1.

We direct G_m as follows: Define the root of G_m to be (arbitrarily chosen) one of the two vertices having only one adjacent edge. In the Figure 1 this could be the leftmost vertex r. Note that in no edge the two endpoints are at the same distance from the root. The edges of G_m are the directed from the endpoint closer to the root, to the endpoint further away from the root. The resulting directed graph is acyclic. We now define $(X_t)_{t=0}^{4^m}$ to be a random walk on the directed graph, starting from the root. This random walk is extended to $t \in \mathbb{Z}$ by assuming that for $t < 0$ $X_t = X_0$, and for $t > 4^m$, $X_t = X_{4^m}$.

Proposition 3.1. *For the above specific random walk,*

$$\sum_{k=0}^{2m}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}d(X_t,\widetilde{X}_t(t-2^k))^p}{2^{kp}} \ge \frac{1}{512\cdot 4^{2p}}\cdot m\sum_{t\in\mathbb{Z}}\mathbb{E}d(X_t,X_{t-1})^p. \quad (19)$$

Proof. Note that on the right hand side of (19),

$$\mathbb{E}d(X_t,X_{t-1})^p = 4^{-m}$$

for $t \in \{0,\dots 4^m-1\}$, and 0 otherwise. So

$$\sum_{t\in\mathbb{Z}}\mathbb{E}d(X_t,X_{t-1})^p = 1.$$

Also note that for $t \in \{0, 4^m - a\}$, $d(X_t, X_{t+a}) = a/4^m$.

Fix $k \in \{0,\dots m-6\}$. Set $h = \lceil k/2\rceil + 2$, and consider the set of integers

$$T_k = \Big\{0,\dots,4^m-1\Big\} \bigcap\Big(\bigcup_{i=0}^{m-h-1}\big[i4^h+4^{h-1}+4^{h-4}, i4^h+4^{h-1}+2\cdot 4^{h-4}\big]\Big).$$

Observe that $|T_k| = 4^{m-h}\cdot 4^{h-4} = \frac{1}{256}\cdot 4^m$. View G_m as composed of $A = G_{m-h}$ where each edge of A is replaced by a copy of G_h. With this view in mind, X_t, when $t \in T_k$ is always between 4^{h-4} to $2\cdot 4^{h-4}$ steps after point "b" (as illustrated in "G_1" in Fig. 1) in a copy of G_h. This means at time $t - 2^k$ the random walk is before that point, and with probability at least 1/2 the two walks X_t, and $\widetilde{X}_t(t-2^k)$ took different outgoing edges at point "b". Since they "walked" at least 4^{h-4} steps after that point, we conclude that with probability at least 1/2, $d(X_t,\widetilde{X}_t(t-2^k)) \ge 4^{h-4-m}$, when $t \in T_k$. Therefore, for $t \in T_k$,

$$\frac{\mathbb{E}d(X_t,\widetilde{X}_t(t-2^k))^p}{2^{kp}} \ge \frac{\frac{1}{2}4^{(h-4-m)p}}{2^{kp}} \ge \tfrac{1}{2}4^{-(m+2)p},$$

and so

$$\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}d(X_t,\widetilde{X}_t(t-2^k))^p}{2^{kp}} \ge \frac{1}{512\cdot 4^{2p}}.$$

Summing over $k \in \{1,\dots,m\}$, we obtain the conclusion of the claim. $\square$

Corollary 3.2. *There exists an n-point doubling and series-parallel metric whose embedding in any p convex space (and in particular in L_p, $p \ge 2$) has $\Omega\left((\log n)^{1/p}\right)$ distortion.*

While the above is a simple corollary of Markov p-convexity, to the best of our knowledge the result has not been known before. It was known before (see, e.g., [18, 17]) that the same graphs has distortion $\Omega(\sqrt{\log n})$ when embedded in 2-convex spaces (and in particular, in L_p, $p \in (1,2]$). However, the proof technique of [18, 17] fails when applied to the case $p > 2$.

3.2 Lipschitz quotients

Here we observe that Markov p-convexity is invariant of Lipschitz quotient.

Proposition 3.3. *If Y is D-Lipschitz quotient of X, then $\Pi_p(Y) \le D\cdot\Pi_p(X)$.*

Proof. Fix a D-Lipschitz quotient $f : X \to Y$ such that $B(f(x), Ar) \supset B(x,r) \supset B(f(x), r/B)$, for $AB \le D$. Also fix a Markov chain X_t on the state space Ω, and a mapping $g : \Omega \to Y$.

Let $\Omega' = \Omega^*$ be the set of finite sequences of elements from Ω. We define a Markov chain X'_t on Ω' as follows

$$\Pr\Big[X'_t = (\omega_1,\dots,\omega_{t-1},\omega_t)\,\Big|\, X'_{t-1} = (\omega_1,\dots,\omega_{t-1})\Big] = \Pr\Big[X_t = \omega_t \,\Big|\, X_{t-1} = \omega_{t-1}\Big],$$

and the rest of the transition probabilities are 0. Also define $g' : \Omega' \to Y$ by $g'((\omega_1,\dots,\omega_t)) = g(\omega_t)$. It is clear that

$g'(X'_t)$ is distributed like $g(X_t)$, so it is sufficient to prove the Markov p-convexity inequality for g'.

We next define a mapping $h' : \Omega' \to X$ by induction on the length of $\omega' \in \Omega'$ as follows: If $\omega' = (\omega_1)$, then we fix $h'(\omega')$ to be an arbitrary element in $f^{-1}(g(\omega_1))$. Next, assume that $\omega' = (\omega_1, \dots, \omega_{i-1}, \omega_i)$. By the co-Lipschitz condition, there exists $x \in X$ such that $f(x) = g(\omega_i)$, and $d_X(x, h'((\omega_1, \dots, \omega_{i-1}))) \le B \cdot d_Y(g(\omega_{i-1}), g(\omega_i))$. We set $h'((\omega_1, \dots, \omega_{i-1}, \omega_i)) := x$. It is clear from the definition that $f \circ h' = g'$

By the Markov p-convexity of X, we have

$$\sum_{k=0}^{m-1} \sum_{t\in\mathbb{Z}} \frac{\mathbb{E} d_X(h'(X'_t) - h'(\widetilde{X}'_t(t-2^k)))^p}{2^{kp}} \le \Pi_p^p(X) \cdot \sum_{t\in\mathbb{Z}} \mathbb{E} d_X(h'(X'_t), h'(X'_{t-1}))^p. \quad (20)$$

Applying f to both sides of (20), we have on the RHS by the choice of h' (and the coLipschitz condition of f),

$$d_X(h'(X'_t), h'(X'_{t-1})) \le B \cdot d_Y(f(h'(X'_t)), f(h'(X'_{t-1}))) = B \cdot d_Y(g'(X'_t), g'(X'_{t-1})).$$

On the LHS of (20) we have using the Lipschitz condition,

$$A \cdot d_X(h'(X'_t), h'(\widetilde{X}'_t(t-2^k))) \ge d_Y(f(h'(X'_t)), f(h'(\widetilde{X}'_t(t-2^k)))) = d_Y(g'(X'_t), g'(\widetilde{X}'_t(t-2^k))).$$

We conclude that

$$\sum_{k=0}^{m-1} \sum_{t\in\mathbb{Z}} \frac{\mathbb{E} d_Y(g'(X'_t), g'(\widetilde{X}'_t(t-2^k)))^p}{2^{kp}} \le (\Pi_p(X) A B)^p \cdot \sum_{t\in\mathbb{Z}} \mathbb{E} d_Y(g'(X'_t), g'(X'_{t-1}))^p.$$

Since this is true for any g, and any Markov chain $\{X_t\}$, the proposition follows. □

4. TREE METRICS DO NOT HAVE THE DICHOTOMY PROPERTY

Theorem 2 is an easy corollary of Theorem 3, derived by setting $s(n) = D$. The rest of this section is therefore devoted to the proof of Theorem 3. The proof is lengthy and technical and we therefore only sketch it in this extended abstract. We begin with an outline (of the sketch) of the proof.

We are given a sequence $s(n)$ such that $s(n)$ is non decreasing, $s(n)/n$ is non increasing, and $s(n) = O(\log n / \log\log n)$. The goal is to construct a metric space X such that $c_X(B_n) \approx s(n)$. Denote by B_∞ the rooted complete binary tree of infinite (i.e. ω) depth. B_∞ will be the underlying space of X. Roughly speaking, the metric on X is based on the regular tree metric on B_∞, but with horizontal distances between vertices at depth h contracted by a factor of $s(h)$. Now it is clear that the natural embedding of B_n in X has distortion at most $s(n)$. We are left to see, in Section 4.1), that the distances defined on X satisfy the triangle inequality, and to prove in Sections 4.2–4.6 an almost matching lower bound on $c_X(B_n)$.

Our approach for proving the lower bound on $c_X(B_n)$ is based on Matoušek's proof [22] that B_n can not be embedded with a constant distortion in uniformly convex spaces. Call an embedding of B_n D-*vertically faithful* if the distances of the images ancestor/descendant pairs in B_n are preserved up to distortion D. Matoušek showed that the class B_n has the dichotomy property with respect to vertical distances. This means that for any $t \in \mathbb{N}$, $\delta > 0$, and $A > 1$, there exists $n = n(t, \delta, A)$ such that for any host space H, if B_n has A-vertically faithful embedding in H, then B_t has $1+\delta$ vertically faithful embedding in H. Matoušek actually proved it only for $t = 2$. His proof extends in straightforward way to any t. But since we need this fact with $t = 4$, for the sake of completeness we reprove it in a different way in Section 4.5.

Matoušek finishes his proof as follows: a $(1+\delta)$ vertically faithful embedding of B_2 contains an image of δ-fork (which is a "half" of $(1+\delta)$ vertically faithful embedding of B_2). It is easy to conclude from (2) that the the distance between the prongs of δ-fork in uniformly convex space has large contraction, and hence the large distortion.

Since the first part of Matoušek's proof is independent of the host space, we can also apply it on X. However, as we shall see, there exist in X δ-forks in which the prongs do not contract much. We solve this problem by studying $(1+\delta)$ vertically faithful embeddings of B_4 and arguing that they must contain a large contraction. This claim, formalized in Lemma 4.3, is proved in Sections 4.2–4.4. We begin in Section 4.2 with studying how the metric P_2 (3-point path) can be approximately embedded in X. We find that there are essentially only two ways to embed it in X, as depicted in Fig. 3.

We then move in Section 4.3 to study δ-forks in X. Since forks are formed by "stitching" two approximate P_2 metrics along a common edge (the handle), we can limit the "search space" using the results of Section 4.2. We find that there are 6 possible types of different approximate forks in X, four of them do not have large contraction of the prongs.

Complete binary trees, and in particular B_4, are composed of forks stitched together, handle to prong. In order to study handle-to-prong stitching, we study (only in the full version) how the metric P_3 (4-point path) can be approximately embedded in X. This is again done by studying how two P_2 metrics can be stitched together, this time bottom edge to top edge. Here we find that there are only three different approximate configurations of P_4 in X.

Using the machinery described above we study in Section 4.4 how the different types of forks can be stitched together in embeddings of B_4, reaching the conclusion that large contraction is unavoidable, and thus completing the proof of Lemma 4.3.

The proof of Theorem 3 is concluded in Section 4.6.

4.1 Horizontally contracted trees

In what follows we denote by B_∞ the rooted infinite complete unweighted binary tree. For $x \in B_\infty$ we let $h(x)$ be its depth, i.e. its distance from the root, and for $x, y \in B_\infty$ we let $\mathsf{lca}(x, y)$ denote their least common ancestor. The tree metric on B_∞ is given by

$$d_{B_\infty}(x, y) = h(x) + h(y) - 2h(\mathsf{lca}(x, y)).$$

We will study the following family of metric spaces which we call *H-trees*. Given a sequence $\varepsilon = \{\varepsilon_n\}_{n=0}^\infty \subseteq (0, 1]$, and assuming $h(x) \le h(y)$, we define $d_\varepsilon : B_\infty \times B_\infty \to [0, \infty)$ by

$$d_\varepsilon(x, y) = h(y) - h(x) + 2\varepsilon_{h(x)} \cdot [h(x) - h(\mathsf{lca}(x, y))].$$

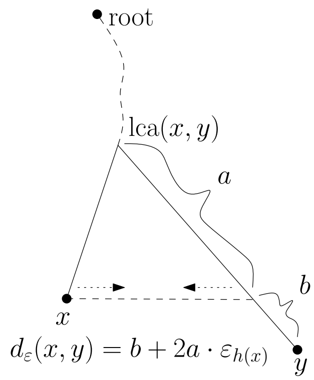

Figure 2: The metric d_ε defined on trees

Note that when $\varepsilon_n = 1$ for all n, we have $d_\varepsilon = d_{B_\infty}$.

LEMMA 4.1. *Assume that $\{\varepsilon_n\}_{n=0}^{\infty} \subseteq (0,1]$ is non-increasing and $\{n\varepsilon_n\}_{n=0}^{\infty}$ is non-decreasing. Then d_ε is a metric on B_∞*

PROOF. Omitted. □

4.2 Approximate midpoints in $(B_\infty, d_\varepsilon)$

From now on we will always assume that $\varepsilon = \{\varepsilon_n\}_{n=0}^{\infty}$ satisfies for all $n \in \mathbb{N}$, $\varepsilon_n \geq \varepsilon_{n+1} > 0$ and $(n+1)\varepsilon_{n+1} \geq n\varepsilon_n$. We recall the concept of *approximate midpoints* which is used frequently in non-linear functional analysis (see [7] and the references therein).

Definition 2. [Approximate midpoints] Let (X, d_X) be a metric space and $\delta \in (0,1)$. For $x, y \in X$ the set of δ-approximate midpoints of x and z is defined as

$$\mathrm{Mid}(x,z,\delta) = \left\{ y \in X : \max\{d_X(x,y), d_X(y,z)\} \leq \frac{1+\delta}{2} \cdot d_X(x,z) \right\}.$$

In what follows given $\eta > 0$ we shall say that two sequences $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ of vertices in B_∞ are η-near if for every $j \in \{1, \ldots, n\}$ we have $d_\varepsilon(u_j, v_j) \leq \eta$.

We shall require the following terminology.

Definition 3. An ordered triple (x, y, z) of vertices in B_∞ will be called a *path-type configuration* if $h(z) \leq h(y) \leq h(x)$, x is a descendant of y, and $h(\mathsf{lca}(z,y)) < h(y)$. The triple (x, y, z) will be called a *tent-type configuration* if $h(y) \leq h(z)$, y is a descendant of x, and $h(\mathsf{lca}(x,z)) < h(x)$. These special configurations are described in Figure 3.

The following useful theorem will be used extensively in the ensuing arguments.

THEOREM 5. *Assume that $\delta \in (0, \frac{1}{16})$, and $\varepsilon_n < \frac{1}{4}$ for all $n \in \mathbb{N}$. Let $x, y, z \in B_\infty$ be such that $y \in \mathrm{Mid}(x,z,\delta)$. Then either (x,y,z) or (z,y,x) is $3\delta d_\varepsilon(x,z)$-near a path-type or tent-type configuration.*

PROOF. Omitted. □

Note that in the notation of Theorem 5, for $\delta < 1/16$, $3d_\varepsilon(x,z) \leq 7d_\varepsilon(x,y)$.

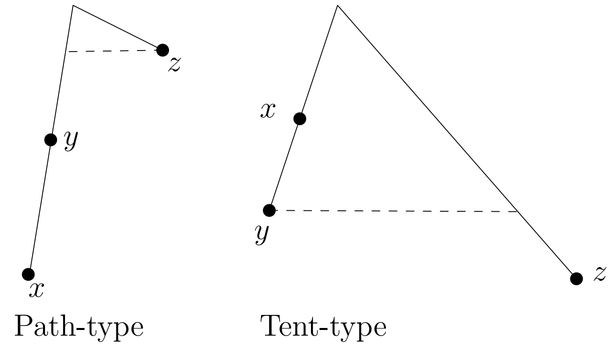

Figure 3: A schematic description of path-type and tent-type configurations.

4.3 Approximate forks in $(B_\infty, d_\varepsilon)$

For $\delta \in (0,1)$, the quadruple (x,y,z,w), $x,y,z,w \in B_\infty$, is called δ-fork, if $y \in \mathrm{Mid}(x,z,\delta) \cap \mathrm{Mid}(x,w,\delta)$. δ-forks in H-trees can be classified using the classification of midpoint configurations. We have four midpoint configurations path-type (P), opposite path-type (p), tent-type (T), and opposite tent-type (t). Thus, there are $\binom{5}{2} = 10$ plausible δ-fork configurations (choose two out of 5: "P","p","T","t","X", where "X" means "the same"). As we shall see, four of these plausible configurations are impossible, two of them have large contraction of the prongs of the forks, i.e. $d_\varepsilon(z,w) \ll d_\varepsilon(x,y)$, which immediately implies large distortion, and the rest of the configurations are problematic in the sense that are not much distorted from the star $K_{1,3}$.

We define the four problematic configurations: A δ-fork (x,y,z,w) is called

- A-near Type I (T‖T)), if both (x,y,z) and (x,y,w) are A near tent-type configurations;
- A near Type II (P‖P), if both (x,y,z) and (x,y,w) are A near path-type configurations;
- A near Type III (p‖T),if (z,y,x) is A near a path-type and (x,y,w) is A near a tent-type, or vice versa;
- A near Type IV (p‖t), if (z,y,x) is A near a path-type and (w,y,x) is A near is $7\delta d_\varepsilon(x,y)$ near a tent-type, or vice versa.

The different configurations are summarized in Table 1.

Midpoint conf.	Type
(T‖T)	Type I
(P‖P)	Type II
(p‖T)	Type III
(p‖t)	Type IV
(p‖p)	prongs contracted
(t‖t)	prongs contracted
(P‖p)	impossible
(P‖t)	possible only as approximate type II
(P‖T)	impossible
(t‖T)	impossible

Table 1: The plausible configurations

The following lemma is the main result of this section.

LEMMA 4.2. *Fix $\delta \in \left(0, \frac{1}{70}\right)$ and assume that $\varepsilon_n < \frac{1}{4}$ for all $n \in \mathbb{N}$. Assume that (x,y,z,w) is a δ-fork in $(B_\infty, d_\varepsilon)$. Let $h_0 = \min\{h(x), h(y), h(z), h(w)\}$. Then either*

- *The δ-fork* (x, y, z, w) *is* $15\delta d_\varepsilon(x, y)$ *near type I, II, III, or IV.*
- *Otherwise* $d_\varepsilon(z, w) \le (83\delta + \varepsilon_{h_0}) 2 d_\varepsilon(y, z)$*, where* $h_0 = \min(h(x), h(y), h(z), h(w))$*.*

The proof of Lemma 4.2 proceeds by checking that the cases marked in Table 1 as "impossible" or "prongs contracted" are indeed so. Details are omitted.

4.4 Inembeddability of vertically faithful B_4

Definition 4. Let T be a rooted tree with root r. An embedding $f : T \to X$ is called an A-vertically faithful embedding, if there exists an $L > 0$ such that for any vertex y and ancestor of y, x we have

$$L \cdot d_T(x, y) \le d_X(f(x), f(y)) \le AL \cdot d_T(x, y).$$

The aim of the current section is to prove the following lemma.

Lemma 4.3. *Let* $\delta < 1/200$*, and let* $f : B_4 \to (B_\infty, d_\varepsilon)$ *be a* $(1+\delta)$ *vertically faithful embedding. Then* $\operatorname{dist}(f) \ge 1/(O(\delta) + \varepsilon_{h_0})$*, where* $h_0 = \min_{x \in B_4} h(f(x))$*.*

The proof is by a contradiction. Assuming the distortion is small, all the δ-forks in the $(1+\delta)$-vertically faithful embedding must be of types *I–IV*. By exploring the constrains on how those δ-forks can be "stitched" together, we reach the conclusion that they are sufficiently severe to force any vertically faithful embedding of B_4 to have a large contraction, and therefore distortion. Details are omitted

4.5 Dichotomy for vertically faithful embedding of binary trees

The aim of this section is to prove the following quantitative bounded distortion (BD) Ramsey result.

Lemma 4.4. *For every* $\delta \in (0, 1)$*,* $t \in \mathbb{N}$*, and* $A \ge 1$*, and for every metric space* X*, if*

$$n \ge \frac{32 \log(2A)}{\delta^2} \cdot \exp\left(\frac{6 \log(2A) t \log t}{\delta}\right),$$

and there exists $f : B_n \to X$ *which is A-vertically faithful embedding, then there exists an isometry (up to scaling)* $\iota : B_t \to B_n$ *for which* $f \circ \iota$ *is* $(1+\delta)$ *vertically faithful embedding.* ≪*Unfortunately, we need the full power of BD-Ramsey here... –Manor*≫

Lemma 4.4 was essentially proved in [22]. However, the proof there is specialized for $t = 2$. The full version contains a somewhat different proof of this fact.

4.6 Non-embeddability of binary trees

Lemma 4.5. *Let* $s(n)$ *be a non-decreasing series, satisfying* $s(n)/n$ *is non increasing, and* $4 < s(n) \le cn^c$*, where* $c > 0$ *is some (small) universal constant. Then there exists a metric space* X *for which*

$$\frac{1}{\frac{C \log(2s(n))}{\log n} + \frac{1}{s(n/40s(n))}} \le c_X(B_n) \le s(n).$$

Proof. Let $\varepsilon_n = 1/s(n)$, and $\varepsilon = (\varepsilon_n)_n$. We choose $X = (B_\infty, d_\varepsilon)$. Clearly, the natural embedding of B_n in $(B_\infty, d_\varepsilon)$ has at most $s(n)$ distortion.

To prove a lower bound on the distortion of any embedding B_n in X, fix an embedding $f : B_n \to X$, such that $\operatorname{dist}(f) \le s(n)$.

Fix $h_0 = n/40s(n)$ and let $\widehat{X} \subset X$ be the induced metric on the subset $\{x \in B_\infty : h(x) > h_0\}$. We next find a "large" complete subtree $\widehat{B}$ of B_n for which $f(\widehat{B}) \subset \widehat{X}$. Let $h_{\min} = \min\{h(x) : x \in f(B_n)\}$, and $h_{\max} = \max\{h(x) : x \in f(B_n)\}$. If $h_{\min} > h_0$, then clearly $f(B_n) \subset \widehat{X}$, and we are done. So assume now that $h_{\min} < h_0$. Since f is injection it must satisfy $h_{\max} \ge n$. Hence $\|f\|_{\mathrm{Lip}} \ge \frac{n - h_0}{2n} \ge 1/4$. Since $\operatorname{dist}(f) \le s(n)$ we conclude that $\|f^{-1}\|_{\mathrm{Lip}} \le 4s(n)$. Since $\operatorname{diam}(X \setminus \widehat{X}) \le 2h_0$, $\operatorname{diam}\big(f^{-1}(X \setminus \widehat{X})\big) \le 8h_0 s(n) = n/5$. Therefore, the levels in B_n occupied by $f^{-1}(X \setminus \widehat{X})$ must appear in either the top $\frac{2}{3}n - 1$ levels or the bottom $\frac{2}{3}n - 1$ levels in B_n. Either way, there exists $\widehat{B} \subset B_n$, a complete binary subtree of height $n/3$ for which $f(\widehat{B}) \subset \hat{X}$.

We now apply Lemma 4.4 on $f|_{\hat{B}}$ — the embedding of $\hat{B}$ in $\hat{X}$ — with $t = 4$, and $A = s(n)$. We obtain that when

$$\frac{n}{3} \ge \exp\left(\frac{12 \log(2s(n)) 4 \log 4}{\delta}\right), \tag{21}$$

there exists an isometry (up to scaling) $\iota : B_4 \to B_n$ for which $f \circ \iota$ is $(1+\delta)$ vertically faithful embedding in $\hat{X}$. By Lemma 4.3,

$$\operatorname{dist}(f \circ \iota) \ge \frac{1}{C\delta + \varepsilon_{h_0}}$$

(for some universal $C > 0$), which implies the same lower bound on $\operatorname{dist}(f)$.

Since ε_{h_0} was already fixed, we are left to choose δ as small as possible, subject to (21). We therefore choose $\delta = \frac{100 \log(2s(n))}{\log n}$. □

Note that Theorem 3 is an immediate corollary of Lemma 4.5.

5. REFERENCES

[1] *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006, Miami, Florida, USA, January 22-26, 2006.* ACM Press, 2006.

[2] S. Arora, L. Lovász, I. Newman, Y. Rabani, Y. Rabinovich, and S. Vempala. Local versus global properties of metric spaces. In *SODA* [1], pages 41–50. Available from: `http://www.cs.princeton.edu/~arora/pubs/localglobal.pdf`.

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

[4] K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.

[5] Y. Bartal and M. Mendel. Multiembedding of metric spaces. *SIAM J. Comput.*, 34(1):248–259 (electronic), 2004.

[6] S. Bates, W. B. Johnson, J. Lindenstrauss, D. Preiss, and G. Schechtman. Affine approximation of Lipschitz functions and nonlinear quotients. *Geom. Funct. Anal.*, 9(6):1092–1127, 1999. Available from:

www.wisdom.weizmann.ac.il/~gideon/papers/LipQuotientsDec9.ps.

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

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

[9] J. Bourgain, V. Milman, and H. Wolfson. On type of metric spaces. *Trans. Amer. Math. Soc.*, 294(1):295–317, 1986.

[10] P. Enflo. Banach spaces which can be given an equivalent uniformly convex norm. In *Proceedings of the International Symposium on Partial Differential Equations and the Geometry of Normed Linear Spaces (Jerusalem, 1972)*, volume 13, pages 281–288 (1973), 1972.

[11] N. Garg, G. Konjevod, and R. Ravi. A polylogarithmic approximation algorithm for the group Steiner tree problem. *J. Algorithms*, 37(1):66–84, 2000.

[12] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston Inc., Boston, MA, english edition, 2007.

[13] A. Gupta. Embedding tree metrics into low-dimensional euclidean spaces. *Discrete & Computational Geometry*, 24(1):105–116, 2000.

[14] R. C. James. Some self-dual properties of normed linear spaces. In *Symposium on Infinite-Dimensional Topology (Louisiana State Univ., Baton Rouge, La., 1967)*, pages 159–175. Ann. of Math. Studies, No. 69. Princeton Univ. Press, Princeton, N.J., 1972.

[15] T. J. Laakso. Ahlfors Q-regular spaces with arbitrary $Q > 1$ admitting weak Poincaré inequality. *Geom. Funct. Anal.*, 10(1):111–123, 2000.

[16] U. Lang and C. Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.

[17] J. R. Lee, M. Mendel, and A. Naor. Metric structures in L_1: dimension, snowflakes, and average distortion. *European J. Combin.*, 26(8):1180–1190, 2005.

[18] 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.

[19] J. R. Lee, A. Naor, and Y. Peres. Trees and markov convexity. In *SODA* [1], pages 1028–1037. Full version to appear in Geom. Funct. Anal.

[20] J. Lindenstrauss. On the modulus of smoothness and divergent series in Banach spaces. *Michigan Math. J.*, 10:241–252, 1963.

[21] J. Matousek. Ramsey-like properties for bi-Lipschitz mappings of finite metric spaces. *Comment. Math. Univ. Carolin*, 33(3):451–463, 1992. Available from: http://kam.mff.cuni.cz/~matousek/rams.ps.gz.

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

[23] M. Mendel and A. Naor. Metric cotype. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 79–88, New York, NY, USA, 2006. ACM.

[24] M. Mendel and A. Naor. Scaled Enflo type is equivalent to Rademacher type. *Bull. London Math. Soc.*, 39(3):493–498, 2007.

[25] M. Mendel and A. Naor. Metric cotype. *Ann. of Math.*, to appear. Preliminary version in SODA '06.

[26] A. Naor, Y. Peres, O. Schramm, and S. Sheffield. Markov chains in smooth Banach spaces and Gromov-hyperbolic metric spaces. *Duke Math. J.*, 134(1):165–197, 2006.

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

[28] I. Newman and Y. Rabinovich. A lower bound on the distortion of embedding planar metrics into euclidean space. In *Symposium on Computational Geometry*, pages 94–96, 2002. Available from: http://cs.haifa.ac.il/~ilan/online-papers/comp-geom-02.ps.

[29] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.

[30] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.

[31] M. Ribe. On uniformly homeomorphic normed spaces. *Ark. Mat.*, 14(2):237–244, 1976.