Markov convexity and local rigidity of distorted metrics

Manor Mendel∗
Open University of Israel
manorme@openu.ac.il

Assaf Naor†
New York University
naor@cims.nyu.edu

Abstract

It is shown that a Banach space admits an equivalent norm whose modulus of uniform convexity has power-type $p$ if and only if it is Markov $p$-convex. Counterexamples are constructed to natural questions related to isomorphic uniform convexity of metric spaces, showing in particular that tree metrics fail to have the dichotomy property.

Contents

1 Introduction 2

1.1 The nonexistence of a metric dichotomy for trees 5

1.1.1 Overview of the proofs of Theorem 1.10 and Theorem 1.12 7

2 Markov $p$-convexity and $p$-convexity coincide 9

3 A doubling space which is not Markov $p$-convex for any $p \in (0, \infty)$ 15

4 Lipschitz quotients 17

5 A dichotomy theorem for vertically faithful embeddings of trees 18

6 Tree metrics do not have the dichotomy property 21

6.1 Horizontally contracted trees 21

6.2 Geometry of $H$-trees 23

6.2.1 Classification of approximate midpoints 28

6.2.2 Classification of approximate forks 30

6.2.3 Classification of approximate 3-paths 36

6.3 Nonembeddability of vertically faithful $B_4$ 42

6.4 Nonembeddability of binary trees 44

7 Discussion and open problems 45

∗Supported by ISF grant 221/07, BSF grant 2006009, and a gift from Cisco research center.
†Supported by NSF grants CCF-0635078 and CCF-0832795, BSF grant 2006009, and the Packard Foundation.
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 \(\|x\|_X \leq \|Tx\|_Y \leq D\|x\|_X\) 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 \(f\) and \(f^{-1}\) 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 research program was put forth by Bourgain in 1986 [5].

Since its inception, the Ribe program attracted the work of many mathematicians, and led to the development of several satisfactory metric theories that extend important concepts and results of Banach space theory; see the introduction of [24] for a historical discussion. So far, progress on the Ribe program has come hand-in-hand with 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, derive some of its metric consequences, and construct unexpected counter-examples which indicate that further progress on the Ribe program can uncover nonlinear phenomena that are markedly different from their Banach space counterparts. 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, which will be explained in detail below, were announced in [23].

For \(p \geq 2\), 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 \(a, b > 0\), \(a\|x\|_X \leq ||x|| \leq b\|x\|_X\) for all \(x \in X\), and a constant \(K > 0\) satisfying:

\[
||x|| = ||y|| = 1 \implies \left\| \frac{x + y}{2} \right\|^p \leq 1 - K\|x - y\|^p.
\]

\(X\) is called superreflexive if it is \(p\)-convex for some \(p \geq 2\) (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 [10] and Enflo [7]). For concreteness, we recall (see, e.g., [2]) that \(L_p\) is \(2\)-convex for \(p \in (1, 2]\) and \(p\)-convex for \(p \in [2, \infty)\). Ribe’s theorem implies that \(p\)-convexity, and hence also superreflexivity, is preserved under uniform homeomorphisms. The first major success of the Ribe program is a famous theorem of Bourgain [5] which obtains a metrical characterization of superreflexivity as follows.

**Theorem 1.1** (Bourgain’s metrical characterization of superreflexivity [5]). 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 \((\mathcal{M}, d_{\mathcal{M}}), (\mathcal{N}, d_{\mathcal{N}})\), the parameter \(c_{\mathcal{M}}(\mathcal{N})\) denotes the smallest bi-Lipschitz distortion with which \(\mathcal{N}\) embeds into \(\mathcal{M}\), i.e., the infimum of

James’ original definition of superreflexivity is that a Banach space \(X\) is superreflexive if “its local structure forces reflexivity”, i.e., if every Banach space \(Y\) that is finitely representable in \(X\) must be reflexive. Enflo’s renorming theorem states that superreflexivity is equivalent to having an equivalent norm \(|| \cdot ||\) that 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 \(||x + y|| \leq 2 - \delta\).
A Banach space is

Theorem 1.3.

in \[16\] if the converse is also true. Here we answer this question positively: Bourgain \[5\], that if a Banach space

and their algorithmic applications. It was also shown in \[16\], via a modification of an argument of

constant factors, the Euclidean distortion of

\( T \)

to order to adapt it to a nonlinear condition such as (3).

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 the Ribe program for \( p \)-convexity requires significant additional work beyond

Bourgain’s superreflexivity theorem. As a first step in this direction, Lee, Naor and Peres \[16\]
defined a bi-Lipschitz invariant of metric spaces called Markov 2-convexity, which is motivated by

Ball’s notion of Markov type \[1\] and Bourgain’s argument in \[5\].

**Definition 1.2** \((16)\). Let \( \{X_t\}_{t \in \mathbb{Z}} \) be a Markov chain on a state space \( \Omega \). Given an integer \( k \geq 0 \),

we denote by \( \{X_t(k)\}_{t \in \mathbb{Z}} \) the process which equals \( X_t \) for time \( t \leq k \), and evolves independently

(with respect to the same transition probabilities) for time \( t > k \). Fix \( p > 0 \). A metric space

\((X, d_X)\) is called Markov \( p \)-convex with constant \( \Pi \) if for every Markov chain \( \{X_t\}_{t \in \mathbb{Z}} \) on a state

space \( \Omega \), and every \( f : \Omega \rightarrow X \),

\[
\sum_{k=0}^{\infty} \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_X \left( f(X_t), f\left( X_t(k) \right) \right) \left/ 2^{kp} \right. \right] \leq \Pi^p \cdot \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_X \left( f(X_t), f(X_{t-1}) \right)^p \right].
\] (3)

The least constant \( \Pi \) for which (3) holds for all Markov chains is called the Markov \( p \)-convexity constant of \( X \), and is denoted \( \Pi_p(X) \). We shall say that \((X, d_X)\) is Markov \( p \)-convex if \( \Pi_p(X) < \infty \).

To gain intuition for Definition 1.2 consider the standard downward random walk starting from

the root of the binary tree \( B_n \) (with absorbing states at the leaves). For an arbitrary mapping \( f \)

from \( B_n \) to a metric space \((X, d_X)\), the triangle inequality implies that for each \( k \in \mathbb{N} \) we have

\[
\sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_X \left( f(X_t), f\left( X_t(k) \right) \right) \left/ 2^{kp} \right. \right] \leq_p \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_X \left( f(X_t), f(X_{t-1}) \right)^p \right],
\] (4)

with asymptotic equality (up to constants depending only on \( p \)) for \( k \leq \frac{\log n}{2} \) when \( X = B_n \) and \( f \)

is the identity mapping. On the other hand, if \( X \) is a Markov \( p \)-convex space then the sum over \( k \) of

the left-hand side of (4) is uniformly bounded by the right-hand side of (4), and therefore Markov

\( p \)-convex spaces cannot contain \( B_n \) with distortion uniformly bounded in \( n \).

We refer to \[16\] for more information on the notion of Markov \( p \)-convexity. In particular, it is

shown in \[16\] 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 \[16\] for \( L_p \) versions of this statement

and their algorithmic applications. It was also shown in \[16\], via a modification of an argument of

Bourgain \[5\], that if a Banach space \( X \) is \( p \)-convex then it is also Markov \( p \)-convex. It was asked

in \[16\] if the converse is also true. Here we answer this question positively:

**Theorem 1.3.** 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 the Ribe

program in this case. Our proof of Theorem 1.3 is based on a renorming method of Pisier \[29\]. It

can be viewed as a nonlinear variant of Pisier’s argument, and several subtle changes are required

in order to adapt it to a nonlinear condition such as (3).
Results similar to Theorem 1.3 have been obtained for the notions of type and cotype of Banach spaces (see [6, 30, 11, 25, 24, 22]), 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 [16], here we show that this invariant is preserved under Lipschitz quotients. The notion of Lipschitz quotient was introduced by Gromov [8, Sec. 1.25]. 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 $r > 0$,

$$f(B_X(x,r)) \supseteq B_Y\left(f(x), \frac{r}{c}\right).$$

Here we show the following result:

**Theorem 1.4.** 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 [3] Bates, Johnson, Lindenstrauss, Preiss and Schechtman investigated in detail Lipschitz quotients of Banach spaces. Their results imply that if $2 \leq p < q$ 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 our new “invariant approach” to the above result of [3] significantly extends it. Note that the method of [3] 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 1.3 it is natural to ask if 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 [16]. Here we show that the answer is negative:

**Theorem 1.5.** There exists a metric space $(X, d_X)$ which is not Markov $p$-convex for any $p \in (0, \infty)$, yet $\lim_{n \to \infty} c_X(B_n) = \infty$. In fact, $(X, d_X)$ can be a doubling metric space, and hence $c_X(B_n) \geq 2^{\kappa n}$ for some constant $\kappa > 0$.

Theorem 1.5 is in sharp contrast to the previously established metric characterizations of the linear notions of type and cotype. Specifically, it was shown by Bourgain, Milman and Wolfson [6] that any metric space with no nontrivial metric type must contain the Hamming cubes $\{0,1\}^n, \|\cdot\|_1$ with distortion independent of $n$. An analogous result was obtained in [24] for metric spaces with no nontrivial metric cotype, with the Hamming cube replaced by the $\ell_\infty$ grid $\{1, \ldots, m\}^n, \|\cdot\|_\infty$.

Our proof of Theorem 1.5 is based on an analysis of the behavior of a certain Markov chain on the Laakso graphs: a sequence of combinatorial graphs whose definition is recalled in Section 3. As a consequence of this analysis, we obtain the following distortion lower bound:

**Theorem 1.6.** For any $p \geq 2$, the Laakso graph of cardinality $n$ incurs distortion $\Omega((\log n)_{1/p})$ in any embedding into a $p$-convex Banach space.

Thus, in particular, for $p > 2$ the $n$-point Laakso graph incurs distortion $\Omega((\log n)_{1/p})$ in any embedding into $L_p$. The case of $L_p$ embeddings of the Laakso graphs when $1 < p \leq 2$ was already solved in [26, 12, 15, 14] using the uniform 2-convexity property of $L_p$. But, these proofs rely
crucially on 2-convexity and do not extend to the case of \( p \)-convexity when \( p > 2 \). Subsequent to the publication of our proof of Theorem 1.6 in the announcement \[23\], an alternative proof of this fact was recently discovered by Johnson and Schechtman in \[11\].

### 1.1 The nonexistence of a metric dichotomy for trees

Bourgain’s metrical characterization of superreflexivity yields the following statement:

**Theorem 1.7** (Bourgain’s tree dichotomy \[5\]). For any Banach space \((X, \| \cdot \|_X)\) one of the following two dichotomic possibilities must hold true:

- either for all \( n \in \mathbb{N} \) we have \( c_X(B_n) = 1 \),
- or there exists \( \alpha = \alpha_X > 0 \) such that for all \( n \in \mathbb{N} \) we have \( c_X(B_n) \geq (\log n)^{\alpha} \).

Thus, there is a gap in the possible rates of growth of the sequence \( \{c_X(B_n)\}_{n=1}^\infty \) when \( X \) is a Banach space; consequently, if we were told that, say, \( c_X(B_n) = O(\log \log n) \), then we would immediately deduce that actually \( c_X(B_n) = 1 \) for all \( n \). Additional gap results of this type are known due to the theory of nonlinear type and cotype:

**Theorem 1.8** (Bourgain-Milman-Wolfson cube dichotomy \[6\]). For any metric space \((X, d_X)\) one of the following two dichotomic possibilities must hold true:

- either for all \( n \in \mathbb{N} \) we have \( c_X(\{0,1\}^n, \| \cdot \|_1) = 1 \),
- or there exists \( \alpha = \alpha_X > 0 \) such that for all \( n \in \mathbb{N} \) we have \( c_X(\{0,1\}^n, \| \cdot \|_1) \geq n^\alpha \).

Theorem 1.7 is a metric analogue of Pisier’s characterization \[28\] of Banach spaces with trivial Rademacher type. A metric analogue of the Maurey-Pisier characterization \[20\] of Banach spaces with finite Rademacher cotype yields the following dichotomy result for \( \ell_\infty \) grids:

**Theorem 1.9** (Grid dichotomy \[21\]). For any metric space \((X, d_X)\) one of the following two dichotomic possibilities must hold true:

- either for all \( n \in \mathbb{N} \) we have \( c_X(\{0, \ldots, n\}^n, \| \cdot \|_\infty) = 1 \),
- or there exists \( \alpha = \alpha_X > 0 \) such that for all \( n \in \mathbb{N} \) we have \( c_X(\{0, \ldots, n\}^n, \| \cdot \|_\infty) \geq n^\alpha \).

We refer to the survey article \[21\] for more information on the theory of metric dichotomies.

Note that Theorem 1.7 is stated for Banach spaces, while Theorem 1.8 and Theorem 1.9 hold for general metric spaces. One might expect that as in the case of previous progress on Ribe’s program, a metric theory of \( p \)-convexity would result in a proof that Theorem 1.7 holds when \( X \) is a general metric space. Surprisingly, we show here that this is not true:

**Theorem 1.10.** There exists a universal constant \( C > 0 \) with the following property. Assume that \( \{s(n)\}_{n=0}^\infty \subseteq [4, \infty) \) is a nondecreasing sequence such that \( \{n/s(n)\}_{n=0}^\infty \) is also nondecreasing. Then there exists a metric space \((X, d_X)\) satisfying for all \( n \geq 2 \),

\[
s \left( \left\lfloor \frac{n}{40s(n)} \right\rfloor \right) \left( 1 - \frac{C s(n) \log s(n)}{\log n} \right) \leq c_X(B_n) \leq s(n). \tag{6}
\]

Thus, assuming that \( s(n) = o\left( \frac{\log n}{\log \log n} \right) \), there exists a subsequence \( \{s(n_k)\}_{k=1}^\infty \) for which

\[
(1 - o(1)) s(n_k) \leq c_X(B_{n_k}) \leq s(n_k). \tag{7}
\]
Theorem 1.10 shows that unlike the case of Banach spaces, for general metric spaces, $c_X(B_n)$ can have an arbitrarily slow growth rate.

Bourgain, Milman and Wolfson also obtained in [6] the following finitary version of Theorem 1.8:

**Theorem 1.11 (Local rigidity of Hamming cubes [6]).** For every $ε > 0$, $D > 1$ and $n \in \mathbb{N}$ there exists $m = m(ε, D, n) \in \mathbb{N}$ such that

$$\lim_{n \to \infty} m(ε, D, n) = \infty,$$

and for every metric $d$ on $\{0, 1\}^n$ which is bi-Lipschitz with distortion $\leq D$ to the $ℓ_1$ (Hamming) metric,

$$c(\{0, 1\}^n, d)(\{0, 1\}^m, \| \cdot \|_1) \leq 1 + ε.$$

We refer to [6] (see also [30]) for bounds on $m(ε, D, n)$. Informally, Theorem 1.11 says that the Hamming cube $(\{0, 1\}^n, \| \cdot \|_1)$ is locally rigid in the following sense: it is impossible to distort the Hamming metric on a sufficiently large hypercube without the resulting metric space containing a hardly distorted copy of an arbitrarily large Hamming cube. Stated in this way, Theorem 1.11 is a metric version of James’ theorem [9] that $ℓ_1$ is not a distortable space. The analogue of Theorem 1.11 with the Hamming cube replaced by the $ℓ_∞$ grid $(\{0, \ldots, n\}^n, \| \cdot \|_∞)$ is Matoušek’s BD-Ramsey theorem [19]; see [24] for quantitative results of this type in the $ℓ_∞$ case. The following variant of Theorem 1.10 shows that a local rigidity statement as above fails to hold true for binary trees; it can also be viewed as a negative solution of the distortion problem for the infinite binary tree (see [27] and [4, Ch. 13, 14] for more information on the distortion problem for Banach spaces).

**Theorem 1.12.** Let $B_∞$ be the complete unweighted infinite binary tree. For every $D \geq 4$ there exists a metric $d$ on $B_∞$ that is $D$-equivalent to the original shortest-path metric on $B_∞$, yet for every $ε \in (0, 1)$ and $m \in \mathbb{N}$,

$$c(B_∞, d)(B_m) \leq D - ε \implies m \leq D^C D^2/ε.$$

The local rigidity problem for binary trees was studied by several mathematicians. In particular, C. Fefferman asked (private communication, 2005) whether $\{B_n\}_{n=1}^\infty$ have the local rigidity property, and Theorem 1.12 answers this question negatively. Fefferman also proved a partial local rigidity result which is a non-quantitative variant of Theorem 1.14 below (see also Section 5). We are very grateful to C. Fefferman for asking us the question that led to the counter-examples of Theorem 1.10 and Theorem 1.12, for sharing with us his partial positive results, and for encouraging us to work on these questions. M. Gromov also investigated the local rigidity problem for binary trees, and proved (via different methods) non-quantitative partial positive results in the spirit of Theorem 1.14. We thank M. Gromov for sharing with us his unpublished work on this topic.

The results of Theorem 1.10 and Theorem 1.12 are quite unexpected. Unfortunately, their proofs are delicate and lengthy, and as such constitute the most involved part of this article. In order to facilitate the understanding of these constructions, we end the introduction with an overview of the main geometric ideas that are used in their proofs. This is done in Section 1.1.1 below—we recommend reading this section first before delving into the technical details presented in Section 6.
1.1.1 Overview of the proofs of Theorem 1.10 and Theorem 1.12

For $x \in B_\infty$ let $h(x)$ be its depth, i.e., its distance from the root. Also, for $x, y \in B_\infty$ let $\text{lca}(x, y)$ denote their least common ancestor. The tree metric on $B_\infty$ is then given by:

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

The metric space $X$ of Theorem 1.10 will be $B_\infty$ as a set, with a new metric defined as follows. Given a sequence $\varepsilon = \{\varepsilon_n\}_{n=0}^\infty \subseteq (0, 1]$ we define $d_\varepsilon : B_\infty \times B_\infty \to [0, \infty)$ by

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

$d_\varepsilon$ does not necessarily satisfy the triangle inequality, but under some simple conditions on the sequence $\{\varepsilon_n\}_{n=0}^\infty$ it does become a metric on $B_\infty$; see Lemma 6.1. A pictorial description of the metric $d_\varepsilon$ is contained in Figure 1. Note that when $\varepsilon_n = 1$ for all $n$, we have $d_\varepsilon = d_{B_\infty}$. Below we call the metric spaces $(B_\infty, d_\varepsilon)$ horizontally distorted trees, or $H$-trees, in short.

![Figure 1: The metric $d_\varepsilon$ defined on $B_\infty$. The arrows indicate horizontal contraction by $\varepsilon_{h(x)}$.](image)

The metric space $(X, d_X)$ of Theorem 1.10 will be $(B_\infty, d_\varepsilon)$, where $\varepsilon_n = 1/s(n)$ for all $n$. The identity mapping of $B_n$ into the top $n$-levels of $B_\infty$ has distortion at most $s(n)$, and therefore $c(X(B_n)) \leq s(n)$. The challenge is to prove the lower bound on $c(X(B_n))$ in (6). Our initial approach to lower-bounding $c(X(B_n))$ was Matoušek’s metric differentiation proof [18] of asymptotically sharp distortion lower bounds for embeddings of $B_n$ into uniformly convex Banach spaces.

Following Matoušek’s terminology [18], for $\delta > 0$ a quadruple of points $(x, y, z, w)$ in a metric space $(X, d_X)$ is called a $\delta$-fork if $y \in \text{Mid}(x, z, \delta) \cap \text{Mid}(x, w, \delta)$, where for $a, b \in X$ the set of $\delta$-approximate midpoints $\text{Mid}(a, b, \delta) \subseteq X$ is defined as the set of all $w \in X$ satisfying $\max\{d_X(x, y), d_X(y, z)\} \leq \frac{1+\delta}{2} \cdot d_X(x, z)$. The points $z, w$ will be called below the prongs of the $\delta$-fork $(x, y, z, w)$. Matoušek starts with the observation that if $X$ is a uniformly convex Banach space then in any $\delta$-fork in $X$ the distance between the prongs must be much smaller (as $\delta \to 0$) than $d_X(x, y)$. Matoušek then shows that for all $D > 0$, any distortion $D$ embedding of $B_n$ into $X$ must map some 0-fork in $B_n$ to a $\delta$-fork in $X$, provided $n$ is large enough (as a function of $D$ and $\delta$). This reasoning immediately implies that $c(X(B_n))$ must be large when $X$ is a uniformly convex Banach space, and a clever argument of Matoušek in [18] turns this qualitative argument into sharp quantitative bounds.
Of course, we cannot hope to use the above argument of Matoušek in order to prove Theorem 1.10, since Bourgain’s tree dichotomy theorem (Theorem 1.7) does hold true for Banach spaces. But, perhaps we can mimic this uniform convexity argument for other target metric spaces? On the face of it, $H$-trees are ideally suited for this purpose, since the horizontal contractions that we introduced shrink distances between the prongs of canonical forks (call $(x, y, z, w) \in B_{\infty}$ a canonical fork if $x$ is an ancestor of $y$ and $z, w$ are descendants of $y$ at depth $h(x) + 2(h(y) - h(x))$). It is for this reason exactly that we defined $H$-trees.

Unfortunately, the situation isn’t so simple. It turns out that $H$-trees do not behave like uniformly convex Banach spaces in terms of the prong-contractons that they impose of $\delta$-forks. $H$-trees can even contain larger problematic configurations that have several undistorted $\delta$-forks; such an example is depicted in Figure 2.

![Figure 2](image)

**Figure 2**: The metric space on the right is the $H$-tree $(B_{\infty}, d_{\varepsilon})$, where $\varepsilon_n = \delta$ for all $n$. The picture describes an embedding of the tree on the left ($B_3$ minus 4 leaves) into $(B_{\infty}, d_{\varepsilon})$ with distortion at most 6, yet all ancestor/descendant distances are distorted by at most $1 + O(\delta)$.

Thus, in order to prove Theorem 1.10 it does not suffice to use Matoušek’s argument that a bi-Lipschitz embedding of a large enough $B_n$ must send some 0-fork to a $\delta$-fork. But, it turns out that this argument applies not only to forks, but also to larger configurations.

**Definition 1.13.** Let $(T, d_T)$ be a tree with root $r$, and let $(X, d_X)$ be a metric space. A mapping $f : T \to X$ is called a $D$-vertically faithful embedding if there exists a (scaling factor) $\lambda > 0$ satisfying for any $x, y \in T$ such that $x$ is an ancestor of $y$,

$$\lambda d_T(x, y) \leq d_X(f(x), f(y)) \leq D\lambda d_T(x, y).$$

(8)

Recall that the distortion of a mapping $\phi : \mathcal{M} \to \mathcal{N}$ between metric spaces $(\mathcal{M}, d_{\mathcal{M}})$ and $(\mathcal{N}, d_{\mathcal{N}})$ is defined as

$$\text{dist}(\phi) \overset{\text{def}}{=} \left(\sup_{x, y \in \mathcal{M}, x \neq y} \frac{d_{\mathcal{N}}(\phi(x), \phi(y))}{d_{\mathcal{M}}(x, y)}\right) \cdot \left(\sup_{x, y \in \mathcal{M}, x \neq y} \frac{d_{\mathcal{M}}(x, y)}{d_{\mathcal{N}}(\phi(x), \phi(y))}\right) \in [1, \infty].$$

With this terminology, we can state the following crucial result.
Theorem 1.14. There exists a universal constant $c > 0$ with the following property. Fix an integer $t \geq 2$, $\delta, \xi \in (0, 1)$, and $D \geq 2$, and assume that $n \in \mathbb{N}$ satisfies

$$n \geq \frac{1}{\xi} D^c (t \log t)^{1/\delta}. \tag{9}$$

Let $(X, d_X)$ be a metric space and $f : B_n \to X$ a $D$-vertically faithful embedding. Then there exists a mapping $\phi : B_t \to B_n$ with the following properties.

- If $x, y \in B_t$ are such that $x$ is an ancestor of $y$, then $\phi(x)$ is an ancestor of $\phi(y)$.
- $\text{dist}(\phi) \leq 1 + \xi$.
- The mapping $f \circ \phi : B_t \to X$ is a $(1 + \delta)$-vertically faithful embedding of $B_t$ in $X$.

Theorem 1.14 is essentially due to Matoušek [18]. Matoušek actually proved this statement only for $t = 2$, since this is all that he needed in order to analyze forks. But, his proof extends in a straightforward way to any $t \in \mathbb{N}$. Since we will use this assertion with larger $t$, for the sake of completeness we reprove it, in a somewhat different way, in Section 5. Note that Theorem 1.14 says that $\{B_n\}_{n=1}^{\infty}$ do have a local rigidity property with respect to vertically faithfully embeddings.

We solve the problem created by the existence of configurations as those depicted in Figure 2 by studying $(1 + \delta)$-vertically faithful embeddings of $B_4$, and arguing that they must contain a large contracted pair of points. This claim, formalized in Lemma 6.27, is proved in Sections 6.2, 6.3.

We begin in Section 6.2.1 with studying how the metric $P_2$ (3-point path) can be approximately embedded in $(B_\infty, d_\varepsilon)$. We find that there are essentially only two ways to embed it in $(B_\infty, d_\varepsilon)$, as depicted in Figure 3. We then proceed in Section 6.2.2 to study $\delta$-forks in $(B_\infty, d_\varepsilon)$. 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 6.2.1. We find that there are six possible types of different approximate forks in $(B_\infty, d_\varepsilon)$, only four of which (depicted in Figure 4) do not have highly contracted 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 investigate in Section 6.2.3 how the metric $P_3$ (4-point path) can be approximately embedded in $(B_\infty, d_\varepsilon)$. This is again done by studying how two $P_2$ metrics can be stitched together, this time bottom edge to top edge. We find that there are only three different approximate configurations of $P_3$ in $(B_\infty, d_\varepsilon)$.

Using the machinery described above, we study in Section 6.3 how the different types of forks can be stitched together in embeddings of $B_4$ into $(B_\infty, d_\varepsilon)$, reaching the conclusion that a large contraction is unavoidable, and thus completing the proof of Lemma 6.27. The proofs of Theorem 1.10 and Theorem 1.12 are concluded in Section 6.4.

2 Markov $p$-convexity and $p$-convexity coincide

In this section we prove Theorem 1.3, 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 [16], based on an argument from [3]. Our argument here is different and simpler.
It was proved in [29] 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 $a, b \in X$,

$$2\|a\|^p + \frac{2}{K^p}\|b\|^p \leq \|a + b\|^p + \|a - b\|^p. \tag{10}$$

**Proposition 2.1.** Let $\{X_t\}_{t \in \mathbb{Z}}$ be random variables taking values in a set $\Omega$. For every $s \in \mathbb{Z}$ let $\{\tilde{X}_t(s)\}_{t \in \mathbb{Z}}$ be random variables taking values in $\Omega$, with the following property:

$$\forall r \leq s \leq t, \ (X_r, X_t) \text{ and } (X_r, \tilde{X}_t(s)) \text{ have the same distribution.} \tag{11}$$

Fix $p \geq 2$ and let $(X, \| \cdot \|)$ be a Banach space whose norm satisfies (10). Then for every $f : \Omega \to X$ we have

$$\sum_{k=0}^{\infty} \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \left\| f(X_t) - f(\tilde{X}_t(t - 2^k)) \right\|^p \right] \leq (4K)^p \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \left\| f(X_t) - f(X_{t-1}) \right\|^p \right]. \tag{12}$$

**Remark 2.2.** Observe that condition (11) holds when $\{X_t\}_{t \in \mathbb{Z}}$ is a Markov chain on a state space $\Omega$, and $\{\tilde{X}_t(s)\}_{t \in \mathbb{Z}}$ is as in Definition 1.2.

We start by proving a useful inequality that is a simple consequence of (11).

**Lemma 2.3.** Let $X$ be a Banach space whose norm satisfies (10). Then for every $x, y, z, w \in X$,

$$\frac{\|x - w\|^p + \|x - z\|^p}{2^{p-1}} + \frac{\|z - w\|^p}{4^{p-1}K^p} \leq \|y - w\|^p + \|z - y\|^p + 2\|y - x\|^p. \tag{13}$$

**Proof.** For every $x, y, z, w \in X$, (10) implies that

$$\frac{\|x - w\|^p}{2^{p-1}} + \frac{2}{K^p} \left\| y - \frac{x + w}{2} \right\|^p \leq \|y - x\|^p + \|y - w\|^p,$$

and

$$\frac{\|z - x\|^p}{2^{p-1}} + \frac{2}{K^p} \left\| y - \frac{z + x}{2} \right\|^p \leq \|z - y\|^p + \|y - x\|^p.$$ 

Summing these two inequalities, and applying the convexity of the map $u \mapsto \|u\|^p$, we see that

$$\|y - w\|^p + \|z - y\|^p + 2\|y - x\|^p \geq \frac{\|x - w\|^p + \|z - x\|^p}{2^{p-1}} + \frac{4}{K^p} \cdot \left( \left\| y - \frac{x + w}{2} \right\|^p + \left\| y - \frac{z + x}{2} \right\|^p \right) \geq \frac{\|x - w\|^p + \|z - x\|^p}{2^{p-1}} + \frac{4}{K^p} \cdot \left\| \frac{z - w}{4} \right\|^p,$$

implying (13). \qed

**Proof of Proposition 2.1.** Using Lemma 2.3 we see that for every $t \in \mathbb{Z}$ and $k \in \mathbb{N}$,

$$\frac{\|f(X_t) - f(X_{t-2^k})\|^p}{2^{p-1}} + \frac{\|f(\tilde{X}_t(t - 2^k)) - f(X_{t-2^k})\|^p}{4^{p-1}K^p} \leq \|f(X_{t-2^k}) - f(X_t)\|^p + \|f(X_{t-2^k-1}) - f(\tilde{X}_t(t - 2^k))\|^p + 2\|f(X_{t-2^k-1}) - f(X_{t-2^k})\|^p.$$ 

10
Taking expectation, and using the assumption (11), we get
\[
\frac{\mathbb{E} \left[ \| f(X_t) - f(X_{t-2^k}) \|^p \right]}{2^{kp-2}} + \frac{\mathbb{E} \left[ \| f(X_t) - f(\tilde{X}_t(t - 2^{k-1})) \|^p \right]}{4^{p-1}K^p} \leq 2\mathbb{E} \left[ \| f(X_{t-2^{k-1}}) - f(X_t) \|^p \right] + 2\mathbb{E} \left[ \| f(X_{t-2^{k-1}}) - f(X_{t-2^k}) \|^p \right].
\]
Dividing by $2^{(k-1)p+2}$ this becomes
\[
\frac{\mathbb{E} \left[ \| f(X_t) - f(X_{t-2^k}) \|^p \right]}{2^{kp}} + \frac{\mathbb{E} \left[ \| f(X_t) - f(\tilde{X}_t(t - 2^{k-1})) \|^p \right]}{2^{(k+1)p}K^p} \leq \frac{\mathbb{E} \left[ \| f(X_{t-2^{k-1}}) - f(X_t) \|^p \right]}{2^{(k-1)p+1}} + \frac{\mathbb{E} \left[ \| f(X_{t-2^{k-1}}) - f(X_{t-2^k}) \|^p \right]}{2^{(k-1)p+1}}.
\]
Summing this inequality over $k = 1, \ldots, m$ and $t \in \mathbb{Z}$ we get
\[
\sum_{k=1}^m \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_t) - f(X_{t-2^k}) \|^p \right] + \sum_{k=1}^m \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_t) - f(\tilde{X}_t(t - 2^{k-1})) \|^p \right] \leq \sum_{k=1}^m \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_{t-2^{k-1}}) - f(X_t) \|^p \right] + \sum_{k=1}^m \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_{t-2^{k-1}}) - f(X_{t-2^k}) \|^p \right] = \sum_{j=0}^{m-1} \sum_{s \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_s) - f(X_{s-2^j}) \|^p \right].
\]

It is only of interest to prove (12) when $\sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_t) - f(X_{t-1}) \|^p \right] < \infty$. By the triangle inequality, this implies that for every $k \in \mathbb{N}$ we have $\sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_t) - f(X_{t-2^k}) \|^p \right] < \infty$. We may therefore cancel terms in (14), arriving at the following inequality:
\[
\sum_{k=1}^m \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_t) - f(\tilde{X}_t(t - 2^{k-1})) \|^p \right] \leq \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_t) - f(X_{t-1}) \|^p \right] - \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_t) - f(X_{t-2^m}) \|^p \right] \leq \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \| f(X_t) - f(X_{t-1}) \|^p \right].
\]

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

Proposition 2.1 now follows by letting $m \to \infty$. \hfill \square

We next prove the more interesting direction of the equivalence of $p$-convexity and Markov $p$-convexity: a Markov $p$-convex Banach space is also $p$-convex.
Theorem 2.4. Let \((X, \| \cdot \|)\) be a Banach space which is Markov \(p\)-convex with constant \(\Pi\). 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

\[
\frac{\|x + y\|}{2} \leq \frac{\|x\|^p + \|y\|^p}{2} - \frac{1 - (1 - \varepsilon)^p}{4\Pi^p(p + 1)} \frac{\|x - y\|^p}{2}.
\]

Thus the norm \(\| \cdot \|\) satisfies (10) with constant \(K = O\left(\frac{\Pi}{\varepsilon^{1/p}}\right)\).

Proof. The fact that \(X\) is Markov \(p\)-convex with constant \(\Pi\) implies that for every Markov chain \(\{X_t\}_{t \in \mathbb{Z}}\) with values in \(X\), and for every \(m \in \mathbb{N}\), we have

\[
\sum_{k=0}^{m} \sum_{t=1}^{2^m} \mathbb{E}\left[\frac{\|X_t - \tilde{X}_t(t - 2^k)\|^p}{2^kp}\right] \leq \Pi^p \sum_{t=1}^{2^m} \mathbb{E}\left[\|X_t - X_{t-1}\|^p\right]. \tag{15}
\]

For \(x \in X\) we shall say that a Markov chain \(\{X_t\}_{t=-\infty}^{2m}\) 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

\[
\|x\|_m = \inf \left\{ \left( \frac{1}{2^m} \sum_{t=1}^{2^m} \mathbb{E}\left[\|X_t - X_{t-1}\|^p\right] - \frac{\eta}{\Pi^p} \sum_{k=0}^{m} \sum_{t=1}^{2^m} \mathbb{E}\left[\frac{\|X_t - \tilde{X}_t(t - 2^k)\|^p}{2^kp}\right] \right)^{1/p} \right\}, \tag{16}
\]

where the infimum in (16) is taken over all \(m\)-admissible representations of \(x\). Observe that an \(m\)-admissible 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 \(\{X_t\}_{t=-\infty}^{2m}\) is an \(m\)-admissible representation of \(x\) then

\[
\sum_{t=1}^{2^m} \mathbb{E}\left[\|X_t - X_{t-1}\|^p\right] - \frac{\eta}{\Pi^p} \sum_{k=0}^{m} \sum_{t=1}^{2^m} \mathbb{E}\left[\frac{\|X_t - \tilde{X}_t(t - 2^k)\|^p}{2^kp}\right] \geq (1 - \eta) \sum_{t=1}^{2^m} \mathbb{E}\left[\|X_t - X_{t-1}\|^p\right] \geq (1 - \varepsilon)^p \sum_{t=1}^{2^m} \|tx - (t - 1)x\|^p = 2^m(1 - \varepsilon)^p \|x\|^p, \tag{17}
\]

where in the first inequality of (17) we used the convexity of the function \(z \mapsto \|z\|^p\). In conclusion, we see that for all \(x \in X\),

\[(1 - \varepsilon)\|x\| \leq \|x\|_m \leq \|x\|. \tag{18}
\]

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

\[
\sum_{t=1}^{2^m} \mathbb{E}\left[\|X_t - X_{t-1}\|^p\right] - \frac{\eta}{\Pi^p} \sum_{k=0}^{m} \sum_{t=1}^{2^m} \mathbb{E}\left[\frac{\|X_t - \tilde{X}_t(t - 2^k)\|^p}{2^kp}\right] \leq 2^m(\|x\|^p_m + \delta), \tag{19}
\]

12
and
\[ \sum_{t=1}^{2^m} \mathbb{E} \left[ \|Y_t - Y_{t-1}\|^p \right] - \frac{\eta}{\Pi p} \sum_{k=0}^{m} \sum_{t=1}^{2^m} \mathbb{E} \left[ \|Y_t - \tilde{Y}_t(t - 2^k)\|^p \right] \leq 2^m \left( \|y\|^p_{m \odot} + \delta \right). \tag{20} \]

Define a Markov chain \( \{Z_t\}_{t=-\infty}^{2^m+1} \subset X \) as follows. For \( t \leq -2^m \) set \( Z_t = 0 \). With probability \( \frac{1}{2} \) let \( (Z_{2^m+1}, Z_{2^m+2}, \ldots, Z_{2^m+1}) \) equal
\[ (0, \ldots, 0, X_1, X_2, \ldots, X_{2^m-1}, X_{2^m} + Y_1, X_{2^m} + Y_2, \ldots, X_{2^m} + Y_{2^m}) \]
and with probability \( \frac{1}{2} \) let \( (Z_{2^m+1}, Z_{2^m-2}, \ldots, Z_{2^m+1}) \) equal
\[ (0, \ldots, 0, Y_1, Y_2, \ldots, Y_{2^m}, X_1 + Y_{2^m}, X_2 + Y_{2^m}, \ldots, X_{2^m} + Y_{2^m}) \]

Hence, \( Z_t = 0 \) for \( t \leq 0 \), for \( t \in \{1, \ldots, 2^m\} \) we have \( \mathbb{E}[Z_t] = \frac{\mathbb{E}[X_t] + \mathbb{E}[Y_t]}{2} = t \cdot \frac{x+y}{2} \), and for \( t \in \{2^m+1, \ldots, 2^{m+1}\} \) we have
\[ \mathbb{E}[Z_t] = \frac{\mathbb{E}[X_{t-2^m}] + \mathbb{E}[X_{t-2^m} + Y_{2^m}]}{2} = \frac{2^m x + (t-2^m) y + (t-2^m) x + 2^m y}{2} = t \cdot \frac{x+y}{2}. \]

Thus \( \{Z_t\}_{t=-\infty}^{2^m+1} \) is an \((m+1)\)-admissible representation of \( \frac{x+y}{2} \). The definition \( (16) \) implies that
\[ 2^{m+1} \left\| \frac{x+y}{2} \right\|_{m+1}^p \leq \sum_{t=1}^{2^{m+1}} \mathbb{E} \left[ \|Z_t - Z_{t-1}\|^p \right] - \frac{\eta}{\Pi p} \sum_{k=0}^{m+1} \sum_{t=1}^{2^{m+1}} \mathbb{E} \left[ \|Z_t - \tilde{Z}_t(t - 2^k)\|^p \right]. \tag{21} \]

Note that by definition,
\[ \sum_{t=1}^{2^{m+1}} \mathbb{E} \left[ \|Z_t - Z_{t-1}\|^p \right] = \sum_{t=1}^{2^m} \mathbb{E} \left[ \|X_t - X_{t-1}\|^p \right] + \sum_{t=1}^{2^m} \mathbb{E} \left[ \|Y_t - Y_{t-1}\|^p \right]. \tag{22} \]

Moreover,
\[ \sum_{k=0}^{m+1} \sum_{t=1}^{2^{m+1}} \mathbb{E} \left[ \|Z_t - \tilde{Z}_t(t - 2^k)\|^p \right] \]
\[ = \frac{1}{2^{(m+1)p}} \sum_{t=1}^{2^{m+1}} \mathbb{E} \left[ \|Z_t - \tilde{Z}_t(t - 2^{m+1})\|^p \right] + \sum_{k=0}^{m+1} \sum_{t=1}^{2^m} \mathbb{E} \left[ \|Z_t - \tilde{Z}_t(t - 2^k)\|^p \right]. \tag{23} \]

We bound each of the terms in \( (23) \) separately. Note that by construction we have for every \( t \in \{1, \ldots, 2^m\} \),
\[ Z_t - \tilde{Z}_t(t - 2^{m+1}) = Z_t - \tilde{Z}_t(1 - 2^{m+1}) = \begin{cases} X_t - Y_t & \text{with probability } 1/4, \\ Y_t - X_t & \text{with probability } 1/4, \\ X_t - \tilde{X}_t(1) & \text{with probability } 1/4, \\ Y_t - \tilde{Y}_t(1) & \text{with probability } 1/4. \end{cases} \]
Thus, the first term in the right hand side of (23) can be bounded from below as follows:

\[
\frac{1}{2^{(m+1)p}} \sum_{t=1}^{2^{m+1}} \mathbb{E} \left[ \left\| Z_t - \tilde{Z}_t(t - 2^{m+1}) \right\|^p \right] \geq \frac{1}{2^{(m+1)p+1}} \sum_{t=1}^{2^m} \mathbb{E} \left[ \left\| X_t - Y_t \right\|^p \right] \\
\geq \frac{1}{2^{(m+1)p+1}} \sum_{t=1}^{2^m} \mathbb{E} \left[ X_t \right] - \mathbb{E} \left[ Y_t \right] \left\|^p \right. = \frac{\|x - y\|^p}{2^{(m+1)p+1}} \sum_{t=1}^{2^m} t^p \geq \frac{2^m \|x - y\|^p}{2^{p+1}(p + 1)}. \tag{24}
\]

We now proceed to bound from below the second term in the right hand side of (23). Note first that for every \( k \in \{0, \ldots, m\} \) and every \( t \in \{2^m + 1, \ldots, 2^{m+1}\} \) we have

\[
Z_t - \tilde{Z}_t(t - 2^k) = \begin{cases} 
(X_{2^m + Y_{t-2^m}}) - (X_{2^m + Y_{t-2^m}}) & \text{with probability 1/2}, \\
(Y_{2^m + X_{t-2^m}}) - (Y_{2^m + X_{t-2^m}}) & \text{with probability 1/2}.
\end{cases}
\]

By Jensen’s inequality, if \( U, V \) are \( X \)-valued independent random variables with \( \mathbb{E}[V] = 0 \), then

\[
\mathbb{E}[\|U + V\|^p] \geq \mathbb{E}[\|U + \mathbb{E}[V]\|^p] = \mathbb{E}[\|U\|^p].
\]

Thus, since \( \{X_t\}_{t=-\infty}^{2^m} \) and \( \{Y_t\}_{t=-\infty}^{2^m} \) are independent,

\[
\mathbb{E} \left[ \left\| Y_{t-2^m} - \tilde{Y}_{t-2^m}(t - 2^m - 2^k) + X_{2^m} - \tilde{X}_{2^m}(t - 2^k) \right\|^p \right] \\
\geq \mathbb{E} \left[ \left\| Y_{t-2^m} - \tilde{Y}_{t-2^m}(t - 2^m - 2^k) \right\|^p \right],
\]

and

\[
\mathbb{E} \left[ \left\| X_{t-2^m} - \tilde{X}_{t-2^m}(t - 2^m - 2^k) + Y_{2^m} - \tilde{Y}_{2^m}(t - 2^k) \right\|^p \right] \\
\geq \mathbb{E} \left[ \left\| X_{t-2^m} - \tilde{X}_{t-2^m}(t - 2^m - 2^k) \right\|^p \right].
\]

It follows that for every \( k \in \{0, \ldots, m\} \) and every \( t \in \{2^m + 1, \ldots, 2^{m+1}\} \) we have

\[
\mathbb{E} \left[ \left\| Z_t - \tilde{Z}_t(t - 2^k) \right\|^p \right] \\
\geq \frac{1}{2} \mathbb{E} \left[ \left\| X_{t-2^m} - \tilde{X}_{t-2^m}(t - 2^m - 2^k) \right\|^p \right] + \frac{1}{2} \mathbb{E} \left[ \left\| Y_{t-2^m} - \tilde{Y}_{t-2^m}(t - 2^m - 2^k) \right\|^p \right]. \tag{25}
\]

Hence,

\[
\sum_{k=0}^{m} \sum_{t=1}^{2^{m+1}} \mathbb{E} \left[ \left\| Z_t - \tilde{Z}_t(t - 2^k) \right\|^p \right] \\
\geq \sum_{k=0}^{m} \sum_{t=1}^{2^m} \frac{1}{2} \mathbb{E} \left[ \left\| X_t - \tilde{X}_t(t - 2^k) \right\|^p \right] + \frac{1}{2} \mathbb{E} \left[ \left\| Y_t - \tilde{Y}_t(t - 2^k) \right\|^p \right] \\
+ \sum_{k=0}^{m} \sum_{t=2^m+1}^{2^{m+1}} \frac{1}{2} \left[ \mathbb{E} \left[ \left\| X_{t-2^m} - \tilde{X}_{t-2^m}(t - 2^m - 2^k) \right\|^p \right] + \frac{1}{2} \mathbb{E} \left[ \left\| Y_{t-2^m} - \tilde{Y}_{t-2^m}(t - 2^m - 2^k) \right\|^p \right] \right] \\
= \sum_{k=0}^{m} \sum_{t=1}^{2^m} \mathbb{E} \left[ \left\| X_t - \tilde{X}_t(t - 2^k) \right\|^p \right] + \sum_{k=0}^{m} \sum_{t=1}^{2^m} \mathbb{E} \left[ \left\| Y_t - \tilde{Y}_t(t - 2^k) \right\|^p \right]. \tag{26}
\]

Combining (19), (20), (21), (22), (23), (24) and (26), and letting \( \delta \) tend to 0, we see that

\[
2^{m+1} \left\| \frac{x + y}{2} \right\|^p \leq 2^m \|x\|^p + 2^m \|y\|^p - \frac{\eta}{2^{m+1}(p + 1)}.
\]
or,
\[
\left\| \frac{x + y}{2} \right\|_m^{p+1} \leq \frac{\|x\|_m^p + \|y\|_m^p}{2} - \frac{\eta}{4p(p+1)} \left\| \frac{x - y}{2} \right\|_m^p.
\]  \hspace{1cm} (27)

Define for \( w \in X \),
\[
\|w\| = \limsup_{m \to \infty} \|w\|_m.
\]

Then a combination of (18) and (27) yields that
\[
(1 - \varepsilon)\|x\| \leq \|w\| \leq \|x\|,
\]
and
\[
\left\| \frac{x + y}{2} \right\|_m^{p+1} \leq \frac{\|x\|_m^p + \|y\|_m^p}{2} - \frac{\eta}{4p(p+1)} \left\| \frac{x - y}{2} \right\|_m^p
\leq \frac{\|x\|_m^p + \|y\|_m^p}{2} - \frac{\eta}{4p(p+1)} \left\| \frac{x - y}{2} \right\|_m^p.
\]  \hspace{1cm} (28)

Note that (28) implies that the set \( \{x \in X : \|w\| \leq 1\} \) is convex, so that \( \| \cdot \| \) is a norm on \( X \). This concludes the proof of Theorem 2.4. \( \square \)

3 A doubling space which is not Markov \( p \)-convex for any \( p \in (0, \infty) \)

Consider the Laakso graphs \( \{G_i\}_{i=0}^\infty \), which are defined as follows. \( G_0 \) is the graph consisting of one edge of unit length. 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. Note that each edge of \( G_i \) has length \( 4^{-i} \); we denoted the resulting shortest path metric on \( G_i \) by \( d_{G_i} \). As shown in [13 Thm. 2.3], the doubling constant of metric space \((G_i, d_{G_i})\) is at most 6.

We direct \( G_m \) as follows. Define the root of \( G_m \) to be (an arbitrarily chosen) one of the two vertices having only one adjacent edge. In the figure 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 then 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 the standard random walk on the directed graph \( G_m \), starting from the root. This random walk is extended to \( t \in \mathbb{Z} \) by stipulating that \( X_t = X_0 \) for \( t < 0 \), and \( X_t = X_t^{4^m} \) for \( t > 4^m \).

**Proposition 3.1.** For the random walk defined above,
\[
\sum_{k=0}^{2m} \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \frac{d_{G_m}(X_t, \tilde{X}_t(t-2^k))^p}{2kp} \right] \geq \frac{m}{8p} \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_{G_m}(X_t, X_{t-1})^p \right].
\]  \hspace{1cm} (29)
Proof. For every $t \in \mathbb{Z}$ we have,
\[
\mathbb{E} \left[ d_{G_m}(X_t, X_{t-1})^p \right] = \begin{cases} 
4^{-mp} & t \in \{0, \ldots, 4^m - 1\}, \\
0 & \text{otherwise}.
\end{cases}
\]
Hence,
\[
\sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_{G_m}(X_t, X_{t-1})^p \right] = 4^{-m(p-1)}.
\] (30)

Fix $k \in \{0, \ldots, 2m - 2\}$ and write $h = \lfloor k/2 \rfloor$. View $G_m$ as being built from $A = G_{m-h}$, where each edge of $A$ has been replaced by a copy of $G_h$. Note that for every $i \in \{0, \ldots, 4^{m-h-1} + 1\}$, at time $t = (4i + 1)4^h$ the walk $X_i$ is at a vertex of $G_m$ which has two outgoing edges, corresponding to distinct copies of $G_h$. To see this it suffices to show that all vertices of $G_{m-h}$ that are exactly $(4i + 1)$ edges away from the root, have out-degree 2. This fact is true since $G_{m-h}$ is obtained from $G_{m-h-1}$ by replacing each edge by a copy of $G_1$, and each such copy of $G_1$ contributes one vertex of out-degree 2, corresponding to the vertex labeled $b$ in the figure describing $G_1$.

Consider the set of times
\[
T_k \overset{\text{def}}{=} \{0, \ldots, 4^m - 1\} \cap \left( \bigcup_{i=0}^{4^{m-h-1}+1} \left( (4i + 1)4^h + 4^h - 2, (4i + 1)4^h + 2 \cdot 4^h - 2 \right) \right).
\]

For $t \in T_k$ find $i \in \{0, \ldots, 4^{m-h-1} + 1\}$ such that $t \in \left( (4i + 1)4^h + 4^h - 2, (4i + 1)4^h + 2 \cdot 4^h - 2 \right]$. Since, by the definition of $h$, we have $t - 2^k \in \left( (4i + 1)4^h - 4^h, (4i + 1)4^h \right)$, the walks $\{X_s\}_{s \in \mathbb{Z}}$ and $\{	ilde{X_s}(t - 2^k)\}_{s \in \mathbb{Z}}$ started evolving independently at some vertex lying in a copy of $G_h$ preceding a vertex $v$ of $G_m$ which has two outgoing edges, corresponding to distinct copies of $G_h$. Thus, with probability at least $\frac{1}{2}$, the walks $X_t$ and $\tilde{X_t}(t - 2^k)$ lie on two distinct copies of $G_h$ in $G_m$, immediately following the vertex $v$, and at distance at least $4^h - 2 \cdot 4^m$ and at most $2 \cdot 4^h - 2 \cdot 4^m$ from $v$. Hence, with probability at least $\frac{1}{2}$ we have $d_{G_m} \left( X_t, \tilde{X_t}(t - 2^k) \right) \geq 2 \cdot 4^h - 2 \cdot 4^m = 2^{2h-3-2m}$, and therefore,
\[
\mathbb{E} \left[ d(X_t, \tilde{X_t}(t - 2^k))^p \right] \geq \frac{1}{2} \left( 2^{2h-3-2m} \right)^p \geq 2^{-(2m+3)p-1}.
\]

We deduce that for all $k \in \{0, \ldots, 2m - 2\}$,
\[
\sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d(X_t, \tilde{X_t}(t - 2^k))^p \right] \geq \sum_{t \in T_k} \mathbb{E} \left[ d(X_t, \tilde{X_t}(t - 2^k))^p \right] \geq |T_k| \cdot 2^{-(2m+3)p-1} \geq 4^{h-2} \cdot 4^{m-h-1} \cdot 2^{-(2m+3)p-1} \geq \frac{1}{8^p} \cdot 4^{-m(p-1)}.
\] (31)

A combination of (30) and (31) implies (29). \hfill \Box

Proof of Theorem 7.3 As explained in [12] [13], by passing to an appropriate Gromov-Hausdorff limit, there exists a doubling metric space $(X, d_X)$ that contains an isometric copy of all the Laakso graphs $\{G_m\}_{m=0}^{\infty}$. Proposition 3.1 therefore implies that $X$ is not Markov $p$-convex for any $p \in (0, \infty)$.

\hfill \Box
Proof of Theorem 1.6. Let \((X, d_X)\) be a Markov \(p\)-convex metric space, i.e., \(\Pi_p(X) < \infty\). Assume that \(f : G_m \to X\) satisfies

\[
x, y \in G_m \implies \frac{1}{A} d_{G_m}(x, y) \leq d_X(f(x), f(y)) \leq Bd_{G_m}(x, y).
\] (32)

Let \(\{X_t\}_{t \in \mathbb{Z}}\) be the random walk from Proposition 4.1. Then

\[
\frac{m}{8^p A^p} \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_{G_m}(X_t, X_{t-1})^p \right] \leq \frac{1}{A^p} \sum_{k=0}^{2m} \sum_{t \in \mathbb{Z}} \frac{\mathbb{E} \left[ d_{G_m}(X_t, \bar{X}_t(t - 2k))^p \right]}{2^{kp}}
\] (33)

\[
\leq \sum_{k=0}^{2m} \sum_{t \in \mathbb{Z}} \frac{\mathbb{E} \left[ d_X(f(X_t), f(\bar{X}_t(t - 2k)))^p \right]}{2^{kp}}
\] (34)

\[
\leq \Pi_p(X)^p \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_X(f(X_t), f(X_{t-1}))^p \right]
\] (35)

\[
\leq \Pi_p(X)^p B^p \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_{G_m}(X_t, X_{t-1})^p \right].
\] (36)

Thus \(AB \geq m^{1/p} \geq (\log |G_m|)^{1/p}\).

\[\Box\]

4 Lipschitz quotients

Say that a metric space \((Y, d_Y)\) is a \(D\)-Lipschitz quotient of a metric space \((X, d_X)\) if there exist \(a, b > 0\) with \(ab \leq D\) and a mapping \(f : X \to Y\) such that for all \(x \in X\) and \(r > 0\),

\[
B_Y \left( f(x), \frac{r}{a} \right) \subseteq f(B_X(x, r)) \subseteq B_Y(f(x), br).
\] (33)

Observe that the last inclusion in (33) is equivalent to the fact that \(f\) is \(b\)-Lipschitz.

The following proposition implies Theorem 1.6.

Proposition 4.1. If \((Y, D_Y)\) is a \(D\)-Lipschitz quotient of \((X, d_X)\) then \(\Pi_p(Y) \leq D \cdot \Pi_p(X)\).

Proof. Fix \(f : X \to Y\) satisfying (34). Also, fix a Markov chain \(\{X_t\}_{t \in \mathbb{Z}}\) on a state space \(\Omega\), and a mapping \(g : \Omega \to Y\).

Fix \(m \in \mathbb{Z}\) and let \(\Omega^*\) be the set of finite sequences of elements of \(\Omega\) starting at time \(m\), i.e., the set of sequences of the form \((\omega_i)_{i=m}^{t-m+1} \in \Omega^{t-m+1}\) for all \(t \geq m\). It will be convenient to consider the Markov chain \(\{X_t^*\}_{t=m}^\infty\) on \(\Omega^*\) which is given by:

\[
\Pr[X_t^* = (\omega_m, \omega_{m+1}, \ldots, \omega_t)] = \Pr[X_m = \omega_m, X_{m+1} = \omega_{m+1}, \ldots, X_t = \omega_t].
\]

Also, define \(g^* : \Omega^* \to Y\) by \(g^*(\omega_1, \ldots, \omega_t) = g(\omega_t)\). By definition, \(\{g^*(X_t^*)\}_{t=m}^\infty\) and \(\{g(X_t)\}_{t=m}^\infty\) are identically distributed.

We next define a mapping \(h^* : \Omega^* \to X\) such that \(f \circ h^* = g^*\) and for all \((\omega_m, \ldots, \omega_t) \in \Omega^*\),

\[
d_X(h^*(\omega_m, \ldots, \omega_{t-1}), h^*(\omega_m, \ldots, \omega_t)) \leq ad_Y(g(\omega_{t-1}), g(\omega_t)).
\] (34)

17
For $\omega^* \in \Omega^*$, we will define $h^*(\omega^*)$ by induction on the length of $\omega^*$. If $\omega^* = (\omega_m)$, then we fix $h^*(\omega^*)$ to be an arbitrary element in $f^{-1}(g(\omega_m))$. Assume that $\omega^* = (\omega_m, \ldots, \omega_{t-1}, \omega_t)$ and that $h^*(\omega_m, \ldots, \omega_{t-1})$ has been defined. Set $x = f(h^*(\omega_m, \ldots, \omega_{t-1})) = g^*(\omega_m, \ldots, \omega_{t-1}) = g(\omega_{t-1})$ and $r = adY(g(\omega_{t-1}), g(\omega_t))$. Since $g(\omega_t) \in B_Y(x, r/a)$, it follows from (33) there exists $y \in X$ such that $f(y) = g(\omega_t)$, and $d_X(x, y) \leq r$. We then define $h^*((\omega_m, \ldots, \omega_{t-1}, \omega_t)) = y$.

Write $X_t^* = X^*_m$ for $t \leq m$. By the Markov $p$-convexity of $(X, d_X)$, we have

$$
\sum_{k=0}^{\infty} \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \frac{d_X(h^*(X_t^*), h^*(\tilde{X}_t^*(t-2^k)))}{2^{kp}} \right] \leq \Pi_p(X)^p \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_X(h^*(X_t^*), h^*(X_{t-1}^*)) \right].
$$

By (34) we have for every $t \geq m + 1$,

$$
d_X(h^*(X_t^*), h^*(X_{t-1}^*)) \leq adY(g(X_t), g(X_{t-1})),
$$

while for $t \leq m$ we have $d_X(h^*(X_t^*), h^*(X_{t-1}^*)) = 0$. Thus,

$$
\sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_X(h^*(X_t^*), h^*(X_{t-1}^*)) \right] \leq a^p \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_Y(g(X_t), g(X_{t-1})) \right].
$$

At the same time, using the fact that $f$ is $b$-Lipschitz and $f \circ h^* = g^*$, we see that if $t \geq m + 2^k$,

$$
d_X(h^*(X_t^*), h^*(\tilde{X}_t^*(t-2^k))) \geq \frac{1}{b}d_Y(f(h^*(X_t^*)), f(h^*(\tilde{X}_t^*(t-2^k))))
= \frac{1}{b}d_Y(g^*(X_t^*), g^*(\tilde{X}_t^*(t-2^k))) = \frac{1}{b}d_Y(g(X_t), g(X_{t-1}(t-2^k)))
$$

Thus,

$$
\sum_{k=0}^{\infty} \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \frac{d_X(h^*(X_t^*), h^*(\tilde{X}_t^*(t-2^k)))}{2^{kp}} \right] \geq \frac{1}{b^p} \sum_{k=0}^{\infty} \sum_{t=m+2^k}^{\infty} \mathbb{E} \left[ \frac{d_Y(g(X_t), g(X_{t-1}(t-2^k)))}{2^{kp}} \right].
$$

By combining (36) and (37) with (35), and letting $m$ tend to $-\infty$, we get the inequality:

$$
\sum_{k=0}^{\infty} \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ \frac{d_Y(g(X_t), g(X_{t-1}(t-2^k)))}{2^{kp}} \right] \leq (ab \Pi_p(X))^p \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_Y(g(X_t), g(X_{t-1})) \right].
$$

Since this inequality holds for every Markov chain $\{X_t\}_{t \in \mathbb{Z}}$ and every $g : \Omega \to Y$, and since $ab \leq D$, we have proved that $\Pi_p(Y) \leq D \Pi_p(X)$, as required.

5 A dichotomy theorem for vertically faithful embeddings of trees

In this section we prove Theorem 1.14. The proof naturally breaks into two parts. The first is the following BD Ramsey property of paths (which can be found non-quantitatively in [19], where also the BD Ramsey terminology is explained).

A mapping $\phi : \mathcal{M} \to \mathcal{N}$ is called a rescaled isometry if $\text{dist}(\phi) = 1$, or equivalently there exists $\lambda > 0$ such that $d_{\mathcal{N}}(\phi(x), \phi(y)) = \lambda d_{\mathcal{M}}(x, y)$ for all $x, y \in \mathcal{M}$. For $n \in \mathbb{N}$ let $P_n$ denote the $n$-path, i.e., the set $\{0, \ldots, n\}$ equipped with the metric inherited from the real line.
Proposition 5.1. Fix $\delta \in (0, 1)$, $D \geq 2$ and $t, n \in \mathbb{N}$ satisfying $n \geq D^{(4t \log t)/\delta}$. If $f : P_n \to X$ satisfies $\text{dist}(f) \leq D$ then there exists a rescaled isometry $\phi : P_t \to P_n$ such that $\text{dist}(f \circ \phi) \leq 1 + \delta$.

Given a metric space $(X, d_X)$ and a nonconstant mapping $f : P_n \to X$, define
\[
T(X, f) = \frac{d_X(f(0), f(n))}{\max_{i \in \{1, \ldots, n\}} d_X(f(i - 1), f(i))} = \frac{d_X(f(0), f(n))}{n\|f\|_{\text{Lip}}}.
\]
If $f$ is a constant mapping (equivalently $\max_{i \in \{1, \ldots, n\}} d_X(f(i - 1), f(i)) = 0$) then we set $T(X, f) = 0$.

Note that by the triangle inequality we always have $T(X, f) \leq 1$.

Lemma 5.2. For every $m, n \in \mathbb{N}$ and $f : P_{mn} \to X$, there exist rescaled isometries $\phi^{(n)} : P_n \to P_{mn}$ and $\phi^{(m)} : P_m \to P_{mn}$, such that
\[
T(X, f) \leq T\left(X, f \circ \phi^{(m)}\right) \cdot T\left(X, f \circ \phi^{(n)}\right).
\]

Proof. Fix $f : P_{mn} \to X$ and define $\phi^{(m)} : P_m \to P_{mn}$ by $\phi^{(m)}(i) = in$. Then,
\[
d_X(f(0), f(mn)) \leq T\left(X, f \circ \phi^{(m)}\right) m \max_{i \in \{1, \ldots, m\}} d_X(f(i - 1)n, f(in)). \tag{38}
\]

Similarly, for every $i \in \{1, \ldots, m\}$ define $\phi^{(n)}_i : P_n \to P_{mn}$ by $\phi^{(n)}_i(j) = (i - 1)n + j$. Then
\[
d_X(f((i - 1)n), f(in)) \leq T\left(X, f \circ \phi^{(n)}_i\right) n \max_{j \in \{1, \ldots, n\}} d_X(f((i - 1)n + j - 1), f((i - 1)n + j)). \tag{39}
\]
Letting $i \in \{1, \ldots, m\}$ be such that $T\left(X, f \circ \phi^{(n)}_i\right)$ is maximal, and $\phi^{(n)} = \phi^{(n)}_i$, we conclude that
\[
d_X(f(0), f(mn)) \leq T\left(X, f \circ \phi^{(m)}\right) T\left(X, f \circ \phi^{(n)}\right) mn \max_{i \in \{1, \ldots, mn\}} d_X(f(i - 1), f(i)). \tag{38} \hfill \square
\]

Lemma 5.3. For every $f : P_m \to X$ we have $\text{dist}(f) \geq 1/T(X, f)$.

Proof. Assuming $a|i - j| \leq d_X(f(i), f(j)) \leq b|i - j|$ for all $i, j \in P_m$, the claim is $bT(X, f) \geq a$. Indeed, $am \leq d_X(f(0), f(m)) \leq T(X, f)m \max_{i = \varepsilon \in \{1, \ldots, m\}} d_X(f(i - 1), f(i)) \leq T(X, f)b$m. \hfill \square

Lemma 5.4. Fix $f : P_m \to X$. If $0 < \varepsilon < 1/m$ and $T(X, f) \geq 1 - \varepsilon$, then $\text{dist}(f) \leq 1/(1 - m\varepsilon)$.

Proof. Denote $b = \max_{i \in \{1, \ldots, m\}} d_X(f(i), f(i - 1)) > 0$. For every $0 \leq i < j \leq m$ we have $d_X(f(i), f(j)) \leq \sum_{\ell = i + 1}^{j} d_X(f(\ell - 1), f(\ell)) \leq b|j - i|$, and
\[
(1 - \varepsilon)mb \leq T(X, f)mb = d_X(f(0), f(m)) \leq d_X(f(0), f(i)) + d_X(f(i), f(j)) + d_X(f(j), f(m)) \leq d_X(f(i), f(j)) + b(m + i - j).
\]
Thus $d_X(f(i), f(j)) \geq b|j - i - m\varepsilon| \geq (1 - m\varepsilon)|j - i|$. \hfill \square

Proof of Proposition 5.1. Set $k = \lfloor \log_t n \rfloor$ and denote by $I$ the identity mapping from $P_k$ to $P_n$. By Lemma 5.3 we have $T(X, f \circ I) \geq 1/D$. An iterative application of Lemma 5.2 implies that there exists a rescaled isometry $\phi : P_t \to P_k$ such that
\[
T(X, f \circ I \circ \phi) \geq D^{-1/k} \geq e^{-2\log D/\log_t n} \geq e^{-\delta/(2t)} \geq 1 - \frac{\delta}{2t}.
\]
By Lemma 5.4 we therefore have $\text{dist}(f \circ I \circ \phi) \leq 1/(1 - \delta/2) \leq 1 + \delta$. \hfill \square
The second part of the proof of Theorem 1.14 uses the following combinatorial lemma due to Matoušek [18]. Denote by $T_{k,m}$ the complete rooted tree of height $m$, in which every non-leaf vertex has $k$ children. For a rooted tree $T$, denote by $\text{SP}(T)$ the set of all unordered pairs $\{x,y\}$ of distinct vertices of $T$ such that $x$ is an ancestor of $y$.

**Lemma 5.5** ([18 Lem. 5]). Let $m, r, k \in \mathbb{N}$ satisfy $k \geq r^{(m+1)^2}$. Suppose that each of the pairs from $\text{SP}(T_{k,m})$ is colored by one of $r$ colors. Then there exists a copy $T'$ of $B_m$ in this $T_{k,m}$ such that the color of any pair $\{x,y\}$ in $\text{SP}(T')$ only depends on the levels of $x$ and $y$.

**Proof of Lemma 1.14** Let $f: B_n \to X$ be a $D$-vertically faithful embedding, i.e., for some $\lambda > 0$ it satisfies

$$\lambda d_{B_n}(x,y) \leq d_X(f(x),f(y)) \leq D\lambda d_{B_n}(x,y)$$

whenever $x,y \in B_n$ are such that $x$ is an ancestor of $y$.

Let $k, \ell \in \mathbb{N}$ be auxiliary parameters to be determined later, and define $m = \lceil n/(k\ell) \rceil$. We first construct a mapping $g: T_{2k,m} \to B_n$ in a top-down manner as follows. If $r$ is the root of $T_{2k,m}$ then $g(r)$ is defined to be the root of $B_n$. Having defined $g(u)$, let $v_1, \ldots, v_{2k} \in T_{2k,m}$ be the children of $u$, and let $w_1, \ldots, w_{2k} \in B_n$ be the descendants of $g(u)$ at depth $k$ below $g(u)$. For each $i \in \{1, \ldots, 2^k\}$ let $g(v_i)$ be an arbitrary descendant of $w_i$ at depth $h(g(u)) + \ell k$. Note that for this construction to be possible we need to have $mtk \leq n$, which is ensured by our choice of $m$.

By construction, if $x, y \in T_{2k,m}$ and $x$ is an ancestor of $y$, then $g(x)$ is an ancestor of $g(y)$ and $d_{B_n}(g(x),g(y)) = \ell kd_{T_{2k,m}}(x,y)$. Also, if $x, y \in T_{2k,m}$ and $\text{lca}(x,y) = u$, then we have

$$h(\text{lca}(g(x),g(y))) \in \{h(g(u)), h(g(u)) + 1, \ldots, h(g(u)) + k - 1\}.$$ 

This implies that

$$\ell d_{T_{2k,m}}(x,y) \leq d_{B_n}(x,y) \leq \ell kd_{T_{2k,m}}(x,y).$$

Thus, assuming $\ell \geq 2$, we have $\text{dist}(g) \leq 1 + 2/\ell$. Moreover, denoting $F = f \circ g$ and using (40), we see that if $x, y \in T_{2k,m}$ are such that $x$ is an ancestor of $y$ then

$$k\ell \lambda d_{T_{2k,m}}(x,y) \leq d_X(F(x),F(y)) \leq D\ell k\lambda d_{T_{2k,m}}(x,y).$$

(41)

Color every pair $\{x,y\} \in \text{SP}(T_{2k,m})$ with the color

$$\chi(\{x,y\}) = \left\lfloor \log_{1+\delta/4} \left( \frac{d_X(F(x),F(y))}{k\ell \lambda d_{T_{2k,m}}(x,y)} \right) \right\rfloor \in \{1, \ldots, r\},$$

where $r = \lceil \log_{1+\delta/4} D \rceil$. Assuming that

$$2^k \geq r^{(m+1)^2},$$

by Lemma 5.5 there exists a copy $T'$ of $B_m$ in $T_{2k,m}$ such that the colors of pairs $\{x,y\} \in \text{SP}(T')$ only depend on the levels of $x$ and $y$.

Let $P$ be a root-leaf path in $T'$ (isometric to $P_m$). The mapping $F|_P: P \to X$ has distortion at most $D$ by (41). Assuming

$$m \geq D^{16(\log t)/\delta},$$

by Proposition 5.1 there are $\{x_i\}_{i=0}^t \subseteq P$ such that for some $a, b \in \mathbb{N}$ with $a, a + tb \in [0,m]$, for all $i$ we have $h(x_i) = a + ib$, and for some $\theta > 0$, for all $i, j \in \{0, \ldots, t\}$,

$$\theta b|i - j| \leq d_X(F(x_i),F(x_j)) \leq \left(1 + \frac{\delta}{4}\right) \theta b|i - j|.$$
Define a rescaled isometry \( \varphi : B_t \rightarrow T' \) in a top-down manner as follows: \( \varphi(r) = x_0 \), and having defined \( \varphi(u) \in T' \), if \( v, w \) are the children of \( u \) in \( B_t \) and \( v', w' \) are the children of \( \varphi(u) \) in \( T' \), the vertices \( \varphi(v), \varphi(w) \) are chosen as arbitrary descendants in \( T' \) of \( v', w' \) (respectively) at depth \( h(\varphi(u)) + b \). Consider the mapping \( G : B_t \rightarrow X \) given by \( G = F \circ \varphi = f \circ g \circ \varphi \). Take \( x, y \in B_t \) such that \( x \) is an ancestor of \( y \). Write \( h(x) = i \) and \( h(y) = j \). Thus \( h(\varphi(x)) = a + ib \) and \( h(\varphi(y)) = a + jb \). It follows that \( \{\varphi(x), \varphi(y)\} \) is colored by the same color as \( \{x_i, x_j\} \), i.e.,

\[
\left\lfloor \log_{1 + \delta/4} \left( \frac{d_X(G(x), G(y))}{k\ell\lambda bd_B(X, y)} \right) \right\rfloor = \chi(\{\varphi(x), \varphi(y)\}) = \chi(\{x_i, x_j\}) = \left\lfloor \log_{1 + \delta/4} \left( \frac{d_X(F(x_i), F(y_j))}{k\ell \lambda bd_B(X, y)} \right) \right\rfloor.
\]

Consequently, using \( (44) \) we deduce that

\[
\frac{\theta b}{1 + \delta/4} d_B(x, y) \leq d_X(G(x), G(y)) \leq \left( 1 + \frac{\delta}{4} \right)^2 \theta bd_B(x, y).
\]

Thus \( G \) is a \((1 + \delta/4)^3 \leq 1 + \delta \) vertically faithful embedding of \( B_t \) into \( X \).

It remains to determine the values of the auxiliary parameters \( \ell, k \), which will lead to the desired restriction on \( n \) given in \( (1) \). First of all, we want to have \( \text{dist}(g \circ \varphi) \leq 1 + \xi \). Since \( \varphi \) is a rescaled isometry and (for \( \ell \geq 2 \)) \( \text{dist}(g) \leq 1 + 2/\ell \), we choose \( \ell = \lceil 2/\xi \rceil \geq 2 \). We will choose \( k \) so that \( 4k \leq n \xi \), so that \( n/(k \ell) \geq 1 \). Since \( m = \lceil n/(k \ell) \rceil \), we have \( m + 1 \leq n \xi/k \) and \( m \geq n \xi/(4k) \). Recall that \( r = \lceil \log_{1 + \delta/4} D \rceil \leq 2 \log_{1 + \delta/4} D \leq 16D/\delta \). Hence the requirement \( (12) \) will be satisfied if

\[
2^{k^3} \geq \left( \frac{16D}{\delta} \right)^{n^2 \xi^2},
\]

and the requirement \( (43) \) will be satisfied if

\[
\frac{n \xi}{4k} \geq D^{16(t \log t)/\delta}.
\]

There exists an integer \( k \) satisfying both \( (45) \) and \( (46) \) provided that

\[
\left( \frac{n^2 \xi^2 \log_2 \left( \frac{16D}{\delta} \right)}{4D^{16(t \log t)/\delta}} \right)^{1/3} + 1 \leq \frac{n \xi}{4D^{16(t \log t)/\delta}}
\]

which holds true provided the constant \( c \) in \( (9) \) is large enough.

\[\blacksquare\]

6 Tree metrics do not have the dichotomy property

This section is devoted to the proofs of Theorem \( \ref{thm1.1} \) and Theorem \( \ref{thm1.2} \). These proofs were outlined in Section \( \ref{sec1.1.1} \) and we will use the notation introduced there.

6.1 Horizontally contracted trees

We start with the following lemma which supplies conditions on \( \{\varepsilon_n\}_{n=0}^\infty \) ensuring that the \( H \)-tree \((B_\infty, d_\varepsilon)\) is a metric space.

Lemma 6.1. Assume that \( \{\varepsilon_n\}_{n=0}^\infty \subseteq (0, 1] \) is non-increasing and \( \{n\varepsilon_n\}_{n=0}^\infty \) is non-decreasing. Then \( d_\varepsilon \) is a metric on \( B_\infty \).
Proof. Take $x, y, z \in B_\infty$ and without loss of generality assume that $h(x) \leq h(y)$. We distinguish between the cases $h(z) > h(y)$, $h(x) \leq h(z) \leq h(y)$ and $h(z) < h(x)$.

If $h(z) > h(y)$ then
\[
d(x, z) + d(y, z) - d(x, y) = 2[h(z) - h(y)] + 2\varepsilon_h \cdot [h(lca(x, y)) - h(lca(x, z))] + 2\varepsilon_h \cdot [h(y) - h(lca(z, y))]
\geq 2\varepsilon_h \cdot [h(lca(x, y)) - h(lca(x, z))] + 2\varepsilon_h \cdot [h(y) - h(lca(z, y))].
\]

To show that (47) is non-negative observe that this is obvious if $h(lca(x, y)) \geq h(lca(x, z))$. So assume that $h(lca(x, y)) < h(lca(x, z))$. In this case necessarily $h(lca(z, y)) = h(lca(x, y))$, so we can bound (47) from below as follows
\[
2\varepsilon_h \cdot [h(lca(x, y)) - h(lca(x, z))] + 2\varepsilon_h \cdot [h(y) - h(lca(z, y))]
\geq 2\varepsilon_h \cdot [h(lca(x, y)) - h(x)] + 2\varepsilon_h \cdot [h(y) - h(lca(x, y))]
\geq 2\varepsilon_h \cdot h(lca(x, y)) \left(1 - \frac{h(x)}{h(y)}\right) \geq 0.
\]

If $h(z) < h(x)$ then
\[
d(x, z) + d(x, y) = h(x) + h(y) - 2h(z) + 2\varepsilon_h \cdot [h(lca(z, x)) - h(lca(z, y))]
\geq h(y) - h(x) + 2\varepsilon_h \cdot [h(x) - h(z)] + 2\varepsilon_h \cdot [h(z) - h(lca(x, z)) - h(lca(y, z))]
= h(y) - h(x) + 2\varepsilon_h \cdot [h(x) + h(z) - h(lca(x, z)) - h(lca(y, z))]
\geq h(y) - h(x) + 2\varepsilon_h \cdot [h(x) - h(lca(x, y))]
= d(x, y).
\]

Where in (48) we used the fact that $h(z) \geq h(lca(x, z)) + h(lca(y, z)) - h(lca(x, y))$, which is true since $h(lca(x, y)) \geq \min\{h(lca(x, z)), h(lca(y, z))\}$.

It remains to deal with the case $h(x) \leq h(z) \leq h(y)$. In this case
\[
d(x, z) + d(x, y) = h(y) - h(x) + 2\varepsilon_h \cdot [h(x) - h(lca(x, z))]) + 2\varepsilon_h \cdot [h(z) - h(lca(y, z))]
\geq h(y) - h(x) + 2\varepsilon_h \cdot [h(x) - h(lca(x, z))]) + 2\varepsilon_h \cdot [h(z) - h(lca(y, z))]
= h(y) - h(x) + 2\varepsilon_h \cdot \left[2h(x) - h(lca(x, z)) - \frac{h(x)}{h(z)}h(lca(y, z))\right]
\geq h(y) - h(x) + 2\varepsilon_h \cdot [h(x) - h(lca(x, y))]
= d(x, y),
\]

where (49) is equivalent to the inequality
\[
h(x) \geq h(lca(x, z)) + \frac{h(x)}{h(z)}h(lca(y, z)) - h(lca(x, y)).
\]

To prove (50), note that it is true if $h(lca(x, y)) \geq h(lca(x, z))$, since clearly $h(lca(y, z)) \leq h(z)$. If, on the other hand, $h(lca(x, y)) < h(lca(x, z))$ then using the assumption that $h(z) \geq h(x)$ it is enough to show that $h(x) \geq h(lca(x, z)) + h(lca(y, z)) - h(lca(x, y))$. Necessarily $h(lca(x, y)) = h(lca(y, z))$, so that the required inequality follows from the fact that $h(x) \geq h(lca(x, z))$. \qed
6.2 Geometry of H-trees

6.2.1 Classification of approximate midpoints

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 important concept of approximate midpoints which is used frequently in nonlinear functional analysis (see [4] and the references therein).

**Definition 6.2** (Approximate midpoints). Let $(X, d_X)$ be a metric space and $\delta \in (0, 1)$. For $x, y \in X$ the set of $\delta$-approximate midpoints of $x$ and $z$ is defined as

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

From now on, whenever we refer to the set $\text{Mid}(x, z, \delta)$, the underlying metric will always be $d_\varepsilon$. 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 $B_\infty$ are $\eta$-near if for every $j \in \{1, \ldots, n\}$ we have $d_\varepsilon(u_j, v_j) \leq \eta$. We shall also require the following terminology:

**Definition 6.3.** An ordered triple $(x, y, z)$ of vertices in $B_\infty$ will be called a path-type configuration if $h(z) \leq h(y) \leq h(x)$, $x$ is a descendant of $y$, and $h(lca(z, y)) < h(y)$. 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 $h(lca(x, z)) < h(x)$. These special configurations are described in Figure 3.

![Figure 3: A schematic description of path-type and tent-type configurations.](image)

The following useful theorem will be used extensively in the ensuing arguments. Its proof will be broken down into several elementary lemmas.

**Theorem 6.4.** Assume that $\delta \in (0, \frac{1}{16})$, and the sequence $\varepsilon = \{\varepsilon_n\}_{n=0}^\infty$ satisfies $\varepsilon_n < \frac{1}{4}$ for all $n \in \mathbb{N}$. Let $x, y, z \in (B_\infty, d_\varepsilon)$ be such that $y \in \text{Mid}(x, z, \delta)$. Then either $(x, y, z)$ or $(z, y, x)$ is $3\delta d_\varepsilon(x, z)$-near a path-type or tent-type configuration.

In what follows, given a vertex $v \in B_\infty$ we denote the subtree rooted at $v$ by $T_v$.

**Lemma 6.5.** Assume that $\varepsilon_n \leq \frac{1}{2}$ for all $n$. Fix $a \in B_\infty$ and let $u, v \in B_\infty$ be its children. For every $x, z \in T_v$ such that $h(x) \geq h(z)$ consider the function $D_{x,z} : \{a\} \cup T_v \to [0, \infty)$ defined by $D_{x,z}(y) = d_\varepsilon(x, y) + d_\varepsilon(z, y)$. Fix an arbitrary vertex $w \in T_v$ such that $h(w) = h(z)$. Then for every $y \in T_v$ we have $D_{x,z}(y) \geq D_{x,z}(w)$.

23
Proof. By the definition of $d_ε$ we have $D_{x,z}(y) = Q(h(y))$ where

$$Q(k) = \max\{h(x), k\} + \max\{k, h(z)\} - \min\{h(x), k\} - \min\{k, h(z)\} + 2ε_{\min\{h(x),k\}}[\min\{h(x),k\} - h(a)] + 2ε_{\min\{k,h(z)\}}[\min\{k,h(z)\} - h(a)].$$

The required result will follow if we show that $Q$ is non-increasing on $\{h(a), h(a) + 1, \ldots, h(z)\}$ and non-decreasing on $\{h(z), h(z) + 1, \ldots\}$. If $k \in \{h(a), h(a) + 1, \ldots, h(z) - 1\}$ then

$$Q(k + 1) - Q(k) = -2 + 4ε_{k+1}[k + 1 - h(a)] - 4ε_k[k - h(a)] \leq -2 + 4ε_k[k + 1 - h(a)] - 4ε_k[k - h(a)] = -2(1 - 2ε_k) \leq 0.$$

If $k \in \{h(x), h(x) + 1, \ldots\}$ then $Q(k + 1) - Q(k) = 2$, and if $k \in \{h(z), \ldots, h(x) - 1\}$ then

$$Q(k + 1) - Q(k) = 2[(k + 1)ε_k - 1 - kε_k] + 2h(a)[ε_k - ε_{k+1}] \geq 0.$$

This completes the proof of Lemma 6.6.

**Lemma 6.6.** Assume that $ε_n < \frac{1}{2}$ for all $n \in \mathbb{N}$. Fix $δ \in (0, \frac{1}{3})$ and $x, y, z ∈ B∞$ such that $h(x) > h(z)$, $y ∈ \text{Mid}(x, z, δ)$ and $h(\text{lca}(x, z)) > h(\text{lca}(x, y))$. Then

$$h(z) + \frac{1 - 3δ}{2}d_ε(x, z) ≤ h(y) < h(x) ≤ h(y) + \frac{1 + 3δ}{1 - 3δ}[h(y) - h(z)].$$

Moreover, if $y' ∈ B∞$ is the point on the segment joining $x$ and $\text{lca}(x, y)$ such that $h(y') = h(y)$ then $d_ε(y, y') ≤ δd_ε(x, z)$. Thus $(x, y', z)$ is a path-type configuration which is $δd_ε(x, z)$-near $(x, y, z)$.

**Proof.** Write $a = \text{lca}(x, y)$. If $u, v$ are the two children of $a$, then without loss of generality $x, z ∈ T_u$ and $y ∈ T_v$. Let $w ∈ T_u$ be such that $h(w) = h(z)$.

By Lemma 6.5,

$$d_ε(x, y) + d_ε(z, y) ≥ d_ε(x, w) + d_ε(z, w) = h(x) - h(z) + 4ε_{h(z)}[h(z) - h(a)] ≥ 2d_ε(x, z) - [h(x) - h(z)]. \quad (51)$$

On the other hand, since $y ∈ \text{Mid}(x, z, δ)$, we have that

$$d_ε(x, y) + d_ε(z, y) ≤ (1 + δ)d_ε(x, z).$$

Additionally, by the definition of $d_ε$ we know that if $h(y) ≤ h(z)$ then

$$\frac{1 + δ}{2}d_ε(x, z) ≥ d_ε(x, y) ≥ h(x) - h(y) ≥ h(x) - h(z).$$

Combining these observations with (51) we get that

$$(1 + δ)d_ε(x, z) ≥ 2d_ε(x, z) - \frac{1 + δ}{2}d_ε(x, z), \quad (52)$$

which is a contradiction since $δ < \frac{1}{3}$. Therefore $h(y) > h(z)$. If $h(y) ≥ h(x)$ then

$$\frac{1 + δ}{2}d_ε(x, z) ≥ d_ε(z, y) ≥ h(y) - h(z) ≥ h(x) - h(z),$$

24
Thus the triple \((z, y, x)\) is a configuration of path-type which is \(\delta d_{\varepsilon}(x, z)\)-near \((z, y, x)\).
If \( h(x) \leq h(z) \leq h(y) \) then since
\[
\frac{1 + \delta}{2} d_{e}(x, z) \geq d_{e}(x, y) = h(y) - h(x) + 2\epsilon h(x)[h(x) - h(a)]
\]
and \( d_{e}(x, z) = h(z) - h(x) + 2\epsilon h(x)[h(x) - h(a)] \) we deduce that
\[
-\frac{1 - \delta}{2} d_{e}(x, z) \geq d_{e}(x, y) - d_{e}(x, z) = h(y) - h(z) \geq 0.
\]
It follows that \( x = z \), in contradiction to our assumption.

If \( h(z) < h(x) \) then let \( z' \) be the point on the segment joining \( a \) and \( y \) such that \( h(z') = h(z) \). We thus have that
\[
d_{e}(z, z') = 2\epsilon h(z)[h(z) - h(lca(y, z))] = d_{e}(z, y) - [h(y) - h(z)].
\]
Moreover,
\[
2\epsilon h(z)[h(lca(z, y)) - h(a)] = d_{e}(x, z) - [h(x) - h(z)] - 2\epsilon h(z)[h(z) - h(lca(z, y))]
\]
\[
\geq \frac{2}{1 + \delta} d_{e}(z, y) - [h(x) - h(z)] - 2\epsilon h(z)[h(z) - h(lca(z, y))]
\]
\[
= \frac{2}{1 + \delta} \left( h(y) - h(z) + 2\epsilon h(z)[h(z) - h(lca(y, z))] \right)
\]
\[
- (h(y) - h(z) + 2\epsilon h(z)[h(z) - h(lca(y, z))]) + [h(y) - h(x)]
\]
\[
= \frac{1 - \delta}{1 + \delta} d_{e}(y, z) + [h(y) - h(x)]
\]
\[
\geq \frac{1 - \delta}{1 + \delta} \cdot \frac{1 - \delta}{2} d_{e}(x, z) + [h(y) - h(x)]
\]
\[
\geq \left( \frac{1 - \delta}{1 + \delta} \right)^{2} d_{e}(x, y) + [h(y) - h(x)]
\]
\[
= d_{e}(x, y) + [h(y) - h(x)] - \frac{4\delta}{(1 + \delta)^{2}} d_{e}(x, y)
\]
\[
= 2|h(y) - h(x)| + 2\epsilon h(x)[h(x) - h(a)] - \frac{4\delta}{(1 + \delta)^{2}} d_{e}(x, y)
\]
\[
\geq 2|h(y) - h(x)| + 2\epsilon h(z)[h(z) - h(lca(y, z))] - \frac{2\delta}{1 + \delta} d_{e}(x, z).
\]
Thus
\[
\frac{2\delta}{1 + \delta} d_{e}(x, z) \geq 2|h(y) - h(x)| + 2\epsilon h(z)[h(z) - h(lca(y, z))] = 2|h(y) - h(x)| + d_{e}(z, z'). \tag{55}
\]
Let \( y' \) be the point on the path from \( a \) to \( y \) such that \( h(y') = h(x) \). It follows from (55) that the triple \((z', y', x)\) is a configuration of tent-type which is \( 2\delta d_{e}(x, z) \)-near \((z, y, x)\).

This completes the proof of Lemma 6.7 when \( h(x) \leq h(y) \). The case \( h(x) > h(y) \) is proved analogously. Here are the details.
Thus the triple $(z, y', x')$ is a configuration of path-type which is $3 \delta d_e(x, z)$-near $(z, y, x)$.

If $h(x) > h(z) \geq h(y)$ then
\[
d_e(x, y') = d_e(x, y) = d_e(x, y) - \frac{2 \delta}{1 + \delta} d_e(x, y) = d_e(x, y) - \frac{2 \delta}{1 + \delta} d_e(x, y)
\]
\[
= d_e(x, y) - \frac{2 \delta}{1 + \delta} d_e(x, y) = h(x) - h(y) + 2 \varepsilon_{h(y)}[h(y) - h(lca(z, y))] = d_e(z, y) \geq \frac{1 - \delta}{1 + \delta} d_e(x, y)
\]
Canceling terms we see that
\[
\frac{2 \delta}{1 + \delta} d_e(x, y) \geq h(x) - h(z) + 2 \varepsilon_{h(y)}[h(lca(z, y)) - h(a)]
\]
\[
= h(x) - h(z) + 2 \varepsilon_{h(y)}[h(z) - h(a)] - 2 \varepsilon_{h(y)}[h(z) - h(lca(z, y))]
\]
\[
\geq h(x) - h(z) + 2 \varepsilon_{h(z)}[h(z) - h(a)] - 2 \varepsilon_{h(y)}[h(z) - h(lca(z, y))]
\]
\[
= d_e(x, z) - 2 \varepsilon_{h(y)}[h(z) - h(lca(z, y))]
\]
\[\geq 2d_\varepsilon(z, y) - \frac{2\delta}{1 + \delta}d_\varepsilon(x, z) - 2\varepsilon h(y)[h(z) - h(lca(z, y))]\]
\[= 2(h(z) - h(y) + 2\varepsilon h(y)[h(y) - h(lca(z, y))]) - \frac{2\delta}{1 + \delta}d_\varepsilon(x, z) - 2\varepsilon h(y)[h(z) - h(lca(z, y)))]\]
\[= 2\left(1 - 2\varepsilon h(y)\right)[h(z) - h(y)] + 2\varepsilon h(y)[h(z) - h(lca(z, y))] - \frac{2\delta}{1 + \delta}d_\varepsilon(x, z)\]
\[\geq [h(z) - h(y)] + 2\varepsilon h(y)[h(z) - h(lca(z, y))] - \frac{2\delta}{1 + \delta}d_\varepsilon(x, z)\]
\[= d_\varepsilon(z, y) - \frac{2\delta}{1 + \delta}d_\varepsilon(x, z)\]
\[\geq \left(1 - \frac{\delta}{2} - \frac{2\delta}{1 + \delta}\right)d_\varepsilon(x, z),\]

which is a contradiction since \(\delta < \frac{1}{16}\).

The only remaining case is when \(h(x) > h(y) > h(z)\). In this case we proceed as follows.

\[d_\varepsilon(x, z) = h(x) - h(z) + 2\varepsilon h(z)[h(z) - h(a)]\]
\[= d_\varepsilon(y, z) + [h(x) - h(y)] + 2\varepsilon h(z)[h(lca(y, z)) - h(a)]\]
\[\geq d_\varepsilon(x, y) - \frac{2\delta}{1 + \delta}d_\varepsilon(x, y) + [h(x) - h(y)] + 2\varepsilon h(z)[h(lca(y, z)) - h(a)]\]
\[\geq h(x) - h(y) + 2\varepsilon h(y)[h(y) - h(a)] - \delta d_\varepsilon(x, z)\]
\[+ [h(x) - h(y)] + 2\varepsilon h(z)[h(lca(y, z)) - h(a)]\]
\[\geq 2[h(x) - h(y)] + 2\varepsilon h(z)[h(z) - h(a)] + 2\varepsilon h(z)[h(lca(y, z)) - h(a)] - \delta d_\varepsilon(x, z)\]
\[= 2[d_\varepsilon(x, z) - 2d_\varepsilon(z, y) + 2\varepsilon h(z)[h(z) - h(lca(y, z))] - \delta d_\varepsilon(x, z)\]
\[\geq (1 - 2\delta)d_\varepsilon(x, z) + 2\varepsilon h(z)[h(z) - h(lca(y, z))].\]  

(58)

Let \(z'\) be the point on the path from \(a\) to \(y\) such that \(h(z') = h(z)\). Then

\[d_\varepsilon(z, z') = 2\varepsilon h(z)[h(z) - h(lca(y, z))] \leq 2\delta d_\varepsilon(x, z).\]

Therefore the triple \((z', y, x)\) is of tent-type and is \(2\delta d_\varepsilon(x, z)\)-near \((z, y, x)\). The proof of Lemma \([6,4]\) is complete.

**Proof of Theorem \([6.4]\)** It remains to check that for every \(x, y, z \in B_\infty\) such that \(y \in \text{Mid}(x, z, \delta)\), at least one of the triples \((x, y, z)\) or \((z, y, x)\) satisfies the conditions of Lemma \([6.6]\) or Lemma \([6.7]\).

Indeed, if \(h(lca(x, y)) = h(lca(x, z))\) then \(lca(x, y) = lca(x, z)\), so Lemma \([6.7]\) applies. If \(h(lca(x, y)) < h(lca(x, z))\) then \(lca(z, y) = lca(x, z)\), so Lemma \([6.7]\) applies to the triple \((z, y, x)\).

If \(h(lca(x, y)) < h(lca(x, z))\) then \(lca(x, y) = lca(z, y)\), and so \(h(lca(x, z)) > h(lca(z, y))\). Hence Lemma \([6.6]\) applies to either the triple \((x, y, z)\) or the triple \((z, y, x)\).

\(\Box\)
We end this subsection with a short discussion on the distance between tent-type and path-type configurations. It turns out that when \( \varepsilon_h \ll \delta \), a \( \delta \)-midpoint configuration \((x, y, z)\) can be close to a path-type configuration, and at the same time the reversed triple \((z, y, x)\) close to a tent-type configuration (or vice versa). However, it is easy to see that this is the only “closeness” possible.

**Lemma 6.8.** Fix \( x, y, z \in B_\infty \) with \( x \neq y \). Then the following statements are impossible:

1. \((x, y, z)\) is \( \frac{1}{5} d_e(x, y) \)-near a path-type configuration and a tent-type configuration.

2. \((x, y, z)\) is \( \frac{1}{5} d_e(x, y) \)-near a path-type configuration and \((z, y, x)\) is \( \frac{1}{11} d_e(x, y) \)-near a path-type configuration.

3. \((x, y, z)\) is \( \frac{1}{11} d_e(x, y) \)-near a tent-type configuration and \((z, y, x)\) is \( \frac{1}{11} d_e(x, y) \)-near a tent-type configuration.

**Proof.** For case 1 of Lemma 6.8, assume for contradiction that \((x, y, z)\) is \( \frac{1}{5} d_e(x, y) \)-near a path-type configuration \((a_1, b_1, c_1)\), and also \( \frac{1}{5} d_e(x, y) \)-near a tent-type configuration \((\alpha_1, \beta_1, \gamma_1)\). By the definitions of path-type and tent-type configurations, \( a_1 \) is a descendant of \( b_1 \) and \( \beta_1 \) is a descendant of \( \alpha_1 \). Hence,

\[
h(a_1) - h(b_1) = d_e(a_1, b_1) \geq d_e(x, y) - d_e(x, a_1) - d_e(y, b_1) \geq \frac{3}{5} d_e(x, y),
\]

(59)

and

\[
h(\beta_1) - h(\alpha_1) = d_e(\alpha_1, \beta_1) \geq d_e(x, y) - d_e(x, \alpha_1) - d_e(y, \beta_1) \geq \frac{3}{5} d_e(x, y).
\]

(60)

By summing (59) and (60) we see that,

\[
\frac{4}{5} d_e(x, y) \geq d_e(a_1, x) + d_e(x, \alpha_1) + d_e(b_1, y) + d_e(y, \beta_1) \geq d_e(a_1, \alpha_1) + d_e(b_1, \beta_1)
\]

\[
\geq h(a_1) - h(\alpha_1) + h(\beta_1) - h(b_1) \geq \frac{6}{5} d_e(x, y),
\]

a contradiction.

For case 2 of Lemma 6.8, assume for contradiction that \((x, y, z)\) is \( \frac{1}{5} d_e(x, y) \)-near a path-type configuration \((a_2, b_2, c_2)\), and also \((z, y, x)\) is \( \frac{1}{11} d_e(x, y) \)-near a path-type configuration \((\alpha_2, \beta_2, \gamma_2)\). By the definitions of path-type and tent-type configurations, \( a_2 \) is a descendant of \( b_2 \) and \( h(\beta_2) > h(\gamma_2) \). Hence,

\[
h(a_2) - h(b_2) = d_e(a_2, b_2) \geq d_e(x, y) - d_e(x, a_2) - d_e(y, b_2) \geq \frac{9}{11} d_e(x, y),
\]

(61)

and

\[
h(\beta_2) - h(\gamma_2) + 2\varepsilon_{h(\gamma_2)} [h(\gamma_2) - h(lca(\beta_2, \gamma_2))] = d_e(\beta_2, \gamma_2)
\]

\[
\geq d_e(x, y) - d_e(x, \gamma_2) - d_e(y, \beta_2) \geq \frac{9}{11} d_e(x, y).
\]

(62)
By summing (61) and (62) we see that
\[
\frac{17}{11} d_\varepsilon(x, y) \geq d_\varepsilon(a_2, x) + d_\varepsilon(x, \gamma_2) + d_\varepsilon(b_2, y) + d_\varepsilon(y, \beta_2) + d_\varepsilon(\beta_2, y) + d_\varepsilon(x, y) + d_\varepsilon(x, \gamma_2)
\]
\[
\geq d_\varepsilon(a_2, \gamma_2) + d_\varepsilon(b_2, \beta_2) + d_\varepsilon(\beta_2, \gamma_2)
\]
\[
\geq (h(a_2) - h(\gamma_2)) + (h(\beta_2) - h(b_2)) + 2\varepsilon_{h(\gamma_2)}[h(\gamma_2) - h(lca(\beta_2, \gamma_2))]
\]
\[
\overset{(61)}{\geq} \frac{18}{11} d_\varepsilon(x, y),
\]
a contradiction.

For case 3 of Lemma 6.8 assume for contradiction that \((x, y, z)\) is \(\frac{1}{11} d_\varepsilon(x, y)\)-near a tent-type configuration \((a_3, b_3, c_3)\), and also \((z, y, x)\) is \(\frac{1}{11} d_\varepsilon(x, y)\)-near a tent-type configuration \((a_3, b_3, \gamma_3)\). Then \(b_3\) is a descendant of \(a_3\) and \(h(\gamma_3) > h(\beta_3)\). Hence,
\[
h(b_3) - h(a_3) = d_\varepsilon(a_3, b_3) \geq d_\varepsilon(x, y) - d_\varepsilon(a_3, x) - d_\varepsilon(y, b_3) \geq \frac{9}{11} d_\varepsilon(x, y),
\]
and
\[
h(\gamma_3) - h(\beta_3) + 2\varepsilon_{h(\gamma_3)}[h(\beta_3) - h(lca(\beta_3, \gamma_3))] = d_\varepsilon(\beta_3, \gamma_3)
\]
\[
\geq d_\varepsilon(x, y) - d_\varepsilon(x, \gamma_3) - d_\varepsilon(y, \beta_3) \geq \frac{9}{11} d_\varepsilon(x, y).
\]
Hence,
\[
\frac{17}{11} d_\varepsilon(x, y) \geq d_\varepsilon(b_3, y) + d_\varepsilon(y, \beta_3) + d_\varepsilon(x, \alpha_3) + d_\varepsilon(x, \gamma_3) + d_\varepsilon(\beta_3, y) + d_\varepsilon(x, y) + d_\varepsilon(x, \gamma_3)
\]
\[
\geq d_\varepsilon(b_3, \beta_3) + d_\varepsilon(a_3, \gamma_3) + d_\varepsilon(\beta_3, \gamma_3)
\]
\[
\overset{(63)}{\geq} (h(b_3) - h(\beta_3)) + (h(\gamma_3) - h(a_3)) + 2\varepsilon_{h(\beta_3)}[h(\beta_3) - h(lca(\beta_3, \gamma_3))]
\]
\[
\overset{(64)}{\geq} \frac{18}{11} d_\varepsilon(x, y),
\]
a contradiction. \(\blacksquare\)

### 6.2.2 Classification of approximate forks

We begin with three “stitching lemmas” that roughly say that given three points \(x, x', y \in (B_\infty, d_\varepsilon)\) such that \(x'\) is near \(x\), there exists \(y'\) near \(y\) such that \(d_\varepsilon(x', y')\) is close to \(d_\varepsilon(x, y)\), and \(y'\) relates to \(x'\) in \(B_\infty\) “in the same way” that \(y\) relates \(x\) in \(B_\infty\).

**Lemma 6.9.** Let \(x, x', y, y' \in B_\infty\) be such that \(y\) is an ancestor of \(x\), and \(y'\) is an ancestor of \(x'\) satisfying \(h(x) - h(y) = h(x') - h(y')\). Then \(d_\varepsilon(y, y') \leq d_\varepsilon(x, x')\).

**Proof.** Assume without loss of generality that \(h(x) \geq h(x')\). So,
\[
d_\varepsilon(x, x') = h(x) - h(x') + 2\varepsilon_{h(x')}[h(x') - h(lca(x, x'))].
\]
Note that \(h(lca(y, y')) = \min\{h(y'), h(lca(x, x'))\}\). Hence,
\[
d_\varepsilon(y, y') = h(y) - h(y') + 2\varepsilon_{h(y')}[h(y') - h(lca(y, y'))]
\]
\[
= h(x) - h(x') + 2\varepsilon_{h(y')}[h(y') - \min\{h(y'), h(lca(x, x'))\}]
\]
\[
= h(x) - h(x') + 2\varepsilon_{h(y')} \max\{0, h(y') - h(lca(x, x'))\}. \quad (65)
\]
If the maximum in (65) is 0, then
\[ d_\varepsilon(y, y') = h(x) - h(x') \leq d_\varepsilon(x, x'). \]

If the maximum in (65) equals \( h(y') - h(lca(x, x')) \), then
\[ d_\varepsilon(y, y') = h(x) - h(x') + 2\varepsilon_{h(y')} [h(y') - h(lca(x, x'))] \]
\[ \leq h(x) - h(x') + 2\varepsilon_{h(x')} [h(x') - h(lca(x, x'))] = d_\varepsilon(x, x'), \]
(66)

where in (66) we used the fact that the sequence \( \{\varepsilon_n(n - a)\}_{n=0}^\infty \) is nondecreasing for all \( a \geq 0 \).

**Lemma 6.10.** Let \( x, x', y \in B_\infty \) be such that \( h(y) \leq h(x) \). Then there exists \( y' \in B_\infty \) which satisfies \( h(y') - h(x') = h(y) - h(x) \),
\[ d_\varepsilon(y, y') \leq d_\varepsilon(x, x'), \]
(67)

and
\[ d_\varepsilon(x, x') - 2d_\varepsilon(x, x') \leq d_\varepsilon(y, y') \leq d_\varepsilon(x, y) + 2d_\varepsilon(x, x'). \]
(68)

**Proof.** Note that (68) follows from (67) by the triangle inequality. Assume first that \( h(x) \geq h(x') \).

In this case choose \( y' \) to be an ancestor of \( y \) satisfying \( h(y') - h(x') = h(y) - h(x) \). Then,
\[ d_\varepsilon(y, y') = h(y) - h(y') = h(x) - h(x') \leq d_\varepsilon(x, x'). \]

We next assume that \( h(x) < h(x') \). If \( h(lca(x, x')) \neq h(lca(x, y)) \) then choose \( y' \) to be an arbitrary descendant of \( y \) such that \( h(y') - h(y) = h(x') - h(x) \). As before, we conclude that \( d_\varepsilon(y, y') = h(y') - h(y) = h(x') - h(x) \leq d_\varepsilon(x, x') \).

It remains to deal with the case \( h(x') > h(x) \) and \( h(lca(x, y)) = h(lca(x', x)) \), which also implies that \( h(lca(x', y)) > h(lca(x, y)) \). In this case, we choose \( y' \) to be an arbitrary point on a branch containing both \( lca(x, y) \) and \( x \), such that \( h(y') - h(y) = h(x') - h(x) \). Then \( lca(y', y) = lca(x, x') \), and therefore,
\[ d_\varepsilon(y, y') = h(y') - h(y) + 2\varepsilon_{h(y')} [h(y') - h(lca(y', y'))] \]
\[ = h(x') - h(x) + 2\varepsilon_{h(y')} [h(y') - h(lca(x, x'))] \]
\[ \leq h(x') - h(x) + 2\varepsilon_{h(x')} [h(x') - h(lca(x, x'))] \]
\[ = d_\varepsilon(x, x'), \]
proving (67) in the last remaining case.

**Lemma 6.11.** Let \( x, x', y \in B_\infty \) be such that \( y \) is a descendant of \( x \). Then for any \( y' \in B_\infty \) which is a descendant of \( x' \) and satisfying \( h(y') - h(x') = h(y) - h(x) \), we have
\[ d_\varepsilon(y, y') \leq d_\varepsilon(x, x') + 2\varepsilon_{\min(h(y'), h(y))} [h(y') - h(x')] \leq d_\varepsilon(x, x') + 2\varepsilon_{h(y)} [h(y) - h(x) + d_\varepsilon(x, x')]. \]

**Proof.** Note that \( h(lca(y, y')) \geq h(lca(x, x')) \). Assume first that \( h(x') \geq h(x) \). Then,
\[ d_\varepsilon(y, y') = h(y') - h(y) + 2\varepsilon_{h(y')} [h(y') - h(lca(y, y'))] \]
\[ = h(x') - h(x) + 2\varepsilon_{h(y)} [h(x') - h(lca(y', y'))] + 2\varepsilon_{h(y)} [h(y) - h(x)] \]
\[ \leq h(x') - h(x) + 2\varepsilon_{h(y)} [h(x) - h(lca(x, x'))] + 2\varepsilon_{h(y)} [h(y) - h(x)] \]
\[ \leq d_\varepsilon(x, x') + 2\varepsilon_{h(y)} [h(y) - h(x)]. \]
When \( h(x') < h(x) \), we similarly obtain the bound:

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

The last inequality in the statement of Lemma 6.11 is proved by observing that when \( h(y') < h(y) \),

\[
\varepsilon_{h(y')}[h(y) - h(x)] = \varepsilon_{h(y')}[h(y') - h(x')] \leq \varepsilon_{h(y)}[h(y) - h(x)] \leq \varepsilon_{h(y)}[h(y) - h(x) + d_\varepsilon(x, x')].
\]

**Definition 6.12.** For \( \delta \in (0, 1) \) and \( x, y, z, w \in B_\infty \), the quadruple \((x, y, z, w)\) is called a \( \delta \)-fork, if

\[
y \in \text{Mid}(x, z, \delta) \cap \text{Mid}(x, w, \delta).
\]

\( \delta \)-forks in H-trees can be approximately classified using the approximate classification of midpoint configurations of Section 6.2.1. We have four types of midpoint configurations (recall Figure 3):

- path-type; denoted (P) in what follows,
- reverse path-type; denoted (p)—\((x, y, z)\) is of type (p) iff \((z, y, x)\) is of type (P),
- tent-type; denoted (T),
- reverse tent-type; denoted (t)—\((x, y, z)\) is of type (t) iff \((z, y, x)\) is of type (T).

Thus, there are \( \binom{5}{2} = 10 \) possible \( \delta \)-fork configurations in \((B_\infty, d_\varepsilon)\) (choose two out of the five symbols “P”, “p”, “T”, “t”, “X”, where “X” means “the same”). As we shall see, four of these possible 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 they are not much distorted from the star \( K_{1,3} \) (the metric \( d \) on four points \( p, q, r, s \) given by \( d(p, q) = d(q, r) = d(q, s) = 1 \) and \( d(p, s) = d(p, r) = d(r, s) = 2 \)). The 10 possible \( \delta \)-fork configurations are summarized in Table 1.

| Midpoint configuration | 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 ten possible fork configurations.

For future reference, we give names to the four problematic configurations:
Definition 6.13. For $\eta, \delta \in (0, 1)$, a $\delta$-fork $(x, y, z, w)$ of $(B_\infty, d_\varepsilon)$ is called

- $\eta$-near Type I (configuration $(T \parallel T)$ in Table 1), if both $(x, y, z)$ and $(x, y, w)$ are $\eta$-near tent-type configurations;
- $\eta$-near Type II (configuration $(P \parallel P)$ in Table 1), if both $(x, y, z)$ and $(x, y, w)$ are $\eta$-near path-type configurations;
- $\eta$-near Type III (configuration $(p \parallel T)$ in Table 1), if $(z, y, x)$ is $\eta$-near a path-type configuration and $(x, y, w)$ is $\eta$-near a tent-type configuration, or vice versa;
- $\eta$-near Type IV (configuration $(p \parallel t)$ in Table 1), if $(z, y, x)$ is $\eta$-near a path-type configuration and $(w, y, x)$ is $\eta$-near a tent-type configuration, or vice versa.

A schematic description of the four problematic configurations is contained in Figure 4.

The following lemma is the main result of this section.

Lemma 6.14. Fix $\delta \in (0, \frac{1}{10})$ and assume that $\varepsilon_n < \frac{1}{4}$ for all $n \in \mathbb{N}$. If $(x, y, z, w)$ is a $\delta$-fork of $(B_\infty, d_\varepsilon)$ then either it is $35\delta d_\varepsilon(x, y)$-near one of the types I, II, III, IV, or we have $d_\varepsilon(z, w) \leq 2(35\delta + \varepsilon h_0)d_\varepsilon(x, y)$, where $h_0 = \min\{h(x), h(y), h(z), h(w)\}$.

Remark 6.15. One can strengthen the statement of Lemma 6.14 so that in the first case the fork $(x, y, z, w)$ is $O(\delta d_\varepsilon(x, y))$ near another fork $(x', y', z', w')$ which is of (i.e. 0-near) one of the types I, II, III, IV. This statement is more complicated to prove, and since we do not actually need it in what follows, we opted to use a weaker property which suffices for our purposes, yet simplifies (the already quite involved) proof.

The proof of Lemma 6.14 proceeds by checking that the cases marked in Table 1 as “impossible” or “prongs contracted” are indeed so—see Figure 5 for a schematic description of the latter case.

We begin with the $(p \parallel p)$ configuration.

Lemma 6.16. Let $(x, y, z, w)$ be a $\delta$-fork of $(B_\infty, d_\varepsilon)$ and assume that both $(z, y, x)$, and $(w, y, x)$ are $\eta d_\varepsilon(x, y)$-near path-type configurations. Then, assuming that $\max\{\delta, \eta\} < 1/8$ and $\varepsilon_n < \frac{1}{4}$ for all $n$, we have

$$d_\varepsilon(z, w) \leq (9\eta + 6\delta + 2\varepsilon h(y))d_\varepsilon(x, y).$$
So, in conclusion, 

\[ h(w) = h(y') + h(w'') - h(y'') = h(y') + d_\varepsilon(w'', y'') \geq h(y) - |h(y) - h(y')| + d_\varepsilon(w, y) - 2\eta d_\varepsilon(x, y) \geq h(y) - d_\varepsilon(y, y') + \frac{1 - \delta}{1 + \delta} d_\varepsilon(x, y) - 2\eta d_\varepsilon(x, y) \geq h(y) + \int d_\varepsilon(x, y) \geq h(y). \]

By Lemma 6.11,

\[ d_\varepsilon(w', w'') \leq d_\varepsilon(y', y'') + 2\varepsilon_{h(w')} [h(w'') - h(y'')] \leq d_\varepsilon(y', y'') + 2\varepsilon_{h(w')} d_\varepsilon(y'', w'') \leq 2\eta d_\varepsilon(x, y) + 2\varepsilon_{h(y)} \left( 2\eta + \frac{1 + \delta}{1 - \delta} \right) d_\varepsilon(x, y). \] (69)

Observe that

\[ \left( \frac{1 - \delta}{1 + \delta} - 2\eta \right) d_\varepsilon(x, y) \leq d_\varepsilon(z, y) - 2\eta d_\varepsilon(x, y) \leq d_\varepsilon(z', y') \leq d_\varepsilon(z, y) + 2\eta d_\varepsilon(x, y) \leq \left( \frac{1 + \delta}{1 - \delta} + 2\eta \right) d_\varepsilon(x, y), \] (70)

Since \( d_\varepsilon(w', y') = d_\varepsilon(w'', y'') \), we obtain similarly the bounds:

\[ \left( \frac{1 - \delta}{1 + \delta} - 2\eta \right) d_\varepsilon(x, y) \leq d_\varepsilon(w', y') \leq d_\varepsilon(z, y) + 2\eta d_\varepsilon(x, y) \leq \left( \frac{1 + \delta}{1 - \delta} + 2\eta \right) d_\varepsilon(x, y). \] (71)

Hence

\[ d_\varepsilon(z', w') = |d_\varepsilon(y', z') - d_\varepsilon(y', w')| \leq \left( \frac{4\delta}{1 - \delta^2} + 4\eta \right) d_\varepsilon(x, y). \] (72)

So, in conclusion,

\[ d_\varepsilon(z, w) \leq d_\varepsilon(z, z') + d_\varepsilon(w, w'') + d_\varepsilon(w'', w') + d_\varepsilon(z', w') \leq \left( 8\eta + \frac{4\delta}{1 - \delta^2} + 2 \left( \frac{1 + \delta}{1 - \delta} \right) \varepsilon_{h(y)} \right) d_\varepsilon(x, y) \leq (9\eta + 6\delta + 2\varepsilon_{h(y)}) d_\varepsilon(x, y). \]
We next consider the $(t \parallel t)$ configuration.

**Lemma 6.17.** Let $(x, y, z, w)$ be a $\delta$-fork of $(B_\infty, d_\varepsilon)$. Assume that both $(z, y, x)$ and $(w, y, x)$ are $\eta d_\varepsilon(x, y)$-near tent-type configurations. Then, assuming that $\max \{\delta, \eta\} < 1/4$, we have

$$d_\varepsilon(z, w) \leq (8\eta + 5\delta)d_\varepsilon(x, y).$$

**Proof.** Let $(z', y', x')$ be a tent-type configuration that is $\eta d_\varepsilon(x, y)$-near $(z, y, x)$, and let $(w'', y'', x'')$ be a tent-type configuration that is $\eta d_\varepsilon(x, y)$-near $(w, y, x)$. Assume without loss of generality that $h(y'') - h(w'') \geq h(y') - h(z')$. Let $\bar{w}$ be a point on the path between $w''$ and $y''$ such that $h(y'') - h(\bar{w}) = h(y') - h(z')$. Then,

$$d_\varepsilon(w'', \bar{w}) = h(y'') - h(w'') - (h(y') - h(z')) = d_\varepsilon(y'', w'') - d_\varepsilon(y', z')$$

$$\leq d_\varepsilon(y, w) - d_\varepsilon(y, z) + 4\eta d_\varepsilon(x, y) \leq \left(\frac{1 + \delta}{1 - \delta} - \frac{1 - \delta}{1 + \delta} + 4\eta\right) d_\varepsilon(x, y). \quad (73)$$

By Lemma 6.9 we have $d_\varepsilon(\bar{w}, z') \leq d_\varepsilon(y', y'') \leq 2\eta d_\varepsilon(x, y)$. Hence we conclude that

$$d_\varepsilon(y, z) \leq d_\varepsilon(z, z') + d_\varepsilon(\bar{w}, z') + d_\varepsilon(\bar{w}, w'') + d_\varepsilon(w'', w) \leq (4\delta \leq 1 - \delta^2 + 8\eta) d_\varepsilon(x, y). \quad \square$$

**Lemma 6.18.** Let $(x, y, z, w)$ be a $\delta$-fork of $B_\infty$. Assume that $(x, y, z)$ is $\eta d_\varepsilon(x, y)$-near a path-type configuration. Assume also that $\delta < 1/30$, $\eta < 1/10$, and $\varepsilon_n < 1/4$ for all $n$. Then $(x, y, w)$ is $(2\eta + 21\delta)d_\varepsilon(x, y)$-near a path-type configuration, i.e., $(x, y, z, w)$ is $(2\eta + 21\delta)d_\varepsilon(x, y)$-near a type II configuration.

**Proof.** Let $(x', y', z')$ be a path-type configuration which is $\eta d_\varepsilon(x, y)$-near $(x, y, z)$. By Theorem 6.4, either $(x, y, w)$ or $(w, y, x)$ must be $3\delta d_\varepsilon(x, z) \leq \frac{6}{1 - \delta}d_\varepsilon(x, y) \leq 7\delta d_\varepsilon(x, y)$-near either a path-type configuration or a tent-type configuration.

Suppose first that $(x, y, w)$ is $7\delta d_\varepsilon(x, y)$-near a path-type configuration $(x'', y'', w'')$. In this case, $x''$ is an ancestor of $y''$ and $h(y'') - h(x'') = d_\varepsilon(x'', y'') \geq (1 - 14\delta)d_\varepsilon(x, y)$. At the same time, $y'$ is an ancestor of $x'$ and $h(x') - h(y') = d_\varepsilon(x', y') \geq (1 - 2\eta)d_\varepsilon(x, y)$. So,

$$2(\eta + 7\delta)d_\varepsilon(x, y) \geq d_\varepsilon(y'', y') + d_\varepsilon(x', x'') \geq h(y'') - h(x'') + h(x') - h(y') \geq (1 - \eta - 7\delta)d_\varepsilon(x, y),$$

which is a contradiction since $\eta + 7\delta < 1/2$.

Next suppose that $(w, y, x)$ is $7\delta d_\varepsilon(x, y)$-near a path-type configuration $(w'', y'', x'')$. Then $|h(x') - h(x'')| \leq d_\varepsilon(x', x'') \leq (\eta + 7\delta)d_\varepsilon(x, y)$. So,

$$|\eta + 7\delta|d_\varepsilon(x, y) \geq d_\varepsilon(y', y'') \geq h(y'') - h(y')$$

$$= (h(y'') - h(x'')) + (h(x'') - h(x')) + h(x') - h(y') > 0 - (\eta + 7\delta)d_\varepsilon(x, y) + (1 - 2\eta)d_\varepsilon(x, y),$$

which is a contradiction.

Lastly, suppose that $(w, y, x)$ is $7\delta d_\varepsilon(x, y)$-near a tent-type configuration $(w'', y'', x'')$. Note that $|h(y') - h(y'')| \leq d_\varepsilon(y', y'') \leq (\eta + 7\delta)d_\varepsilon(x, y)$. So, $h(y') \geq h(y'') - (\eta + 7\delta)d_\varepsilon(x, y)$. Also,

$$h(y'') - h(w'') = d_\varepsilon(y'', w'') \geq d_\varepsilon(y, w) - 14\delta d_\varepsilon(x, y) \geq \left(\frac{1 - \delta}{1 + \delta} - 14\delta\right) d_\varepsilon(x, y) \geq (\eta + 7\delta)d_\varepsilon(x, y).$$

35
Consider the point \( \bar{w} \) defined as the ancestor of \( y' \) at distance \( h(y'') - h(w'') - (\eta + 7\delta)d_ε(x, y) \) from \( y' \). Let also \( w''' \) be the ancestor of \( y'' \) at distance \( h(y'') - h(w'') - (\eta + 7\delta)d_ε(x, y) \) from \( y'' \). By Lemma 6.9, we have \( d_ε(\bar{w}, w''') \leq d_ε(y', y'') \leq (\eta + 7\delta)d_ε(x, y) \). Therefore,

\[
d_ε(\bar{w}, w) \leq d_ε(\bar{w}, w''') + d_ε(w''', w'') + d_ε(w'', w) \leq (2\eta + 21\delta)d_ε(x, y).
\]

Hence \((x, y, w)\) is \((2\eta + 21\delta)d_ε(x, y)\)-near the path-type configuration \((x', y', \bar{w})\).

**Lemma 6.19.** Let \((x, y, z, w)\) be a \(\delta\)-fork of \(B_∞\). Assume that \((x, y, z)\) is \(\eta d_ε(x, y)\)-near a tent-type configuration. Assume also that \(\eta < 1/10\) and \(\varepsilon_n < 1/4\) for all \(n\). Then \((w, y, x)\) cannot be \(\eta d_ε(x, y)\)-near a tent-type configuration.

**Proof.** Let \((x', y', z')\) be a tent type configuration that is \(\eta d_ε(x, y)\)-near \((x, y, z)\). Suppose for contradiction that there exists a tent type configuration \((w'', y'', x'')\) that is \(\eta d_ε(x, y)\)-near \((w, y, x)\). Note that \(h(y'') \geq h(y') - \varepsilon_n \geq h(y') - 2\eta d_ε(x, y)\) and \(h(y') - h(x') \geq (1 - 2\eta)d_ε(x, y) > 2\eta d_ε(x, y)\). Let \(x^*\) be the ancestor of \(y''\) at distance \(h(y'') - h(x') - 2\eta d_ε(x, y)\) from \(y''\), and let \(\bar{x}\) be the ancestor of \(y''\) at distance \(h(y') - h(x') - 2\eta d_ε(x, y)\) from \(y''\). An application of Lemma 6.9 yields the estimate \(d_ε(\bar{x}, x^*) \leq d_ε(y', y'') \leq 2\eta d_ε(x, y)\). But, since \(h(x^*) \geq h(y'')\), we also know that \(d_ε(\bar{x}, x^*) = h(y'') - h(\bar{x}) = d_ε(y', x') - 2\eta d_ε(x, y)\). Hence,

\[
2\eta d_ε(x, y) \geq d_ε(\bar{x}, x^*) \geq d_ε(\bar{x}, x'') = d_ε(x*, x') - d_ε(x', x'') \geq d_ε(x', y') - 6\eta d_ε(x, y) \geq (1 - 8\eta)d_ε(x, y),
\]

which is a contradiction, since \(\eta < 1/10\).

**Proof of Lemma 6.14.** Since \((x, y, z, w)\) is a \(\delta\)-fork, by Theorem 6.4, both \((x, y, z)\) and \((x, y, w)\) are \(7\delta d_ε(x, y)\)-near a tent-type configuration, a path-type configuration, or the corresponding reverse configurations. We have 10 possible combinations of these pairs, as appearing in Table 1. By applying Lemmas 6.18 and 6.19 with \(\eta = 7\delta\), we rule out three of these configurations, and a fourth configuration is possible but only as \(35\delta d_ε(x, y)\)-near a type II configuration.

We are left with six possible configurations. By applying Lemmas 6.16 and 6.17 with \(\eta = 7\delta\) we conclude that in two of those configurations we have \(d_ε(w, z) \leq (69\delta + 2\varepsilon_{h_0})d_ε(x, y)\), and the rest are configurations that are \(7\delta d_ε(x, y)\)-near one of the types I–IV.

**6.2.3 Classification of approximate 3-paths**

We start with the following natural notion:

**Definition 6.20.** For \(x_0, x_1, x_2, x_3 \in B_∞\) the quadruple \((x_0, x_1, x_2, x_3)\) is called a \((1+\delta)\)-approximate \(P_3\) if there exists \(L > 0\) such that for every \(0 \leq i \leq j \leq 3\) we have

\[(j - i)L \leq d_ε(x_i, x_j) \leq (1 + \delta)(j - i)L.\]

Note that in this case \(x_1 \in \text{Mid}(x_0, x_2, \delta)\) and \(x_2 \in \text{Mid}(x_1, x_3, \delta)\).

As in the case of \(\delta\)-forks, there are 10 possible concatenations of two midpoints configurations (path-type or tent-type): P-P, P-p, P-T, P-t, p-P, p-T, T-T, T-t, t-T (the midpoint configurations p-p, p-p, t-p, t-p, P-P, P-T, T-P, T-P, t-t, t-t are respectively such concatenations with the order of \(x_0, x_1, x_2, x_3\) reversed). We will rule out some of these possibilities, and obtain some stronger properties for the rest. See Table 2.

As in the case of \(\delta\)-forks, it will be beneficial to give names to three special types approximate 3-paths:
### Definition 6.21.
For \( x_0, x_1, x_2, x_3 \in B_\infty \) and \( \eta > 0 \), a quadruple \( (x_0, x_1, x_2, x_3) \) is called:

- \( \eta \)-near a type A configuration if both \( (x_0, x_1, x_2) \) and \( (x_1, x_2, x_3) \) are \( \eta \)-near path-type configurations,

- \( \eta \)-near a type B configuration if \( (x_0, x_1, x_2) \) is \( \eta \)-near a path-type configuration, and \( (x_3, x_2, x_1) \) is \( \eta \)-near tent-type configuration,

- \( \eta \)-near type C configuration if \( (x_2, x_1, x_0) \) is \( \eta \)-near a path-type configuration, and \( (x_1, x_2, x_3) \) is \( \eta \)-near a tent-type configuration.

See also Figure 6.

#### Figure 6: The three possible types of approximate 3-paths.

The following lemma is the main result of this subsection.

### Lemma 6.22.
Assume that \( \varepsilon_n < \frac{1}{4} \) for all \( n \) and fix \( \delta < 1/200 \). Assume that \( (x_0, x_1, x_2, x_3) \) is a \( (1 + \delta) \)-approximate \( P_3 \). Then either \( (x_0, x_1, x_2, x_3) \) or \( (x_3, x_2, x_1, x_0) \) is \( 35\delta d_\varepsilon(x_0, x_1) \)-near a configuration of type A, B or C.
The proof of Lemma 6.22 is again a case analysis that examines all 10 possible ways (up to symmetry) to concatenate two midpoint configurations. The proof is divided into a few lemmas according to the cases, and is completed at the end of this subsection.

**Lemma 6.23.** Assume that \( \varepsilon_n < \frac{1}{2} \) for all \( n \) and that \((x_0, x_1, x_2, x_3)\) is a \((1 + \delta)\)-approximate \(P_3\) such that \((x_0, x_1, x_2)\) is \( \eta \delta \varepsilon (x_0, x_1) \)-near a path-type configuration. If \( \max \{ \delta, \eta \} < 1/200 \) then either \((x_1, x_2, x_3)\) is \( 7 \delta \varepsilon (x_0, x_1) \)-near a path-type configuration (type A), or \((x_3, x_2, x_1)\) is \( 7 \delta \varepsilon (x_0, x_1) \)-near a tent-type configuration (type B).

**Proof.** Due to Theorem 6.4 we only need to rule out the possibility that \((x_3, x_2, x_1)\) is \( 7 \delta \varepsilon (x_0, x_1) \)-near a path-type configuration, or that \((x_1, x_2, x_3)\) is \( 7 \delta \varepsilon (x_1, x_2) \)-near a tent-type configuration. Let \((x'_0, x'_1, x'_2)\) be a path-type configuration that is \( \eta \delta \varepsilon (x_0, x_1) \)-near \((x_0, x_1, x_2)\).

Suppose first that \((x_3, x_2, x_1)\) is \( 7 \delta \varepsilon (x_1, x_2) \)-near the path-type configuration \((x''_3, x''_2, x''_1)\). Since \( h(x'_1) \geq h(x'_2) \) and \( h(x'_2) \geq h(x'_3) \) we have,

\[
|h(x_1) - h(x_2)| \leq |h(x_1) - h(x'_1)| + (h(x'_1) - h(x'_2)) + |h(x'_2) - h(x_2)|
\]

and similarly,

\[
|h(x_1) - h(x_2)| \leq |h(x_1) - h(x''_1)| + (h(x''_1) - h(x''_2)) + |h(x''_2) - h(x_2)|
\]

By summing (74) and (75) we obtain the bound

\[
2|h(x_1) - h(x_2)| \leq (2 \eta + 14 \delta) d_\varepsilon(x_0, x_1) + d_\varepsilon(x'_1, x''_1) + d_\varepsilon(x'_2, x''_2) \leq (4 \eta + 28 \delta) d_\varepsilon(x_0, x_1).
\]

Thus

\[
|h(x_1) - h(x_2)| \leq (2 \eta + 14 \delta) d_\varepsilon(x_0, x_1).
\]

Since \(x'_0\) is a descendant of \(x'_1\),

\[
|h(x_0) - h(x_1) - d_\varepsilon(x_0, x_1)| \leq |h(x'_0) - h(x'_1)| - d_\varepsilon(x_0, x_1)| + 2 \eta d_\varepsilon(x_0, x_1)
\]

\[
= |d_\varepsilon(x'_0, x'_1)| - d_\varepsilon(x_0, x_1)| + 2 \eta d_\varepsilon(x_0, x_1) \leq 4 \eta d_\varepsilon(x_0, x_1).
\]

Similarly, since \(x''_3\) is a descendant of \(x''_2\),

\[
|h(x_3) - h(x_2) - d_\varepsilon(x_0, x_1)| \leq |h(x''_3) - h(x''_2)| - d_\varepsilon(x_0, x_1)| + 14 \delta d_\varepsilon(x_1, x_2)
\]

\[
= |d_\varepsilon(x''_3, x''_2)| - d_\varepsilon(x_0, x_1)| + 14 \delta d_\varepsilon(x_0, x_1) \leq 28 \delta d_\varepsilon(x_0, x_1).
\]

Hence,

\[
|h(x''_3) - h(x'_0)| \leq |h(x''_3) - h(x_3)| + |h(x_3) - h(x_2) - d_\varepsilon(x_0, x_1)| + |h(x_2) - h(x_1)|
\]

\[
+ |h(x_0) - h(x_1) - d_\varepsilon(x_0, x_1)| + |h(x_0) - h(x'_0)|
\]

\[
\leq d_\varepsilon(x''_3, x'_2) + 28 \delta d_\varepsilon(x_0, x_1) + (2 \eta + 14 \delta) d_\varepsilon(x_0, x_1) + 4 \eta d_\varepsilon(x_0, x_1) + d_\varepsilon(x_0, x'_0)
\]

\[
\leq (49 \delta + 7 \eta) d_\varepsilon(x_0, x_1).
\]
We record for future reference the following consequence of (76) and (79):

\[
\min\{h(x'_0), h(x''_3)\} - \min\{h(x'_1), h(x''_2)\} \leq \max\{h(x'_0) - h(x'_1), h(x'_0) - h(x''_2)\}
\]

\[
\leq \max\{d_e(x'_0, x'_1), h(x''_3) - h(x''_2) + (49\delta + 7\eta) d_e(x_0, x_1)\}
\]

\[
\leq \max\{(1 + 2\eta)d_e(x_0, x_1), d_e(x''_3, x''_2) + (49\delta + 7\eta) d_e(x_0, x_1)\}
\]

\[
\leq (1 + 64\delta + 7\eta) d_e(x_0, x_1). \tag{80}
\]

We next claim that

\[
lca(x'_0, x''_3) = lca(x'_1, x''_2). \tag{81}
\]

Indeed, since \(x'_1\) is an ancestor of \(x'_0\) and \(x''_2\) is an ancestor of \(x''_3\), if \(lca(x'_0, x''_3) \neq lca(x'_1, x''_2)\) then either \(x'_1\) is a descendant of \(x''_2\), or \(x''_2\) is a descendant of \(x'_1\). If \(x'_1\) is a descendant of \(x''_2\) then

\[
(\eta + 7\delta) d_e(x_0, x_1) \geq d_e(x'_1, x''_2) \geq d_e(x''_2, x'_1) \geq d_e(x_2, x_1) - (\eta + 7\delta) d_e(x_0, x_1)
\]

\[
\geq \frac{1}{1 + \delta} d_e(x_0, x_1) - (\eta + 7\delta) d_e(x_0, x_1),
\]

which is a contradiction since \(\delta, \eta < 1/200\). Similarly, if \(x''_2\) is a descendant of \(x'_1\) then

\[
(\eta + 7\delta) d_e(x_0, x_1) \geq d_e(x'_2, x''_3) \geq d_e(x''_3, x'_2) \geq d_e(x_2, x_1) - (\eta + 7\delta) d_e(x_0, x_1)
\]

arriving once more at a contradiction. This proves (81).

Now,

\[
\frac{3}{1 + \delta} d_e(x_0, x_1) \leq d_e(x_0, x_3)
\]

\[
\leq d_e(x''_3, x'_0) + (\eta + 7\delta) d_e(x_0, x_1)
\]

\[
\leq 2\varepsilon_{\min\{h(x'_0), h(x''_3)\}} \max\{h(x'_0), h(x''_3)\} - h(lca(x'_0, x''_3)) + (8\eta + 56\delta) d_e(x_0, x_1) \quad \text{(81)}
\]

\[
\leq 2\varepsilon_{\min\{h(x'_0), h(x''_3)\}} \max\{h(x'_1), h(x''_2)\} - h(lca(x'_1, x''_2)) + (8\eta + 56\delta) d_e(x_0, x_1)
\]

\[
+ 2\varepsilon_{\min\{h(x''_3), h(x''_2)\}} \max\{h(x''_3), h(x'_3)\} - h(lca(x''_3, x'_3)) \quad \text{(82)}
\]

\[
\leq 2\varepsilon_{\min\{h(x'_1), h(x''_2)\}} \max\{h(x'_1), h(x''_2)\} - h(lca(x'_1, x''_2))
\]

\[
+ \left(8\eta + 56\delta + \frac{1 + 64\delta + 7\eta}{2}\right) d_e(x_0, x_1)
\]

\[
\leq d_e(x'_1, x''_2) + \left(\frac{1}{2} + 88\delta + 12\eta\right) d_e(x_0, x_1)
\]

\[
\leq \left(\frac{3}{2} + 96\delta + 13\eta\right) d_e(x_0, x_1), \tag{83}
\]

where in (82) we used \(\min\{h(x'_0), h(x''_3)\} \geq \min\{h(x'_1), h(x''_2)\}\) and \(\varepsilon_{\min\{h(x'_0), h(x''_3)\}} < 1/4\). Since \(\max\{\eta, \delta\} < 1/200\), the bound (83) is a contradiction.
Next suppose that \((x_1, x_2, x_3)\) is \(7\delta d_\varepsilon(x_0, x_1)\)-near a tent-type configuration \((x''_1, x''_2, x''_3)\). Since \(h(x''_2) \leq h(x'_1)\) and \(h(x''_3) \geq h(x''_1)\), we have
\[
|h(x_1) - h(x_2)| \leq |h(x_1) - h(x'_1)| + (h(x'_1) - h(x''_2)) + |h(x''_2) - h(x_2)|
\leq d_{\varepsilon}(x_1, x'_1) + (h(x'_1) - h(x''_2)) + d_{\varepsilon}(x''_2, x_2)
\leq d_{\varepsilon}(x_1, x'_1) + d_{\varepsilon}(x'_1, x''_1) + d_{\varepsilon}(x''_2, x''_2) + d_{\varepsilon}(x''_2, x_2)
\leq (4\eta + 14\delta)d_{\varepsilon}(x_0, x_1).
\] (84)
On the other hand, \(x''_2\) is an ancestor of \(x''_2\), and therefore we have
\[
\left(\frac{1}{1 + \delta} - 14\delta\right) d_{\varepsilon}(x_0, x_1) \leq d_{\varepsilon}(x''_1, x''_2) = h(x''_2) - h(x''_1)
\leq |h(x_1) - h(x_2)| + 14\delta d_{\varepsilon}(x_0, x_1) \leq (4\eta + 28\delta)d_{\varepsilon}(x_0, x_1),
\] (85)
which is a contradiction since \(\max\{\eta, \delta\} < 1/200\).

**Lemma 6.24.** Assume that \(\varepsilon_n < \frac{1}{4}\) for all \(n\) and that \((x_0, x_1, x_2, x_3)\) is a \((1 + \delta)\)-approximate \(P_3\) such that \((x_2, x_1, x_0)\) is \(\eta d_{\varepsilon}(x_0, x_1)\)-near a path-type configuration. If \(\max\{\delta, \eta\} < 1/200\) then either \((x_3, x_2, x_1)\) is \(7\delta d_{\varepsilon}(x_0, x_1)\)-near a path-type configuration \((\text{reverse type A})\), or \((x_1, x_2, x_3)\) is \(7\delta d_{\varepsilon}(x_0, x_1)\)-near a tent-type configuration \((\text{type C})\).

**Proof.** Let \((x'_2, x'_1, x'')\) be in path-type configuration that is \(\eta d_{\varepsilon}(x_0, x_1)\)-near \((x_2, x_1, x_0)\). First, assume for contradiction that \((x_3, x_2, x_1)\) is \(7\delta d_{\varepsilon}(x_1, x_2)\)-near a tent-type configuration \((x''_3, x''_2, x''_1)\). Then \(h(x''_3) \geq h(x''_2)\), where as \(h(x'_2) - h(x'_1) = d_{\varepsilon}(x'_2, x'_1)\). Arguing as in (84), it follows that
\[
|h(x_1) - h(x_2)| \leq (2\eta + 28\delta)d_{\varepsilon}(x_0, x_1),
\] and we arrive at a contradiction by arguing similarly to (85).

Next, assume for contradiction that \((x_1, x_2, x_3)\) is \(7\delta d_{\varepsilon}(x_1, x_2)\)-near a tent-type configuration \((x''_1, x''_2, x''_3)\). Then \(h(x''_1) - h(x''_3) = d_{\varepsilon}(x''_1, x''_2)\), whereas \(h(x'_2) - h(x'_1) = d_{\varepsilon}(x'_1, x'_2)\). By summing these two identities, we arrive at a contradiction as follows:
\[
\left(\frac{2}{1 + \delta} - 2\eta - 14\delta\right) d_{\varepsilon}(x_0, x_1) \leq d_{\varepsilon}(x''_1, x''_2) + d_{\varepsilon}(x''_1, x''_3) = (h(x'_2) - h(x'_1)) + (h(x''_3) - h(x''_3))
\leq d_{\varepsilon}(x'_2, x''_3) + d_{\varepsilon}(x'_1, x''_1) \leq (2\eta + 14\delta)d_{\varepsilon}(x_0, x_1).
\]  

**Lemma 6.25.** Assume that \(\varepsilon_n < \frac{1}{4}\) for all \(n\) and that \((x_0, x_1, x_2, x_3)\) is a \((1 + \delta)\)-approximate \(P_3\) such that \((x_0, x_1, x_2)\) is \(\eta d_{\varepsilon}(x_0, x_1)\)-near a path-type configuration. If \(\max\{\delta, \eta\} < 1/200\) then either \((x_2, x_1, x_0)\) is \((14\delta + 3\eta)d_{\varepsilon}(x_0, x_1)\)-near a path-type configuration and \((x_1, x_2, x_3)\) is \(7\delta d_{\varepsilon}(x_1, x_2)\)-near a tent-type configuration \((\text{type C})\), or \((x_3, x_2, x_1)\) is \(7\delta d_{\varepsilon}(x_0, x_1)\)-near a path-type configuration \((\text{reverse type B})\).

**Proof.** Let \((x''_0, x'_1, x'_2)\) be a tent-type configuration that is \(\eta d_{\varepsilon}(x_0, x_1)\)-near \((x_0, x_1, x_2)\). First, suppose that \((x_1, x_2, x_3)\) is \(7\delta d_{\varepsilon}(x_1, x_2)\)-near a tent-type configuration \((x''_1, x''_2, x''_3)\). Note that \(h(x''_1) - h(x'_1) \leq d_{\varepsilon}(x''_1, x'_1) \leq (\eta + 7\delta)d_{\varepsilon}(x_0, x_1)\). So, let \(x''_0\) be an ancestor of \(x''_1\) at distance \(h(x''_1) - h(x'_1) = (\eta + 7\delta)d_{\varepsilon}(x_0, x_1)\) from \(x''_1\), and let \(x''_0\) be an ancestor of \(x'_1\) at distance \(h(x''_1) - h(x'_0) = (\eta + 7\delta)d_{\varepsilon}(x_0, x_1)\) from \(x'_1\). Then \(h(x'_1) - h(x''_0) = h(x''_1) - h(x'_0)\) and \(d_{\varepsilon}(x''_0, x'_0) \leq (\eta + 7\delta)d_{\varepsilon}(x_0, x_1)\). By Lemma 6.38,
\[
d_{\varepsilon}(x_0, x''_0) - (2\eta + 7\delta)d_{\varepsilon}(x_0, x_1) \leq d_{\varepsilon}(x_0, x''_0) - d_{\varepsilon}(x'_0, x'_0) - d_{\varepsilon}(x''_0, x_0) \leq d_{\varepsilon}(x'_0, x'_0) \leq d_{\varepsilon}(x'_1, x''_1) \leq (\eta + 7\delta)d_{\varepsilon}(x_0, x_1).
\]
Hence $(x''_2, x''_3, x''_0)$ is a path-type configuration that is $(14\delta + 3\eta)d_\varepsilon(x_0, x_1)$-near $(x_2, x_1, x_0)$.

Next assume for contradiction that $(x_3, x_2, x_1)$ is $7\delta d_\varepsilon(x_1, x_2)$ near a tent-type configuration $(x''_3, x''_2, x''_1)$. Then

\[(1 - 15\delta) d_\varepsilon(x_0, x_1) \leq \left( \frac{1}{1 + \delta} - 14\delta \right) d_\varepsilon(x_0, x_1) \leq h(x''_2) - h(x''_3) \leq (1 + 15\delta)d_\varepsilon(x_0, x_1), \tag{86}\]

and

\[(1 - \delta - 2\eta) d_\varepsilon(x_0, x_1) \leq \left( \frac{1}{1 + \delta} - 2\eta \right) d_\varepsilon(x_0, x_1) \leq h(x'_1) - h(x'_0) \leq (1 + \delta + 2\eta)d_\varepsilon(x_0, x_1). \tag{87}\]

So, let $x''_3$ be an ancestor of $x'_2$ at distance $h(x''_2) - h(x''_3) - (16\delta + 2\eta)d_\varepsilon(x_0, x_1) \in [0, h(x''_2)]$ from $x''_2$, and let $x'_0$ be an ancestor of $x'_1$ at distance $h(x''_2) - h(x''_3) - (16\delta + 2\eta)d_\varepsilon(x_0, x_1) \in [0, h(x'_1)]$ from $x'_1$. Then

\[d_\varepsilon(x''_3, x''_3) \leq (16\delta + 2\eta)d_\varepsilon(x_0, x_1), \tag{88}\]

and

\[d_\varepsilon(x'_0, x'_0) = \left| h(x'_0) - h(x'_0) \right| = \left| h(x'_0) - (h(x'_1) - h(x'_2) + h(x''_3) + (16\delta + 2\eta)d_\varepsilon(x_0, x_1)) \right| \leq 2(16\delta + 2\eta)d_\varepsilon(x_0, x_1). \tag{89}\]

Moreover, $h(x'_1) - h(x'_0) = h(x''_2) - h(x''_3)$, so by Lemma 6.9 we have

\[\left( \frac{3}{1 + \delta} - 55\delta - 7\eta \right) d_\varepsilon(x_0, x_1) \leq d_\varepsilon(x_0, x_3) - (55\delta + 7\eta)d_\varepsilon(x_0, x_1) \leq d_\varepsilon(x'_0, x''_3) \leq d_\varepsilon(x'_1, x'_2) \leq (1 + 8\delta + \eta)d_\varepsilon(x_0, x_1),\]

which is a contradiction since $\max\{\delta, \eta\} < 1/200$.

Lastly, assume for contradiction that $(x_1, x_2, x_3)$ is $7\delta d_\varepsilon(x_1, x_2)$-near a path-type configuration $(x''_1, x''_2, x''_3)$. Then since $h(x'_1) \leq h(x'_2)$ we have

\[\left( \frac{1}{1 + \delta} - 14\delta \right) d_\varepsilon(x_0, x_1) \leq d_\varepsilon(x''_1, x''_2) = h(x''_1) - h(x''_2) \leq (h(x''_1) - h(x''_2)) + (h(x''_2) - h(x'_1)) \leq d_\varepsilon(x'_1, x'_1) + d_\varepsilon(x''_2, x''_2) \leq (14\delta + 2\eta)d_\varepsilon(x_0, x_1),\]

a contradiction. \hfill \Box

**Lemma 6.26.** Assume that $\varepsilon_n < \frac{1}{4}$ for all $n$ and that $(x_0, x_1, x_2, x_3)$ is a $(1 + \delta)$-approximate $P_3$ such that $(x_2, x_1, x_0)$ is $\eta d_\varepsilon(x_0, x_1)$-near a tent-type configuration. If $\max\{\delta, \eta\} < 1/200$ then $(x_1, x_2, x_3)$ cannot be $7\delta d_\varepsilon(x_0, x_1)$ near a tent-type configuration.

**Proof.** Let $(x''_2, x''_3, x''_0)$ be a tent-type configuration that is $\eta d_\varepsilon(x_0, x_1)$-near $(x_2, x_1, x_0)$. Suppose for contradiction that $(x_1, x_2, x_3)$ is $7\delta d_\varepsilon(x_0, x_1)$-near a tent-type configuration $(x''_1, x''_2, x''_3)$. Then $h(x'_1) - h(x'_2) = d_\varepsilon(x'_1, x'_2)$, whereas $h(x''_2) - h(x''_3)$ is $d_\varepsilon(x'_1, x''_2)$. Taking the sum of these two inequalities we conclude that

\[d_\varepsilon(x''_1, x''_2) + d_\varepsilon(x''_1, x''_3) \leq d_\varepsilon(x'_1, x'_2) + d_\varepsilon(x'_2, x'_3) \leq (2\eta + 14\delta)d_\varepsilon(x_0, x_1).\]

At the same time, \(\left( \frac{2}{1 + \delta} - 2\eta - 14\delta \right) d_\varepsilon(x_0, x_1) \leq d_\varepsilon(x''_1, x''_2) + d_\varepsilon(x'_1, x'_2),\) which leads to the desired contradiction. \hfill \Box
Lemma 6.28. Fix $w$ with $D$ ration. If both $(f, u_2, x_3)$ and $x_2 \in \text{Mid}(x_1, x_3, \delta)$. Since the assumptions of Theorem 6.4 hold, we can apply with $\eta = 7\delta$ Lemmas 6.23, 6.24, 6.25, 6.26 and conclude that either $(x_0, x_1, x_2, x_3)$ or $(x_3, x_2, x_1, x_0)$ must be $35\delta d_\varepsilon(x_0, x_1)$-near a configuration of type $A$, $B$ or $C$. $\square$

6.3 Nonembeddability of vertically faithful $B_4$

In what follows we need some standard notation on trees. As before, $B_n$ is the complete binary tree of height $n$; the root of $B_n$ is denoted by $r$. Denote by $I(B_n)$ the set of internal vertices of $B_n$, i.e., vertices of $B_n$ which are not the root or a leaf. For a vertex $v$ in $\{r\} \cup I(B_n)$ we denote by $v_0$ and $v_1$ its children. For $a \in \{0, 1\}^*$ (the set of finite sequences of ’0’ and ’1’) and $x \in \{0, 1\}$ we denote by $v_{ax} = (v_a)_x$.

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

Lemma 6.27. Fix $0 < \delta < 1/400$ and let $f : B_4 \rightarrow (B_\infty, d_\varepsilon)$ be a $(1 + \delta)$-vertically faithful embedding. Then the distortion of $f$ satisfies

$$\text{dist}(f) \geq \frac{1}{500 \delta + \varepsilon h_0},$$

where $h_0 = \min_{x \in B_4} h(f(x))$.

The proof of Lemma 6.27 is by a contradiction. By Lemma 6.14 assuming the distortion of $f$ is small, all the $\delta$-forks in the $(1 + \delta)$-vertically faithful embedding must be of types I–IV. By exploring the constrains implied by Lemma 6.22 on how those $\delta$-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 high distortion.

Fix $f : B_4 \rightarrow (B_\infty, d_\varepsilon)$. For $u \in I(B_4)$ we denote by $\mathcal{F}(u)$ the fork in which $u$ is the center point, i.e., if $v$ be the parent of $u$ in $B_4$, then

$$\mathcal{F}(u) \overset{\text{def}}{=} (f(v), f(u), f(u_0), f(u_1)).$$

We shall assume from now on that $f$ satisfies the assumptions of Lemma 6.27 i.e., that it satisfies (8) with $D = 1 + \delta$ for some $\delta < 1/400$ and $\lambda > 0$.

Lemma 6.28. Fix $u \in B_4$ with $h(u) \in \{1, 2\}$. If the fork $\mathcal{F}(u)$ is $37\delta \lambda$-near a type I or type III configuration, then there exists $w \in I(B_4)$ satisfying

$$d_\varepsilon(f(u_0), f(u_1)) \leq (170\delta + \varepsilon h_0) \cdot 2\lambda. \quad (90)$$

Proof. Let $v$ be the parent of $u$. Hence, $(f(v), f(u), f(u_0), f(u_1))$ is $35\delta(1 + \delta)\lambda$-near a type I or a type III configuration. Assume first that $(f(v), f(u), f(u_0), f(u_1))$ is $37\delta \lambda$-near a type I configuration. If both $(f(u_0), f(u), f(v))$ and $(f(u_1), f(u), f(v))$ were $37\delta \lambda$-near a path type configuration then by Lemma 6.16 (with $\eta = 37\delta$) we would have

$$d_\varepsilon(f(u_0), f(u_1)) \leq (339\delta + 2\varepsilon h_0)(1 + \delta)\lambda \leq (170\delta + \varepsilon h_0) \cdot 2\lambda, \quad (91)$$

proving (90) with $w = u$. The same conclusion holds when $(f(v), f(u), f(u_0), f(u_1))$ is $37\delta \lambda$-near a type III configuration: in this case without loss of generality $(f(v), f(u), f(u_0))$ is $37\delta \lambda$-near
a tent-type configuration and \((f(u_1), f(u), f(v))\) is \(37\delta \lambda\)-near a path-type configuration. Using Lemma 6.16 as above we would arrive at the conclusion (91) if \((f(u_0), f(u), f(v))\) were \(37\delta \lambda\)-near a path-type configuration. Thus, in both the type I and type III cases of Lemma 6.28 we may assume that \((f(v), f(u), f(u_0))\) is \(37\delta \lambda\)-near a tent-type configuration, and that, by Lemma 6.8 \((f(v), f(u), f(u_0))\) is not \(37\delta \lambda\)-near a path-type configuration, and \((f(u_0), f(u), f(v))\) is not \(37\delta \lambda\)-near a path-type configuration or a tent-type configuration.

By Lemma 6.22 and Table 2 \((f(u_0c), f(u_0), f(u), f(v))\) must be \(35\delta(1 + \delta)\lambda\)-near a type B configuration for both \(c \in \{0, 1\}\). This means that \((f(u_0c), f(u_0), f(u))\) are both \(35\delta(1 + \delta)\lambda\)-near a path-type configuration, and so by Lemma 6.16 (with \(\eta = 35\delta(1 + \delta)\)) we deduce that \(d_c(f(u_00), f(u_01)) \leq (170\delta + \varepsilon h_0) \cdot 2\lambda\).

**Lemma 6.29.** Fix \(u \in B_4\) with \(h(u) \in \{1, 2\}\). If \(\mathcal{F}(u)\) is \(37\delta \lambda\)-near a type II configuration then for both \(b \in \{0, 1\}\) either \(\mathcal{F}(u_b)\) is \(99\delta \lambda\)-near a type II configuration, or \(d_c(f(u_00), f(u_01)) \leq 400\delta\).

**Proof.** Let \(v\) be the parent of \(u\). For both \(c \in \{0, 1\}\) we know that \((f(v), f(u), f(u_0), f(u_0c))\) is a \((1 + \delta)\)-approximate \(P_3\), and therefore by Lemma 6.22 either either \((f(v), f(u), f(u_0), f(u_0c))\) or \((f(u_0c), f(u_0), f(u), f(v))\) is \(35\delta(1 + \delta)\lambda\)-near a configuration of type \(A, B\) or \(C\). Note that since \((f(v), f(u), f(u_0))\) is assumed to be \(37\delta \lambda\)-near a path-type configuration, we rule out the possibility that \((f(v), f(u), f(u_0), f(u_0c))\) is \(35\delta(1 + \delta)\lambda\)-near a configuration of type \(C\), since otherwise both \((f(v), f(u), f(u_0))\) and \((f(u_0c), f(u_0), f(u), f(v))\) would be \(37\delta \lambda\)-near path-type configurations, contradicting Lemma 6.8. For the same reason we rule out the possibility that \((f(u_0c), f(u_0), f(u), f(v))\) is \(35\delta(1 + \delta)\lambda\)-near a configuration of type \(A\) or type \(B\). An inspection of the three remaining possibilities shows that either \((f(u), f(u_0), f(u_0c))\) is \(37\delta \lambda\)-near a path-type configuration, or \((f(u_0c), f(u_0), f(u))\) is \(37\delta \lambda\)-near a tent-type configuration.

Now,

- If for both \(c \in \{0, 1\}\) we have that \((f(u), f(u_0), f(u_0c))\) is \(37\delta \lambda\)-near a path-type configuration, then \(\mathcal{F}(u_0)\) is \(37\delta \lambda\)-near a type II configuration.

- If for both \(c \in \{0, 1\}\) we have that \((f(u_0c), f(u_0), f(u))\) are \(37\delta \lambda\)-near a tent-type configuration, then by Lemma 6.17 we have \(d_c(f(u_01), f(u_00)) \leq 400\delta\).

- By Lemma 6.18 the only way that \((f(u), f(u_0), f(u_00))\) could be \(37\delta \lambda\)-near a path type configuration while at the same time \((f(u_01), f(u_0), f(u))\) is \(37\delta \lambda\)-near a tent-type configuration (or vice versa), is that \(\mathcal{F}(u_0)\) is \(99\delta \lambda\)-near a type II configuration.

**Lemma 6.30.** Fix \(u \in B_4\) with \(h(u) \in \{1, 2\}\). If \(\mathcal{F}(u)\) is \(35\delta(1 + \delta)\lambda\)-near a type IV configuration, then there exists \(b \in \{0, 1\}\) such that \(\mathcal{F}(u_b)\) is \(37\delta \lambda\)-near a type II configuration.

**Proof.** Let \(v\) be the parent of \(u\). Without loss of generality \((f(u_0), f(u), f(v))\) is \(35\delta(1 + \delta)\lambda\)-near a tent-type configuration. By Lemma 6.22 (using Lemma 6.8 to rule out the remaining possibilities), this means that for both \(c \in \{0, 1\}\) the quadruple \((f(u_0c), f(u_0), f(u), f(v))\) is \(35\delta(1 + \delta)^2\lambda\)-near a type \(C\) configuration, and therefore \(\mathcal{F}(u_0)\) is \(35\delta(1 + \delta)^2\lambda\)-near a type II configuration.

**Lemma 6.31.** Fix \(u \in B_4\) with \(h(u) \in \{0, 1, 2\}\). If \(\mathcal{F}(u_0)\) and \(\mathcal{F}(u_1)\) are both \(99\delta \lambda\)-near type II configuration then \(d_c(f(u_0), f(u_1)) \leq 1000\delta\).
Proof. By our assumptions, \((f(u), f(u_0), f(u_0))\) is \(99\delta\lambda\)-near a path type configuration \((u', u'_1, u''_1)\) and \((f(u), f(u_1), f(u_{10}))\) is \(99\delta\lambda\)-near a path-type configuration \((u'', u''_1, u''_1)\). We may assume without loss of generality that \(h(u'') - h(u'_1) \leq h(u') - h(u'_0)\). We may therefore consider the ancestor \(u''_1\) of \(u'\) such that \(h(u') - h(u'_1) = h(u'') - h(u''_1)\), implying in particular that \(h(u'_1) \geq h(u'_0)\) (recall that \(u'_0\) is an ancestor of \(u''\), and \(u''_1\) is ancestor of \(u''_1\)). By Lemma 6.31 we have

\[
d_{x}(u''_1, u''_1) \leq d_{x}(u', u'') \leq 198\delta\lambda.
\] (92)

Hence,

\[
h(u') - h(u'_1) = d_{x}(u', u'_1) \geq d_{x}(u', u'') - 198\delta\lambda \geq d_{x}(f(u), f(u_0)) - 394\delta\lambda \geq (1 - 394\delta)\lambda.
\] (93)

But, we also know that

\[
h(u') - h(u'_0) = d_{x}(u', u'_0) \leq d_{x}(f(u), f(u_0)) + 198\delta\lambda \leq (1 + 200\delta)\lambda.
\] (94)

It follows from (93) and (94) that \(d_{x}(u_0, u'_1) = h(u'_1) - h(u'_0) \leq 601\delta\lambda\). Therefore,

\[
d_{x}(f(u_1), f(u_0)) \leq d_{x}(f(u_0), u'_0) + d_{x}(u_0, u'_1) + d_{x}(u'_1, u''_1) + d_{x}(u''_1, f(u_1)) \leq 1000\delta\lambda.
\]

\[\square\]

Proof of Lemma 6.27. We may assume that for all \(u \in I(B_1)\) the fork \(F(u)\) is \(35\delta(1 + \delta)\lambda\)-near a configuration of type I, II, III, or IV. Indeed, otherwise the proof is complete by Lemma 6.14. If \(F(r_0)\) is \(35\delta(1 + \delta)\lambda\)-near a type I or type III configuration, then by Lemma 6.28 the proof is complete. If \(F(r_0)\) is \(35\delta(1 + \delta)\lambda\)-near a type IV configuration then by Lemma 6.30 there exists \(b \in \{0, 1\}\) such that \(F(r_{0b})\) is \(37\delta\lambda\)-near a type II configuration. It therefore remains to deal with the case in which for some \(u \in \{r_0, r_{0b}\}\) the fork \(F(u)\) is \(37\delta\lambda\)-near a type II configuration. Applying Lemma 6.29 either we are done, or both \(F(u_0)\) and \(F(u_1)\) are \(99\delta\lambda\)-near a type II configuration, but then by Lemma 6.31 the proof of Lemma 6.27 is complete.

\[\square\]

6.4 Nonembeddability of binary trees

We are now in position to complete the proof of Theorem 1.10.

Proof of Theorem 1.10. Write \(\varepsilon_n = 1/s(n)\), and \(\varepsilon = \{\varepsilon_n\}_{n=0}^{\infty}\). Thus \(\{\varepsilon_n\}_{n=0}^{\infty}\) is non-increasing, \(\{n\varepsilon_n\}_{n=0}^{\infty}\) is non-decreasing, and \(\varepsilon_n \leq 1/4\). We can therefore choose the metric space \((X, d_X) = (B_\infty, d_\varepsilon)\). The identity embedding of \(B_n\) into the top \(n\)-levels of \(B_\infty\) shows that \(c_X(B_n) \leq s(n)\). It remains to prove the lower bound on \(c_X(B_n)\). To this end take an arbitrary injection \(f : B_n \to X\) satisfying \(\text{dist}(f) \leq s(n)\), and we will now prove that

\[
\text{dist}(f) \geq s \left( \frac{n}{40s(n)} \right) \left( 1 - \frac{Cs(n) \log s(n)}{\log n} \right).
\] (95)

By adjusting the constant \(C\) in (95), we may assume below that \(n\) is large enough, say, \(n \geq 100\). Write \(h_0 = \lfloor n/(40s(n)) \rfloor\) and define \(X_{>h_0} = \{x \in B_\infty : h(x) > h_0\}\). We claim that there exists a complete binary subtree \(T \subseteq B_n\) of height at least \([n/3]\), such that we have \(f(T) \subseteq X_{>h_0}\). Indeed, 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 \(f(B_n) \subseteq X_{>h_0}\), and we can take \(T = B_n\). So assume that \(h_{\min} < h_0\). Since \(f\) is an injection it must satisfy \(h_{\max} \geq n\). Hence \(\|f\|_{\text{Lip}} \geq \frac{h_{\max} - h_{\min}}{2n} \geq \frac{n-h_0}{2n} \geq \frac{1}{10}\). Since \(\text{dist}(f) \leq s(n)\)
we conclude that $\|f^{-1}\|_{\text{Lip}} \leq 4s(n)$. It follows that, since $\text{diam}(X \setminus X_{>h_0}) \leq 2h_0$, we have $\text{diam}(f^{-1}(X \setminus X_{>h_0})) \leq 8h_0s(n) \leq n/5$. If the top $\lceil n/3 \rceil$ levels of $B_n$ are mapped into $X_{>h_0}$ then we are done, so assume that there exists $u \in f^{-1}(X \setminus X_{>h_0})$ of depth at most $\leq \lfloor n/3 \rfloor$. In this case $f^{-1}(X \setminus X_{>h_0})$ must be contained in the first $\lfloor n/3 \rfloor + n/5 < 2n/3 - 1$ levels of $B_n$, so we can take $T$ to be any subtree of $B_n$ contained in the last $\lfloor n/3 \rfloor$ levels of $B_n$.

Fix $\delta \in (0, 1)$. By Theorem 1.11 (with $t = 4$, $D = s(n)$ and $\xi = \delta$), there exists a universal constant $\kappa > 0$ such that if $n \geq s(n)^{\kappa/\delta}$ then there exists a mapping $\phi : B_4 \rightarrow B_n$ with $\text{dist}(\phi) \leq 1 + \delta$ such that $f \circ \phi$ is a $(1 + \delta)$-vertically faithful embedding of $B_4$ into $X_{>h_0}$. Choosing $\delta = \frac{\log s(n)}{\log n}$, by increasing $C$ in (95) if necessary, we may assume that $\delta < 1/400$. Lemma 6.27 then implies

$$(1 + \delta) \text{dist}(f) \geq \text{dist}(f \circ \phi) \geq \frac{1}{500\delta + \epsilon_{h_0}} = \frac{1}{500\kappa \frac{\log s(n)}{\log n} + \frac{1}{\log n} \frac{1}{s((n/(40s(n)))}. \quad (96)$$

The deduction of (7) from (6) is a simple exercise: if $s(n) = \omega(\log n / \log \log n)$ then we have $(s(n) \log s(n))/\log n = o(1)$. The desired claim will then follow once we check that

$$\limsup_{n \rightarrow \infty} \frac{s(\lfloor n/(40s(n)) \rfloor)}{s(n)} = 1. \quad (97)$$

Indeed, if (96) failed then there would exist $\epsilon_0 \in (0, 1)$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$s(\lfloor n/\log n \rfloor) \leq s(\lfloor n/(40s(n)) \rfloor) \leq (1 - \epsilon_0)s(n). \quad (97)$$

Iterating (97), it would follow that $s(n_j) \geq n_j^{1/(40)}$ for some subsequence $\{n_j\}_{j=1}^\infty$, a contradiction. \hfill $\square$

**Proof of Theorem 1.12.** The proof is identical to the above argument: all one has to notice is that when $s(n) = D$ for all $n \in \mathbb{N}$ the resulting metric $d_x$ on $B_\infty$ is $D$-equivalent to the original shortest path metric on $B_\infty$. In this case, if $c_X(B_n) \leq D - \epsilon$ then the bound (95) implies that $n \leq D^{CD^2/\epsilon}$.

\hfill $\square$

## 7 Discussion and open problems

A very interesting question that arises naturally from Theorem 1.3 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:

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

A Banach space $X$ is $p$-convex if and only if its dual space $X^*$ is $q$-smooth, where $\frac{1}{p} + \frac{1}{q} = 1$ [11]. 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 [10, 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. Trees are natural candidates for finite metric obstructions to $p$-convexity, but it is unclear what would be the possible finite metric witnesses to the “non-$q$-smoothness” of a metric space.

$H$-trees are geometric objects that are quite simple combinatorially, yet as we have seen, they have interesting bi-Lipschitz properties. It would therefore be of interest to investigate the geometry of $H$-trees for its own right. In particular, what is the $L_1$ distortion of an $H$-tree? How close can an $H$-tree be to a metric of negative type?
References

[1] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992. doi:10.1007/BF01896971

[2] 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. doi:10.1007/BF01231769

[3] 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.

[4] 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.

[5] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986. doi:10.1007/BF02766125

[6] J. Bourgain, V. Milman, and H. Wolfson. On type of metric spaces. *Trans. Amer. Math. Soc.*, 294(1):295–317, 1986. doi:10.2307/2000132

[7] 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. doi:10.1007/BF02762802

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

[9] R. C. James. Uniformly non-square Banach spaces. *Ann. of Math. (2)*, 80:542–550, 1964.

[10] 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.

[11] W. B. Johnson and G. Schechtman. Diamond graphs and super-reflexivity. *J. Topol. Anal.*, 1(2):177–189, 2009. doi:10.1142/S1793525309000114

[12] T. J. Laakso. Plane with $A_\infty$-weighted metric not bi-Lipschitz embeddable to $\mathbb{R}^N$. *Bull. London Math. Soc.*, 34(6):667–676, 2002. doi:10.1112/S0024609302001200

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

[14] 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. arXiv:math.MG/0407278

[15] J. R. Lee and A. Naor. Embedding the diamond graph in $L_p^\mathbb{N}$ and dimension reduction in $L_1$. *Geom. Funct. Anal.*, 14(4):745–747, 2004. arXiv:math.FA/0407520, doi:10.1007/s00039-004-0473-8
[16] J. R. Lee, A. Naor, and Y. Peres. Trees and Markov convexity. *Geom. Funct. Anal.*, 18(5):1609–1659, 2009. [arXiv:0706.0545](https://arxiv.org/abs/0706.0545), doi:10.1007/s00039-008-0689-0.

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

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

[19] J. Matoušek. 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](http://kam.mff.cuni.cz/~matousek/rams.ps.gz).

[20] B. Maurey and G. Pisier. Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. *Studia Math.*, 58(1):45–90, 1976.

[21] M. Mendel. Metric dichotomies. In *Limits of graphs in group theory and computer science*, pages 59–76. EPFL Press, Lausanne, 2009. [arXiv:arXiv:0710.1994](https://arxiv.org/abs/0710.1994).

[22] M. Mendel and A. Naor. Scaled Enflo type is equivalent to Rademacher type. *Bull. London Math. Soc.*, 39(3):493–498, 2007. [arXiv:math.FA/0506215](https://arxiv.org/abs/math.FA/0506215), doi:10.1112/blms/bdm016.

[23] M. Mendel and A. Naor. Markov convexity and local rigidity of distorted metrics [extended abstract]. In *Computational geometry (SCG'08)*, pages 49–58. ACM, New York, 2008. [arXiv:arXiv:0803.1697v1](https://arxiv.org/abs/0803.1697v1), doi:10.1145/1377676.1377686.

[24] M. Mendel and A. Naor. Metric cotype. *Ann. of Math. (2)*, 168(1):247–298, 2008. [arXiv:arXiv:math/0506201](https://arxiv.org/abs/math/0506201), doi:10.4007/annals.2008.168.247.

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

[26] I. Newman and Y. Rabinovich. A lower bound on the distortion of embedding planar metrics into Euclidean space. *Discrete Comput. Geom.*, 29(1):77–81, 2003. doi:10.1007/s00454-002-2813-5.

[27] E. Odell and T. Schlumprecht. The distortion problem. *Acta Math.*, 173(2):259–281, 1994. doi:10.1007/BF02398436.

[28] G. Pisier. Sur les espaces de Banach qui ne contiennent pas uniformément de $l^n_1$. *C. R. Acad. Sci. Paris Sér. A-B*, 277:A991–A994, 1973.

[29] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975. doi:10.1007/BF02760337.

[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.