Markov convexity and local rigidity of distorted metrics

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

Assaf Naor†
Courant Institute
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 23
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 $\|x\|_X \leqslant \|Tx\|_Y \leqslant D\|x\|_X$ 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 $p \geqslant 2$, 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 $a, b > 0$, $a\|x\|_X \leqslant |||x||| \leqslant b\|x\|_X$ for all $x \in X$), and a constant $K > 0$ satisfying:

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

X is called superreflexive if it is p-convex for some $p \geqslant 2$ (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*

$$\lim_{n\to\infty} c_X(B_n) = \infty. \tag{2}$$

Here, and in what follows, given two metric spaces $(\mathscr{M}, d_{\mathscr{M}})$, $(\mathscr{N}, d_{\mathscr{N}})$, the parameter $c_{\mathscr{M}}(\mathscr{N})$ denotes the smallest bi-Lipschitz distortion with which $\mathscr{N}$ embeds into $\mathscr{M}$, i.e., the infimum of

[1] 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 $|||x + y||| \leqslant 2 - \delta$.

those $D > 0$ such that there exists a scaling factor $r > 0$ and a mapping $f : \mathscr{N} \to \mathscr{M}$ satisfying $rd_{\mathscr{N}}(x, y) \leqslant d_{\mathscr{M}}(x, y) \leqslant Drd_{\mathscr{N}}(x, y)$ for all $x, y \in \mathscr{N}$ (if no such f exists then set $c_{\mathscr{M}}(\mathscr{N}) = \infty$).

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 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 Ω. Given an integer $k \geqslant 0$, we denote by $\{\widetilde{X}_t(k)\}_{t\in\mathbb{Z}}$ the process which equals X_t for time $t \leqslant k$, and evolves independently (with respect to the same transition probabilities) for time $t > k$. Fix $p > 0$. A metric space (X, d_X) is called Markov p-convex with constant Π if for every Markov chain $\{X_t\}_{t\in\mathbb{Z}}$ on a state space Ω, and every $f : \Omega \to X$,*

$$\sum_{k=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_X\left(f(X_t), f\left(\widetilde{X}_t\left(t-2^k\right)\right)\right)^p\right]}{2^{kp}} \leqslant \Pi^p \cdot \sum_{t\in\mathbb{Z}} \mathbb{E}\big[d_X(f(X_t), f(X_{t-1}))^p\big]. \tag{3}$$

The least constant Π 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

$$\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_X\left(f(X_t), f\left(\widetilde{X}_t\left(t-2^k\right)\right)\right)^p\right]}{2^{kp}} \lesssim_p \sum_{t\in\mathbb{Z}} \mathbb{E}\big[d_X(f(X_t), f(X_{t-1}))^p\big], \tag{4}$$

with asymptotic equality (up to constants depending only on p) for $k \leqslant \frac{\log n}{2}$ 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, 1, 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 $r > 0$,

$$f\left(B_X(x, r)\right) \supseteq B_Y\left(f(x), \frac{r}{c}\right). \tag{5}$$

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 $2 \leqslant p < q$ 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], where it was shown that the answer is positive if (X, d_X) is a metric tree. Here we show that in general 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 $c_X(B_n) \geqslant 2^{\kappa n}$ 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 ℓ_∞ 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 $p \geqslant 2$, the Laakso graph of cardinality n incurs distortion $\Omega\left((\log n)^{1/p}\right)$ in any embedding into a p-convex Banach space.*

Thus, in particular, for $p > 2$ the n-point Laakso graph incurs distortion $\Omega\left((\log n)^{1/p}\right)$ in any embedding into L_p. The case of L_p embeddings of the Laakso graphs when $1 < p \leqslant 2$ 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 $c_X(B_n) \geqslant (\log n)^\alpha$.*

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 $c_X(\{0,1\}^n, \|\cdot\|_1) \geqslant n^\alpha$.*

Theorem 1.8 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 ℓ_∞ grids:

Theorem 1.9 (Grid dichotomy [24])**.** *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 $c_X(\{0,\ldots,n\}^n, \|\cdot\|_\infty) \geqslant n^\alpha$.*

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 $n \geqslant 2$,*

$$s\left(\left\lfloor \frac{n}{40s(n)} \right\rfloor\right)\left(1 - \frac{Cs(n)\log s(n)}{\log n}\right) \leqslant c_X(B_n) \leqslant s(n). \tag{6}$$

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

$$(1 - o(1))s(n_k) \leqslant c_X(B_{n_k}) \leqslant s(n_k). \tag{7}$$

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* $\varepsilon > 0$, $D > 1$ *and* $n \in \mathbb{N}$ *there exists* $m = m(\varepsilon, D, n) \in \mathbb{N}$ *such that*

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

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

$$c_{(\{0,1\}^n,d)}\left(\{0,1\}^m, \|\cdot\|_1\right) \leqslant 1 + \varepsilon.$$

We refer to [6] (see also [30]) for bounds on $m(\varepsilon, D, n)$. 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 $(\{0,\ldots,n\}^n, \|\cdot\|_\infty)$ 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* $D \geqslant 4$ *there exists a metric* d *on* B_∞ *that is* D*-equivalent to the original shortest-path metric on* B_∞*, yet for every* $\varepsilon \in (0,1)$ *and* $m \in \mathbb{N}$*,*

$$c_{(B_\infty,d)}(B_m) \leqslant D - \varepsilon \implies m \leqslant D^{CD^2/\varepsilon}.$$

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 $\mathsf{lca}(x, y)$ denote their least common ancestor. The tree metric on B_∞ is then given by:

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

The metric space X of Theorem 1.10 will be B_∞ 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

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

d_ε 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_∞; see Lemma 6.1. A pictorial description of the metric d_ε 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.

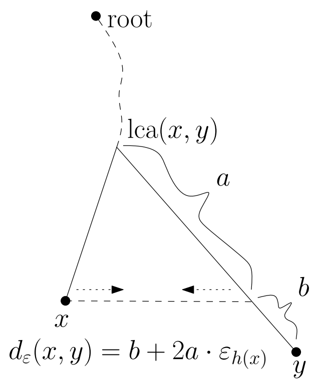


Figure 1: *The metric d_ε defined on B_∞. The arrows indicate horizontal contraction by $\varepsilon_{h(x)}$.*

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_∞ has distortion at most $s(n)$, and therefore $c_X(B_n) \leqslant s(n)$. 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 δ-fork if $y \in \mathrm{Mid}(x, z, \delta) \cap \mathrm{Mid}(x, w, \delta)$, where for $a, b \in X$ the set of δ-approximate midpoints $\mathrm{Mid}(a, b, \delta) \subseteq X$ is defined as the set of all $w \in X$ satisfying $\max\{d_X(x, y), d_X(y, z)\} \leqslant \frac{1+\delta}{2} \cdot d_X(x, z)$. The points z, w will be called below the prongs of the δ-fork (x, y, z, w). Matoušek starts with the observation that if X is a uniformly convex Banach space then in any δ-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 δ-fork in X, provided n is large enough (as a function of D and δ). 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-contractions that they impose of δ-forks. H-trees can even contain larger problematic configurations that have several undistorted δ-forks; such an example is depicted in Figure 2.

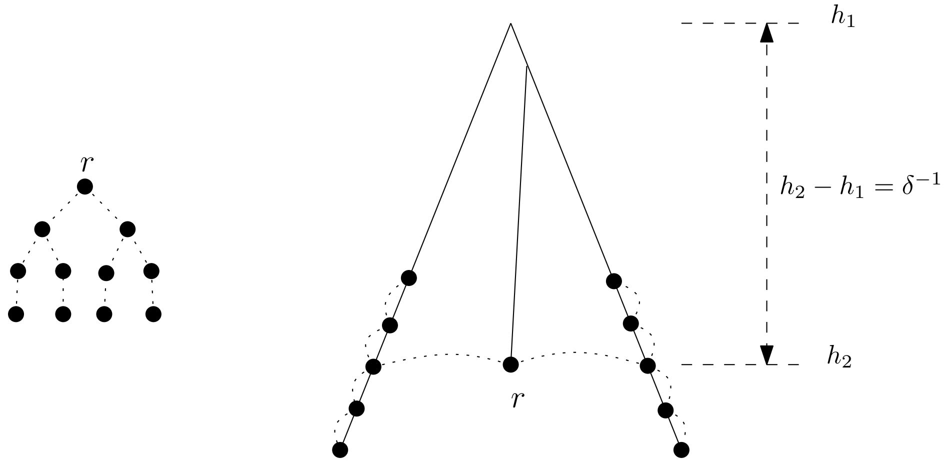


Figure 2: *The metric space on the right is the H-tree $(B_\infty, d_\varepsilon)$, where $\varepsilon_n = \delta$ for all n. The picture describes an embedding of the tree on the left (B_3 minus 4 leaves) into $(B_\infty, d_\varepsilon)$ with distortion at most 6, yet all ancestor/descendant distances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

$$\lambda d_T(x, y) \leqslant d_X(f(x), f(y)) \leqslant D\lambda d_T(x, y). \tag{8}$$

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

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

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 $t \geqslant 2$, $\delta, \xi \in (0,1)$, and $D \geqslant 2$, and assume that $n \in \mathbb{N}$ satisfies*

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

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 $\phi(y)$.*
- $\mathrm{dist}(\phi) \leqslant 1 + \xi$.
- *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 δ-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_4 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 [5]. 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 $a,b\in X$,

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

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

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

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

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

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

We start by proving a useful inequality that is a simple consequence of (10).

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

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

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

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

and

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

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

$$\begin{aligned}\|y-w\|^p+\|z-y\|^p+2\|y-x\|^p&\geqslant \frac{\|x-w\|^p+\|z-x\|^p}{2^{p-1}}+\frac{4}{K^p}\cdot\frac{\left\|y-\frac{x+w}{2}\right\|^p+\left\|y-\frac{z+x}{2}\right\|^p}{2}\\ &\geqslant \frac{\|x-w\|^p+\|z-x\|^p}{2^{p-1}}+\frac{4}{K^p}\cdot\left\|\frac{z-w}{4}\right\|^p,\end{aligned}$$

implying (13). $\square$

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

$$\begin{aligned}&\frac{\|f(X_t)-f(X_{t-2^k})\|^p+\|f(\widetilde{X}_t(t-2^{k-1}))-f(X_{t-2^k})\|^p}{2^{p-1}}+\frac{\|f(X_t)-f(\widetilde{X}_t(t-2^{k-1}))\|^p}{4^{p-1}K^p}\\ &\quad\leqslant \|f(X_{t-2^{k-1}})-f(X_t)\|^p+\|f(X_{t-2^{k-1}})-f(\widetilde{X}_t(t-2^{k-1}))\|^p+2\|f(X_{t-2^{k-1}})-f(X_{t-2^k})\|^p.\end{aligned}$$

Taking expectation, and using the assumption (11), we get

$$\frac{\mathbb{E}\left[\|f(X_t)-f(X_{t-2^k})\|^p\right]}{2^{p-2}}+\frac{\mathbb{E}\left[\|f(X_t)-f(\widetilde{X}_t(t-2^{k-1}))\|^p\right]}{4^{p-1}K^p}$$
$$\leqslant 2\mathbb{E}\left[\|f(X_{t-2^{k-1}})-f(X_t)\|^p\right]+2\mathbb{E}\left[\|f(X_{t-2^{k-1}})-f(X_{t-2^k})\|^p\right].$$

Dividing by $2^{(k-1)p+2}$ this becomes

$$\frac{\mathbb{E}\left[\|f(X_t)-f(X_{t-2^k})\|^p\right]}{2^{kp}}+\frac{\mathbb{E}\left[\|f(X_t)-f(\widetilde{X}_t(t-2^{k-1}))\|^p\right]}{2^{(k+1)p}K^p}$$
$$\leqslant \frac{\mathbb{E}\left[\|f(X_{t-2^{k-1}})-f(X_t)\|^p\right]}{2^{(k-1)p+1}}+\frac{\mathbb{E}\left[\|f(X_{t-2^{k-1}})-f(X_{t-2^k})\|^p\right]}{2^{(k-1)p+1}}.$$

Summing this inequality over $k=1,\ldots,m$ and $t\in\mathbb{Z}$ we get

$$\begin{aligned}
&\sum_{k=1}^{m}\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}\left[\|f(X_t)-f(X_{t-2^k})\|^p\right]}{2^{kp}}+\sum_{k=1}^{m}\sum_{t\in\mathbb{Z}}\frac{\left[\mathbb{E}\|f(X_t)-f(\widetilde{X}_t(t-2^{k-1}))\|^p\right]}{2^{(k+1)p}K^p}\\
\leqslant\ &\sum_{k=1}^{m}\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}\left[\|f(X_{t-2^{k-1}})-f(X_t)\|^p\right]}{2^{(k-1)p+1}}+\sum_{k=1}^{m}\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}\left[\|f(X_{t-2^{k-1}})-f(X_{t-2^k})\|^p\right]}{2^{(k-1)p+1}}\\
=\ &\sum_{j=0}^{m-1}\sum_{s\in\mathbb{Z}}\frac{\mathbb{E}\left[\|f(X_s)-f(X_{s-2^j})\|^p\right]}{2^{jp}}. \qquad (14)
\end{aligned}$$

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:

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

Equivalently,

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

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

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 Π. Then for every $\varepsilon \in (0,1)$ there exists a norm $|||\cdot|||$ on X such that for all $x, y \in X$,*

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

and

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

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

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

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

For $x \in X$ we shall say that a Markov chain $\{X_t\}_{t=-\infty}^{2^m}$ is an m-admissible representation of x if $X_t = 0$ for $t \leqslant 0$ 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

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

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 $t \leqslant 0$ and $X_t = tx$ for $t \in \{1, \ldots, 2^m\}$. This example shows that $|||x|||_m \leqslant \|x\|$. On the other hand, if $\{X_t\}_{t=-\infty}^{2^m}$ is an m-admissible representation of x then

$$\begin{aligned}\sum_{t=1}^{2^m}\mathbb{E}\left[\|X_t - X_{t-1}\|^p\right] - \frac{\eta}{\Pi^p}\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\mathbb{E}\left[\left\|X_t - \widetilde{X}_t(t-2^k)\right\|^p\right]}{2^{kp}} &\overset{(15)}{\geqslant} (1-\eta)\sum_{t=1}^{2^m}\mathbb{E}\left[\|X_t - X_{t-1}\|^p\right] \\ \geqslant (1-\varepsilon)^p\sum_{t=1}^{2^m}\left\|\mathbb{E}[X_t] - \mathbb{E}[X_{t-1}]\right\|^p = (1-\varepsilon)^p\sum_{t=1}^{2^m}\|tx-(t-1)x\|^p &= 2^m(1-\varepsilon)^p\|x\|^p, \end{aligned} \tag{17}$$

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 $x \in X$,

$$(1-\varepsilon)\|x\| \leqslant |||x|||_m \leqslant \|x\|. \tag{18}$$

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

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

and

$$\sum_{t=1}^{2^m}\mathbb{E}\left[\|Y_t-Y_{t-1}\|^p\right]-\frac{\eta}{\Pi^p}\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\mathbb{E}\left[\left\|Y_t-\widetilde{Y}_t(t-2^k)\right\|^p\right]}{2^{kp}}\leqslant 2^m(\|\|y\|\|_m^p+\delta). \tag{20}$$

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

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

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

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

Hence, $Z_t=0$ for $t\leqslant 0$, for $t\in\{1,\ldots,2^m\}$ we have $\mathbb{E}\left[Z_t\right]=\frac{\mathbb{E}[X_t]+\mathbb{E}[Y_t]}{2}=t\cdot\frac{x+y}{2}$, and for $t\in\{2^m+1,\ldots,2^{m+1}\}$ we have

$$\mathbb{E}\left[Z_t\right]=\frac{\mathbb{E}\left[X_{2^m}+Y_{t-2^m}\right]+\mathbb{E}\left[X_{t-2^m}+Y_{2^m}\right]}{2}=\frac{2^mx+(t-2^m)y+(t-2^m)x+2^my}{2}=t\cdot\frac{x+y}{2}.$$

Thus $\{Z_t\}_{t=-\infty}^{2^{m+1}}$ is an $(m+1)$-admissible representation of $\frac{x+y}{2}$. The definition (16) implies that

$$2^{m+1}\left\|\left\|\frac{x+y}{2}\right\|\right\|_{m+1}^p\leqslant\sum_{t=1}^{2^{m+1}}\mathbb{E}\left[\|Z_t-Z_{t-1}\|^p\right]-\frac{\eta}{\Pi^p}\sum_{k=0}^{m+1}\sum_{t=1}^{2^{m+1}}\frac{\mathbb{E}\left[\left\|Z_t-\widetilde{Z}_t(t-2^k)\right\|^p\right]}{2^{kp}}. \tag{21}$$

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,

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

We bound each of the terms in (23) separately. Note that by construction we have for every $t\in\{1,\ldots,2^m\}$,

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

Thus, the first term in the right hand side of (23) can be bounded from below as follows:

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

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

$$Z_t-\widetilde{Z}_t(t-2^k)=\begin{cases}(X_{2^m}+Y_{t-2^m})-\left(\widetilde{X}_{2^m}(t-2^k)+\widetilde{Y}_{t-2^m}(t-2^m-2^k)\right) & \text{with probability } 1/2,\\ (Y_{2^m}+X_{t-2^m})-\left(\widetilde{Y}_{2^m}(t-2^k)+\widetilde{X}_{t-2^m}(t-2^m-2^k)\right) & \text{with probability } 1/2.\end{cases}$$

By Jensen's inequality, if U,V are X-valued independent random variables with $\mathbb{E}[V]=0$, then $\mathbb{E}\left[\|U+V\|^p\right]\geqslant\mathbb{E}\left[\|U+\mathbb{E}[V]\|^p\right]=\mathbb{E}\left[\|U\|^p\right]$. Thus, since $\{X_t\}_{t=-\infty}^{2^m}$ and $\{Y_t\}_{t=-\infty}^{2^m}$ are independent,

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

and

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

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

$$\mathbb{E}\left[\left\|Z_t-\widetilde{Z}_t(t-2^k)\right\|^p\right]$$
$$\geqslant\frac12\mathbb{E}\left[\left\|X_{t-2^m}-\widetilde{X}_{t-2^m}(t-2^m-2^k)\right\|^p\right]+\frac12\mathbb{E}\left[\left\|Y_{t-2^m}-\widetilde{Y}_{t-2^m}(t-2^m-2^k)\right\|^p\right]. \quad (25)$$

Hence,

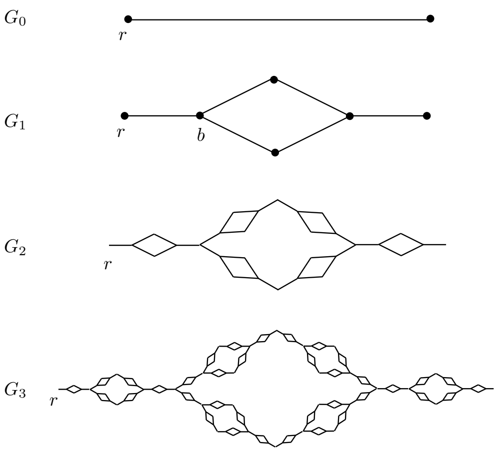
$$\sum_{k=0}^{m}\sum_{t=1}^{2^{m+1}}\frac{\mathbb{E}\left[\left\|Z_t-\widetilde{Z}_t(t-2^k)\right\|^p\right]}{2^{kp}}\overset{(25)}{\geqslant}\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\frac12\mathbb{E}\left[\left\|X_t-\widetilde{X}_t(t-2^k)\right\|^p\right]+\frac12\mathbb{E}\left[\left\|Y_t-\widetilde{Y}_t(t-2^k)\right\|^p\right]}{2^{kp}}$$
$$+\sum_{k=0}^{m}\sum_{t=2^m+1}^{2^{m+1}}\frac{\frac12\left[\mathbb{E}\left\|X_{t-2^m}-\widetilde{X}_{t-2^m}(t-2^m-2^k)\right\|^p\right]+\frac12\mathbb{E}\left[\left\|Y_{t-2^m}-\widetilde{Y}_{t-2^m}(t-2^m-2^k)\right\|^p\right]}{2^{kp}}$$
$$=\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\mathbb{E}\left[\left\|X_t-\widetilde{X}_t(t-2^k)\right\|^p\right]}{2^{kp}}+\sum_{k=0}^{m}\sum_{t=1}^{2^m}\frac{\mathbb{E}\left[\left\|Y_t-\widetilde{Y}_t(t-2^k)\right\|^p\right]}{2^{kp}}. \quad (26)$$

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

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

or,
$$\left|\left|\left|\frac{x+y}{2}\right|\right|\right|_{m+1}^p \leqslant \frac{|||x|||_m^p + |||y|||_m^p}{2} - \frac{\eta}{4\Pi^p(p+1)} \cdot \left\|\frac{x-y}{2}\right\|^p. \tag{27}$$
Define for $w \in X$,
$$|||w||| = \limsup_{m\to\infty} |||w|||_m.$$
Then a combination of (18) and (27) yields that
$$(1-\varepsilon)\|x\| \leqslant |||x||| \leqslant \|x\|,$$
and
$$\begin{aligned}\left|\left|\left|\frac{x+y}{2}\right|\right|\right|^p &\leqslant \frac{|||x|||^p + |||y|||^p}{2} - \frac{\eta}{4\Pi^p(p+1)} \cdot \left\|\frac{x-y}{2}\right\|^p \\ &\leqslant \frac{|||x|||^p + |||y|||^p}{2} - \frac{\eta}{4\Pi^p(p+1)} \cdot \left|\left|\left|\frac{x-y}{2}\right|\right|\right|^p. \end{aligned} \tag{28}$$
Note that (28) implies that the set $\{x \in X : \ |||x||| \leqslant 1\}$ is convex, so that $|||\cdot|||$ is a norm on X. This concludes the proof of Theorem 2.4. □

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



Consider the Laakso graphs [12], $\{G_i\}_{i=0}^{\infty}$, 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 $X_t = X_{4^m}$ for $t > 4^m$.

Proposition 3.1. *For the random walk defined above,*
$$\sum_{k=0}^{2m}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_{G_m}\left(X_t, \widetilde{X}_t(t-2^k)\right)^p\right]}{2^{kp}} \gtrsim \frac{m}{8^p}\sum_{t\in\mathbb{Z}} \mathbb{E}\left[d_{G_m}(X_t, X_{t-1})^p\right]. \tag{29}$$

Proof. For every $t \in \mathbb{Z}$ we have,

$$\mathbb{E}\left[d_{G_m}(X_t, X_{t-1})^p\right] = \begin{cases} 4^{-mp} & t \in \{0, \dots 4^m - 1\}, \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$\sum_{t\in\mathbb{Z}} \mathbb{E}\left[d_{G_m}(X_t, X_{t-1})^p\right] = 4^{-m(p-1)}. \tag{30}$$

Fix $k \in \{0, \dots, 2m-2\}$ and write $h = \lceil k/2 \rceil$. 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 $i \in \{0, \dots, 4^{m-h-1}+1\}$, at time $t = (4i+1)4^h$ the walk X_t 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

$$T_k \stackrel{\mathrm{def}}{=} \{0, \dots, 4^m - 1\} \bigcap \left(\bigcup_{i=0}^{4^{m-h-1}+1} \left[(4i+1)4^h + 4^{h-2}, (4i+1)4^h + 2\cdot 4^{h-2}\right] \right).$$

For $t \in T_k$ find $i \in \{0, \dots, 4^{m-h-1}+1\}$ 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 $\{\widetilde{X}_s(t-2^k)\}_{s\in\mathbb{Z}}$ 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 $\widetilde{X}_t(t-2^k)$ 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 $d_{G_m}\left(X_t, \widetilde{X}_t(t-2^k)\right) \geqslant 2\cdot 4^{h-2}\cdot 4^{-m} = 2^{2h-3-2m}$, and therefore,

$$\frac{\mathbb{E}\left[d(X_t, \widetilde{X}_t(t-2^k))^p\right]}{2^{kp}} \geqslant \frac{\frac{1}{2}2^{(2h-3-2m)p}}{2^{kp}} \geqslant 2^{-(2m+3)p-1}.$$

We deduce that for all $k \in \{0, \dots, 2m-2\}$,

$$\begin{aligned}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d(X_t, \widetilde{X}_t(t-2^k))^p\right]}{2^{kp}} &\geqslant \sum_{t\in T_k} \frac{\mathbb{E}\left[d(X_t, \widetilde{X}_t(t-2^k))^p\right]}{2^{kp}} \geqslant |T_k|\cdot 2^{-(2m+3)p-1} \\ &\gtrsim 4^{h-2}\cdot 4^{m-h-1}\cdot 2^{-(2m+3)p-1} \gtrsim \frac{1}{8^p}4^{-m(p-1)}. \end{aligned}\tag{31}$$

A combination of (30) and (31) implies (29). $\square$

Proof of Theorem 1.5. 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 $p \in (0, \infty)$. $\square$

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

$$x, y \in G_m \implies \frac{1}{A} d_{G_m}(x, y) \leqslant d_X(f(x), f(y)) \leqslant B d_{G_m}(x, y). \tag{32}$$

Let $\{X_t\}_{t\in\mathbb{Z}}$ be the random walk from Proposition 3.1. Then

$$\begin{aligned}
\frac{m}{8^p A^p} \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d_{G_m}(X_t, X_{t-1})^p\right] &\overset{(29)}{\lesssim} \frac{1}{A^p} \sum_{k=0}^{2m} \sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_{G_m}\big(X_t, \widetilde{X}_t(t-2^k)\big)^p\right]}{2^{kp}} \\
&\overset{(32)}{\leqslant} \sum_{k=0}^{2m} \sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_X\big(f(X_t), f(\widetilde{X}_t(t-2^k))\big)^p\right]}{2^{kp}} \\
&\overset{(3)}{\leqslant} \Pi_p(X)^p \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d_X(f(X_t), f(X_{t-1}))^p\right] \\
&\overset{(32)}{\leqslant} \Pi_p(X)^p B^p \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d_{G_m}(X_t, X_{t-1})^p\right].
\end{aligned}$$

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

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 $ab \leqslant D$ and a mapping $f : X \to Y$ such that for all $x \in X$ and $r > 0$,

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

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

The following proposition implies Theorem 1.4.

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

Proof. Fix $f : X \to Y$ satisfying (33). Also, fix a Markov chain $\{X_t\}_{t\in\mathbb{Z}}$ on a state space Ω, and a mapping $g : \Omega \to Y$.

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

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

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 $(\omega_m, \ldots, \omega_t) \in \Omega^*$,

$$d_X\left(h^*(\omega_m, \ldots, \omega_{t-1}), h^*(\omega_m, \ldots, \omega_t)\right) \leqslant a d_Y(g(\omega_{t-1}), g(\omega_t)). \tag{34}$$

For $\omega^* \in \Omega^*$, we will define $h^*(\omega^*)$ by induction on the length of ω^*. 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 $r = ad_Y(g(\omega_{t-1}), g(\omega_t))$. 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 $d_X(x,y) \leqslant r$. We then define $h^*((\omega_m, \ldots, \omega_{t-1}, \omega_t)) \stackrel{\mathrm{def}}{=} y$.

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

$$\sum_{k=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_X\big(h^*(X_t^*), h^*(\widetilde{X}_t^*(t-2^k))\big)^p\right]}{2^{kp}} \leqslant \Pi_p(X)^p \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d_X(h^*(X_t^*), h^*(X_{t-1}^*))^p\right]. \tag{35}$$

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

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

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

$$\sum_{t\in\mathbb{Z}} \mathbb{E}\left[d_X(h^*(X_t^*), h^*(X_{t-1}^*))^p\right] \leqslant a^p \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d_Y(g(X_t), g(X_{t-1}))^p\right]. \tag{36}$$

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

$$\begin{aligned} d_X\big(h^*(X_t^*), h^*(\widetilde{X}_t^*(t-2^k))\big) &\geqslant \frac{1}{b} d_Y\big(f(h^*(X_t^*)), f(h^*(\widetilde{X}_t^*(t-2^k)))\big) \\ &= \frac{1}{b} d_Y\big(g^*(X_t^*), g^*(\widetilde{X}_t^*(t-2^k))\big) = \frac{1}{b} d_Y\big(g(X_t), g(\widetilde{\ }X_t(t-2^k))\big) \end{aligned}$$

Thus,

$$\sum_{k=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_X\big(h^*(X_t^*), h^*(\widetilde{X}_t^*(t-2^k))\big)^p\right]}{2^{kp}} \geqslant \frac{1}{b^p} \sum_{k=0}^{\infty} \sum_{t=m+2^k}^{\infty} \frac{\mathbb{E}\left[d_Y\big(g(X_t), g(\widetilde{X}_t(t-2^k))\big)^p\right]}{2^{kp}}. \tag{37}$$

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

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

Since this inequality holds for every Markov chain $\{X_t\}_{t\in\mathbb{Z}}$ and every $g : \Omega \to Y$, and since $ab \leqslant D$, we have proved that $\Pi_p(Y) \leqslant D\Pi_p(X)$, 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 $\phi : \mathscr{M} \to \mathscr{N}$ is called a *rescaled isometry* if $\mathrm{dist}(\phi) = 1$, or equivalently there exists $\lambda > 0$ such that $d_{\mathscr{N}}(\phi(x), \phi(y)) = \lambda d_{\mathscr{M}}(x,y)$ for all $x, y \in \mathscr{M}$. 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 $\delta \in (0,1)$, $D \geqslant 2$ and $t, n \in \mathbb{N}$ satisfying $n \geqslant D^{(4t\log t)/\delta}$. If $f : P_n \to X$ satisfies* $\mathrm{dist}(f) \leqslant D$ *then there exists a rescaled isometry $\phi : P_t \to P_n$ such that* $\mathrm{dist}(f \circ \phi) \leqslant 1 + \delta$.

Given a metric space (X, d_X) and a nonconstant mapping $f : P_n \to X$, define

$$T(X,f) \stackrel{\mathrm{def}}{=} \frac{d_X(f(0), f(n)}{n \max_{i\in\{1,\ldots,n\}} d_X(f(i-1), f(i))} = \frac{d_X(f(0), f(n))}{n\|f\|_{\mathrm{Lip}}}.$$

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 $T(X,f) \leqslant 1$.

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*

$$T(X,f) \leqslant T\left(X, f\circ\phi^{(m)}\right) \cdot T\left(X, f\circ\phi^{(n)}\right).$$

Proof. Fix $f : P_{mn} \to X$ and define $\phi^{(m)} : P_m \to P_{mn}$ by $\phi^{(m)}(i) = in$. Then,

$$d_X(f(0), f(mn)) \leqslant T\left(X, f\circ\phi^{(m)}\right) m \max_{i\in\{1,\ldots,m\}} d_X(f((i-1)n), f(in)). \tag{38}$$

Similarly, for every $i \in \{1,\ldots,m\}$ define $\phi_i^{(n)} : P_n \to P_{mn}$ by $\phi_i^{(n)}(j) = (i-1)n + j$. Then

$$d_X(f((i-1)n), f(in)) \leqslant T\left(X, f\circ\phi_i^{(n)}\right) n \max_{j\in\{1,\ldots,n\}} d_X(f((i-1)n+j-1), f((i-1)n+j)). \tag{39}$$

Letting $i \in \{1,\ldots,m\}$ be such that $T\left(X, f\circ\phi_i^{(n)}\right)$ is maximal, and $\phi^{(n)} = \phi_i^{(n)}$, we conclude that

$$d_X(f(0), f(mn)) \stackrel{(38)\wedge(39)}{\leqslant} T\left(X, f\circ\phi^{(m)}\right) T\left(X, f\circ\phi^{(n)}\right) mn \max_{i\in\{1,\ldots,mn\}} d_X(f(i-1), f(i)). \qquad \square$$

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

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

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

Proof. Denote $b = \max_{i\in\{1,\ldots,n\}} d_X(f(i), f(i-1)) > 0$. For every $0 \leqslant i < j \leqslant m$ we have $d_X(f(i), f(j)) \leqslant \sum_{\ell=i+1}^{j} d_X(f(\ell-1), f(\ell)) \leqslant b|j-i|$, and

$$\begin{aligned}(1-\varepsilon)mb &\leqslant T(X,f)mb = d_X(f(0), f(m)) \\ &\leqslant d_X(f(0), f(i)) + d_X(f(i), f(j)) + d_X(f(j), f(m)) \leqslant d_X(f(i), f(j)) + b(m+i-j).\end{aligned}$$

Thus $d_X(f(i), f(j)) \geqslant b(j-i-m\varepsilon) \geqslant (1-m\varepsilon)b|j-i|$. $\square$

Proof of Proposition 5.1. Set $k = \lfloor \log_t n \rfloor$ and denote by I the identity mapping from P_{t^k} to P_n. By Lemma 5.3 we have $T(X, f\circ I) \geqslant 1/D$. An iterative application of Lemma 5.2 implies that there exists a rescaled isometry $\phi : P_t \to P_{t^k}$ such that

$$T(X, f\circ I\circ\phi) \geqslant D^{-1/k} \geqslant e^{-2\log D/\log_t n} \geqslant e^{-\delta/(2t)} \geqslant 1 - \frac{\delta}{2t}.$$

By Lemma 5.4 we therefore have $\mathrm{dist}(f\circ I\circ\phi) \leqslant 1/(1-\delta/2) \leqslant 1+\delta$. $\square$

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 $\mathrm{SP}(T)$ 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 $k \geqslant r^{(m+1)^2}$. Suppose that each of the pairs from $\mathrm{SP}(T_{k,m})$ 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 $\{x, y\} \in \mathrm{SP}(T')$ 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

$$\lambda d_{B_n}(x, y) \leqslant d_X(f(x), f(y)) \leqslant D\lambda d_{B_n}(x, y) \tag{40}$$

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 $m = \lfloor n/(k\ell)\rfloor$. We first construct a mapping $g : T_{2^k,m} \to B_n$ in a top-down manner as follows. If r is the root of $T_{2^k,m}$ then $g(r)$ is defined to be the root of B_n. Having defined $g(u)$, let $v_1, \ldots, v_{2^k} \in T_{2^k,m}$ be the children of u, and let $w_1, \ldots, w_{2^k} \in B_n$ 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 $m\ell k \leqslant n$, which is ensured by our choice of m.

By construction, if $x, y \in T_{2^k,m}$ and x is an ancestor of y, then $g(x)$ is an ancestor of $g(y)$ and $d_{B_n}(g(x), g(y)) = \ell k d_{T_{2^k,m}}(x, y)$. Also, if $x, y \in T_{2^k,m}$ and $\mathsf{lca}(x, y) = u$, then we have $h(\mathsf{lca}(g(x), g(y))) \in \{h(g(u)), h(g(u)) + 1, \ldots, h(g(u)) + k - 1\}$. This implies that

$$((\ell - 1)k + 1)d_{T_{2^k,m}}(x, y) \leqslant d_{B_n}(x, y) \leqslant \ell k d_{T_{2^k,m}}(x, y).$$

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

$$k\ell\lambda d_{T_{2^k,m}}(x, y) \leqslant d_X(F(x), F(y)) \leqslant D\ell k\lambda d_{T_{2^k,m}}(x, y). \tag{41}$$

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

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

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

$$2^k \geqslant r^{(m+1)^2}, \tag{42}$$

by Lemma 5.5 there exists a copy T' of B_m in $T_{2^k,m}$ such that the colors of pairs $\{x, y\} \in \mathrm{SP}(T')$ 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

$$m \geqslant D^{16(t\log t)/\delta}, \tag{43}$$

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 $i, j \in \{0, \ldots, t\}$,

$$\theta b|i - j| \leqslant d_X(F(x_i), F(x_j)) \leqslant \left(1 + \frac{\delta}{4}\right)\theta b|i - j|. \tag{44}$$

Define a rescaled isometry $\varphi : B_t \to T'$ 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 $G : B_t \to X$ 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.,

$$\left\lfloor \log_{1+\frac{\delta}{4}} \left(\frac{d_X(G(x), G(y))}{k\ell\lambda b d_{B_t}(x, y)} \right) \right\rfloor = \chi(\{\varphi(x), \varphi(y)\}) = \chi(\{x_i, x_j\}) = \left\lfloor \log_{1+\frac{\delta}{4}} \left(\frac{d_X(F(x_i), F(y_j))}{k\ell\lambda b d_{B_t}(x, y)} \right) \right\rfloor .$$

Consequently, using (44) we deduce that

$$\frac{\theta b}{1+\delta/4} d_{B_t}(x, y) \leqslant d_X(G(x), G(y)) \leqslant \left(1 + \frac{\delta}{4}\right)^2 \theta b d_{B_t}(x, y).$$

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

It remains to determine the values of the auxiliary parameters ℓ, k, which will lead to the desired restriction on n given in (9). First of all, we want to have $\mathrm{dist}(g \circ \varphi) \leqslant 1 + \xi$. Since φ is a rescaled isometry and (for $\ell \geqslant 2$) $\mathrm{dist}(g) \leqslant 1 + 2/\ell$, we choose $\ell = \lceil 2/\xi \rceil \geqslant 2$. We will choose k so that $4k \leqslant n\xi$, so that $n/(k\ell) \geqslant 1$. Since $m = \lfloor n/(k\ell) \rfloor$, we have $m + 1 \leqslant n\xi/k$ and $m \geqslant n\xi/(4k)$. Recall that $r = \lceil \log_{1+\delta/4} D \rceil \leqslant 2 \log_{1+\delta/4} D \leqslant 16D/\delta$. Hence the requirement (42) will be satisfied if

$$2^{k^3} \geqslant \left(\frac{16D}{\delta} \right)^{n^2\xi^2}, \tag{45}$$

and the requirement (43) will be satisfied if

$$\frac{n\xi}{4k} \geqslant D^{16(t \log t)/\delta}. \tag{46}$$

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

$$\left(n^2\xi^2 \log_2 \left(\frac{16D}{\delta} \right) \right)^{1/3} + 1 \leqslant \frac{n\xi}{4D^{16(t \log t)/\delta}},$$

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

6 Tree metrics do not have the dichotomy property

This section is devoted to the proofs of Theorem 1.10 and Theorem 1.12. These proofs were outlined in Section 1.1.1, 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_ε is a metric on B_∞*

Proof. Take $x, y, z \in B_\infty$ and without loss of generality assume that $h(x) \leqslant h(y)$. We distinguish between the cases $h(z) > h(y)$, $h(x) \leqslant h(z) \leqslant h(y)$ and $h(z) < h(x)$.

If $h(z) > h(y)$ then

$$\begin{aligned}
& d_\varepsilon(x,z) + d_\varepsilon(z,y) - d_\varepsilon(x,y) \\
&\quad = \; 2[h(z) - h(y)] + 2\varepsilon_{h(x)} \cdot [h(\mathsf{lca}(x,y)) - h(\mathsf{lca}(x,z))] + 2\varepsilon_{h(y)} \cdot [h(y) - h(\mathsf{lca}(z,y))] \\
&\quad \geqslant \; 2\varepsilon_{h(x)} \cdot [h(\mathsf{lca}(x,y)) - h(\mathsf{lca}(x,z))] + 2\varepsilon_{h(y)} \cdot [h(y) - h(\mathsf{lca}(z,y))]\,. \qquad (47)
\end{aligned}$$

To show that (47) is non-negative observe that this is obvious if $h(\mathsf{lca}(x,y)) \geqslant h(\mathsf{lca}(x,z))$. So assume that $h(\mathsf{lca}(x,y)) < h(\mathsf{lca}(x,z))$. In this case necessarily $h(\mathsf{lca}(z,y)) = h(\mathsf{lca}(x,y))$, so we can bound (47) from below as follows

$$\begin{aligned}
& 2\varepsilon_{h(x)} \cdot [h(\mathsf{lca}(x,y)) - h(\mathsf{lca}(x,z))] + 2\varepsilon_{h(y)} \cdot [h(y) - h(\mathsf{lca}(z,y))] \\
&\quad \geqslant \; 2\varepsilon_{h(x)} \cdot [h(\mathsf{lca}(x,y)) - h(x)] + 2\frac{h(x)}{h(y)}\varepsilon_{h(x)} \cdot [h(y) - h(\mathsf{lca}(x,y))] \\
&\quad = \; 2\varepsilon_{h(x)} \cdot h(\mathsf{lca}(x,y)) \left(1 - \frac{h(x)}{h(y)}\right) \geqslant 0.
\end{aligned}$$

If $h(z) < h(x)$ then

$$\begin{aligned}
& d_\varepsilon(x,z) + d_\varepsilon(z,y) = h(x) + h(y) - 2h(z) + 2\varepsilon_{h(z)} \cdot [2h(z) - h(\mathsf{lca}(x,z)) - h(\mathsf{lca}(z,y))] \\
&\quad \geqslant \; h(y) - h(x) + 2\varepsilon_{h(x)} \cdot [h(x) - h(z)] + 2\varepsilon_{h(x)} \cdot [2h(z) - h(\mathsf{lca}(x,z)) - h(\mathsf{lca}(y,z))] \\
&\quad = \; h(y) - h(x) + 2\varepsilon_{h(x)} \cdot [h(x) + h(z) - h(\mathsf{lca}(x,z)) - h(\mathsf{lca}(y,z))] \\
&\quad \geqslant \; h(y) - h(x) + 2\varepsilon_{h(x)} \cdot [h(x) - h(\mathsf{lca}(x,y))] \qquad (48) \\
&\quad = \; d_\varepsilon(x,y).
\end{aligned}$$

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

It remains to deal with the case $h(x) \leqslant h(z) \leqslant h(y)$. In this case

$$\begin{aligned}
& d_\varepsilon(x,z) + d_\varepsilon(z,y) = h(y) - h(x) + 2\varepsilon_{h(x)} \cdot [h(x) - h(\mathsf{lca}(x,z))] + 2\varepsilon_{h(z)} \cdot [h(z) - h(\mathsf{lca}(y,z))] \\
&\quad \geqslant \; h(y) - h(x) + 2\varepsilon_{h(x)} \cdot [h(x) - h(\mathsf{lca}(x,z))] + 2\frac{h(x)}{h(z)}\varepsilon_{h(x)} \cdot [h(z) - h(\mathsf{lca}(y,z))] \\
&\quad = \; h(y) - h(x) + 2\varepsilon_{h(x)} \cdot \left[2h(x) - h(\mathsf{lca}(x,z)) - \frac{h(x)}{h(z)}h(\mathsf{lca}(y,z))\right] \\
&\quad \geqslant \; h(y) - h(x) + 2\varepsilon_{h(x)} \cdot [h(x) - h(\mathsf{lca}(x,y))] \qquad (49) \\
&\quad = \; d_\varepsilon(x,y),
\end{aligned}$$

where (49) is equivalent to the inequality

$$h(x) \geqslant h(\mathsf{lca}(x,z)) + \frac{h(x)}{h(z)}h(\mathsf{lca}(y,z)) - h(\mathsf{lca}(x,y)). \qquad (50)$$

To prove (50), note that it is true if $h(\mathsf{lca}(x,y)) \geqslant h(\mathsf{lca}(x,z))$, since clearly $h(\mathsf{lca}(y,z)) \leqslant h(z)$. If, on the other hand, $h(\mathsf{lca}(x,y)) < h(\mathsf{lca}(x,z))$ then using the assumption that $h(z) \geqslant h(x)$ it is enough to show that $h(x) \geqslant h(\mathsf{lca}(x,z)) + h(\mathsf{lca}(y,z)) - h(\mathsf{lca}(x,y))$. Necessarily $h(\mathsf{lca}(x,y)) = h(\mathsf{lca}(y,z))$, so that the required inequality follows from the fact that $h(x) \geqslant h(\mathsf{lca}(x,z))$. $\square$

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 $n \in \mathbb{N}$, $\varepsilon_n \geqslant \varepsilon_{n+1} > 0$ and $(n+1)\varepsilon_{n+1} \geqslant n\varepsilon_n$. 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 δ-approximate midpoints of x and z is defined as*

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

¿From now on, whenever we refer to the set $\mathrm{Mid}(x,z,\delta)$, the underlying metric will always be understood to be d_ε. In what follows, given $\eta > 0$ we shall say that two sequences $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ of vertices in B_∞ are η-near if for every $j \in \{1, \ldots, n\}$ we have $d_\varepsilon(u_j, v_j) \leqslant \eta$. We shall also require the following terminology:

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

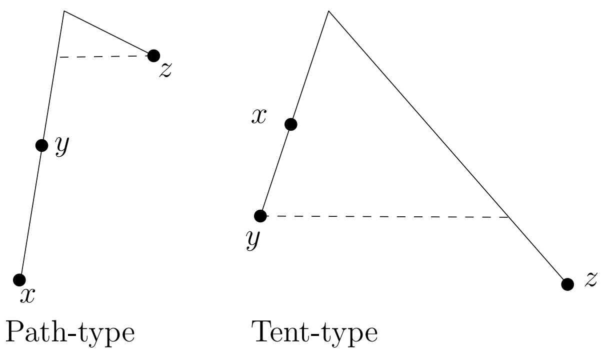


Figure 3: *A schematic description of path-type and tent-type configurations.*

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 $y \in \mathrm{Mid}(x,z,\delta)$. Then either (x,y,z) or (z,y,x) is $3\delta d_\varepsilon(x,z)$-near a path-type or tent-type configuration.*

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 $\varepsilon_n \leqslant \frac{1}{2}$ for all n. Fix $a \in B_\infty$ and let $u, v \in B_\infty$ be its children. For every $x, z \in T_u$ such that $h(x) \geqslant h(z)$ 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 $D_{x,z}(y) \geqslant D_{x,z}(w)$.*

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

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

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

$$Q(k + 1) - Q(k) = -2 + 4\varepsilon_{k+1}[k + 1 - h(a)] - 4\varepsilon_k[k - h(a)] \\ \leqslant -2 + 4\varepsilon_k[k + 1 - h(a)] - 4\varepsilon_k[k - h(a)] = -2(1 - 2\varepsilon_k) \leqslant 0.$$

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

$$Q(k + 1) - Q(k) = 2 [(k + 1)\varepsilon_{k+1} - k\varepsilon_k] + 2h(a)[\varepsilon_k - \varepsilon_{k+1}] \geqslant 0.$$

This completes the proof of Lemma 6.5. □

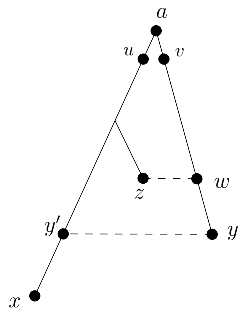
Lemma 6.6. *Assume that $\varepsilon_n < \frac{1}{2}$ for all $n \in \mathbb{N}$. Fix $\delta \in \left(0, \frac{1}{3}\right)$ and $x, y, z \in B_\infty$ such that $h(x) \geqslant h(z)$, $y \in \mathrm{Mid}(x, z, \delta)$ and $h(\mathsf{lca}(x, z)) > h(\mathsf{lca}(x, y))$. Then*

$$h(z) + \frac{1 - 3\delta}{2} d_\varepsilon(x, z) \leqslant h(y) < h(x) \leqslant h(y) + \frac{1 + 3\delta}{1 - 3\delta}[h(y) - h(z)].$$

Moreover, if $y' \in B_\infty$ is the point on the segment joining x and $\mathsf{lca}(x, y)$ such that $h(y') = h(y)$ then $d_\varepsilon(y, y') \leqslant \delta d_\varepsilon(x, z)$. Thus (x, y', z) is a path-type configuration which is $\delta d_\varepsilon(x, z)$-near (x, y, z)

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

By Lemma 6.5,



$$d_\varepsilon(x, y) + d_\varepsilon(z, y) \geqslant d_\varepsilon(x, w) + d_\varepsilon(z, w) \\ = h(x) - h(z) + 4\varepsilon_{h(z)}[h(z) - h(a)] \geqslant 2d_\varepsilon(x, z) - [h(x) - h(z)]. \quad (51)$$

On the other hand, since $y \in \mathrm{Mid}(x, z, \delta)$, we have that

$$d_\varepsilon(x, y) + d_\varepsilon(z, y) \leqslant (1 + \delta)d_\varepsilon(x, z).$$

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

$$\frac{1 + \delta}{2} d_\varepsilon(x, z) \geqslant d_\varepsilon(x, y) \geqslant h(x) - h(y) \geqslant h(x) - h(z).$$

Combining these observations with (51) we get that

$$(1 + \delta)d_\varepsilon(x, z) \geqslant 2d_\varepsilon(x, z) - \frac{1 + \delta}{2} d_\varepsilon(x, z), \tag{52}$$

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

$$\frac{1 + \delta}{2} d_\varepsilon(x, z) \geqslant d_\varepsilon(z, y) \geqslant h(y) - h(z) \geqslant h(x) - h(z),$$

so that we arrive at a contradiction as in (52). We have thus shown that $h(z) < h(y) < h(x)$.

Now, since $y \in \mathrm{Mid}(x,z,\delta)$,

$$\begin{aligned} h(x)-h(z)+2\varepsilon_{h(y)}[h(y)-h(a)]+2\varepsilon_{h(z)}[h(z)-h(a)] &= d_\varepsilon(x,y)+d_\varepsilon(z,y) \\ &\leqslant (1+\delta)d_\varepsilon(x,z) = (1+\delta)\left(h(x)-h(z)+2\varepsilon_{h(z)}[h(z)-h(a)]\right). \end{aligned}$$

Thus, letting y' be the point on the segment joining x and a such that $h(y')=h(y)$, we see that

$$d_\varepsilon(y,y') = 2\varepsilon_{h(y)}[h(y)-h(a)] \leqslant \delta\left(h(x)-h(z)+2\varepsilon_{h(z)}[h(z)-h(a)]\right) = \delta d_\varepsilon(x,z),$$

Moreover

$$\begin{aligned} \frac{1-\delta}{2}d_\varepsilon(x,z) \leqslant d_\varepsilon(y,z) &= h(y)-h(z)+2\varepsilon_{h(z)}h(z)-2\varepsilon_{h(z)}h(a) \\ &\leqslant h(y)-h(z)+2\varepsilon_{h(y)}h(y)-2\varepsilon_{h(y)}h(a) \leqslant h(y)-h(z)+\delta d_\varepsilon(x,z). \end{aligned}$$

Thus

$$h(y)-h(z) \geqslant \frac{1-3\delta}{2}d_\varepsilon(x,z). \tag{53}$$

Hence,

$$\frac{2}{1-3\delta}[h(y)-h(z)] \overset{(53)}{\geqslant} d_\varepsilon(x,z) = [h(x)-h(y)]+[h(y)-h(z)].$$

It follows that

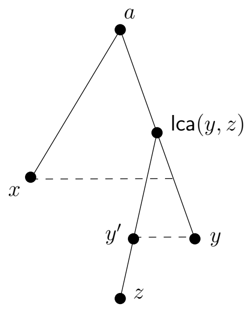
$$h(x)-h(y) \leqslant \frac{1+3\delta}{1-3\delta}[h(y)-h(z)].$$

This completes the proof of Lemma 6.6. $\square$

Lemma 6.7. *Assume that $\varepsilon_n < \frac{1}{4}$ for all $n\in\mathbb{N}$. Fix $\delta \in \left(0,\frac{1}{16}\right)$ and assume that $x,y,z \in B_\infty$ are distinct vertices such that $\mathsf{lca}(x,y)=\mathsf{lca}(x,z)$, and $y \in \mathrm{Mid}(x,z,\delta)$. Then either (x,y,z) or (z,y,x) is $3\delta d_\varepsilon(x,z)$-near a path-type or tent-type configuration.*

Proof. Denote $a=\mathsf{lca}(x,y)$. Our assumption implies that $h(\mathsf{lca}(z,y)) \geqslant h(a)$. We perform a case analysis on the relative heights of x,y,z. Assume first that $h(x)\leqslant h(y)$.

If $h(x)\leqslant h(y)\leqslant h(z)$ then



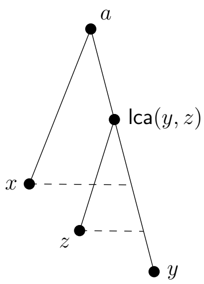
$$\begin{aligned} (1+\delta)d_\varepsilon(x,z) &\geqslant d_\varepsilon(x,y)+d_\varepsilon(y,z) \\ &= h(y)-h(x)+2\varepsilon_{h(x)}[h(x)-h(a)] \\ &\quad + h(z)-h(y)+2\varepsilon_{h(y)}[h(y)-h(\mathsf{lca}(z,y))] \\ &= d_\varepsilon(x,z)+2\varepsilon_{h(y)}[h(y)-h(\mathsf{lca}(z,y))]. \end{aligned} \tag{54}$$

Let y' be the point on the path from $\mathsf{lca}(y,z)$ to z such that $h(y')=h(y)$. Then (54) implies that

$$d_\varepsilon(y,y') = 2\varepsilon_{h(y)}[h(y)-h(\mathsf{lca}(z,y))] \leqslant \delta d_\varepsilon(x,z).$$

Thus the triple (z,y',x) is a configuration of path-type which is $\delta d_\varepsilon(x,z)$-near (z,y,x).

If $h(x) \leqslant h(z) \leqslant h(y)$ then since



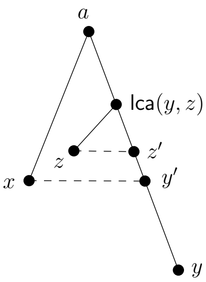
$$\frac{1+\delta}{2} d_\varepsilon(x,z) \geqslant d_\varepsilon(x,y) = h(y) - h(x) + 2\varepsilon_{h(x)}[h(x) - h(a)]$$

and $d_\varepsilon(x,z) = h(z) - h(x) + 2\varepsilon_{h(x)}[h(x) - h(a)]$ we deduce that

$$-\frac{1-\delta}{2} d_\varepsilon(x,z) \geqslant d_\varepsilon(x,y) - d_\varepsilon(x,z) = h(y) - h(z) \geqslant 0.$$

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



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

Moreover,

$$\begin{aligned}
2\varepsilon_{h(z)}[h(\mathsf{lca}(z,y)) - h(a)] &= d_\varepsilon(x,z) - [h(x) - h(z)] - 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(z,y))] \\
&\geqslant \frac{2}{1+\delta} d_\varepsilon(z,y) - [h(x) - h(z)] - 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(z,y))] \\
&= \frac{2}{1+\delta}\big(h(y) - h(z) + 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(y,z))]\big) \\
&\quad - \big(h(y) - h(z) + 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(y,z))]\big) + [h(y) - h(x)] \\
&= \frac{1-\delta}{1+\delta} d_\varepsilon(y,z) + [h(y) - h(x)] \\
&\geqslant \frac{1-\delta}{1+\delta} \cdot \frac{1-\delta}{2} d_\varepsilon(x,z) + [h(y) - h(x)] \\
&\geqslant \left(\frac{1-\delta}{1+\delta}\right)^2 d_\varepsilon(x,y) + [h(y) - h(x)] \\
&= d_\varepsilon(x,y) + [h(y) - h(x)] - \frac{4\delta}{(1+\delta)^2} d_\varepsilon(x,y) \\
&= 2[h(y) - h(x)] + 2\varepsilon_{h(x)}[h(x) - h(a)] - \frac{4\delta}{(1+\delta)^2} d_\varepsilon(x,y) \\
&\geqslant 2[h(y) - h(x)] + 2\varepsilon_{h(z)}h(z) - 2\varepsilon_{h(z)}h(a) - \frac{2\delta}{1+\delta} d_\varepsilon(x,z).
\end{aligned}$$

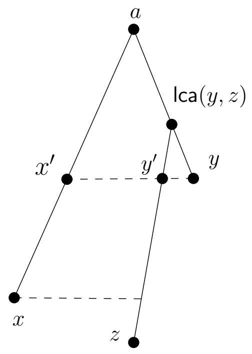
Thus

$$\frac{2\delta}{1+\delta} d_\varepsilon(x,z) \geqslant 2[h(y) - h(x)] + 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(y,z))] = 2[h(y) - h(x)] + d_\varepsilon(z,z'). \tag{55}$$

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 $2\delta d_\varepsilon(x,z)$-near (z,y,x).

This completes the proof of Lemma 6.7 when $h(x) \leqslant h(y)$. The case $h(x) > h(y)$ is proved analogously. Here are the details.

Assume first of all that $h(z) \geqslant h(x) > h(y)$. Then,



$$\begin{aligned} d_\varepsilon(x,z) &\geqslant \frac{2}{1+\delta} d_\varepsilon(z,y) = 2d_\varepsilon(z,y) - \frac{2\delta}{1+\delta} d_\varepsilon(z,y) \\ &= 2\left(h(z) - h(y) + 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(z,y))]\right) - \frac{2\delta}{1+\delta} d_\varepsilon(z,y) \\ &\geqslant 2\left(h(z) - h(y) + 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(z,y))]\right) - \delta d_\varepsilon(x,z). \qquad (56) \end{aligned}$$

On the other hand, since $h(x) > h(y)$,

$$\begin{aligned} d_\varepsilon(x,z) &= h(z) - h(x) + 2\varepsilon_{h(x)}[h(x) - h(a)] \\ &\leqslant h(z) - h(x) + 2\varepsilon_{h(y)}[h(x) - h(a)] \\ &= \left(h(x) - h(y) + 2\varepsilon_{h(y)}[h(y) - h(a)]\right) + h(y) + h(z) - 2h(x) + 2\varepsilon_{h(y)}[h(x) - h(y)] \\ &= d_\varepsilon(x,y) + h(y) + h(z) - 2h(x) + 2\varepsilon_{h(y)}[h(x) - h(y)] \\ &\leqslant \frac{1+\delta}{1-\delta} d_\varepsilon(y,z) + h(y) + h(z) - 2h(x) + 2\varepsilon_{h(y)}[h(x) - h(y)] \\ &= \left(h(z) - h(y) + 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(z,y))]\right) + \frac{2\delta}{1-\delta} d_\varepsilon(y,z) \\ &\quad + h(y) + h(z) - 2h(x) + 2\varepsilon_{h(y)}[h(x) - h(y)] \\ &\leqslant 2[h(z) - h(x)] + 2\varepsilon_{h(y)}[h(x) - h(\mathsf{lca}(z,y))] + \tfrac{1+\delta}{1-\delta}\delta d_\varepsilon(x,z). \end{aligned}$$

Combining this bound with (56), and canceling terms, gives

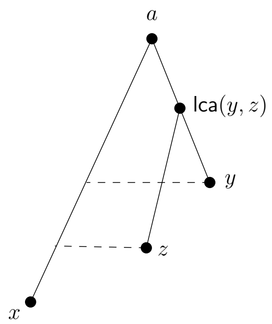
$$\begin{aligned} \frac{2\delta}{1-\delta} d_\varepsilon(x,z) &\geqslant 2[h(x) - h(y)] - 4\varepsilon_{h(y)}[h(x) - h(y)] + 2\varepsilon_{h(y)}[h(x) - h(\mathsf{lca}(z,y))] \\ &\geqslant 2\left(1 - 2\varepsilon_{h(y)}\right)[h(x) - h(y)] + 2\varepsilon_{h(y)}[h(x) - h(\mathsf{lca}(z,y))] \\ &> [h(x) - h(y)] + 2\varepsilon_{h(y)}[h(x) - h(\mathsf{lca}(z,y))], \qquad (57) \end{aligned}$$

where we used the fact that $\varepsilon_{h(y)} < \frac{1}{4}$. Let x' be the point on the path from x to a such that $h(x') = h(y)$, and let y' be the point on the path from a to z such that $h(y') = h(y)$. Then by (57) $d_\varepsilon(x,x') = h(x) - h(y) \leqslant 3\delta d_\varepsilon(x,z)$ and

$$d_\varepsilon(y,y') = 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(z,y))] \leqslant 2\varepsilon_{h(y)}[h(x) - h(\mathsf{lca}(z,y))] \leqslant 3\delta d_\varepsilon(x,z).$$

Thus the triple (z, y', x') is a configuration of path-type which is $3\delta d_\varepsilon(x,z)$-near (z,y,x).

If $h(x) > h(z) \geqslant h(y)$ then



$$h(z) - h(y) + 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(z,y))] = d_\varepsilon(z,y) \geqslant \frac{1-\delta}{1+\delta} d_\varepsilon(x,y)$$

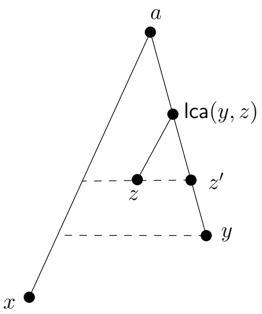
$$= d_\varepsilon(x,y) - \frac{2\delta}{1+\delta} d_\varepsilon(x,y) = h(x) - h(y) + 2\varepsilon_{h(y)}[h(y) - h(a)] - \frac{2\delta}{1+\delta} d_\varepsilon(x,y).$$

Canceling terms we see that

$$\begin{aligned} \frac{2\delta}{1+\delta} d_\varepsilon(x,y) &\geqslant h(x) - h(z) + 2\varepsilon_{h(y)}[h(\mathsf{lca}(z,y)) - h(a)] \\ &= h(x) - h(z) + 2\varepsilon_{h(y)}[h(z) - h(a)] - 2\varepsilon_{h(y)}[h(z) - h(\mathsf{lca}(z,y))] \\ &\geqslant h(x) - h(z) + 2\varepsilon_{h(z)}[h(z) - h(a)] - 2\varepsilon_{h(y)}[h(z) - h(\mathsf{lca}(z,y))] \\ &= d_\varepsilon(x,z) - 2\varepsilon_{h(y)}[h(z) - h(\mathsf{lca}(z,y))] \end{aligned}$$

$$
\begin{aligned}
&\geqslant 2d_\varepsilon(z,y) - \frac{2\delta}{1+\delta} d_\varepsilon(x,z) - 2\varepsilon_{h(y)}[h(z) - h(\mathsf{lca}(z,y))] \\
&= 2\left(h(z) - h(y) + 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(z,y))]\right) - \frac{2\delta}{1+\delta} d_\varepsilon(x,z) \\
&\quad - 2\varepsilon_{h(y)}[h(z) - h(\mathsf{lca}(z,y))] \\
&= 2\left(1 - 2\varepsilon_{h(y)}\right)[h(z) - h(y)] + 2\varepsilon_{h(y)}[h(z) - h(\mathsf{lca}(z,y))] - \frac{2\delta}{1+\delta} d_\varepsilon(x,z) \\
&\geqslant [h(z) - h(y)] + 2\varepsilon_{h(y)}[h(z) - h(\mathsf{lca}(z,y))] - \frac{2\delta}{1+\delta} d_\varepsilon(x,z) \\
&= d_\varepsilon(z,y) - \frac{2\delta}{1+\delta} d_\varepsilon(x,z) \\
&\geqslant \left(\frac{1-\delta}{2} - \frac{2\delta}{1+\delta}\right) d_\varepsilon(x,z),
\end{aligned}
$$

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.

$$
\begin{aligned}
d_\varepsilon(x,z) &= h(x) - h(z) + 2\varepsilon_{h(z)}[h(z) - h(a)] \\
&= d_\varepsilon(y,z) + [h(x) - h(y)] + 2\varepsilon_{h(z)}[h(\mathsf{lca}(y,z)) - h(a)] \\
&\geqslant d_\varepsilon(x,y) - \frac{2\delta}{1+\delta} d_\varepsilon(x,y) + [h(x) - h(y)] + 2\varepsilon_{h(z)}[h(\mathsf{lca}(y,z)) - h(a)] \\
&\geqslant h(x) - h(y) + 2\varepsilon_{h(y)}[h(y) - h(a)] - \delta d_\varepsilon(x,z) \\
&\quad + [h(x) - h(y)] + 2\varepsilon_{h(z)}[h(\mathsf{lca}(y,z)) - h(a)] \\
&\geqslant 2[h(x) - h(y)] + 2\varepsilon_{h(z)}[h(z) - h(a)] + 2\varepsilon_{h(z)}[h(\mathsf{lca}(y,z)) - h(a)] - \delta d_\varepsilon(x,z) \\
&= 2[h(x) - h(y)] + 4\varepsilon_{h(z)}[h(z) - h(a)] - 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(y,z))] - \delta d_\varepsilon(x,z) \\
&= 2d_\varepsilon(x,z) - 2d_\varepsilon(z,y) + 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(y,z))] - \delta d_\varepsilon(x,z) \\
&\geqslant (1-2\delta)\, d_\varepsilon(x,z) + 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(y,z))]. \qquad (58)
\end{aligned}
$$

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

$$
d_\varepsilon(z,z') = 2\varepsilon_{h(z)}[h(z) - h(\mathsf{lca}(y,z))] \overset{(58)}{\leqslant} 2\delta d_\varepsilon(x,z).
$$

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 6.7 is complete. □

Proof of Theorem 6.4. It remains to check that for every $x, y, z \in B_\infty$ such that $y \in \mathrm{Mid}(x,z,\delta)$, 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 $h(\mathsf{lca}(x,y)) = h(\mathsf{lca}(x,z))$ then $\mathsf{lca}(x,y) = \mathsf{lca}(x,z)$, so Lemma 6.7 applies. If $h(\mathsf{lca}(x,y)) < h(\mathsf{lca}(x,z))$ then $\mathsf{lca}(z,y) = \mathsf{lca}(x,z)$, so Lemma 6.7 applies to the triple (z,y,x). If $h(\mathsf{lca}(x,y)) < h(\mathsf{lca}(x,z))$ then $\mathsf{lca}(x,y) = \mathsf{lca}(z,y)$, and so $h(\mathsf{lca}(x,z)) > h(\mathsf{lca}(z,y))$. Hence Lemma 6.6 applies to either the triple (x,y,z) or the triple (z,y,x). □

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 δ-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 $\frac{1}{5}d_\varepsilon(x, y)$-near a path-type configuration and a tent-type configuration.*
2. *(x, y, z) is $\frac{1}{11}d_\varepsilon(x, y)$-near a path-type configuration and (z, y, x) is $\frac{1}{11}d_\varepsilon(x, y)$-near a path-type configuration.*
3. *(x, y, z) is $\frac{1}{11}d_\varepsilon(x, y)$-near a tent-type configuration and (z, y, x) is $\frac{1}{11}d_\varepsilon(x, y)$-near a tent-type configuration.*

Proof. For case 1 of Lemma 6.8, assume for contradiction that (x, y, z) is $\frac{1}{5}d_\varepsilon(x, y)$-near a path-type configuration (a_1, b_1, c_1), and also $\frac{1}{5}d_\varepsilon(x, y)$-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 β_1 is a descendant of α_1. Hence,

$$h(a_1) - h(b_1) = d_\varepsilon(a_1, b_1) \geqslant d_\varepsilon(x, y) - d_\varepsilon(x, a_1) - d_\varepsilon(y, b_1) \geqslant \frac{3}{5}d_\varepsilon(x, y), \tag{59}$$

and

$$h(\beta_1) - h(\alpha_1) = d_\varepsilon(\alpha_1, \beta_1) \geqslant d_\varepsilon(x, y) - d_\varepsilon(x, \alpha_1) - d_\varepsilon(y, \beta_1) \geqslant \frac{3}{5}d_\varepsilon(x, y). \tag{60}$$

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

$$\begin{aligned}\frac{4}{5}d_\varepsilon(x, y) &\geqslant d_\varepsilon(a_1, x) + d_\varepsilon(x, \alpha_1) + d_\varepsilon(b_1, y) + d_\varepsilon(y, \beta_1) \geqslant d_\varepsilon(a_1, \alpha_1) + d_\varepsilon(b_1, \beta_1) \\ &\geqslant h(a_1) - h(\alpha_1) + h(\beta_1) - h(b_1) \overset{(59)\wedge(60)}{\geqslant} \frac{6}{5}d_\varepsilon(x, y),\end{aligned}$$

a contradiction.

For case 2 of Lemma 6.8, assume for contradiction that (x, y, z) is $\frac{1}{11}d_\varepsilon(x, y)$-near a path-type configuration (a_2, b_2, c_2), and also (z, y, x) is $\frac{1}{11}d_\varepsilon(x, y)$-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,

$$h(a_2) - h(b_2) = d_\varepsilon(a_2, b_2) \geqslant d_\varepsilon(x, y) - d_\varepsilon(x, a_2) - d_\varepsilon(y, b_2) \geqslant \frac{9}{11}d_\varepsilon(x, y), \tag{61}$$

and

$$\begin{aligned}h(\beta_2) - h(\gamma_2) + 2\varepsilon_{h(\gamma_2)}[h(\gamma_2) - h(\mathsf{lca}(\beta_2, \gamma_2))] &= d_\varepsilon(\beta_2, \gamma_2) \\ &\geqslant d_\varepsilon(x, y) - d_\varepsilon(x, \gamma_2) - d_\varepsilon(y, \beta_2) \geqslant \frac{9}{11}d_\varepsilon(x, y).\end{aligned} \tag{62}$$

By summing (61) and (62) we see that

$$\begin{aligned}
\frac{17}{11}d_\varepsilon(x,y) \quad &\geqslant \quad d_\varepsilon(a_2,x)+d_\varepsilon(x,\gamma_2)+d_\varepsilon(b_2,y)+d_\varepsilon(y,\beta_2)+d_\varepsilon(\beta_2,y)+d_\varepsilon(x,y)+d_\varepsilon(x,\gamma_2)\\
&\geqslant \quad d_\varepsilon(a_2,\gamma_2)+d_\varepsilon(b_2,\beta_2)+d_\varepsilon(\beta_2,\gamma_2)\\
&\geqslant \quad \big(h(a_2)-h(\gamma_2)\big)+\big(h(\beta_2)-h(b_2)\big)+2\varepsilon_{h(\gamma_2)}[h(\gamma_2)-h(\mathsf{lca}(\beta_2,\gamma_2))]\\
&\overset{(61)\wedge(62)}{\geqslant} \quad \frac{18}{11}d_\varepsilon(x,y),
\end{aligned}$$

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 $(\alpha_3,\beta_3,\gamma_3)$. Then b_3 is a descendant of a_3 and $h(\gamma_3)>h(\beta_3)$. Hence,

$$h(b_3)-h(a_3)=d_\varepsilon(a_3,b_3)\geqslant d_\varepsilon(x,y)-d_\varepsilon(x,a_3)-d_\varepsilon(y,b_3)\geqslant\frac{9}{11}d_\varepsilon(x,y), \tag{63}$$

and

$$\begin{aligned}
h(\gamma_3)-h(\beta_3)+2\varepsilon_{h(\beta_3)}[h(\beta_3)-h(\mathsf{lca}(\beta_3,\gamma_3))]&=d_\varepsilon(\beta_3,\gamma_3)\\
&\geqslant d_\varepsilon(x,y)-d_\varepsilon(x,\gamma_3)-d_\varepsilon(y,\beta_3)\geqslant\frac{9}{11}d_\varepsilon(x,y).
\end{aligned}\tag{64}$$

Hence,

$$\begin{aligned}
\frac{17}{11}d_\varepsilon(x,y) \quad &\geqslant \quad d_\varepsilon(b_3,y)+d_\varepsilon(y,\beta_3)+d_\varepsilon(x,\alpha_3)+d_\varepsilon(x,\gamma_3)+d_\varepsilon(\beta_3,y)+d_\varepsilon(x,y)+d_\varepsilon(x,\gamma_3)\\
&\geqslant \quad d_\varepsilon(b_3,\beta_3)+d_\varepsilon(a_3,\gamma_3)+d_\varepsilon(\beta_3,\gamma_3)\\
&\geqslant \quad \big(h(b_3)-h(\beta_3)\big)+\big(h(\gamma_3)-h(a_3)\big)+2\varepsilon_{h(\beta_3)}[h(\beta_3)-h(\mathsf{lca}(\beta_3,\gamma_3))]\\
&\overset{(63)\wedge(64)}{\geqslant} \quad \frac{18}{11}d_\varepsilon(x,y),
\end{aligned}$$

a contradiction. □

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_∞ "in the same way" that y relates x in B_∞.

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 $d_\varepsilon(y,y')\leqslant d_\varepsilon(x,x')$.*

Proof. Assume without loss of generality that $h(x)\geqslant h(x')$. So,

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

Note that $h(\mathsf{lca}(y,y'))=\min\{h(y'),h(\mathsf{lca}(x,x'))\}$. Hence,

$$\begin{aligned}
d_\varepsilon(y,y')&=h(y)-h(y')+2\varepsilon_{h(y')}[h(y')-h(\mathsf{lca}(y,y'))]\\
&=h(x)-h(x')+2\varepsilon_{h(y')}[h(y')-\min\{h(y'),h(\mathsf{lca}(x,x'))\}]\\
&=h(x)-h(x')+2\varepsilon_{h(y')}\max\{0,h(y')-h(\mathsf{lca}(x,x'))\}.
\end{aligned}\tag{65}$$

If the maximum in (65) is 0, then

$$d_\varepsilon(y, y') = h(x) - h(x') \leqslant d_\varepsilon(x, x').$$

If the maximum in (65) equals $h(y') - h(\mathsf{lca}(x, x'))$, then

$$\begin{aligned} d_\varepsilon(y, y') &= h(x) - h(x') + 2\varepsilon_{h(y')}[h(y') - h(\mathsf{lca}(x, x'))] \\ &\leqslant h(x) - h(x') + 2\varepsilon_{h(x')}[h(x') - h(\mathsf{lca}(x, x'))] = d_\varepsilon(x, x'), \end{aligned} \tag{66}$$

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

Lemma 6.10. *Let $x, x', y \in B_\infty$ be such that $h(y) \leqslant h(x)$. Then there exists $y' \in B_\infty$ which satisfies $h(y') - h(x') = h(y) - h(x)$,*

$$d_\varepsilon(y, y') \leqslant d_\varepsilon(x, x'), \tag{67}$$

and

$$d_\varepsilon(x, y) - 2d_\varepsilon(x, x') \leqslant d_\varepsilon(y', x') \leqslant d_\varepsilon(x, y) + 2d_\varepsilon(x, x'). \tag{68}$$

Proof. Note that (68) follows from (67) by the triangle inequality. Assume first that $h(x) \geqslant h(x')$. In this case choose y' to be an ancestor of y satisfying $h(y) - h(y') = h(x) - h(x')$. Then,

$$d_\varepsilon(y, y') = h(y) - h(y') = h(x) - h(x') \leqslant d_\varepsilon(x, x').$$

We next assume that $h(x) < h(x')$. If $h(\mathsf{lca}(x, x')) \neq h(\mathsf{lca}(x, y))$ 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 $d_\varepsilon(y, y') = h(y') - h(y) = h(x') - h(x) \leqslant d_\varepsilon(x, x')$.

It remains to deal with the case $h(x') > h(x)$ and $h(\mathsf{lca}(x, y)) = h(\mathsf{lca}(x', x))$, which also implies that $h(\mathsf{lca}(x', y)) > h(\mathsf{lca}(x, y))$. In this case, we choose y' to be an arbitrary point on a branch containing both $\mathsf{lca}(x, y)$ and x, such that $h(y') - h(y) = h(x') - h(x)$. Then $\mathsf{lca}(y', y) = \mathsf{lca}(x, x')$, and therefore,

$$\begin{aligned} d_\varepsilon(y, y') &= h(y') - h(y) + 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(y, y')] \\ &= h(x') - h(x) + 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(x, x'))] \\ &\leqslant h(x') - h(x) + 2\varepsilon_{h(x)}[h(x) - h(\mathsf{lca}(x, x'))] \\ &= d_\varepsilon(x, x'), \end{aligned}$$

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*

$$d_\varepsilon(y, y') \leqslant d_\varepsilon(x, x') + 2\varepsilon_{\min\{h(y'),h(y)\}}[h(y) - h(x)] \leqslant d_\varepsilon(x, x') + 2\varepsilon_{h(y)}[h(y) - h(x) + d_\varepsilon(x, x')].$$

Proof. Note that $h(\mathsf{lca}(y, y')) \geqslant h(\mathsf{lca}(x, x'))$. Assume first that $h(x') \geqslant h(x)$. Then,

$$\begin{aligned} d_\varepsilon(y, y') &= h(y') - h(y) + 2\varepsilon_{h(y)}[h(y) - h(\mathsf{lca}(y, y'))] \\ &= h(x') - h(x) + 2\varepsilon_{h(y)}[h(x) - h(\mathsf{lca}(y, y'))] + 2\varepsilon_{h(y)}[h(y) - h(x)] \\ &\leqslant h(x') - h(x) + 2\varepsilon_{h(y)}[h(x) - h(\mathsf{lca}(x, x')] + 2\varepsilon_{h(y)}[h(y) - h(x)] \\ &\leqslant d_\varepsilon(x, x') + 2\varepsilon_{h(y)}[h(y) - h(x)]. \end{aligned}$$

When $h(x') < h(x)$, we similarly obtain the bound:

$$\begin{aligned}d_\varepsilon(y,y') &= h(y)-h(y')+2\varepsilon_{h(y')}[h(y')-h(\mathsf{lca}(y,y'))]\\ &= h(x)-h(x')+2\varepsilon_{h(y')}[h(x')-h(\mathsf{lca}(y,y'))]+2\varepsilon_{h(y')}[h(y')-h(x')]\\ &\leqslant d_\varepsilon(x,x')+2\varepsilon_{h(y')}[h(y)-h(x)].\end{aligned}$$

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

$$\varepsilon_{h(y')}[h(y)-h(x)] = \varepsilon_{h(y')}[h(y')-h(x')] \leqslant \varepsilon_{h(y)}[h(y)-h(x')] \leqslant \varepsilon_{h(y)}[h(y)-h(x)+d_\varepsilon(x,x')]. \quad \square$$

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 δ-fork, if*

$$y \in \mathrm{Mid}(x,z,\delta) \cap \mathrm{Mid}(x,w,\delta).$$

δ-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 δ-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 δ-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 δ-fork (x,y,z,w) of $(B_\infty, d_\varepsilon)$ is called*

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

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

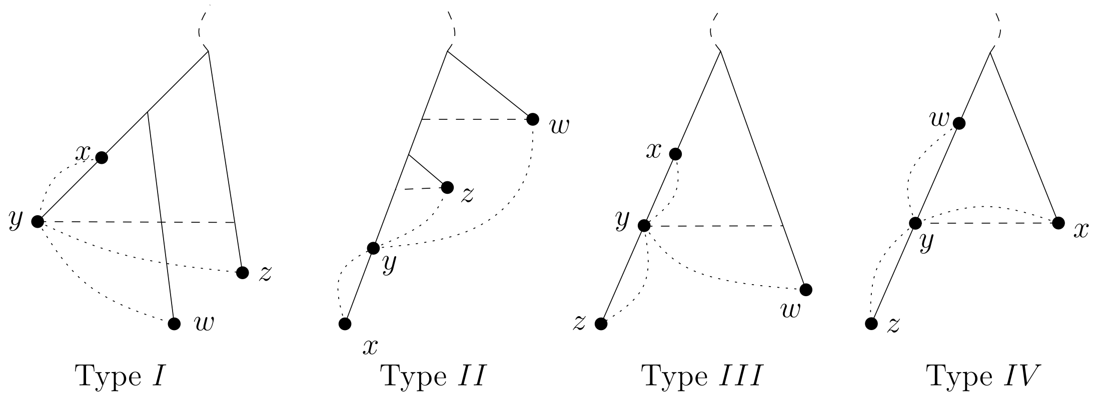


Figure 4: *The four "problematic" types of δ-forks.*

The following lemma is the main result of this section.

Lemma 6.14. *Fix $\delta \in \left(0, \frac{1}{70}\right)$ and assume that $\varepsilon_n < \frac{1}{4}$ for all $n \in \mathbb{N}$. If (x,y,z,w) is a δ-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 $d_\varepsilon(z,w) \leqslant 2(35\delta + \varepsilon_{h_0}) d_\varepsilon(x,y)$, 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 $(p\|p)$ configuration.

Lemma 6.16. *Let (x,y,z,w) be a δ-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*

$$d_\varepsilon(z,w) \leqslant \left(9\eta + 6\delta + 2\varepsilon_{h(y)}\right) d_\varepsilon(x,y).$$

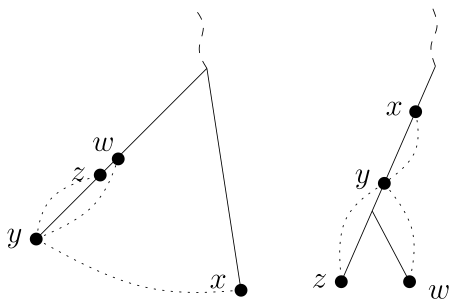


Figure 5: *The two configurations of δ-forks with large contraction of the prongs.*

Proof. Let (z', y', x') be a path-type configuration that is $\eta d_\varepsilon(x,y)$-near (z,y,x), and let (w'', y'', x'') be a path-type configuration that is $\eta d_\varepsilon(x,y)$-near (w,y,x). Without loss of generality assume that $h(y'') \geqslant h(y')$. Let w' be the descendant of y' satisfying $h(w') - h(y') = h(w'') - h(y'')$ such that w' is either an ancestor or an arbitrary descendant of z'. Note that $h(w') \geqslant h(y)$. Indeed,

$$\begin{aligned} h(w') = h(y') + h(w'') - h(y'') = h(y') + d_\varepsilon(w'', y'') \geqslant h(y) - |h(y) - h(y')| + d_\varepsilon(w, y) - 2\eta d_\varepsilon(x,y) \\ \geqslant h(y) - d_\varepsilon(y, y') + \frac{1-\delta}{1+\delta} d_\varepsilon(x,y) - 2\eta d_\varepsilon(x,y) \geqslant h(y) + \left(\frac{1-\delta}{1+\delta} - 3\eta\right) d_\varepsilon(x,y) \geqslant h(y). \end{aligned}$$

By Lemma 6.11,

$$\begin{aligned} d_\varepsilon(w', w'') \leqslant d_\varepsilon(y', y'') + 2\varepsilon_{h(w')} \left[h\left(w''\right) - h\left(y''\right)\right] \leqslant d_\varepsilon(y', y'') + 2\varepsilon_{h(w')} d_\varepsilon(y'', w'') \\ \leqslant 2\eta d_\varepsilon(x,y) + 2\varepsilon_{h(y)} \left(2\eta + \frac{1+\delta}{1-\delta}\right) d_\varepsilon(x,y). \end{aligned} \tag{69}$$

Observe that

$$\begin{aligned} \left(\frac{1-\delta}{1+\delta} - 2\eta\right) d_\varepsilon(x,y) \leqslant d_\varepsilon(z,y) - 2\eta d_\varepsilon(x,y) \leqslant d_\varepsilon(z', y') \\ \leqslant d_\varepsilon(z,y) + 2\eta d_\varepsilon(x,y) \leqslant \left(\frac{1+\delta}{1-\delta} + 2\eta\right) d_\varepsilon(x,y), \end{aligned} \tag{70}$$

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

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

Hence

$$d_\varepsilon(z', w') = \left|d_\varepsilon(y', z') - d_\varepsilon(y', w')\right| \overset{(70)\wedge(71)}{\leqslant} \left(\frac{4\delta}{1-\delta^2} + 4\eta\right) d_\varepsilon(x,y). \tag{72}$$

So, in conclusion,

$$\begin{aligned} d_\varepsilon(z,w) \leqslant d_\varepsilon(z,z') + d_\varepsilon(w,w'') + d_\varepsilon(w'', w') + d_\varepsilon(z', w') \\ \overset{(69)\wedge(72)}{\leqslant} \left(8\eta + \frac{4\delta}{1-\delta^2} + 2\left(2\eta + \frac{1+\delta}{1-\delta}\right)\varepsilon_{h(y)}\right) d_\varepsilon(x,y) \leqslant \left(9\eta + 6\delta + 2\varepsilon_{h(y)}\right) d_\varepsilon(x,y). \end{aligned} \qquad \square$$

We next consider the (t$\|$t) configuration.

Lemma 6.17. *Let (x,y,z,w) be a δ-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*

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

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 $h(y'') - h(w'') \geqslant h(y') - h(z')$. Let $\tilde{w}$ be a point on the path between w'' and y'' such that $h(y'') - h(\tilde{w}) = h(y') - h(z')$. Then,

$$\begin{aligned} d_\varepsilon(w'',\tilde{w}) &= h(y'') - h(w'') - (h(y') - h(z')) = d_\varepsilon(y'',w'') - d_\varepsilon(y',z') \\ &\leqslant d_\varepsilon(y,w) - d_\varepsilon(y,z) + 4\eta d_\varepsilon(x,y) \leqslant \left(\frac{1+\delta}{1-\delta} - \frac{1-\delta}{1+\delta} + 4\eta\right) d_\varepsilon(x,y). \end{aligned} \tag{73}$$

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

$$d_\varepsilon(y,z) \leqslant d_\varepsilon(z,z') + d_\varepsilon(\tilde{w},z') + d_\varepsilon(\tilde{w},w'') + d_\varepsilon(w'',w) \overset{(73)}{\leqslant} \left(\frac{4\delta}{1-\delta^2} + 8\eta\right) d_\varepsilon(x,y). \qquad \square$$

Lemma 6.18. *Let (x,y,z,w) be a δ-fork of B_∞. 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 $3\delta d_\varepsilon(x,z) \leqslant \frac{6}{1-\delta} d_\varepsilon(x,y) \leqslant 7\delta d_\varepsilon(x,y)$-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 tent-type configuration (x'',y'',w''). In this case, x'' is an ancestor of y'' and $h(y'') - h(x'') = d_\varepsilon(x'',y'') \geqslant (1-14\delta) d_\varepsilon(x,y)$. At the same time, y' is an ancestor of x' and $h(x') - h(y') = d_\varepsilon(x',y') \geqslant (1-2\eta) d_\varepsilon(x,y)$. So,

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

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

$$\begin{aligned} &(\eta+7\delta) d_\varepsilon(x,y) \geqslant d_\varepsilon(y',y'') \geqslant h(y'') - h(y') \\ &\quad = (h(y'') - h(x'')) + (h(x'') - h(x')) + h(x') - h(y') > 0 - (\eta + 7\delta) d_\varepsilon(x,y) + (1-2\eta) d_\varepsilon(x,y), \end{aligned}$$

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 $|h(y') - h(y'')| \leqslant d_\varepsilon(y',y'') \leqslant (\eta + 7\delta) d_\varepsilon(x,y)$. So, $h(y') \geqslant h(y'') - (\eta + 7\delta) d_\varepsilon(x,y)$. Also,

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

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

$$d_\varepsilon(\bar{w}, w) \leqslant d_\varepsilon(\bar{w}, w''') + d_\varepsilon(w''', w'') + d_\varepsilon(w'', w) \leqslant (2\eta + 21\delta)d_\varepsilon(x, y).$$

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

Lemma 6.19. *Let (x, y, z, w) be a δ-fork of B_∞. Assume that (x, y, z) is $\eta d_\varepsilon(x, y)$-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 $\eta d_\varepsilon(x, y)$-near a tent-type configuration.*

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

$$2\eta d_\varepsilon(x, y) \geqslant d_\varepsilon(\tilde{x}, x^*) \geqslant d_\varepsilon(\tilde{x}, x'') - d_\varepsilon(x^*, x') - d_\varepsilon(x', x'') \geqslant d_\varepsilon(x', y') - 6\eta d_\varepsilon(x, y) \geqslant (1 - 8\eta)d_\varepsilon(x, y),$$

which is a contradiction, since $\eta < 1/10$. □

Proof of Lemma 6.14. Since (x, y, z, w) is a δ-fork, by Theorem 6.4, both (x, y, z) and (x, y, w) are $7\delta d_\varepsilon(x, y)$-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 $35\delta d_\varepsilon(x, y)$-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 $d_\varepsilon(w, z) \leqslant (69\delta + 2\varepsilon_{h_0})d_\varepsilon(x, y)$, and the rest are configurations that are $7\delta d_\varepsilon(x, y)$-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 $x_0, x_1, x_2, x_3 \in B_\infty$ 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 $0 \leqslant i \leqslant j \leqslant 3$ we have*

$$(j - i)L \leqslant d_\varepsilon(x_i, x_j) \leqslant (1 + \delta)(j - i)L.$$

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

As in the case of δ-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, p-t, T-T, T-t, t-T (the midpoint configurations p-p, P-p, t-p, T-p, p-P, t-P, T-P, t-t, 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 δ-forks, it will be beneficial to give names to three special types approximate 3-paths:

Midpoint configuration	Reverse configuration	Type
(P-P)	(p-p)	type A
(P-p)	(P-p)	impossible
(P-T)	(t-p)	impossible
(P-t)	(T-p)	type B
(p-P)	(p-P)	impossible
(p-T)	(t-P)	type C
(p-t)	(T-P)	impossible
(T-T)	(t-t)	possible only as type C
(T-t)	(T-t)	impossible
(t-T)	(t-T)	impossible

Table 2: *The possible configurations of 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:*

- *η-near a type A configuration if both (x_0, x_1, x_2) and (x_1, x_2, x_3) are η-near path-type configurations,*
- *η-near a type B configuration if (x_0, x_1, x_2) is η-near a path-type configuration, and (x_3, x_2, x_1) is η-near tent-type configuration,*
- *η near type C configuration if (x_2, x_1, x_0) is η-near a path-type configuration, and (x_1, x_2, x_3) is η-near a tent-type configuration.*

See also Figure 6.

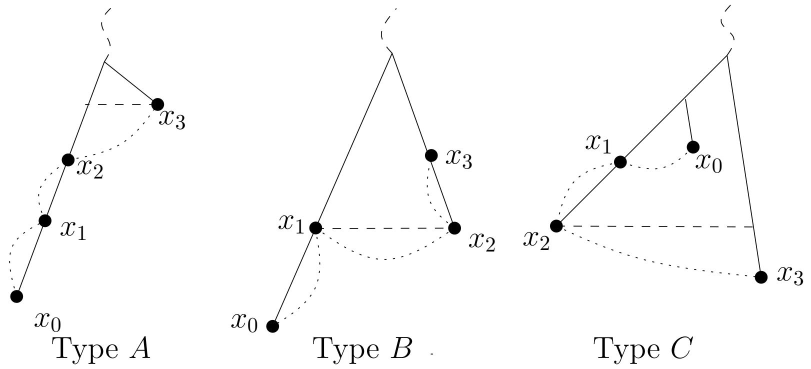


Figure 6: *The three possible types of approximate 3-paths.*

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 $\varepsilon_n < \frac{1}{4}$ 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_1, x_2, x_3) is $7\delta d_\varepsilon(x_0, x_1)$-near a path-type configuration (type A), or (x_3, x_2, x_1) is $7\delta d_\varepsilon(x_0, x_1)$-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 $7\delta d_\varepsilon(x_1, x_2)$-near a path-type configuration, or that (x_1, x_2, x_3) is $7\delta d_\varepsilon(x_1, x_2)$-near a tent-type configuration. Let (x_0', x_1', x_2') be a path-type configuration that is $\eta d_\varepsilon(x_0, x_1)$-near (x_0, x_1, x_2).

Suppose first that (x_3, x_2, x_1) is $7\delta d_\varepsilon(x_1, x_2)$-near the path-type configuration (x_3'', x_2'', x_1''). Since $h(x_1') \geqslant h(x_2')$ and $h(x_2'') \geqslant h(x_1'')$ we have,

$$\begin{aligned} |h(x_1) - h(x_2)| &\leqslant |h(x_1) - h(x_1')| + \big(h(x_1') - h(x_2')\big) + |h(x_2') - h(x_2)| \\ &\leqslant d_\varepsilon(x_1, x_1') + \big(h(x_1') - h(x_2')\big) + d_\varepsilon(x_2', x_2) \leqslant 2\eta d_\varepsilon(x_0, x_1) + \big(h(x_1') - h(x_2')\big), \end{aligned} \tag{74}$$

and similarly,

$$\begin{aligned} |h(x_1) - h(x_2)| &\leqslant |h(x_1) - h(x_1'')| + \big(h(x_2'') - h(x_1'')\big) + |h(x_2'') - h(x_2)| \\ &\leqslant d_\varepsilon(x_1, x_1'') + \big(h(x_2'') - h(x_1'')\big) + d_\varepsilon(x_2', x_2) \leqslant 14\delta d_\varepsilon(x_0, x_1) + \big(h(x_2'') - h(x_1'')\big). \end{aligned} \tag{75}$$

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

$$2|h(x_1) - h(x_2)| \leqslant (2\eta + 14\delta)\, d_\varepsilon(x_0, x_1) + d_\varepsilon(x_1', x_1'') + d_\varepsilon(x_2', x_2'') \leqslant (4\eta + 28\delta)\, d_\varepsilon(x_0, x_1).$$

Thus

$$|h(x_1) - h(x_2)| \leqslant (2\eta + 14\delta) d_\varepsilon(x_0, x_1). \tag{76}$$

Since x_0' is a descendant of x_1',

$$\begin{aligned} |h(x_0) - h(x_1) - d_\varepsilon(x_0, x_1)| &\leqslant |h(x_0') - h(x_1') - d_\varepsilon(x_0, x_1)| + 2\eta d_\varepsilon(x_0, x_1) \\ &= |d_\varepsilon(x_0', x_1') - d_\varepsilon(x_0, x_1)| + 2\delta d_\varepsilon(x_0, x_1) \leqslant 4\eta d_\varepsilon(x_0, x_1). \end{aligned} \tag{77}$$

Similarly, since x_3'' is a descendant of x_2'',

$$\begin{aligned} |h(x_3) - h(x_2) - d_\varepsilon(x_0, x_1)| &\leqslant |h(x_3'') - h(x_2'') - d_\varepsilon(x_0, x_1)| + 14\delta d_\varepsilon(x_1, x_2) \\ &= |d_\varepsilon(x_3'', x_2'') - d_\varepsilon(x_0, x_1)| + 14\delta d_\varepsilon(x_0, x_1) \leqslant 28\delta d_\varepsilon(x_0, x_1). \end{aligned} \tag{78}$$

Hence,

$$\begin{aligned} |h(x_3'') - h(x_0')| &\leqslant |h(x_3'') - h(x_3)| + |h(x_3) - h(x_2) - d_\varepsilon(x_0, x_1)| + |h(x_2) - h(x_1)| \\ &\quad + |h(x_0) - h(x_1) - d_\varepsilon(x_0, x_1)| + |h(x_0) - h(x_0')| \\ &\overset{(76)\wedge(77)\wedge(78)}{\leqslant} d_\varepsilon(x_3'', x_3) + 28\delta d_\varepsilon(x_0, x_1) + (2\eta + 14\delta) d_\varepsilon(x_0, x_1) + 4\eta d_\varepsilon(x_0, x_1) + d_\varepsilon(x_0, x_0') \\ &\leqslant (49\delta + 7\eta) d_\varepsilon(x_0, x_1). \end{aligned} \tag{79}$$

We record for future reference the following consequence of (76) and (79):

$$
\begin{aligned}
\min\{h(x_0'),h(x_3'')\}-\min\{h(x_1'),h(x_2'')\} &\leqslant \max\left\{h(x_0')-h(x_1'),h(x_0')-h(x_2'')\right\}\\
&\overset{(79)}{\leqslant} \max\left\{d_\varepsilon(x_0',x_1'),h(x_3'')-h(x_2'')+(49\delta+7\eta)d_\varepsilon(x_0,x_1)\right\}\\
&\leqslant \max\left\{(1+2\eta)d_\varepsilon(x_0,x_1),d_\varepsilon(x_3'',x_2'')+(49\delta+7\eta)d_\varepsilon(x_0,x_1)\right\}\\
&\leqslant (1+64\delta+7\eta)d_\varepsilon(x_0,x_1). \qquad (80)
\end{aligned}
$$

We next claim that

$$
\mathsf{lca}(x_0',x_3'')=\mathsf{lca}(x_1',x_2''). \qquad (81)
$$

Indeed, since x_1' is an ancestor of x_0' and x_2'' is an ancestor of x_3'', if $\mathsf{lca}(x_0',x_3'')\neq\mathsf{lca}(x_1',x_2'')$ 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

$$
\begin{aligned}
(\eta+7\delta)d_\varepsilon(x_0,x_1)\geqslant d_\varepsilon(x_1',x_1'')\geqslant d_\varepsilon(x_2'',x_1')&\geqslant d_\varepsilon(x_2,x_1)-(\eta+7\delta)d_\varepsilon(x_0,x_1)\\
&\geqslant \frac{1}{1+\delta}d_\varepsilon(x_0,x_1)-(\eta+7\delta)d_\varepsilon(x_0,x_1),
\end{aligned}
$$

which is a contradiction since $\delta,\eta<1/200$. Similarly, if x_2'' is a descendant of x_1' then

$$
(\eta+7\delta)d_\varepsilon(x_0,x_1)\geqslant d_\varepsilon(x_2',x_2'')\geqslant d_\varepsilon(x_2'',x_1')\geqslant \frac{1}{1+\delta}d_\varepsilon(x_0,x_1)-(\eta+7\delta)d_\varepsilon(x_0,x_1),
$$

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

Now,

$$
\begin{aligned}
\frac{3}{1+\delta}d_\varepsilon(x_0,x_1) &\leqslant d_\varepsilon(x_0,x_3)\\
&\leqslant d_\varepsilon(x_3'',x_0')+(\eta+7\delta)d_\varepsilon(x_0,x_1)\\
&\overset{(79)}{\leqslant} 2\varepsilon_{\min\{h(x_0'),h(x_3'')\}}\left[\min\{h(x_0'),h(x_3'')\}-h(\mathsf{lca}(x_0',x_3''))\right]+(8\eta+56\delta)d_\varepsilon(x_0,x_1)\\
&\overset{(81)}{=} 2\varepsilon_{\min\{h(x_0'),h(x_3'')\}}\left[\min\{h(x_1'),h(x_2'')\}-h(\mathsf{lca}(x_1',x_2''))\right]+(8\eta+56\delta)d_\varepsilon(x_0,x_1)\\
&\qquad +2\varepsilon_{\min\{h(x_0'),h(x_3'')\}}\left[\min\{h(x_0'),h(x_3'')\}-\min\{h(x_1'),h(x_2'')\}\right]\\
&\overset{(80)}{\leqslant} 2\varepsilon_{\min\{h(x_1'),h(x_2'')\}}\left[\min\{h(x_1'),h(x_2'')\}-h(\mathsf{lca}(x_1',x_2''))\right]\\
&\qquad +\left(8\eta+56\delta+\frac{1+64\delta+7\eta}{2}\right)d_\varepsilon(x_0,x_1) \qquad (82)\\
&\leqslant d_\varepsilon(x_1',x_2'')+\left(\frac{1}{2}+88\delta+12\eta\right)d_\varepsilon(x_0,x_1)\\
&\leqslant \left(\frac{3}{2}+96\delta+13\eta\right)d_\varepsilon(x_0,x_1), \qquad (83)
\end{aligned}
$$

where in (82) we used $\min\{h(x_0'),h(x_3'')\}\geqslant\min\{h(x_1'),h(x_2'')\}$ 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 $h(x_2') \leqslant h(x_1')$ and $h(x_2'') \geqslant h(x_1'')$, we have

$$\begin{aligned} |h(x_1) - h(x_2)| &\leqslant |h(x_1) - h(x_1')| + \big(h(x_1') - h(x_2')\big) + |h(x_2') - h(x_2)| \\ &\leqslant d_\varepsilon(x_1, x_1') + \big(h(x_1') - h(x_2')\big) + \big(h(x_2'') - h(x_1'')\big) + d_\varepsilon(x_2', x_2) \\ &\leqslant d_\varepsilon(x_1, x_1') + d_\varepsilon(x_1', x_1'') + d_\varepsilon(x_2', x_2'') + d_\varepsilon(x_2', x_2) \\ &\leqslant (4\eta + 14\delta) d_\varepsilon(x_0, x_1). \end{aligned} \tag{84}$$

On the other hand, x_1'' is an ancestor of x_2'', and therefore we have

$$\begin{aligned} \left(\frac{1}{1+\delta} - 14\delta\right) d_\varepsilon(x_0, x_1) &\leqslant d_\varepsilon(x_1'', x_2'') = h(x_2'') - h(x_1'') \\ &\leqslant |h(x_1) - h(x_2)| + 14\delta d_\varepsilon(x_0, x_1) \overset{(84)}{\leqslant} (4\eta + 28\delta) d_\varepsilon(x_0, x_1), \end{aligned} \tag{85}$$

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

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 (reverse type A), or (x_1, x_2, x_3) is $7\delta d_\varepsilon(x_0, x_1)$-near a tent-type configuration (type C).*

Proof. Let (x_2', x_1', x_0') 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 $h(x_1'') \geqslant h(x_2'')$, where as $h(x_2') - h(x_1') = d_\varepsilon(x_2', x_1')$. Arguing as in (84), it follows that $|h(x_1) - h(x_2)| \leqslant (2\eta + 28\delta) d_\varepsilon(x_0, x_1)$, 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 path-type configuration (x_1'', x_2'', x_3''). Then $h(x_1'') - h(x_2'') = d_\varepsilon(x_1'', x_2'')$, whereas $h(x_2') - h(x_1') = d_\varepsilon(x_1', x_2')$. By summing these two identities, we arrive at a contradiction as follows:

$$\begin{aligned} \left(\frac{2}{1+\delta} - 2\eta - 14\delta\right) d_\varepsilon(x_0, x_1) &\leqslant d_\varepsilon(x_1', x_2') + d_\varepsilon(x_1'', x_2'') = \big(h(x_2') - h(x_2'')\big) + \big(h(x_1'') - h(x_1')\big) \\ &\leqslant d_\varepsilon(x_2', x_2'') + d_\varepsilon(x_1', x_1'') \leqslant (2\eta + 14\delta) d_\varepsilon(x_0, x_1). \quad \square \end{aligned}$$

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 tent-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 (type C), or (x_3, x_2, x_1) is $7\delta d_\varepsilon(x_0, x_1)$-near a path-type configuration (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 $|h(x_1') - h(x_1'')| \leqslant d_\varepsilon(x_1', x_1'') \leqslant (\eta + 7\delta) d_\varepsilon(x_0, x_1)$. So, let x_0'' be an ancestor of x_1'' at distance $h(x_1') - h(x_0') - (\eta + 7\delta) d_\varepsilon(x_0, x_1) \in [0, h(x_1'')]$ 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 $d_\varepsilon(x_0^*, x_0') \leqslant (\eta + 7\delta) d_\varepsilon(x_0, x_1)$. By Lemma 6.9,

$$\begin{aligned} d_\varepsilon(x_0, x_0'') - (2\eta + 7\delta) d_\varepsilon(x_0, x_1 &\leqslant d_\varepsilon(x_0, x_0'') - d_\varepsilon(x_0^*, x_0') - d_\varepsilon(x_0', x_0) \leqslant d_\varepsilon(x_0^*, x_0'') \\ &\leqslant d_\varepsilon(x_1', x_1'') \leqslant (\eta + 7\delta) d_\varepsilon(x_0, x_1). \end{aligned}$$

Hence (x_2'', x_1'', 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

$$(1-15\delta)\,d_\varepsilon(x_0,x_1) \leqslant \left(\frac{1}{1+\delta}-14\delta\right)d_\varepsilon(x_0,x_1) \leqslant h(x_2'')-h(x_3'') \leqslant (1+15\delta)d_\varepsilon(x_0,x_1), \tag{86}$$

and

$$(1-\delta-2\eta)\,d_\varepsilon(x_0,x_1) \leqslant \left(\frac{1}{1+\delta}-2\eta\right)d_\varepsilon(x_0,x_1) \leqslant h(x_1')-h(x_0') \leqslant (1+\delta+2\eta)d_\varepsilon(x_0,x_1). \tag{87}$$

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

$$d_\varepsilon(x_3'',x_3^{\#\#}) \leqslant (16\delta+2\eta)\,d_\varepsilon(x_0,x_1), \tag{88}$$

and

$$\begin{aligned} d_\varepsilon(x_0',x_0^{\#}) = \left|h(x_0')-h(x_0^{\#})\right| &= \big|h(x_0')-\big(h(x_1')-h(x_2'')+h(x_3'')+(16\delta+2\eta)\,d_\varepsilon(x_0,x_1)\big)\big| \\ &\overset{(86)\wedge(87)}{\leqslant} 2(16\delta+2\eta)d_\varepsilon(x_0,x_1). \end{aligned} \tag{89}$$

Moreover, $h(x_1)-h(x_0^{\#})=h(x_2'')-h(x_3^{\#\#})$, so by Lemma 6.9 we have

$$\begin{aligned} \left(\frac{3}{1+\delta}-55\delta-7\eta\right)d_\varepsilon(x_0,x_1) \leqslant d_\varepsilon(x_0,x_3)-(55\delta+7\eta)d_\varepsilon(x_0,x_1) &\overset{(88)\wedge(89)}{\leqslant} d_\varepsilon(x_0^{\#},x_3^{\#\#}) \\ &\leqslant d_\varepsilon(x_1',x_2'') \leqslant (1+8\delta+\eta)d_\varepsilon(x_0,x_1), \end{aligned}$$

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 $h(x_1')\leqslant h(x_2')$ we have

$$\begin{aligned} \left(\frac{1}{1+\delta}-14\delta\right)d_\varepsilon(x_0,x_1) \leqslant d_\varepsilon(x_1'',x_2'') = h(x_1'')-h(x_2'') &\leqslant \big(h(x_1'')-h(x_2'')\big)+\big(h(x_2')-h(x_1')\big) \\ &\leqslant d_\varepsilon(x_1'',x_1')+d_\varepsilon(x_2'',x_2') \leqslant (14\delta+2\eta)d_\varepsilon(x_0,x_1), \end{aligned}$$

a contradiction. $\square$

Lemma 6.26. *Assume that $\varepsilon_n<\frac14$ 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_1',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 $h(x_2'')-h(x_1'')=d_\varepsilon(x_1'',x_2'')$. Taking the sum of these two inequalities we conclude that

$$d_\varepsilon(x_1'',x_2'')+d_\varepsilon(x_1',x_2') \leqslant d_\varepsilon(x_1',x_1'')+d_\varepsilon(x_2',x_2'') \leqslant (2\eta+14\delta)d_\varepsilon(x_0,x_1).$$

At the same time, $\left(\frac{2}{1+\delta}-2\eta-14\delta\right)d_\varepsilon(x_0,x_1)\leqslant d_\varepsilon(x_1'',x_2'')+d_\varepsilon(x_1',x_2')$, which leads to the desired contradiction. $\square$

Proof of Lemma 6.22. Since (x_0, x_1, x_2, x_3) is a $(1+\delta)$-approximate P_3, we have $x_1 \in \mathrm{Mid}(x_0, x_2, \delta)$, and $x_2 \in \mathrm{Mid}(x_1, x_3, \delta)$. 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. □

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 $\alpha \in \{0,1\}^*$ (the set of finite sequences of '0' and '1') and $a \in \{0,1\}$ we denote by $v_{\alpha a} = (v_\alpha)_a$.

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

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

$$\mathrm{dist}(f) \geqslant \frac{1}{500\delta + \varepsilon_{h_0}},$$

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 δ-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 δ-forks can be "stitched" together, we reach the conclusion that they are sufficiently severe to force any vertically faithful embedding of B_4 to have a large contraction, and therefore high distortion.

Fix $f : B_4 \to (B_\infty, d_\varepsilon)$. 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

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

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*

$$d_\varepsilon(f(w_0), f(w_1)) \leqslant (170\delta + \varepsilon_{h_0}) \cdot 2\lambda. \tag{90}$$

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

$$d_\varepsilon(f(u_0), f(u_1)) \leqslant (339\delta + 2\varepsilon_{h_0})(1+\delta)\lambda \leqslant (170\delta + \varepsilon_{h_0}) \cdot 2\lambda, \tag{91}$$

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 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 $d_\varepsilon(f(u_{00}), f(u_{01})) \leqslant (170\delta + \varepsilon_{h_0}) \cdot 2\lambda$. □

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 $d_\varepsilon(f(u_{b0}), f(u_{b1})) \leqslant 400\delta\lambda$.*

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 $(f(v), f(u), f(u_0), f(u_{0c}))$ 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 $(f(u_0), f(u), f(v))$ 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 $d_\varepsilon(f(u_{01}), f(u_{00})) \leqslant 400\delta$.
- 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 a II configuration then $d_\varepsilon(f(u_0), f(u_1)) \leqslant 1000\delta\lambda$.*

Proof. By our assumptions, $(f(u), f(u_0), f(u_{00}))$ is $99\delta\lambda$-near a path type configuration (u', u'_0, u'_{00}) and $(f(u), f(u_1), f(u_{10}))$ is $99\delta\lambda$-near a path-type configuration (u'', u''_1, u''_{10}). We may assume without loss of generality that $h(u'') - h(u'_1) \leqslant h(u') - h(u'_0)$. 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 $h(u^*_1) \geqslant h(u'_0)$ (recall that u'_0 is an ancestor of u', and u''_1 is ancestor of u''). By Lemma 6.9 we have

$$d_\varepsilon(u^*_1, u''_1) \leqslant d_\varepsilon(u', u'') \leqslant 198\delta\lambda. \tag{92}$$

Hence,

$$h(u') - h(u^*_1) = d_\varepsilon(u', u^*_1) \overset{(92)}{\geqslant} d_\varepsilon(u', u''_1) - 198\delta\lambda \geqslant d_\varepsilon(f(u), f(u_1)) - 394\delta\lambda \geqslant (1 - 394\delta)\lambda. \tag{93}$$

But, we also know that

$$h(u') - h(u'_0) = d_\varepsilon(u', u'_0) \leqslant d_\varepsilon(f(u), f(u_0)) + 198\delta\lambda \leqslant (1 + 200\delta)\lambda. \tag{94}$$

It follows from (93) and (94) that $d_\varepsilon(u'_0, u^*_1) = h(u^*_1) - h(u'_0) \leqslant 601\delta\lambda$. Therefore,

$$d_\varepsilon(f(u_1), f(u_0)) \leqslant d_\varepsilon(f(u_0), u'_0) + d_\varepsilon(u'_0, u^*_1) + d_\varepsilon(u^*_1, u''_1) + d_\varepsilon(u''_1, f(u_1)) = 1000\delta\lambda. \qquad \square$$

Proof of Lemma 6.27. We may assume that for all $u \in I(B_4)$ the fork $\mathcal{F}(u)$ 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 $\mathcal{F}(r_0)$ 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 $\mathcal{F}(u)$ is $37\delta\lambda$-near a type II configuration. Applying Lemma 6.29, either we are done, or both $\mathcal{F}(u_0)$ and $\mathcal{F}(u_1)$ 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 $\varepsilon_n \leqslant 1/4$. 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_∞ shows that $c_X(B_n) \leqslant s(n)$. 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 $\mathrm{dist}(f) \leqslant s(n)$, and we will now prove that

$$\mathrm{dist}(f) \geqslant s\left(\left\lfloor \frac{n}{40s(n)} \right\rfloor\right)\left(1 - \frac{Cs(n)\log s(n)}{\log n}\right). \tag{95}$$

By adjusting the constant C in (95), we may assume below that n is large enough, say, $n \geqslant 100$. 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 $\lceil n/3 \rceil$, 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 $h_{\max} \geqslant n$. Hence $\|f\|_{\mathrm{Lip}} \geqslant \frac{h_{\max} - h_{\min}}{2n} \geqslant \frac{n - h_0}{2n} \geqslant \frac{1}{4}$. Since $\mathrm{dist}(f) \leqslant s(n)$

we conclude that $\|f^{-1}\|_{\mathrm{Lip}} \leqslant 4s(n)$. It follows that, since $\operatorname{diam}(X \smallsetminus X_{>h_0}) \leqslant 2h_0$, we have $\operatorname{diam}\big(f^{-1}(X \smallsetminus X_{>h_0})\big) \leqslant 8h_0 s(n) \leqslant n/5$. 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 $u \in f^{-1}(X \smallsetminus X_{>h_0})$ of depth at most $\leqslant \lceil n/3 \rceil$. In this case $f^{-1}(X \smallsetminus X_{>h_0})$ must be contained in the first $\lceil n/3 \rceil + n/5 < 2n/3 - 1$ levels of B_n, so we can take T to be any subtree of B_n contained in the last $\lceil n/3 \rceil$ levels of B_n.

Fix $\delta \in (0,1)$. By Theorem 1.14 (with $t = 4$, $D = s(n)$ and $\xi = \delta$), there exists a universal constant $\kappa > 0$ such that if $n \geqslant s(n)^{\kappa/\delta}$ then there exists a mapping $\phi : B_4 \to B_n$ with $\operatorname{dist}(\phi) \leqslant 1+\delta$ such that $f \circ \phi$ is a $(1+\delta)$-vertically faithful embedding of B_4 into $X_{>h_0}$. Choosing $\delta = \kappa \frac{\log s(n)}{\log n}$, by increasing C in (95) if necessary, we may assume that $\delta < 1/400$. Lemma 6.27 then implies

$$(1+\delta)\operatorname{dist}(f) \geqslant \operatorname{dist}(f \circ \phi) \geqslant \frac{1}{500\delta + \varepsilon_{h_0}} = \frac{1}{500\kappa \frac{\log s(n)}{\log n} + \frac{1}{s(\lfloor n/(40 s(n)) \rfloor)}}.$$

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

$$\limsup_{n\to\infty} \frac{s\left(\lfloor n/(40s(n)) \rfloor\right)}{s(n)} = 1. \tag{96}$$

Indeed, if (96) failed then there would exist $\varepsilon_0 \in (0,1)$ and $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$,

$$s\left(\lfloor n/\log n \rfloor\right) \leqslant s\left(\lfloor n/(40s(n)) \rfloor\right) \leqslant (1-\varepsilon_0)s(n). \tag{97}$$

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

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_ε on B_∞ is D-equivalent to the original shortest path metric on B_∞. In this case, if $c_X(B_n) \leqslant D - \varepsilon$ then the bound (95) implies that $n \leqslant D^{CD^2/\varepsilon}$. $\square$

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:

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

A Banach space X is p-convex if and only if its dual space X^* is q-smooth, where $\frac{1}{p} + \frac{1}{q} = 1$ [17]. 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.

[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.

[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.

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

[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.

[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.

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

[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.

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

[16] J. R. Lee, A. Naor, and Y. Peres. Trees and Markov convexity. *Geom. Funct. Anal.*, 18(5):1609–1659, 2009.

[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.

[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.

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

[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.

[24] M. Mendel and A. Naor. Metric cotype. *Ann. of Math. (2)*, 168(1):247–298, 2008.

[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.

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

[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.

[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.