ESTIMATING DISTORTION VIA METRIC INVARIANTS

SEAN LI

1. Introduction

In the latter half of the 20th century, researchers started realizing the importance of understanding how well certain metric spaces can quantitatively embed into other metric spaces. Here "well" depends on the context at hand. In essence, the better a metric space X embeds into Y, the closer geometric properties of X correspond to that of Y. If one continues on this train of thought, then one can try to embed a relatively not well understood metric space X into another more well understood space Y. If such a good embedding exists, then one can try to use the properties and techniques developed for Y to understand the geometry of X. If no such good embedding exists, then one can try to deduce the obstruction of such an embedding, which would give more information for both X and Y.

While part of the motivation came from purely mathematical considerations, the same philosophy also found use in the development of approximation algorithms in theoretical computer science. Indeed, one may be asked to solve a computational problem on a data sets that comes with a natural metric. There are some untractable problems that become much easier to solve when the data set contains some further structure—being Euclidean for example. Thus, be embedding the data set into this easier to solve space, one can speed up the computation at the loss of accuracy that can be bound by the fidelity of the embedding. While the topic of such applications are intresting on their own and could easily (and have) fill books, they are beyond the scope of these notes, and we will not pursue this line any further. See [18] for more information on approximation algorithms.

We now introduce the the quantity to which we measure how well a metric space embeds into one another. Recall that $f : (X, d_X) \to (Y, d_Y)$ is called a biLipschitz embedding if there exists some $D \geq 1$ so that there exists some $s \in \mathbb{R}$ for which

$$sd_X(x, y) \leq d_Y(f(x), f(y)) \leq Dsd_X(x, y).$$

Thus, up to rescalings of the metric, f preserves the metric of X in Y up to a multiplicative factor of D. Here, D is called the biLipschitz distortion (or just distortion) of f. Two metric spaces are said to be biLipschitz equivalent if there exists a surjective biLipschitz embedding between them. We will typically be calculating distortions of embeddings of metric spaces in Banach spaces and so by rescaling the function, we can usually suppose that $s = 1$ or $1/D$.

Given metric spaces X and Y, we let

$$c_Y(X) := \inf\{D : \exists f : X \to Y \text{ with distortion } D\},$$

with the understanding that $c_Y(X) = \infty$ if no such biLipschitz embedding exist. We can then say that $c_Y(X)$ is the distortion of X into Y without referencing any map.

Upper bounding $c_Y(X)$ typically entails constructing an explicit embedding for which you bound the distortion. We will be more interested in lower bounding $c_Y(X)$, for which one has to show that *all* embeddings must have biLipschitz distortion greater than our lower bound. There are many ways to achieve this. In these notes, we will use biLipschitz metric invariants (or just metric invariants) to accomplish such a task. We now introduce our first metric invariant, the Enflo type, which will give a nice simple example to show how one can use such properties to estimate distortion lower bounds.

2. Enflo type

We define the quantity

$$c_2(n) = \sup\{c_{\ell_2}(X) : X \text{ is an } n\text{-point metric space}\}.$$

Thus, every n-point metric space can embed into Hilbert space with distortion no more than $c_2(n)$.

Amazingly, it wasn't until 1970 that it was shown that $\sup_{n\in\mathbb{N}} c_2(n) = \infty$. The first proof of this fact was given by Enflo in [6]. Nowadays, it is known that $c_2(n) \asymp \log n$.[1] The upper bound was found by an explicit embedding by Bourgain [4] in 1985 and the lower bound was first matched by expander graphs as shown in [9] in 1995. The lower bound established in [6] is the following.

Theorem 2.1.

$$c_2(n) \geq \sqrt{\log n}.$$

We now describe the metric space used by Enflo. The Hamming cubes are the metric spaces

$$D_n = (\{0,1\}^n, \|\cdot\|_1).$$

Thus, elements of D_n are strings 0 and 1 of length n. These are just the corners of a cube of the ℓ_1^n normed space. We call two pairs of points in D_n an edge if they are of distance 1 (*i.e.* if their strings differ in only one place) and a diagonal if they are of distance n (*i.e.* if their strings differ at every place). Note that the metric of D_n can also be viewed as a graph path metric based on the set of edges. Each point x has n other points that form edges with x and 1 other point forms a diagonal with x.

To establish our lower bound, we need to show that $c_2(D_n) \geq \sqrt{n}$. One can easily verify that D_n also embed into ℓ_2 with distortion no more than $\sqrt{n}$ if one just embeds the points to the corresponding points of the unit cube in $\mathbb{R}^n$ so our lower bound will actually tight for this specific example.

Enflo proved Theorem 2.1 using the following proposition.

Proposition 2.2. *Let $f : D_n \to \ell_2$ be any map. Then,*

$$\sum_{\{x,y\}\in\mathbf{diags}} \|f(x) - f(y)\|^2 \leq \sum_{\{u,v\}\in\mathbf{edges}} \|f(u) - f(v)\|^2. \tag{1}$$

Note that we are not really using the metric structure of D_n here, just the graph structure. We will first need the following lemma.

Lemma 2.3 (Short diagonals lemma)**.** *Let x, y, z, w be arbitrary points in ℓ_2. Then*

$$\|x - z\|^2 + \|y - w\|^2 \leq \|x - y\|^2 + \|y - z\|^2 + \|z - w\|^2 + \|w - x\|^2.$$

[1] In these notes, we will say $a \lesssim b$ (resp. $a \gtrsim b$) if there exists some absolute constant $C > 0$ so that $a \leq Cb$ (resp. $a \geq Cb$). We write $a \asymp b$ if $a \lesssim b \lesssim a$.

Proof. As the norms are raised to the power 2, they break apart according to their coordinates. Thus, it suffices to prove that

$$(x-z)^2 + (y-w)^2 \le (x-y)^2 + (y-z)^2 + (z-w)^2 + (w-x)^2.$$

But

$$(x-y)^2 + (y-z)^2 + (z-w)^2 + (w-x)^2 - (x-z)^2 - (y-w)^2 = (x-y+z-w)^2 \ge 0.$$

□

Proof of Proposition 2.2. We will induct on n. For the base case when $n = 2$, we simply set x, y, z, w to be the images of the points D_n so that $\{x, z\}$ and $\{y, w\}$ correspond to the images of diagonals. Lemma 2.3 then gives our needed inequality.

Now suppose we have shown the statement for D_{n-1} and consider D_n. Note that D_n can be viewed as two separate copies of D_{n-1}. Indeed, the set of points of D_{n-1} that correspond to strings all beginning with 0 form one such D_{n-1} and the subset corresponding to strings all beginning with 1 form the other. Let $D^{(0)}$ and $D^{(1)}$ denote these two subsets in D_{n-1}. It easily follows that, for each $v \in D^{(0)}$, there exists a unique $w \in D^{(1)}$ for which $\{v, w\}$ form an edge and vice versa. Let $\mathbf{edges}' \subset \mathbf{edges}$ denote this collection of edges. Let $\mathbf{edges}_0$ and $\mathbf{edges}_1$ denote the edges of $D^{(0)}$ and $D^{(1)}$, respectively. Note that these are still edges of D_n and

$$\mathbf{edges} = \mathbf{edges}_0 \cup \mathbf{edges}_1 \cup \mathbf{edges}', \tag{2}$$

where the union above is disjoint. Let $\mathbf{diags}_0$ and $\mathbf{diags}_1$ denote the diagonals of $D^{(0)}$ and $D^{(1)}$. Note that these are *not* diagonals of D_n as their distances are only $n-1$. However, if $\{u, v\}$ is a diagonal of $\mathbf{diags}_0$ and $u', v' \in D^{(1)}$ so that $\{u, u'\}$ and $\{v, v'\}$ are edges of $\mathbf{edges}'$, then $\{u', v'\} \in \mathbf{diags}_1$ and $\{u, v'\}, \{u', v\} \in \mathbf{diags}$.

By the induction hypothesis, we have that

$$\sum_{\{x,y\}\in\mathbf{diags}_0} \|f(x)-f(y)\|^2 \le \sum_{\{u,v\}\in\mathbf{edges}_0} \|f(u)-f(v)\|^2, \tag{3}$$

$$\sum_{\{x,y\}\in\mathbf{diags}_1} \|f(x)-f(y)\|^2 \le \sum_{\{u,v\}\in\mathbf{edges}_1} \|f(u)-f(v)\|^2. \tag{4}$$

Let $\{u, v\} \in \mathbf{diags}_0$. As was stated above, there exists a unique $\{u', v'\} \in \mathbf{diags}_1$ so that $\{u, v'\}, \{u', v\} \in \mathbf{diags}$. Using Lemma 2.3, we get that

$$\begin{aligned}&\|f(u)-f(v')\|^2 + \|f(u')-f(v)\|^2 \\ &\qquad \le \|f(u)-f(v)\|^2 + \|f(v)-f(v')\|^2 + \|f(v')-f(u')\|^2 + \|f(u')-f(v)\|^2.\end{aligned}$$

As all diagonals of $\mathbf{diags}$ can be expressed in such manner, we get that

$$\begin{aligned}&\sum_{\{x,y\}\in\mathbf{diags}} \|f(x)-f(y)\|^2 \\ &\qquad \le \sum_{\{u,v\}\in\mathbf{edges}'} \|f(u)-f(v)\|^2 + \sum_{i=0}^{1} \sum_{\{w,z\}\in\mathbf{diags}_i} \|f(w)-f(z)\|^2. \end{aligned} \tag{5}$$

The proposition now follows immediately from (2), (3), (4), and (5). □

We can now prove our main theorem.

Proof of Theorem 2.1. Let $f : D_n \to \ell_2$ be any embedding satisfying the biLipschitz bounds

$$sd(x,y) \leq \|f(x) - f(y)\| \leq Ds \cdot d(x,y), \tag{6}$$

for some $s \in \mathbb{R}$. We then get from Proposition (2.2) and the biLipschitz bounds of f that

$$\begin{aligned} s^2 n^2 |\mathbf{diags}| = s^2 \sum_{\{x,y\} \in \mathbf{diags}} d(x,y)^2 &\overset{(6)}{\leq} \sum_{\{x,y\} \in \mathbf{diags}} \|f(x) - f(y)\|^2 \\ &\overset{(1)}{\leq} \sum_{\{u,v\} \in \mathbf{edges}} \|f(u) - f(v)\|^2 \overset{(6)}{\leq} D^2 s^2 \sum_{\{u,v\} \in \mathbf{edges}} d(u,v)^2 = D^2 s^2 |\mathbf{edges}|. \end{aligned}$$

In the first and last equalities, we used the fact that edges and diagonals have distance 1 and n, respectively. One easily calculates that $|\mathbf{diags}| = 2^{n-1}$ and $|\mathbf{edges}| = n2^{n-1}$. This gives that

$$D \geq \sqrt{\frac{n^2 |\mathbf{diags}|}{|\mathbf{edges}|}} = \sqrt{\frac{n^2 2^{n-1}}{n 2^{n-1}}} = \sqrt{n}.$$

This shows that

$$c_2(D_n) \geq \sqrt{n},$$

which finishes the proof as $|D_n| = 2^n$. $\square$

Looking back at the proof of Theorem (2.1), we see that the crucial property that allowed everything to work was the fact that ℓ_2 satisfied (1) for every embedding $f : D_n \to \ell_2$. Thus, any metric space (X, d) satisfying (1) for every embedding $f : D_n \to X$ satisfies

$$c_X(D_n) \geq \sqrt{n}.$$

For any $p > 1$, we say that a metric space has *Enflo type* p if there exists some $T > 0$ so that for every $f : D_n \to X$,

$$\sum_{\{x,y\} \in \mathbf{diags}} d(f(x), f(y))^p \leq T^p \sum_{\{u,v\} \in \mathbf{edges}} d(f(u), f(v))^p. \tag{7}$$

We let $T_p(X)$ be the best possible T such that (7) is satisfied is called the Enflo type p constant. We usually do not care about its specific value other than the fact that it exists. A superficial modification to the proof of Theorem (2.1) immediately shows that there exists some $C > 0$ depending on T and p so that

$$c_X(D_n) \geq C n^{1-\frac{1}{p}}$$

and so $c_X(n) \geq C(\log n)^{1-\frac{1}{p}}$ also.

We make a few important observations before moving on from Enflo type.

Observe that having Enflo type p is a biLipschitz invariant, that is, if $f : (X, d_X) \to (Y, d_Y)$ is a bijective biLipschitz map between two metric spaces and one has Enflo type p, then so does the other. Letting $D \geq 1$ be the distortion of f, one can further bound the Enflo constants

$$\frac{1}{D} T_p(X) \leq T_p(Y) \leq D T_p(X). \tag{8}$$

Also, if a metric space X biLipschitz embeds into another metric space Y that has Enflo type p, then X also has Enflo type p.

Note also how the proof of the distortion lower bound comes from the statement of the property. The property gives us a ratio bound of distances in the image. The first step then is to apply the biLipschitz bounds of the embedding to translate that into a ratio bound of distances in the domain along with the distortion constant. The distortion lower bound then follows from using the geometry of the domain to estimate showing its ratio bound of distances. A more succinct way of expressing this comes from (8). One gets from Proposition 2.2 that $T_2(\ell_2) \leq 1$. One can also calculate (as we did) that $T_2(D_n) \geq \sqrt{n}$. Thus, if D is the distortion of $F : D_n \to \ell_2$, one gets

$$D \overset{(8)}{\geq} \frac{T_2(D_n)}{T_2(\ell_2)} \geq \sqrt{n}.$$

Thus, we see that this method follows the philosophy of metric embeddings described in the introduction as getting a good distortion lower bound will come from the fact that the domain's geometry does not allow for the distance ratio to be as good as that in the image. In the case of Enflo type, diagonals in Hilbert space can be much shorter than they are in D_n.

The distortion bound for Hamming cubes is one of the simplest and straightforward bounds one can derive from metric invariants. There are other metric invariant that allow you to calculate distortion bounds of other spaces, but they may not always follow so quickly and easily. The next metric invariant we discuss will also give a simple distortion bound for a different family of metric spaces. But before we introduce it this new metric invariant, we make a brief detour into nonlinear functional analysis to show how certain linear invariants can give rise to metric invariants.

3. The Ribe program

Let X be an infinite dimensional Banach space. Recall that X has Enflo type p if there exists some $T > 0$ so that for all $n \in \mathbb{N}$ and all embeddings $f : D_n \to X$, we have

$$\sum_{\{x,y\} \in \mathbf{diags}} \|f(x) - f(y)\|^p \leq T^p \sum_{\{u,w\} \in \mathbf{edges}} \|f(u) - f(v)\|^p.$$

The reason that this is called Enflo type is because it is a generalization of the linear property Rademacher type, which says that for some $T' > 0$, for any $n \in \mathbb{N}$ and $x_1, ..., x_n \in X$, we have that

$$\mathbb{E}_\varepsilon \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|^p \leq T'^p \sum_{i=1}^n \|x_i\|^p. \tag{9}$$

Here, $\mathbb{E}$ is taking the expectation with respect to uniformly chosen $\varepsilon \in \{-1, 1\}^n$. To see how Enflo type p implies Rademacher type p, simply take f to be the linear function

$$f : D_n \to X$$
$$\{a_1, ..., a_n\} \mapsto \sum_{i=1}^n (2a_i - 1)x_i.$$

Pisier proved a partial converse [16] which says that if X is a Banach space with Rademacher type $p > 1$, then X has Enflo type p' for any $p' < p$. Thus, the Banach spaces with Rademacher type $p > p' > 1$ give us a rich class of Enflo type p' metric spaces. Whether Rademacher type p implies Enflo type p is still an open question.

Note that Rademacher type is a linear isomorphic invariant much like Enflo type is a metric biLipschitz invariant. One can also observe that Rademacher type only depends on the finite dimensional linear substructure of X. Indeed, the defining inequality (9) only needs to be verified on finite subsets of vectors. Such isomorphic properties that depend only on the finite dimensional substructure of a Banach space are called *local* properties.

We recall that a Banach space X is said to be finitely representable in another Banach space Y if there exists some $K < \infty$ so that for each finite dimensional subspace $Z \subset X$, there exists some subspace $Z' \subset X$ so that $d_{BM}(Z, Z') \leq K$. Here, d_{BM} is the Banach-Mazur distance. We have thus shown that local properties are invariant under finitely representability. Examples of local properties include Rademacher type, Rademacher cotype, superreflexivity, uniform convexity, and uniform smoothness.

Ribe proved in [17] the following theorem, which gives a sufficient condition for mutual finite representability.

Theorem 3.1. *Let X and Y be infinite dimensional separable Banach spaces that are uniformly homeomorphic. Then X and Y are mutually finitely representable.*

Ribe's theorem should be compared to the Mazur-Ulam theorem [12], which shows that any bijective isometry between Banach spaces is affine, and Kadets's theorem [7], which states that any two separable Banach spaces are homeomorphic. These two theorems state that the super-rigid world of isometries and the super-relaxed world of homeomorphisms are completely trivial for completely opposite reasons when applied to Banach spaces. Thus, Ribe's theorem states that interesting phenomena exist inbetween these two extremes.

Thus, if two Banach spaces are equivalent in some metrically quantitative way (as expressed by the modulus of continuity for the uniform homeomorphism), then their finite dimensional linear substructures are isomorphically equivalent. In particular, uniform homeomorphisms preserve local properties.

Thus, as uniform homeomorphisms only deal with a Banach space's metric structure, Ribe's theorem suggests that local properties may be recast in purely metric terms. This endeavor to do so is called the Ribe program and has produced a great number of metric invariants, including Enflo type (although Enflo type predates Ribe's theorem). The next metric invariant we will cover is Markov p-convexity, which characterized the local property of having a modulus of uniform convexity of power type p. We will not go into any more details about the rest of the Ribe program, but we refer the interested reader to the surveys [1, 14] and the references that lie therein.

4. Markov convexity

Recall that a Banach space X is said to be uniformly convex if for every $\varepsilon > 0$, there exists some $\delta = \delta(\varepsilon) > 0$ so that for any $x, y \in X$ such that $\|x\|, \|y\| \leq 1$ and $\|x - y\| < \delta$, we have

$$1 - \left\| \frac{x+y}{2} \right\| < \varepsilon.$$

As the name suggests, the unit ball of a uniformly convex Banach space is convex in a uniform fashion. Here, $\delta(\varepsilon)$ is called the modulus of uniform convexity. A Banach space is said to be uniformly convex of power type $p > 1$ (or just p-convex) if there is some $C > 0$ so that the modulus satisfies

$$\delta(\varepsilon) \geq C\varepsilon^p.$$

It easily follows that a p-convex Banach space is p'-convex for all $p' > p$. It was proven in [2] that a Banach space X is p-convex if and only if there exists some $K > 0$ and an equivalent norm $|\cdot|$ so that

$$|x|^p + |y|^p \geq 2\left|\frac{x-y}{2}\right|^p + 2\left|\frac{x+y}{2K}\right|^p, \qquad \forall x, y \in X. \tag{10}$$

This is a one-sided parallelogram inequality with power p. We will use this characterization of p-convex Banach spaces from now on.

Pisier prove in [15] the striking fact that any uniformly convex Banach spaces can be renormed to be p-convex for some $p \in [2, \infty)$. He also proved that p-convexity is preserved by isomorphism and so is actually a local property as it only depends on 2-subspaces of X. Thus, the Ribe program suggests that there is a metric invariant characterizing p-convexity. In [10], the authors introduced a metric invariant known as Markov p-convexity that was shown to be implied by p-convexity. In [13], the authors completed the characterization by showing that any Banach space that was Markov p-convex had an equivalent norm that was p-convex. Before we describe Markov p-convexity, we first must establish some notation.

Given some Markov chain $\{X_t\}_{t\in\mathbb{Z}}$ on a state space Ω and some integer $k \in \mathbb{Z}$, we let $\{\widetilde{X}_t(k)\}_{k\in\mathbb{Z}}$ denote the Markov chain on Ω so that for $t \leq k$, $\widetilde{X}_t(k) = X_t$ and for $t > k$, $\widetilde{X}_t(k)$ evolves independently (but with respect to the same transition probabilities) to X_t. We never specify that the Markov chain has to be time homogeneous. We can now describe Markov p-convexity.

Let $p > 0$. We say that a metric space (X, d) is Markov p-convex if there exists some $\Pi > 0$ so that for every $f : \Omega \to (X, d)$ and every Markov chain $\{X_t\}_{t\in\mathbb{Z}}$ on Ω,

$$\sum_{k=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d\left(f(X_t), f\left(\widetilde{X}_t(t-2^k)\right)\right)^p\right]}{2^{kp}} \leq \Pi^p \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d(f(X_t), f(X_{t-1})^p\right]. \tag{11}$$

It follows immediately that this is indeed a biLipschitz metric invariant.

The full proof of the equivalence of Markov p-convexity with p-convexity is beyond the scope of these notes. We will just prove the easy direction of p-convexity implying Markov p-convexity later. First, we will try to make sense of exactly what Markov convexity is saying. For this, it will be more illuminating to see what spaces are *not* Markov convex.

Markov p-convexity says in essence that independent Markov chains do not drift too far apart compared to how far they travel at all places and all scales. An example of a simple metric space that does not satisfy this property—and the one that motivated the definition of Markov convexity—are complete binary trees. Indeed, the branching nature of trees allow for Markov chains to diverge linearly.

Let $\{X_t\}_{t\in\mathbb{Z}}$ be the standard downward random walk on B_n, the complete binary tree of depth n, where each branching is taken independently with probability $1/2$ and the walk stops completely after it reaches a leaf (thus, B_n is our state space). We can set X_t to be at the root for $t \le 0$. Here, we are using time inhomogeneity.

Proposition 4.1. *Let X_t be the random walk as described above on B_n. Then there exists some constant $C > 0$ depending only on p so that*

$$\sum_{k=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d\left(X_t, \widetilde{X}_t(t-2^k)\right)^p\right]}{2^{kp}} \ge C\log n \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d(X_t, X_{t-1})^p\right].$$

Proof. From the description of the random walk, we can easily compute

$$\sum_{t\in\mathbb{Z}} \mathbb{E}\left[d(X_t, X_{t+1})^p\right] = n. \tag{12}$$

To compute the left hand side of (11), we have when $k \in \left\{0, ..., \left\lfloor \frac{1}{2}\log n\right\rfloor\right\}$ and $t \in \{2^k, ..., n\}$ that

$$\mathbb{E}\left[d\left(X_t, \widetilde{X}_t(t-2^k)\right)^p\right] = 2^{p-1}2^{kp}.$$

Indeed, this is simply because there is a 1/2 chance that X_{t-2^k+1} and $X_{t-2^k+1}(t-2^k)$ are different in which case X_t and $X_t(t-2^k)$ would differ by 2^{k+1}. Thus, we have the lower bound

$$\sum_{k=0}^{\left\lfloor \frac{1}{2}\log n\right\rfloor}\sum_{t=2^k}^{n} \frac{\mathbb{E}\left[d\left(X_t, \widetilde{X}_t(t-2^k)\right)^p\right]}{2^{kp}} = \sum_{k=0}^{\left\lfloor \frac{1}{2}\log n\right\rfloor}\sum_{t=2^k}^{n} 2^{p-1} \ge Cn\log n, \tag{13}$$

where $C > 0$ is some constant depending only on n. By (12) and (13), we have finish the proof. $\square$

We can now prove the following theorem.

Theorem 4.2. *Let $p > 1$ and suppose (X, d) is a metric space that is Markov p-convex. Then there exists some $C > 0$ depending only on X so that $c_X(B_n) \ge C(\log\log|B_n|)^{1/p}$.*

Proof. Let X_t be the random walk on B_n as described above. Let $f : B_n \to X$ be a Lipschitz map with distortion D. Then we have by definition of Markov p-convexity and distortion that there exists some $\Pi > 0$ so that

$$s^p \sum_{k=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d\left(X_t, \widetilde{X}_t(t-2^k)\right)^p\right]}{2^{kp}} \le \sum_{k=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d\left(f(X_t), f\left(\widetilde{X}_t(t-2^k)\right)\right)^p\right]}{2^{kp}}$$

$$\overset{(11)}{\le} 2^{p-1}\Pi^p \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d(f(X_t), f(X_{t-1}))^p\right] \le 2^{p-1}D^p s^p \Pi^p \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d(X_t, X_{t-1})^p\right].$$

Appealing to Proposition 4.1, we see that we must have $D \ge \frac{(2\log n)^{1/p}}{2\Pi}$, which establishes the claim once we remember that $|B_n| = 2^{n+1} - 1$. $\square$

We now prove the following theorem.

Theorem 4.3. *Let $p \in [2, \infty)$ and let X be a p-convex Banach space. Then X is Markov p-convex.*

As an immediate corollary of Theorems 4.2 and 4.3, we get

Corollary 4.4. *Let X be a p-convex Banach space. Then there exists some constant $C > 0$ depending only on X so that $c_X(B_n) \geq C(\log n)^{1/p}$.*

We will follow the proof of Theorem 4.3 as done in [13]. We first need the following fork lemma.

Lemma 4.5. *Let X be a Banach space whose norm $|\cdot|$ satisfies* (10). *Then for every $x, y, z, w \in X$,*

$$\frac{|x-w|^p + |x-z|^p}{2^{p-1}} + \frac{|z-w|^p}{4^{p-1}K^p} \leq |y-w|^p + |z-y|^p + 2|y-x|^p. \tag{14}$$

Proof. We have by (10) that for every $x, y, z, w \in X$ that

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

Thus, adding these two inequalities together and using convexity to $|\cdot|^p$, we get

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

□

This lemma says that the tips of the fork z, w cannot be too far apart if $\{x, y, z\}$ and $\{x, y, w\}$ are almost geodesic. Thus, if z and w are independently evolved Markov chains, this will property essentially tells us then that they cannot diverge far.

We can now prove Theorem 4.3. The only property concerning p-convex Banach spaces we will use in the following proof is (14). However, (14) is a purely metric statement (although it is not biLipschitz invariant). Thus, the following proof shows that any metric space satisfying (14) is Markov p-convex.

Proof of Theorem 4.3. We get from (14) that for every Markov chain $\{X_t\}_{t\in\mathbb{Z}}$, $f : \Omega \to X$, $t \in \mathbb{Z}$, and $k \geq 0$ that

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

Note that $(X_{t-2^k}, \widetilde{X}_t(t-2^{k-1}))$ and (X_{t-2^k}, X_t) have the same distribution by definition of $\widetilde{X}_t(t-2^{k-1})$. Thus, taking expectation, we get that

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

We divide this inequality by $2^{(k-1)p+2}$ to get

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

Sum this inequality over $k=1,...,m$ and $t\in\mathbb{Z}$ to get

$$\begin{aligned}\sum_{k=1}^{m}\sum_{t\in\mathbb{Z}}&\frac{\mathbb{E}\left[|f(X_t)-f(X_{t-2^k})|^p\right]}{2^{kp}}+\sum_{k=1}^{m}\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}\left[|f(X_t)-f(\widetilde{X}_t(t-2^{k-1})|^p\right]}{2^{(k+1)p}K^p}\\
&\leq \sum_{k=1}^{m}\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}\left[|f(X_{t-2^{k-1}})-f(X_t)|^p\right]}{2^{(k-1)p+1}}+\sum_{k=1}^{m}\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}\left[|f(X_{t-2^{k-1}})-f(X_{t-2^k})|^p\right]}{2^{(k-1)p+1}}\\
&=\sum_{j=0}^{m-1}\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}\left[|f(X_t)-f(X_{t-2^j})|^p\right]}{2^{jp}}.\end{aligned}\tag{15}$$

By the triangle inequality, we have that

$$\sum_{t\in\mathbb{Z}}\frac{\mathbb{E}\left[|f(X_t)-f(X_{t-2^j})|^p\right]}{2^{jp}}\leq\sum_{t\in\mathbb{Z}}\mathbb{E}\left[|f(X_t)-f(X_{t+1})|^p\right].$$

We can clearly assume that $\sum_{t\in\mathbb{Z}}\mathbb{E}\left[|f(X_t)-f(X_{t+1})|^p\right]<\infty$ as otherwise the statement of the proposition is trivial. Thus, we have that the summation on the right hand side of (15) is finite for every $m\geq 1$. We can thus subtract the left hand side from the right hand side in (15) to get

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

This is the same as the following inequality

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

Taking $m\to\infty$ then finishes the proof. □

Corollary 4.4 was first proven by Matoušek in [11] using a metric differentiation argument. The result of [11] was itself a sharpening of a result of Bourgain in [5] which says that the finite complete binary trees embed with uniformly bounded distortion into a Banach space X if and only if X is *not* isomorphic to any uniformly convex space. This is actually the first result of the Ribe program giving a metrical characterization of the local property of a space being isomorphic to a uniformly convex space (also called superreflexivity, although this was not the original formulation). It is now known that the statement of Bourgain also holds with the infinite complete binary tree [3].

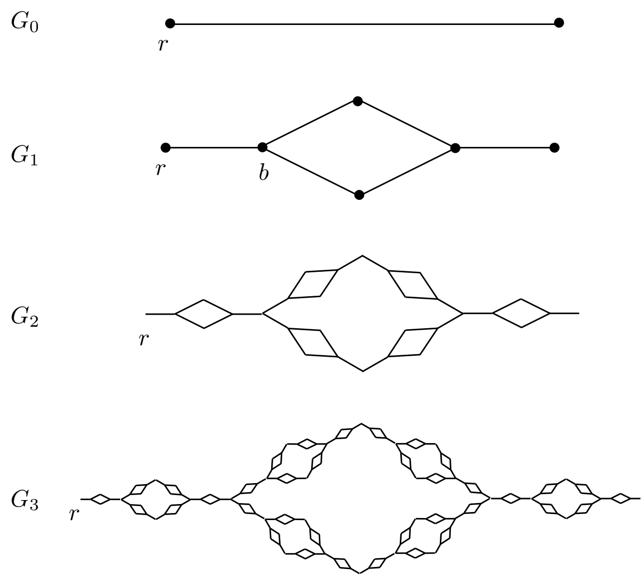

FIGURE 1. The first four Laakso graphs

Another class of metric spaces that lend well for using Markov convexity to estimate distortion bounds are the Laakso graphs. Laakso graphs were described in [8]. We define the graphs $\{G_k\}_{k=0}^{\infty}$ as follows. The first stage $G_0 = 0$ is simply an edge and G_1 is as pictured in Figure 1. To get G_k from G_{k-1}, one replaces all the edges of G_{k-1} with a copy of G_1. The metric is the shortest path metric. For each G_k, let r be the left-most vertex as shown in Figure 1. We can define the random walk $\{X_t\}_{t\in\mathbb{Z}}$ on each G_k where $X_t = r$ for $t \leq 0$ and for $t > 0$, X_t is the standard rightward random walk along the graph of G_k where each branch is taken independently with probability 1/2. Once X_t hits the right-most vertex (at $t = 6^n$), it stays there forever.

We have the following proposition.

Proposition 4.6. *Let G_n be the Laakso graphs of stage n and let X_t be the random walk on G_n as described above. Then there exists some constant $C > 0$ depending only on p so that*

$$\sum_{k=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d\left(X_t, \widetilde{X}_t(t-2^k)\right)^p\right]}{2^{kp}} \geq Cn \sum_{t\in\mathbb{Z}} \mathbb{E}\left[d(X_t, X_{t-1})^p\right].$$

The proof is similar to the proof of Proposition 4.1 although it does require a little more work. The reader can either attempt to prove it as an exercise or consult Proposition 3.1 of [13].

Analogous to Theorem 4.2, we get the following distortion bounds for embeddings of G_n into Markov p-convex metric spaces:

Theorem 4.7. *Let $p > 1$ and suppose (X, d) is a metric space that is Markov p-convex. Then there exists some $C > 0$ depending only on X so that $c_X(G_n) \geq C(\log|G_n|)^{1/p}$.*

REFERENCES

[1] K. Ball, *The Ribe programme*, Séminaire Bourbaki, 2012. exposé 1047.

[2] K. Ball, E. A. Carlen, and E. H. Lieb, *Sharp uniform convexity and smoothness inequalities for trace norms*, Invent. Math. **115** (1994), no. 3, 463-482.
[3] F. Baudier, *Metrical characterization of super-reflexivity and linear type of Banach spaces*, Archiv Math. **89** (2007), 419-429.
[4] J. Bourgain, *On Lipschitz embedding of finite metric spaces in Hilbert spaces*, Israel J. Math **52** (1985), no. 1-2, 46-52.
[5] ______, *The metrical interpretation of superreflexivity in Banach spaces*, Israel J. Math **56** (1986), no. 2, 222-230.
[6] P. Enflo, *On the nonexistence of uniform homeomorphisms between L_p-spaces*, Ark. Mat. **8** (1970), no. 2, 103-105.
[7] M. I. Kadets, *Proof of the topological equivalence of all separable infinite-dimensional Banach spaces*, Funct. Anal. Appl. **1** (1967), no. 1, 53-62.
[8] T. Laakso, *Plane with A_∞-weighted metric not biLipschitz embeddable to $\mathbb{R}^n$*, Bull. London Math. Soc. **34** (2002), no. 6, 667-676.
[9] N. Linial, E. London, and Y. Rabinovich, *The geometry of graphs and some of its algorithmic applications*, Combinatorica **15** (1995), 215-245.
[10] J. Lee, A. Naor, and Y. Peres, *Trees and Markov convexity*, Geom. Funct. Anal. **18** (2009), no. 5, 1609-1659.
[11] J. Matoušek, *On embedding trees into uniformly convex Banach spaces*, Israel J. Math. **114** (1999), 221-237.
[12] S. Mazur and S. Ulam, *Sur les transformations isométriques d'espaces vectoriels normés*, C. R. Math. Acad. Sci. Paris **194** (1932), 946-948.
[13] M. Mendel and A. Naor, *Markov convexity and local rigidity of distorted metrics*, J. Eur. Math. Soc. **15** (2013), no. 1, 287-337.
[14] A. Naor, *An introduction to the Ribe program*, Jpn. J. Math. **7** (2012), no. 2, 167-233.
[15] G. Pisier, *Martingales with values in uniformly convex spaces*, Israel J. Math. (1975), 326-350.
[16] ______, *Probabilistic methods in the geometry of Banach spaces*, Probability and analysis (Varenna, 1985), Lecture Notes in Math., Springer, 1986, pp. 167-241.
[17] M. Ribe, *On uniformly homeomorphic normed spaces*, Ark. Mat. **14** (1976), no. 2, 237-244.
[18] D. P. Williamson and D. B. Shmoys, *The design of approximation algorithms*, Cambridge University Press, 2011.