AN INTRODUCTION TO THE RIBE PROGRAM

ASSAF NAOR

Contents

Date: May 7, 2012.

Joram Lindenstrauss, who was my Ph.D. advisor, passed away on April 29, 2012. He was an enormously influential mathematician and the founder of the field of research that is surveyed here. This article is dedicated to his memory.

1. Introduction

A 1932 theorem of Mazur and Ulam [123] asserts that if X and Y are Banach spaces and $f : X \to Y$ is an *onto* isometry then f must be an affine mapping. The assumption that $f(X) = Y$ is needed here, as exhibited by, say, the mapping $t \mapsto (t, \sin t)$ from $\mathbb{R}$ to $(\mathbb{R}^2, \|\cdot\|_\infty)$. However, a major strengthening of the Mazur-Ulam theorem due to Figiel [60] asserts that if $f : X \to Y$ is an isometry and $f(0) = 0$ then there is a unique linear operator $T : \overline{\mathrm{span}}(f(X)) \to X$ such that $\|T\| = 1$ and $T(f(x)) = x$ for every $x \in X$. Thus, when viewed as metric spaces in the isometric category, Banach spaces are highly rigid: their linear structure is completely preserved under isometries, and, in fact, isometries between Banach spaces are themselves rigid.

At the opposite extreme to isometries, the richness of Banach spaces collapses if one removes all quantitative considerations by treating them as topological spaces. Specifically, answering a question posed in 1928 by Fréchet [62] and again in 1932 by Banach [18], Kadec [92, 93] proved that any two separable infinite dimensional Banach spaces are homeomorphic. See [91, 28, 27, 4] for more information on this topic, as well as its treatment in the monographs [29, 55]. An extension of the Kadec theorem to non-separable spaces was obtained by Toruńczyk in [183].

If one only considers homeomorphisms between Banach spaces that are "quantitatively continuous" rather than just continuous, then one recovers a rich and subtle category that exhibits deep rigidity results but does not coincide with the linear theory of Banach spaces. We will explain how this suggests that, despite having no a priori link to Banach spaces, general metric spaces have a hidden structure. Using this point of view, insights from Banach space theory can be harnessed to solve problems in seemingly unrelated disciplines, including group theory, algorithms, data structures, Riemannian geometry, harmonic analysis and probability theory. The purpose of this article is to describe a research program that aims to expose this hidden structure of metric spaces, while highlighting some achievements that were obtained over the past five decades as well as challenging problems that remain open.

In order to make the previous paragraph precise one needs to define the concept of a quantitatively continuous homeomorphisms. While there are several meaningful and nonequivalent ways to do this, we focus here on *uniform homeomorphisms*. Given two metric spaces $(\mathcal{M}, d_\mathcal{M})$ and $(\mathcal{N}, d_\mathcal{N})$, a bijection $f : \mathcal{M} \to \mathcal{N}$ is called a uniform homeomorphism if both f and f^{-1} are uniformly continuous, or equivalently if there exist nondecreasing functions $\alpha, \beta : [0, \infty) \to (0, \infty]$ with

$\lim_{t\to 0}\beta(t)=0$ such that $\alpha(d_{\mathcal{M}}(a,b))\leqslant d_{\mathcal{N}}(f(a),f(b))\leqslant\beta(d_{\mathcal{M}}(a,b))$ for all distinct $a,b\in\mathcal{M}$.

In the seminal 1964 paper [107] Lindenstrauss proved that, in contrast to the Kadec theorem, there exist many pairs of separable infinite dimensional Banach spaces, including $L_p(\mu)$ and $L_q(\nu)$ if $p\neq q$ and $\max\{p,q\}\geqslant 2$, that are not uniformly homeomorphic. Henkin proved in [78] that if $n\geqslant 2$ then $C^k([0,1]^n)$ is not uniformly homeomorphic to $C^1([0,1])$ for all $k\in\mathbb{N}$ (this result was previously announced by Grothendieck [71] with some indication of a proof). Important work of Enflo [48, 49, 50], which was partly motivated by his profound investigation of Hilbert's fifth problem in infinite dimensions, obtained additional results along these lines. In particular, in [49] Enflo completed Lindenstrauss' work [107] by proving that that $L_p(\mu)$ and $L_q(\nu)$ are not uniformly homeomorphic if $p\neq q$ and $p,q\in[1,2]$, and in [50] he proved that a Banach space $(X,\|\cdot\|_X)$ which is uniformly homeomorphic to a Hilbert space $(H,\|\cdot\|_H)$ must be isomorphic to H, i.e., there exists a bounded *linear* operator $T:X\to H$ such that $\|Tx\|_H\geqslant\|x\|_X$ for all $x\in X$. A later deep theorem of Johnson, Lindenstrauss and Schechthman [89] makes the same assertion with Hilbert space replaced by ℓ_p, $p\in(0,\infty)$, i.e., any Banach space that is uniformly homeomorphic to ℓ_p must be isomorphic to ℓ_p. At the same time, as shown by Aharoni and Lindenstrauss [1] and Ribe [171], there exist pairs of uniformly homeomorphic Banach spaces that are not isomorphic.

In 1976 Martin Ribe proved [169] that if two Banach spaces are uniformly homeomorphic then they have the same finite dimensional subspaces. To make this statement precise, recall James' [85] notion of (crude) finite representability: a Banach space $(X,\|\cdot\|_X)$ is said to be *finitely representable* in a Banach space $(Y,\|\cdot\|_Y)$ if there exists $K\in[1,\infty)$ such that for every finite dimensional linear subspace $F\subseteq X$ there exists a linear operator $T:F\to Y$ satisfying $\|x\|_X\leqslant\|Tx\|_Y\leqslant K\|x\|_X$ for all $x\in F$. For example, for all $p\in[1,\infty]$ any $L_p(\mu)$ space is finitely representable in ℓ_p, and the classical Dvoretzky theorem [47] asserts that Hilbert space is finitely representable in any infinite dimensional Banach space. If $p,q\in[1,\infty]$ and $p\neq q$ then at least one of the spaces $L_p(\mu)$, $L_q(\nu)$ is not finitely representable in the other; see, e.g. [190].

Theorem 1.1 (Ribe's rigidity theorem [170])**.** *If X and Y are uniformly homeomorphic Banach spaces then X is finitely representable in Y and Y is finitely representable in X.*

Influential alternative proofs of Ribe's theorem were obtained by Heinrich and Mankiewicz [77] and Bourgain [33]. See also the treatment in the surveys [52, 25] and Chapter 10 of the book [26]. In [170] Ribe obtained a stronger version of Theorem 1.1 under additional geometric assumptions on the spaces X and Y. The converse to Ribe's theorem fails, since for $p \in [1,\infty) \smallsetminus \{2\}$ the spaces $L_p(\mathbb{R})$ and ℓ_p are finitely representable in each other but not uniformly homeomorphic: for $p = 1$ this was proved by Enflo [25]; for $p \in (1,2)$ this was proved by Bourgain [33]; for $p \in (2,\infty)$ this was proved by Gorelik [65].

Theorem 1.1 (informally) says that isomorphic finite dimensional linear properties of Banach spaces are preserved under uniform homeomorphisms, and are thus in essence "metric properties". For concreteness, suppose that X satisfies the following property: for every $n \in \mathbb{N}$ and every $x_1,\ldots,x_n \in X$ the average of $\|\pm x_1 \pm x_2 \pm \ldots \pm x_n\|_X^2$ over all the 2^n possible choices of signs is at most $K(\|x_1\|_X^2 + \ldots + \|x_n\|_X^2)$, where $K \in (0,\infty)$ may depend on the geometry of X but not on n and $x_1,\ldots,x_n$. Ribe's theorem asserts that if Y is uniformly homeomorphic to X then it also has the same property. Rather than giving a formal definition, the reader should keep properties of this type in mind: they are "finite dimensional linear properties" since they are given by inequalities between lengths of linear combinations of finitely many vectors, and they are "isomorphic" in the sense that they are insensitive to a loss of a constant factor. Ribe's theorem is thus a remarkable rigidity statement, asserting that uniform homeomorphisms between Banach spaces cannot alter their finite dimensional structure.

Ribe's theorem indicates that in principle any isomorphic finite dimensional linear property of Banach spaces can be equivalently formulated using only distances between points and making no reference whatsoever to the linear structure. Recent work of Ostrovskii [156, 157, 159] can be viewed as making this statement formal in a certain abstract sense. The *Ribe program*, as formulated by Bourgain in 1985 (see [31] and mainly [32]), aims to explicitly study this phenomenon. If parts of the finite dimensional linear theory of Banach spaces are in fact a "nonlinear theory in disguise" then if one could understand how to formulate them using only the metric structure this would make it possible to study them in the context of general metric spaces. As a first step in the Ribe program one would want to discover metric reformulations of key concepts of Banach space theory. Bourgain's famous metric characterization of when a Banach space admits an equivalent uniformly convex norm [32] was the first successful completion of a step in this plan. By doing so, Bourgain kick-started the Ribe program, and this was quickly followed by efforts of several researchers leading to satisfactory progress on key steps of the Ribe program.

The Ribe program does not limit itself to reformulating aspects of Banach space theory using only metric terms. Indeed, this should be viewed as only a first (usually highly nontrivial) step. Once this is achieved, one has an explicit "dictionary" that translates concepts that a priori made sense only in the presence of linear structure to the language of general metric spaces. The next important step in the Ribe program is to investigate the extent to which Banach space phenomena, after translation using the new "dictionary", can be proved for general metric spaces. Remarkably, over the past decades it turned out that this approach is very successful, and it uncovers structural properties of metric spaces that have major impact on areas which do not have any a priori link to Banach space theory. Examples of such successes of the Ribe program will be described throughout this article.

A further step in the Ribe program is to investigate the role of the metric reformulations of Banach space concepts, as provided by the first step of the Ribe program, in metric space geometry. This step is not limited to metric analogues of Banach space phenomena, but rather it aims to use the new "dictionary" to solve problems that are inherently nonlinear (examples include the use of nonlinear type in group theory; see Section 9.4). Moreover, given the realization that insights from Banach space theory often have metric analogues, the Ribe program aims to uncover metric phenomena that mirror Banach space phenomena but are not strictly speaking based on metric reformulations of isomorphic finite dimensional linear properties. For example, Bourgain's embedding theorem was discovered due to the investigation of a question raised by Johnson and Lindenstrauss [88] on a metric analogue of John's theorem [87]. Another example is the investigation, as initiated by Bourgain, Figiel and Milman [34], of nonlinear versions of Dvoretzky's theorem [47] (in this context Milman also asked for a nonlinear version of his Quotient of Subspace Theorem [140], a question that is studied in [125]). Both of the examples above led to the discovery of theorems on metric spaces that are truly nonlinear and do not have immediate counterparts in Banach space theory (e.g., the appearance of ultrametrics in the context of nonlinear Dvoretzky theory; see Section 8), and they had major impact on areas such as approximation algorithms and data structures. Yet another example is Ball's nonlinear version [15] of Maurey's extension theorem [120], based on nonlinear type and cotype (see Section 4). Such developments include some of the most challenging and influential aspects of the Ribe program. In essence, Ribe's theorem pointed the way to a certain analogy between linear and nonlinear metric spaces. One of the main features of the Ribe program is that this analogy is a source of new meaningful questions in metric geometry that probably would not have been raised if it weren't for the Ribe program.

Remark 1.1. A rigidity theorem asserts that a deformation of a certain object preserves more structure than one might initially expect. In other words, equivalence in a weak category implies the existence of an equivalence in a stronger category. Rigidity theorems are naturally important since they say much about the structure of the stronger category (i.e., that it is rigid). However, the point of view of the Ribe program is that a rigidity theorem opens the door to a new research direction whose goal is to uncover hidden structures in the weaker category: perhaps the rigidity exhibited by the stronger category is actually an indication that concepts and theorems of the stronger category are "shadows" of a theory that has a significantly wider range of applicability? This philosophy has been very successful in the context of the Ribe program, but similar investigations were also initiated in response to rigidity theorems in other disciplines. For example, it follows from the Mostow rigidity theorem [143] that if two closed hyperbolic n-manifolds ($n > 2$) are homotopically equivalent then they are isometric. This suggests that the volume of a hyperbolic manifold may be generalized to a homotopy invariant quantity defined for arbitrary manifolds: an idea that was investigated by Milnor and Thurston [141] and further developed by Gromov [66] (see also Sections 5.34–5.36 and 5.43 in [69]). These investigations led to the notion of *simplicial volume*, a purely topological notion associated to a closed oriented manifold that remarkably coincides with the usual volume in the case of hyperbolic manifolds. This notion is very helpful for studying general continuous maps between hyperbolic manifolds.

Historical note. Despite the fact that it was first formulated by Bourgain, the Ribe program is called this way because it is inspired by Ribe's rigidity theorem. I do not know the exact origin of this name. In [32] Bourgain explains the program and its motivation from Ribe's theorem, describes the basic "dictionary" that relates Banach space concepts to metric space concepts, presents examples of natural steps of the program, raises some open questions, and proves his metric characterization of isomorphic uniform convexity as the first successful completion of a step in the program. Bourgain also writes in [32] that "A detailed exposition of this program will appear in J. Lindenstrauss's forthcoming survey paper [5]." Reference [5] in [32] is cited as "J. Lindenstrauss, *Topics in the geometry of metric spaces*, to appear." Probably referring to the same unpublished survey, in [31] Bourgain also discusses the Ribe program and writes "We refer the reader to the survey of J. Lindenstrauss [4] for a detailed exposition of this theme", where reference [4] of [31] is "J. Lindenstrauss, *Proceedings Missouri Conf., Missouri – Columbia (1984)*, to appear." Unfortunately, Lindenstrauss' paper was never published.

This article is intended to serve as an introduction to the Ribe program, targeted at nonspecialists. Aspects of this research direction have been previously surveyed in [108, 25, 164, 109, 26, 130, 96, 129] and especially in Ball's Bourbaki exposé [16]. While the material surveyed here has some overlap with these paper, we cover a substantial amount of additional topics. We also present sketches of arguments as an indication of the type of challenges that the Ribe program raises, and we describe examples of applications to areas which are far from Banach space theory in order to indicate the versatility of this approach to metric geometry.

Asymptotic notation. Throughout this article we will use the notation $\lesssim, \gtrsim$ to denote the corresponding inequalities up to universal constant factors. We will also denote equivalence up to universal constant factors by $\asymp$, i.e., $A \asymp B$ is the same as $(A \lesssim B) \wedge (A \gtrsim B)$.

Acknowledgements. This article accompanies the 10th Takagi Lectures delivered by the author at RIMS, Kyoto, on May 26 2012. I am grateful to Larry Guth and Manor Mendel for helpful suggestions. The research presented here is supported in part by NSF grant CCF-0832795, BSF grant 2010021, and the Packard Foundation.

2. Metric type

Fix a Banach space $(X, \|\cdot\|_X)$. By the triangle inequality we have $\|\varepsilon_1 x_1 + \ldots + \varepsilon_n x_n\|_X \leqslant \|x_1\|_X + \ldots + \|x_n\|_X$ for every $x_1, \ldots, x_n \in X$ and every $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$. By averaging this inequality over all possible choices of signs $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$ we obtain the following *randomized triangle inequality*.

$$\frac{1}{2^n} \sum_{\varepsilon_1,\ldots,\varepsilon_n \in \{-1,1\}} \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|_X \leqslant \sum_{i=1}^n \|x_i\|_X. \tag{1}$$

For $p \geqslant 1$, the Banach space X is said to have Rademacher type p if there exists a constant $T \in (0, \infty)$ such that for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in X$ we have

$$\frac{1}{2^n} \sum_{\varepsilon_1,\ldots,\varepsilon_n \in \{-1,1\}} \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|_X \leqslant T \left(\sum_{i=1}^n \|x_i\|_X^p \right)^{1/p}. \tag{2}$$

It is immediate to check (from the case of collinear $x_1,\ldots,x_n$) that if (2) holds then necessarily $p \leqslant 2$. If $p>1$ and (2) holds then X is said to have nontrivial type. Note that if this happens then in most cases, e.g. if $x_1,\ldots,x_n$ are all unit vectors, (2) constitutes an asymptotic improvement of the triangle inequality (1). For concreteness, we recall that $L_p(\mu)$ has Rademacher type $\min\{p,2\}$.

Remark 2.1. A classical inequality of Kahane [94] asserts that for every $q \geqslant 1$ we have

$$\left(\frac{1}{2^n}\sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{-1,1\}}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|_X^q\right)^{1/q} \leqslant \frac{c(q)}{2^n}\sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{-1,1\}}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|_X,$$

where $c(q)\in(0,\infty)$ depends on q but not on n, the choice of vectors $x_1,\ldots,x_n\in X$, and the Banach space X itself. Therefore the property (2) is equivalent to the requirement

$$\left(\frac{1}{2^n}\sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{-1,1\}}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|_X^q\right)^{1/q} \leqslant T\left(\sum_{i=1}^n \|x_i\|_X^p\right)^{1/p}, \qquad (3)$$

with perhaps a different constant $T\in(0,\infty)$.

The improved triangle inequality (2) is of profound importance to the study of geometric and analytic questions in Banach space theory and harmonic analysis; see [121] and the references therein for more information on this topic.

The Ribe theorem implies that the property of having type p is preserved under uniform homeomorphism of Banach spaces. According to the philosophy of the Ribe program, the next goal is to reformulate this property while using only distances between points and making no reference whatsoever to the linear structure of X. We shall now explain the ideas behind the known results on this step of the Ribe program as an illustrative example of the geometric and analytic challenges that arise when one endeavors to address such questions.

2.1. **Type for metric spaces.** The basic idea, due to Enflo [49], and later to Gromov [67] and Bourgain, Milman and Wolfson [35], is as follows. Given $x_1,\ldots,x_n\in X$ define $f:\{-1,1\}^n\to X$ by

$$\forall\, \varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)\in\{-1,1\}^n,\quad f(\varepsilon)=\sum_{i=1}^n \varepsilon_i x_i. \qquad (4)$$

With this notation, the definition of Rademacher type appearing in (2) is the same as the inequality

$$\mathbb{E}_\varepsilon\left[\|f(\varepsilon)-f(-\varepsilon)\|_X\right] \leqslant T\left(\sum_{i=1}^n \mathbb{E}_\varepsilon\left[\|f(\varepsilon)-f(\varepsilon_1,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)\|_X^p\right]\right)^{1/p}, \quad (5)$$

where $\mathbb{E}[\cdot]$ denotes expectation with respect to a uniformly random choice of $\varepsilon \in \{-1,1\}^n$.

Inequality (5) seems to involve only distances between points, except for the crucial fact that the function f itself is the *linear* function appearing in (4). Enflo's (bold) idea [49] (building on his earlier work [49, 50]) is to drop the linearity requirement of f and to demand that (5) holds for *all* functions $f:\{-1,1\}^n \to X$. Thus, for $p \geqslant 1$ we say that a *metric space* $(\mathcal{M},d_\mathcal{M})$ has type p if there exists a constant $T\in(0,\infty)$ such that for every $n\in\mathbb{N}$ and every $f:\{-1,1\}^n\to\mathcal{M}$,

$$\mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(-\varepsilon)\right)\right] \leqslant T\left(\sum_{i=1}^n \mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(\varepsilon_1,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)\right)^p\right]\right)^{1/p}. \quad (6)$$

Remark 2.2. The above definition of type of a metric space is ad hoc: it was chosen here for the sake of simplicity of exposition. While this definition is sufficient for the description of the key ideas and it is also strong enough for the ensuing geometric applications, it differs from the standard definitions of type for metric spaces that appear in the literature. Specifically, motivated by the fact that Rademacher type p for a Banach space $(X,\|\cdot\|_X)$ is equivalent to (3) for any $q\geqslant 1$, combining the above reasoning with the case $q=p$ in (3) leads to Enflo's original definition: say that a metric space $(\mathcal{M},d_\mathcal{M})$ has *Enflo type* p if if there exists a constant $T\in(0,\infty)$ such that for every $n\in\mathbb{N}$ and every $f:\{-1,1\}^n\to\mathcal{M}$,

$$\mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(-\varepsilon)\right)^p\right] \leqslant T^p\sum_{i=1}^n \mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(\varepsilon_1,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)\right)^p\right]. \quad (7)$$

Analogously, by (3) with $q=2$ and Hölder's inequality, if $(X,\|\cdot\|_X)$ has Rademacher type $p\in[1,2]$ then there exists a constant $T\in(0,\infty)$

such that for every $n\in\mathbb{N}$ and every $x_1,\ldots,x_n\in X$,

$$\mathbb{E}_\varepsilon\left[\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|_X^2\right]\leqslant T^2 n^{\frac{2}{p}-1}\sum_{i=1}^n \|x_i\|_i^2.$$

Hence, following the above reasoning, Bourgain, Milman and Wolfson [35] suggested the following definition of type of metric spaces, which is more convenient than Enflo type for certain purposes: say that a metric space $(\mathcal{M},d_\mathcal{M})$ has *BMW type* p if if there exists a constant $T\in(0,\infty)$ such that for every $n\in\mathbb{N}$ and every $f:\{-1,1\}^n\to\mathcal{M}$,

$$\begin{aligned}&\mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(-\varepsilon)\right)^2\right]\\ &\leqslant T^2 n^{\frac{2}{p}-1}\sum_{i=1}^n \mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(\varepsilon_1,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)\right)^2\right].\end{aligned}\quad (8)$$

In [67] Gromov suggested the above definitions of type of metric spaces, but only when $p=2$, in which case (7) and (8) coincide.

Remark 2.3. For the same reason that Rademacher type $p>1$ should be viewed as an improved (randomized) triangle inequality, i.e., an improvement over (1), the above definitions of type of metric spaces should also be viewed as an improved triangle inequality. Indeed, it is straightforward to check that every metric space $(\mathcal{M},d_\mathcal{M})$ satisfies

$$\begin{aligned}&\mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(-\varepsilon)\right)\right]\\ &\qquad\leqslant\sum_{i=1}^n \mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(\varepsilon_1,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)\right)\right]\end{aligned}\quad (9)$$

for every $n\in\mathbb{N}$ and every $f:\{-1,1\}^n\to\mathcal{M}$. Thus every metric space has type 1 (equivalently Enflo type 1) with $T=1$. A similar application of the triangle inequality shows that every metric space has BMW type 1 with $T=1$. Our definition (6) of type of a metric space $(\mathcal{M},d_\mathcal{M})$ is not formally stronger than (9), and with this in mind one might prefer to consider the following variant of (6):

$$\begin{aligned}&\mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon),f(-\varepsilon)\right)\right]\\ &\lesssim\mathbb{E}_\varepsilon\left[\left(\sum_{i=1}^n d_\mathcal{M}\left(f(\varepsilon),f(\varepsilon_1,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)\right)^p\right)^{1/p}\right].\end{aligned}\quad (10)$$

Note that, by Jensen's inequality, (10) implies (6). We chose to work with the definition appearing in (6) only for simplicity of notation and exposition; the argument below will actually yield (10).

2.2. **The geometric puzzle.** One would be justified to be concerned about the "leap of faith" that was performed in Section 2.1. Indeed, if a Banach space $(X, \|\cdot\|_X)$ satisfies (5) for all linear functions as in (4) there is no reason to expect that it actually satisfies (5) for all $f : \{-1,1\}^n \to X$ whatsoever. Thus, for the discussion in Section 2.1 to be most meaningful one needs to prove that if a Banach space has Rademacher type p then it also has type p as a metric space (resp. Enflo type p or BMW type p). This question, posed in 1976 by Enflo [52] (for the case of Enflo type), remains open.

Question 1 (Enflo's problem)**.** *Is it true that if a Banach space has Rademacher type p then it also has Enflo type p?*

We will present below an argument that leads to the following slightly weaker fact: if a Banach space has Rademacher type p then for every $\varepsilon \in (0,1)$ it also has type $p-\varepsilon$ as a metric space. We will follow an elegant argument of Pisier [164], who almost solved Enflo's problem by showing that if a Banach space has Rademacher type p then it also has Enflo type $p-\varepsilon$ for every $\varepsilon \in (0,1)$. Earlier, and via a different argument, Bourgain, Milman and Wolfson proved [35] that if a Banach space has Rademacher type p then it also has BMW type $p-\varepsilon$ for every $\varepsilon \in (0,1)$. More recently, [128] gave a different, more complicated (and less useful), definition of type of a metric space, called *scaled Enflo type*, and showed that a Banach space has Rademacher type p if and only if it has scaled Enflo type p. This completes the Ribe program for Rademacher type, but it leaves much to be understood, as we conjecture that the answer to Question 1 is positive. In [150, 99, 149, 83] it is proved that the answer to Question 1 is positive for certain classes of Banach spaces (including all $L_p(\mu)$ spaces).

To better understand the geometric meaning of the above problems and results consider the following alternative description of the definition of type of a metric space $(\mathcal{M}, d_\mathcal{M})$. Call a subset of 2^n points in $\mathcal{M}$ that is indexed by $\{-1,1\}^n$ a *geometric cube* in $\mathcal{M}$. A diagonal of the geometric cube $\{x_\varepsilon\}_{\varepsilon\in\{-1,1\}^n} \subseteq \mathcal{M}$ is a pair $\{x_\varepsilon, x_\delta\}$ where $\varepsilon, \delta \in \{-1,1\}^n$ differ in all the coordinates (equiv. $\delta = -\varepsilon$). An edge of this geometric cube is a pair $\{x_\varepsilon, x_\delta\}$ where $\varepsilon, \delta \in \{-1,1\}^n$ differ in exactly one coordinate. Then (6) is the following statement

$$\frac{\sum diagonal}{2^n} \leqslant T\left(\frac{\sum edge^p}{2^n}\right)^{1/p}, \tag{11}$$

where in the left hand side of (11) we have the sum of the lengths of all the diagonals of the geometric cube, and in the right hand side of (11) we have the sum of the pth power of the lengths of all the edges of the geometric cube. The assertion that $(\mathcal{M}, d_{\mathcal{M}})$ has type p means that (11) holds for *all* geometric cubes in $\mathcal{M}$.

If $(X, \|\cdot\|_X)$ is a Banach space with Rademacher type p then we know that (11) holds true for all *parallelepipeds* in X, as depicted in Figure 1.

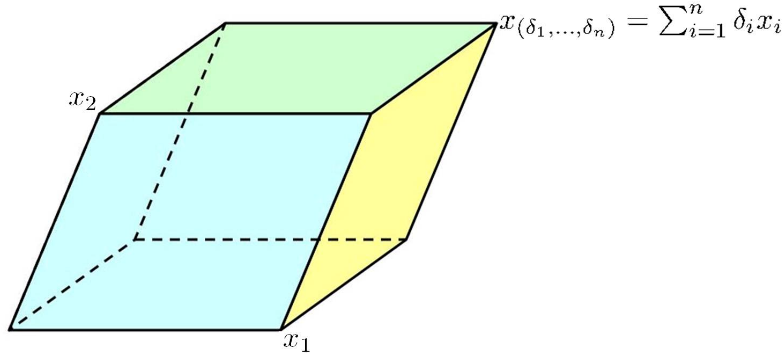

FIGURE 1. X having Rademacher type p is equivalent to the requirement that (11) holds true for every parallelepiped in X, i.e., a set of vectors $\{x_\delta\}_{\delta\in\{0,1\}^n}$ where for some $x_1,\ldots,x_n \in X$ we have $x_\delta = \sum_{i=1}^n \delta_i x_i$ for all $\delta = (\delta_1,\ldots,\delta_n) \in \{0,1\}^n$.

The geometric "puzzle" is therefore to deduce the validity of (11) for all geometric cubes in X (perhaps with p replaced by $p-\varepsilon$) from the assumption that it holds for all parallelepipeds. In other words, given $x_1,\ldots,x_{2^n} \in X$, index these points arbitrarily by $\{-1,1\}^n$. Once this is done, some pairs of these points have been declared as diagonals, and other pairs have been declared as edges, in which case (11) has to hold true for these pairs; see Figure 2.

2.3. **Pisier's argument.** Our goal here is to describe an approach, devised by Pisier in 1986, to deduce metric type from Rademacher type. Before doing so we recall some basic facts related to vector-valued Fourier analysis on $\{-1,1\}^n$. The characters of the group $\{-1,1\}^n$ (equipped with coordinate-wise multiplication) are the Walsh functions $\{W_A\}_{A\subseteq\{1,\ldots,n\}}$, where $W_A(\varepsilon) = \prod_{i\in A}\varepsilon_i$. Fix a Banach space $(X, \|\cdot\|_X)$. Any function $f : \{-1,1\}^n \to X$ has the Fourier expansion

$$f(\varepsilon) = \sum_{A\subseteq\{1,\ldots,n\}} \widehat{f}(A) W_A(\varepsilon),$$

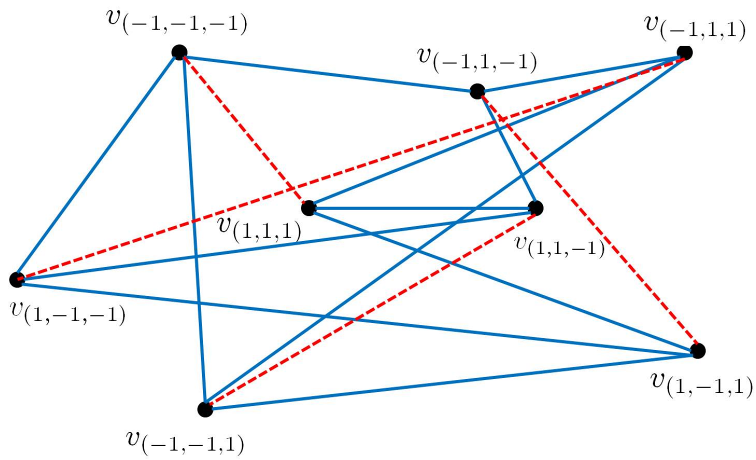

FIGURE 2. A schematic illustration of the problem when $n = 3$. Given $x_1, \ldots, x_8 \in X$, we index them using the labels $\{(\varepsilon_1, \varepsilon_2, \varepsilon_3) : \varepsilon_1, \varepsilon_2, \varepsilon_3 \in \{-1, 1\}\}$ as depicted above. Once this is done, the dotted lines represent diagonals and the full lines represent edges.

where

$$\widehat{f}(A) = \mathbb{E}_\varepsilon\left[f(\varepsilon)W_A(\varepsilon)\right] = \frac{1}{2^n}\sum_{\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)\in\{-1,1\}^n} f(\varepsilon)\prod_{i\in A}\varepsilon_i \in X.$$

For $j \in \{1, \ldots, n\}$ define $\partial_j f : \{-1,1\}^n \to X$ by

$$\begin{aligned}\partial_j f(\varepsilon) &= \frac{f(\varepsilon) - f(\varepsilon_1, \ldots, \varepsilon_{j-1}, -\varepsilon_j, \varepsilon_{j+1}, \ldots, \varepsilon_n)}{2} \\ &= \sum_{\substack{A\subseteq\{1,\ldots,n\}\\ j\in A}} \widehat{f}(A)W_A(\varepsilon). \qquad (12)\end{aligned}$$

The hypercube Laplacian of f is given by

$$\Delta f(\varepsilon) = \sum_{j=1}^n \partial_j f(\varepsilon) = \sum_{A\subseteq\{1,\ldots,n\}} |A|\widehat{f}(A)W_A(\varepsilon).$$

The associated time-t evolute of f under the heat semigroup is

$$e^{-t\Delta}f(\varepsilon) = \sum_{A\subseteq\{1,\ldots,n\}} e^{-t|A|}\widehat{f}(A)W_A(\varepsilon). \qquad (13)$$

Since the operator $e^{-t\Delta}$ coincides with convolution with the Riesz kernel $R_t(\varepsilon) = \prod_{i=1}^n(1 + e^{-t}\varepsilon_i)$, which for $t \geqslant 0$ is the density of a probability measure on $\{-1,1\}^n$, we have by convexity

$$t \geqslant 0 \implies \mathbb{E}_\varepsilon \left[\left\|e^{-t\Delta} f(\varepsilon)\right\|_X\right] \leqslant \mathbb{E}_\varepsilon \left[\|f(\varepsilon)\|_X\right]. \tag{14}$$

It immediately follows from (13) that

$$e^{-t\Delta}\left(W_{\{1,\ldots,n\}} e^{-t\Delta} f\right) = e^{-tn} W_{\{1,\ldots,n\}} f. \tag{15}$$

Consequently, we deduce from (15) and (14) that

$$t \geqslant 0 \implies \mathbb{E}_\varepsilon \left[\left\|e^{-t\Delta} f(\varepsilon)\right\|_X\right] \geqslant e^{-nt} \mathbb{E}_\varepsilon \left[\|f(\varepsilon)\|_X\right]. \tag{16}$$

Fix $s>0$ that will be determined later. Let $g_s^* : \{-1,1\}^n \to X^*$ be a normalizing functional of $e^{-s\Delta} f - \widehat{f}(\emptyset) \in L_1(\{-1,1\}^n, X)$, i.e.,

$$\forall\, \varepsilon \in \{-1,1\}^n, \quad \|g_s^*(\varepsilon)\|_{X^*} \leqslant 1, \tag{17}$$

and

$$\begin{aligned}\mathbb{E}_\varepsilon \left[\left\|e^{-s\Delta}\left(f(\varepsilon) - \widehat{f}(\emptyset)\right)\right\|_X\right] &= \mathbb{E}_\varepsilon \left[g_s^*(\varepsilon)\left(e^{-s\Delta}\left(f(\varepsilon) - \widehat{f}(\emptyset)\right)\right)\right] \\ &= \sum_{\substack{A \subseteq \{1,\ldots,n\} \\ A \neq \emptyset}} e^{-s|A|} \widehat{g_s^*}(A)\left(\widehat{f}(A)\right). \end{aligned} \tag{18}$$

In [164], Pisier succeeds to relate general geometric cubes in X to parallelepipeds in X by interpolating g_s^* between two hypercubes. Specifically, for every $t \geqslant 0$ consider the function

$$(g_s^*)_t : \{-1,1\}^n \times \{-1,1\}^n \to X^*$$

given by

$$(g_s^*)_t(\varepsilon,\delta) = \sum_{A \subseteq \{1,\ldots,n\}} \widehat{g_s^*}(A) \prod_{i \in A}\left(e^{-t}\varepsilon_i + (1-e^{-t})\delta_i\right). \tag{19}$$

Equivalently, $(g_s^*)_t(\varepsilon,\delta) = g_s^*\left(e^{-t}\varepsilon + (1-e^{-t})\delta\right)$, where we interpret the substitution of the vector $e^{-t}\varepsilon + (1-e^{-t})\delta \in \mathbb{R}^n$ into the function g_s^*, which is defined a priori only on $\{-1,1\}^n$, by formally substituting this vector into the Fourier expansion of g_s^*.

Yet another way to interpret $(g_s^*)_t(\varepsilon,\delta)$ is to note that for every $A \subseteq \{1,\ldots,n\}$,

$$\begin{aligned}\prod_{i\in A}\left(e^{-t}\varepsilon_i + (1-e^{-t})\delta_i\right) &= W_A(\varepsilon)\prod_{i=1}^n\left(e^{-t} + (1-e^{-t})(\varepsilon_i\delta_i)^{\mathbf{1}_A(i)}\right) \\ &= W_A(\varepsilon)\sum_{B\subseteq\{1,\ldots,n\}} e^{-t|B|}(1-e^{-t})^{n-|B|} W_{A\smallsetminus B}(\varepsilon\delta) \\ &= \sum_{B\subseteq\{1,\ldots,n\}} e^{-t|B|}(1-e^{-t})^{n-|B|} W_{A\cap B}(\varepsilon) W_{A\smallsetminus B}(\delta). \end{aligned} \tag{20}$$

Hence, by substituting (20) into (19) we have

$$\begin{aligned}&(g_s^*)_t(\varepsilon,\delta)\\ &= \sum_{B\subseteq\{1,\ldots,n\}} e^{-t|B|}(1-e^{-t})^{n-|B|} g_s^*\left(\sum_{i\in B}\varepsilon_i e_i + \sum_{i\in\{1,\ldots,n\}\smallsetminus B}\delta_i e_i\right), \quad (21)\end{aligned}$$

where $e_1,\ldots,e_n$ is the standard basis of $\mathbb{R}^n$. In particular, it follows from (17) and (21) that for every $\varepsilon,\delta\in\{-1,1\}^n$,

$$\|(g_s^*)_t(\varepsilon,\delta)\|_{X^*} \leqslant \sum_{k=1}^n \binom{n}{k} e^{-kt}(1-e^{-t})^{n-k} = 1. \quad (22)$$

By directly expanding the products in (19) and collecting the terms that are linear in the variables $(\delta_1,\ldots,\delta_n)$, we see that

$$\begin{aligned}&(g_s^*)_t(\varepsilon,\delta)\\ &= (e^t-1)\sum_{i=1}^n \delta_i \sum_{\substack{A\subseteq\{1,\ldots,n\}\\ i\in A}} e^{-|A|t}\widehat{g_s^*}(A) W_{A\smallsetminus\{i\}}(\varepsilon) + \Phi_{s,t}^*(\varepsilon,\delta), \quad (23)\end{aligned}$$

where the error term $\Phi_{s,t}^*(\varepsilon,\delta)\in X^*$ satisfies

$$\mathbb{E}_\delta\left[\Phi_{s,t}^*(\varepsilon,\delta)\left(\sum_{i=1}^n \delta_i x_i\right)\right] = 0 \quad (24)$$

for all $\varepsilon\in\{-1,1\}^n$ and all choices of vectors $x_1,\ldots,x_n\in X$. By substituting $x_i=\varepsilon_i\partial_i f(\varepsilon)$ into (24), and recalling (12), we deduce from (23) that

$$\begin{aligned}&\mathbb{E}_\varepsilon\mathbb{E}_\delta\left[(g_s^*)_t(\varepsilon,\delta)\left(\sum_{i=1}^n \delta_i\varepsilon_i\partial_i f(\varepsilon)\right)\right]\\ &\quad= (e^t-1)\sum_{i=1}^n \sum_{\substack{A\subseteq\{1,\ldots,n\}\\ i\in A}} e^{-t|A|}\widehat{g_s^*}(A)\left(\widehat{f}(A)\right)\\ &\quad= (e^t-1)\sum_{A\subseteq\{1,\ldots,n\}} |A| e^{-t|A|}\widehat{g_s^*}(A)\left(\widehat{f}(A)\right). \quad (25)\end{aligned}$$

Recalling (22) we see that

$$\mathbb{E}_\varepsilon\mathbb{E}_\delta\left[(g_s^*)_t(\varepsilon,\delta)\left(\sum_{i=1}^n \delta_i\varepsilon_i\partial_i f(\varepsilon)\right)\right] \leqslant \mathbb{E}_\varepsilon\mathbb{E}_\delta\left[\left\|\sum_{i=1}^n \delta_i\partial_i f(\varepsilon)\right\|_X\right]. \quad (26)$$

Hence,

$$\begin{aligned}
\mathbb{E}_\varepsilon\left[\left\|e^{-s\Delta}\left(f(\varepsilon)-\mathbb{E}_\delta\left[f(\delta)\right]\right)\right\|_X\right] &\stackrel{(18)}{=} \sum_{\substack{A\subseteq\{1,\ldots,n\}\\ A\neq\emptyset}} e^{-s|A|}\widehat{g_s^*}(A)\left(\widehat{f}(A)\right)\\
&= \int_s^\infty\left(\sum_{A\subseteq\{1,\ldots,n\}}|A|e^{-t|A|}\widehat{g_s^*}(A)\left(\widehat{f}(A)\right)\right)dt\\
&\stackrel{(25)\wedge(26)}{\leqslant} \left(\int_s^\infty\frac{dt}{e^t-1}\right)\mathbb{E}_\varepsilon\mathbb{E}_\delta\left[\left\|\sum_{i=1}^n\delta_i\partial_i f(\varepsilon)\right\|_X\right]\\
&= \log\left(\frac{e^s}{e^s-1}\right)\mathbb{E}_\varepsilon\mathbb{E}_\delta\left[\left\|\sum_{i=1}^n\delta_i\partial_i f(\varepsilon)\right\|_X\right]. \qquad (27)
\end{aligned}$$

Recalling (16), it follows from (27) that

$$\mathbb{E}_\varepsilon\left[\|f(\varepsilon)-\mathbb{E}_\delta[f(\delta)]\|_X\right] \leqslant e^{ns}\log\left(\frac{e^s}{e^s-1}\right)\mathbb{E}_\varepsilon\mathbb{E}_\delta\left[\left\|\sum_{i=1}^n\delta_i\partial_i f(\varepsilon)\right\|_X\right]. \qquad (28)$$

By choosing $s\asymp\frac{\log\log n}{n\log n}$ so as to minimize the right hand side of (28),

$$\mathbb{E}_\varepsilon\left[\|f(\varepsilon)-\mathbb{E}_\delta[f(\delta)]\|_X\right] \leqslant \left(\log n+O(\log\log n)\right)\mathbb{E}_\varepsilon\mathbb{E}_\delta\left[\left\|\sum_{i=1}^n\delta_i\partial_i f(\varepsilon)\right\|_X\right]. \qquad (29)$$

If X has Rademacher type $p>1$, i.e., it satisfies (2), then

$$\begin{aligned}
\mathbb{E}_\varepsilon\left[\|f(\varepsilon)-f(-\varepsilon)\|_X\right] &\leqslant 2\mathbb{E}_\varepsilon\left[\|f(\varepsilon)-\mathbb{E}_\delta[f(\delta)]\|_X\right]\\
&\lesssim T(\log n)\mathbb{E}_\varepsilon\left[\left(\sum_{i=1}^n\|\partial_i f(\varepsilon)\|_X^p\right)^{1/p}\right]\\
&\lesssim T(\log n)\left(\sum_{i=1}^n\mathbb{E}_\varepsilon\left[\|f(\varepsilon)-f(\varepsilon_1,\ldots,-\varepsilon_i,\ldots,\varepsilon_n)\|_X^p\right]\right)^{1/p}. \qquad (30)
\end{aligned}$$

This proves that if X has Rademacher type p then it almost has type p as a metric space: inequality (6) holds with an additional logarithmic factor. We have therefore managed to deduce the fully metric "diagonal versus edge" inequality (11) from the corresponding inequality for parallelepipeds, though with a (conjecturally) redundant factor of $\log n$. Using similar ideas, for every $\varepsilon\in(0,1)$ one can also deduce the validity of the Enflo type p condition (7) without the $\log n$ term but with p replaced by $p-\varepsilon$ and the implied constant depending on ε. See Pisier's paper [164] for the proof of this alternative tradeoff. A similar tradeoff was previously proved for BMW type using a different method by Bourgain, Milman and Wolfson [35].

Remark 2.4. An inspection of the above argument reveals that there exists a universal constant $C\in(0,\infty)$ such that for every Banach space $(X,\|\cdot\|_X)$, every $q\in[1,\infty]$, every $n\in\mathbb{N}$, and every $f:\{-1,1\}^n\to X$ we have

$$\left(\mathbb{E}_\varepsilon\left[\|f(\varepsilon)-\mathbb{E}_\delta[f(\delta)]\|_X^q\right]\right)^{1/q} \leqslant C(\log n)\left(\mathbb{E}_\varepsilon\mathbb{E}_\delta\left[\left\|\sum_{i=1}^n \delta_i\partial_i f(\varepsilon)\right\|_X^q\right]\right)^{1/q}. \quad (31)$$

Inequality (31) was proved in 1986 by Pisier [164], and is known today as *Pisier's inequality*. Removal of the $\log n$ factor from (31) for Banach spaces with nontrivial Rademacher type would yield a positive solution Enflo's problem (Question 1). Talagrand proved [179] that there exist Banach spaces for which the $\log n$ term in (31) cannot be removed, but we conjecture that if $(X,\|\cdot\|_X)$ has Rademacher type $p>1$ then the $\log n$ term in (31) can be replaced by a constant that may depend on X but not on n. In [179] it was shown that the $\log n$ term in (31) can be replaced by a universal constant if $X=\mathbb{R}$, and in [188] it was shown that this is true for a general Banach space X if $q=\infty$. In [150, 83] it is shown that the $\log n$ term in (31) can be replaced by a constant that is independent of n for certain classes of Banach spaces that include all $L_p(\mu)$ spaces, $p\in(1,\infty)$.

2.4. **Unique obstructions to type.** There is an obvious obstruction preventing a Banach space $(X,\|\cdot\|_X)$ from having any Rademacher type $p>1$: if X contains well-isomorphic copies of $\ell_1^n=(\mathbb{R}^n,\|\cdot\|_1)$ for all $n\in\mathbb{N}$ then its Rademacher type must be trivial. Indeed, assume that $(X,\|\cdot\|_X)$ satisfies (2) and for $n\in\mathbb{N}$ and $D\in(0,\infty)$ suppose that there exists a linear operator $A:\ell_1^n\to X$ satisfying $\|x\|_1\leqslant\|Ax\|_X\leqslant D\|x\|_1$ for all $x\in\ell_1^n$. Letting $\varepsilon_1,\ldots,\varepsilon_n$ be the standard basis of $\mathbb{R}^n$, it follows that for $x_i=Ae_i$ we have $\|x_i\|_X\leqslant D$ and

$$\forall\,\varepsilon\in\{-1,1\}^n,\quad \|\varepsilon_1x_i+\ldots+\varepsilon_nx_n\|_X=\|A(\varepsilon_1e_1+\cdots+\varepsilon_ne_n)\|_X\geqslant n.$$

These facts are in conflict with (2), since they force the constant T appearing in (2) to satisfy

$$T\geqslant\frac{n^{1-\frac1p}}{D}. \quad (32)$$

Pisier proved [161] that the well-embeddability of $\{\ell_1^n\}_{n=1}^\infty$ is the *only obstruction* to nontrivial Rademacher type: a Banach space $(X,\|\cdot\|_X)$ fails to have nontrivial type if and only if for every $\varepsilon\in(0,1)$ and every $n\in\mathbb{N}$ there exists a linear operator $A:\ell_1^n\to X$ satisfying $\|x\|_1\leqslant\|Ax\|_X\leqslant(1+\varepsilon)\|x\|_1$ for all $x\in\ell_1^n$. In other words, once we know that X does not contain isomorphic copies of $\{\ell_1^n\}_{n=1}^\infty$ we immediately deduce that the norm on X must satisfy the asymptotically stronger randomized triangle inequality (2).

As one of the first examples of the applicability of Banach space insights to general metric spaces, Bourgain, Milman and Wolfson [35] proved the only obstruction preventing a *metric space* $(\mathcal{M},d_\mathcal{M})$ from having any BMW type $p>1$ is that $\mathcal{M}$ contains bi-Lipschitz copies of the Hamming cubes $\{(\{-1,1\}^n,\|\cdot\|_1)\}_{n=1}^\infty$.

To make this statement precise it would be useful to recall the following standard notation from bi-Lipschitz embedding theory: given two metric space $(\mathcal{M},d_\mathcal{M})$ and $(\mathcal{N},d_\mathcal{N})$, denote by

$$c_{(\mathcal{M},d_\mathcal{M})}(\mathcal{N},d_\mathcal{N}) \tag{33}$$

(or $c_\mathcal{M}(\mathcal{N})$ if the metrics are clear from the context) the infimum over those $D\in[1,\infty]$ for which there exists $f:\mathcal{N}\to\mathcal{M}$ and a scaling factor $\lambda\in(0,\infty)$ satisfying

$$\forall\, x,y\in\mathcal{N},\quad \lambda d_\mathcal{N}(x,y)\leqslant d_\mathcal{M}(f(x),f(y))\leqslant D\lambda d_\mathcal{N}(x,t).$$

This parameter is called the $\mathcal{M}$ distortion of $\mathcal{N}$. When $\mathcal{M}$ is a Hilbert space, this parameter is called the Euclidean distortion of $\mathcal{N}$.

Suppose that $p>1$ and $(\mathcal{M},d_\mathcal{M})$ satisfies any of the type p inequalities (6), (7) or (8) (i.e., our definition of metric type, Enflo type, or BMW type, respectively). If $c_\mathcal{M}(\{-1,1\}^n,\|\cdot\|_1)<D$ then there exists $f:\{-1,1\}^n\to\mathcal{M}$ and $\lambda>0$ such that

$$\forall\,\varepsilon,\delta\in\{-1,1\}^n,\quad \lambda\|\varepsilon-\delta\|_1\leqslant d_\mathcal{M}(f(\varepsilon),f(\delta))\leqslant D\lambda\|\varepsilon-\delta\|_1.$$

It follows that $d_\mathcal{M}(f(\varepsilon),f(\varepsilon_1,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)\leqslant 2D\lambda$ for all $\varepsilon\in\{-1,1\}^n$ and $i\in\{1,\ldots,n\}$. Also, $d_\mathcal{M}(f(\varepsilon),f(-\varepsilon))\geqslant 2n\lambda$ for all $\varepsilon\in\{-1,1\}^n$. Hence any one of the nonlinear type conditions (6), (7) or (8) implies that

$$c_\mathcal{M}(\{-1,1\}^n,\|\cdot\|_1)\geqslant\frac{n^{1-\frac1p}}{T}. \tag{34}$$

Bourgain, Milman and Wolfson proved [35] (see also the exposition in [164]) that a metric space $(\mathcal{M},d_\mathcal{M})$ fails to satisfy the improved randomized triangle inequality (8) if and only if $c_\mathcal{M}(\{-1,1\}^n,\|\cdot\|_1)=1$ for all $n\in\mathbb{N}$. It is open whether the same “unique obstruction” result holds true for Enflo type as well.

We note in passing that it follows from (30) and (34) that if $(X,\|\cdot\|_X)$ is a normed space with type $p>1$ then

$$c_X(\{-1,1\}^n,\|\cdot\|_1)\gtrsim_X \frac{n^{1-\frac{1}{p}}}{\log n}, \tag{35}$$

where the implied constant may depend on the geometry of X but not on n. In combination with (32), we deduce that $c_X(\ell_1^n)$ and $c_X(\{-1,1\}^n)$ have the same asymptotic order of magnitude, up to a logarithmic term which we conjecture can be removed. This logarithmic term is indeed not needed if X is an $L_p(\mu)$ space, as shown by Enflo [49] for $p\in(1,2]$ and in [150] for $p\in(2,\infty)$ (alternative proofs are given in [99, 149]). It is tempting to guess that $(\{-1,1\}^n,\|\cdot\|_1)$ has (up to constant factors) the largest ℓ_p distortion among all subsets of ℓ_1 of size 2^n. This stronger statement remains a challenging open problem; it has been almost solved (again, up to a logarithmic factor) only for $p=2$ in [6].

3. Metric cotype

The natural "dual" notion to Rademacher type, called *Rademacher cotype*, arises from reversing the inequalities in (2) or (3) (formally, duality is a subtle issue in this context; see [122, 163]). Specifically, say that a Banach space $(X,\|\cdot\|_X)$ has Rademacher cotype $q\in[1,\infty]$ if there exists a constant $C\in(0,\infty)$ such that for every $n\in\mathbb{N}$ and every $x_1,\ldots,x_n\in X$ we have

$$\bigg(\sum_{i=1}^n\|x_i\|_X^q\bigg)^{1/q}\leqslant C\mathbb{E}\left[\bigg\|\sum_{i=1}^n\varepsilon_ix_i\bigg\|_X\right]. \tag{36}$$

It is simple to check that if (36) holds then necessarily $q\in[2,\infty]$, and that every Banach space has Rademacher cotype ∞ (with $C=1$). As in the case of Rademacher type, the notion of Rademacher cotype is of major importance to Banach space theory; e.g. it affects the dimension of almost spherical sections of convex bodies [61]. For more information on the notion of Rademacher cotype (including a historical discussion), see the survey [121] and the references therein.

As explained in Remark 2.1, Kahane's inequality implies that the requirement (36) is equivalent (with a different constant C) to the requirement

$$\sum_{i=1}^n\|x_i\|_X^q\leqslant C^q\mathbb{E}\left[\bigg\|\sum_{i=1}^n\varepsilon_ix_i\bigg\|_X^q\right]. \tag{37}$$

For simplicity of notation we will describe below metric variants of (37), though the discussion carries over mutatis mutandis also to the natural analogues of (36).

In Banach spaces it is meaningful to reverse the inequality in the definition of Rademacher type, but in metric spaces reversing the the inequality in the definition of Enflo type results in a requirement that no metric space can satisfy unless it consists of a single point (the same assertion holds true for our definition of metric type (6) and BMW type (8), but we will only discuss Enflo type from now on). Indeed, assume that a metric space $(\mathcal{M}, d_\mathcal{M})$ satisfies

$$\sum_{i=1}^{n}\mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon), f(\varepsilon_1,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)\right)^q\right] \leqslant C^q\mathbb{E}_\varepsilon\left[d_\mathcal{M}\left(f(\varepsilon), f(-\varepsilon)\right)^q\right]. \quad (38)$$

For all $f:\{-1,1\}^n\to X$. If $\mathcal{M}$ contains two distinct point x_0, y_0 then apply (38) to a function $f:\{-1,1\}^n\to\{x_0,y_0\}$ chosen uniformly at random from the 2^{2^n} possible functions of this type. The right hand side of (38) will always be bounded by $C^q d_\mathcal{M}(x_0,y_0)^q$, while the expectation over the random function f of the left hand side of (38) is $\frac{n}{2}d_\mathcal{M}(x_0,y_0)^q$. Thus necessarily $C\gtrsim n^{1/q}$.

In [130] the following definition of *metric cotype* was introduced. A metric space $(\mathcal{M}, d_\mathcal{M})$ has metric cotype q if there exists a constant $C\in(0,\infty)$ such that for every $n\in\mathbb{N}$ there exists an even integer $m\in\mathbb{N}$ such that every $f:\mathbb{Z}_m^n\to\mathcal{M}$ satisfies

$$\sum_{j=1}^{n}\sum_{x\in\mathbb{Z}_m^n} d_\mathcal{M}\left(f\left(x+\frac{m}{2}e_j\right), f(x)\right)^q \leqslant \frac{(Cm)^q}{3^n}\sum_{\varepsilon\in\{-1,0,1\}^n}\sum_{x\in\mathbb{Z}_m^n} d_\mathcal{M}(f(x+\varepsilon), f(x))^q. \quad (39)$$

Here $e_1,\ldots,e_n$ are the standard basis of the discrete torus $\mathbb{Z}_m^n$ and addition is performed modulo m. The average over $\varepsilon\in\{-1,0,1\}^n$ on the right hand side of (39) is natural here, as it corresponds to the ℓ_∞ edges of the discrete torus.

It turns out that it is possible to complete the step of the Ribe program corresponding to Rademacher cotype via the above definition of metric cotype. Specifically, the following theorem was proved in [130].

Theorem 3.1. *A Banach space $(X,\|\cdot\|_X)$ has Rademacher cotype q if and only if it has metric cotype q.*

The definition of metric cotype stipulates that for every $n \in \mathbb{N}$ there exists an even integer $m \in \mathbb{N}$ for which (39) holds true, but for certain applications it is important to have good bounds on m. The argument that was used above to rule out (38) shows that if $(\mathcal{M}, d_{\mathcal{M}})$ contains at least two points then the validity of (39) implies that $m \gtrsim n^{1/q}$. In [130] it was proved that one can ensure that m has this order of magnitude if X is Banach space with nontrivial Rademacher type.

Theorem 3.2. *Let $(X, \|\cdot\|_X)$ be a Banach space with Rademacher cotype $q < \infty$ and Rademacher type $p > 1$. Then* (39) *holds true for some even integer $m \leqslant \kappa n^{1/q}$, where $\kappa \in (0,\infty)$ depends on the geometry of X but not on n.*

As an example of an application of Theorem 3.2, the following characterization of the values of $p, q \in [1,\infty)$ for which $L_p[0,1]$ is uniformly homeomorphic to a subset of $L_q[0,1]$ was obtained in [130], answering a question posed by Enflo [52] in 1976.

Theorem 3.3. *Fix $p, q \in [1,\infty)$. Then $L_p[0,1]$ is uniformly homeomorphic to a subset of $L_q[0,1]$ if and only if either $p \leqslant q$ or $p, q \in [1,2]$.*

An analogous result was proved for coarse embeddings in [130] and for quasisymmetric embeddings in [145], answering a question posed by Väisälä [185]. The link between Theorem 3.2 and these results is that one can argue that if $(\mathcal{M}, d_{\mathcal{M}})$ satisfies (39) with $m \lesssim n^{1/q}$ then any Banach space that embeds into $\mathcal{M}$ in one of these senses inherits the cotype of $\mathcal{M}$. Thus, metric cotype (with appropriate dependence of m on n) is an obstruction to a variety of weak notions of metric embeddings. The following natural open question is of major importance.

Question 2. *Is it possible to obtain the conclusion of Theorem 3.2 without the assumption that X has nontrivial Rademacher type? In other words, is it true that any Banach space $(X, \|\cdot\|_X)$ with Rademacher cotype $q < \infty$ satisfies* (39) *with $m \lesssim_X n^{1/q}$?*

We conjecture that the answer to Question 2 is positive, in which case metric cotype itself, without additional assumptions, would be an invariant for uniform, coarse and quasisymmetric embeddings. For a general Banach space $(X, \|\cdot\|_X)$ of Rademacher cotype q the best known bound on m in terms of n in (39), due to [63], is $m \lesssim n^{1+1/q}$.

There are additional applications of metric cotype for which the tight dependence of m on n in (39) has no importance. In analogy to the discussion in Section 2.4, it was proved by Maurey and Pisier [122] that the only obstruction that can prevent a Banach space $(X, \|\cdot\|_X)$ from having finite Rademacher cotype is the presence of well-isomorphic copies of $\{\ell_\infty^n\}_{n=1}^\infty$. In [130] a variant of the definition of metric cotype was given, in analogy to the Bourgain-Milman-Wolfson variant of Enflo type, and it was shown that a *metric space* has finite metric cotype in this sense if and only if $c_{\mathcal{M}}(\{1,\ldots,m\}^n, \|\cdot\|_\infty) = 1$ for every $m,n \in \mathbb{N}$. This nonlinear Maurey-Pisier theorem was used in [130] to prove the following dichotomy result for general metric spaces, answering a question posed by Arora, Lovász, Newman, Rabani, Rabinovich and Vempala [7] and improving a Ramsey-theoretical result of Matoušek [116].

Theorem 3.4 (General metric dichotomy [130])**.** *Let $\mathcal{F}$ be a family of metric spaces. Then one of the following dichotomic possibilities must hold true.*

- *For every finite metric space $(\mathcal{M}, d_{\mathcal{M}})$ and for every $\varepsilon \in (0,\infty)$ there exists $\mathcal{N} \in \mathcal{F}$ such that*
$$c_{\mathcal{N}}(\mathcal{M}) \leqslant 1 + \varepsilon.$$
- *There exists $\alpha(\mathcal{F}), \kappa(\mathcal{F}) \in (0,\infty)$ and for each $n \in \mathbb{N}$ there exists an n-point metric space $(\mathcal{M}_n, d_{\mathcal{M}_n})$ such that for every $\mathcal{N} \in \mathcal{F}$ we have*
$$c_{\mathcal{N}}(\mathcal{M}_n) \geqslant \kappa(\mathcal{F})(\log n)^{\alpha(\mathcal{F})}.$$

We refer to [129, Sec. 1.1] and [133], as well as the survey paper [124], for more information on the theory of metric dichotomies. Theorem 3.4 leaves the following fundamental question open.

Question 3 (Metric cotype dichotomy problem [130, 133])**.** *Can one replace the constant $\alpha(\mathcal{F})$ of Theorem 3.4 by a constant $\alpha \in (0,1]$ that is independent of the family $\mathcal{F}$? It isn't even known if one can take $\alpha(\mathcal{F}) = 1$ for all families of metric spaces $\mathcal{F}$. If true, this would be sharp due to Bourgain's embedding theorem* [31]*.*

4. Markov type and cotype

As part of his investigation of the Lipschitz extension problem [15], K. Ball introduced a stronger version of type of metric spaces called *Markov type*. Other than its applications to Lipschitz extension, the notion of Markov type has found many applications in embedding theory, some of which will be described in Section 9.4.

Recall that a stochastic process $\{Z_t\}_{t=0}^\infty$ taking values in $\{1,\ldots,n\}$ is called a *stationary reversible Markov chain* if there exists an n by n stochastic matrix $A = (a_{ij})$ such that for every $t \in \mathbb{N} \cup \{0\}$ and every $i,j \in \{1,\ldots,n\}$ we have $\Pr[Z_{t+1} = j | Z_t = i] = a_{ij}$, for every $i \in \{1,\ldots,n\}$ the probability $\pi_i = \Pr[Z_t = i]$ does not depend on t, and $\pi_i a_{ij} = \pi_j a_{ji}$ for all $i,j \in \{1,\ldots,n\}$.

A metric space $(\mathcal{M}, d_{\mathcal{M}})$ is said to have Markov type $p \in (0,\infty)$ with constant $M \in (0,\infty)$ if for every $n \in \mathbb{N}$, every stationary reversible Markov chain on $\{1,\ldots,n\}$, every $f:\{1,\ldots,n\} \to \mathcal{M}$ and every time $t \in \mathbb{N}$ we have

$$\mathbb{E}\left[d_{\mathcal{M}}(f(Z_t), f(Z_0))^p\right] \leqslant M^p t \mathbb{E}\left[d_{\mathcal{M}}(f(Z_1), f(Z_0))^p\right]. \qquad (40)$$

Note that the triangle inequality implies that every metric space has Markov type 1 with constant 1. Ball proved [15] that if $p \in [1,2]$ then any $L_p(\mu)$ space has Markov type p with constant 1. Thus, while it is well-known that the standard random walk on the integers is expected to be at distance at most $\sqrt{t}$ from the origin after t steps, Ball established the less well-known fact that *any* stationary reversible random walk in Hilbert space has this property. If a metric space has Markov type p then it also has Enflo type p, as proved in [150]. In essence, Enflo type p corresponds to (40) in the special case when the Markov chain is the standard random walk on the Hamming cube $\{-1,1\}^n$. Thus the Markov type p condition is a strengthening of Enflo type, its power arising in part from the flexibility to choose any stationary reversible Markov chain whatsoever.

Remark 4.1. We do not know to what extent Enflo type $p > 1$ implies Markov type p (or perhaps Markov type q for some $1 < q < p$). When the metric space $(\mathcal{M}, d_{\mathcal{M}})$ is an unweighted graph equipped with the shortest path metric (as is often the case in applications), it is natural to introduce an intermediate notion of Markov type in which the Markov chains are only allowed to "move" along edges, i.e., by considering (40) under the additional restriction that if $a_{ij} > 0$ then $\{f(i), f(j)\}$ is an edge. Call this notion "edge Markov type p". For some time it was unclear whether edge Markov type p implies Markov type p. However, in [147] it was shown that there exists a Cayley graph with edge Markov type p for every $1 < p < \frac{4}{3}$ that does not have nontrivial Enflo type. It is unknown whether a similar example exists with edge Markov type 2.

In [149] it was shown that for $p \in [2,\infty)$ any $L_p(\mu)$ space has Markov type 2 (with constant $M \asymp \sqrt{p}$). More generally, it is proved in [149] that any p-uniformly smooth Banach space has Markov type p. Uniform smoothness, and its dual notion uniform convexity, are defined as follows. Let $(X, \|\cdot\|_X)$ be a normed space with unit sphere $S_X = \{x \in X : \|x\|_X = 1\}$. The *modulus of uniform convexity* of X is defined for $\varepsilon \in [0,2]$ as

$$\delta_X(\varepsilon) = \inf\left\{1 - \frac{\|x+y\|_X}{2} :\ x, y \in S_X,\ \|x-y\|_X = \varepsilon\right\}. \qquad (41)$$

X is said to be *uniformly convex* if $\delta_X(\varepsilon) > 0$ for all $\varepsilon \in (0,2]$. X is said to have modulus of uniform convexity of power type q if there exists a constant $c \in (0,\infty)$ such that $\delta_X(\varepsilon) \geqslant c\,\varepsilon^q$ for all $\varepsilon \in [0,2]$. It is straightforward to check that in this case necessarily $q \geqslant 2$. The *modulus of uniform smoothness* of X is defined for $\tau \in (0,\infty)$ as

$$\rho_X(\tau) \stackrel{\mathrm{def}}{=} \sup\left\{\frac{\|x+\tau y\|_X+\|x-\tau y\|_X}{2}-1:\ x,y\in S_X\right\}. \tag{42}$$

X is said to be *uniformly smooth* if $\lim_{\tau\to 0}\rho_X(\tau)/\tau = 0$. X is said to have modulus of uniform smoothness of power type p if there exists a constant $C \in (0,\infty)$ such that $\rho_X(\tau) \leqslant C\tau^p$ for all $\tau \in (0,\infty)$. It is straightforward to check that in this case necessarily $p \in [1,2]$.

For concreteness, we recall [74] (see also [17]) that if $p \in (1,\infty)$ then $\delta_{\ell_p}(\varepsilon) \gtrsim_p \varepsilon^{\max\{p,2\}}$ and $\rho_{\ell_p}(\tau) \lesssim_p \tau^{\min\{p,2\}}$. The moduli appearing in (41) and (42) relate to each other via the following classical duality formula of Lindenstrauss [106]:

$$\rho_{X^*}(\tau) = \sup\left\{\frac{\tau\varepsilon}{2}-\delta_X(\varepsilon):\ \varepsilon\in[0,2]\right\}. \tag{43}$$

An important theorem of Pisier [162] asserts that X admits an equivalent uniformly convex norm if and only if it admits an equivalent norm whose modulus of uniform convexity is of power type q for some $q \in [2,\infty)$. Similarly, X admits an equivalent uniformly smooth norm if and only if it admits an equivalent norm whose modulus of uniform smoothness is of power type p for some $p \in (1,2]$.

We will revisit these notions later, but at this point it suffices to say that, as proved in [149], any Banach space that admits an equivalent norm whose modulus of uniform smoothness is of power type p also has Markov type p. The relation between Rademacher type p and Markov type p is unclear. While for every $p \in (1,2]$ there exist Banach spaces with Rademacher p that do not admit any equivalent uniformly smooth norm [86, 165], the following question remains open.

Question 4. *Does there exists a Banach space $(X,\|\cdot\|_X)$ with Markov type $p>1$ yet $(X,\|\cdot\|_X)$ does not admit a uniformly smooth norm?*

In addition to uniformly smooth Banach spaces, the Markov type of several spaces of interest has been computed. For example, the following classes of metric spaces are known to have Markov type 2: weighted graph theoretical trees [149], series parallel graphs [37], hyperbolic groups [149], simply connected Riemmanian manifolds with pinched negative sectional curvature [149], Alexandrov spaces of nonnegative curvature [154]. Also, the Markov type of certain p-Wasserstein spaces was computed in [5].

Recall that a metric space $(M, d_{\mathcal{M}})$ is *doubling* if there exists $K \in \mathbb{N}$ such that for every $x \in \mathcal{M}$ and $r \in (0,\infty)$ there exist $y_1,\ldots,y_K \in \mathcal{M}$ such that $B(x,r) \subseteq B(y_1, r/2) \cup \ldots \cup B(y_K, r/2)$, i.e., every ball in $\mathcal{M}$ can be covered by K balls of half the radius. Here, and in what follows, $B(z,\rho) = \{w \in \mathcal{M} :\ d_{\mathcal{M}}(z,w) \leqslant \rho\}$ for all $z \in \mathcal{M}$ and $\rho \geqslant 0$. The parameter K is called a *doubling constant* of $(\mathcal{M}, d_{\mathcal{M}})$.

Question 5. *Does every doubling metric space have Markov type 2? Specifically, does the Heisenberg group have Markov type 2?*

Assouad's embedding theorem [9] says that if $(\mathcal{M}, d_{\mathcal{M}})$ is doubling then the metric space $(\mathcal{M}, d_{\mathcal{M}}^{1-\varepsilon})$ admits a bi-Lipschitz embedding into Hilbert space for every $\varepsilon \in (0,1)$. As observed in [149], this implies that if $(\mathcal{M}, d_{\mathcal{M}})$ is doubling then it has Markov type p for all $p < 2$. It was also shown in [149] that if $(\mathcal{M}, d_{\mathcal{M}})$ is doubling with constant $K \in (1,\infty)$ then for every $n \in \mathbb{N}$, every stationary reversible Markov chain on $\{1,\ldots,n\}$, every $f : \{1,\ldots,n\} \to \mathcal{M}$ and every time $t \in \mathbb{N}$,

$$\forall\, u > 0, \quad \Pr\left[d_{\mathcal{M}}(f(Z_t), f(Z_0)) \geqslant u\sqrt{t}\right] \leqslant \frac{O((\log K)^2)}{u^2}\mathbb{E}\left[d_{\mathcal{M}}(f(Z_1), f(Z_0))^2\right]. \quad (44)$$

Thus, one can say that doubling spaces have "weak Markov type 2". Using the method of [166] it is also possible to show that doubling spaces have Enflo type 2.

Further support of a positive answer to Question 5 was obtained in [149], where it was shown that the Laakso graphs $\{G_k\}_{k=0}^{\infty}$ have Markov type 2. These graphs are defined [101] iteratively by letting G_0 be a single edge and G_{i+1} is obtained by replacing the middle third of each edge of G_i by a quadrilateral; see Figure 4. Equipped with their shortest path metric, each Laakso graph G_k is doubling with constant 6 (see the proof of [102, Thm. 2.3]), yet, as proved by Laakso [101], we have $\lim_{k\to\infty} c_{\ell_2}(G_k) = \infty$ (in fact [102, Thm. 2.3] asserts that $c_{\ell_2}(G_k) \gtrsim \sqrt{k}$). The graphs $\{G_k\}_{k=0}^{\infty}$ are among the standard examples of doubling spaces that do not well-embed into Hilbert space, yet, as proved in [149], they do have Markov type 2. The Heisenberg group, i.e., the group of all 3 by 3 matrices generated by the set

$$S = \left\{ \begin{pmatrix} 1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}, \begin{pmatrix} 1 & -1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}, \begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{pmatrix}, \begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & -1\\ 0 & 0 & 1\end{pmatrix} \right\},$$

and equipped with the associated word metric, is another standard example of a doubling space that does not admit a bi-Lipschitz embedding into Hilbert space [160, 175]. However, as indicated in Question 5,

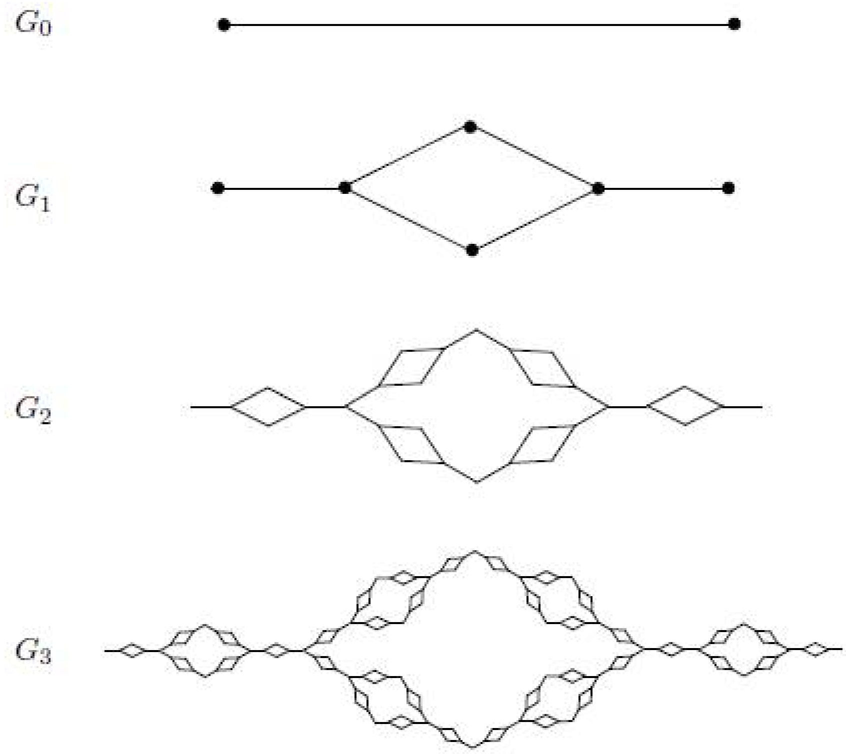

FIGURE 3. The first four Laakso graphs.

the intriguing problem whether the Heisenberg group has Markov type 2 remains open.

Note that by the nonlinear Maurey-Pisier theorem [130], as discussed in Section 3, a doubling metric space must have finite metric cotype. The Laakso graphs $\{G_k\}_{k=0}^{\infty}$, being examples of series parallel graphs, admit a bi-Lipschitz embedding into ℓ_1 with distortion bounded by a constant independent of k, as proved in [73]. Since ℓ_1 has Rademacher cotype 2, it follows from Theorem 3.1 that the Laakso graphs have metric cotype 2 (with the constant C in (39) taken to be independent of k). We do not know if all doubling metric spaces have metric cotype 2. The Heisenberg group is a prime example for which this question remains open. Note that the Heisenberg group does not embed into any $L_1(\mu)$ space [41]. Therefore the above reasoning for the Laakso graphs does not apply to the Heisenberg group.

Metric trees and the Laakso graphs are nontrivial examples of planar graphs that have Markov type 2. This result of [149] was extended to all series parallel graphs in [37]. It was also shown in [149] that any planar graph satisfies the weak Markov type 2 inequality (44), and using [166] one can show that planar graphs have Enflo type 2. It remains open whether all planar graphs have Markov type 2.

4.1. **Lipschitz extension via Markov type and cotype.** Here we explain Ball's original motivation for introducing Markov type.

Ball also introduced in [15] a *linear* property of Banach spaces that he called Markov cotype 2, and he indicated a two-step definition that could be used to extend this notion to general metric spaces. Motivated by Ball's ideas, the following variant of his definition was introduced in [132]. A metric space $(\mathcal{M}, d_{\mathcal{M}})$ has *metric Markov cotype* $q \in (0,\infty)$ with constant $C \in (0,\infty)$ if for every $m, n \in \mathbb{N}$, every n by n symmetric stochastic matrix $A = (a_{ij})$, and every $x_1,\ldots,x_n \in \mathcal{M}$, there exist $y_1,\ldots,y_n \in \mathcal{M}$ satisfying

$$\sum_{i=1}^n d_{\mathcal{M}}(x_i,y_i)^q + m\sum_{i=1}^n\sum_{j=1}^n a_{ij} d_{\mathcal{M}}(y_i,y_j)^q \leqslant C^q \sum_{i=1}^n\sum_{j=1}^n \left(\frac{1}{m}\sum_{t=0}^{m-1} A^t\right)_{ij} d_{\mathcal{M}}(x_i,x_j)^q. \quad (45)$$

To better understand the meaning of (45), observe that the Markov type p condition for $(\mathcal{M}, d_{\mathcal{M}})$ implies that

$$\sum_{i=1}^n\sum_{j=1}^n (A^m)_{ij} d_{\mathcal{M}}(x_i,x_j)^p \leqslant M^p m \sum_{i=1}^n\sum_{j=1}^n a_{ij} d_{\mathcal{M}}(x_i,x_j)^p. \quad (46)$$

Thus (45) aims to reverse the direction of the inequality in (46), with the following changes. One is allowed to pass from the initial points $x_1,\ldots,x_n \in \mathcal{M}$ to new points $y_1,\ldots,y_m \in \mathcal{M}$. The first summand in the left hand side of (45) ensures that on average y_i is close to x_i. The remaining terms in (45) correspond to the reversal of (46), with $\{x_i\}_{i=1}^n$ replaced by $\{y_i\}_{i=1}^n$ in the left hand side, and the power A^m replaced by the Cesàro average $\frac{1}{m}\sum_{t=0}^{m-1} A^t$.

Although (45) was inspired by Ball's ideas, the formal relation between the above definition of metric Markov cotype and Ball's original definition in [15] is unclear. We chose to work with the above definition since it suffices for the purpose of Ball's original application, and in addition it can be used for other purposes. Specifically, metric Markov cotype is key to the development of calculus for nonlinear spectral gaps and the construction of super-expanders; an aspect of the Ribe program that we will not describe here for lack of space (see [132]).

For $q \in [1,\infty)$, a metric space $(\mathcal{M}, d_\mathcal{M})$ is called W_q-barycentric with constant $\Gamma \in (0,\infty)$ if for every finitely supported probability measure μ on $\mathcal{M}$ there exists a point $\beta_\mu \in \mathcal{M}$ (a barycenter of μ) such that $\beta_{\delta_x} = x$ for all $x \in X$ and for every two finitely supported probability measures μ, ν we have $d_\mathcal{M}(\beta_\mu, \beta_\nu) \leqslant \Gamma W_q(\mu,\nu)$, where $W_q(\cdot,\cdot)$ denotes the q-Wasserstein metric (see [187, Sec. 7.1]). Note that by convexity every Banach space is W_q-barycentric with constant 1.

The following theorem from [135] is a metric space variant of Ball's Lipschitz extension theorem [15] (the proof follows the same ideas as in [15] with some technical differences of lesser importance).

Theorem 4.1. *Fix $q \in (0,\infty)$ and let $(\mathcal{M}, d_\mathcal{M})$, $(\mathcal{N}, d_\mathcal{N})$ be two metric spaces. Assume that $\mathcal{M}$ has Markov type q with constant M and $\mathcal{N}$ has metric Markov cotype q with constant C. Assume also that $\mathcal{N}$ is W_q-barycentric with constant Γ. Then for every $A \subseteq \mathcal{M}$, every finite $S \subseteq \mathcal{M} \smallsetminus A$, and every Lipschitz mapping $f : A \to \mathbb{N}$ there exists $F : A \cup S \to \mathbb{N}$ satisfying $F(x) = f(x)$ for all $x \in A$ and*

$$\|F\|_{\mathrm{Lip}} \lesssim_{\Gamma,M,C} \|f\|_{\mathrm{Lip}},$$

where the implied constant depends only on Γ, M, C.

Ball proved Theorem 4.1 when $\mathcal{N}$ is a Banach space, $q = 2$, and the metric Markov cotype assumption is replaced by his linear notion of Markov cotype. He proved that every Banach space that admits an equivalent norm with modulus of uniform convexity of power type 2 satisfies his notion of Markov cotype 2. In combination with [149], it follows that the conclusion of Theorem 4.1 holds true if $\mathcal{M}$ is a Banach space that admits an equivalent norm with modulus of uniform smoothness of power type 2 and $\mathcal{N}$ is a Banach space that admits an equivalent norm with modulus of uniform convexity of power type 2. In particular, for $1 < q \leqslant 2 \leqslant p < \infty$ we can take $\mathcal{M} = \ell_p$ and $\mathcal{N} = \ell_q$. This answers positively a 1983 conjecture of Johnson and Lindenstrauss [88]. The motivation of the question of Johnson and Lindenstrauss belongs to the Ribe program (see also [115]): to obtain a metric analogue of a classical theorem of Maurey [120] that implies this result for linear operators, i.e., in Maurey's setting A is a closed linear subspace and f is a linear operator, in which case the conclusion is that $F : \mathcal{M} \to \mathcal{N}$ is a bounded linear operator with $\|F\| \lesssim_{M,C} \|f\|$. Examples of applications of Ball's extension theorem can be found in [144, 126].

In [135] it is shown that if $(X, \|\cdot\|_X)$ is a Banach space that admits an equivalent norm with modulus of uniform convexity of power type q then it has metric Markov cotype q. Also, it is shown in [135] that certain barycentric metric spaces have metric Markov cotype q; this is true in particular for $CAT(0)$ spaces, and hence also all simply connected manifolds of nonpositive sectional curvature (see [36]). These facts, in conjunction with Theorem 4.1, yield new Lipschitz extension theorems; see [135].

For Banach spaces the notion of metric Markov cotype q does not coincide with Rademacher cotype: one can deduce from a clever construction of Kalton [95] that there exists a closed linear subspace X of L_1 (hence X has Rademacher cotype 2) that does not have metric Markov cotype q for any $q < \infty$. The following natural question remains open.

Question 6. *Does ℓ_1 have metric Markov cotype 2?*

By Theorem 4.1, a positive solution of Question 6 would answer a well known question of Ball [15], by showing that every Lipschitz function from a subset of ℓ_2 to ℓ_1 can be extended to a Lipschitz function defined on all of ℓ_2. See [114] for ramifications of this question in theoretical computer science.

5. Markov convexity

Deep work of James [84, 85] and Enflo [51] implies that a Banach space $(X, \|\cdot\|_X)$ admits an equivalent uniformly convex norm if and only if it admits an equivalent uniformly smooth norm, and these properties are equivalent to the assertion that any Banach space $(Y, \|\cdot\|_Y)$ that is finitely representable in X must be reflexive. Such spaces are called *superreflexive* Banach spaces. The Ribe program suggests that superreflexivity has a purely metric reformulation. This is indeed the case, as proved by Bourgain [32].

For $k, n \in \mathbb{N}$ let T_n^k denote the complete k-regular tree of depth n, i.e., the finite unweighted rooted tree such that the length of any root-leaf path equals n and every non-leaf vertex has exactly k adjacent vertices. We shall always assume that T_n^k is equipped with the shortest path metric $d_{T_r^k}(\cdot,\cdot)$, i.e., the distance between any two vertices is the sum of their distances to their least common ancestor. Bourgain's characterization of superreflexivity [32] asserts that a Banach space $(X, \|\cdot\|_X)$ admits an equivalent uniformly convex norm if and only if for all $k \geqslant 3$ we have

$$\lim_{n\to\infty} c_X(T_n^k) = \infty. \tag{47}$$

Bourgain's proof also yields the following asymptotic computation of the Euclidean distortion of T_n^k:

$$k \geqslant 3 \implies c_{\ell_2}\left(T_n^k\right) \asymp \sqrt{\log n}. \tag{48}$$

All known proofs of the lower bound $c_{\ell_2}\left(T_n^k\right) \gtrsim \sqrt{\log n}$ are non-trivial (in addition to the original proof of [32], alternative proofs appeared in [117, 113, 104]). In this section we will describe a proof of (48) from the viewpoint of random walks.

It is a nontrivial consequence of the work of Pisier [162] that the Banach space property of admitting an equivalent norm whose modulus of uniform convexity has power type p is an isomorphic local linear property. As such, the Ribe program calls for a purely metric reformulation of this property. Since Pisier proved [162] that a Banach space is superrreflexive if and only if it admits an equivalent norm whose modulus of uniform convexity has power type p for some $p \in [2,\infty)$, this question should be viewed as asking for a quantitative refinement of Bourgain's metric characterization of superreflexivity.

The following definition is due to [104]. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a Markov chain on a state space Ω. Given integers $k, s \geqslant 0$, denote by $\{\widetilde{Z}_t(s)\}_{t\in\mathbb{Z}}$ the process that equals Z_t for time $t \leqslant s$, and evolves independently (with respect to the same transition probabilities) for time $t > s$. Fix $p > 0$. A metric space $(\mathcal{M}, d_{\mathcal{M}})$ is called Markov p-convex with constant Π if for every Markov chain $\{Z_t\}_{t\in\mathbb{Z}}$ on a state space Ω, and every mapping $f : \Omega \to \mathcal{M}$,

$$\sum_{s=0}^{\infty}\sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_{\mathcal{M}}\left(f(Z_t), f\left(\widetilde{Z}_t\left(t-2^s\right)\right)\right)^p\right]}{2^{sp}} \leqslant \Pi^p \cdot \sum_{t\in\mathbb{Z}} \mathbb{E}\big[d_{\mathcal{M}}(f(Z_t), f(Z_{t-1}))^p\big]. \quad (49)$$

The infimum over those $\Pi \in [0,\infty]$ for which (49) holds for all Markov chains is called the Markov p-convexity constant of $\mathcal{M}$, and is denoted $\Pi_p(\mathcal{M})$. We say that $(\mathcal{M}, d_{\mathcal{M}})$ is Markov p-convex if $\Pi_p(\mathcal{M}) < \infty$.

We will see in a moment how to work with (49), but we first state the following theorem, which constitutes a completion of the step of the Ribe program that corresponds to the Banach space property of admitting an equivalent norm whose modulus of uniform convexity has power type p. The "only if" part of this statement is due to [104] and the "if" part is due to [129].

Theorem 5.1. *Fix $p \in [2,\infty)$. A Banach space $(X, \|\cdot\|_X)$ admits an equivalent norm whose modulus of uniform convexity has power type p if and only if $(X, \|\cdot\|_X)$ is Markov p-convex.*

The meaning of (49) will become clearer once we examine the following example. Fix an integer $k \geqslant 3$ and let $\{Z_t\}_{t\in\mathbb{Z}}$ be the following

Markov chain whose state space is T_n^k. Z_t equals the root of T_n^k for $t \leqslant 0$, and $\{Z_t\}_{t\in\mathbb{N}}$ is the standard *outward* random walk (i.e., if $0 \leqslant t < n$ then Z_{t+1} is distributed uniformly over the $k-1$ neighbors of Z_t that are further away from the root than Z_t), with absorbing states at the leaves. Suppose that $(\mathcal{M}, d_{\mathcal{M}})$ is a metric space that is Markov p-convex with constant Π, and for some $\lambda, D \in (0,\infty)$ we are given an embedding $f : T_n^k \to \mathcal{M}$ that satisfies $\lambda d_{T_n^k}(x,y) \leqslant d_{\mathcal{M}}(f(x), f(y)) \leqslant D\lambda d_{T_n^k}(x,y)$ for all $x, y \in T_n^k$. For every $s, t \in \mathbb{N}$ such that $2^s \leqslant t \leqslant n$, with probability at least $1 - 1/(k-1)$ the vertices Z_{t-2^s+1} and $\widetilde{Z}_{t-2^s+1}(t-2^s)$ are distinct, in which case $d_{T_n^k}(Z_t, \widetilde{Z}_t(t-2^s)) = 2^{s+1}$. It therefore follows from (49) that

$$\lambda^p n \log n \lesssim \sum_{s=0}^{\infty} \sum_{t\in\mathbb{Z}} \frac{\mathbb{E}\left[d_{\mathcal{M}}\left(f(Z_t), f\left(\widetilde{Z}_t\left(t-2^s\right)\right)\right)^p\right]}{2^{sp}}$$
$$\leqslant \Pi^p \cdot \sum_{t\in\mathbb{Z}} \mathbb{E}\big[d_{\mathcal{M}}(f(Z_t), f(Z_{t-1}))^p\big] \leqslant \Pi^p D^p \lambda^p n.$$

Consequently,

$$c_{\mathcal{M}}(T_n^k) \gtrsim \frac{1}{\Pi_p(\mathcal{M})} (\log n)^{1/p}.$$

In particular, when $\mathcal{M} = \ell_2$ this explains (48).

A different choice of Markov chain can be used in combination with Markov convexity to compute the asymptotic behavior of the Euclidean distortion of the *lamplighter group* over $\mathbb{Z}_n$; see [104, 14]. Similar reasoning also applies to the Laakso graphs $\{G_k\}_{k=0}^{\infty}$, as depicted in Figure 4. In this case let $\{Z_t\}_{t=0}^{\infty}$ be the Markov chain that starts at the leftmost vertex of G_k (see Figure 4), and at each step moves to the right. If Z_t is a vertex of degree 3 then Z_{t+1} equals one of the two vertices on the right of Z_t, each with probability $\frac{1}{2}$. An argument along the above lines (see [129, Sec. 3]) yields

$$c_{\mathcal{M}}(G_k) \gtrsim \frac{1}{\Pi_p(\mathcal{M})} k^{1/p}.$$

This estimate is sharp when $\mathcal{M} = \ell_q$ for all $q \in (1,\infty)$. Note that since the Laakso graphs are doubling, they do not contain bi-Lipschitz copies of T_n^3 with distortion bounded independently of n. Thus the Markov convexity invariant applies equally well to trees and Laakso graphs, despite the fact that these examples are very different from each other as metric spaces. Recently Johnson and Schechtman [90] proved that if for a Banach space $(X, \|\cdot\|_X)$ we have $\lim_{k\to\infty} c_X(G_k) = \infty$ then X is superreflexive. Thus the nonembeddability of the Laakso graphs is a metric characterization of superreflexivity that is different from Bourgain's characterization (47).

In addition to uniformly convex Banach spaces, other classes of metric spaces for which Markov convexity has been computed include Alexandrov spaces of nonnegative curvature [12] (they are Markov 2-convex) and the Heisenberg group (it is Markov 4-convex, as shown by Sean Li). Markov convexity has several applications to metric geometry, including a characterization of tree metrics that admit a bi-Lipschitz embedding into Euclidean space [104], a polynomial time approximation algorithm to compute the ℓ_p distortion of tree metrics [104], and applications to the theory of Lipschitz quotients [129].

6. Metric smoothness?

Since a Banach space admits an equivalent uniformly convex norm if and only if it admits an equivalent uniformly smooth norm, Bourgain's characterization of superreflexivity implies that, for every $k \geqslant 3$, a Banach space X admits an equivalent uniformly smooth norm if and only if $\lim_{n\to\infty} c_X(T_n^k) = \infty$. Nevertheless, a subtlety of this problem appears if one is interested in equivalent norms whose modulus of uniform smoothness has a given power type. Specifically, a Banach space X admits an equivalent norm whose modulus of uniform smoothness has power type p if and only if X^* admits an equivalent norm whose modulus of uniform convexity has power type $p/(p-1)$; this is an immediate consequence of (43). Despite this fact, and in contrast to Theorem 5.1, we do not know how to complete the Ribe program for the property of admitting an equivalent norm whose modulus of uniform smoothness has power type p. The presence of Trees and Laakso graphs is a natural obstruction to uniform convexity, but it remains open to isolate a natural (and useful) family of metric spaces whose presence is an obstruction to uniform smoothness of power type p.

7. Bourgain's discretization problem

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces with unit balls B_X and B_Y, respectively. For $\varepsilon \in (0,1)$ let $\delta_{X\hookrightarrow Y}(\varepsilon)$ be the supremum over those $\delta \in (0,1)$ such that every δ-net $\mathcal{N}_\delta$ in B_X satisfies $c_Y(\mathcal{N}_\delta) \geqslant (1-\varepsilon)c_Y(X)$. $\delta_{X\hookrightarrow Y}(\cdot)$ is called the *discretization modulus* corresponding to X, Y. Ribe's theorem follows from the assertion that if $\dim(X) < \infty$ then $\delta_{X\hookrightarrow Y}(\varepsilon) > 0$ for all $\varepsilon \in (0,1)$. This implication follows from the classical observation [44] that uniformly continuous mappings on Banach spaces are bi-Lipschitz for large distances, and a

w^* differentiation argument of Heinrich and Mankiewicz [77]; see [64] for the details.

In [33] Bourgain found a new proof of Ribe's theorem that furnished an explicit bound on $\delta_{X\hookrightarrow Y}(\cdot)$. Specifically, if $\dim(X)=n$ then

$$\forall\,\varepsilon\in(0,1),\quad \delta_{X\hookrightarrow Y}(\varepsilon)\geqslant e^{-(n/\varepsilon)^{Cn}}, \tag{50}$$

where $C\in(0,\infty)$ is a universal constant. (50) should be viewed as a quantitative version of Ribe's theorem, and it yields an abstract and generic way to obtain a family of finite metric spaces that serve as obstructions whose presence characterizes the failure of any given isomorphic finite dimensional linear property of Banach spaces; see [157, 159].

In light of the Ribe program it would be of great interest to determine the asymptotic behavior in n of, say, $\delta_{X\hookrightarrow Y}(1/2)$. However, the bound (50) remains the best known estimate, while the known (simple) upper bounds on $\delta_{X\hookrightarrow Y}(1/2)$ decay like a power of n; see [64]. This question is of interest even when X,Y are restricted to certain subclasses of Banach spaces, in which the following improvement is known [64]: for all $p\in[1,\infty)$ we have $\delta_{X\hookrightarrow L_p(\mu)}(1/2)\gtrsim(\dim(X))^{-5/2}$ (the implied constant is universal). We refer to [64] for a more general statement along these lines, as well as to [105, 82] for alternative approaches to this question.

8. Nonlinear Dvoretzky theorems

A classical theorem of Dvoretzky [47] asserts, in confirmation of a conjecture of Grothendieck [70], that for every $k\in\mathbb{N}$ and $D>1$ there exists $n=n(k,D)\in\mathbb{N}$ such that every n-dimensional normed space has a k-dimensional linear subspace that embeds into Hilbert space with distortion D; see [139, 138, 174] for the best known bounds on $n(k,D)$. In accordance with the Ribe program, Bourgain, Figiel and Milman asked in 1986 if there is an analogue of the Dvoretzky phenomenon which holds for general metric spaces. Specifically, they investigated the largest $m\in\mathbb{N}$ such that *any* finite metric space $(\mathcal{M},d_{\mathcal{M}})$ of cardinality n has a subset $S\subseteq\mathcal{M}$ with $|S|\geqslant m$ such that the metric space $(S,d_{\mathcal{M}})$ embeds with distortion D into Hilbert space. Twenty years later, Tao asked an analogous question in terms of Hausdorff dimension: given $\alpha>0$ and $D>1$, what is the supremum over those $\beta\geqslant 0$ such that every compact metric space $\mathcal{M}$ with $\dim_H(\mathcal{M})\geqslant\alpha$ has a subset $S\subseteq\mathcal{M}$ with $\dim_H(\mathcal{M})\geqslant\beta$ that embeds into Hilbert space with distortion D? Here $\dim_H(\cdot)$ denotes Hausdorff dimension.

A pleasing aspect of the Ribe program is that sometimes we get more than we asked for. In our case, we asked for almost Euclidean subsets, but the known answers to the above questions actually provide subsets that are even more structured: they are approximately ultrametric. Before describing these answers to the above questions, we therefore first discuss the structure of ultrametric spaces, since this additional structure is crucial for a variety of applications.

8.1. **The structure of ultrametric spaces.** Let $(\mathcal{M}, d_\mathcal{M})$ be an ultrametric space, i.e.,

$$\forall\, x,y,z \in \mathcal{M}, \quad d_\mathcal{M}(x,y) \leqslant \max\left\{d_\mathcal{M}(x,z), d_\mathcal{M}(y,z)\right\}. \tag{51}$$

In the discussion below, assume for simplicity that $\mathcal{M}$ is finite: this case contains all the essential ideas, and the natural extensions to infinite ultrametric spaces can be found in e.g. [81, 134, 98]. Define an equivalence relation $\sim$ on $\mathcal{M}$ by

$$\forall\, x,y \in \mathcal{M}, \quad x \sim y \iff d_\mathcal{M}(x,y) < \operatorname{diam}(\mathcal{M}) = \max_{z,w\in\mathcal{M}} d_\mathcal{M}(z,w).$$

Observe that it is the ultra-triangle inequality (51) that makes $\sim$ be indeed an equivalence relation. Let $A_1,\ldots,A_k$ be the corresponding equivalence classes. Thus $d_\mathcal{M}(x,y) < \operatorname{diam}(\mathcal{M})$ if $(x,y) \in \bigcup_{i=1}^k A_i \times A_i$ and $d_\mathcal{M}(x,y) = \operatorname{diam}(\mathcal{M})$ if $(x,y) \in \mathcal{M} \smallsetminus \bigcup_{i=1}^k A_i \times A_i$.

By applying this construction to each equivalence class separately, and iterating, one obtains a sequence of partitions $\mathscr{P}_0,\ldots,\mathscr{P}_n$ of $\mathcal{M}$ such that $\mathscr{P}_0 = \{\mathcal{M}\}$, $\mathscr{P}_n = \{\{x\}\}_{x\in\mathcal{M}}$, and $\mathscr{P}_{i+1}$ is a refinement of $\mathscr{P}_i$ for all $i \in \{0,\ldots,n-1\}$. Moreover, for every $x,y \in \mathcal{M}$, if we let $i \in \{0,\ldots,n\}$ be the maximal index such that $x,y \in A$ for some $A \in \mathscr{P}_i$, then $d_\mathcal{M}(x,y) = \operatorname{diam}(A)$. Alternatively, consider the following graph-theoretical tree whose vertices are labeled by subsets of $\mathcal{M}$. The root is labeled by $\mathcal{M}$ and the ith level of the tree is in one-to-one correspondence with the elements of the partition $\mathscr{P}_i$. The descendants of an i level vertex whose label is $A \in \mathscr{P}_i$ are declared to be the $i+1$ level vertices whose labels are $\{B \in \mathscr{P}_{i+1} :\ B \subseteq A\}$. With this combinatorial picture in mind, $\mathcal{M}$ can be identified as the leaves of the tree and the metric on $\mathcal{M}$ has the following simple description: the distance between any two leaves is the diameter of the set corresponding to their least common ancestor in the tree. This simple combinatorial structure of ultrametric spaces will be harnessed extensively in the ensuing discussion. See [81, 134, 98] for an extension of this picture to infinite compact ultrametric spaces (in which case the points of $\mathcal{M}$ are in one-to-one correspondence with the ends of an infinite tree).

We record two more consequences of the above discussion. First of all, by considering the natural lexicographical order that is induced on the leaves of the tree, we obtain a linear order $\prec$ on $\mathcal{M}$ such that if $x, y \in \mathcal{M}$ satisfy $x \preceq y$ then

$$\mathrm{diam}([x,y]) = \mathrm{diam}(\{z \in \mathcal{M} :\ x \preceq z \preceq y\}) = d_{\mathcal{M}}(x,y). \tag{52}$$

See [98] for a proof of the existence of a linear order satisfying (52) for every compact ultrametric space $(\mathcal{M}, d_{\mathcal{M}})$, in which case the order interval $[x,y]$ is always a Borel subsets of $\mathcal{M}$.

The second consequence that we wish to record here is that $(\mathcal{M}, d_{\mathcal{M}})$ admits an isometric embedding into the sphere of radius $\mathrm{diam}(\mathcal{M})/\sqrt{2}$ of Hilbert space. This is easily proved by induction on $\mathcal{M}$ as follows. Letting $A_1, \ldots, A_k$ be the equivalence classes as above, by the induction hypothesis there exist isometric embeddings $f_i : A_i \to H_i$, where $H_1, \ldots, H_k$ are Hilbert spaces and $\|f_i(x)\|_{H_i} = \mathrm{diam}(A_i)/\sqrt{2}$ for all $i \in \{1, \ldots, k\}$. Now define

$$f : \mathcal{M} \to \left(\bigoplus_{i=1}^{k} H_i\right) \oplus \ell_2^k \stackrel{\mathrm{def}}{=} H$$

by

$$x \in A_i \implies f(x) = f_i(x) + \sqrt{\frac{\mathrm{diam}(\mathcal{M})^2 - \mathrm{diam}(A_i)^2}{2}} e_i,$$

where $e_1, \ldots, e_k$ is the standard basis of $\ell_2^k = (\mathbb{R}^k, \|\cdot\|_2)$. One deduces directly from this definition, and the fact that $d_{\mathcal{M}}(x,y) = \mathrm{diam}(\mathcal{M})$ if $(x,y) \in \mathcal{M} \smallsetminus \bigcup_{i=1}^k A_i \times A_i$, that $\|f(x)\|_H = \mathrm{diam}(\mathcal{M})/\sqrt{2}$ for all $x \in \mathcal{M}$ and $\|f(x) - f(y)\|_H = d_{\mathcal{M}}(x,y)$ for all $x, y \in \mathcal{M}$. See [186] for more information on Hilbertian isometric embeddings of ultrametric spaces.

Thus, the reader should keep the following picture in mind when considering a finite ultrametric space $(\mathcal{M}, d_{\mathcal{M}})$: it corresponds to the leaves of a tree that are isometrically embedded in Hilbert space. Moreover, for every node of the tree the distinct subtrees that are rooted at its children are, after translation, mutually orthogonal.

8.2. **Ultrametric spaces are ubiquitous.** The following theorem is equivalent to the main result of [136], the original formulation of which will not be stated here; the formulation below is due to [134].

Theorem 8.1 (Ultrametric skeleton theorem)**.** *For every $\varepsilon \in (0,1)$ there exists $c_\varepsilon \in [1,\infty)$ with the following property. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a compact metric space and let μ be a Borel probability measure on $\mathcal{M}$. Then there exists a compact subset $S \subseteq \mathcal{M}$ and a Borel probability measure ν that is supported on S, such that $(S, d_{\mathcal{M}})$ embeds into an* *ultrametric space with distortion at most* $9/\varepsilon$ *and*

$$\forall (x,r)\in \mathcal{M}\times[0,\infty),\quad \nu\left(B(x,r)\cap S\right)\leqslant \left(\mu\left(B(x,c_\varepsilon r)\right)\right)^{1-\varepsilon}. \quad (53)$$

The subset $S\subseteq \mathcal{M}$ of Theorem 8.1 is called an *ultrametric skeleton* of $\mathcal{M}$ since, as we shall see below and is explained further in [134], it must be "large" and "spread out", and, more importantly, its main use is to deduce global information about the initial metric space $(\mathcal{M},d_{\mathcal{M}})$.

By Theorem 8.1 we know that despite the fact that ultrametric spaces have a very restricted structure, every metric measure space has an ultrametric skeleton. We will now describe several consequences of this fact. Additional examples of consequences of Theorem 8.1 are contained in Sections 9.1, 9.2, 9.3 below.

Our first order of business is to relate Theorem 8.1 to the above nonlinear Dvoretzky problems. Theorem 8.1 was discovered in the context of investigations on nonlinear Dvoretzky theory, and as such it constitutes another example of a metric space phenomenon that was uncovered due to the Ribe program.

Theorem 8.2. *For every* $\varepsilon\in(0,1)$ *and* $n\in\mathbb{N}$*, any* n*-point metric space has a subset of size at least* $n^{1-\varepsilon}$ *that embeds into an ultrametric space with distortion* $O(1/\varepsilon)$*.*

Proof. This is a simple corollary of the ultrametric skeleton theorem, which does not use its full force. Specifically, right now we will only care about the case $r=0$ in (53), though later we will need (53) in its entirety. So, let $(\mathcal{M},d_{\mathcal{M}})$ be an n-point metric space and let μ be the uniform probability measure on $\mathcal{M}$. An application of Theorem 8.1 to the metric measure space $(\mathcal{M},d_{\mathcal{M}},\mu)$ yields an ultrametric skeleton (S,ν). Thus $(S,d_{\mathcal{M}})$ embeds into an ultrametric space with distortion $O(1/\varepsilon)$. Since ν is a probability measure that is supported on S, there must exist a point $x\in S$ with $\nu(\{x\})\geqslant 1/|S|$. By (53) (with $r=0$) we have $\nu(\{x\})\leqslant \mu(\{x\})^{1-\varepsilon}=1/n^{1-\varepsilon}$. Thus $|S|\geqslant n^{1-\varepsilon}$, □

Theorem 8.2 was first proved in [127], as a culmination of the investigations in [34, 97, 30, 20, 22]. The best known bound for this problem is due to [152], where it is shown that if $\varepsilon\in(0,1)$ then any n-point metric space has a subset of size $n^{1-\varepsilon}$ that embeds into an ultrametric space with distortion at most

$$D(\varepsilon)=\frac{2}{\varepsilon(1-\varepsilon)^{\frac{1-\varepsilon}{\varepsilon}}}. \quad (54)$$

Theorem 8.2 belongs to the nonlinear Dvoretzky framework of Bourgain, Figiel and Milman because we have seen that ultrametric spaces admit an isometric embedding into Hilbert space. Moreover, the following matching impossibility result was proved in [22].

Theorem 8.3. *There exist universal constants $K,\kappa\in(0,\infty)$ and for every $n\in\mathbb{N}$ there exists an n-point metric space $(\mathcal{M}_n,d_{\mathcal{M}_n})$ such that for every $\varepsilon\in(0,1)$ we have*

$$\forall\, S\subseteq \mathcal{M}_n,\quad |S|\geqslant Kn^{1-\varepsilon}\implies c_{\ell_2}(S,d_{\mathcal{M}_n})\geqslant \frac{\kappa}{\varepsilon}.$$

In addition to showing that Theorem 8.2, and hence also Theorem 8.1, is asymptotically sharp, Theorem 8.3 establishes that, in general, the best way (up to constant factors) to find a large approximately Euclidean subset is to actually find a subset satisfying the more stringent requirement of being almost ultrametric.

Turning to the Hausdorff dimensional nonlinear Dvoretzky problem, we have the following consequence of the ultrametric skeleton theorem due to [136].

Theorem 8.4. *For every $\varepsilon\in(0,1)$ and $\alpha\in(0,\infty)$, any compact metric space of Hausdorff dimension greater than α has a closed subset of Hausdorff dimension greater than $(1-\varepsilon)\alpha$ that embeds into an ultrametric space with distortion $O(1/\varepsilon)$.*

Proof. Let $(\mathcal{M},d_{\mathcal{M}})$ be a compact metric space with $\dim_H(\mathcal{M})>\alpha$. By the Frostman lemma (see [80, 119]) it follows that there exists a Borel probability measure μ on $\mathcal{M}$ and $K\in(0,\infty)$ such that

$$\forall (x,r)\in\mathcal{M}\times[0,\infty),\quad \mu(B(x,r))\leqslant Kr^{\alpha}. \tag{55}$$

An application of Theorem 8.1 to the metric measure space $(\mathcal{M},d_{\mathcal{M}},\mu)$ yields an ultrametric skeleton (S,ν). If $\{B(x_i,r_i)\}_{i=1}^{\infty}$ is a collection of balls that covers S then

$$\begin{aligned}1=\nu(S)=\nu\left(\bigcup_{i=1}^{\infty}B(x_i,r_i)\right)&\leqslant\sum_{i=1}^{\infty}\nu(B(x_i,r_i))\\ &\overset{(53)}{\leqslant}\sum_{i=1}^{\infty}\mu\left(B(x_i,c_\varepsilon r_i)\right)^{1-\varepsilon}\overset{(55)}{\leqslant}K^{1-\varepsilon}c_\varepsilon^{(1-\varepsilon)\alpha}\sum_{i=1}^{\infty}r_i^{(1-\varepsilon)\alpha}.\end{aligned}$$

Having obtained an absolute positive lower bound on $\sum_{i=1}^{\infty}r_i^{(1-\varepsilon)\alpha}$ for all the covers of S by balls $\{B(x_i,r_i)\}_{i=1}^{\infty}$, we conclude the desired dimension lower bound $\dim_H(S)\geqslant(1-\varepsilon)\alpha$. □

Remark 8.1. It is also proved in [136] that there is a universal constant $\kappa\in(0,\infty)$ such that for every $\alpha>0$ and $\varepsilon\in(0,1)$ there exists a compact metric space $(\mathcal{M}, d_{\mathcal{M}})$ with $\dim_H(\mathcal{M}) = \alpha$ such that

$$\forall\, S \subseteq \mathcal{M}, \quad \dim_H(S) \geqslant (1-\varepsilon)\alpha \implies c_{\ell_2}(S, d_{\mathcal{M}}) \geqslant \frac{\kappa}{\varepsilon}.$$

Therefore, as in the case of the nonlinear Dvoretzky problem for finite metric spaces, the question of finding in a general metric space a high-dimensional subset which is approximately Euclidean is the same (up to constants) as the question of finding a high-dimensional subset which is approximately an ultrametric space. This phenomenon helps explain how investigations that originated in Dvoretzky's theorem led to a theorem such as 8.1 whose conclusion seems to be far from its initial Banach space motivation: the Ribe program indicated a natural question to ask, but the answer itself turned out to be a truly nonlinear phenomenon involving subsets which are approximately ultrametric spaces; a (perhaps unexpected) additional feature that is more useful than just the extraction of approximately Euclidean subsets.

Remark 8.2. As mentioned above, the best known distortion bound in Theorem 8.2 is given in (54). When $\varepsilon \to 1$ this bound tends to 2 from above. Distortion 2 is indeed a barrier here: the nonlinear Dvoretzky problem exhibits a phase transition at distortion 2 between power-type and logarithmic behavior of the largest Euclidean subset that can be extracted in general metric spaces of cardinality n. This phenomenon was discovered in [22]; see also [21, 23, 38] for related threshold phenomena. In their original paper [34] that introduced the nonlinear Dvoretzky problem, Bourgain Figiel and Milman proved that for every $D > 1$ any n-point metric space has a subset of size at least $c(D)\log n$ that embeds with distortion D into Hilbert space. They also proved that there exists constants $D_0 = 1.023...$, $\kappa \in (0,\infty)$ and for every $n \in \mathbb{N}$ there exists an n-point metric space $(\mathcal{M}_n, d_{\mathcal{M}_n})$ such that every $S \subseteq \mathcal{M}_n$ with $|S| \geqslant \kappa \log n$ satisfies $c_{\ell_2}(S, d_{\mathcal{M}_n}) \geqslant D_0$. In [22] this impossibility result was extended to any distortion in $(1,2)$, thus establishing the above phase transition phenomenon. The asymptotic behavior of the nonlinear Dvoretzky problem at distortion $D = 2$ remains unknown. For the Hausdorff dimensional version of this question the phase transition at distortion 2 becomes more extreme: for every $\delta \in (0,1/2)$ one can obtain [136] a version of Theorem 8.4 with the resulting subset S having ultrametric distortion $2+\delta$ and $\dim_H(S) \gtrsim \frac{\delta}{\log(1/\delta)}\alpha$. In contrast, for every $\alpha \in (0,\infty)$ there exists [136] a compact metric space $(\mathcal{M}, d_{\mathcal{M}})$ of Hausdorff dimension α such that if $S \subseteq \mathcal{M}$ embeds into Hilbert space with distortion strictly smaller than 2 then $\dim_H(S) = 0$.

9. Examples of applications

Several applications of the Ribe program have already been discussed throughout this article. In this section we describe some additional applications of this type. We purposefully chose examples of applications to areas which are far from Banach space theory, as an indication of the relevance of the Ribe program to a variety of fields.

9.1. **Majorizing measures.** A (centered) Gaussian process is a family of random variables $\{G_x\}_{x\in X}$, where X is an abstract index set and for every $x_1,\ldots,x_n \in X$ and $s_1,\ldots,s_n \in \mathbb{R}$ the random variable $\sum_{i=1}^n s_i G_{x_i}$ is a mean zero Gaussian random variable. To avoid technicalities that will obscure the key geometric ideas we will assume throughout the ensuing discussion that X is finite.

Given a centered Gaussian process $\{G_x\}_{x\in X}$, it is of great interest to compute (or estimate up to constants) the quantity $\mathbb{E}\left[\max_{x\in X} G_x\right]$. The process induces the metric $d(x,y)=\sqrt{\mathbb{E}\left[(G_x-G_y)^2\right]}$ on X, and this metric determines $\mathbb{E}\left[\max_{x\in X} G_x\right]$. Indeed, if $X=\{x_1,\ldots,x_n\}$ then consider the n by n matrix $D=(d(x_i,x_j)^2)$ and observe that D is negative semidefinite on the subspace $\{x\in\mathbb{R}^n:\ \sum_{i=1}^n x_i=0\}$ of $\mathbb{R}^n$. Then,

$$\mathbb{E}\left[\max_{i\in\{1,\ldots,n\}} G_{x_i}\right] = \frac{1}{(2\pi)^{n/2}}\int_{\left\{x\in\mathbb{R}^n:\ \sum_{i=1}^n x_i=0\right\}}\left(\max_{i\in\{1,\ldots,n\}}\left(\sqrt{-D}x\right)_i\right)e^{-\frac12\|x\|_2^2}dx.$$

More importantly, $\mathbb{E}\left[\max_{x\in X} G_x\right]$ is well-behaved under bi-Lipschitz deformations of (X,d): by the classical Slepian lemma (see e.g. [57, 178]), if $\{G_x\}_{x\in X}$ and $\{H_x\}_{x\in X}$ are Gaussian processes satisfying

$$\alpha\sqrt{\mathbb{E}\left[(G_x-G_y)^2\right]}\leqslant\sqrt{\mathbb{E}\left[(H_x-H_y)^2\right]}\leqslant\beta\sqrt{\mathbb{E}\left[(G_x-G_y)^2\right]}$$

for all $x,y\in X$, then

$$\alpha\mathbb{E}\left[\max_{x\in X} G_x\right]\leqslant\mathbb{E}\left[\max_{x\in X} H_x\right]\leqslant\beta\mathbb{E}\left[\max_{x\in X} G_x\right].$$

These facts suggest that one could "read" the value of $\mathbb{E}\left[\max_{x\in X} G_x\right]$ (up to universal constant factors) from the geometry of the metric space (X,d). How to do this explicitly has been a long standing mystery until Talagrand proved [178] in 1987 his celebrated *majorizing measure theorem*, which solved this question and, based on his investigations over the ensuing two decades, led to a systematic geometric method to estimate $\mathbb{E}\left[\max_{x\in X} G_x\right]$, with many important applications (see the books [103, 180, 181] and the references therein). We will now explain the majorizing measure theorem itself, and how it is a consequence of the ultrametric skeleton theorem; this deduction is due to [134].

For a finite metric space (X,d) let $\mathrm{Prob}(X)$ denote the space of all probability measures on X. Consider the quantity

$$\gamma_2(X,d)=\inf_{\mu\in\mathrm{Prob}(X)}\sup_{x\in X}\int_0^\infty\sqrt{\log\left(\frac{1}{\mu(B(x,r))}\right)}dr.$$

The parameter $\gamma_2(X,d)$ should be viewed as a Gaussian version of a covering number. Indeed, the integral $\int_0^\infty\sqrt{\log\left(1/\mu(B(x,r))\right)}dr$ is large if μ has a small amount of mass near x, so $\gamma_2(X,d)$ measures the extent to which one can spread unit mass over X so that all the points are "close" to this mass distribution in the sense that $\max_{x\in X}\int_0^\infty\sqrt{\log\left(1/\mu(B(x,r))\right)}dr$ is as small as possible.

Fernique introduced $\gamma_2(X,d)$ in [57], where he proved that every Gaussian process $\{G_x\}_{x\in X}$ satisfies $\mathbb{E}\left[\sup_{x\in X}G_x\right]\lesssim\gamma_2(X,d)$. Under additional assumptions, he also obtained a matching lower bound $\mathbb{E}\left[\sup_{x\in X}G_x\right]\gtrsim\gamma_2(X,d)$. Notably, Fernique proved in 1975 (see [58] and also [59, Thm. 1.2]) that if the metric $d(x,y)=\sqrt{\mathbb{E}\left[(G_x-G_y)^2\right]}$ happens to be an ultrametric then $\mathbb{E}\left[\sup_{x\in X}G_x\right]\asymp\gamma_2(X,d)$. By the Slepian lemma, the same conclusion holds true also if (X,d) embeds with distortion $O(1)$ into an ultrametric space.

It is simple to see how the ultrametric structure is relevant to such probabilistic considerations: in Section 8.1 we explained that an ultrametric space can be represented as a subset of Hilbert space corresponding to leaves of a tree in which the subtrees rooted at a given vertex are mutually orthogonal. In the setting of Gaussian processes orthogonality is equivalent to (stochastic) independence, so the geometric assumption of ultrametricity in fact has strong probabilistic ramifications. Specifically, the problem reduces to the estimation of the expected supremum of the following special type of Gaussian process, indexed by leaves of a graph theoretical tree $T=(V,E)$: to each edge $e\in E(T)$ we associated a mean zero Gaussian random variable H_e, the variables $\{H_e\}_{e\in E}$ are independent, and for every leaf x we have $G_x=\sum_{e\in E(P_x)}H_e$, where P_x is the unique path joining x and the root of T. This additional independence that the ultrametric structure provides allowed Fernique to directly prove that $\mathbb{E}\left[\sup_{x\in X}G_x\right]\gtrsim\gamma_2(X,d)$.

Due in part to the above evidence, Fernique conjectured in 1974 that $\mathbb{E}\left[\sup_{x\in X}G_x\right]\asymp\gamma_2(X,d)$ for every Gaussian process $\{G_x\}_{x\in X}$. Talagrand's majorizing measure theorem [178] is the positive resolution of this conjecture. By Fernique's work as described above, this amounts to the assertion that $\mathbb{E}\left[\sup_{x\in X} G_x\right] \gtrsim \gamma_2(X,d)$ for every Gaussian process $\{G_x\}_{x\in X}$. Talagrand's strategy was to show that there is $S \subseteq X$ that embeds into an ultrametric space with distortion $O(1)$, and $\gamma_2(S,d) \gtrsim \gamma_2(X,d)$. It would then follow from Fernique's original proof of the majorizing measure theorem for ultrametric spaces that $\mathbb{E}\left[\sup_{x\in S} G_x\right] \gtrsim \gamma_2(S,d) \gtrsim \gamma_2(X,d)$. Since trivially $\mathbb{E}\left[\sup_{x\in X} G_x\right] \geqslant \mathbb{E}\left[\sup_{x\in S} G_x\right]$, this strategy will indeed prove the majorizing measures theorem.

Consider the following quantity

$$\delta_2(X,d) = \sup_{\mu\in \mathrm{Prob}(X)} \inf_{x\in X} \int_0^\infty \sqrt{\log\left(\frac{1}{\mu(B(x,r))}\right)}dr.$$

For the same reason that $\gamma_2(X,d)$ is in essence a Gaussian covering number, $\delta_2(X,d)$ should be viewed as a Gaussian version of a packing number. A short argument (see [134]) shows that $\delta_2(X,d) \asymp \gamma_2(X,d)$ for every finite metric space (X,d).

Take $\mu \in \mathrm{Prob}(X)$ at which $\delta_2(X,d)$ is attained, i.e., for every $x \in X$ we have $\int_0^\infty \sqrt{\log\left(1/\mu(B(x,r))\right)}dr \geqslant \delta_2(X,d)$. An application of the ultrametric skeleton theorem to the metric measure space (X,d,μ) with, say, $\varepsilon = 3/4$, yields an ultrametric skeleton (S,ν). Thus $S \subseteq X$ embeds into an ultrametric space with distortion $O(1)$ and $\nu \in \mathrm{Prob}(S)$ satisfies $\nu(B(x,r)) \leqslant \sqrt[4]{\mu(B(x,Cr))}$ for all $x \in X$ and $r > 0$, where $C > 0$ is a universal constant. It follows that for every $x \in S$ the integral $\int_0^\infty \sqrt{\log\left(1/\nu(B(x,r))\right)}dr$ is at least $\frac{1}{2}\int_0^\infty \sqrt{\log\left(1/\mu(B(x,Cr))\right)}dr$, which by a change of variable equals $\frac{1}{2C}\int_0^\infty \sqrt{\log\left(1/\mu(B(x,r))\right)}dr$. But $\int_0^\infty \sqrt{\log\left(1/\mu(B(x,r))\right)}dr \gtrsim \delta_2(X,d)$ by our choice of μ. By the definition of $\delta_2(S,d)$ we have $\delta_2(S,d) \geqslant \int_0^\infty \sqrt{\log\left(1/\nu(B(x,r))\right)}dr$, so $\delta_2(S,d) \gtrsim \delta_2(X,d)$. Since $\delta_2(\cdot) \asymp \gamma_2(\cdot)$, the proof is complete.

Remark 9.1. The use of ultrametric constructions in metric spaces in order to prove maximal inequalities is a powerful paradigm in analysis. The original work of Fernique and Talagrand on majorizing measures is a prime example of the success of such an approach, and methods related to (parts of the proof of) the ultrametric skeleton theorem have been used in the context of certain maximal inequalities in [131, 151]. Other notable examples of related ideas include [43, 19, 153, 56].

9.2. **Lipschitz maps onto cubes.** Keleti, Máthé and Zindulka [98] proved the following theorem using the nonlinear Dvoretzky theorem for Hausdorff dimension (Theorem 8.4), thus answering a question of Urbański [184].

Theorem 9.1. *Fix $n \in \mathbb{N}$ and let $(\mathcal{M}, d_{\mathcal{M}})$ be a compact metric space of Hausdorff dimension bigger than n. Then there exists a Lipschitz mapping from $\mathcal{M}$ onto the cube $[0,1]^n$.*

If, in addition to the assumptions of Theorem 9.1, $(\mathcal{M}, d_{\mathcal{M}})$ is an ultrametric space, then Theorem 9.1 is proved as follows. By Frostman's lemma there exists a Borel probability measure μ on $\mathcal{M}$ and $K \in (0,\infty)$ such that $\mu(A) \leqslant K(\mathrm{diam}(A))^n$ for all Borel $A \subseteq \mathcal{M}$. Moreover, as explained in Section 8.1, there exists a linear order $\prec$ on $\mathcal{M}$ satisfying (52). Define $\varphi : \mathcal{M} \to [0,1]$ by $\varphi(x) = \mu(\{y \in \mathcal{M} : \ y \prec x\})$. Then $|\varphi(x) - \varphi(y)| \leqslant K d_{\mathcal{M}}(x,y)^n$ for all $x, y \in X$. Thus φ is continuous, and since μ is atom-free and $\mathcal{M}$ is compact, it follows that $\varphi(\mathcal{M}) = [0,1]$. Letting P be a $1/n$-Hölder Peano curve from $[0,1]$ onto $[0,1]^n$ (see e.g. [173]), the mapping $f = P \circ \varphi$ has the desired properties.

To prove Theorem 9.1, start with a general compact metric space $(\mathcal{M}, d_{\mathcal{M}})$ with $\dim_H(\mathcal{M}) > n$. By Theorem 8.4 there exists a compact subset $S \subseteq \mathcal{M}$ with $\dim_H(S) > n$ that admits a bi-Lipschitz embedding into an ultrametric space. By the above reasoning there exists a Lipschitz mapping f from S onto $[0,1]^n$. We now conclude the proof of Theorem 9.1 by extending f to a Lipschitz mapping $F : \mathcal{M} \to [0,1]$ (e.g. via the nonlinear Hahn-Banach theorem [26, Lem. 1.1]).

The above reasoning exemplifies the role of ultrametric skeletons: S was used as a tool, but the conclusion makes no mention of ultrametric spaces. Moreover, S itself admits an n-Hölder mapping onto $[0,1]$, something which is impossible to do for general $\mathcal{M}$. Only after composition with a Peano curve do we get a Lipschitz mapping to which the nonlinear Hahn-Banach theorem applies, allowing us to deduce a theorem about $\mathcal{M}$ with no mention of the ultrametric skeleton S.

9.3. **Approximate distance oracles and approximate ranking.** Here we explain applications of nonlinear Dvoretzky theory to computer science. By choosing to discuss only a couple examples we are doing an injustice to the impact that the Ribe program has had on theoretical computer science. We refer to [111, 10, 110, 118, 146, 189] for a more thorough (but still partial) description of the role of ideas that are motivated by the Ribe program in approximation algorithms. Even if we only focus attention on nonlinear Dvoretzky theorems, the full picture is omitted below: Theorem 8.2 also yields the best known lower bound [20, 22] on the competitive ratio of the randomized k-server problem; a central question in the field of online algorithms.

An n-point metric space (X, d_X) is completely determined by the numbers $\{d_X(x,y)\}_{x,y \in X}$. One can therefore store $\binom{n}{2}$ numbers, so that when one is asked the distance between two points $x, y \in X$ it is possible to output the number $d_X(x, y)$ in constant time[1]. The *approximate distance oracle* problem asks for a way to store $o(n^2)$ numbers so that given (a distance query) $x, y \in X$ one can quickly output a number that is guaranteed to be within a prescribed factor of the true distance $d_X(x, y)$. The following theorem was proved in [127] as a consequence of the nonlinear Dvoretzky theorem 8.2.

Theorem 9.2. *Fix $D > 1$. Every n-point metric space $(\{1, \ldots, n\}, d)$ can be preprocessed in time $O(n^2)$ to yield a data structure of size $O(n^{1+O(1/D)})$ so that given $i, j \in \{1, \ldots, n\}$ one can output in $O(1)$ time a number $E(i, j)$ that is guaranteed to satisfy*

$$d(i, j) \leqslant E(i, j) \leqslant Dd(i, j). \tag{56}$$

Here, and in what follows, all the implied constants in the $O(\cdot)$ notation are universal constants. The preprocessing time of Theorem 9.2 is due to Mendel and Schwob [137], improving over the original preprocessing time of $O(n^{2+O(1/D)})$ that was obtained in [127].

In their important paper [182], Thorup and Zwick constructed approximate distance oracles as in Theorem 9.2, but with query time $O(D)$. Their preprocessing time is $O(n^2)$, and the size of their data structure is $O(Dn^{1+2(1+O(1/D))/D})$. The key feature of 9.2 is that it yields constant query time, i.e., a true oracle. In addition, the proof of Theorem 9.2 is via a new geometric method that we will sketch below, based on nonlinear Dvoretzky theory.

Note that the exponent of n in the size of the Thorup-Zwick oracle is at most $1 + 2(1 + o(1))/D$, while in Theorem 9.2 it is $1 + C/D$ for some universal constant C (which can be shown to be at most 20). This difference in constants can be important for applications, but recently Wulff-Nilsen proved [191] that one can use the oracle of Theorem 9.2 as a black box (irrespective of the constant C) to construct an oracle of size $O(n^{1+2(1+\varepsilon)/D})$ whose query time depends only on ε. The significance of the constant 2 here is that [182] establishes that it

[1]For the sake of the discussion in this survey one should think of "time" as the number of locations in the data structure that are probed plus the number of arithmetic operations that are performed. "Size" refers to the number of floating point numbers that are stored. The computational model in which we will be working is the RAM model, although weaker computational models such as the "Unit cost floating-point word RAM model" will suffice. See [75, 127] for a discussion of these computational issues. The preprocessing algorithms below are randomized, in which case "preprocessing time" refers to "expected preprocessing time". All other algorithms are deterministic.

is sharp conditioned on the validity of a positive solution to a certain well-known combinatorial open question of Erdős [53].

Sommer, Verbin and Yu [177] have shown that Theorem 9.2 is sharp in the sense of the following lower bound in the cell-probe model[2]. Any data structure that, given a query $i,j\in\{1,\ldots,n\}$, outputs in time t a number $E(i,j)$ satisfying (56) must have size at least $n^{1+c/(tD)}/\log n$. This lower bound works even when the oracle's performance is measured only on metric spaces corresponding to sparse graphs. The fact that the query time t of Theorem 9.2 is a universal constant thus makes this theorem asymptotically sharp. Nonlinear Dvoretzky theory is the only currently known method that yields such sharp results.

It turns out that the proof of Theorem 8.2 in [127] furnishes a randomized polynomial time algorithm that, given an n-point metric space (X,d_X), outputs a subset $S\subseteq X$ with $|S|\geqslant n^{1-\varepsilon}$ such that (S,d_X) embeds into an ultrametric space with distortion $O(1/\varepsilon)$. Moreover, we can ensure that there exists an ultrametric ρ on X such that for every $x\in X$ and $s\in S$ we have $d_X(x,s)\leqslant\rho(x,s)\leqslant\frac{c}{\varepsilon}d_X(x,s)$, where $c\in(0,\infty)$ is a universal constant. The latter statement follows from the following general *ultrametric extension lemma* [127], though the proof of Theorem 8.2 in [127] actually establishes this fact directly without invoking Lemma 9.3 below (this is important if one cares about constant factors).

Lemma 9.3 (Extension lemma for approximate ultrametrics)**.** *Let (X,d_X) be a finite metric space and fix $S\subseteq X$ and $D\geqslant 1$. Suppose that that $\rho_0:S\times S\to[0,\infty)$ is an ultrametric on S satisfying $d_X(x,y)\leqslant\rho_0(x,y)\leqslant Dd_X(x,y)$ for all $x,y\in S$. Then there exists an ultrametric $\rho:X\times X\to[0,\infty)$ such that $\rho(x,y)=\rho_0(x,y)$ if $x,y\in S$, for every $x,y\in X$ we have $\rho(x,y)\geqslant d_X(x,y)/3$, and for every $x\in X$ and $y\in S$ we have $\rho(x,y)\leqslant 2Dd_X(x,y)$.*

We are now in position to apply Theorem 8.2 iteratively as follows. Set $S_0=\emptyset$ and let $S_1\subseteq X$ be the subset whose existence is stipulated in Theorem 8.2. Thus there exists an ultrametric ρ_1 on X satisfying $d_X(x,y)\leqslant\rho_1(x,y)\leqslant\frac{c}{\varepsilon}d_X(x,y)$ for all $x\in X$ and $y\in S_1$. Apply the same procedure to $X\smallsetminus S_1$, and continue inductively until the entire space X is exhausted. We obtain a partition $\{S_1,\ldots,S_m\}$ of X with the following properties holding for every $k\in\{1,\ldots,m\}$.

- $|S_k|\geqslant\left(n-\sum_{j=0}^{k-1}|S_j|\right)^{1-\varepsilon}$.

[2]See [142] for more information on the cell probe computational model. It suffice to say here that it is a weak model, so cell probe lower bounds should be viewed as strong impossibility results.

- There exists an ultrametric ρ_k on $X \smallsetminus \bigcup_{j=0}^{k-1} S_j$ satisfying
$$d_X(x,y) \leqslant \rho_k(x,y) \leqslant \frac{c}{\varepsilon} d_X(x,y)$$
for all $x \in X \smallsetminus \bigcup_{j=0}^{k-1} S_j$ and $y \in S_k$.

As we have seen in Section 8.1, for every $k \in \{1,\ldots,m\}$ the ultrametric ρ_k corresponds to a combinatorial tree whose leaves are $X \smallsetminus \bigcup_{j=0}^{k-1} S_j$ and each vertex of which is labeled by a nonegative number such that for $x,y \in X \smallsetminus \bigcup_{j=0}^{k-1} S_j$ the label of their least common ancestor is exactly $\rho_k(x,y)$. A classical theorem of Harel and Tarjan [76] (see also [24]) states that any N-vertex tree can be preprocessed in time $O(N)$ so as to yield a data structure of size $O(N)$ which, given two nodes as a query, returns their least common ancestor in time $O(1)$. By applying the Harel-Tarjan data structure to each of the trees corresponding to ρ_k we obtain an array of data structures (see Figure 4) that can answer distance queries as follows. Given distinct $x,y \in X$ let $k \in \{1,\ldots,m\}$ be the minimal index for which $\{x,y\} \cap S_k \neq \emptyset$. Thus $x,y \in X \smallsetminus \bigcup_{j=0}^{k-1} S_j$, and, using the Harel-Tarjan data structure corresponding to ρ_k, output in $O(1)$ time the label of the least common ancestor of x,y in the tree corresponding to ρ_k. This output equals $\rho_k(x,y)$, which, since $\{x,y\} \cap S_k \neq \emptyset$, satisfies $d_X(x,y) \leqslant \rho_k(x,y) \leqslant \frac{c}{\varepsilon} d_X(x,y)$. Setting $D = c/\varepsilon$ and analyzing the size of the data structure thus obtained (using the recursion for the cardinality of S_k), yields Theorem 9.2; the details of this computation can be found in [127].

The ideas presented above are used in [127] to solve additional data structure problems. For example, we have the following theorem that addresses the *approximate ranking problem*, in which the goal is to compress the natural "n proximity orders" (or "rankings") induced on each of the points in an n-point metric space (i.e., each $x \in X$ orders the points of X by increasing distance from itself).

Theorem 9.4. *Fix $D > 1$, $n \in \mathbb{N}$ and an n-point metric space (X, d_X). Then there exists a data structure which can be preprocessed in time $O\left(Dn^{2+O(1/D)} \log n\right)$, has size $O\left(Dn^{1+O(1/D)}\right)$, and supports the following type of queries. Given $x \in X$, have "fast access" to a bijection $\pi^{(x)} : \{1,\ldots,n\} \to X$ satisfying*
$$\forall\, 1 \leqslant i < j \leqslant n, \quad d_X\left(x, \pi^{(x)}(i)\right) \leqslant D d_X\left(x, \pi^{(x)}(j)\right).$$
By "fast access" to $\pi^{(x)}$ we mean that we can do the following in $O(1)$ time:

(1) Given $x \in X$ and $i \in \{1,\ldots,n\}$ output $\pi^{(x)}(i)$.

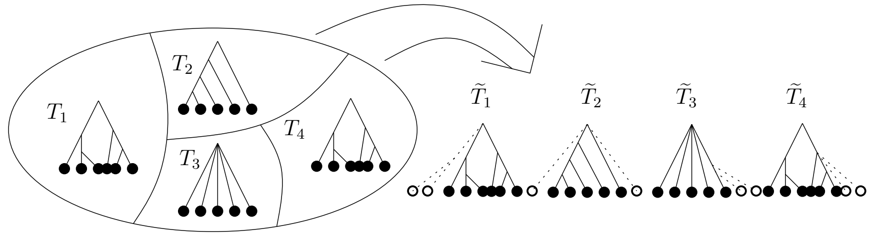

FIGURE 4. In the approximate distance oracle problem an iterative application of Theorem 8.2 yields an array of trees, which are then transformed into an array of Harel-Tarjan data structures. For the approximate ranking problem we also need to extend each tree to a tree whose leaves are the entire space X using Lemma 9.3. The nodes that were added to these trees are illustrated by empty circles, and the dotted lines are their connections to the original tree.

(2) Given $x, u \in X$ output $j \in \{1, \ldots, n\}$ satisfying $\pi^{(x)}(j) = u$.

The proof of Theorem 9.4 follows the same procedure as above, with the following differences: at each stage we extend the ultrametric ρ_k from $X \smallsetminus \bigcup_{j=0}^{k-1} S_j$ to X using Lemma 9.3, and we replace the Harel-Tarjan data structure by a new data structure that is custom-made for the approximate ranking problem. The details are contained in [127].

9.4. **Random walks and quantitative nonembeddability.** While Ball introduced the notion of Markov type in order to investigate the Lipschitz extension problem, this notion has proved to be a versatile tool for the purpose of proving nonembeddability results. The use of Markov type in the context of embedding problems was introduced in [112], and this method has been subsequently developed in [22, 149, 13, 147, 148]. Somewhat curiously, Markov type can also be used as a tool to prove Lipschitz non-extendability results; see [148]. Markov type is therefore a good example of the impact of ideas originating in the Ribe program on metric geometry.

In this section we illustrate how one can use the notion of Markov type to reason that certain metric spaces must be significantly distorted in any embedding into certain Banach spaces. Since our goal here is to explain in the simplest possible terms this way of thinking about nonembeddability, we will mostly deal with model problems, which might not necessarily be the most general/difficult/important problems of this type. For example, we will almost always state our results for embeddings into Hilbert space, though it will be obvious how to extend our statements to general target spaces with Markov type $p \in (1, \infty)$. Also, we will present proofs in the case of finite graphs with large girth. While these geometric objects are somewhat exotic, they serve as a suitable model case for other spaces of interest, to which Markov type techniques also apply (e.g., certain Cayley graphs, including the discrete hypercube), since the large girth assumption simplifies the arguments, while preserving the essential ideas. We stress, however, that finite graphs with large girth are interesting geometric objects in their own right. Their existence is established with essentially complete freedom in the choice of certain governing parameters (such as the girth and degree; see [172]), yet understanding their geometry is difficult: this is illustrated by the fact that several basic problems on the embeddability properties of such graphs remain open. We will present some of these open problems later.

Fix an integer $k \geqslant 3$. Let $G = (V, E)$ be an n-vertex k-regular connected graph, equipped with its associated shortest path metric d_G. Let g be the girth of G, i.e., the length of the shortest closed cycle in G. Fix an integer $r < \frac{g}{4}$. For any ball B of radius r in G, the metric space (B, d_G) is isometric to $\left(T_r^k, d_{T_r^k}\right)$ (the tree T_r^k is defined in Section 5); see Figure 5. Thus Bourgain's lower bound (48) implies that

$$c_{\ell_2}(G) \gtrsim \sqrt{\log g}. \tag{57}$$

Can we do better than (57)? It seems reasonable to expect that we should be able to say more about the geometry of G than that it contains a large tree. When one tries to imagine what does a finite graph with large girth look like, one quickly realizes that it must be a complicated object: while it is true that small enough balls in such a graph are trees, these local trees must somehow be glued together to create a finite k-regular graph. It seems natural to expect that the interaction between these local trees induces a geometry which is far more complicated than what is suggested by the lower bound (57). This question was raised in 1995 by Linial, London and Rabinovich [111]. Our ultimate goal is to argue that *all* large enough subsets of G must be significantly distorted when embedded into Hilbert space, but as a warmup we will start with an argument of [112] which shows how the fact that Hilbert space has Markov type 2 easily implies the following exponential improvement to (57):

$$c_{\ell_2}(G) \gtrsim \sqrt{g}. \tag{58}$$

To prove (58) we shall use the fact that G has large girth as follows: it isn't only the case that G contains large trees, in fact *every* small enough ball in G is isometric to a tree. This information can be harnessed to our advantage as follows. Let $\{Z_t\}_{t=0}^{\infty}$ be the standard random walk on G, i.e., Z_0 is uniformly distributed on V and Z_{t+1} conditioned on Z_t is uniformly distributed on the k-neighbors of Z_t. Then $\{Z_t\}_{t=0}^{\infty}$ is a stationary reversible Markov chain on V. We claim that for every $t < \frac{g}{2} - 1$ we have

$$\mathbb{E}\left[d_G(Z_t, Z_0)\right] \gtrsim t. \tag{59}$$

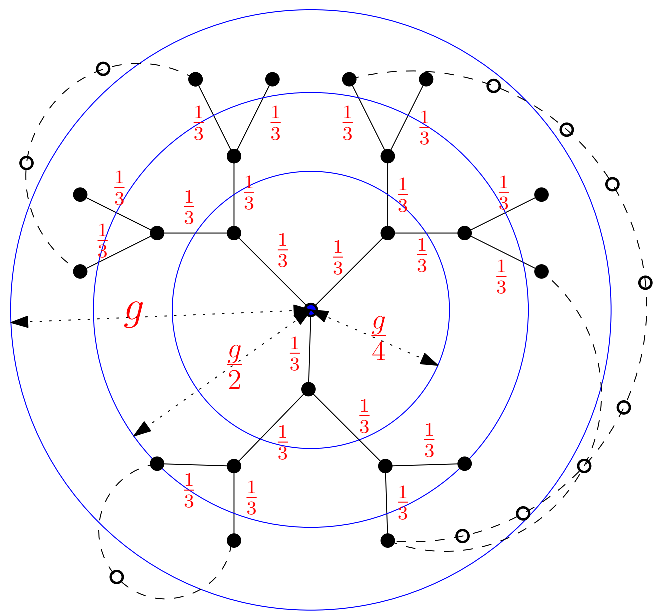

FIGURE 5. A 3-regular graph with girth g. Balls of radius $\frac{g}{2}$ look like trees in the sense that the distance of any vertex to the center of the ball is the same as the corresponding distance in a 3-regular tree rooted at the center of the ball. Any ball of radius $\frac{g}{4}$ is isometric to a 3-regular tree. If we pick the center of the ball uniformly at random, and then perform a standard random walk, then up to time $< \frac{g}{2}$, at each step there is probability $\frac{2}{3}$ to step further away from the center in the next step.

The proof of (59) is simple. Z_0 is chosen uniformly among the vertices of G. But, once Z_0 has been chosen, the walk $\left\{Z_s : \ s < \frac{g}{2} - 1\right\}$ is simply the standard walk on a k-regular tree starting from its root.

At each step of this walk, if $Z_s \neq Z_0$ then with probability $1-\frac{1}{k}$ the vertex Z_{s+1} is one of the $k-1$ neighbors of Z_s which are further away from Z_0 than Z_s, and with probability $\frac{1}{k}$ the vertex Z_{s+1} is the unique neighbor of Z_s that lies on the (unique) path joining Z_s and Z_0. If it happens to be the case that $Z_s = Z_0$, then Z_{s+1} is further away from Z_0 than Z_s with probability 1. Since $1-\frac{1}{k} > \frac{1}{k}$, we see that even though $\left\{Z_s : \ s < \frac{g}{2}-1\right\}$ is a stationary reversible Markov chain, in terms of the distance from Z_0 it is effectively a one dimensional random walk with positive drift, implying the required lower bound (59).

Suppose that $f : V \to L_2$ satisfies

$$\forall\, x,y \in V, \quad d_G(x,y) \leqslant \|f(x)-f(y)\|_2 \leqslant D d_G(x,y). \tag{60}$$

Our goal is to bound D from below. The fact that Hilbert space has Markov type 2 implies that for all times $t < \frac{g}{2}-1$ we have

$$\begin{aligned} t^2 &\overset{(59)}{\lesssim} \left(\mathbb{E}\left[d_G(Z_t,Z_0)\right]\right)^2 \leqslant \mathbb{E}\left[d_G(Z_t,Z_0)^2\right] \overset{(60)}{\leqslant} \mathbb{E}\left[\|f(Z_t)-f(Z_0)\|_2^2\right] \\ &\overset{(40)}{\leqslant} t\mathbb{E}\left[\|f(Z_1)-f(Z_0)\|_2^2\right] \overset{(60)}{\leqslant} tD^2\mathbb{E}\left[d_G(Z_1,Z_0)^2\right] = tD^2. \end{aligned} \tag{61}$$

Taking $t \asymp g$ in (61) yields (58).

The above argument can be extended to the case when G is not necessarily a regular graph. All we need is that the *average degree* of G is greater than 2. Recall that the average degree of G is

$$\frac{1}{|V|}\sum_{x\in V} \deg_G(x) = \frac{2|E|}{|V|},$$

where $\deg_G(x)$ denotes the number of edges in E emanating from x. Since we will soon be forced to deal with graphs of large girth which are not necessarily regular, we record here the following lemma from [22]:

Lemma 9.5. *Let $G=(V,E)$ be a connected graph with girth g and average degree k. Then*

$$c_{\ell_2}(G) \gtrsim \left(1-\frac{2}{k}\right)\sqrt{g}. \tag{62}$$

The proof of Lemma 9.5 follows the lines of the above proof of (58), with the following changes. For $x \in V$ define

$$\pi(x) = \frac{\deg_G(x)}{\sum_{y\in V}\deg_G(y)} = \frac{\deg_G(x)}{2|E|}. \tag{63}$$

Now, let $\{Z_t\}_{t=0}^\infty$ be the standard random walk on G, where Z_0 is distributed on V according to the probability distribution π. Then $\{Z_t\}_{t=0}^\infty$ is a stationary reversible Markov chain on V, so that the Markov type

2 inequality still applies to it. A short computation now yields (62); the details are contained in Theorem 6.1 of [22].

This type of use of random walks is quite flexible. For example, consider the case of the Hamming cube $(\{-1,1\}^n, \|\cdot\|_1)$. Let $\{Z_t\}_{t=0}^\infty$ be the standard random walk on $\{-1,1\}^n$, where Z_0 is distributed uniformly on $\{-1,1\}^n$. At each step, one of the n coordinates of Z_t is chosen uniformly at random, and its sign is flipped. For $t < \frac{n}{3}$ we have $\mathbb{E}[\|Z_t - Z_0\|_1] \gtrsim t$, since at each step with probability at least $\frac{2}{3}$ the coordinate being flipped has not been flipped in any previous step of the walk, and the walk therefore went further away from its starting point, whereas the probability that it got closer to its starting point is at most $\frac{1}{3}$. As we have argued above, this implies that $c_{\ell_2}(\{-1,1\}^n) \gtrsim \sqrt{n}$. This lower bound is sharp up to the implied multiplicative constant; in fact, a classical result of Enflo [49] states that $c_{\ell_2}(\{-1,1\}^n) = \sqrt{n}$. Enflo's proof of this fact uses a tensorization argument (i.e., induction on dimension while relying on the product structure of the Hamming cube). Another proof [99] of Enflo's theorem can be deduced from a Fourier analytic argument (both known proofs of the equality $c_{\ell_2}(\{-1,1\}^n) = \sqrt{n}$ are nicely explained in the book [118]). These proofs rely heavily on the structure of the Hamming cube, while, as we shall see below, the random walk proof that we presented here is more robust: e.g. it applies to negligibly small subsets of the Hamming cube which may be highly unstructured.

Before passing to a more sophisticated application of Markov type, we recall the following interesting open question [112].

Question 7. *Let $c_2(g)$ be the infimum of $c_{\ell_2}(G)$ over all finite 3-regular connected graphs G with girth g. What is the growth rate of $c_2(g)$ as $g \to \infty$? In particular, does $c_2(g)$ grow asymptotically faster than $\sqrt{g}$?*

In order to prove that $\lim_{g\to\infty} c_2(g)/\sqrt{g} = \infty$ (if true), we would need to use more about the structure of G than the fact that a ball of radius $\asymp g$ around each vertex is isometric to a 3-regular tree. One would need to understand the complicated regime in which these local trees interact. Our understanding of the geometry of these interactions is currently quite poor, which is why Question 7 is meaningful. On the other hand, if for arbitrarily large $g \in \mathbb{N}$ there were 3-regular graphs G of girth g with $c_{\ell_2}(G) \lesssim \sqrt{g}$, this would also have interesting consequences, as explained in [112]. Note that one could also ask a variant of Question 7, when g depends on the cardinality of V. The case $g \asymp \log|V|$ is of particular importance (see [112]).

Letting $c_1(g)$ denote the infimum of $c_{\ell_1}(G)$ over all finite 3-regular connected graphs G with girth g, it was also asked in [112] whether or not $c_1(g)$ tends to ∞ with g. This question was recently solved by Ostrovskii [158], who showed that for arbitrarily large $n \in \mathbb{N}$ there exists a 3-regular graph G_n of girth at least a constant multiple of $\log\log n$ yet $c_{\ell_1}(G_n) = O(1)$. Since trees admit an isometric embedding into ℓ_1, such questions address the issue of how the local geometry of a metric space affects its global geometry (see [7, 39, 100, 167] for related investigations along these lines). It remains an interesting open question whether there exist arbitrarily large graphs of logarithmic girth that admit a bi-Lipschitz embedding into ℓ_1; see [112] for ramifications of this question.

9.4.1. *Impossibility results for nonlinear Dvoretzky problems.* Our goal here is to explain the relevance of Markov type techniques to proving impossibility results for nonlinear Dvoretzky problems, i.e., to show that certain metric spaces cannot have large subsets that well-embed into Hilbert space (or a metric space with nontrivial Markov type). Everything presented here is part of the investigation in [22] of the nonlinear Dvoretzky problem in concrete examples. Additional results of this type are contained in [22].

We have already seen that $c_{\ell_2}\left(\{-1,1\}^n\right) \gtrsim \sqrt{n}$. Assume now that we are given a subset $S \subseteq \{-1,1\}^n$. If we only knew that the cardinality of S is large, would it then be possible to show that $c_{\ell_2}(S)$ is also large? It is not clear how to proceed if $|S| = o(2^n)$ (this isn't clear even when $|S|$ is, say, one tenth of the cube). The random walk technique turns out to be robust enough to yield almost sharp bounds on the Euclidean distortion of a large subset of the Hamming cube, without any a priori assumption on the structure of the subset. Namely, it was proved in [22] that for every $S \subseteq \{-1,1\}^n$ we have

$$c_{\ell_2}(S) \gtrsim \sqrt{\frac{n}{1+\log\left(\frac{2^n}{|S|}\right)}}. \tag{64}$$

Thus, in particular, if $|S| = 2^{n(1-\varepsilon)} = |\{-1,1\}^n|^{1-\varepsilon}$, then (64) becomes

$$c_{\ell_2}(S) \gtrsim \min\left\{\frac{1}{\sqrt{\varepsilon}}, \sqrt{n}\right\}.$$

This bound is tight up to logarithmic factors: it was shown in [22] that for every $\varepsilon \in (0,1)$ there exists $S \subseteq \{-1,1\}^n$ with $|S| \geqslant 2^{n(1-\varepsilon)}$ and

$$c_{\ell_2}(S) \lesssim \sqrt{\frac{1+\log(1/\varepsilon)}{\varepsilon}}.$$

The proof of (64) uses Markov type in a crucial way. Here, in order to illustrate the main ideas, we will deal with the analogous problem for subsets of graphs with large girth. Namely, let $G = (V, E)$ be a finite k-regular ($k \geqslant 3$) connected graph with girth g. Assume that $S \subseteq V$ is equipped with the metric d_G inherited from G. We will prove the following lower bound on $c_{\ell_2}(S)$, which is also due to [22]:

$$c_{\ell_2}(S) \gtrsim \sqrt{\frac{g}{1 + \log_k\left(\frac{|V|}{|S|}\right)}}. \tag{65}$$

Note that when $S = V$ we return to (58), but the proof of (65) is more subtle than the proof of (58). This proof uses more heavily the fact that in (40) we are free to choose the stationary reversible Markov chain as we wish. Our plan is to construct a special stationary reversible Markov chain on S, which in conjunction with the Markov type 2 property of Hilbert space, will establish (65).

Ideally, we would like our Markov chain to be something like the standard random walk on G, restricted to S. Lemma 9.5 indicates that for this approach to work we need S to have large average degree, or equivalently to contain many edges of G. But, S might be very small, and need not contain any edge of G. We will overcome this problem by considering a different set of edges E' on V, which is nevertheless closely related to the geometry of G, such that S contains sufficiently many edges from E'. Before proceeding to carry out this plan, we therefore need to make a small digression which explains a spectral method for showing that a subset of a graph contains many edges.

9.4.2. λ_n *and self mixing.* Let $H = (\{1, \ldots, n\}, E_H)$ be a d-regular loop-free graph on $\{1, \ldots, n\}$. We denote by $A_H = (a_{ij})$ its adjacency matrix, i.e., the $n \times n$ matrix whose entries are in $\{0, 1\}$, and $a_{ij} = 1$ if and only if $ij \in E_H$. Let $\lambda_1(H) \geqslant \lambda_2(H) \geqslant \cdots \geqslant \lambda_n(H)$ be the eigenvalues of A_H. Thus $\lambda_1(H) = d$, and since the diagonal entries of H vanish, $\mathrm{trace}(A_H) = \sum_{i=1}^n \lambda_i(H) = 0$. In particular we are ensured that $\lambda_n(H)$ is negative.

Let $\{v_1, \ldots, v_n\}$ be an eigenbasis of A_H, which is orthonormal with respect to the standard scalar product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$. We can choose the labeling so that $v_1 = \frac{1}{\sqrt{n}} \mathbf{1}_{\{1,\ldots,n\}}$, and the eigenvalue corresponding to v_i is $\lambda_i(H)$. For every $S \subseteq \{1, \ldots, n\}$ let $E_H(S)$ denote the number of edges in E_H that are incident to two vertices in S. Observe that

$$\langle A_H \mathbf{1}_S, \mathbf{1}_S\rangle = \sum_{i=1}^n \lambda_i(H)\langle v_i, \mathbf{1}_S\rangle^2 = \frac{d|S|^2}{n} + \sum_{i=2}^n \lambda_i(H)\langle v_i, \mathbf{1}_S\rangle^2$$
$$\geqslant \frac{d|S|^2}{n} + \lambda_n(H)\sum_{i=2}^n \langle v_i, \mathbf{1}_S\rangle^2 = \frac{d|S|^2}{n} + \lambda_n(H)\left(|S| - \frac{|S|^2}{n}\right).$$

Thus, since $\lambda_n(H) < 0$,

$$2E_H(S) = \langle A_H \mathbf{1}_S, \mathbf{1}_S\rangle \geqslant \frac{d|S|^2}{n} + \lambda_n(H)|S|. \tag{66}$$

We can use (66) to deduce that $E_H(S)$ is large provided that $\lambda_n(H)$ is not too negative (in [22] such a bound is called a *self mixing inequality*). The bound in (66) is perhaps less familiar than Cheeger's inequality [40, 3], which relates the number of edges joining S and its complement to $\lambda_2(H)$, but these two inequalities are the same in spirit. We refer to the survey [79] for more information on the connection between the second largest eigenvalue and graph expansion. While bounds on $\lambda_2(H)$ would have been very useful for us to have in the ensuing argument to prove (65) (and the corresponding proof of (64) in [22]), we will only obtain bounds on $|\lambda_n(H)|$ (for an appropriately chosen graph H), which will nevertheless suffice for our purposes.

9.4.3. *The spectral argument in the case of large girth.* Returning to the proof of (66), let $G = (V, E)$ be an n-vertex k-regular connected graph ($k \geqslant 3$) with girth g. We assume throughout that G is loop-free and contains no multiple edges. As before, the shortest path metric on G is denoted by d_G. Fix $m \in \mathbb{N}$ and let $G^{(m)} = (V, E_{G^{(m)}})$ denote the distance m graph of G, i.e., the graph on V in which two vertices $u, v \in V$ are joined by an edge if and only if $d_G(u, v) = m$.

Recall that $A_{G^{(m)}}$ denotes the adjacency matrix of $G^{(m)}$. Thus we have $A_{G^{(0)}} = I_V$ (the identity matrix on V) and $A_{G^{(1)}} = A_G$. Moreover, $A_G^2 = kI_V + A_{G^{(2)}}$, and $A_G A_{G^{(m-1)}} = (k-1)A_{G^{(m-2)}} + A_{G^{(m)}}$ for all $2 < m < \frac{g}{2}$. Indeed, write $(A_G A_{G^{(m-1)}})_{uv} = \sum_{w\in V} (A_G)_{uw} (A_{G^{(m-1)}})_{wv}$ for all $u, v \in V$. There are only two types of possible contributions to this sum: either $d_G(u, v) = m$ and w is on the unique path joining u and v such that $uw \in E$, or $d_G(u, v) = m - 2$ and w is one of the neighbors of v which is not on the path joining u and v (the number of such w equals k if $m = 2$, and equals $k - 1$ if $m > 2$).

The above discussion shows that if we define a sequence of polynomials $\{P_m^k(x)\}_{m=0}^\infty$ by

$$P_0^k(x) = 1, \quad P_1^k(x) = x, \quad P_2^k(x) = x^2 - k, \tag{67}$$

and recursively,

$$P_m^k(x) = xP_{m-1}^k(x) - (k-1)P_{m-2}^k(x), \tag{68}$$

then for all integers $0 \leqslant m < \frac{g}{2}$,

$$A_{G^{(m)}} = P_m^k\left(A_G\right). \tag{69}$$

The polynomials $\{P_m^k(x)\}_{m=0}^{\infty}$ are known as the Geronimus polynomials (see [176] and the references therein). By (69), when $m < \frac{g}{2}$ the eigenvalues of $A_{G^{(m)}}$ are $\left\{P_m^k\left(\lambda_i(A_G)\right)\right\}_{i=1}^n$. For the purpose of bounding the negative number $\lambda_n\left(A_{G^{(m)}}\right)$ from below, it therefore suffices to use the bound

$$\lambda_n\left(A_{G^{(m)}}\right) \geqslant \min_{x\in\mathbb{R}} P_m^k(x). \tag{70}$$

A simple induction shows that $P_m^k(x)$ is a polynomial of degree m with leading coefficient 1, and it is an even function for even m, and an odd function for odd m. Moreover, we have the following trigonometric identity (see [176]):

$$P_m^k\left(2\sqrt{k-1}\cos\vartheta\right) = (k-1)^{\frac{m}{2}-1}\cdot\frac{(k-1)\sin((m+1)\vartheta)-\sin((m-1)\vartheta)}{\sin\vartheta}. \tag{71}$$

The proof of (71) is a straightforward induction: check the validity of (71) for $m = 1, 2$ using (67), and verify by induction that (71) holds using the recursion (68).

Define $\vartheta_q = \frac{\frac{\pi}{2}+q\pi}{m+1}$. For every $q \in \{0,\ldots,m\}$ we have $\vartheta_q \in (0,\pi)$, and the sign of $P_m^k\left(2\sqrt{k-1}\cos\vartheta_q\right)$ is equal to the sign of

$$(k-1)\sin((m+1)\vartheta_q) - \sin((m-1)\vartheta_q) = (-1)^q(k-1) - \sin\left(\frac{m-1}{m+1}\left(\frac{\pi}{2}+q\pi\right)\right).$$

Thus for every $q \in \{0,\ldots,m\}$ the value $P_m^k\left(2\sqrt{k-1}\cos\vartheta_q\right)$ is positive if q is even, and negative if q is odd. It follows that P_m^k must have a zero in each of the m intervals $\left\{\left[2\sqrt{k-1}\cos\vartheta_q, 2\sqrt{k-1}\cos\vartheta_{q+1}\right]\right\}_{q=0}^{m-1}$. Since P_m^k is a polynomial of degree m, we deduce that the zeros of P_m^k are contained in the interval $\left[-2\sqrt{k-1}, 2\sqrt{k-1}\right]$. In particular, if m is even, since $P_m^k(x)$ is an even function which tends to ∞ as $x \to \infty$, it can take negative values only in the interval $\left[-2\sqrt{k-1}, 2\sqrt{k-1}\right]$. See Figure 6 for the case $k = 3$, $m = 8$.

It follows from the above discussion, combined with (70) and (71), that for every even integer $0 < m < \frac{g}{2}$ we have

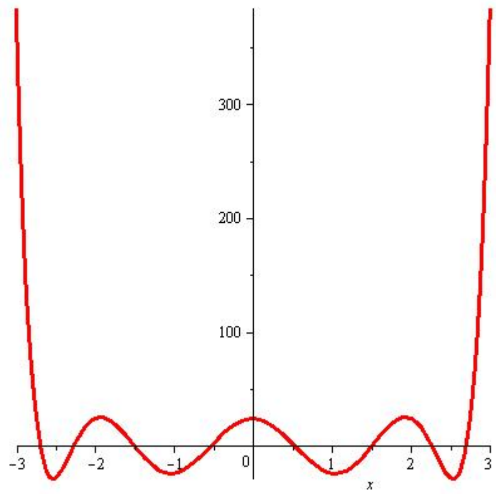

FIGURE 6. A plot of the polynomial
$$P_8^3(x) = x^8 - 15x^6 + 70x^4 - 104x^2 + 24.$$

$$\begin{aligned}
&\lambda_n\left(A_{G^{(m)}}\right)\\
\geqslant\ & (k-1)^{\frac{m}{2}-1} \min_{\vartheta\in[-\pi,\pi]} \frac{(k-1)\sin((m+1)\vartheta)-\sin((m-1)\vartheta)}{\sin\vartheta}\\
=\ & (k-1)^{\frac{m}{2}-1} \min_{\vartheta\in[-\pi,\pi]} \left((k-1)e^{-m\vartheta i}\sum_{r=0}^{m} e^{2\vartheta r i} - e^{-(m-2)\vartheta i}\sum_{r=0}^{m-2} e^{2\vartheta r i}\right)\\
\geqslant\ & -(k-1)^{\frac{m}{2}-1}\left((k-1)(m+1)+m-1\right)\\
\geqslant\ & -(k-1)^{\frac{m}{2}-1}k(m+1). \qquad (72)
\end{aligned}$$

Since the degree of $G^{(m)}$ is $k(k-1)^{m-1}$, the following corollary is a combination of (66) and (72).

Corollary 9.6. *Let $G=(V,E)$ be a k-regular graph with girth g. Then for all even integers $0<m<\frac{g}{2}$ and for all $S\subseteq V$, the average degree in the graph induced by $G^{(m)}$ on S satisfies*

$$\frac{2E_{G^{(m)}}(S)}{|S|} \geqslant \frac{|S|}{n}k(k-1)^{m-1} - (k-1)^{\frac{m}{2}-1}k(m+1).$$

In particular, if

$$\frac{|S|}{n} \geqslant \frac{2m+2}{(k-1)^{\frac{m}{2}}}, \qquad (73)$$

then

$$\frac{2E_{G^{(m)}}(S)}{|S|} \geqslant k(k-1)^{m-1}\frac{|S|}{2n}. \tag{74}$$

9.4.4. *Completion of the proof of* (65). Corollary 9.6, in combination with Lemma 9.5, suggests that we should consider the stationary reversible random walk on the graph induced by $G^{(m)}$ on S. We will indeed do so, and by judiciously choosing m, (65) will follow.

For each $v \in S$ we denote by $\deg_{G^{(m)}[S]}(v)$ its degree in the graph induced by $G^{(m)}$ on S, i.e., the number of vertices $u \in S$ that are at distance m from v, where the distance is measured according to the original shortest path metric on G. As in (63), for $v \in S$ we write

$$\pi(v) = \frac{\deg_{G^{(m)}[S]}(v)}{2E_{G^{(m)}}(S)}. \tag{75}$$

Let $\{Z_t\}_{t=0}^{\infty}$ be the following Markov chain on S: Z_0 is distributed according to π, and Z_{t+1} is distributed uniformly on the $\deg_{G^{(m)}[S]}(Z_t)$ vertices of S at distance m from Z_t (note that $\deg_{G^{(m)}[S]}(Z_t) > 0$, since Z_t is distributed only on those $v \in S$ for which $\pi(v) > 0$).

At time $t \in \mathbb{N}$ we clearly have $d_G(Z_0, Z_t) \leqslant tm$. In order to remain in the local "tree range", we will therefore impose the assumption

$$tm < \frac{g}{4}. \tag{76}$$

Assume from now on that m is divisible by 6. We first observe that for t as in (76), the number of neighbors $w \in V$ of Z_{t-1} in the graph $G^{(m)}$ which satisfy $d_G(w, Z_0) < d_G(Z_0, Z_{t-1}) + \frac{m}{3}$ is at most $(k-1)^{\frac{2m}{3}-1}$. Indeed, we may assume that $d_G(Z_0, Z_{t-1}) > \frac{m}{3}$, since otherwise for any such w we have

$$\begin{aligned} d_G(w, Z_0) &\geqslant d_G(w, Z_{t-1}) - d_G(Z_0, Z_{t-1}) \\ &= m - d_G(Z_0, Z_{t-1}) \geqslant d_G(Z_0, Z_{t-1}) + \frac{m}{3}. \end{aligned}$$

So, assuming $d_G(Z_0, Z_{t-1}) > \frac{m}{3}$ and $d_G(w, Z_0) < d_G(Z_0, Z_{t-1}) + \frac{m}{3}$, let v be the point on the unique path joining Z_0 and Z_{t-1} such that $d_G(v, Z_{t-1}) = \frac{m}{3} + 1$. The path in G (whose length is m) joining w and Z_{t-1} must pass through v. See Figure 7 for an explanation of this simple fact. Note that $d_G(w, v) = \frac{2m}{3} - 1$, and hence the number of such w is at most $(k-1)^{\frac{2m}{3}-1}$.

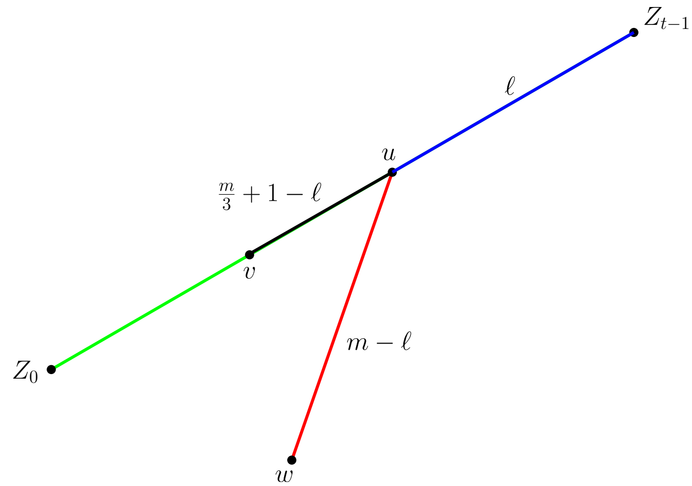

FIGURE 7. If the path joining w and Z_{t-1} does not pass through v then it must touch the path joining Z_0 and Z_{t-1} at a vertex u, as depicted above. Denoting $\ell = d_G(u, Z_{t-1})$, we have $\ell \leqslant \frac{m}{3}$. Hence, since $d_G(w, Z_0) = (d_G(Z_0, Z_{t-1}) - \ell) + (m - \ell)$, we have $d_G(w, Z_0) \geqslant d_G(Z_0, Z_{t-1}) + \frac{m}{3}$.

Let $N(Z_{t-1})$ denote the number of $w \in S$ with $d_G(w, Z_{t-1}) = m$ and $d_G(w, Z_0) < d_G(Z_0, Z_{t-1}) + \frac{m}{3}$. Then,

$$\begin{aligned}
&\mathbb{E}\left[d_G(Z_0, Z_t)\right] \\
&\quad\geqslant \mathbb{E}\Bigg[\frac{\deg_{G^{(m)}[S]}(Z_{t-1}) - N(Z_{t-1})}{\deg_{G^{(m)}[S]}(Z_{t-1})}\left(d_G(Z_0, Z_{t-1}) + \frac{m}{3}\right) \\
&\qquad\qquad + \frac{N(Z_{t-1})}{\deg_{G^{(m)}[S]}(Z_{t-1})}\left(d_G(Z_0, Z_{t-1}) - m\right)\Bigg] \\
&\quad= \mathbb{E}\left[d_G(Z_0, Z_{t-1})\right] + \frac{m}{3} - \frac{4m}{3}\mathbb{E}\left[\frac{N(Z_{t-1})}{\deg_{G^{(m)}[S]}(Z_{t-1})}\right]. \qquad (77)
\end{aligned}$$

We will estimate the last term appearing in (77) via the point-wise bound $N(Z_{t-1}) \leqslant (k-1)^{\frac{2m}{3}-1}$ that we proved above, together with (74), for which we need to assume (73).

$$\mathbb{E}\left[\frac{N(Z_{t-1})}{\deg_{G^{(m)}[S]}(Z_{t-1})}\right] \leqslant (k-1)^{\frac{2m}{3}-1} \sum_{\substack{v\in S\\ \deg_{G^{(m)}[S]}(v)>0}} \frac{\pi(v)}{\deg_{G^{(m)}[S]}(v)}$$

$$\overset{(75)}{\leqslant} (k-1)^{\frac{2m}{3}-1}\frac{|S|}{2E_{G^{(m)}}(S)} \overset{(74)}{\leqslant} \frac{1}{k(k-1)^{\frac{m}{3}}}\cdot\frac{2n}{|S|}. \quad (78)$$

By combining (77) and (78) we get the bound

$$\mathbb{E}\left[d_G(Z_0,Z_t)\right] \geqslant \mathbb{E}\left[d_G(Z_0,Z_{t-1})\right] + \frac{m}{3} - \frac{8mn}{3k(k-1)^{\frac{m}{3}}|S|}$$
$$\geqslant \mathbb{E}\left[d_G(Z_0,Z_{t-1})\right] + \frac{m}{6}, \quad (79)$$

provided that

$$\frac{|S|}{n} \geqslant \frac{16}{k(k-1)^{\frac{m}{3}}}. \quad (80)$$

We can ensure that our restrictions on m, namely (73) and (80), are satisfied for some $m \asymp 1+\log_k(n/|S|)$ that is divisible by 6. For such a value of m, we know that (79) is valid as long as t satisfies (76). Thus, by iterating (79) we see that for some $t \asymp g/m$ we have

$$\mathbb{E}\left[d_G(Z_0,Z_t)^2\right] \geqslant \left(\mathbb{E}\left[d_G(Z_0,Z_t)\right]\right)^2 \gtrsim (tm)^2 \gtrsim g^2. \quad (81)$$

If $f: S \to \ell_2$ satisfies

$$d_G(x,y) \leqslant \|f(x)-f(y)\|_2 \leqslant D d_G(x,y) \quad (82)$$

for all $x,y\in S$, then it follows from the Markov type 2 property of Hilbert space that

$$g^2 \overset{(81)}{\lesssim} \mathbb{E}\left[d_G(Z_0,Z_t)^2\right] \overset{(82)\wedge(40)}{\leqslant} t\mathbb{E}\left[\|f(Z_1)-f(Z_0)\|_2^2\right]$$
$$\overset{(82)}{\leqslant} tD^2\mathbb{E}\left[d_G(Z_0,Z_1)^2\right] = D^2tm^2 \asymp D^2 g\left(1+\log_k\left(\frac{n}{|S|}\right)\right).$$

This completes the proof of (65). □

9.4.5. *Discrete groups.* Let G be an infinite group which is generated by a finite symmetric subset $S=S^{-1}\subseteq G$. Let d_S denote the left invariant word metric induced by S on G, i.e., $d_S(x,y)$ is the smallest integer $k\geqslant 0$ such that there exist $s_1,\ldots,s_k\in S$ with $x^{-1}y=s_1s_2\cdots s_k$. It has long been established that it is fruitful to study finitely generated groups as geometric objects, i.e., as metric spaces when equipped with a word metric (see [68, 46, 42] and the references therein for an indication of the large amount of literature on this topic). Here we will describe the role of Markov type in this context.

Assume that the metric space (G, d_S) does not admit a bi-Lipschitz embedding into Hilbert space, i.e., $c_{\ell_2}(G, d_S) = \infty$. Based on the experience of researchers thus far, this assumption is not restrictive: it is conjectured in [45] that if (G, d_S) does admit a bi-Lipschitz embedding into Hilbert space then G has an Abelian subgroup of finite index. Fix a mapping $f : G \to \ell_2$. Note that if f is not a Lipschitz function then the mapping $x \mapsto \max_{s \in S} \|f(xs) - f(x)\|_2$ must be unbounded on G. If we consider only mappings f which have bounded displacement on edges of the Cayley graph induced by S on G, then the fact that $c_{\ell_2}(G, d_S) = \infty$ must mean that if we set $\omega_f(t) = \inf_{d_S(x,y) \geqslant t} \|f(x) - f(y)\|_2$ then $\omega_f(t) = o(t)$ as $t \to \infty$. To see this consider the mapping $\psi : G \to \ell_2 \oplus \ell_2(G) \cong \ell_2$ given by $\psi(x) = f(x) \oplus \delta_x$. The fact ψ has infinite distortion implies that f must asymptotically compress arbitrarily large distances in G.

The modulus $\omega_f(t)$ is called the compression function of f. If we manage to show that for any Lipschitz function $f : G \to \ell_2$ the rate at which $\omega_f(t)/t$ tends to zero must be "fast", then we might deduce valuable structural information on the group G. This general approach (including the terminology that we are using) is due to Gromov (see Section 7.E in [68]). Here we will study a further refinement of this idea, which will yield a numerical invariant of infinite groups called the compression exponent. This elegant definition is due to Guentner and Kaminker [72], and it was extensively studied in recent years (see the introduction to [148] for background and references). We will focus here on the use of random walk techniques in the task of computing (or estimating) this invariant.

The Guentner-Kaminker definition is simple to state. Given a metric space (Y, d_Y), the Y-compression exponent of G, denoted $\alpha_Y^*(G)$, is the supremum of those $\alpha \geqslant 0$ for which there exists a Lipschitz function $f : G \to Y$ which satisfies $d_Y(f(x), f(y)) \gtrsim d_S(x, y)^\alpha$ for all $x, y \in X$. We remark that in the notation $\alpha_Y^*(G)$ we dropped the explicit reference to the generating set S. This is legitimate since, if we switch to a different finite symmetric generating set $S' \subseteq G$, then the resulting word metric $d_{S'}$ is bi-Lipschitz equivalent to the original word metric d_S, and therefore the Y-compression exponents of S and S' coincide. In other words, the number $\alpha_Y^*(G) \in [0, 1]$ is a true algebraic invariant of the group G, which does not depend on the particular choice of a finite symmetric set of generators. The parameter $\alpha_{\ell_2}^*(G)$ is called the Hilbert compression exponent of G. It was shown in [8] that any

$\alpha \in [0,1]$ is the Hilbert compression exponent of some finitely generated group G (see [11, 155] for the related question for amenable groups). Nevertheless, there are relatively few concrete examples of groups G for which $\alpha^*_{\ell_2}(G)$ (and $\alpha^*_{\ell_p}(G)$) has been computed. We will demonstrate how Markov type is relevant to the problem of estimating $\alpha^*_Y(G)$. This approach was introduced in [13], and further refined in [147, 148].

We will examine the applicability of random walks to the computation of compression exponents of discrete groups via an illustrative example: the wreath product of the group of integers $\mathbb{Z}$ with itself. Before doing so, we recall for the sake of completeness the definition of the wreath product of two general groups G, H. Readers who are not accustomed to this concept are encouraged to focus on the case $G = H = \mathbb{Z}$, as it contains the essential ideas that we wish to convey.

Let G, H be groups which are generated by the finite symmetric sets $S_G \subseteq G$, $S_H \subseteq H$. We denote by e_G, e_H the identity elements of G, H, respectively. We also denote by e_{G^H} the function from H to G which takes the value e_G at all points $x \in H$. The (restricted) wreath product of G with H, denoted $G \wr H$, is defined as the group of all pairs (f, x) where $f : H \to G$ has finite support (i.e., $f(z) = e_{G^H}(z) = e_G$ for all but finitely many $z \in H$) and $x \in H$, equipped with the product

$$(f,x)(g,y) = \left(z \mapsto f(z)g(x^{-1}z), xy\right).$$

$G \wr H$ is generated by the set $\{(e_{G^H}, x) : x \in S_H\} \cup \{(\delta_y, e_H) : y \in S_G\}$, where $\delta_y : H \to G$ is the function which takes the value y at e_H and the value e_G on $H \smallsetminus \{e_H\}$.

When $G = C_2 = \{0,1\}$, the cyclic group of order 2, then the group $C_2 \wr H$ is often called the lamplighter group on H. In this case imagine that at every site $x \in H$ there is a lamp, which can either be on or off. An element $(f, x) \in C_2 \wr H$ can be thought of as indicating that a "lamplighter" is located at $x \in H$, and f represents the locations of those (finitely many) lamps which are on (these locations are the sites $y \in H$ where $f(y) = 1$). The distance in $C_2 \wr H$ between (f, x) and (g, y) is the minimum number of steps required for the lamplighter to start at x, visit all the sites $z \in H$ for which $f(z) \neq g(z)$, change $f(z)$ to $g(z)$, and end up at the site y. Here, by a "step" we mean a move from x to xs for some $s \in S_H$, or a change of the state of the lamp (from on to off or vice versa) at the current location of the lamplighter. Thus, the distance between (f, x) and (g, y) is, up to a factor of 2, the shortest (in the metric d_{S_H}) traveling salesman tour starting at x, covering the symmetric difference of the supports of f and g, and terminating at y. For a general group G, the description of the metric on $G \wr H$ is similar, the only difference being that the lamps can have G different states (not just on or off), and the cost of changing the state of a lamp from $a \in G$ to $b \in G$ is $d_{S_G}(a, b)$. See Figure 9.4.5 for a schematic description of the case $G = H = \mathbb{Z}$.

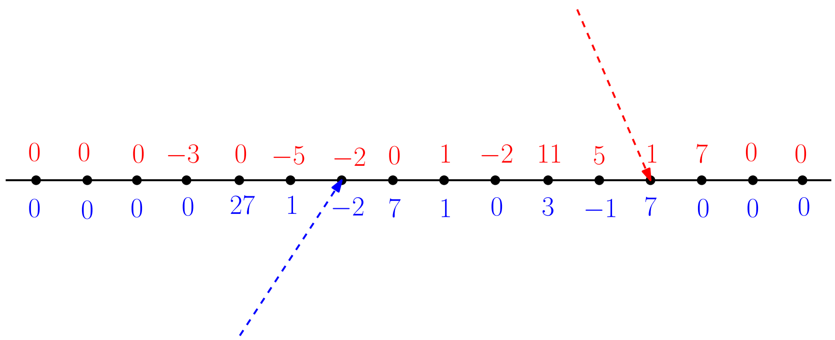

FIGURE 8. An example of two elements in $\mathbb{Z}\wr\mathbb{Z}$. The numbers above and below the $\mathbb{Z}$-axis represent two $\mathbb{Z}$-valued finitely supported functions. The arrows indicate the location of the lamplighter in each of the corresponding elements of $\mathbb{Z}\wr\mathbb{Z}$. In order to compute the distance between these elements, the lamplighter of the top configuration must visit the locations where the top values differ from the bottom values, and at these locations the top value must be changed to the bottom value. At the end of the process, the lamplighter must end up at the location indicated by the bottom arrow. Each movement of the lamplighter to a neighboring integer adds a unit cost to this process, and an increment or decrement of 1 to the value at the location of the lamplighter also incurs a unit cost. The distance in $\mathbb{Z}\wr\mathbb{Z}$ is the minimum cost of such a process which transforms the top configuration to the bottom configuration. Thus, in the above example, the top lamplighter will first move one step to the right, incurring a unit cost, change the 7 in the top row to 0, incurring a cost of 7 units, take one step to the left (incurring a unit cost), change the 1 in the top row to a 7 (incurring a cost of 6 units), and so on.

We shall now describe an argument using random walks, showing that $\alpha^*_{\ell_2}(\mathbb{Z}\wr\mathbb{Z}) \leqslant \frac{2}{3}$. This approach is due to [13]. In fact, as shown in [147], $\alpha^*_{\ell_2}(\mathbb{Z}\wr\mathbb{Z}) \geqslant \frac{2}{3}$, and therefore the argument below is sharp, and yields the exact computation $\alpha^*_{\ell_2}(\mathbb{Z}\wr\mathbb{Z}) = \frac{2}{3}$. More generally, it is shown in [148] that for every $p \in [1,2]$ we have

$$\alpha^*_{\ell_p}(\mathbb{Z}\wr\mathbb{Z}) = \frac{p}{2p-1}. \tag{83}$$

The proof of (83) when $p \neq 2$ requires an additional idea that we will not work out in detail here: instead of examining the standard random walk on $\mathbb{Z}\wr\mathbb{Z}$ one studies a discrete version of a q-stable random walk for every $q \in (p,2]$. This yields a new twist of the Markov type method: it is beneficial to adapt the random walk to the geometry of the target space, and to use random walks with unbounded increments (though, we have already seen the latter occur in Section 9.4.4). We refer to [147, 148] for more general results that go beyond that case of $\mathbb{Z}\wr\mathbb{Z}$, as well as an explanation of the background, history, and applications of these types of problems. It suffices to say here that we chose to focus on the group $\mathbb{Z}\wr\mathbb{Z}$ since before the introduction of random walk techniques, it was the simplest concrete group which resisted the attempts to compute its ℓ_p compression exponents.

Consider the standard random walk $\{W_t\}_{t=0}^\infty$ on $\mathbb{Z}\wr\mathbb{Z}$, starting at the identity element. Namely, we start at $e_{\mathbb{Z}\wr\mathbb{Z}}$, i.e., the configuration corresponding to all the lamps being turned off, and the lamplighter being at 0. At each step a fair coin is tossed, and depending on the outcome of the coin toss, either the lamplighter moves to one of its two neighboring locations uniformly at random, or the value at the current location of the lamplighter is changed by $+1$ or -1 uniformly at random.

After t steps, we expect that a constant fraction of the coin tosses resulted in a movement of the lamplighter, which is just a standard random walk on the integers $\mathbb{Z}$. Thus, at time t we expect the lamplighter to be located at $\pm \asymp \sqrt{t}$. One might also expect that during the walk the lamplighter spent roughly (up to constant factors) the same amount of total time at a definite fraction of the sites between 0 and its location at time t. There are $\asymp \sqrt{t}$ such sites, and therefore, if this intuition is indeed correct, we expect the time spent at each of these sites to be $\asymp \frac{t}{\sqrt{t}} = \sqrt{t}$. At each such site the value of the lamp is also the result of a random walk on $\mathbb{Z}$, and therefore at time t we expect W_t to have $\asymp \sqrt{t}$ sites at which the value of the lamp is $\pm \asymp \sqrt{\sqrt{t}} = \pm\sqrt[4]{t}$. This heuristic argument suggests that

$$\mathbb{E}\left[d_{\mathbb{Z}\wr\mathbb{Z}}\left(W_t, e_{\mathbb{Z}\wr\mathbb{Z}}\right)\right] \gtrsim \sqrt{t}\cdot\sqrt[4]{t} = t^{\frac{3}{4}}. \tag{84}$$

These considerations can indeed be made to yield a rigorous proof of (84); see [54] and also [168], as well as Section 6 in [147] for an extension to the case of general wreath products.

The fact that ℓ_2 has Markov type 2 suggests that if $f : \mathbb{Z}\wr\mathbb{Z} \to \ell_2$ satisfies $d_{\mathbb{Z}\wr\mathbb{Z}}(x,y)^\alpha \lesssim \|f(x)-f(y)\|_2 \lesssim d_{\mathbb{Z}\wr\mathbb{Z}}(x,y)$, then

$$\mathbb{E}\left[\|f(W_t)-f(W_0)\|_2^2\right] \lesssim t,$$

yet due to (84),

$$\mathbb{E}\left[\|f(W_t)-f(W_0)\|_2^2\right] \gtrsim t^{2\alpha\cdot\frac{3}{4}}.$$

This implies that $\alpha \leqslant \frac{2}{3}$, as required. But, this argument is flawed: we are only allowed to use the Markov type 2 inequality (40) for stationary reversible Markov chains. The Markov chain $\{W_t\}_{t=0}^\infty$ starts at the deterministic point $e_{\mathbb{Z}\wr\mathbb{Z}}$ rather than at a point chosen uniformly at random over $\mathbb{Z}\wr\mathbb{Z}$. Of course, since $\mathbb{Z}\wr\mathbb{Z}$ is an infinite set, there is no way to make W_0 be uniformly distributed over it. The above argument can be salvaged by either considering instead an appropriately truncated random walk starting at a uniformly chosen point from a large enough Følner set of $\mathbb{Z}\wr\mathbb{Z}$, or by applying an argument of Aharoni, Maurey and Mityagin [2] and Gromov [45] (see also [148]) to reduce the problem to equivariant embeddings, and then to prove that the Markov type inequality does hold true for images of the random walk $\{W_t\}_{t=0}^\infty$ (starting at $e_{\mathbb{Z}\wr\mathbb{Z}}$) under equivariant mappings. See [13] for the former approach and [147] for the latter approach.

References

[1] I. Aharoni and J. Lindenstrauss. Uniform equivalence between Banach spaces. *Bull. Amer. Math. Soc.*, 84(2):281–283, 1978.

[2] I. Aharoni, B. Maurey, and B. S. Mityagin. Uniform embeddings of metric spaces and of Banach spaces into Hilbert spaces. *Israel J. Math.*, 52(3):251–265, 1985.

[3] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[4] R. D. Anderson. Hilbert space is homeomorphic to the countable infinite product of lines. *Bull. Amer. Math. Soc.*, 72:515–519, 1966.

[5] A. Andoni, O. Neiman, and A. Naor. Snowflake universality of Wasserstein spaces. Preprint, 2010.

[6] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21(1):1–21 (electronic), 2008.

[7] S. Arora, L. Lovász, I. Newman, Y. Rabani, Y. Rabinovich, and S. Vempala. Local versus global properties of metric spaces (extended abstract). In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 41–50, New York, 2006. ACM.

[8] G. Arzhantseva, C. Drutu, and M. Sapir. Compression functions of uniform embeddings of groups into Hilbert and Banach spaces. *J. Reine Angew. Math.*, 633:213–235, 2009.
[9] P. Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.
[10] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301 (electronic), 1998.
[11] T. Austin. Amenable groups with very poor compression into Lebesgue spaces. *Duke Math. J.*, 159(2):187–222, 2011.
[12] T. Austin and A. Naor. On the bi-Lipschitz structure of Wasserstein spaces. Preprint., 2009.
[13] T. Austin, A. Naor, and Y. Peres. The wreath product of $\mathbb{Z}$ with $\mathbb{Z}$ has Hilbert compression exponent $\frac{2}{3}$. *Proc. Amer. Math. Soc.*, 137(1):85–90, 2009.
[14] T. Austin, A. Naor, and A. Valette. The Euclidean distortion of the lamplighter group. *Discrete Comput. Geom.*, 44(1):55–74, 2010.
[15] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.
[16] K. Ball. The Ribe programme. Séminaire Bourbaki, exposé 1047, 2012.
[17] 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.
[18] S. Banach. *Théorie des opérations linéaires*. PWN—Polish Scientific Publishers, Warsaw, 1932. Monografie Matematyczne, Tom 1. [Mathematical Monographs, Vol. 1].
[19] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *STOC '98 (Dallas, TX)*, pages 161–168. ACM, New York, 1999.
[20] Y. Bartal, B. Bollobás, and M. Mendel. Ramsey-type theorems for metric spaces with applications to online problems. *J. Comput. System Sci.*, 72(5):890–921, 2006.
[21] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type dichotomies. *J. London Math. Soc. (2)*, 71(2):289–303, 2005.
[22] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005.
[23] Y. Bartal, N. Linial, M. Mendel, and A. Naor. Some low distortion metric Ramsey problems. *Discrete Comput. Geom.*, 33(1):27–41, 2005.
[24] M. A. Bender and M. Farach-Colton. The LCA problem revisited. In G. H. Gonnet, D. Panario, and A. Viola, editors, *LATIN*, volume 1776 of *Lecture Notes in Computer Science*, pages 88–94. Springer, 2000.
[25] Y. Benyamini. The uniform classification of Banach spaces. In *Texas functional analysis seminar 1984–1985 (Austin, Tex.)*, Longhorn Notes, pages 15–38. Univ. Texas Press, Austin, TX, 1985.
[26] 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.
[27] C. Bessaga. On topological classification of complete linear metric spaces. *Fund. Math.*, 56:251–288, 1964/1965.

[28] C. Bessaga and A. Pełczyński. Some remarks on homeomorphisms of Banach spaces. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 8:757–761, 1960.

[29] C. Bessaga and A. Pełczyński. *Selected topics in infinite-dimensional topology*. PWN—Polish Scientific Publishers, Warsaw, 1975. Monografie Matematyczne, Tom 58. [Mathematical Monographs, Vol. 58].

[30] A. Blum, H. Karloff, Y. Rabani, and M. Saks. A decomposition theorem for task systems and bounds for randomized server problems. *SIAM J. Comput.*, 30(5):1624–1661 (electronic), 2000.

[31] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

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

[33] J. Bourgain. Remarks on the extension of Lipschitz maps defined on discrete sets and uniform homeomorphisms. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 157–167. Springer, Berlin, 1987.

[34] J. Bourgain, T. Figiel, and V. Milman. On Hilbertian subsets of finite metric spaces. *Israel J. Math.*, 55(2):147–152, 1986.

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

[36] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[37] B. Brinkman, A. Karagiozova, and J. R. Lee. Vertex cuts, random walks, and dimension reduction in series-parallel graphs. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 621–630. ACM, New York, 2007.

[38] M. Charikar and A. Karagiozova. A tight threshold for metric Ramsey phenomena. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 129–136 (electronic), New York, 2005. ACM.

[39] M. Charikar, K. Makarychev, and Y. Makarychev. Local global tradeoffs in metric embeddings. In *Proceedings of 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007)*, pages 713–723, 2007.

[40] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pages 195–199. Princeton Univ. Press, Princeton, N. J., 1970.

[41] J. Cheeger and B. Kleiner. Differentiating maps into L^1, and the geometry of BV functions. *Ann. of Math. (2)*, 171(2):1347–1385, 2010.

[42] P.-A. Cherix, M. Cowling, P. Jolissaint, P. Julg, and A. Valette. *Groups with the Haagerup property*, volume 197 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 2001. Gromov's a-T-menability.

[43] M. Christ. A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral. *Colloq. Math.*, 60/61(2):601–628, 1990.

[44] H. Corson and V. Klee. Topological classification of convex sets. In *Proc. Sympos. Pure Math., Vol. VII*, pages 37–51. Amer. Math. Soc., Providence, R.I., 1963.

[45] Y. de Cornulier, R. Tessera, and A. Valette. Isometric group actions on Hilbert spaces: growth of cocycles. *Geom. Funct. Anal.*, 17(3):770–792, 2007.

[46] P. de la Harpe. *Topics in geometric group theory.* Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2000.

[47] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem, 1961.

[48] P. Enflo. On a problem of Smirnov. *Ark. Mat.*, 8:107–109, 1969.

[49] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105 (1969), 1969.

[50] P. Enflo. Uniform structures and square roots in topological groups. I, II. *Israel J. Math. 8 (1970), 230-252; ibid.*, 8:253–272, 1970.

[51] P. Enflo. Banach spaces which can be given an equivalent uniformly convex norm. *Israel J. Math.*, 13:281–288, 1973.

[52] P. Enflo. Uniform homeomorphisms between Banach spaces. In *Séminaire Maurey-Schwartz (1975–1976), Espaces, L^p, applications radonifiantes et géométrie des espaces de Banach, Exp. No. 18*, page 7. Centre Math., École Polytech., Palaiseau, 1976.

[53] P. Erdős. Extremal problems in graph theory. In *Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963)*, pages 29–36. Publ. House Czechoslovak Acad. Sci., Prague, 1964.

[54] A. G. Èrschler. On the asymptotics of the rate of departure to infinity. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 283(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 6):251–257, 263, 2001.

[55] M. Fabian, P. Habala, P. Hájek, V. Montesinos Santalucía, J. Pelant, and V. Zizler. *Functional analysis and infinite-dimensional geometry.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 8. Springer-Verlag, New York, 2001.

[56] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.*, 69(3):485–497, 2004.

[57] X. Fernique. Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École d'Été de Probabilités de Saint-Flour, IV-1974*, pages 1–96. Lecture Notes in Math., Vol. 480. Springer, Berlin, 1975.

[58] X. Fernique. Évaluations de processus gaussiens composés. In *Probability in Banach spaces (Proc. First Internat. Conf., Oberwolfach, 1975)*, pages 67–83. Lecture Notes in Math., Vol. 526. Springer, Berlin, 1976.

[59] X. Fernique. Caractérisation de processus à trajectoires majorées ou continues. In *Séminaire de Probabilités, XII (Univ. Strasbourg, Strasbourg, 1976/1977)*, volume 649 of *Lecture Notes in Math.*, pages 691–706. Springer, Berlin, 1978.

[60] T. Figiel. On nonlinear isometric embeddings of normed linear spaces. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 16:185–188, 1968.

[61] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.

[62] M. Fréchet. *Les espaces abstraits et leur théorie considérée comme introduction à l'analyse générale.* Paris: Gauthier-Villars (Collection de monographies sur la théorie des fonctions). XII, 296 p. , 1928.

[63] O. Giladi, M. Mendel, and A. Naor. Improved bounds in the metric cotype inequality for Banach spaces. *J. Funct. Anal.*, 260(1):164–194, 2011.

[64] O. Giladi, A. Naor, and G. Schechtman. Bourgain's discretization theorem. Preprint available at `http://arxiv.org/abs/1110.5368`, to appear in *Ann. Fac. Sci. Toulouse Math.*, 2011.

[65] E. Gorelik. The uniform nonequivalence of L_p and l_p. *Israel J. Math.*, 87(1-3):1–8, 1994.

[66] M. Gromov. Volume and bounded cohomology. *Inst. Hautes Études Sci. Publ. Math.*, (56):5–99 (1983), 1982.

[67] M. Gromov. Filling Riemannian manifolds. *J. Differential Geom.*, 18(1):1–147, 1983.

[68] M. Gromov. Asymptotic invariants of infinite groups. In *Geometric group theory, Vol. 2 (Sussex, 1991)*, volume 182 of *London Math. Soc. Lecture Note Ser.*, pages 1–295. Cambridge Univ. Press, Cambridge, 1993.

[69] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces.* Modern Birkhäuser Classics. Birkhäuser Boston Inc., Boston, MA, english edition, 2007. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.

[70] A. Grothendieck. Sur certaines classes de suites dans les espaces de Banach et le théorème de Dvoretzky-Rogers. *Bol. Soc. Mat. São Paulo*, 8:81–110 (1956), 1953.

[71] A. Grothendieck. Erratum au mémoire: Produits tensoriels topologiques et espaces nucléaires. *Ann. Inst. Fourier, Grenoble*, 6:117–120, 1955–1956.

[72] E. Guentner and J. Kaminker. Exactness and uniform embeddability of discrete groups. *J. London Math. Soc. (2)*, 70(3):703–718, 2004.

[73] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and l_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004.

[74] O. Hanner. On the uniform convexity of L^p and l^p. *Ark. Mat.*, 3:239–244, 1956.

[75] S. Har-Peled and M. Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM J. Comput.*, 35(5):1148–1184 (electronic), 2006.

[76] D. Harel and R. E. Tarjan. Fast algorithms for finding nearest common ancestors. *SIAM J. Comput.*, 13(2):338–355, 1984.

[77] S. Heinrich and P. Mankiewicz. Applications of ultrapowers to the uniform and Lipschitz classification of Banach spaces. *Studia Math.*, 73(3):225–251, 1982.

[78] G. M. Henkin. The lack of a uniform homeomorphism between the spaces of smooth functions of one and of n variables ($n \geqslant 2$). *Mat. Sb. (N.S.)*, 74 (116):595–607, 1967.

[79] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.

[80] J. D. Howroyd. On dimension and on the existence of sets of finite positive Hausdorff measure. *Proc. London Math. Soc. (3)*, 70(3):581–604, 1995.

[81] B. Hughes. Trees and ultrametric spaces: a categorical equivalence. *Adv. Math.*, 189(1):148–191, 2004.

[82] T. Hytönen, S. Li, and A. Naor. Quantitative affine approximation for UMD targets. Preprint, 2012.

[83] T. Hytönen and A. Naor. Pisier's inequality revisited. Preprint, 2012.

[84] 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.

[85] R. C. James. Super-reflexive Banach spaces. *Canad. J. Math.*, 24:896–904, 1972.

[86] R. C. James. Nonreflexive spaces of type 2. *Israel J. Math.*, 30(1-2):1–13, 1978.

[87] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[88] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[89] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. Banach spaces determined by their uniform structures. *Geom. Funct. Anal.*, 6(3):430–470, 1996.

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

[91] M. Ĭ. Kadec′. On strong and weak convergence. *Dokl. Akad. Nauk SSSR*, 122:13–16, 1958.

[92] M. Ĭ. Kadec′. Topological equivalence of all separable Banach spaces. *Dokl. Akad. Nauk SSSR*, 167:23–25, 1966.

[93] M. I. Kadec. A proof of the topological equivalence of all separable infinite-dimensional Banach spaces. *Funkcional. Anal. i Priložen.*, 1:61–70, 1967.

[94] J.-P. Kahane. Sur les sommes vectorielles $\sum \pm u_n$. *C. R. Acad. Sci. Paris*, 259:2577–2580, 1964.

[95] N. Kalton. The uniform structure of Banach spaces. To appear in *Math. Ann.*, 2011.

[96] N. J. Kalton. The nonlinear geometry of Banach spaces. *Rev. Mat. Complut.*, 21(1):7–60, 2008.

[97] H. Karloff, Y. Rabani, and Y. Ravid. Lower bounds for randomized k-server and motion-planning algorithms. *SIAM J. Comput.*, 23(2):293–312, 1994.

[98] T. Keleti, A. Máthé, and O. Zindulka. Hausdorff dimension of metric spaces and Lipschitz maps onto cubes. Preprint, available at `http://arxiv.org/abs/1203.0686`, 2012.

[99] S. Khot and A. Naor. Nonembeddability theorems via Fourier analysis. *Math. Ann.*, 334(4):821–852, 2006.

[100] S. Khot and R. Saket. SDP integrality gaps with local ℓ_1-embeddability. In *Proceedings of 50th Symposium on Foundations of Computer Science (FOCS 2009)*, pages 565–574, 2009.

[101] 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.

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

[103] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

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

[105] S. Li and A. Naor. Discretization and affine approximation in high dimensions. Preprint available at `http://arxiv.org/abs/1202.2567`. To appear in Israel J. Math., 2012.

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

[107] J. Lindenstrauss. On nonlinear projections in Banach spaces. *Michigan Math. J.*, 11:263–287, 1964.

[108] J. Lindenstrauss. Some aspects of the theory of Banach spaces. *Advances in Math.*, 5:159–180 (1970), 1970.

[109] J. Lindenstrauss. Uniform embeddings, homeomorphisms and quotient maps between Banach spaces (a short survey). *Topology Appl.*, 85(1-3):265–279, 1998. 8th Prague Topological Symposium on General Topology and Its Relations to Modern Analysis and Algebra (1996).

[110] N. Linial. Finite metric-spaces—combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pages 573–586, Beijing, 2002. Higher Ed. Press.

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

[112] N. Linial, A. Magen, and A. Naor. Girth and Euclidean distortion. *Geom. Funct. Anal.*, 12(2):380–394, 2002.

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

[114] K. Makarychev and Y. Makarychev. Metric extension operators, vertex sparsifiers and Lipschitz extendability. In *51th Annual IEEE Symposium on Foundations of Computer Science*, pages 255–264, 2010.

[115] M. B. Marcus and G. Pisier. Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. *Acta Math.*, 152(3-4):245–301, 1984.

[116] J. Matoušek. Ramsey-like properties for bi-Lipschitz mappings of finite metric spaces. *Comment. Math. Univ. Carolin.*, 33(3):451–463, 1992.

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

[118] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[119] P. Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.

[120] B. Maurey. *Théorèmes de factorisation pour les opérateurs linéaires à valeurs dans les espaces L^p*. Société Mathématique de France, Paris, 1974. With an English summary, Astérisque, No. 11.

[121] B. Maurey. Type, cotype and K-convexity. In *Handbook of the geometry of Banach spaces, Vol. 2*, pages 1299–1332. North-Holland, Amsterdam, 2003.

[122] 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.
[123] S. Mazur and S. Ulam. Sur les transformations isométriques d'espaces vectoriels normés. *C. R. Math. Acad. Sci. Paris*, 194:946–948, 1932.
[124] M. Mendel. Metric dichotomies. In *Limits of graphs in group theory and computer science*, pages 59–76. EPFL Press, Lausanne, 2009.
[125] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Adv. Math.*, 189(2):451–494, 2004.
[126] M. Mendel and A. Naor. Some applications of Ball's extension theorem. *Proc. Amer. Math. Soc.*, 134(9):2577–2584 (electronic), 2006.
[127] M. Mendel and A. Naor. Ramsey partitions and proximity data structures. *J. Eur. Math. Soc. (JEMS)*, 9(2):253–275, 2007.
[128] M. Mendel and A. Naor. Scaled Enflo type is equivalent to Rademacher type. *Bull. Lond. Math. Soc.*, 39(3):493–498, 2007.
[129] 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. Full version to appear in J. Eur. Math. Soc. (JEMS).
[130] M. Mendel and A. Naor. Metric cotype. *Ann. of Math. (2)*, 168(1):247–298, 2008.
[131] M. Mendel and A. Naor. Maximum gradient embeddings and monotone clustering. *Combinatorica*, 30(5):581–615, 2010.
[132] M. Mendel and A. Naor. Towards a calculus for non-linear spectral gaps. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 236–255, 2010.
[133] M. Mendel and A. Naor. A note on dichotomies for metric transforms. Available at `http://arxiv.org/abs/1102.1800`, 2011.
[134] M. Mendel and A. Naor. Ultrametric skeletons. Preprint available at `http://arxiv.org/abs/1112.3416`. To appear in *Proc. Natl. Acad. Sci. USA*, 2011.
[135] M. Mendel and A. Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. Preprint, 2012.
[136] M. Mendel and A. Naor. Ultrametric subsets with large Hausdorff dimension. *Invent. Math.* DOI 10.1007/s00222-012-0402-7, 2012.
[137] M. Mendel and C. Schwob. Fast C-K-R partitions of sparse graphs. *Chic. J. Theoret. Comput. Sci.*, pages Article 2, 15, 2009.
[138] V. Milman and G. Schechtman. An "isomorphic" version of Dvoretzky's theorem. II. In *Convex geometric analysis (Berkeley, CA, 1996)*, volume 34 of *Math. Sci. Res. Inst. Publ.*, pages 159–164. Cambridge Univ. Press, Cambridge, 1999.
[139] V. D. Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen.*, 5(4):28–37, 1971.
[140] V. D. Milman. Almost Euclidean quotient spaces of subspaces of a finite-dimensional normed space. *Proc. Amer. Math. Soc.*, 94(3):445–449, 1985.
[141] J. Milnor and W. Thurston. Characteristic numbers of 3-manifolds. *Enseignement Math. (2)*, 23(3-4):249–254, 1977.
[142] P. B. Miltersen. Cell probe complexity - a survey. In *19th Conference on the Foundations of Software Technology and Theoretical Computer Science*, 1999.

[143] G. D. Mostow. Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms. *Inst. Hautes Études Sci. Publ. Math.*, (34):53–104, 1968.

[144] A. Naor. A phase transition phenomenon between the isometric and isomorphic extension problems for Hölder functions between L_p spaces. *Mathematika*, 48(1-2):253–271 (2003), 2001.

[145] A. Naor. An application of metric cotype to quasisymmetric embeddings. Available at http://arxiv.org/abs/math/0607644, 2006.

[146] A. Naor. L_1 embeddings of the Heisenberg group and fast estimation of graph isoperimetry. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1549–1575, New Delhi, 2010. Hindustan Book Agency.

[147] A. Naor and Y. Peres. Embeddings of discrete groups and the speed of random walks. *Int. Math. Res. Not. IMRN*, pages Art. ID rnn 076, 34, 2008.

[148] A. Naor and Y. Peres. L_p compression, traveling salesmen, and stable walks. *Duke Math. J.*, 157(1):53–108, 2011.

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

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

[151] A. Naor and T. Tao. Random martingales and localization of maximal inequalities. *J. Funct. Anal.*, 259(3):731–779, 2010.

[152] A. Naor and T. Tao. Scale-oblivious metric fragmentation and the nonlinear Dvoretzky theorem. Preprint available at http://arxiv.org/abs/1003.4013, to appear in *Israel J. Math.*, 2010.

[153] F. Nazarov, S. Treil, and A. Volberg. The Tb-theorem on non-homogeneous spaces. *Acta Math.*, 190(2):151–239, 2003.

[154] S.-I. Ohta. Markov type of Alexandrov spaces of non-negative curvature. *Mathematika*, 55(1-2):177–189, 2009.

[155] A. Olshanskii and D. Osin. A quasi-isometric embedding theorem for groups. Preprint available at http://arxiv.org/abs/1202.6437, 2011.

[156] M. I. Ostrovskii. On metric characterizations of some classes of Banach spaces. *C. R. Acad. Bulgare Sci.*, 64(6):775–784, 2011.

[157] M. I. Ostrovskii. Different forms of metric characterizations of classes of Banach spaces. Preprint available at http://arxiv.org/abs/1112.0801, 2012.

[158] M. I. Ostrovskii. Low-distortion embeddings of graphs with large girth. *J. Funct. Anal.*, 262(8), 2012.

[159] M. I. Ostrovskii. Test-space characterizations of some classes of Banach spaces. Preprint available at http://arxiv.org/abs/1112.3086, 2012.

[160] P. Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math. (2)*, 129(1):1–60, 1989.

[161] 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.

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

[163] G. Pisier. Holomorphic semigroups and the geometry of Banach spaces. *Ann. of Math. (2)*, 115(2):375–392, 1982.

[164] 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.
[165] G. Pisier and Q. H. Xu. Random series in the real interpolation spaces between the spaces v_p. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 185–209. Springer, Berlin, 1987.
[166] Y. Rabinovich. On average distortion of embedding metrics into the line. *Discrete Comput. Geom.*, 39(4):720–733, 2008.
[167] P. Raghavendra and D. Steurer. Integrality gaps for strong SDP relaxations of unique games. In *Proceedings of 50th Symposium on Foundations of Computer Science (FOCS 2009)*, pages 575–585, 2009.
[168] D. Revelle. Rate of escape of random walks on wreath products and related groups. *Ann. Probab.*, 31(4):1917–1934, 2003.
[169] M. Ribe. On uniformly homeomorphic normed spaces. *Ark. Mat.*, 14(2):237–244, 1976.
[170] M. Ribe. On uniformly homeomorphic normed spaces. II. *Ark. Mat.*, 16(1):1–9, 1978.
[171] M. Ribe. Existence of separable uniformly homeomorphic nonisomorphic Banach spaces. *Israel J. Math.*, 48(2-3):139–147, 1984.
[172] H. Sachs. Regular graphs with given girth and restricted circuits. *J. London Math. Soc.*, 38:423–429, 1963.
[173] H. Sagan. *Space-filling curves*. Universitext. Springer-Verlag, New York, 1994.
[174] G. Schechtman. Two observations regarding embedding subsets of Euclidean spaces in normed spaces. *Adv. Math.*, 200(1):125–135, 2006.
[175] S. Semmes. On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞-weights. *Rev. Mat. Iberoamericana*, 12(2):337–410, 1996.
[176] P. Solé. The second eigenvalue of regular graphs of given girth. *J. Combin. Theory Ser. B*, 56(2):239–249, 1992.
[177] C. Sommer, E. Verbin, and W. Yu. Distance oracles for sparse graphs. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009)*, pages 703–712. IEEE Computer Soc., Los Alamitos, CA, 2009.
[178] M. Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987.
[179] M. Talagrand. Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem. *Geom. Funct. Anal.*, 3(3):295–314, 1993.
[180] M. Talagrand. *The generic chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005. Upper and lower bounds of stochastic processes.
[181] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes*. 2011. Modern Methods and Classical Problems. Forthcoming book.
[182] M. Thorup and U. Zwick. Approximate distance oracles. *J. ACM*, 52(1):1–24 (electronic), 2005.
[183] H. Toruńczyk. Characterizing Hilbert space topology. *Fund. Math.*, 111(3):247–262, 1981.
[184] M. Urbański. Transfinite Hausdorff dimension. *Topology Appl.*, 156(17):2762–2771, 2009.

[185] J. Väisälä. The free quasiworld. Freely quasiconformal and related maps in Banach spaces. In *Quasiconformal geometry and dynamics (Lublin, 1996)*, volume 48 of *Banach Center Publ.*, pages 55–118. Polish Acad. Sci., Warsaw, 1999.

[186] I. A. Vestfrid and A. F. Timan. A universality property of Hilbert spaces. *Dokl. Akad. Nauk SSSR*, 246(3):528–530, 1979.

[187] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[188] R. Wagner. Notes on an inequality by Pisier for functions on the discrete cube. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 263–268. Springer, Berlin, 2000.

[189] D. P. Williamson and D. B. Shmoys. *The design of approximation algorithms*. Cambridge University Press, Cambridge, 2011.

[190] P. Wojtaszczyk. *Banach spaces for analysts*, volume 25 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1991.

[191] C. Wulff-Nilsen. Approximate distance oracles with improved query time. Preprint, available at `http://arxiv.org/abs/1202.2336`, 2011.

Courant Institute, New York University, New York NY 10012
E-mail address: `naor@cims.nyu.edu`