METRIC DIMENSION REDUCTION: A SNAPSHOT OF THE RIBE PROGRAM

ASSAF NAOR

1. INTRODUCTION

The purpose of this article is to survey some of the context, achievements, challenges and mysteries of the field of *metric dimension reduction*, including new perspectives on major older results as well as recent advances.

From the point of view of theoretical computer science, mathematicians "stumbled upon" metric dimension reduction in the early 1980s, as exemplified by the following quote [**?**].

> *Two decades ago, analysts stumbled upon a surprising fact [...], the Johnson–Lindenstrauss Lemma, as a crucial tool in their project of extending functions in continuous ways. This result [...] says that, if you project n points in some high dimensional space down to a random $O(\log n)$-dimensional plane, the chances are overwhelming that all distances will be preserved within a small relative error. So, if distance is all you care about, there is no reason to stay in high dimensions!*
>
> C. Papadimitriou, 2004 (forward to *The random projection method* by S. Vempala).

The above use of the term "stumbled upon" is justified, because it would be fair to say that at the inception of this research direction mathematicians did not anticipate the remarkable swath of its later impact on algorithms. However, rather than being discovered accidentally, the investigations that will be surveyed here can be motivated by classical issues in metric geometry. From the internal perspective of pure mathematics, it would be more befitting to state that the aforementioned early work stumbled upon the unexpected depth, difficulty and richness of basic questions on the relation between "rough quantitative geometry" and dimension. Despite major efforts by many mathematicians over the past four decades, such questions remain by and large stubbornly open.

We will explain below key ideas of major developments in metric dimension reduction, and also describe the larger mathematical landscape that partially motivates these investigations, most notably the *bi-Lipschitz embedding problem into* $\mathbb{R}^n$ and the *Ribe program*. By choosing to focus on aspects of this area within pure mathematics, we will put aside the large (and growing) literature that investigates algorithmic ramifications of metric dimension reduction. Such applications warrant a separate treatment that is far beyond the scope of the present exposition; some aspects of that material are covered in the monographs [**?**, **?**, **?**] and the surveys [**?**, **?**], as well as the articles of Andoni–Indyk–Razenshteyn and Arora in the present volume.

Remark 1. The broader term *dimension reduction* is used ubiquitously in statistics and machine learning, with striking applications whose full rigorous understanding sometimes awaits the scrutiny of mathematicians (see e.g. [**?**]). A common (purposefully vague) description of this term is the desire to decrease the degrees of freedom of a high-dimensional data set while approximately preserving some of its pertinent features; formulated in such great generality, the area includes topics such as neural networks (see e.g. [**?**]). The commonly used term *curse of dimensionality* refers to the perceived impossibility of this goal in many settings, and that the performance (running time, storage space) of certain algorithmic tasks must deteriorate exponentially as the underlying dimension grows. But, sometimes it does seem that certain high-dimensional data sets can be realized faithfully using a small number of latent variables as auxiliary "coordinates." Here we restrict ourselves exclusively to *metric* dimension reduction, i.e., to notions of faithfulness of low-dimensional representations that require the (perhaps quite rough) preservation of pairwise distances, including ways to prove the impossibility thereof.

Roadmap. The rest of the Introduction is an extensive and detailed account of the area of metric dimension reduction, including statements of most of the main known results, background and context, and many important open questions. The Introduction is thus an expository account of the field, so those readers who do not wish to delve into some proofs, could read it separately from the rest of the text. The remaining sections contain further details and complete justifications of those statements that have not appeared in the literature.

1.1. **Bi-Lipschitz embeddings.** Fix $\alpha \geqslant 1$. A metric space $(\mathcal{M}, d_\mathcal{M})$ is said to embed with distortion α into a metric space $(\mathcal{N}, d_\mathcal{N})$ if there is a mapping (an embedding) $f : \mathcal{M} \to \mathcal{N}$ and (a scaling factor) $\tau > 0$ such that

$$\forall\, x, y \in \mathcal{M}, \qquad \tau d_\mathcal{M}(x, y) \leqslant d_\mathcal{N}\big(f(x), f(y)\big) \leqslant \alpha\tau d_\mathcal{M}(x, y). \tag{1}$$

The infimum over $\alpha \in [1, \infty]$ for which $(\mathcal{M}, d_\mathcal{M})$ embeds with distortion α into $(\mathcal{N}, d_\mathcal{N})$ is denoted $\mathsf{c}_{(\mathcal{N}, d_\mathcal{N})}(\mathcal{M}, d_\mathcal{M})$, or $\mathsf{c}_\mathcal{N}(\mathcal{M})$ if the underlying metrics are clear from the context. If $\mathsf{c}_\mathcal{N}(\mathcal{M}) < \infty$, then $(\mathcal{M}, d_\mathcal{M})$ is said to admit a bi-Lipschitz embedding into $(\mathcal{N}, d_\mathcal{N})$. Given $p \in [1, \infty)$, if $\mathcal{N}$ is an $L_p(\mu)$ space into which $\mathcal{M}$ admits a bi-Lipschitz embedding, then we use the notation $\mathsf{c}_{L_p(\mu)}(\mathcal{M}) = \mathsf{c}_p(\mathcal{M})$. The numerical invariant $\mathsf{c}_2(\mathcal{M})$, which measures the extent to which $\mathcal{M}$ is close to being a (subset of a) Euclidean geometry, is called the *Euclidean distortion* of $\mathcal{M}$.

A century of intensive research into bi-Lipschitz embeddings led to a rich theory with many deep achievements, but the following problem, which is one of the first questions that arise naturally in this direction, remains a major longstanding mystery; see e.g. [**?**, **?**, **?**, **?**]. Analogous issues in the context of topological dimension, differentiable manifolds and Riemannian manifolds were famously settled by Menger [**?**] and Nöbeling [**?**], Whitney [**?**] and Nash [**?**], respectively.

Problem 2 (the bi-Lipschitz embedding problem into $\mathbb{R}^k$). Obtain an intrinsic characterization of those metric spaces $(\mathcal{M}, d_\mathcal{M})$ that admit a bi-Lipschitz embedding into $\mathbb{R}^k$ for some $k \in \mathbb{N}$.

Problem 2 is one of the qualitative underpinnings of the issues that will be surveyed here. We say that it is "qualitative" because it ignores the magnitude of the distortion altogether, and therefore one does not need to specify which norm on $\mathbb{R}^k$ is considered in Problem 2, since all the norms on $\mathbb{R}^k$ are (bi-Lipschitz) equivalent. Problem 2 is also (purposefully) somewhat vague, because the notion of "intrinsic characterization" is not well-defined. We will return to this matter in Section 3 below, where candidates for such a characterization are discussed. At this juncture, it suffices to illustrate what Problem 2 aims to achieve through the following useful example. If one does not impose any restriction on the target dimension and allows for a bi-Lipschitz embedding into an infinite dimensional Hilbert space, then the following intrinsic characterization is available. A metric space $(\mathcal{M}, d_\mathcal{M})$ admits a bi-Lipschitz embedding into a Hilbert space if and only if there exists $C = C_\mathcal{M} \in [0, 1)$ such that for every $n \in \mathbb{N}$ and every positive semidefinite symmetric matrix $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$ all of whose rows sum to zero (i.e., $\sum_{j=1}^n a_{ij} = 0$ for every $i \in \{1, \ldots, n\}$), the following quadratic distance inequality holds true.

$$\forall\, x_1, \ldots, x_n \in \mathcal{M}, \quad \sum_{i=1}^n \sum_{j=1}^n a_{ij} d_\mathcal{M}(x_i, x_j)^2 \leqslant C \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| d_\mathcal{M}(x_i, x_j)^2. \tag{2}$$

In fact, one can refine this statement quantitatively as follows. A metric space embeds with distortion $\alpha \in [1, \infty)$ into a Hilbert space if and only if in the setting of (2) we have

$$\forall\, x_1, \ldots, x_n \in \mathcal{M}, \quad \sum_{i=1}^n \sum_{j=1}^n a_{ij} d_\mathcal{M}(x_i, x_j)^2 \leqslant \frac{\alpha^2 - 1}{\alpha^2 + 1} \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| d_\mathcal{M}(x_i, x_j)^2. \tag{3}$$

The case $\alpha = 1$ of (3), i.e., the case of isometric embeddings, is a famous classical theorem of Schoenberg [**?**], and the general case of (3) is due to Linial, London and Rabinovich [**?**, Corollary 3.5]. The above characterization is clearly intrinsic, as it is a family of finitary distance inequalities among points of $\mathcal{M}$ that do not make reference to any other auxiliary/external object. With such a characterization at hand, one could examine the internal structure of a given metric space so as to determine if it can be represented faithfully as a subset of a Hilbert space. Indeed, [**?**] uses (3) to obtain an algorithm that takes as input an n-point metric space $(\mathcal{M}, d_\mathcal{M})$ and outputs in polynomial time an arbitrarily good approximation to its Euclidean distortion $\mathsf{c}_2(\mathcal{M})$.

A meaningful answer to Problem 2 could in principle lead to a method for determining if a member of a family $\mathcal{F}$ of metric spaces admits an embedding with specified distortion into a member of a family $\mathcal{G}$ of low dimensional normed spaces. Formulated in such great generality, this type of question encompasses all of the investigations into metric dimension reduction that will be discussed in what follows, except that we will also examine analogous issues for embeddings with guarantees that are substantially weaker (though natural and useful) than the "vanilla" bi-Lipschitz requirement (1).

Remark 3. Analogues of the above questions are very natural also when the target low-dimensional geometries are not necessarily normed spaces. Formulating meaningful goals in such a setting is not as straightforward as it is for normed spaces, e.g. requiring that the target space is a manifold of low topological dimension is not

very useful, so one must impose geometric restrictions on the target manifold. As another example (to which we will briefly return later), one could ask about embeddings into spaces of probability measures that are equipped with a Wasserstein (transportation cost) metric, with control on the size of the underlying metric space. At present, issues of this type are largely an uncharted terrain whose exploration is likely to be interesting and useful.

1.2. **Local theory and the Ribe program.**

Besides being motivated by the bi-Lipschitz embedding problem into $\mathbb{R}^k$, much of the inspiration for the studies that will be presented below comes from a major endeavour in metric geometry called the *Ribe program*. This is a large and active research area that has been (partially) surveyed in [**?**, **?**, **?**, **?**, **?**]. It would be highly unrealistic to attempt to cover it comprehensively here, but we will next present a self-contained general introduction to the Ribe program that is intended for non-experts, including aspects that are relevant to the ensuing discussion on metric dimension reduction.

Martin Ribe was a mathematician who in the 1970s obtained a few beautiful results in functional analysis, prior to leaving mathematics. Among his achievements is a very influential rigidity theorem [**?**] which shows that the local linear theory of Banach spaces could in principle be described using only distances between points, and hence it could potentially apply to general metric spaces.

Before formulating the above statement precisely, it is instructive to consider a key example. The *infimal cotype* [**?**] q_X of a Banach space $(X, \|\cdot\|)$ is the infimum over those $q \in [2, \infty]$ for which[1]

$$\forall\, n \in \mathbb{N}, \ \forall\, x_1, \ldots, x_n \in X, \qquad \sum_{i=1}^{n} \|x_i\|^2 \lesssim_{X,q} \frac{n^{1-\frac{2}{q}}}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \left\| \sum_{i=1}^{n} \varepsilon_i x_i \right\|^2. \tag{4}$$

In the special case $x_1 = \ldots = x_n = x \in X \smallsetminus \{0\}$, the left hand side of (4) is equal to $n\|x\|^2$ and by expanding the squares one computes that the right hand side of (4) is equal to $n^{2(1-1/q)}\|x\|^2$. Hence (4) necessitates that $q \geqslant 2$, which explains why we imposed this restriction on q at the outset. Note also that (4) holds true in any Banach space when $q = \infty$. This is a quick consequence of the convexity of the mapping $x \mapsto \|x\|^2$, since for every $\varepsilon \in \{-1,1\}^n$ and $i \in \{1, \ldots, n\}$ we have

$$\|x_i\|^2 = \left\| \frac{(\varepsilon_1 x_1 + \ldots + \varepsilon_n x_n) + (-\varepsilon_1 x_1 - \ldots - \varepsilon_{i-1} x_{i-1} + \varepsilon_i x_i - \varepsilon_{i+1} x_{i+1} \ldots - \varepsilon_n x_n)}{2} \right\|^2$$

$$\leqslant \frac{\|\varepsilon_1 x_1 + \ldots + \varepsilon_n x_n\|^2 + \| - \varepsilon_1 x_1 - \ldots - \varepsilon_{i-1} x_{i-1} + \varepsilon_i x_i - \varepsilon_{i+1} x_{i+1} \ldots - \varepsilon_n x_n\|^2}{2}. \tag{5}$$

By averaging (5) over $\varepsilon \in \{-1,1\}^n$ and $i \in \{1, \ldots, n\}$ we see that (4) holds if $q = \infty$. So, one could view (4) for $q < \infty$ as a requirement that the norm $\|\cdot\| : X \to [0, \infty)$ has a property that is asymptotically stronger than mere convexity. When $X = \ell_\infty$, this requirement does not hold for any $q < \infty$, since if $\{x_i\}_{i=1}^n$ are the first n elements of the coordinate basis, then the left hand side of (4) equals n while its right hand side equals $n^{1-1/q}$.

Maurey and Pisier proved [**?**] that the above obstruction to having $q_X < \infty$ is actually the *only possible* such obstruction. Thus, by ruling out the presence of copies of $\{\ell_\infty^n\}_{n=1}^\infty$ in X one immediately deduces the "upgraded" (asymptotically stronger as $n \to \infty$) randomized convexity inequality (4) for some $q < \infty$.

Theorem 4. *The following conditions are equivalent for every Banach space $(X, \|\cdot\|)$.*

- *There is no $\alpha \in [1, \infty)$ such that ℓ_∞^n is α-isomorphic to a subspace of X for every $n \in \mathbb{N}$.*
- *$q_X < \infty$.*

The (standard) terminology that is used in Theorem 4 is that given $\alpha \in [1, \infty)$, a Banach space $(Y, \|\cdot\|_Y)$ is said to be α-isomorphic to a subspace of a Banach space $(Z, \|\cdot\|_Z)$ if there is a linear operator $T : Y \to Z$ satisfying $\|y\|_Y \leqslant \|Ty\|_Z \leqslant \alpha\|y\|_Y$ for every $y \in Y$; this is the same as saying that Y embeds into Z with distortion α via an embedding that is a linear operator.

Suppose that X and Y are Banach spaces that are uniformly homeomorphic, i.e., there is a bijection $f : X \to Y$ such that both f and f^{-1} are uniformly continuous. By the aforementioned rigidity theorem of Ribe

---

[1] In addition to the standard "O" notation, we will use throughout this article the following standard and convenient asymptotic notation. Given two quantities $Q, Q' > 0$, the notations $Q \lesssim Q'$ and $Q' \gtrsim Q$ mean that $Q \leqslant CQ'$ for some universal constant $C > 0$. The notation $Q \asymp Q'$ stands for $(Q \lesssim Q') \wedge (Q' \lesssim Q)$. If we need to allow for dependence on parameters, we indicate this by subscripts. For example, in the presence of auxiliary objects (e.g. numbers or spaces) $\phi, \mathfrak{z}$, the notation $Q \lesssim_{\phi,\mathfrak{z}} Q'$ means that $Q \leqslant C(\phi, \mathfrak{z})Q'$, where $C(\phi, \mathfrak{z}) > 0$ is allowed to depend only on $\phi, \mathfrak{z}$; similarly for the notations $Q \gtrsim_{\phi,\mathfrak{z}} Q'$ and $Q \asymp_{\phi,\mathfrak{z}} Q'$.

(which will be formulated below in full generality), this implies in particular that $q_X = q_Y$. So, despite the fact that the requirement (4) involves linear operations (summation and sign changes) that do not make sense in general metric spaces, it is in fact preserved by purely metric (quantitatively continuous, though potentially very complicated) deformations. Therefore, in principle (4) could be characterized while only making reference to distances between points in X. More generally, Ribe's rigidity theorem makes an analogous assertion for *any* isomorphic local linear property of a Banach space; we will define formally those properties in a moment, but, informally, they are requirements in the spirit of (4) that depend only on the finite dimensional subspaces of the given Banach space and are stable under linear isomorphisms that could potentially incur a large error.

The purely metric reformulation of (4) about which we speculated above is only suggested but not guaranteed by Ribe's theorem. From Ribe's statement we will only infer an indication that there might be a "hidden dictionary" for translating certain linear properties into metric properties, but we will not be certain that any specific "entry" of this dictionary (e.g. the entry for, say, "$q_X = \pi$") does in fact exist, and even if it does exist, we will not have an indication what it says. A hallmark of the Ribe program is that at its core it is a search for a family of analogies and definitions, rather than being a collection of specific conjectures. Once such analogies are made and the corresponding questions are formulated, their value is of course determined by the usefulness/depth of the phenomena that they uncover and the theorems that could be proved about them. Thus far, not all of the steps of this endeavour turned out to have a positive answer, but the vast majority did. This had major impact on the study of metric spaces that a priori have nothing to do with Banach spaces, such as graphs, manifolds, groups, and metrics that arise in algorithms (e.g. as continuous relaxations).

The first written formulation of the plan to uncover a hidden dictionary between normed spaces and metric spaces is the following quote of Bourgain [**?**], a decade after Ribe's theorem appeared.

> *It follows in particular from Ribe's result [...] that the notions from local theory of normed spaces are determined by the metric structure of the space and thus have a purely metrical formulation. The next step consists in studying these metrical concepts in general metric spaces in an attempt to develop an analogue of the linear theory. A detailed exposition of this program will appear in J. Lindenstrauss's forthcoming survey [...] in our "dictionary" linear operators are translated in Lipschitz maps, the operator norm by the Lipschitz constant of the map [...] The translations of "Banach-Mazur distance" and "finite-representability" in linear theory are immediate. At the roots of the local theory of normed spaces are properties such as type, cotype, superreflexivity [...] The analogue of type in the geometry of metric spaces is [...] A simple metrical invariant replacing the notion of cotype was not yet discovered.*

<div align="right">J. Bourgain, 1986.</div>

Unfortunately, the survey of Lindenstrauss that is mentioned above never appeared. Nonetheless, Lindenstrauss had massive impact on this area as a leader who helped set the course of the Ribe program, as well as due to the important theorems that he proved in this direction. In fact, the article [**?**] of Johnson and Lindenstrauss, where the aforementioned metric dimension reduction lemma was proved, appeared a few years before [**?**] and contained inspirational (even prophetic) ideas that had major subsequent impact on the Ribe program (including on Bourgain's works in this area). In the above quote, we removed the text describing "the analogue of type in the geometry of metric spaces" so as to not digress; it refers to the influential work of Bourgain, Milman and Wolfson [**?**] (see also the earlier work of Enflo [**?**] and the subsequent work of Pisier [**?**]). "Superreflexivity" was the main focus of [**?**], where the corresponding step of the Ribe program was completed (we will later discuss and use a refinement of this solution). An answer to the above mentioned question on cotype, which we will soon describe, was subsequently found by Mendel and the author [**?**]. We will next explain the terminology "finite-representability" in the above quote, so as to facilitate the ensuing discussion.

1.2.1. *Finite representability.* The first decades of work on the geometry of Banach spaces focused almost entirely on an inherently infinite dimensional theory. This was governed by Banach's partial ordering of Banach spaces [**?**, Chapter 7], which declares that a Banach space X has "linear dimension" at most that of a Banach space Y if there exists $\alpha \geqslant 1$ such that X is α-isomorphic to a subspace of Y. In a remarkable feat of foresight, the following quote of Grothendieck [**?**] heralded the *local theory of Banach spaces*, by shifting attention to the geometry of the finite dimensional subspaces of a Banach space as a way to understand its global structure.

> *assouplissons la notion de "dimension linéaire" de Banach, en disant que l'espace normé E a un* type *linéaire inférieur à celui d'un espace normé F, si on peut trouver un M > 0 fixe tel que tout sous-espace de dimension finie E_1 de E soit isomorphe "à M près" à un sous-espace F_1 de F (i.e. il existe une*

<div align="center">4</div>

application linéaire biunivoque de E_1 sur F_1 , de norme $\leqslant 1$, dont lápplication inverse a une norme $\leqslant 1 + M$).

<div align="right">A. Grothendieck, 1953.</div>

Grothendieck's work in the 1950s exhibited astounding (technical and conceptual) ingenuity and insight that go well-beyond merely defining a key concept, as he did above. In particular, in [?] he conjectured an important phenomenon[2] that was later proved by Dvoretzky [?] (see the discussion in [?]), and his contributions in [?] were transformative (e.g. [?, ?, ?, ?]). The above definition set the stage for decades of (still ongoing) work on the local theory of Banach spaces which had major impact on a wide range of mathematical areas.

The above "softening" of Banach's "linear dimension" is called today *finite representability*, following the terminology of James [?] (and his important contributions on this topic). Given $\alpha \in [1, \infty)$, a Banach space X is said to be α-finitely representable in a Banach space Y if for any $\beta > \alpha$, any finite dimensional subspace of X is β-isomorphic to a subspace of Y (in the notation of the above quote, $\beta = 1 + M$); X is (crudely) finitely representable in Y if there is some $\alpha \in [1, \infty)$ such that X is α-finitely representable in Y. This means that the finite dimensional subspaces of X are not very different from subspaces of Y; if each of X and Y is finitely representable in the other, then this should be viewed as saying that X and Y have the same finite dimensional subspaces (up to a global allowable error that does not depend on the finite dimensional subspace in question). As an important example of the "taming power" of this definition, the *principle of local reflexivity* of Lindenstrauss and Rosenthal [?] asserts that even though sometimes $X^{**} \neq X$, it is always true that X^{**} is 1-finitely representable in X. Thus, while in infinite dimensions X^{**} can be much larger than X, passing to the bidual cannot produce substantially new finite dimensional structures. The aforementioned Dvoretzky theorem [?] asserts that ℓ_2 is 1-finitely representable in any infinite dimensional Banach space. As another example of a landmark theorem on finite representability, Maurey and Pisier strengthened [?] Theorem 4 by showing that ℓ_{q_X} is 1-finitely representable in any infinite dimensional Banach space X.

Isomorphic local linear properties of Banach spaces are defined to be those properties that are preserved under finite representability. As an example, one should keep in mind finitary inequalities such as the cotype condition (4). The formal statement of Ribe's rigidity theorem [?] is

Theorem 5. *Uniformly homeomorphic Banach spaces X and Y are finitely representable in each other.*

The "immediate translation" of finite representability in the above quoted text from [?] is to define that for $\alpha \in [1, \infty)$ a metric space $\mathcal{M}$ is α-finitely representable in a metric space $\mathcal{N}$ if $c_{\mathcal{N}}(\mathcal{C}) \leqslant \alpha$ for *every* finite subset $\mathcal{C} \subseteq \mathcal{M}$. By doing so one does not induce any terminological conflict, because one can show that a Banach space X is (linearly) α-finitely representable in a Banach space Y if and only if X is α-finitely representable in Y when X and Y are viewed as metric spaces. This statement follows from "soft" reasoning that is explained in [?] (relying on a w^*-differentiation argument of Heinrich and Mankiewicz [?] as well as properties of ultrapowers of Banach spaces and the aforementioned principle of local reflexivity), though it also follows from Ribe's original proof of Theorem 5 in [?], and a different (quantitative) approach to this statement was obtained in [?].

1.2.2. *Universality and dichotomies.* Say that a metric space $\mathcal{M}$ is (finitarily) universal if there is $\alpha \geqslant 1$ such that $c_{\mathcal{M}}(\mathcal{F}) \leqslant \alpha$ for *every* finite metric space $\mathcal{F}$. By [?, ?], this requirement holds for some $\alpha \geqslant 1$ if and only if it holds for $\alpha = 1$ (almost-isometric embeddings), so the notion of universality turns out to be insensitive to the underlying distortion bound. Since for every $n \in \mathbb{N}$, any n-point metric space $(\mathcal{F} = \{x_1, \ldots, x_n\}, d_{\mathcal{F}})$ is isometric to a subset of ℓ_∞ via the embedding $(x \in \mathcal{F}) \mapsto (d_{\mathcal{F}}(x, x_i))_{i=1}^n$ (Fréchet embedding [?]), a different way to state the notion of universality is to say that $\mathcal{M}$ is universal if ℓ_∞ is finitely representable in $\mathcal{M}$.

Determining whether a given metric space is universal is a subtle matter. By Theorem 4, for a Banach space X this is the same as asking to determine whether $q_X = \infty$. Such questions include major difficult issues in functional analysis that have been studied for a long time; as notable examples, see the works [?, ?] on the (non)universality of the dual of the disc algebra, and the characterization [?] of Sidon subsets of the dual of a compact Abelian group G in terms of the universality of their span is the space of continuous functions on G. Here are three famous concrete situations in which it is unknown if a certain specific space is universal.

---

[2]This phenomenon was situated within the Ribe program by Bourgain, Figiel and Milman [?], and as such it eventually had ramifications to a well-studied (algorithmic) form of metric dimension reduction through its use to "compress" a finite metric space into a data structure called an *approximate distance oracle* [?]. To date, the only known way to construct such a data structure with constant query time (and even, by now, conjecturally sharp approximation factor [?]) is via the nonlinear Dvoretzky theorem of [?], and thus through the Ribe program. For lack of space, we will not discuss this direction here; see the survey [?].

Question 6 (Pisier's dichotomy problem). For each $n \in \mathbb{N}$ let X_n be an arbitrary linear subspace of ℓ_∞^n satisfying

$$\limsup_{n \to \infty} \frac{\dim(X_n)}{\log n} = \infty. \tag{6}$$

Pisier conjectured [**?**] that (6) forces the ℓ_2 (Pythagorean) direct sum $(X_1 \oplus X_2 \oplus \ldots)_2$ to be universal. By duality, a positive answer to this question is equivalent to the following appealing statement on the geometry of polytopes. For $n \in \mathbb{N}$, suppose that $K \subseteq \mathbb{R}^n$ is an origin-symmetric polytope with $e^{o(n)}$ faces. Then, for each $\delta > 0$ there is $k = k(n, \delta) \in \{1, \ldots, n\}$ with $\lim_{n \to \infty} k(n, \delta) = \infty$, a subspace $F = F(n, \delta)$ of $\mathbb{R}^n$ with $\dim(F) = k$ and a parallelepiped $Q \subseteq F$ (thus, Q is an image of $[-1, 1]^k$ under an invertible linear transformation) such that $Q \subseteq K \cap F \subseteq (1 + \delta)Q$. Hence, a positive answer to Pisier's dichotomy conjecture implies that every centrally symmetric polytope with $e^{o(n)}$ faces has a central section of dimension k going to ∞ (as a function of the specific $o(n)$ dependence in the underlying assumption), which is $(1+\delta)$-close to a polytope (a parallelepiped) with only $O(k)$ faces. The use of "dichotomy" in the name of this conjecture is due to the fact that this conclusion does not hold with $o(n)$ replaced by $O(n)$, as seen by considering polytopes that approximate the Euclidean ball. More generally, by the "isomorphic version" of the Dvoretzky theorem due to Milman and Schechtman [**?**], for every sequence of normed spaces $\{Y_n\}_{n=1}^\infty$ with $\dim(Y_n) = n$ (not only $Y_n = \ell_\infty^n$, which is the case of interest above), and every $k(n) \in \{1, \ldots, n\}$ with $k(n) = O(\log n)$, there is a subspace $X_n \subseteq Y_n$ with $\dim(X_n) = k(n)$ such that the space $(X_1 \oplus X_2 \oplus \ldots)_2$ is isomorphic to a Hilbert space, and hence in particular it is not universal. The best-known bound in Pisier's dichotomy conjecture appears in the forthcoming work of Schechtman and Tomczak-Jaegermann [**?**], where it is shown that the desired conclusion does indeed hold true if (6) is replaced by the stronger assumption $\limsup_{n \to \infty} \dim(X_n)/((\log n)^2 (\log \log n)^2) = \infty$; this is achieved [**?**] by building on ideas of Bourgain [**?**], who obtained the same conclusion if $\limsup_{n \to \infty} \dim(X_n)/(\log n)^4 > 0$. Thus, due to [**?**] the above statement about almost-parallelepiped central sections of centrally symmetric polytopes does hold true if the initial polytope is assumed to have $\exp(o(\sqrt{n}/\log n))$ faces.

Prior to stating the next question on universality (which, apart from its intrinsic interest, plays a role in the ensuing discussion on metric dimension reduction), we need to very briefly recall some basic notation and terminology from optimal transport (see e.g. [**?**, **?**]). Suppose that $(\mathcal{M}, d_{\mathcal{M}})$ is a separable complete metric space and fix $p \in [1, \infty)$. Denote by $\mathsf{P}_1(\mathcal{M})$ the set of all Borel probability measures μ on $\mathcal{M}$ of finite p'th moment, i.e., those Borel probability measure μ on $\mathcal{M}$ for which $\int_{\mathcal{M}} d_{\mathcal{M}}(x, y)^p \, d\mu(y) < \infty$ for all $x \in \mathcal{M}$. A probability measure $\pi \in \mathsf{P}_p(\mathcal{M} \times \mathcal{M})$ is a called a coupling of $\mu, \nu \in \mathsf{P}_p(\mathcal{M})$ if $\mu(A) = \pi(A \times \mathcal{M})$ and $\nu(A) = \pi(\mathcal{M} \times A)$ for every Borel measurable subset $A \subseteq \mathcal{M}$. The Wasserstein-p distance between $\mu, \nu \in \mathsf{P}_p(\mathcal{M})$, denoted $\mathsf{W}_p(\mu, \nu)$, is defined to be the infimum of $(\iint_{\mathcal{M} \times \mathcal{M}} d_{\mathcal{M}}(x, y)^p \, d\pi(x, y))^{1/p}$ over all couplings π of μ, ν. Below, $\mathsf{P}_p(\mathcal{M})$ is always assumed to be endowed with the metric W_p. The following question is from [**?**].

Question 7 (Bourgain's universality problem). Is $\mathsf{P}_1(\mathbb{R}^2)$ universal? This formulation may seem different from the way it is asked in [**?**], but, as explained [**?**, Section 1.5], it is equivalent to it. More generally, is $\mathsf{P}_1(\mathbb{R}^k)$ universal for some integer $k \geqslant 2$ (it is simple to see that $\mathsf{P}_1(\mathbb{R})$ is not universal)? In [**?**] it was proved that $\mathsf{P}_1(\ell_1)$ is universal (see also the exposition in [**?**]). So, it is important here that the underlying space is finite dimensional, though to the best of our knowledge it is also unknown whether $\mathsf{P}_1(\ell_2)$ is universal, or, for that matter, if $\mathsf{P}_1(\ell_p)$ is universal for any $p \in (1, \infty)$. See [**?**] for a (sharp) universality property of $\mathsf{P}_p(\mathbb{R}^3)$ if $p \in (1, 2]$.

For the following open question about universality (which will also play a role in the subsequent discussion on metric dimension reduction), recall the notion [**?**, **?**] of *projective tensor product* of Banach spaces. Given two Banach spaces $(X, \| \cdot \|_X)$ and $(Y, \| \cdot \|_Y)$, their projective tensor product $X \widehat{\otimes} Y$ is the completion of their algebraic tensor product $X \otimes Y$ under the norm whose unit ball is the convex hull of the simple tensors of vectors of norm at most 1, i.e., the convex hull of the set $\{x \otimes y \in X \otimes Y : \|x\|_X, \|y\|_Y \leqslant 1\}$. For example, $\ell_1 \widehat{\otimes} X$ can be naturally identified with $\ell_1(X)$, and $\ell_2 \widehat{\otimes} \ell_2$ can be naturally identified with Schatten–von Neumann trace class S_1 (recall that for $p \in [1, \infty]$, the Schatten–von Neumann trace class S_p is the Banach space [**?**] of those compact linear operators $T : \ell_2 \to \ell_2$ for which $\|T\|_{\mathsf{S}_p} = (\sum_{j=1}^\infty \mathsf{s}_j(T)^p)^{1/p} < \infty$, where $\{\mathsf{s}_j(T)\}_{j=1}^\infty$ are the singular values of T); see the monograph [**?**] for much more on tensor products of Banach spaces.

It is a longstanding endeavour in Banach space theory to understand which properties of Banach spaces are preserved under projective tensor products; see [**?**, **?**] and the references therein for more on this research direction. Deep work of Pisier [**?**] shows that there exist two Banach spaces X, Y that are not universal (even of cotype 2) such that $X \widehat{\otimes} Y$ is universal. The following question was posed by Pisier in [**?**].

Question 8 (universality of projective tensor products). Suppose that $p \in (1,2)$. Is $\ell_p \widehat{\otimes} \ell_2$ universal? We restricted the range of p here because it is simple to check that $\ell_1 \widehat{\otimes} \ell_2 \cong \ell_1(\ell_2)$ is not universal, Tomczak-Jaegermann [?] proved that $\ell_2 \otimes \ell_2 \cong \mathsf{S}_1$ is not universal, and Pisier proved [?] that $\ell_p \widehat{\otimes} \ell_q$ is not universal when $p, q \in [2, \infty)$. It was also asked in [?] if $\ell_2 \widehat{\otimes} \ell_2 \widehat{\otimes} \ell_2$ is universal. The best currently available result in this direction (which will be used below) is that, using the local theory of Banach spaces and recent work on locally decodable codes, it was shown in [?] that if $a, b, c \in (1, \infty)$ satisfy $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} \leqslant 1$, then $\ell_a \widehat{\otimes} \ell_b \widehat{\otimes} \ell_c$ is universal.

The following theorem is a union of several results of [?].

Theorem 9. *The following conditions are equivalent for a metric space* $(\mathcal{M}, d)$.

- $\mathcal{M}$ *is not universal.*
- *There exists* $q = q(\mathcal{M}) \in (0, \infty)$ *with the following property. For every* $n \in \mathbb{N}$ *there is* $m = m(n, \mathcal{M}, q) \in \mathbb{N}$ *such that any collection of points* $\{x_w\}_{w \in \mathbb{Z}_{2m}^n}$ *in* $\mathcal{M}$ *satisfies the inequality*

$$\sum_{i=1}^{n} \sum_{w \in \mathbb{Z}_{2m}^n} \frac{d(x_{w+me_i}, x_w)^2}{m^2} \lesssim_{X,q} \frac{n^{1-\frac{2}{q}}}{3^n} \sum_{\varepsilon \in \{-1,0,1\}^n} \sum_{w \in \mathbb{Z}_{2m}^n} d(x_{w+\varepsilon}, x)^2. \tag{7}$$

  Here $e_1, \ldots, e_n$ *are the standard basis of* $\mathbb{R}^n$ *and addition (in the indices) is modulo* $2m$.
- *There is* $\theta(\mathcal{M}) \in (0, \infty)$ *with the following property. For arbitrarily large* $n \in \mathbb{N}$ *there exists an* n-point *metric space* $(\mathcal{B}_n, d_n)$ *such that*

$$\mathsf{c}_{\mathcal{M}}(\mathcal{B}_n) \gtrsim_{\mathcal{M}} (\log n)^{\theta(\mathcal{M})}. \tag{8}$$

Moreover, if we assume that $\mathcal{M}$ *is a Banach space rather than an arbitrary metric space, then for any* $q \in (0, \infty]$ *the validity of* (7) *(as stated, i.e., for each* $n \in \mathbb{N}$ *there is* $m \in \mathbb{N}$ *for which* (7) *holds for every configuration* $\{x_w\}_{w \in \mathbb{Z}_{2m}^n}$ *of points in* $\mathcal{M}$) *is equivalent to* (4), *and hence, in particular, the infimum over those* $q \in (0, \infty]$ *for which* (7) *holds true is equal to the infimal cotype* $q_{\mathcal{M}}$ *of* $\mathcal{M}$.

The final sentence of Theorem 9 is an example of a successful step in the Ribe program, because it reformulates the local linear invariant (4) in purely metric terms, namely as the quadratic geometric inequality (7) that imposes a restriction on the behavior of pairwise distances within any configuration of $(2m)^n$ points $\{x_w\}_{w \in \mathbb{Z}_{2m}^n}$ (indexed by the discrete torus $\mathbb{Z}_{2m}^n$) in the given Banach space.

With this at hand, one can consider (7) to be a property of a metric space, while initially (4) made sense only for a normed space. As in the case of (4), if $q = \infty$, then (7) holds in any metric space $(\mathcal{M}, d_{\mathcal{M}})$ (for any $m \in \mathbb{Z}$, with the implicit constant in (7) being universal); by its general nature, such a statement must of course be nothing more than a formal consequence of the triangle inequality, as carried out in [?]. So, the validity of (7) for $q < \infty$ could be viewed as an asymptotic (randomized) enhancement of the triangle inequality in $(\mathcal{M}, d_{\mathcal{M}})$; by considering the the canonical realization of $\mathbb{Z}_{2m}^n$ in $\mathbb{C}^n$, namely the points $\{(\exp(\pi i w_1/m), \ldots, \exp(\pi i w_n/m))\}_{w \in \mathbb{Z}_{2m}^n}$, equipped with the metric inherited from $\ell_\infty^n(\mathbb{C})$, one checks that not every metric space satisfies this requirement. The equivalence of the first two bullet points in Theorem 9 shows that once one knows that a metric space is not universal, one deduces the validity of such an enhancement of the triangle inequality. This is an analogue of Theorem 4 of Maurey and Pisier for general metric spaces.

The equivalence of the first and third bullet points in Theorem 9 yields the following dichotomy. If one finds a finite metric space $\mathcal{F}$ such that $\mathsf{c}_{\mathcal{F}}(\mathcal{M}) > 1$, then there are arbitrary large finite metric spaces whose minimal distortion in $\mathcal{M}$ is at least a fixed positive power (depending on $\mathcal{M}$) of the logarithm their cardinality. Hence, for example, if every n-point metric space embeds into $\mathcal{M}$ with distortion $O(\log \log n)$, then actually for any $\delta > 0$, every finite metric space embeds into $\mathcal{M}$ with distortion $1 + \delta$. See [?, ?, ?, ?] for more on metric dichotomies of this nature, as well as a quite delicate counterexample [?] for a natural variant for trees (originally asked by C. Fefferman). It remains a mystery [?] if the power of the logarithm $\theta(\mathcal{M})$ in Theorem 9 could be bounded from below by a universal positive constant, as formulated in the following open question.

Question 10 (metric cotype dichotomy problem). Is there a universal constant $\theta > 0$ such that in Theorem 9 one could take $\theta(\mathcal{M}) > \theta$. All examples that have been computed thus far leave the possibility that even $\theta(\mathcal{M}) \geqslant 1$, which would be sharp (for $\mathcal{M} = \ell_2$) by Bourgain's embedding theorem [?]. Note, however, that in [?] it is asked whether for the Wasserstein space $\mathsf{P}_p(\mathbb{R}^3)$ we have $\liminf_{p \to 1} \theta(\mathsf{P}_p(\mathbb{R}^3)) = 0$. If this were true, then it would resolve the metric cotype dichotomy problem negatively. It would be interesting to understand the bi-Lipschitz structure of these spaces of measures on $\mathbb{R}^3$ regardless of this context, due to their independent importance.

Theorem 9 is a good illustration of a "vanilla" accomplishment of the Ribe program, since it obtains a metric reformulation of a key isomorphic linear property of metric spaces, and also proves statements about general metric spaces which are inspired by the analogies with the linear theory that the Ribe program is aiming for. However, even in this particular setting of metric cotype, Theorem 9 is only a part of the full picture, as it has additional purely metric ramifications. Most of these rely on a delicate issue that has been suppressed in the above statement of Theorem 9, namely that of understanding the asymptotic behavior of $m = m(n, \mathcal{M}, q)$ in (7). This matter is not yet fully resolved even when $\mathcal{M}$ is a Banach space [**?**, **?**], and generally such questions seem to be quite challenging (see [**?**, **?**, **?**] for related issues). Thus far, whenever this question was answered for specific (classes of) metric spaces, it led to interesting geometric applications; e.g. its resolution for certain Banach spaces in [**?**] was used in [**?**] to answer a longstanding question [**?**] about quasisymmetric embeddings, and its resolution for Alexandrov spaces of (global) nonpositive curvature [**?**] (see e.g. [**?**, **?**] for the relevant background) in the forthcoming work [**?**] is used there to answer a longstanding question about the coarse geometry of such Alexandrov spaces.

1.3. **Metric dimension reduction.** By its nature, many aspects of the local theory of Banach spaces involve describing phenomena that rely on dimension-dependent estimates. In the context of the Ribe program, the goal is to formulate/conjecture analogous phenomena for metric spaces, which is traditionally governed by asking Banach space-inspired questions about a finite metric space $(\mathcal{M}, d_\mathcal{M})$ in which $\log |\mathcal{M}|$ serves as a replacement for the dimension. This analogy arises naturally also in the context of the bi-Lipschitz embedding problem into $\mathbb{R}^k$ (Problem 2); see Remark 39 below. Early successful instances of this analogy can be found in the work of Marcus and Pisier [**?**], as well as the aforementioned work of Johnson and Lindenstrauss [**?**]. However, it should be stated at the outset that over the years it became clear that while making this analogy is the right way to get "on track" toward the discovery of fundamental metric phenomena, from the perspective of the Ribe program the reality is much more nuanced and, at times, even unexpected and surprising.

Johnson and Lindenstrauss asked [**?**, Problem 3] whether every finite metric space $\mathcal{M}$ embeds with distortion $O(1)$ into some normed space $X_\mathcal{M}$ (which is allowed to depend on $\mathcal{M}$) of dimension $\dim(X_\mathcal{M}) \lesssim \log |\mathcal{M}|$. In addition to arising from the above background, this question is motivated by [**?**, Problem 4], which asks if the Euclidean distortion of every finite metric space $\mathcal{M}$ satisfies $\mathsf{c}_2(\mathcal{M}) \lesssim \sqrt{\log |\mathcal{M}|}$. If so, this would have served as a very satisfactory metric analogue of John's theorem [**?**], which asserts that any finite dimensional normed space X is $\sqrt{\dim(X)}$-isomorphic to a subspace of ℓ_2. Of course, John's theorem shows that a positive answer to the former question [**?**, Problem 3] formally implies a positive answer to the latter question [**?**, Problem 4].

The aforementioned Johnson–Lindenstrauss lemma [**?**] (JL lemma, in short) shows that, at least for finite subsets of a Hilbert space, the answer to the above stated [**?**, Problem 3] is positive.

Theorem 11 (JL lemma). *For each $n \in \mathbb{N}$ and $\alpha \in (1, \infty)$, there is $k \in \{1, \ldots, n\}$ with $k \lesssim_\alpha \log n$ such that any n-point subset of ℓ_2 embeds into ℓ_2^k with distortion α.*

We postpone discussion of this fundamental geometric fact to Section 2 below, where it is examined in detail and its proof is presented. Beyond Hilbert spaces, there is only one other example (and variants thereof) of a Banach space for which it is currently known that [**?**, Problem 3] has a positive answer for any of its finite subsets, as shown in the following theorem from [**?**].

Theorem 12. *There is a Banach space $\mathfrak{T}^{(2)}$ which is not isomorphic to a Hilbert space yet it has the following property. For every finite subset $\mathcal{C} \subseteq \mathfrak{T}^{(2)}$ there is $k \in \{1, \ldots, n\}$ with $k \lesssim \log |\mathcal{C}|$ and a k-dimensional linear subspace F of $\mathfrak{T}^{(2)}$ such that $\mathcal{C}$ embeds into F with $O(1)$ distortion.*

The space $\mathfrak{T}^{(2)}$ of Theorem 12 is not very quick to describe, so we refer to [**?**] for the details. It suffices to say here that this space is the 2-convexification of the classical Tsirelson space [**?**, **?**], and that the proof that it satisfies the stated dimension reduction result is obtained in [**?**] via a concatenation of several (substantial) structural results in the literature; see Section 4 in [**?**] for a discussion of variants of this construction, as well as related open questions. The space $\mathfrak{T}^{(2)}$ of Theorem 12 is not isomorphic to a Hilbert space, but barely so: it is explained in [**?**] that for every $n \in \mathbb{N}$ there exists an n-dimensional subspace F_n of $\mathfrak{T}^{(2)}$ with $\mathsf{c}_2(F_n) \geqslant e^{c\mathsf{Ack}^{-1}(n)}$, where $c > 0$ is a universal constant and $\mathsf{Ack}^{-1}(\cdot)$ is the inverse of the Ackermann function from computability theory (see e.g. [**?**, Appendix B]). So, indeed $\lim_{n \to \infty} \mathsf{c}_2(F_n) = \infty$, but at a tremendously slow rate.

Remarkably, despite major scrutiny for over 3 decades, it remains unknown if [**?**, Problem 3] has a positive answer for subsets of *any* non-universal classical Banach space. In particular, the following question is open.

Question 13. Suppose that $p \in [1, \infty) \smallsetminus \{2\}$. Are there $\alpha = \alpha(p), \beta = \beta(p) \in [1, \infty)$ such that for any $n \in \mathbb{N}$, every n-point subset of ℓ_p embeds with distortion α into some k-dimensional normed space with $k \leqslant \beta \log n$?

It is even open if in Question 13 one could obtain a bound of $k = o(n)$ for any fixed $p \in [1, \infty) \smallsetminus \{2\}$. Using John's theorem as above, a positive answer to Question 13 would imply that $c_2(\mathcal{C}) \lesssim_p \sqrt{\log |\mathcal{C}|}$ for any finite subset $\mathcal{C}$ of ℓ_p. At present, such an embedding statement is not known for *any* $p \in [1, \infty) \smallsetminus \{2\}$, though for $p \in [1, 2]$ it is known [**?**] that any n-point subset of ℓ_p embeds into ℓ_2 with distortion $(\log n)^{1/2+o(1)}$; it would be interesting to obtain any $o(\log n)$ bound here for any fixed $p \in (2, \infty)$, which would be a "nontrivial" asymptotic behavior in light of the following general theorem [**?**].

Theorem 14 (Bourgain's embedding theorem). $c_2(\mathcal{M}) \lesssim \log |\mathcal{M}|$ *for every finite metric space* $\mathcal{M}$.

The above questions from [**?**] were the motivation for the influential work [**?**], where Theorem 14 was proved. Using a probabilistic construction and the JL lemma, it was shown in [**?**] that Theorem 14 is almost sharp in the sense that there are arbitrarily large n-point metric spaces $\mathcal{M}_n$ for which $c_2(\mathcal{M}_n) \gtrsim (\log n)/\log\log n$. By John's theorem, for every $\alpha \geqslant 1$, if X is a finite dimensional normed space and $c_X(\mathcal{M}_n) \leqslant \alpha$, then $c_2(\mathcal{M}_n) \leqslant \alpha\sqrt{\dim(X)}$. Therefore the above lower bound on $c_2(\mathcal{M}_n)$ implies that $\dim(X) \gtrsim (\log n)^2/(\alpha^2(\log\log n)^2)$.

The achievement of [**?**] is thus twofold. Firstly, it discovered Theorem 14 (via the introduction of an influential randomized embedding method), which is the "correct" metric version of John's theorem in the Ribe program. The reality turned out to be more nuanced in the sense that the answer is not quite as good as the $O(\sqrt{\log n})$ that was predicted in [**?**], but the $O(\log n)$ of Theorem 14 is still a strong and useful phenomenon that was discovered through the analogy that the Ribe program provided. Secondly, we saw above that [**?**, Problem 3] was *disproved* in [**?**], though the "bad news" that follows from [**?**] is only mildly worse than the $O(\log n)$ dimension bound that [**?**, Problem 3] predicted, namely a dimension lower bound that grows quite slowly, not faster than $(\log n)^{O(1)}$. Curiously, the very availability of strong dimension reduction in ℓ_2 through the JL lemma is what was harnessed in [**?**] to deduce that any "host normed space" that contains $\mathcal{M}_n$ with $O(1)$ distortion must have dimension at least of order $(\log n/\log\log n)^2 \gg \log n$. Naturally, in light of these developments, the question of understanding what is the correct asymptotic behavior of the smallest $k(n) \in \mathbb{N}$ such that any n-point metric space embeds with distortion $O(1)$ into a $k(n)$-dimensional normed space was raised in [**?**].

In order to proceed, it would be convenient to introduce some notation and terminology.

Definition 15 (metric dimension reduction modulus). *Fix* $n \in \mathbb{N}$ *and* $\alpha \in [1, \infty)$. *Suppose that* $(X, \|\cdot\|_X)$ *is a normed space. Denote by* $k_n^\alpha(X)$ *the minimum* $k \in \mathbb{N}$ *such that for* any $\mathcal{C} \subseteq X$ *with* $|\mathcal{C}| = n$ *there exists a* k-dimensional linear subspace $F_{\mathcal{C}}$ of X into which $\mathcal{C}$ embeds with distortion α.

The quantity $k_n^\alpha(\ell_\infty)$ was introduced by Bourgain [**?**] under the notation $\psi_\alpha(n) = k_n^\alpha(\ell_\infty)$; see also [**?**, **?**] where this different notation persists, though for the sake of uniformity of the ensuing discussion we prefer not to use it here because we will treat $X \neq \ell_\infty$ extensively. [**?**] focused for concreteness on the arbitrary value $\alpha = 2$, and asked for the asymptotic behavior $k_n^2(\ell_\infty)$ as $n \to \infty$.

An n-point subset of $X = \ell_\infty$ is nothing more than a general n-point metric space, via the aforementioned isometric Fréchet embedding. In the same vein, a k-dimensional linear subspace of ℓ_∞ is nothing more than a general k-dimensional normed space $(F, \|\cdot\|_F)$ via the linear isometric embedding $(x \in F) \mapsto (x_i^*(x))_{i=1}^\infty$, where $\{x_i^*\}_{i=1}^\infty$ is an arbitrary sequence of linear functionals on F that are dense in the unit sphere of the dual space F^*. Thus, the quantity $k_n^\alpha(\ell_\infty)$ is the smallest $k \in \mathbb{N}$ such that every n-point metric space $(\mathcal{M}, d_{\mathcal{M}})$ can be realized with distortion at most α as a subset of $(\mathbb{R}^k, \|\cdot\|_m)$ for some norm $\|\cdot\|_m : \mathbb{R}^k \to \mathbb{R}^k$ on $\mathbb{R}^k$ (which, importantly, is allowed to be adapted to the initial metric space $\mathcal{M}$), i.e., $c_{(\mathbb{R}^k, \|\cdot\|_m)}(\mathcal{M}, d_{\mathcal{M}}) \leqslant \alpha$. This is precisely the quantity that [**?**, Problem 3] asks about, and above we have seen that [**?**] gives the lower bound

$$k_n^\alpha(\ell_\infty) \gtrsim \left(\frac{\log n}{\alpha \log\log n} \right)^2. \tag{9}$$

Theorem 16 below is a summary of the main nontrivial[3] bounds on the modulus $k_n^\alpha(X)$ that are currently known for specific Banach spaces X. Since such a "combined statement" contains a large amount information and covers a lot of the literature on this topic, we suggest reading it in tandem with the subsequent discussion, which includes further clarifications and explanation of the history of the respective results. A "take home"

---

[3]Trivially $k_n^\alpha(X) \leqslant n - 1$, by considering in Definition 15 the subspace $F_{\mathcal{C}} = \mathbf{span}(\mathcal{C} - x_0)$ for any fixed $x_0 \in \mathcal{C}$.

message from the statements below is that despite major efforts by many researchers, apart from information on metric dimension reduction for ℓ_2, the space $\mathcal{T}^{(2)}$ of Theorem 12, ℓ_∞, ℓ_1 and S_1, nothing is known for other spaces (even for ℓ_1 and S_1 more remains to be done, notably with respect to bounding $\mathsf{k}_n^\alpha(\ell_1), \mathsf{k}_n^\alpha(\mathsf{S}_1)$ from above).

Theorem 16 (summary of the currently known upper and lower bounds on metric dimension reduction)*. There exist universal constants $c, C > 0$ such that the following assertions hold true for every integer $n \geqslant 20$.*

(1) In the Hilbertian setting, we have the sharp bounds

$$\forall\, \alpha \geqslant 1 + \frac{1}{\sqrt[3]{n}}, \qquad \mathsf{k}_n^\alpha(\ell_2) \asymp \frac{\log n}{\log(1 + (\alpha - 1)^2)} \asymp \max\left\{ \frac{\log n}{(\alpha - 1)^2}, \frac{\log n}{\log \alpha} \right\}. \tag{10}$$

(2) For the space $\mathcal{T}^{(2)}$ of Theorem 12, there exists $\alpha_0 \in [1, \infty)$ for which $\mathsf{k}_n^{\alpha_0}(\mathcal{T}^{(2)}) \asymp \log n$.

(3) For ℓ_∞, namely in the setting of $[?, \text{Problem } 3]$, we have

$$\forall\, \alpha \in [1, 2), \qquad \mathsf{k}_n^\alpha(\ell_\infty) \asymp n, \tag{11}$$

and

$$\forall\, \alpha \geqslant 2, \qquad n^{\frac{c}{\alpha}} + \frac{\log n}{\log\left(\frac{\log n}{\alpha} + \frac{\alpha \log \log n}{\log n} \right)} \lesssim \mathsf{k}_n^\alpha(\ell_\infty) \lesssim \frac{n^{\frac{C}{\alpha}} \log n}{\log\left(1 + \frac{\alpha}{\log n}\right)}. \tag{12}$$

(4) For ℓ_1, we have

$$\forall\, \alpha \geqslant 1, \qquad n^{\frac{c}{\alpha^2}} + \frac{\log n}{\log(\alpha + 1)} \lesssim \mathsf{k}_n^\alpha(\ell_1) \lesssim \frac{n}{\alpha}. \tag{13}$$

Moreover, if $\alpha \geqslant 2C\sqrt{\log n} \log \log n$, then we have the better upper bound

$$\mathsf{k}_n^\alpha(\ell_1) \lesssim \frac{\log n}{\log\left(\frac{\alpha}{C\sqrt{\log n} \log \log n}\right)}. \tag{14}$$

(5) For the Schatten–von Neumann trace class S_1, we have

$$\forall\, \alpha \geqslant 1, \qquad \mathsf{k}_n^\alpha(\mathsf{S}_1) \gtrsim n^{\frac{c}{\alpha^2}} + \frac{\log n}{\log(\alpha + 1)}. \tag{15}$$

The bound $\mathsf{k}_n^\alpha(\ell_2) \lesssim (\log n)/\log(1 + (\alpha - 1)^2)$ in (10) restates Theorem 11 (the JL lemma) with the implicit dependence on α now stated explicitly; it actually holds for every $\alpha > 1$, as follows from the original proof in $[?]$ and explained in Section 2 below. The restriction $\alpha \geqslant 1 + 1/\sqrt[3]{n}$ in (10) pertains only to the corresponding lower bound on $\mathsf{k}_n^\alpha(\ell_2)$, which exhibits different behaviors in the low-distortion and high-distortion regimes.

Despite scrutiny of many researchers over the past 3 decades, only very recently the dependence on α in the JL lemma when α is arbitrarily close to 1 but independent of n was proved to be sharp by Larsen and Nelson $[?]$ (see $[?, ?, ?]$ for earlier results in this direction, as well as the subsequent work $[?]$). This is so even when α is allowed to tend to 1 with n, and even in a somewhat larger range than the requirement $\alpha \geqslant 1 + 1/\sqrt[3]{n}$ in (10) (see $[?]$ for the details), though there remains a small range of values of α (n-dependent, very close to 1) for which it isn't currently known what is the behavior of $\mathsf{k}_n^\alpha(\ell_2)$. The present article is focused on embeddings that permit large errors, and in particular in ways to prove impossibility results even if large errors are allowed. For this reason, we will not describe here the ideas of the proof in $[?]$ that pertains to the almost-isometric regime.

For, say, $\alpha \geqslant 2$, it is much simpler to see that the $\mathsf{k}_n^\alpha(\ell_2) \gtrsim (\log n)/\log \alpha$, in even greater generality that also explains the appearance of the term $(\log n)/\log(\alpha + 1)$ in (13) and (15). One could naturally generalize Definition 15 so as to introduce the following notation for *relative metric dimension reduction moduli*. Let $\mathcal{F}$ be a family of metric spaces and $\mathcal{Y}$ be a family of normed spaces. For $n \in \mathbb{N}$ and $\alpha \in [1, \infty)$, denote by $\mathsf{k}_n^\alpha(\mathcal{F}, \mathcal{Y})$ the minimum $k \in \mathbb{N}$ such that for every $\mathcal{M} \in \mathcal{F}$ with $|\mathcal{M}| = n$ there exists $Y \in \mathcal{Y}$ with $\dim(Y) = k$ such that $\mathsf{c}_Y(\mathcal{M}) \leqslant \alpha$. When $\mathcal{F}$ is the collection of all the finite subsets of a fixed Banach space X, and $\mathcal{Y}$ is the collection of all the finite-dimensional linear subspaces of a fixed Banach space Y, we use the simpler notation $\mathsf{k}_n^\alpha(\mathcal{F}, \mathcal{Y}) = \mathsf{k}_n^\alpha(X, Y)$. Thus, the modulus $\mathsf{k}_n^\alpha(X)$ of Definition 15 coincides with $\mathsf{k}_n^\alpha(X, X)$. Also, under this notation Question (13) asks if for $p \in [1, \infty) \smallsetminus \{2\}$ we have $\mathsf{k}_n^\alpha(\ell_p, \ell_\infty) \lesssim_p \log n$ for some $1 \leqslant \alpha \lesssim_p 1$. The study of the modulus $\mathsf{k}_n^\alpha(\mathcal{F}, \mathcal{Y})$ is essentially a completely unexplored area, partially because even our understanding

of the "vanilla" dimension reduction modulus $\mathsf{k}_n^{\alpha}(X)$ is currently very limited. By a short volumetric argument that is presented in Section 2 below, every infinite dimensional Banach space X satisfies

$$\forall (n, \alpha) \in \mathbb{N} \times [1, \infty), \qquad \mathsf{k}_n^{\alpha}(X, \ell_{\infty}) \geqslant \frac{\log n}{\log(\alpha + 1)}. \tag{16}$$

Hence also $\mathsf{k}_n^{\alpha}(X) \geqslant (\log n)/\log(\alpha + 1)$, since (16) rules out embeddings into any normed space of dimension less than $(\log n)/\log(\alpha + 1)$, rather than only into such spaces that are also subspaces of X.

Using an elegant Fourier-analytic argument, Arias-de-Reyna and Rodríguez-Piazza proved in [?] that for every $\alpha \in [1, 2)$ we have $\mathsf{k}_n^{\alpha}(\ell_{\infty}) \gtrsim (2-\alpha)n$. This was slightly improved by Matoušek [?] to $\mathsf{k}_n^{\alpha}(\ell_{\infty}) \gtrsim n$, i.e., he showed that the constant multiple of n actually remains bounded below by a positive constant as $\alpha \to 2^-$ (curiously, the asymptotic behavior of $\mathsf{k}_n^2(\ell_{\infty})$ remains unknown). These results establish (11). So, for sufficiently small distortions one cannot hope to embed every n-point metric space into some normed space of dimension $o(n)$. For larger distortions (our main interest), it was conjectured in [?] that $\mathsf{k}_n^{\alpha}(\ell_{\infty}) \lesssim (\log n)^{O(1)}$ if $\alpha > 2$.

The bounds in (12) refute this conjecture of [?], since they include the lower bound $\mathsf{k}_n^{\alpha}(\ell_{\infty}) \gtrsim n^{c/\alpha}$, which is a landmark achievement of Matoušek [?] (obtained a decade after Bourgain asked about the asymptotics here and over a decade after Johnson and Lindenstrauss posed the question whether $\mathsf{k}_n^{\alpha}(\ell_{\infty}) \lesssim_{\alpha} \log n$). It is, of course, an exponential improvement over Bourgain's bound (9). Actually, in the intervening period Linial, London and Rabinovich [?] removed the iterated logarithm in the lower bound of [?] by showing that Theorem 14 (Bourgain's embedding) is sharp up to the value of the implicit universal constant. By the same reasoning as above (using John's theorem), this also removed the iterated logarithm from the denominator in (9), i.e., [?] established that $\mathsf{k}_n^{\alpha}(\ell_{\infty}) \gtrsim (\log n)^2/\alpha^2$. This was the best-known bound prior to [?].

Beyond proving a fundamental geometric theorem, which, as seen in (12), is optimal up to the constant in the exponent, this work of Matoušek is important because it injected a refreshing approach from real algebraic geometry into this area, which was previously governed by considerations from analysis, geometry, probability and combinatorics. Section 4 covers this outstanding contribution in detail, and obtains the following stronger statement that wasn't previously noticed in the literature but follows from an adaptation of Matoušek's ideas.

Theorem 17 (impossibility of coarse dimension reduction). *There is a universal constant $c \in (0, \infty)$ with the following property. Suppose that $\omega, \Omega : [0, \infty) \to [0, \infty)$ are increasing functions that satisfy $\omega(s) \leqslant \Omega(s)$ for all $s \in [0, \infty)$, as well as $\lim_{s \to \infty} \omega(s) = \infty$. Define*

$$\beta(\omega, \Omega) \stackrel{\text{def}}{=} \sup_{s \in (0, \infty)} \frac{s}{\omega^{-1}(2\Omega(s))} \in (0, 1). \tag{17}$$

For arbitrarily large $n \in \mathbb{N}$ there is a metric space $(\mathcal{M}, d_{\mathcal{M}}) = (\mathcal{M}(n, \omega, \Omega), d_{\mathcal{M}(n, \omega, \Omega)})$ with $|\mathcal{M}| = 3n$ such that for any normed space $(X, \|\cdot\|_X)$, if there exists $f : \mathcal{M} \to X$ which satisfies

$$\forall x, y \in \mathcal{M}, \qquad \omega(d_{\mathcal{M}}(x, y)) \leqslant \|f(x) - f(y)\|_X \leqslant \Omega(d_{\mathcal{M}}(x, y)), \tag{18}$$

then necessarily

$$\dim(X) \gtrsim n^{c\beta(\omega, \Omega)}. \tag{19}$$

A mapping that satisfies (18) is called a *coarse embedding* (with moduli ω, Ω), as introduced in Gromov's seminal work [?] and studied extensively ever since, with a variety of interesting applications (see the monographs [?, ?, ?] and the references therein). The bi-Lipschitz requirement (1) corresponds to $\omega(s) = \tau s$ and $\Omega(s) = \alpha\tau s$ in (18), in which case (19) becomes Matoušek's aforementioned lower bound on $\mathsf{k}_n^{\alpha}(\ell_{\infty})$. Theorem 17 asserts that there exist arbitrarily large finite metric spaces that cannot be embedded even with a very weak (coarse) guarantee into any low-dimensional normed space, with the dimension of the host space being forced to be at least a power of their cardinality, which is exponentially larger than the logarithmic behavior that one would predict from the natural ball-covering requirement that is induced by low-dimensionality (see the discussion of the doubling condition in Section 3, as well as the proof of (16) in Section 2).

Remark 18. Consider the following special case of Theorem 17. Fix $\theta \in (0, 1]$ and let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space. It is straightforward to check that $d_{\mathcal{M}}^{\theta} : \mathcal{M} \times \mathcal{M} \to [0, \infty)$ is also a metric on $\mathcal{M}$. The metric space $(\mathcal{M}, d_{\mathcal{M}}^{\theta})$ is commonly called the θ-*snowflake* of $\mathcal{M}$ (in reference to the von Koch snowflake curve; see e.g. [?]) and it is denoted $\mathcal{M}^{\theta}$. Given $\alpha \geqslant 1$, the statement that $\mathcal{M}^{\theta}$ embeds with distortion α into a normed space $(X, \|\cdot\|_X)$ is the same as the requirement (18) with $\omega(s) = s^{\theta}$ and $\Omega(s) = \alpha s^{\theta}$. Hence, by Theorem 17 there exist arbitrarily large n-point metric spaces $\mathcal{M}_n = \mathcal{M}_n(\alpha, \theta)$ such that if $\mathcal{M}_n^{\theta}$ embeds with distortion

α into some k-dimensional normed space, then $k \geqslant n^{c/(2\alpha)^{1/\theta}}$. Conversely, Remark 21 below shows that for every $n \in \mathbb{N}$ and $\alpha > 1$, the θ-snowflake of any n-point metric space embeds with distortion α into a normed space X with $\dim(X) \lesssim_{\alpha,\theta} n^{C/\alpha^{1/\theta}}$. So, the bound (19) of Theorem 17 is quite sharp even for embeddings that are not bi-Lipschitz, though we did not investigate the extent of its sharpness for more general moduli $\omega, \Omega : [0, \infty) \to [0, \infty)$.

At this juncture, it is natural to complement the (coarse) strengthening in Theorem 17 of Matoušek's bound $\mathsf{k}_n^\alpha(\ell_\infty) \geqslant n^{c/\alpha}$ by stating the following different type of strengthening, which we recently obtained in [**?**].

Theorem 19 (impossibility of average dimension reduction). *There is a universal constant $c \in (0, \infty)$ with the following property. For arbitrarily large $n \in \mathbb{N}$ there is an n-point metric space $(\mathcal{M}, d_{\mathcal{M}})$ such that for any normed space $(X, \|\cdot\|_X)$ and any $\alpha \in [1, \infty)$, if there exists $f : \mathcal{M} \to X$ which satisfies $\|f(x) - f(y)\|_X \leqslant \alpha d_{\mathcal{M}}(x, y)$ for all $x, y \in X$, yet $\frac{1}{n^2} \sum_{x,y \in X} \|f(x) - f(y)\|_X \geqslant \frac{1}{n^2} \sum_{x,y \in X} d_{\mathcal{M}}(x, y)$, then necessarily $\dim(X) \geqslant n^{c/\alpha}$.*

An n-point metric space $\mathcal{M}$ as in Theorem 19 is intrinsically high dimensional even on average, in the sense that if one wishes to assign in an α-Lipschitz manner to each point in $\mathcal{M}$ a vector in some normed space X such that the average distance in the image is the same as the average distance in $\mathcal{M}$, then this forces the abient dimension to satisfy $\dim(X) \geqslant n^{c/\alpha}$. Prior to Theorem 19, the best-known bound here was $\dim(X) \gtrsim (\log n)^2/\alpha^2$, namely the aforementioned lower bound $\mathsf{k}_n^\alpha(\ell_\infty) \gtrsim (\log n)^2/\alpha^2$ of Linial, London and Rabinovich [**?**] actually treated the above "average distortion" requirement rather than only the (pairwise) bi-Lipschitz requirement.

Remark 20. The significance of Theorem 19 will be discussed further in Section 5 below; see also [**?**, **?**]. In Section 5 we will present a new proof of Theorem 19 that is different from (though inspired by) its proof in [**?**]. It suffices to say here that the proof of Theorem 19 is conceptually different from Matoušek's approach [**?**]. Namely, in contrast to the algebraic/topological argument of [**?**], the proof of Theorem 19 relies on the theory of nonlinear spectral gaps, which is also an outgrowth of the Ribe program; doing justice to this theory and its ramifications is beyond the scope of the present article (see [**?**] and the references therein), but the basics are recalled in Section 5. Importantly, the proof of Theorem 19 obtains a criterion for determining if a given metric space $\mathcal{M}$ satisfies its conclusion, namely $\mathcal{M}$ can be taken to be the shortest-path metric of any bounded degree graph with a spectral gap. This information is harnessed in the forthcoming work [**?**] to imply that finite-dimensional normed spaces have a structural proprty (a new type of hierarchical partitioning scheme) which has implications to the design of efficient data structures for approximate nearest neighbor search, demonstrating that the omnipresent "curse of dimensionality" is to some extent absent from this fundamental algorithmic task.

In the intervening period between Bourgain's work [**?**] and Matousek's solution [**?**], the question of determining the asymptotic behavior of $\mathsf{k}_n^\alpha(\ell_\infty)$ was pursued by Johnson, Lindenstrauss and Schechtman, who proved in [**?**] that $\mathsf{k}_n^\alpha(\ell_\infty) \lesssim_\alpha n^{C/\alpha}$ for some universal constant $C > 0$. They demonstrated this by constructing for every n-point metric space $\mathcal{M}$ a normed space $X_{\mathcal{M}}$, which they (probabilistically) tailored to the given metric space $\mathcal{M}$, with $\dim(X_{\mathcal{M}}) \lesssim_\alpha n^{C/\alpha}$ and such that $\mathcal{M}$ embeds into $X_{\mathcal{M}}$ with distortion α. Subsequently, Matoušek showed [**?**] via a different argument that one could actually work here with $X_{\mathcal{M}} = \ell_\infty^k$ for $k \in \mathbb{N}$ satisfying $k \lesssim_\alpha n^{C/\alpha}$, i.e., in order to obtain this type of upper bound on the asymptotic behavior of $\mathsf{k}_n^\alpha(\ell_\infty)$ one does not need to adapt the target normed space to the metric space $\mathcal{M}$ that is being embedded. The implicit dependence on α here, as well as the constant C in the exponent, were further improved in [**?**]. For $\alpha = O((\log n)/\log \log n)$, the upper bound on $\mathsf{k}_n^\alpha(\ell_\infty)$ that appears in (12) is that of [**?**], and for the remaining values of α it is due to a more recent improvement over [**?**] by Abraham, Bartal and Neiman [**?**] (specifically, the upper bound in (12) is a combination of Theorem 5 and Theorem 6 of [**?**]).

Remark 21. An advantage of the fact [**?**] that one could take $X_{\mathcal{M}} = \ell_\infty^k$ rather than the more general normed space of [**?**] is that it quickly implies the optimality of the lower bound from Remark 18 on dimension reduction of snowflakes. Fix $n \in \mathbb{N}$, $\alpha > 1$ and $\theta \in (0, 1]$. Denote $\delta = \min\{\sqrt{\alpha} - 1, 1\}$, so that $\alpha \asymp \alpha/(1 + \delta) > 1$. By [**?**], given an n-point metric space $(\mathcal{M}, d_{\mathcal{M}})$ there is an integer $k \lesssim n^{c/\alpha^{1/\theta}}$ and $f = (f_1, \ldots, f_k) : \mathcal{M} \to \mathbb{R}^k$ such that

$$\forall\, x, y \in \mathcal{M}, \qquad d_{\mathcal{M}}(x, y) \leqslant \|f(x) - f(y)\|_{\ell_\infty^k} \leqslant \left(\frac{\alpha}{1 + \delta}\right)^{\frac{1}{\theta}} d_{\mathcal{M}}(x, y).$$

Hence (here it becomes useful that we are dealing with the ℓ_∞^k norm, as it commutes with powering),

$$\forall\, x, y \in \mathcal{M}, \qquad d_\mathcal{M}(x,y)^\theta \leqslant \max_{i \in \{1,\dots,k\}} |f_i(x) - f_i(y)|^\theta \leqslant \frac{\alpha}{1+\delta} d_\mathcal{M}(x,y)^\theta.$$

By works of Kahane [**?**] and Talagrand [**?**], there is $m = m(\delta, \theta)$ and a mapping (a quasi-helix) $h : \mathbb{R} \to \mathbb{R}^m$ such that $|s - t|^\theta \leqslant \|h(s) - h(t)\|_{\ell_\infty^m} \leqslant (1+\delta)|s - t|^\theta$ for all $s, t \in \mathbb{R}$. The mapping

$$(x \in \mathcal{M}) \mapsto \bigoplus_{i=1}^k h \circ f_i(x) \in \bigoplus_{i=1}^k \ell_\infty^m$$

is a distortion-α embedding of the θ-snowflake $(\mathcal{M}, d_\mathcal{M}^\theta)$ into a normed space of dimension $mk \lesssim_{\alpha,\theta} n^{c/\alpha^{1/\theta}}$. The implicit dependence on α, θ that [**?**, **?**] imply here is quite good, but likely not sharp as $\alpha \to 1^+$ when $\theta \neq \frac{1}{2}$.

Since the expressions in (12) are somewhat involved, it is beneficial to restate them on a case-by-case basis as follows. For sufficiently large α, we have a bound[4] that is sharp up to universal constant factors.

$$\alpha \geqslant (\log n) \log \log n \implies \mathsf{k}_n^\alpha(\ell_\infty) \asymp \frac{\log n}{\log\left(\frac{\alpha}{\log n}\right)}. \tag{20}$$

For a range of smaller values of α, including those α that do not tend to ∞ with n, we have

$$1 \leqslant \alpha \leqslant \frac{\log n}{\log \log n} \implies n^{\frac{c}{\alpha}} \lesssim \mathsf{k}_n^\alpha(\ell_\infty) \lesssim n^{\frac{C}{\alpha}}. \tag{21}$$

(21) satisfactorily shows that the asymptotic behavior of $\mathsf{k}_n^\alpha(\ell_\infty)$ is of power-type, but it is not as sharp as (20). We suspect that determining the correct exponent of n in the power-type dependence of $\mathsf{k}_n^\alpha(\ell_\infty)$ would be challenging (there is indication [**?**, **?**], partially assuming a positive answer to a difficult conjecture of Erdős [**?**, **?**], that this exponent has infinitely many jump discontinuities as a function of α). In an intermediate range $(\log n)/\log \log n \lesssim \alpha \lesssim (\log n) \log \log n$ the bounds (12) are less satisfactory. The case $\alpha \asymp \log n$, corresponding to the distortion in Bourgain's embedding theorem, is especially intriguing, with (12) becoming

$$\frac{\log n}{\log \log \log n} \lesssim \mathsf{k}_n^{\Theta(\log n)}(\ell_\infty) \lesssim \log n. \tag{22}$$

The first inequality in (22) has not been stated in the literature, and we justify it in Section 5 below. A more natural lower bound here would be a constant multiple of $(\log n)/\log \log n$, as this corresponds to the volumetric restriction (16), and moreover by the upper bound (10) in the JL lemma we know that any n-point subset of a Hilbert space does in fact embed with distortion $\log n$ into ℓ_2^k with $k \lesssim (\log n)/\log \log n$. The triple logarithm in (22) is therefore quite intriguing/surprising, thus leading to the following open question.

Question 22. Given an integer $n \geqslant 2$, what is the asymptotic behavior of the smallest $k = k_n \in \mathbb{N}$ for which any n-point metric space $\mathcal{M}$ embeds with distortion $O(\log n)$ into some k-dimensional normed space $X_\mathcal{M}$.

There is a dearth of available upper bounds on $\mathsf{k}_n^\alpha(\cdot)$, i.e., positive results establishing that metric dimension reduction is possible. This is especially striking in the case of $\mathsf{k}_n^\alpha(\ell_1)$, due to the importance of ℓ_1 from the perspective of pure mathematics and algorithms. The upper bound on $\mathsf{k}_n^\alpha(\ell_1)$ in the large distortion regime (14) follows from combining the Euclidean embedding of [**?**] with the JL lemma. The only general dimension reduction result in ℓ_1 that lowers the dimension below the trivial bound $\mathsf{k}_n^\alpha(\ell_1) \leqslant n - 1$ is the forthcoming work [**?**], where the estimate $\mathsf{k}_n^\alpha(\ell_1) \lesssim n/\alpha$ in (13) is obtained; even this modest statement requires effort (among other things, it relies on the sparsification method of Batson, Spielman and Srivastava [**?**]).

The bound $\mathsf{k}_n^\alpha(\ell_1) \geqslant n^{c/\alpha^2}$ in (13) is a remarkable theorem of Brinkman and Charikar [**?**] which answered a question that was at the time open for many years. To avoid any possible confusion, it is important to note that [**?**] actually exhibits an n-point subset $\mathcal{C}_{\mathsf{BC}}$ of ℓ_1 for which it is shown in [**?**] that if $\mathcal{C}_{\mathsf{BC}}$ embeds with distortion α into ℓ_1^k, then necessarily $k \geqslant n^{c/\alpha^2}$. On the face of it, this seems weaker than (13), because the lower

---

[4]One can alternatively justify the upper bound in (20) (for sufficiently large n) by first using Theorem 14 (Bourgain's embedding theorem) to embed an n-point metric space $\mathcal{M}$ into ℓ_2 with distortion $A \log n$ for some universal constant $A \geqslant 1$, and then using Theorem 11 (the JL lemma) with the dependence on the distortion as stated in (10) to reduce the dimension of the image of $\mathcal{M}$ under Bourgain's embedding to $O((\log n)/\log(\alpha/(A \log n)))$ while incurring a further distortion of $\alpha/(A \log n)$, thus making the overall distortion be at most α. The right hand side of (20) is therefore in fact an upper bound on $\mathsf{k}_n^\alpha(\ell_\infty, \ell_2)$; see also Corollary 50.

bound on $\mathsf{k}_n^\alpha(\ell_1)$ in (13) requires showing that if $\mathcal{C}_{\mathsf{BC}}$ embeds into an *arbitrary* finite-dimensional linear subspace F of ℓ_1, then necessarily $\dim(F) \geqslant n^{c/\alpha^2}$. However, Talagrand proved [?] that in this setting for every $\beta > 1$ the subspace F embeds with distortion β into ℓ_1^k, where $k \lesssim_\beta \dim(F) \log \dim(F)$. From this, an application of the above stated result of [?] gives that $\dim(F) \log \dim(F) \gtrsim n^{c/\alpha^2}$, and so the lower bound in (13) follows from the formulation in [?]. Satisfactory analogues of the above theorem of Talagrand are known [?, ?, ?] (see also the survey [?] for more on this subtle issue) when ℓ_1 is replaced by ℓ_p for some $p \in (1, \infty)$, but such reductions to "canonical" linear subspaces are not available elsewhere, so the above reasoning is a rare "luxury" and in general one must treat arbitrary low-dimensional linear subspaces of the Banach space in question.

The above difficulty was overcome for S_1 in [?], where (15) was proven. The similarity of the lower bounds in (13) and (15) is not coincidental. One can view the Brinkman–Charikar example $\mathcal{C}_{\mathsf{BC}} \subseteq \ell_1$ also as a collection of diagonal matrices in S_1, and [?] treats this very same subset by strengthening the assertion of [?] that $\mathcal{C}_{\mathsf{BC}}$ does not well-embed into low-dimensional subspaces of S_1 which consist entirely of diagonal matrices, to the same assertion for low-dimensional subspaces of S_1 which are now allowed to consist of any matrices whatsoever. Using our notation for relative dimension reduction moduli, this gives the stronger assertion $\mathsf{k}_n^\alpha(\ell_1, \mathsf{S}_1) \geqslant n^{c/\alpha^2}$.

A geometric challenge of the above discussion is that, even after one isolates a candidate n-point subset $\mathcal{C}$ of ℓ_1 that is suspected not to be realizable with $O(1)$ distortion in low-dimensions (finding such a suspected intrinsically high-dimensional set is of course a major challenge in itself), one needs to devise a way to somehow argue that if one could find a configuration of n points in a low-dimensional subspace F of ℓ_1 (or S_1) whose pairwise distances are within a fixed, but potentially very large, factor $\alpha \geqslant 1$ of the corresponding pairwise distances within $\mathcal{C}$ itself, then this would force the ambient dimension $\dim(F)$ to be very large. In [?] this was achieved via a clever proof that relies on linear programming; see also [?] for a variant of this linear programming approach in the almost isometric regime $\alpha \to 1^+$. In [?] a different proof of the Brinkman–Charikar theorem was found, based on information-theoretic reasoning. Another entirely different geometric method to prove that theorem was devised in [?]; see also [?, ?] for more applications of the approach of [?].

Very recently, a further geometric approach was obtained in [?], where it was used to derive a stronger statement that, as shown in [?], cannot follow from the method of [?] (the statement is that the n-point subset $\mathcal{C}_{\mathsf{BC}} \subseteq \ell_1$ is not even an α-Lipschitz quotient of any subset of a low dimensional subspace of S_1; see [?] for the relevant definition an a complete discussion). The approach of [?] relies on an invariant that arose in the Ribe program and is called *Markov convexity*. Fix $q > 0$. Let $\{\chi_t\}_{t \in \mathbb{Z}}$ be a Markov chain on a state space Ω. Given an integer $k \geqslant 0$, denote by $\{\widetilde{\chi}_t(k)\}_{t \in \mathbb{Z}}$ the process that equals χ_t for time $t \leqslant k$, and evolves independently of χ_t (with respect to the same transition probabilities) for time $t > k$. Following [?], the Markov q-convexity constant of a metric space $(\mathcal{M}, d_\mathcal{M})$, denoted $\Pi_q(\mathcal{M})$, is the infimum over those $\Pi \in [0, \infty]$ such that for every Markov chain $\{\chi_t\}_{t \in \mathbb{Z}}$ on a state space Ω and every $f : \Omega \to \mathcal{M}$ we have

$$\left(\sum_{k=1}^{\infty} \sum_{t \in \mathbb{Z}} \frac{1}{2^{qk}} \mathbb{E}\left[d_\mathcal{M}\big(f\big(\widetilde{\chi}_t(t - 2^k)\big), f(\chi_t)\big)^q \right] \right)^{\frac{1}{q}} \leqslant \Pi \left(\sum_{t \in \mathbb{Z}} \mathbb{E}\left[d_\mathcal{M}\big(f(\chi_t), f(\chi_{t-1})\big)^q \right] \right)^{\frac{1}{q}}.$$

By [?, ?], a Banach space X satisfies $\Pi_q(X) < \infty$ if and only if it has an equivalent norm $||| \cdot ||| : X \to [0, \infty)$ whose modulus of uniform convexity has power type q, i.e., $|||x+y||| \leqslant 2 - \Omega_X(|||x-y|||^q)$ for every $x, y \in X$ with $|||x||| = |||y||| = 1$. This completes the step in the Ribe program which corresponds to the local linear property "X admits an equivalent norm whose modulus of uniform convexity has power type q," and it is a refinement of the aforementioned characterization of superreflexivity in [?] (which by deep results of [?, ?] corresponds to the cruder local linear property "there is a finite $q \geqslant 2$ for which X admits an equivalent norm whose modulus of uniform convexity has power type q"). By [?, ?], the Brinkman–Charikar subset $\mathcal{C}_{\mathsf{BC}} \subseteq \ell_1$ (as well as a variant of it due to Laakso [?] which has [?] the same non-embeddability property into low-dimensional subspaces of ℓ_1) satisfies $\Pi_q(\mathcal{C}_{\mathsf{BC}}) \gtrsim (\log n)^{1/q}$ for every $q \geqslant 2$ (recall that in our notation $|\mathcal{C}_{\mathsf{BC}}| = n$). At the same time, it is proved in [?] that $\Pi_2(F) \lesssim \sqrt{\log \dim(F)}$ for every finite dimensional subset of S_1. It remains to contrast these asymptotic behaviors (for $q = 2$) to deduce that if $\mathsf{c}_F(\mathcal{C}_{\mathsf{BC}}) \leqslant \alpha$, then necessarily $\dim(F) \geqslant n^{c/\alpha^2}$.

Prior to the forthcoming work [?], the set $\mathcal{C}_{\mathsf{BC}}$ (and variants thereof of a similar nature) was the only known example that demonstrates that there is no ℓ_1 analogue of the JL-lemma. The following theorem is from [?].

Theorem 23. *There is a universal constant $c \in (0, \infty)$ with the following property. For arbitrarily large $n \in \mathbb{N}$ there exists an n-point $O(1)$-doubling subset $\mathcal{H}_n$ of ℓ_1 satisfying $\mathsf{c}_4(\mathcal{H}_n) \lesssim 1$, such that for every $\alpha \in [1, \infty)$ and*

14

every finite-dimensional linear subspace F of S_1, if $\mathcal{H}_n$ embeds into F with distortion α, then necessarily

$$\dim(F) \geqslant \exp\left(\frac{c}{\alpha^2}\sqrt{\log n}\right). \tag{23}$$

See Section 3 for the (standard) terminology "doubling" that is used in Theorem 23. While (23) is weaker than the lower bound of Brinkman–Charikar in terms of the dependence on n, it nevertheless rules out metric dimension reduction in ℓ_1 (or S_1) in which the target dimension is, say, a power of $\log n$. The example $\mathcal{H}_n$ of Theorem 23 embeds with distortion $O(1)$ into ℓ_4, and hence in particular $\sup_{n\in\mathbb{N}}\Pi_4(\mathcal{H}_n) \lesssim \Pi_4(\ell_4) < \infty$, by [?]. This makes $\mathcal{H}_n$ be qualitatively different from all the previously known examples which exhibit the impossibility of metric dimension reduction in ℓ_1, and as such its existence has further ramifications that answer longstanding questions; see [?] for a detailed discussion. The proof of Theorem 23 is markedly different from (and more involved than) previous proofs of impossibility of dimension reduction in ℓ_1, as it relies on new geometric input (a subtle property of the 3-dimensional Heisenberg group which fails for the 5-dimensional Heisenberg group) that is obtained in [?], in combination with results from [?, ?, ?, ?]; full details appear in [?].

1.4. **Spaces admitting bi-Lipschitz and average metric dimension reduction.** Say that an infinite dimensional Banach space $(X, \|\cdot\|_X)$ *admits metric dimension reduction* if there is $\alpha = \alpha_X \in [1,\infty)$ such that

$$\lim_{n\to\infty}\frac{\log\mathsf{k}_n^\alpha(X)}{\log n} = 0.$$

In other words, the requirement here is that for some $\alpha = \alpha_X \in [1,\infty)$ and every $n \in \mathbb{N}$, any n-point subset $\mathcal{C} \subseteq X$ embeds with (bi-Lipschitz) distortion α into some linear subspace F of X with $\dim(F) = n^{o_X(1)}$.

Analogously, we say that $(X, \|\cdot\|_X)$ *admits average metric dimension reduction* if there is $\alpha = \alpha_X \in [1,\infty)$ such that for any $n \in \mathbb{N}$ there is $k_n = n^{o_X(1)}$, i.e., $\lim_{n\to\infty}(\log k_n)/\log n = 0$, such that for any n-point subset $\mathcal{C}$ of X there is a linear subspace F of X with $\dim(F) = k_n$ and a mapping $f : \mathcal{C} \to F$ which is α-Lipschitz, i.e., $\|f(x) - f(y)\|_X \leqslant \alpha\|x - y\|_X$ for all all $x, y \in \mathcal{C}$, yet

$$\frac{1}{n^2}\sum_{x\in\mathcal{C}}\sum_{y\in\mathcal{C}}\|f(x) - f(y)\|_X \geqslant \frac{1}{n^2}\sum_{x\in\mathcal{C}}\sum_{y\in\mathcal{C}}\|x - y\|_X. \tag{24}$$

Our choice here of the behavior $n^{o_X(1)}$ for the target dimension is partially motivated by the available results, based on which this type of asymptotic behavior appears to be a benchmark. We stress, however, that since the repertoire of spaces that are known to admit metric dimension reduction is currently very limited, finding any new setting in which one could prove that reducing dimension to $o_X(n)$ is possible would be a highly sought after achievement. In the same vein, finding new spaces for which one could prove a metric dimension reduction lower bound that tends to ∞ faster than $\log n$ (impossibility of a JL-style guarantee) would be very interesting.

Remark 24. In the above definition of spaces admitting average metric dimension reduction we imposed the requirement (24) following the terminology that was introduced by Rabinovich in [?], and due to the algorithmic usefulness of this notion of embedding. However, one could also consider natural variants such as $(\frac{1}{n^2}\sum_{x\in\mathcal{C}}\sum_{y\in\mathcal{C}}\|f(x) - f(y)\|_X^p)^{1/p} \geqslant (\frac{1}{n^2}\sum_{x\in\mathcal{C}}\sum_{y\in\mathcal{C}}\|x - y\|_X^q)^{1/q}$ for any $p, q \in (0,\infty]$, and much of the ensuing discussion holds mutatis mutandis in this setting as well.

The only Banach spaces that are currently known to admit metric dimension reduction are Hilbert spaces and the space $\mathcal{T}^{(2)}$ of Theorem 12 (and variants thereof). These examples allow for the possibility that if $(X, \|\cdot\|_X)$ admits metric dimension reduction, i.e., $\mathsf{k}_n^{O_X(1)}(X) = n^{o_X(1)}$, then actually $\mathsf{k}_n^{O_X(1)}(X) = O_X(\log n)$. Such a dichotomy would of course be remarkable, but there is insufficient evidence to conjecture that this is so.

The available repertoire of spaces that admit average metric dimension reduction is larger, since if $p \in [2,\infty)$, then ℓ_p and even S_p satisfy the assumption of the following theorem, by [?] and [?], respectively.

Theorem 25. *Let $(X, \|\cdot\|_X)$ be an infinite dimensional Banach space with unit ball $B_X = \{x \in X : \|x\|_X \leqslant 1\}$. Suppose that there is a Hilbert space $(H, \|\cdot\|_H)$ and a one-to-one mapping $f : B_X \to H$ such that f is Lipschitz and $f^{-1} : f(B_X) \to X$ is uniformly continuous. Then X admits average metric dimension reduction. In fact, this holds for embeddings into a subspace of logarithmic dimension, i.e., there is $\alpha = \alpha_X \in [1,\infty)$ such that for any $n \in \mathbb{N}$ and any n-point subset $\mathcal{C}$ of X there is a linear subspace F of X with $\dim(F) \lesssim \log n$ and a mapping $f : \mathcal{C} \to F$ which satisfies both (24) and $\|f(x) - f(y)\|_X \leqslant \alpha\|x - y\|_X$ for all $x, y \in \mathcal{C}$.*

Proof. This statement is implicit in [**?**]. By combining [**?**, Proposition 7.5] and [**?**, Lemma 7.6] there is a $O_X(1)$-Lipschitz mapping $f : \mathcal{C} \to \ell_2$ which satisfies $\frac{1}{n^2} \sum_{x \in \mathcal{C}} \sum_{y \in \mathcal{C}} \|f(x) - f(y)\|_2 \geqslant \frac{1}{n^2} \sum_{x \in \mathcal{C}} \sum_{y \in \mathcal{C}} \|x - y\|_X$. By the JL lemma we may assume that f actually takes values in ℓ_2^k for some $k \lesssim \log n$. Since X is infinite dimensional, Dvoretzky's theorem [**?**] ensures that ℓ_2^k is 2-isomorphic to a k-dimensional subspace F of X. □

Remark 26. By [**?**], for $p \in [2, \infty)$ the assumption of Theorem 25 holds for $X = \ell_p$. An inspection of the proofs in [**?**] reveals that the dependence of the Lipschitz constant $\alpha = \alpha_p$ on p that Theorem 25 provides for $X = \ell_p$ grows to ∞ exponentially with p. As argued in [**?**, Section 5.1] (using metric cotype), this exponential behavior is unavoidable using the above proof. However, in this special case a more sophisticated argument of [**?**] yields $\alpha_p \lesssim p^{5/2}$; see equation (7.40) in [**?**]. Motivated by [**?**, Corollary 1.6], we conjecture that this could be improved to $\alpha_p \lesssim p$, and there is some indication (see [**?**, Lemma 1.11]) that this would be sharp.

Prior to [**?**], it was not known if there exists a Banach space which fails to admit average metric dimension reduction. Now we know (Theorem 19) that ℓ_∞ fails to admit average metric dimension reduction, and therefore also any universal Banach space fails to admit average metric dimension reduction. A fortiori, the same is true also for (non-average) metric dimension reduction, but this statement follows from the older work [**?**]. Failure of average metric dimension reduction is not known for *any* non-universal (finite cotype) Banach space, and it would be very interesting to provide such an example. By [**?**, **?**] we know that ℓ_1 and S_1 fail to admit metric dimension reduction, but this is not known for average distortion, thus leading to the following question.

Question 27. Does ℓ_1 admit average metric dimension reduction? Does ℓ_p have this property for any $p \in [1, 2)$?

All of the available examples of n-point subsets of ℓ_1 for which the ℓ_1 analogue of the JL lemma fails (namely if $k = O(\log n)$, then they do not embed with $O(1)$ distortion into ℓ_1^k) actually embed into the real line $\mathbb{R}$ with $O(1)$ average distortion; this follows from [**?**]. Specifically, the examples in [**?**, **?**] are the shortest-path metric on planar graphs, and the example in Theorem 23 is $O(1)$-doubling, and both of these classes of metric spaces are covered by [**?**]; see also [**?**, Section 7] for generalizations. Thus, the various known proofs which demonstrate that the available examples cannot be embedded into a low dimensional subspace of ℓ_1 argue that any such low-dimensional embedding must highly distort some distance, but this is not so for a typical distance. A negative answer to Question 27 would therefore require a substantially new type of construction which exhibits a much more "diffuse" intrinsic high-dimensionality despite it being a subset of ℓ_1. In the reverse direction, a positive answer to Question 27, beyond its intrinsic geometric/structural interest, could have algorithmic applications.

1.4.1. *Lack of stability under projective tensor products.* Prior to the recent work [**?**], it was unknown whether the property of admitting metric dimension reduction is preserved under projective tensor products.

Corollary 28. *There exist Banach spaces X, Y that admit metric dimension reduction yet $X \widehat{\otimes} Y$ does not.*

Since S_1 is isometric to $\ell_2 \widehat{\otimes} \ell_2$ and [**?**] establishes that S_1 fails to admit metric dimension reduction, together with the JL lemma this implies Corollary 28 (we can thus even have $X = Y$ and $\mathsf{k}_n^\alpha(n) \lesssim_\alpha \log n$ for all $\alpha > 1$).

Since we do not know whether S_1 admits average metric dimension reduction (the above comments pertaining to Question 27 are valid also for S_1), the analogue of Corollary 28 for average metric dimension reduction was previously unknown. Here we note the following statement, whose proof is a somewhat curious argument.

Theorem 29. *There exist Banach spaces X, Y that admit average metric dimension reduction yet $X \widehat{\otimes} Y$ does not. Moreover, for every $p \in (2, \infty)$ we can take here $X = \ell_p$.*

Proof. By [**?**] (which relies on major input from the theory of locally decodable codes [**?**] and an important inequality of Pisier [**?**]), the 3-fold product $\ell_3 \widehat{\otimes} \ell_3 \widehat{\otimes} \ell_3$ is universal. So, by the recent work [**?**] (Theorem 19), $\ell_3 \widehat{\otimes} \ell_3 \widehat{\otimes} \ell_3$ does not admit average metric dimension reduction. At the same time, by Theorem 25 we know that ℓ_3 admits average metric dimension reduction. So, if $\ell_3 \widehat{\otimes} \ell_3$ fails to admit average metric dimension reduction, then we can take $X = Y = \ell_3$ in Theorem 29. Otherwise, $\ell_3 \widehat{\otimes} \ell_3$ does admit average metric dimension reduction, in which case we can take $X = \ell_3$ and $Y = \ell_3 \widehat{\otimes} \ell_3$. Thus, in either of the above two cases, the first assertion of Theorem 29 holds true. The second assertion of Theorem 29 follows by repeating this argument using the fact [**?**] that $\ell_p \widehat{\otimes} \ell_p \widehat{\otimes} \ell_q$ is universal if $2/p + 1/q \leqslant 1$, or equivalently $q \geqslant p/(p-2)$. If we choose, say, $q = \max\{2, p/(p-2)\}$, then by Theorem 25 we know that both ℓ_p and ℓ_q admit average metric dimension reduction, while $\ell_p \widehat{\otimes} \ell_p \widehat{\otimes} \ell_q$ does not. So, the second assertion of Theorem 29 holds for either $Y = \ell_p$ or $Y = \ell_p \widehat{\otimes} \ell_q$. □

The proof of Theorem 29 establishes that at least one of the pairs $(X = \ell_3, Y = \ell_3)$ or $(X = \ell_3, Y = \ell_3 \widehat{\otimes} \ell_3)$ satisfies its conclusion, but it gives no indication which of these two options occurs. This naturally leads to

Question 30. Does $\ell_3 \widehat{\otimes} \ell_3$ admit average metric dimension reduction?

A positive answer to Question 30 would yield a new space that admits average metric dimension reduction. In order to claim that $\ell_3 \widehat{\otimes} \ell_3$ is indeed new in this context, one must show that it does not satisfy the assumption of Theorem 25. This is so because S_1 (hence also ℓ_1) is finitely representable in $\ell_3 \widehat{\otimes} \ell_3$; see e.g. [**?**, page 61]. The fact that no Banach space in which ℓ_1 is finitely representably satisfies the assumption of Theorem 25 follows by combining [**?**, Lemma 1.12], [**?**, Proposition 7.5], and [**?**, Lemma 7.6]. This also shows that a positive answer to Question 30 would imply that any n-point subset of ℓ_1 (or S_1) embeds with $O(1)$ average distortion into some normed space (a subspace of $\ell_3 \widehat{\otimes} \ell_3$) of dimension $n^{o(1)}$, which is a statement in the spirit of Question 27. If the answer to Question 30 were negative, then $\ell_3 \widehat{\otimes} \ell_3$ would be the first example of a non-universal space which fails to admit average metric dimension reduction, because Pisier proved [**?**, **?**] that $\ell_3 \widehat{\otimes} \ell_3$ is not universal.

Another question that arises naturally from Theorem 29 is whether its conclusion holds true also for $p = 2$.

Question 31. Is there a Banach space Y that admits average metric dimension reduction yet $\ell_2 \widehat{\otimes} Y$ does not?

1.4.2. *Wasserstein spaces.* Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space and $p \in [1, \infty)$. The Wasserstein space $\mathsf{P}_p(\mathcal{M})$ is not a Banach space, but there is a natural version of the metric dimension reduction question in this context.

Question 32. Fix $\alpha > 1$, $n \in \mathbb{N}$ and $\mu_1, \ldots, \mu_n \in \mathsf{P}_p(\mathcal{M})$. What is the asymptotic behavior of the smallest $k \in \mathbb{N}$ for which there is $\mathcal{S} \subseteq \mathcal{M}$ with $|\mathcal{S}| \leqslant k$ such that $(\{\mu_1, \ldots, \mu_n\}, \mathsf{W}_p)$ embeds with distortion α into $\mathsf{P}_p(\mathcal{S})$?

Spaces of measures with the Wasserstein metric W_p are of major importance in pure and applied mathematics, as well as in computer science (mainly for $p = 1$, where they are used in graphics and vision, but also for other values of p; see e.g. the discussion in [**?**]). However, their bi-Lipschitz structure is poorly understood, especially so in the above context of metric dimension reduction. If k were small in Question 32, then this would give a way to "compress" collections of measures using measures with small support while approximately preserving Wasserstein distances. In the context of, say, image retrieval (mainly $\mathcal{M} = \{1, \ldots, n\}^2 \subseteq \mathbb{R}^2$ and $p = 1$), this could be viewed as obtaining representations of images using a small number of "pixels."

Charikar [**?**] and Indyk–Thaper [**?**] proved that if $\mathcal{M}$ is a finite metric space, then $\mathsf{P}_1(\mathcal{M})$ embeds into ℓ_1 with distortion $O(\log |\mathcal{M}|)$. Hence, if the answer to Question 32 were $k = n^{o(1)}$ for some $\alpha = O(1)$, then it would follow that any n-point subset of $\mathsf{P}_1(\mathcal{M})$ embeds into ℓ_1 with distortion $o(\log n)$, i.e., better distortion than the general bound that is provided by Bourgain's embedding theorem (actually the ℓ_1-variant of that theorem, which is also known to be sharp in general [**?**]). This shows that one cannot hope to answer Question 32 with $k = n^{o(1)}$ and $\alpha = O(1)$ without imposing geometric restrictions on the underlying metric space $\mathcal{M}$, since if $(\mathcal{M} = \{x_1, \ldots, x_n\}, d_{\mathcal{M}})$ is a metric space for which $\mathsf{c}_1(\mathcal{M}) \asymp \log n$, then we can take $\mu_1, \ldots, \mu_n$ to be the point masses $\delta_{x_1}, \ldots, \delta_{x_n}$, so that $(\{\mu_1, \ldots, \mu_n\}, \mathsf{W}_1)$ is isometric to $(\mathcal{M}, d_{\mathcal{M}})$. The pertinent issue is therefore to study Question 32 when the $\mathcal{M}$ is "nice." For example, sufficiently good bounds here for $\mathcal{M} = \mathbb{R}^2$ would be relevant to Question 7, but at this juncture such a potential approach to Question 7 is quite speculative.

The above "problematic" example relied inherently on the fact that the underlying metric space $\mathcal{M}$ is itself far from being embeddable in ℓ_1, but the difficulty persists even when $\mathcal{M} = \ell_1$. Indeed, we recalled in Question 7 that Bourgain proved [**?**] that $\mathsf{P}_1(\ell_1)$ is universal, and hence the spaces of either Theorem 17 or Theorem 19 embed into $\mathsf{P}_1(\ell_1)$ with $O(1)$ distortion. So, for arbitrarily large $n \in \mathbb{N}$ we can find probability measures $\mu_1, \ldots, \mu_n$ on ℓ_1 (actually on a sufficiently high dimensional Hamming cube $\{0, 1\}^N$) such that $(\{\mu_1, \ldots, \mu_n\}, \mathsf{W}_1)$ does not admit a good embedding into any normed space of dimension $n^{o(1)}$. This rules out an answer of $k = n^{o(1)}$ to Question 32 (even for average distortion) for $(\mu_1, \ldots, \mu_n, \mathsf{W}_1)$, because in the setting of Question 32, while $\mathsf{P}_1(\mathcal{S})$ is not a normed space, it embeds isometrically into a normed space of dimension $|\mathcal{S}| - 1$ (the dual of the mean-zero Lipschitz functions on $(\mathcal{S}, d_{\mathcal{M}})$; see e.g. [**?**, **?**] for an explanation of this standard fact). In the case of average distortion, one could see this using a different approach of Khot and the author [**?**] which constructs a collection of $n = e^{O(d)}$ probability measures on the Hamming cube $\{0, 1\}^d$ that satisfy the conclusion of Theorem 19, as explained in [**?**, Remark 5]. This shows that even though these n probability measures reside on a Hamming cube of dimension $O(\log n)$, one cannot realize their Wasserstein-1 geometry with $O(1)$ distortion (even on average) in any normed space of dimension $n^{o(1)}$, let alone in $\mathsf{P}_1(\mathcal{S})$ with $|\mathcal{S}| = n^{o(1)}$.

It is therefore natural to investigate Question 32 when $\mathcal{M}$ is low-dimensional. When $p = 1$, this remains an (important) uncharted terrain. When $p > 1$ and $\mathcal{M} = \mathbb{R}^3$, partial information on Question 32 follows from [**?**].

To see this, focus for concreteness on the case $p = 2$. Fix $\alpha \geqslant 1$ and $n \in \mathbb{N}$. Suppose that $(\mathcal{N}, d_{\mathcal{N}})$ is an n-point metric space for which the conclusion of Theorem 17 holds true with $\omega(t) = \sqrt{t}$ and $\Omega(t) = 2\alpha\sqrt{t}$. By [?], the metric space $(\mathcal{N}, \sqrt{d_{\mathcal{N}}})$ embeds with distortion 2 into $\mathsf{P}_2(\mathbb{R}^3)$, where $\mathbb{R}^3$ is equipped with the standard Euclidean metric. Hence, if the image under this embedding of $\mathcal{N}$ in $\mathsf{P}_2(\mathbb{R}^3)$ embedded into some k-dimensional normed space with distortion α, then by Theorem 17 necessarily $k \geqslant n^{c/\alpha^2}$ for some universal constant c. This does not address Question 32 as stated, because to the best of our knowledge it is not known whether $\mathsf{P}_2(\mathcal{S})$ embeds with $O(1)$ distortion into some "low-dimensional" normed space for every "small" $\mathcal{S} \subseteq \mathbb{R}^3$ (the relation between "small" and "low-dimensional" remains to be studied). In the case of average distortion, repeat this argument with $\mathcal{N}$ now being the metric space of Theorem 19. By Remark 48 below, if the image in $\mathsf{P}_2(\mathbb{R}^3)$ of $(\mathcal{N}, \sqrt{d_{\mathcal{N}}})$ embedded with average distortion α into some k-dimensional normed space, then necessarily $k \geqslant \exp(\frac{c}{\alpha}\sqrt{\log n})$.

2. Finite subsets of Hilbert space

The article [?] of Johnson and Lindenstrauss is devoted to proving a theorem on the extension of Lipschitz functions from finite subsets of metric spaces.[5] Over the ensuing decades, the classic [?] attained widespread prominence outside the rich literature on the *Lipschitz extension problem*, due to two components of [?] that had major conceptual significance and influence, but are technically simpler than the proof of its main theorem.

The first of these components is the JL lemma, which we already stated in the Introduction. Despite its wide acclaim and applicability, this result is commonly called a "lemma" rather than a "theorem" because within the context of [?] it was just that, i.e., a relatively simple step toward the proof of the main theorem of [?].

The second of these components is a section of [?] that is devoted to formulating open problems in the context of the Ribe program; we already described a couple of the questions that were raised there, but it contains more questions that proved to be remarkably insightful and had major impact on subsequent research (see e.g. [?, ?]). Despite its importance, the impact of [?] on the Ribe program will not be pursued further in the present article, but we will next proceed to study the JL lemma in detail (including some new observations).

Recalling Theorem 11, the JL lemma [?] asserts that for every integer $n \geqslant 2$ and (distortion/error tolerance) $\alpha \in (1, \infty)$, if $x_1, \ldots, x_n$ are distinct vectors in a Hilbert space $(H, \|\cdot\|_H)$, then there exists (a target dimension) $k \in \{1, \ldots, n\}$ and a new n-tuple of k-dimensional vectors $y_1, \ldots, y_n \in \mathbb{R}^k$ such that

$$k \lesssim_\alpha \log n, \tag{25}$$

and the assignment $x_i \mapsto y_i$, viewed as a mapping into ℓ_2^k, has distortion at most α, i.e.,

$$\forall i, j \in \{1, \ldots, n\}, \qquad \|x_i - x_j\|_H \leqslant \|y_i - y_j\|_{\ell_2^k} \leqslant \alpha\|x_i - x_j\|_H. \tag{26}$$

It is instructive to take note of the "compression" that this statement entails. By tracking the numerical value of the target dimension k that the proof in Section 2.1 below yields (see Remark 38), one concludes that given an arbitrary collection of, say, a billion vectors of length a billion (i.e., $1\,000\,000\,000$ elements of $\mathbb{R}^{1\,000\,000\,000}$), one can find a billion vectors of length 329 (i.e., elements of $\mathbb{R}^{329}$), all of whose pairwise distances are within a factor 2 of the corresponding pairwise distances among the initial configuration of billion-dimensional vectors. Furthermore, if one wishes to maintain the pairwise distances of those billion vectors within a somewhat larger constant factor, say a factor of 10 or 450, then one could do so in dimension 37 or 9, respectively.

The logarithmic dependence on n in (25) is optimal, up to the value of the implicit (α-dependent) constant factor. This is so even when one considers the special case when $x_1, \ldots, x_n \in H$ are the vertices of the standard $(n-1)$-simplex, i.e., $\|x_i - x_j\|_H = 1$ for all distinct $i, j \in \{1, \ldots, n\}$, and even when one allows the Euclidean norm in (26) to be replaced by any norm $\|\cdot\| : \mathbb{R}^k \to [0, \infty)$, namely if instead of (26) we have $1 \leqslant \|y_i - y_j\| \leqslant \alpha$ for all distinct $i, j \in \{1, \ldots, n\}$. Indeed, denote the unit ball of $\|\cdot\|$ by $B = \{z \in \mathbb{R}^k : \|z\| \leqslant 1\}$ and let $\mathbf{vol}_k(\cdot)$ be the Lebesgue measure on $\mathbb{R}^k$. If $i, j \in \{1, \ldots, n\}$ are distinct, then by the triangle inequality the assumed lower bound $\|y_i - y_j\| \geqslant 1$ implies that the interiors of $y_i + \frac{1}{2}B$ and $y_j + \frac{1}{2}B$ are disjoint. Hence, if we denote $A = \bigcup_{i=1}^n (y_i + \frac{1}{2}B)$, then $\mathbf{vol}_k(A) = \sum_{i=1}^n \mathbf{vol}_k(y_i + \frac{1}{2}B) = \frac{n}{2^k}\mathbf{vol}_k(B)$. At the same time, for every $u, v \in A$ there are $i, j \in \{1, \ldots, n\}$ for which $u \in y_i + \frac{1}{2}B$ and $v \in y_j + \frac{1}{2}B$, so by another application of the triangle inequality we have $\|u - v\| \leqslant \|y_i - y_j\| + 1 \leqslant \alpha + 1$. This implies that $A - A \subseteq (\alpha + 1)B$. Hence,

$$(\alpha + 1)\sqrt[k]{\mathbf{vol}_k(B)} = \sqrt[k]{\mathbf{vol}_k((\alpha+1)B)} \geqslant \sqrt[k]{\mathbf{vol}_k(A - A)} \geqslant 2\sqrt[k]{\mathbf{vol}_k(A)} = \sqrt[k]{n\mathbf{vol}_k(B)},$$

---

[5]Stating this theorem here would be an unnecessary digression, but we highly recommend examining the accessible geometric result of [?]; see [?] for a review of the current state of the art on Lipschitz extension from finite subsets.

where the penultimate step uses the Brunn–Minkowski inequality [?]. This simplifies to give

$$k \geqslant \frac{\log n}{\log(\alpha + 1)}. \tag{27}$$

By [?, ?], the vertices of $(n-1)$-simplex embed isometrically into any infinite dimensional Banach space, so we have thus justified the bound (16), and hence in particular the first lower bound on $\mathsf{k}_n^\alpha(\ell_2)$ in (10). As we already explained, the second lower bound (for the almost-isometric regime) on $\mathsf{k}_n^\alpha(\ell_2)$ in (10) is due to the very recent work [?]. The upper bound on $\mathsf{k}_n^\alpha(\ell_2)$ in (10), namely that in (25) we can take

$$k \lesssim \frac{\log n}{\log\left(1 + (\alpha - 1)^2\right)} \asymp \max\left\{\frac{\log n}{(\alpha - 1)^2}, \frac{\log n}{\log \alpha}\right\}, \tag{28}$$

follows from the original proof of the JL lemma in [?]. A justification of (28) appears in Section 2.1 below.

Question 33 (dimension reduction for the vertices of the simplex). Fix $\delta \in (0, \frac{1}{2})$. What is the order of magnitude (up to universal constant factors) of the smallest $\mathfrak{S}(\delta) \in (0, \infty)$ such that for every $n \in \mathbb{N}$ there is $k \in \mathbb{N}$ with $k \leqslant \mathfrak{S}(\delta) \log n$ and $y_1, \ldots, y_n \in \mathbb{R}^k$ that satisfy $1 \leqslant \|y_i - y_j\|_2 \leqslant 1 + \delta$ for all distinct $i, j \in \{1, \ldots, n\}$? By (28) we have $\mathfrak{S}(\delta) \lesssim 1/\delta^2$. The best-known lower bound here is $\mathfrak{S}(\delta) \gtrsim 1/(\delta^2 \log(1/\delta))$, due to Alon [?].

Remark 34. The upper bound (28) treats the target dimension in the JL lemma for an *arbitrary* subset of a Hilbert space. The lower bound (27) was derived in the special case of the vertices of the regular simplex, but it is also more general as it is valid for embeddings of these vertices into an *arbitrary* k-dimensional norm. In this (both special, and more general) setting, the bound (27) is quite sharp for large α. Indeed, by [?] (see also [?, Corollary 2.4]), for each $n \in \mathbb{N}$ and $\alpha > \sqrt{2}$, if we write $k = \lceil (\log(4n))/\log(\alpha^2/(2\sqrt{\alpha^2 - 1})) \rceil$, then for *every* norm $\|\cdot\|$ on $\mathbb{R}^k$ there exist $y_1, \ldots, y_n \in \mathbb{R}^k$ satisfying $1 \leqslant \|y_i - y_j\| \leqslant \alpha$ for distinct $i, j \in \{1, \ldots, n\}$. See [?, Theorem 4.3] for an earlier result in this direction. See also [?] and the references therein (as well as [?, Problem 2.5]) for partial results towards understanding the analogous issue (which is a longstanding open question) in the small distortion regime $\alpha \in (1, \sqrt{2}]$.

2.1. **Optimality of re-scaled random projections.** To set the stage for the proof of the JL lemma, note that by translation-invariance we may assume without loss of generality that one of the vectors $\{x_i\}_{i=1}^n$ vanishes, and then by replacing the Hilbert space H with the span of $\{x_i\}_{i=1}^n$, we may further assume that $H = \mathbb{R}^{n-1}$.

Let $\mathsf{Proj}_{\mathbb{R}^k} \in \mathsf{M}_{k \times (n-1)}(\mathbb{R})$ be the k by $n-1$ matrix of the orthogonal projection from $\mathbb{R}^{n-1}$ onto $\mathbb{R}^k$, i.e., $\mathsf{Proj}_{\mathbb{R}^k} z = (z_1, \ldots, z_k) \in \mathbb{R}^k$ is the first k coordinates of $z = (z_1, \ldots, z_{n-1}) \in \mathbb{R}^{n-1}$. One could attempt to simply truncate the vectors vectors $x_1, \ldots, x_n$ so as to obtain k-dimensional vectors, namely to consider the vectors $\{y_i = \mathsf{Proj}_{\mathbb{R}^k} x_i\}_{i=1}^n$ in (26). This naive (and heavy-handed) way of forcing low-dimensionality can obviously fail miserably, e.g. we could have $\mathsf{Proj}_k x_i = 0$ for all $i \in \{1, \ldots, n\}$. Such a simplistic idea performs poorly because it makes two arbitrary and unnatural choices, namely it does not take advantage of rotation-invariance and scale-invariance. To remedy this, let $\mathsf{O}_{n-1} \subseteq \mathsf{M}_{n-1}(\mathbb{R})$ denote the group of $n-1$ by $n-1$ orthogonal matrices, and fix (a scaling factor) $\sigma \in (0, \infty)$. Let $\mathsf{O} \in \mathsf{O}_{n-1}$ be a random orthogonal matrix distributed according to the Haar probability measure on O_{n-1}. In [?] it was shown that if k is sufficiently large (yet satisfying (25)), then for an appropriate $\sigma > 0$ with positive probability (26) holds for the following random vectors.

$$\{y_i = \sigma \mathsf{Proj}_{\mathbb{R}^k} \mathsf{O} x_i\}_{i=1}^n \subseteq \mathbb{R}^k. \tag{29}$$

We will do more than merely explain why the randomly projected vectors in (29) satisfy the desired conclusion (26) of the JL lemma with positive probability. We shall next demonstrate that such a procedure is the *best possible* (in a certain sense that will be made precise) among all the possible choices of random assignments of $x_1, \ldots, x_n$ to $y_1, \ldots, y_n$ via multiplication by a random matrix in $M_{k \times (n-1)}(\mathbb{R})$, provided that we optimize so as to use the best scaling factor $\sigma \in (0, \infty)$ in (29).

Let μ be any Borel probability measure on $\mathsf{M}_{k \times (n-1)}(\mathbb{R})$, i.e., μ represents an arbitrary (reasonably measurable) distribution over $k \times (n-1)$ random matrices $\mathsf{A} \in \mathsf{M}_{k \times (n-1)}(\mathbb{R})$. For $\alpha \in (1, \infty)$ define

$$\mathfrak{p}_\mu^\alpha \stackrel{\text{def}}{=} \inf_{z \in \mathbf{S}^{n-2}} \mu\left[\left\{\mathsf{A} \in \mathsf{M}_{k \times (n-1)}(\mathbb{R}) : \ 1 \leqslant \|\mathsf{A}z\|_{\ell_2^k} \leqslant \alpha\right\}\right], \tag{30}$$

where $\mathbf{S}^{n-2} = \{z \in \mathbb{R}^{n-1} : \ \|z\|_{\ell_2^{n-1}} = 1\}$ denotes the unit Euclidean sphere in $\mathbb{R}^{n-1}$. Then

$$
\mu\bigg[\bigcap_{i,j\in\{1,\ldots,n\}}\Big\{\mathsf{A}\in\mathsf{M}_{k\times(n-1)}(\mathbb{R}) : \|x_i - x_j\|_{\ell_2^{n-1}} \leqslant \|\mathsf{A}x_i - \mathsf{A}x_j\|_{\ell_2^k} \leqslant \alpha\|x_i - x_j\|_{\ell_2^{n-1}}\Big\}\bigg]
$$

$$
= 1 - \mu\bigg[\bigcup_{i=1}^{n}\bigcup_{j=i+1}^{n}\Big(\mathsf{M}_{k\times(n-1)}(\mathbb{R})\smallsetminus\Big\{\mathsf{A}\in\mathsf{M}_{k\times(n-1)}(\mathbb{R}) : 1\leqslant \Big\|\mathsf{A}\frac{x_i - x_j}{\|x_i - x_j\|_{\ell_2^{n-1}}}\Big\|_{\ell_2^k}\leqslant\alpha\Big\}\Big)\bigg]
$$

$$
\geqslant 1 - \sum_{i=1}^{n}\sum_{j=i+1}^{n}\bigg(1 - \mu\Big[\Big\{\mathsf{A}\in\mathsf{M}_{k\times(n-1)}(\mathbb{R}) : 1\leqslant \Big\|\mathsf{A}\frac{x_i - x_j}{\|x_i - x_j\|_{\ell_2^{n-1}}}\Big\|_{\ell_2^k}\leqslant\alpha\Big\}\Big]\bigg)
$$

$$
\geqslant 1 - \binom{n}{2}\big(1 - \mathfrak{p}_\mu^\alpha\big). \tag{31}
$$

Hence, the random vectors $\{y_i = \mathsf{A}x_i\}_{i=1}^n$ will satisfy (26) with positive probability if $\mathfrak{p}_\mu^\alpha > 1 - \frac{2}{n(n-1)}$.

In order to succeed to embed the largest possible number of vectors into $\mathbb{R}^k$ via the above randomized procedure while using the estimate (31), it is in our best interest to work with a probability measure μ on $\mathsf{M}_{k\times(n-1)}(\mathbb{R})$ for which $\mathfrak{p}_\mu^\alpha$ is as large as possible. To this end, define

$$
\mathfrak{p}_{n,k}^\alpha \stackrel{\text{def}}{=} \sup\Big\{\mathfrak{p}_\mu^\alpha : \ \mu \text{ is a Borel probability measure on } \mathsf{M}_{k\times(n-1)}(\mathbb{R})\Big\}. \tag{32}
$$

Then, the conclusion (26) of the JL lemma will be valid provided $k \in \{1,\ldots,n\}$ satisfies

$$
\mathfrak{p}_{n,k}^\alpha > 1 - \frac{2}{n(n-1)}. \tag{33}
$$

The following proposition asserts that the supremum in the definition (32) of $\mathfrak{p}_{n,k}^\alpha$ is attained at a distribution over random matrices that has the aforementioned structure (29).

Proposition 35 (multiples of random orthogonal projections are JL-optimal)**.** *Fix $\alpha \in (1,\infty)$, an integer $n \geqslant 4$ and $k \in \{1,\ldots,n-3\}$. Let $\mu = \mu_{n,k}^\alpha$ be the probability distribution on $\mathsf{M}_{k\times(n-1)}(\mathbb{R})$ of the random matrix*

$$
\sqrt{\frac{\alpha^{\frac{2n-6}{n-k-3}} - 1}{\alpha^{\frac{2k}{n-k-3}} - 1}} \cdot \mathsf{Proj}_{\mathbb{R}^k}\mathsf{O}, \tag{34}
$$

that is obtained by choosing $\mathsf{O} \in \mathsf{O}_{n-1}$ according to the normalized Haar measure on O_{n-1}. Then $\mathfrak{p}_\mu^\alpha = \mathfrak{p}_{n,k}^\alpha$.

Obviously (34) is *not* a multiple of a uniformly random rank k orthogonal projection $\mathsf{Proj} : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ (chosen according to the normalized Haar measure on the appropriate Grassmannian). To obtain such a distribution, one should multiply the matrix in (34) on the left by O^*. That additional rotation does not influence the Euclidean length of the image, and hence it does not affect the quantity (30). For this reason and for simplicity of notation, we prefer to work with (34) rather than random projections as was done in [**?**].

Proof of Proposition 35. Given $\mathsf{A} \in \mathsf{M}_{k\times(n-1)}(\mathbb{R})$, denote its singular values by $\mathsf{s}_1(\mathsf{A}) \geqslant \ldots \geqslant \mathsf{s}_k(\mathsf{A})$, i.e., they are the eigenvalues (with multiplicity) of the symmetric matrix $\sqrt{\mathsf{A}\mathsf{A}^*} \in \mathsf{M}_k(\mathbb{R})$. Then,

$$
\mathfrak{H}^{\mathsf{O}_{n-1}}\Big[\{\mathsf{O}\in\mathsf{O}_{n-1} : \ 1\leqslant\|\mathsf{A}\mathsf{O}z\|_{\ell_2^k}\leqslant\alpha\}\Big] = \int_{\mathbf{S}^{k-1}}\psi_{n,k}^\alpha\bigg(\Big(\sum_{i=1}^{k}\mathsf{s}_i(\mathsf{A})^2\omega_i^2\Big)^{\frac{1}{2}}\bigg)\,\mathrm{d}\mathfrak{H}^{\mathbf{S}^{k-1}}(\omega), \tag{35}
$$

where $\mathfrak{H}^{\mathsf{O}_{n-1}}$ and $\mathfrak{H}^{\mathbf{S}^{k-1}}$ are the Haar probability measures on the orthogonal group O_{n-1} and the unit Euclidean sphere $\mathbf{S}^{k-1}$, respectively, and the function $\psi_{n,k}^\alpha : [0,\infty) \to \mathbb{R}$ is defined by

$$
\forall\sigma\in[0,\infty), \qquad \psi_{n,k}^\alpha(\sigma) \stackrel{\text{def}}{=} \frac{2\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\int_{\max\{1,\frac{\sigma}{\alpha}\}}^{\max\{1,\sigma\}}\frac{(s^2-1)^{\frac{n-k-3}{2}}}{s^{n-2}}\,\mathrm{d}s. \tag{36}
$$

To verify the identity (35), consider the singular value decomposition

$$
\mathsf{A} = \mathsf{U} \begin{pmatrix} \mathsf{s}_1(\mathsf{A}) & 0 & \dots & \dots & 0 \\ 0 & \mathsf{s}_2(\mathsf{A}) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \mathsf{s}_k(\mathsf{A}) \end{pmatrix} \mathsf{Proj}_{\mathbb{R}^k} \mathsf{V}, \tag{37}
$$

where $\mathsf{U} \in \mathsf{O}_k$ and $\mathsf{V} \in \mathsf{O}_{n-1}$. If $\mathsf{O} \in \mathsf{O}_{n-1}$ is distributed according to $\mathfrak{H}^{\mathsf{O}_{n-1}}$, then by the left-invariance of $\mathfrak{H}^{\mathsf{O}_{n-1}}$ we know that VO is distributed according to $\mathfrak{H}^{\mathsf{O}_{n-1}}$. By rotation-invariance and uniqueness of Haar measure on $\mathbf{S}^{n-2}$ (e.g. [?]), it follows that for every $z \in \mathbf{S}^{n-1}$ the random vector $\mathsf{VO}z$ is distributed according to the normalized Haar measure on $\mathbf{S}^{n-2}$. So, $\mathsf{Proj}_{\mathbb{R}^k} \mathsf{VO}z$ is distributed on the Euclidean unit ball of $\mathbb{R}^k$, with density

$$
(u \in \mathbb{R}^k) \mapsto \frac{\Gamma\left(\frac{n-1}{2}\right)}{\pi^{\frac{k}{2}} \Gamma\left(\frac{n-1-k}{2}\right)} \left(1 - \|u\|_{\ell_2^k}^2\right)^{\frac{n-k-3}{2}} \mathbf{1}_{\left\{\|u\|_{\ell_2^k} \leqslant 1\right\}}. \tag{38}
$$

See [?] for a proof of this distributional identity (or [?, Corollary 4] for a more general derivation); in codimension 2, namely $k = n - 3$, this is a higher-dimensional analogue of Archimedes' theorem that the projection to $\mathbb{R}$ of the uniform surface area measure on the unit Euclidean sphere in $\mathbb{R}^3$ is the Lebesgue measure on $[-1, 1]$. Recalling (37), it follows from this discussion that the Euclidean norm of $\mathsf{AO}z$ has the same distribution as $\left(\sum_{i=1}^k \mathsf{s}_i(\mathsf{A})^2 u_i^2\right)^{1/2}$, where $u = (u_1, \dots, u_k) \in \mathbb{R}^k$ is distributed according to the density (38). The identity (35) now follows by integration in polar coordinates $(\omega, r) \in \mathbf{S}^{k-1} \times [0, \infty)$, followed by the change of variable $s = 1/r$.

Next, $\psi_{n,k}^\alpha$ vanishes on $[0, 1]$, increases on $[1, \alpha]$, and is smooth on $[\alpha, \infty)$. The integrand in (36) is at most s^{-k-1}, so $\lim_{\sigma \to \infty} \psi_{n,k}^\alpha(\sigma) = 0$. By directly differentiating (36) and simplifying the resulting expression, one sees that if $\sigma \in [\alpha, \infty)$, then $(\psi_{n,k}^\alpha)'(\sigma) = 0$ if and only if $\sigma = \sigma_{\max}(n, k, \alpha)$, where

$$
\sigma_{\max}(n, k, \alpha) \stackrel{\text{def}}{=} \sqrt{\frac{\alpha^{\frac{2n-6}{n-k-3}} - 1}{\alpha^{\frac{2k}{n-k-3}} - 1}}. \tag{39}
$$

Therefore, the global maximum of $\psi_{n,k}^\alpha$ is attained at $\sigma_{\max}(n, k, \alpha)$, and by (35) we have

$$
\forall \mathsf{A} \in \mathsf{M}_{k \times (n-1)}(\mathbb{R}), \quad \mathfrak{H}^{\mathsf{O}_{n-1}}\left[\left\{\mathsf{O} \in \mathsf{O}_{n-1} : 1 \leqslant \|\mathsf{AO}z\|_{\ell_2^k} \leqslant \alpha\right\}\right] \leqslant \psi_{n,k}^\alpha\big(\sigma_{\max}(n, k, \alpha)\big) = \mathfrak{p}_\mu^\alpha. \tag{40}
$$

The final step of (40) is another application (35), this time in the special case $\mathsf{A} = \sigma_{\max}(n, k, \alpha) \mathsf{Proj}_{\mathbb{R}^k}$, while recalling (30) and (39), and that μ is the distribution of the random matrix appearing in (34).

To conclude the proof of Proposition (35), take any Borel probability measure ν on $\mathsf{M}_{k \times (n-1)}(\mathbb{R})$ and integrate (40) while using Fubini's theorem to obtain the estimate

$$
\mathfrak{p}_\mu^\alpha \geqslant \int_{\mathsf{M}_{k \times (n-1)}(\mathbb{R})} \mathfrak{H}^{\mathsf{O}_{n-1}}\left[\left\{\mathsf{O} \in \mathsf{O}_{n-1} : 1 \leqslant \|\mathsf{AO}z_0\|_{\ell_2^k} \leqslant \alpha\right\}\right] \mathrm{d}\nu(\mathsf{A})
$$

$$
= \int_{\mathsf{O}_{n-1}} \nu\left[\left\{\mathsf{A} \in \mathsf{M}_{k \times (n-1)}(\mathbb{R}) : 1 \leqslant \|\mathsf{AO}z_0\|_{\ell_2^k} \leqslant \alpha\right\}\right] \mathrm{d}\mathfrak{H}^{\mathsf{O}_{n-1}}(\mathsf{O}) \stackrel{(30)}{\geqslant} \int_{\mathsf{O}_{n-1}} \mathfrak{p}_\nu^\alpha \, \mathrm{d}\mathfrak{H}^{\mathsf{O}_{n-1}}(\mathsf{O}) = \mathfrak{p}_\nu^\alpha.
$$

So, the maximum of $\mathfrak{p}_\nu^\alpha$ over the Borel probability measures ν on $\mathsf{M}_{k \times (n-1)}(\mathbb{R})$ is attained at μ. $\qquad\square$

Remark 36. Recalling (33), due to (40) the conclusion (26) of the JL lemma holds if k satisfies

$$
\frac{2\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} \int_{\frac{1}{\alpha} \sigma_{\max}(n, k, \alpha)}^{\sigma_{\max}(n, k, \alpha)} \frac{(s^2 - 1)^{\frac{n-k-3}{2}}}{s^{n-2}} \, \mathrm{d}s > 1 - \frac{2}{n(n-1)}, \tag{41}
$$

where $\sigma_{\max}(n, k, \alpha)$ is given in (39). This is the best-known bound on k in the JL lemma, which, due to Proposition 35, is the best-possible bound that is obtainable through the reasoning (31). In particular, the asymptotic estimate (28) follows from (41) via straightforward elementary calculus.

Remark 37. The JL lemma was reproved many times; see [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**], though we make no claim that this is a comprehensive list of references. There were several motivations for these further investigations, ranging from the desire to obtain an overall better understanding of the JL phenomenon, to obtain better bounds, and to obtain distributions on random matrices A as in (31) with certain additional properties that are favorable from the computational perspective, such as ease of simulation, use of fewer random bits, sparsity, and the ability to evaluate the mapping $(z \in \mathbb{R}^{n-1}) \mapsto \mathsf{A}z$ quickly (akin to the fast Fourier transform). This body of work represents ongoing efforts by computer scientists and applied mathematicians to further develop improved "JL transforms," driven by their usefulness as a tool for data-compression. We will not survey these ideas here, partially because we established that using random projections yields the best-possible bound on the target dimension k (moreover, this procedure is natural and simple). We speculate that working with the Haar measure on the orthogonal group O_{n-1} as in (29) could have benefits that address the above computational issues, but leave this as an interesting open-ended direction for further research. A specific conjecture towards this goal appears in [**?**, page 320], and we suspect that the more recent work [**?**] on the spectral gap of Hecke operators of orthogonal Cayley graphs should be relevant in this context as well (e.g. for derandomization and fast implementation of (29); see [**?**, **?**] for steps in this direction).

Remark 38. In the literature there is often a preference to use random matrices with independent entries in the context of the JL lemma, partially because they are simple to generate, though see the works [**?**, **?**, **?**] on generating elements of the orthogonal group O_{n-1} that are distributed according to its Haar measure. In particular, the best bound on k in (25) that was previously available in the literature [**?**] arose from applying (31) when A is replaced by the random matrix $\sigma\mathsf{G}$, where $\sigma = 1/\sqrt{k}$ and the entries of $\mathsf{G} = (\mathsf{g}_{ij}) \in M_{k \times (n-1)}(\mathbb{R})$ are independent standard Gaussian random variables. We can, however, optimize over the scaling factor σ in this setting as well, in analogy to the above optimization over the scaling factor in (29), despite the fact that we know that working with the Gaussian matrix G is inferior to using a random rotation. A short calculation reveals that the optimal scaling factor is now $\sqrt{(\alpha^2 - 1)/(2k \log \alpha)}$, i.e., the best possible re-scaled Gaussian matrix for the purpose of reasoning as in (31) is not $\frac{1}{\sqrt{k}}\mathsf{G}$ but rather the random matrix

$$\mathsf{G}_k^\alpha \overset{\text{def}}{=} \sqrt{\frac{\alpha^2 - 1}{2k \log \alpha}} \cdot \mathsf{G}. \tag{42}$$

For this optimal multiple of a Gaussian matrix, one computes that for every $z \in \mathbf{S}^{n-2}$ we have

$$1 - \mathbb{P}\Big[1 \leqslant \|\mathsf{G}_k^\alpha z\|_{\ell_2^k} \leqslant \alpha\Big] = \frac{2k^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} \int_{\log \alpha}^{\infty} \left(\frac{\beta}{e^{2\beta}-1}\right)^{\frac{k}{2}} \exp\left(-\frac{k\beta}{e^{2\beta}-1}\right) \mathrm{d}\beta$$

$$< \frac{4k^{\frac{k}{2}-1}}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{\alpha^2-1}{\log\alpha}\alpha^{\frac{2}{\alpha^2-1}}\right)^{-\frac{k}{2}} \frac{(\alpha^2-1)^2 \log \alpha}{2\alpha^4 \log\alpha + 2\alpha^2 - \alpha^4 - 4\alpha^2(\log\alpha)^2 - 2\log\alpha - 1}. \tag{43}$$

The first step in (43) follows from a straightforward computation using the fact that the squared Euclidean length of $\mathsf{G}_k^\alpha z$ is distributed according to a multiple of the χ^2 distribution with k degrees of freedom (see e.g. [**?**]), i.e., one can write the leftmost term of (43) explicitly as a definite integral, and then check that it indeed equals the middle term of (43), e.g., by verifying the the derivatives with respect to α of both expressions coincide. The final estimate in (43) can be justified via a modicum of straightforward calculus. We deduce from this that the conclusion (26) of the JL lemma is holds with positive probability if for each $i \in \{1, \ldots, n\}$ we take y_i to be the image of x_i under the re-scaled Gaussian matrix in (42), provided that k is sufficiently large so as to ensure that

$$\frac{\Gamma\left(\frac{k}{2}\right)}{k^{\frac{k}{2}-1}} \left(\frac{\alpha^2-1}{\log\alpha}\alpha^{\frac{2}{\alpha^2-1}}\right)^{\frac{k}{2}} \geqslant \frac{2n^2(\alpha^2-1)^2 \log \alpha}{2\alpha^4 \log\alpha + 2\alpha^2 - \alpha^4 - 4\alpha^2(\log\alpha)^2 - 2\log\alpha - 1}. \tag{44}$$

The values that we stated for the target dimension k in the JL lemma with a billion vectors were obtained by using (44), though even better bounds arise from an evaluation of the integral in (43) numerically, which is what we recommend to do for particular settings of the parameters. As $\alpha \to 1$, the above bounds improve over those of [**?**] only in the second-order terms. For larger α these bounds yield substantial improvements that might matter in practice, e.g. for embedding a billion vectors with distortion 2, the target dimension that is required using the best-available estimate in the literature [**?**] is $k = 768$, while (44) shows that $k = 329$ suffices.

3. Infinite subsets of Hilbert space

The JL lemma provides a quite complete understanding of the metric dimension reduction problem for finite subsets of Hilbert space. For infinite subsets of Hilbert space, the research splits into two strands. The first is to understand those subsets $\mathcal{C} \subseteq \mathbb{R}^n$ for which certain random matrices in $\mathsf{M}_{k \times n}(\mathbb{R})$ (e.g. random projections, or matrices whose entries are i.i.d. independent sub-Gaussian random variables) yield with positive probability an embedding of $\mathcal{C}$ into $\mathbb{R}^k$ of a certain pre-specified distortion; this was pursued in [**?, ?, ?, ?, ?, ?, ?, ?**], yielding a satisfactory answer which relies on multi-scale chaining criteria [**?, ?**] .

The second (and older) research strand focuses on the mere *existence* of a low-dimensional embedding rather than on the success of the specific embedding approach of (all the known proofs of) the JL lemma. Specifically, given a subset $\mathcal{C}$ of a Hilbert space and $\alpha \in [1, \infty)$, could one understand when does $\mathcal{C}$ admit an embedding with distortion α into ℓ_2^k for some $k \in \mathbb{N}$? If one ignores the dependence on the distortion α, then this qualitative question coincides with Problem 2 (the bi-Lipschitz embedding problem into $\mathbb{R}^k$), since if a metric space $(\mathcal{M}, d_{\mathcal{M}})$ satisfies $\inf_{k \in \mathbb{N}} \mathsf{c}_{\mathbb{R}^k}(\mathcal{M}) < \infty$, then in particular it admits a bi-Lipschitz embedding into a Hilbert space.

We shall next describe an obvious necessary condition for bi-Lipschitz embeddability into $\mathbb{R}^k$ for some $k \in \mathbb{N}$. In what follows, all balls in a metric space $(\mathcal{M}, d_{\mathcal{M}})$ will be closed balls, i.e., for $x \in \mathcal{M}$ and $r \in [0, \infty)$ we write $B_{\mathcal{M}}(x, r) = \{y \in \mathcal{M} : d_{\mathcal{M}}(x, y) \leqslant r\}$. Given $K \in [2, \infty)$, a metric space $(\mathcal{M}, d_{\mathcal{M}})$ is said to be K-*doubling* (e.g. [**?, ?**]) if every ball in $\mathcal{M}$ (centered anywhere in $\mathcal{M}$ and of any radius) can be covered by at most K balls of half its radius, i.e., for every $x \in \mathcal{M}$ and $r \in [0, \infty)$ there is $m \in \mathbb{N}$ with $m \leqslant K$ and $y_1, \ldots, y_m \in \mathcal{M}$ such that $B_{\mathcal{M}}(x, r) \subseteq B_{\mathcal{M}}(y_1, \frac{1}{2}r) \cup \ldots \cup B_{\mathcal{M}}(y_m, \frac{1}{2}r)$. A metric space is doubling if it is K-doubling for some $K \in [2, \infty)$.

Fix $k \in \mathbb{N}$ and $\alpha \geqslant 1$. If a metric space $(\mathcal{M}, d_{\mathcal{M}})$ embeds with distortion α into a normed space $(\mathbb{R}^k, \|\cdot\|)$, then $\mathcal{M}$ is $(4\alpha + 1)^k$-doubling. Indeed, fix $x \in \mathcal{M}$ and $r > 0$. Let $\{z_1, \ldots, z_n\} \subseteq B_{\mathcal{M}}(x, r)$ be a maximal subset (with respect to inclusion) of $B_{\mathcal{M}}(x, r)$ satisfying $d_{\mathcal{M}}(z_i, z_j) > \frac{1}{2}r$ for distinct $i, j \in \{1, \ldots, n\}$. The maximality of $\{z_1, \ldots, z_n\}$ ensures that for any $w \in B_{\mathcal{M}}(x, r) \smallsetminus \{z_1, \ldots, z_n\}$ we have $\min_{i \in \{1, \ldots, n\}} d_{\mathcal{M}}(w, z_i) \leqslant \frac{1}{2}r$, i.e., $B_{\mathcal{M}}(x, r) \subseteq B_{\mathcal{M}}(z_1, \frac{1}{2}r) \cup \ldots \cup B_{\mathcal{M}}(z_n, \frac{1}{2}r)$. We are assuming that there is an embedding $f : \mathcal{M} \to \mathbb{R}^k$ that satisfies $d_{\mathcal{M}}(u, v) \leqslant \|f(u) - f(v)\| \leqslant \alpha d_{\mathcal{M}}(u, v)$ for all $u, v \in \mathcal{M}$. So, for distinct $i, j \in \{1, \ldots, n\}$ we have $\frac{r}{2} < d_{\mathcal{M}}(z_i, z_j) \leqslant \|f(z_i) - f(z_j)\| \leqslant \alpha d_{\mathcal{M}}(z_i, z_j) \leqslant \alpha(d_{\mathcal{M}}(z_i, x) + d_{\mathcal{M}}(x, z_j)) \leqslant 2\alpha r$. The reasoning that led to (27) with $y_1 = \frac{2}{r} f(z_1), \ldots, y_n = \frac{2}{r} f(z_n)$ and α replaced by 4α gives $k \geqslant (\log n) / \log(4\alpha + 1)$, i.e., $n \leqslant (4\alpha + 1)^k$.

Remark 39. In Section 1.3 we recalled that in the context of the Ribe program $\log |\mathcal{M}|$ was the initial (in hindsight somewhat naive, though still very useful) replacement for the "dimension" of a finite metric space $\mathcal{M}$. This arises naturally also from the above discussion. Indeed, $\mathcal{M}$ is trivially $|\mathcal{M}|$-doubling (simply cover each ball in $\mathcal{M}$ by singletons), and this is the best bound that one could give on the doubling constant of $\mathcal{M}$ in terms of $|\mathcal{M}|$. So, from the perspective of the doubling property, the natural restriction on $k \in \mathbb{N}$ for which there exists an embedding of $\mathcal{M}$ into some k-dimensional normed space with $O(1)$ distortion is that $k \gtrsim \log |\mathcal{M}|$.

Using terminology that was recalled in Remark 18, the definition of the doubling property directly implies that for every $\theta \in (0, 1)$ a metric space $\mathcal{M}$ is doubling if and only if its θ-snowflake $\mathcal{M}^\theta$ is doubling. With this in mind, Theorem 40 below is a very important classical achievement of Assouad [**?**].

Theorem 40. *The following assertions are equivalent for every metric space $(\mathcal{M}, d_{\mathcal{M}})$.*

- $\mathcal{M}$ *is doubling.*
- *For every* $\theta \in (0, 1)$ *there is* $k \in \mathbb{N}$ *such that* $\mathcal{M}^\theta$ *admits a bi-Lipschitz embedding into* $\mathbb{R}^k$.
- *Some snowflake of* $\mathcal{M}$ *admits a bi-Lipschitz embedding into* $\mathbb{R}^k$ *for some* $k \in \mathbb{N}$.

Theorem 40 is a qualitative statement, but its proof in [**?**] shows that for every $K \in [2, \infty)$ and $\theta \in (0, 1)$, there are $\alpha(K, \theta) \in [1, \infty)$ and $k(K, \theta) \in \mathbb{N}$ such that if $\mathcal{M}$ is K-doubling, then $\mathcal{M}^\theta$ embeds into $\mathbb{R}^{k(K, \theta)}$ with distortion $\alpha(K, \theta)$; the argument of [**?**] inherently gives that as $\theta \to 1$, i.e., as the θ-snowflake $\mathcal{M}^\theta$ approaches the initial metric space $\mathcal{M}$, we have $\alpha(K, \theta) \to \infty$ and $k(K, \theta) \to \infty$. A meaningful study of the best-possible asymptotic behavior of the distortion $\alpha(K, \theta)$ here would require specifying which norm on $\mathbb{R}^k$ is being considered. Characterizing the quantitative dependence in terms of geometric properties of the target norm on $\mathbb{R}^k$ has not been carried out yet (it isn't even clear what should the pertinent geometric properties be), though see [**?**] for an almost isometric version when one considers the ℓ_∞ norm on $\mathbb{R}^k$ (with the dimension k tending to ∞ as the distortion approaches 1); see also [**?**] for a further partial step in this direction. In [**?**] it was shown that for $\theta \in [\frac{1}{2}, 1)$ one could take $k(K, \theta) \leqslant k(K)$ to be bounded by a constant that depends only on K; the

proof of this fact in [**?**] relies on a probabilistic construction, but in [**?**] a clever and instructive deterministic proof of this phenomenon was found (though, yielding asymptotically worse estimates on $\alpha(K,\theta), k(K)$ than those of [**?**]).

Assouad's theorem is a satisfactory characterization of the doubling property in terms of embeddability into finite-dimensional Euclidean space. However, it is a "near miss" as an answer to Problem 2: the same statement with $\theta = 1$ would have been a wonderful resolution of the bi-Lipschitz embedding problem into $\mathbb{R}^k$, showing that a simple intrinsic ball covering property is equivalent to bi-Lipschitz embeddability into some $\mathbb{R}^k$. It is important to note that while the snowflaking procedure does in some sense "tend to" the initial metric space as $\theta \to 1$, for $\theta < 1$ it deforms the initial metric space substantially (e.g. such a θ-snowflake does not contain any non-constant rectifiable curve). So, while Assouad's theorem with the stated snowflaking is useful (examples of nice applications appear in [**?**, **?**]), its failure to address the bi-Lipschitz category is a major drawback.

Alas, more than a decade after the publication of Assouad's theorem, it was shown in [**?**] (relying a on a rigidity theorem of [**?**]) that Assouad's theorem does not hold with $\theta = 1$, namely there exists a doubling metric space that does not admit a bi-Lipschitz embedding into $\mathbb{R}^k$ for any $k \in \mathbb{N}$. From the qualitative perspective, we now know that the case $\theta = 1$ of Assouad's theorem fails badly in the sense that there exists a doubling metric space (the continuous 3-dimensional Heisenberg group, equipped with the Carnot–Carathéodory metric) that does not admit a bi-Lipschitz embedding into any Banach space with the Radon–Nikodým property [**?**, **?**] (in particular, it does not admit a bi-Lipschitz embedding into any reflexive or separable dual Banach space, let alone a finite dimensional Banach space), into any $L_1(\mu)$ space [**?**], or into any Alexandrov space of curvature bounded above or below [**?**] (a further strengthening appears in the forthcoming work [**?**]). From the quantitative perspective, by now we know that balls in the discrete 5-dimensional Heisenberg group equipped with the word metric (which is doubling) have the asymptotically worst-possible bi-Lipschitz distortion (as a function of their cardinality) in uniformly convex Banach spaces [**?**] (see also [**?**]) and $L_1(\mu)$ spaces [**?**, **?**]; interestingly, the latter assertion is not true for the 3-dimensional Heisenberg group [**?**], while the former assertion does hold true for the 3-dimensional Heisenberg group [**?**].

All of the known "bad examples" (including, in addition to the Heisenberg group, those that were subsequently found in [**?**, **?**, **?**, **?**]) which show that the doubling property is not the sought-after answer to Problem 2 do not even embed into an infinite-dimensional Hilbert space. This leads to the following natural and intriguing question that was stated by Lang and Plaut in [**?**].

Question 41. Does every doubling subset of a Hilbert admit a bi-Lipschitz embedding into $\mathbb{R}^k$ for some $k \in \mathbb{N}$?

As stated, Question 41 is qualitative, but by a compactness argument (see [**?**, Section 4]) if its answer were positive, then for every $K \in [2, \infty)$ there would exist $d_K \in \mathbb{N}$ and $\alpha_K \in [1, \infty)$ such that any K-doubling subset of a Hilbert space would embed into $\ell_2^{d_K}$ with distortion α_K. If Question 41 had a positive answer, then it would be very interesting to determine the asymptotic behavior of d_K and α_K as $K \to \infty$. A positive answer to Question 41 would be a solution of Problem 2, though the intrinsic criterion that it would provide would be quite complicated, namely it would say that a metric space $(\mathcal{M}, d_\mathcal{M})$ admits a bi-Lipschitz embedding into $\mathbb{R}^k$ for some $k \in \mathbb{N}$ if and only if it is doubling and satisfies the family of quadratic distance inequalities (2). More importantly, it seems that any positive answer to Question 41 would devise a procedure that starts with a subset in a very high-dimensional Euclidean space and, if that subset is $O(1)$-doubling, produce a bi-Lipschitz embedding into $\mathbb{R}^{O(1)}$; such a procedure, if possible, would be a quintessential metric dimension reduction result that is bound to be of major importance. It should be noted that, as proved in [**?**, Remark 4.1], any such general procedure cannot be an embedding into low-dimensions via a linear mapping as in the JL lemma, i.e., Question 41 calls for a genuinely nonlinear dimension reduction technique.[6]

Despite the above reasons why a positive answer to Question 41 would be very worthwhile, we suspect that Question 41 has a negative answer. A specific doubling subset of a Hilbert space which is a potential counterexample to Question 41 was constructed in [**?**, Question 3], but to date it remains unknown whether or not this subset admits a bi-Lipschitz embedding into $\mathbb{R}^{O(1)}$. If the answer to Question 41 is indeed negative, then the next challenge would be to formulate a candidate conjectural characterization for resolving the bi-Lipschitz embedding problem into $\mathbb{R}^k$.

---

[6]On its own, the established necessity of obtaining a genuinely nonlinear embedding method into low dimensions should not discourage attempts to answer Question 41, because some rigorous nonlinear dimension reduction methods have been devised in the literature; see e.g. [**?**, **?**]. However, all of these approaches seem far from addressing Question 41.

The analogue of Question 41 is known to fail in some non-Hilbertian settings. Specifically, it follows from [**?**, **?**, **?**] that for every $p \in (2, \infty)$ there exists a doubling subset $\mathcal{D}_p$ of $L_p(\mathbb{R})$ that does not admit a bi-Lipschitz embedding into any $L_q(\mu)$ space for any $q \in [1, p)$. So, in particular there is no bi-Lipschitz embedding of $\mathcal{D}_p$ into any finite-dimensional normed space, and a fortiori there is no such embedding into any finite-dimensional subspace of $L_p(\mathbb{R})$. Note that in [**?**] this statement is made for embeddings of $\mathcal{D}_p$ into $L_q(\mu)$ in the reflexive range $q \in (1, p)$, and the case $q = 1$ is treated in [**?**] only when $p \geqslant p_0$ for some universal constant $p_0 > 2$. The fact that $\mathcal{D}_p$ does not admit a bi-Lipschitz embedding into any $L_1(\mu)$ space follows by combining the argument of [**?**] with the more recent result[7] of [**?**, **?**] when the underlying group in the construction of [**?**] is the 5-dimensional Heisenberg group; interestingly we now know [**?**] that if one carries out the construction of [**?**] for the 3-dimensional Heisenberg group, then the reasoning of [**?**] would yield the above conclusion only when $p > 4$. A different example of a doubling subset of $L_p(\mathbb{R})$ that fails to embed bi-Lipschitzly into ℓ_p^k for any $k \in \mathbb{N}$ was found in [**?**]. In $L_1(\mathbb{R})$, there is an even stronger counterexample [**?**, Remark 1.4]: By [**?**], the spaces considered in [**?**, **?**] yields a doubling subset of $L_1(\mathbb{R})$ that by [**?**] (see also [**?**]) does not admit a bi-Lipschitz embedding into any Banach space with the Radon–Nikodým property [**?**, **?**], hence it does not admit a bi-Lipschitz embedding into any reflexive or separable dual Banach space. The potential validity of the above statement for $p \in (1, 2)$ remains an intriguing open problem, and the case $p = 2$ is of course Question 41.

4. Matoušek's random metrics, Milnor–Thom, and coarse dimension reduction

Fix two moduli $\omega, \Omega : [0, \infty) \to [0, \infty)$ as in Theorem 17, i.e., they are increasing functions and $\omega \leqslant \Omega$ point-wise. For a metric space $(\mathcal{M}, d_{\mathcal{M}})$ define $\dim_{(\omega, \Omega)}(\mathcal{M}, d_{\mathcal{M}})$ to be the smallest dimension $k \in \mathbb{N}$ for which there exists a k-dimensional normed space $(X, \|\cdot\|_X) = (X(\mathcal{M}), \|\cdot\|_{X(\mathcal{M})})$ and a mapping $f : \mathcal{M} \to X$ that satisfies (18). If no such $k \in \mathbb{N}$ exists, then write $\dim_{(\omega, \Omega)}(\mathcal{M}, d_{\mathcal{M}}) = \infty$. For $\alpha \in [1, \infty)$, this naturally generalizes the notation $\dim_\alpha(\mathcal{M}, d_{\mathcal{M}})$ of [**?**] in the bi-Lipschitz setting, which coincides with $\dim_{(t, \alpha t)}(\mathcal{M}, d_{\mathcal{M}})$.

Recalling (17), the goal of this section is to show that $\dim_{(\omega, \Omega)}(\mathcal{M}, d_{\mathcal{M}}) \geqslant n^{c\beta(\omega, \Omega)}$ for arbitrarily large $n \in \mathbb{N}$, some universal constant $c \in (0, \infty)$ and some metric space $(\mathcal{M}, d_{\mathcal{M}})$ with $|\mathcal{M}| = 3n$, thus proving Theorem 17. We will do so by following Matoušek's beautiful ideas in [**?**], yielding a probabilistic argument for the existence of such an intrinsically (coarsely) high-dimensional metric space $(\mathcal{M}, d_{\mathcal{M}})$.

The collections of subsets of a set S of size $\ell \in \mathbb{N}$ will be denoted below $\binom{S}{\ell} = \{\mathfrak{e} \subseteq S : |\mathfrak{e}| = \ell\}$. Fix $n \in \mathbb{N}$ and a bipartite graph $\mathsf{G} = (\mathsf{L}, \mathsf{R}, \mathsf{E})$ with $|\mathsf{L}| = |\mathsf{R}| = n$. Thus, L and R are disjoint n-point sets (the "left side" and "right side" of G) and E is a subset of $\binom{\mathsf{L} \cup \mathsf{R}}{2}$ consisting only of $\mathfrak{e} \subseteq \mathsf{L} \cup \mathsf{R}$ such that $|\mathsf{L} \cap \mathfrak{e}| = |\mathsf{R} \cap \mathfrak{e}| = 1$. Following Matoušek [**?**], any such graph G can used as follows as a "template" for obtaining a family $2^{|\mathsf{E}|}$ graphs, each of which having $3n$ vertices. For each $\lambda \in \mathsf{L}$ introduce two new elements λ^+, λ^-. Denote $\mathsf{L}^+ = \{\lambda^+ : \lambda \in \mathsf{L}\}$ and $\mathsf{L}^- = \{\lambda^- : \lambda \in \mathsf{L}\}$. Assume that the sets $\mathsf{L}^+, \mathsf{L}^-, \mathsf{R}$ are disjoint. For every $\sigma : \mathsf{E} \to \{-, +\}$ define

$$\mathsf{E}_\sigma \stackrel{\text{def}}{=} \left\{ \{\lambda^{\sigma(\{\lambda, \rho\})}, \rho\} : (\lambda, \rho) \in \mathsf{L} \times \mathsf{R} \ \wedge \ \{\lambda, \rho\} \in \mathsf{E} \right\} \subseteq \binom{\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}}{2}. \tag{45}$$

We thus obtain a bipartite graph $\mathsf{G}_\sigma = (\mathsf{L}^+ \cup \mathsf{L}^-, \mathsf{R}, \mathsf{E}_\sigma)$. By choosing $\sigma : \mathsf{E} \to \{-, +\}$ uniformly at random, we think of G_σ as a random graph; let $\mathbb{P}$ denote the uniform probability measure on the set of all such σ. In other words, consider $\sigma : \mathsf{E} \to \{-, +\}$ to be independent tosses of a fair coin, one for each edge of G. Given an outcome of the coin tosses σ, each edge $\mathfrak{e} \in \mathsf{E}$ of G induces an element of E_σ as follows. If λ is the endpoint of $\mathfrak{e}$ in L and ρ is the endpoint of $\mathfrak{e}$ in R, then E_σ contains exactly one of the unordered pairs $\{\lambda^+, \rho\}, \{\lambda^-, \rho\}$ depending on whether $\sigma(\mathfrak{e}) = +$ or $\sigma(\mathfrak{e}) = -$, respectively; see Figure 1 below for a schematic depiction of this construction.

Let $d_{\mathsf{G}_\sigma} : (\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}) \times (\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}) \to [0, \infty]$ be the shortest-path metric corresponding to G_σ, with the convention that $d_{\mathsf{G}_\sigma}(x, y) = \infty$ if $x, y \in \mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}$ belong to different connected components of G_σ. We record for convenience of later use the following very simple observation.

Claim 42. *Fix $\lambda \in \mathsf{L}$ and $\sigma : \mathsf{E} \to \{-, +\}$. Suppose that $k \stackrel{\text{def}}{=} d_{\mathsf{G}_\sigma}(\lambda^+, \lambda^-) < \infty$. Then the original "template graph" G contains a cycle of length at most k.*

Proof. Denote by $\pi : \mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R} \to \mathsf{L} \cup \mathsf{R}$ the canonical "projection," i.e., π is the identity mapping on R and $\pi(\lambda^+) = \pi(\lambda^-) = \lambda$ for every $\lambda \in \mathsf{L}$. The natural induced mapping $\pi : \mathsf{E}_\sigma \to \mathsf{E}$ (given by $\pi(\{x, y\}) = \{\pi(x), \pi(y)\}$
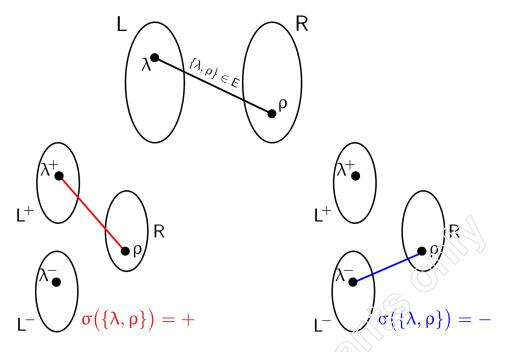
---

[7]When [**?**] was written, only a weaker bound of [**?**] was known.

Figure 1 – The random bipartite graph $\mathsf{G}_\sigma = (\mathsf{L}^+ \cup \mathsf{L}^-, \mathsf{R}, \mathsf{E}_\sigma)$ that is associated to the bipartite graph $\mathsf{G} = (\mathsf{L}, \mathsf{R}, \mathsf{E})$ and coin flips $\sigma : \mathsf{E} \to \{-, +\}$. Suppose that $(\lambda, \rho) \in \mathsf{L} \times \mathsf{R}$ and $\mathfrak{e} = \{\lambda, \rho\} \in \mathsf{E}$. If the outcome of the coin that was flipped for the edge $\mathfrak{e}$ is $+$, then include in E_σ the red edge between λ^+ and ρ and do not include an edge between λ^- and ρ. If the outcome of the coin that was flipped for the edge $\mathfrak{e}$ is $-$, then include in E_σ the blue edge between λ^- and ρ and do not include an edge between λ^+ and ρ.

for each $\{x, y\} \in \mathsf{E}_\sigma$) is one-to-one, because by construction E_σ contains one and only one of the unordered pairs $\{\mu^+, \rho\}, \{\mu^-, \rho\}$ for each $(\mu, \rho) \in \mathsf{L} \times \mathsf{R}$ with $\{\mu, \rho\} \in \mathsf{E}$.

Let $\gamma : \{0, \ldots, k\} \to \mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}$ be a geodesic in G_σ that joins λ^+ and λ^-. Thus $\gamma(0) = \lambda^+$, $\gamma(k) = \lambda^-$ and $\{\{\gamma(i-1), \gamma(i)\}\}_{i=1}^k$ are distinct edges in E_σ (they are distinct because γ is a shortest path joining λ^+ and λ^- in G_σ). By the injectivity of π on E_σ, the unordered pairs $\{\{\pi(\gamma(i-1)), \pi(\gamma(i))\}\}_{i=1}^k$ are distinct edges in E. So, the subgraph H of G that is induced on the vertices $\{\pi(\gamma(i))\}_{i=0}^k$ has at least k edges. But, H has at most k vertices, because $\pi(\gamma(0)) = \pi(\gamma(k)) = \lambda$. Hence H is not a tree, i.e., it contains a cycle of length at most k. $\square$

Even though d_{G_σ} is not necessarily a metric due to its possible infinite values, for every $s, T \in (0, \infty)$ we can re-scale and truncate it so as to obtain a (random) metric $d_\sigma^{s,T} : (\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}) \times (\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}) \to [0, \infty]$ by defining

$$\forall\, x, y \in \mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}, \qquad d_\sigma^{s,T}(x, y) \overset{\text{def}}{=} \min\left\{ sd_{\mathsf{G}_\sigma}(x, y), T \right\}. \tag{46}$$

The following lemma shows that if in the above construction G has many edges and no short cycles, then with overwhelmingly high probability the random metric in (46) has large coarse metric dimension.

Lemma 43. *There is a universal constant $\eta > 0$ with the following property. For every $\omega, \Omega : [0, \infty) \to [0, \infty)$ as above, every $n \in \mathbb{N}$ and every template graph $\mathsf{G} = (\mathsf{L}, \mathsf{R}, \mathsf{E})$ as above, suppose that $g \in \mathbb{N}$ and $s, T > 0$ satisfy*

$$\frac{\omega^{-1}(2\Omega(s))}{s} < g \leqslant \frac{T}{s}, \tag{47}$$

and that the shortest cycle in G has length at least g. Then for every $\delta \in (0, \frac{1}{3}]$ we have

$$\mathbb{P}\left[\sigma : \mathsf{E} \to \{-, +\} : \ \dim_{\omega,\Omega}\left(\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}, d_\sigma^{s,T} \right) \leqslant \delta\eta \frac{|\mathsf{E}|}{n} \right] < \left(2\delta^\delta \right)^{-|\mathsf{E}|}. \tag{48}$$

In particular, by choosing $\delta = \frac{1}{3}$ in (48) we deduce that

$$\mathbb{P}\left[\sigma : \mathsf{E} \to \{-, +\} : \ \dim_{\omega,\Omega}\left(\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}, d_\sigma^{s,T} \right) > \frac{\eta|\mathsf{E}|}{3n} \right] > 1 - e^{-\frac{1}{5}|\mathsf{E}|}. \tag{49}$$

Prior to proving Lemma 43, we shall now explain how it implies Theorem 17.

Proof of Theorem 17 assuming Lemma 43. Recalling (17), we can fix $s \in (0, \infty)$ such that

$$g \stackrel{\text{def}}{=} \left\lfloor \frac{\omega^{-1}(2\Omega(s))}{s} \right\rfloor + 1 \leqslant \frac{2}{\beta(\omega, \Omega)}. \tag{50}$$

There is a universal constant $\kappa \in (0, \infty)$ such that for arbitrarily large $n \in \mathbb{N}$ there exists a bipartite graph $\mathsf{G} = (\mathsf{L}, \mathsf{R}, \mathsf{E})$ with $|\mathsf{L}| = |\mathsf{R}| = n$, girth at least g (i.e., G does not contain any cycle of length smaller than g) and $|\mathsf{E}| \geqslant n^{1+\kappa/g}$. Determining the largest possible value of κ here is a well-studied and longstanding open problem in graph theory (see e.g. the discussions in [?, ?, ?]), but for the present purposes any value of κ suffices. For the latter (much more modest) requirement, one can obtain G via a simple probabilistic construction (choosing each of the edges independently at random and deleting an arbitrary edge from each cycle of length at most $g - 1$). See [?] for the best known lower bound on κ here (arising from an algebraic construction).

We shall use the above graph G as the template graph for the random graphs $\{\mathsf{G}_\sigma\}_{\sigma:\mathsf{E}\to\{-,+\}}$. Our choice of g in (50) ensures that if we write $T = sg$, then (47) holds true and we obtain a distribution over metric spaces $(\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}, d_\sigma^{s,T})$ for which the conclusion (49) of Lemma 43 holds true. Hence, by choosing $c = \kappa/2$ and substituting the bound $|\mathsf{E}| \geqslant n^{1+\kappa/g}$ into (49) while using (50) we have

$$\mathbb{P}\left[\sigma : \mathsf{E} \to \{-, +\} : \ \dim_{\omega,\Omega}\left(\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}, d_\sigma^{s,T}\right) \gtrsim n^{c\beta(\omega,\Omega)} \right] \geqslant 1 - \exp\left(-\frac{1}{5} n^{1+c\beta(\omega,\Omega)}\right).$$

Consequently, by the definition of $\dim_{\omega,\Omega}(\cdot)$, with probability exponentially close to 1 the random metric space $(\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}, d_\sigma^{s,T})$ satisfies the assertion of Theorem 17. $\qquad\square$

The proof of Lemma 43 relies on the following lemma that was obtained implicitly by Matoušek [?]. Its proof takes as input a clever argument of Alon [?] which uses the classical bound of Milnor [?] and Thom [?] on the number of connected components of a real algebraic variety.

Lemma 44. *Fix $m, n \in \mathbb{N}$ and $\mathsf{E} \subseteq \{1, \ldots, n\}^2$. Suppose that $\mathsf{A}_1 = (a_{ij}^1), \ldots, \mathsf{A}_m = (a_{ij}^m) \in \mathsf{M}_n(\mathbb{R})$ are matrices that satisfy $a_{ij}^k \neq 0$ for all $(i, j) \in \mathsf{E}$ and $k \in \{1, \ldots, m\}$, and that the sign vectors*

$$\left(\mathbf{sign}(a_{ij}^1)\right)_{(i,j)\in\mathsf{E}}, \left(\mathbf{sign}(a_{ij}^2)\right)_{(i,j)\in\mathsf{E}} \ldots, \left(\mathbf{sign}(a_{ij}^m)\right)_{(i,j)\in\mathsf{E}} \subseteq \{-1, 1\}^{\mathsf{E}} \tag{51}$$

are distinct. Then there exists $k \in \{1, \ldots, m\}$ such that

$$\mathbf{rank}(\mathsf{A}_k) \gtrsim \frac{\log m}{n \log\left(\frac{|\mathsf{E}|}{\log m}\right)}. \tag{52}$$

Proof. Let $\alpha \in \mathbb{N}$ be an auxiliary parameter that will be specified later so as to optimize the ensuing argument. Write $h = \lceil |\mathsf{E}|/\alpha \rceil$ and fix any partition of E into subsets $\mathsf{J}_1, \ldots, \mathsf{J}_h \neq \varnothing$ (i.e., $\mathsf{J}_1, \ldots, \mathsf{J}_h \subseteq \mathsf{E}$ are pairwise disjoint and $\mathsf{E} = \mathsf{J}_1 \cup \ldots \cup \mathsf{J}_h$) that satisfy $|\mathsf{J}_u| \leqslant \alpha$ for all $u \in \{1, \ldots, h\}$.

Denote

$$\mu \stackrel{\text{def}}{=} \min_{\substack{(i,j)\in\mathsf{E} \\ k\in\{1,\ldots,m\}}} |a_{ij}^k| \qquad \text{and} \qquad r \stackrel{\text{def}}{=} \max_{k\in\{1,\ldots,m\}} \mathbf{rank}(\mathsf{A}_k). \tag{53}$$

Lemma 44 assumes that $\mu > 0$, and its goal is to show that r is at least a universal constant multiple of the quantity that appears in the right hand side on (52). The definition of r means that for every $k \in \{1, \ldots, m\}$ there exist n-by-r and r-by-n matrices $\mathsf{B}_k \in \mathsf{M}_{n\times r}(\mathbb{R})$ and $\mathsf{C}_k \in \mathsf{M}_{r\times n}(\mathbb{R})$, respectively, such that $\mathsf{A}_k = \mathsf{B}_k\mathsf{C}_k$. Define vectors $\{\zeta_k = (\zeta_1^k, \ldots, \zeta_h^k) \in \mathbb{R}^h\}_{k=1}^m$ by setting

$$\forall (k, u) \in \{1, \ldots, m\} \times \{1, \ldots, h\}, \qquad \zeta_u^k \stackrel{\text{def}}{=} \sqrt{\prod_{(i,j)\in\mathsf{J}_u} (a_{ij}^k)^2 - \frac{1}{2}\mu^{2|\mathsf{J}_u|}}. \tag{54}$$

Observe that the definition of μ in (53) ensures that the quantity under the square root in (54) is positive, so $\zeta_u \in (0, \infty)$. Define polynomials $\{\mathfrak{p}_u : \mathsf{M}_{n\times r}(\mathbb{R}) \times \mathsf{M}_{r\times n}(\mathbb{R}) \times \mathbb{R}^h \to \mathbb{R}\}_{u=1}^h$ by setting

$$\mathfrak{p}_u(X, Y, z) \stackrel{\text{def}}{=} \prod_{(i,j)\in\mathsf{J}_u} (XY)_{ij}^2 - z_u^2 - \frac{1}{2}\mu^{2|\mathsf{J}_u|} = \prod_{(i,j)\in\mathsf{J}_u} \left(\sum_{k=1}^n x_{ik}y_{kj}\right)^2 - z_u^2 - \frac{1}{2}\mu^{2|\mathsf{J}_u|}, \tag{55}$$

27

for all $u \in \{1, \dots, h\}$, $X = (x_{is}) \in \mathsf{M}_{n \times r}(\mathbb{R})$, $Y = (y_{sj}) \in \mathsf{M}_{r \times n}(\mathbb{R})$ and $z = (z_i) \in \mathbb{R}^h$. The above notation ensures that $\mathfrak{p}_u(\mathsf{B}_k, \mathsf{C}_k, \zeta_k) = 0$ for all $k \in \{1, \dots, m\}$ and $u \in \{1, \dots, h\}$. In other words, $\{(\mathsf{B}_k, \mathsf{C}_k, \zeta_k)\}_{k=1}^m \subseteq \mathcal{V}$, where $\mathcal{V} \subseteq \mathsf{M}_{n \times r}(\mathbb{R}) \times \mathsf{M}_{r \times n}(\mathbb{R}) \times \mathbb{R}^h$ is the variety

$$\mathcal{V} \stackrel{\mathrm{def}}{=} \bigcap_{u=1}^h \Big\{ (X, Y, z) \in \mathsf{M}_{n \times r}(\mathbb{R}) \times \mathsf{M}_{r \times n}(\mathbb{R}) \times \mathbb{R}^h; \ \mathfrak{p}_u(X, Y, z) = 0 \Big\}. \tag{56}$$

We claim that $\mathcal{V}$ has at least m connected components. In fact, if $k, \ell \in \{1, \dots, m\}$ are distinct, then $(\mathsf{B}_k, \mathsf{C}_k, \zeta_k)$ and $(\mathsf{B}_\ell, \mathsf{C}_\ell, \zeta_\ell)$ belong to different connected component of $\mathcal{V}$. Indeed, suppose for the sake of obtaining a contradiction that $\mathcal{C} \subseteq \mathcal{V}$ is a connected subset of $\mathcal{V}$ and $(\mathsf{B}_k, \mathsf{C}_k, \zeta_k), (\mathsf{B}_\ell, \mathsf{C}_\ell, \zeta_\ell) \in \mathcal{C}$. Since $k \neq \ell$, by switching the roles of k and ℓ if necessary, the assumption of Lemma 44 ensures that there exists $(i, j) \in \mathsf{E}$ such that $(\mathsf{B}_k \mathsf{C}_k)_{ij} = a_{ij}^k < 0 < a_{ij}^\ell = (\mathsf{B}_\ell \mathsf{C}_\ell)_{ij}$. So, if we denote $\psi : \mathcal{C} \to \mathbb{R}$ by $\psi(X, Y, z) = (XY)_{ij}$, then $\psi(\mathsf{B}_k, \mathsf{C}_k, \zeta_k) < 0 < \psi(\mathsf{B}_\ell, \mathsf{C}_\ell, \zeta_\ell)$. Since $\mathcal{C}$ is connected and ψ is continuous, it follows that $\psi(X, Y, z) = 0$ for some $(X, Y, z) \in \mathcal{C}$. Let $u \in \{1, \dots, h\}$ be the index for which $(i, j) \in \mathsf{J}_u$. By the definition (55) of $\mathfrak{p}_u$, the fact that $\psi(X, Y, z) = (XY)_{ij} = 0$ implies that $\mathfrak{p}_u(X, Y, z) = -z_u^2 - \frac{1}{2}\mu^{|\mathsf{J}_u|} \leqslant -\frac{1}{2}\mu^{|\mathsf{J}_u|} < 0$, since $\mu > 0$. Hence $(X, Y, z) \notin \mathcal{V}$, in contradiction to our choice of (X, Y, z) as an element of $\mathcal{C} \subseteq \mathcal{V}$.

Recalling (55), for all $u \in \{1, \dots, h\}$ the degree of $\mathfrak{p}_u$ is $4|\mathsf{J}_u| \leqslant 4\alpha$. So, the variety $\mathcal{V}$ in (56) is defined using h polynomials of degree at most 4α in $2nr + h$ variables. By (a special case of) a theorem of Milnor [**?**] and Thom [**?**], the number of connected components of $\mathcal{V}$ is at most $4\alpha(8\alpha - 1)^{2nr+h-1} = 4\alpha(8\alpha - 1)^{2nr + \lceil |\mathsf{E}|/\alpha \rceil - 1}$. Since we already established that this number of connected components is at least m, it follows that

$$m \leqslant 4\alpha(8\alpha - 1)^{2nr + \left\lceil \frac{|\mathsf{E}|}{\alpha} \right\rceil - 1} \iff r \geqslant \frac{1}{2n}\left(\frac{\log\left(\frac{m}{4\alpha}\right)}{\log(8\alpha - 1)} - \left\lceil \frac{|\mathsf{E}|}{\alpha} \right\rceil + 1 \right). \tag{57}$$

The value of $\alpha \in \mathbb{N}$ that maximizes the rightmost quantity in (57) satisfies

$$\alpha \asymp \frac{|\mathsf{E}|}{\log(2m)}\left(\log\left(\frac{|\mathsf{E}|}{\log(2m)} \right) \right)^2.$$

For this α the estimate (57) simplifies to imply the desired bound $r \gtrsim \log m / (n \log(|\mathsf{E}| / \log m))$. $\qquad \square$

Proof of Lemma 43. For notational convenience, write $\mathsf{L} = \{\lambda_1, \dots, \lambda_n\}$ and $\mathsf{R} = \{\rho_1, \dots, \rho_n\}$, and think of E as a subset of $\{1, \dots, n\}^2$ (i.e., $(i, j) \in \mathsf{E}$ if and only if $\{\lambda_i, \rho_j\}$ is an edge of G).

For every $\Delta \in \mathbb{N}$ denote

$$\mathcal{B}_\Delta \stackrel{\mathrm{def}}{=} \Big\{ \sigma : \mathsf{E} \to \{-, +\} : \ \dim_{\omega, \Omega}\left(\mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}, d_\sigma^{s, T} \right) \leqslant \Delta \Big\}. \tag{58}$$

Then, by the definition of $\dim_{\omega, \Omega}(\cdot)$, if $\sigma \in \mathcal{B}_\Delta$, we can fix a normed space $(X_\sigma, \|\cdot\|_{X_\sigma})$ with $\dim(X_\sigma) = \Delta$ and a mapping $f_\sigma : \mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R} \to X_\sigma$ that satisfies

$$\forall x, y \in \mathsf{L}^+ \cup \mathsf{L}^- \cup \mathsf{R}, \qquad \omega\big(d_\sigma^{s,T}(x, y)\big) \leqslant \|f_\sigma(x) - f_\sigma(y)\|_{X_\sigma} \leqslant \Omega\big(d_\sigma^{s,T}(x, y)\big). \tag{59}$$

Using the Hahn–Banach theorem, for each $i \in \{1, \dots, n\}$ and $\sigma \in \mathcal{B}_\Delta$ we can fix a linear functional $z_{\sigma, i}^* \in X_\sigma^*$ of unit norm that normalizes the vector $f_\sigma(\lambda_i^+) - f_\sigma(\lambda_i^-) \in X_\sigma$, i.e.,

$$z_{\sigma, i}^*\big(f_\sigma(\lambda_i^+) - f_\sigma(\lambda_i^-)\big) = \big\|f_\sigma(\lambda_i^+) - f_\sigma(\lambda_i^+)\big\|_{X_\sigma} \qquad \text{and} \qquad \|z_{\sigma, i}^*\|_{X_\sigma^*} = \sup_{w \in X_\sigma \smallsetminus \{0\}} \frac{|z_{\sigma, i}^*(w)|}{\|w\|_{X_\sigma}} = 1. \tag{60}$$

Using these linear functionals, define an $n \times n$ matrix $\mathsf{A}_\sigma = (a_{ij}^\sigma) \in \mathsf{M}_n(\mathbb{R})$ by setting

$$\forall (i, j) \in \{1, \dots, n\}^2, \qquad a_{ij}^\sigma \stackrel{\mathrm{def}}{=} z_{\sigma, i}^*\big(f_\sigma(\rho_j)\big) - \frac{1}{2} z_{\sigma, i}^*\big(f_\sigma(\lambda_i^+) + f_\sigma(\lambda_i^-)\big). \tag{61}$$

Observe in passing that the following identity holds true for every $i, j \in \{1, \dots, n\}$ and $\sigma \in \mathcal{B}_\Delta$.

$$\sigma(i, j)a_{ij}^\sigma = \frac{1}{2} z_{\sigma, i}^*\big(f_\sigma(\lambda_i^+) - f_\sigma(\lambda_i^-)\big) + \sigma(i, j)z_{\sigma, i}^*\Big(f_\sigma(\rho_j) - f_\sigma\Big(\lambda_i^{\sigma(i,j)}\Big) \Big). \tag{62}$$

(Simply verify (62) for the cases $\sigma(i, j) = +$ and $\sigma(i, j) = -$ separately, using the linearity of $z_{\sigma, i}^*$.)

Since we are assuming in Lemma 43 that the shortest cycle in the template graph G has length at least g, it follows from Claim 42 that $d_{\mathsf{G}_\sigma}(\lambda_i^+, \lambda_i^-) \geqslant g$ for all $i \in \{1, \ldots, n\}$ and $\sigma : \mathsf{E} \to \{-, +\}$. So,

$$d_\sigma^{s,T}(\lambda_i^+, \lambda_i^-) \overset{(46)}{=} \min\left\{ sd_{\mathsf{G}_\sigma}(\lambda_i^+, \lambda_i^-), T \right\} \geqslant \min\{sg, T\} \overset{(47)}{=} sg. \tag{63}$$

Recalling (45), we have $\{\lambda_i^{\sigma(i,j)}, \rho_j\} \in \mathsf{E}_\sigma$ for all $(i,j) \in \mathsf{E}$. Hence $d_{\mathsf{G}_\sigma}(\lambda_i^{\sigma(i,j)}, \rho_j) = 1$ and therefore

$$d_\sigma^{s,T}\left(\lambda_i^{\sigma(i,j)}, \rho_j\right) \overset{(46)}{\leqslant} sd_{\mathsf{G}_\sigma}\left(\lambda_i^{\sigma(i,j)}, \rho_j\right) = s. \tag{64}$$

Consequently, for every $(i,j) \in \mathsf{E}$ and $\sigma \in \mathcal{B}_\Delta$ we have

$$
\begin{aligned}
\sigma(i,j)a_{ij}^\sigma \quad &\overset{(62)}{\geqslant} \quad \frac{1}{2}z_{\sigma,i}^*\left(f_\sigma(\lambda_i^+) - f_\sigma(\lambda_i^-)\right) - \left\|z_{\sigma,i}^*\right\|_{X_\sigma^*} \cdot \left\|f_\sigma(\rho_j) - f_\sigma\left(\lambda_i^{\sigma(i,j)}\right)\right\|_{X_\sigma} \\
&\overset{(60)}{=} \quad \frac{1}{2}\left\|f_\sigma(\lambda_i^+) - f_\sigma(\lambda_i^-)\right\|_{X_\sigma} - \left\|f_\sigma(\rho_j) - f_\sigma\left(\lambda_i^{\sigma(i,j)}\right)\right\|_{X_\sigma} \\
&\overset{(59)}{\geqslant} \quad \frac{1}{2}\omega\left(d_\sigma^{s,T}(\lambda_i^+, \lambda_i^-)\right) - \Omega\left(d_\sigma^{s,T}\left(\lambda_i^{\sigma(i,j)}, \rho_j\right)\right) \\
&\overset{(63)\wedge(64)}{\geqslant} \quad \frac{1}{2}\omega(sg) - \Omega(s) \overset{(47)}{>} 0.
\end{aligned}
$$

Hence, $a_{ij}^\sigma \neq 0$ and $\mathbf{sign}(a_{ij}^\sigma) = \sigma(i,j)$ for all $(i,j) \in \mathsf{E}$ and $\sigma \in \mathcal{B}_\Delta$. This is precisely the setting of Lemma 44 (with $m = |\mathcal{B}_\Delta|$), from which we conclude that there exists $\tau \in \mathcal{B}_\Delta$ such that

$$\mathbf{rank}(\mathsf{A}_\tau) \geqslant \frac{c \log |\mathcal{B}_\Delta|}{n \log\left(\frac{|\mathsf{E}|}{\log |\mathcal{B}_\Delta|}\right)}, \tag{65}$$

where $c \in (0, \infty)$ is a universal constant. Henceforth, we shall fix a specific $\tau \in \mathcal{B}_\Delta$ as in (65).

Since $\tau \in \mathcal{B}_\Delta$ we have $\dim(X_\tau) = \Delta$, so we can fix a basis $e_\tau^1, \ldots, e_\tau^\Delta$ of X_τ and for every $j \in \{1, \ldots, n\}$ write $f_\tau(\rho_j) = \gamma_{\tau,j}^1 e_\tau^1 + \ldots + \gamma_{\tau,j}^\Delta e_\tau^\Delta$ for some scalars $\gamma_{\tau,j}^1, \ldots, \gamma_{\tau,j}^\Delta \in \mathbb{R}$. Hence,

$$(a_{ij}^\tau)_{i=1}^n \overset{(61)}{=} \gamma_{\tau,j}^1\left(z_{\tau,i}^*(e_\tau^1)\right)_{i=1}^n + \ldots + \gamma_{\tau,j}^\Delta\left(z_{\tau,i}^*(e_\tau^\Delta)\right)_{i=1}^n - \frac{1}{2}\left(z_{\sigma,i}^*\left(f_\tau(\lambda_i^+) + f_\tau(\lambda_i^-)\right)\right)_{i=1}^n.$$

We have thus expressed the columns of the matrix A_τ as elements of the span of the $\Delta + 1$ vectors

$$\left(z_{\tau,i}^*(e_\tau^1)\right)_{i=1}^n, \left(z_{\tau,i}^*(e_\tau^2)\right)_{i=1}^n, \ldots, \left(z_{\tau,i}^*(e_\tau^\Delta)\right)_{i=1}^n, \left(z_{\sigma,i}^*\left(f_\tau(\lambda_i^+) + f_\tau(\lambda_i^-)\right)\right)_{i=1}^n \in \mathbb{R}^n.$$

Consequently, the rank of A_τ is at most $\Delta + 1 \leqslant 2\Delta$. By contrasting this with (65), we see that

$$\frac{|\mathsf{E}|}{\log |\mathcal{B}_\Delta|} \log\left(\frac{|\mathsf{E}|}{\log |\mathcal{B}_\Delta|}\right) \geqslant \frac{c|\mathsf{E}|}{2\Delta n}. \tag{66}$$

We shall now conclude by showing that Lemma 43 holds true with $\eta = c/2$. Indeed, fix $\delta \in (0, \frac{1}{3}]$ and observe that we may assume also that $\delta\eta|\mathsf{E}|/n \geqslant 1$, since otherwise the left hand side of (48) vanishes. Then, by choosing $\Delta = \lfloor \delta\eta|\mathsf{E}|/n \rfloor \in \mathbb{N}$ in the above reasoning it follows from (66) that

$$\frac{|\mathsf{E}|}{\log |\mathcal{B}_\Delta|} \log\left(\frac{|\mathsf{E}|}{\log |\mathcal{B}_\Delta|}\right) \geqslant \frac{c}{2\delta\eta} = \frac{1}{\delta} \geqslant 3 > e.$$

This implies that $|\mathcal{B}_\Delta| \leqslant \delta^{-\delta|\mathsf{E}|}$. Equivalently, $\mathbb{P}[\mathcal{B}_\Delta] \leqslant (2\delta)^{-\delta|\mathsf{E}|}$, which is the desired bound (48). $\qquad\square$

5. Nonlinear spectral gaps and impossibility of average dimension reduction

Fix $n \in \mathbb{N}$ and an irreducible reversible row-stochastic matrix $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$. This implies that there is a unique[8] A-stationary probability measure $\pi = (\pi_1, \ldots, \pi_n) \in [0,1]^n$ on $\{1, \ldots, n\}$, namely $\pi\mathsf{A} = \pi$, and we have the reversibility condition $\pi_i a_{ij} = \pi_j a_{ji}$ for all $i, j \in \{1, \ldots, n\}$. Then A is a self-adjoint contraction on $L_2(\pi)$, and we denote by $1 = \lambda_1(\mathsf{A}) \geqslant \lambda_2(\mathsf{A}) \geqslant \ldots \geqslant \lambda_n(\mathsf{A}) \geqslant -1$ the decreasing rearrangement of its eigenvalues.

---

[8]We are assuming irreducibility only for notational convenience, namely so that π will be unique and could therefore be suppressed in the ensuing notation. Our arguments work for any stochastic matrix and any probability measure π on $\{1, \ldots, n\}$ with respect to which A is reversible. We suggest focusing initially on the case when A is symmetric and π is the uniform measure on $\{1, \ldots, n\}$, though the general case is useful for treating graphs that are not regular, e.g. those of Section 4. See [**?**] for the relevant background.

Given a metric space $(\mathcal{M}, d_{\mathcal{M}})$ and $p > 0$, define $\gamma(\mathsf{A}, d_{\mathcal{M}}^p)$ to be the infimum over those $\gamma > 0$ such that

$$\forall\, x_1, \ldots, x_n \in \mathcal{M}, \qquad \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i \pi_j d_{\mathcal{M}}(x_i, x_j)^p \leqslant \gamma \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i a_{ij} d_{\mathcal{M}}(x_i, x_j)^p. \tag{67}$$

This definition is implicit in [?], and appeared explicitly in [?]; see [?, ?, ?] for a detailed treatment. It suffices to note here that if $(H, \|\cdot\|_H)$ is a Hilbert space and $p = 2$, then by expanding the squares one directly sees that $\gamma(\mathsf{A}, \|\cdot\|_H^2) = 1/(1 - \lambda_2(\mathsf{A}))$ is the reciprocal of the *spectral gap* of A. In general, we think of $\gamma(\mathsf{A}, d_{\mathcal{M}}^p)$ as measuring the magnitude of the nonlinear spectral gap of A with respect to the kernel $d_{\mathcal{M}}^p : \mathcal{M} \times \mathcal{M} \to [0, \infty)$.

Using the notation that was recalled in Section 1.1, the definition (67) immediately implies that nonlinear spectral gaps are bi-Lipschitz invariants in the sense that $\gamma(\mathsf{A}, d_{\mathcal{M}}^p) \leqslant \mathsf{c}_n(\mathcal{M})^p \gamma(\mathsf{A}, d_{\mathcal{N}}^p)$ for every two metric spaces $(\mathcal{M}, d_{\mathcal{M}})$ and $(\mathcal{N}, d_{\mathcal{N}})$, every matrix A as above and every $p > 0$. In particular, if $(H, \|\cdot\|_H)$ is a Hilbert space into which $(\mathcal{M}, d_{\mathcal{M}})$ admits a bi-Lipschitz embedding, then we have the following general (trivial) bound.

$$\sqrt{\gamma(\mathsf{A}, d_{\mathcal{M}}^2)} \leqslant \mathsf{c}_2(\mathcal{M}) \sqrt{\gamma(\mathsf{A}, \|\cdot\|_H^2)}. \tag{68}$$

In the recent work [?] we proved the following theorem, which improves over (68) when $\mathcal{M}$ is a Banach space.

Theorem 45. *Suppose that $(X, \|\cdot\|_X)$ is a Banach space and that $(H, \|\cdot\|_H)$ is a Hilbert space. Then for every $M \in (0, \infty)$ and every matrix A as above for which $\lambda_2(\mathsf{A}) \leqslant 1 - M^2/\mathsf{c}_2(X)^2$ we have*

$$\sqrt{\gamma(\mathsf{A}, \|\cdot\|_X^2)} \lesssim \frac{\log(M+1)}{M} \mathsf{c}_2(X) \sqrt{\gamma(\mathsf{A}, \|\cdot\|_H^2)}. \tag{69}$$

In the setting of Theorem 45, since $\gamma(\mathsf{A}, \|\cdot\|_H^2) = 1/(1 - \lambda_2(\mathsf{A}))$, the bound (69) can be rewritten as

$$\gamma(\mathsf{A}, \|\cdot\|_X^2) \lesssim \left(\frac{\log(\mathsf{c}_2(X)\sqrt{1 - \lambda_2(\mathsf{A})} + 1)}{1 - \lambda_2(\mathsf{A})} \right)^2,$$

which is how Theorem 45 was stated in [?]. Note that (69) coincides (up to the implicit constant factor) with the trivial bound (68) if $M = O(1)$, but (69) is an asymptotic improvement over (68) as $M \to \infty$.

The proof of Theorem 45 in [?] is a short interpolation argument that takes as input a theorem from [?]. While we do not know of a different proof of (69), below we will present a new and self-contained derivation of the following weaker estimate (using the same notation as in Theorem 45) that suffices for deducing Theorem 19.

$$\sqrt{\gamma(\mathsf{A}, \|\cdot\|_X^2)} \lesssim \frac{\log(\mathsf{c}_2(X) + 1)}{M} \mathsf{c}_2(X) \sqrt{\gamma(\mathsf{A}, \|\cdot\|_H^2)}. \tag{70}$$

The proof of (70) appears in Section 5.1 below. We will next show how (70) implies Theorem 19 as well as the leftmost inequality in (12), which includes as a special case the triple logarithmic estimate in (22).

Given $n \in \mathbb{N}$ and an n-vertex connected graph $\mathsf{G} = (\{1, \ldots, n\}, \mathsf{E_G})$, let $\mathsf{A_G}$ be its random walk matrix, i.e., if $\deg_{\mathsf{G}}(i)$ is the degree in G of the vertex $i \in \{1, \ldots, n\}$, then $(\mathsf{A_G})_{ij} = \mathbf{1}_{\{i,j\} \in \mathsf{E_G}} / \deg_{\mathsf{G}}(i)$ for all $i, j \in \{1, \ldots, n\}$. For $i \in \{1, \ldots, n\}$ we also denote $\pi_i^{\mathsf{G}} = \deg_{\mathsf{G}}(i)/(2|\mathsf{E_G}|)$. Thus, $\pi^{\mathsf{G}} \in \mathbb{R}^n$ is the probability measure on $\{1, \ldots, n\}$ with respect to which $\mathsf{A_G}$ is reversible. We will use the simpler notation $\lambda_i(\mathsf{A_G}) = \lambda_i(\mathsf{G})$ for every $i \in \{1, \ldots, n\}$. For $p \in (0, \infty)$ and a metric space $(\mathcal{M}, d_{\mathcal{M}})$, we will write $\gamma(\mathsf{G}, d_{\mathcal{M}}^p) = \gamma(\mathsf{A_G}, d_{\mathcal{M}}^p)$. The shortest-path metric that is induced by G on $\{1, \ldots, n\}$ will be denoted $d_{\mathsf{G}} : \{1, \ldots, n\} \times \{1, \ldots, n\} \to \mathbb{N} \cup \{0\}$.

Theorem 46. *There is a universal constant $K > 1$ with the following property. Fix $n \in \mathbb{N}$ and $\alpha \geqslant 1$. Let $\mathsf{A} \in \mathsf{M}_n(\mathbb{R})$ and $\pi \in [0, 1]^n$ be as above. For every normed space $(X, \|\cdot\|_X)$, if $f : \{1, \ldots, n\} \to X$ satisfies*

$$\left(\sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i a_{ij} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \leqslant \alpha, \tag{71}$$

then necessarily

$$\dim(X) \gtrsim K^{\frac{1 - \lambda_2(\mathsf{A})}{\alpha}} \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i \pi_j \|f(i) - f(j)\|_X^2}. \tag{72}$$

In particular, in the special case when $\mathsf{G} = (\{1, \ldots, n\}, \mathsf{E_G})$ is a connected graph we have

$$\left(\frac{1}{|\mathsf{E_G}|} \sum_{\{i,j\} \in \mathsf{E_G}} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \leqslant \alpha \implies \dim(X) \gtrsim K^{\frac{1 - \lambda_2(\mathsf{G})}{\alpha}} \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i^{\mathsf{G}} \pi_j^{\mathsf{G}} \|f(i) - f(j)\|_X^2}.$$

In the case of regular graphs with a spectral gap, Theorem 46 has the following corollary.

Corollary 47. *Fix two integers* $n, r \geqslant 3$ *and let* $\mathsf{G} = (\{1, \ldots, n\}, \mathsf{E}_\mathsf{G})$ *be a connected* r-*regular graph. If* $(X, \|\cdot\|_X)$ *is a normed space into which there is a mapping* $f : \{1, \ldots, n\} \to X$ *that satisfies*

$$\left(\frac{1}{|\mathsf{E}_\mathsf{G}|} \sum_{\{i,j\} \in \mathsf{E}_\mathsf{G}} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \leqslant \alpha \qquad \text{and} \qquad \left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \geqslant \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_\mathsf{G}(i,j), \quad (73)$$

then necessarily

$$\dim(X) \gtrsim n^{\frac{c(1-\lambda_2(\mathsf{G}))}{\alpha \log r}},$$

where $c \in (0, \infty)$ *is a universal constant.*

Proof. This is nothing more than a special case of Theorem 46 once we note that by a straightforward and standard counting argument (see e.g. [**?**]) we have $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} d_\mathsf{G}(i,j) \gtrsim \log_r n$. □

For every integer $r \geqslant 3$ there exist arbitrarily large r-regular graphs G with $\lambda_2(\mathsf{G}) = 1 - \Omega(1)$; see [**?**] for this and much more on such *spectral expanders*. Corollary 47 shows that the shortest-path metric on any such graph with $r = O(1)$ satisfies the conclusion of Theorem 19, because the α-Lipschitz assumption of Theorem 19 implies the first inequality in (73) and the assumption $\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|f(i) - f(j)\|_X \geqslant \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_\mathsf{G}(i,j)$ of Theorem 19 implies the second inequality in (73) (using Jensen's inequality).

Note that we actually proved above that any expander is "metrically high dimensional" in a stronger sense. Specifically, if $\mathsf{G} = (\{1, \ldots, n\}, \mathsf{E}_\mathsf{G})$ is a $O(1)$-spectral expander, i.e., it is $O(1)$-regular and $\lambda_2(\mathsf{G}) \leqslant 1 - \Omega(1)$, and one finds vectors $x_1, \ldots, x_n$ in a normed space $(X, \|\cdot\|_X)$ for which the averages $\frac{1}{|\mathsf{E}_\mathsf{G}|} \sum_{\{i,j\} \in \mathsf{E}_\mathsf{G}} \|x_i - x_j\|_X^2$ and $\frac{1}{n^2} \sum_{i,j=1}^{n} \|x_i - x_j\|_X^2$ are within a $O(1)$ factor of the averages $\frac{1}{|\mathsf{E}_\mathsf{G}|} \sum_{\{i,j\} \in \mathsf{E}_\mathsf{G}} d_\mathsf{G}(i,j)^2 = 1$ and $\frac{1}{n^2} \sum_{i,j=1}^{n} d_\mathsf{G}(i,j)^2$, respectively, then this "finitary average distance information" (up to a fixed but potentially very large multiplicative error) forces the ambient space X to be very high (worst-possible) dimensional, namely $\dim(X) \geqslant n^{\Omega(1)}$.

Remark 48. If one replaces (73) by the requirement that for an increasing modulus $\omega : [0, \infty) \to [0, \infty)$ we have,

$$\left(\frac{1}{|\mathsf{E}_\mathsf{G}|} \sum_{\{i,j\} \in \mathsf{E}_\mathsf{G}} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \leqslant 1 \qquad \text{and} \qquad \left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \geqslant \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega\big(d_\mathsf{G}(i,j)\big), \quad (74)$$

then the above argument applies mutatis mutandis to yield the conclusion

$$\dim(X) \gtrsim e^{c(1-\lambda_2(\mathsf{G}))\omega(c \log_r n)}. \quad (75)$$

Indeed, the aforementioned counting argument shows that least 50% of the pairs $(i,j) \in \{1, \ldots, n\}^2$ satisfy $d_\mathsf{G}(i,j) \gtrsim \log_r n$. Compare (75) to Theorem 17 which provides a stronger bound if the average requirement (74) is replaced by its pairwise counterpart (18). Nevertheless, the bound (75) is quite sharp (at least when $r = O(1)$ and $\lambda_2(\mathsf{G}) = 1 - \Omega(1)$), in the sense that there is a normed space $(X, \|\cdot\|_X)$ for which (74) holds and

$$\dim(X) \lesssim e^{C(\log r)\omega\left(\frac{C \log n}{\sqrt{1-\lambda_2(\mathsf{G})}} \right)} \log n, \quad (76)$$

where $C > 0$ is a universal constant. Indeed, by [**?**] the diameter of the metric space $(\{1, \ldots, n\}, d_\mathsf{G})$ satisfies $\mathbf{diam}(\mathsf{G}) \lesssim (\log n)/\sqrt{1 - \lambda_2(\mathsf{G})}$. By an application of (12) with $\alpha \asymp (\log_r n)/\omega(\mathbf{diam}(\mathsf{G}))$ there exists a normed space $(X, \|\cdot\|_X)$ with $\dim(X) \lesssim n^{O(1/\alpha)} \log n$, thus (76) holds, and a mapping $f : \{1, \ldots, n\} \to X$ that satisfies $d_\mathsf{G}(i,j)/\alpha \leqslant \|f(i) - f(j)\|_X \leqslant d_\mathsf{G}(i,j)$ for all $i, j \in \{1, \ldots, n\}$. Hence, the first inequality in (74) holds, and

$$\left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \geqslant \frac{1}{\alpha} \left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_\mathsf{G}(i,j)^2 \right)^{\frac{1}{2}} \gtrsim \frac{\log_r n}{\alpha} \asymp \omega\big(\mathbf{diam}(\mathsf{G}) \big) \geqslant \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega\big(d_\mathsf{G}(i,j)\big).$$

31

Proof of Theorem 46 assuming (70). Let $C \in (0, \infty)$ be the implicit universal constant in (70). Then

$$\left(\sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i \pi_j \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \overset{(67)}{\leqslant} \sqrt{\gamma(\mathsf{A}, \|\cdot\|_X^2)} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i a_{ij} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}}$$

$$\overset{(71)}{\leqslant} \alpha \sqrt{\gamma(\mathsf{A}, \|\cdot\|_X^2)} \overset{(70)}{\leqslant} \frac{C\alpha \log(\mathsf{c}_2(X) + 1)}{1 - \lambda_2(\mathsf{A})}, \quad (77)$$

where the last step of (77) is an application of (70) with $M = \mathsf{c}_2(X)\sqrt{1 - \lambda_2(\mathsf{A})}$, while using that for a Hilbert space H we have $\gamma(\mathsf{A}, \|\cdot\|_H^2) = 1/(1 - \lambda_2(\mathsf{A}))$. It follows that

$$2\sqrt{\dim(X)} \geqslant 2\mathsf{c}_2(X) \geqslant \mathsf{c}_2(X) + 1 \overset{(77)}{\geqslant} e^{\frac{1 - \lambda_2(\mathsf{A})}{C\alpha} \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i \pi_j \|f(i) - f(j)\|_X^2}}, \quad (78)$$

where the first step of (78) uses John's theorem [?]. This establishes (72) with $K = e^{2/C} > 1$. $\qquad\square$

For non-contracting embeddings (in particular, for bi-Lipschitz embedding), the proof of the following lemma is an adaptation of the proof of [?, Theorem 13].

Lemma 49. *Fix two integers $n, r \geqslant 3$ and let $\mathsf{G} = (\{1, \ldots, n\}, \mathsf{E}_\mathsf{G})$ be a connected r-regular graph. If $(X, \|\cdot\|_X)$ is a normed space into which there is a mapping $f : \{1, \ldots, n\} \to X$ that satisfies*

$$\min_{\substack{i,j \in \{1,\ldots,n\} \\ i \neq j}} \frac{\|f(i) - f(j)\|_X}{d_\mathsf{G}(i,j)} \geqslant 1, \qquad \text{and} \qquad \left(\frac{1}{|\mathsf{E}_\mathsf{G}|} \sum_{\{i,j\} \in \mathsf{E}_\mathsf{G}} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \leqslant \alpha. \quad (79)$$

Then necessarily

$$\frac{\log n}{\log r} n^{\frac{1}{2\dim(X)}} \lesssim \alpha \sqrt{\gamma(\mathsf{G}, \|\cdot\|_X^2)}. \quad (80)$$

Prior to proving Lemma 49, we will derive some of its corollaries. For the terminology of Corollary 50 below, recall that a Banach space Y is said to be B-convex [?] if ℓ_1 is *not* finitely representable in Y; see the survey [?] for more on this important notion, including useful analytic, geometric and probabilistic characterizations.

Corollary 50. *There is a universal constant $C \in (0, \infty)$ for which the following assertion holds true. Let Y be an infinite dimensional B-convex Banach space. For arbitrarily large $n \in \mathbb{N}$, if $\alpha \geqslant C \log n$, then we have*

$$\mathsf{k}_n^\alpha(\ell_\infty, Y) \asymp_Y \frac{\log n}{\log\left(\frac{\alpha}{\log n}\right)}. \quad (81)$$

Thus, we have in particular $\mathsf{k}_n^{C \log n}(\ell_\infty, Y) \asymp_Y \log n$.

Proof. The upper bound $\mathsf{k}_n^\alpha(\ell_\infty, Y) \lesssim (\log n)/\log(\alpha/\log n)$ actually holds for *any* infinite dimensional Banach space Y. Indeed, by Bourgain's embedding theorem [?] any n-point metric space $\mathcal{M}$ admits an embedding f into ℓ_2 with distortion $A \log n$, where $A \in (0, \infty)$ is a universal constant. If $\alpha \geqslant 4A \log n$, then by applying the JL-Lemma [?] we know that $f(\mathcal{M})$ embeds with distortion $\alpha/(2A \log n)$ into ℓ_2^k, where $k \lesssim (\log n)/\log(\alpha/\log n)$. By Dvoretzky's theorem [?], we know that ℓ_2^k embeds with distortion 2 into Y, so overall we obtain an embedding of $\mathcal{M}$ into a k-dimensional subspace of Y with distortion at most $2(A \log n)(\alpha/(2A \log n)) = \alpha$.

Conversely, suppose that $\alpha \geqslant 2 \log n$ and that Y is a B-convex Banach space. By a theorem of V. Lafforgue [?] (see also [?] for a different approach), for arbitrarily large $n \in \mathbb{N}$ there is a $O(1)$-regular graph $\mathsf{G} = (\{1, \ldots, n\}, \mathsf{E}_\mathsf{G})$ such that $\gamma(\mathsf{G}, \|\cdot\|_Y^2) \lesssim_Y 1$. If $(\{1, \ldots, n\}, d_\mathsf{G})$ embeds with distortion α into a k-dimensional subspace of Y, then by Lemma 49 we have $n^{1/(2k)} \lesssim_Y \alpha/\log n$. Thus $k \gtrsim_Y (\log n)/\log(\alpha/\log n)$, as required. $\qquad\square$

Question 51. Is the assumption of B-convexity needed for the conclusion (81) of Corollary 50? Perhaps finite cotype suffices for this purpose? This matter is of course closely related to Question 22.

Corollary 52. *Under the assumptions and notation of Lemma 49, we have*

$$\frac{\log n}{\log r} n^{\frac{1}{2\dim(X)}} \lesssim \frac{\alpha \log(\mathsf{c}_2(X) + 1)}{1 - \lambda_2(\mathsf{G})}.$$

Proof. This is nothing more than a substitution of (70) with $M = \mathsf{c}_2(X)\sqrt{1 - \lambda_2(\mathsf{G})}$ into (80). $\qquad\square$

Since by John's theorem [?] we have $c_2(X) \leqslant \sqrt{\dim(X)}$ and for every $n \in \mathbb{N}$ there exists a graph G as in Lemma 49 with $r = O(1)$ and $\lambda_2(\mathsf{G}) = 1 - \Omega(1)$, it follows from Corollary 52 that

$$\alpha \log(\mathsf{k}_n^\alpha(\ell_\infty) + 1) \gtrsim n^{\frac{1}{2\mathsf{k}_n^\alpha(\ell_\infty)}} \log n.$$

This implies the lower bound on $\mathsf{k}_n^\alpha(\ell_\infty)$ in (12). In particular, for $\alpha \asymp \log n$ it gives the first inequality in (22).

Proof of Lemma 49. Denote $\gamma = \gamma(\mathsf{G}, \|\cdot\|_X^2)$. For $i \in \{1, \ldots, n\}$ write

$$\mathcal{U}_i = f^{-1}\Big(B_X\big(f(i), \alpha\sqrt{2\gamma}\big)\Big) = \Big\{ j \in \{1, \ldots, n\}: \ \|f(i) - f(j)\|_X \leqslant \alpha\sqrt{2\gamma} \Big\}. \tag{82}$$

Let $m \in \{1, \ldots, n\}$ satisfy $|\mathcal{U}_m| = \max_{i \in \{1, \ldots, n\}} |\mathcal{U}_i|$. Then

$$n^2 \gamma \alpha^2 \overset{(67) \wedge (79)}{\geqslant} \sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_X^2 \geqslant \sum_{i=1}^n \sum_{j \in \{1, \ldots, n\} \smallsetminus \mathcal{U}_i} \|f(i) - f(j)\|_X^2$$

$$\overset{(82)}{>} \sum_{i=1}^n \big(\alpha\sqrt{2\gamma}\big)^2 (n - |\mathcal{U}_i|) \geqslant 2n\gamma\alpha^2(n - |\mathcal{U}_m|). \tag{83}$$

This simplifies to $|\mathcal{U}_m| \geqslant \frac{1}{2}n$. Also, since $\frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_X^2 \geqslant \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n d_\mathsf{G}(i,j)^2 \gtrsim (\log_r n)^2$, the first inequality in (83) implies the a priori lower bound $\alpha\sqrt{\gamma} \gtrsim \log_r n$.

Next, fix $\rho \in (0, \infty)$ and let $\mathcal{N}_{2\rho} \subseteq \mathcal{U}_m$ be a maximal (with respect to inclusion) 2ρ-separated subset of $\mathcal{U}_m$. Then $\mathcal{U}_m \subseteq \cup_{i \in \mathcal{N}_\Delta} B_\mathsf{G}(i, 2\rho)$, where $B_\mathsf{G}(i, 2\rho)$ denotes the ball centered at i of radius 2ρ in the shortest-path metric d_G. Since G is r-regular, for each $i \in \{1, \ldots, n\}$ we have the (crude) bound $|B_\mathsf{G}(i, 2\rho)| \leqslant 2r^{2\rho}$. Hence, $\frac{1}{2}n \leqslant |\mathcal{U}_m| \leqslant 2r^{2\rho}|\mathcal{N}_{2\rho}|$. So, if we choose $\rho = \frac{1}{4}\log_r n$, then $|\mathcal{N}_{2\rho}| \gtrsim \sqrt{n}$. Since by (79) distinct $i, j \in \mathcal{N}_{2\rho}$ satisfy $\|f(i) - f(j)\|_X \geqslant d_\mathsf{G}(i,j) \geqslant 2\rho$, the X-balls $\{B_X(f(i), \rho) : \ i \in \mathcal{N}_{2\rho}\}$ have pairwise disjoint interiors. At the same time, since each $i \in \mathcal{N}_{2\rho}$ belongs to $\mathcal{U}_m$, we have $\|f(i) - f(m)\|_X \leqslant \alpha\sqrt{2\gamma}$ (by the definition of $\mathcal{U}_m$), and hence $B_X(f(i), \rho) \subseteq B_X(f(m), \alpha\sqrt{2\gamma} + \rho)$. So, writing $\dim(X) = k$, we have the following volume comparison.

$$(\alpha\sqrt{2\gamma} + \rho)^k \mathbf{vol}_k\big(B_X(0,1)\big) = \mathbf{vol}_k\big(B_X(f(m), \alpha\sqrt{2\gamma} + \rho)\big) \geqslant \mathbf{vol}_k\bigg(\bigcup_{i \in \mathcal{N}_{2\rho}} B_X(f(i), \rho) \bigg)$$

$$= \sum_{i \in \mathcal{N}_{2\rho}} \mathbf{vol}_k\big(B_X(f(i), \rho)\big) = \rho^k \mathbf{vol}_k\big(B_X(0,1)\big)|\mathcal{N}_{2\rho}| \gtrsim \rho^k \mathbf{vol}_k\big(B_X(0,1)\big)\sqrt{n}.$$

This simplifies to give $n^{\frac{1}{2k}} \lesssim \frac{\alpha\sqrt{2\gamma}}{\rho} + 1 \asymp \frac{\alpha\sqrt{\gamma}}{\log_r n}$, where we used the definition of ρ, and that $\alpha\sqrt{\gamma} \gtrsim \log_r n$. $\quad\square$

5.1. **Nonlinear Rayleigh quotient inequalities.**

Our goal in this section is to present a proof of (70). As we stated earlier, the proof that appears below is different from the proof of Theorem 45 in [?]. However, the reason that underlies its validity is the same as that of the original argument in [?]. Specifically, we arrived at the ensuing proof because we were driven by an algorithmic need that arose in [?]. This need required proving a point-wise strengthening of an upper bound on nonlinear spectral gaps, which is called in [?] a "nonlinear Rayleigh quotient inequality." We will clarify what we mean by this later; a detailed discussion appears in [?].

The need to make the interpolation-based proof in [?] constructive/algorithmic led us to merge the argument in [?] with the *proof of* a theorem from [?], rather than quoting and using the latter as a "black box" as we did in [?]. In doing so, we realized that for the purpose of obtaining only the weaker bound (70) one could more efficiently combine [?] and [?] so as to skip the use of complex interpolation and to obtain the estimate (70) as well as its nonlinear Rayleigh quotient counterpart. Thus, despite superficial differences, the argument below amounts to unravelling the proofs in [?, ?] and removing steps that are needed elsewhere but not for (70). At present, we do not have a proof of the stronger inequality (69) that differs from its proof in [?], and the interpolation-based approach of [?] is used for more refined algorithmic results in the forthcoming work [?].

We will continue using the notation/conventions that were set at the beginning of Section 5. Fix $p \geqslant 1$ and a metric space $(\mathcal{M}, d_\mathcal{M})$. Let $L_p(\pi; \mathcal{M})$ be the metric space $(\mathcal{M}^n, d_{L_p(\pi;\mathcal{M})})$, where $d_{L_p(\pi;\mathcal{M})}: \mathcal{M}^n \times \mathcal{M}^n \to [0, \infty)$ is

$$\forall\, x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathcal{M}^n, \qquad d_{L_p(\pi;\mathcal{M})}(x,y) \overset{\text{def}}{=} \bigg(\sum_{i=1}^n \pi_i d_\mathcal{M}(x_i, y_i)^p \bigg)^{\frac{1}{p}}.$$

Throughout what follows, it will be notationally convenient to slightly abuse notation by considering $\mathcal{M}$ as a subset of $L_p(\pi; \mathcal{M})$ through its identification with the *diagonal* subset of $\mathcal{M}^n$, which is an isometric copy of $\mathcal{M}$ in $L_p(\pi; \mathcal{M})$. Namely, we identify each $x \in \mathcal{M}$ with the n-tuple $(x, x \ldots, x) \in \mathcal{M}^n$.

If $x = (x_1, \ldots, x_n) \in L_p(\pi; \mathcal{M}) \smallsetminus \mathcal{M}$, then the corresponding *nonlinear Rayleigh quotient* is defined to be

$$\mathcal{R}(x; \mathsf{A}, d_{\mathcal{M}}^p) \overset{\text{def}}{=} \frac{\sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_{\mathcal{M}}(x_i, x_j)^p}{\sum_{i=1}^n \sum_{j=1}^n \pi_i \pi_j d_{\mathcal{M}}(x_i, x_j)^p}. \tag{84}$$

The restriction $x \notin \mathcal{M}$ was made here only to ensure that the denominator in (84) does not vanish. By definition,

$$\gamma(\mathsf{A}, d_{\mathcal{M}}^p) = \sup_{x \in L_p(\pi; \mathcal{M}) \smallsetminus \mathcal{M}} \frac{1}{\mathcal{R}(x; \mathsf{A}, d_{\mathcal{M}}^p)}. \tag{85}$$

Note that $L_p(\pi; X)$ is a Banach space for every Banach space $(X, \|\cdot\|_X)$. In this case, the matrix $\mathsf{A} \in \mathsf{M}_n(\mathbb{R})$ induces a linear operator $\mathsf{A} \otimes \mathsf{Id}_X : L_p(\pi; X) \to L_p(\pi; X)$ that is given by $(\mathsf{A} \otimes \mathsf{Id}_X)(x_1, \ldots x_n) = (\sum_{j=1}^n a_{ij} x_j)_{i=1}^n$.

The following lemma records some simple and elementary general properties of nonlinear Rayleigh quotients.

Lemma 53. *Suppose that $(\mathcal{M}, d_{\mathcal{M}})$ is a metric space, $n \in \mathbb{N}$, $p \in [1, \infty)$ and $\delta \in [0, 1]$. Let $\pi = (\pi_1, \ldots, \pi_n)$ be a probability measure on $\{1, \ldots, n\}$ and $\mathsf{A}, \mathsf{B} \in \mathsf{M}_n(\mathbb{R})$ be row-stochastic matrices that are reversible with respect to π. For any $x \in L_p^n(\pi; \mathcal{M}) \smallsetminus \mathcal{M}$ we have*

(1) $\mathcal{R}(x; \delta \mathsf{A} + (1 - \delta)\mathsf{B}, d_{\mathcal{M}}^p) = \delta \mathcal{R}(x; \mathsf{A}, d_{\mathcal{M}}^p) + (1 - \delta)\mathcal{R}(x; \mathsf{B}, d_{\mathcal{M}}^p)$
(2) $\mathcal{R}(x; (1 - \delta)\mathsf{Id}_n + \delta \mathsf{A}, d_{\mathcal{M}}^p) = \delta \mathcal{R}(x; \mathsf{A}, d_{\mathcal{M}}^p)$, where $\mathsf{Id}_n \in \mathsf{M}_n(\mathbb{R})$ is the identity matrix.
(3) $\mathcal{R}(x; \mathsf{A}, d_{\mathcal{M}}^p) \leqslant 2^p$.
(4) $\mathcal{R}(x; \mathsf{AB}, d_{\mathcal{M}}^p)^{\frac{1}{p}} \leqslant \mathcal{R}(x; \mathsf{A}, d_{\mathcal{M}}^p)^{\frac{1}{p}} + \mathcal{R}(x; \mathsf{B}, d_{\mathcal{M}}^p)^{\frac{1}{p}}$.
(5) $\mathcal{R}(x; \mathsf{A}^{\mathsf{t}}, d_{\mathcal{M}}^p) \leqslant \mathsf{t}^p \mathcal{R}(x; \mathsf{A}, d_{\mathcal{M}}^p)$ for every $\mathsf{t} \in \mathbb{N}$.

Proof. The first assertion is an immediate consequence of the definition of nonlinear Rayleigh quotients. The second assertion is a special case of the first assertion, since by definition $\mathcal{R}(x; \mathsf{Id}_n, d_{\mathcal{M}}^p) = 0$. The third assertion is justified by noting that by the triangle inequality, for every $i, j, k \in \{1, \ldots, n\}$ we have

$$d_{\mathcal{M}}(x_i, x_j)^p \leqslant \left(d_{\mathcal{M}}(x_i, x_k) + d_{\mathcal{M}}(x_k, x_j)\right)^p \leqslant 2^{p-1} d_{\mathcal{M}}(x_i, x_k)^p + 2^{p-1} d_{\mathcal{M}}(x_k, x_j)^p. \tag{86}$$

where the last step of (86) uses the convexity of the function $(t \in [0, \infty)) \mapsto t^p$. By multiplying (86) by $\pi_i a_{ij}/n^2$, summing over $i, j, k \in \{1, \ldots, n\}$ and using the fact that A is reversible with respect to π, we get

$$\sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} d_{\mathcal{M}}(x_i, x_j)^p \leqslant 2^p \sum_{i=1}^n \sum_{j=1}^n \pi_i \pi_j d_{\mathcal{M}}(x_i, x_j)^p.$$

Recalling the notation (84), this is precisely the third assertion of Lemma 53.

It remains to justify the fourth assertion of Lemma 53, because its fifth assertion follows from iterating its fourth assertion $\mathsf{t} - 1$ times (with B a power of A). To this end, writing $\mathsf{A} = (a_{ij})$ and $\mathsf{B} = (b_{ij})$, we have

$$\left(\sum_{i=1}^n \sum_{j=1}^n \pi_i (\mathsf{AB})_{ij} d_{\mathcal{M}}(x_i, x_j)^p\right)^{\frac{1}{p}} \leqslant \left(\sum_{i=1}^n \sum_{j=1}^n \pi_i \left(\sum_{k=1}^n a_{ik} b_{kj}\right)\left(d_{\mathcal{M}}(x_i, x_k) + d_{\mathcal{M}}(x_k, x_j)\right)^p\right)^{\frac{1}{p}}$$

$$\leqslant \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \pi_i a_{ik} b_{kj} d_{\mathcal{M}}(x_i, x_k)^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \pi_i a_{ik} b_{kj} d_{\mathcal{M}}(x_k, x_j)^p\right)^{\frac{1}{p}}$$

$$= \left(\sum_{i=1}^n \sum_{k=1}^n \pi_i a_{ik} d_{\mathcal{M}}(x_i, x_k)^p\right)^{\frac{1}{p}} + \left(\sum_{j=1}^n \sum_{k=1}^n \pi_k b_{kj} d_{\mathcal{M}}(x_k, x_j)^p\right)^{\frac{1}{p}}, \tag{87}$$

where the first step of (87) uses the triangle inequality in $\mathcal{M}$, the second step of (87) uses the triangle inequality in $L_p(\mu)$ with μ being the measure on $\{1, \ldots, n\}^3$ given by $\mu(i, j, k) = \pi_i a_{ik} b_{kj}$ for all $i, j, k \in \{1, \ldots, n\}$, and the final step of (87) uses the fact that A and B are both row-stochastic and reversible with respect to π. $\square$

The identity in the following claim is a consequence of a very simple and standard Hilbertian computation that we record here for ease of later references.

Claim 54. *For every Hilbert space $(H, \| \cdot \|_H)$ and every $x \in L_2(\pi; H) \smallsetminus H$ we have $\mathcal{R}(x; \mathsf{A}^2, \| \cdot \|_H^2) \leqslant 1$. Moreover, if $\sum_{i=1}^n \pi_i x_i = 0$, then*

$$\frac{\|(\mathsf{A} \otimes \mathsf{Id}_H) x\|_{L_2(\pi; H)}}{\|x\|_{L_2(\pi; H)}} = \sqrt{1 - \mathcal{R}(x; \mathsf{A}^2, \| \cdot \|_H^2)}.$$

Proof. Let $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{R}$ be the scalar product that induces the Hilbertian norm $\| \cdot \|_H$. Then, the scalar product that induces the norm $\| \cdot \|_{L_2(\pi; H)}$ is given by $\langle y, z \rangle_{L_2(\pi; H)} = \sum_{i=1}^n \pi_i \langle y_i, z_i \rangle$. By expanding the squares while using the fact that A is row-stochastic, reversible relative to π, and $\sum_{i=1}^n \pi_i x_i = 0$, we get that

$$\sum_{i=1}^n \sum_{j=1}^n \pi_i (\mathsf{A}^2)_{ij} \|x_i - x_j\|_H^2 = 2\|x\|_{L_2(\pi; H)}^2 - 2\sum_{i=1}^n \pi_i \Big\langle x_i, \sum_{j=1}^n (\mathsf{A}^2)_{ij} x_j \Big\rangle$$

$$= 2\|x\|_{L_2(\pi; H)}^2 - 2\big\langle x, (\mathsf{A}^2 \otimes \mathsf{Id}_H) x \big\rangle_{L_2(\pi; H)} = 2\|x\|_{L_2(\pi; H)}^2 - 2\big\|(\mathsf{A} \otimes \mathsf{Id}_H) x\big\|_{L_2(\pi; H)}^2,$$

and

$$\sum_{i=1}^n \sum_{j=1}^n \pi_i \pi_j \|x_i - x_j\|_H^2 = 2\sum_{i=1}^n \pi_i \|x_i\|_H^2 - 2\Big\| \sum_{i=1}^n \pi_i x_i \Big\|_H^2 = 2\|x\|_{L_2(\pi; H)}^2.$$

Therefore, recalling the definition (84), we have

$$\mathcal{R}(x; \mathsf{A}^2, \| \cdot \|_H^2) = 1 - \frac{\big\|(\mathsf{A} \otimes \mathsf{Id}_H) x\big\|_{L_2(\pi; H)}^2}{\|x\|_{L_2(\pi; H)}^2} \leqslant 1. \qquad \square$$

Lemma 55 (point-wise Rayleigh quotient estimate for Hilbert isomorphs). *Let $(X, \| \cdot \|_X)$ be a normed space and fix $\mathsf{d} \in [1, \infty)$. Suppose that $\| \cdot \|_H : X \to [0, \infty)$ is a Hilbertian norm on X that satisfies*

$$\forall y \in X, \qquad \|y\|_H \leqslant \|y\|_X \leqslant \mathsf{d} \|y\|_H. \tag{88}$$

For every $x \in L_2(\pi; X) \smallsetminus X$ define a quantity $\mathsf{t}(x, \mathsf{A}) = \mathsf{t}(x; \mathsf{A}, \| \cdot \|_H, \mathsf{d})$ to be the minimum $\mathsf{t} \in \mathbb{N}$ such that

$$\mathcal{R}\bigg(x; \Big(\frac{1}{2} \mathsf{Id}_n + \frac{1}{2} \mathsf{A}\Big)^{2\mathsf{t}}, \| \cdot \|_H^2 \bigg) \geqslant 1 - \frac{1}{4\mathsf{d}^2}, \tag{89}$$

with the convention that $\mathsf{t}(x; \mathsf{A}) = \infty$ if no such t exists. Then,

$$\frac{1}{\mathcal{R}(x; \mathsf{A}, \| \cdot \|_X^2)} \lesssim \mathsf{t}(x; \mathsf{A})^2. \tag{90}$$

Proof. We may assume without loss of generality that $\sum_{i=1}^n \pi_i x_i = 0$ and $\mathsf{t}(x; \mathsf{A}) < \infty$. Define a matrix

$$\mathsf{B}_x \overset{\text{def}}{=} \Big(\frac{1}{2} \mathsf{Id}_n + \frac{1}{2} \mathsf{A}\Big)^{\mathsf{t}(x; \mathsf{A})} \in \mathsf{M}_n(\mathbb{R}). \tag{91}$$

Then B_x is also a row-stochastic matrix which is reversible with respect to π, and, by the definition of $\mathsf{t}(x; \mathsf{A})$,

$$\mathcal{R}(x; \mathsf{B}_x^2, \| \cdot \|_H^2) \geqslant 1 - \frac{1}{4\mathsf{d}^2}.$$

By Claim 54, since $\sum_{i=1}^n \pi_i x_i = 0$, this implies that

$$\frac{\|(\mathsf{B}_x \otimes \mathsf{Id}_H) x\|_{L_2(\pi; H)}}{\|x\|_{L_2(\pi; H)}} \leqslant \sqrt{1 - \Big(1 - \frac{1}{4\mathsf{d}^2}\Big)} = \frac{1}{2\mathsf{d}}. \tag{92}$$

At the same time, due to (88) we have

$$\frac{\|(\mathsf{B}_x \otimes \mathsf{Id}_X) x\|_{L_2(\pi; X)}}{\|x\|_{L_2(\pi; X)}} \leqslant \mathsf{d} \frac{\|(\mathsf{B}_x \otimes \mathsf{Id}_H) x\|_{L_2(\pi; H)}}{\|x\|_{L_2(\pi; H)}}. \tag{93}$$

By combining (92) and (93) we see that $\|(\mathsf{B}_x \otimes \mathsf{Id}_H) x\|_{L_2(\pi; X)} \leqslant \frac{1}{2} \|x\|_{L_2(\pi; X)}$. Consequently,

$$\|x - (\mathsf{B}_x \otimes \mathsf{Id}_X) x\|_{L_2(\pi; X)} \geqslant \|x\|_{L_2(\pi; X)} - \|(\mathsf{B}_x \otimes \mathsf{Id}_X) x\|_{L_2(\pi; X)} \geqslant \frac{1}{2} \|x\|_{L_2(\pi; X)}. \tag{94}$$

35

Observe that

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\pi_j\|x_i-x_j\|_X^2\right)^{\frac12} \leqslant \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\pi_j\big(\|x_i\|_X+\|x_j\|_X\big)^2\right)^{\frac12} \leqslant 2\|x\|_{L_2(\pi;X)}, \tag{95}$$

where in the first step of (95) we used the triangle inequality in X and the second step of (95) is an application of the triangle inequality in $L_2(\pi\otimes\pi)$. Also, since B_x is row-stochastic,

$$\|x-(\mathsf{B}_x\otimes\mathsf{Id}_X)\,x\|_{L_2(\pi;X)} = \left(\sum_{i=1}^{n}\pi_i\left\|\sum_{j=1}^{n}(\mathsf{B}_x)_{ij}(x_i-x_j)\right\|_X^2\right)^{\frac12} \leqslant \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i(\mathsf{B}_x)_{ij}\|x_i-x_j\|_X^2\right)^{\frac12}, \tag{96}$$

where in the final step of (96) we used the convexity of the function $\|\cdot\|_X^2 : X \to \mathbb{R}$.

Recalling the definition (84), by substituting (95) and (96) into (94) we see that

$$\mathcal{R}\left(x;\left(\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A}\right)^{\mathsf{t}(x;\mathsf{A})},\|\cdot\|_X^2\right) \overset{(91)}{=} \mathcal{R}\big(x;\mathsf{B}_x,\|\cdot\|_X^2\big) \geqslant \frac{1}{16}. \tag{97}$$

We now conclude the proof of the desired estimate (90) as follows.

$$1 \overset{(97)}{\lesssim} \mathcal{R}\left(x;\left(\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A}\right)^{\mathsf{t}(x;\mathsf{A})},\|\cdot\|_X^2\right) \leqslant \mathsf{t}(x;\mathsf{A})^2\mathcal{R}\left(x;\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A},\|\cdot\|_X^2\right) = \frac12\mathsf{t}(x;\mathsf{A})^2\mathcal{R}\big(x;\mathsf{A},\|\cdot\|_X^2\big), \tag{98}$$

where the second step of (98) uses the fifth assertion of Lemma 53, and the final step uses its second assertion. $\square$

The quantity $\mathsf{t}(x;\mathsf{A})$ of Lemma 55 can be bounded as follows in terms of the spectral gap of A.

Lemma 56. *Continuing with the notation of Lemma 55, the following estimate holds true.*

$$\mathsf{t}(x;\mathsf{A}) \leqslant \left\lceil\frac{\log(2\mathsf{d})}{\log\left(\frac{2}{1+\lambda_2(\mathsf{A})}\right)}\right\rceil \lesssim \frac{\log(2\mathsf{d})}{1-\lambda_2(\mathsf{A})}. \tag{99}$$

Proof. Since A is row-stochastic, $\lambda_n(\mathsf{A}) \geqslant -1$. Therefore $\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A}$ is a positive semidefinite self-adjoint operator on $L_2(\pi)$ that preserves the hyperplane $L_2^0(\pi) = \{u \in \mathbb{R}^n; \sum_{i=1}^{n}\pi_i u_i = 0\}$. The largest eigenvalue of $\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A}$ on $L_2^0(\pi)$ is $\tfrac12+\tfrac12\lambda_2(\mathsf{A})$, and therefore $\|(\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A})^{\mathsf{t}}u\|_{L_2(\pi)} \leqslant (\tfrac12+\tfrac12\lambda_2(\mathsf{A}))^{\mathsf{t}}\|u\|_{L_2(\pi)}$ for $u \in L_2^0(\pi)$ and $\mathsf{t} \in \mathbb{N}$.

If $x \in L_2(\pi;X)\smallsetminus X$ satisfies $\sum_{i=1}^{n}\pi_i x_i = 0$, then we may apply the above observation to the coordinates of x with respect to some orthonormal basis of H, each of which is an element of $L_2^0(\pi)$, and deduce that

$$\left(\frac12+\frac12\lambda_2(\mathsf{A})\right)^{\mathsf{t}} \geqslant \frac{\left\|\left(\left(\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A}\right)^{\mathsf{t}}\otimes\mathsf{Id}_H\right)x\right\|_{L_2(\pi;H)}}{\|x\|_{L_2(\pi;H)}} = \sqrt{1-\mathcal{R}\left(x;\left(\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A}\right)^{2\mathsf{t}},\|\cdot\|_H^2\right)}, \tag{100}$$

where in the second step of (100) we applied Lemma 54 with A replaced by $(\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A})^{\mathsf{t}}$. Hence

$$\mathcal{R}\left(x;\left(\frac12\mathsf{Id}_n+\frac12\mathsf{A}\right)^{2\mathsf{t}},\|\cdot\|_H^2\right) \geqslant 1 - \left(\frac12+\frac12\lambda_2(\mathsf{A})\right)^{2\mathsf{t}}.$$

Consequently, if $\mathsf{t} \geqslant (\log(2\mathsf{d}))/\log(2/(\lambda_2(\mathsf{A})+1))$, then $\mathcal{R}(x;(\tfrac12\mathsf{Id}_n+\tfrac12\mathsf{A})^{2\mathsf{t}},\|\cdot\|_H^2) \geqslant 1-\frac{1}{4\mathsf{d}^2}$. By the definition of $\mathsf{t}(x;\mathsf{A})$, this implies the first inequality in (99). The second inequality in (99) follows by elementary calculus. $\square$

Proof of (70). By a classical linearization argument [**?**] (see [**?**, Chapter 7] for a modern treatment), for every $\mathsf{d} > \mathsf{c}_2(X)$ there is a Hilbertian norm $\|\cdot\|_H$ on X that satisfies (88). We therefore see that for every A as above

$$\sqrt{\gamma(\mathsf{A},\|\cdot\|_X^2)} \overset{(85)}{=} \sup_{x\in L_2(\pi;X)\smallsetminus X}\frac{1}{\sqrt{\mathcal{R}(x;\mathsf{A},\|\cdot\|_X^2)}} \overset{(90)}{\leqslant} \sup_{x\in L_2(\pi;X)\smallsetminus X}\mathsf{t}(x;\mathsf{A})$$

$$\overset{(99)}{\lesssim} \frac{\log(\mathsf{c}_2(X)+1)}{1-\lambda_2(\mathsf{A})} = \frac{\log(\mathsf{c}_2(X)+1)}{\sqrt{1-\lambda_2(\mathsf{A})}}\sqrt{\gamma(\mathsf{A},\|\cdot\|_H^2)} \leqslant \frac{\log(\mathsf{c}_2(X)+1)}{M}\mathsf{c}_2(X)\sqrt{\gamma(\mathsf{A},\|\cdot\|_H^2)}, \tag{101}$$

where, for the final step of (101) recall that in the context of (70) we assume that $\lambda_2(\mathsf{A}) \leqslant 1 - M^2/\mathsf{c}_2(X)^2$. $\square$

5.1.1. *Structural implications of nonlinear Rayleigh quotient inequalities.* Fix integers $n, k, r \geqslant 3$ (think of n as much larger than k). Let $(X, \|\cdot\|_X)$ be a k-dimensional normed space. Suppose that $\mathsf{G} = (\{1, \ldots, n\}, \mathsf{E_G})$ is a connected r-regular graph. Although we phrased (and used) Corollary 47 as an impossibility result that provides an obstruction (spectral gap) for faithfully realizing (on average) the metric space $(\{1, \ldots, n\}, d_{\mathsf{G}})$ in X, a key insight of the recent work [?] by Andoni, Nikolov, Razenshteyn, Waingarten and the author is that one could "flip" this point of view to deduce from Corollary 47 useful information on those graphs that do happen to admit such a faithful geometric realization in X, namely they satisfy (73). Clearly there are plenty of graphs with this property, including those graphs that arise from discrete approximations of subsets of X (as a "vanilla" example to keep in mind, fix a small parameter $\delta > 0$, consider a δ-net in the unit ball of X as the vertices, and join two net points by an edge if their distance in X is $O(\delta)$). The conclusion of Corollary 47 for any such graph is that it cannot have a large spectral gap, and by Cheeger's inequality [?, ?, ?, ?] it follows that this graph can be partitioned into two pieces with a small (relative) "discrete boundary." On the other hand, if we are given a mapping $f : \{1, \ldots, n\} \to X$ that satisfies the first condition in (73) but not the second condition in (73), then there must be a ball in X of relatively small radius that contains a substantial fraction of the vectors $\{f(i)\}_{i=1}^n$. The partition of $\{1, \ldots, n\}$ that corresponds to this dense ball and its complement encodes useful geometric "clustering" information. We have thus observed a dichotomic behavior that allows one to partition geometrically-induced graphs using either a "spectral partition" or a "dense ball partition."

In [?], the above idea is used iteratively to construct a hierarchical partition of X. Our overview suppresses important technical steps, which include both randomization and a re-weighting procedure of the graphs that arise at later stages of the construction (the start of the construction is indeed the above "net graph"); see [?] for the full details. In particular, one needs to use general row-stochastic matrices due to the re-weighting procedure, i.e., one uses the full strength of Theorem 46 rather than only the case of graphs as in Corollary 47.

In summary, one can use the bound (70) on nonlinear spectral gaps to provide a "cutting rule" that governs an iterative partitioning procedure in which each inductive step is either geometric (a ball and its complement) or a less explicit existential step that follows from spectral information which is deduced from a contrapositive assumption of (rough, average) embeddability. This structural information is used in [?] to design a new data structure for approximate nearest neighbor search in arbitrary norms (see the article of Andoni, Indyk and Razenshteyn in the present volume for an extensive account of approximate nearest neighbor search). Although this yields important (and arguably unexpected) progress on an algorithmic question of central importance, the non-explicitness and potential high complexity of the spectral partitioning step raises issues of efficiency that are not yet fully resolved. Specifically, the most general data structure that is designed in [?] is efficient only in the so-called "cell probe model," but not in the full polynomial-time sense; we refer to [?] for an explanation of these complexity-theoretic issues and their significance, because they are beyond the scope of the present article.

While the above issue of efficiency does not occur in our initial investigation within pure mathematics, it is very important from the algorithmic perspective. This is what initially led to the desire to obtain a nonlinear Rayleigh quotient inequality rather than to merely bound the nonlinear spectral gap, though (in hindsight) such inequalities are interesting from the mathematical perspective as well. We did not formally define what we mean by a "nonlinear Rayleigh quotient inequality" because there is some flexibility here, but the basic desire is, given $x_1, \ldots, x_n \in X$, to bound their Rayleigh quotient in X by a Rayleigh quotient of points in a *Euclidean space*.

The inequality (90) of Lemma 55 is of the above form, because the parameter $\mathsf{t}(x; \mathsf{A})$ defined by (89) involves only examining a certain Rayleigh quotient in a Hilbert space. It should be noted, however, that to date we have not succeeded to use the specific nonlinear Rayleigh quotient inequality of Lemma 55 for algorithmic purposes (though with more work this may be possible), despite the fact that it was found with this motivation in mind.

Other nonlinear Rayleigh quotient inequalities were obtained in [?, ?] and used to address issues of algorithmic efficiency. Very roughly, the drawback of (90) is that the matrix A is changed in the Hilbertian Rayleigh quotient of (89) (the main problem is the potentially high power $2\mathsf{t}$). A more directly algorithmically-useful nonlinear Rayleigh quotient inequality would be to change the point $x \in L_2(\pi; X)$ but not change the matrix A. Namely, suppose that we could control $\mathcal{R}(x; \mathsf{A}, \|\cdot\|_X^2)$ from below by a function of $\mathcal{R}(\varphi_{\mathsf{A}}(x); \mathsf{A}, \|\cdot\|_H^2)$, for some mapping

$\phi_{\mathsf{A}} : L_2(\pi; X) \to L_2(\pi; H)$. Nonlinear Rayleigh quotient inequalities of this type are proved in [**?**, **?**], though the associated mappings ϕ_{A} turn out to be highly nonlinear and quite complicated.[9]

The upshot of the latter type of nonlinear Rayleigh quotient inequality is that if (due to existence of a faithful embedding into X) we know that $\mathcal{R}(x; \mathsf{A}, \|\cdot\|_X^2)$ is small, then it follows that also $\mathcal{R}(\phi_{\mathsf{A}}(x); \mathsf{A}, \|\cdot\|_H^2)$ is small. The *proof of* Cheeger's inequality (via examination of level sets of the second eigenvector) would now provide a sparse "spectral partition" of $\{1, \ldots, n\}$ that has the following auxiliary structure: The partition is determined by thresholding one of the coordinates of H (in some fixed orthonormal basis), namely the part to which each $i \in \{1, \ldots, n\}$ belongs depends only on whether the coordinate in question of the transformed vector $\phi_{\mathsf{A}}(x)_i \in H$ is above or below a certain value. If in addition $(\mathsf{A}, x) \mapsto \phi_{\mathsf{A}}(x)$ has favorable computational properties (see [**?**] for a formulation; roughly, what is important here is that after a "preprocessing step" one can decide quickly to which piece of the partition each $i \in \{1, \ldots, n\}$ belongs), then this would lead to fast "query time."

The above description of the algorithmic role of nonlinear Rayleigh quotient inequalities is impressionistic, but it conveys the core ideas while not delving into (substantial) details. Such inequalities are interesting in their own right, partially because they necessitate making mathematical arguments constructive, thus leading to new proofs, as we did for (70), and also leading to intrinsically meaningful studies, such as obtaining [**?**] algorithmic versions of existential statements that arise from the use of the maximum principle in complex interpolation.

Acknowledgements. I am grateful to Daniel Kane for a conversation that led to a simplification of Proposition 35. I also thank Alexandros Eskenazis, Manor Mendel, Ilya Razenshteyn and Gideon Schechtman for their helpful suggestions.

REFERENCES

[1] I. Abraham, Y. Bartal, and O. Neiman. Advances in metric embedding theory. *Adv. Math.*, 228(6):3026–3126, 2011.

[2] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. System Sci.*, 66(4):671–687, 2003. Special issue on PODS 2001 (Santa Barbara, CA).

[3] N. Ailon and B. Chazelle. The fast Johnson-Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009.

[4] N. Ailon and E. Liberty. An almost optimal unrestricted fast Johnson-Lindenstrauss transform. *ACM Trans. Algorithms*, 9(3):Art. 21, 12, 2013.

[5] A. D. Aleksandrov. A theorem on triangles in a metric space and some of its applications. In *Trudy Mat. Inst. Steklov., v 38*, Trudy Mat. Inst. Steklov., v 38, pages 5–23. Izdat. Akad. Nauk SSSR, Moscow, 1951.

[6] N. Alon. The number of polytopes, configurations and real matroids. *Mathematika*, 33(1):62–71, 1986.

[7] N. Alon. Problems and results in extremal combinatorics. I. *Discrete Math.*, 273(1-3):31–53, 2003. EuroComb'01 (Barcelona).

[8] N. Alon. Perturbed identity matrices have high rank: proof and applications. *Combin. Probab. Comput.*, 18(1-2):3–15, 2009.

[9] N. Alon, H. Kaplan, G. Nivasch, M. Sharir, and S. Smorodinsky. Weak ϵ-nets and interval chains. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1194–1203. ACM, New York, 2008.

[10] N. Alon and B. Klartag. Optimal compression of approximate inner products and dimension reduction. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 639–650, 2017.

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

[12] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[13] A. Andoni, M. S. Charikar, O. Neiman, and H. L. Nguyen. Near linear lower bound for dimension reduction in ℓ_1. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS 2011*, pages 315–323. IEEE Computer Soc., Los Alamitos, CA, 2011.

[14] A. Andoni, A. Naor, and O. Neiman. Snowflake universality of Wasserstein spaces. To appear in *Ann. Sci. Éc. Norm. Supér. (4)*. Preprint available at https://arxiv.org/abs/1509.08677, 2015.

[15] A. Andoni, A. Naor, and O. Neiman. Impossibility of sketching of the 3D transportation metric with quadratic cost. In *43rd International Colloquium on Automata, Languages, and Programming*, volume 55 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 83, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.

[16] A. Andoni, A. Naor, and O. Neiman. On isomorphic dimension reduction in ℓ_1. Preprint, 2017.

[17] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Data-dependent hashing via nonlinear spectral gaps. To appear in *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, 2018.

---

[9]Specifically, in [**?**] such a mapping ϕ_{A} is constructed for Schatten-von Neumann trace classes using the Brouwer fixed-point theorem and estimates from [**?**]. In [**?**], ϕ_{A} is constructed for general normed spaces using, in addition to Brouwer's theorem, convex programming and (algorithmic variants of) complex interpolation. These lead to data structures that are efficient in all respects other than the "preprocessing stage," which at present remains potentially time-consuming due to the complexity of ϕ_{A}.

[18] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Hölder homeomorphisms and approximate nearest neighbors. Forthcoming manuscript, 2018.

[19] A. Andoni, H. L. Nguyen, A. Nikolov, I. Razenshteyn, and E. Waingarten. Approximate near neighbors for general symmetric norms. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 902–913. ACM, New York, 2017.

[20] J. Arias-de Reyna, K. Ball, and R. Villa. Concentration of the distance in finite-dimensional normed spaces. *Mathematika*, 45(2):245–252, 1998.

[21] J. Arias-de Reyna and L. Rodríguez-Piazza. Finite metric spaces needing high dimension for Lipschitz embeddings in Banach spaces. *Israel J. Math.*, 79(1):103–111, 1992.

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

[23] R. I. Arriaga and S. Vempala. An algorithmic theory of learning: robust concepts and random projection. In *40th Annual Symposium on Foundations of Computer Science (New York, 1999)*, pages 616–623. IEEE Computer Soc., Los Alamitos, CA, 1999.

[24] P. Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[25] T. Austin and A. Naor. On the bi-Lipschitz structure of Wasserstein spaces. Preprint, 2017.

[26] T. Austin, A. Naor, and R. Tessera. Sharp quantitative nonembeddability of the Heisenberg group into superreflexive Banach spaces. *Groups Geom. Dyn.*, 7(3):497–522, 2013.

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

[28] K. Ball. The Ribe programme. *Astérisque*, (352):Exp. No. 1047, viii, 147–159, 2013. Séminaire Bourbaki. Vol. 2011/2012. Exposés 1043–1058.

[29] S. Banach. *Théorie des opérations linéaires.* Éditions Jacques Gabay, Sceaux, 1993. Reprint of the 1932 original.

[30] Y. Bartal and L.-A. Gottlieb. Dimension reduction techniques for ℓ_p ($1 \leqslant p \leqslant 2$), with applications. In *32nd International Symposium on Computational Geometry*, volume 51 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. 16, 15. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.

[31] Y. Bartal, L.-A. Gottlieb, and O. Neiman. On the impossibility of dimension reduction for doubling subsets of ℓ_p. *SIAM J. Discrete Math.*, 29(3):1207–1222, 2015.

[32] Y. Bartal, N. Linial, M. Mendel, and A. Naor. Low dimensional embeddings of ultrametrics. *European J. Combin.*, 25(1):87–92, 2004.

[33] Y. Bartal and M. Mendel. Dimension reduction for ultrametrics. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 664–665. ACM, New York, 2004.

[34] Y. Bartal, B. Recht, and L. J. Schulman. Dimensionality reduction: beyond the Johnson-Lindenstrauss bound. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 868–887. SIAM, Philadelphia, PA, 2011.

[35] F. Barthe, O. Guédon, S. Mendelson, and A. Naor. A probabilistic approach to the geometry of the l_p^n-ball. *Ann. Probab.*, 33(2):480–513, 2005.

[36] J. Batson, D. A. Spielman, and N. Srivastava. Twice-Ramanujan sparsifiers. *SIAM J. Comput.*, 41(6):1704–1721, 2012.

[37] A. Beck. A convexity condition in Banach spaces and the strong law of large numbers. *Proc. Amer. Math. Soc.*, 13:329–334, 1962.

[38] 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.

[39] B. Bollobás. *Extremal graph theory*, volume 11 of *London Mathematical Society Monographs*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1978.

[40] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[41] M. Bonk and O. Schramm. Embeddings of Gromov hyperbolic spaces. *Geom. Funct. Anal.*, 10(2):266–306, 2000.

[42] G. Bouligand. Ensembles impropres et nombre dimensionnel. I, II. *Bull. Sci. Math., II. Sér.*, 52:320–344, 1928.

[43] M. Bourdon and H. Pajot. Poincaré inequalities and quasiconformal structure on the boundary of some hyperbolic buildings. *Proc. Amer. Math. Soc.*, 127(8):2315–2324, 1999.

[44] J. Bourgain. New Banach space properties of the disc algebra and H^∞. *Acta Math.*, 152(1-2):1–48, 1984.

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

[46] J. Bourgain. Subspaces of l_N^∞, arithmetical diameter and Sidon sets. In *Probability in Banach spaces, V (Medford, Mass., 1984)*, volume 1153 of *Lecture Notes in Math.*, pages 96–127. Springer, Berlin, 1985.

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

[48] 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.

[49] J. Bourgain, S. Dirksen, and J. Nelson. Toward a unified theory of sparse dimensionality reduction in Euclidean space. *Geom. Funct. Anal.*, 25(4):1009–1088, 2015.

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

[51] J. Bourgain and A. Gamburd. A spectral gap theorem in SU(d). *J. Eur. Math. Soc. (JEMS)*, 14(5):1455–1511, 2012.

[52] J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. *Acta Math.*, 162(1-2):73–141, 1989.

[53] J. Bourgain and V. Milman. Dichotomie du cotype pour les espaces invariants. *C. R. Acad. Sci. Paris Sér. I Math.*, 300(9):263–266, 1985.

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

[55] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki. Local random quantum circuits are approximate polynomial-designs. *Comm. Math. Phys.*, 346(2):397–434, 2016.

[56] P. Braß. On equilateral simplices in normed spaces. *Beiträge Algebra Geom.*, 40(2):303–307, 1999.

[57] J. Briët, A. Naor, and O. Regev. Locally decodable codes and the failure of cotype for projective tensor products. *Electron. Res. Announc. Math. Sci.*, 19:120–130, 2012.

[58] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1. *J. ACM*, 52(5):766–788 (electronic), 2005.

[59] 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.

[60] Y. Burago, M. Gromov, and G. Perelman. A. D. Aleksandrov spaces with curvatures bounded below. *Uspekhi Mat. Nauk*, 47(2(284)):3–51, 222, 1992.

[61] C. J. C. Burges. Dimension reduction: A guided tour. *Foundations and Trends in Machine Learning*, 2(4):275–365, 2010.

[62] T.-H. H. Chan, A. Gupta, and K. Talwar. Ultra-low-dimensional embeddings for doubling metrics. *J. ACM*, 57(4):Art. 21, 26, 2010.

[63] M. Charikar and A. Sahai. Dimension reduction in the \ell _1 norm. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*, pages 551–560, 2002.

[64] M. S. Charikar. Similarity estimation techniques from rounding algorithms. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 380–388. ACM, New York, 2002.

[65] S. Chechik. Approximate distance oracles with constant query time. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing*, pages 654–663. ACM, New York, 2014.

[66] 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.

[67] J. Cheeger. Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.*, 9(3):428–517, 1999.

[68] J. Cheeger and B. Kleiner. On the differentiability of Lipschitz maps from metric measure spaces to Banach spaces. In *Inspired by S. S. Chern*, volume 11 of *Nankai Tracts Math.*, pages 129–152. World Sci. Publ., Hackensack, NJ, 2006.

[69] J. Cheeger and B. Kleiner. Differentiability of Lipschitz maps from metric measure spaces to Banach spaces with the Radon-Nikodým property. *Geom. Funct. Anal.*, 19(4):1017–1028, 2009.

[70] 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.

[71] J. Cheeger, B. Kleiner, and A. Naor. Compression bounds for Lipschitz maps from the Heisenberg group to L_1. *Acta Math.*, 207(2):291–373, 2011.

[72] F. R. K. Chung, V. Faber, and T. A. Manteuffel. An upper bound on the diameter of a graph from eigenvalues associated with its Laplacian. *SIAM J. Discrete Math.*, 7(3):443–457, 1994.

[73] B. S. Cirel'son. It is impossible to imbed ℓ_p or c_0 into an arbitrary Banach space. *Funkcional. Anal. i Priložen.*, 8(2):57–60, 1974.

[74] R. R. Coifman and G. Weiss. *Analyse harmonique non-commutative sur certains espaces homogènes.* Lecture Notes in Mathematics, Vol. 242. Springer-Verlag, Berlin-New York, 1971. Étude de certaines intégrales singulières.

[75] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures Algorithms*, 22(1):60–65, 2003.

[76] G. David and S. Semmes. *Fractured fractals and broken dreams*, volume 7 of *Oxford Lecture Series in Mathematics and its Applications.* The Clarendon Press, Oxford University Press, New York, 1997. Self-similar geometry through metric and measure.

[77] G. David and M. Snipes. A non-probabilistic proof of the Assouad embedding theorem with bounds on the dimension. *Anal. Geom. Metr. Spaces*, 1:36–41, 2013.

[78] B. V. Dekster. Simplexes with prescribed edge lengths in Minkowski and Banach spaces. *Acta Math. Hungar.*, 86(4):343–358, 2000.

[79] J. Diestel, J. Fourie, and J. Swart. The projective tensor product. I. In *Trends in Banach spaces and operator theory (Memphis, TN, 2001)*, volume 321 of *Contemp. Math.*, pages 37–65. Amer. Math. Soc., Providence, RI, 2003.

[80] J. Diestel, J. H. Fourie, and J. Swart. *The metric theory of tensor products.* American Mathematical Society, Providence, RI, 2008. Grothendieck's résumé revisited.

[81] S. Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *Found. Comput. Math.*, 16(5):1367–1396, 2016.

[82] A. Dmitriyuk and Y. Gordon. Generalizing the Johnson-Lindenstrauss lemma to k-dimensional affine subspaces. *Studia Math.*, 195(3):227–241, 2009.

[83] A. Dmitriyuk and Y. Gordon. Randomized large distortion dimension reduction. *Positivity*, 18(4):767–784, 2014.

[84] R. Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, fourth edition, 2010.

[85] 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.

[86] K. Efremenko. 3-query locally decodable codes of subexponential length. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 39–44. ACM, New York, 2009.

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

[88] P. Enflo. Banach spaces which can be given an equivalent uniformly convex norm. *Israel J. Math.*, 13:281–288 (1973), 1972.

[89] P. Enflo. On infinite-dimensional topological groups. In *Séminaire sur la Géométrie des Espaces de Banach (1977–1978)*, pages Exp. No. 10–11, 11. École Polytech., Palaiseau, 1978.

[90] 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.

[91] A. Eskenazis, M. Mendel, and A. Naor. Diamond convexity: A bifurcation in the Ribe program. Preprint, 2017.

[92] A. Eskenazis, M. Mendel, and A. Naor. Nonpositive curvature is not coarsely universal. Preprint, 2017.

[93] T. Figiel and W. B. Johnson. A uniformly convex Banach space which contains no l_p. *Compositio Math.*, 29:179–190, 1974.

[94] P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44(3):355–362, 1988.

[95] M. Fréchet. Les dimensions d'un ensemble abstrait. *Math. Ann.*, 68(2):145–168, 1910.

[96] Z. Füredi and P. A. Loeb. On the best constant for the Besicovitch covering theorem. *Proc. Amer. Math. Soc.*, 121(4):1063–1073, 1994.

[97] A. Genz. Methods for generating random orthogonal matrices. In *Monte Carlo and quasi-Monte Carlo methods 1998 (Claremont, CA)*, pages 199–213. Springer, Berlin, 2000.

[98] 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.

[99] O. Giladi and A. Naor. Improved bounds in the scaled Enflo type inequality for Banach spaces. *Extracta Math.*, 25(2):151–164, 2010.

[100] O. Giladi, A. Naor, and G. Schechtman. Bourgain's discretization theorem. *Ann. Fac. Sci. Toulouse Math. (6)*, 21(4):817–837, 2012.

[101] G. Godefroy. De Grothendieck à Naor: une promenade dans l'analyse métrique des espaces de Banach. *Gaz. Math.*, (151):13–24, 2017.

[102] Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pages 84–106. Springer, Berlin, 1988.

[103] L.-A. Gottlieb and R. Krauthgamer. A nonlinear approach to dimension reduction. *Discrete Comput. Geom.*, 54(2):291–315, 2015.

[104] 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.

[105] M. Gromov. Random walk in random groups. *Geom. Funct. Anal.*, 13(1):73–146, 2003.

[106] A. Grothendieck. Résumé des résultats essentiels dans la théorie des produits tensoriels topologiques et des espaces nucléaires. *Ann. Inst. Fourier Grenoble*, 4:73–112 (1954), 1952.

[107] A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.

[108] 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.

[109] A. Grothendieck. Produits tensoriels topologiques et espaces nucléaires. *Mem. Amer. Math. Soc.*, No. 16:140, 1955.

[110] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, pages 534–543, 2003.

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

[112] A. Gupta and K. Talwar. Making doubling metrics geodesic. *Algorithmica*, 59(1):66–80, 2011.

[113] S. Har-Peled. *Geometric approximation algorithms*, volume 173 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

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

[115] J. Heinonen. *Geometric embeddings of metric spaces*, volume 90 of *Report. University of Jyväskylä Department of Mathematics and Statistics*. University of Jyväskylä, Jyväskylä, 2003.

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

[117] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

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

[119] T. Hytönen and A. Naor. Heat flow and quantitative differentiation. To appear in *J. Eur. Math. Soc. (JEMS)*, preprint available at `https://arxiv.org/abs/1608.01915`, 2016.

[120] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 10–33. IEEE Computer Soc., Los Alamitos, CA, 2001.

[121] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC '98 (Dallas, TX)*, pages 604–613. ACM, New York, 1999.

[122] P. Indyk and A. Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):Art. 31, 12, 2007.

[123] P. Indyk and N. Thaper. Fast image retrieval via embeddings. In *ICCV '03: Proceedings of the 3rd International Workshop on Statistical and Computational Theories of Vision*, 2003.

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

[125] 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.

[126] 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.

[127] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. On Lipschitz embedding of finite metric spaces in low-dimensional normed spaces. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 177–184. Springer, Berlin, 1987.

[128] W. B. Johnson and A. Naor. The Johnson-Lindenstrauss lemma almost characterizes Hilbert space, but not quite. *Discrete Comput. Geom.*, 43(3):542–553, 2010.

[129] W. B. Johnson and G. Schechtman. Finite dimensional subspaces of L_p. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 837–870. North-Holland, Amsterdam, 2001.

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

[131] J.-P. Kahane. Hélices et quasi-hélices. In *Mathematical analysis and applications, Part B*, volume 7 of *Adv. in Math. Suppl. Stud.*, pages 417–433. Academic Press, New York-London, 1981.

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

[133] D. M. Kane and J. Nelson. Sparser Johnson-Lindenstrauss transforms. *J. ACM*, 61(1):Art. 4, 23, 2014.

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

[135] S. Khot and A. Naor. Grothendieck-type inequalities in combinatorial optimization. *Comm. Pure Appl. Math.*, 65(7):992–1035, 2012.

[136] B. Klartag and S. Mendelson. Empirical processes and random projections. *J. Funct. Anal.*, 225(1):229–245, 2005.

[137] P. K. Kothari and R. Meka. Almost optimal pseudorandom generators for spherical caps [extended abstract]. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 247–256. ACM, New York, 2015.

[138] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[139] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: a new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005.

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

[141] 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.

[142] V. Lafforgue. Propriété (T) renforcée Banachique et transformation de Fourier rapide. *J. Topol. Anal.*, 1(3):191–206, 2009.

[143] V. Lafforgue and A. Naor. A doubling subset of L_p for $p > 2$ that is inherently infinite dimensional. *Geom. Dedicata*, 172:387–398, 2014.

[144] V. Lafforgue and A. Naor. Vertical versus horizontal Poincaré inequalities on the Heisenberg group. *Israel J. Math.*, 203(1):309–339, 2014.

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

[146] U. Lang and T. Schlichenmaier. Nagata dimension, quasisymmetric embeddings, and Lipschitz extensions. *Int. Math. Res. Not.*, (58):3625–3655, 2005.

[147] K. G. Larsen and J. Nelson. The Johnson-Lindenstrauss lemma is optimal for linear dimensionality reduction. In *43rd International Colloquium on Automata, Languages, and Programming*, volume 55 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 82, 11. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.

[148] K. G. Larsen and J. Nelson. Optimality of the Johnson–Lindenstrauss lemma. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 633–638, 2017.

[149] F. Lazebnik, V. A. Ustimenko, and A. J. Woldar. A new series of dense graphs of high girth. *Bull. Amer. Math. Soc. (N.S.)*, 32(1):73–79, 1995.

[150] J. R. Lee, A. de Mesmay, and M. Moharrami. Dimension reduction for finite trees in ℓ_1. *Discrete Comput. Geom.*, 50(4):977–1032, 2013.

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

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

[153] J. R. Lee and A. Naor. L_p metrics on the Heisenberg group and the Goemans-Linial conjecture. In *Proceedings of 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 99–108, 2006. Available at https://web.math.princeton.edu/~naor/homepage%20files/L_pHGL.pdf.

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

[155] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[156] J. Lindenstrauss and A. Pełczyński. Absolutely summing operators in L_p-spaces and their applications. *Studia Math.*, 29:275–326, 1968.

[157] J. Lindenstrauss and H. P. Rosenthal. The $\mathscr{L}_p$ spaces. *Israel J. Math.*, 7:325–349, 1969.

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

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

[160] 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.

[161] J. Matoušek. Note on bi-Lipschitz embeddings into normed spaces. *Comment. Math. Univ. Carolin.*, 33(1):51–55, 1992.

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

[163] J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.

[164] J. Matoušek. On embedding expanders into ℓ_p spaces. *Israel J. Math.*, 102:189–197, 1997.

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

[166] J. Matoušek. On variants of the Johnson-Lindenstrauss lemma. *Random Structures Algorithms*, 33(2):142–156, 2008.

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

[168] B. Maurey and G. Pisier. Caractérisation d'une classe d'espaces de Banach par des propriétés de séries aléatoires vectorielles. *C. R. Acad. Sci. Paris Sér. A-B*, 277:A687–A690, 1973.

[169] 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.

[170] S. Mazur. Une remarque sur l'homéomorphie des champs fonctionels. *Stud. Math.*, 1:83–85, 1929.

[171] M. Mendel. Metric dichotomies. In *Limits of graphs in group theory and computer science*, pages 59–76. EPFL Press, Lausanne, 2009.

[172] M. Mendel and A. Naor. Ramsey partitions and proximity data structures. *J. Eur. Math. Soc. (JEMS)*, 9(2):253–275, 2007.

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

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

[175] M. Mendel and A. Naor. A note on dichotomies for metric transforms. Available at http://arxiv.org/abs/1102.1800, 2011.

[176] M. Mendel and A. Naor. Markov convexity and local rigidity of distorted metrics. *J. Eur. Math. Soc. (JEMS)*, 15(1):287–337, 2013.

[177] M. Mendel and A. Naor. Nonlinear spectral calculus and super-expanders. *Publ. Math. Inst. Hautes Études Sci.*, 119:1–95, 2014.

[178] M. Mendel and A. Naor. Expanders with respect to Hadamard spaces and random graphs. *Duke Math. J.*, 164(8):1471–1548, 2015.

[179] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.*, 17(4):1248–1282, 2007.

[180] S. Mendelson and N. Tomczak-Jaegermann. A subgaussian embedding theorem. *Israel J. Math.*, 164:349–364, 2008.

[181] K. Menger. Dimensionstheorie. Leipzig: B. G. Teubner. iv, 318 S. (1928)., 1928.

[182] F. Mezzadri. How to generate random matrices from the classical compact groups. *Notices Amer. Math. Soc.*, 54(5):592–604, 2007.

[183] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[184] V. D. Milman and G. Schechtman. An "isomorphic" version of Dvoretzky's theorem. *C. R. Acad. Sci. Paris Sér. I Math.*, 321(5):541–544, 1995.

[185] J. Milnor. On the Betti numbers of real varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.

[186] A. Naor. An application of metric cotype to quasisymmetric embeddings. In *Metric and differential geometry*, volume 297 of *Progr. Math.*, pages 175–178. Birkhäuser/Springer, Basel, 2012.

[187] A. Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.

[188] A. Naor. Comparison of metric spectral gaps. *Anal. Geom. Metr. Spaces*, 2:1–52, 2014.

[189] A. Naor. Discrete Riesz transforms and sharp metric X_p inequalities. *Ann. of Math. (2)*, 184(3):991–1016, 2016.

[190] A. Naor. A spectral gap precludes low-dimensional embeddings. In *33rd International Symposium on Computational Geometry*, volume 77 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 50, 16. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.

[191] A. Naor and O. Neiman. Assouad's theorem with dimension independent of the snowflaking. *Rev. Mat. Iberoam.*, 28(4):1123–1142, 2012.

[192] 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.

[193] A. Naor, G. Pisier, and G. Schechtman. Impossibility of dimension reduction in the nuclear norm. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 1345–1352, 2018. Available at https://arxiv.org/abs/1710.08896.

[194] A. Naor and Y. Rabani. On Lipschitz extension from finite subsets. *Israel J. Math.*, 219(1):115–161, 2017.

[195] A. Naor and G. Schechtman. Planar earthmover is not in L_1. *SIAM J. Comput.*, 37(3):804–826, 2007.

[196] A. Naor and L. Silberman. Poincaré inequalities, embeddings, and wild groups. *Compos. Math.*, 147(5):1546–1572, 2011.

[197] A. Naor and R. Young. Foliated corona decompositions and the L_1 distortion of balls in the 3-dimensional Heisenberg group. Preprint, 2017.

[198] A. Naor and R. Young. The integrality gap of the Goemans-Linial SDP relaxation for sparsest cut is at least a constant multiple of $\sqrt{\log n}$. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 564–575. ACM, New York, 2017.

[199] A. Naor and R. Young. Vertical perimeter versus horizontal perimeter. To appear in Ann. of Math., available at `https://arxiv.org/abs/1701.00620`, 2017.

[200] J. Nash. C^1 isometric imbeddings. *Ann. of Math. (2)*, 60:383–396, 1954.

[201] O. Neiman. Low dimensional embeddings of doubling metrics. *Theory Comput. Syst.*, 58(1):133–152, 2016.

[202] J. Nelson. Chaining introduction with some computer science applications. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (120):42–65, 2016.

[203] I. Newman and Y. Rabinovich. On multiplicative λ-approximations and some geometric applications. *SIAM J. Comput.*, 42(3):855–883, 2013.

[204] G. Nöbeling. über eine n-dimensionale Universalmenge im R^{2n+1}. *Math. Ann.*, 104(1):71–80, 1931.

[205] P. W. Nowak and G. Yu. *Large scale geometry*. EMS Textbooks in Mathematics. European Mathematical Society (EMS), Zürich, 2012.

[206] M. Ostrovskii and B. Randrianantoanina. Metric spaces admitting low-distortion embeddings into all n-dimensional Banach spaces. *Canad. J. Math.*, 68(4):876–907, 2016.

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

[208] M. I. Ostrovskii. *Metric embeddings*, volume 49 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2013. Bilipschitz and coarse embeddings into Banach spaces.

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

[210] S. D. Pauls. The large scale geometry of nilpotent Lie groups. *Comm. Anal. Geom.*, 9(5):951–982, 2001.

[211] A. Peł czyński. *Banach spaces of analytic functions and absolutely summing operators*. American Mathematical Society, Providence, R.I., 1977. Expository lectures from the CBMS Regional Conference held at Kent State University, Kent, Ohio, July 11–16, 1976, Conference Board of the Mathematical Sciences Regional Conference Series in Mathematics, No. 30.

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

[213] G. Pisier. Un théorème sur les opérateurs linéaires entre espaces de Banach qui se factorisent par un espace de Hilbert. *Ann. Sci. École Norm. Sup. (4)*, 13(1):23–43, 1980.

[214] G. Pisier. Remarques sur un résultat non publié de B. Maurey. In *Seminar on Functional Analysis, 1980–1981*, pages Exp. No. V, 13. École Polytech., Palaiseau, 1981.

[215] G. Pisier. Counterexamples to a conjecture of Grothendieck. *Acta Math.*, 151(3-4):181–208, 1983.

[216] 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.

[217] G. Pisier. Factorization of operator valued analytic functions. *Adv. Math.*, 93(1):61–125, 1992.

[218] G. Pisier. Random series of trace class operators. In *Proceedings Cuarto CLAPEM Mexico 1990. Contribuciones en probabilidad y estadistica matematica*, pages 29–42, 1992. Available at `http://arxiv.org/abs/1103.2090`.

[219] G. Pisier. Grothendieck's theorem, past and present. *Bull. Amer. Math. Soc. (N.S.)*, 49(2):237–323, 2012.

[220] G. Puy, M. E. Davies, and R. Gribonval. Recipes for stable linear embeddings from Hilbert spaces to $\mathbb{R}^m$. *IEEE Trans. Inform. Theory*, 63(4):2171–1287, 2017.

[221] Y. Rabinovich. On average distortion of embedding metrics into the line. *Discrete Comput. Geom.*, 39(4):720–733, 2008.

[222] O. Regev. Entropy-based bounds on dimension reduction in L^1. *Israel J. Math.*, 195(2):825–832, 2013.

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

[224] E. Ricard. Hölder estimates for the noncommutative Mazur maps. *Arch. Math. (Basel)*, 104(1):37–45, 2015.

[225] J. Roe. *Lectures on coarse geometry*, volume 31 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2003.

[226] R. A. Ryan. *Introduction to tensor products of Banach spaces*. Springer Monographs in Mathematics. Springer-Verlag London, Ltd., London, 2002.

[227] G. Schechtman. More on embedding subspaces of L_p in l_r^n. *Compositio Math.*, 61(2):159–169, 1987.

[228] G. Schechtman. Asymptotic geometric analysis, Part I [book review of MR3331351]. *Bull. Amer. Math. Soc. (N.S.)*, 54(2):341–345, 2017.

[229] G. Schechtman and N. Tomczak-Jaegermann. Polylog dimensional subspaces of ℓ_∞^N. Preprint, 2018.

[230] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, expanded edition, 2014.

[231] I. J. Schoenberg. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, 44(3):522–536, 1938.

[232] S. Semmes. On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞-weights. *Rev. Mat. Iberoamericana*, 12(2):337–410, 1996.

[233] S. Semmes. Bilipschitz embeddings of metric spaces into Euclidean spaces. *Publ. Mat.*, 43(2):571–653, 1999.

[234] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.*, 82(1):93–133, 1989.

[235] A. J. Stam. Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces. *J. Appl. Probab.*, 19(1):221–228, 1982.

[236] G. W. Stewart. The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM J. Numer. Anal.*, 17(3):403–409 (loose microfiche suppl.), 1980.

[237] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002)*, volume 338 of *Contemp. Math.*, pages 357–390. Amer. Math. Soc., Providence, RI, 2003.

[238] M. Talagrand. Embedding subspaces of L_1 into l_1^N. *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.

[239] M. Talagrand. Approximating a helix in finitely many dimensions. *Ann. Inst. H. Poincaré Probab. Statist.*, 28(3):355–363, 1992.

[240] M. Talagrand. Embedding subspaces of L_p in l_p^N. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 311–325. Birkhäuser, Basel, 1995.

[241] M. Talagrand. *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Heidelberg, 2014. Modern methods and classical problems.

[242] R. M. Tanner. Explicit concentrators from generalized N-gons. *SIAM J. Algebraic Discrete Methods*, 5(3):287–293, 1984.

[243] R. Thom. Sur l'homologie des variétés algébriques réelles. In *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)*, pages 255–265. Princeton Univ. Press, Princeton, N.J., 1965.

[244] M. Thorup and U. Zwick. Approximate distance oracles. *J. ACM*, 52(1):1–24, 2005.

[245] N. Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of trace classes $S_p(1 \leqslant p < \infty)$. *Studia Math.*, 50:163–182, 1974.

[246] 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. Inst. Math., Warsaw, 1999.

[247] S. S. Vempala. *The random projection method*, volume 65 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, Providence, RI, 2004. With a foreword by Christos H. Papadimitriou.

[248] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[249] J. von Neumann. Some matrix-inequalities and metrization of matric-space. *Tomsk Univ. Rev.*, 1:286–300, 1937. Reprinted in Collected Works (Pergamon Press, 1962), iv, 205–219.

[250] H. Whitney. Differentiable manifolds. *Ann. of Math. (2)*, 37(3):645–680, 1936.

Mathematics Department, Princeton University, Princeton, New Jersey 08544-1000, USA.

E-mail address: naor@math.princeton.edu