

# 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  $\epsilon$ -approximation, 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 [?].

$$\begin{array}{ccccccc} o & & o & & o & & [Jo - L] \\ & oo & & o & o & & o \\ o & o & n & o & & o & o \\ & & o & & & & o \\ o & & o & & o & o & o \end{array}$$

C. Papadimitriou, 2004 (forward to  $O(\log n)$  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  $L_\infty$ -approximation of  $\mathbb{R}^n$  and the  $R_\infty$ -approximation. 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.

$R_\infty$  1 The broader term  $\epsilon$ -approximation 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  $\epsilon$ -approximation 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 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.

---

Supported in part by NSF grant CCF-1412958, the Packard Foundation and the Simons Foundation. This work was carried out under the auspices of the Simons Algorithms and Geometry (A&G) Think Tank.

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

$$\forall x, y \in M, \quad \tau d_M(x, y) \leq d_N(f(x), f(y)) \leq \alpha \tau d_M(x, y). \quad (1)$$

The infimum over  $\alpha \in [1, \infty]$  for which  $(M, d_M)$  embeds with distortion  $\alpha$  into  $(N, d_N)$  is denoted  $c_{(N, d_N)}(M, d_M)$ , or  $c_N(M)$  if the underlying metrics are clear from the context. If  $c_N(M) < \infty$ , then  $(M, d_M)$  is said to admit a bi-Lipschitz embedding into  $(N, d_N)$ . Given  $p \in [1, \infty)$ , if  $N$  is an  $L_p(\mu)$  space into which  $M$  admits a bi-Lipschitz embedding, then we use the notation  $c_{L_p(\mu)}(M) = c_p(M)$ . The numerical invariant  $c_2(M)$ , which measures the extent to which  $M$  is close to being a (subset of a) Euclidean geometry, is called the *distortion* of  $M$ .

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.

**P 2** (the bi-Lipschitz embedding problem into  $\mathbb{R}^k$ ) Obtain an intrinsic characterization of those metric spaces  $(M, d_M)$  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  $(M, d_M)$  admits a bi-Lipschitz embedding into a Hilbert space if and only if there exists  $C = C_M \in [0, 1]$  such that for every  $n \in \mathbb{N}$  and every positive semidefinite symmetric matrix  $A = (a_{ij}) \in M_n(\mathbb{R})$  all of whose rows sum to zero (i.e.,  $\sum_{j=1}^n a_{ij} = 0$  for every  $i \in \{1, \dots, n\}$ ), the following quadratic distance inequality holds true.

$$\forall x_1, \dots, x_n \in M, \quad \sum_{i=1}^n \sum_{j=1}^n a_{ij} d_M(x_i, x_j)^2 \leq C \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| d_M(x_i, x_j)^2. \quad (2)$$

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

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

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  $M$  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  $(M, d_M)$  and outputs in polynomial time an arbitrarily good approximation to its Euclidean distortion  $c_2(M)$ .

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).

**R 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 quantity  $q_X$  of a Banach space  $(X, \|\cdot\|)$  is the infimum over those  $q \in [2, \infty]$  for which<sup>1</sup>

$$\forall n \in \mathbb{N}, \forall x_1, \dots, x_n \in X, \quad \sum_{i=1}^n \|x_i\|^2 \lesssim_{X,q} \frac{n^{1-2/q}}{2^{n/q}} \sum_{\varepsilon \in \{-1,1\}^n} \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|^2. \quad (4)$$

In the special case  $x_1 = \dots = x_n = x \in X \setminus \{0\}$ , 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  $q \geq 2$ , 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  $i \in \{1, \dots, n\}$  we have

$$\begin{aligned} \|x_i\|^2 &= \left\| \frac{(\varepsilon_1 x_1 + \dots + \varepsilon_n x_n) + (-\varepsilon_1 x_1 - \dots - \varepsilon_{i-1} x_{i-1} + \varepsilon_i x_i - \varepsilon_{i+1} x_{i+1} \dots - \varepsilon_n x_n)}{2} \right\|^2 \\ &\leq \frac{\|\varepsilon_1 x_1 + \dots + \varepsilon_n x_n\|^2 + \|-\varepsilon_1 x_1 - \dots - \varepsilon_{i-1} x_{i-1} + \varepsilon_i x_i - \varepsilon_{i+1} x_{i+1} \dots - \varepsilon_n x_n\|^2}{2}. \end{aligned} \quad (5)$$

By averaging (5) over  $\varepsilon \in \{-1,1\}^n$  and  $i \in \{1, \dots, n\}$  we see that (4) holds if  $q = \infty$ . So, one could view (4) for  $q < \infty$  as a requirement that the norm  $\|\cdot\| : X \rightarrow [0, \infty)$  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 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  $n \rightarrow \infty$ ) randomized convexity inequality (4) for some  $q < \infty$ .

**Theorem 4.** Let  $(X, \|\cdot\|)$  be a Banach space. Then the following are equivalent:

- $\alpha \in [1, \infty)$  and  $\ell_\infty^n \hookrightarrow_\alpha X$  for all  $n \in \mathbb{N}$
- $q_X < \infty$

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  $\alpha$ -isomorphic to a subspace of a Banach space  $(Z, \|\cdot\|_Z)$  if there is a linear operator  $T : Y \rightarrow Z$  satisfying  $\|y\|_Y \leq \|Ty\|_Z \leq \alpha\|y\|_Y$  for every  $y \in Y$ ; this is the same as saying that  $Y$  embeds into  $Z$  with distortion  $\alpha$  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  $f : X \rightarrow Y$  such that both  $f$  and  $f^{-1}$  are uniformly continuous. By the aforementioned rigidity theorem of Ribe

<sup>1</sup>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  $Q \leq CQ'$  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)  $\phi, \beta$ , the notation  $Q \lesssim_{\phi, \beta} Q'$  means that  $Q \leq C(\phi, \beta)Q'$ , where  $C(\phi, \beta) > 0$  is allowed to depend only on  $\phi, \beta$ ; similarly for the notations  $Q \gtrsim_{\phi, \beta} Q'$  and  $Q \asymp_{\phi, \beta} Q'$ .

(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 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.

I o o o E [ o o o o o o o  
o o o o o o o o o o  
o o o o A o o o o o J L  
o o L [ o o o o o B L x o o  
o o o o A oo o o o o o o  
o [ A o o o o o o o o o

J. Bourgain, 1986.

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. 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  $\alpha \geq 1$  such that  $X$  is  $\alpha$ -isomorphic to a subspace of  $Y$ . In a remarkable feat of foresight, the following quote of Grothendieck [?] heralded the  $o \quad o \quad o \quad B$ , by shifting attention to the geometry of the finite dimensional subspaces of a Banach space as a way to understand its global structure.

o o o o B o type  
linéaire o o o o o M > 0 o o  
o E<sub>1</sub> o o o o F<sub>1</sub> F (

$$\leq 1 + M)^o \qquad o \qquad E_1 \qquad F_1 \qquad o \qquad \leq 1 \qquad o \qquad o \qquad o$$

A. Grothendieck, 1953.

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<sup>2</sup> 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 , 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  $\alpha$ -finitely representable in a Banach space  $Y$  if for any  $\beta > \alpha$ , any finite dimensional subspace of  $X$  is  $\beta$ -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  $\alpha$ -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

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

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

Pisier conjectured [?] that (6) forces the  $\ell_2$  (Pythagorean) direct sum  $(X_1 \oplus X_2 \oplus \dots)_2$  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  $k = k(n, \delta) \in \{1, \dots, n\}$  with  $\lim_{n \rightarrow \infty} k(n, \delta) = \infty$ , 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

For ICM 2018 participants only

o 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  $\ell_2 \otimes \ell_2 \cong S_1$  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  $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} \leq 1$ , 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.**  $(M, d)$

- $M$   $o$
- $\mathbb{N}$   $q = q(M) \in (0, \infty)$   $\{x_w\}_{w \in \mathbb{Z}_{2m}^\ell}$   $m$   $n \in \mathbb{N}$   $m = m(n, M, q) \in$

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

- $H$   $e_1, \dots, e_n$   $\theta(M) \in (0, \infty)$   $(\mathcal{B}_n, d_n)$   $\mathbb{R}^n$   $($   $)$   $n \in \mathbb{N}$   $2m$   $n$   $o$

$$c_m(\mathcal{B}_n) \gtrsim_m (\log n)^{\theta(M)}. \quad (8)$$

$o$   $o$   $m$   $B$   $q \in (0, \infty]$

$\{x_w\}_{w \in \mathbb{Z}_{2m}^\ell}$   $(7)$   $($   $)$   $n \in \mathbb{N}$   $m \in \mathbb{N}$   $(7)$   $o$   $o$   $o$   $o$

$(7)$   $o$   $m$   $(4)$   $o$   $qm$   $o$   $m$   $q \in (0, \infty]$   $o$

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  $\{x_w\}_{w \in \mathbb{Z}_{2m}^\ell}$  (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  $(M, d_m)$  (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  $(M, d_m)$ ; by considering the canonical realization of  $\mathbb{Z}_{2m}^n$  in  $\mathbb{C}^n$ , namely the points  $\{(\exp(\pi i w_1/m), \dots, \exp(\pi i w_n/m))\}_{w \in \mathbb{Z}_{2m}^\ell}$ , 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  $c_{\mathcal{F}}(M) > 1$ , then there are arbitrary large finite metric spaces whose minimal distortion in  $M$  is at least a fixed positive power (depending on  $M$ ) of the logarithm their cardinality. Hence, for example, if every  $n$ -point metric space embeds into  $M$  with distortion  $O(\log \log n)$ , then actually for any  $\delta > 0$ , every finite metric space embeds into  $M$  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  $\theta(M)$  in Theorem 9 could be bounded from below by a universal positive constant, as formulated in the following open question.

o 10 (metric cotype dichotomy problem) Is there a universal constant  $\theta > 0$  such that in Theorem 9 one could take  $\theta(M) > \theta$ . All examples that have been computed thus far leave the possibility that even  $\theta(M) \geq 1$ , which would be sharp (for  $M = \ell_2$ ) by Bourgain's embedding theorem [?]. Note, however, that in [?] it is asked whether for the Wasserstein space  $P_p(\mathbb{R}^3)$  we have  $\liminf_{p \rightarrow 1} \theta(P_p(\mathbb{R}^3)) = 0$ . 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  $c_2(\mathcal{M}) \lesssim \sqrt{\log |\mathcal{M}|}$ . 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  $\ell_2$ . 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).  $\begin{array}{ccccccc} n & o & n \in \mathbb{N} & \alpha \in (1, \infty) & k \in \{1, \dots, n\} & k \lesssim_{\alpha} \log n \\ o & \ell_2 & o \ell_2^k & o \ o \ \alpha \end{array}$

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.**  $\begin{array}{ccccccc} B & \mathcal{T}^{(2)} & o & o & o & o & H \\ o & o & \mathcal{C} \subseteq \mathcal{T}^{(2)} & k \in \{1, \dots, n\} & k \lesssim \log |\mathcal{C}| & k & o \\ F & o \ \mathcal{T}^{(2)} & \mathcal{C} & o \ F & O(1) & o & o \end{array}$

The space  $\mathcal{T}^{(2)}$  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  $\mathcal{T}^{(2)}$  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  $\mathcal{T}^{(2)}$  with  $c_2(F_n) \geq e^{c \text{Ack}^{-1}(n)}$ , where  $c > 0$  is a universal constant and  $\text{Ack}^{-1}(\cdot)$  is the inverse of the Ackermann function from computability theory (see e.g. [?, Appendix B]). So, indeed  $\lim_{n \rightarrow \infty} c_2(F_n) = \infty$ , 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 non-universal classical Banach space. In particular, the following question is open.



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

It is even open if in Question 13 one could obtain a bound of  $k = o(n)$  for any fixed  $p \in [1, \infty) \setminus \{2\}$ . 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  $\ell_p$ . At present, such an embedding statement is not known for  $p \in [1, \infty) \setminus \{2\}$ , though for  $p \in [1, 2]$  it is known [?] that any  $n$ -point subset of  $\ell_p$  embeds into  $\ell_2$  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).  $c_2(M) \lesssim \log |M|$  o  $m$

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  $M_n$  for which  $c_2(M_n) \gtrsim (\log n)/\log \log n$ . By John's theorem, for every  $\alpha \geq 1$ , if  $X$  is a finite dimensional normed space and  $c_X(M_n) \leq \alpha$ , then  $c_2(M_n) \leq \alpha \sqrt{\dim(X)}$ . Therefore the above lower bound on  $c_2(M_n)$  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 o 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  $\ell_2$  through the JL lemma is what was harnessed in [?] to deduce that any "host normed space" that contains  $M_n$  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).  $n \in \mathbb{N}$   $\alpha \in [1, \infty)$  o  $(X, \|\cdot\|_X)$   
o o  $k_n^\alpha(X)$   $k \in \mathbb{N}$  o any  $\mathcal{C} \subseteq X$   $|\mathcal{C}| = n$   
 $k$  o  $F_{\mathcal{C}} \circ X$  o  $\mathcal{C}$  o o  $\alpha$

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  $\alpha = 2$





$\alpha$  into some  $k$ -dimensional normed space, then  $k \geq n^{c/(2\alpha)^{1/\theta}}$ . Conversely, Remark 21 below shows that for every  $n \in \mathbb{N}$  and  $\alpha > 1$ , the  $\theta$ -snowflake of any  $n$ -point metric space embeds with distortion  $\alpha$  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  $\omega, \Omega : [0, \infty) \rightarrow [0, \infty)$ .

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

**Theorem 19** (impossibility of average dimension reduction).

$$\begin{array}{c} \begin{array}{ccc} \begin{array}{c} (X, \|\cdot\|_X) \\ x, y \in X \end{array} & \begin{array}{c} \alpha \in [1, \infty) \\ \frac{1}{n^2} \sum_{x, y \in X} \|f(x) - f(y)\|_X \geq \frac{1}{n^2} \sum_{x, y \in X} d_m(x, y) \end{array} & \begin{array}{c} n \in \mathbb{N} \\ f : \mathcal{M} \rightarrow X \end{array} \end{array} \quad \begin{array}{c} \begin{array}{c} o \\ (M, d_m) \end{array} \quad \begin{array}{c} c \in (0, \infty) \\ \|f(x) - f(y)\|_X \leq \alpha d_m(x, y) \\ \dim(X) \geq n^{c/\alpha} \end{array} \end{array}$$

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  $\alpha$ -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 ambient dimension to satisfy  $\dim(X) \geq n^{c/\alpha}$ . Prior to Theorem 19, the best-known bound here was  $\dim(X) \gtrsim (\log n)^2/\alpha^2$ , namely the aforementioned lower bound  $k_n^\alpha(\ell_\infty) \gtrsim (\log n)^2/\alpha^2$  of Linial, London and Rabinovich [?] actually treated the above "average distortion" requirement rather than only the (pairwise) bi-Lipschitz requirement.

*R* 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 property (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  $k_n^\alpha(\ell_\infty)$  was pursued by Johnson, Lindenstrauss and Schechtman, who proved in [?] that  $k_n^\alpha(\ell_\infty) \lesssim_\alpha n^{C/\alpha}$  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  $\alpha$ . 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  $k_n^\alpha(\ell_\infty)$  one does not need to adapt the target normed space to the metric space  $\mathcal{M}$  that is being embedded. The implicit dependence on  $\alpha$  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  $k_n^\alpha(\ell_\infty)$  that appears in (12) is that of [?], and for the remaining values of  $\alpha$  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 [?]).

*R* 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  $(\mathcal{M}, d_m)$  there is an integer  $k \lesssim n^{c/\alpha^{1/\theta}}$  and  $f = (f_1, \dots, f_k) : \mathcal{M} \rightarrow \mathbb{R}^k$  such that

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

Hence (here it becomes useful that we are dealing with the  $\ell_\infty^k$  norm, as it commutes with powering),

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

By works of Kahane [?] and Talagrand [?], there is  $m = m(\delta, \theta)$  and a mapping (a quasi-helix)  $h : \mathbb{R} \rightarrow \mathbb{R}^m$  such that  $|s - t|^\theta \leq \|h(s) - h(t)\|_{\ell_\infty^m} \leq (1 + \delta)|s - t|^\theta$  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- $\alpha$  embedding of the  $\theta$ -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  $\alpha, \theta$  that [?, ?] imply here is quite good, but likely not sharp as  $\alpha \rightarrow 1^+$  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  $\alpha$ , we have a bound<sup>4</sup> that is sharp up to universal constant factors.

$$\alpha \geq (\log n) \log \log n \implies k_n^\alpha(\ell_\infty) \asymp \frac{\log n}{\log \left( \frac{\alpha}{\log n} \right)}. \quad (20)$$

For a range of smaller values of  $\alpha$ , including those  $\alpha$  that do not tend to  $\infty$  with  $n$ , we have

$$1 \leq \alpha \leq \frac{\log n}{\log \log n} \implies n^{\frac{c}{\alpha}} \lesssim k_n^\alpha(\ell_\infty) \lesssim n^{\frac{c}{\alpha}}. \quad (21)$$

(21) satisfactorily shows that the asymptotic behavior of  $k_n^\alpha(\ell_\infty)$  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  $k_n^\alpha(\ell_\infty)$  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  $\alpha$ ). 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

$$\frac{\log n}{\log \log \log n} \lesssim k_n^{\Theta(\log n)}(\ell_\infty) \lesssim \log n. \quad (22)$$

The first inequality in (22) has not been stated in the literature, and we justify it in Section beto co227(1)27(er-t840(t

bound on  $k_n^\alpha(\ell_1)$  in (13) requires showing that if  $\mathcal{C}_{\text{BC}}$  embeds into an finite-dimensional linear subspace  $F$  of  $\ell_1$ , then necessarily  $\dim(F) \geq n^{c/\alpha^2}$ . However, Talagrand proved [?] that in this setting for every  $\beta > 1$  the subspace  $F$  embeds with distortion  $\beta$  into  $\ell_1^k$ , 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  $\ell_1$  is replaced by  $\ell_p$  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  $\mathcal{C}_{\text{BC}} \subseteq \ell_1$  also as a collection of diagonal matrices in  $S_1$ , and [?] treats this very same subset by strengthening the assertion of [?] that  $\mathcal{C}_{\text{BC}}$  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  $k_n^{\text{r}}(\ell_1, S_1) \geq n^{c/\alpha^2}$ .

A geometric challenge of the above discussion is that, even after one isolates a candidate  $n$ -point subset  $\mathcal{C}$  of  $\ell_1$  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  $\ell_1$  (or  $S_1$ ) whose pairwise distances are within a fixed, but potentially very large, factor  $\alpha \geq 1$  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  $\alpha \rightarrow 1^+$ . 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  $\mathcal{C}_{\text{BC}} \subseteq \ell_1$  is not even an  $\alpha$ -Lipschitz quotient of any subset of a low dimensional subspace of  $\mathbf{S}_1$ ; see [?] for the relevant definition and a complete discussion). The approach of [?] relies on an invariant that arose in the Ribe program and is called  $\text{Ribe}_q$ . Fix  $q > 0$ . Let  $\{\chi_t\}_{t \in \mathbb{Z}}$  be a Markov chain on a state space  $\Omega$ . Given an integer  $k \geq 0$ , denote by  $\{\tilde{\chi}_t(k)\}_{t \in \mathbb{Z}}$  the process that equals  $\chi_t$  for time  $t \leq k$ , and evolves independently of  $\chi_t$  (with respect to the same transition probabilities) for time  $t > k$ . Following [?], the Markov  $q$ -convexity constant of a metric space  $(M, d_M)$ , denoted  $\Pi_q(M)$ , 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  $\Omega$  and every  $f : \Omega \rightarrow M$  we have

$$\left( \sum_{k=1}^{\infty} \sum_{t \in \mathbb{Z}} \frac{1}{2^{qk}} \mathbb{E} \left[ d_m(f(\tilde{X}_t(t-2^k)), f(X_t))^q \right] \right)^{\frac{1}{q}} \leq \Pi \left( \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_m(f(X_t), f(X_{t-1}))^q \right] \right)^{\frac{1}{q}}.$$

By [?, ?], a Banach space  $X$  satisfies  $\Pi_q(X) < \infty$  if and only if it has an equivalent norm  $||| \cdot ||| : X \rightarrow [0, \infty)$  whose modulus of uniform convexity has power type  $q$ , i.e.,  $|||x+y||| \leq 2 - \Omega_X(|||x-y|||^q)$  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  $q \geq 2$  for which  $X$  admits an equivalent norm whose modulus of uniform convexity has power type  $q$ "). By [?, ?], the Brinkman–Charikar subset  $\mathcal{C}_{\text{BC}} \subseteq \ell_1$  (as well as a variant of it due to Laakso [?] which has [?] the same non-embeddability property into low-dimensional subspaces of  $\ell_1$ ) satisfies  $\Pi_q(\mathcal{C}_{\text{BC}}) \gtrsim (\log n)^{1/q}$  for every  $q \geq 2$  (recall that in our notation  $|\mathcal{C}_{\text{BC}}| = n$ ). At the same time, it is proved in [?] that  $\Pi_2(F) \lesssim \sqrt{\log \dim(F)}$  for every finite dimensional subset of  $\mathbf{S}_1$ . It remains to contrast these asymptotic behaviors (for  $q = 2$ ) to deduce that if  $\mathbf{c}_F(\mathcal{C}_{\text{BC}}) \leq \alpha$ , then necessarily  $\dim(F) \geq n^{c/\alpha^2}$ .

Prior to the forthcoming work [?], the set  $\mathbb{C}_{\text{BC}}$  (and variants thereof of a similar nature) was the only known example that demonstrates that there is no  $\ell_1$  analogue of the JL-lemma. The following theorem is from [?].

**Theorem 23.** *Let  $c \in (0, \infty)$  and  $n \in \mathbb{N}$ . Let  $\mathcal{H}_n$  be a  $c$ -Hankel class on  $\ell_1^n$ . Then*

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

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  $\ell_1$  (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  $\ell_4$ , 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  $\ell_1$ , 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  $\ell_1$ , 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  $(X, \|\cdot\|_X)$   $\mathcal{O}$  if there is  $\alpha = \alpha_X \in [1, \infty)$  such that

$$\lim_{n \rightarrow \infty} \frac{\log k_n^\alpha(X)}{\log n} = 0.$$

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  $\alpha$  into some linear subspace  $F$  of  $X$  with  $\dim(F) = n^{\mathcal{O}(1)}$ .

Analogously, we say that  $(X, \|\cdot\|_X)$   $\mathcal{O}$  if there is  $\alpha = \alpha_X \in [1, \infty)$  such that for any  $n \in \mathbb{N}$  there is  $k_n = n^{\mathcal{O}(1)}$ , i.e.,  $\lim_{n \rightarrow \infty} (\log k_n)/\log n = 0$ , 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  $f : \mathcal{C} \rightarrow F$  which is  $\alpha$ -Lipschitz, i.e.,  $\|f(x) - f(y)\|_X \leq \alpha \|x - y\|_X$  for all  $x, y \in \mathcal{C}$ , yet

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

Our choice here of the behavior  $n^{\mathcal{O}(1)}$  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  $\mathcal{O}_X(n)$  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  $\infty$  faster than  $\log n$  (impossibility of a JL-style guarantee) would be very interesting.

**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  $(\frac{1}{n^2} \sum_{x \in \mathcal{C}} \sum_{y \in \mathcal{C}} \|f(x) - f(y)\|_X^p)^{1/p}$  distortion

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

*R* 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  $\infty$  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  $\ell_\infty$  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 non-universal (finite cotype) Banach space, and it would be very interesting to provide such an example. By [?, ?] we know that  $\ell_1$  and  $S_1$  fail to admit metric dimension reduction, but this is not known for average distortion, thus leading to the following question.

*o* 27 Does  $\ell_1$  admit average metric dimension reduction? Does  $\ell_p$  have this property for any  $p \in [1, 2)$ ?

All of the available examples of  $n$ -point subsets of  $\ell_1$  for which the  $\ell_1$  analogue of the JL lemma fails (namely if  $k = O(\log n)$ , then they do not embed with  $O(1)$  distortion into  $\ell_1^k$ ) 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  $\ell_1$  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  $\ell_1$ . In the reverse direction, a positive answer to Question 27, beyond its intrinsic geometric/structural interest, could have algorithmic applications.

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

**Corollary 28.**  $B \quad X, Y \quad o \quad o \quad X \hat{\otimes} Y \quad o \quad o$

Since  $S_1$  is isometric to  $\ell_2 \hat{\otimes} \ell_2$  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  $k_n^\alpha(n) \lesssim_\alpha \log n$  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.**  $B \quad X, Y \quad o \quad o \quad X \hat{\otimes} Y \quad o$   
 $o \quad o \quad o \quad o \quad p \in (2, \infty) \quad X = \ell_p$

*P oo* By [?] (which relies on major input from the theory of locally decodable codes [?] and an important inequality of Pisier [?]), the 3-fold product  $\ell_3 \hat{\otimes} \ell_3 \hat{\otimes} \ell_3$  is universal. So, by the recent work [?] (Theorem 19),  $\ell_3 \hat{\otimes} \ell_3 \hat{\otimes} \ell_3$  does not admit average metric dimension reduction. At the same time, by Theorem 25 we know that  $\ell_3$  admits average metric dimension reduction. So, if  $\ell_3 \hat{\otimes} \ell_3$  fails to admit average metric dimension reduction, then we can take  $X = Y = \ell_3$  in Theorem 29. Otherwise,  $\ell_3 \hat{\otimes} \ell_3$  does admit average metric dimension reduction, in which case we can take  $X = \ell_3$  and  $Y = \ell_3 \hat{\otimes} \ell_3$ . 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  $\ell_p \hat{\otimes} \ell_p \hat{\otimes} \ell_q$  is universal if  $2/p + 1/q \leq 1$ , or equivalently  $q \geq p/(p-2)$ . If we choose, say,  $q = \max\{2, p/(p-2)\}$ , then by Theorem 25 we know that both  $\ell_p$  and  $\ell_q$  admit average metric dimension reduction, while  $\ell_p \hat{\otimes} \ell_p \hat{\otimes} \ell_q$  does not. So, the second assertion of Theorem 29 holds for either  $Y = \ell_p$  or  $Y = \ell_p \hat{\otimes} \ell_q$ .  $\square$



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

o 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  $\ell_1$ ) is finitely representable in  $\ell_3 \widehat{\otimes} \ell_3$ ; see e.g. [?, page 61]. The fact that no Banach space in which  $\ell_1$  is finitely representable 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  $\ell_1$  (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  $p = 2$ .

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

1.4.2.  $W$  Let  $(M, d_M)$  be a metric space and  $p \in [1, \infty)$ . The Wasserstein space  $P_p(M)$  is not a Banach space, but there is a natural version of the metric dimension reduction question in this context.

o 32 Fix  $\alpha > 1$ ,  $n \in \mathbb{N}$  and  $\mu_1, \dots, \mu_n \in P_p(M)$ . What is the asymptotic behavior of the smallest  $k \in \mathbb{N}$  for which there is  $\mathcal{S} \subseteq M$  with  $|\mathcal{S}| \leq k$  such that  $(\{\mu_1, \dots, \mu_n\}, W_p)$  embeds with distortion  $\alpha$  into  $P_p(\mathcal{S})$ ?

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  $M = \{1, \dots, n\}^2 \subseteq \mathbb{R}^2$  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  $M$  is a finite metric space, then  $P_1(M)$  embeds into  $\ell_1$  with distortion  $O(\log |M|)$ . 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  $P_1(M)$  embeds into  $\ell_1$  with distortion  $o(\log n)$ , i.e., better distortion than the general bound that is provided by Bourgain's embedding theorem (actually the  $\ell_1$ -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  $M$ , since if  $(M = \{x_1, \dots, x_n\}, d_M)$  is a metric space for which  $c_1(M) \asymp \log n$ , then we can take  $\mu_1, \dots, \mu_n$  to be the point masses  $\delta_{x_1}, \dots, \delta_{x_n}$ , so that  $(\{\mu_1, \dots, \mu_n\}, W_1)$  is isometric to  $(M, d_M)$ . The pertinent issue is therefore to study Question 32 when the  $M$  is "nice." For example, sufficiently good bounds here for  $M = \mathbb{R}^2$  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  $M$  is itself far from being embeddable in  $\ell_1$ , but the difficulty persists even when  $M = Td\ell TFTbutnTftheTfiieingethespiritofQuesETBT$

To see this, focus for concreteness on the case  $p = 2$ . Fix  $\alpha \geq 1$  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  $P_2(\mathbb{R}^3)$ , where  $\mathbb{R}^3$  is equipped with the standard Euclidean metric. Hence, if the image under this embedding of  $\mathcal{N}$  in  $P_2(\mathbb{R}^3)$  embedded into some  $k$ -dimensional normed space with distortion  $\alpha$ , then by Theorem 17 necessarily  $k \geq n^{c/\alpha^2}$  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  $P_2(\mathcal{S})$  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  $P_2(\mathbb{R}^3)$  of  $(\mathcal{N}, \sqrt{d_{\mathcal{N}}})$  embedded with average distortion  $\alpha$  into some  $k$ -dimensional normed space, then necessarily  $k \geq \exp(\frac{c}{\alpha}\sqrt{\log n})$ .

## 2. FINE BEHANCE

The article [?] of Johnson and Lindenstrauss is devoted to proving a theorem on the extension of Lipschitz functions from finite subsets of metric spaces.<sup>5</sup> Over the ensuing decades, the classic [?] attained widespread prominence outside the rich literature on the  $L_2$  norm, 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  $n \geq 2$  and (distortion/error tolerance)  $\alpha \in (1, \infty)$ , if  $x_1, \dots, x_n$  are distinct vectors in a Hilbert space  $(H, \|\cdot\|_H)$ , then there exists (a target dimension)  $k \in \{1, \dots, n\}$  and a new  $n$ -tuple of  $k$ -dimensional vectors  $y_1, \dots, y_n \in \mathbb{R}^k$  such that

$$k \lesssim_{\alpha} \log n, \quad (25)$$

and the assignment  $x_i \mapsto y_i$ , viewed as a mapping into  $\ell_2^k$ , has distortion at most  $\alpha$ , i.e.,

$$\forall i, j \in \{1, \dots, n\}, \quad \|x_i - x_j\|_H \leq \|y_i - y_j\|_{\ell_2} \leq \alpha \|x_i - x_j\|_H. \quad (26)$$

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

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

<sup>5</sup>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

$$k \geq \frac{\log n}{\log(\alpha + 1)}. \quad (27)$$

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  $k_n^\alpha(\ell_2)$  in (10). As we already explained, the second lower bound (for the almost-isometric regime) on  $k_n^\alpha(\ell_2)$  in (10) is due to the very recent work [?]. The upper bound on  $k_n^\alpha(\ell_2)$  in (10), namely that in (25) we can take

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

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

*o* 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  $k \leq \mathfrak{S}(\delta) \log n$  and  $y_1, \dots, y_n \in \mathbb{R}^k$  that satisfy  $1 \leq \|y_i - y_j\|_2 \leq 1 + \delta$  for all distinct  $i, j \in \{1, \dots, n\}$ ? 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 [?].

*R* 34 The upper bound (28) treats the target dimension in the JL lemma for an 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  $k$ -dimensional norm. In this (both special, and more general) setting, the bound (27) is quite sharp for large  $\alpha$ . 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 norm  $\|\cdot\|$  on  $\mathbb{R}^k$  there exist  $y_1, \dots, y_n \in \mathbb{R}^k$  satisfying  $1 \leq \|y_i - y_j\| \leq \alpha$  for distinct  $i, j \in \{1, \dots, n\}$ . 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  $\text{Proj}_{\mathbb{R}} \in M_{k \times (n-1)}(\mathbb{R})$  be the  $k$  by  $n - 1$  matrix of the orthogonal projection from  $\mathbb{R}^{n-1}$  onto  $\mathbb{R}^k$ , i.e.,  $\text{Proj}_{\mathbb{R}} z = (z_1, \dots, z_k) \in \mathbb{R}^k$  is the first  $k$  coordinates of  $z = (z_1, \dots, z_{n-1}) \in \mathbb{R}^{n-1}$ . One could attempt to simply truncate the vectors  $x_1, \dots, x_n$  so as to obtain  $k$ -dimensional vectors, namely to consider the vectors  $\{y_i = \text{Proj}_{\mathbb{R}} x_i\}_{i=1}^n$  in (26). This naive (and heavy-handed) way of forcing low-dimensionality can obviously fail miserably, e.g. we could have  $\text{Proj}_{\mathbb{R}} x_i = 0$  for all  $i \in \{1, \dots, n\}$ . 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  $O_{n-1} \subseteq M_{n-1}(\mathbb{R})$  denote the group of  $n - 1$  by  $n - 1$  orthogonal matrices, and fix (a scaling factor)  $\sigma \in (0, \infty)$ . Let  $O \in O_{n-1}$  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.

$$\{y_i = \sigma \text{Proj}_{\mathbb{R}} O x_i\}_{i=1}^n \subseteq \mathbb{R}^k. \quad (29)$$

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

*o* (in a certain sense that will be made precise) among all the possible choices of random assignments of  $x_1, \dots, x_n$  to  $y_1, \dots, y_n$  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  $\mu$  be any Borel probability measure on  $M_{k \times (n-1)}(\mathbb{R})$ , i.e.,  $\mu$  represents an arbitrary (reasonably measurable) distribution over  $k \times (n - 1)$  random matrices  $A \in M_{k \times (n-1)}(\mathbb{R})$ . For  $\alpha \in (1, \infty)$  define

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

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

$$\begin{aligned}
& \mu \left[ \bigcap_{i,j \in \{1, \dots, n\}} \left\{ A \in M_{k \times (n-1)}(\mathbb{R}) : \|x_i - x_j\|_{\ell_2^{\ell-1}} \leq \|Ax_i - Ax_j\|_{\ell_2} \leq \alpha \|x_i - x_j\|_{\ell_2^{\ell-1}} \right\} \right] \\
&= 1 - \mu \left[ \bigcup_{i=1}^n \bigcup_{j=i+1}^n \left( M_{k \times (n-1)}(\mathbb{R}) \setminus \left\{ A \in M_{k \times (n-1)}(\mathbb{R}) : 1 \leq \left\| A \frac{x_i - x_j}{\|x_i - x_j\|_{\ell_2^{\ell-1}}} \right\|_{\ell_2} \leq \alpha \right\} \right) \right] \\
&\geq 1 - \sum_{i=1}^n \sum_{j=i+1}^n \left( 1 - \mu \left[ \left\{ A \in M_{k \times (n-1)}(\mathbb{R}) : 1 \leq \left\| A \frac{x_i - x_j}{\|x_i - x_j\|_{\ell_2^{\ell-1}}} \right\|_{\ell_2} \leq \alpha \right\} \right] \right) \\
&\geq 1 - \binom{n}{2} (1 - \mathfrak{p}_\mu^\alpha).
\end{aligned} \tag{31}$$

Hence, the random vectors  $\{y_i = Ax_i\}_{i=1}^n$  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  $\mu$  on  $M_{k \times (n-1)}(\mathbb{R})$  for which  $\mathfrak{p}_\mu^\alpha$  is as large as possible. To this end, define

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

Then, the conclusion (26) of the JL lemma will be valid provided  $k \in \{1, \dots, n\}$  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).  $\alpha \in (1, \infty) \quad n \geq 4$   
 $k \in \{1, \dots, n-3\} \quad L \quad \mu = \mu_{n,k}^\alpha \quad o \quad o \quad o \quad M_{k \times (n-1)}(\mathbb{R}) \quad o \quad o$

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

$$o \quad oo \quad O \in O_{n-1} \quad o \quad o \quad o \quad \neq \quad H \quad o \quad O_{n-1} \quad \mathfrak{p}_\mu^\alpha = \mathfrak{p}_{n,k}^\alpha$$

Obviously (34) is a multiple of a uniformly random rank  $k$  orthogonal projection  $\text{Proj} : \mathbb{R}^{n-1} \rightarrow \mathbb{R}^{n-1}$  (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 [?].

$P \quad oo \quad o \quad P \quad o \quad o \quad o \quad 5$  Given  $A \in M_{k \times (n-1)}(\mathbb{R})$ , denote its singular values by  $s_1(A) \geq \dots \geq s_k(A)$ , i.e., they are the eigenvalues (with multiplicity) of the symmetric matrix  $\sqrt{AA^*} \in M_k(\mathbb{R})$ . Then,

$$\mathfrak{H}^{O_{\ell-1}} \left[ \left\{ O \in O_{n-1} : 1 \leq \|AOz\|_{\ell_2} \leq \alpha \right\} \right] = \int_{\mathbf{S}^{-1}} \psi_{n,k}^\alpha \left( \left( \sum_{i=1}^k s_i(A)^2 \omega_i^2 \right)^{\frac{1}{2}} \right) d\mathfrak{H}^{\mathbf{S}^{-1}}(\omega), \tag{35}$$

where  $\mathfrak{H}^{O_{\ell-1}}$  and  $\mathfrak{H}^{\mathbf{S}^{-1}}$  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  $\psi_{n,k}^\alpha : [0, \infty) \rightarrow \mathbb{R}$  is defined by

$$\forall \sigma \in [0, \infty), \quad \psi_{n,k}^\alpha(\sigma) \stackrel{\text{def}}{=} \frac{2\pi^{\frac{2}{\ell}}}{\Gamma(\frac{k}{2})} \frac{\max\{1, \sigma\}}{\max\{1, \frac{\sigma}{\alpha}\}} \frac{(s^2 - 1)^{\frac{\ell-3}{2}}}{s^{n-2}} ds. \tag{36}$$

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

$$A = U \begin{pmatrix} s_1(A) & 0 & \dots & \dots & 0 \\ 0 & s_2(A) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & s_k(A) \end{pmatrix} \text{Proj}_{\mathbb{R}} V, \quad (37)$$

where  $U \in O_k$  and  $V \in O_{n-1}$ . If  $O \in O_{n-1}$  is distributed according to  $\mathfrak{H}^{O_{\ell}-1}$ , then by the left-invariance of  $\mathfrak{H}^{O_{\ell}-1}$  we know that  $VO$  is distributed according to  $\mathfrak{H}^{O_{\ell}-1}$ . 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  $VOz$  is distributed according to the normalized Haar measure on  $\mathbf{S}^{n-2}$ . So,  $\text{Proj}_{\mathbb{R}} VOz$  is distributed on the Euclidean unit ball of  $\mathbb{R}^k$ , with density

$$(u \in \mathbb{R}^k) \mapsto \frac{\Gamma(\frac{n-1}{2})}{\pi^{\frac{k}{2}} \Gamma(\frac{n-1-k}{2})} \left(1 - \|u\|_{\ell_2}^2\right)^{\frac{\ell-3}{2}} \mathbf{1}_{\{\|u\|_2 \leq 1\}}. \quad (38)$$

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  $AOz$  has the same distribution as  $(\sum_{i=1}^k s_i(A)^2 u_i^2)^{1/2}$ , 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  $\lim_{\sigma \rightarrow \infty} \psi_{n,k}^\alpha(\sigma) = 0$ . 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

$$\sigma_{\max}(n, k, \alpha) \stackrel{\text{def}}{=} \sqrt{k}$$

*R* 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  $\mathbf{A}$  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  $(z \in \mathbb{R}^{n-1}) \mapsto \mathbf{A}z$  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  $\mathbf{O}_{n-1}$  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).

*R* 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  $\mathbf{O}_{n-1}$  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  $\mathbf{A}$  is replaced by the random matrix  $\sigma \mathbf{G}$ , where  $\sigma = 1/\sqrt{k}$  and the entries of  $\mathbf{G} = (g_{ij}) \in M_{k \times (n-1)}(\mathbb{R})$  are independent standard Gaussian random variables. We can, however, optimize over the scaling factor  $\sigma$  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  $\mathbf{G}$  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  $\frac{1}{\sqrt{k}}\mathbf{G}$  but rather the random matrix

$$\mathbf{G}_k^\alpha \stackrel{\text{def}}{=} \sqrt{\frac{\alpha^2 - 1}{2k \log \alpha}} \cdot \mathbf{G}. \quad (42)$$

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

$$\begin{aligned} 1 - \mathbb{P}\left[1 \leq \|\mathbf{G}_k^\alpha z\|_{\ell_2} \leq \alpha\right] &= \frac{2k^{\frac{1}{2}}}{\Gamma(\frac{k}{2})} \int_{\log \alpha}^{\infty} \left(\frac{\beta}{e^{2\beta} - 1}\right)^{\frac{1}{2}} \exp\left(-\frac{k\beta}{e^{2\beta} - 1}\right) d\beta \\ &< \frac{4k^{\frac{1}{2}-1}}{\Gamma(\frac{k}{2})} \left(\frac{\alpha^2 - 1}{\log \alpha} \alpha^{\frac{2}{\alpha^2 - 1}}\right)^{-\frac{1}{2}} \frac{(\alpha^2 - 1)^2 \log \alpha}{2\alpha^4 \log \alpha + 2\alpha^2 - \alpha^4 - 4\alpha^2(\log \alpha)^2 - 2 \log \alpha - 1}. \end{aligned} \quad (43)$$

The first step in (43) follows from a straightforward computation using the fact that the squared Euclidean length of  $\mathbf{G}_k^\alpha z$  is distributed according to a multiple of the  $\chi^2$  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  $\alpha$  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  $i \in \{1, \dots, n\}$  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

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

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  $\alpha \rightarrow 1$ , the above bounds improve over those of [?] only in the second-order terms. For larger  $\alpha$  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.

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  $M_{k \times n}(\mathbb{R})$  (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 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  $\alpha$  into  $\ell_2^k$  for some  $k \in \mathbb{N}$ ? If one ignores the dependence on the distortion  $\alpha$ , then this qualitative question coincides with Problem 2 (the bi-Lipschitz embedding problem into  $\mathbb{R}^k$ ), since if a metric space  $(M, d_M)$  satisfies  $\inf_{k \in \mathbb{N}} c_{\mathbb{R}}(M) < \infty$ , 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  $(M, d_M)$  will be closed balls, i.e., for  $x \in M$  and  $r \in [0, \infty)$  we write  $B_M(x, r) = \{y \in M : d_M(x, y) \leq r\}$ . Given  $K \in [2, \infty)$ , a metric space  $(M, d_M)$  is said to be  $K$ -o (e.g. [?, ?]) if every ball in  $M$  (centered anywhere in  $M$  and of any radius) can be covered by at most  $K$  balls of half its radius, i.e., for every  $x \in M$  and  $r \in [0, \infty)$  there is  $m \in \mathbb{N}$  with  $m \leq K$  and  $y_1, \dots, y_m \in M$  such that  $B_M(x, r) \subseteq B_M(y_1, \frac{1}{2}r) \cup \dots \cup B_M(y_m, \frac{1}{2}r)$ . A metric space is doubling if it is  $K$ -doubling for some  $K \in [2, \infty)$ .

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

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

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  $M$  is doubling if and only if its  $\theta$ -snowflake  $M^\theta$  is doubling. With this in mind, Theorem 40 below is a very important classical achievement of Assouad [?].

**Theorem 40.**  $(M, d_M)$

- $M$  o
- o  $\theta \in (0, 1)$   $k \in \mathbb{N}$   $M^\theta$   $L$   $\mathbb{R}^k$  o  $\mathbb{R}^k$  o  $k \in \mathbb{N}$
- o o  $M$   $L$   $\mathbb{R}^k$  o  $k \in \mathbb{N}$

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  $M$  is  $K$ -doubling, then  $M^\theta$  embeds into  $\mathbb{R}^{k(K, \theta)}$  with distortion  $\alpha(K, \theta)$ ; the argument of [?] inherently gives that as  $\theta \rightarrow 1$ , i.e., as the  $\theta$ -snowflake  $M^\theta$  approaches the initial metric space  $M$ , we have  $\alpha(K, \theta) \rightarrow \infty$  and  $k(K, \theta) \rightarrow \infty$ . 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  $\ell_\infty$  norm on  $\mathbb{R}^k$  (with the dimension  $k$  tending to  $\infty$  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  $k(K, \theta) \leq k(K)$  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  $\theta \rightarrow 1$ , for  $\theta < 1$  it deforms the initial metric space substantially (e.g. such a  $\theta$ -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 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 [?].

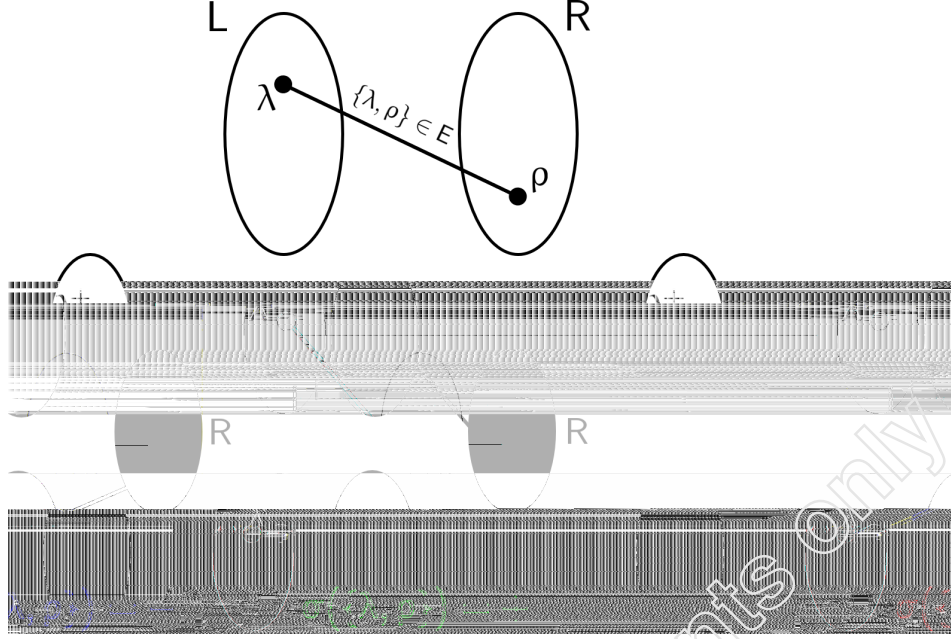
All of the known "bad examples" (including, in addition to the Heisenberg group, those that were subsequently found in [?, ?, ?, ?]) which show that the doubling property is not the sought-after answer to Problem 2 do not even embed into an infinite-dimensional Hilbert space. This leads to the following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

o 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  $\alpha_K$ . If Question 41 had a positive answer, then it would be very interesting to determine the asymptotic behavior of  $d_K$  and  $\alpha_K$  as  $K \rightarrow \infty$ . 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  $(M, d_M)$  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, /F1127(51T1b)-28(ovide)edhatthe



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  $p \geq p_0$  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<sup>7</sup> 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  $\ell_p^k$  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 [



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

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

Let  $\gamma : \{0, \dots, k\} \rightarrow L^+ \cup L^- \cup R$  be a geodesic in  $G_\sigma$  that joins  $\lambda^+$  and  $\lambda^-$ . Thus  $\gamma(0) = \lambda^+$ ,  $\gamma(k) = \lambda^-$  and  $\{\{\gamma(i-1), \gamma(i)\}\}_{i=1}^k$  are distinct edges in  $E_\sigma$  (they are distinct because  $\gamma$  is a shortest path joining  $\lambda^+$  and  $\lambda^-$  in  $G_\sigma$ ). By the injectivity of  $\pi$  on  $E_\sigma$ , 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  $d_{G_\sigma}$  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  $d_\sigma^{s,T} : (L^+ \cup L^- \cup R) \times (L^+ \cup L^- \cup R) \rightarrow [0, \infty]$  by defining

$$\forall x, y \in L^+ \cup L^- \cup R, \quad d_\sigma^{s,T}(x, y) \stackrel{\text{def}}{=} \min \{s d_{G_\sigma}(x, y), T\}. \quad (46)$$

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.**  $\eta > 0$   $\omega, \Omega : [0, \infty) \rightarrow [0, \infty)$   $G = (L, R, E)$   $g \in \mathbb{N}$   $s, T > 0$

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

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

$$\delta = \frac{1}{3} \quad (48)$$

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

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

$$P \ll o \ll o \ll \frac{1}{\gamma} \ll L$$

Recalling (17), we can fix  $s \in (0, \infty)$  such that

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

There is a universal constant  $\kappa \in (0, \infty)$  such that for arbitrarily large  $n \in \mathbb{N}$  there exists a bipartite graph  $G = (L, R, E)$  with  $|L| = |R| = n$ , girth at least  $g$  (i.e.,  $G$  does not contain any cycle of length smaller than  $g$ ) and  $|E| \geq n^{1+\kappa/g}$ . Determining the largest possible value of  $\kappa$  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  $\kappa$  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  $\kappa$  here (arising from an algebraic construction).

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

$$\mathcal{V} \stackrel{\text{def}}{=} \bigcap_{u=1}^h \left\{ (X, Y, z) \in \mathbf{M}_{n \times r}(\mathbb{R}) \times \mathbf{M}_{r \times n}(\mathbb{R}) \times \mathbb{R}^h; \mathbf{p}_u(X, Y, z) = 0 \right\}. \quad (56)$$

We claim that  $\mathcal{V}$  has at least  $m$  connected components. In fact, if  $k, \ell \in \{1, \dots, m\}$  are distinct, then  $(\mathbf{B}_k, \mathbf{C}_k, \zeta_k)$  and  $(\mathbf{B}_\ell, \mathbf{C}_\ell, \zeta_\ell)$  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  $(\mathbf{B}_k, \mathbf{C}_k, \zeta_k), (\mathbf{B}_\ell, \mathbf{C}_\ell, \zeta_\ell) \in \mathcal{C}$ . Since  $k \neq \ell$ , by switching the roles of  $k$  and  $\ell$  if necessary, the assumption of Lemma 44 ensures that there exists  $(i, j) \in \mathbf{E}$  such that  $(\mathbf{B}_k \mathbf{C}_k)_{ij} = a_{ij}^k < 0 < a_{ij}^\ell = (\mathbf{B}_\ell \mathbf{C}_\ell)_{ij}$ . So, if we denote  $\psi : \mathcal{C} \rightarrow \mathbb{R}$  by  $\psi(X, Y, z) = (XY)_{ij}$ , then  $\psi(\mathbf{B}_k, \mathbf{C}_k, \zeta_k) < 0 < \psi(\mathbf{B}_\ell, \mathbf{C}_\ell, \zeta_\ell)$ . Since  $\mathcal{C}$  is connected and  $\psi$  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  $(i, j) \in \mathbf{J}_u$ . By the definition (55) of  $\mathbf{p}_u$ , the fact that  $\psi(X, Y, z) = (XY)_{ij} = 0$  implies that  $\mathbf{p}_u(X, Y, z) = -z_u^2 - \frac{1}{2}\mu^{|J|} \leq -\frac{1}{2}\mu^{|J|} < 0$ , 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  $\mathbf{p}_u$  is  $4|\mathbf{J}_u| \leq 4\alpha$ . So, the variety  $\mathcal{V}$  in (56) is defined using  $h$  polynomials of degree at most  $4\alpha$  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  $4\alpha(8\alpha - 1)^{2nr+h-1} = 4\alpha(8\alpha - 1)^{2nr+\lceil |\mathbf{E}|/\alpha \rceil - 1}$ . Since we already established that this number of connected components is at least  $m$ , it follows that

$$m \leq 4\alpha(8\alpha - 1)^{2nr+\lceil \frac{|\mathbf{E}|}{\alpha} \rceil - 1} \iff r \geq \frac{1}{2n} \left( \frac{\log\left(\frac{m}{4\alpha}\right)}{\log(8\alpha - 1)} - \left\lceil \frac{|\mathbf{E}|}{\alpha} \right\rceil + 1 \right). \quad (57)$$

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

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

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

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

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

$$\mathcal{B}_\Delta \stackrel{\text{def}}{=} \left\{ \sigma : \mathbf{E} \rightarrow \{-, +\} : \dim_{\omega, \Omega}(\mathbf{L}^+ \cup \mathbf{L}^- \cup \mathbf{R}, d_\sigma^{s, T}) \leq \Delta \right\}. \quad (58)$$

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  $f_\sigma : \mathbf{L}^+ \cup \mathbf{L}^- \cup \mathbf{R} \rightarrow X_\sigma$  that satisfies

$$\forall x, y \in \mathbf{L}^+ \cup \mathbf{L}^- \cup \mathbf{R}, \quad \omega(d_\sigma^{s, T}(x, y)) \leq \|f_\sigma(x) - f_\sigma(y)\|_{X_\sigma} \leq \Omega(d_\sigma^{s, T}(x, y)). \quad (59)$$

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.,

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

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

$$\forall (i, j) \in \{1, \dots, n\}^2, \quad a_{ij}^\sigma$$

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  $d_{G_\sigma}(\lambda_i^+, \lambda_i^-) \geq g$  for all  $i \in \{1, \dots, n\}$  and  $\sigma : E \rightarrow \{-, +\}$ . So,

$$d_{G_\sigma}^{s,T}(\lambda_i^+, \lambda_i^-) \stackrel{(46)}{=} \min\{sd_{G_\sigma}(\lambda_i^+, \lambda_i^-), T\} \geq \min\{sg, T\} \stackrel{(47)}{=} sg. \quad (63)$$

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

$$d_{G_\sigma}^{s,T}(\lambda_i^{\sigma(i,j)}, \rho_j) \stackrel{(46)}{\leq} sd_{G_\sigma}(\lambda_i^{\sigma(i,j)}, \rho_j) = s. \quad (64)$$

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

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

Hence,  $a_{ij}^\sigma \neq 0$  and  $\text{sign}(a_{ij}^\sigma) = \sigma(i, j)$  for all  $(i, j) \in E$  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

$$\text{rank}(A_\tau) \geq \frac{c \log |\mathcal{B}_\Delta|}{n \log \left( \frac{|E|}{\log |\mathcal{B}_\Delta|} \right)}, \quad (65)$$

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  $e_\tau^1, \dots, e_\tau^\Delta$  of  $X_\tau$  and for every  $j \in \{1, \dots, n\}$  write  $f_\tau(\rho_j) = \gamma_{\tau,j}^1 e_\tau^1 + \dots + \gamma_{\tau,j}^\Delta e_\tau^\Delta$  for some scalars  $\gamma_{\tau,j}^1, \dots, \gamma_{\tau,j}^\Delta \in \mathbb{R}$ . Hence,

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

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

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

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

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

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  $\delta\eta|E|/n \geq 1$ , since otherwise the left hand side of (48) vanishes. Then, by choosing  $\Delta = \lfloor \delta\eta|E|/n \rfloor \in \mathbb{N}$  in the above reasoning it follows from (66) that

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

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

5. N I EA EC\* A GA A D I IBI I\* F A E AGE DI E I ED C\* I

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

<sup>8</sup>We are assuming irreducibility only for notational convenience, namely so that  $\pi$  will be unique and could therefore be suppressed in the ensuing notation. Our arguments work for any stochastic matrix and any probability measure  $\pi$  on  $\{1, \dots, n\}$  with respect to which  $A$  is reversible. We suggest focusing initially on the case when  $A$  is symmetric and  $\pi$  is the uniform measure on  $\{1, \dots, n\}$ , 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  $\gamma(\mathbf{A}, d_{\mathcal{M}}^p)$  to be the infimum over those  $\gamma > 0$  such that

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

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  $\gamma(\mathbf{A}, \|\cdot\|_H^2) = 1/(1 - \lambda_2(\mathbf{A}))$  is the reciprocal of the  $\lambda_2$  of  $\mathbf{A}$ . In general, we think of  $\gamma(\mathbf{A}, d_{\mathcal{M}}^p)$  as measuring the magnitude of the nonlinear spectral gap of  $\mathbf{A}$  with respect to the kernel  $d_{\mathcal{M}}^p : \mathcal{M} \times \mathcal{M} \rightarrow [0, \infty)$ .

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  $\gamma(\mathbf{A}, d_{\mathcal{M}}^p) \leq c_n(\mathcal{M})^p \gamma(\mathbf{A}, d_{\mathcal{N}}^p)$  for every two metric spaces  $(\mathcal{M}, d_{\mathcal{M}})$  and  $(\mathcal{N}, d_{\mathcal{N}})$ , every matrix  $\mathbf{A}$  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.

$$\sqrt{\gamma(\mathbf{A}, d_{\mathcal{M}}^2)} \leq c_2(\mathcal{M}) \sqrt{\gamma(\mathbf{A}, \|\cdot\|_H^2)}. \quad (68)$$

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

**Theorem 45.**  $\begin{matrix} o & (X, \|\cdot\|_X) & B & (H, \|\cdot\|_H) & R & o \\ M \in (0, \infty) & \mathbf{A} & o & o & \lambda_2(\mathbf{A}) \leq 1 - M^2/c_2(X)^2 \end{matrix}$

$$\sqrt{\gamma(\mathbf{A}, \|\cdot\|_X^2)} \lesssim \frac{\log(M+1)}{M} c_2(X) \sqrt{\gamma(\mathbf{A}, \|\cdot\|_H^2)}. \quad (69)$$

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

$$\gamma(\mathbf{A}, \|\cdot\|_X^2) \lesssim \left( \frac{\log(c_2(X) \sqrt{1 - \lambda_2(\mathbf{A})}) + 1}{1 - \lambda_2(\mathbf{A})} \right)^2,$$

which is how Theorem 45 was stated in [?]. Note that (69) coincides (up to the implicit constant factor) with

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

**Corollary 47.** *Let  $n, r \geq 3$  and  $G = (\{1, \dots, n\}, E_G)$  be an  $r$ -regular graph. Let  $f : \{1, \dots, n\} \rightarrow X$  be a function. Then*

$$\left( \frac{1}{|E_G|} \sum_{\{i,j\} \in E_G} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \leq \alpha \quad \text{and} \quad \left( \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \geq \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_G(i, j), \quad (73)$$

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

$$c \in (0, \infty)$$

*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  $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n d_G(i, j) \gtrsim \log_r n$ .  $\square$

For every integer  $r \geq 3$  there exist arbitrarily large  $r$ -regular graphs  $G$  with  $\lambda_2(G) = 1 - \Omega(1)$ ; see [?] for this and much more on such graphs. Corollary 47 shows that the shortest-path metric on any such graph with  $r = O(1)$  satisfies the conclusion of Theorem 19, because the  $\alpha$ -Lipschitz assumption of Theorem 19 implies the first inequality in (73) and the assumption  $\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_X \geq \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_G(i, j)$  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  $G = (\{1, \dots, n\}, E_G)$  is a  $O(1)$ -spectral expander, i.e., it is  $O(1)$ -regular and  $\lambda_2(G) \leq 1 - \Omega(1)$ , and one finds vectors  $x_1, \dots, x_n$  in a normed space  $(X, \|\cdot\|_X)$  for which the averages  $\frac{1}{|E_G|} \sum_{\{i,j\} \in E_G} \|x_i - x_j\|_X^2$  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  $\frac{1}{|E_G|} \sum_{\{i,j\} \in E_G} d_G(i, j)^2 = 1$  and  $\frac{1}{n^2} \sum_{i,j=1}^n d_G(i, j)^2$ , 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  $\dim(X) \geq n^{\Omega(1)}$ .

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

$$\left( \frac{1}{|E_G|} \sum_{\{i,j\} \in E_G} \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \leq 1 \quad \text{and} \quad \left( \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \geq \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \omega(d_G(i, j)), \quad (74)$$

then the above argument applies mutatis mutandis to yield the conclusion

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

Indeed, the aforementioned counting argument shows that least 50% of the pairs  $(i, j) \in \{1, \dots, n\}^2$  satisfy  $d_G(i, j) \gtrsim \log_r n$ . 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  $\lambda_2(G) = 1 - \Omega(1)$ ), in the sense that there is a normed space  $(X, \|\cdot\|_X)$  for which (74) holds and

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

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

$$\left( \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_X^2 \right)^{\frac{1}{2}} \geq \frac{1}{\alpha} \left( \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_G(i, j)^2 \right)^{\frac{1}{2}} \gtrsim \frac{\log_r n}{\alpha} \asymp \omega(\text{diam}(G)) \geq \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \omega(d_G(i, j)).$$

(70) Let  $C \in (0, \infty)$  be the implicit universal constant in (70). Then

$$\begin{aligned} \left(\sum_{i=1}^n\sum_{j=1}^n\pi_i\pi_j\|f(i)-f(j)\|_X^2\right)^{\frac{1}{2}} &\stackrel{(67)}{\leqslant}\sqrt{\gamma(\mathbf{A},\|\cdot\|_X^2)}\left(\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|f(i)-f(j)\|_X^2\right)^{\frac{1}{2}} \\ &\stackrel{(71)}{\leqslant}\alpha\sqrt{\gamma(\mathbf{A},\|\cdot\|_X^2)}\stackrel{(70)}{\leqslant}\frac{C\alpha\log(\mathsf{c}_2(X)+1)}{1-\lambda_2(\mathbf{A})}, \end{aligned}\tag{77}$$

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

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

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

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.** *Let  $n, r \geqslant 3$ . Let  $\mathsf{G} = (\{1, \dots, n\}, \mathsf{E}_{\mathsf{G}})$  be a graph with  $r$  edges and let  $\mathbf{I} = (X, \|\cdot\|_X)$  be a normed space. Suppose that*

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



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

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

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

Denote  $\gamma = \gamma(G, \|\cdot\|_X^2)$ . For  $i \in \{1, \dots, n\}$  write

$$\mathcal{U}_i = f^{-1}\left(B_X(f(i), \alpha\sqrt{2\gamma})\right) = \left\{j \in \{1, \dots, n\} : \|f(i) - f(j)\|_X \leq \alpha\sqrt{2\gamma}\right\}. \quad (82)$$

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

$$\begin{aligned} n^2 \gamma \alpha^2 &\stackrel{(67) \wedge (79)}{\geq} \sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_X^2 \geq \sum_{i=1}^n \sum_{j \in \{1, \dots, n\} \setminus \mathcal{U}_i} \|f(i) - f(j)\|_X^2 \\ &\stackrel{(82)}{>} \sum_{i=1}^n (\alpha\sqrt{2\gamma})^2 (n - |\mathcal{U}_i|) \geq 2n\gamma\alpha^2(n - |\mathcal{U}_m|). \end{aligned} \quad (83)$$

This simplifies to  $|\mathcal{U}_m| \geq \frac{1}{2}n$ . Also, since  $\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_X^2 \geq \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_G(i, j)^2 \gtrsim (\log_r n)^2$ , 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)  $2\rho$ -separated subset of  $\mathcal{U}_m$ . Then  $\mathcal{U}_m \subseteq \cup_{i \in \mathcal{N}_{2\rho}} B_G(i, 2\rho)$ , where  $B_G(i, 2\rho)$  denotes the ball centered at  $i$  of radius  $2\rho$  in the shortest-path metric  $d_G$ . Since  $G$  is  $r$ -regular, for each  $i \in \{1, \dots, n\}$  we have the (crude) bound  $|B_G(i, 2\rho)| \leq 2r^{2\rho}$ . Hence,  $\frac{1}{2}n \leq |\mathcal{U}_m| \leq 2r^{2\rho}|\mathcal{N}_{2\rho}|$ . 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  $\|f(i) - f(j)\|_X \geq d_G(i, j) \geq 2\rho$ , 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  $\|f(i) - f(m)\|_X \leq \alpha\sqrt{2\gamma}$  (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.

$$\begin{aligned} (\alpha\sqrt{2\gamma} + \rho)^k \text{vol}_k(B_X(0, 1)) &= \text{vol}_k(B_X(f(m), \alpha\sqrt{2\gamma} + \rho)) \geq \text{vol}_k\left(\bigcup_{i \in \mathcal{N}_{2\rho}} B_X(f(i), \rho)\right) \\ &= \sum_{i \in \mathcal{N}_{2\rho}} \text{vol}_k(B_X(f(i), \rho)) = \rho^k \text{vol}_k(B_X(0, 1)) |\mathcal{N}_{2\rho}| \gtrsim \rho^k \text{vol}_k(B_X(0, 1)) \sqrt{n}. \end{aligned}$$

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

**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  $\alpha$ -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  $p \geq 1$  and a metric space  $(M, d_M)$ . Let  $L_p(\pi; M)$  be the metric space  $(M^n, d_{L_\bullet(\pi; M)})$ , where  $d_{L_\bullet(\pi; M)} : M^n \times M^n \rightarrow [0, \infty)$  is

$$\forall x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in M^n, \quad d_{L_\bullet(\pi; M)}(x, y) \stackrel{\text{def}}{=} \left( \sum_{i=1}^n \pi_i d_M(x_i, y_i)^p \right)^{\frac{1}{p}}.$$

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 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  $(x, x, \dots, x) \in \mathcal{M}^n$ .

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

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

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

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

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

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

**Lemma 53.** Let  $(\mathcal{M}, d_{\mathcal{M}})$  be a metric space,  $n \in \mathbb{N}$ ,  $p \in [1, \infty)$ ,  $\delta \in [0, 1]$ ,  $L$  a linear operator on  $L_p(\pi; \mathcal{M})$ , and  $\pi = (\pi_1, \dots, \pi_n)$  a probability measure on  $\{1, \dots, n\}$ . Let  $\mathbf{A}, \mathbf{B} \in \mathbf{M}_n(\mathbb{R})$  be matrices. Then for any  $x \in L_p(\pi; \mathcal{M}) \setminus \mathcal{M}$ ,

- (1)  $\mathcal{R}(x; \delta \mathbf{A} + (1 - \delta) \mathbf{B}, d_{\mathcal{M}}^p) = \delta \mathcal{R}(x; \mathbf{A}, d_{\mathcal{M}}^p) + (1 - \delta) \mathcal{R}(x; \mathbf{B}, d_{\mathcal{M}}^p)$
- (2)  $\mathcal{R}(x; (1 - \delta) \text{Id}_n + \delta \mathbf{A}, d_{\mathcal{M}}^p) = \delta \mathcal{R}(x; \mathbf{A}, d_{\mathcal{M}}^p)$  if  $\text{Id}_n \in \mathbf{M}_n(\mathbb{R})$
- (3)  $\mathcal{R}(x; \mathbf{A}, d_{\mathcal{M}}^p) \leq 2^p$
- (4)  $\mathcal{R}(x; \mathbf{AB}, d_{\mathcal{M}}^p)^{\frac{1}{p}} \leq \mathcal{R}(x; \mathbf{A}, d_{\mathcal{M}}^p)^{\frac{1}{p}} + \mathcal{R}(x; \mathbf{B}, d_{\mathcal{M}}^p)^{\frac{1}{p}}$
- (5)  $\mathcal{R}(x; \mathbf{A}^t, d_{\mathcal{M}}^p) \leq t^p \mathcal{R}(x; \mathbf{A}, d_{\mathcal{M}}^p)$  if  $t \in \mathbb{N}$

*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  $\mathcal{R}(x; \text{Id}_n, d_{\mathcal{M}}^p) = 0$ . The third assertion is justified by noting that by the triangle inequality, for every  $i, j, k \in \{1, \dots, n\}$  we have

$$d_{\mathcal{M}}(x_i, x_j)^p \leq (d_{\mathcal{M}}(x_i, x_k) + d_{\mathcal{M}}(x_k, x_j))^p \leq 2^{p-1} d_{\mathcal{M}}(x_i, x_k)^p + 2^{p-1} d_{\mathcal{M}}(x_k, x_j)^p. \quad (86)$$

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  $i, j, k \in \{1, \dots, n\}$  and using the fact that  $\mathbf{A}$  is reversible with respect to  $\pi$ , we get

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

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  $t-1$  times (with  $\mathbf{B}$  a power of  $\mathbf{A}$ ). To this end, writing  $\mathbf{A} = (a_{ij})$  and  $\mathbf{B} = (b_{ij})$ , we have

$$\begin{aligned} \left( \sum_{i=1}^n \sum_{j=1}^n \pi_i (\mathbf{AB})_{ij} d_{\mathcal{M}}(x_i, x_j)^p \right)^{\frac{1}{p}} &\leq \left( \sum_{i=1}^n \sum_{j=1}^n \pi_i \left( \sum_{k=1}^n a_{ik} b_{kj} \right) (d_{\mathcal{M}}(x_i, x_k) + d_{\mathcal{M}}(x_k, x_j))^p \right)^{\frac{1}{p}} \\ &\leq \left( \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \pi_i a_{ik} b_{kj} d_{\mathcal{M}}(x_i, x_k)^p \right)^{\frac{1}{p}} + \left( \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \pi_i a_{ik} b_{kj} d_{\mathcal{M}}(x_k, x_j)^p \right)^{\frac{1}{p}} \\ &= \left( \sum_{i=1}^n \sum_{k=1}^n \pi_i a_{ik} d_{\mathcal{M}}(x_i, x_k)^p \right)^{\frac{1}{p}} + \left( \sum_{j=1}^n \sum_{k=1}^n \pi_k b_{kj} d_{\mathcal{M}}(x_k, x_j)^p \right)^{\frac{1}{p}}, \end{aligned} \quad (87)$$

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  $\mu$  being the measure on  $\{1, \dots, n\}^3$  given by  $\mu(i, j, k) = \pi_i a_{ik} b_{kj}$  for all  $i, j, k \in \{1, \dots, n\}$ , and the final step of (87) uses the fact that  $\mathbf{A}$  and  $\mathbf{B}$  are both row-stochastic and reversible with respect to  $\pi$ .  $\square$

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.**  $\sum_{i=1}^n \pi_i x_i = 0$   $(H, \|\cdot\|_H)$   $x \in L_2(\pi; H) \setminus H$   $\mathcal{R}(x; \mathbf{A}^2, \|\cdot\|_H^2) \leq 1$

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

*Proof.* Let  $\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{R}$  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  $\mathbf{A}$  is row-stochastic, reversible relative to  $\pi$ , and  $\sum_{i=1}^n \pi_i x_i = 0$ , we get that

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n \pi_i (\mathbf{A}^2)_{ij} \|x_i - x_j\|_H^2 &= 2\|x\|_{L_2(\pi; H)}^2 - 2 \sum_{i=1}^n \pi_i \left\langle x_i, \sum_{j=1}^n (\mathbf{A}^2)_{ij} x_j \right\rangle \\ &= 2\|x\|_{L_2(\pi; H)}^2 - 2 \langle x, (\mathbf{A}^2 \otimes \text{Id}_H)x \rangle_{L_2(\pi; H)} = 2\|x\|_{L_2(\pi; H)}^2 - 2\|(\mathbf{A} \otimes \text{Id}_H)x\|_{L_2(\pi; H)}^2, \end{aligned}$$

and

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

Therefore, recalling the definition (84), we have

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

**Lemma 55** (point-wise Rayleigh quotient estimate for Hilbert isomorphs).  $L(X, \|\cdot\|_X)$   $\mathbf{d} \in [1, \infty)$   $\|\cdot\|_H : X \rightarrow [0, \infty)$   $\forall y \in X, \quad \|y\|_H \leq \|y\|_X \leq \mathbf{d}\|y\|_H.$

$$\forall y \in X, \quad \|y\|_H \leq \|y\|_X \leq \mathbf{d}\|y\|_H. \quad (88)$$

$x \in L_2(\pi; X) \setminus X$   $\mathbf{t}(x; \mathbf{A}) = \mathbf{t}(x; \mathbf{A}, \|\cdot\|_H, \mathbf{d})$   $\mathbf{t} \in \mathbb{N}$

$$\mathcal{R}\left(x; \left(\frac{1}{2}\text{Id}_n + \frac{1}{2}\mathbf{A}\right)^{2\mathbf{t}}, \|\cdot\|_H^2\right) \geq 1 - \frac{1}{4\mathbf{d}^2}, \quad (89)$$

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

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

$$\mathbf{B}_x \stackrel{\text{def}}{=} \left(\frac{1}{2}\text{Id}_n + \frac{1}{2}\mathbf{A}\right)^{\mathbf{t}(x; \mathbf{A})} \in \mathbf{M}_n(\mathbb{R}). \quad (91)$$

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

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

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

$$\frac{\|(\mathbf{B}_x \otimes \text{Id}_H)x\|_{L_2(\pi; H)}}{\|x\|_{L_2(\pi; H)}} \leq \sqrt{1 - \left(1 - \frac{1}{4\mathbf{d}^2}\right)} = \frac{1}{2\mathbf{d}}. \quad (92)$$

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

$$\frac{\|(\mathbf{B}_x \otimes \text{Id}_X)x\|_{L_2(\pi; X)}}{\|x\|_{L_2(\pi; X)}} \leq \mathbf{d} \frac{\|(\mathbf{B}_x \otimes \text{Id}_H)x\|_{L_2(\pi; H)}}{\|x\|_{L_2(\pi; H)}}. \quad (93)$$

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

$$\|x - (\mathbf{B}_x \otimes \text{Id}_X)x\|_{L_2(\pi; X)} \geq \|x\|_{L_2(\pi; X)} - \|(\mathbf{B}_x \otimes \text{Id}_X)x\|_{L_2(\pi; X)} \geq \frac{1}{2}\|x\|_{L_2(\pi; X)}. \quad (94)$$

Observe that

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

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,

$$\|x - (B_x \otimes \text{Id}_X) x\|_{L_2(\pi; X)} = \left( \sum_{i=1}^n \pi_i \left\| \sum_{j=1}^n (B_x)_{ij} (x_i - x_j) \right\|_X^2 \right)^{\frac{1}{2}} \leq \left( \sum_{i=1}^n \sum_{j=1}^n \pi_i (B_x)_{ij} \|x_i - x_j\|_X^2 \right)^{\frac{1}{2}}, \quad (96)$$

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

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

$$\mathcal{R}\left(x; \left(\frac{1}{2}\text{Id}_n + \frac{1}{2}\mathbf{A}\right)^{\mathbf{t}(x; \mathbf{A})}, \|\cdot\|_X^2\right) \stackrel{(91)}{=} \mathcal{R}(x; B_x, \|\cdot\|_X^2) \geq \frac{1}{16}. \quad (97)$$

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

$$1 \stackrel{(97)}{\lesssim} \mathcal{R}\left(x; \left(\frac{1}{2}\text{Id}_n + \frac{1}{2}\mathbf{A}\right)^{\mathbf{t}(x; \mathbf{A})}, \|\cdot\|_X^2\right) \leq \mathbf{t}(x; \mathbf{A})^2 \mathcal{R}\left(x; \frac{1}{2}\text{Id}_n + \frac{1}{2}\mathbf{A}, \|\cdot\|_X^2\right) = \frac{1}{2} \mathbf{t}(x; \mathbf{A})^2 \mathcal{R}(x; \mathbf{A}, \|\cdot\|_X^2), \quad (98)$$

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

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

**Lemma 56.**

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

*P oo* Since  $\mathbf{A}$  is row-stochastic,  $\lambda_n(\mathbf{A}) \geq -1$ . Therefore  $\frac{1}{2}\text{Id}_n + \frac{1}{2}\mathbf{A}$  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  $\frac{1}{2}\text{Id}_n + \frac{1}{2}\mathbf{A}$  on  $L_2^0(\pi)$  is  $\frac{1}{2} + \frac{1}{2}\lambda_2(\mathbf{A})$ , and therefore  $\|(\frac{1}{2}\text{Id}_n + \frac{1}{2}\mathbf{A})^{\mathbf{t}} u\|_{L_2(\pi)} \leq (\frac{1}{2} + \frac{1}{2}\lambda_2(\mathbf{A}))^{\mathbf{t}} \|u\|_{L_2(\pi)}$  for  $u \in L_2^0(\pi)$  and  $\mathbf{t} \in \mathbb{N}$ .

If  $x \in L_2(\pi; X) \setminus X$  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

$$\left( \frac{1}{2} \|\mathbf{A}\|, \frac{1}{2} \right)_{x; \frac{1}{2}}$$

5.1.1.  $\begin{matrix} o & o & o \\ & E & \\ & & o \end{matrix}$  Fix integers  $n, k, r \geq 3$  (think of  $n$  as much larger than  $k$ ). Let  $(X, \|\cdot\|_X)$  be a  $k$ -dimensional normed space. Suppose that  $G = (\{1, \dots, n\}, E_G)$  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  $(\{1, \dots, n\}, d_G)$  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  $\delta$ -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  $f : \{1, \dots, n\} \rightarrow X$  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  $\{1, \dots, n\}$  that corresponds to this dense ball and its complement

$\phi_A : L_2(\pi; X) \rightarrow L_2(\pi; H)$ . Nonlinear Rayleigh quotient inequalities of this type are proved in [?, ?], though the associated mappings  $\phi_A$  turn out to be highly nonlinear and quite complicated.<sup>9</sup>

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  $\mathcal{R}(x; A, \|\cdot\|_X^2)$  is small, then it follows that also  $\mathcal{R}(\phi_A(x); A, \|\cdot\|_H^2)$  is small. The  $\circ \circ \circ$  Cheeger's inequality (via examination of level sets of the second eigenvector) would now provide a sparse "spectral partition" of  $\{1, \dots, n\}$  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  $i \in \{1, \dots, n\}$  belongs depends only on whether the coordinate in question of the transformed vector  $\phi_A(x)_i \in H$  is above or below a certain value. If in addition  $(A, x) \mapsto \phi_A(x)$  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  $i \in \{1, \dots, n\}$  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  $\epsilon$ -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.  $\lambda_1$ , 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  $\ell_1$ . 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  $\ell_1$ . 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.

<sup>9</sup>Specifically, in [?] such a mapping  $\phi_A$  is constructed for Schatten-von Neumann trace classes using the Brouwer fixed-point theorem and estimates from [?]. In [?],  $\phi_A$  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  $\phi_A$ .

- [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  $\mathbf{R}^k$ . *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  $\ell_p$ , ( $1 \leq p \leq 2$ ), 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  $\ell_p$ . *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  $\ell_p$ -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  $H^\infty$ . *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  $l_N^\infty$ , 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  $\ell_\infty$  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.



the R. hv. red bling 401 - G.E.(G. J. and H. prin) 28 tion

- [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  $l_1$ . *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  $\mathbf{R}^k$ . 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 Aspects of Fractals*. Cambridge University Press, 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  $L_{\bullet}$ . 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  $A_\infty$ -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  $L_{\bullet}$  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  $\ell_1$ . *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  $L_{\bullet}$  and dimension reduction in  $L_1$ . *Geom. Funct. Anal.*, 14(4):745–747, 2004.
- [153] J. R. Lee and A. Naor.  $L_{\bullet}$  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](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  $L_{\bullet}$ -spaces and their applications. *Studia Math.*, 29:275–326, 1968.
- [157] J. Lindenstrauss and H. P. Rosenthal. The  $\mathcal{L}_{\bullet}$  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  $\ell_\infty$  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 fonctionnels. *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  $X_\infty$  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  $\lambda$ -approximations and some geometric applications. *SIAM J. Comput.*, 42(3):855–883, 2013.
- [204] G. Nöbeling. über eine  $n$ -dimensionale Universalmenge im  $R^{2^n+1}$ . *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. Pełczyński. *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 estadística matemática*, 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. Gribenval. Recipes for stable linear embeddings from Hilbert spaces to  $\mathbb{R}^n$ . *IEEE Trans. Inform. Theory*, 63(4):1271–1287, 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  $L_\infty$  in  $l^\infty$ . *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  $\ell_\infty^N$ . 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  $A_\infty$ -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  $L_\bullet$  in  $l_\bullet^N$ . 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  $S_\bullet(1 \leq p < \infty)$ . *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.

MAHEAIC DE A E S, P I CE U I E I S, P I CF, NE JE E 08544-1000, USA.  
 E-mail address: naor@math.princeton.edu