RANDOM ZERO SETS WITH LOCAL GROWTH GUARANTEES

ALAN CHANG, ASSAF NAOR, AND KEVIN REN

ABSTRACT. We prove that if $(\mathcal{M}, d)$ is an n-point metric space that embeds quasisymmetrically into a Hilbert space, then for every $\tau > 0$ there is a random subset $\mathcal{Z}$ of $\mathcal{M}$ such that for any pair of points $x, y \in \mathcal{M}$ with $d(x, y) \geqslant \tau$, the probability that both $x \in \mathcal{Z}$ and $d(y, \mathcal{Z}) \geqslant \beta\tau / \sqrt{1 + \log(|B(y, \kappa\beta\tau)|/|B(y, \beta\tau)|)}$ is $\Omega(1)$, where $\kappa > 1$ is a universal constant and $\beta > 0$ depends only on the modulus of the quasisymmetric embedding. The proof relies on a refinement of the Arora–Rao–Vazirani rounding technique. Among the applications of this result is that the largest possible Euclidean distortion of an n-point subset of ℓ_1 is $\Theta(\sqrt{\log n})$, and the integrality gap of the Goemans–Linial semidefinite program for the Sparsest Cut problem on inputs of size n is $\Theta(\sqrt{\log n})$. Multiple further applications are given.

CONTENTS

A. C. was supported by NSF grant DMS-2247233. A. N. was supported by NSF grant DMS-2054875, BSF grant 2018223, and the Simons Investigator award. K. R. was supported by an NSF GRFP fellowship.

An extended abstract announcing some of the algorithmic aspects of this work, and titled "Optimal rounding for Sparsest Cut," will appear in the proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC 2025).

1. INTRODUCTION

The field of *metric embeddings* is a mathematical discipline with interest in its own right, but it is also notable for its rich and multifaceted interactions with a broad swath of pure and applied mathematics, ranging from its origins in geometry and functional analysis, through its profound connections to group theory, topology, harmonic analysis, extension of functions, probability, combinatorics, statistics, approximation algorithms, computational complexity, and data structures. The main contribution of the present work resolves a longstanding question in metric embeddings, which has implications for several of the aforementioned directions. Given that the natural readership has varied backgrounds, the results that are obtained herein are presented below in a self-contained manner that introduces all of the relevant background, and explains in context the motivations and the meanings of what is accomplished. This inevitably leads to a longer and more gradual Introduction. We will therefore open with a brief section (Section 1.1) that quickly covers a selection of the main results assuming familiarity with some history, background, notation, and terminology, while postponing the presentation of precise technical details and the discussion of the history to subsequent sections. After this initial overview, we will restart the Introduction in Section 1.2 while presenting our results more formally and with relevant background, and we will also later describe (mainly in Section 2) the history and key ideas of our proofs.

1.1. **Informal overview.** Bourgain's embedding theorem [Bou85] asserts that any n-point metric space embeds into a Hilbert space with (bi-Lipschitz) distortion $O(\log n)$. This is optimal up to the value of the implicit universal constant, as discovered independently by Linial– London–Rabinovich [LLR95] and Aumann–Rabani [AR98]; the "culprits" which exhibit this optimality are n-vertex expander graphs.

When a metric space $\mathcal{M}$ does not admit a bi-Lipschitz embedding into a Hilbert space (which is typically the case), one can quantify the extent to which $\mathcal{M}$ is non-Euclidean by studying its Euclidean distortion growth sequence $\{\mathsf{c}_2^n(\mathcal{M})\}_{n=1}^\infty$, which is defined by letting $\mathsf{c}_2^n(\mathcal{M})$ be the supremum of the Euclidean distortion $\mathsf{c}_2(\mathcal{C})$ over all n-point subsets $\mathcal{C}$ of $\mathcal{M}$. The aforementioned embedding theorem of Bourgain says that $\mathsf{c}_2^n(\mathcal{M}) \lesssim \log n$ for every metric space $\mathcal{M}$. As any n-point metric space is isometric to a subset of ℓ_∞^n, the aforementioned nonembeddability result of Linial– London–Rabinovich and Aumann–Rabani says that $\mathsf{c}_2^n(\ell_\infty) \asymp \log n$. Of course, one also obviously has $\mathsf{c}_2^n(\ell_2) = 1$. Prior to the present work and despite major efforts over the past decades, the growth rate of $\mathsf{c}_2^n(\mathbf{X})$ as $n \to \infty$ was not known up to positive constant factors for *any* infinite dimensional normed space $\mathbf{X}$ other than the above trivial cases of ℓ_2 and ℓ_∞, and obvious variants thereof (namely, $\mathsf{c}_2^n(\mathbf{X}) = O(1)$ if (and only if) $\mathbf{X}$ is $O(1)$-isomorphic to a Hilbert space, and for $\mathsf{c}_2^n(\mathbf{X}) \asymp \log n$ to hold it suffices that $\{\ell_\infty^n\}_{n=1}^\infty$ embed into $\mathbf{X}$ with distortion $O(1)$, which is equivalent to $\mathbf{X}$ not having finite cotype by a theorem of Maurey–Pisier [MP73]).

Here we will prove that $\mathsf{c}_2^n(\ell_1) \asymp \sqrt{\log n}$. This resolves a famous folklore question in metric geometry and functional analysis that goes back to Enflo's seminal work [Enf69], though to the best of our knowledge it was first asked in print by Goemans [Goe97]. The new result herein is the positive embedding statement $\mathsf{c}_2^n(\ell_1) \lesssim \sqrt{\log n}$, while multiple n-point subsets of ℓ_1 have long been known to have Euclidean distortion at least a positive constant multiple of $\sqrt{\log n}$, including the Hamming cubes [Enf69] and the Laakso and diamond graphs [Laa00, LP01, NR03]; the fact that our estimate $\mathsf{c}_2^n(\ell_1) \lesssim \sqrt{\log n}$ is saturated by substantially different n-point subsets of ℓ_1 indicates the difficulty of proving such an embedding result, which, as we will soon explain, holds in much greater generality and thus provides a positive answer within a large class of metric spaces to an old and influential question of Johnson–Lindenstrauss [JL84] (reiterated in [Bou85, page 47]) on the validity of a natural nonlinear version of John's theorem [Joh48].

Given $0 < s, \varepsilon < 1$, say that a metric space $(\mathcal{M}, d_\mathcal{M})$ is (s,ε)-quasisymmetrically Hilbertian if there is a one-to-one function f from $\mathcal{M}$ to a Hilbert space $(\mathbf{H}, \|\cdot\|_\mathbf{H})$ such that $\|f(x)-f(y)\|_\mathbf{H} \leqslant (1-\varepsilon)\|f(x)-f(z)\|_\mathbf{H}$ for every $x,y,z \in \mathcal{M}$ satisfying $d_\mathcal{M}(x,y) \leqslant s d_\mathcal{M}(x,z)$. Call $\mathcal{M}$ quasisymetrically Hilbertian if there are $0 < s, \varepsilon < 1$ for which it is (s,ε)-quasisymmetrically Hilbertian. The above requirement is satisfied trivially for every metric space $\mathcal{M}$ when $\varepsilon = 0$ and any $s > 0$ (by assigning points in $\mathcal{M}$ to an orthonormal system). Thus, we are asking here for an embedding into a Hilbert space that is better—however slightly—than this trivial embedding, a requirement that holds when $\mathcal{M}$ is quasisymmetrically equivalent to a subset of a Hilbert space in the sense of Ahlfors–Beurling [BA56] and Tukia–Väisälä [TV80], which is a classical and extensively studied notion. The class of quasisymmetrically Hilbertian metric spaces encompasses disparate geometries, though not all metric spaces (e.g., any space that contains arbitrarily large expander graphs is not quasisymmetrically Hilbertian). One of our main embedding results (which includes the above statement on $\mathsf{c}_2^n(\ell_1)$ as a special case), is that if $\mathcal{M}$ is (s,ε)-quasisymmetrically Hilbertian, then $\mathsf{c}_2^n(\mathcal{M}) \lesssim_{s,\varepsilon} \sqrt{\log n}$. This more general setup is geometrically valuable in its own right, but also it includes all metric spaces of negative type, which is crucial for the following application to computer science.

The Sparsest Cut problem is a central and extensively studied topic in approximation algorithms due to its intrinsic interest, its deep connections to pure mathematics, and its varied use as a subroutine for many other algorithmic tasks [Shm96]. It takes as its input an integer n (the "universe size"), and two collections of symmetric nonnegative pairwise weights $\{C_{ij}\}_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,n\}}$ and $\{D_{ij}\}_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,n\}}$ (called "capacities" and "demands," respectively), and its goal is to compute (or approximate) in polynomial time the minimum of $(\sum_{i\in S}\sum_{j\in\{1,\ldots,n\}\smallsetminus S} C_{ij})/\sum_{i\in S}\sum_{j\in\{1,\ldots,n\}\smallsetminus S} D_{ij}$ over all possible nonempty proper subsets S of $\{1,\ldots,n\}$ (this ratio should be thought of as the total capacity that crosses the boundary of the bipartition $(S,\{1,\ldots,n\}\smallsetminus S)$ of $\{1,\ldots,n\}$, divided by the total demand that crosses its boundary). In the mid-1990s, Goemans and Linial (independently) proposed an algorithm for Sparsest Cut, called the Goemans–Linial semidefinite program, but despite intensive efforts prior to the present work the asymptotic growth rate as $n\to\infty$ (up to positive universal constant factors) of the performance of this famous algorithm remained unknown (no other algorithm is currently conjectured to have asymptotically better performance than the Goemans–Linial semidefinite program). Letting $\alpha_{\mathsf{GL}}(n)$ denote the smallest $\alpha\geqslant 1$ such that the Goemans–Linial semidefinite program outputs a number that is at most α times the aforementioned minimum of the ratio of total capacity to total demand across the boundary of a bipartition, we deduce herein that $\alpha_{\mathsf{GL}}(n)$ is bounded from above and from below by positive universal constant multiples of $\sqrt{\log n}$. Our new contribution is the positive algorithmic statement $\alpha_{\mathsf{GL}}(n)\lesssim\sqrt{\log n}$, as the matching impossibility result was already obtained (quite laboriously) in [NY18].

Prior to passing to an overview of our main geometric structural result, which answers a question from [ALN05], we will mention one more of its consequences; more applications will be described later in the Introduction. Gromov classically associated a quantity called the observable diameter to a metric probability space $(\mathcal{M},d_{\mathcal{M}},\mu)$. Briefly and informally (the precise definition is recalled in Section 1.3.2), the observable diameter quantifies the extent to which one could measure the size of the metric space using sufficiently smooth real-valued functions as observations, while accounting for possible observational errors by discarding at most a fixed fraction (with respect to the given probability measure μ) of the observations. Lévy's spherical isoperimetric theorem [Lév51] implies that the ratio between the observable diameter of the Euclidean n-sphere to its actual (metric) diameter is of order $1/\sqrt{n}$. Here we prove that the Euclidean n-sphere is extremal in this regard among all nondegenerate n-dimensional metric probability spaces that are quasisymmetrically Hilbertian (both "nondegenerate" and "n-dimensional" need to be suitably defined here; see Section 1.3.2 for the precise formulation), namely, up to dimension-independent positive constant factors, the Euclidean n-sphere has the smallest possible observable diameter among all such spaces (this is new even for arbitrary Borel probability measures on $\mathbb{R}^n$).

1.1.1. *Random zero sets.* Consider the unit integer grid in $\mathbb{R}^2$, colored black and white as a checker board. By randomly rotating and shifting this grid (say, rotating by a uniformly random angle and shifting by a vector chosen uniformly from a large bounded region), and then considering the union of all the black squares, one obtains a random subset of $\mathbb{R}^2$ which is an archetypical example of what is called (following the terminology of [ALN05]) a random zero set (at unit scale; by dilating the grid one obtains the corresponding object at any given scale). There are multiple ways to perform such constructions in $\mathbb{R}^n$ for any $n\in\mathbb{N}$ (e.g., flip an independent coin for each integer cell and consider the union of all of those cells whose outcome is "tails"). Another example of a useful random zero set for the Euclidean space $\mathbb{R}^n$ is as follows. Choose a direction $v \in S^{n-1}$ uniformly at random (i.e., v is distributed according to the normalized surface area measure on the unit Euclidean sphere in $\mathbb{R}^n$), and consider the slabs $\{\{x\in\mathbb{R}^n:\ k\leqslant \langle x,v\rangle < k+1\}:\ k\in\mathbb{Z}\}$. After coloring those slabs back and white in an alternating fashion, shifting them randomly in the direction of v and considering the union of the white slabs, one obtains another archetypical example of a random zero set. Constructions of this type are commonly used for organizing Euclidean point sets in harmonic analysis, geometry, algorithms and data structures.

The above procedures rely on properties of $\mathbb{R}^n$ that are not purely metric (coordinate structure, random rotations). We wish to obtain meaningful analogues of such random zero sets in an arbitrary quasisymmetrically Hilbertian metric measure space $(\mathcal{M}, d_{\mathcal{M}}, \mu)$. We will consider only finite $\mathcal{M}$ to circumvent measurability issues, and only μ of full support (e.g., the counting measure on $\mathcal{M}$). The most involved and innovative part of the present work is a construction when $\mathcal{M}$ is (s,ε)-quasisymmetrically Hilbertian for any scale $\tau>0$ a random subset $\mathscr{Z}_\tau$ of $\mathcal{M}$ such that for *every* pair of points $x,y\in\mathcal{M}$ with $d_{\mathcal{M}}(x,y)\geqslant\tau$ and any $0<\lambda\leqslant 1/2$, the probability that both x belongs to $\mathscr{Z}_\tau$ and the distance of y to $\mathscr{Z}_\tau$ is at least $\lambda\tau$ is greater than a positive universal constant multiple of $(\mu(B(y,19\beta\tau))/\mu(B(y,\beta\tau)))^{-\lambda^2}$, where $\beta=\beta(s,\varepsilon)$ depends only on s,ε and $B(y,r)$ denotes the closed ball in $\mathcal{M}$ of radius $r>0$ centered at $z\in\mathcal{M}$.

A crucial feature of the above random zero set is that the stated super-Gaussian probability lower bound depends (optimally) on the local growth of the measure μ near y at scale τ: If the measure of the ball $B(y,19\beta\tau)$ is not much larger than the measure of $B(y,\beta\tau)$, then the bound improves. This is critical for our applications as it leads to a cancellation that yields estimates that are sharp up to constant factors, while previously known bounds included a redundant lower order factor that tends to ∞ as $|\mathcal{M}|\to\infty$.

Random zero sets as above were previously available with the local growth $\mu(B(y,19\beta\tau))/\mu(B(y,\beta\tau))$ replaced by the larger quantity $\mu(\mathcal{M})/\min_{z\in\mathcal{M}}\mu(\{z\})$ by the deep work [ARV04] of Arora–Rao–Vazirani in combination with the clever use of duality by Chawla–Gupta–Räcke [CGR05]; see [ALN05]. The utility to metric embeddings of possibly improving this to random zero sets with the above local growth guarantee was realized long ago, due to its compatibility with the measured descent embedding method [KLMN04]. Indeed, the question whether one could construct such random zero sets was posed in [ALN05] while expressing skepticism that this is possible due to the nonlocal nature of the Arora–Rao–Vazirani reasoning. The fact that they do exist is an unexpected turn of events which leads to a conceptual change of perspective on a major line work in the literature on metric embeddings, as will be clarified in subsequent sections. Our proof introduces novel enhancements of the Arora–Rao–Vazirani approach to rounding semidefinite programs (via the perspective that has been subsequently developed by Rothvoss [Rot16]), and we expect that more applications of this will be found beyond those that are obtained herein.

1.2. **Detailed technical statements and background.** Unless stated otherwise, all the metric spaces that are discussed herein will be tacitly assumed to contain at least two points. Our main result is the following geometric structural theorem for metric spaces:

Theorem 1. *There is a universal constant $\kappa>1$ with the following property. Given $\eta:[0,\infty)\to[0,\infty)$ and a finite metric space $(\mathcal{M},d)$ that has a modulus-η quasisymmetric embedding into a Hilbert space, there is $\beta=\beta(\eta)>0$ that depends only on η such that for every nondegenerate measure μ on $\mathcal{M}$ and for every $\tau>0$ there exists a probability measure $\mathbb{P}=\mathbb{P}^{\tau,\mathcal{M}}$ on the nonempty subsets $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ of $\mathcal{M}$ that satisfies*[1]

$$\forall x,y\in\mathcal{M},\qquad d(x,y)\geqslant\tau \implies \mathbb{P}\left[\emptyset\neq\mathscr{Z}\subseteq\mathcal{M}:\ d(y,\mathscr{Z})\geqslant\frac{\beta\tau}{\sqrt{1+\log\frac{\mu(B(y,\kappa\beta\tau))}{\mu(B(y,\beta\tau))}}}\quad\text{and}\quad x\in\mathscr{Z}\right]\gtrsim 1. \qquad (1)$$

[1]We will use throughout the ensuing text the following (standard) conventions for asymptotic notation, in addition to the usual $O(\cdot),o(\cdot),\Omega(\cdot),\Theta(\cdot)$ notation. Given $a,b>0$, by writing $a\lesssim b$ or $b\gtrsim a$ we mean that $a\leqslant\kappa b$ for some universal constant $\kappa>0$, and $a\asymp b$ stands for $(a\lesssim b)\wedge(b\lesssim a)$. Thus, we are asserting in (1) that the probability is at least some positive universal constant. When we will need to allow for dependence on parameters, we will indicate it by subscripts. For example, in the presence of auxiliary objects q,U,ϕ, the notation $a\lesssim_{q,U,\phi} b$ means that $a\leqslant\kappa(q,U,\phi)b$, where $\kappa(q,U,\phi)>0$ may depend only on q,U,ϕ, and similarly for the notations $a\gtrsim_{q,U,\phi} b$ and $a\asymp_{q,U,\phi} b$. Thus, in Theorem 1 we can also write $\beta\asymp_\eta 1$.

The statement of Theorem 1 uses the following (standard) conventions for notation and terminology, which we will also maintain throughout what follows. Balls in a metric space $(\mathcal{M}, d_{\mathcal{M}})$ will always be closed balls, namely $B(x,r) = \{y \in \mathcal{M} : \ d_{\mathcal{M}}(x,y) \leqslant r\}$ for $x \in \mathcal{M}$ and $r \geqslant 0$. The distance of a point $y \in \mathcal{M}$ from a nonempty subset $\mathcal{Z}$ of $\mathcal{M}$ is $d(y,\mathcal{Z}) = \inf_{z\in\mathcal{Z}} d_{\mathcal{M}}(y,z)$. A Borel measure on $\mathcal{M}$ is nondegenerate if $0 < \mu(B(x,r)) < \infty$ for every $x \in \mathcal{M}$ and every $r > 0$. Finally (and mainly), one says that $(\mathcal{M}, d_{\mathcal{M}})$ embeds quasisymmetrically into a metric space $(\mathcal{N}, d_{\mathcal{N}})$ if there is an injection $\varphi : \mathcal{M} \to \mathcal{N}$ and an increasing modulus $\eta : [0,\infty) \to [0,\infty)$ with $\lim_{s\to 0} \eta(s) = 0$, for which every distinct $x, y, z \in \mathcal{M}$ satisfy

$$\frac{d_{\mathcal{N}}(\varphi(x),\varphi(y))}{d_{\mathcal{N}}(\varphi(x),\varphi(z))} \leqslant \eta\left(\frac{d_{\mathcal{M}}(x,y)}{d_{\mathcal{M}}(x,z)}\right). \tag{2}$$

One calls such φ a modulus-η quasisymmetric embedding of $\mathcal{M}$ into $\mathcal{N}$.

The notion of quasisymmetric embedding can be traced back to the classical work [BA56], though it has been formally introduced in [TV80]. This important concept has been investigated extensively; see for examples the survey [Väi99] and the monograph [Hei01] for an indication of the large body of work on this topic. It suffices to say here that (2) encodes the fact that φ preserves the "thinness" of triangles, and hence it roughly preserve "shape." In contrast, a bi-Lipschitz embedding (see (20) below) roughly preserves "size," hence a fortiori it preserves shape. The requirement that $(\mathcal{M}, d_{\mathcal{M}})$ embeds quasisymmetrically into $(\mathcal{N}, d_{\mathcal{N}})$ is therefore quite weak,[2] so Theorem 1 applies to a rich class of metric spaces, including notably those which have a snowflake that admits a bi-Lipschitz embedding into a Hilbert space, and in particular metric spaces of negative type (we will discuss these examples further below).

Theorem 1 settles a problem that was considered by experts (for reasons that will soon become evident) ever since the appearance of the Arora–Rao–Vazirani rounding algorithm [ARV04] and the measured descent embedding technique [KLMN04]. The question whether Theorem 1 holds was first posed in the literature in [ALN05] (specifically, see the paragraph just before Section 3 there), though [ALN05] expresses implicit doubt that Theorem 1 could be valid. The availability of Theorem 1 is indeed a somewhat surprising development as it leads to a change of perspective on (and a sharp improvement of) a line of work that straddles metric geometry, functional analysis and computer science. We will return to this in Section 2, which is devoted to a discussion of the history and prior results that were obtained in this direction, and a description of the conceptual innovations that occur in our proof of Theorem 1. An explanation of the meaning and origin of (1), which is commonly called a "random zero set" requirement, appears in Section 1.3.2 and Section 2. We will next explain some corollaries of Theorem 1.

1.3. **Geometric and algorithmic consequences of Theorem 1.** Here, we will explain some applications of Theorem 1; doing this will also demonstrate why it was sought-after over the past 20 years by researchers in this area. Much of what we will describe belongs to the well-known intended consequences of this line of research (see Section 2 for the history). Other parts are indeed corollaries of Theorem 1 that have a short justification (modulo the available literature), but to the best of our knowledge they were not previously anticipated. We will recall all of the needed terminology when it will become relevant as we go along, thus making the ensuing discussion self-contained.

1.3.1. *Quasisymmetrically Hilbertian metric spaces.* If $(\mathcal{M}, d)$ is a metric space and $\{e_x\}_{x\in\mathcal{M}}$ denotes the standard coordinate basis of $\ell_2(\mathcal{M})$,[3] then for every increasing function $\eta : [0,\infty) \to [0,\infty)$ that satisfies

[2]As an indication of this, it has long been unknown (see [Väi99, problem 8.3.1]) whether any two separable infinite dimensional Banach spaces are quasisymmetrically equivalent to each other, i.e., there exists a quasisymmetric bijection between them, and this possibility has been ruled out only relatively recently [Nao12a]. As another illustration of the subtle restrictions that a quasisymmmetry can impose, by [DS89] there are two closed 4-dimensional manifolds that are homeomorphic but not quasisymmetrically homeomorphic. The discussion herein provides more obstructions to the existence of a quasisymmetry.

[3]The functional-analytic notation that is used herein is entirely standard, as in e.g. the classical treatise [LT77].

$\lim_{s\to 0^+}\eta(s)\geqslant 1$ and for every distinct $x,y,z\in\mathcal{M}$ we have

$$\frac{\|e_x-e_y\|_{\ell_2(\mathcal{M})}}{\|e_x-e_z\|_{\ell_2(\mathcal{M})}}=\frac{\sqrt{2}}{\sqrt{2}}=1\leqslant\lim_{s\to 0^+}\eta(s)<\eta\left(\frac{d(x,y)}{d(x,z)}\right).$$

This observation shows that the existence of an embedding for which (2) holds with $\mathcal{N}$ a Hilbert space is a vacuous requirement (it is always satisfied) whenever $\lim_{s\to 0^+}\eta(s)\geqslant 1$. Of course, $\lim_{s\to 0^+}\eta(s)=0$ is part of the definition of a quasisymmetric embedding, so this trivial setting is excluded by the assumption of Theorem 1. Nevertheless, our proof of Theorem 1 will show that its conclusion holds even if we require only that $\lim_{s\to 0^+}\eta(s)<1$. In other words, existence of an embedding into a Hilbert space that satisfies anything beyond the above trivial guarantee suffices for our purposes. It seems worthwhile to introduce the following terminology for this (perhaps somewhat deceptively) non-stringent requirement:

Definition 2 (quasisymmetrically Hilbertian metric space)**.** *For $0<s,\varepsilon<1$, say that a metric space $(\mathcal{M},d)$ is (s,ε)-quasisymmetrically Hilbertian if there is an injection $\varphi:\mathcal{M}\to\mathbf{H}$ into a Hilbert space $\mathbf{H}$ such that*

$$\forall x,y,z\in\mathcal{M},\qquad d(x,y)\leqslant sd(x,z)\implies\|\varphi(x)-\varphi(y)\|_{\mathbf{H}}\leqslant(1-\varepsilon)\|\varphi(x)-\varphi(z)\|_{\mathbf{H}}.\tag{3}$$

Say that $(\mathcal{M},d)$ is quasisymmetrically Hilbertian if there exists such a φ for some $0<s,\varepsilon<1$.

Using the terminology of Definition 2, we can now state the following strengthening of Theorem 1:

Theorem 3 (generalization of Theorem 1)**.** *There is a universal constant $\alpha>1$ with the following property. Given $0<s,\varepsilon\leqslant 1/2$, denote $\beta=s^{\alpha/\varepsilon}$. Suppose that a finite metric space $(\mathcal{M},d)$ is (s,ε)-quasisymmetrically Hilbertian and that μ is a nondegenerate measure on $\mathcal{M}$. Then, for every $\tau>0$ there is a probability measure $\mathbb{P}^\tau$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ such that for every $0<\lambda\leqslant 1/2$ and every $x,y\in\mathcal{M}$ with $d(x,y)\geqslant\tau$ we have*[4]

$$\mathbb{P}^\tau\left[\emptyset\neq\mathcal{Z}\subseteq\mathcal{M}:\ d(y,\mathcal{Z})\geqslant\lambda\beta\tau\quad\text{and}\quad x\in\mathcal{Z}\right]\gtrsim\left(\frac{\mu(B(y,19\beta\tau))}{\mu(B(y,\beta\tau))}\right)^{-\lambda^2}.\tag{4}$$

Theorem 1 coincides with the special case $\lambda\asymp 1/\sqrt{1+\log(\mu(B(y,19\beta\tau))/\mu(B(y,\beta\tau)))}$ of Theorem 3, now with κ and the dependence of β on η given explicitly. To the best of our knowledge, the distributional estimate (4) in the rest of the range $0<\lambda\leqslant 1/2$ was not previously conjectured in the literature.

In contrast to Theorem 3, for general metric spaces the best analogous result is the following theorem:

Theorem 4. *Let $(\mathcal{M},d)$ be a finite metric space and μ a nondegenerate measure on $\mathcal{M}$. For every $\tau>0$ there is a probability measure $\mathbb{P}^\tau$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ such that for every $0<\lambda\leqslant 1/8$ and every $x,y\in\mathcal{M}$ with $d(x,y)\geqslant\tau$,*

$$\mathbb{P}^\tau\left[\emptyset\neq\mathcal{Z}\subseteq\mathcal{M}:\ d(y,\mathcal{Z})\geqslant\lambda\tau\quad\text{and}\quad x\in\mathcal{Z}\right]\geqslant\frac{1}{4}\left(\frac{\mu(B(y,\frac{5}{8}\tau))}{\mu(B(y,\frac{1}{8}\tau))}\right)^{-8\lambda}.\tag{5}$$

Even though Theorem 4 does not appear in the literature, it follows from a generalization of the proof of the main result of [MN07], combined with an idea from [Rao99]; we will include the proof of Theorem 4 in Section 11. The fact that Theorem 4 cannot be improved follows from Remark 12 below.

In summary, Theorem 3 states that the additional information that $\mathcal{M}$ is quasisymmetrically Hilbertian can be used to improve the linear dependence on λ in the exponent in right hand side of (5) to the quadratic dependence on λ in the exponent in the right hand side of (4), which we will soon see is the best one can hope for.

[4] The constant 19 that appears in (4) is included for concreteness as it arises from the ensuing proof, but it is neither claimed to be sharp nor is it important for the applications of (4) that are obtained herein. The same is true for the constants in (5).

1.3.2. *Observable diameter, embeddings, nonlinear spectral gaps, and Lipschitz extension.* It is natural to start by recalling the following definition, which was introduced in [ALN05, Definition 2.2]:

Definition 5 (spreading random zero set)**.** *Fix $\zeta > 0$ and $0 < \delta \leqslant 1$. A finite metric space $(\mathcal{M}, d)$ is said to admit a random zero set which is ζ-spreading with probability δ if for every $\tau > 0$ there is a probability measure $\mathbb{P}^\tau$ on the nonempty subsets $2^{\mathcal{M}} \smallsetminus \{\emptyset\}$ of $\mathcal{M}$ such that*

$$\forall x, y \in \mathcal{M}, \qquad d(x,y) \geqslant \tau \implies \mathbb{P}^\tau \left[\emptyset \neq \mathcal{Z} \subseteq \mathcal{M} :\ d(y, \mathcal{Z}) \geqslant \frac{\tau}{\zeta} \quad \text{and} \quad x \in \mathcal{Z} \right] \geqslant \delta. \tag{6}$$

The requirement (6) of Definition 5 is simpler than the conclusion (1) of Theorem 1, which involves a spreading parameter $\zeta = \zeta(y,\tau)$ that depends on the point y and the distance scale τ, suitably encoding what one should think of as the "local growth rate of the measure μ near y at scale τ." In a similar vein, the requirement (6) of Definition 5 is simpler than the conclusion (4) of Theorem 3, which allows the lower bound δ on the "separation probability" to depend on the local growth rate of the measure μ near y at scale τ. Our first example of the utility of such more refined geometric information deduces the existence of a traditional random zero set in sense of Definition 5, when the metric space is doubling.

Given $K \in \mathbb{N}$, a metric space $(\mathcal{M}, d)$ is said to be K-doubling if for every $x \in \mathcal{M}$ and every $r \geqslant 0$ there are $x_1, \ldots, x_K \in \mathcal{M}$ such that $B(x, 2r) \subseteq B(x_1, r) \cup \cdots \cup B(x_K, r)$; note that necessarily $K \geqslant 2$ if $\mathcal{M}$ contains at least two points. By [VK84], if $\mathcal{C} \subseteq \mathcal{M}$ is finite (it suffices to only assume here is that $(\mathcal{C}, d)$ is complete [LS98]), then there is a measure μ whose support is equal to $\mathcal{C}$ such that for every $x \in \mathcal{C}$ and every $r \geqslant 0$,

$$0 < \mu\big(B(x, 2r)\big) \leqslant K^{O(1)} \mu\big(B(x, r)\big). \tag{7}$$

By applying Theorem 3 to a measure μ as in (7), we obtain the following statement:

Theorem 6. *There is a universal constant $\kappa > 1$ with the following property. Fix $0 < s, \varepsilon \leqslant \frac{1}{2}$, $p \geqslant 1$, $K \in \mathbb{N}$. Let $(\mathcal{M}, d)$ be a K-doubling metric space that is (s,ε)-quasisymmetrically Hilbertian. Then, every finite subset of $\mathcal{M}$ admits a random zero set that is $(1/s)^{\kappa/\varepsilon}\sqrt{\max\{1, (\log K)/p\}}$-spreading with probability $e^{-\kappa p}$.*

Proof of Theorem 6 assuming Theorem 3 . If $\mathcal{C} \subseteq \mathcal{M}$ is finite, then use [VK84] to get a doubling measure μ satisfying (7) whose support is equal to $\mathcal{C}$. Let $\alpha > 1$ and $\beta = s^{\alpha/\varepsilon}$ be as in Theorem 3. By iterating (7) five times one sees that $\mu(B(y, 19\beta\tau)) \leqslant \mu(B(y, 2^5\beta\tau)) \leqslant K^{O(1)}\mu(B(y, \beta\tau))$ for every $\tau > 0$ and $y \in \mathcal{C}$. Theorem 3 therefore provides a probability measure $\mathbb{P}^\tau$ on $2^{\mathcal{C}} \smallsetminus \{\emptyset\}$ and a universal constant $\gamma > 1$ such that for every $\lambda \in [1/\sqrt{\log K}, 1/2]$ and every $x, y \in \mathcal{C}$ with $d(x,y) \geqslant \tau$ we have

$$\mathbb{P}^\tau \left[\emptyset \neq \mathcal{Z} \subseteq \mathcal{C} :\ x \in \mathcal{Z} \quad \text{and} \quad d(y, \mathcal{Z}) \geqslant \lambda\beta\tau \right] \geqslant K^{-\gamma\lambda^2}. \tag{8}$$

If $1 \leqslant p \leqslant (\log K)/4$, i.e., $\sqrt{p/\log K} \in [1/\sqrt{\log K}, 1/2]$, we may use (8) with $\lambda = \sqrt{p/\log K}$ to see that $\mathbb{P}^\tau$ is ζ-spreading with probability $e^{-\gamma p}$, where $\zeta = 1/(\lambda\beta) = (1/s)^{\alpha/\varepsilon}\sqrt{(\log K)/p}$. If $p > (\log K)/4$, then use (8) with $\lambda = 1/2$ to get that $\mathbb{P}^\tau$ is ζ-spreading with probability $K^{-\gamma/4} > e^{-\gamma p}$, where $\zeta = 2/\beta = 2(1/s)^{\alpha/\varepsilon}$. The desired conclusion now follows provided $\kappa > 1$ is a sufficiently large universal constant. □

We will next describe some consequences of Theorem 6, starting with resolving a problem that was left open in [NRS05]. The geometric phenomenon that [NRS05] aimed to establish is:

> *"Up to dimension-independent constant factors, the n-dimensional Euclidean sphere has the smallest possible observable diameter among all the normalized $e^{O(n)}$-doubling metric probability spaces that are quasisymmetrically Hilbertian."*

For a metric space $(\mathcal{M}, d)$, a Borel probability measure μ on $\mathcal{M}$, and $\theta > 0$ the θ-observable diameter

$$\mathrm{ObsDiam}_\mu^{(\mathcal{M},d)}(\theta)$$

of the metric measure space $(\mathcal{M}, d, \mu)$ is defined [Gro07] to be the supremum over all possible 1-Lipschitz functions $f : \mathcal{M} \to \mathbb{R}$ of the infimum of $\mathrm{diam}_{\mathbb{R}}(f(\mathcal{S})) = \sup\{|f(x) - f(y)| :\ x, y \in \mathcal{S}\}$ over all possible Borel subsets $\mathcal{S} \subseteq \mathcal{M}$ with $\mu(\mathcal{S}) \geqslant 1-\theta$. If the metric is clear from the context, then we will use the notation

$$\mathrm{ObsDiam}_{\mu}^{\mathcal{M}}(\theta) = \mathrm{ObsDiam}_{\mu}^{(\mathcal{M},d)}(\theta).$$

The observable diameter is extensively-studied; see e.g. [Led01, Gro07, Shi16, BF22]. Its qualitative interpretation (e.g. [Ber00, page 336]) is to view each real-valued 1-Lipschitz function as a way to "observe" the size of a given metric space while accounting for possible observational errors by discarding at most a θ-fraction (with respect to a given probability measure) of the observations. As explained in [Gro07], for every $n \in \mathbb{N}$ and every $0<\theta<1$, the θ-observable diameter of the unit Euclidean sphere in $\mathbb{R}^n$ equipped with its normalized surface area measure is at most $1/\sqrt{n}$ times a positive factor that depends only on θ.

For $n \in \mathbb{N}$, denote the standard unit Euclidean sphere in $\mathbb{R}^n$ by $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. Throughout what follows, S^{n-1} will be equipped with the metric that is inherited from ℓ_2^n, and σ^{n-1} will denote the normalized surface-area (probability) measure on S^{n-1}. Since ℓ_2^n is K_n-doubling for some $K_n \leqslant 5^n$ (e.g. [Sem01, Section 2.2]; a better upper bound on K_n appears in [VG05]), also $S^{n-1} \subseteq \ell_2^n$ is 5^n-doubling.

If $(\mathcal{M},d)$ is a metric space and μ is a Borel probability measure on $\mathcal{M}$, then we say that μ is normalized if the median of $d(x,y)$ is equal to 1 when $(x,y) \in \mathcal{M}\times\mathcal{M}$ is distributed according to $\mu\times\mu$. Note that σ^{n-1} is not normalized per this terminology, but one could normalize it by rescaling the metric by a factor that is bounded from above and from below by positive universal constants.

Theorem 7. *For every $0<s,\varepsilon<1$ and every $n\in\mathbb{N}$, if $(\mathcal{M},d)$ is a metric space that is (s,ε)-quasisymmetrically Hilbertian and 5^n-doubling, then every normalized Borel probability measure μ on $\mathcal{M}$ satisfies*

$$\forall 0<\theta\leqslant\theta_0,\qquad \mathrm{ObsDiam}_{\mu}^{\mathcal{M}}(\theta) \gtrsim_{s,\varepsilon} \mathrm{ObsDiam}_{\sigma^{n-1}}^{S^{n-1}}(\theta), \tag{9}$$

where $\theta_0>0$ is a universal constant.

Theorem 7 establishes that the above extremal property of the observable diameter of the Euclidean sphere indeed holds; its (simple) derivation from Theorem 6 appears in Section 9.

It would be worthwhile to investigate the extent to which the phenomenon that is expressed in Theorem 7 holds for metric spaces that need not be quasisymmetrically Hilbertian. For example, we do not know if when $2<p<\infty$, any $e^{O(n)}$-doubling subset $\mathcal{M}$ of ℓ_p satisfies (9) (for any Borel probability measure μ on $\mathcal{M}$), with the implicit dependence on s,ε now replaced by a dependence on p; if this is not the case, then it would be interesting to determine the asymptotic rate (as $n\to\infty$) at which the smallest possible observable diameter of such spaces tends to 0. Theorem 7 does not cover this setting because ℓ_p is not quasisymmetrically Hilbertian when $p>2$; this follows from [Nao12a], as explained in Remark 56 below. It is simple to see that some assumption on the $e^{O(n)}$-doubling metric space is needed for the conclusion (9) of Theorem 6 to hold with a dimension-independent constant factor (i.e., Theorem 6 does not hold for subsets of ℓ_∞); indeed, expander graphs demonstrate this, as explained in Remark 66 below.

Theorem 7 is a manifestation of the following impossibility result for super-Gaussian concentration of any quasisymmetrically Hilbertian doubling metric probability space. Given a metric space $(\mathcal{M},d)$ and a Borel probability measure μ on $\mathcal{M}$, the isoperimetric (or concentration) function of the metric probability space $(\mathcal{M},d,\mu)$ is defined as follows, where $\mathcal{B}or(\mathcal{M},d)\subseteq 2^{\mathcal{M}}$ denotes the Borel subsets of $\mathcal{M}$:

$$\forall t\geqslant 0,\qquad I_{\mu}^{(\mathcal{M},d)}(t) \stackrel{\mathrm{def}}{=} \sup_{\substack{\emptyset\neq\mathcal{C}\in\mathcal{B}or(\mathcal{M},d)\\ \mu(\mathcal{C})\geqslant\frac12}} \mu\big(\{x\in\mathcal{M}:\ d(x,\mathcal{C})\geqslant t\}\big). \tag{10}$$

We will use the notation $I_{\mu}^{\mathcal{M}}(t) = I_{\mu}^{(\mathcal{M},d)}(t)$ when the metric is clear from the context.

Concentration of measure is related to the observable diameter through the following relations, which hold for every complete metric space $(\mathcal{M},d)$ and every Borel probability measure μ on $\mathcal{M}$; their simple derivations can be found in e.g. [Led01, Proposition 1.12] and [Shi16, Remark 2.28].

$$\forall\theta,t>0,\qquad I_{\mu}^{\mathcal{M}}(t)>\theta \implies \mathrm{ObsDiam}_{\mu}^{\mathcal{M}}(\theta)\geqslant t \qquad\text{and}\qquad I_{\mu}^{\mathcal{M}}(t)\leqslant\theta \implies \mathrm{ObsDiam}_{\mu}^{\mathcal{M}}(2\theta)\leqslant 2t. \tag{11}$$

The solution [Lév51] of the isoperimetric problem for the Euclidean sphere implies that for any $n\in\mathbb{N}$,

$$\forall\, t>0,\qquad I_{\sigma^{n-1}}^{S^{n-1}}(t)\lesssim e^{-O(n)t^2}. \tag{12}$$

Alternatively, by the solution [Bor75, ST78] of the isoperimetric problem for the Gaussian measure, if we rescale the metric on ℓ_2^n by a factor of order $1/\sqrt{n}$ so that the standard Gaussian measure on ℓ_2^n will be normalized, the isoperimetric function of the resulting metric probability space is also bounded from above by the right hand side of (12). More such examples are available in the literature; see e.g. [Mau79, Gro80, MS86, GM87], and a systematic treatment can be found in [Led01]. Inspired by [Cha14],[5] we say that a metric probability space exhibits the super-concentration phenomenon if its isoperimetric function is asymptotically smaller (as $n\to\infty$) than the right hand side of (12). This phenomenon is useful when available, as demonstrated in [Cha14], but Theorem 8 below, which is also a simple consequence of Theorem 6 that we will derive in Section 9, asserts the following "no measure is super-concentrated" phenomenon: If $(\mathcal{M},d)$ is a quasisymmetrically Hilbertian metric space that is $e^{O(n)}$-doubling, then $(\mathcal{M},d,\mu)$ does not have super-concentration for any Borel probability measure μ on $\mathcal{M}$ whatsoever.

Theorem 8. *For every $0<s,\varepsilon<1$ there exists $c=c(s,\varepsilon)>0$ with the following property.*[6] *If $n\in\mathbb{N}$ and $(\mathcal{M},d)$ is a metric space that is (s,ε)-quasisymmetrically Hilbertian and 5^n-doubling, then every normalized Borel probability measure μ on $\mathcal{M}$ satisfies*

$$\forall\, 0<\phi\leqslant 1,\qquad I_\mu^{\mathcal{M}}(c\phi)\gtrsim e^{-n\phi^2}. \tag{13}$$

Following [Rab08, Nao21a], given $p,D\geqslant 1$ we say that a metric space $(\mathcal{M},d_{\mathcal{M}})$ embeds with p-average distortion D into a metric space $(\mathcal{N},d_{\mathcal{N}})$ if for every Borel probability measure μ on $\mathcal{M}$ there are $s>0$ and $f=f_\mu:\mathcal{M}\to\mathcal{N}$ that is sD-Lipschitz, namely, $d_{\mathcal{N}}(f(x),f(y))\leqslant sDd_{\mathcal{M}}(x,y)$ for every $x,y\in\mathcal{M}$, such that

$$\bigg(\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{N}}(f(x),f(y))^p\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac1p}\geqslant s\bigg(\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^p\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac1p}.$$

If this occurs for $p=1$, then one omits p by saying that $\mathcal{M}$ embeds with average distortion D into $\mathcal{N}$, and if this occurs for $p=2$, then one says that $\mathcal{M}$ embeds with quadratic average distortion D into $\mathcal{N}$.

By substituting Theorem 6 into [Nao14, Proposition 7.1], and then substituting the resulting statement into [Nao21a, Lemma 60], we arrive at the following corollary of Theorem 1:

Theorem 9. *There exists a universal constant $\kappa>1$ with the following property. Suppose that $K\in\mathbb{N}$ and $0<s,\varepsilon\leqslant 1/2$. Let $(\mathcal{M},d)$ be a K-doubling metric space that is (s,ε)-quasisymmetrically Hilbertian. Then, for every $p\geqslant 1$ one can embed $(\mathcal{M},d)$ into the real line $\mathbb{R}$ with p-average distortion at most*

$$\Big(\frac1s\Big)^{\frac{\kappa}{\varepsilon}}\max\bigg\{\sqrt{\frac{\log K}{p}},1\bigg\}. \tag{14}$$

The upper bound (14) on the average distortion in Theorem 9 is new even for subsets of a Hilbert space.[7] Furthermore, Theorem 9 is optimal in the entire range of possible values of $K\in\mathbb{N}$ and $p\geqslant 1$, up to the dependence on s,ε (which we did not investigate in the present work), as seen by the Fact 10 below, which we will prove in Section 12; this also demonstrates the optimality of Theorem 3.

Fact 10. *For $n\in\mathbb{N}$ and $p\geqslant 1$, the smallest $D\geqslant 1$ such that ℓ_2^n embeds into $\mathbb{R}$ with p-average distortion D is bounded from above and from below by positive universal constant multiples of $\sqrt{\max\{1,n/p\}}$.*

[5]In [Cha14], even weaker estimates about variance rather than concentration are said to be super-concentrated.

[6]An inspection of the ensuing proofs reveals that if $0<s,\varepsilon\leqslant\frac12$, then both the constant c in (13) and the implicit constant factor in (9) can be taken to be a positive universal constant multiple of $s^{\kappa/\varepsilon}$, where κ is the constant in Theorem 6.

[7]Nevertheless, we expect that for the (very) special case of subsets of Hilbert space one could prove Theorem 9 using more standard (and simpler) chaining-style reasoning à la Fernique–Talagrand, along the lines of the proof in [IN07] which treats a related, but different, setting. The main value of Theorem 9 is its applicability to metric spaces that are far from Euclidean.

By standard volumetric reasoning [CW71, Hei01], every n-dimensional normed space is K_n-doubling for some $K_n = e^{O(n)}$. As the quantity $\sqrt{\max\{1, n/p\}}$ in Fact 10 is bounded from above and from below by positive universal constant multiples of $\sqrt{\max\{1,(\log K_n)/p\}}$, Fact 10 shows that Theorem 9 is sharp.

Theorem 11 below shows that Theorem 9 is, in fact, sharp in the following much stronger sense: Even if one relaxes the stringent goal of Theorem 9 to embed $\mathcal{M}$ into $\mathbb{R}$ by allowing the embedding to be into any Banach space of type 2, then it is still impossible to do so with p-average distortion that is asymptotically better (as $p, K \to \infty$) than what is provided by Theorem 9. For this (as well as for ensuing discussions), we recall the (standard) terminology that a Banach space $(\mathbf{X}, \|\cdot\|)$ is said [HJ74, MP76] to have (Rademacher) type $1 \leqslant r \leqslant 2$ if there is $T > 0$ such that for every $n \in \mathbb{N}$, every $x_1, \ldots, x_n \in \mathbf{X}$ satisfy

$$\mathbb{E}\Big[\Big\|\sum_{i=1}^{n} \varepsilon_i x_i\Big\|^r\Big] \leqslant T^r \sum_{i=1}^{n} \|x_i\|^r, \tag{15}$$

where the expectation in (15) is with respect to i.i.d. symmetric Bernoulli random variables $\varepsilon_1, \ldots, \varepsilon_n$, i.e., they are independent and $\Pr[\varepsilon_i = 1] = \Pr[\varepsilon_i = -1] = 1/2$ for every $i \in [n] = \{1, \ldots, n\}$. The infimum over those T for which the above requirement holds is denoted $T_r(\mathbf{X})$. We then have, for example, $T_q(L_q) \asymp 1$ when $1 \leqslant q \leqslant 2$, and $T_2(L_q) \asymp \sqrt{q}$ when $q \geqslant 2$, as explained in, say, Chapter 6 of the textbook [AK16].

The proof of the following theorem was shown to us by Alexandros Eskenazis; we thank him for allowing us to include it in Section 12, which also treats target spaces of type r for any $1 \leqslant r \leqslant 2$.

Theorem 11. *Suppose that $(\mathbf{X}, \|\cdot\|)$ is a Banach space of type 2. For every $n \in \mathbb{N}$ and $p \geqslant 1$, if ℓ_1^n embeds into $\mathbf{X}$ with p-average distortion D, then necessarily $D \gtrsim \sqrt{n}/\max\{\sqrt{p}, T_2(\mathbf{X})\}$.*

Theorem 11 implies the aforementioned strong optimality of Theorem 9 as ℓ_1 is an instance of a metric space that embeds quasisymmetrically into a Hilbert space (two different examples of closed-form formulae of embeddings that exhibit this appear in e.g. [MN04, Remark 5.10] and [Nao10, Section 3]).

Remark 12. *By repeating the above reasoning mutatis mutandis while using Theorem 4 instead of Theorem 3, one sees that for every $p \geqslant 1$, any finite K-doubling metric space admits a random zero set that is $O(\max\{1,(\log K)/p\})$-spreading with probability e^{-p}. Consequently, any K-doubling metric space embeds with p-average distortion $O(\max\{1,(\log K)/p\})$ into $\mathbb{R}$. This cannot be improved even if one allows the embedding to be into any target normed space $\mathbf{X}$ that has type $r > 1$. Indeed, by combining the deep result of* [Laf09] *with the first bound in equation (131) of* [Nao21a] *and the reasoning in* [LLR95, Mat97], *one sees that for arbitrarily large $n \in \mathbb{N}$ there is an n-point metric space such if it embeds with p-average distortion D into $\mathbf{X}$, then necessarily $D \gtrsim_{r, T_r(\mathbf{X})} \max\{1, (\log n)/p\}$. It remains to note that any n-point metric space is (vacuously) n-doubling. Theorem 9 thus shows that for doubling metric spaces that are also quasisymmetrically Hilbertian the aforementioned upper bound on the p-average distortion into $\mathbb{R}$ can be (sharply) improved to the square root of the best that one can hope for in the setting of arbitrary metric spaces.*

Given $n \in \mathbb{N}$, let $\triangle^{n-1} = \{\pi = (\pi_1, \ldots, \pi_n) \in [0,1]^n : \sum_{i=1}^n \pi_i = 1\}$ denote the simplex of probability measures on $[n]$. When we say that a matrix $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$ is stochastic we always mean row-stochastic, i.e., $(a_{i1}, \ldots, a_{in}) \in \triangle^{n-1}$ for every $i \in \{1, \ldots, n\}$. Given $\pi \in \triangle^{n-1}$, a stochastic matrix $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$ is π-reversible if $\pi_i a_{ij} = \pi_j a_{ji}$ for every $i, j \in \{1, \ldots, n\}$. In this case, A is a self-adjoint contraction on $L_2(\pi)$ and the decreasing rearrangement of the eigenvalues of A is denoted $1 = \lambda_1(\mathsf{A}) \geqslant \ldots \geqslant \lambda_n(\mathsf{A}) \geqslant -1$. The spectral gap of A is $1 - \lambda_2(\mathsf{A})$. The conductance (with respect to A, π) of a subset $\emptyset \neq S \subsetneq [n]$ is

$$\Phi_{\mathsf{A}}(S) \stackrel{\mathrm{def}}{=} \frac{1}{\pi(S)\pi([n] \smallsetminus S)} \sum_{i \in S} \sum_{j \in [n] \smallsetminus S} \pi_i a_{ij},$$

where $\pi(S) = \sum_{i \in S} \pi_i$. The Cheeger constant of A is defined by

$$h(\mathsf{A}) \stackrel{\mathrm{def}}{=} \min_{\emptyset \neq S \subsetneq [n]} \Phi_{\mathsf{A}}(S).$$

For $p \geqslant 1$, the p-Poincaré constant (reciprocal of the p-spectral gap) of A with respect to a metric space $(\mathcal{M}, d)$, denoted $\gamma(\mathsf{A}, d^p)$, is the infimum over $\gamma \in [0,\infty]$ such that for every $x_1, \ldots, x_n \in \mathcal{M}$ we have:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i \pi_j d(x_i, x_j)^p \leqslant \gamma \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i a_{ij} d(x_i, x_j)^p. \tag{16}$$

The works [Mat97, Laf08, NS11, MN14, Nao14, Nao18, ANN+18a, dLdlS21, Nao21a, Esk22, Nao24b] include information on nonlinear spectral gaps and their applications. Because it is simple to check that $\gamma(\mathsf{A}, d_{\mathbb{R}}^2) = 1/(1-\lambda_2(\mathsf{A}))$ and $\gamma(\mathsf{A}, d_{\mathbb{R}}) = 1/h(\mathsf{A})$, where $d_{\mathbb{R}}$ is the standard metric on $\mathbb{R}$, Theorem 9 implies:

Theorem 13. *Fix $n, K \in \mathbb{N}$ and $0 < s, \varepsilon \leqslant 1/2$. Suppose that $(\mathcal{M}, d)$ is a metric space that is both K-doubling and (s,ε)-quasisymmetrically Hilbertian. Then, every stochastic matrix $\mathsf{A} \in \mathsf{M}_n(\mathbb{R})$ satisfies*

$$\gamma(\mathsf{A}, d) \lesssim_{s,\varepsilon} \frac{\sqrt{\log K}}{h(\mathsf{A})} \qquad \text{and} \qquad \gamma(\mathsf{A}, d^2) \lesssim_{s,\varepsilon} \frac{\log K}{1-\lambda_2(\mathsf{A})}. \tag{17}$$

Furthermore, the implicit constants in (17) *can be taken to be* $(1/s)^{O(1/\varepsilon)}$.

Theorem 13 complements the bounds in [Nao14, Section 7.1.1] by showing that they can be improved if one considers quasisymmetrically Hilbertian doubling metric spaces rather than arbitrary doubling metric spaces; in the latter case, the dependence on K in the two estimates in (17) can be taken to be, respectively, $\log K$ and $(\log K)^2$, and this is optimal. The second inequality in (17) is optimal, as seen by taking $\mathsf{A} = \mathsf{Q}_n$ to be the transition matrix of the standard random walk on the Hamming cube, because it is straightforward to check that Q_n satisfies the following bounds:

$$\frac{1}{1-\lambda_1(\mathsf{Q}_n)} \asymp n \qquad \text{yet} \qquad \gamma(\mathsf{Q}_n, d_{\ell_1^n}^2) \asymp n^2.$$

Whether the first inequality in (17) can be improved is perhaps the most important question that remains open in the context of the present work; see Question 14 below. The best lower bound that is currently available is the example in [KM13], which shows that one cannot replace the $\sqrt{\log K}$ term in (17) by anything smaller than $\exp(c\sqrt{\log\log K})$ for some universal constant $c > 0$. If it were possible to improve the dependence on K in the first inequality in (17), then this would have major algorithmic implications; we will explain this matter after presenting the relevant algorithmic setup in Section 1.3.3.

Question 14. *For $0 < s, \varepsilon \leqslant 1/2$ and $K \in \mathbb{N}$, what is the growth rate as $K \to \infty$ (up to constant factors that depend on s, ε) of the smallest $\alpha = \alpha(K, s, \varepsilon)$ such that $\gamma(\mathsf{A}, d) \leqslant \alpha / h(\mathsf{A})$ for every (s,ε)-quasisymmetrically Hilbertian K-doubling metric space $(\mathcal{M}, d)$, every $n \in \mathbb{N}$ and every stochastic matrix $\mathsf{A} \in \mathsf{M}_n(\mathbb{R})$?*

Among the consequences of the nonlinear spectral gap bounds of Theorem 13 is that they quickly imply the following bound on the cutting modulus, which is a parameter that was introduced in [ANN+18a] as a tool for hierarchically partitioning metric spaces. It was used in [ANN+18a] to construct a data structure for approximate nearest neighbor search; we will return to this algorithmic setting in Section 1.3.3.

Given $\Phi > 0$, the cutting modulus $\Xi_{\mathcal{M}}(\Phi)$ of a metric space $(\mathcal{M}, d)$ was defined in [ANN+18a] to be the infimum over those $\Xi \in (0,\infty]$ with the following property. For every $n \in \mathbb{N}$, every $r > 0$, every $\pi \in \Delta^{n-1}$, every stochastic π-reversible matrix $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$, and every $x_1, \ldots, x_n \in \mathcal{M}$ that satisfy

$$\forall\, i, j \in [n], \qquad a_{ij} > 0 \implies d(x_i, x_j) \leqslant r, \tag{18}$$

at least one of the following two (non-dichotomic) scenarios must occur:

- Either there exists $\emptyset \neq S \subsetneq [n]$ such that $\Phi_{\mathsf{A}}(S) \leqslant \Phi$,
- or, there exists $i \in [n]$ such that $\pi\big(\{j \in [n] :\ d(x_i, x_j) \leqslant \Xi r\}\big) \geqslant \frac{1}{2}$.

Theorem 15. *Fix $K \in \mathbb{N}$ and $0 < s, \varepsilon \leqslant 1$. Suppose that $(\mathcal{M}, d)$ is a metric space that is both K-doubling and (s,ε)-quasisymmetrically Hilbertian. Then, for every $\Phi > 0$ we have*

$$\Xi_{\mathcal{M}}(\Phi) \lesssim_{s,\varepsilon} \frac{\sqrt{\log K}}{\Phi}.$$

Proof of Theorem 15 assuming Theorem 13. Fix $\Phi > 0$ and take any $0 < \Xi < \Xi_{\mathcal{M}}(\Phi)$. By the definition of $\Xi_{\mathcal{M}}(\Phi)$ as an infimum, this means that there are $n \in \mathbb{N}$ and $r > 0$ for which we can find a probability vector $\pi \in \Delta^{n-1}$, a stochastic π-reversible matrix $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$, and points $x_1, \ldots, x_n \in \mathcal{M}$ that satisfy (18) yet

$$\min_{\emptyset \neq S \subsetneq [n]} \Phi_{\mathsf{A}}(S) > \Phi \qquad \text{and} \qquad \min_{j \in [n]} \pi\big(\{j \in [n] :\ d(x_i, x_j) > \Xi r\}\big) > \frac{1}{2}. \tag{19}$$

The first assertion in (19) is the same as saying that $h(\mathsf{A}) > \Phi$. Hence, $\gamma(\mathsf{A}, d) \lesssim_{s,\varepsilon} \sqrt{\log K}/\Phi$ by the first inequality in (17). An application of the definition of $\gamma(\mathsf{A}, d)$ to the points $x_1, \ldots, x_n$ now shows that

$$\frac{\Xi r}{2} \overset{(19)}{\leqslant} \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i \pi_j d(x_i, x_j) \lesssim_{s,\varepsilon} \frac{\sqrt{\log K}}{\Phi} \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i a_{ij} d(x_i, x_j) \overset{(18)}{\leqslant} \frac{\sqrt{\log K}}{\Phi} r. \qquad \square$$

Remark 16. *The above proof of Theorem 15 used the first inequality in* (17). *One can also deduce Theorem 15 from Theorem 13 by repeating this reasoning mutatis mutandis while using the second inequality in* (17) *and incorporating Cheeger's inequality for Markov chains* [JS88, LS88]. *Indeed, the cutting modulus was bounded in* [ANN+18a] *using bounds on quadratic nonlinear spectral gaps in this way. However, it is worthwhile to deduce Theorem 15 using only the first inequality in* (17) *since the second inequality in* (17) *is sharp yet the first inequality in* (17) *could possibly be improved, as expressed in Question 14. Algorithmic ramifications of this possibility will be discussed in Section 1.3.3.*

Following [LLR95], the (bi-Lipschitz) distortion of a finite metric space $(\mathcal{M}, d_{\mathcal{M}})$ in an infinite metric space $(\mathcal{N}, d_{\mathcal{N}})$, which is denoted $\mathsf{c}_{(\mathcal{N}, d_{\mathcal{N}})}(\mathcal{M}, d_{\mathcal{M}})$ or simply $\mathsf{c}_{\mathcal{N}}(\mathcal{M})$ when the underlying metrics are clear from the context, is the infimum over those $D > 0$ for which there are $f : \mathcal{M} \to \mathcal{N}$ and $s > 0$ such that

$$\forall\, x, y \in \mathcal{M}, \qquad s d_{\mathcal{M}}(x, y) \leqslant d_{\mathcal{N}}\big(f(x), f(y)\big) \leqslant D s d_{\mathcal{M}}(x, y). \tag{20}$$

If $p \geqslant 1$ and $\mathcal{N}$ is an infinite dimensional $L_p(\mu)$ space, then $\mathsf{c}_{L_p(\mu)}(\mathcal{M})$ does not depend on μ (in fact, this holds if $\dim(L_p(\mu)) \geqslant |\mathcal{M}|(|\mathcal{M}| - 1)/2$ by [Bal90]), so one commonly uses the shorter notation:

$$\mathsf{c}_p(\mathcal{M}) \overset{\mathrm{def}}{=} \mathsf{c}_{L_p(\mu)}(\mathcal{M}). \tag{21}$$

The especially important and useful parameters $\mathsf{c}_2(\mathcal{M})$ and $\mathsf{c}_1(\mathcal{M})$ are naturally called, respectively, the Euclidean distortion of $\mathcal{M}$ and the L_1 distortion of $\mathcal{M}$.

The classical Fréchet embedding $\Phi_{(\mathcal{M}, d_{\mathcal{M}})}$ of $\mathcal{M}$ into the set $\mathbb{R}^{2^{\mathcal{M}} \smallsetminus \{\emptyset\}}$ of real-valued functions on the nonempty subsets of $\mathcal{M}$ is defined by setting for every $x \in \mathcal{M}$ and $\emptyset \neq \mathcal{Z} \subseteq \mathcal{M}$,

$$\Phi_{(\mathcal{M}, d_{\mathcal{M}})}(x)(\mathcal{Z}) \overset{\mathrm{def}}{=} d_{\mathcal{M}}(x, \mathcal{Z}). \tag{22}$$

Given $p, D \geqslant 1$, one says that $(\mathcal{M}, d_{\mathcal{M}})$ embeds into L_p with distortion D via the Fréchet embedding if there exists a probability measure $\mathbb{P}$ on $2^{\mathcal{M}} \smallsetminus \{\emptyset\}$ such that for every $x, y \in \mathcal{M}$ we have

$$d_{\mathcal{M}}(x, y) \leqslant D \|\Phi_{(\mathcal{M}, d_{\mathcal{M}})}(x) - \Phi_{(\mathcal{M}, d_{\mathcal{M}})}(y)\|_{L_p(\mathbb{P})}. \tag{23}$$

For example, the famous embedding of [Bou85] is of this form. This terminology is consistent with what we recalled above because (23) implies that $\mathsf{c}_p(\mathcal{M}) \leqslant D$. Indeed, contrast (23) with the trivial estimate

$$\|\Phi_{(\mathcal{M}, d_{\mathcal{M}})}(x) - \Phi_{(\mathcal{M}, d_{\mathcal{M}})}(y)\|_{L_p(\mathbb{P})} \leqslant \|\Phi_{(\mathcal{M}, d_{\mathcal{M}})}(x) - \Phi_{(\mathcal{M}, d_{\mathcal{M}})}(y)\|_{L_\infty(\mathbb{P})} \leqslant d_{\mathcal{M}}(x, y), \tag{24}$$

where the first step of (24) holds as $\mathbb{P}$ is a probability measure and the second step of (24) is a straightforward consequence of the triangle inequality for $d_{\mathcal{M}}$. However, knowing that $(\mathcal{M}, d_{\mathcal{M}})$ embeds into L_p with distortion D via the Fréchet embedding provides more information than mere embeddability into L_p, which is sometimes needed in applications (e.g. [FHL08, MM16]); furthermore, the former embedding may not be possible when it is known that the latter embedding does exist (e.g. [MR01, BLMN06]).

The measured descent embedding technique [KLMN05] yields the following result:

Theorem 17 (special case of measured descent [KLMN05]). *Fix $n\in\mathbb{N}$, $\alpha\geqslant\beta>0$, and $0<\varepsilon,\delta,\theta\leqslant 1$. Let $(\mathcal{M},d)$ be an n-point metric space with the property that for every $\tau>0$ there is a probability measure $\mathbb{P}^\tau$ on the nonempty subsets of $\mathcal{M}$ such that for every $x,y\in\mathcal{M}$ that satisfy $\tau\leqslant d(x,y)\leqslant(1+\theta)\tau$ we have*

$$\mathbb{P}^\tau\left[\emptyset\neq\mathcal{Z}\subseteq\mathcal{M}:\ d(y,\mathcal{Z})\geqslant\frac{\varepsilon\tau}{\sqrt{1+\log\frac{|B(y,\alpha\tau)|}{|B(y,\beta\tau)|}}}\quad\text{and}\quad x\in\mathcal{Z}\right]\geqslant\delta. \tag{25}$$

Then,

$$\mathsf{c}_2(\mathcal{M})\lesssim_{\varepsilon,\delta,\theta,\alpha,\beta}\sqrt{\log n}. \tag{26}$$

Furthermore, the bound (26) *on the Euclidean distortion of $\mathcal{M}$ is obtained via the Fréchet embedding.*

Even though those who are familiar with [KLMN05] will recognize Theorem 17 as a special case of measured descent, Theorem 17 does not appear in [KLMN05] as a standalone statement. Instead, Theorem 17 follows from part of[8] the proof of Lemma 1.8 in [KLMN05]: the random zero set appears in the second displayed equation on page 847 of [KLMN05] and the corresponding Fréchet embedding appears in the line immediately following it. As Theorem 17 is crucial for the applications of Theorem 17 that we will explain next, we will include in Section 10 a self-contained proof of Theorem 17 which builds on the above ideas while incorporating further enhancements so as to yield both a more general statement and the best dependence that we currently have of the implicit constant factor in (26) on $\varepsilon,\delta,\theta,\alpha,\beta$; see (287).

In the influential work [JL84], Johnson and Lindenstrauss asked if $\mathsf{c}_2(\mathcal{M})\lesssim\sqrt{\log n}$ for any n-point metric space $(\mathcal{M},d)$. If this were true, then it would have been a satisfactory analogue of John's theorem [Joh48] that $\mathsf{c}_2(\mathbf{X})\leqslant\sqrt{\dim\mathbf{X}}$ for any finite-dimensional normed space $\mathbf{X}$, in accordance with the predictions of the Ribe program [Bou86, Kal08, Nao12b, Bal13, Ost13, God17, Nao18]. Bourgain famously answered [Bou85] the aforementioned Johnson–Lindenstrauss question negatively; an asymptotically stronger (sharp) impossibility result was subsequently obtained (by another method) in [LLR95, AR98]. Hence, a positive answer would necessitate imposing restrictions on the metric space $(\mathcal{M},d)$. The following theorem follows by substituting Theorem 1 into Theorem 17. It answers the Johnson–Lindenstrauss problem positively within the class of quasisymmetrically Hilbertian metric spaces.

Theorem 18 (Johnson–Lindenstrauss problem/nonlinear John theorem). *For every $0<s,\varepsilon\leqslant 1/2$ and $n\in\mathbb{N}$, if $(\mathcal{M},d)$ is an n-point (s,ε)-quasisymmetrically Hilbertian metric space, then $\mathsf{c}_2(\mathcal{M})\lesssim_{s,\varepsilon}\sqrt{\log n}$.*

The class of metric spaces that admit a quasisymmetric embedding into a Hilbert space encompasses metric spaces of finite Assouad–Nagata dimension [LS05], or more generally [NS11] metric spaces that admit a padded stochastic decomposition (thus, it includes, say, planar graphs [KPR93] and doubling metric spaces [Ass83, GKL03]), as well as $L_p(\mu)$ spaces for $1\leqslant p\leqslant 2$ [BDCK66, WW75], and the infinite dimensional Heisenberg group $\mathbb{H}^\infty$ [LN06].

Remark 19. *Ignoring much of the information that is provided by Theorem 9, one can consider the real line as a subset of ℓ_2 to deduce that for any $K\in\mathbb{N}$ and $0<s,\varepsilon\leqslant 1/2$, a K-doubling (s,ε)-quasisymmetrically Hilbertian metric space embeds into ℓ_2 with average distortion $O(\sqrt{\log K})$, which is sharp*[9] *by Fact 10. This nonlinear version of John's theorem is quite satisfactory for the following reasons. Firstly, it does not hold for general K-doubling metric spaces, for which the best bound that one can get on the Euclidean average distortion is $O(\log K)$, as follows from the proofs in* [LLR95, AR98, Mat97]*; a stronger impossibility result appears in Remark 12. Secondly, unlike Theorem 18, such a result does not hold for bi-Lipschitz*

[8] The proof of [KLMN05, Lemma 1.8] derives further facts that are relevant to the setup of [KLMN05], but not for Theorem 17.

[9] We warn of the following somewhat confusing pitfall here. Assouad famously proved [Ass83] that any doubling metric space embeds quasisymmetrically into ℓ_2 (even into $\mathbb{R}^n$ for some $n\in\mathbb{N}$). So, superficially it might seem that our assumption that the space is quasisymmetrically Hilbertian is redundant. However, Assouad's theorem yields a quasisymmetric embedding whose modulus (necessarily) depends on the doubling constant.

embeddings, so one must somehow relax the requirement from the embedding, as we did here by considering average distortion. Indeed, by [Laa00, LP01] *there exists a doubling metric space $\mathcal{M}$ that embeds quasisymmetrically into ℓ_2 by* [Ass83]*, yet $\mathcal{M}$ does not admit any bi-Lipschitz embedding into ℓ_2; the subsequent works* [CK06, LN06]*, which are natural extensions to infinite dimensional targets of the important contributions* [Pan89, Sem96]*, show that one can also take $\mathcal{M}$ here to be the Heisenberg group.*

A longstanding open question asks for the growth rate of the largest possible Euclidean distortion of a finite subset of ℓ_1. This question arose from Enflo's influential proof [Enf69] that the d-dimensional Hamming cube $\mathcal{C}_d = \{0,1\}^d \subseteq \ell_1^d$ has Euclidean distortion $\sqrt{d}$, i.e., it is of order $\sqrt{\log n}$ for $n = 2^d = |\mathcal{C}_d|$.

To the best of our knowledge, the aforementioned question remained a famous unpublished folklore problem for many years, though it was eventually published by Goemans in [Goe97, page 157], who stated the elegant conjecture that for every $n \in \mathbb{N}$, any n-point subset of ℓ_1 embeds into ℓ_2 with distortion $O(\sqrt{\log n})$; see also Linial's Open Problem 4 in [Lin02]. By contrasting the conjectural asymptotic upper bound in [Goe97] with the lower bound of [Enf69], it would then follow that the d-dimensional Hamming cube has the asymptotically largest possible growth rate among all subsets of ℓ_1 of size 2^d.

As an indication of the difficulty here, we mention that arbitrarily large n-point subsets of ℓ_1 that are different from the Hamming cube have been shown to have Euclidean distortion $\Theta(\sqrt{\log n})$. E.g., these can be planar graphs, which can even be taken to be $O(1)$-doubling [Laa00, LP01, NR03]. Thus, the above question has asymptotic extremizers that are markedly different from each other.

Theorem 18 provides the following positive resolution of the above conjecture:

Theorem 20 (finite subsets of ℓ_1 that are furthest from being Euclidean)**.** *For every $n \in \mathbb{N}$, the maximal Euclidean distortion of an n-point subset of ℓ_1 is bounded from above and from below by positive universal constant multiples of $\sqrt{\log n}$.*

For $0 < \theta \leqslant 1$ and a metric space $(\mathcal{M}, d)$, it is common to call (in reference to the Koch snowflake; see e.g. [DS97]) the metric space $(\mathcal{M}, d^\theta)$ the θ-snowflake of $(\mathcal{M}, d)$, or simply the θ-snowflake of $\mathcal{M}$ if the metric is clear from the context. If the $\frac{1}{2}$-snowflake of $\mathcal{M}$ embeds isometrically into a Hilbert space, then one says that $\mathcal{M}$ is a metric space of negative type; see e.g. the monograph [DL10] or the survey [Nao10] for more on this important and useful notion, including the reason for the nomenclature.

In [Goe97, page 158], Goemans conjectured that for every $n \in \mathbb{N}$, any n-point metric space of negative type embeds into ℓ_2 with distortion $O(\sqrt{\log n})$. By definition, a metric space of negative type has a modulus-η quasisymmetric emebdding into ℓ_2 with $\eta(t) = \sqrt{t}$ for every $t \geqslant 0$, and with (2) holding as equality. Hence, Theorem 18 provides the following positive resolution of Goemans' conjecture (the corresponding lower bound follows by considering the Hamming cube $\{0,1\}^d \subseteq \ell_1^d$ again and using [Enf69], as its image in ℓ_2^d under the formal identity mapping exhibits that it is a metric space of negative type):

Theorem 21 (finite metrics of negative type that are furthest from being Euclidean)**.** *For every $n \in \mathbb{N}$, the largest possible Euclidean distortion of an n-point metric space of negative type is bounded from above and from below by positive universal constant multiples of $\sqrt{\log n}$.*

In fact, we also deduce from Theorem 18 the following result that answers the major problem of evaluating the growth rate of the largest possible L_1 distortion of a finite metric space of negative type (this formulation appears as Open Problem 3 in [Lin02], and see also e.g. [Goe97, Mat02]):

Theorem 22 (finite metrics of negative type that are furthest from subsets of L_1)**.** *For every $n \in \mathbb{N}$, the largest possible L_1 distortion of an n-point metric space of negative type is bounded from above and from below by positive universal constant multiples of $\sqrt{\log n}$.*

The upper bound on the distortion in Theorem 22 follows from Theorem 18 since $\mathsf{c}_1(\mathcal{M}) \leqslant \mathsf{c}_2(\mathcal{M})$ for every finite metric space $(\mathcal{M}, d_\mathcal{M})$; there are multiple ways to justify this, but perhaps the quickest (though overkill) is to apply Dvoretzky's theorem [Dvo61]. The matching lower bound is due to [NY18], where it is exhibited by considering a sufficiently dense net in the unit ball of the 5-dimensional Heisenberg group $\mathbb{H}^5$, equipped with a suitable equivalent metric of negative type that was constructed in [LN06] (by [NY22], the 3-dimensional Heisenberg group $\mathbb{H}^3$ is insufficient for this purpose).

Remark 23. *Until its landmark negative answer in* [KV05]*, it was a major open question (due, independently, to Goemans and Linial) whether every metric space of negative type admits a bi-Lipschitz embedding into L_1. Theorem 22 completes the efforts by many researchers in mathematics and theoretical computer science (including* [LN06, ALN08, CGR08, KR09, CK10a, CK10b, CKN11, KV15, NY18]*, as well as multiple unpublished works) to understand the rate (for n-point metric spaces of negative type, as $n \to \infty$) of the failure of this possibility. It is worthwhile to recall in this context the (still open) conjecture of* [ANV10] *that any invariant metric of negative type on an Abelian group admits a bi-Lipschitz embedding into L_1.*

Thus far, we applied Theorem 1 by combining it with Theorem 17 without utilizing the further knowledge from Theorem 17 that the corresponding distortion bound (26) is obtained via the Fréchet embedding. We will next present an application to the Lipschitz extension problem that incorporates this additional information. In order to do so, we first need to recall basic notation that originates in [Mat90].

Suppose that $(\mathcal{S}, d_\mathcal{S})$ is a (source) metric space and $(\mathcal{T}, d_\mathcal{T})$ is a (target) metric space. For a nonempty subset $\mathcal{M}$ of $\mathcal{S}$, let $\mathsf{e}(\mathcal{S}, \mathcal{M}; \mathcal{T})$ denote the infimum over those $K \in [1,\infty]$ such that for every $L \geqslant 0$ and every L-Lipschitz function $f : \mathcal{M} \to \mathcal{T}$ there is a KL-Lipschitz function $F : \mathcal{S} \to \mathcal{N}$ that extends f.

The following observation is a straightforward consequence of the above definition:

Observation 24. *Fix $\alpha, \beta \geqslant 0$. Suppose that $(\mathcal{S}, d_\mathcal{S})$ is a metric space and that $\emptyset \neq \mathcal{M} \subseteq \mathcal{S}$. Let $(\mathcal{N}, d_\mathcal{N})$ be a metric space such that there is an α-Lipschitz function $\varphi : \mathcal{S} \to \mathcal{N}$ satisfying $d_\mathcal{N}(\varphi(x), \varphi(y)) \geqslant d_\mathcal{S}(x, y)/\beta$ for every $x, y \in \mathcal{M}$. Then, $\mathsf{e}(\mathcal{S}, \mathcal{M}; \mathcal{T}) \leqslant \alpha\beta \mathsf{e}(\mathcal{N}, \varphi(\mathcal{M}); \mathcal{T})$ for every metric space $(\mathcal{T}, d_\mathcal{T})$.*

The Lipschitz extension modulus $\mathsf{e}(\mathcal{S}; \mathcal{T})$ of a pair $(\mathcal{S}, \mathcal{T})$ of metric spaces is defined to be the supremum of $\mathsf{e}(\mathcal{S}, \mathcal{M}; \mathcal{T})$ over all the nonempty subsets $\mathcal{M}$ of $\mathcal{S}$. Analogously, following [LN04] we define the absolute Lipschitz extension modulus $\mathsf{ae}(\mathcal{M}; \mathcal{T})$ of a pair $(\mathcal{M}, \mathcal{T})$ of metric spaces to be the supremum of $\mathsf{e}(\mathcal{S}, \mathcal{M}; \mathcal{T})$ over all the possible metric spaces $\mathcal{S}$ that (isometrically) contain $\mathcal{M}$ as a subset.

The study of the above moduli has a rich history due to their intrinsic geometric and analytic interest as well as their applications and connections to multiple areas of mathematics; a (partial) description of this background appears in the monographs [BB12a, BB12b, CMN19], and in [MN13, NR17, Nao24a]. Nevertheless, basic problems about these moduli have stubbornly resisted many efforts, including the following example of a longstanding open question due to Ball [Bal92], to which we will return later in an algorithmic context, following the insights of Makarychev and Makarychev in [MM16]:

Question 25 (Ball's extension problem)**.** *Is it true that $\mathsf{e}(\ell_2; \ell_1) < \infty$?*

More generally, we pose the following question which we think is of fundamental importance:

Question 26. *Characterize the class $\mathcal{T}(\ell_2)$ of those target metric spaces $(\mathcal{T}, d_\mathcal{T})$ for which $\mathsf{e}(\ell_2; \mathcal{T}) < \infty$.*

Kirszbraun's theorem [Kir34] says that $\ell_2 \in \mathcal{T}(\ell_2)$; in fact, $\mathsf{e}(\ell_2; \ell_2) = 1$. Question 25 is of course the special case of Question 26 that asks whether $\ell_1 \in \mathcal{T}(\ell_2)$. One could also aim to characterize for any source metric space $(\mathcal{S}, d_\mathcal{S})$ the class $\mathcal{T}(\mathcal{S})$ of those target metric spaces $(\mathcal{T}, d_\mathcal{T})$ for which $\mathsf{e}(\mathcal{S}, \mathcal{T}) < \infty$. This seems to be a challenging though potentially fruitful research direction. As a start, we do not know if given two Banach spaces $\mathbf{X}, \mathbf{Y}$, the equality $\mathcal{T}(\mathbf{X}) = \mathcal{T}(\mathbf{Y})$ implies that $\mathbf{X}$ and $\mathbf{Y}$ are bi-Lipschitz equivalent.

When $\mathsf{e}(\mathcal{S}; \mathcal{T}) = \infty$, which is typically the case, one still studies the Lipschitz extension problem for the pair $(\mathcal{S}, \mathcal{T})$ by considering the asymptotic growth rate of the following finitary moduli (which are finite if a quite mild assumption on the target space $\mathcal{T}$ holds; see [JLS86] and Section 5.3 in [Nao24a]):

$$\forall n \in \mathbb{N}, \qquad \mathsf{e}_n(\mathcal{S}; \mathcal{T}) \stackrel{\mathrm{def}}{=} \sup_{\substack{\emptyset \neq \mathcal{M} \subseteq \mathcal{S} \\ |\mathcal{M}| \leqslant n}} \mathsf{e}(\mathcal{S}, \mathcal{M}; \mathcal{T}).$$

The introduction of [NR17] (particularly Section 1.3 there) surveys known bounds on these parameters.[10] The following consequence of Theorem 1 provides a modest amount of further information on this topic:

Corollary 27. *There exists a universal constant $C>0$ with the following property. Fix $0<s,\varepsilon\leqslant 1/2$ and an integer $n\geqslant 2$. Suppose that $(\mathcal{M},d_{\mathcal{M}})$ is an n-point metric space that is also (s,ε)-quasisymmetrically Hilbertian. Then, for every $0<\delta\leqslant 1$ and every target metric space $(\mathcal{T},d_{\mathcal{T}})$ we have*

$$\mathsf{ae}(\mathcal{M};\mathcal{T})\lesssim_{s,\varepsilon} e^{\frac{C}{\delta}}\,\mathsf{e}_n\big(\ell_2^{\lceil \delta\log n\rceil};\mathcal{T}\big)\sqrt{\log n}. \tag{27}$$

The main result of [JL84] is that $\mathsf{ae}(\mathcal{M};\ell_2)\lesssim\sqrt{\log n}$ for any n-point metric space $\mathcal{M}$. By Corollary 27 for $\delta=1$ we see that if $\mathcal{M}$ is also (s,ε)-quasisymmetrically Hilbertian, then $\mathsf{ae}(\mathcal{M};\mathcal{T})\lesssim_{s,\varepsilon,\mathcal{T}}\sqrt{\log n}$ for any metric space $(\mathcal{T},d_{\mathcal{T}})$ that belongs to the class $\mathcal{T}(\ell_2)$ from Question 26.

Proof of Corollary 27 assuming Theorem 1. Fix a metric space $(\mathcal{S},d_{\mathcal{S}})$ such that $\mathcal{M}\subseteq\mathcal{S}$ and the restriction of $d_{\mathcal{S}}$ to $\mathcal{M}\times\mathcal{M}$ coincides with $d_{\mathcal{M}}$. By combining Theorem 1 with Theorem 17 there is a probability measure $\mathbb{P}$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ such that the Fréchet embedding $\Phi=\Phi_{(\mathcal{M},d_{\mathcal{M}})}:\mathcal{M}\to L_2(\mathbb{P})$ in (22) satisfies

$$\forall\, x,y\in\mathcal{M},\qquad \|\Phi(x)-\Phi(y)\|_{L_2(\mathbb{P})}\gtrsim_{s,\varepsilon}\frac{1}{\sqrt{\log n}}d_{\mathcal{M}}(x,y). \tag{28}$$

The key observation now is that each of the coordinates of the Fréchet embedding is a distance from a nonempty subset of $\mathcal{M}$, so one can define Φ on the super space $\mathcal{S}$ and not just on $\mathcal{M}$, and the 1-Lipschitz condition (24), which is nothing more than an application of the triangle inequality for $d_{\mathcal{S}}$, still holds. Therefore, we may assume from now that $\Phi:\mathcal{S}\to L_2(\mathbb{P})$ is 1-Lipschitz and satisfies (28).

Denote $m=\lceil\delta\log n\rceil$. The (proof of the) Johnson–Lindenstrauss dimension reduction lemma [JL84] (see equation (28) in [Nao18] for the version that we are using here) implies that there exists a universal constant $C>0$ and a 1-Lipschitz function $g:\Phi(\mathcal{M})\to\ell_2^m$ that satisfies

$$\forall\, u,v\in\Phi(\mathcal{M}),\qquad \|g(u)-g(v)\|_2\geqslant e^{-\frac{C}{\delta}}\|u-v\|_{L_2(\mathbb{P})}. \tag{29}$$

By Kirszbraun's extension theorem [Kir34] (also e.g. [BL00, Chapter 1]) there is a 1-Lipschitz function $G:L_2(\mathbb{P})\to\ell_2^m$ that extends g. Denote $\varphi=G\circ\Phi:\mathcal{S}\to\ell_2^m$. Then φ is 1-Lipschitz and for every $x,y\in\mathcal{M}$,

$$\|\varphi(x)-\varphi(y)\|_2=\|g(\Phi(x))-g(\Phi(y))\|_2\overset{(29)}{\gtrsim} e^{-\frac{C}{\delta}}\|\Phi(x)-\Phi(y)\|_{L_2(\mathbb{P})}\overset{(28)}{\gtrsim}_{s,\varepsilon}\frac{e^{-\frac{C}{\delta}}}{\sqrt{\log n}}d_{\mathcal{M}}(x,y).$$

It remains to apply Observation 24 with $\mathcal{N}=\ell_2^m=\ell_2^{\lceil\delta\log n\rceil}$, $\alpha=1$ and $\beta\lesssim_{s,\varepsilon}e^{\frac{C}{\delta}}\sqrt{\log n}$, and then to take the supremum over all possible super-spaces $(\mathcal{S},d_{\mathcal{S}})$ of $(\mathcal{M},d_{\mathcal{M}})$ to arrive at the desired estimate (27). □

1.3.3. *Sparsest Cut, Max-Cut, sparsification, and nearest neighbor search.* Given $n\in\mathbb{N}$, the Sparsest Cut problem (with general capacities and demands) on n-vertices takes as its input two n-by-n symmetric matrices with nonnegative entries $\mathsf{C}=(c_{ij}),\mathsf{D}=(d_{ij})\in\mathsf{M}_n([0,\infty))$ and aims to evaluate (or estimate) in polynomial time the following quantity:

$$\mathrm{SparsestCut}(\mathsf{C},\mathsf{D})\overset{\mathrm{def}}{=}\min_{\emptyset\neq S\subsetneq[n]}\frac{\sum_{i\in S}\sum_{j\in[n]\smallsetminus S}c_{ij}}{\sum_{i\in S}\sum_{j\in[n]\smallsetminus S}d_{ij}}. \tag{30}$$

In the mid-1990s Goemans and Linial introduced a semidefinite program (SDP) that computes (with $o(1)$ precision) in polynomial a number $\mathrm{SDP}_{\mathsf{GL}}(\mathsf{C},\mathsf{D})\geqslant 0$ that satisfies:

$$\mathrm{SDP}_{\mathsf{GL}}(\mathsf{C},\mathsf{D})\leqslant\mathrm{SparsestCut}(\mathsf{C},\mathsf{D}).$$

[10] Note that [NR17] failed to state the best known bound on $\mathsf{e}_n(\ell_1,\ell_p)$ for $1<p<2$, which is $\mathsf{e}_n(\ell_1,\ell_p)\lesssim\sqrt{(\log n)/(p-1)}$. This is a special case of [MN06, Theorem 2.1], i.e., it was actually available at the time [NR17] was written but overlooked there.

See e.g. [NY18] and the references therein for a relatively recent account of this extensively studied topic, as well as Section 2. The pertinent question is therefore to understand the growth rate as $n \to \infty$ of the integrality gap of the Goemans–Linial SDP for Sparsest Cut, which is defined to be the following quantity:

$$\sup_{\substack{\mathsf{C},\mathsf{D}\in\mathsf{M}_n([0,\infty))\\ \mathsf{C},\mathsf{D}\ \text{symmeric}}} \frac{\mathrm{SparsestCut}(\mathsf{C},\mathsf{D})}{\mathrm{SDP}_{\mathsf{GL}}(\mathsf{C},\mathsf{D})}.$$

The algorithm that outputs $\mathrm{SDP}_{\mathsf{GL}}(\mathsf{C},\mathsf{D})$ is then guaranteed to estimate $\mathrm{SparsestCut}(\mathsf{C},\mathsf{D})$ within a factor that is at most this integrality gap. No other algorithm is currently known to (or is conjectured to) perform asymptotically better then this algorithm of Goemans and Linial.

Theorem 28 (integrality gap of the Goemans–Linial SDP for Sparsest Cut)**.** *For every $n \in \mathbb{N}$, the n-vertex integrality gap of the Goemans–Linial semidefinite program for the Sparsest Cut problem is bounded from above and from below by positive universal constant multiples of $\sqrt{\log n}$.*

The lower bound of $\Omega(\sqrt{\log n})$ in Theorem 28 on the integrality gap of the Goemans–Linial SDP for Sparsest Cut is from [NY18]. Thus, our contribution to Theorem 28 is in terms of algorithm design rather than proving an impossibility result, i.e., we derive an improved (sharp) upper bound on the integrality gap of this SDP, and hence we provide the best known polynomial time approximation algorithm for Sparsest Cut. The fact that Theorem 22 implies a $O(\sqrt{\log n})$ upper bound on the integrality gap of the Goemans–Linial SDP for Sparsest Cut is a standard result that has been known for a long time (going back at least to [Goe97]); see [Nao10, Lemma 4.5] for its detailed derivation (based on the duality argument in [Mat02, Proposition 15.5.2], which is attributed there to unpublished work of Rabinovich).

We focused above on the algorithmic task of efficiently approximating the number $\mathrm{SparsestCut}(\mathsf{C},\mathsf{D})$, but the fact that our $O(\sqrt{\log n})$-embedding of any n-point metric space of negative type is actually (per Theorem 21) into ℓ_2 rather than "merely" into ℓ_1 implies formally that there is also an algorithm which outputs a subset $\emptyset \neq S \subsetneq [n]$ that is a near-minimizer of the right hand side of (30), up to the aforementioned $O(\sqrt{\log n})$ error tolerance. This deduction utilizes the observation from the seminal work [LLR95] of Linial, London and Rabinovich that optimal embeddings into a Hilbert space can themselves be found in polynomial time (unlike embeddings into L_1, see [IM04, DL10]), as this task itself can be cast as a semidefinite program; the (standard) deduction of this assertion is worked out in e.g. [ALN08, Section 5].

One can refine the above discussion for each $k \in \{2,\ldots,n\}$ to obtain a $O(\sqrt{\log k})$-factor approximation algorithm if the matrix D has support of size at most k, i.e., $|\{(i,j)\in[n]\times[n] : d_{ij}>0\}| \leqslant k$. Also this fact is a formal consequence of what we have seen, as Theorem 21 is proved by substituting Theorem 1 into Theorem 17, which yields the stated distortion guarantee via the Fréchet embedding. That embedding can be automatically extended to any super-space while remaining 1-Lipschitz (as we have seen above in the proof of Corollary 27), which is all that is needed in order to obtain the aforementioned $O(\sqrt{\log k})$-factor approximation guarantee. Again, the standard deduction of this assertion is worked out in [ALN08], though note that [ALN08, Section 5] incorporates a step that is irrelevant for our purposes because the embedding of [ALN08] is not Fréchet, so [ALN08] must justify why it could be extended.

It is worthwhile to summarize the above discussion as the following separate algorithmic statement:

Theorem 29 (algorithm for finding an approximate Sparsest Cut when there are k demand pairs)**.** *There exists a polynomial time algorithm with the following property. Suppose that $n \in \mathbb{N}$ and $k \in \{2,\ldots,n\}$. Let $\mathsf{C} = (c_{ij}), \mathsf{D} = (d_{ij}) \in \mathsf{M}_n([0,\infty))$ be symmetric matrices such that $|\{(i,j)\in[n]\times[n] : d_{ij}>0\}| \leqslant k$. Then, the aforementioned algorithm outputs a subset $\emptyset \neq S \subsetneq [n]$ that satisfies:*

$$\frac{\sum_{i\in S}\sum_{j\in[n]\smallsetminus S} c_{ij}}{\sum_{i\in S}\sum_{j\in[n]\smallsetminus S} d_{ij}} \lesssim \mathrm{SparsestCut}(\mathsf{C},\mathsf{D})\sqrt{\log k}. \qquad (31)$$

The next algorithmic application of Theorem 1 that we will describe in this section follows from the work of Makarychev, Makarychev and Vijayaraghavan [MMV14] which beautifully relates the Sparsest

Cut problem to perturbation resilience of the Max-Cut problem (in the sense of Bilu–Linial [BL12]; see also the MaxCut-specific work [BDLS13], and the survey [MM16]).

The Max-Cut problem on n-vertices takes as its input an n-by-n symmetric matrix with nonnegative entries $\mathsf{W} = (w_{ij}) \in \mathsf{M}_n([0,\infty))$, which one should think of as an edge-weighted graph whose vertex set is $[n]$, and aims to compute (or approximate) in polynomial time the following quantity:

$$\mathrm{MaxCut}(\mathsf{W}) \stackrel{\mathrm{def}}{=} \max_{\emptyset \neq S \subsetneq [n]} \sum_{i\in S} \sum_{j \in [n]\setminus S} w_{ij}. \tag{32}$$

Given $\gamma \geqslant 1$, a matrix W as above is Bilu–Linial γ-stable (γ-stable, in short) for Max-Cut if there is a unique $\emptyset \neq S \subsetneq [n]$ such that $\mathrm{MaxCut}(\mathsf{W}) = \sum_{i\in S}\sum_{j\in[n]\setminus S} w_{ij}$, namely, the right hand side of (32) has only one maximizer, and furthermore we have $\mathrm{MaxCut}(\mathsf{W}') = \sum_{i\in S}\sum_{j\in[n]\setminus S} w'_{ij}$ for every symmetric matrix $\mathsf{W}' = (w'_{ij}) \in \mathsf{M}_n([0,\infty))$ that satisfies $w_{ij} \leqslant w'_{ij} \leqslant \gamma w_{ij}$ for every $i,j \in [n]$.

Makarychev, Makarychev and Vijayaraghavan studied [MMV14] an SDP that can be evaluated (with $o(1)$ precision) in polynomial time and outputs a number $\mathrm{SDP}_{\mathsf{MMV}}(\mathsf{W}) \geqslant 0$ that satisfies

$$\mathrm{SDP}_{\mathsf{MMV}}(\mathsf{W}) \geqslant \mathrm{MaxCut}(\mathsf{W}). \tag{33}$$

The Makarychev–Makarychev–Vijayaraghavan SDP for Max-Cut coincides with the Goemans–Williamson SDP for Max-Cut [GW94] with the (by now standard) added squared-ℓ_2 triangle inequality constraints, except for the "twist" that those constraints are imposed on the symmetrized version of the vector solution, namely, if the SDP outputs unit vectors $v_1,\ldots,v_n \in S^{n-1}$, then Makarychev, Makarychev and Vijayaraghavan require that $\|u-v\|_2^2 \leqslant \|u-w\|_2^2 + \|w-v\|_2^2$ for every $u,v,w \in \{v_1,\ldots,v_n,-v_1,\ldots,-v_n\}$.[11]

The Makarychev–Makarychev–Vijayaraghavan SDP for Max-Cut is said to be integral on the input W (see [MMV14, Definition 2.3]) if for every optimal solution $v_1,\ldots,v_n \in S^{n-1}$ of this SDP there is a unit vector $e \in S^{n-1}$ such that $v_i \in \{e,-e\}$ for every $i \in [n]$. By (33), $\mathrm{SDP}_{\mathsf{MMV}}(\mathsf{W}) = \mathrm{MaxCut}(\mathsf{W})$ when this occurs, i.e., the Makarychev–Makarychev–Vijayaraghavan algorithm outputs an exact solution for Max-Cut on the input W, in contrast to the fact that (under standard complexity-theoretic assumptions) one cannot hope to always obtain such an exact solution on general inputs [PY91, Hås01, KKMO07, MOO10].

By substituting Theorem 22 into [MMV14, Theorem 3.1] and [MMV14, Theorem 5.2], we get the following evaluation of the growth rate of the critical $\gamma = \gamma(n)$ such that the Makarychev–Makarychev–Vijayaraghavan SDP for Max-Cut is integral on every n-vertex instance of Max-Cut that is γ-stable:

Theorem 30 (stable integrality of the Makarychev–Makarychev–Vijayaraghavan SDP for Max-Cut)**.** *There are universal constants $C > c > 0$ with the following property for any integer $n \geqslant 2$. If $\gamma \geqslant C\sqrt{\log n}$, then the Makarychev–Makarychev–Vijayaraghavan SDP relaxation of Max-Cut is integral on any n-vertex input that is γ-stable for Max-Cut. If $1 \leqslant \gamma \leqslant c\sqrt{\log n}$, then there is an n-vertex input that is γ-stable for Max-Cut on which the Makarychev–Makarychev–Vijayaraghavan SDP relaxation of Max-Cut is not integral.*

We will next combine Corollary 27 with [MM16, Theorem 5.1] to get an asymptotic improvement of [MM16, Theorem 5.11], which is an intriguing link between the (major) classical problem on Lipschitz extension that we recalled in Question 25, and the vertex sparsification paradigm of Moitra [Moi09].

Fix $n \in \mathbb{N}$ and $Q \geqslant 1$. Let $\mathsf{A} = (\alpha_{ij}) \in \mathsf{M}_n([0,\infty))$ be a symmetric matrix (which, again, we think of as an edge-weighted graph whose vertex set is $[n]$). Following [Moi09], given $U \subseteq [n]$ and a symmetric matrix

$$\mathsf{B} = (\beta_{ij})_{(i,j)\in U\times U} \in M_{U\times U}\big([0,\infty)\big),$$

one says that B is a Q-quality vertex cut sparsifier of A if the following holds for every $\emptyset \neq S \subsetneq U$:

$$\min_{\substack{T\subseteq[n]\\ T\cap U = S}} \sum_{i\in T}\sum_{j\in[n]\setminus T} \alpha_{ij} \leqslant \sum_{p\in S}\sum_{q\in U\setminus S} \beta_{pq} \leqslant Q \min_{\substack{T\subseteq[n]\\ T\cap U = S}} \sum_{i\in T}\sum_{j\in[n]\setminus T} \alpha_{ij}. \tag{34}$$

[11] See [Kar05, ACMM05] for earlier incarnations of this idea, along with demonstrations of its utility in other aspects of combinatorial optimization.

The combinatorial meaning of (34) is that if the weighted graph B on the (potentially much) smaller subset U of the vertices $[n]$ is a Q-quality vertex cut sparsifier of a weighted graph A, then the size with respect to B of the edge-boundary of any partition of U into two sets S and $U\smallsetminus S$ is up to the specified error tolerance Q the minimal size with respect to A of the edge-boundary of a bipartition of $[n]$ that separates S and $U\smallsetminus S$. If $|U|$ and Q are small, then this is a valuable tool for reducing the dependence on n in combinatorial optimization problems; see [Moi09, LM10] for concrete examples of such applications.

For $k\in\mathbb{N}$, let Q_k^{cut} denote the infimum over all those $Q\geqslant 1$ such that for every $n\in\mathbb{N}$, for every symmetric matrix $\mathsf{A}=(\alpha_{ij})\in\mathsf{M}_n([0,\infty))$ and every $U\subseteq[n]$ with $|U|=k$ there exists a Q-quality cut sparsifier $\mathsf{B}=(\beta_{ij})_{(i,j)\in U\times U}\in M_{U\times U}([0,\infty))$. Moitra proved in [Moi09] that

$$Q_k^{\mathrm{cut}}\lesssim\frac{\log k}{\log\log k},\tag{35}$$

which remains the best known upper bound despite substantial efforts [CLLM10, EGK$^+$10, MM16]. Using the notation for Lipschitz extension moduli we recalled above, Theorem 5.1 in [MM16] states that

$$Q_k^{\mathrm{cut}}=\mathsf{e}(\ell_1;\ell_1).\tag{36}$$

By (36), Theorem D.2 in [MM16] (which is due to Johnson and Schechtman, applying [FJS88]) gives

$$Q_k^{\mathrm{cut}}\gtrsim\frac{\sqrt{\log k}}{\log\log k},\tag{37}$$

which is currently the best known lower bound on Q_k^{cut}.

By substituting the identity (36) into Corollary 27 we conclude that

$$Q_k^{\mathrm{cut}}\lesssim\mathsf{e}_k\big(\ell_2^{\lceil\log k\rceil};\ell_1\big)\sqrt{\log k}.\tag{38}$$

Thus, a lower bound on Q_k^{cut} that grows faster than a positive universal constant multiple of $\sqrt{\log k}$ would imply a negative answer to Question 25; this would hold in particular if the old upper bound (35) were sharp. Conversely, if Question 25 had a positive answer, then by (38) we would get that $Q_k^{\mathrm{cut}}\lesssim\sqrt{\log k}$, i.e., we would then have an upper bound on Q_k^{cut} that almost matches the lower bound (37).

Remark 31. *The best known upper bound* [LN05] *on* $\mathsf{e}(\ell_2^n,\ell_1)$*, in fact, on* $\mathsf{e}(\ell_2^n;\mathbf{Z})$ *for any Banach space* $\mathbf{Z}$*, is* $\mathsf{e}(\ell_2^n;\mathbf{Z})\lesssim\sqrt{n}$*. The best known lower bound is* $\mathsf{e}(\ell_2^n;\mathbf{Z})\gtrsim\sqrt[4]{n}$ *for a suitably chosen Banach space* $\mathbf{Z}$*. This is due to* [MN13]*, building on Kalton's important contributions* [Kal04, Kal12]*; a different proof was found in* [Nao21b]*. In both cases the target space* $\mathbf{Z}$ *is not* ℓ_1 *and it is conceivable that an upper bound on* $\mathsf{e}(\ell_2^n;\ell_1)$ *that is better than* $O(\sqrt{n})$ *is available (indeed, Question 25 might have a positive answer). The forthcoming work* [BN24] *can be viewed as some evidence that perhaps* $\mathsf{e}(\ell_2^n;\mathbf{Z})\lesssim\sqrt[4]{n}$ *for every Banach space* $\mathbf{Z}$*. If this were true, then its special case* $\mathbf{Z}=\ell_1$ *together with* (38) *would imply that* $Q_k^{\mathrm{cut}}\lesssim(\log k)^{3/4}$.

Our final example of an application of Theorem 1 (through its corollaries in Section 1.3.2) is to the Approximate Nearest Neighbor Search problem, which is of importance to both theory and practice; the relatively recent survey [AIR18] and the references therein are a good entry point to this active area.

Let $(\mathcal{M},d)$ be a metric space, $n\in\mathbb{N}$ and $D>1$. The goal of the D-approximate nearest neighbor search problem is to preprocess $\mathcal{C}$ so as to obtain a data structure with the following property. There is an algorithm that takes as input a query point $x\in\mathcal{M}\smallsetminus\mathcal{C}$, examines only the data structure, and outputs a point $y\in\mathcal{C}$ that is guaranteed to satisfy $d(x,y)\leqslant Dd(x,\mathcal{C})$. Both the preprocessing and the query algorithm are allowed to use randomness, in which case the aforementioned approximation guarantee has to hold with probability that is at least a positive universal constant (say, $2/3$). The pertinent parameters which one hopes to make small are the preprocessing time (how long it takes to construct the data structure), the size of the data structure (how much space is required to store it), and the query time (given the query point x, how long it takes the algorithm to output its approximate nearest neighbor y in $\mathcal{C}$).

The above formulation sets aside computational issues regarding how the metric space is given; these are important but secondary in the present context. It suffices to consider the metric in the black-box model, i.e., through oracle calls to the distance function. In concrete examples (e.g. if there is an extrinsic coordinate structure) more realistic computational models for the input metric are available, but those are case-specific while herein we wish to focus on more general principles.

By [HPIM12], the approximate nearest neighbor search problem can be reduced to solving $(\log n)^{O(1)}$ instances of its "decision version," called the near neighbor search problem. Given $D > 1$ and $r > 0$, the (D, r)-near neighbor search problem aims to (randomly) preprocess the data set $\mathcal{C}$ to build a data structure for which there is a (randomized) algorithm that takes as input a query point $x \in \mathcal{M} \setminus \mathcal{C}$, examines only the data structure and outputs a point $y \in \mathcal{C}$ such that if $d(x, \mathcal{C}) \leqslant r$, then with $\Omega(1)$ probability we have $d(x, y) \leqslant Dr$. Due to this reduction, we will focus from now on near neighbor search.

The following result is a substitution of Theorem 15 into [ANN$^+$18a, Theorem 4.1]:

Theorem 32 (cell-probe data structure for near neighbor on quasisymmetrically Hilbertian spaces)**.** *Suppose that $0 < s, \varepsilon \leqslant 1/2$ and that $K, n, w \geqslant 2$ are integers. Let $(\mathcal{M}, d)$ be a finite metric space whose size satisfies $\log|\mathcal{M}| \lesssim \log\log n$ that is both K-doubling and (s, ε)-quasisymmetrically Hilbertian. Fix $\mathcal{C} \subseteq \mathcal{M}$ with $|\mathcal{C}| = n$. Also, fix $r > 0$ that can be specified using w bits. For every $(\log\log n)/\log n \leqslant \delta \leqslant 1$ one can randomly preprocess this datum (we make no claim about the preprocessing time) and eventually store some memory as a (random) data structure consisting of a sequence of cells, each of which contains a single bit. The total number of cells (the space of the data structure) is $O(w + n^{1+\delta}\log|\mathcal{M}|)$. There is a randomized algorithm that takes as input a query point $x \in \mathcal{M}$, probes at most $O(w + n^{\delta}\log|\mathcal{M}|)$ cells of the data structure (possibly adaptively), performs unbounded auxiliary computations and uses unbounded auxiliary memory, and outputs a point $y \in \mathcal{C}$ such that if $d(x, \mathcal{C}) \leqslant r$, then with probability $\Omega(1)$ we have*

$$d(x, y) \lesssim_{s,\varepsilon} \frac{\sqrt{\log K}}{\delta} r. \tag{39}$$

Background on the cell-probe model for data structures can be found in e.g. [Yao81, Mil99]. Theorem 32 plainly does not address all of the algorithmic issues that we mentioned above. It does not control the preprocessing time for constructing the data structure, and it only bounds the number of cells of the data structure that the algorithm examines without controlling its overall running time. As in [ANN$^+$18a], the value of Theorem 32 is that it serves as a "proof of concept" rather than a final result, or at least it can be thought of as a barrier for proving impossibility of efficient approximate near neighbor data structures with the parameters described in Theorem 32. This is so because all of the known unconditional data structure lower bounds proceed by proving a cell-probe lower bound [Mil99].

Theorem 32 raises the challenge to make the iterative partitioning scheme of [ANN$^+$18a] that is powered by the cutting modulus estimate of Theorem 15 computationally efficient, which would entail making our construction in the proof of Theorem 1 computationally efficient. We expect that this goal is achievable (perhaps only for a subclass of quasisymmetrically Hilbertian metric spaces that are represented in a suitably succinct manner), but would require effort and new ideas, analogously to the transition from [ANN$^+$18a] to [ANN$^+$18b]; we defer this interesting research direction to future work.

By combining [ANN$^+$18a, Theorem 4.1] with the facts that were recalled immediately after the statement of Theorem 13, we see that for arbitrary K-doubling metric spaces Theorem 32 holds mutatis mutandis, but with the factor $\sqrt{\log K}$ in (39) replaced by $\log K$. The available D-approximate near neighbor data structure for general K-doubling metric spaces with approximation $D = O((\log K)^{\theta})$ for some $0 < \theta < 1$ suffers from the "curse of dimensionality" (the intrinsic dimension in this context is naturally of order $\log K$), in the sense that its query time grows super-polynomially in $\log K$. Specifically, the query time obtained in [Fil23] when $D = C(\log K)^{\theta}$ is at least $\exp(\Omega((\log K)^{1-\theta})/C)$, albeit in a tailor-made (non-standard) computational model (when $D = 1+\varepsilon$ for some $0 < \varepsilon < 1$, which is a very important range of parameters but not the one that concerns us here, the same issue occurs in the older works [KL05, HPM06] that use a standard computational model). If the given metric space is also (s, ε)-quasisymmetrically

Hilbertian and $n=\exp(o_{s,\varepsilon}(\sqrt{\log K}))$, then Theorem 32 applies with $D\lesssim_{s,\varepsilon}\sqrt{\log K}/\delta$ and its query time is better than that of [Fil23].

If the answer to Question 14 were such that the $\sqrt{\log K}$ term in (17) could be improved to $o(\sqrt{\log K})$, then we would correspondingly improve the dependence on K in (39). This is an intriguing possibility of potential practical significance in addition to its great theoretical interest. As an example of a consequence of this possiblity that does not relate to (17), we note that because for every $n\in\mathbb{N}$ any n-point metric space is (trivially) n-doubling, in combination with [Nao14, Theorem 1.3] this would imply in particular that any n-point metric space of negative type embeds with average distortion $o(\sqrt{\log n})$ into ℓ_1, thus improving over the main embedding result of [ARV09] and consequently yielding an asymptotic improvement over the best-known polynomial time approximation algorithm for the Sparsest Cut problem with uniform capacities and demands, i.e., the algorithmic optimization problem that corresponds to the special case of (30) in which $c_{ij}\in\{0,1\}$ and $d_{ij}=1$ for every $i,j\in[n]$.

2. History and proof overview

The precursor to Theorem 1 appeared as Theorem 3.1 of [ALN08] (first announced in [ALN05]); using the terminology that we recalled as Definition 5, it states that any n-point metric space $(\mathcal{M},d)$ of negative type admits a random zero set that is $O(\sqrt{\log n})$-spreading with probability $\Omega(1)$, i.e., for some universal constant $c>0$ and any $\tau>0$ there exists a probability measure $\mathbb{P}=\mathbb{P}^{\tau,\mathcal{M}}$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ such that

$$\forall x,y\in\mathcal{M},\qquad d(x,y)\geqslant\tau\implies\mathbb{P}\left[\emptyset\neq\mathscr{Z}\subseteq\mathcal{M}:\ d(y,\mathscr{Z})\geqslant\frac{c\tau}{\sqrt{\log n}}\quad\text{and}\quad x\in\mathscr{Z}\right]\gtrsim 1. \tag{40}$$

The main result of [ALN05, ALN08] used (40) to prove that $\mathsf{c}_2(\mathcal{M})\lesssim\sqrt{\log n}\log\log n$, and [ALN07] showed that the same bound on the Euclidean distortion of $\mathcal{M}$ can be obtained via the Fréchet embedding.

Evidently, conclusion (1) of Theorem 1 is quantitatively stronger than (40). However, its main contribution is a qualitative enhancement that allows one to exploit cancelation to derive the upper bounds[12] of Theorem 20, Theorem 21, Theorem 22 and Theorem 28 without any unbounded lower order factors, thus resolving the respective open questions; the previously best known bounds here [ALN08] had a redundant $\log\log n$ factor, and correspondingly conclusion (31) of Theorem 29 was known with the factor $\sqrt{\log k}$ replaced by $\sqrt{\log k}\log\log k$. The connection between the Sparsest Cut problem and perturbation resilience of the MaxCut problem was discovered in [MMV14], where Theorem 30 was obtained with the lower bound $\gamma\geqslant C\sqrt{\log n}$ replaced by $\gamma\geqslant C\sqrt{\log n}\log\log n$, as [MMV14] appealed to [ALN08]; the impact of Theorem 1 in this context is to obtain the sharp rate of growth of the threshold for stable integrality.[13] The connection between Lipschitz extension and vertex sparsification was discovered in [MM16], where the bound (38) was obtained with the term $\sqrt{\log k}$ replaced by $\sqrt{\log k}\log\log k$, as [MM16] appealed to [ALN07]; the above derivation of Corollary 27 follows the approach of [MM16].

The phenomenon that Theorem 7 establishes was first broached in [NRS05]. While [NRS05] did not succeed to obtain the desired extremal property of the Euclidean sphere that is independent of dimension, Theorem 1.7 of [NRS05] is a result towards this goal that is off by a logarithmic factor. Specifically, [NRS05, Theorem 1.7] shows that for any increasing $\eta:[0,\infty)\to[0,\infty)$ with $\lim_{s\to0}\eta(s)=0$ and any $n\in\mathbb{N}$, if $(\mathcal{M},d)$ is a metric space of diameter $O(1)$ that has a modulus-η quasisymmetric embedding into ℓ_2 and μ is a normalized $e^{O(n)}$-doubling Borel probability measure on $\mathcal{M}$, then there exists $0<\theta_\eta<1$ such that

$$\mathrm{ObsDiam}_{\mu}^{\mathcal{M}}(\theta_\eta)\gtrsim_\eta\frac{1}{\sqrt{\log n}}\mathrm{ObsDiam}_{\sigma^{n-1}}^{S^{n-1}}(\theta_\eta). \tag{41}$$

[12]Recall that the new content of these theorems is their upper bounds, namely their positive embedding/algorithmic assertions; the matching impossibility results are due to [Enf69, NY18].

[13]Up to universal constant factors; perhaps here one could hope to discover the exact threshold. Obtaining sharp constants in Theorem 18, Theorem 20, Theorem 21, Theorem 22 and Theorem 28 seems very far beyond reach, likely an unrealistic goal.

The other applications to doubling metric spaces that are obtained herein did not appear in the literature even up to lower-order factors, though for some of those applications we expect that one could suitably adapt the reasoning of [NRS05, ALN08] to derive similarly non-sharp statements.

The utility of Theorem 1 for removing altogether any unbounded lower order factor from [ALN08] is obvious, as it is nothing more than a direct substitution of (1) into measured descent [KLMN05] (Theorem 17 herein). This possibility is discussed in [ALN05], though with skepticism that the statement of Theorem 1 could be valid due to the nonlocal nature of the (deep) ARV rounding algorithm [ARV09], which is the central input to (40). In contrast, a key feature of (1) is that the "performance" of the random zero set at the given scale τ for each given pair of points $x, y \in \mathcal{M}$ depends only the local "snapshot" $B(y, \kappa\beta\tau)$ of $\mathcal{M}$ near y at scale $O_\eta(\tau)$, and moreover that it depends only on the extent to which the measure of $B(y, \kappa\beta\tau)$ increases relative to the measure of the proportionally smaller snapshot $B(y, \beta\tau)$ of $\mathcal{M}$.

In absence of such locality, [ALN08] starts out with (40), from which point it does not make any further appeal to [ARV09] and instead it proceeds by adapting measured descent in order to obtain the aforementioned distortion bound, i.e., while incurring a lower order yet unbounded multiplicative loss. The reasoning that is used in [ALN07] to show that this can, in fact, be achieved via the Fréchet embedding also takes (40) as a "black box" without any further appeal to [ARV09] and proceeds by yet another enhancement of measured descent.

We do not see how to derive the aforementioned sharp embedding results via the above route, which is purely metric/analytic in contrast to the more structural nature of ARV. Instead, the present work "flips" the approach of [ALN08] by leaving measured decent untouched (it can now be simply quoted), and enhancing, as we will next outline, the structural insights that are provided by the ARV framework.

After the appearance of the ARV algorithm in [ARV04], simplifications, refinements, extensions, and reformulations of it were developed in multiple works, including notably its full journal version [ARV09] and [CGR05, Lee05, NRS05, ACMM05, ALS11, Rot16]. All of those contributions were valuable to us in the process of developing the ensuing enhancement of ARV. Rothvoss' lecture notes [Rot16] stand out here since they creatively redo the setup and reasoning in a natural, and, as it turns out, more flexible way. In particular, we introduce Definition 34 below of compatibility of a labelled graph with a mapping into $\mathbb{R}^n$, which arose from our efforts to understand the extent to which the proof in [Rot16] can be strengthened.

2.1. **Directional Euclidean sparsification of graphs.** The above referenced versions of the ARV algorithm study (explicitly or implicitly) a natural way to "sparsify" certain (combinatorial) graphs, i.e., a procedure that removes in a meaningful way some of the edges of a given graph. Our proof of Theorem 1 investigates a more involved version of this procedure, applied to graphs that encode more geometric information than the proximity graphs that were used in those works. This section is devoted to explaining these ideas in suitable generality.

The aforementioned literature considers a finite negative type metric space $(\mathcal{M}, d)$ and a scale $\tau > 0$, and studies the proximity graph $\mathsf{G} = (\mathcal{M}, E)$ whose vertex set is $\mathcal{M}$ and $\{x, y\} \in E$ if and only if $d(x, y) \leqslant \Delta$, where $\Delta = c\tau/\sqrt{\log|\mathcal{M}|}$ for some universal constant $c > 0$. In the present setting, we are led to consider a certain graph (specified below) whose vertex set is still $\mathcal{M}$, yet if a pair of points $x, y \in \mathcal{M}$ are joined by an edge, then this will encode information that combines both their proximity and the local growth rate near x, y of a given measure μ on $\mathcal{M}$.[14] The sparsification procedure in our context will involve a pairwise thresholding criterion (we will soon explain what this means) that is nonconstant, while previous works considered a fixed threshold that is independent of the given pair of points in $\mathcal{M}$.

Because the ensuing reasoning needs to consider multiple metrics simultaneously, including multiple metrics on the same set (arising from both the original metric and the shortest-path metrics of graphs as above), it will be beneficial to use subscripts when denoting distances, balls, diameters.[15] Thus, given

[14]In fact, we will need to analyse separately the connected components of such a graph, but we will initially suppress that subtlety for the purpose of this overview.

[15]In parts that treat only a single metric, such as Section 10 and some of the Introduction, we will drop this convention.

a metric space $(\mathcal{M},d_{\mathcal{M}})$, we will write $\mathrm{diam}_{\mathcal{M}}(A)=\sup_{a,b\in A} d_{\mathcal{M}}(a,b)$ and $d_{\mathcal{M}}(x,A)=\inf_{a\in A} d_{\mathcal{M}}(x,a)$ for, respectively, the $d_{\mathcal{M}}$-diameter of $\emptyset\neq A\subseteq \mathcal{M}$ and the $d_{\mathcal{M}}$-distance of $x\in\mathcal{M}$ from A. We will also write $B_{\mathcal{M}}(x,r)=\{y\in\mathcal{M}:\ d_{\mathcal{M}}(x,y)\leqslant r\}$ for the $d_{\mathcal{M}}$-ball centered at x of radius $r\geqslant 0$. Thus, given $n\in\mathbb{N}$, $p\geqslant 1$ and $x\in\mathbb{R}^n$, we will use the notation $B_{\ell_p^n}(x,r)=\{y\in\mathbb{R}^n:\ \|x-y\|_p\leqslant r\}$, where $\|z\|_p=(|z_1|^p+\ldots+|z_n|^p)^{1/p}$ for $z=(z_1,\ldots,z_n)\in\mathbb{R}^n$. The the standard scalar product on $\mathbb{R}^n$ will be denoted $\langle\cdot,\cdot\rangle:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$, i.e., $\langle z,w\rangle=z_1w_1+\ldots+z_nw_n$ for $z=(z_1,\ldots,z_n), w=(w_1,\ldots,w_n)\in\mathbb{R}^n$. The standard Gaussian measure on $\mathbb{R}^n$ will be denoted (as usual) by γ_n, i.e., the density of γ_n at $z\in\mathbb{R}^n$ is proportional to $\exp(-\|z\|_2^2/2)$.

Throughout what follows, all graphs will be tacitly assumed to be finite and will be allowed to have self-loops. Given a (possibly disconnected) graph $\mathsf{G}=(V,E)$, denote the shortest-path/geodesic (extended) metric that it induces on its vertex set V by $d_{\mathsf{G}}:V\times V\to[0,\infty]$, under the natural convention that $d_{\mathsf{G}}(x,y)=\infty$ if and only if $(x,y)\in\Gamma\times\Gamma'$ for distinct connected components $\Gamma,\Gamma'\subseteq V$ of G. For $r\geqslant 0$ and $x\in V$, the corresponding ball in (V,d_{G}) will be denoted $B_{\mathsf{G}}(x,r)=\{y\in V:\ d_{\mathsf{G}}(x,y)\leqslant r\}$. In particular, $B_{\mathsf{G}}(x,1)=\{x\}\cup\{y\in V:\ \{x,y\}\in E\}$, which is also denoted $N_{\mathsf{G}}(x)$, is the neighborhood in G of the vertex x. Correspondingly, the neighborhood in G of a vertex subset $U\subseteq V$ will be denoted $N_{\mathsf{G}}(U)$, i.e.,

$$N_{\mathsf{G}}(U)\stackrel{\mathrm{def}}{=}\bigcup_{x\in U}N_{\mathsf{G}}(x)=\big\{y\in V:\ d_{\mathsf{G}}(y,U)\leqslant 1\big\}.$$

When a graph $\mathsf{G}=(V,E)$ is accompanied by a mapping $f:V\to\mathbb{R}^n$, which we think of as a geometric representation of G, and an edge-labelling $\sigma:E\to\mathbb{R}$, which we think of as a thresholding criterion for determining which edges will be retained in the ensuing sparsification, for each vector $v\in\mathbb{R}^n$ we can consider the sub-graph of G that is obtained by deleting those $\{x,y\}\in E$ with $|\langle f(x)-f(y),v\rangle|\leqslant 4\sigma(\{x,y\})$:

Definition 33 (directional Euclidean sparsification)**.** *Let $\mathsf{G}=(V,E)$ be a graph. If $n\in\mathbb{N}$ and $f:V\to\mathbb{R}^n$, then for every $\sigma:E\to\mathbb{R}$ and $v\in\mathbb{R}^n$ define*[16]

$$E(v;f,\sigma)\stackrel{\mathrm{def}}{=}\big\{\{x,y\}\in E:\ |\langle f(x)-f(y),v\rangle|>4\sigma(\{x,y\})\big\}\subseteq E. \tag{42}$$

We thus obtain the following subgraph[17] *of G, which we call the Euclidean sparsification of G in direction v corresponding to its Euclidean representation f and the thresholding function σ:*

$$\mathsf{G}(v;f,\sigma)\stackrel{\mathrm{def}}{=}\big(V,E(v;f,\sigma)\big). \tag{43}$$

Understanding typical properties of $\mathsf{G}(v;f,\sigma)$ when v is chosen randomly according to the Gaussian measure γ_n is interesting in its own right. Here, we will investigate its matching number, which is also what [Rot16] studies (as do other ARV-related works, usually implicitly), in the case when $\sigma\equiv 1/2$ is constant and $\mathsf{G}=(\mathcal{M},E)$ is the above "vanilla" proximity graph on a finite metric space $(\mathcal{M},d_{\mathcal{M}})$ of negative type, i.e., $\{x,y\}\in E$ if and only if $d(x,y)\leqslant\Delta$ for some fixed $\Delta>0$. Recall that the matching number $\nu(\mathsf{G})$ of a graph $\mathsf{G}=(V,E)$ is the maximum cardinality of a pairwise-disjoint collection of its edges. We will need only rudimentary properties of this basic combinatorial notion (that will be recalled when they will arise in proofs), which are covered in e.g. [LP09].

Since both the graphs $\mathsf{G}=(V,E)$ that we will investigate herein, and their labelings $\sigma:E\to\mathbb{R}$ that we will use as thresholds for Euclidean sparsification, are more complicated than the aforementioned special case, it is beneficial to first study the above setting abstractly. We arrived at the following definition, that will have an important role below, by examining the elegant proof in [Rot16] with this goal in mind:

Definition 34 (compatibility of a graph with its Euclidean realization and edge labeling)**.** *Fix $n\in\mathbb{N}$ and $C>0$. Let $\mathsf{G}=(V,E)$ be a graph, $f:V\to\mathbb{R}^n$ and $\sigma:E\to[0,\infty)$. We say that G is C-compatible with f and σ if there exist $\Delta:V\to[0,\infty)$ and $K:V\to\mathbb{N}$ that have the following three properties:*

[16]We decided to insert the factor 4 in the definition (42) of Euclidean sparsification for convenience only, as this yields a slight simplification of expressions in the ensuing reasoning. This choice is, of course, nothing more than a superficial normalization which could be removed if so desired by replacing throughout what follows the thresholding function σ by $\sigma/4$.

[17]Observe that even though graphs are allowed herein to have self-loops, if σ takes values in $[0,\infty)$, then the strict inequality in (42) implies that $\mathsf{G}(v;f,\sigma)$ will never have self-loops, i.e., $\{x,x\}\notin E(v;f,\sigma)$ for every $x\in V$.

(1) *For every* $x \in V$ *and* $y \in B_{\mathsf{G}}(x, K(x)-1)$, *if* $z \in V$ *is such that* $\{y,z\} \in E$, *then*

$$\Delta(x) \leqslant \sigma(\{y,z\}). \tag{44}$$

(2) *For every* $x \in V$ *and every* $y \in N_{\mathsf{G}}(x)$ *we have*

$$\int_{\mathbb{R}^n} \bigg(\max_{z \in B_{\mathsf{G}}(y,K(y))} \langle f(z) - f(y), v\rangle \bigg) \mathrm{d}\gamma_n(v) \leqslant K(x)\Delta(y). \tag{45}$$

(3) *For every* $x \in V$ *we have*

$$f\Big(B_{\mathsf{G}}\big(x, K(x)\big)\Big) \subseteq B_{\ell_2^n}\Big(f(x), \frac{1}{C}\Delta(x)\Big). \tag{46}$$

Definition 34 is scale-invariant in the following sense: if G is C-compatible with f and σ, then G is also C-compatible with λf and $\lambda\sigma$ for every $\lambda \geqslant 0$, as seen by considering $\lambda\Delta$. The only part of Definition 34 that involves the parameter C is the inclusion (46), which implies in particular that C-compatibility becomes a more stringent property as C grows. The precise role of the properties that Definition 34 requires from K, Δ will become apparent upon examining the details of how they are applied in the ensuing proofs. In (very) broad strokes, their significance is that K, Δ take as input a single vertex, which we think of as a consistent choice of "local scales" at that vertex (for, respectively, the domain and range of f), yet they control the pairwise interactions σ through (44), and the oscillations of f through (45); both of these controls occur on (combinatorial) balls in G whose radius is determined by K.

Note the following feature of Definition 34: it stipulates the existence of Δ, K with the stated properties, but these are auxiliary objects that occur internally to the definition and are not part of the notion of G being C-compatible with f and σ. Thus, any statement about C-compatibility will not refer to K, Δ, i.e., they will only be used as tools within its proof. This is exemplified by the following theorem, which is central to our proof of Theorem 1. It asserts that C-compatibility for large $C > 0$ implies that the expected matching number $\nu(\mathsf{G}(v; f, \sigma))$ is small when v is distributed according to γ_n. This result is a generalization of the ARV reasoning as it was recast by Rothvoss, and its proof (which appears in Section 3 below), is an adaptation of the strategy that he introduced in [Rot16, Section 6.1].

Theorem 35 (expected matching number of Euclidean sparsification)**.** *Fix* $C \geqslant 1$ *and* $n \in \mathbb{N}$. *Suppose that* $\mathsf{G} = (V,E)$ *is a graph that is* C*-compatible with* $f: V \to \mathbb{R}^n$ *and* $\sigma: E \to [0,\infty)$. *Then,*[18]

$$\int_{\mathbb{R}^n} \nu\big(\mathsf{G}(v; f, \sigma)\big)\,\mathrm{d}\gamma_n(v) < 6e^{-\frac{1}{4}C^2}|V|. \tag{47}$$

The proof of Theorem 35 is the only part of our proof of Theorem 3 that elaborates on the central innovation of [ARV09], namely, what is known today as the "ARV chaining argument" (through the approach to it that was devised in [Rot16]). We will suitably implement this argument in Section 3 to prove Theorem 35. That implementation involves work that is of a more technical nature, and the conceptually new contribution here is the mere introduction of Definition 34, which makes Theorem 35 possible.

2.2. **From Theorem 35 to Theorem 3.** The rest of the proof of Theorem 3 does not relate to ARV chaining. We will next describe the steps that remain in the derivation of Theorem 3 from Theorem 35. Proposition 36 below will be used for that purpose; its elementary proof appears in Section 4. Like Theorem 35, Proposition 36 is about graph theory/probability: Notice that we have not yet used the assumption that $(\mathcal{M}, d_{\mathcal{M}})$ is a metric space, or that it is quasisymmetrically Hilbertian; these will occur only in subsequent stages of the reasoning.

[18] As V is finite, it is straightforward to check from the definition (42) that the function $v \mapsto \nu(\mathsf{G}(v; f, \sigma))$ from $\mathbb{R}^n$ to $\mathbb{N} \cup \{0\}$ is Borel-measurable, i.e., $\{v \in \mathbb{R}^n : \nu(\mathsf{G}(v; f, \sigma)) = m\}$ is a Borel subset of $\mathbb{R}^n$ for fixed $m \in \mathbb{N} \cup \{0\}$. So, the integral in (47) is defined.

Proposition 36. *There is a universal constant $\kappa > 1$ with the following property. Fix $n \in \mathbb{N}$ and let $\mathsf{G} = (V, E)$ be a graph. Suppose that $\Lambda : V \to (0, \infty]$ is moderately varying*[19] *along edges of G in the following sense:*

$$\forall \{x, y\} \in E, \qquad \Lambda(y) \leqslant 2\Lambda(x). \tag{48}$$

Assume also that we are given a mapping $f : V \to \mathbb{R}^n$ and a probability measure ω on $V \times V$ such that the following inequality holds for every $x, y \in V$ for which (x, y) belongs to the support of ω, i.e., $\omega(x, y) > 0$, and also x and y belong to same connected component of G:

$$\|f(x) - f(y)\|_2 \geqslant \min\big\{\Lambda(x), \Lambda(y)\big\}. \tag{49}$$

Then, for every $C \geqslant 1$ and every $v \in \mathbb{R}^n$, there are $A(v) = A^{\omega,C}(v), B(v) = B^{\omega,C}(v) \subseteq V$ such that the mapping

$$(v \in \mathbb{R}^n) \mapsto \big(A(v), B(v)\big) \in 2^V \times 2^V \tag{50}$$

is Borel-measurable,[20] *and the following point-wise estimate holds:*

$$\forall\, v \in \mathbb{R}^n,\ \forall (x, y) \in A(v) \times B(v), \qquad \{x, y\} \in E \implies |\langle v, f(x) - f(y)\rangle| > C \max\big\{\Lambda(x), \Lambda(y)\big\}. \tag{51}$$

Furthermore, the γ_n-expected ω-measure of the product $A(v) \times B(v) \subseteq V \times V$ satisfies

$$\int_{\mathbb{R}^n} \omega\big(A(v) \times B(v)\big)\, \mathrm{d}\gamma_n(v) \gtrsim e^{-\kappa C^2}. \tag{52}$$

The following proposition is an especially important step in the proof of Theorem 3:

Proposition 37. *There is a universal constant $\zeta \geqslant 1$ with the following property.*[21] *Let $(\mathcal{M}, d_{\mathcal{M}})$ be a finite metric space, equipped with a nondegenerate measure μ. For $\tau, C > 0$ define $\rho = \rho_{d_{\mathcal{M}},\mu,C,\tau} : \mathcal{M} \to [1, \infty)$ by*

$$\forall x \in \mathcal{M}, \qquad \rho(x) \stackrel{\mathrm{def}}{=} 1 + \frac{\zeta}{C}\sqrt{\log \frac{\mu(B_{\mathcal{M}}(x, 19\tau))}{\mu(B_{\mathcal{M}}(x, \tau))}}. \tag{53}$$

Then, there exists a mapping $q = q_{d_{\mathcal{M}},\mu,C,\tau} : \mathcal{M} \to \mathcal{M}$ satisfying

$$\forall x \in \mathcal{M}, \qquad d_{\mathcal{M}}(q(x), x) \leqslant 7\tau, \tag{54}$$

with the following properties.

Let $\mathsf{G} = \mathsf{G}_{d_{\mathcal{M}},\mu,C,\tau}$ be the graph whose vertex set is $\mathcal{M}$ and whose edge set $E = E_{d_{\mathcal{M}},\mu,C,\tau}$ is given by

$$\forall x, y \in \mathcal{M}, \qquad \{x, y\} \in E \iff d_{\mathcal{M}}(x, y) \leqslant \frac{\tau}{\min\{\rho(x), \rho(y)\}}. \tag{55}$$

For every $x \in \mathcal{M}$, let $\Gamma(x) \subseteq \mathcal{M}$ denote the connected component of x in G. Then, $q(x) \in \Gamma(x)$ for every $x \in \mathcal{M}$. Furthermore, for any $n \in \mathbb{N}$ and any mapping $\varphi : \mathcal{M} \to \mathbb{R}^n$, the graph G is C-compatible with $\varphi \circ q : \mathcal{M} \to \mathbb{R}^n$

---

[19] The constant factor 2 in (48) was chosen arbitrarily for simplicity as this suffices for the present purposes, but the proof herein of Proposition 36 shows that replacing it by any factor that is bigger than 1 only impacts the value of κ in (52).

[20] Proposition 36 must state the measurability of the mapping in (50) in order to allow us to freely consider (at multiple junctures throughout the ensuing proofs) probabilities of various events, as well as expectations such as (52). The classical treatment of (Borel-)measurability of set-valued mappings between topological spaces is covered in e.g. the survey [Wag77], but we do not need to recall it here: As V is a finite set, the mapping in (50) can take finitely many values, so its measurability in Proposition 36 means that all of its fibers are Borel, i.e., for every $A, B \subseteq V$ the set $\{v \in \mathbb{R}^n :\ A(v) = A \text{ and } B(v) = B\}$ is a Borel subset of $\mathbb{R}^n$. This is also how one should interpret below measurability statements about set-valued mappings that are defined on topological spaces that differ from $\mathbb{R}^n$; we will need to consider such situations later in this article (all of the topological spaces that will occur below will be Polish, i.e., separable complete metric spaces). Measurability is thus a benign matter in the present finitary setting, but conceivably in the future one could need to consider variants for infinite spaces, in which case we expect that the nature of the concrete and simple constructions herein to easily lead to favorable measurability properties.

[21] An inspection of the reasoning herein reveals that $\zeta = 2$ would work for Proposition 37, though this value is not sharp.

and $\sigma : E \to [0,\infty)$*, where* $\sigma = \sigma_{d_{\mathcal{M}},C,\varphi\circ q} : E \to [0,\infty)$ *is defined by*[22]

$$\forall \{x,y\} \in E, \qquad \sigma(\{x,y\}) \stackrel{\mathrm{def}}{=} C\Bigg(\max_{\substack{a\in B_{\mathcal{M}}(x,2\tau)\cap B_{\mathcal{M}}(y,2\tau)\cap\Gamma(x)\\ b\in B_{\mathcal{M}}(a,2\tau)\cap\Gamma(x)}} \|\varphi\circ q(a)-\varphi\circ q(b)\|_2\Bigg). \tag{56}$$

Before proceeding to the rest of the steps of the proof of Theorem 3, will next discuss the significance of Proposition 37, whose proof, which appears in Section 5, contains a key idea of the present work.

Note that Proposition 37 is the first time that the assumption that $(\mathcal{M},d_{\mathcal{M}})$ is a metric space is used in the proof of Theorem 3, but this proposition works for any metric space and we do not yet need to know that it is quasisymmetrically Hilbertian. Also, Proposition 37 introduces the type of graphs to which the general combinatorial/probabilisitc statements of Theorem 35 and Proposition 36 will be applied: their vertices are points in $\mathcal{M}$ and their edges are given by (55). The relevance of Proposition 37 to Theorem 35 is evident, as Proposition 37 produces situations in which the C-compatibility assumption of Theorem 35 is satisfied. The link to Proposition 36 will be made later, in a subsequent step of the proof of Theorem 3.

The crucial geometric contribution of Proposition 37 is constructing the mapping $q:\mathcal{M}\to\mathcal{M}$, which should be viewed as a way to "compress" a given metric space: its image $q(\mathcal{M})\subseteq\mathcal{M}$ will typically be much smaller than $\mathcal{M}$, yet it will encode geometric properties of $\mathcal{M}$ and μ that will be important for the purpose of working with ratios of measures of balls as in (53), which is our main goal. In fact, Proposition 37 states that q is a universally compatible compression scheme in the sense that C-compatibility arises upon composition with q of any function whatsoever from $\mathcal{M}$ to $\mathbb{R}^n$ (for a suitable choice of edge-labelling σ).

Proposition 37 treats the specific function ρ in (53) because this is what is needed below, and also since this is what arises from the use of the Gaussian measure in requirement (45) of Definition 34. Nevertheless, the compression scheme in Section 5 works for any edge set as in (55) when $\rho:\mathcal{M}\to[1,\infty)$ is arbitrary. The construction in Section 5 is therefore more general than what is used for Proposition 37, and hence it could be useful for other purposes (perhaps in settings to which appropriate non-Gaussian or non-Euclidean versions of compatibility are pertinent). The idea is to consider and suitably analyse hierarchically nested (2τ)-nets in the level sets of ρ, arranged in increasing order; see Construction 47.

The fact that the function q of Proposition 37 preserves connected components of G will be used to apply Proposition 37 when φ is the mapping from Definition 2 of what it means for $(\mathcal{M},d_{\mathcal{M}})$ to be (s,ε)-quasisymemtrically Hilbertian; recall (3). The point is that since by (55) we know that $\{x,y\}\in E$ implies that $d_{\mathcal{M}}(x,y)\leqslant\tau$, as $\rho\geqslant 1$ by (53), any connected component Γ of G is τ-discretely path connected as a subset of $\mathcal{M}$, i.e., for every $x,y\in\Gamma$ there exist $k\in\mathbb{N}$ and a discrete path $z_0=x,z_1,\ldots,z_k=y\in\Gamma$ joining x to y such that $d_{\mathcal{M}}(z_i,z_{i-1})\leqslant\tau$ for every $i\in[k]$. This will allow us to iterate (3) within each connected component of G, leading to the following proposition, which is the only way that the assumption that $(\mathcal{M},d_{\mathcal{M}})$ is quasisymmetrically Hilbertian will be used in the proof of Theorem 3:

Proposition 38. *Fix* $0<\varepsilon,s\leqslant\frac12$ *and* $r\geqslant 1$, *and suppose that* $\beta>0$ *satisfies*

$$\beta\leqslant\beta(s,\varepsilon,r)\stackrel{\mathrm{def}}{=} s^{\frac{3\log(8r)}{\varepsilon}}. \tag{57}$$

For a finite (s,ε)*-quasisymmetrically Hilbertian metric space* $(\mathcal{M},d_{\mathcal{M}})$*, a nondegenerate measure* μ *on* $\mathcal{M}$*, and* $\tau,C>0$*, let* $\mathsf{G}=(\mathcal{M},E)$ *be the graph from Proposition* (37) *with* (C,τ) *replaced by* $(rC,\beta\tau)$*, i.e.,*

$$\forall x,y\in\mathcal{M}, \qquad \{x,y\}\in E \iff d_{\mathcal{M}}(x,y)\leqslant\frac{\beta\tau}{\min\{\rho(x),\rho(y)\}}, \tag{58}$$

where, with $\zeta\geqslant 1$ *the universal constant from Proposition* (37)*, we denote*

$$\forall x\in\mathcal{M}, \qquad \rho(x)\stackrel{\mathrm{def}}{=}1+\frac{\zeta}{rC}\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}. \tag{59}$$

---

[22]This σ is well-defined because if $\{x,y\}\in E$, then x and y belong to the same connected component of G, i.e., $\Gamma(x)=\Gamma(y)=\Gamma$. Also, because $\rho\geqslant 1$ by (53), definition (55) implies that if $\{x,y\}\in E$, then $d_{\mathcal{M}}(x,y)\leqslant\tau$, so the maximum in (56) is over a nonempty subset of $\mathcal{M}$ (e.g., one can take there $a=x\in\Gamma$ and $b=y\in\Gamma$).

Then, there are $f : \mathcal{M} \to \mathbb{R}^{|\mathcal{M}|}$, $\sigma : E \to [0,\infty)$, and $\Lambda : \mathcal{M} \to (0,\infty]$ that have the following properties:

- G *is (rC)-compatible with f and σ;*
- *If $x, y \in \mathcal{M}$ belong to the same connected component of G, then*

$$d_{\mathcal{M}}(x,y) \geqslant \tau \implies C\|f(x)-f(y)\|_2 \geqslant \max\{\Lambda(x),\Lambda(y)\}, \tag{60}$$

and

$$\{x,y\} \in E \implies \Lambda(y) \leqslant 2\Lambda(x) \qquad \text{and} \qquad 4\sigma(\{x,y\}) \leqslant \min\{\Lambda(x),\Lambda(y)\}. \tag{61}$$

The proof of Proposition 38 appears in Section 6: it consists of an application of Proposition 37 to φ as in (3), together with studying some basic implications of being (s,ε)-quasisymmetrically Hilbertian, in the spirit of (but not identical to) the foundational work in Section 2 of [TV80].

The upshot of Proposition 38 is that it provides the compatibility of the graph that we care about with f and σ, which is the assumption of Theorem 35, but now we can ignore the specific choice of σ from Proposition 37 that appears in (56), and instead we know that σ is controlled as in the second inequality in (61) by a function Λ that satisfies the assumptions of Proposition 36. Thus, Proposition 38 will allow us to combine Theorem 35 and Proposition 36 to obtain the following theorem:

Theorem 39. *There exists a universal constant $\alpha_0 \geqslant 1$ with the following properties. Fix $0 < s,\varepsilon \leqslant \frac{1}{2}$ and $\alpha \geqslant \alpha_0$. Suppose that $\beta > 0$ satisfies*

$$\beta \leqslant \beta_\alpha(s,\varepsilon) \stackrel{\mathrm{def}}{=} s^{\frac{\alpha}{\varepsilon}}. \tag{62}$$

Let $(\mathcal{M}, d_{\mathcal{M}})$ be a finite metric space that is (s,ε)-Hilbertian. Suppose that $0 < \tau \leqslant \mathrm{diam}(\mathcal{M})$ and that ω is a symmetric probability measure on $\mathcal{M}\times\mathcal{M}$ whose support is contained in $\{(x,y) \in \mathcal{M}\times\mathcal{M} : d_{\mathcal{M}}(x,y) \geqslant \tau\}$. In other words, $\omega(\mathcal{M}\times\mathcal{M}) = 1$, for every $x,y \in \mathcal{M}$ we have $\omega(x,y) = \omega(y,x)$, and $\omega(x,y) > 0 \implies d_{\mathcal{M}}(x,y) \geqslant \tau$. Then, for every $C \geqslant 1$ and $v \in \mathbb{R}^{|\mathcal{M}|}$ there exist nonempty subsets $A^*(v) = A^*_{\omega,C}(v), B^*(v) = B^*_{\omega,C}(v)$ of $\mathcal{M}$ such that the set-valued mapping $(v \in \mathbb{R}^n) \mapsto (A^*(v), B^*(v))$ is Borel-measurable, they satisfy

$$\forall\, v \in \mathbb{R}^{|\mathcal{M}|},\ \forall (x,y) \in A^*(v)\times B^*(v), \qquad d_{\mathcal{M}}(x,y) > \frac{\beta\tau}{\min\{\rho(x),\rho(y)\}}, \tag{63}$$

where $\rho : \mathcal{M} \to [1,\infty)$ is defined by

$$\forall\, x \in \mathcal{M}, \qquad \rho(x) \stackrel{\mathrm{def}}{=} 1 + \frac{1}{\alpha C}\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}, \tag{64}$$

and furthermore, if $\kappa > 1$ is the universal constant from Proposition 36, then

$$\int_{\mathbb{R}^n} \omega\big(A^*(v)\times B^*(v)\big)\,\mathrm{d}\gamma_n(v) \gtrsim e^{-\kappa C^2}. \tag{65}$$

The conclusion of Theorem 3 about random zero sets is a simple formal consequence of the fact that for every fixed probability measure ω as in Theorem 39, there exist random pairs of sets as in Theorem 39 that are separated per (63) and ω-large per (65). The reason for this is mainly duality (minimax theorem), together with a simple scale gluing argument; the (short) details of this deduction appear in Section 8. The idea to insert into the ARV reasoning a "weighting" such as ω on pairs of points of $\mathcal{M}$ is a key insight of [CGR05], were it was introduced in order to prove that any n-point metric space of negative type embeds into ℓ_2 with distortion $O((\log n)^{3/4})$; this idea played the same role in [ALN05], as well as herein.

The proof that Theorem 39 follows from Theorem 35, Proposition 36, and Proposition 38, which, as we explained above, is all that remains to complete the proof of Theorem 3, appears in Section 7. We will next conclude this section by sketching in broad strokes the reason why this works.

Write $n = |\mathcal{M}|$. Start by applying Proposition 38 with $r = \zeta\alpha$, where ζ is the universal constant in (59), thus ensuring that (64) coincides with (59). Henceforth in this sketch, $\mathsf{G} = (\mathcal{M}, E)$ will stand for the graph from this application of Proposition 38, i.e., its edges are given by (58) where ρ is defined in (59). The first bullet point in the conclusion of Proposition 38 makes it possible to apply Theorem 35 to get that

$$\int_{\mathbb{R}^n} \nu\big(\mathsf{G}(v; f,\sigma)\big)\,\mathrm{d}\gamma_n(v) \lesssim e^{-\frac{r^2}{4}C^2} n = e^{-\frac{\zeta^2\alpha^2}{4}C^2} n. \tag{66}$$

Think of the expectation estimate (66) as expressing the following structural information about the Euclidean sparsification $\mathsf{G}(v; f,\sigma)$ of G in a typical direction $v \in \mathbb{R}^n$: for such v the graph $\mathsf{G}(v; f,\sigma)$ is "clustered" in the sense that it cannot have a large collection of disjoint edges, and hence there is a small set of vertices (i.e., a small subset of $\mathcal{M}$) which is incident to all of the edges in $\mathsf{G}(v; f,\sigma)$. Even though this is not quite how we will use Theorem 35 in Section 7, namely, we will actually use a similar statement about fractional matchings of $\mathsf{G}(v; f,\sigma)$ that is a simple formal consequence of Theorem 35, for the purpose of intuitively understanding within the present sketch the reason why the deduction of Theorem 39 works, it suffices to initially consider the above combinatorial implication of (66).

Conclusion (60) of Proposition 38 is stronger than assumption (49) of Proposition 36 with f replaced by Cf. Also, the first inequality in conclusion (61) of Proposition 38 coincides with assumption (48) of Proposition 36. We may therefore proceed to apply Proposition 36 with f replaced by Cf to get subsets $A(v), B(v) \subseteq \mathcal{M}$ for each $v \in \mathbb{R}^n$ such that (52) holds and, by canceling C in (51) with f replaced by Cf,

$$\forall\, v \in \mathbb{R}^n,\ \forall (x,y) \in A(v)\times B(v), \qquad \{x,y\}\in E \implies |\langle v, f(x)-f(y)\rangle| > \max\big\{\Lambda(x),\Lambda(y)\big\}. \tag{67}$$

The next observation is crucial. Recalling Definition 33, by the second inequality in conclusion (61) of Proposition 38, it follows from (67) that for every $v \in \mathbb{R}^n$, if $(x,y) \in A(v)\times B(v)$ and $\{x,y\}\in E$, then $\{x,y\}$ is also an edge of $\mathsf{G}(v; f,\sigma)$. Equivalently, if $(x,y) \in A(v)\times B(v)$ yet $\{x,y\}$ is not an edge of $\mathsf{G}(v; f,\sigma)$, then necessarily $\{x,y\}\notin E$, i.e., by the definition (58) of E the desired inequality in (63) holds.

Per the above discussion, for typical $v \in \mathbb{R}^n$ there is a small subset of $\mathcal{M}$ that is incident to all of the edges in $\mathsf{G}(v; f,\sigma)$, so by removing it we get large subsets $A^*(v) \subseteq A(v)$ and $B^*(v) \subseteq B^*(v)$ such that (63) holds. The notion of "large" here must be interpreted as the size of $A^*(v)\times B^*(v)$ with respect to the given measure ω on $\mathcal{M}\times\mathcal{M}$, since the input to this reasoning (supplied by Proposition 36) is (52); this is why we will actually work with a weighted version of (66) for fractional matchings, but, as we stated above, it is a simple formal consequence Theorem 35 that is quickly deduced in Section 7. Finally, for the above reasoning to succeed, the lower bound in (52) needs to dominate the upper bound in (66); this is why in Theorem 39 α is assumed to be at least a sufficiently large universal constant α_0 (given in (241) below). The corresponding number-crunching is carried out in Section 7.

3. Proof of Theorem 35

As we explained in Section 2, in this section we will prove Theorem 35 by following the strategy in [Rot16]. To start with, the following simple lemma is a straightforward generalization of [Rot16, Lemma 10]:

Lemma 40. *For every $x \in V$ and $L, R \geqslant 0$ define $F^{\mathsf{G},f}_{x,R}: \mathbb{R}^n \to \mathbb{R}$ by setting*

$$\forall\, v \in \mathbb{R}^n, \qquad F^{\mathsf{G},f}_{x,R}(v) \stackrel{\mathrm{def}}{=} \max_{y\in B_{\mathsf{G}}(x,R)} \langle f(y)-f(x), v\rangle. \tag{68}$$

Then, $F^{\mathsf{G},f}_{x,R}$ is L-Lipschitz (as a function from ℓ_2^n to $\mathbb{R}$) provided that the following inclusion holds:

$$f\big(B_{\mathsf{G}}(x,R)\big) \subseteq B_{\ell_2^n}\big(f(x),L\big), \tag{69}$$

Proof. For every $u, v \in \mathbb{R}^n$ we have

$$
\begin{aligned}
F^{\mathsf{G},f}_{x,R}(u) &= \max_{y\in B_{\mathsf{G}}(x,R)} \big(\langle f(y)-f(x),v\rangle + \langle f(y)-f(x),u-v\rangle\big)\\
&\leqslant \max_{y\in B_{\mathsf{G}}(x,R)} \langle f(y)-f(x),v\rangle + \max_{y\in B_{\mathsf{G}}(x,R)} \langle f(y)-f(x),u-v\rangle\\
&= F^{\mathsf{G},f}_{x,R}(v) + \max_{y\in B_{\mathsf{G}}(x,R)} \langle f(y)-f(x),u-v\rangle\\
&\leqslant F^{\mathsf{G},f}_{x,R}(v) + \left(\max_{y\in B_{\mathsf{G}}(x,R)} \|f(x)-f(y)\|_2\right)\|u-v\|_2\\
&\leqslant F^{\mathsf{G},f}_{x,R}(v) + L\|u-v\|_2,
\end{aligned}
\tag{70}
$$

where we used the definition (68) in the first and third steps of (70), the fourth step of (70) is an application of Cauchy–Schwarz, and the final step of (70) is a restatement of the assumption (69). □

We will henceforth denote the γ_n-mean of the function $F^{\mathsf{G},f}_{x,R}$ that is given in (68) by $\mathbb{E}^{\mathsf{G},f}_{x,R}$, i.e.,

$$
\mathbb{E}^{\mathsf{G},f}_{x,R} \stackrel{\mathrm{def}}{=} \int_{\mathbb{R}^n} F^{\mathsf{G},f}_{x,R}\,\mathrm{d}\gamma_n. \tag{71}
$$

Note in passing that the definition (68) implies the following point-wise monotonicity in the parameter R that holds for every fixed vertex $x\in V$ and every fixed vector $v\in\mathbb{R}^n$:

$$
\forall 0\leqslant R_1\leqslant R_2,\qquad F^{\mathsf{G},f}_{x,R_1}(v)\leqslant F^{\mathsf{G},f}_{x,R_2}(v). \tag{72}
$$

By integrating (72) with respect to γ_n, we record for ease of later use the following basic property of (71):

$$
\forall x\in V,\ \forall 0\leqslant R_1\leqslant R_2,\qquad \mathbb{E}^{\mathsf{G},f}_{x,R_1}\leqslant \mathbb{E}^{\mathsf{G},f}_{x,R_2}. \tag{73}
$$

We will also need some control on how (71) varies as we change the vertex x along edges of G; this appears in the following simple lemma, which is a straightforward generalization of [Rot16, Lemma 11]:

Lemma 41. *For every $R\geqslant 0$, every $x\in V$, every $y\in N_{\mathsf{G}}(x)$, and every $v\in\mathbb{R}^n$ we have*

$$
F^{\mathsf{G},f}_{x,R+1}(v)\geqslant F^{\mathsf{G},f}_{y,R}(v)+\langle f(y)-f(x),v\rangle.
$$

Consequently, $\mathbb{E}^{\mathsf{G},f}_{x,R+1}\geqslant \mathbb{E}^{\mathsf{G},f}_{y,R}$, as seen by integrating this point-wise inequality with respect to γ_n.

Proof. Simply observe that $B_{\mathsf{G}}(y,R)\subseteq B_{\mathsf{G}}(x,R+1)$, as $y\in N_{\mathsf{G}}(x)$, and therefore

$$
\begin{aligned}
F^{\mathsf{G},f}_{x,R+1}(v) &\stackrel{(68)}{=} \max_{z\in B_{\mathsf{G}}(x,R+1)}\langle f(z)-f(x),v\rangle \geqslant \max_{z\in B_{\mathsf{G}}(y,R)}\langle f(z)-f(x),v\rangle\\
&= \langle f(y)-f(x),v\rangle + \max_{z\in B_{\mathsf{G}}(y,R)}\langle f(z)-f(y),v\rangle \stackrel{(68)}{=} \langle f(y)-f(x),v\rangle + F^{\mathsf{G},f}_{y,R}(v).
\end{aligned}
$$
□

The following technical lemma contains a trichotomy that is crucial for a subsequent induction step:

Lemma 42. *Fix $C>0$, $k\in\mathbb{N}\cup\{0\}$, and $n\in\mathbb{N}$. Let $\mathsf{G}=(V,E)$ be a graph, equipped with mappings*

$$
f:V\to\mathbb{R}^n,\qquad \Delta:V\to[0,\infty),\qquad K:V\to\mathbb{N},\qquad \sigma:E\to[0,\infty),
$$

for which

$$
\forall x\in V,\qquad f\Big(B_{\mathsf{G}}\big(x,K(x)\big)\Big)\subseteq B_{\ell_2^n}\Big(f(x),\frac{1}{C}\Delta(x)\Big). \tag{74}
$$

Assume that for every $v\in\mathbb{R}^n$ there exists a set of vertices $A_v\subseteq V$ and a one-to-one function $\phi_v:A_v\to V$ that satisfy the following properties. Firstly, we require that

$$
\forall v\in\mathbb{R}^n,\ \forall x\in A_v,\qquad \{\phi_v(x),x\}\in E. \tag{75}
$$

Secondly, we require that

$$\forall v \in \mathbb{R}^n,\ \forall x \in A_v, \qquad \left\langle f(x) - f\big(\phi_v(x)\big), v\right\rangle > \Delta(x) + \Delta\big(\phi_v(x)\big) + 2\sigma\big(\{x, \phi_v(x)\}\big). \tag{76}$$

Thirdly, we require that for every $x \in V$ the set $\{v \in \mathbb{R}^n : x \in A_v\}$ is Lebesgue-measurable and it satisfies

$$\gamma_n\big(\{v \in \mathbb{R}^n : x \in A_v\}\big) \geqslant 3e^{-\frac{1}{4}C^2}. \tag{77}$$

Then, for every $\emptyset \neq U \subseteq V$ at least one of the following three (mutually exclusive) properties must hold:

(1) There exists $x \in N_{\mathsf{G}}(U)$ for which $K(x) \leqslant k$;

(2) We have

$$|N_{\mathsf{G}}(U)| > e^{\frac{C^2}{4}}|U| \qquad \text{and} \qquad \forall x \in N_{\mathsf{G}}(U), \qquad K(x) \geqslant k+1, \tag{78}$$

and furthermore[23]

$$\forall x \in N_{\mathsf{G}}(U), \qquad \mathbb{E}^{\mathsf{G},f}_{x,k+1} \geqslant \max_{y \in U \cap N_G(x)} \mathbb{E}^{\mathsf{G},f}_{y,k}. \tag{79}$$

(3) We have

$$|N_{\mathsf{G}}(U)| \leqslant e^{\frac{C^2}{4}}|U| \qquad \text{and} \qquad \forall x \in N_{\mathsf{G}}(U), \qquad K(x) \geqslant k+1, \tag{80}$$

and furthermore there exists a subset $\emptyset \neq W \subseteq N_{\mathsf{G}}(U)$ for which

$$|W| > e^{-\frac{C^2}{4}}|U| \qquad \text{and} \qquad \forall x \in W, \qquad \mathbb{E}^{\mathsf{G},f}_{x,k+1} > \min_{\substack{y \in U \\ \{x,y\} \in E}} \Big(\mathbb{E}^{\mathsf{G},f}_{y,k} + 2\sigma(\{x,y\})\Big). \tag{81}$$

Proof. Suppose that Case *(1)* does not hold, which means that (as K takes values in $\mathbb{N}$) we have

$$\forall x \in N_{\mathsf{G}}(U), \qquad K(x) \geqslant k+1. \tag{82}$$

In other words, the second requirement in (78) and (80) holds automatically if Case *(1)* does not hold.

If $|N_{\mathsf{G}}(U)| > e^{C^2/4}|U|$ in addition to (82), then we claim that Case *(2)* holds. For this, we need to verify (79). Indeed, suppose that $x \in N_{\mathsf{G}}(U)$ and consider any $y \in U \cap N_{\mathsf{G}}(x)$. By Lemma 41 we have

$$\mathbb{E}^{\mathsf{G},f}_{x,k+1} \geqslant \mathbb{E}^{\mathsf{G},f}_{y,k}. \tag{83}$$

Since y is an arbitrary vertex in $U \cap N_{\mathsf{G}}(x)$, this proves (79).

It thus remains to assume that (80) holds and to then demonstrate the rest of Case *(3)*, i.e., to show that there exists $W \subseteq N_{\mathsf{G}}(U)$ that satisfies (81). To this end, for each $v \in \mathbb{R}^n$ define $S_v \subseteq V$ by

$$S_v \stackrel{\mathrm{def}}{=} \Big\{x \in A_v \cap U : F^{\mathsf{G},f}_{x,k}(v) \geqslant \mathbb{E}^{\mathsf{G},f}_{x,k} - \Delta(x)\Big\}. \tag{84}$$

Observe that the following inclusion holds for every $x \in U$:

$$\big\{v \in \mathbb{R}^n : x \in A_v\big\} \subseteq \big\{v \in \mathbb{R}^n : x \in S_v\big\} \bigcup \Big\{v \in \mathbb{R}^n : F^{\mathsf{G},f}_{x,k}(v) < \mathbb{E}^{\mathsf{G},f}_{x,k} - \Delta(x)\Big\}. \tag{85}$$

We therefore have

$$3e^{-\frac{1}{4}C^2} \stackrel{(77)}{\leqslant} \gamma_n\big(\{v \in \mathbb{R}^n : x \in A_v\}\big) \stackrel{(85)}{\leqslant} \gamma_n\big(\{v \in \mathbb{R}^n : x \in S_v\}\big) + \gamma_n\big(\{v \in \mathbb{R}^n : F^{\mathsf{G},f}_{x,k}(v) < \mathbb{E}^{\mathsf{G},f}_{x,k} - \Delta(x)\}\big). \tag{86}$$

By the second part of our current assumption (80), every $x \in N_{\mathsf{G}}(U)$ satisfies $K(x) \geqslant k+1$. Hence

$$\forall x \in N_{\mathsf{G}}(U), \qquad f\big(B_{\mathsf{G}}(x,k)\big) \subseteq f\big(B_{\mathsf{G}}(x,k+1)\big) \subseteq f\Big(B_{\mathsf{G}}\big(x,K(x)\big)\Big) \stackrel{(74)}{\subseteq} B_{\ell_2^n}\Big(f(x), \frac{1}{C}\Delta(x)\Big).$$

By Lemma 40 this implies that

$$\forall x \in N_{\mathsf{G}}(U), \qquad \big\|F^{\mathsf{G},f}_{x,k}\big\|_{\mathrm{Lip}(\ell_2^n)} \leqslant \frac{\Delta(x)}{C} \quad \text{and} \quad \big\|F^{\mathsf{G},f}_{x,k+1}\big\|_{\mathrm{Lip}(\ell_2^n)} \leqslant \frac{\Delta(x)}{C}. \tag{87}$$

[23]The maximum in (79) is well-defined because $U \cap N_G(x) \neq \emptyset$ if (and only if) $x \in N_{\mathsf{G}}(U)$.

The Gaussian isoperimetric theorem [Bor75, ST78] (see [Led01, Corollary 2.6]) therefore implies[24] that

$$\forall x \in N_{\mathsf{G}}(U), \qquad \gamma_n\Big(\{v \in \mathbb{R}^n :\ F^{\mathsf{G},f}_{x,k}(v) < \mathbb{E}^{\mathsf{G},f}_{x,k} - \Delta(x)\}\Big) \leqslant e^{-\frac12 C^2}, \tag{88}$$

and also

$$\forall x \in N_{\mathsf{G}}(U), \qquad \gamma_n\Big(\{v \in \mathbb{R}^n :\ F^{\mathsf{G},f}_{x,k+1}(v) > \mathbb{E}^{\mathsf{G},f}_{x,k+1} + \Delta(x)\}\Big) \leqslant e^{-\frac12 C^2}. \tag{89}$$

In particular, a substitution of (88) into (86) gives the following statement:

$$\forall x \in U, \qquad \gamma_n\Big(\{v \in \mathbb{R}^n :\ x \in S_v\}\Big) \geqslant 3e^{-\frac14 C^2} - e^{-\frac12 C^2} > 2e^{-\frac14 C^2}. \tag{90}$$

Define $T_v \subseteq V$ for every $v \in \mathbb{R}^n$ by

$$T_v \stackrel{\mathrm{def}}{=} \phi_v(S_v). \tag{91}$$

As $\phi_v : A_v \to V$ is assumed to be one-to-one, we have

$$\forall v \in \mathbb{R}^n, \qquad |T_v| = |S_v|. \tag{92}$$

The definition (84) ensures that $S_v \subseteq A_v$, so $T_v \subseteq N_{\mathsf{G}}(S_v)$ by (75). By (84) also $S_v \subseteq U$, so

$$T_v \subseteq N_{\mathsf{G}}(U). \tag{93}$$

Define $W \subseteq N_{\mathsf{G}}(U)$ by

$$W \stackrel{\mathrm{def}}{=} \Big\{x \in N_{\mathsf{G}}(U) :\ \gamma_n\Big(\{v \in \mathbb{R}^n :\ x \in T_v\}\Big) > e^{-\frac12 C^2}\Big\}. \tag{94}$$

With this choice, we can reason as follows (the second step below is an application of Fubini):

$$\begin{aligned}
2e^{-\frac14 C^2}|U| &\stackrel{(90)}{\leqslant} \sum_{x\in U} \gamma_n\Big(\{v \in \mathbb{R}^n :\ x \in S_v\}\Big)\\
&= \int_{\mathbb{R}^n} |S_v|\,\mathrm{d}\gamma_n\\
&\stackrel{(92)}{=} \int_{\mathbb{R}^n} |T_v|\,\mathrm{d}\gamma_n\\
&\stackrel{(93)}{=} \sum_{x\in N_{\mathsf{G}}(U)\smallsetminus W} \gamma_n\Big(\{v \in \mathbb{R}^n :\ x \in T_v\}\Big) + \sum_{x\in W} \gamma_n\Big(\{v \in \mathbb{R}^n :\ x \in T_v\}\Big)\\
&\stackrel{(94)}{\leqslant} e^{-\frac12 C^2}|N_{\mathsf{G}}(U)\smallsetminus W| + |W|\\
&= e^{-\frac12 C^2}|N_{\mathsf{G}}(U)| + \Big(1 - e^{-\frac12 C^2}\Big)|W|\\
&\stackrel{(80)}{\leqslant} e^{-\frac14 C^2}|U| + \Big(1 - e^{-\frac12 C^2}\Big)|W|.
\end{aligned}$$

This simplifies to give the following justification of the first requirement in (81) for our choice (94) of W:

$$|W| \geqslant \frac{e^{-\frac14 C^2}}{1 - e^{-\frac12 C^2}}|U| > e^{-\frac14 C^2}|U|.$$

To complete the proof of Lemma 42, it remains to derive the second requirement in (81); note that we did not use thus far the remaining assumption (76), but we will do so for this purpose. Consider $x \in W$. The definition (94) of W ensures that also $x \in N_{\mathsf{G}}(U)$, and therefore the estimate in (89) holds. Hence,

$$\gamma_n\Big(\{v \in \mathbb{R}^n :\ F^{\mathsf{G},f}_{x,k+1}(v) > \mathbb{E}^{\mathsf{G},f}_{x,k+1} + \Delta(x)\}\Big) \stackrel{(89)}{\leqslant} e^{-\frac12 C^2} \stackrel{(94)}{<} \gamma_n\Big(\{v \in \mathbb{R}^n :\ x \in T_v\}\Big).$$

This implies in particular that

$$\exists v \in \mathbb{R}^n, \qquad x \in T_v \qquad \text{and} \qquad F^{\mathsf{G},f}_{x,k+1}(v) \leqslant \mathbb{E}^{\mathsf{G},f}_{x,k+1} + \Delta(x). \tag{95}$$

[24]Formally, the way Gaussian isoperimetry is typically stated in the literature implies (88) and (89) when the upper bounds on the Lipschitz constants in (87) are strictly positive, i.e., when $\Delta(x) > 0$. Nevertheless, when $\Delta(x) = 0$ by (87) we get that the functions $F^{\mathsf{G},f}_{x,k}$ and $F^{\mathsf{G},f}_{x,k+1}$ are constant, so the probabilities in the left hand sides of (88) and (89) vanish.

From now, fix $v \in \mathbb{R}^n$ for which the two requirements in (95) hold. By the definition (91) of T_v, there exists (a unique) $y \in S_v$ such that $x = \phi_v(y)$. By the definition (84) of S_v we know that

$$F_{y,k}^{\mathsf{G},f}(v) \geqslant \mathbb{E}_{y,k}^{\mathsf{G},f} - \Delta(y).$$

Recalling the definition (68) of $F_{y,k}^{\mathsf{G},f}$, it follows that

$$\exists z \in B_{\mathsf{G}}(y,k), \qquad \langle f(z) - f(y), v\rangle \geqslant \mathbb{E}_{y,k}^{\mathsf{G},f} - \Delta(y). \tag{96}$$

We will also fix from now any $z \in V$ that satisfies (96). The definition (76) of S_v ensures that $y \in A_v$, so we may apply the assumption 76 while recalling that y was chosen to satisfy $x = \phi_v(y)$, to get the estimate

$$\langle f(y) - f(x), v\rangle > \Delta(y) + \Delta(x) + 2\sigma(\{y,x\}). \tag{97}$$

Consequently,

$$\langle f(z) - f(x), v\rangle \overset{(96)\wedge(97)}{>} \mathbb{E}_{y,k}^{\mathsf{G},f} + \Delta(x) + 2\sigma(\{y,x\}). \tag{98}$$

But $z \in B_{\mathsf{G}}(x, k+1)$, as $z \in B_{\mathsf{G}}(y,k)$ and $\{x,y\} \in E$. So, by the definition (68) of $F_{x,k+1}^{\mathsf{G},f}$, from (98) we get

$$F_{x,k+1}^{\mathsf{G},f}(v) > \mathbb{E}_{y,k}^{\mathsf{G},f} + \Delta(x) + 2\sigma(\{y,x\}). \tag{99}$$

In combination with the second condition in (95), it follows from (99) that

$$\mathbb{E}_{x,k+1}^{\mathsf{G},f} > \mathbb{E}_{y,k}^{\mathsf{G},f} + 2\sigma(\{y,x\}).$$

Because $y \in S_v \subseteq U$ and $\{x,y\} \in E$, this completes the proof of the second requirement in (81). $\square$

In the proof of Theorem 35 we will use the following convenient notation for every $x \in V$ and $R \geqslant 1$:

$$m_\sigma(x,R) \overset{\text{def}}{=} \min_{\substack{\{y,z\}\in E \\ y\in B_{\mathsf{G}}(x,R-1)}} \sigma(\{y,z\}) \tag{100}$$

Also, for $R < 1$ we denote $m_\sigma(x,R) = 0$. Observe that (by definition) we have

$$\forall x \in V,\ \forall 0 \leqslant R_1 \leqslant R_2, \qquad m_\sigma(x,R_2) \leqslant m_\sigma(x,R_1) \qquad \text{and} \qquad \forall \{x,y\} \in E, \qquad m_\sigma(x,2) \leqslant \sigma(\{x,y\}). \tag{101}$$

Thus, in particular, we have

$$\forall R \geqslant 2,\ \forall \{x,y\} \in E, \qquad m_\sigma(x,R) \leqslant \sigma(\{x,y\}). \tag{102}$$

Furthermore, the definition (100) immediately implies that

$$\forall R \geqslant 1,\ \forall x \in V,\ \forall y \in N_{\mathsf{G}}(x), \qquad m_\sigma(x,R+1) \leqslant m_\sigma(y,R). \tag{103}$$

Remark 43. *It is worthwhile to observe for ease of later reference that using the notions in* (71) *and* (100) *we can restate conditions* (44) *and* (45) *from Definition 34 of C-compatibility as the following requirements:*

$$\forall x \in V,\ \forall y \in N_{\mathsf{G}}(x), \qquad \Delta(x) \leqslant m_\sigma\big(x, K(x)\big) \qquad \text{and} \qquad \mathbb{E}_{x,K(y)}^{\mathsf{G},f} \leqslant K(x)\Delta(y).$$

We are now ready to prove Theorem 35:

Proof of Theorem 35. We will begin with quick setup that (straightforwardly) provides measurability and symmetry requirements that are needed for the ensuing reasoning. Definition (42) ensures the mapping

$$E(\cdot; f,\sigma) : \mathbb{R}^n \to 2^E$$

is Borel measurable. Also, by definition $E(-v; f,\sigma) = E(v; f,\sigma)$ for every $v \in \mathbb{R}^n$. For each $v \in \mathbb{R}^n$, let $\mathcal{M}(v)$ be the collection of all the maximal-size matchings of the graph $\mathsf{G}(v; f,\sigma)$. Since V is finite, the mapping

$$\mathcal{M} : \mathbb{R}^n \to 2^{2^E} \subseteq 2^{2^{2^V}}$$

is also Borel-measurable and satisfies $\mathcal{M}(-v) = \mathcal{M}(v)$ for every $v \in \mathbb{R}^n$. Fix any linear ordering $\prec$ on the (finite) set of matchings of G and for each $v \in \mathbb{R}^n$ let $M(v)$ be the element of $\mathcal{M}(v)$ that is minimal with respect to $\prec$. The mapping $v \mapsto M(v)$ is Borel-measurable and satisfies $M(-v) = M(v)$ for every $v \in \mathbb{R}^n$. Furthermore, this construction ensures that for every fixed edge $e \in E$, the set $\{v \in \mathbb{R}^n : e \in M(v)\}$ is Borel.

Suppose for the purpose of obtaining a contradiction that

$$\int_{\mathbb{R}^n} |M(v)| \,\mathrm{d}\gamma_n(v) \geqslant 6e^{-\frac{1}{4}C^2}|V|. \tag{104}$$

We will start by performing the following iterative procedure. Denote $V_1 = V$ and $M_1(v) = M(v)$ for every $v \in \mathbb{R}^n$. Suppose inductively that $V_j \subseteq V$ and $\{M_j(v)\}_{v\in\mathbb{R}^n}$ have been defined for some $j \in \mathbb{N}$, where for each $v \in \mathbb{R}^n$ we require that $M_j(v) \subseteq M(v)$ is a matching of the following graph:

$$\mathsf{G}_j(v; f, \sigma) \stackrel{\mathrm{def}}{=} \Big(V_j, E_j(v; f, \sigma) \stackrel{\mathrm{def}}{=} \big\{e \in E(v; f, \sigma) : e \subseteq V_j\big\}\Big). \tag{105}$$

Furthermore, we require that the mapping M_j is Borel-measurable, satisfies $M_j(-v) = M_j(v)$ for every $v \in \mathbb{R}^n$, and for every fixed edge $e \in E$, the set $\{v \in \mathbb{R}^n : e \in M_j(v)\}$ is Borel.

If there does not exist $x_j \in V_j$ that satisfies

$$\gamma_n\big(\{v \in \mathbb{R}^n : \exists e \in M_j(v) \text{ such that } x_j \in e\}\big) < 6e^{-\frac{1}{4}C^2}, \tag{106}$$

then terminate the construction. Otherwise, fix $x_j \in V_j$ for which (106) holds and denote

$$V_{j+1} \stackrel{\mathrm{def}}{=} V_j \smallsetminus \{x_j\} \qquad \text{and} \qquad \forall\, v \in \mathbb{R}^n, \qquad M_{j+1}(v) \stackrel{\mathrm{def}}{=} M_j(v) \smallsetminus \big\{e \in M_j(v) : x_j \in e\big\}.$$

The above measurability requirements for M_{j+1} are immediate from this definition.

Note that for every $v \in \mathbb{R}^n$, since $M(v)$ is a matching of G and $M_j(v) \subseteq M(v)$, there is at most one edge $e \in M_j(v)$ for which $x_j \in e$. Consequently, we have the following point-wise estimate:

$$\forall\, v \in \mathbb{R}^n, \qquad |M_{j+1}(v)| \geqslant |M_j(v)| - \mathbf{1}_{\{\exists e \in M_j(v) \text{ such that } x_j \in e\}}.$$

By integrating this inequality with respect to γ_n and using our choice of x_j, we conclude that

$$\begin{aligned}\int_{\mathbb{R}^n} |M_{j+1}(v)| \,\mathrm{d}\gamma_n(v) &\geqslant \int_{\mathbb{R}^n} |M_j(v)| \,\mathrm{d}\gamma_n(v) - \gamma_n\big(\{v \in \mathbb{R}^n : \exists e \in M_j(v) \text{ such that } x_j \in e\}\big)\\ &> \int_{\mathbb{R}^n} |M_j(v)| \,\mathrm{d}\gamma_n(v) - 6e^{-\frac{1}{4}C^2}.\end{aligned} \tag{107}$$

Suppose that this procedure terminates at the $J \in \mathbb{N}$ step; this must occur because V is finite and each step removes a vertex, whence $J \leqslant |V| + 1$. By applying (107) inductively we see that

$$\int_{\mathbb{R}^n} |M_J(v)| \,\mathrm{d}\gamma_n(v) > \int_{\mathbb{R}^n} |M(v)| \,\mathrm{d}\gamma_n(v) - 6e^{-\frac{1}{4}C^2}(J-1) \geqslant \int_{\mathbb{R}^n} |M(v)| \,\mathrm{d}\gamma_n(v) - 6e^{-\frac{1}{4}C^2}|V| \stackrel{(104)}{\geqslant} 0. \tag{108}$$

Thanks to the strict inequality in the first step of (108), it follows in particular that $V_J \neq \emptyset$.[25] Furthermore, because the above procedure terminated at the $J \in \mathbb{N}$ step, we have

$$\forall\, x \in V_J, \qquad \gamma_n\big(\{v \in \mathbb{R}^n : \exists e \in M_J(v) \text{ such that } x \in e\}\big) \geqslant 6e^{-\frac{1}{4}C^2}. \tag{109}$$

We will next proceed to apply Lemma 42 (multiple times) to the following induced subgraph of G:

$$\mathsf{G}_J \stackrel{\mathrm{def}}{=} \big(V_J, E_J \stackrel{\mathrm{def}}{=} \{e \in E : e \subseteq V_J\}\big).$$

Theorem 35 assumes that G is C-compatible with f and σ, so let $K : V \to \mathbb{N}$ and $\Delta : V \to [0, \infty)$ be mappings as in Definition 34. Henceforth, we will slightly abuse notation by also denoting the restrictions of f, K, Δ to V_J by f, K, Δ, and correspondingly denoting the restriction of σ to E_J by σ. Observe that because G_J is an induced subgraph of G, we have $B_{\mathsf{G}_J}(x, R) \subseteq B_{\mathsf{G}}(x, R)$ for every $x \in V_J$ and $R \geqslant 0$. Thanks

[25] By increasing the factor 6 in (104) to, say, 7, and then repeating the above reasoning, we can ensure that the mean of $|M_J(v)|$ with respect to γ_n is at least $\exp(-C^2/4)|V|$, and hence also $|V_J| \gtrsim \exp(-C^2/4)|V|$, since $M_J(v)$ is a matching of a graph whose vertex set is V_J. We do not need such a guarantee below, but it is worthwhile to note it in case this will occur in the future.

to these inclusions, an inspection of Definition 34 reveals that as G is C-compatible with f and σ, also G_J is C-compatible with f and σ.

For each vector $v \in \mathbb{R}^n$ define $A_v \subseteq V_J$ by

$$A_v \stackrel{\mathrm{def}}{=} \big\{x \in V_J : \exists \{x,y\} \in M_J(v) \text{ such that } \langle f(x)-f(y), v\rangle > 4\sigma(\{x,y\})\big\}. \tag{110}$$

Because $M_J(v)$ is a matching of G, if $x \in A_v$, then there is a unique $y \in V_J$ such that $\{x,y\} \in M_J(v)$. We can therefore define $\phi_v(x) = y$. The fact that $M_J(v)$ is a matching of G also implies that $\phi_v : A_v \to V_J$ is one-to-one, and by design this definition ensures that assumption (75) of Lemma 42 is satisfied.

We next claim that the following equality of events holds for every fixed $x \in V_J$:

$$\big\{v \in \mathbb{R}^n : \exists e \in M_J(v) \text{ such that } x \in e\big\} = \big\{v \in \mathbb{R}^n :\ x \in A_v\big\} \cup \big\{v \in \mathbb{R}^n :\ x \in A_{-v}\big\}. \tag{111}$$

Indeed, it is immediate from the definition (110) of $\{A_v\}_{v\in\mathbb{R}^n}$ that the right hand side of (111) is contained in the left hand side of (111). The reverse inclusion follows from the fact that $M_J(v)$ is a matching of the graph $\mathsf{G}_J(v; f,\sigma)$ in (105), which is a subgraph of the Euclidean sparsification $\mathsf{G}(v; f,\sigma)$ of G. So, if $v \in \mathbb{R}^n$ and there is $e \in M_J(v)$ with $x \in e$, then $e \in E(v; f,\sigma)$. By the definition (42) of Euclidean sparsification, this means that if $y \in V_J$ is such that $e = \{x,y\}$, then $|\langle f(x)-f(y), v\rangle| > 4\sigma(\{x,y\})$. If $\langle f(x)-f(y), v\rangle > 4\sigma(\{x,y\})$, then $x \in A_v$, and otherwise $\langle f(x)-f(y), v\rangle < -4\sigma(\{x,y\})$, which means that $x \in A_{-v}$.

By (110), the sets in the right hand side of (111) are disjoint unless $v = 0$. Therefore, for every $x \in V_J$

$$\begin{aligned} 6e^{-\frac14 C^2} &\stackrel{(109)}{\leqslant} \gamma_n\big(\{v \in \mathbb{R}^n : \exists e \in M_J(v) \text{ such that } x \in e\}\big) \\ &\stackrel{(111)}{=} \gamma_n\big(\{v \in \mathbb{R}^n :\ x \in A_v\}\big) + \gamma_n\big(\{v \in \mathbb{R}^n :\ x \in A_{-v}\}\big) = 2\gamma_n\big(\{v \in \mathbb{R}^n :\ x \in A_v\}\big). \end{aligned}$$

Hence,

$$\forall x \in V_J, \qquad \gamma_n\big(\{v \in \mathbb{R}^n :\ x \in A_v\}\big) \geqslant 3e^{-\frac14 C^2}.$$

In other words, we checked that condition (77) of Lemma 42 is satisfied for G_J (the corresponding measurability requirement follows from the measurability that we ensured in the above construction).

For every $v \in \mathbb{R}^n$ and $x \in A_v$, by (a special case of) the condition (44) of Definition 34, which is part of the C-compatibility assumption of Theorem 35, we know that

$$2\sigma\big(\{x, \phi_v(x)\}\big) \geqslant \Delta(x) + \Delta\big(\phi_v(x)\big). \tag{112}$$

Consequently, we deduce as follows that condition (76) of Lemma 42 is satisfied:

$$\big\langle f(x) - f\big(\phi_v(x)\big), v\big\rangle \stackrel{(110)}{>} 4\sigma\big(\{x,\phi_v(x)\}\big) \stackrel{(112)}{\geqslant} \Delta(x) + \Delta\big(\phi_v(x)\big) + 2\sigma\big(\{x,\phi_v(x)\}\big).$$

Condition (74) of Lemma 42 coincides with condition (46) of Definition 34 with G replaced by G_J, so it holds thanks to the C-compatibility of G_J with f and σ. This completes the verification of all of the assumptions of Lemma 42.

We will next apply Lemma 42 iteratively as follows to obtain $T \in \mathbb{N}$, as well as nonempty subsets $\emptyset \neq U_0, U_1, \ldots, U_T \subseteq V_J$ and indices $a_1, \ldots, a_T \in \{2,3\}$, for which the following requirements hold. Firstly, at the start of the induction set $U_0 = V_J$ and $a_1 = 3$. Secondly, at the end of the iteration we have

$$\min_{x \in N_{\mathsf{G}_J}(U_T)} K(x) \leqslant T. \tag{113}$$

Thirdly, for every $t \in [T]$ (so, excluding $t = 0$), if we denote

$$d_t \stackrel{\mathrm{def}}{=} \big|\{s \in [t] :\ a_s = 3\}\big|, \tag{114}$$

then

$$\forall x \in U_t, \qquad K(x) \geqslant t \qquad \text{and} \qquad \mathbb{E}^{\mathsf{G}_J, f}_{x,t} > 2d_t m_\sigma(x,t). \tag{115}$$

Finally, we require that the following estimate holds:

$$\forall\, t \in [T], \qquad \frac{|U_t|}{|U_{t-1}|} > e^{\frac{C^2}{2}\left(\frac{1}{2}-\mathbf{1}_{\{a_t=3\}}\right)} = \begin{cases} e^{\frac{1}{4}C^2} & \text{if } a_t = 2, \\ e^{-\frac{1}{4}C^2} & \text{if } a_t = 3. \end{cases} \tag{116}$$

Supposing for the moment that the above construction has already been carried out, we will next see how it can be used to complete the proof of Theorem 35. By (113) we can fix $x \in N_{\mathsf{G}_J}(U_T)$, thus we can also fix $y \in U_T \cap N_{\mathsf{G}_J}(x)$, for which $K(x) \leqslant T$. As $y \in N_{\mathsf{G}_J}(x)$, by combining $K(x) \leqslant T$ with condition (44) of Definition 34, per its formulation in Remark 43, we see that

$$\mathbb{E}^{\mathsf{G}_J,f}_{y,K(y)} \leqslant T\Delta(y). \tag{117}$$

Because $y \in U_T$, the case $t = T$ of the first condition in (115) gives the bound $K(y) \leqslant T$, so by the (very simple) monotonicity properties that we recorded in (73) and in (101), we have

$$\mathbb{E}^{\mathsf{G}_J,f}_{y,T} \leqslant \mathbb{E}^{\mathsf{G}_J,f}_{y,K(y)}, \qquad \text{and} \qquad m_\sigma\big(y,K(y)\big) \leqslant m_\sigma(y,T). \tag{118}$$

Also, by Remark 43, the (first part of the) C-compatibility assumption of Theorem 35 gives the estimate

$$\Delta(y) \leqslant m_\sigma\big(y,K(y)\big). \tag{119}$$

By combining (117), (118) and (119), we conclude that

$$\mathbb{E}^{\mathsf{G}_J,f}_{y,T} \leqslant T m_\sigma(y,T). \tag{120}$$

At the same time, by applying (116) inductively we see that

$$\frac{|U_T|}{|U_0|} = \prod_{t=1}^{T} \frac{|U_t|}{|U_{t-1}|} \overset{(116)}{>} e^{\frac{C^2}{2}\left(\frac{T}{2}-\sum_{t=1}^{T}\mathbf{1}_{\{a_t=3\}}\right)} \overset{(114)}{=} e^{\frac{C^2}{2}\left(\frac{T}{2}-d_T\right)}. \tag{121}$$

Since $U_T \subseteq V_J = U_0$, the left hand side of (121) is at most 1, so, since $C > 0$, (121) implies that

$$d_T > \frac{T}{2}. \tag{122}$$

Because $y \in U_T$, a substitution of (122) into the case $t = T$ of the second part of (115) gives

$$\mathbb{E}^{\mathsf{G}_J,f}_{y,T} > T m_\sigma(y,T). \tag{123}$$

We thus arrive at the desired contradiction by contrasting (120) with (123), i.e., the contrapositive assumption (104), which is the premise of the current discussion, cannot hold. This completes the proof of Theorem 35 assuming that the aforementioned construction (of $T, U_0, U_1, \ldots, U_T, a_1, \ldots, a_T$ with the above specifications) can indeed be carried out; this is what we will justify next.

Recall that we already defined $U_0 = V_J$ and $a_1 = 3$. We will start by applying Lemma 42 with $U = U_0$ and $d = k = 0$. Case *(1)* of Lemma 42 does not hold because $k = 0$ and K takes values in $\mathbb{N}$. Case *(2)* of Lemma 42 does not hold as $U = U_0 = V_J$ and $C > 0$, so $|N_{\mathsf{G}_J}(V_J)| = |V_J| < e^{C^2/4}|V_J|$. Lemma 42 therefore ensures that its Case *(3)* holds, thus producing a set $\varnothing \neq W \subseteq N_{\mathsf{G}_J}(V_J) = V_J$. We will then define $U_1 = W$. With this notation, we thus know from the first inequality in (81) that

$$\frac{|U_1|}{|U_0|} > e^{-\frac{1}{4}C^2}.$$

Because $a_1 = 3$ (by definition), this coincides with the case $t = 1$ of (116). Furthermore, by (81) we have

$$\forall\, x \in U_1, \qquad \mathbb{E}^{\mathsf{G}_J,f}_{x,1} > 2 \min_{\substack{y \in V_J \\ \{x,y\} \in E_J}} \sigma(\{x,y\}) \overset{(100)}{=} 2m_\sigma(x,1).$$

As $d_1 = 1$ thanks to (114), this coincides with the case $t = 1$ of the second requirement in (115). The case $t = 1$ of the first requirement in (115) is automatic because K takes values in $\mathbb{N}$.

Assume inductively that for $t \in \mathbb{N}$ we already defined $\varnothing \neq U_0, U_1, \ldots, U_t \subseteq V_J$ and $a_1, \ldots, a_t \in \{2,3\}$ for which both (115) and (116) hold, with $d_1, \ldots, d_t$ given by (114). If there is $x \in N_{\mathsf{G}_J}(U_t)$ for which $K(x) \leqslant t$, then define $T = t$, thus ensuring that (113) is satisfied. Otherwise, by Lemma 42 applied with $U = U_t$ and $d = d_t$, either Case *(2)* of Lemma 42 holds, or Case *(3)* of Lemma 42 holds. We will encode this dichotomy by defining $a_{t+1} = 2$ if Case *(2)* holds, and defining $a_{t+1} = 3$ if Case *(3)* holds.

If Case *(2)* holds, then set $U_{t+1} = N_{\mathsf{G}_J}(U_t)$. By (114) we have $d_{t+1} = d_t$, as $a_{t+1} = 2$, so (116) holds with t replaced by $t+1$ thanks to the first condition in the conclusion (78) of Lemma 42 in this case. The rest of (78) is a restatement of the first part of (115) with t replaced by $t+1$, and the second part of (115) (also with t replaced by $t+1$) is derived as follows from the remaining conclusion (79) of Lemma 42 in Case *(2)*:

$$\mathbb{E}_{x,t+1}^{\mathsf{G}_J,f} \overset{(79)}{\geqslant} \min_{y\in U_t\cap N_{\mathsf{G}_J}(x)} \mathbb{E}_{y,t}^{\mathsf{G}_J,f} \overset{(115)}{>} \min_{y\in U_t\cap N_{\mathsf{G}_J}(x)} 2d_t m_\sigma(y,t) \overset{(103)}{\geqslant} 2d_{t+1}m_\sigma(x,t+1).$$

If Case *(3)* holds, then define $U_{t+1} = W$, where W is the subset of $N_{\mathsf{G}_J}(U_t)$ that Case ((3)) produces. Then, the second condition in (80) shows that the first requirement in (115) holds with t replaced by $t+1$. To show that the second requirement in (115) is satisfied with t replaced by $t+1$, observe that by (114) we have $d_{t+1} = d_t + 1$, as $a_{t+1} = 3$, so by the second condition in (81), we see that

$$\mathbb{E}_{x,t+1}^{\mathsf{G}_J,f} \overset{(81)}{>} \min_{\substack{y\in U_t\\ \{x,y\}\in E_J}} \left(\mathbb{E}_{y,t}^{\mathsf{G}_J,f} + 2\sigma(\{x,y\})\right) \overset{(115)}{>} \min_{\substack{y\in U_t\\ \{x,y\}\in E_J}} \left(2d_t m_\sigma(y,t) + 2\sigma(\{x,y\})\right) \overset{(102)\wedge(103)}{\geqslant} 2d_{t+1}m_\sigma(x,t+1).$$

This completes the derivation of (115) when Case *(3)* holds, thus completing the inductive step.

By the first condition in (115), the above iterative procedure cannot continue indefinitely, as it entails that $t \leqslant \max_{x\in V} K(x)$; the desired construction is completed when this iteration terminates. □

4. Probabilistic groundwork

Our goal here is to present the proof of Proposition 36, which consists of elementary probabilistic considerations. We start with the following simple general lemma:

Lemma 44. *Fix $\beta > 0$. Let X be a random variable, defined on some probability space $(\Omega,\mathbb{P})$, which has a density φ that is nondecreasing on $(-\infty,0]$ and nonincreasing on $[0,\infty)$. Suppose that $f:\mathbb{R}\to[0,\infty)$ is β-periodic, i.e, $f(t+\beta) = f(t)$ for every $t\in\mathbb{R}$, and also that f is integrable on the interval $[0,\beta]$. Then,*

$$\mathbb{E}\big[f(\mathsf{X})\big] \geqslant \left(\frac{1}{\beta}\int_0^\beta f(t)\,\mathrm{d}t\right)\mathbb{P}\big[|\mathsf{X}|\geqslant\beta\big].$$

Proof. Simply note that for every $k\in\mathbb{N}$ we have

$$\begin{aligned}\int_{(k-1)\beta}^{k\beta} f(x)\varphi(x)\,\mathrm{d}x &\geqslant \int_{(k-1)\beta}^{k\beta} f(x)\varphi(k\beta)\,\mathrm{d}x\\ &= \left(\int_0^\beta f(t)\,\mathrm{d}t\right)\varphi(k\beta) \geqslant \left(\int_0^\beta f(t)\,\mathrm{d}t\right)\frac{1}{\beta}\int_{k\beta}^{(k+1)\beta}\varphi(x)\,\mathrm{d}x,\end{aligned}\tag{124}$$

and

$$\begin{aligned}\int_{-k\beta}^{-(k-1)\beta} f(x)\varphi(x)\,\mathrm{d}x &\geqslant \int_{k\beta}^{-(k-1)\beta} f(x)\varphi(-k\beta)\,\mathrm{d}x\\ &= \left(\int_0^\beta f(t)\,\mathrm{d}t\right)\varphi(-k\beta) \geqslant \left(\int_0^\beta f(t)\,\mathrm{d}t\right)\frac{1}{\beta}\int_{-(k+1)\beta}^{-k\beta}\varphi(x)\,\mathrm{d}x,\end{aligned}\tag{125}$$

where in (124) and (125) the first and third steps use the monotonicity assumptions on φ, and the second steps use the β-periodicity of f. By summing (124) and (125) over $k\in\mathbb{N}$ we conclude that

$$\mathbb{E}\big[f(\mathsf{X})\big] = \int_{-\infty}^{\infty} f(x)\varphi(x)\,\mathrm{d}x \geqslant \left(\frac{1}{\beta}\int_0^\beta f(t)\,\mathrm{d}t\right)\int_{\mathbb{R}\smallsetminus(-\beta,\beta)}\varphi(x)\,\mathrm{d}x = \left(\frac{1}{\beta}\int_0^\beta f(t)\,\mathrm{d}t\right)\mathbb{P}\big[|\mathsf{X}|\geqslant\beta\big].$$ □

For every $\theta \in [0,1]$ define two (periodic) subsets $L_\theta, R_\theta \subseteq \mathbb{R}$ of the real line by

$$L_\theta \stackrel{\mathrm{def}}{=} \left\{ r \in \mathbb{R} :\ r-\theta \in \bigcup_{k\in\mathbb{Z}} \left[k, k+\frac{1}{4}\right) \right\} \qquad \text{and} \qquad R_\theta \stackrel{\mathrm{def}}{=} \left\{ r \in \mathbb{R} :\ r-\theta \in \bigcup_{k\in\mathbb{Z}} \left[k+\frac{1}{2}, k+\frac{3}{4}\right) \right\}. \tag{126}$$

By definition, we then have

$$\forall \theta \in [0,1],\ \forall (a,b) \in L_\theta \times R_\theta, \qquad |a-b| > \frac{1}{4}. \tag{127}$$

Throughout what follows, we will denote the Lebesgue probability measure on $[0,1]$ by $\mathbb{U}$. Observe that

$$\forall a \in \mathbb{R}, \qquad \mathbb{U}\big(\{\theta \in [0,1] :\ a \in L_\theta\}\big) = \mathbb{U}\big(\{\theta \in [0,1] :\ a \in R_\theta\}\big) = \frac{1}{4}, \tag{128}$$

and, if we consider the periodic tent function $\tau : \mathbb{R} \to [0,\infty)$ that is given by

$$\tau(s) \stackrel{\mathrm{def}}{=} \max\left\{ \frac{1}{4} - \left| \frac{1}{2} - (s - \lfloor s \rfloor) \right|, 0 \right\} = \begin{cases} 0 & \text{if } 0 \leqslant s - \lfloor s \rfloor \leqslant \frac{1}{4}, \\ s - \lfloor s \rfloor - \frac{1}{4} & \text{if } \frac{1}{4} \leqslant s - \lfloor s \rfloor \leqslant \frac{1}{2}, \\ \frac{3}{4} - (s - \lfloor s \rfloor) & \text{if } \frac{1}{2} \leqslant s - \lfloor s \rfloor \leqslant \frac{3}{4}, \\ 0 & \text{if } \frac{3}{4} \leqslant s - \lfloor s \rfloor \leqslant 1, \end{cases} \tag{129}$$

then

$$\forall (a,b) \in \mathbb{R}^2, \qquad \mathbb{U}\big(\{\theta \in [0,1] :\ (a,b) \in L_\theta \times R_\theta\}\big) = \tau(|a-b|). \tag{130}$$

Both (128) and (130) follow from straightforward elementary computations, which we omit.

Lemma 45. *Fix $C > 0$. For every $v \in \mathbb{R}^n$ and $\theta \in [0,1]$ define $L_C(v,\theta), R_C(v,\theta) \subseteq \mathbb{R}^n$ by*

$$L_C(v,\theta) \stackrel{\mathrm{def}}{=} \left\{ z \in \mathbb{R}^n :\ \frac{\langle v, z\rangle}{4C} \in L_\theta \right\} \qquad \text{and} \qquad R_C(v,\theta) \stackrel{\mathrm{def}}{=} \left\{ z \in \mathbb{R}^n :\ \frac{\langle v, z\rangle}{4C} \in R_\theta \right\}, \tag{131}$$

where $L_\theta, R_\theta \subseteq \mathbb{R}$ are as in (126)*. Then, both $(v,\theta) \mapsto L_C(v,\theta)$ and $(v,\theta) \mapsto R_C(v,\theta)$ are Borel-measurable set-valued mappings from (from $\mathbb{R}^n \times [0,1]$ to the Borel subsets of $\mathbb{R}^n$) that satisfy*

$$\forall (v,\theta) \in \mathbb{R}^n \times [0,1],\ \forall (x,y) \in L_C(v,\theta) \times R_C(v,\theta), \qquad |\langle v, x-y\rangle| > C. \tag{132}$$

Furthermore, recalling that $\mathbb{U}$ denotes the Lebesgue measure on $[0,1]$, we have

$$\forall v, z \in \mathbb{R}^n, \qquad \mathbb{U}\big(\{\theta \in [0,1] :\ z \in L_C(v,\theta)\}\big) = \mathbb{U}\big(\{\theta \in [0,1] :\ z \in R_C(v,\theta)\}\big) = \frac{1}{4}, \tag{133}$$

and

$$\forall x, y \in \mathbb{R}^n, \qquad (\gamma_n \times \mathbb{U})\big(\{(v,\theta) \in \mathbb{R}^n \times [0,1] :\ (x,y) \in L_C(v,\theta) \times R_C(v,\theta)\}\big) \gtrsim e^{-\frac{9C^2}{\|x-y\|_2^2}}. \tag{134}$$

Proof. The measurability assertion is immediate from the definitions (126) and (131). Requirement (132) follows directly from (131) and (127). Requirement (133) follows directly from (131) and (128). To justify the remaining property (134), fix distinct $x, y \in \mathbb{R}^n$ and observe that

$$(\gamma_n \times \mathbb{U})\big(\{(v,\theta) \in \mathbb{R}^n \times [0,1] :\ (x,y) \in L_C(v,\theta) \times R_C(v,\theta)\}\big)$$
$$\stackrel{(134)}{=} \int_{\mathbb{R}^n} \mathbb{U}\left(\left\{ \theta \in [0,1] :\ \frac{\langle v, x\rangle}{4C} \in L_\theta \quad \text{and} \quad \frac{\langle v, y\rangle}{4C} \in R_\theta \right\} \right) \mathrm{d}\gamma_n(v) \stackrel{(130)}{=} \int_{\mathbb{R}^n} \tau\left(\frac{|\langle v, x-y\rangle|}{4C} \right) \mathrm{d}\gamma_n(v),$$

where we recall the definition of τ in (129). By rotation-invariance, $\langle v, x-y\rangle / \|x-y\|_2$ is distributed over $\mathbb{R}$ according to γ_1 when v is distributed over $\mathbb{R}^n$ according to γ_n. By Lemma 44, we therefore have

$$\int_{\mathbb{R}^n} \tau\left(\frac{|\langle v, x-y\rangle|}{4C} \right) \mathrm{d}\gamma_n(v) \geqslant \left(\int_0^1 \tau(t)\,\mathrm{d}t \right) \sqrt{\frac{2}{\pi}} \int_{\frac{4C}{\|x-y\|_2}}^{\infty} e^{-\frac{s^2}{2}}\,\mathrm{d}s \asymp \min\left\{ \frac{\|x-y\|_2}{C}, 1 \right\} e^{-\frac{8C^2}{\|x-y\|_2^2}} \gtrsim e^{-\frac{9C^2}{\|x-y\|_2^2}},$$

where the second step is valid as the integral of τ over $[0,1]$ is a positive universal constant, and the following standard asymptotic identity holds (its elementary proof can be found in e.g. [Kom55, Dur19]):

$$\int_a^\infty e^{-\frac{s^2}{2}}\,\mathrm{d}s \asymp \min\left\{ \frac{1}{a}, 1 \right\} e^{-\frac{a^2}{2}}. \qquad \square$$

Lemma 46. *Fix $\alpha, C > 0$ and $n \in \mathbb{N}$. Suppose that X is a finite set, equipped with mappings $f : X \to \mathbb{R}^n$ and $\Lambda : X \to (0,\infty]$. Then, there is a Polish probability space $(\Omega, \mathbb{P})$[26] and for every $v \in \mathbb{R}^n$ and $\chi \in \Omega$ there are*

$$A_C(v,\chi) = A_C^{f,\Lambda,\alpha}(v,\chi), B_C(v,\chi) = B_C^{f,\Lambda,\alpha}(v,\chi) \subseteq X,$$

such that the set-valued mappings $(v,\chi) \mapsto A_C(v,\chi)$ and $(v,\chi) \mapsto B_C(v,\chi)$ from $\mathbb{R}^n \times \Omega$ to 2^X are Borel-measurable, and have the following properties. Firstly, for any $x, y \in X$, any $v \in \mathbb{R}^n$ and any $\chi \in \Omega$, if

$$(x,y) \in A_C(v,\chi) \times B_C(v,\chi) \qquad \text{and} \qquad e^{-\alpha}\Lambda(x) \leqslant \Lambda(y) \leqslant e^{\alpha}\Lambda(x), \tag{135}$$

then

$$|\langle v, f(x) - f(y)\rangle| > C \max\{\Lambda(x), \Lambda(y)\}. \tag{136}$$

Secondly, for every $v \in \mathbb{R}^n$ and $x \in X$ we have

$$\mathbb{P}\big(\{\chi \in \Omega :\ x \in A_C(v,\chi)\}\big) = \mathbb{P}\big(\{\chi \in \Omega :\ x \in B_C(v,\chi)\}\big) = \frac{1}{6}. \tag{137}$$

Finally, every $x, y \in X$ for which $\min\{\Lambda(x), \Lambda(y)\} < \infty$ satisfy

$$\gamma_n \times \mathbb{P}\big(\{(v,\chi) \in \mathbb{R}^n \times \Omega :\ (x,y) \in A_C(v,\chi) \times B_C(v,\chi)\}\big) \gtrsim \exp\left(-9e^{4\alpha}\frac{\min\{\Lambda(x)^2, \Lambda(y)^2\}}{\|f(x) - f(y)\|_2^2} C^2\right). \tag{138}$$

Thus, $\gamma_n \times \mathbb{P}\big(\{(v,\chi) \in \mathbb{R}^n \times \Omega :\ (x,y) \in A_C(v,\chi) \times B_C(v,\chi)\}\big) \gtrsim e^{-9e^{4\alpha}C^2}$ if $\|f(x) - f(y)\|_2 \geqslant \min\{\Lambda(x), \Lambda(y)\}$.

Proof. Lemma 46 permits Λ to take the value ∞ because this will be convenient when Lemma 46 will be applied, but it corresponds to a degenerate situation. To isolate the main point of Lemma 46, denote

$$X_\infty \stackrel{\mathrm{def}}{=} \{x \in X :\ \Lambda(x) = \infty\} \qquad \text{and} \qquad X_{<\infty} \stackrel{\mathrm{def}}{=} X \smallsetminus X_\infty. \tag{139}$$

We will first treat the (more meaningful) subset $X_{<\infty}$ of X.

For every $r \in [0,1]$ and $i \in \mathbb{Z}$ define $Y_i(r) = Y_i^\alpha(r) \subseteq X_{<\infty}$ by

$$Y_i(r) \stackrel{\mathrm{def}}{=} \left\{x \in X_{<\infty} :\ e^{3\alpha(i+r)-2\alpha} \leqslant \Lambda(x) < e^{3\alpha(i+r)}\right\} = \left\{x \in X_{<\infty} :\ \frac{1}{3\alpha}\log\Lambda(x) - r \in \left[i - \frac{2}{3}, i\right)\right\}. \tag{140}$$

Note that $\{Y_i(r)\}_{i\in\mathbb{Z}}$ are pairwise disjoint subsets of $X_{<\infty}$. Writing also

$$Y(r) = Y^\alpha(r) \stackrel{\mathrm{def}}{=} \bigcup_{i\in\mathbb{Z}} Y_i(r), \tag{141}$$

it is straightforward to compute that

$$\forall x \in X_{<\infty}, \qquad \mathbb{U}\big(\{r \in [0,1] :\ x \in Y(r)\}\big) = \frac{2}{3}. \tag{142}$$

Recalling the notation introduced in Lemma 45, every $\theta = (\theta_i)_{i\in\mathbb{Z}} \in [0,1]^{\mathbb{Z}}$ and $r \in [0,1]$, we now define

$$\begin{aligned} E_C(v,\theta,r) &\stackrel{\mathrm{def}}{=} \bigcup_{i\in\mathbb{Z}} E_C^i(v,\theta_i,r), \qquad \text{where} \qquad E_C^i(v,\theta_i,r) \stackrel{\mathrm{def}}{=} Y_i(r) \cap f^{-1}\big(L_{e^{3\alpha(i+r)}C}(v,\theta_i)\big) \subseteq X_{<\infty},\\ F_C(v,\theta,r) &\stackrel{\mathrm{def}}{=} \bigcup_{i\in\mathbb{Z}} F_C^i(v,\theta_i,r), \qquad \text{where} \qquad F_C^i(v,\theta_i,r) \stackrel{\mathrm{def}}{=} Y_i(r) \cap f^{-1}\big(R_{e^{3\alpha(i+r)}C}(v,\theta_i)\big) \subseteq X_{<\infty}. \end{aligned} \tag{143}$$

Fix $\theta = (\theta_i)_{i\in\mathbb{Z}} \in [0,1]^{\mathbb{Z}}$, $v \in \mathbb{R}^n$, and $r \in [0,1]$, as well as $x, y \in X_{<\infty}$ that satisfy

$$(x,y) \in E_C(v,\theta,r) \times F_C(v,\theta,r) \qquad \text{and} \qquad e^{-\alpha}\Lambda(x) \leqslant \Lambda(y) \leqslant e^{\alpha}\Lambda(x), \tag{144}$$

The first condition in (144) and the definitions (143) imply in particular that there are $i, j \in \mathbb{Z}$ such that

$$(x,y) \in Y_i(r) \times Y_j(r) \qquad \text{and} \qquad \big(f(x), f(y)\big) \in L_{e^{3\alpha(i+r)}C}(v,\theta_i) \times R_{e^{3\alpha(j+r)}C}(v,\theta_j). \tag{145}$$

[26] Thus, Ω is a separable complete metric space and $\mathbb{P}$ is a Borel probability measure on Ω.

Together with the second condition in (144), the first condition in (145) entails that necessarily $i = j$. Indeed, if $j \geqslant i+1$, then we see from the definition (140) that

$$\Lambda(y) \overset{y\in Y_j(r)}{>} e^{3\alpha(j+r)-2\alpha} \geqslant e^{3\alpha(i+1+r)-2\alpha} = e^{\alpha}\cdot e^{3\alpha(i+r)} \overset{x\in Y_i(r)}{\geqslant} e^{\alpha}\Lambda(x),$$

in contradiction to (135). Analogously, we cannot have $j \leqslant i-1$. Thus necessarily $i = j$, and therefore

$$\max\{\Lambda(x),\Lambda(y)\} < e^{3\alpha(i+r)}, \tag{146}$$

by the definition 140 of $Y_i(r)$. Using Lemma 45, the second condition in (145) now implies that

$$|\langle v, f(x)-f(y)\rangle| \overset{(132)}{>} e^{3\alpha(i+r)}C \overset{(146)}{>} C\max\big\{\Lambda(x),\Lambda(y)\big\}. \tag{147}$$

We have thus checked that the following implication holds:

$$\forall x,y\in X_{<\infty},\qquad (144) \implies |\langle v, f(x)-f(y)\rangle| > C\max\big\{\Lambda(x),\Lambda(y)\big\}. \tag{148}$$

Next, for every fixed $v\in\mathbb{R}^n$ and $x\in X_{<\infty}$ we have

$$\begin{aligned}
&(\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(\theta,r)\in[0,1]^{\mathbb{Z}}\times[0,1]:\ x\in E_C(v,\theta,r)\}\big)\\
&\overset{(143)}{=} \sum_{i\in\mathbb{Z}} (\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(\theta,r)\in[0,1]^{\mathbb{Z}}\times[0,1]:\ x\in Y_i(r) \text{ and } f(x)\in L_{e^{3\alpha(i+r)}C}(v,\theta_i)\}\big)\\
&= \sum_{i\in\mathbb{Z}}\int_0^1 \mathbf{1}_{\{x\in Y_i(r)\}}\bigg(\int_0^1 \mathbf{1}_{\{f(x)\in L_{e^{3\alpha(i+r)}C}(v,\theta_i)\}}\,\mathrm{d}\theta_i\bigg)\mathrm{d}r\\
&= \sum_{i\in\mathbb{Z}}\int_0^1 \mathbf{1}_{\{x\in Y_i(r)\}}\mathbb{U}\big(\{\theta\in[0,1]:\ f(x)\in L_{e^{3\alpha(i+r)}C}(v,\theta)\}\big)\,\mathrm{d}r\\
&\overset{(133)}{=} \sum_{i\in\mathbb{Z}}\frac{1}{4}\mathbb{U}\big(\{r\in[0,1]:\ x\in Y_i(r)\}\big)\\
&\overset{(141)}{=} \frac{1}{4}\mathbb{U}\big(\{r\in[0,1]:\ x\in Y(r)\}\big) \overset{(142)}{=} \frac{1}{6}.
\end{aligned} \tag{149}$$

By the analogous reasoning, for every $x\in X_{<\infty}$ we also have

$$(\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(\theta,r)\in[0,1]^{\mathbb{Z}}\times[0,1]:\ x\in F_C(v,\theta,r)\}\big) = \frac{1}{6}. \tag{150}$$

Finally, suppose that $i,j\in\mathbb{Z}$ satisfy $i\neq j$. If $x,y\in X_{<\infty}$, then for every $r\in[0,1]$ and $v\in\mathbb{R}^n$ we have

$$\begin{aligned}
&\mathbb{U}^{\mathbb{Z}}\Big(\big\{\theta\in[0,1]^{\mathbb{Z}}:\ (x,y)\in f^{-1}\big(L_{e^{3\alpha(i+r)}C}(v,\theta_i)\big)\times f^{-1}\big(R_{e^{3\alpha(j+r)}C}(v,\theta_j)\big)\big\}\Big)\\
&\quad= \mathbb{U}\big(\{\theta_i\in[0,1]:\ f(x)\in L_{e^{3\alpha(i+r)}C}(v,\theta_i)\}\big)\mathbb{U}\big(\{\theta_j\in[0,1]:\ f(y)\in R_{e^{3\alpha(j+r)}C}(v,\theta_j)\}\big) \overset{(133)}{=} \frac{1}{16}.
\end{aligned} \tag{151}$$

By multiplying (151) by $\mathbf{1}_{\{(x,y)\in Y_i(r)\times Y_j(r)\}}$ and then integrating with respect to $\gamma_n\times\mathbb{U}$, we see that

$$\begin{aligned}
&(\gamma_n\times\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(v,\theta,r)\in\mathbb{R}^n\times[0,1]^{\mathbb{Z}}\times[0,1]:(x,y)\in E_C^i(v,\theta,r)\times F_C^j(v,\theta,r)\}\big)\\
&\qquad\qquad\overset{(143)\wedge(151)}{=} \frac{1}{16}\mathbb{U}\big(\{r\in[0,1]:\ (x,y)\in Y_i(r)\times Y_j(r)\}\big),
\end{aligned} \tag{152}$$

where we recall that the identity (152) holds for every $x,y\in X_{<\infty}$ whenever $i,j\in\mathbb{Z}$ satisfy $i\neq j$.

At the same time, for every $i\in\mathbb{Z}$ and $r\in[0,1]$, if $x,y\in Y_i(r)$, then

$$e^{3\alpha(i+r)} \overset{(140)}{\leqslant} e^{2\alpha}\min\big\{\Lambda(x),\Lambda(y)\big\}. \tag{153}$$

Therefore, for every $x,y\in X_{<\infty}$ and every fixed $r\in[0,1]$ we have

$$
\begin{aligned}
&\mathbb{1}_{\{x,y\in Y_i(r)\}}(\gamma_n\times \mathbb{U}^{\mathbb{Z}})\Big(\big\{(v,\theta)\in\mathbb{R}^n\times[0,1]^{\mathbb{Z}}:\ (x,y)\in f^{-1}\big(L_{e^{3\alpha(i+r)}C}(v,\theta_i)\big)\times f^{-1}\big(R_{e^{3\alpha(i+r)}C}(v,\theta_i)\big)\big\}\Big)\\
&=\mathbb{1}_{\{x,y\in Y_i(r)\}}(\gamma_n\times \mathbb{U})\Big(\big\{(v,\theta_i)\in\mathbb{R}^n\times[0,1]:\ \big(f(x),f(y)\big)\in L_{e^{3\alpha(i+r)}C}(v,\theta_i)\times R_{e^{3\alpha(i+r)}C}(v,\theta_i)\big\}\Big)\\
&\overset{(134)}{\gtrsim} \mathbb{1}_{\{x,y\in Y_i(r)\}}\exp\left(-\frac{9e^{6\alpha(i+r)}C^2}{\|f(x)-f(y)\|_2^2}\right)\\
&\overset{(153)}{\geqslant} \mathbb{1}_{\{x,y\in Y_i(r)\}}\exp\left(-9e^{4\alpha}\frac{\min\{\Lambda(x)^2,\Lambda(y)^2\}}{\|f(x)-f(y)\|_2^2}C^2\right).
\end{aligned}\tag{154}
$$

By integrating (154) with respect to $\mathbb{U}$, we conclude that for every $i\in\mathbb{Z}$ and every $x,y\in X_{<\infty}$,

$$
\begin{aligned}
(\gamma_n\times\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})&\big(\{(v,\theta,r)\in\mathbb{R}^n\times[0,1]^{\mathbb{Z}}\times[0,1]:\ (x,y)\in E_C^i(v,\theta,r)\times F_C^i(v,\theta,r)\}\big)\\
&\overset{(143)\wedge(154)}{\gtrsim}\exp\left(-9e^{4\alpha}\frac{\min\{\Lambda(x)^2,\Lambda(y)^2\}}{\|f(x)-f(y)\|_2^2}C^2\right)\mathbb{U}\big(\{r\in[0,1]:\ x,y\in Y_i(r)\}\big).
\end{aligned}\tag{155}
$$

By combining (152) and (155), we see that the following estimate holds for every $x,y\in X_{<\infty}$:

$$
\begin{aligned}
&(\gamma_n\times\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(v,\theta,r)\in\mathbb{R}^n\times[0,1]^{\mathbb{Z}}\times[0,1]:\ (x,y)\in E_C(v,\theta,r)\times F_C(v,\theta,r)\}\big)\\
&\overset{(143)}{=}\sum_{\substack{i,j\in\mathbb{Z}\\ i\neq j}}(\gamma_n\times\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(v,\theta,r)\in\mathbb{R}^n\times[0,1]^{\mathbb{Z}}\times[0,1]:\ (x,y)\in E_C^i(v,\theta,r)\times F_C^j(v,\theta,r)\}\big)\\
&\qquad+\sum_{i\in\mathbb{Z}}(\gamma_n\times\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(v,\theta,r)\in\mathbb{R}^n\times[0,1]^{\mathbb{Z}}\times[0,1]:\ (x,y)\in E_C^i(v,\theta,r)\times F_C^i(v,\theta,r)\}\big)\\
&\overset{(152)\wedge(155)}{\gtrsim}\exp\left(-9e^{4\alpha}\frac{\min\{\Lambda(x)^2,\Lambda(y)^2\}}{\|f(x)-f(y)\|_2^2}C^2\right)\sum_{i,j\in\mathbb{Z}}\mathbb{U}\big(\{r\in[0,1]:\ (x,y)\in Y_i(r)\times Y_j(r)\}\big)\\
&\overset{(141)}{=}\exp\left(-9e^{4\alpha}\frac{\min\{\Lambda(x)^2,\Lambda(y)^2\}}{\|f(x)-f(y)\|_2^2}C^2\right)\mathbb{U}\big(\{r\in[0,1]:\ x,y\in Y(r)\}\big)\\
&\geqslant\exp\left(-9e^{4\alpha}\frac{\min\{\Lambda(x)^2,\Lambda(y)^2\}}{\|f(x)-f(y)\|_2^2}C^2\right)\Big(\mathbb{U}\big(\{r\in[0,1]:\ x\in Y(r)\}\big)+\mathbb{U}\big(\{r\in[0,1]:\ y\in Y(r)\}\big)-1\Big)\\
&\overset{(142)}{=}\frac{1}{3}\exp\left(-9e^{4\alpha}\frac{\min\{\Lambda(x)^2,\Lambda(y)^2\}}{\|f(x)-f(y)\|_2^2}C^2\right).
\end{aligned}\tag{156}
$$

The above considerations yield for $E_C(\cdot),F_C(\cdot)$ the conclusions of Lemma 46 when $x,y\in X_{<\infty}$, but it is simple to extend those to every $x,y\in X$ as follows. Let $\mathbb{Q}$ be the probability measure on the 3-point set $\{1,2,3\}$ that is given by $\mathbb{Q}(1)=2/3$ and $\mathbb{Q}(2)=\mathbb{Q}(3)=1/6$. Consider the (Polish) probability space

$$
(\Omega,\mathbb{P})\overset{\mathrm{def}}{=}\big([0,1]^{\mathbb{Z}}\times[0,1]\times\{1,2,3\},\mathbb{U}^{\mathbb{Z}}\times\mathbb{U}\times\mathbb{Q}\big),\tag{157}
$$

and, recalling (139), define $A_C,B_C:\mathbb{R}^n\times\Omega\to 2^X$ by setting for every $(v,\theta,r,k)\in\mathbb{R}^n\times[0,1]^{\mathbb{Z}}\times[0,1]\times\{1,2,3\}$,

$$
\big(A_C(v,\theta,r,k),B_C(v,\theta,r,k)\big)\overset{\mathrm{def}}{=}\begin{cases}\big(E_C(v,\theta,r),F_C(v,\theta,r)\big) & \text{if } k=1,\\ \big(E_C(v,\theta,r)\cup X_\infty,F_C(v,\theta,r)\big) & \text{if } k=2,\\ \big(E_C(v,\theta,r),F_C(v,\theta,r)\cup X_\infty\big) & \text{if } k=3.\end{cases}\tag{158}
$$

For these definitions, the measurability requirements of Lemma 46 are immediate[27].

Suppose first that (135) is satisfied for some $x,y\in X$, $v\in\mathbb{R}^n$ and $\chi=(\theta,r,k)\in\Omega$. If $\{x,y\}\cap X_\infty\neq\emptyset$, then by the second requirement in (135) we have $\Lambda(x)=\Lambda(y)=\infty$, i.e., $x,y\in X_\infty$, but then by (158) we see that

[27] Recall that we assumed in Lemma 46 that X is finite, but we did this only to remove the need to mention any assumption on f and Λ; if X is an infinite Polish space, then (158) will define Borel-measurable set valued mappings if we impose the assumptions that f and Λ are Borel measurable.

the first requirement in (135) cannot hold. We therefore necessarily have $x,y\in X_{<\infty}$, in which case (144) holds thanks to (158) and (135), so the desired conclusion (136) follows from (147).

Next, to verify (137) fix $v\in\mathbb{R}^n$ and $x\in X$. If $x\in X_{<\infty}$, then

$$\mathbb{P}\big(\{(\theta,r,k)\in\Omega:\ x\in A_C(v,\theta,r,k)\}\big)\overset{(157)\wedge(158)}{=}(\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(\theta,r)\in[0,1]^{\mathbb{Z}}\times[0,1]:\ x\in E_C(v,\theta,r)\}\big)\overset{(149)}{=}\frac{1}{6},$$

and analogously thanks to (150) we also have

$$\mathbb{P}\big(\{(\theta,r,k)\in\Omega:\ x\in B_C(v,\theta,r,k)\}\big)=\frac{1}{6}.$$

On the other hand, if $x\in X_{\infty}$, then

$$\mathbb{P}\big(\{(\theta,r,k)\in\Omega:\ x\in A_C(v,\theta,r,k)\}\big)\overset{(157)\wedge(158)}{=}\mathbb{Q}[k=2]=\frac{1}{6},$$

and

$$\mathbb{P}\big(\{(\theta,r,k)\in\Omega:\ x\in B_C(v,\theta,r,k)\}\big)\overset{(157)\wedge(158)}{=}\mathbb{Q}[k=3]=\frac{1}{6}.$$

Altogether, this completes the verification of (137).

To complete the proof of Lemma 46, it remains to check that (138) holds for every $x,y\in X$ such that $\min\{\Lambda(x),\Lambda(y)\}<\infty$, i.e., $|\{x,y\}\cap X_{<\infty}|\geqslant 1$. Indeed, if $x,y\in X_{<\infty}$, then thanks to (157) and (158), we see that (138) coincides with (156). If $x\in X_{<\infty}$ and $y\in X_{\infty}$, then

$$\begin{aligned}\gamma_n\times\mathbb{P}&\big(\{(v,\theta,r,k)\in\mathbb{R}^n\times\Omega:(x,y)\in A_C(v,\theta,r,k)\times B_C(v,\theta,r,k)\}\big)\\&\overset{(158)}{=}\gamma_n\times\mathbb{P}\big(\{(v,\theta,r,k)\in\mathbb{R}^n\times\Omega:\ x\in E_C(v,\theta,r)\quad\text{and}\quad k=3\}\big)\\&\overset{(157)}{=}\frac{1}{6}\int_{\mathbb{R}^n}(\mathbb{U}^{\mathbb{Z}}\times\mathbb{U})\big(\{(\theta,r)\in[0,1]^{\mathbb{Z}}\times[0,1]:\ x\in E_C(v,\theta,r)\}\big)\,\mathrm{d}\gamma_n(v)\\&\overset{(149)}{=}\frac{1}{36}\gtrsim\exp\bigg(-9e^{4\alpha}\frac{\min\{\Lambda(x)^2,\Lambda(y)^2\}}{\|f(x)-f(y)\|_2^2}C^2\bigg),\end{aligned}$$

i.e., (138) holds in this case. The remaining case $(x,y)\in X_{<\infty}\times X_{\infty}$ is derived analogously using (150). □

Proof of Proposition 36. Set $\alpha=\log 2$, for assumption (48) of Proposition 36 to match assumption (135) of Lemma 46, which we will next apply.[28] Let $V^1,\ldots,V^m\subseteq V$ be the connected components of G. For each $s\in[m]$, let $A_C^s,B_C^s:\mathbb{R}^n\times\Omega^s\to 2^{V_s}\subseteq 2^V$ the random subsets of V^s that Lemma 46 provides (for $X=V^s$, the C from the statement of Proposition 36, the above α, and slightly abusing notation by identifying f,Λ with their restrictions to V^s), where the underlying probability space is now denoted $(\Omega^s,\mathbb{P}^s)$.

Consider the probability space

$$(\Omega,\mathbb{P})\overset{\mathrm{def}}{=}(\Omega^1\times\ldots\times\Omega^s,\mathbb{P}^1\times\ldots\times\mathbb{P}^s).$$

Define random subsets $A_C,B_C:\mathbb{R}^n\times\mathbb{P}\to 2^X$ of V by setting

$$\forall\, v\in\mathbb{R}^n,\ \forall\,\chi=(\chi_1,\ldots,\chi_m)\in\Omega,\qquad A_C(v,\chi)\overset{\mathrm{def}}{=}\bigcup_{s=1}^m A_C^s(v,\chi_s)\qquad\text{and}\qquad B_C(v,\chi)\overset{\mathrm{def}}{=}\bigcup_{s=1}^m B_C^s(v,\chi_s).\tag{159}$$

In other words, for each fixed $v\in\mathbb{R}^n$ the random pairs of subsets $(A_C^1(v,\cdot),B_C^1(v,\cdot)),\ldots,(A_C^m(v,\cdot),B_C^m(v,\cdot))$ are chosen independently, and then $A_C(v,\cdot)$ and $B_C(v,\cdot)$ are the unions of their first and second coordinates, respectively. We will proceed to check that with positive probability (indeed, in expectation), the sets in (159) when χ is distributed according to $\mathbb{P}$ satisfy the stated conclusions of Proposition 36.

In fact, requirement (51) holds for every $\chi\in\Omega$, since if $(x,y)\in A_C(v,\chi)\times B_C(v,\chi)$ and $\{x,y\}\in E$, then x,y are in the same connected component of G, so from (159) we know that $(x,y)\in A_C^s(v,\chi)\times B_C^s(v,\chi)$ for some $s\in[m]$. Also (135) holds by (48), so conclusion (136) of Lemma 46 for $(\Omega^s,\mathbb{P}^s)$ gives (51).

[28]The proof shows that if we replace the factor 2 in (48) by e^{α} for any $\alpha>0$, then the conclusions of Proposition 36 hold with the universal constant κ in (52) replaced by $9e^{4\alpha}$.

For each $x\in V$ let $s(x)$ denote the unique $s\in[m]$ to which x belongs. We then have

$$\begin{aligned}&\int_\Omega\bigg(\int_{\mathbb{R}^n}\omega\big(A_C(v,\chi)\times B_C(v,\chi)\big)\,\mathrm{d}\gamma_n(v)\bigg)\,\mathrm{d}\mathbb{P}(\chi)\\ &\qquad=\sum_{(x,y)\in V}\omega(x,y)(\gamma_n\times\mathbb{P})\big(\{(v,\chi)\in\mathbb{R}^n\times\Omega:\ (x,y)\in A_C(v,\chi)\times B_C(v,\chi)\}\big)\\ &\qquad\stackrel{(159)}{=}\sum_{(x,y)\in V}\omega(x,y)(\gamma_n\times\mathbb{P})\big(\{(v,\chi)\in\mathbb{R}^n\times\Omega:\ (x,y)\in A_C^{s(x)}(v,\chi_{s(x)})\times B_C^{s(y)}(v,\chi_{s(y)})\}\big)\\ &\qquad\geqslant\min_{x,y\in V}(\gamma_n\times\mathbb{P})\big(\{(v,\chi)\in\mathbb{R}^n\times\Omega:\ (x,y)\in A_C^{s(x)}(v,\chi_{s(x)})\times B_C^{s(y)}(v,\chi_{s(y)})\}\big),\end{aligned}\tag{160}$$

where the last step of (160) holds because ω is a probability measure. For $x,y\in V$, if $s(x)=s(y)=s$, then

$$\begin{aligned}&(\gamma_n\times\mathbb{P})\big(\{(v,\chi)\in\mathbb{R}^n\times\Omega:\ (x,y)\in A_C^{s(x)}(v,\chi_{s(x)})\times B_C^{s(y)}(v,\chi_{s(y)})\}\big)\\ &\qquad\stackrel{(159)}{=}(\gamma_n\times\mathbb{P}^s)\big(\{(v,\chi_s)\in\mathbb{R}^n\times\Omega:\ (x,y)\in A_C^{s}(v,\chi_s)\times B_C^{s}(v,\chi_s)\}\big)\stackrel{(138)\wedge(49)}{\gtrsim}e^{-9e^{4\alpha}C^2}.\end{aligned}\tag{161}$$

On the other hand, if $s(x)\neq s(y)$, then $A_C^{s(x)}(v,\chi_{s(x)})$ and $B_C^{s(y)}(v,\chi_{s(y)})$ are independent random subsets of X for each fixed $v\in\mathbb{R}^n$, so we have the following crude estimate:

$$\begin{aligned}&(\gamma_n\times\mathbb{P})\big(\{(v,\chi)\in\mathbb{R}^n\times\Omega:\ (x,y)\in A_C^{s(x)}(v,\chi_{s(x)})\times B_C^{s(y)}(v,\chi_{s(y)})\}\big)\\ &\stackrel{(159)}{=}\int_{\mathbb{R}^n}\mathbb{P}^{s(x)}\left(\{\chi_{s(x)}\in\Omega^{s(x)}:\ x\in A_C^{s(x)}(v,\chi_{s(x)})\}\right)\mathbb{P}^{s(y)}\left(\{\chi_{s(y)}\in\Omega^{s(y)}:\ y\in A_C^{s(y)}(v,\chi_{s(y)})\}\right)\mathrm{d}\gamma_n(v)\\ &\stackrel{(137)}{=}\frac{1}{36}\gtrsim e^{-9e^{4\alpha}C^2}.\end{aligned}\tag{162}$$

A substitution of (161) and (162) into (160) shows that

$$\int_\Omega\bigg(\int_{\mathbb{R}^n}\omega\big(A_C(v,\chi)\times B_C(v,\chi)\big)\,\mathrm{d}\gamma_n(v)\bigg)\,\mathrm{d}\mathbb{P}(\chi)\gtrsim e^{-9e^{4\alpha}C^2}.$$

Hence there exists $\chi^\omega\in\Omega$ such that if we denote $A(v)=A_C(v,\chi^\omega)$ and $B(v)=B_C(v,\chi^\omega)$ for $v\in\mathbb{R}^n$, then

$$\int_{\mathbb{R}^n}\omega\big(A_C^\omega(v)\times B_C^\omega(v)\big)\,\mathrm{d}\gamma_n(v)\gtrsim e^{-9e^{4\alpha}C^2}.$$

This completes the proof of (52), with the (concrete, but far from optimal) value $\kappa=9e^{4\log 2}=144$. $\square$

5. UNIVERSALLY COMPATIBLE COMPRESSION OF LOCAL GROWTH PROXIMITY GRAPHS

The following is a key construction that underlies our proof of Proposition 37:

Construction 47. *Let $(\mathcal{S},d_{\mathcal{S}})$ be a finite metric space and fix a function $\vartheta:\mathcal{S}\to\mathbb{R}$. For each $\xi\in\mathbb{R}$ denote*

$$S_\xi=S_\xi(\vartheta)\stackrel{\mathrm{def}}{=}\vartheta^{-1}\big((-\infty,\xi]\big)=\big\{x\in\mathcal{S}:\vartheta(x)\leqslant\xi\big\},\tag{163}$$

namely, S_ξ is the ξ-sublevel set of ϑ. Let $\ell\in\mathbb{N}$ and $\xi_1<\xi_2<\ldots<\xi_\ell$ be the increasing rearrangement of the values of ϑ, i.e., $\vartheta(\mathcal{S})=\{\xi_1,\ldots,\xi_\ell\}$. Thus,

$$S_{\xi_1}\subsetneq S_{\xi_2}\subsetneq\ldots\subsetneq S_{\xi_\ell}=\mathcal{S}.$$

Fix $\tau>0$. For each $i\in[\ell]$ define inductively $\mathcal{N}_{\xi_i}\subseteq S_{\xi_i}$ as follows. Let $\mathcal{N}_{\xi_1}$ be a subset of S_{ξ_1} that is maximal with respect to inclusion among all those $\mathcal{N}\subseteq S_{\xi_1}$ such that $d_{\mathcal{S}}(a,b)>2\tau$ for every distinct $a,b\in\mathcal{N}$. If $i\in\{2,\ldots,\ell\}$, then assume inductively that $\mathcal{N}_{\xi_{i-1}}$ has already been defined and let $\mathcal{N}_{\xi_i}$ be a subset of S_{ξ_i} that is maximal with respect to inclusion among all those $\mathcal{N}\subseteq S_{\xi_i}$ that contain $\mathcal{N}_{\xi_{i-1}}$ and such that $d_{\mathcal{S}}(a,b)>2\tau$ for every distinct $a,b\in\mathcal{N}$. We thus have (by design),

$$\mathcal{N}_{\xi_1}\subseteq\mathcal{N}_{\xi_2}\subseteq\ldots\subseteq\mathcal{N}_{\xi_\ell}.\tag{164}$$

Observe also that $\mathcal{N}_{\xi_i}$ is (2τ)-dense in S_{ξ_i} for every $i \in [\ell]$, i.e., $d_{\mathcal{S}}(w, \mathcal{N}_{\xi_i}) \leqslant 2\tau$ for every $w \in S_{\xi_i}$ (otherwise, we could add w to $\mathcal{N}_{\xi_i}$, in contradiction to its maximality with respect to inclusion).

Next, define a rounding function $q : \mathcal{S} \to \mathcal{S}$ as follows. Given $w \in \mathcal{S}$, fix any minimizer $w_{\min}$ of ϑ on $B_{\mathcal{S}}(w, 5\tau)$, i.e., $w_{\min} \in B_{\mathcal{S}}(w, 5\tau)$ and $\vartheta(w_{\min}) \leqslant \vartheta(z)$ for every $z \in B_{\mathcal{S}}(w, 5\tau)$. Because $w_{\min} \in S_{\vartheta(w_{\min})}$ and $\mathcal{N}_{\vartheta(w_{\min})}$ is (2τ)-dense in $S_{\vartheta(w_{\min})}$, we can define $q(w)$ to be any point in $\mathcal{N}_{\vartheta(w_{\min})}$ with $d_{\mathcal{S}}(q(w), w_{\min}) \leqslant 2\tau$.

The following lemma derives basic properties of the objects defined in Construction 47:

Lemma 48. *Under the definitions of Construction 47, q and $\{\mathcal{N}_{\xi_i}\}_{i=1}^{\ell}$ have the following properties:*

$$\forall\, w \in \mathcal{S}, \qquad d_{\mathcal{S}}(q(w), w) \leqslant 7\tau \qquad \text{and} \qquad q\big(B_{\mathcal{S}}(w, 5\tau)\big) \subseteq \mathcal{N}_{\vartheta(w)}. \tag{165}$$

Consequently,

$$\forall\, x, y \in \mathcal{S}, \qquad d_{\mathcal{S}}(x, y) \leqslant 3\tau \implies q\big(B_{\mathcal{S}}(y, 2\tau)\big) \subseteq B_{\mathcal{S}}(y, 9\tau) \cap \mathcal{N}_{\vartheta(x)}. \tag{166}$$

Proof. For (165), fix $w \in \mathcal{S}$. The first part of (165) is a direct consequence of Construction 47 and the triangle inequality, since for every $w \in \mathcal{S}$ we have $d_{\mathcal{S}}(w_{\min}, w) \leqslant 5\tau$ and $d_{\mathcal{S}}(q(w), w_{\min}) \leqslant 2\tau$, by definition. For the second part of (165), observe that for every $z \in B_{\mathcal{S}}(w, 5\tau)$ we have $w \in B_{\mathcal{S}}(z, 5\tau)$ and $z_{\min}$ is the minimizer of ϑ on $B_{\mathcal{S}}(z, 5\tau)$, so $\vartheta(z_{\min}) \leqslant \vartheta(w)$. Hence, $\mathcal{N}_{\vartheta(z_{\min})} \subseteq \mathcal{N}_{\vartheta(w)}$ by (164). But $q(z) \in \mathcal{N}_{\vartheta(z_{\min})}$ by definition, so $q(z) \in \mathcal{N}_{\vartheta(w)}$. As z is an arbitrary point in $B_{\mathcal{S}}(w, 5\tau)$, this proves the second part of (165).

(166) is a formal consequence of (165) and the triangle inequality, since if $x, y \in \mathcal{S}$ satisfy $d_{\mathcal{S}}(x, y) \leqslant 3\tau$, then $B_{\mathcal{S}}(y, 2\tau) \subseteq B_{\mathcal{S}}(x, 5\tau)$, whence $q(B_{\mathcal{S}}(y, 2\tau)) \subseteq q(B_{\mathcal{S}}(x, 5\tau)) \subseteq \mathcal{N}_{\vartheta(x)}$ by the case $w = x$ of the second part of (165), and for every $w \in B_{\mathcal{S}}(y, 2\tau)$ we have $d_{\mathcal{S}}(q(w), y) \leqslant d_{\mathcal{S}}(q(w), w) + d_{\mathcal{S}}(w, y) \leqslant 7\tau + 2\tau = 9\tau$ using the first part of (165), so we also have $q(B_{\mathcal{S}}(y, 2\tau)) \subseteq B_{\mathcal{S}}(y, 9\tau)$. □

The significance of incorporating the rounding function q of Construction 47 arises from the following lemma, which bounds the sizes of sets such as those that appear in the right hand side of (166):

Lemma 49. *Fix $\tau > 0$. Let $(\mathcal{S}, d_{\mathcal{S}})$ be a finite metric space. Suppose that $\mathfrak{d} : \mathcal{S} \to (0, \infty)$ and $\mathfrak{n} : \mathcal{S} \to (0, \infty)$ are functions that have the following property. For every $x \in \mathcal{S}$ and every subset $\mathcal{T}$ of $B_{\mathcal{S}}(x, 18\tau)$ with the property that every distinct $y, z \in \mathcal{T}$ satisfy $d_{\mathcal{S}}(y, z) > 2\tau$, we have*[29]

$$\sum_{y \in \mathcal{T}} \mathfrak{d}(y) \leqslant \mathfrak{n}(x). \tag{167}$$

Consider the setting of Construction 47 applied to $(\mathcal{S}, d_{\mathcal{S}})$ and τ, and with $\vartheta = \vartheta_{\mathfrak{d},\mathfrak{n}} : \mathcal{S} \to [1, \infty)$ given by

$$\forall\, x \in \mathcal{S}, \qquad \vartheta(x) \stackrel{\mathrm{def}}{=} \frac{\mathfrak{n}(x)}{\mathfrak{d}(x)}. \tag{168}$$

Then, for every $\xi \in \vartheta(\mathcal{S})$ and every $\emptyset \neq \mathcal{C} \subseteq \mathcal{M}$ with $\mathrm{diam}_{\mathcal{S}}(\mathcal{C}) \leqslant 18\tau$ we have

$$|\mathcal{C} \cap \mathcal{N}_{\xi}| \leqslant \xi. \tag{169}$$

Before passing to the proof of Lemma 49, we record in the following simple remark the examples that we will use in the ensuing proof of Proposition 37 of (numerator functions) $\mathfrak{n} : \mathcal{S} \to (0, \infty)$ and (denominator functions) $\mathfrak{d} : \mathcal{S} \to (0, \infty)$ as in (168) that satisfy assumption (167) of Lemma 49:

Remark 50. *In the setting of Lemma 49, for every $\mathfrak{d} : \mathcal{S} \to (0, \infty)$ and $x \in \mathcal{S}$ define $\mathfrak{d}_*(x)$ to be the maximum of $\sum_{y \in \mathcal{T}} \mathfrak{d}(y)$ over all possible $\mathcal{T} \subseteq B_{\mathcal{S}}(x, 18\tau)$ that satisfy $d_{\mathcal{S}}(y, z) > 2\tau$ for every distinct $y, z \in \mathcal{T}$. Then, by design the smallest $\mathfrak{n} : \mathcal{S} \to (0, \infty)$ that satisfies* (167) *is $\mathfrak{d}_* : \mathcal{S} \to (0, \infty)$, i.e., assumption* (167) *of Lemma 49 holds if and only if $\mathfrak{n} \geqslant \mathfrak{d}_*$ point-wise. A simple way to achieve this occurs when $(\mathcal{M}, d_{\mathcal{M}})$ is a metric space that contains $(\mathcal{S}, d_{\mathcal{S}})$ isometrically, i.e., $\mathcal{S} \subseteq \mathcal{M}$ and the restriction of $d_{\mathcal{M}}$ to $\mathcal{S} \times \mathcal{S}$ coincides with $d_{\mathcal{S}}$. If μ is any nondegenerate measure on the super-space $\mathcal{M}$, then consider the functions $\mathfrak{d}, \mathfrak{n} : \mathcal{S} \to (0, \infty)$ that are given by setting $\mathfrak{d}(x) = \mu(B_{\mathcal{M}}(x, \tau))$ and $\mathfrak{n}(x) = \mu(B_{\mathcal{M}}(x, 19\tau))$ for each $x \in \mathcal{S}$; note that we are evaluating here the measure μ on balls of the super-space $\mathcal{M}$ which could contain points that do not belong to $\mathcal{S}$ (this*

[29]Thus, in particular, $\mathfrak{d}(x) \leqslant \mathfrak{n}(x)$ for every $x \in \mathcal{S}$, as seen by considering $\mathcal{T} = \{x\}$.

nuance will be needed later). For any $x \in \mathcal{S}$, if $\mathcal{T} \subseteq \mathcal{S}$ is such that $d_{\mathcal{S}}(y,z) > 2\tau$ for every distinct $y,z \in \mathcal{T}$, then by the triangle inequality for $d_{\mathcal{M}}$ the balls $\{B_{\mathcal{M}}(y,\tau)\}_{y\in\mathcal{T}}$ (in the super-space $\mathcal{M}$) are pairwise disjoint. If also $\mathcal{T} \subseteq B_{\mathcal{S}}(x,18\tau)$, then $\bigcup_{y\in\mathcal{T}} B_{\mathcal{M}}(y,\tau) \subseteq B_{\mathcal{M}}(x,19\tau)$ by the triangle inequality for $d_{\mathcal{M}}$, so (167) *holds.*

Proof of Lemma 49. As $\xi \in \vartheta(\mathcal{S}) \subseteq [1,\infty)$, the conclusion of Lemma 49 is immediate if $|\mathcal{C}\cap\mathcal{N}_\xi| \leqslant 1$. We may therefore assume that $\mathcal{C}\cap\mathcal{N}_\xi \neq \emptyset$. By Construction 47 we know that $d_{\mathcal{S}}(y,z) > 2\tau$ every distinct $y,z \in \mathcal{N}_\xi$. Since $\operatorname{diam}_{\mathcal{S}}(\mathcal{C}) \leqslant 18\tau$, we also have $\mathcal{C} \subseteq B_{\mathcal{S}}(x,18\tau)$ for every $x \in \mathcal{C}$. We can therefore apply the assumption on $\mathfrak{d},\mathfrak{n}$ in Lemma 49 for $\mathcal{T} = \mathcal{C}\cap\mathcal{N}_\xi$ to get that

$$\forall\, x \in \mathcal{C}, \qquad \sum_{y\in\mathcal{C}\cap\mathcal{N}_\xi} \mathfrak{d}(y) \overset{(167)}{\leqslant} \mathfrak{n}(x). \tag{170}$$

But, Construction 47 ensures that $\mathcal{N}_\xi \subseteq S_\xi$, so recalling (163) we have $\vartheta(x) \leqslant \xi$ for every $x \in \mathcal{N}_\xi$. Thus,

$$\forall\, x \in \mathcal{N}_\xi, \qquad \mathfrak{n}(x) \overset{(168)}{\leqslant} \xi\mathfrak{d}(x). \tag{171}$$

Consequently,

$$\forall\, x \in \mathcal{C}\cap\mathcal{N}_\xi, \qquad \sum_{y\in\mathcal{C}\cap\mathcal{N}_\xi} \mathfrak{d}(y) \overset{(170)\wedge(171)}{\leqslant} \xi\mathfrak{d}(x). \tag{172}$$

Let $x_* \in \mathcal{C}\cap\mathcal{N}_\xi$ be the minimizer of $\mathfrak{d}$ on $\mathcal{C}\cap\mathcal{N}_\xi$, i.e., $\mathfrak{d}(x_*) \leqslant \mathfrak{d}(y)$ for every $y \in \mathcal{C}\cap\mathcal{N}_\xi$. Then, (172) applied to x_* implies that $|\mathcal{C}\cap\mathcal{N}_\xi|\mathfrak{d}(x_*) \leqslant \xi\mathfrak{d}(x_*)$. As $\mathfrak{d}(x_*) > 0$, we conclude that $|\mathcal{C}\cap\mathcal{N}_\xi| \leqslant \xi$. □

We will use Lemma 49 via the following corollary of it which provides a bound on Gaussian processes that arise from the composition of arbitrary mappings $\varphi : \mathcal{S} \to \mathbb{R}^m$ with the rounding function q:

Corollary 51. *Continuing with the notation and assumptions of Lemma 49, fix $m \in \mathbb{N}$ and $\varphi : \mathcal{S} \to \mathbb{R}^m$. Then, the following estimate holds for every $y \in \mathcal{S}$ and every $\mathcal{U} \subseteq B_{\mathcal{S}}(y,2\tau)$:*

$$\int_{\mathbb{R}^m} \Big(\max_{z\in\mathcal{U}} |\langle v, \varphi\circ q(z) - \varphi\circ q(y)\rangle|\Big)\,\mathrm{d}\gamma_n(v) \lesssim \Big(\max_{z\in\mathcal{U}} \|\varphi\circ q(z) - \varphi\circ q(y)\|_2\Big) \min_{w\in B_{\mathcal{S}}(y,3\tau)} \sqrt{\log\frac{\mathfrak{n}(w)}{\mathfrak{d}(w)}}. \tag{173}$$

Proof. For every $y,w \in \mathcal{S}$, by applying Lemma 49 with $\xi = \vartheta(w)$ and $\mathcal{C} = B_{\mathcal{S}}(y,9\tau)$, we see that

$$\forall\, y,w \in \mathcal{S}, \qquad \big|B_{\mathcal{S}}(y,9\tau)\cap\mathcal{N}_{\vartheta(w)}\big| \overset{(169)}{\leqslant} \vartheta(w) \overset{(168)}{=} \frac{\mathfrak{n}(w)}{\mathfrak{d}(w)}. \tag{174}$$

By combining (174) with conclusion (166) of Lemma 48, we see that

$$\forall\, y,w \in \mathcal{S}, \qquad d_{\mathcal{S}}(y,w) \leqslant 3\tau \implies \big|q\big(B_{\mathcal{S}}(y,2\tau)\big)\big| \leqslant \frac{\mathfrak{n}(w)}{\mathfrak{d}(w)}. \tag{175}$$

Fixing $y \in \mathcal{S}$ and $\mathcal{U} \subseteq B_{\mathcal{S}}(y,2\tau)$ as in Corollary 51, we get from (175) the following cardinality estimate:

$$|\varphi\circ q(\mathcal{U})| \leqslant |q(\mathcal{U})| \leqslant \big|q\big(B_{\mathcal{S}}(y,2\tau)\big)\big| \leqslant \min_{w\in B_{\mathcal{S}}(y,3\tau)} \frac{\mathfrak{n}(w)}{\mathfrak{d}(w)}. \tag{176}$$

By a simple bound on the expected maximum of a finite Gaussian process which appears in many places in the literature, including specifically in e.g. [Pis89, Lemma 4.14], we have[30]

$$\int_{\mathbb{R}^m} \Big(\max_{z\in\mathcal{U}} |\langle v, \varphi\circ q(z) - \varphi\circ q(y)\rangle|\Big)\,\mathrm{d}\gamma_n(v) \lesssim \Big(\max_{z\in\mathcal{U}} \|\varphi\circ q(z) - \varphi\circ q(y)\|_2\Big)\sqrt{\log|\varphi\circ q(\mathcal{U})|}. \tag{177}$$

The desired conclusion (173) follows from a substitution of (176) into (177). □

[30] A mechanical inspection of the (elementary) proof of [Pis89, Lemma 4.14] shows that the implicit universal constant in (177) can be taken to be less than 2, and, in fact, that it can be taken to be arbitrarily close to $\sqrt{2}$ as $|\varphi\circ q(\mathcal{U})| \to \infty$.

Observe that the proof of Corollary 51 makes the value of the compression (via composition with the rounding function q) that Construction 47 provides evident: If we would have considered φ rather than $\varphi\circ q$ in (173), then the same reasoning would have led to the factor $\min_{x\in B_{\mathcal{S}}(y,3\tau)}\sqrt{\log(\mathfrak{n}(x)/\mathfrak{d}(x))}$ in (173) being replaced by $\sqrt{\log|\mathcal{U}|}\leqslant\sqrt{\log|B_{\mathcal{S}}(y,2\tau)|}$, which is insufficient for our proof of Theorem 1.

Remark 52. *If $h:[1,\infty)\to\mathbb{R}$ is monotone increasing, then one could consider Lemma 49 with the function ϑ in* (168) *replaced by the function $(x\in\mathcal{S})\mapsto h(\mathfrak{n}(x)/\mathfrak{d}(x))$. As Construction 47 only depends on the ordering of the values in ϑ, it will output the same rounding function ρ, independently of h, and therefore Corollary 51 will remain unchanged. In the overview of the proof that we presented in Section 2, we sketched the above compression step this way while referring to nets in the level sets of the function ρ that is given in equation* (53) *of Proposition 37; using the notation of the statement Proposition 37, this corresponds to considering $h(t)=1+(\zeta/C)\sqrt{\log t}$ above, with $\mathfrak{d}(x)=\mu(B_{\mathcal{M}}(x,\tau))$ and $\mathfrak{n}(x)=\mu(B_{\mathcal{M}}(x,19\tau))$.*

We will soon explain how Construction 47 facilitates compatibility (per Definition 34) of proximity graphs of a metric space (as in e.g. definition (55) of Proposition 37, though we will work in greater generality). The following lemma is the main way that we will use such proximity information:

Lemma 53. *Fix $\tau>0$. Suppose that $(\mathcal{S},d_{\mathcal{S}})$ is a finite metric space. Given $\rho:\mathcal{S}\to[1,\infty)$, consider the graph $\mathsf{G}=\mathsf{G}_{d_{\mathcal{S}},\rho,\tau}=(\mathcal{S},E)$ whose vertex set is $\mathcal{S}$ and whose edge set $E=E_{d_{\mathcal{S}},\rho,\tau}$ is defined by*

$$\forall\, x,y\in\mathcal{S},\qquad \{x,y\}\in E\iff d_{\mathcal{S}}(x,y)\leqslant\frac{\tau}{\min\{\rho(x),\rho(y)\}}. \tag{178}$$

Then,

$$\forall\, x\in\mathcal{S},\qquad B_{\mathsf{G}}\big(x,\widetilde{\rho}(x)+1\big)\subseteq B_{\mathcal{S}}(x,2\tau)\qquad\text{where}\qquad \widetilde{\rho}(x)\stackrel{\mathrm{def}}{=}\min_{z\in B_{\mathcal{S}}(x,2\tau)}\rho(z). \tag{179}$$

Proof. We will actually prove the following local Lipschitz property of the formal identity mapping from $(\mathcal{S},d_{\mathsf{G}})$ to $(\mathcal{S},d_{\mathcal{S}})$, which holds for every $\alpha\geqslant1$ and every $x,y\in\mathcal{S}$:

$$d_{\mathsf{G}}(x,y)\leqslant(\alpha-1)\bigg(\min_{z\in B_{\mathcal{S}}(x,\alpha\tau)}\rho(z)\bigg)+1\implies d_{\mathcal{S}}(x,y)\leqslant\frac{\tau}{\min_{z\in B_{\mathcal{S}}(x,\alpha\tau)}\rho(z)}d_{\mathsf{G}}(x,y). \tag{180}$$

This is stronger than the desired inclusion in (179) since if $\alpha=2$, then the minimum that appears in (180) coincides with the function $\widetilde{\rho}$ that we introduced in (179), so (180) implies for $\alpha=2$ that

$$\forall\, x,y\in\mathcal{S},\qquad d_{\mathsf{G}}(x,y)\leqslant\widetilde{\rho}(x)+1\leqslant2\widetilde{\rho}(x)\implies d_{\mathcal{S}}(x,y)\leqslant\frac{\tau}{\widetilde{\rho}(x)}d_{\mathsf{G}}(x,y)\leqslant2\tau. \tag{181}$$

The second inequality in (181)uses the assumption $\rho\geqslant1$.

If $d_{\mathsf{G}}(x,y)=0$, then $x=y$, so (180) holds. We will next prove (180) by induction on $d_{\mathsf{G}}(x,y)$. If

$$1\leqslant d_{\mathsf{G}}(x,y)\leqslant(\alpha-1)\bigg(\min_{z\in B_{\mathcal{S}}(x,\alpha\tau)}\rho(z)\bigg)+1, \tag{182}$$

then there is $w\in V$ such that $d_{\mathsf{G}}(x,w)=d_{\mathsf{G}}(x,y)-1$ and $\{w,y\}\in E$. Hence, the induction hypothesis gives

$$d_{\mathcal{S}}(x,w)\leqslant\frac{\tau}{\min_{z\in B_{\mathcal{S}}(x,\alpha\tau)}\rho(z)}d_{\mathsf{G}}(x,w)=\frac{\tau}{\min_{z\in B_{\mathcal{S}}(x,\alpha\tau)}\rho(z)}\big(d_{\mathsf{G}}(x,w)-1\big)\stackrel{(182)}{\leqslant}(\alpha-1)\tau<\alpha\tau, \tag{183}$$

and the definition (178) of E gives

$$d_{\mathcal{S}}(w,y)\leqslant\frac{\tau}{\min\{\rho(w),\rho(y)\}}\leqslant\tau, \tag{184}$$

where we used the assumption $\rho\geqslant1$. Hence, $d_{\mathcal{S}}(x,y)\leqslant d_{\mathcal{S}}(x,w)+d_{\mathcal{S}}(w,y)\leqslant(\alpha-1)\tau+\tau=\alpha\tau$, by (183) and (184). Thus, $y\in B_{\mathcal{S}}(x,\alpha\tau)$. Also $w\in B_{\mathcal{S}}(x,\alpha\tau)$, by (183). So, $\min\{\rho(w),\rho(y)\}\geqslant\min_{z\in B_{\mathcal{S}}(x,\alpha\tau)}\rho(z)$. A substitution of this into (184) gives

$$d_{\mathcal{S}}(w,y)\leqslant\frac{\tau}{\min_{z\in B_{\mathcal{S}}(x,\alpha\tau)}\rho(z)}. \tag{185}$$

The inductive step now concludes as follows:

$$d_{\mathbb{S}}(x,y) \leqslant d_{\mathbb{S}}(x,w)+d_{\mathbb{S}}(w,y) \overset{(183)\wedge(185)}{\leqslant} \frac{\tau}{\min_{z\in B_{\mathbb{S}}(x,\alpha\tau)}\rho(z)}\big(d_{\mathsf{G}}(x,w)-1\big)+\frac{\tau}{\min_{z\in B_{\mathbb{S}}(x,\alpha\tau)}\rho(z)} = \frac{\tau}{\min_{z\in B_{\mathbb{S}}(x,\alpha\tau)}\rho(z)}d_{\mathsf{G}}(x,w). \qquad \square$$

We are now ready to perform the main purpose of this section, which is to prove Proposition 37:

Proof of Proposition 37. Take the universal constant $\zeta \geqslant 1$, which is part of the definition (53) of the function $\rho:\mathcal{M}\to[1,\infty)$ in Proposition 37, to be the implicit factor in conclusion (173) of Corollary 49.

We will apply Lemma 53 separately to each of the connected component of Γ of the graph $\mathsf{G}=(\mathcal{M},E)$ from Proposition 37 whose edges are defined in(55), with $(\mathbb{S},d_{\mathbb{S}})=(\Gamma,d_{\mathcal{M}})$ and with $\rho:\Gamma\to[1,\infty)$ replaced by $\rho_\Gamma=\rho|_\Gamma$, the restriction of the function ρ given in (53) to Γ. In accordance with (179), we have

$$\forall\, x\in\Gamma, \qquad \widetilde{\rho_\Gamma}(x)=\min_{z\in B_\Gamma(x,2\tau)}\rho(z), \tag{186}$$

where we denote $B_\Gamma(x,r)\stackrel{\mathrm{def}}{=}B_{\mathcal{M}}(x,r)\cap\Gamma$ for every $x\in\Gamma$ and $r\geqslant 0$.

The above application of Lemma 53 to $(\Gamma,d_{\mathcal{M}})$ and ρ_Γ produces a graph G_Γ whose vertex set is Γ. The definition of the edge set of G_Γ coincides per (178)with the restriction to Γ of the edge set of the graph G on $\mathcal{M}$ that is given in (55), i.e., $E_{d_\Gamma,\rho_\Gamma,\tau}=\{\{x,y\}\in E_{d_{\mathcal{M}},\rho,\tau}:\ \{x,y\}\subseteq\Gamma\}$. Consequently, as Γ is a connected component of G, we have

$$\forall\, r\in[0,\infty),\ \forall\, x\in\Gamma, \qquad B_{\mathsf{G}}(x,r)=B_{\mathsf{G}}(x,r)\cap\Gamma=B_{\mathsf{G}_\Gamma}(x,r). \tag{187}$$

So, the inclusion (179) from Lemma 53 can be written as follows for every connected component Γ of G:

$$\forall\, x\in\Gamma, \qquad B_{\mathsf{G}}(x,\widetilde{\rho_\Gamma}(x)+1)=B_{\mathsf{G}_\Gamma}(x,\widetilde{\rho_\Gamma}(x)+1)\overset{(179)}{\subseteq}B_\Gamma(x,2\tau). \tag{188}$$

Next, as in Remark 50, define $\mathfrak{d},\mathfrak{n}:\mathcal{M}\to(0,\infty)$ by setting

$$\forall\, x\in\mathcal{M}, \qquad \mathfrak{d}(x)\stackrel{\mathrm{def}}{=}\mu\big(B_{\mathcal{M}}(x,\tau)\big) \qquad \text{and} \qquad \mathfrak{n}(x)\stackrel{\mathrm{def}}{=}\mu\big(B_{\mathcal{M}}(x,19\tau)\big). \tag{189}$$

Let $\vartheta:\mathcal{M}\to[1,\infty)$ be given as in (168) for the choices of $\mathfrak{d},\mathfrak{n}$ in (189). For every connected component Γ of G denote the restrictions of $\mathfrak{d},\mathfrak{n},\vartheta$ to Γ by $\mathfrak{d}_\Gamma,\mathfrak{n}_\Gamma,\vartheta_\Gamma$, respectively. Apply Construction 47 to $(\Gamma,d_{\mathcal{M}})$ and ϑ_Γ, thus obtaining a rounding function $q_\Gamma:\Gamma\to\Gamma$. Let $q:\mathcal{M}\to\mathcal{M}$ be given by defining it to coincide with q_Γ on each connected component Γ of G, i.e.,

$$\forall\, x\in\mathcal{M}, \qquad q(x)=q_{\Gamma(x)}(x). \tag{190}$$

The requirement $q(x)\in\Gamma(x)$ of Proposition 37 holds by design. Also, (190) ensures that requirement (54) of Proposition 37 holds, by applying Lemma 48 separately to each connected component of G.

Verifying Definition 34 of C-compatibility (for $\varphi\circ q$) requires introducing two auxiliary functions $K:\mathcal{M}\to\mathbb{N}$ and $\Delta:\mathcal{M}\to[0,\infty)$; recalling (186), we will define those functions as follows:

$$\forall\, x\in\mathcal{M}, \qquad K(x)\stackrel{\mathrm{def}}{=}\big\lceil\widetilde{\rho_{\Gamma(x)}}(x)\big\rceil \qquad \text{and} \qquad \Delta(x)\stackrel{\mathrm{def}}{=}C\bigg(\max_{z\in B_{\Gamma(x)}(x,2\tau)}\|\varphi\circ q(x)-\varphi\circ q(z)\|_2\bigg). \tag{191}$$

With these choices (and the above preparation), it is quite simple to check that Definition 34 is satisfied.

For requirement (44) of Definition 34, we are given $x\in\mathcal{M}$ and $y\in B_{\mathsf{G}}(x,K(x)-1)$, as well as $z\in\mathcal{M}$ such that $\{y,z\}\in E$. The goal is to check that $\Delta(x)\leqslant\sigma(\{y,z\})$, with σ given in (56). This indeed holds since

$$y,z\in B_{\mathsf{G}}\big(x,K(x)\big)\overset{(191)}{\subseteq}B_{\mathsf{G}}\big(x,\widetilde{\rho_{\Gamma(x)}}(x)+1\big)\overset{(188)}{\subseteq}B_{\Gamma(x)}(x,2\tau). \tag{192}$$

Therefore, $x \in B_{\mathcal{M}}(y, 2\tau) \cap B_{\mathcal{M}}(z, 2\tau) \cap \Gamma$, where $\Gamma \stackrel{\mathrm{def}}{=} \Gamma(x) = \Gamma(y) = \Gamma(z)$, so we have

$$\begin{aligned}\Delta(x) &\stackrel{(191)}{=} C\left(\max_{b \in B_{\mathcal{M}}(x,2\tau)\cap\Gamma} \|\varphi\circ q(x) - \varphi\circ q(b)\|_2\right)\\ &\leqslant C\Bigg(\max_{\substack{a\in B_{\mathcal{M}}(y,2\tau)\cap B_{\mathcal{M}}(z,2\tau)\cap\Gamma\\ b\in B_{\mathcal{M}}(a,2\tau)\cap\Gamma}} \|\varphi\circ q(a) - \varphi\circ q(b)\|_2\Bigg) \stackrel{(56)}{=} \sigma(\{y,z\}).\end{aligned}$$

Requirement (45) of Definition 34 is proved as follows. If $x \in \mathcal{M}$ and $y \in N_{\mathsf{G}}(x)$, then $\Gamma(x) = \Gamma(y) \stackrel{\mathrm{def}}{=} \Gamma$, and also $d_{\mathcal{M}}(x, y) \leqslant \tau$ by (55), as $\rho \geqslant 1$. By the triangle inequality we therefore have

$$B_\Gamma(x, 2\tau) \subseteq B_\Gamma(y, 3\tau). \tag{193}$$

Apply Corollary 51 to $(\Gamma, d_{\mathcal{M}})$, the functions $\mathfrak{d}_\Gamma, \mathfrak{n}_\Gamma : \Gamma :\to (0,\infty)$,[31] and $\mathcal{U} \stackrel{\mathrm{def}}{=} B_\Gamma(y, 2\tau)$. Recalling that at the start of the proof of Proposition 37 we defined ζ to be the implicit factor in the right hand side of (173), we thus obtain the following estimate:

$$\begin{aligned}&\int_{\mathbb{R}^m}\left(\max_{z\in B_\Gamma(y,2\tau)} |\langle v, \varphi\circ q(z) - \varphi\circ q(y)\rangle|\right) \mathrm{d}\gamma_n(v)\\ &\qquad \stackrel{(173)}{\leqslant} \zeta\Big(\max_{z\in B_\Gamma(y,2\tau)} \|\varphi\circ q(z) - \varphi\circ q(y)\|_2\Big) \min_{w\in B_\Gamma(y,3\tau)} \sqrt{\log\frac{\mu(B_{\mathcal{M}}(w,19\tau))}{\mu(B_{\mathcal{M}}(w,\tau))}}\\ &\qquad \stackrel{(193)}{\leqslant} \zeta\Big(\max_{z\in B_\Gamma(y,2\tau)} \|\varphi\circ q(z) - \varphi\circ q(y)\|_2\Big) \min_{w\in B_\Gamma(x,2\tau)} \sqrt{\log\frac{\mu(B_{\mathcal{M}}(w,19\tau))}{\mu(B_{\mathcal{M}}(w,\tau))}}\\ &\qquad \stackrel{(191)}{=} \Delta(y) \min_{w\in B_\Gamma(x,2\tau)} \frac{\zeta}{C}\sqrt{\log\frac{\mu(B_{\mathcal{M}}(w,19\tau))}{\mu(B_{\mathcal{M}}(w,\tau))}}.\end{aligned} \tag{194}$$

Consequently,

$$\begin{aligned}\int_{\mathbb{R}^n}\left(\max_{z\in B_{\mathsf{G}}(y,K(y))} |\langle v, \varphi\circ q(z) - \varphi\circ q(y)\rangle|\right) \mathrm{d}\gamma_n(v) &\stackrel{(192)}{\leqslant} \int_{\mathbb{R}^n}\left(\max_{z\in B_\Gamma(y,2\tau)} |\langle v, \varphi\circ q(z) - \varphi\circ q(y)\rangle|\right) \mathrm{d}\gamma_n(v)\\ &\stackrel{(53)\wedge(194)}{<} \Delta(y) \min_{w\in B_\Gamma(x,2\tau)} \rho(w)\\ &\stackrel{(186)}{=} \Delta(y)\widetilde{\rho_\Gamma}(x)\\ &\stackrel{(191)}{\leqslant} \Delta(y)K(x).\end{aligned} \tag{195}$$

Requirement (45) of Definition 34 coincides with (195). The second step of (195) is the only place in the proof of Proposition 37 in which the specific choice of ρ in (53) is used.

Finally, the remaining requirement (46) of Definition 34 is established as follows for every $x \in \mathcal{M}$:

$$\begin{aligned}\varphi\circ q\Big(B_{\mathsf{G}}\big(x, K(x)\big)\Big) &\stackrel{(192)}{\subseteq} \varphi\circ q\big(B_{\Gamma(x)}(x, 2\tau)\big)\\ &\subseteq B_{\ell_2^n}\Big(\varphi\circ q(x), \max_{z\in B_{\Gamma(x)}(x,2\tau)} \|\varphi\circ q(x) - \varphi\circ q(z)\|_2\Big)\\ &\stackrel{(191)}{=} B_{\ell_2^n}\Big(\varphi\circ q(x), \frac{1}{C}\Delta(x)\Big).\end{aligned}$$

This concludes the proof of Proposition 37. □

[31] Recall that this is how we denoted the restrictions to Γ of the functions given in (189); as explained in Remark 50, they satisfy assumption (167) of Lemma 49.

6. A STRUCTURAL IMPLICATION OF QUASISYMMETRY

The ultimate goal of this section is to prove Proposition 38. We will start by studying some basic implications of the notion of a metric space $(\mathcal{M},d)$ being (s,ε)-quasisymmetrically Hilbertian (given in Definition 2), though many of the ensuing elementary considerations do not require the target space to be a Hilbert space. Similar issues are covered in the foundational work [TV80, Section 2].

It is convenient to introduce the following notion as it arises naturally when one wishes to iterate (3):

Terminology 54. *Given $0<s\leqslant 1$ and $\delta\geqslant 0$, say that a metric space $(\mathcal{M},d)$ has (s,δ)-non-discrete distances if for every $x,y\in\mathcal{M}$ there exists $z\in\mathcal{M}$ such that $sd(x,y)-\delta\leqslant d(x,z)\leqslant sd(x,y)$.*

The next lemma is a simple iterative application of a condition such as the requirement (3) of being quasisymmetrically Hilbertian.[32]

Lemma 55. *Fix $\alpha,\delta\geqslant 0$ and $0<s<1$. Suppose that $(\mathcal{M},d_{\mathcal{M}})$ is a metric space that has (s,δ)-non-discrete distances. Suppose furthermore that $(\mathcal{N},d_{\mathcal{N}})$ is a metric space and that $\varphi:\mathcal{M}\to\mathcal{N}$ satisfies*

$$\forall (x,y,z)\in\mathcal{M}^3,\qquad d_{\mathcal{M}}(x,y)\leqslant sd_{\mathcal{M}}(x,z)\implies d_{\mathcal{N}}\big(\varphi(x),\varphi(y)\big)\leqslant \alpha d_{\mathcal{N}}\big(\varphi(x),\varphi(z)\big). \tag{196}$$

Then, for every $\ell\in\mathbb{N}$ we have

$$\forall (x,y,z)\in\mathcal{M}^3,\qquad d_{\mathcal{M}}(x,y)\leqslant s^{\ell}d_{\mathcal{M}}(x,z)-\frac{s(1-s^{\ell-1})}{1-s}\delta\implies d_{\mathcal{N}}\big(\varphi(x),\varphi(y)\big)\leqslant \alpha^{\ell} d_{\mathcal{N}}\big(\varphi(x),\varphi(z)\big). \tag{197}$$

Proof. Observe that (196) and (197) coincide when $\ell=1$. We will proceed by induction on ℓ. Assuming that (197) holds for $\ell\in\mathbb{N}$, suppose that $x,y,z\in\mathcal{M}$ satisfy

$$d_{\mathcal{M}}(x,y)\leqslant s^{\ell+1}d_{\mathcal{M}}(x,z)-\frac{s(1-s^{\ell})}{1-s}\delta. \tag{198}$$

As $(\mathcal{M},d_{\mathcal{M}})$ is assumed to have (s,δ)-non-discrete distances, we can fix $w\in\mathcal{M}$ such that

$$sd_{\mathcal{M}}(x,z)-\delta\leqslant d_{\mathcal{M}}(x,w)\leqslant sd_{\mathcal{M}}(x,z). \tag{199}$$

The second inequality in (199) makes it possible for us to apply (196) to the triple (x,w,z), which gives

$$d_{\mathcal{N}}\big(\varphi(x),\varphi(w)\big)\leqslant \alpha d_{\mathcal{N}}\big(\varphi(x),\varphi(z)\big). \tag{200}$$

Furthermore, observe that

$$\begin{aligned} s^{\ell}d_{\mathcal{M}}(x,w)-\frac{s(1-s^{\ell-1})}{1-s}\delta &\overset{(199)}{\geqslant} s^{\ell}\big(sd_{\mathcal{M}}(x,z)-\delta\big)-\frac{s(1-s^{\ell-1})}{1-s}\delta\\ &\overset{(198)}{\geqslant} d_{\mathcal{M}}(x,y)+\frac{s(1-s^{\ell})}{1-s}\delta-s^{\ell}\delta-\frac{s(1-s^{\ell-1})}{1-s}\delta=d_{\mathcal{M}}(x,y).\end{aligned}$$

We can therefore apply the induction hypothesis (197) to the triple (x,y,w) to get that

$$d_{\mathcal{N}}\big(\varphi(x),\varphi(y)\big)\leqslant \alpha^{\ell} d_{\mathcal{N}}\big(\varphi(x),\varphi(w)\big).$$

Thanks to (200), this implies $d_{\mathcal{N}}(\varphi(x),\varphi(y))\leqslant \alpha^{\ell+1}d_{\mathcal{N}}(\varphi(x),\varphi(z))$, completing the induction step. $\square$

Remark 56. *Given $\delta>0$, recall that a metric space $(\mathcal{M},d)$ is said to be δ-discretely path connected if for every $x,y\in\mathcal{M}$ there exist $k\in\mathbb{N}$ and a k-step discrete path $z_0=x,z_1,\ldots,z_k=y\in\mathcal{M}$ joining x to y such that $d_{\mathcal{M}}(z_i,z_{i-1})\leqslant\delta$ for every $i\in[k]$. Any such space has (s,δ)-non-discrete distances for every $0<s\leqslant 1$. Indeed, if $sd(x,y)<\delta$, then choose $z=x$ in Terminology 54. If $sd(x,y)>\delta$, then let i be the largest element of $[k]\cup\{0\}$ for which $d(x,z_i)<sd(x,y)-\delta$; there is such i by the assumption $sd(x,y)-\delta>0=d(x,z_0)$. Now, $i<k$ as $d(x,z_k)=d(x,y)\geqslant sd(x,y)>sd(x,y)-\delta$. So, $d(x,z_{i+1})\geqslant sd(x,y)-\delta$ by the maximality of i. Also, $d(x,z_{i+1})\leqslant d(x,z_i)+d(z_i,z_{i+1})<sd(x,y)-\delta+\delta=sd(x,y)$. Thus $z=z_{i+1}$ works for Terminology 54.*

[32]Lemma 55 assumes that the spaces in question are metric spaces because this is the context of the present investigation, but the triangle inequality is never used in its proof.

Furthermore, if $(\mathcal{M}, d)$ *is path connected, then by continuity it has* $(s,0)$*-non-discrete distances for any* $0<s<1$*. Lemma 55 therefore implies that if* (196) *holds, then for every* $\ell\in\mathbb{N}$ *and every* $x,y,z\in\mathcal{M}$ *we have*

$$d_{\mathcal{M}}(x,y)\leqslant s^{\ell}d_{\mathcal{M}}(x,z)\implies d_{\mathcal{N}}\big(\varphi(x),\varphi(y)\big)\leqslant \alpha^{\ell}d_{\mathcal{N}}\big(\varphi(x),\varphi(z)\big). \tag{201}$$

Suppose that $d_{\mathcal{M}}(x,y)\leqslant sd_{\mathcal{M}}(x,z)$ *and write* $\ell\stackrel{\mathrm{def}}{=}\lfloor(\log(d(x,z)/d(x,y)))/\log(1/s)\rfloor\in\mathbb{N}$*. By design we have* $d_{\mathcal{M}}(x,y)\leqslant s^{\ell}d_{\mathcal{M}}(x,z)$*, so* $d_{\mathcal{N}}(\varphi(x),\varphi(y))\leqslant\alpha^{\ell}d_{\mathcal{N}}(\varphi(x),\varphi(z))$ *by* (201)*. If* $0<\alpha<1$*, then*

$$\alpha^{\ell}\leqslant\alpha^{\frac{\log\frac{d(x,z)}{d(x,y)}}{\log\frac{1}{s}}-1}=\frac{1}{\alpha}\left(\frac{d(x,y)}{d(x,z)}\right)^{\frac{\log\frac{1}{\alpha}}{\log\frac{1}{s}}}.$$

Consequently, for every distinct $x,y,z\in\mathcal{M}$ *that satisfy* $d_{\mathcal{M}}(x,y)\leqslant sd_{\mathcal{M}}(x,z)$ *we have*

$$\frac{d_{\mathcal{N}}(\varphi(x),\varphi(y))}{d_{\mathcal{N}}(\varphi(x),\varphi(z))}\leqslant\eta\bigg(\frac{d_{\mathcal{M}}(x,y)}{d_{\mathcal{M}}(x,z)}\bigg),$$

where $\eta=\eta_{\alpha,s}:[0,\infty)\to[0,\infty)$ *is given by setting* $\eta(t)=t^{(\log(1/\alpha))/\log(1/s)}/\alpha$ *for* $t\geqslant0$*, so* $\lim_{t\to0^+}\eta(t)=0$*.*

We have thus shown that for every path connected metric space $(\mathcal{M},d_{\mathcal{M}})$ *and* $0<\alpha,s<1$*, if* $\varphi:\mathcal{M}\to\mathcal{N}$ *satisfies* (196)*, then it also satisfies the condition* (2) *of being quasisymmetric, except that now* (2) *is required to hold for every distinct* $x,y,z\in\mathcal{M}$ *that satisfy* $d_{\mathcal{M}}(x,y)\leqslant sd_{\mathcal{M}}(x,z)$*, rather than for any triple of distinct points in* $\mathcal{M}$ *whatsoever. A straightforward inspection of the proof of the main result of* [Nao12a] *reveals that even though its statement assumes the existence of a quasisymmetric embedding in the classical sense* (2)*, it actually uses this weaker condition. Therefore, thanks to the above reasoning* [Nao12a] *implies in particular that if* $2<p<\infty$*, then* ℓ_p *is not* (s,ε)*-quasisymmetrically Hilbertian for any* $0<s,\varepsilon<1$*.*

The following simple lemma will also be used in the proof of Proposition 38:

Lemma 57. *Fix* $\tau,s,\alpha>0$*. Let* $(\mathcal{M},d_{\mathcal{M}})$ *and* $(\mathcal{N},d_{\mathcal{N}})$ *be metric spaces. Suppose that* $f:\mathcal{M}\to\mathcal{N}$ *satisfies*

$$\forall\, x,y,z\in\mathcal{M},\qquad d_{\mathcal{M}}(x,y)\leqslant s\tau\leqslant 2sd_{\mathcal{M}}(x,z)\implies d_{\mathcal{N}}\big(f(x),f(y)\big)\leqslant\alpha d_{\mathcal{N}}\big(f(x),f(z)\big). \tag{202}$$

Given a nonempty subset $\mathcal{E}$ *of* $\mathcal{M}\times\mathcal{M}$ *with*

$$\forall(x,y)\in\mathcal{E},\qquad d_{\mathcal{M}}(x,y)\geqslant\tau, \tag{203}$$

define $\Lambda:\mathcal{M}\to[0,\infty)$ *by setting*

$$\forall\, x\in\mathcal{M},\qquad \Lambda(x)\stackrel{\mathrm{def}}{=}\inf_{(a,b)\in\mathcal{E}}\max\big\{d_{\mathcal{N}}\big(f(x),f(a)\big),d_{\mathcal{N}}\big(f(x),f(b)\big)\big\}. \tag{204}$$

Then, $\Lambda(y)\leqslant(\alpha+1)\Lambda(x)$ *for every* $x,y\in\mathcal{M}$ *with* $d_{\mathcal{M}}(x,y)\leqslant s\tau$ *(by symmetry, also* $\Lambda(y)\geqslant\Lambda(x)/(\alpha+1)$*).*

Proof. Suppose that $x,y\in\mathcal{M}$ satisfy $d_{\mathcal{M}}(x,y)\leqslant s\tau$. If $(w,z)\in\mathcal{E}$, then in particular $d_{\mathcal{M}}(w,z)\geqslant\tau$, by (203). By the triangle inequality we therefore have $\max\{d_{\mathcal{M}}(x,w),d_{\mathcal{M}}(x,z)\}\geqslant\tau/2$; without loss of generality we may assume that $d_{\mathcal{M}}(x,z)\geqslant\tau/2$. We can now use (202) to get that

$$d_{\mathcal{N}}\big(f(x),f(y)\big)\leqslant\alpha d_{\mathcal{N}}\big(f(x),f(z)\big). \tag{205}$$

Consequently,

$$\begin{aligned}\Lambda(y)&\leqslant\max\big\{d_{\mathcal{N}}\big(f(y),f(w)\big),d_{\mathcal{N}}\big(f(y),f(z)\big)\big\}\\&\leqslant d_{\mathcal{N}}\big(f(x),f(y)\big)+\max\big\{d_{\mathcal{N}}\big(f(x),f(w)\big),d_{\mathcal{N}}\big(f(x),f(z)\big)\big\}\\&\leqslant(\alpha+1)\max\big\{d_{\mathcal{N}}\big(f(x),f(w)\big),d_{\mathcal{N}}\big(f(x),f(z)\big)\big\},\end{aligned}$$

where the last step uses (205). By taking the infimum over $(w,z)\in\mathcal{E}$, we get that $\Lambda(y)\leqslant(\alpha+1)\Lambda(x)$. □

We can now prove Proposition 38, which, as explained in Section 2, is a structural implication of $(\mathcal{M},d_{\mathcal{M}})$ being (s,ε)-quasisymmetrically Hilbertian that is used in the proof of Theorem 3 (the assumption that $(\mathcal{M},d_{\mathcal{M}})$ is (s,ε)-quasisymmetrically Hilbertian is not used elsewhere in the proof of Theorem 3).

Proof of Proposition 38. Write $n=|\mathcal{M}|$. The assumption that $(\mathcal{M},d_\mathcal{M})$ is (s,ε)-quasisymmetrically Hilbertian means that there is an injection $\varphi:\mathcal{M}\to\mathbb{R}^n$ such that

$$\forall\, x,y,z\in\mathcal{M},\qquad d_\mathcal{M}(x,y)\leqslant s d_\mathcal{M}(x,z) \implies \|\varphi(x)-\varphi(y)\|_2\leqslant (1-\varepsilon)\|\varphi(x)-\varphi(z)\|_2. \tag{206}$$

Apply Proposition 37 with τ replaced by $\beta\tau$ and C replaced by rC to get $q:\mathcal{M}\to\mathcal{M}$ that satisfies

$$\forall\, x\in\mathcal{M},\qquad d_\mathcal{M}(q(x),x)\leqslant 7\beta\tau, \tag{207}$$

and such that, recalling that the statement of Proposition 37 denotes the connected component in G of each $x\in\mathcal{M}$ by $\Gamma(x)$, if we define, in accordance with (56),

$$f\stackrel{\mathrm{def}}{=}\varphi\circ q \qquad\text{and}\qquad \forall\{x,y\}\in E,\qquad \sigma(\{x,y\})\stackrel{\mathrm{def}}{=} rC\Bigg(\max_{\substack{a\in B_\mathcal{M}(x,2\beta\tau)\cap B_\mathcal{M}(y,2\beta\tau)\cap\Gamma(x)\\ b\in B_\mathcal{M}(a,2\beta\tau)\cap\Gamma(x)}}\|f(a)-f(b)\|_2\Bigg), \tag{208}$$

then G is (rC)-compatible with f and σ. So, the first requirement of Proposition 38 holds by construction.

It remains to define Λ and show that it satisfies the rest of the assertions of Proposition 38. Firstly,

$$\forall\, x\in\mathcal{M},\qquad \mathrm{diam}_\mathcal{M}\big(\Gamma(x)\big)\geqslant\tau \implies \Lambda(x)\stackrel{\mathrm{def}}{=} C\Bigg(\min_{\substack{w,z\in\Gamma(x)\\ d_\mathcal{M}(w,z)\geqslant\tau}}\max\big\{\|f(x)-f(w)\|_2,\|f(x)-f(z)\|_2\big\}\Bigg). \tag{209}$$

The assumption $\mathrm{diam}_\mathcal{M}(\Gamma(x))\geqslant\tau$ in (209) ensures that the minimum is over a nonempty set; we complement this definition to the case $\mathrm{diam}_\mathcal{M}(\Gamma(x))<\tau$, which is uninteresting for Proposition 38, as follows:

$$\forall\, x\in\mathcal{M},\qquad \mathrm{diam}_\mathcal{M}\big(\Gamma(x)\big)<\tau \implies \Lambda(x)\stackrel{\mathrm{def}}{=}\infty.$$

This ensures that (60) holds (vacuously) and (61) holds (trivially) when $x,y\in\mathcal{M}$ are both members of the same connected component of G whose diameter is strictly smaller than τ.

Passing to the more meaningful case of $x,y\in\mathcal{M}$ with $\Gamma(x)=\Gamma(y)\stackrel{\mathrm{def}}{=}\Gamma$ and $\mathrm{diam}_\mathcal{M}(\Gamma)\geqslant\tau$, we will first check that $\Lambda(x)\neq 0$. If $w,z\in\Gamma$ satisfy $d_\mathcal{M}(w,z)\geqslant\tau$, then $d_\mathcal{M}(q(w),q(z))\geqslant\tau-14\beta\tau>0$, thanks to (207) and since $\beta<1/14$ by (57). Hence $q(w)\neq q(z)$, so because φ is injective and $f=\varphi\circ q$, either $f(x)\neq f(w)$ or $f(x)\neq f(z)$, which implies that $\Lambda(x)>0$ by (209). If $d_\mathcal{M}(x,y)\geqslant\tau$, then we can consider $(w,z)=(x,y)$ in (209) to get $\Lambda(x)\leqslant C\|f(x)-f(y)\|_2$. By symmetry, also $\Lambda(y)\leqslant C\|f(x)-f(y)\|_2$, so (60) holds.

It remains to verify (61). We will first check that the assumptions of Lemma 57 are satisfied with $\alpha=1$, s replaced by $s/8$, f replaced by Cf, and the target space $(\mathcal{N},d_\mathcal{N})$ taken to be ℓ_2^n. So, suppose that $x,y,z\in\mathcal{M}$ satisfy $d_\mathcal{M}(x,y)\leqslant s\tau/8$ and $d_\mathcal{M}(x,z)\geqslant\tau/2$. Then, by (207) and the triangle inequality for $d_\mathcal{M}$,

$$d_\mathcal{M}\big(q(x),q(y)\big)\leqslant\frac{s}{8}\tau+14\beta\tau\leqslant\frac{s}{4}\tau\qquad\text{and}\qquad d_\mathcal{M}\big(q(x),q(z)\big)\geqslant\frac{1}{2}\tau-14\beta\tau\geqslant\frac{1}{4}\tau, \tag{210}$$

where the two last steps in (210) hold since $\beta\leqslant s/112<1/56$ by (57), using $0<s,\varepsilon\leqslant\frac12$ and $r\geqslant 1$. By contrasting the two estimates in (210) we see that $d_\mathcal{M}(q(x),q(y))\leqslant s d_\mathcal{M}(q(x),q(z))$, so we may apply (206) with (x,y,z) replaced by $(q(x),q(y),q(z))$ to get $\|f(x)-f(y)\|_2\leqslant(1-\varepsilon)\|f(x)-f(z)\|_2\leqslant\|f(x)-f(z)\|_2$, as $f=\varphi\circ q$. This establishes assumption (202) of Lemma 57 with the above parameters.

If $\{x,y\}\in E$, which implies $\Gamma(x)=\Gamma(y)\stackrel{\mathrm{def}}{=}\Gamma$, and if $\mathrm{diam}_\mathcal{M}(\Gamma)\geqslant\tau$, then as $\rho\geqslant 1$ by (59), definition (58) of E gives $d_\mathcal{M}(x,y)\leqslant\beta\tau\leqslant(s/8)\tau$, using (57). Applying Lemma 57 to $\mathcal{E}=\{(w,z)\in\Gamma\times\Gamma: d_\mathcal{M}(w,z)\geqslant\tau\}$, notice that the restriction of (204) to Γ coincides with the restriction of (209) to Γ,[33] so $\Lambda(y)\leqslant 2\Lambda(x)$.

It remains to prove the second part of (61) when $\{x,y\}\in E$ and $\mathrm{diam}(\Gamma)\geqslant\tau$, where $\Gamma\stackrel{\mathrm{def}}{=}\Gamma(x)=\Gamma(y)$. By (208), we need to show that if $a\in B_\mathcal{M}(x,2\beta\tau)\cap B_\mathcal{M}(y,2\beta\tau)\cap\Gamma$ and $b\in B_\mathcal{M}(a,2\beta\tau)\cap\Gamma$, then

$$4rC\|f(a)-f(b)\|_2\leqslant\min\big\{\Lambda(x),\Lambda(y)\big\}. \tag{211}$$

[33] Recall that we chose above to apply Lemma 57 with f replaced by Cf.

To see why (211) holds, fix from now such $a,b\in \mathcal{M}$. Then, $d_{\mathcal{M}}(x,b)\leqslant d_{\mathcal{M}}(x,a)+d_{\mathcal{M}}(a,b)\leqslant 4\beta\tau$, so

$$\begin{aligned}\max\left\{d_{\mathcal{M}}\big(q(x),q(a)\big),d_{\mathcal{M}}\big(q(x),q(a)\big)\right\} &\overset{(207)}{\leqslant} \max\left\{d_{\mathcal{M}}(x,a),d_{\mathcal{M}}(x,b)\right\}+14\beta\tau\\ &\leqslant 18\beta\tau\leqslant \frac14 s^{\ell}\tau-\frac{s}{1-s}\beta\tau<\frac14 s^{\ell}\tau-\frac{s(1-s^{\ell-1})}{1-s}\beta\tau,\end{aligned}\tag{212}$$

where we introduce the notation

$$\ell=\ell(r,\varepsilon)\overset{\mathrm{def}}{=}\left\lceil\frac{\log(8r)}{\varepsilon}\right\rceil\in\mathbb{N},\tag{213}$$

using which the penultimate inequality in (212) is straightforward to verify thanks to the assumption (57) on β and because $0<s,\varepsilon\leqslant 1/2$ and $r\geqslant 1$.

Because Γ is the connected component of x (and y) in the graph G, and thanks to (58) and the fact that $\rho\geqslant 1$ by (59), the $d_{\mathcal{M}}$-distance between the endpoints of every edge of G is at most $\beta\tau$, the metric space $(\Gamma,d_{\mathcal{M}})$ is $\beta\tau$-discretely path connected, and hence it has $(s,\beta\tau)$-non-discrete distances (recall Terminology 54 and Remark 56). We can therefore apply Lemma 55 to the metric space $(\Gamma,d_{\mathcal{M}})$ with s,φ as above, $\alpha=1-\varepsilon$ and $\delta=\beta\tau$, in which case assumption (196) of Lemma 55 for $\mathcal{N}=\ell_2^n$ coincides with (206).

The assumption $\mathrm{diam}(\Gamma)\geqslant\tau$ means that there are $w,z\in\Gamma$ for which $d_{\mathcal{M}}(w,z)\geqslant\tau$. By the triangle inequality for $d_{\mathcal{M}}$, it follows that $\max\{d_{\mathcal{M}}(x,w),d_{\mathcal{M}}(x,z)\}\geqslant\tau/2$, so we may assume without loss of generality that $d_{\mathcal{M}}(x,z)\geqslant\tau/2$. Using (207) and the triangle inequality for $d_{\mathcal{M}}$, we therefore have

$$d_{\mathcal{M}}\big(q(x),q(z)\big)\geqslant\frac12\tau-14\beta\tau\geqslant\frac14\tau.\tag{214}$$

By combining (212) and (214), we see that

$$\max\left\{d_{\mathcal{M}}\big(q(x),q(a)\big),d_{\mathcal{M}}\big(q(x),q(b)\big)\right\}<s^{\ell}d_{\mathcal{M}}\big(q(x),q(z)\big)-\frac{s(1-s^{\ell-1})}{1-s}\beta\tau.\tag{215}$$

Since $a,b,x,z\in\Gamma$ and Proposition 37 asserts that $q(\Gamma)\subseteq\Gamma$, we have $q(a),q(b),q(x),q(z)\in\Gamma$. Because we checked that the assumptions of Lemma 55 hold for $(\Gamma,d_{\mathcal{M}})$ with the above parameters, and (215) corresponds to the assumption in (197), then using that $f=\varphi\circ q$ we get from Lemma 55 (applied to φ twice, once to the triple $(q(x),q(a),q(z))\in\Gamma^3$, and once to the triple $(q(x),q(b),q(z))\in\Gamma^3$) that

$$\begin{aligned}\max\left\{\|f(x)-f(a)\|_2,\|f(x)-f(b)\|_2\right\}&\leqslant(1-\varepsilon)^{\ell}\|f(x)-f(z)\|_2\leqslant e^{-\varepsilon\ell}\|f(x)-f(z)\|_2\\ &\overset{(213)}{\leqslant}\frac{1}{8r}\|f(x)-f(z)\|_2\leqslant\frac{1}{8r}\max\left\{\|f(x)-f(w)\|_2,\|f(x)-f(z)\|_2\right\}.\end{aligned}\tag{216}$$

By taking the minimum over all $w,z\in\Gamma$ with $d_{\mathcal{M}}(w,z)\geqslant\tau$ while recalling (209), we get from (216) that

$$\max\left\{\|f(x)-f(a)\|_2,\|f(x)-f(b)\|_2\right\}\leqslant\frac{\Lambda(x)}{8rC}.$$

By the triangle inequality for ℓ_2^n, this implies that $4rC\|f(a)-f(b)\|_2\leqslant\Lambda(x)$. The symmetric reasoning shows that also $4rC\|f(a)-f(b)\|_2\leqslant\Lambda(y)$. This completes the proof of the desired inequality (211). □

7. Proof of Theorem 39

As Theorem 39 treats a general class of probability measures on $\mathcal{M}\times\mathcal{M}$, which is needed for the duality step that will be performed in Section 8, a fractional version of Theorem 35 is required. Prior to passing to the proof of Theorem 39, we will start this section by explaining this formal consequence of Theorem 35.

Recall that $\nu(\mathsf{G})$ denotes the matching number of a graph $\mathsf{G}=(V,E)$, i.e., it is the maximum size of a pairwise disjoint collection of edges in E. The fractional matching number of G, denoted $\nu^*(\mathsf{G})$, is the maximum of $\sum_{e\in E}\phi(e)$ over all possible $\phi:E\to[0,1]$ that satisfy the following constraints:

$$\forall x\in V,\qquad \sum_{\substack{e\in E\\ x\in e}}\phi(e)\leqslant 1.\tag{217}$$

Clearly $\nu^*(\mathsf{G}) \geqslant \nu(\mathsf{G})$, since $\nu(\mathsf{G})$ is the maximum of the same optimization problem as above if we add to it the further requirement that ϕ only takes values in $\{0,1\}$. Conversely, by [Lov74] we have $\nu^*(\mathsf{G}) \leqslant \frac{3}{2}\nu(\mathsf{G})$, with equality holding if and only if G is a disjoint union of triangles; see [Für81, Corollary 1] (also, e.g. the recent exposition [CKO16, Corollary 9]). We therefore have the following corollary of Theorem 35:

Corollary 58. *Fix $C \geqslant 1$ and $n \in \mathbb{N}$. Suppose that $\mathsf{G} = (V,E)$ is a graph that is C-compatible with $f : V \to \mathbb{R}^n$ and $\sigma : E \to [0,\infty)$. Then,*[34]

$$\int_{\mathbb{R}^n} \nu^*\big(\mathsf{G}(v; f,\sigma)\big)\,\mathrm{d}\gamma_n(v) < 9e^{-\frac{1}{4}C^2}|V|. \tag{218}$$

We will, in fact, need below the following vertex-weighted version of Corollary 58, which is a quick consequence of it thanks to the immediate fact that the notion of C-compatibility is preserved by blowups (we soon will explain what this means). Given a graph $\mathsf{G} = (V,E)$ and vertex weights $Q : V \to [0,\infty)$, let $\nu^*_Q(\mathsf{G})$ be the maximum of $\sum_{e\in E}\phi(e)$ over all possible $\phi : E \to [0,\infty)$ that satisfy the following constraints:

$$\forall\, x \in V, \qquad \sum_{\substack{e\in E\\ x\in e}} \phi(e) \leqslant Q(x). \tag{219}$$

Thus, $\nu^*(\mathsf{G})$ corresponds to the special $Q \equiv 1$.

Suppose that $Q(V) \subseteq \mathbb{N}\cup\{0\}$. Write $V' \stackrel{\mathrm{def}}{=} V \smallsetminus Q^{-1}(0)$ and $E' \stackrel{\mathrm{def}}{=} \{e \in E :\ e \subseteq V'\}$. The graph $\mathsf{G}' \stackrel{\mathrm{def}}{=} (V',E')$ is obtained from G by deleting all the vertices on which Q vanishes and all of the edges that are incident to such a vertex. For every $x \in V'$ introduce $Q(x)$-copies $\{x_1,\ldots,x_{Q(x)}\}$ of x, and define

$$U \stackrel{\mathrm{def}}{=} \bigcup_{x\in V'} \big\{x_1,\ldots,x_{Q(x)}\big\} \qquad \text{and} \qquad F \stackrel{\mathrm{def}}{=} \bigcup_{\{x,y\}\in E'} \big\{\{x_i,y_j\} : (i,j) \in [Q(x)]\times[Q(y)]\big\}.$$

Thus, $\mathsf{H} \stackrel{\mathrm{def}}{=} (U,F)$ is the standard $Q|_{V'}$-blowup of G' in which each vertex x of G' is replaced by a "cloud" consisting of $Q(x)$ copies of x, and every edge of G' is replaced by the complete bipartite graph between the cloud of x and the cloud of y. It is straightforward to check as follows that $\nu^*_Q(\mathsf{G}) = \nu^*(\mathsf{H})$.

Suppose that $\phi : E \to [0,\infty)$ satisfies the constraints (219). We "lift" ϕ to $\phi^* : F \to [0,1]$ by setting

$$\forall \{x,y\} \in E',\ \forall (i,j) \in [Q(x)]\times[Q(y)], \qquad \phi^*(\{x_i,y_j\}) = \frac{\phi(\{x,y\})}{Q(x)Q(y)}.$$

Then, $\sum_{f\in F}\phi^*(f) = \sum_{e\in E}\phi(e)$, and by (219) we know that ϕ^* satisfies the analogue of the constraints (217) for H. This shows that $\nu^*_Q(\mathsf{G}) \leqslant \nu^*(\mathsf{H})$. In the reverse direction, for $\Phi : F \to [0,1]$ define $\Phi_* : E \to [0,\infty)$ by

$$\forall \{x,y\} \in E, \qquad \Phi_*(\{x,y\}) = \sum_{i=1}^{Q(x)}\sum_{j=1}^{Q(y)} \Phi(\{x_i,y_j\}).$$

By definition, $\sum_{f\in F}\Phi(f) = \sum_{e\in E}\Phi_*(e)$, and if Φ satisfies the analogue of the constraints (217) for H, then Φ_* satisfies (219). Therefore, $\nu^*(\mathsf{H}) \leqslant \nu^*_Q(\mathsf{G})$, so we indeed have $\nu^*_Q(\mathsf{G}) = \nu^*(\mathsf{H})$.

Continuing with the above setting and notation, suppose that G is C-compatible with $f : V \to \mathbb{R}^n$ and $\sigma : E \to [0,\infty)$ for some $n \in \mathbb{N}$ and $C > 0$, per Definition 34. Lift f to $f^* : U \to \mathbb{R}^n$ by setting $f^*(x_i) = f(x)$ for every $x \in V'$ and $i \in [Q(x)]$. Also lift σ to $\sigma^* : F \to [0,\infty)$ by setting $\sigma^*(\{x_i,y_j\}) = \sigma(\{x,y\})$ for every $\{x,y\} \in E$ and $(i,j) \in [Q(x)]\times[Q(y)]$. Then H is C-compatible with f^* and σ^*, as seen immediately by lifting the mappings $K : V \to \mathbb{R}^n$ and $\Delta : V \to [0,\infty)$ from Definition 34 in the same manner, i.e., defining $K^*(x_i) = K(x)$ and $\Delta^*(x_i) = \Delta(x)$ for every $x \in V'$ and $i \in [Q(x)]$. Therefore, by Corollary 58 we have

$$\int_{\mathbb{R}^n} \nu^*\big(\mathsf{H}(v; f^*,\sigma^*)\big)\,\mathrm{d}\gamma_n(v) < 9e^{-\frac{1}{4}C^2}|U| = 9e^{-\frac{1}{4}C^2}\sum_{x\in V} Q(x).$$

[34]The function $(v \in \mathbb{R}^n) \mapsto \nu^*(\mathsf{G}(v; f,\sigma))$ in (218) is Borel-measurable because by (42) the set $\{v \in \mathbb{R}^n :\ E(v; f,\sigma) = E'\}$ is Borel for every fixed $E' \subseteq E$, and there are finitely many such E'.

For each $v\in\mathbb{R}^n$, an inspection of the definition of the Euclidean sparsification $\mathsf{H}(v;f^*,\sigma^*)$ reveals that it coincides with the $Q|_{V'}$-blowup of $\mathsf{G}(v;f,\sigma)$. Consequently, $\nu^*(\mathsf{H}(v;f^*,\sigma^*))=\nu^*_Q(\mathsf{G}(v;f,\sigma))$. Hence,

$$\int_{\mathbb{R}^n}\nu^*_Q\big(\mathsf{G}(v;f,\sigma)\big)\,\mathrm{d}\gamma_n(v)<9e^{-\frac14 C^2}\sum_{x\in V}Q(x). \tag{220}$$

We derived (220) under the assumption that Q takes values in the integers, but it follows formally from this that (220) holds for any $Q:V\to[0,\infty)$. Indeed, by considering an arbitrarily precise approximation of Q by rational-valued vertex weights, we may assume that Q takes values in $\mathbb{Q}$, and then by rescaling by the common denominator of these values, the general case follows from the case $Q(V)\subseteq\mathbb{N}\cup\{0\}$.

Corollary 59. *Fix $C\geqslant 1$ and $n\in\mathbb{N}$. Suppose that $\mathsf{G}=(V,E)$ is a graph that is C-compatible with $f:V\to\mathbb{R}^n$ and $\sigma:E\to[0,\infty)$. Then, for every $Q:V\to[0,\infty)$ we have*

$$\int_{\mathbb{R}^n}\nu^*_Q\big(\mathsf{G}(v;f,\sigma)\big)\,\mathrm{d}\gamma_n(v)<9e^{-\frac14 C^2}\sum_{x\in V}Q(x). \tag{221}$$

We will also use the following general lemma about vertex-weighted fractional matching numbers:

Lemma 60. *Suppose that V is a finite set and that ω is a symmetric nonnegative measure on $V\times V$. Define $Q=Q_\omega:V\to[0,\infty)$ by setting*

$$\forall\, x\in V,\qquad Q(x)\stackrel{\mathrm{def}}{=}\sum_{y\in V}\omega(x,y). \tag{222}$$

Let $L,R\subseteq V$ be disjoint subsets of V. Suppose that $\mathsf{G}=(L\cup R,E)$ is a bipartite graph whose sides are L,R (thus, $|e\cap L|=|e\cap R|=1$ for every $e\in E$). Then, there exist subsets $L_0^*\subseteq L$ and $R_0^*\subseteq R$ such that

$$\omega(L_0^*\times R_0^*)\geqslant\omega(L\times R)-2\nu^*_{Q|_{L\cup R}}(\mathsf{G})\qquad\text{and}\qquad\forall\,(x,y)\in L_0^*\times R_0^*,\qquad\{x,y\}\notin E. \tag{223}$$

Proof. We will slightly abuse notation by letting Q also denote the restriction $Q|_{L\cup R}$ of Q to $L\cup R$; thus, the Q-weighted fractional matching number $\nu^*_{Q|_{L\cup R}}(\mathsf{G})$ that appears in (223) will be denoted $\nu^*_Q(\mathsf{G})$ below

Let $\mathsf{FM}_Q(\mathsf{G})\subseteq\mathbb{R}^E$ be the Q-weighted fractional matching polytope of G, i.e.,

$$\mathsf{FM}_Q(\mathsf{G})\stackrel{\mathrm{def}}{=}\Big\{\phi:E\to[0,\infty):\sum_{\substack{e\in E\\ x\in e}}\phi(e)\leqslant Q(x)\text{ for all }x\in V\Big\}.$$

In other words, $\mathsf{FM}_Q(\mathsf{G})$ consists of all of the edge weights $\phi:E\to[0,\infty)$ that satisfy the constraints that appear in (219). By the compactness of $\mathsf{FM}_Q(\mathsf{G})$, we may fix from now a maximizer ϕ^* of the fractional matching problem for G with vertex weights Q, i.e., $\phi^*:E\to[0,\infty)$ belongs to $\mathsf{FM}_Q(\mathsf{G})$ and

$$\sum_{e\in E}\phi^*(e)=\nu^*_Q(\mathsf{G})=\sup_{\phi\in\mathsf{FM}_Q(\mathsf{G})}\sum_{e\in E}\phi(e). \tag{224}$$

Define $Q^*:L\cup R\to[0,\infty)$ by setting

$$\forall\, x\in L\cup R,\qquad Q^*(x)\stackrel{\mathrm{def}}{=}\sum_{\substack{e\in E\\ x\in e}}\phi^*(e)\leqslant Q(x), \tag{225}$$

where the last inequality in (225) holds because $\phi^*\in\mathsf{FM}_Q(\mathsf{G})$. Then, since G is bipartite we have

$$\sum_{x\in L}Q^*(x)+\sum_{y\in R}Q^*(y)\stackrel{(225)}{=}2\sum_{e\in E}\phi^*(e)\stackrel{(224)}{=}2\nu^*_Q(\mathsf{G}). \tag{226}$$

Next, define $L_0^*\subseteq L$ and $R_0^*\subseteq B$ as follows:

$$L_0^*\stackrel{\mathrm{def}}{=}\big\{x\in L:\ Q^*(x)<Q(x)\big\}\qquad\text{and}\qquad R_0^*\stackrel{\mathrm{def}}{=}\big\{x\in R:\ Q^*(x)<Q(x)\big\}. \tag{227}$$

In other words, thanks to the last inequality in (225), we can also write

$$L\smallsetminus L_0^*=\big\{x\in L:\ Q^*(x)=Q(x)\big\}\qquad\text{and}\qquad R\smallsetminus R_0^*=\big\{x\in R:\ Q^*(x)=Q(x)\big\}. \tag{228}$$

The second requirement in (223) is satisfied because if there were $x \in L_0^*$ and $y \in R_0^*$ with $\{x,y\} \in E$, then by (227) there would be $\varepsilon > 0$ such that

$$Q^*(x) + \varepsilon \leqslant Q(x) \qquad \text{and} \qquad Q^*(y) + \varepsilon \leqslant Q(y). \tag{229}$$

Hence, if we consider $\phi : E \to [0,\infty)$ by setting

$$\forall e \in E, \qquad \phi(e) \stackrel{\mathrm{def}}{=} \begin{cases} \phi^*(e) & \text{if } e \in E \smallsetminus \{x,y\}, \\ \phi^*(e) + \varepsilon & \text{if } e = \{x,y\}, \end{cases}$$

then $\phi \in \mathsf{FM}_Q(\mathsf{G})$ thanks to the fact that $\phi^* \in \mathsf{FM}_Q(\mathsf{G})$ and (229), yet $\sum_{e\in E} \phi(e) = \sum_{e \in E} \phi^*(e) + \varepsilon > \nu_Q^*(\mathsf{G})$ by the first equality in (224). This is in contradiction to the second equality in (224).

Finally, the first requirement in (223) is justified as follows:

$$\begin{aligned}
\omega(L_0^* \times R_0^*) &= \omega(L \times R) - \sum_{x \in L} \sum_{y \in R \smallsetminus R_0^*} \omega(x,y) - \sum_{x \in L \smallsetminus L_0^*} \sum_{y \in R_0^*} \omega(x,y) \\
&\geqslant \omega(L \times R) - \sum_{y \in R \smallsetminus R_0^*} \sum_{x \in V} \omega(x,y) - \sum_{x \in L \smallsetminus L_0^*} \sum_{y \in V} \omega(x,y) \\
&= \omega(L \times R) - \sum_{y \in R \smallsetminus R_0^*} \sum_{x \in V} \omega(y,x) - \sum_{x \in L \smallsetminus L_0^*} \sum_{y \in V} \omega(x,y) \\
&\stackrel{(222)}{=} \omega(L \times R) - \sum_{y \in R \smallsetminus R_0^*} Q(y) - \sum_{x \in L \smallsetminus L_0^*} Q(x) \\
&\stackrel{(228)}{=} \omega(L \times R) - \sum_{y \in R \smallsetminus R_0^*} Q^*(y) - \sum_{x \in L \smallsetminus L_0^*} Q^*(x) \\
&\geqslant \omega(L \times R) - \sum_{y \in R} Q^*(y) - \sum_{x \in L} Q^*(x) \\
&\stackrel{(226)}{=} \omega(L \times R) - 2\nu_Q^*(\mathsf{G}),
\end{aligned}$$

where the third step above is where the symmetry assumption on ω is used. □

With the above preparation at hand, we can now prove Theorem 39:

Proof of Theorem 39. Denote $n \stackrel{\mathrm{def}}{=} |\mathcal{M}|$. Let $\alpha_0 \geqslant 1$ be a universal constant whose value will be specified at the end of the ensuing reasoning, to satisfy constraints that will arise as we go along; see (241). Suppose that $\alpha \geqslant \alpha_0$ and, letting $\zeta \geqslant 1$ be the universal constant that appears in (59), define

$$r \stackrel{\mathrm{def}}{=} \zeta\alpha \geqslant 1. \tag{230}$$

The statement of Theorem 39 assumes that $\beta > 0$ satisfies (62). Our next step will be to apply Proposition 38 with this choice of β, for which we need to know that the assumption (57) of Proposition 38 holds. This requirement simplifies to $\alpha \geqslant 3\log(8r) = 3\log(8\zeta\alpha)$ using (230), and since our only assumption on α here is that $\alpha \geqslant \alpha_0$, for this to be satisfied it suffices to impose the following first restriction on α_0:

$$\alpha_0 \geqslant 6\log(48\zeta). \tag{231}$$

So, apply Proposition 38 to $(\mathcal{M}, d_{\mathcal{M}})$ with the datum $s, \varepsilon, r, \tau, C, \beta$. This produces a graph $\mathsf{G} = (\mathcal{M}, E)$ that is given by (58), and the function $\rho : \mathcal{M} \to [1,\infty)$ that appears in (58) is defined in (59). Proposition 38 also produces $f : \mathcal{M} \to \mathbb{R}^n$, $\sigma : E \to [0,\infty)$ and $\Lambda : \mathcal{M} \to (0,\infty]$, such that G is (rC)-compatible with f and σ, and the requirements in (60) and (61) are satisfied for every $x, y \in \mathcal{M}$ that belong to the same connected component of G. The choice of r in (230) was made to ensure that the function ρ in (59) coincides with the function ρ in (64), i.e., the output of Proposition 38 is consistent with the statement of Theorem 39.

Part of the above consequences of Proposition 38 makes it possible for us to use Proposition 36 as follows. The first requirement in (61) coincides with assumption (48) of Proposition 36. Also, because the measure ω of Theorem 39 is assumed to be supported on $\{(x,y)\in \mathcal{M}\times\mathcal{M}:\ d_{\mathcal{M}}(x,y)\geqslant \tau\}$ and the second inequality in (60) holds under the assumption $d_{\mathcal{M}}(x,y)\geqslant\tau$, we have

$$\forall\, x,y\in\mathcal{M},\qquad \omega(x,y)>0 \implies \|Cf(x)-Cf(y)\|_2\geqslant \max\big\{\Lambda(x),\Lambda(y)\big\}.$$

Therefore, the function Cf satisfies[35] assumption (49) of Proposition 36.

We may thus apply Proposition 36 to obtain for every $v\in\mathbb{R}^n$ subsets $A(v),B(v)$ of $\mathcal{M}$ that satisfy the stated (rudimentary) measurability requirements, and for which we have, by rewriting conclusion (51) of Proposition 36 with f replaced by Cf, the following point-wise inequality:

$$\forall\, v\in\mathbb{R}^n,\ \forall (x,y)\in A(v)\times B(v),\qquad \{x,y\}\in E \implies |\langle v, f(x)-f(y)\rangle|>\max\big\{\Lambda(x),\Lambda(y)\big\}. \tag{232}$$

Conclusion (52) of Proposition 36 also holds, i.e., for some universal constants $0<\theta\leqslant 1\leqslant\kappa$, we have

$$\int_{\mathbb{R}^n}\omega\big(A(v)\times B(v)\big)\,\mathrm{d}\gamma_n(v)\geqslant \theta e^{-\kappa C^2}. \tag{233}$$

Observe that $A(v)$ and $B(v)$ are disjoint for every $v\in\mathbb{R}^n$. Indeed, the definition (58) of E implies that $\{x,x\}\in E$ for every $x\in\mathcal{M}$ (i.e., by definition G has a self-loop at each of its vertices), so if it were the case that $x\in A(v)\cap B(v)$, then we would get a contradiction to the strict inequality in (232) in its special case $x=y$. We can therefore consider the bipartite graph $\mathsf{G}_v=(A(v)\cup B(v),E_v)$ that is defined by

$$\forall\, x,y\in A(v)\cup B(v),\qquad \{x,y\}\in E_v \iff \{x,y\}\in E \text{ and } (x,y)\in\big(A(v)\times B(v)\big)\cup\big(B(v)\times A(v)\big). \tag{234}$$

We will next make the following observation about G_v that will be especially important to the ensuing reasoning. For every $v\in\mathbb{R}^n$, the graph G_v is a subgraph of the Euclidean sparsification $\mathsf{G}(v;f,\sigma)$ per its definition in (42) and (43), for the f and σ that were obtained above from Proposition 38. Indeed, by (234), if $\{x,y\}\in E_v$, then $\{x,y\}\in E$ and either $(x,y)\in A(v)\times B(v)$ or $(y,x)\in A(v)\times B(v)$. Hence, $|\langle v,f(x)-f(y)\rangle|>\max\{\Lambda(x),\Lambda(y)\}$ by (232). But, $\max\{\Lambda(x),\Lambda(y)\}\geqslant\min\{\Lambda(x),\Lambda(y)\}\geqslant 4\sigma(\{x,y\})$ using the second inequality in conclusion (61) of Proposition 36. So, $|\langle v,f(x)-f(y)\rangle|>4\sigma(\{x,y\})$, which, per (42), is precisely the requirement for $\{x,y\}\in E$ to be an edge of the Euclidean sparsification $\mathsf{G}(v;f,\sigma)$.

Consider the function $Q=Q_\omega:\mathcal{M}\to[0,\infty)$ that is given by setting $Q(x)=\sum_{y\in\mathcal{M}}\omega(x,y)$ for every $x\in\mathcal{M}$. The fact that G_v is a subgraph of $\mathsf{G}(v;f,\sigma)$ for every $v\in\mathbb{R}^n$ implies in particular the following point-wise estimate on its maximal fractional matching with respect to the vertex weights Q:

$$\forall\, v\in\mathbb{R}^n,\qquad \nu^*_{Q|_{A(v)\cup B(v)}}(\mathsf{G}_v)\leqslant \nu^*_Q\big(\mathsf{G}(v;f,\sigma)\big). \tag{235}$$

Both sides of the inequality in (235) are Borel-measurable functions of v since there are finitely many subgraphs H of G and the sets $\{v\in\mathbb{R}^n:\ \mathsf{G}_v=\mathsf{H}\}$ and $\{v\in\mathbb{R}^n:\ \mathsf{G}(v;f,\sigma)=\mathsf{H}\}$ are Borel subsets of $\mathbb{R}^n$ by the measurability of $v\mapsto(A(v),B(v))$ in Proposition 36, and the definition of $\mathsf{G}(v;f,\sigma)$. Integrate (235) with respect to γ_n while substituting the fact that G is (rC)-compatible with f and σ into Corollary 59, to get

$$\int_{\mathbb{R}^n}\nu^*_{Q|_{A(v)\cup B(v)}}(\mathsf{G}_v)\,\mathrm{d}\gamma_n(v)\overset{(221)\wedge(235)}{<} 9e^{-\frac14 r^2C^2}\sum_{x\in\mathcal{M}}Q(x)\overset{(222)}{=}9e^{-\frac14 r^2C^2}\omega(\mathcal{M}\times\mathcal{M})=9e^{-\frac14 r^2C^2}. \tag{236}$$

Next, for each $v\in\mathbb{R}^n$ apply Lemma 60 to G_v. This yields $A_0^*(v)\subseteq A(v)$ and $B_0^*(v)\subseteq B(v)$ such that

$$\omega\big(A_0^*(v)\times B_0^*(v)\big)\geqslant \omega\big(A(v)\times B(v)\big)-2\nu^*_{Q|_{A(v)\cup B(v)}}(\mathsf{G}_v), \tag{237}$$

and

$$\forall (x,y)\in A_0^*(v)\times B_0^*(v),\qquad \{x,y\}\notin E_v. \tag{238}$$

The mapping $(v\in\mathbb{R}^n)\mapsto(A_0^*(v),B_0^*(v))$ is Borel-measurable because it is the composition of the mappings $v\mapsto(A(v),B(v))$ and $(A(v),B(v))\mapsto(A_0^*(v),B_0^*(v))$, the first of which is Borel-measurable by Proposition 36 and the second of which is a mapping between finite sets.

[35] In fact, we get here that Cf satisfies a stronger form of (49) with $\max\{\cdot,\cdot\}$ in place of $\min\{\cdot,\cdot\}$ in the right hand side, but we do not need to take advantage of this herein.

By the definition (58) of E, it follows from (234) and (238) that

$$\forall (x,y)\in A_0^*(v)\times B_0^*(v),\qquad d_{\mathcal{M}}(x,y)>\frac{\beta\tau}{\min\{\rho(x),\rho(y)\}}. \tag{239}$$

Furthermore,

$$\begin{aligned}\int_{\mathbb{R}^n}\omega\big(A_0^*(v)\times B_0^*(v)\big)\,\mathrm{d}\gamma_n(v)&\overset{(237)}{\geqslant}\int_{\mathbb{R}^n}\omega\big(A(v)\times B(v)\big)\,\mathrm{d}\gamma_n(v)-2\int_{\mathbb{R}^n}\nu^*_{Q|_{A(v)\cup B(v)}}(\mathsf{G}_v)\,\mathrm{d}\gamma_n(v)\\&\overset{(233)\wedge(236)}{\geqslant}\theta e^{-\kappa C^2}-9e^{-\frac14 r^2C^2}\geqslant\frac{\theta}{2}e^{-\kappa C^2},\end{aligned} \tag{240}$$

provided $r\geqslant 2\sqrt{\kappa+\log(18/\theta)}$, so that the penultimate step of (240) is valid (this is where the assumption $C\geqslant 1$ is used). Recalling the dependence (230) of r on α, and that we are assuming $\alpha\geqslant\alpha_0$, this will be satisfied if $\alpha_0\geqslant(2/\zeta)\sqrt{\kappa+\log(18/\theta)}$. Earlier we also required that the restriction (231) on α_0 holds, so altogether the entirety of the requirements that we impose on α_0 will be satisfied by the following choice:

$$\alpha_0\overset{\mathrm{def}}{=}\max\left\{\frac{2}{\zeta}\sqrt{\kappa+\log\frac{18}{\theta}},6\log(48\zeta)\right\}. \tag{241}$$

We have thus almost established the assertions of Theorem 39, except that $A_0^*(v)$ or $B_0^*(v)$ could possibly be empty for some $v\in\mathbb{R}^n$, while the subsets of $\mathcal{M}$ that Theorem 39 outputs are required (for convenience in subsequent applications) to be nonempty. This further requirement can be ensured since we are assuming in Theorem 39 that $\tau\geqslant\mathrm{diam}(\mathcal{M})$, so we can fix $x_0,y_0\in\mathcal{M}$ with $d_{\mathcal{M}}(x_0,y_0)\geqslant\tau$ and define

$$\forall v\in\mathbb{R}^n,\qquad \big(A^*(v),B^*(v)\big)\overset{\mathrm{def}}{=}\begin{cases}\big(A_0^*(v),B_0^*(v)\big) & \text{if }\emptyset\notin\big\{A_0^*(v),B_0^*(v)\big\},\\ \big(\{x_0\},\{y_0\}\big), & \text{if }\emptyset\in\big\{A_0^*(v),B_0^*(v)\big\}.\end{cases}$$

Then, the mapping $(v\in\mathbb{R}^n)\mapsto(A^*(v),B^*(v))$ is Borel-measurable as it is the composition of the mappings $v\mapsto(A_0^*(v),B_0^*(v))$ and $(A_0^*(v),B_0^*(v))\mapsto(A^*(v),B^*(v))$, the first of which we checked is measurable and the second of which is between finite sets. Because $0<\beta\leqslant 1$ and $\rho\geqslant 1$, it follows from (239) that the first desired conclusion (63) of Theorem 39 holds, and it follows from (240) that the second desired conclusion (65) of Theorem 39 holds. □

8. Duality

Our goal here is to deduce Theorem 3 from Theorem 39, which is primarily straightforward duality:

Lemma 61. *Fix $0\leqslant\mathfrak{p}\leqslant 1$. Suppose that X is a finite set. Let $\phi:X\times X\to\mathbb{R}$ be a symmetric real-valued function, and let S be a nonempty subset of $X\times X$. Fix also any function $\psi:X\to\mathbb{R}$ and define*

$$\mathcal{F}\overset{\mathrm{def}}{=}\Big\{(A,B)\in\big(2^X\setminus\{\emptyset\}\big)\times\big(2^X\setminus\{\emptyset\}\big):\ \forall(x,y)\in A\times B,\quad \phi(x,y)\geqslant\max\big\{\psi(x),\psi(y)\big\}\Big\}. \tag{242}$$

Assume that for every probability measure ω on $X\times X$ whose support is contained in S there exists a probability measure μ_ω on $\mathcal{F}$ that satisfies:

$$\int_{\mathcal{F}}\frac{\omega(A\times B)+\omega(B\times A)}{2}\,\mathrm{d}\mu_\omega(A,B)\geqslant\mathfrak{p}. \tag{243}$$

Then, there exists a probability measure $\mathbb{P}$ on $2^X\setminus\{\emptyset\}$ such that

$$\forall(x,y)\in S,\qquad \mathbb{P}\Big[\emptyset\neq\mathcal{Z}\subseteq X:\ x\in\mathcal{Z}\quad\text{and}\quad\min_{z\in\mathcal{Z}}\phi(y,z)\geqslant\psi(y)\Big]\geqslant\mathfrak{p}. \tag{244}$$

Proof. For a finite set U we denote the simplex of all the probability measures on U by Δ_U. Observe that

$$\min_{\omega\in\Delta_S}\max_{\mu\in\Delta_{\mathcal{F}}}\int_{\mathcal{F}}\frac{\omega(A\times B)+\omega(B\times A)}{2}\,\mathrm{d}\mu(A,B)\geqslant\min_{\omega\in\Delta_S}\int_{\mathcal{F}}\frac{\omega(A\times B)+\omega(B\times A)}{2}\,\mathrm{d}\mu_\omega(A,B)\overset{(243)}{\geqslant}\mathfrak{p}.$$

By the minimax theorem [vN28, Sim95], we therefore have

$$\begin{aligned}
p &\leqslant \max_{\mu\in\Delta_{\mathcal{F}}}\min_{\omega\in\Delta_S}\int_{\mathcal{F}}\frac{\omega(A\times B)+\omega(B\times A)}{2}\,\mathrm{d}\mu(A,B)\\
&=\max_{\mu\in\Delta_{\mathcal{F}}}\min_{\omega\in\Delta_S}\left(\frac12\sum_{(A,B)\in\mathcal{F}}\sum_{(x,y)\in A\times B}\omega(x,y)\mu(A,B)+\frac12\sum_{(A,B)\in\mathcal{F}}\sum_{(x,y)\in B\times A}\omega(x,y)\mu(A,B)\right)\\
&=\max_{\mu\in\Delta_{\mathcal{F}}}\min_{\omega\in\Delta_S}\sum_{(x,y)\in S}\omega(x,y)\frac{\mu\big(\{(A,B)\in\mathcal{F}:\ (x,y)\in A\times B\}\big)+\mu\big(\{(A,B)\in\mathcal{F}:\ (y,x)\in A\times B\}\big)}{2}\\
&=\max_{\mu\in\Delta_{\mathcal{F}}}\min_{(x,y)\in S}\frac{\mu\big(\{(A,B)\in\mathcal{F}:\ (x,y)\in A\times B\}\big)+\mu\big(\{(A,B)\in\mathcal{F}:\ (y,x)\in A\times B\}\big)}{2}.
\end{aligned}$$

Thus, there exists a probability measure μ on $\mathcal{F}$ that satisfies

$$\forall(x,y)\in S,\qquad \frac12\mu\big(\{(A,B)\in\mathcal{F}:\ (x,y)\in A\times B\}\big)+\frac12\mu\big(\{(A,B)\in\mathcal{F}:\ (y,x)\in A\times B\}\big)\geqslant \mathfrak{p}. \tag{245}$$

Suppose that $(\mathcal{A},\mathcal{B})$ is a random element of $\mathcal{F}$ that is distributed according to μ. Let ε be a standard $\{0,1\}$-valued Bernoulli random variable (thus, the probabilities that $\varepsilon=0$ and $\varepsilon=1$ are both equal $1/2$) that is independent of $(\mathcal{A},\mathcal{B})$. Denote the probability space on which $(\mathcal{A},\mathcal{B})$ and ε are defined by $(\Omega,\mathbb{P})$, and let $\mathcal{Z}$ be the random nonempty subset of X which is the following function of Ω: if $\varepsilon=0$, then $\mathcal{Z}=\mathcal{A}$ and if $\varepsilon=1$, then $\mathcal{Z}=\mathcal{B}$. This definition means that for every fixed $x,y\in X$ we have

$$\begin{aligned}
&\mathbb{P}\Big[x\in\mathcal{Z}\quad\text{and}\quad \min_{z\in\mathcal{Z}}\phi(y,z)\geqslant\psi(y)\Big]\\
&\qquad=\frac12\mathbb{P}\Big[x\in\mathcal{A}\quad\text{and}\quad\min_{z\in\mathcal{A}}\phi(y,z)\geqslant\psi(y)\Big]+\frac12\mathbb{P}\Big[x\in\mathcal{B}\quad\text{and}\quad\min_{z\in\mathcal{B}}\phi(y,z)\geqslant\psi(y)\Big].
\end{aligned}\tag{246}$$

As ϕ is a symmetric function, by the definition (242) of $\mathcal{F}$ the following inclusions of events hold:

$$\begin{cases}\{x\in\mathcal{A}\quad\text{and}\quad y\in\mathcal{B}\}\subseteq\big\{x\in\mathcal{A}\quad\text{and}\quad\min_{z\in\mathcal{A}}\phi(y,z)\geqslant\psi(y)\big\},\\ \{x\in\mathcal{B}\quad\text{and}\quad y\in\mathcal{A}\}\subseteq\big\{x\in\mathcal{B}\quad\text{and}\quad\min_{z\in\mathcal{B}}\phi(y,z)\geqslant\psi(y)\big\}.\end{cases}\tag{247}$$

Therefore,

$$\begin{aligned}
\mathbb{P}\Big[x\in\mathcal{Z}\quad\text{and}\quad\min_{z\in\mathcal{Z}}\phi(y,z)\geqslant\psi(y)\Big]&\overset{(246)\wedge(247)}{\geqslant}\frac12\mathbb{P}\big[x\in\mathcal{A}\quad\text{and}\quad y\in\mathcal{B}\big]+\frac12\mathbb{P}\big[x\in\mathcal{B}\quad\text{and}\quad y\in\mathcal{A}\big]\\
&=\frac12\mu\big(\{(A,B)\in\mathcal{F}:\ (x,y)\in A\times B\}\big)+\frac12\mu\big(\{(A,B)\in\mathcal{F}:\ (y,x)\in A\times B\}\big).
\end{aligned}\tag{248}$$

The desired probabilistic estimate (244) now follows by combining (245) and (248). □

We can now explain how to deduce Theorem 3 from Theorem 39; as Theorem 39 was already proven, this will complete the proof of Theorem 3. The reasoning below will freely use the notations and assumptions from the statements of Theorem 39.

Proof of Theorem 3. We will use the notations and assumptions of Theorem 3 and Theorem 39, with the following choice of universal constant α in the definition $\beta=s^{\alpha/\varepsilon}$ in Theorem 3, which uses the universal constant $\kappa>1$ from Proposition 36; this is consistent with requirement (62) of Theorem 39, as $\alpha\geqslant\alpha_0$:

$$\alpha\overset{\text{def}}{=}\max\Big\{2e\sqrt{2\kappa},\alpha_0\Big\}.\tag{249}$$

Our goal is to apply Lemma 61 with $X=\mathcal{M}$, $\phi=d_{\mathcal{M}}$ and $S=\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d_{\mathcal{M}}(x,y)\geqslant\tau\}$; note that the conlusion of Theorem 3 is vacuous if $\operatorname{diam}(\mathcal{M})<\tau$, so we may assume from now that $S\neq\emptyset$. We will also choose $\psi=\beta\tau/\rho$, where ρ is from (64). With these choices, the set $\mathcal{F}$ in (242) becomes

$$\mathcal{F}=\left\{(A,B)\in\big(2^{\mathcal{M}}\smallsetminus\{\emptyset\}\big)\times\big(2^{\mathcal{M}}\smallsetminus\{\emptyset\}\big):\ \forall(x,y)\in A\times B,\quad d_{\mathcal{M}}(x,y)\geqslant\frac{\beta\tau}{\min\{\rho(x),\rho(y)\}}\right\}.\tag{250}$$

Fix $C \geqslant 1$ and a probability measure ω on $\mathcal{M}\times\mathcal{M}$ whose support is contained in S, which by the definition of S means that every $x,y\in\mathcal{M}$ with $\omega(x,y)>0$ satisfy $d_{\mathcal{M}}(x,y)\geqslant\tau$. We can thus apply Theorem 39 to the symmetrization of ω, i.e., to the probability measure that assigns the mass $(\omega(x,y)+\omega(y,x))/2$ to each $(x,y)\in\mathcal{M}\times\mathcal{M}$, to get for every $v\in\mathbb{R}^n$ nonempty subsets $A^*(v),B^*(v)$ of $\mathcal{M}$. Let $\mu_{\omega,C}$ be the law of the pair $(A^*(v),B^*(v))$ as $v\in\mathbb{R}^n$ is distributed according to γ_n. By (250) and (63) we know that $(A^*(v),B^*(v))\in\mathcal{F}$ for all $v\in\mathbb{R}^n$, i.e., $\mu_{\omega,C}$ is supported on $\mathcal{F}$. By (65) applied to the symmetrization of ω,

$$\int_{\mathcal{F}}\frac{\omega(A\times B)+\omega(B\times A)}{2}\,\mathrm{d}\mu_{\omega,C}(A,B)\gtrsim e^{-\kappa C^2}. \tag{251}$$

We are thus in position to use Lemma 61 to get a probability measure $\mathbb{P}^C$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ that satisfies

$$\forall (x,y)\in S,\qquad \mathbb{P}^C\big[\emptyset\neq\mathcal{Z}\subseteq\mathcal{M}:\ d_{\mathcal{M}}(x,\mathcal{Z})\geqslant\psi(y)\quad\text{and}\quad y\in\mathcal{Z}\big]\gtrsim e^{-\kappa C^2}.$$

Recalling our choice of S and ψ, as well as the notation for ρ in (64), this conclusion coincides with the requirement that every $x,y\in\mathcal{M}$ with $d_{\mathcal{M}}(x,y)\geqslant\tau$ satisfy

$$\mathbb{P}^C\left[\emptyset\neq\mathcal{Z}\subseteq\mathcal{M}:\ d_{\mathcal{M}}(x,\mathcal{Z})\geqslant\frac{\beta\tau}{1+\frac{1}{\alpha C}\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}}\quad\text{and}\quad y\in\mathcal{Z}\right]\gtrsim e^{-\kappa C^2}. \tag{252}$$

Observe that the special case $C=O(1)$ of (252) already implies Theorem 1. To deduce Theorem 3, let $\mathcal{Z}$ be the following random nonempty subset of $\mathcal{M}$: First choose $k\in\mathbb{N}$ with probability 2^{-k}, and then select a random subset of $\mathcal{M}$ according to $\mathbb{P}^C$ for $C=e^{k-1}$. Denote the law of this random subset $\mathcal{Z}$ by $\mathbb{P}$.

Fix $x,y\in\mathcal{M}$ with $d_{\mathcal{M}}(x,y)\geqslant\tau$. Suppose first that

$$\frac{\alpha}{\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}}\leqslant\lambda\leqslant\frac12. \tag{253}$$

As $1/\lambda-1>0$,

$$\frac{\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}}{\alpha\left(\frac{1}{\lambda}-1\right)}\geqslant\frac{\lambda}{\alpha}\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}\overset{(253)}{\geqslant}1.$$

Hence, there exists $k\in\mathbb{N}$ such that

$$\frac{\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}}{\alpha\left(\frac{1}{\lambda}-1\right)}\leqslant e^{k-1}<\frac{e\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}}{\alpha\left(\frac{1}{\lambda}-1\right)}\leqslant\frac{2e\lambda}{\alpha}\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}, \tag{254}$$

where the third inequality in (254) is valid because $0<\lambda\leqslant 1/2$, so $1/\lambda-1\geqslant 1/(2\lambda)$. Now,

$$\begin{aligned}\mathbb{P}\big[d_{\mathcal{M}}(x,\mathcal{Z})\geqslant\lambda\beta\tau\quad\text{and}\quad y\in\mathcal{Z}\big]&\geqslant 2^{-k}\mathbb{P}^{e^{k-1}}\big[d_{\mathcal{M}}(x,\mathcal{Z})\geqslant\lambda\beta\tau\quad\text{and}\quad y\in\mathcal{Z}\big]\\ &\geqslant 2^{-k}\mathbb{P}^{e^{k-1}}\left[d_{\mathcal{M}}(x,\mathcal{Z})\geqslant\frac{\beta\tau}{1+\frac{1}{\alpha e^{k-1}}\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(y,\beta\tau))}}}\quad\text{and}\quad y\in\mathcal{Z}\right]\gtrsim e^{-\left(k\log 2+\kappa e^{2k-2}\right)},\end{aligned} \tag{255}$$

where the first step of (255) is a direct consequence of the definition of $\mathbb{P}$, the second step of (255) is valid by the first inequality in (254), which can be rewritten as

$$\lambda\leqslant\frac{1}{1+\frac{1}{\alpha e^{k-1}}\sqrt{\log\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}},$$

and the third step of (255) is an application of (252) with $C=e^{k-1}\geqslant 1$. Next, because $\kappa,k\geqslant 1$, it is straightforward to check that $k\log 2+\kappa e^{2k-2}\leqslant 2\kappa e^{2k-2}$. Consequently,

$$\mathbb{P}\big[d_{\mathcal{M}}(x,\mathcal{Z})\geqslant\lambda\beta\tau\quad\text{and}\quad y\in\mathcal{Z}\big]\overset{(255)}{\gtrsim}e^{-2\kappa e^{2k-2}}\overset{(254)}{\geqslant}\left(\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}\right)^{-\frac{8\kappa e^2}{\alpha^2}\lambda^2}\geqslant\left(\frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}\right)^{-\lambda^2},$$

where the last step holds as $\alpha \geqslant 2e\sqrt{2\kappa}$ by (249). This proves (4) for λ satisfying (253). The validity of (4) in the entire range $0 < \lambda \leqslant 1/2$ follows formally from its validity when λ belongs to the restricted range (253): if $0 < \lambda < \alpha/\sqrt{\log(\mu(B_{\mathcal{M}}(x,19\beta\tau))/\mu(B_{\mathcal{M}}(x,\beta\tau))}$, then by using the endpoint case of what we just proved,

$$\mathbb{P}\left[d_{\mathcal{M}}(x,\mathscr{Z}) \geqslant \lambda\beta\tau \quad \text{and} \quad y \in \mathscr{Z}\right] \geqslant \mathbb{P}\left[d_{\mathcal{M}}(y,\mathscr{Z}) \geqslant \frac{\alpha}{\sqrt{\log \frac{\mu(B_{\mathcal{M}}(x,19\beta\tau))}{\mu(B_{\mathcal{M}}(x,\beta\tau))}}}\beta\tau \quad \text{and} \quad x \in \mathscr{Z}\right] \gtrsim e^{-\alpha^2} \asymp 1. \qquad \square$$

9. From a random zero set to a "no measure is concentrated phenomenon"

Our goal here is to show how Theorem 7 and Theorem 8 follow from Theorem 6. We will, in fact, obtain the following more precise result, from which it is quick to deduce Theorem 7 and Theorem 8:

Theorem 62. *For every $0 < s,\varepsilon < 1$ there exists $c = c(s,\varepsilon) > 0$ with the following property. Given $n \in \mathbb{N}$, let $(\mathcal{M}, d)$ be a 5^n-doubling (s,ε)-quasisymmetrically Hilbertian metric. Suppose that μ is a Borel probability measure on $\mathcal{M}$ and fix $0 < \mathfrak{p} \leqslant 1$ and $\mathfrak{d} > 0$ for which*

$$(\mu \times \mu)\big(\{(x,y) \in \mathcal{M} \times \mathcal{M} :\ d(x,y) \geqslant \mathfrak{d}\}\big) \geqslant \mathfrak{p}. \tag{256}$$

Then, for every $\sqrt{\kappa/n} \leqslant \phi \leqslant 1$, where $\kappa > 1$ is the universal constant from Theorem 6 , we have

$$I_\mu^{\mathcal{M}}(c\phi\mathfrak{d}) \geqslant \mathfrak{p}e^{-n\phi^2}. \tag{257}$$

Prior to proving Theorem 62, we will next assume its validity and explain how it implies Theorem 7 and Theorem 8. If the assumptions and notation of Theorem 62 hold, then

$$\forall 0 \leqslant \phi \leqslant 1, \qquad I_\mu^{\mathcal{M}}(c\phi\mathfrak{d}) \gtrsim \mathfrak{p}e^{-n\phi^2}. \tag{258}$$

Indeed, conclusion (257) of Theorem 62 implies (258) if $\sqrt{\kappa/n} \leqslant \phi \leqslant 1$ (with the implicit constant in (258) equal to 1). If $0 < \phi \leqslant \sqrt{\kappa/n}$, then since per (10) the isoperimetric function is non-increasing,

$$I_\mu^{\mathcal{M}}(c\phi\mathfrak{d}) \geqslant I_\mu^{\mathcal{M}}\Big(c\sqrt{\frac{\kappa}{n}}\mathfrak{d}\Big) \overset{(257)}{\geqslant} \mathfrak{p}e^{-\kappa} \gtrsim_\kappa \mathfrak{p}e^{-n\phi^2}.$$

The normalization assumption of Theorem 7 and Theorem 8 is that (256) holds for $\mathfrak{d} = 1$ and $\mathfrak{p} = \frac{1}{2}$, as we say that μ is normalized if the median of $d(x,y)$ when $(x,y) \in \mathcal{M} \times \mathcal{M}$ is distributed according to $\mu \times \mu$ is equal to 1. Thus, Theorem 8 is a special case of (258).

To prove (a similar strengthening of) Theorem 7, suppose first that $\mathfrak{p}/(2e^n) \leqslant \theta \leqslant \min\{\mathfrak{p}/(2e^\kappa), \mathfrak{p}^2/4\}$, where we are continuing to reason under the assumptions and notation of Theorem 62. Define

$$\phi_0 = \phi_0(\mathfrak{p},\theta,n) \overset{\mathrm{def}}{=} \sqrt{\frac{\log\frac{\mathfrak{p}}{2\theta}}{n}} \asymp \sqrt{\frac{\log\frac{1}{\theta}}{n}}, \tag{259}$$

where the last equivalence in (259) holds as $\theta \leqslant \mathfrak{p}^2/4$. The rest of the above assumptions on θ are equivalent to $\sqrt{\kappa/n} \leqslant \phi_0 \leqslant 1$, so (257) holds, i.e., $I_\mu^{\mathcal{M}}(c\phi_0\mathfrak{d}) \geqslant 2\theta > \theta$. Therefore, by the first implication in (11),

$$\forall \frac{\mathfrak{p}}{2e^n} \leqslant \theta \leqslant \min\Big\{\frac{\mathfrak{p}}{2e^\kappa}, \frac{\mathfrak{p}^2}{4}\Big\}, \qquad \mathrm{ObsDiam}_\mu^{\mathcal{M}}(\theta) \geqslant c\phi_0\mathfrak{d} \overset{(259)}{\asymp} c\mathfrak{d}\sqrt{\frac{\log\frac{1}{\theta}}{n}}. \tag{260}$$

However, the following estimate on the observable diameter of the Euclidean sphere is a standard consequence of Lévy's spherical isoperimetric theorem [Lév51] (see e.g. equation (1.21) in [Led01]):

$$\forall \theta > 0, \qquad \mathrm{ObsDiam}_{\sigma^{n-1}}^{S^{n-1}}(\theta) \lesssim \sqrt{\frac{\log\frac{1}{\theta}}{n}}. \tag{261}$$

We therefore conclude from (260) and (261) that

$$\forall \frac{\mathfrak{p}}{2e^n} \leqslant \theta \leqslant \min\Big\{\frac{\mathfrak{p}}{2e^\kappa}, \frac{\mathfrak{p}^2}{4}\Big\}, \qquad \frac{1}{\mathfrak{d}}\mathrm{ObsDiam}_\mu^{\mathcal{M}}(\theta) \gtrsim_{s,\varepsilon} \mathrm{ObsDiam}_{\sigma^{n-1}}^{S^{n-1}}(\theta).$$

If $0<\theta\leqslant \mathfrak{p}/(2e^n)$, then because the definition of the θ-observable diameter (of any metric probability space) immediately implies that it is non-increasing in θ, the special case $\theta=\mathfrak{p}/(2e^n)$ of (260) gives

$$\frac{1}{\mathfrak{d}}\mathrm{ObsDiam}_\mu^{\mathcal{M}}(\theta)\geqslant \frac{1}{\mathfrak{d}}\mathrm{ObsDiam}_\mu^{\mathcal{M}}\Big(\frac{\mathfrak{p}}{2e^n}\Big)\overset{(260)}{\gtrsim}_{s,\varepsilon} 1\asymp \mathrm{diam}_{\ell_2^n}\big(S^{n-1}\big)\gtrsim \mathrm{ObsDiam}_{\sigma^{n-1}}^{S^{n-1}}(\theta).$$

Altogether, we have verified that the assumptions of Theorem 62 imply that

$$\forall 0<\theta\leqslant \min\Big\{\frac{\mathfrak{p}}{2e^\kappa},\frac{\mathfrak{p}^2}{4}\Big\},\qquad \frac{1}{\mathfrak{d}}\mathrm{ObsDiam}_\mu^{\mathcal{M}}(\theta)\gtrsim_{s,\varepsilon}\mathrm{ObsDiam}_{\sigma^{n-1}}^{S^{n-1}}(\theta). \tag{262}$$

Theorem 7 is the special case $\mathfrak{p}=\frac12$ and $\mathfrak{d}=1$ of (262), so we get (9) with $\theta_0=\min\{1/(4e^\kappa),1/16\}$.

Note that [NRS05] proved the (suboptimal, dimension-dependent) estimate (41) under the assumption that $\mathrm{diam}(\mathcal{M})<\infty$ and the following expectation bound holds for some $\alpha>0$:

$$\iint_{\mathcal{M}\times\mathcal{M}} d(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\geqslant \alpha\mathrm{diam}(\mathcal{M}). \tag{263}$$

This setting is covered by the above derivation of (262) assuming the probability bound (256) because

$$\alpha\mathrm{diam}(\mathcal{M})\overset{(263)}{\leqslant}(\mu\times\mu)\Big(\big\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant\frac{\alpha}{2}\mathrm{diam}(\mathcal{M})\big\}\Big)\mathrm{diam}(\mathcal{M})+\frac{\alpha}{2}\mathrm{diam}(\mathcal{M}).$$

So, (256) holds with $\mathfrak{p}=\alpha/2$ and $\mathfrak{d}=\alpha\mathrm{diam}(\mathcal{M})/2$, in which case (262) becomes

$$\forall 0<\theta\leqslant \min\Big\{\frac{\alpha}{4e^\kappa},\frac{\alpha^2}{16}\Big\},\qquad \frac{\mathrm{ObsDiam}_\mu^{\mathcal{M}}(\theta)}{\mathrm{diam}(\mathcal{M})}\gtrsim_{s,\varepsilon}\alpha\mathrm{ObsDiam}_{\sigma^{n-1}}^{S^{n-1}}(\theta),$$

which is the form of the conclusion of [NRS05, Theorem 1.7], though now it is dimension-independent.

Passing to the proof of Theorem 62, we will first record Lemma 63 below. It shows that the existence of large well-separated subsets yields a lower bound on the isoperimetric function; a proof of this simple fact is implicit in the proof of [Shi16, Proposition 2.26]. Recall that, as in (10), the Borel subsets of a metric space $(\mathcal{M},d)$ are denoted $\mathscr{Bor}(\mathcal{M},d)$. The distance between $\emptyset\neq\mathcal{A},\mathcal{B}\subseteq\mathcal{M}$ is denoted (as usual) by

$$d(\mathcal{A},\mathcal{B})\stackrel{\mathrm{def}}{=}\inf\big\{d(a,b):\ (a,b)\in\mathcal{A}\times\mathcal{B}\big\}.$$

Lemma 63. *For every metric space $(\mathcal{M},d)$, every Borel probability measure μ on $\mathcal{M}$ satisifes*

$$\forall t>0,\qquad I_\mu^{\mathcal{M}}(t)\geqslant \sup_{\substack{\mathcal{A},\mathcal{B}\in\mathscr{Bor}(\mathcal{M},d)\smallsetminus\{\emptyset\}\\ d(\mathcal{A},\mathcal{B})\geqslant 2t}}\min\big\{\mu(\mathcal{A}),\mu(\mathcal{B})\big\}.$$

Proof. Fix $t>0$. Let $\mathcal{A},\mathcal{B}$ be nonempty Borel subsets of $\mathcal{M}$ with $d(\mathcal{A},\mathcal{B})\geqslant 2t$. By the triangle inequality, $2t\leqslant d(\mathcal{A},\mathcal{B})\leqslant d(x,\mathcal{A})+d(x,\mathcal{B})\leqslant 2\max\{d(x,\mathcal{A}),d(x,\mathcal{B})\}$ for every $x\in\mathcal{M}$. Thus, $\max\{d(x,\mathcal{A}),d(x,\mathcal{B})\}\geqslant t$ for every $x\in\mathcal{M}$, i.e., $\mathcal{M}=\{x\in\mathcal{M}:\ d(x,\mathcal{A})\geqslant t\}\cup\{x\in\mathcal{M}:\ d(x,\mathcal{B})\geqslant t\}$. As μ is a probability measure, this implies that $\max\{\mu(\{x\in\mathcal{M}:\ d(x,\mathcal{A})\geqslant t\}),\mu(\{x\in\mathcal{M}:\ d(x,\mathcal{B})\geqslant t\})\}\geqslant\frac12$. If $\mu(\{x\in\mathcal{M}:\ d(x,\mathcal{A})\geqslant t\})\geqslant\frac12$, then denote $\mathcal{C}=\{x\in\mathcal{M}:\ d(x,\mathcal{A})\geqslant t\}$ and observe that (by design) we have $\{x\in\mathcal{M}:\ d(x,\mathcal{C})\geqslant t\}\supseteq\mathcal{A}$. So,

$$I_\mu^{\mathcal{M}}(t)\overset{(10)}{\geqslant}\mu\big(\{x\in\mathcal{M}:\ d(x,\mathcal{C})\geqslant t\}\big)\geqslant\mu(\mathcal{A}).$$

The symmetric reasoning gives $I_\mu^{\mathcal{M}}(t)\geqslant\mu(\mathcal{B})$ in the remaining case $\mu(\{x\in\mathcal{M}:\ d(x,\mathcal{B})\geqslant t\})\geqslant\frac12$. □

The following lemma shows that a spreading random zero set is an obstruction to isoperimetry:

Lemma 64. *If a finite metric space $(\mathcal{M},d)$ admits a random zero set which is ζ-spreading with probability δ (per Definition 5) for some $\zeta>0$ and $0<\delta\leqslant 1$, then every probability measure μ on $\mathcal{M}$ satisfies*

$$\forall t>0,\qquad I_\mu^{\mathcal{M}}(t)\geqslant\delta(\mu\times\mu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant 2t\zeta\}\big). \tag{264}$$

Proof. Fix $t>0$ and let $\mathbb{P}^{2t\zeta}$ be as in Definition 5 with $\tau=2t\zeta$, i.e.,

$$\forall x,y\in\mathcal{M},\qquad d(x,y)\geqslant 2t\zeta \implies \mathbb{P}^{2t\zeta}\left[\emptyset\neq\mathscr{Z}\subseteq\mathcal{M}:\ d(y,\mathscr{Z})\geqslant 2t \quad\text{and}\quad x\in\mathscr{Z}\right]\geqslant\delta. \tag{265}$$

Then,

$$\begin{aligned}
&\int_{2^{\mathcal{M}}\setminus\{\emptyset\}}\mu\big(\{y\in\mathcal{M}:\ d(y,\mathscr{Z})\geqslant 2t\}\big)\mu(\mathscr{Z})\,\mathbb{P}^{2t\zeta}(\mathscr{Z})\\
&\qquad=\iiint_{(2^{\mathcal{M}}\setminus\{\emptyset\})\times\mathcal{M}\times\mathcal{M}}\mathbf{1}_{\{d(y,\mathscr{Z})\geqslant 2t\ \text{ and }\ x\in\mathscr{Z}\}}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\,\mathrm{d}\mathbb{P}^{2t\zeta}(\mathscr{Z})\\
&\qquad=\iint_{\mathcal{M}\times\mathcal{M}}\mathbb{P}^{2t\zeta}\left[\emptyset\neq\mathscr{Z}\subseteq\mathcal{M}:\ d(y,\mathscr{Z})\geqslant 2t \quad\text{and}\quad x\in\mathscr{Z}\right]\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\\
&\qquad\geqslant\iint_{\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant 2t\zeta\}}\mathbb{P}^{2t\zeta}\left[\emptyset\neq\mathscr{Z}\subseteq\mathcal{M}:\ d(y,\mathscr{Z})\geqslant 2t \quad\text{and}\quad x\in\mathscr{Z}\right]\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\\
&\qquad\overset{(265)}{\geqslant}\delta(\mu\times\mu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant 2t\zeta\}\big).
\end{aligned}$$

Therefore, there exists $\emptyset\neq\mathscr{Z}=\mathscr{Z}_\mu\subseteq\mathcal{M}$ such that

$$\begin{aligned}
\min\big\{\mu\big(\{y\in\mathcal{M}:\ d(y,\mathscr{Z})\geqslant 2t\}\big),\mu(\mathscr{Z})\big\}&\geqslant\mu\big(\{y\in\mathcal{M}:\ d(y,\mathscr{Z})\geqslant 2t\}\big)\mu(\mathscr{Z})\\
&\geqslant\delta(\mu\times\mu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant 2t\zeta\}\big).
\end{aligned}\tag{266}$$

The desired estimate (264) holds vacuously if its right hand side vanishes, so we may assume that $(\mu\times\mu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant 2t\zeta\}\big)>0$, whence $\{y\in\mathcal{M}:\ d(y,\mathscr{Z})\geqslant 2t\},\mathscr{Z}\neq\emptyset$ by (266). Furthermore, $d(\{y\in\mathcal{M}:\ d(y,\mathscr{Z})\geqslant 2t\},\mathscr{Z})\geqslant 2t$, so (264) follows by combining Lemma 63 with (266). □

Lemma 65 below is an approximation step for passing from Lemma 64 to the analogous statement for infinite spaces; we include its proof for completeness as we did not locate it in the literature, though it is straightforward and one could take it an easy exercise rather than reading it.

Lemma 65. *Let $(\mathcal{M},d)$ be a metric space in which every ball is totally bounded, and let μ a Borel probability measure on $\mathcal{M}$. For any $0<\sigma<1$ and $t,D>0$ there exists a finitely supported probability measure ν on $\mathcal{M}$ such that*

$$(\mu\times\mu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)>D\}\big)\leqslant(\nu\times\nu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)>D\}\big)+\sigma, \tag{267}$$

and,

$$I_\mu^{\mathcal{M}}(t)\geqslant I_\nu^{\mathcal{M}}(t+\sigma)-\sigma. \tag{268}$$

Prior to justifying Lemma 65, we will show how it quickly implies Theorem 62:

Proof of Theorem 62 assuming Lemma 65. We may assume without loss of generality that $0<s,\varepsilon\leqslant 1/2$. Since balls in doubling metric spaces are totally bounded, by Lemma 65 it suffices to prove Theorem 62 when $(\mathcal{M},d)$ is a finite metric space. Because $\phi\geqslant\sqrt{\kappa/n}$, we may apply Theorem 6 with $p=n\phi^2/\kappa\geqslant 1$ and $K=5^n$. We have $p\leqslant\log K$ as $\kappa>1\geqslant\phi$, so $(\mathcal{M},d)$ has a random zero set that is ζ-spreading with probability δ, where $\delta=e^{-\kappa p}=e^{-n\phi^2}$ and $\zeta=s^{-\kappa/\varepsilon}\sqrt{(n\log 5)/p}$. Next, apply Lemma 64 with $t=c\phi\mathfrak{d}$, where $c=c(s,\varepsilon)=s^{\kappa/\varepsilon}/(2\sqrt{\kappa\log 5})$. For the above choices of parameters we have $2t\zeta=\mathfrak{d}$, so (264) implies the desired estimate (257) by (256). □

Proof of Lemma 65. Fix any $x_0\in\mathcal{M}$. As $\lim_{R\to\infty}\mu\big(B(x_0,R)\big)=\mu(\mathcal{M})=1$, we can also fix $R>t$ such that

$$\mu\big(B(x_0,R-t)\big)\geqslant 1-\frac{1}{2}\sigma. \tag{269}$$

Finally, as $\lim_{\omega\to 0^+}(\mu\times\mu)(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant D+3\omega\})=(\mu\times\mu)(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)>D\})$ we may fix $0<\omega\leqslant\sigma/2$ for which

$$(\mu\times\mu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)>D\}\big)\leqslant(\mu\times\mu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant D+3\omega\}\big)+\frac{1}{4}\sigma^2. \tag{270}$$

Observe in passing that the above choices imply in particular the following estimate:

$$\begin{aligned}(\mu\times\mu)&\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)>D\}\big)\\ &\stackrel{(270)}{\leqslant}(\mu\times\mu)\big(\{(x,y)\in B(x_0,R)\times B(x_0,R):\ d(x,y)\geqslant D+3\omega\}\big)+\frac{1}{4}\sigma^2\\ &\qquad+(\mu\times\mu)\Big((\mathcal{M}\times\mathcal{M})\smallsetminus\big(B(x_0,R)\times B(x_0,R)\big)\Big)\\ &\stackrel{(269)}{\leqslant}(\mu\times\mu)\big(\{(x,y)\in B(x_0,R)\times B(x_0,R):\ d(x,y)\geqslant D+3\omega\}\big)+\sigma.\end{aligned}\tag{271}$$

As every ball in $\mathcal{M}$ is assumed to be totally bounded, there are $n\in\mathbb{N}$ and $x_1,\ldots,x_n\in B(x_0,R)$ such that

$$B(x_0,R)\subseteq\bigcup_{i=1}^{n}B(x_i,\omega).\tag{272}$$

We will next consider the Voronoi tessellation $\mathcal{V}_1,\ldots,\mathcal{V}_n\subseteq B(x_0,R)$ of $B(x_0,R)$ that is induced by the (ordered) ω-dense subset $\{x_1,\ldots,x_n\}$ of $B(x_0,R)$, which is defined (as usual) by setting

$$\mathcal{V}_1\stackrel{\mathrm{def}}{=}\big\{x\in B(x_0,R):\ d(x,x_1)=d(x,\{x_1,\ldots,x_n\})\big\},$$

and then proceeding inductively by setting

$$\forall\, i\in\{1,\ldots,n-1\},\qquad \mathcal{V}_{i+1}\stackrel{\mathrm{def}}{=}\big\{x\in B(x_0,R):\ d(x,x_{i+1})=d(x,\{x_1,\ldots,x_n\})\big\}\smallsetminus\bigcup_{j=1}^{i}\mathcal{V}_j.$$

Thus, $d(x,x_i)\leqslant\omega$ for every $i\in[n]$ and every $x\in\mathcal{V}_i$, thanks to (272), and because, by design, $\{\mathcal{V}_i\}_{i\in[n]}$ is a (Borel) partition of $B(x_0,R)$, the following defines a finitely supported probability measure on $\mathcal{M}$:

$$\nu\stackrel{\mathrm{def}}{=}\sum_{i=1}^{n}\frac{\mu(\mathcal{V}_i)}{\mu\big(B(x_0,R)\big)}\boldsymbol{\delta}_{x_i}.\tag{273}$$

To check (267), denote

$$J\stackrel{\mathrm{def}}{=}\Big\{(i,j)\in[n]\times[n]:\ d(x_i,x_j)\geqslant D+\omega\Big\},\tag{274}$$

so that

$$\begin{aligned}(\nu\times\nu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)>D\}\big)&\geqslant(\nu\times\nu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)\geqslant D+\omega\}\big)\\ &\stackrel{(273)\wedge(274)}{=}\sum_{(i,j)\in J}\frac{\mu(\mathcal{V}_i)\mu(\mathcal{V}_j)}{\mu\big(B(x_0,R)\big)^2}\geqslant\mu\Big(\bigcup_{(i,j)\in J}\mathcal{V}_i\times\mathcal{V}_j\Big),\end{aligned}\tag{275}$$

where the last step of (275) is valid as $\{\mathcal{V}_i\times\mathcal{V}_j\}_{(i,j)\in[n]\times[n]}$ are pairwise disjoint. Also, we will show that the following inclusion holds:

$$\big\{(x,y)\in B(x_0,R)\times B(x_0,R):\ d(x,y)\geqslant D+3\omega\big\}\subseteq\bigcup_{(i,j)\in J}\mathcal{V}_i\times\mathcal{V}_j,\tag{276}$$

from which we will get that

$$\begin{aligned}(\mu\times\mu)&\big(\{(x,y)\in B(x_0,R)\times B(x_0,R):\ d(x,y)\geqslant D+3\omega\}\big)\\ &\stackrel{(275)\wedge(276)}{\leqslant}(\nu\times\nu)\big(\{(x,y)\in\mathcal{M}\times\mathcal{M}:\ d(x,y)>D\}\big).\end{aligned}\tag{277}$$

To verify (276), if $x,y\in B(x_0,R)$, then fix $i,j\in[n]$ such that $(x,y)\in\mathcal{V}_i\times\mathcal{V}_j$, whence $d(x,x_i)\leqslant\omega$ and $d(y,x_j)\leqslant\omega$. If also $d(x,y)\geqslant D+3\omega$, then $d(x_i,x_j)\geqslant d(x,y)-d(x,x_i)-d(y,y_j)\geqslant D+\omega$, so $(i,j)\in J$ by (274), thus proving (276). The desired bound (267) now follows by substituting (277) into (271).

It remains to prove (268). As ν is supported on the finite set $\{x_1,\ldots,x_n\}$, we can fix $K\subseteq[n]$ such that

$$\nu\big(\{x_k\}_{k\in K}\big)\geqslant\frac{1}{2}\qquad\text{and}\qquad I_\nu^{\mathcal{M}}(t+\sigma)=\nu\big(\{x\in\mathcal{M}:\ d(x,\{x_k\}_{k\in K})\geqslant t+\sigma\}\big).\tag{278}$$

If we introduce the notation

$$\mathcal{C}_0 \stackrel{\mathrm{def}}{=} \bigcup_{k\in K} \mathcal{V}_k \qquad \text{and} \qquad \mathcal{S}_0 \stackrel{\mathrm{def}}{=} \bigcup_{\ell\in L} \mathcal{V}_\ell, \quad \text{where} \quad L \stackrel{\mathrm{def}}{=} \big\{\ell\in[n]:\ d(x_\ell,\{x_k\}_{k\in K})\geqslant t+\sigma\big\}, \tag{279}$$

then by the definition (273) of ν we see that (278) becomes

$$\mu(\mathcal{C}_0)\geqslant \frac12 \mu\big(B(x_0,R)\big) \qquad \text{and} \qquad \mu(\mathcal{S}_0)=I_\nu^{\mathcal{M}}(t+\sigma)\mu\big(B(x_0,R)\big). \tag{280}$$

Consider the Borel subsets $\mathcal{C},\mathcal{S}$ of $\mathcal{M}$ that are given by

$$\mathcal{C}\stackrel{\mathrm{def}}{=}\mathcal{C}_0\cup\big(\mathcal{M}\smallsetminus B(x_0,R)\big) \qquad \text{and} \qquad \mathcal{S}\stackrel{\mathrm{def}}{=}\mathcal{S}_0\cap B(x_0,R-t). \tag{281}$$

Observe that the following simple inclusion holds:

$$\mathcal{S}\subseteq\big\{x\in\mathcal{M}: d(x,\mathcal{C})\geqslant t\big\}. \tag{282}$$

Indeed, take $x\in\mathcal{S}$ and $y\in\mathcal{C}$. If $y\in\mathcal{M}\smallsetminus B(x_0,R)$, then $d(x,y)\geqslant d(y,x_0)-d(x,x_0)>t$ since $x\in B(x_0,R-t)$ by (281). Otherwise $y\in\mathcal{C}_0$ and $x\in\mathcal{S}_0$ by (281), so by (279) there are $\ell\in L$ and $k\in K$ such that $x\in\mathcal{V}_\ell$ and $y\in\mathcal{V}_k$, whence $d(x,x_\ell)\leqslant\omega$ and $d(y,x_k)\leqslant\omega$. Because $d(x_\ell,x_k)\geqslant t+\sigma$ by the definition (279) of L, it follows that $d(x,y)\geqslant d(x_\ell,x_k)-d(x,x_\ell)-d(y,x_k)\geqslant t+\sigma-2\omega\geqslant t$, since $\omega\leqslant\sigma/2$. This proves (282).

Because $\mathcal{C}_0\subseteq B(x_0,R)$, the two sets in the union that defines $\mathcal{C}$ in (279) are disjoint, so

$$\mu(\mathcal{C})=\mu(\mathcal{C}_0)+1-\mu\big(B(x_0,R)\big)\overset{(280)}{\geqslant}1-\frac12\mu\big(B(x_0,R)\big)\geqslant\frac12.$$

We may therefore apply the definition (10) of $I_\mu^{\mathcal{M}}(t)$ to $\mathcal{C}$ and conclude the proof of (268) as follows:

$$\begin{aligned} I_\mu^{\mathcal{M}}(t)&\geqslant\mu\big(\{x\in\mathcal{M}:d(x,\mathcal{C})\geqslant t\}\big)\overset{(282)}{\geqslant}\mu(\mathcal{S})\overset{(281)}{\geqslant}\mu(\mathcal{S}_0)-\mu\big(\mathcal{M}\smallsetminus B(x_0,R-t)\big)\\ &\overset{(280)}{=}I_\nu^{\mathcal{M}}(t+\sigma)+\big(\mu\big(B(x_0,R)\big)-1\big)I_\nu^{\mathcal{M}}(t+\sigma)-1+\mu\big(B(x_0,R-t)\big)\overset{(269)}{\geqslant}I_\nu^{\mathcal{M}}(t+\sigma)-\sigma. \qquad \square\end{aligned}$$

Remark 66. *Fix $n\in\mathbb{N}$ and a 4-regular graph $\mathsf{G}=(V,E)$ with $|V|=5^n$ such the second-largest eigenvalue of the transition matrix of its standard random walk is at most $1-\gamma$, where $\gamma>0$ is a universal constant; the survey* [HLW06] *discusses such expander graphs, including their existence. Let μ be the uniform probability measure on V and let $\mathfrak{m}>0$ be a median of $d_\mathsf{G}(x,y)$ with respect to $(x,y)\in V\times V$ distributed uniformly at random. Then $\mathfrak{m}\asymp n$ (one direction of this asymptotic equivalence follows from* [Chu89] *and the other direction is a consequence of a simple counting argument, as in e.g.* [Mat97]*). By* [GM83] *(its discrete counterpart which we are using here is treated in* [AM85, Alo86, BHT00]*), we have*

$$\forall\, t>0,\qquad I_\mu^{\left(V,\frac{1}{\mathfrak{m}}d_\mathsf{G}\right)}(t)\lesssim e^{-O(n)t}. \tag{283}$$

By e.g. [Led01, Proposition 1.12]*,* (283) *implies the following estimate on the observable diameter:*

$$\forall\, 0<\theta\leqslant\frac12,\qquad \mathrm{ObsDiam}_\mu^{\left(V,\frac{1}{\mathfrak{m}}d_\mathsf{G}\right)}(\theta)\lesssim\frac{\log\frac{1}{\theta}}{n}.$$

The metric space $(V,\mathfrak{m}^{-1}d_\mathsf{G})$ is trivially 5^n-doubling, so this demonstrates that some assumption must be imposed on metric a space so that conclusion (9) *of Theorem 7 will hold (with the implicit dependence on s,ε replaced by dependence on the aforementioned assumption). Finding "useful minimal assumptions" that are needed here is an interesting research direction (purposefully formulated somewhat vaguely).*

The main result of this section is the following theorem, which generalizes part of [KLMN05]. It describes a way to obtain a multi-scale stochastic mixture of distributions over random subsets of a metric space.

Theorem 67. *Let $(\mathcal{M},d)$ be a finite metric space and let μ be a nondegenerate measure on $\mathcal{M}$. Denote*

$$\Phi = \Phi_\mu \stackrel{\mathrm{def}}{=} \frac{\mu(\mathcal{M})}{\min_{x\in\mathcal{M}}\mu(x)}. \tag{284}$$

Suppose that for every $n\in\mathbb{Z}$ we are given a probability measure $\mathbb{P}^n$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$. Then, for every $a>b$ there exists a probability measure $\mathbb{P}$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ such that for every $x,y\in\mathcal{M}$, every $\ell,u\geqslant 0$, and every $n\in\mathbb{Z}$,[36]

$$\begin{aligned}&\mathbb{P}\left[d(x,\mathcal{Z})\geqslant \min\{2^{b+n-2},\ell\}\ \text{ and }\ d(y,\mathcal{Z})\leqslant u\right]\\ &\qquad\gtrsim \frac{\left\lfloor \log\frac{\mu(B(x,2^{a+n}))}{\mu(B(x,2^{b+n}))}\right\rfloor}{(a-b+1)\log\Phi}\,\mathbb{P}^n\left[d(x,\mathcal{Z})\geqslant \ell\ \text{ and }\ d(y,\mathcal{Z})\leqslant u\right].\end{aligned} \tag{285}$$

Prior to proving Theorem 67, we will use it to deduce the following theorem, which is a restatement of Theorem 17, except that the implicit dependence on the parameters in (26) is now stated explicitly:

Theorem 68. *Fix $\alpha\geqslant\beta>0$ and $0<\varepsilon,\delta,\theta\leqslant 1$. Let $(\mathcal{M},d)$ be a finite metric space such that for every $\tau>0$ there is a probability measure $\mathbb{P}^\tau$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ that satisfies the following estimate:*

$$\forall x,y\in\mathcal{M},\qquad \tau\leqslant d(x,y)\leqslant(1+\theta)\tau \implies \mathbb{P}^\tau\left[d(x,\mathcal{Z})\geqslant \frac{\varepsilon\tau}{\sqrt{\log\frac{e|B(x,\alpha\tau)|}{|B(x,\beta\tau)|}}}\quad\text{and}\quad y\in\mathcal{Z}\right]\geqslant\delta. \tag{286}$$

Then,

$$\mathsf{c}_2(\mathcal{M})\lesssim\frac{1}{\min\{\beta,\varepsilon\}}\left(\frac{1+\log\frac{\max\{\alpha,1\}}{\min\{\beta,1\}}}{\delta\theta}\right)^{\frac12}\sqrt{\log|\mathcal{M}|}. \tag{287}$$

Furthermore, the bound (287) *on the Euclidean distortion of $\mathcal{M}$ is obtained via the Fréchet embedding.*

Proof of Theorem 68 assuming Theorem 67. Since the assumption (286) of Theorem 68 becomes weaker if we increase α and decrease β, we may assume from now without loss of generality that $\alpha\geqslant 2$ and $\beta<1$.

For each $n\in\mathbb{Z}$, define as follows a probability measure $\mathbb{Q}^n$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$. Let τ be chosen uniformly at random from $\{2^n,(1+\theta)2^n,\ldots,(1+\theta)^{\lceil 1/\theta\rceil}2^n\}$, and then choose $\mathcal{Z}\subseteq\mathcal{M}$ according to $\mathbb{P}^\tau$. We will denote the law of the resulting random subset $\mathcal{Z}$ by $\mathbb{Q}^n$. If $x,y\in\mathcal{M}$ satisfy $2^n\leqslant d(x,y)\leqslant 2^{n+1}$, then with probability that is at least a positive universal constant multiple of θ we have $\tau\leqslant d(x,y)\leqslant(1+\theta)\tau$, which implies in particular that $2^{n-1}\leqslant\tau\leqslant 2^{n+1}$, as $2^n\leqslant d(x,y)\leqslant 2^{n+1}$ and $0<\theta\leqslant 1$, so we may use (286) to deduce that

$$\forall x,y\in\mathcal{M},\qquad 2^n\leqslant d(x,y)\leqslant 2^{n+1}\implies \mathbb{Q}^n\left[d(x,\mathcal{Z})\geqslant\frac{\varepsilon 2^{n-1}}{\sqrt{\log\frac{e|B(x,\alpha 2^{n+1})|}{|B(x,\beta 2^{n-1})|}}}\quad\text{and}\quad y\in\mathcal{Z}\right]\gtrsim\delta\theta. \tag{288}$$

Writing $2\alpha=2^a$ and $\beta/2=2^b$, so $a\geqslant 2$ and $b<-1$ and $1+\log(\alpha/\beta)\asymp a-b+1\asymp a-b$, (288) becomes

$$\forall x,y\in\mathcal{M},\qquad 2^n\leqslant d(x,y)\leqslant 2^{n+1}\implies \mathbb{Q}^n\left[d(x,\mathcal{Z})\geqslant\frac{\varepsilon 2^{n-1}}{\sqrt{\log\frac{e|B(x,2^{a+n})|}{|B(x,2^{b+n})|}}}\quad\text{and}\quad y\in\mathcal{Z}\right]\gtrsim\delta\theta. \tag{289}$$

Of course, by symmetry we also have

$$\forall x,y\in\mathcal{M},\qquad 2^n\leqslant d(x,y)\leqslant 2^{n+1}\implies \mathbb{Q}^n\left[d(y,\mathcal{Z})\geqslant\frac{\varepsilon 2^{n-1}}{\sqrt{\log\frac{e|B(y,2^{a+n})|}{|B(y,2^{b+n})|}}}\quad\text{and}\quad x\in\mathcal{Z}\right]\gtrsim\delta\theta. \tag{290}$$

[36] Note that the definition (284) of the aspect ratio Φ of μ implies that $\Phi\geqslant|\mathcal{M}|$. In particular, $\Phi\geqslant 2$ per our convention that all metric spaces are not singletons, so the denominator in the right hand side of (285) does not vanish.

Apply Theorem 67 to the measures $\{\mathbb{Q}^n\}_{n\in\mathbb{Z}}$ and with μ being the counting measure on $\mathcal{M}$ (thus, recalling (284), we have $\Phi=|\mathcal{M}|$ in this case) to get a probability measure $\mathbb{P}$ on $2^{\mathcal{M}}\smallsetminus\{\emptyset\}$ that satisfies the following inequality for every $x,y\in\mathcal{M}$, every $n\in\mathbb{Z}$, and every $\ell,u\geqslant 0$:

$$\begin{aligned}\mathbb{P}\big[d(x,\mathscr{Z})\geqslant \min\{2^{b+n-2},\ell\}\ \text{ and }\ d(y,\mathscr{Z})\leqslant u\big]&\\ \gtrsim \frac{\left\lfloor \log\frac{\mu(B(x,2^{a+n}))}{\mu(B(x,2^{b+n}))}\right\rfloor}{(a-b)\log|\mathcal{M}|}\mathbb{Q}^n\big[d(x,\mathscr{Z})\geqslant \ell\ \text{ and }\ d(y,\mathscr{Z})\leqslant u\big].&\end{aligned}\tag{291}$$

We will proceed to bound the distortion into $L_2(\mathbb{P})$ of the Fréchet embedding $\Phi_{(\mathcal{M},d)}$ that is given in (22).

Fixing distinct $x,y\in\mathcal{M}$, define $n=n(x,y)\in\mathbb{Z}$ by

$$n=n(x,y)\stackrel{\mathrm{def}}{=}\lfloor \log_2 d(x,y)\rfloor \iff 2^n\leqslant d(x,y)<2^{n+1}.\tag{292}$$

Apply (291) with the parameters $\ell=\varepsilon 2^{n-1}/\sqrt{\log\frac{e|B(x,2^{a+n})|}{|B(x,2^{b+n})|}}$ and $u=0$ to get, using (289), that

$$\mathbb{P}\left[d(x,\mathscr{Z})\geqslant \min\left\{2^{b-2},\frac{\varepsilon}{2\sqrt{\log\frac{e|B(x,2^{a+n})|}{|B(x,2^{b+n})|}}}\right\}2^n\ \text{ and }\ y\in\mathscr{Z}\right]\gtrsim \frac{\left\lfloor\log\frac{|B(x,2^{a+n})|}{|B(x,2^{b+n})|}\right\rfloor}{(a-b)\log|\mathcal{M}|}\delta\theta.\tag{293}$$

In the same vein, using (290) we also get the symmetric estimate

$$\mathbb{P}\left[d(y,\mathscr{Z})\geqslant \min\left\{2^{b-2},\frac{\varepsilon}{2\sqrt{\log\frac{e|B(y,2^{a+n})|}{|B(y,2^{b+n})|}}}\right\}2^n\ \text{ and }\ x\in\mathscr{Z}\right]\gtrsim \frac{\left\lfloor\log\frac{|B(y,2^{a+n})|}{|B(y,2^{b+n})|}\right\rfloor}{(a-b)\log|\mathcal{M}|}\delta\theta.$$

Next, observe that because $a\geqslant 2$ and $b<-1$,

$$B\big(x,2^{b+n}\big)\cup B\big(y,2^{b+n}\big)\subseteq B\big(x,2^{a+n}\big)\cap B\big(y,2^{a+n}\big)\qquad\text{and}\qquad B\big(x,2^{b+n}\big)\cap B\big(y,2^{b+n}\big)=\emptyset.\tag{294}$$

Indeed, if $z\in B(x,2^{b+n})$, then $d(z,y)\leqslant d(z,x)+d(x,y)\leqslant 2^{b+n}+d(x,y)<2^{b+n}+2^{n+1}<5\cdot 2^{n-1}<2^{a+n}$, where the third step uses (292), the penultimate step uses $b<-1$ and the final step uses $a\geqslant 2$. Thus, $B(x,2^{b+n})\subseteq B(y,2^{a+n})$. By symmetry also $B(y,2^{b+n})\subseteq B(x,2^{a+n})$, so the first assertion in (294) holds. For the second assertion of (294), if there were $w\in B(x,2^{b+n})\cap B(y,2^{b+n})$, then we get the contradiction $d(x,y)\leqslant d(x,w)+d(w,y)\leqslant 2^{b+1+n}<2^n\leqslant d(x,y)$, where the penultimate step uses $b<-1$ and the final step uses (292). Having verified that (294) is indeed satisfied, we deduce from it that

$$2\min\big\{|B(x,2^{b+n})|,|B(y,2^{b+n})|\big\}\leqslant |B(x,2^{b+n})|+|B(y,2^{b+n})|\leqslant \min\big\{|B\big(x,2^{a+n}\big)|,|B(y,2^{a+n})|\big\}.$$

By interchanging the roles of x and y, if necessary, we may therefore assume without loss of generality that $|B(x,2^{a+n})|/|B(x,2^{b+n})|\geqslant 2$, in which case (293) implies (recalling that $2^b=\beta/2$) that

$$\mathbb{P}\left[d(x,\mathscr{Z})\geqslant \frac18\min\left\{\beta,\frac{\varepsilon}{\sqrt{\log\frac{|B(x,2^{a+n})|}{|B(x,2^{b+n})|}}}\right\}2^n\ \text{ and }\ y\in\mathscr{Z}\right]\gtrsim \frac{\log\frac{|B(x,2^{a+n})|}{|B(x,2^{b+n})|}}{(a-b)\log|\mathcal{M}|}\delta\theta.\tag{295}$$

Consequently, the Fréchet embedding $\Phi_{(\mathcal{M},d)}$ yields the distortion bound (287) into $L_2(\mathbb{P})$ because

$$\begin{aligned}&\|\Phi_{(\mathcal{M},d)}(x)-\Phi_{(\mathcal{M},d)}(y)\|_{L_2(\mathbb{P})}=\|d(x,\mathscr{Z})-d(y,\mathscr{Z})\|_{L_2(\mathbb{P})}\\ &\stackrel{(295)}{\gtrsim}\left(\frac{\log\frac{|B(x,2^{a+n})|}{|B(x,2^{b+n})|}}{(a-b)\log|\mathcal{M}|}\delta\theta\right)^{\frac12}\min\left\{\beta,\frac{\varepsilon}{\sqrt{\log\frac{|B(x,2^{a+n})|}{|B(x,2^{b+n})|}}}\right\}2^n\stackrel{(292)}{\gtrsim}\min\{\varepsilon,\beta\}\left(\frac{\delta\theta}{(a-b)\log|\mathcal{M}|}\right)^{\frac12}d(x,y).\qquad\square\end{aligned}$$

In preparation for the proof of Theorem 67, we introduce the following notation for a metric space $(\mathcal{M},d)$ and a nondegenerate measure μ on $\mathcal{M}$:

$$\forall\, x\in\mathcal{M},\ \forall\, t\in\mathbb{R},\qquad \mathcal{k}(x,t)=\mathcal{k}_{\mu,d}(x,t)\stackrel{\mathrm{def}}{=}\max\Big\{k\in\mathbb{Z}:\ \mu\big(B(x,2^k)\big)\leqslant e^t\Big\}.\tag{296}$$

In other words, if we define

$$\forall x \in \mathcal{M},\ \forall \theta \in \mathbb{R}, \qquad s_\theta(x) = s_{\theta,\mu,d}(x) \stackrel{\mathrm{def}}{=} \log \mu\big(B(x,2^\theta)\big), \tag{297}$$

then for every $x \in \mathcal{M}$ the function $t \mapsto \hbar(x,t)$ is piecewise constant and given by

$$\forall x \in \mathcal{M},\ \forall i \in \mathbb{Z}, \qquad s_i(x) \leqslant t < s_{i+1}(x) \implies \hbar(x,t) = i. \tag{298}$$

We record for ease of later use the following simple variant of an observation from [KLMN05]:

Lemma 69. *For every $x \in \mathcal{M}$ and $t \in \mathbb{Z}$ we have*

$$\forall y \in B\big(x, 2^{\hbar(x,t)-1}\big), \qquad \hbar(y,t) \in \big\{\hbar(x,t)-1, \hbar(x,t), \hbar(x,t)+1\big\}. \tag{299}$$

Proof. Denote $k = \hbar(x,t)$. If $y \in B(x,2^{k-1})$, then $B(y,2^{k-1}) \subseteq B(x,2^k)$. Because $\mu(B(x,2^k)) \leqslant e^t$ by the definition (296) of $k = \hbar(x,t)$, it follows that $\mu(B(y,2^{k-1})) \leqslant e^t$, i.e., $k(y,t) \in \{k-1,k,\ldots\}$ by (296). Similarly, $B(x,2^{k+1}) \subseteq B(y,2^{k+1}+2^{k-1}) \subseteq B(y,2^{k+2})$ while $\mu(B(x,2^{k+1})) > e^t$ by the definition (296) of $k = \hbar(x,t)$, so $\mu(B(y,2^{k+2})) > e^t$, which implies by (296) that $\hbar(y,t) \in \{\ldots,k,k+1\}$. $\square$

We can now prove Theorem 67:

Proof of Theorem 67. Because Theorem 67 only discusses ratios of values of the measure μ, we may assume without loss of generality that μ is normalized so that $\min_{x\in\mathcal{M}} \mu(x) = 1$, in which case $\mu(\mathcal{M}) = \Phi$.

Fix $a > b$. Define two subsets I, T of $\mathbb{Z}$ by

$$I \stackrel{\mathrm{def}}{=} \big\{\lceil b\rceil, \ldots, \lceil a\rceil\big\} \qquad \text{and} \qquad T \stackrel{\mathrm{def}}{=} \big\{0, \ldots, \lceil \log \Phi\rceil - 1\big\}. \tag{300}$$

The ensuing construction involves auxiliary random variables which we will next introduce. Firstly, let i be a random variable that is distributed uniformly over I, and let t be a random variable that is distributed uniformly over T. Secondly, for every $i \in \mathbb{Z}$ let σ_i and η_i be random variables that are distributed uniformly over $\{0,1\}$ and $\{0,1,2\}$, respectively. Finally, for every $n \in \mathbb{Z}$ let $\mathscr{Z}_n$ be a random subset of $\mathcal{M}$ that is distributed according to $\mathbb{P}^n$. We also require that all of the above random variables/subsets, namely

$$\mathsf{i}, \mathsf{t}, \{\sigma_i\}_{i\in\mathbb{Z}}, \{\eta_i\}_{i\in\mathbb{Z}}, \{\mathscr{Z}_n\}_{n\in\mathbb{Z}}, \tag{301}$$

are independent, and we denote the (product) probability space on which they are defined by $(\Omega, \mathbb{P})$.

Consider the following random subset $\mathscr{Z}$ of $\mathcal{M}$, which is a function of the random variables in (301):

$$\mathscr{Z} \stackrel{\mathrm{def}}{=} \big\{z \in \mathcal{M} :\ z \in \mathscr{Z}_{\hbar(z,\mathsf{t})-\mathsf{i}+\eta_{\hbar(z,\mathsf{t})-\mathsf{i}}}\big\} \cup \big\{z \in \mathcal{M} :\ \sigma_{\hbar(z,\mathsf{t})-\mathsf{i}} = 1\big\}, \tag{302}$$

where we recall the notation (296). We will prove that for every $x, y \in \mathcal{M}$, every $\ell, u \geqslant 0$, and every $n \in \mathbb{Z}$,

$$\begin{aligned}\mathbb{P}\big[\mathscr{Z} \neq \emptyset \ \text{ and } \ d(x,\mathscr{Z}) \geqslant \min\{2^{b+n-2},\ell\} \ \text{ and } \ d(y,\mathscr{Z}) \leqslant u\big]& \\ \gtrsim \frac{\Big\lfloor \log \frac{\mu(B(x,2^{a+n}))}{\mu(B(x,2^{b+n}))}\Big\rfloor}{(a-b+1)\log\Phi}\,\mathbb{P}^n\big[d(x,\mathscr{Z}) \geqslant \ell \ \text{ and } \ d(y,\mathscr{Z}) \leqslant u\big].&\end{aligned} \tag{303}$$

This will imply the desired conclusion (285) of Theorem 67 by considering the random subset $\mathscr{Z}$ in (302) conditioned on the event $\{\mathscr{Z} \neq \emptyset\}$, i.e., the restriction of $\mathscr{Z}$ to the subset $\{\mathscr{Z} \neq \emptyset\}$ of Ω, equipped with the probability measure $\mathbb{P}[\mathscr{Z} \neq \emptyset]^{-1}\,\mathbb{P}$ (note that (303) implies in particular that $\mathbb{P}[\mathscr{Z} \neq \emptyset] > 0$).

Given $n \in \mathbb{Z}$, define two (cylinder) events $C_n \subseteq C_n^* \subseteq \Omega$ by

$$C_n^* \stackrel{\mathrm{def}}{=} \big\{(\eta_{n-2},\eta_{n-1},\eta_n) = (2,1,0)\big\} \quad \text{and} \quad C_n \stackrel{\mathrm{def}}{=} \big\{(\sigma_{n-2},\sigma_{n-1},\sigma_n,\eta_{n-2},\eta_{n-1},\eta_n) = (0,0,0,2,1,0)\big\}. \tag{304}$$

Thus,

$$\Pr[C_n^*] = \frac{1}{3^3} \asymp 1 \qquad \text{and} \qquad \Pr[C_n] = \frac{1}{2^3 3^3} \asymp 1. \tag{305}$$

Also, recalling the notation (296), for every $n \in \mathbb{Z}$ and $w \in \mathcal{M}$ define events $E_n(w), F_n(w), G(w) \subseteq \Omega$ by

$$E_n(w) \stackrel{\mathrm{def}}{=} \big\{w \in \mathscr{Z}_n\big\}, \qquad F_n(w) \stackrel{\mathrm{def}}{=} \big\{\hbar(w,\mathsf{t}) - \mathsf{i} \in \{n-2,n-1,n\}\big\}, \qquad G(w) \stackrel{\mathrm{def}}{=} \big\{\sigma_{\hbar(w,\mathsf{t})-\mathsf{i}} = 1\big\}. \tag{306}$$

Claim 70. *For every $n \in \mathbb{Z}$, $y \in \mathcal{M}$, and $u \geqslant 0$, the following inclusion of events holds:*

$$\{\mathscr{Z} \neq \emptyset \text{ and } d(y,\mathscr{Z}) \leqslant u\} \supseteq \bigg(\bigcup_{w\in B(y,u)} \Big(E_n(w) \cap \big(F_n(w) \cup G(w)\big)\Big)\bigg) \cap C_n^*. \tag{307}$$

Proof. If the event in the right-hand side of (307) occurs, then the event C_n^* occurs and, recalling the notations in (306), there exists $w \in \mathscr{Z}_n$ with $d(w,y) \leqslant u$ such that $\mathscr{k}(w,\mathrm{t})-\mathrm{i} \in \{n-2,n-1,n\}$ or $\sigma_{\mathscr{k}(w,\mathrm{t})-\mathrm{i}} = 1$. If $\mathscr{k}(w,\mathrm{t})-\mathrm{i} \in \{n-2,n-1,n\}$, then $\mathscr{k}(w,\mathrm{t})-\mathrm{i}+\eta_{\mathscr{k}(w,\mathrm{t})-\mathrm{i}} = n$ thanks to the assumed occurrence of the event C_n^* that is given in (304), so $w \in \mathscr{Z}$ by (302). On the other hand, if $\sigma_{\mathscr{k}(w,\mathrm{t})-\mathrm{i}} = 1$, then (302) implies directly that $w \in \mathscr{Z}$. In both cases $w \in \mathscr{Z}$, whence $\mathscr{Z} \neq \emptyset$ and $d(y,\mathscr{Z}) \leqslant d(y,w) \leqslant u$, as required. □

Next, recalling the notation (297), for every $n \in \mathbb{Z}$ and $x \in \mathcal{M}$ define an event $H_n(x) \subseteq \Omega$ by

$$H_n(x) \stackrel{\mathrm{def}}{=} \big\{s_{\mathrm{i}+n-1}(x) \leqslant \mathrm{t} < s_{\mathrm{i}+n}(x)\big\}. \tag{308}$$

Claim 71. *For every $m \in \mathbb{Z}$ and $x \in \mathcal{M}$ we have*

$$\mathbb{P}[H_m(x)] \gtrsim \frac{\Big\lfloor \log \frac{\mu(B(x,2^{a+m}))}{\mu(B(x,2^{b+m}))} \Big\rfloor}{(a-b+1)\log\Phi}. \tag{309}$$

Proof. Simply compute that, since i is distributed uniformly over I and t is distributed uniformly over T, where I,T are given in (300), we have

$$\begin{aligned}\mathbb{P}[H_m(x)] &\stackrel{(308)}{=} \frac{1}{\lceil a\rceil - \lceil b\rceil + 1} \sum_{i=\lceil b\rceil}^{\lceil a\rceil} \frac{|\{t \in T:\ s_{i+m-1}(x) \leqslant t < s_{i+m}(x)\}|}{\lceil\log\Phi\rceil} \\ &= \frac{|T \cap \{\lceil s_{\lceil b\rceil+m-1}(x)\rceil,\ldots,\lceil s_{\lceil a\rceil+m}(x)\rceil - 1\}|}{(\lceil a\rceil - \lceil b\rceil + 1)\lceil\log\Phi\rceil} = \frac{\lceil s_{\lceil a\rceil+m}(x)\rceil - \lceil s_{\lceil b\rceil+m-1}(x)\rceil}{(\lceil a\rceil - \lceil b\rceil + 1)\lceil\log\Phi\rceil} \gtrsim \frac{\Big\lfloor \log \frac{\mu(B(x,2^{a+m}))}{\mu(B(x,2^{b+m}))} \Big\rfloor}{(a-b+1)\log\Phi},\end{aligned} \tag{310}$$

where the penultimate step of (310) is valid thanks to our normalization $\min_{z\in\mathcal{M}} \mu(z) = 1$, which ensures that $T \supseteq \{\lceil s_{\lceil b\rceil+m-1}(x)\rceil,\ldots,\lceil s_{\lceil a\rceil+m}(x)\rceil - 1\}$, and the last step of (310) holds since $\lceil a\rceil - \lceil b\rceil + 1 \asymp a-b+1$ and $\lceil\log\Phi\rceil \asymp \log\Phi$, as $a > b$ and $\Phi \geqslant 2$, and furthermore

$$\begin{aligned}\lceil s_{\lceil a\rceil+m}(x)\rceil - \lceil s_{\lceil b\rceil+m-1}(x)\rceil &\gtrsim \lfloor s_{\lceil a\rceil+m}(x) - s_{\lceil b\rceil+m-1}(x)\rfloor \\ &\stackrel{(297)}{=} \bigg\lfloor \log\frac{\mu(B(x,2^{\lceil a\rceil+m}))}{\mu(B(x,2^{\lceil b\rceil-1+m}))}\bigg\rfloor \geqslant \bigg\lfloor \log\frac{\mu(B(x,2^{a+m}))}{\mu(B(x,2^{b+m}))}\bigg\rfloor.\end{aligned}$$ □

For the next (and final) claim, we define for every $n \in \mathbb{Z}$, $x \in \mathcal{M}$ and $\ell \geqslant 0$ an event $L_n(x,\ell) \subseteq \Omega$ by

$$L_n(x,\ell) \stackrel{\mathrm{def}}{=} \big\{d(x,\mathscr{Z}_n) \geqslant \ell\big\}. \tag{311}$$

Claim 72. *For every $n \in \mathbb{Z}$, $x \in \mathcal{M}$, and $\ell \geqslant 0$, the following inclusion of events holds:*

$$\big\{\mathscr{Z} \neq \emptyset \text{ and } d(x,\mathscr{Z}) \geqslant \min\{2^{b+n-2},\ell\}\big\} \supseteq H_n(x) \cap L_n(x,\ell) \cap C_n \cap \big\{\mathscr{Z} \neq \emptyset\big\}. \tag{312}$$

Proof. Suppose that $z \in \mathcal{M}$ satisfies $d(z,x) < \min\{2^{b+n-2},\ell\}$, and also that the event in the right hand side of (312) occurs. Our goal is to deduce that $z \notin \mathscr{Z}$. Indeed, $d(z,\mathscr{Z}_n) \geqslant d(x,\mathscr{Z}_n) - d(z,x) > d(x,\mathscr{Z}_n) - \ell \geqslant 0$, because part of our assumption on z is that $d(x,z) < \ell$ and the occurrence of the event in the right hand side of (312) ensures in particular that $L_n(x,\ell)$ occurs, i.e., $d(x,\mathscr{Z}_n) \geqslant \ell$. Thus, $z \notin \mathscr{Z}_n$. By (302), we will therefore be done if we will show that the occurrence of the event in the right hand side of (312) implies that $\mathscr{k}(z,\mathrm{t}) - \mathrm{i} + \eta_{\mathscr{k}(z,\mathrm{t})-\mathrm{i}} = n$ and $\sigma_{\mathscr{k}(z,\mathrm{t})-\mathrm{i}} = 0$. This indeed holds since $s_{\mathrm{i}+n-1}(x) \leqslant \mathrm{t} < s_{\mathrm{i}+n}(x)$ as $H_n(x)$ occurs, so $\mathscr{k}(x,\mathrm{t}) = \mathrm{i}+n-1$ by (298). But $\mathrm{i} \geqslant \lceil b\rceil \geqslant b$ since $\mathrm{i} \in I$, where we recall that $I \subseteq \mathbb{Z}$ is given in (300), so $\mathscr{k}(x,\mathrm{t}) \geqslant b+n-1$. Hence, $d(z,x) < 2^{b+n-2} \leqslant 2^{\mathscr{k}(x,\mathrm{t})-1}$ using the rest of our assumption on z, so by Lemma 69 we see that $\mathscr{k}(z,\mathrm{t}) - \mathrm{i} \in \{n-2,n-1,n\}$. As the occurrence of the right hand side of (312) ensures that the event C_n in (304) occurs, it follows that $\sigma_{\mathscr{k}(z,\mathrm{t})-\mathrm{i}} = 0$ and $\mathscr{k}(z,\mathrm{t}) - \mathrm{i} + \eta_{\mathscr{k}(z,\mathrm{t})-\mathrm{i}} = n$, as required. □

Next, fix $n \in \mathbb{Z}$, $x, y \in \mathcal{M}$ and $\ell, u \geqslant 0$. Denote $|B(y,u)| = p$ and write

$$B(y,u) = \{w_1, \dots, w_p\}, \tag{313}$$

i.e., we are fixing an arbitrary linear order on $B(y,u)$. For every $q \in [p]$ define an event $A_q \subseteq \Omega$ by

$$A_q \stackrel{\mathrm{def}}{=} \Big(E_n(w_q) \smallsetminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap \big(F_n(w_q) \cup G(w_q)\big) \cap H_n(x) \cap L_n(x,\ell) \cap C_n. \tag{314}$$

Thus, $A_1, \dots, A_p$ are pairwise disjoint and by combining Claim 72 and Claim 70 we see that

$$\big\{\mathscr{Z} \neq \varnothing \ \text{ and } \ d(x,\mathscr{Z}) \geqslant \min\{2^{b+n-2}, \ell\} \ \text{ and } \ d(y,\mathscr{Z}) \leqslant u\big\} \supseteq \bigcup_{q=1}^{p} A_q.$$

Consequently,

$$\mathbb{P}\big[\mathscr{Z} \neq \varnothing \ \text{ and } \ d(x,\mathscr{Z}) \geqslant \min\{2^{b+n-2}, \ell\} \ \text{ and } \ d(y,\mathscr{Z}) \leqslant u\big] \geqslant \sum_{q=1}^{p} \mathbb{P}[A_q]. \tag{315}$$

In order to understand the right hand side of (315), define for every $q \in [p]$ events $A'_q, A''_q \subseteq \Omega$ by

$$A'_q \stackrel{\mathrm{def}}{=} \Big(E_n(w_q) \smallsetminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap F_n(w_q) \cap H_n(x) \cap L_n(x,\ell) \cap C_n, \tag{316}$$

and,

$$A''_q \stackrel{\mathrm{def}}{=} \Big(E_n(w_q) \smallsetminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap \big(\Omega \smallsetminus F_n(w_q)\big) \cap G(w_q) \cap H_n(x) \cap L_n(x,\ell) \cap C_n \stackrel{(314)\wedge(316)}{=} A_q \smallsetminus A'_q. \tag{317}$$

We therefore have

$$\forall\, q \in [p], \qquad \mathbb{P}[A_q] = \mathbb{P}[A'_q] + \mathbb{P}[A''_q]. \tag{318}$$

Recalling (306) and (308), the event $(\Omega \smallsetminus F_n(w_q)) \cap H_n(x)$ in the right hand side of (317) means that $\mathsf{s}_{\mathsf{i}+n-1}(x) \leqslant \mathsf{t} < \mathsf{s}_{\mathsf{i}+n}(x)$ and $\mathscr{k}(w_q, \mathsf{t}) - \mathsf{i} \notin \{n-2, n-1, n\}$. So, by integrating with respect to i and t we get

$$\mathbb{P}[A''_q] = \frac{1}{|I| \cdot |T|} \sum_{\substack{(i,t) \in I \times T \\ s_{i+n-1}(x) \leqslant t < s_{i+n}(x) \\ \mathscr{k}(w_q,t) - i \notin \{n-2,n-1,n\}}} \mathbb{P}\bigg[\Big(E_n(w_q) \smallsetminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap \big\{\sigma_{\mathscr{k}(w_q,t)-i} = 1\big\} \cap L_n(x,\ell) \cap C_n\bigg], \tag{319}$$

where we used the facts that the events $E_n(w_1), \dots, E_n(w_p)$ and $L_n(x,\ell)$, defined in (306) and (311), respectively, depend only on $\mathscr{Z}_n$, the event C_n from (304) depends only on $(\sigma_{n-2}, \sigma_{n-1}, \sigma_n, \eta_{n-2}, \eta_{n-1}, \eta_n)$, and the event $G(w_q)$, that does depend on i, t per (306), is equal to $\{\sigma_{\mathscr{k}(w_q,\mathsf{t})-\mathsf{i}} = 1\}$. The crucial point to observe now is that if $(i,t) \in I \times T$ are such that $\mathscr{k}(w_q, t) - i \notin \{n-2, n-1, n\}$, then $\sigma_{\mathscr{k}(w_q,t)-i}$ and C_n are independent and $\mathbb{P}[\sigma_{\mathscr{k}(w_q,t)-i} = 1] = 1/2$, so each of the summands in the right hand side (319) satisfy

$$\mathbb{P}\bigg[\Big(E_n(w_q) \smallsetminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap \big\{\sigma_{\mathscr{k}(w_q,t)-i} = 1\big\} \cap L_n(x,\ell) \cap C_n\bigg] \stackrel{(305)}{=} \frac{1}{2^4 3^3} \mathbb{P}\bigg[\Big(E_n(w_q) \smallsetminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap L_n(x,\ell)\bigg].$$

Thus,

$$\mathbb{P}[A''_q] = \frac{1}{|I| \cdot |T|} \sum_{\substack{(i,t) \in I \times T \\ s_{i+n-1}(x) \leqslant t < s_{i+n}(x) \\ \mathscr{k}(w_q,t) - i \notin \{n-2,n-1,n\}}} \frac{1}{2^4 3^3} \mathbb{P}\bigg[\Big(E_n(w_q) \smallsetminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap L_n(x,\ell)\bigg]. \tag{320}$$

The analogous consideration shows mutatis mutandis that if for $q \in [p]$ we define $A_q''' \subseteq \Omega$ by

$$A_q''' \stackrel{\mathrm{def}}{=} \Big(E_n(w_q) \setminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap \big(\Omega \setminus F_n(w_q)\big) \cap H_n(x) \cap L_n(x,\ell) \cap C_n, \tag{321}$$

then

$$\Pr[A_q'''] = \frac{1}{|I|\cdot|T|} \sum_{\substack{(i,t)\in I\times T\\ s_{i+n-1}(x)\leqslant t<s_{i+n}(x)\\ \hbar(w_q,t)-i\notin\{n-2,n-1,n\}}} \frac{1}{2^3 3^3}\mathbb{P}\left[\Big(E_n(w_q)\setminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big)\cap L_n(x,\ell)\right] \stackrel{(320)}{=} 2\mathbb{P}[A_q'']. \tag{322}$$

Consequently, for every $q \in [p]$ we have

$$\mathbb{P}[A_q] \stackrel{(318)\wedge(322)}{\geqslant} \mathbb{P}[A_q'] + \frac{1}{2}\mathbb{P}[A_q'''] \geqslant \frac{1}{2}\big(\mathbb{P}[A_q'] + \mathbb{P}[A_q''']\big) = \frac{1}{2}\mathbb{P}[A_q''''], \tag{323}$$

where we introduce the notation

$$A_q'''' \stackrel{\mathrm{def}}{=} \Big(E_n(w_q) \setminus \bigcup_{r=1}^{q-1} E_n(w_r)\Big) \cap H_n(x) \cap L_n(x,\ell) \cap C_n, \tag{324}$$

and observe that by (316), (321) and (324), the events A_q', A_q''' are disjoint and they satisfy $A_q' \cup A_q''' = A_q''''$.

Finally, the events $\{A_q''''\}_{q=1}^p$ are pairwise disjoint, so

$$\mathbb{P}\big[\mathscr{Z} \neq \emptyset \text{ and } d(x,\mathscr{Z}) \geqslant \min\{2^{b+n-2},\ell\} \text{ and } d(y,\mathscr{Z}) \leqslant u\big] \stackrel{(315)\wedge(324)}{\geqslant} \frac{1}{2}\sum_{q=1}^p \mathbb{P}[A_q''''] = \frac{1}{2}\mathbb{P}\Big(\bigcup_{q=1}^p A_q''''\Big). \tag{325}$$

Observing the following simple identity of events,

$$\begin{aligned}\bigcup_{q=1}^p A_q'''' &\stackrel{(324)}{=} \Big(\bigcup_{q=1}^p E_n(w_q)\Big) \cap H_n(x) \cap L_n(x,\ell) \cap C_n \stackrel{(306)\wedge(313)}{=} \big\{B(y,u)\cap \mathscr{Z}_n \neq \emptyset\big\} \cap H_n(x) \cap L_n(x,\ell) \cap C_n\\ &\stackrel{(311)}{=} H_n(x) \cap \big\{d(x,\mathscr{Z}_n) \geqslant \ell \text{ and } d(y,\mathscr{Z}_n) \leqslant u\big\} \cap C_n,\end{aligned}$$

we get from the independence of the events $H_n(x), \{d(x,\mathscr{Z}_n) \geqslant \ell \text{ and } d(y,\mathscr{Z}_n) \leqslant u\}, C_n$ (recalling (304) and (308), they depend on $(\mathsf{i},\mathsf{t}), \mathscr{Z}_n, (\sigma_{n-2},\sigma_{n-1},\sigma_n,\eta_{n-2},\eta_{n-1},\eta_n)$, respectively) and (325) that

$$\begin{aligned}&\mathbb{P}\big[\mathscr{Z} \neq \emptyset \text{ and } d(x,\mathscr{Z}) \geqslant \min\{2^{b+n-2},\ell\} \text{ and } d(y,\mathscr{Z}) \leqslant u\big]\\ &\qquad\gtrsim \mathbb{P}\Big[H_n(x) \cap \big\{d(x,\mathscr{Z}_n) \geqslant \ell \text{ and } d(y,\mathscr{Z}_n) \leqslant u\big\} \cap C_n\Big]\\ &\qquad= \mathbb{P}[H_n(x)]\cdot \mathbb{P}[d(x,\mathscr{Z}_n) \geqslant \ell \text{ and } d(y,\mathscr{Z}_n) \leqslant u]\cdot \mathbb{P}[C_n]\\ &\qquad\gtrsim \frac{\Big\lfloor \log \frac{\mu(B(x,2^{a+n}))}{\mu(B(x,2^{b+n}))}\Big\rfloor}{(a-b+1)\log\Phi}\mathbb{P}[d(x,\mathscr{Z}_n) \geqslant \ell \text{ and } d(y,\mathscr{Z}_n) \leqslant u],\end{aligned}$$

where the last step is an application of (305) and conclusion (309) of Claim 71. This proves (303). □

11. Proof of Theorem 4

We preferred to state Theorem 4 for finite metric spaces to avoid the need to consider measurability issues about distributions over random subsets. Nevertheless, all that one needs here is that the event that appears in the left hand side of (5) is $\mathbb{P}^\tau$-measurable, and it is possible to ensure that the ensuing proof of Theorem 4 achieves this under mild assumptions on a separable metric space $(\mathcal{M},d)$ and any nondegenerate Borel measure μ on $\mathcal{M}$ with $\mu(\mathcal{M}) < \infty$; such a treatment appears in [Nao24a, Section 3.3].

The ensuing proof of Theorem 4 generalizes the proof of the main result of [MN07]; the incorporation of the selectors $\{\sigma_t\}_{t=1}^\infty$ (see the paragraph after equation (328)) originates from [Rao99].

Proof of Theorem 4. As the right hand side of (5) involves only a ratio of two values of μ, by normalizing we may assume that μ is a probability measure. Let $\{\mathsf{z}_t\}_{t=1}^{\infty}$ be i.i.d. points distributed according to μ, and denote their law by $\mathbb{P}^{\mu}=\mu^{\otimes\mathbb{N}}$. For every $r>0$ and $x\in\mathcal{M}$ consider the following random variable:

$$T(x,r)\stackrel{\mathrm{def}}{=}\inf\big\{t\in\mathbb{N}:\ d(\mathsf{z}_t,x)\leqslant r\big\}, \tag{326}$$

with the convention that $T(x,r)=\infty$ if there is no $t\in\mathbb{N}$ for which $d(\mathsf{z}_t,x)\leqslant r$. Observe that because μ is assumed to be a nondegenerate measure, $T(x,r)$ is finite almost everywhere. Indeed,

$$\begin{aligned}\mathbb{P}^{\mu}\big[T(x,r)=\infty\big]&=\mathbb{P}^{\mu}\big[\forall\, t\in\mathbb{N},\ d(\mathsf{z}_t,x)>r)\big]\\&=\lim_{t\to\infty}\mathbb{P}^{\mu}\Big[\bigcap_{s=1}^{t}\big\{\mathsf{z}_s\in\mathcal{M}\smallsetminus B(x,r)\big\}\Big]=\lim_{t\to\infty}\big(1-\mu(B(x,r))\big)^t=0,\end{aligned}$$

where in the penultimate step uses the independence of $\{\mathsf{z}_s\}_{s=1}^{\infty}$ and the final step holds as $\mu(B(x,r))>0$.

Next, we claim that the following inclusion of events holds:

$$\big\{\forall\, w\in B(x,\lambda\tau),\ T(w,r)=t\big\}\supseteq\left(\bigcap_{s=1}^{t-1}\big\{\mathsf{z}_s\in\mathcal{M}\smallsetminus B(x,r+\lambda\tau)\big\}\right)\cap\big\{\mathsf{z}_t\in B(x,r-\lambda\tau)\big\}. \tag{327}$$

Indeed, suppose that the event in the right hand side of (327) occurs, i.e., for every $s\in\{1,\ldots,t-1\}$ we have $d(\mathsf{z}_s,x)>r+\lambda\tau$ and also $d(\mathsf{z}_t,x)\leqslant r-\lambda\tau$. Therefore, if $w\in B(x,\lambda\tau)$ and $s\in\{1,\ldots,t-1\}$, then

$$d(\mathsf{z}_s,w)\geqslant d(\mathsf{z}_s,x)-d(x,w)>(r+\lambda\tau)-\lambda\tau=r,$$

which means that $T(w,r)\geqslant t$, by the definition (326). At the same time,

$$d(\mathsf{z}_t,w)\leqslant d(\mathsf{z}_t,x)+d(x,w)\leqslant(r-\lambda\tau)+\lambda\tau=r,$$

so, in fact, $T(w,r)=t$. As this holds for every $w\in B(x,\lambda\tau)$, we thus checked that the event in the left hand side of (327) occurs. Using (327) and the independence of $\{\mathsf{z}_s\}_{s=1}^{t}$, we therefore get the following bound:

$$\mathbb{P}^{\mu}\big[\forall\, w\in B(x,\lambda\tau),\ T(w,r)=t\big]\geqslant\Big(1-\mu\big(B(x,r+\lambda\tau)\big)\Big)^{t-1}\mu\big(B(x,r-\lambda\tau)\big). \tag{328}$$

Let R be a random variable that is distributed uniformly over the interval $(\tau/4,\tau/2)$. Also, let $\{\sigma_t\}_{t=1}^{\infty}$ be standard Bernoulli random variables, i.e., $\Pr[\sigma_t=1]=\Pr[\sigma_t=0]=1/2$ for every $t\in\mathbb{N}$. We will require below that the random variables $\{\mathsf{z}_t\}_{t=1}^{\infty},\mathsf{R},\{\sigma_t\}_{t=1}^{\infty}$ are independent. Denote by $\mathbb{P}^{\tau}$ their joint law (note that the dependence on the scale τ here arises only from the presence of the random variable R). Finally, define a random subset $\mathscr{Z}\subseteq\mathcal{M}$ as follows:

$$\mathscr{Z}\stackrel{\mathrm{def}}{=}\big\{x\in\mathcal{M}:\ \sigma_{T(x,\mathsf{R})}=1\big\}. \tag{329}$$

We will next demonstrate that (5) holds, thus proving Theorem 4 by conditioning on the event $\{\mathscr{Z}\neq\emptyset\}$.

Fix $x,y\in\mathcal{M}$ with $d(x,y)\geqslant\tau$ and observe the following inclusion of events:

$$\begin{aligned}&\big\{d(y,\mathscr{Z})\geqslant\lambda\tau\quad\text{and}\quad x\in\mathscr{Z}\big\}\\&\qquad\supseteq\bigcup_{\substack{(s,t)\in\mathbb{N}\\ s\neq t}}\big(\{T(x,\mathsf{R})=s\}\cap\{\sigma_s=1\}\cap\{\forall\, w\in B(y,\lambda\tau),\ T(w,\mathsf{R})=t\}\cap\{\sigma_t=0\}\big).\end{aligned} \tag{330}$$

To check (330), if the right hand side of (330) occurs, then $x\in\mathscr{Z}$ and $B(y,\lambda r)\subseteq\mathcal{M}\smallsetminus\mathscr{Z}$ by (329). The latter inclusion implies that $d(y,\mathscr{Z})\geqslant\lambda r$, i.e., the left hand side of (330) occurs. As the events in right hand side of (330), namely, $\{\{T(x,\mathsf{R})=s\}\cap\{\sigma_s=1\}\cap\{\forall\, w\in B(y,\lambda\tau),\ T(w,\mathsf{R})=t\}\cap\{\sigma_t=0\}\}_{s,t=1}^{\infty}$ are pairwise disjoint (as membership in each of them implies $s=T(x,\mathsf{R})$ and $t=T(y,\mathsf{R})$), we deduce from (330) that

$$\mathbb{P}^{\tau}\big[d(y,\mathscr{Z})\geqslant\lambda\tau\quad\text{and}\quad x\in\mathscr{Z}\big]\geqslant\sum_{\substack{(s,t)\in\mathbb{N}\\ s\neq t}}\frac14\mathbb{P}^{\tau}\big[T(x,\mathsf{R})=s\quad\text{and}\quad\forall\, w\in B(y,\lambda\tau),\ T(w,\mathsf{R})=t\big], \tag{331}$$

where we used the fact that the random variables $T(x,\mathsf{R}),T(y,\mathsf{R})$ depend only on $\{\mathsf{z}_k\}_{k=1}^{\infty}$ and R, so they are independent of σ_s,σ_t for each $s,t\in\mathbb{N}$, and σ_s,σ_t are independent if $s\neq t$. For every $t\in\mathbb{N}$, the events

$\{T(x,\mathsf{R})=t\}$ and $\{\forall\, w\in B(y,\lambda\tau),\ T(w,\mathsf{R})=t\}\subseteq\{T(y,\mathsf{R})=t\}$ are disjoint, since if $T(x,\mathsf{R})=T(y,\mathsf{R})=t$, then by the definition (326) of $T(\cdot,\cdot)$ we would have $d(\mathsf{z}_t,x),d(\mathsf{z}_t,y)\leqslant\mathsf{R}<\tau/2$, so $d(x,y)<\tau$ by the triangle inequality, in contradiction to our assumption. Thus, (331) can be rewritten as follows:

$$\begin{aligned}\mathbb{P}^\tau\big[d(y,\mathscr{Z})\geqslant\lambda\tau\quad\text{and}\quad x\in\mathscr{Z}\big]&\geqslant\frac14\sum_{s=1}^\infty\sum_{t=1}^\infty\mathbb{P}^\tau\big[T(x,\mathsf{R})=s\quad\text{and}\quad\forall\, w\in B(y,\lambda\tau),\ T(w,\mathsf{R})=t\big]\\&=\frac14\sum_{t=1}^\infty\mathbb{P}^\tau\big[\forall\, w\in B(y,\lambda\tau),\ T(w,\mathsf{R})=t\big]\\&=\frac14\sum_{t=1}^\infty\frac4\tau\int_{\frac14\tau}^{\frac12\tau}\mathbb{P}^\mu\big[\forall\, w\in B(y,\lambda\tau),\ T(w,r)=t\big]\,\mathrm{d}r,\end{aligned}\tag{332}$$

where the last step of (332) uses the definition of R and its independence from $\{\mathsf{z}_k\}_{k=1}^\infty$.

We can now conclude the justification (4) as follows:

$$\begin{aligned}&\mathbb{P}^\tau\big[d(y,\mathscr{Z})\geqslant\lambda\tau\quad\text{and}\quad x\in\mathscr{Z}\big]\\&\geqslant\frac1\tau\int_{\frac14\tau}^{\frac12\tau}\Big(\sum_{t=1}^\infty\big(1-\mu(B(x,r+\lambda\tau))\big)^{t-1}\mu(B(x,r-\lambda\tau))\Big)\,\mathrm{d}r \qquad (333)\\&=\frac1\tau\int_{\frac14\tau}^{\frac12\tau}e^{\log\mu(B(x,r-\lambda\tau))-\log\mu(B(x,r+\lambda\tau))}\,\mathrm{d}r\\&\geqslant\frac14\exp\left(\frac4\tau\Big(\int_{\frac14\tau}^{\frac12\tau}\log\mu(B(x,r-\lambda\tau))\,\mathrm{d}r-\int_{\frac14\tau}^{\frac12\tau}\log\mu(B(x,r+\lambda\tau))\,\mathrm{d}r\Big)\right) \qquad (334)\\&=\frac14\exp\left(\frac4\tau\Big(\int_{(\frac14-\lambda)\tau}^{(\frac14+\lambda)\tau}\log\mu(B(x,\rho))\,\mathrm{d}\rho-\int_{(\frac12-\lambda)\tau}^{(\frac12+\lambda)\tau}\log\mu(B(x,\rho))\,\mathrm{d}\rho\Big)\right)\\&\geqslant\frac14\exp\Big(8\lambda\Big(\log\mu\big(B(x,\tfrac18\tau)\big)-\log\mu\big(B(x,\tfrac58\tau)\big)\Big)\Big)\\&=\frac14\left(\frac{\mu(B(y,\frac58\tau))}{\mu(B(y,\frac18\tau))}\right)^{-8\lambda},\end{aligned}$$

where (333) is a combination of (328) and (332), in (334) we used Jensen's inequality, and the rest of the steps above hold because $\lambda\leqslant 1/8$. This completes the proof of Theorem 4. □

12. IMPOSSIBILITY RESULTS FOR AVERAGE DISTORTION

Fact 10 is the special case $q=2$ of the following proposition:

Proposition 73. *Fix $p\geqslant 1$ and $n\in\mathbb{N}$. If $1\leqslant q\leqslant 2$, then the smallest $D\geqslant 1$ for which ℓ_q^n embeds into $\mathbb{R}$ with p-average distortion D is bounded from above and below by positive universal constant multiples of $\sqrt{\max\{1,n/p\}}$. If $q\geqslant 2$, then the smallest $D\geqslant1$ for which ℓ_q^n embeds into $\mathbb{R}$ with p-average distortion D is at most a universal constant multiple of $(\max\{1,n/p\})^{1-1/q}$, and it is at least a positive universal constant multiple of the following quantity:*

$$\begin{cases}\frac qp & \text{if } 1\leqslant p\leqslant\frac{q}{n^{\frac1q}}\text{ and } q\leqslant\log n,\\ n^{\frac1q} & \text{if }\max\Big\{1,\frac{q}{n^{\frac1q}}\Big\}\leqslant p\leqslant q\leqslant\log n,\\ \frac{\log n}{p} & \text{if } p\leqslant\log n\leqslant q,\\ 1 & \text{if }\log n\leqslant p\leqslant q\text{ or }p\geqslant\max\{q,\frac{n}{e^q}\},\\ \left(\frac np\right)^{\frac1q} & \text{if } q\leqslant p\leqslant\frac{n}{e^q}.\end{cases}\tag{335}$$

Since ℓ_q embeds quasisymmetrically into ℓ_2 when $1 \leqslant q \leqslant 2$ (specifically, by [BDCK66] its $\frac{q}{2}$-snowflake is isometric to a subset of ℓ_2), for this range of q the upper bound on D in Proposition 73 is a special case of Theorem 9. But, Theorem 9 has a much simpler proof in its special case $\mathcal{M} = \ell_q^n$, which we will include below. When $q > 2$ the upper and lower bounds on D in Proposition 73 do not match, and they do not belong to the framework that we study herein because ℓ_q does not admit a quasisymmetric mebedding into a Hilbert space [Nao12a]. Nevertheless, it is beneficial to derive below the best bounds that we currently have for q in this range, as a small step towards the following independently interesting open question that arises naturally from the above discussion:

Question 74. *Given $n \in \mathbb{N}$ and $p \geqslant 1$, what is the growth rate as $n \to \infty$ of the smallest $D \geqslant 1$ such that ℓ_∞^n embeds with p-average distortion D into $\mathbb{R}$? We currently do not know the answer to this question even in the (most interesting) case $p = 2$. More generally, for every $q > 2$, what is the order of magnitude (up to universal constant factors) of the smallest $D \geqslant 1$ such that ℓ_q^n embeds with p-average distortion D into $\mathbb{R}$?*

By [Nao14], when $q > 2$ the smallest $D \geqslant 1$ such that ℓ_q embeds into ℓ_2 with quadratic[37] average distortion D is bounded from above and from below by positive universal constant multiples of q (the proof of this fact in [Nao14] is nonconstructive; in [KNT21] an algorithmic proof was found). Proposition 73 shows that the situation is markedly different if one wishes to study the average distortion of ℓ_q in $\mathbb{R}$, and Question 74 aims to understand this phenomenon, with possible algorithmic ramifications, depending on the answer. Regardless, it is likely that a substantially new idea will be needed to answer Question 74.

Proof of Proposition 73. We will first derive the asserted upper bounds on D. For $q > 1$ let $\mathsf{H} = \mathsf{H}^{(q)}$ be the symmetric real-valued random variable whose density at each $s \in \mathbb{R}$ is equal to

$$\frac{1}{2\Gamma\left(2-\frac{1}{q}\right)} e^{-|s|^{\frac{q}{q-1}}}.$$

We extend this notation to $q = 1$ by letting $\mathsf{H} = \mathsf{H}^{(1)}$ be a symmetric Bernoulli random variable, namely, $\Pr[\mathsf{H} = 1] = \Pr[\mathsf{H} = -1] = 1/2$. If we take $\mathsf{H}_1, \ldots, \mathsf{H}_n$ to be i.i.d. copies of H and consider the random vector

$$\mathsf{U} \stackrel{\mathrm{def}}{=} (\mathsf{H}_1, \ldots, \mathsf{H}_n) \in \mathbb{R}^n,$$

then the random vector $\mathsf{U}/\|\mathsf{U}\|_{q/(q-1)}$ takes (by design) values in the unit sphere of the dual $\ell_{q/(q-1)}^n$ of ℓ_q^n, and furthermore the random variable $\|\mathsf{U}\|_{q/(q-1)}$ is independent of $\mathsf{U}/\|\mathsf{U}\|_{q/(q-1)}$; these useful probabilistic facts are due to [SZ90, RR91] (generalizations that could pertain to future extensions of the ensuing reasoning were found in [BLMN06, Section 2] and [Nao24a, Lemma 156]).[38]

By e.g. equation (6.48) in [Nao24a], in combination with Stirling's formula, we have

$$\left(\mathbb{E}\Big[\|\mathsf{U}\|_{\frac{q}{q-1}}^p\Big]\right)^{\frac{1}{p}} = \left(\frac{\Gamma\left(\frac{(n+p)(q-1)}{q}\right)}{\Gamma\left(\frac{n(q-1)}{q}\right)}\right)^{\frac{1}{p}} \asymp \big(\max\{p,n\}\big)^{1-\frac{1}{q}}. \tag{336}$$

Also, by [BLMN06, Proposition 7] (whose key input is [GK95]; for $q = 1$ it suffices to use [Hit93] here), for every $a_1, \ldots, a_n \in \mathbb{R}$, if we let $a_1^* \geqslant a_2^* \geqslant \ldots \geqslant a_n^* \geqslant 0$ be the decreasing rearrangement of $|a_1|, \ldots, |a_n|$, then

$$\left(\mathbb{E}\bigg[\Big|\sum_{i=1}^n a_i \mathsf{H}_i\Big|^p\bigg]\right)^{\frac{1}{p}} \asymp p^{1-\frac{1}{q}}\left(\sum_{i=1}^{\lfloor p\rfloor}(a_i^*)^q\right)^{\frac{1}{q}} + \sqrt{p}\left(\sum_{i=\lfloor p\rfloor}^n (a_i^*)^2\right)^{\frac{1}{2}} \gtrsim \min\left\{p^{1-\frac{1}{q}}, \frac{\sqrt{p}}{n^{\max\left\{0,\frac{1}{q}-\frac{1}{2}\right\}}}\right\}\|a\|_q, \tag{337}$$

[37] By [Nao21a, Proposition 6], this statement formally implies bounds on the p-average distortion of ℓ_q into a Hilbert space.

[38] Note that when $q = 1$ the random vector U is distributed uniformly over $\{-1,1\}^n$ and $\|\mathsf{U}\|_\infty = 1$ is constant. If one does not mind obtaining worse bounds on the universal constants that the ensuing proof provides, then in the range $1 \leqslant q \leqslant 2$ one could work instead with the random vector $\mathsf{U}^{(1)}/n^{(q-1)/q}$ when $\mathsf{U}^{(1)}$ is distributed uniformly over $\{-1,1\}^n$.

where the last step of (337) follows from a straightforward application of Hölder's inequality. Hence,

$$\left(\mathbb{E}\left[\left|\left\langle a,\frac{1}{\|\mathsf{U}\|_{\frac{q}{q-1}}}\mathsf{U}\right\rangle\right|^p\right]\right)^{\frac{1}{p}}=\left(\frac{\mathbb{E}\left[|\langle a,\mathsf{U}\rangle|^p\right]}{\mathbb{E}\left[\|\mathsf{U}\|_{\frac{q}{q-1}}^p\right]}\right)^{\frac{1}{p}}\gtrsim\left(\min\left\{1,\frac{n}{p}\right\}\right)^{1-\frac{1}{\max\{2,q\}}},\tag{338}$$

where in (338) we continue to denote the standard scalar product on $\mathbb{R}^n$ by $\langle\cdot,\cdot\rangle:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$, the first step of (338) holds because $\|\mathsf{U}\|_{q/(q-1)}$ is independent of $\mathsf{U}/\|\mathsf{U}\|_{q/(q-1)}$, and the last step of (338) consists of a substitution of (336) and (337) and simplification of the resulting expression.

Thanks to (338), there is a universal constant $C>0$ such that if we consider the random linear functional $\mathsf{F}:\ell_q^n\to\mathbb{R}$ that is given by setting for every $x\in\mathbb{R}^n$,

$$\mathsf{F}(x)\stackrel{\mathrm{def}}{=}\frac{C\left(\max\left\{1,\frac{n}{p}\right\}\right)^{1-\frac{1}{\max\{2,q\}}}}{\|\mathsf{U}\|_{\frac{q}{q-1}}}\langle x,\mathsf{U}\rangle,$$

then the Lipschitz constant of F with respect to the ℓ_q^n metric is at most $C(\max\{1,n/p\})^{1-1/\max\{2,q\}}$, and

$$\forall\, x,y\in\mathbb{R}^n,\qquad \left(\mathbb{E}\left[\|\mathsf{F}(x)-\mathsf{F}(y)\|_q^p\right]\right)^{\frac{1}{p}}\geqslant\|x-y\|_q.\tag{339}$$

By Fubini, it follows from (339) that for every Borel probability measure μ on $\mathbb{R}^n$ we have

$$\mathbb{E}\left[\iint_{\mathbb{R}^n\times\mathbb{R}^n}|\mathsf{F}(x)-\mathsf{F}(y)|^p\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right]\geqslant\iint_{\mathbb{R}^n\times\mathbb{R}^n}\|x-y\|_q^p\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y).$$

Hence, with positive probability the $C(\max\{1,n/p\})^{1-1/\max\{2,q\}}$-Lipschitz function F satisfies

$$\left(\iint_{\mathbb{R}^n\times\mathbb{R}^n}|\mathsf{F}(x)-\mathsf{F}(y)|^p\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{p}}\geqslant\left(\iint_{\mathbb{R}^n\times\mathbb{R}^n}\|x-y\|_q^p\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{p}},$$

thus proving the upper bound that is asserted in Proposition 73 (in the entire range $q\geqslant 1$) on the smallest $D\geqslant 1$ for which ℓ_q^n embeds with p-average distortion D into $\mathbb{R}$.

It remains to prove the lower bounds on the average distortion that are asserted in Proposition 73. Observe first that it suffices to treat only the case $q\geqslant 2$. Indeed, if $1\leqslant q\leqslant 2$ then by [FLM77] there exists an integer $m\asymp n$ such that ℓ_2^n embeds with (bi-Lipschitz) distortion $O(1)$ into ℓ_q^n. Hence, if ℓ_q^n embeds with p-average distortion D into $\mathbb{R}$, then also ℓ_2^n embeds with p-average distortion $O(D)$ into $\mathbb{R}$, which implies that $D\gtrsim\sqrt{\max\{1,n/p\}}$ by the (yet to be proven) case $q=2$ of Proposition 73. Thus, suppose that $q\geqslant 2$ and that ℓ_q^n embeds with p-average distortion $D\geqslant 1$ into $\mathbb{R}$. The goal is to show that necessarily

$$D\gtrsim\begin{cases}\frac{q}{p} & \text{if } 1\leqslant p\leqslant\frac{q}{n^{\frac{1}{q}}}\text{ and } q\leqslant\log n,\\ n^{\frac{1}{q}} & \text{if }\max\left\{1,\frac{q}{n^{\frac{1}{q}}}\right\}\leqslant p\leqslant q\leqslant\log n,\\ \frac{\log n}{p} & \text{if } p\leqslant\log n\leqslant q,\\ 1 & \text{if }\log n\leqslant p\leqslant q\text{ or } p\geqslant\max\left\{q,\frac{n}{e^q}\right\},\\ \left(\frac{n}{p}\right)^{\frac{1}{q}} & \text{if } q\leqslant p\leqslant\frac{n}{e^q},\end{cases}\tag{340}$$

Let $k=k(n)\in\mathbb{N}$ be the largest integer such that $k(k-1)/2\leqslant n$. Thus $k\asymp\sqrt{n}$, so $\log k\asymp\log n$. By combining part (i) of Theorem 1 in [Mat97] with Proposition 1 in [Bal90], every k-point metric space embeds into ℓ_q^n with (bi-Lipschitz) distortion $O(\max\{1,(\log k)/q\})\asymp\max\{1,(\log n)/q\}$. At the same time, by the *proof of* part (ii) of Theorem 1 in [Mat97] there exists a k-point metric space $\mathcal{M}$ (namely, any bounded degree k-vertex $\Omega(1)$-expander graph works here) such that every embedding of $\mathcal{M}$ into $\mathbb{R}$ incurs p-average distortion that is at least a positive universal constant multiple of $\max\{1,(\log k)/p\}\asymp\max\{1,(\log n)/p\}$.

Our assumption on D therefore implies that

$$D \gtrsim \max\left\{1, \frac{\max\{1, \frac{\log n}{p}\}}{\max\{1, \frac{\log n}{q}\}}\right\} \asymp \begin{cases} \frac{q}{p} & \text{if } 1 \leqslant p, q \leqslant \log n, \\ 1 & \text{if } p, q \geqslant \log n \text{ or } q \leqslant \log n \leqslant p, \\ \frac{\log n}{p} & \text{if } 1 \leqslant p \leqslant \log n \leqslant q. \end{cases} \tag{341}$$

Next, denote the unit sphere of ℓ_q^n by $B_q^n = \{x \in \mathbb{R}^n : \|x\|_q \leqslant 1\}$. Let κ_q^n be the cone measure [GM87] on the sphere ∂B_q^n, i.e., $\kappa_q^n(E) = \mathrm{vol}_n([0,1]E)$ for every Lebesgue measurable $E \subseteq \partial B_q^n$. By our assumption on D, there exists $f : \mathbb{R}^n \to \mathbb{R}$ that is D-Lipschitz with respect to the ℓ_q^n-metric such that

$$\begin{aligned} \left(\iint_{\partial B_q^n \times \partial B_q^n} |f(x) - f(y)|^p \,\mathrm{d}\kappa_q^n(x)\,\mathrm{d}\kappa_q^n(y)\right)^{\frac{1}{p}} &\geqslant \left(\iint_{\partial B_q^n \times \partial B_q^n} \|x - y\|_q^p \,\mathrm{d}\kappa_q^n(x)\,\mathrm{d}\kappa_q^n(y)\right)^{\frac{1}{p}} \\ &\geqslant \left(\int_{\partial B_q^n} \left\|x - \int_{\partial B_q^n} y\,\mathrm{d}\kappa_q^n(y)\right\|_q^p \mathrm{d}\kappa_q^n(x)\right)^{\frac{1}{p}} = \left(\int_{\partial B_q^n} \|x\|_q^p \,\mathrm{d}\kappa_q^n(x)\right)^{\frac{1}{p}} = 1 \end{aligned} \tag{342}$$

where the second step of (342) uses Jensen's inequality. As f is D-Lipschitz with respect to the ℓ_q^n-metric and $q \geqslant 2$, by the Gromov–Milman concentration inequality [GM87] (see [AdRBV98] for a different simple proof; the case $p = 2$ follows from Lévy's classical isoperimetric theorem, as in e.g. [Led01, Chapter 2]) combined with the estimate on the modulus of uniform convexity of ℓ_q in [Cla36], we have

$$\forall s \geqslant 0, \qquad \kappa_q^n\Big[\big\{x \in \partial B_q^n : |f(x) - M_f| \geqslant s\big\}\Big] \lesssim e^{-n\left(\frac{s}{CD}\right)^q},$$

where M_f is the median of f and $C > 0$ is a universal constant. Consequently,

$$\forall s \geqslant 0, \qquad \big(\kappa_q^n \times \kappa_q^n\big)\Big[\big\{(x,y) \in \partial B_q^n \times \partial B_q^n : |f(x) - f(y)| \geqslant s\big\}\Big] \lesssim e^{-n\left(\frac{s}{CD}\right)^q}. \tag{343}$$

Hence,

$$\begin{aligned} 1 &\leqslant \left(\iint_{\partial B_q^n \times \partial B_q^n} |f(x) - f(y)|^p \,\mathrm{d}\kappa_q^n(x)\,\mathrm{d}\kappa_q^n(y)\right)^{\frac{1}{p}} \\ &= \left(\int_0^\infty p s^{p-1} \big(\kappa_q^n \times \kappa_q^n\big)\Big[\big\{(x,y) \in \partial B_q^n \times \partial B_q^n : |f(x) - f(y)| \geqslant s\big\}\Big] \,\mathrm{d}s\right)^{\frac{1}{p}} \\ &\lesssim \left(\int_0^\infty p s^{p-1} e^{-n\left(\frac{s}{CD}\right)^q} \,\mathrm{d}s\right)^{\frac{1}{p}} = C\frac{\Gamma\left(1 + \frac{p}{q}\right)^{\frac{1}{p}}}{n^{\frac{1}{q}}} D \asymp \left(\frac{p}{n}\right)^{\frac{1}{q}} D, \end{aligned} \tag{344}$$

where the first step of (344) uses (343), the third step of (344) uses (342), and the final step of (344) is an application of Stirling's formula. We therefore conclude that D must obey the following lower estimate:

$$D \gtrsim \left(\frac{n}{p}\right)^{\frac{1}{q}}. \tag{345}$$

The desired estimate (340) follows by combining (342) with (345), and a straightforward case analysis. □

Next, the case $r = 2$ of the following theorem coincides with Theorem 11:

Theorem 75. *Let $(\mathbf{X}, \|\cdot\|)$ be a Banach space of type $1 \leqslant r \leqslant 2$. Given $n \in \mathbb{N}$ and $p \geqslant 1$, if ℓ_1^n embeds into $\mathbf{X}$ with p-average distortion $D \geqslant 1$, then necessarily*

$$D \gtrsim \begin{cases} \frac{n^{1-\frac{1}{r}}}{T_r(\mathbf{X})} & \text{if } 1 \leqslant p \leqslant T_r(\mathbf{X})^2 n^{\frac{2}{r}-1}, \\ \sqrt{\frac{n}{p}} & \text{if } T_r(\mathbf{X})^2 n^{\frac{2}{r}-1} \leqslant p \leqslant n. \end{cases} \tag{346}$$

Theorem 76 below is the main ingredient in the proof of Theorem 75. Its formulation uses the convention (to which we will adhere later as well) that all the expectations of Banach space-valued functions whose domain is a finite set are with respect to the uniform probability measure on that set. Thus, if S is a finite set and $(\mathbf{X},\|\cdot\|_{\mathbf{X}})$ is a Banach space, then for every $f:S\to\mathbf{X}$ we will use the notations

$$\mathbb{E}[f]=\frac{1}{|S|}\sum_{s\in S}f(s)\in\mathbf{X}\qquad\text{and}\qquad\forall\, p\geqslant 1,\quad \|f\|_{L_p(S;\mathbf{X})}=\Big(\mathbb{E}\big[\|f\|_{\mathbf{X}}^p\big]\Big)^{\frac1p}=\bigg(\frac{1}{|S|}\sum_{s\in S}\|f(s)\|_{\mathbf{X}}^p\bigg)^{\frac1p}\geqslant 0.$$

Given $n\in\mathbb{N}$ and a Banach space $(\mathbf{X},\|\cdot\|_{\mathbf{X}})$, for $f:\{-1,1\}^n\to\mathbf{X}$ and $i\in[n]$ define $\partial_i f:\{-1,1\}^n\to\mathbf{X}$ by

$$\forall\,\varepsilon\in\{-1,1\}^n,\qquad \partial_i f(\varepsilon)\stackrel{\mathrm{def}}{=}f(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\ldots,\varepsilon_n)-f(\varepsilon). \tag{347}$$

Theorem 76. *Fix $p\geqslant 2\geqslant r\geqslant 1$ and $n\in\mathbb{N}$. Suppose that $(\mathbf{X},\|\cdot\|_{\mathbf{X}})$ is a Banach space of type r. Then, every $f:\{-1,1\}^n\to\mathbf{X}$ with $\mathbb{E}[f]=0$ satisfies the following estimate:*

$$\|f\|_{L_p(\{-1,1\}^n;\mathbf{X})}\lesssim T_r(\mathbf{X})\bigg(\sum_{i=1}^n\|\partial_i f\|_{L_r(\{-1,1\}^n;\mathbf{X})}^r\bigg)^{\frac1r}+\sqrt{p}\bigg(\sum_{i=1}^n\|\partial_i f\|_{L_p(\{-1,1\}^n;\mathbf{X})}^2\bigg)^{\frac12}. \tag{348}$$

The proof of Theorem 76 was found by Alexandros Eskenazis, answering positively a question that we posed in an earlier draft of the present work. We are grateful to him for allowing us to include it here.

Prior to proving Theorem 76, we will next show how it implies Theorem 75:

Proof of Theorem 75 assuming Theorem 76. It suffices to prove Theorem 75 when $p\geqslant 2$, because its conclusion when $1\leqslant p\leqslant 2$ follows formally from its special case $p=2$ by [Nao21a, equation (105)].

So, assume from now that $p\geqslant 2$ and suppose that $f:\{-1,1\}^n\to\mathbf{X}$ satisfies[39]

$$\Big(\mathbb{E}\big[\|f(\varepsilon)-f(\delta)\|_{\mathbf{X}}^p\big]\Big)^{\frac1p}\geqslant\Big(\mathbb{E}\big[\|\varepsilon-\delta\|_1^p\big]\Big)^{\frac1p}\asymp n, \tag{349}$$

and that also for some $D\geqslant 1$ we have

$$\forall\,\varepsilon,\delta\in\{-1,1\}^n,\qquad \|f(\varepsilon)-f(\delta)\|_{\mathbf{X}}\leqslant D\|\varepsilon-\delta\|_1=2D|\{i\in[n]:\ \varepsilon_i\neq\delta_i\}|, \tag{350}$$

i.e., f is D-Lipschitz with respect to the metric on $\{-1,1\}^n$ that is inherited from ℓ_1^n and the metric on $\mathbf{X}$ that is induced by $\|\cdot\|_{\mathbf{X}}$. Theorem 75 will be proven once we will show that D must satisfy (346), which is a simple consequence of (348). Indeed, by translating f we may assume that $\mathbb{E}[f]=0$. By Theorem 76,

$$\begin{aligned} n&\stackrel{(349)}{\lesssim}\Big(\mathbb{E}\big[\|f(\varepsilon)-f(\delta)\|_{\mathbf{X}}^p\big]\Big)^{\frac1p}\leqslant 2\|f\|_{L_p(\{-1,1\}^n;\mathbf{X})}\\ &\stackrel{(348)}{\lesssim}T_r(\mathbf{X})\bigg(\sum_{i=1}^n\|\partial_i f\|_{L_p(\{-1,1\}^n;\mathbf{X})}^r\bigg)^{\frac1r}+\sqrt{p}\bigg(\sum_{i=1}^n\|\partial_i f\|_{L_p(\{-1,1\}^n;\mathbf{X})}^2\bigg)^{\frac12}\stackrel{(347)\wedge(350)}{\leqslant}2D\Big(T_r(\mathbf{X})n^{\frac1r}+\sqrt{pn}\Big).\end{aligned}$$

Therefore,

$$D\gtrsim\frac{n}{T_r(\mathbf{X})n^{\frac1r}+\sqrt{pn}}\asymp\begin{cases}\frac{n^{1-\frac1r}}{T_r(\mathbf{X})} & \text{if } 1\leqslant p\leqslant T_r(\mathbf{X})^2n^{\frac2r-1},\\ \sqrt{\frac{n}{p}} & \text{if } T_r(\mathbf{X})^2n^{\frac2r-1}\leqslant p\leqslant n.\end{cases}$$ □

The proof of Theorem 76 is a quick concatenation of the (substantial) results of [BELP08] and [IvHV20]:

Proof of Theorem 76. The main result of [IvHV20] is well-known to yield the following estimate:

$$\|f\|_{L_r(\{-1,1\}^n;\mathbf{X})}\lesssim T_r(\mathbf{X})\bigg(\sum_{i=1}^n\|\partial_i f\|_{L_r(\{-1,1\}^n;\mathbf{X})}^r\bigg)^{\frac1r}. \tag{351}$$

[39]Note that the last equivalence in the assumption (349) has a quick justification: For one direction use the triangle inequality in $L_p(\{-1,1\}^n;\ell_1^n)$, and the other direction use Jensen's inequality for the averaging over δ.

(Specifically, deduce (351) from [IvHV20] by applying [IvHV20, Theorem 1.2] to the function $\Phi(x) = \|x\|_{\mathbf{X}}^r$, and then combining the resulting estimate with the second and third displayed equations on page 674 of [IvHV20].) Furthermore, by the (real-valued) Poincaré inequality of [BELP08, Theorem 1.1(1)], we have

$$\Big\| \|f\|_{\mathbf{X}} - \|f\|_{L_1(\{-1,1\}^n;\mathbf{X})} \Big\|_{L_p(\{-1,1\}^n)} \lesssim \sqrt{p} \bigg\| \Big(\sum_{i=1}^n \big(\partial_i \|f\|_{\mathbf{X}}\big)^2 \Big)^{\frac12} \bigg\|_{L_p(\{-1,1\}^n)}. \qquad (352)$$

Since $\|f\|_{L_1(\{-1,1\}^n;\mathbf{X})} \leqslant \|f\|_{L_r(\{-1,1\}^n;\mathbf{X})}$ and by the triangle inequality for $\|\cdot\|_{\mathbf{X}}$ we have $\partial_i \|f\|_{\mathbf{X}}(\varepsilon) \leqslant \|\partial_i f(\varepsilon)\|_{\mathbf{X}}$ for every $\varepsilon \in \{-1,1\}^n$ and every $i \in [n]$, it follows from (351) and (352) that

$$\begin{aligned} \|f\|_{L_p(\{-1,1\}^n;\mathbf{X})} &\leqslant \Big\| \|f\|_{\mathbf{X}} - \|f\|_{L_1(\{-1,1\}^n;\mathbf{X})} \Big\|_{L_p(\{-1,1\}^n)} + \|f\|_{L_1(\{-1,1\}^n;\mathbf{X})} \\ &\lesssim T_r(\mathbf{X}) \bigg(\sum_{i=1}^n \|\partial_i f\|_{L_r(\{-1,1\}^n;\mathbf{X})}^r \bigg)^{\frac1r} + \sqrt{p} \bigg\| \Big(\sum_{i=1}^n \big(\|\partial_i f\|_{\mathbf{X}}\big)^2 \Big)^{\frac12} \bigg\|_{L_p(\{-1,1\}^n)}, \end{aligned} \qquad (353)$$

where the first step of (353) is an application of the triangle inequality in $L_p(\{-1,1\}^n)$. It remains to note that final term in (353) is at most the second term in (348) by the triangle inequality in $L_{\frac{p}{2}}(\{-1,1\}^n)$. □

References

[ACMM05] A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. $O(\sqrt{\log n})$ approximation algorithms for Min UnCut, Min 2CNF deletion, and directed cut problems. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 573–581. ACM, New York, 2005. ISBN 1-58113-960-8. doi:10.1145/1060590.1060675.

[AdRBV98] J. Arias-de Reyna, K. Ball, and R. Villa. Concentration of the distance in finite-dimensional normed spaces. *Mathematika*, 45(2):245–252, 1998.

[AIR18] A. Andoni, P. Indyk, and I. Razenshteyn. Approximate nearest neighbor search in high dimensions. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited lectures*, pages 3287–3318. World Sci. Publ., Hackensack, NJ, 2018. ISBN 978-981-3272-93-4; 978-981-3272-87-3.

[AK16] F. Albiac and N. J. Kalton. *Topics in Banach space theory*, volume 233 of *Graduate Texts in Mathematics*. Springer, [Cham], second edition, 2016. ISBN 978-3-319-31555-3; 978-3-319-31557-7. doi:10.1007/978-3-319-31557-7. With a foreword by Gilles Godefory.

[ALN05] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut [extended abstract]. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 553–562. ACM, New York, 2005. ISBN 1-58113-960-8. doi:10.1145/1060590.1060673.

[ALN07] S. Arora, J. R. Lee, and A. Naor. Fréchet embeddings of negative type metrics. *Discrete Comput. Geom.*, 38(4):726–739, 2007.

[ALN08] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21(1):1–21, 2008.

[Alo86] N. Alon. Eigenvalues and expanders. volume 6, pages 83–96. 1986. doi:10.1007/BF02579166. Theory of computing (Singer Island, Fla., 1984).

[ALS11] S. Arora, J. R. Lee, and S. Sachdeva. A reformulation of the Arora–Rao–Vazirani structure theorem, 2011. Unpublished manuscript, available at `https://arxiv.org/abs/1102.1456`.

[AM85] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[ANN+18a] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Data-dependent hashing via nonlinear spectral gaps. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 787–800. ACM, New York, 2018. ISBN 978-1-4503-5559-9. doi:10.1145/3188745.3188846.

[ANN+18b] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Hölder homeomorphisms and approximate nearest neighbors. In *59th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2018*, pages 159–169. IEEE Computer Soc., Los Alamitos, CA, 2018. ISBN 978-1-5386-4230-6. doi:10.1109/FOCS.2018.00024.

[ANV10] T. Austin, A. Naor, and A. Valette. The Euclidean distortion of the lamplighter group. *Discrete Comput. Geom.*, 44(1):55–74, 2010.

[AR98] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301, 1998.

[ARV04] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 222–231. ACM, New York, 2004. ISBN 1-58113-852-0. doi:10.1145/1007352.1007355.

[ARV09] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):1–37, 2009.

[Ass83] P. Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[BA56] A. Beurling and L. Ahlfors. The boundary correspondence under quasiconformal mappings. *Acta Math.*, 96:125–142, 1956.

[Bal90] K. Ball. Isometric embedding in l_p-spaces. *European J. Combin.*, 11(4):305–311, 1990.

[Bal92] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[Bal13] K. Ball. The Ribe programme. 352, pages Exp. No. 1047, viii, 147–159. 2013. ISBN 978-2-85629-371-3. Séminaire Bourbaki. Vol. 2011/2012. Exposés 1043–1058.

[BB12a] A. Brudnyi and Y. Brudnyi. *Methods of geometric analysis in extension and trace problems. Volume 1*, volume 102 of *Monographs in Mathematics*. Birkhäuser/Springer Basel AG, Basel, 2012. ISBN 978-3-0348-0208-6.

[BB12b] A. Brudnyi and Y. Brudnyi. *Methods of geometric analysis in extension and trace problems. Volume 2*, volume 103 of *Monographs in Mathematics*. Birkhäuser/Springer Basel AG, Basel, 2012. ISBN 978-3-0348-0211-6.

[BDCK66] J. Bretagnolle, D. Dacunha-Castelle, and J.-L. Krivine. Lois stables et espaces L^p. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 2:231–259, 1965/66.

[BDLS13] Y. Bilu, A. Daniely, N. Linial, and M. Saks. On the practically interesting instances of MAXCUT. In *30th International Symposium on Theoretical Aspects of Computer Science*, volume 20 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 526–537. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2013. ISBN 978-3-939897-50-7.

[BELP08] L. Ben Efraim and F. Lust-Piquard. Poincaré type inequalities on the discrete cube and in the CAR algebra. *Probab. Theory Related Fields*, 141(3-4):569–602, 2008.

[Ber00] M. Berger. Encounter with a geometer. II. *Notices Amer. Math. Soc.*, 47(3):326–340, 2000.

[BF22] J. Bertrand and M. Fathi. Stability of eigenvalues and observable diameter in RCD$(1,\infty)$ spaces. *J. Geom. Anal.*, 32(11):Paper No. 270, 38, 2022.

[BHT00] S. Bobkov, C. Houdré, and P. Tetali. λ_∞, vertex isoperimetry and concentration. *Combinatorica*, 20(2):153–172, 2000.

[BL00] 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. ISBN 0-8218-0835-4. doi:10.1090/coll/048.

[BL12] Y. Bilu and N. Linial. Are stable instances easy? *Combin. Probab. Comput.*, 21(5):643–660, 2012.

[BLMN06] Y. Bartal, N. Linial, M. Mendel, and A. Naor. Limitations to Fréchet's metric embedding method. *Israel J. Math.*, 151:111–124, 2006.

[BN24] M. Braverman and A. Naor. Quantitative Wasserstein rounding, 2024. Preprint.

[Bor75] C. Borell. The brunn-minkowski inequality in gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975.

[Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[Bou86] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986.

[CGR05] S. Chawla, A. Gupta, and H. Räcke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 102–111. ACM, New York, 2005. ISBN 0-89871-585-7.

[CGR08] S. Chawla, A. Gupta, and H. Räcke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. *ACM Transactions on Algorithms (TALG)*, 4(2):1–18, 2008.

[Cha14] S. Chatterjee. *Superconcentration and related topics*. Springer Monographs in Mathematics. Springer, Cham, 2014. ISBN 978-3-319-03885-8; 978-3-319-03886-5. doi:10.1007/978-3-319-03886-5.

[Chu89] F. R. K. Chung. Diameters and eigenvalues. *J. Amer. Math. Soc.*, 2(2):187–196, 1989.

[CK06] 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. ISBN 978-981-270-062-9; 981-270-062-5. doi:10.1142/9789812772688_0006.

[CK10a] 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.

[CK10b] J. Cheeger and B. Kleiner. Metric differentiation, monotonicity and maps to L^1. *Invent. Math.*, 182(2):335–370, 2010.

[CKN11] 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.

[CKO16] I. Choi, J. Kim, and S. O. The difference and ratio of the fractional matching number and the matching number of graphs. *Discrete Math.*, 339(4):1382–1386, 2016.

[Cla36] J. A. Clarkson. Uniformly convex spaces. *Trans. Amer. Math. Soc.*, 40(3):396–414, 1936.

[CLLM10] M. Charikar, T. Leighton, S. Li, and A. Moitra. Vertex sparsifiers and abstract rounding algorithms. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010*, pages 265–274. IEEE Computer Soc., Los Alamitos, CA, 2010. ISBN 978-0-7695-4244-7.

[CMN19] Ş. Cobzaş, R. Miculescu, and A. Nicolae. *Lipschitz functions*, volume 2241 of *Lecture Notes in Mathematics*. Springer, Cham, 2019. ISBN 978-3-030-16488-1; 978-3-030-16489-8. doi:10.1007/978-3-030-16489-8.

[CW71] R. R. Coifman and G. Weiss. *Analyse harmonique non-commutative sur certains espaces homogènes*, volume Vol. 242 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin-New York, 1971. Étude de certaines intégrales singulières.

[DL10] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer, Heidelberg, 2010. ISBN 978-3-642-04294-2. doi:10.1007/978-3-642-04295-9. First softcover printing of the 1997 original [MR1460488].

[dLdlS21] T. de Laat and M. de la Salle. Banach space actions and L^2-spectral gap. *Anal. PDE*, 14(1):45–76, 2021.

[DS89] S. K. Donaldson and D. P. Sullivan. Quasiconformal 4-manifolds. *Acta Math.*, 163(3-4):181–252, 1989.

[DS97] 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. ISBN 0-19-850166-8. Self-similar geometry through metric and measure.

[Dur19] R. Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fifth edition, 2019. ISBN 978-1-108-47368-2. doi:10.1017/9781108591034.

[Dvo61] 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.

[EGK+10] M. Englert, A. Gupta, R. Krauthgamer, H. Räcke, I. Talgam-Cohen, and K. Talwar. Vertex sparsifiers: new results from old techniques. In *Approximation, randomization, and combinatorial optimization*, volume 6302 of *Lecture Notes in Comput. Sci.*, pages 152–165. Springer, Berlin, 2010. ISBN 978-3-642-15368-6; 3-642-5368-2. doi:10.1007/978-3-642-15369-3_12.

[Enf69] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105, 1969.

[Esk22] A. Eskenazis. Average distortion embeddings, nonlinear spectral gaps, and a metric john theorem. *Astérisque*, (438):No. 1188, 333, 2022.

[FHL08] U. Feige, M. Hajiaghayi, and J. R. Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM J. Comput.*, 38(2):629–657, 2008.

[Fil23] A. Filtser. Labeled nearest neighbor search and metric spanners via locality sensitive orderings. In *39th International Symposium on Computational Geometry*, volume 258 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 33, 18. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2023. ISBN 978-3-95977-273-0. doi:10.4230/lipics.socg.2023.33.

[FJS88] T. Figiel, W. B. Johnson, and G. Schechtman. Factorizations of natural embeddings of l_p^n into L_r. I. *Studia Math.*, 89(1):79–103, 1988.

[FLM77] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.

[Für81] Z. Füredi. Maximum degree and fractional matchings in uniform hypergraphs. *Combinatorica*, 1(2):155–162, 1981.

[GK95] E. D. Gluskin and S. Kwapień. Tail and moment estimates for sums of independent random variables with logarithmically concave tails. *Studia Math.*, 114(3):303–309, 1995.

[GKL03] 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. IEEE Computer Society, 2003. doi:10.1109/SFCS.2003.1238226.

[GM83] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.*, 105(4):843–854, 1983.

[GM87] M. Gromov and V. D. Milman. Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces. *Compositio Math.*, 62(3):263–282, 1987.

[God17] G. Godefroy. De Grothendieck à Naor: une promenade dans l'analyse métrique des espaces de Banach. *Gaz. Math.*, (151):13–24, 2017.

[Goe97] M. X. Goemans. Semidefinite programming in combinatorial optimization. volume 79, pages 143–161. 1997. doi:10.1016/S0025-5610(97)00051-8. Lectures on mathematical programming (ismp97) (Lausanne, 1997).

[Gro80] M. Gromov. Paul Lévy's isoperimetric inequality, 1980. Preprint I.H.E.S., available at `https://www.ihes.fr/~gromov/wp-content/uploads/2018/08/1133.pdf`.

[Gro07] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA, english edition, 2007. ISBN 978-0-8176-4582-3; 0-8176-4582-9. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.

[GW94] M. X. Goemans and D. P. Williamson. .879-approximation algorithms for MAX CUT and MAX 2sat. In F. T. Leighton and M. T. Goodrich, editors, *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 422–431. ACM, 1994. doi:10.1145/195058.195216.

[Hås01] J. Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.

[Hei01] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001. ISBN 0-387-95104-0. doi:10.1007/978-1-4613-0131-8.

[Hit93] P. Hitczenko. Domination inequality for martingale transforms of a Rademacher sequence. *Israel J. Math.*, 84(1-2):161–178, 1993.

[HJ74] J. Hoffmann-Jørgensen. Sums of independent Banach space valued random variables. *Studia Math.*, 52:159–186, 1974.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.

[HPIM12] S. Har-Peled, P. Indyk, and R. Motwani. Approximate nearest neighbor: towards removing the curse of dimensionality. *Theory Comput.*, 8:321–350, 2012.

[HPM06] 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.

[IM04] P. Indyk and J. Matoušek. Low-distortion embeddings of finite metric spaces. In J. E. Goodman and J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry, Second Edition*, pages 177–196. Chapman and Hall/CRC, 2004. doi:10.1201/9781420035315.CH8.

[IN07] P. Indyk and A. Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):Art. 31, 12, 2007.

[IvHV20] P. Ivanisvili, R. van Handel, and A. Volberg. Rademacher type and Enflo type coincide. *Ann. of Math. (2)*, 192(2):665–678, 2020.

[JL84] 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. ISBN 0-8218-5030-X. doi:10.1090/conm/026/737400.

[JLS86] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. Extensions of Lipschitz maps into Banach spaces. *Israel J. Math.*, 54(2):129–138, 1986.

[Joh48] 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, New York, 1948.

[JS88] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for markov chains: the approximation of the permanent resolved (preliminary version). In J. Simon, editor, *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 235–244. ACM, 1988. doi:10.1145/62212.62234.

[Kal04] N. J. Kalton. Spaces of Lipschitz and Hölder functions and their applications. *Collect. Math.*, 55(2):171–217, 2004.

[Kal08] N. J. Kalton. The nonlinear geometry of Banach spaces. *Rev. Mat. Complut.*, 21(1):7–60, 2008.

[Kal12] N. J. Kalton. The uniform structure of Banach spaces. *Math. Ann.*, 354(4):1247–1288, 2012.

[Kar05] G. Karakostas. A better approximation ratio for the vertex cover problem. In L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi, and M. Yung, editors, *Automata, Languages and Programming, 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005, Proceedings*, volume 3580 of *Lecture Notes in Computer Science*, pages 1043–1050. Springer, 2005. doi:10.1007/11523468_84.

[Kir34] M. D. Kirszbraun. Über die zusammenziehenden und Lipschitzchen Transformationen. *Fund. Math.*, 22:77–108, 1934.

[KKMO07] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.

[KL05] R. Krauthgamer and J. R. Lee. The black-box complexity of nearest-neighbor search. *Theoret. Comput. Sci.*, 348(2-3):262–276, 2005.

[KLMN04] R. Krauthgamer, J. Lee, M. Mendel, and A. Naor. Measured descent: a new embedding method for finite metrics [extended abstract]. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 434–443. 2004. doi:10.1109/FOCS.2004.41.

[KLMN05] 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.

[KM13] D. Kane and R. Meka. A PRG for Lipschitz functions of polynomials with applications to sparsest cut. In *STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing*, pages 1–10. ACM, New York, 2013. ISBN 978-1-4503-2029-0. doi:10.1145/2488608.2488610.

[KNT21] D. Kush, A. Nikolov, and H. Tang. Near neighbor search via efficient average distortion embeddings. In *37th International Symposium on Computational Geometry*, volume 189 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 50, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021. ISBN 978-3-95977-184-9.

[Kom55] Y. Komatu. Elementary inequalities for Mills' ratio. *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs.*, 4:69–70, 1955.

[KPR93] P. N. Klein, S. A. Plotkin, and S. Rao. Excluded minors, network decomposition, and multicommodity flow. In S. R. Kosaraju, D. S. Johnson, and A. Aggarwal, editors, *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing, May 16-18, 1993, San Diego, CA, USA*, pages 682–690. ACM, 1993. doi:10.1145/167088.167261.

[KR09] R. Krauthgamer and Y. Rabani. Improved lower bounds for embeddings into L_1. *SIAM J. Comput.*, 38(6):2487–2498, 2009.

[KV05] S. Khot and N. K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into l_1. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 53–62. IEEE Computer Society, 2005. doi:10.1109/SFCS.2005.74.

[KV15] S. A. Khot and N. K. Vishnoi. The unique games conjecture, integrability gap for cut problems and embeddability of negative-type metrics into ℓ_1. *J. ACM*, 62(1):Art. 8, 39, 2015.

[Laa00] T. J. Laakso. Ahlfors Q-regular spaces with arbitrary $Q > 1$ admitting weak Poincaré inequality. *Geom. Funct. Anal.*, 10(1):111–123, 2000.

[Laf08] V. Lafforgue. Un renforcement de la propriété (T). *Duke Math. J.*, 143(3):559–602, 2008.

[Laf09] V. Lafforgue. Propriété (T) renforcée banachique et transformation de Fourier rapide. *J. Topol. Anal.*, 1(3):191–206, 2009.

[Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. ISBN 0-8218-2864-9. doi:10.1090/surv/089.

[Lee05] J. R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 92–101. SIAM, 2005.

[Lév51] P. Lévy. *Problèmes concrets d'analyse fonctionnelle. Avec un complément sur les fonctionnelles analytiques par F. Pellegrino*. Gauthier-Villars, Paris, 1951. 2d ed.

[Lin02] 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. ISBN 7-04-008690-5.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[LM10] F. T. Leighton and A. Moitra. Extensions and limits to vertex sparsification. In *STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 47–56. ACM, New York, 2010. ISBN 978-1-60558-817-9.

[LN04] J. R. Lee and A. Naor. Absolute Lipschitz extendability. *C. R. Math. Acad. Sci. Paris*, 338(11):859–862, 2004.

[LN05] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[LN06] J. R. Lee and A. Naor. L_p metrics on the Heisenberg group and the Goemans–Linial conjecture. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 99–108. IEEE Computer Society, 2006. doi:10.1109/FOCS.2006.47.

[Lov74] L. Lovász. Minimax theorems for hypergraphs. In *Hypergraph Seminar (Proc. First Working Sem., Ohio State Univ., Columbus, Ohio, 1972; dedicated to Arnold Ross)*, volume Vol. 411 of *Lecture Notes in Math.*, pages 111–126. Springer, Berlin-New York, 1974.

[LP01] U. Lang and C. Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.

[LP09] L. Lovász and M. D. Plummer. *Matching theory*. AMS Chelsea Publishing, Providence, RI, 2009. ISBN 978-0-8218-4759-6. doi:10.1090/chel/367. Corrected reprint of the 1986 original [MR0859549].

[LS88] G. F. Lawler and A. D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309(2):557–580, 1988.

[LS98] J. Luukkainen and E. Saksman. Every complete doubling metric space carries a doubling measure. *Proc. Amer. Math. Soc.*, 126(2):531–534, 1998.

[LS05] U. Lang and T. Schlichenmaier. Nagata dimension, quasisymmetric embeddings, and Lipschitz extensions. *Int. Math. Res. Not.*, (58):3625–3655, 2005.

[LT77] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. I*, volume Band 92 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin-New York, 1977. ISBN 3-540-08072-4. Sequence spaces.

[Mat90] J. Matoušek. Extension of Lipschitz mappings on metric trees. *Comment. Math. Univ. Carolin.*, 31(1):99–104, 1990.

[Mat97] J. Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[Mat02] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. ISBN 0-387-95373-6. doi:10.1007/978-1-4613-0039-7.

[Mau79] B. Maurey. Construction de suites symétriques. *C. R. Acad. Sci. Paris Sér. A-B*, 288(14):A679–A681, 1979.

[Mil99] P. B. Miltersen. Cell probe complexity-a survey. In *Proceedings of the 19th conference on the foundations of software technology and theoretical computer science, advances in data structures workshop*. 1999.

[MM16] K. Makarychev and Y. Makarychev. Metric extension operators, vertex sparsifiers and Lipschitz extendability. *Israel J. Math.*, 212(2):913–959, 2016.

[MMV14] K. Makarychev, Y. Makarychev, and A. Vijayaraghavan. Bilu-Linial stable instances of max cut and minimum multiway cut. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 890–906. ACM, New York, 2014. ISBN 978-1-611973-38-9. doi:10.1137/1.9781611973402.67.

[MN04] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Adv. Math.*, 189(2):451–494, 2004.

[MN06] M. Mendel and A. Naor. Some applications of Ball's extension theorem. *Proc. Amer. Math. Soc.*, 134(9):2577–2584, 2006.

[MN07] M. Mendel and A. Naor. Ramsey partitions and proximity data structures. *J. Eur. Math. Soc. (JEMS)*, 9(2):253–275, 2007.

[MN13] M. Mendel and A. Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. *Anal. Geom. Metr. Spaces*, 1:163–199, 2013.

[MN14] M. Mendel and A. Naor. Nonlinear spectral calculus and super-expanders. *Publ. Math. Inst. Hautes Études Sci.*, 119:1–95, 2014.

[Moi09] A. Moitra. Approximation algorithms for multicommodity-type problems with guarantees independent of the graph size. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2009*, pages 3–12. IEEE Computer Soc., Los Alamitos, CA, 2009. ISBN 978-0-7695-3850-1. doi:10.1109/FOCS.2009.28.

[MOO10] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. of Math. (2)*, 171(1):295–341, 2010.

[MP73] 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.

[MP76] 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.

[MR01] J. Matoušek and Y. Rabinovich. On dominated l_1 metrics. *Israel J. Math.*, 123:285–301, 2001.

[MS86] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. ISBN 3-540-16769-2. With an appendix by M. Gromov.

[Nao10] A. Naor. L_1 embeddings of the Heisenberg group and fast estimation of graph isoperimetry. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1549–1575. Hindustan Book Agency, New Delhi, 2010. ISBN 978-81-85931-08-3; 978-981-4324-33-5; 981-4324-33-7.

[Nao12a] 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. ISBN 978-3-0348-0256-7; 978-3-0348-0257-4. doi:10.1007/978-3-0348-0257-4_7.

[Nao12b] A. Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.

[Nao14] A. Naor. Comparison of metric spectral gaps. *Anal. Geom. Metr. Spaces*, 2(1):1–52, 2014.

[Nao18] A. Naor. Metric dimension reduction: a snapshot of the Ribe program. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. I. Plenary lectures*, pages 759–837. World Sci. Publ., Hackensack, NJ, 2018. ISBN 978-981-3272-90-3; 978-981-3272-87-3.

[Nao21a] A. Naor. An average John theorem. *Geom. Topol.*, 25(4):1631–1717, 2021.

[Nao21b] A. Naor. Impossibility of almost extension. *Adv. Math.*, 384:Paper No. 107761, 34, 2021.

[Nao24a] A. Naor. *Extension, separation and isomorphic reverse isoperimetry*, volume 11 of *Memoirs of the European Mathematical Society*. European Mathematical Society (EMS), Berlin, 2024. ISBN 978-3-98547-069-3; 978-3-98547-569-8. doi:10.4171/mems/11.

[Nao24b] A. Naor. Mixing convexity, 2024. Preprint.

[NR03] I. Newman and Y. Rabinovich. A lower bound on the distortion of embedding planar metrics into Euclidean space. *Discrete Comput. Geom.*, 29(1):77–81, 2003.

[NR17] A. Naor and Y. Rabani. On Lipschitz extension from finite subsets. *Israel J. Math.*, 219(1):115–161, 2017.

[NRS05] A. Naor, Y. Rabani, and A. Sinclair. Quasisymmetric embeddings, the observable diameter, and expansion properties of graphs. *Journal of Functional Analysis*, 227(2):273–303, 2005.

[NS11] A. Naor and L. Silberman. Poincaré inequalities, embeddings, and wild groups. *Compos. Math.*, 147(5):1546–1572, 2011.

[NY18] A. Naor and R. Young. Vertical perimeter versus horizontal perimeter. *Annals of Mathematics*, 188(1):171–279, 2018.

[NY22] A. Naor and R. Young. Foliated corona decompositions. *Acta Math.*, 229(1):55–200, 2022.

[Ost13] M. I. Ostrovskii. *Metric embeddings*, volume 49 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2013. ISBN 978-3-11-026340-4; 978-3-11-026401-2. doi:10.1515/9783110264012. Bilipschitz and coarse embeddings into Banach spaces.

[Pan89] P. Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math. (2)*, 129(1):1–60, 1989.

[Pis89] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989. ISBN 0-521-36465-5; 0-521-66635-X. doi:10.1017/CBO9780511662454.

[PY91] C. H. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. *J. Comput. System Sci.*, 43(3):425–440, 1991.

[Rab08] Y. Rabinovich. On average distortion of embedding metrics into the line. *Discrete Comput. Geom.*, 39(4):720–733, 2008.

[Rao99] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry (Miami Beach, FL, 1999)*, pages 300–306. ACM, New York, 1999. doi:10.1145/304893.304983.

[Rot16] T. Rothvoss. Lecture notes on the ARV algorithm for sparsest cut, 2016. Unpublished lecture notes, available at `https://arxiv.org/pdf/1607.00854`.

[RR91] S. T. Rachev and L. Rüschendorf. Approximate independence of distributions on spheres and their stability properties. *Ann. Probab.*, 19(3):1311–1337, 1991.

[Sem96] S. Semmes. On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞-weights. *Rev. Mat. Iberoamericana*, 12(2):337–410, 1996.

[Sem01] S. Semmes. *Some novel types of fractal geometry*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2001. ISBN 0-19-850806-9.

[Shi16] T. Shioya. *Metric measure geometry*, volume 25 of *IRMA Lectures in Mathematics and Theoretical Physics*. EMS Publishing House, Zürich, 2016. ISBN 978-3-03719-158-3. doi:10.4171/158. Gromov's theory of convergence and concentration of metrics and measures.

[Shm96] D. B. Shmoys. Cut problems and their application to divide and conquer. In D. S. Hochbaum, editor, *Approximation algorithms for NP-hard problems*, chapter 5, pages 192–235. Boston, MA: PWS Publishing, 1996. ISBN 0-534-94968-1.

[Sim95] S. Simons. Minimax theorems and their proofs. In *Minimax and applications*, volume 4 of *Nonconvex Optim. Appl.*, pages 1–23. Kluwer Acad. Publ., Dordrecht, 1995. ISBN 0-7923-3615-1. doi:10.1007/978-1-4613-3557-3_1.

[ST78] V. N. Sudakov and B. S. Tsirel'son. Extremal properties of half-spaces for spherically invariant measures. *Journal of Soviet Mathematics*, 9(1):9–18, 1978.

[SZ90] G. Schechtman and J. Zinn. On the volume of the intersection of two L_p^n balls. *Proc. Amer. Math. Soc.*, 110(1):217–224, 1990.

[TV80] P. Tukia and J. Väisälä. Quasisymmetric embeddings of metric spaces. *Ann. Acad. Sci. Fenn. Ser. A I Math.*, 5(1):97–114, 1980.

[Väi99] 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.

[VG05] J.-L. Verger-Gaugry. Covering a ball with smaller equal balls in $\mathbb{R}^n$. *Discrete Comput. Geom.*, 33(1):143–155, 2005.

[VK84] A. L. Vol'berg and S. V. Konyagin. A homogeneous measure exists on any compactum in $\mathbf{R}^n$. *Dokl. Akad. Nauk SSSR*, 278(4):783–786, 1984.

[vN28] J. v. Neumann. Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100(1):295–320, 1928.

[Wag77] D. H. Wagner. Survey of measurable selection theorems. *SIAM J. Control Optim.*, 15(5):859–903, 1977.

[WW75] J. H. Wells and L. R. Williams. *Embeddings and extensions in analysis*, volume Band 84 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, New York-Heidelberg, 1975.

[Yao81] A. C. C. Yao. Should tables be sorted? *J. Assoc. Comput. Mach.*, 28(3):615–628, 1981.

DEPARTMENT OF MATHEMATICS, WASHINGTON UNIVERSITY IN ST. LOUIS, ST. LOUIS MO 63130-4899
Email address: `alanchang@math.wustl.edu`

DEPARTMENT OF MATHEMATICS, PRINCETON NJ 08544-1000
Email address: `naor@math.princeton.edu`

DEPARTMENT OF MATHEMATICS, PRINCETON NJ 08544-1000
Email address: `kr5621@princeton.edu`