AN AVERAGE JOHN THEOREM

ASSAF NAOR

Dedicated to the memory of Eli Stein

Abstract. We prove that the $\frac{1}{2}$-snowflake of any finite-dimensional normed space X embeds into a Hilbert space with quadratic average distortion

$$O\Big(\sqrt{\log \dim(X)}\Big).$$

We deduce from this (optimal) statement that if an n-vertex expander embeds with average distortion $D \geqslant 1$ into X, then necessarily $\dim(X) \geqslant n^{\Omega(1/D)}$, which is sharp by the work of Johnson, Lindenstrauss and Schechtman (1987). This improves over the previously best-known bound $\dim(X) \gtrsim (\log n)^2/D^2$ of Linial, London and Rabinovich (1995), strengthens a theorem of Matoušek (1996) which resolved questions of Johnson and Lindenstrauss (1982), Bourgain (1985) and Arias-de-Reyna and Rodríguez-Piazza (1992), and answers negatively a question that was posed (for algorithmic purposes) by Andoni, Nguyen, Nikolov, Razenshteyn and Waingarten (2016).

1. Introduction

Given $D \geqslant 1$, we say that an infinite metric space $(\mathcal{M}, d_{\mathcal{M}})$ embeds into a normed space $(Z, \|\cdot\|_Z)$ with quadratic average distortion D if for *every* Borel probability measure μ on $\mathcal{M}$ there exists a D-Lipschitz mapping $f = f_\mu : \mathcal{M} \to Z$ that satisfies

$$\iint_{\mathcal{M}\times\mathcal{M}} \|f(x) - f(y)\|_Z^2 \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant \iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^2 \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y). \tag{1}$$

In comparison, the requirement that $(\mathcal{M}, d_{\mathcal{M}})$ embeds with bi-Lipschitz distortion D into $(Z, \|\cdot\|_Z)$ means that there exists a D-Lipschitz mapping $f : \mathcal{M} \to Z$ that satisfies

$$\forall\, x, y \in \mathcal{M}, \qquad \|f(x) - f(y)\|_Z \geqslant d_{\mathcal{M}}(x,y). \tag{2}$$

Thus (1) is a natural average-case counterpart to the worst-case condition (2) where, in lieu of a canonical probability measure on $\mathcal{M}$, one demands that the notion of "average" is with respect to any Borel probability measure on $\mathcal{M}$ while allowing the embedding to depend on the given measure.

The following theorem is (a special case of) our main result. Its statement uses the terminology (e.g [44]) that for $\omega \in (0,1]$, the ω-snowflake of a metric space $(\mathcal{M}, d_{\mathcal{M}})$ is the metric space $(\mathcal{M}, d_{\mathcal{M}}^{\omega})$.

Theorem 1. *For every integer $k \geqslant 2$, the $\frac{1}{2}$-snowflake of any k-dimensional normed space embeds into a Hilbert space with quadratic average distortion $\mathsf{C}\sqrt{\log k}$, where $\mathsf{C} > 0$ is a universal constant.*

Compare Theorem 1 with John's classical theorem [67] that X embeds into a Hilbert space with bi-Lipschitz[1] distortion $\sqrt{k}$. This is sharp, as exhibited by $X = \ell_\infty^k$ or $X = \ell_1^k$. Power-type behavior is necessary also for bi-Lipschitz embeddings of snowflaked norms, as shown by the following lemma.

Lemma 2. *Fix $\omega \in (0,1]$ and $k \in \mathbb{N}$. The ω-snowflake of any k-dimensional normed space embeds with bi-Lipschitz distortion $k^{\frac{\omega}{2}}$ into a Hilbert space. Conversely, any embedding of the ω-snowflake of ℓ_∞^k into a Hilbert space incurs bi-Lipschitz distortion at least a universal constant multiple of $k^{\frac{\omega}{2}}$.*

Supported by the Packard Foundation and the Simons Foundation. The research that is presented here was conducted under the auspices of the Simons Algorithms and Geometry (A&G) Think Tank. An extended abstract [110] announcing parts of this work appeared in the 33rd International Symposium on Computational Geometry.

[1] John's theorem is often stated in the literature with the embedding being a linear transformation, but this is equivalent to the way we stated it by passing to a derivative of the embedding, which exists almost everywhere [54, 26].

In contrast to Lemma 2, in Theorem 1 we establish that if one wishes to obtain an embedding into a Hilbert space which is $\frac{1}{2}$-Hölder and preserves the $\frac{1}{2}$-snowflaked distances only on average rather than the worst-case pairwise distance preservation requirement of John's theorem (or its version for snowflakes that appears in Lemma 2, which shows that for $\frac{1}{2}$-snowflakes the best bi-Lipschitz distortion that one could hope for is of order $\sqrt[4]{k}$), then the distortion can be improved dramatically to a universal constant multiple of $\sqrt{\log k}$. Importantly, the notion of "average" here can be taken to be with respect to any Borel probability measure on X whatsoever.

Theorem 1 is sharp in two ways. Firstly, we will see that its $\mathsf{C}\sqrt{\log k}$ bound is sharp (this occurs when $X = \ell_\infty^k$ and the probability measure is uniform over an isometrically embedded k-vertex expander). Secondly, one cannot perform a lesser amount of snowflaking of the norm while still obtaining average distortion $k^{o(1)}$. Namely, we will see that if $\varepsilon \in (0, \frac{1}{2}]$ and one aims to embed the $\left(\frac{1}{2}+\varepsilon\right)$-snowflake of every such X into a Hilbert space with quadratic average distortion D, then necessarily $D \geqslant k^\varepsilon$ (this occurs when $X = \ell_1^k$ and the probability measure is uniform over $\{0,1\}^k$).

Thus, the exponential improvement over John's distortion bound that we obtain in Theorem 1 is made possible by allowing the distances to be preserved only on average, and simultaneously introducing an inherent nonlinearity through snowflaking; performing only one of these two modifications of John's theorem does not suffice. We will soon see that, despite the fact that the distance preservation guarantee that is furnished by Theorem 1 is (necessarily) weaker than that of John's theorem, it has quite substantial implications. It is also worthwhile to note that unlike John's embedding, which is achieved explicitly by considering the ellipsoid of maximal volume that is inscribed in the unit ball of X, our proof of Theorem 1 establishes the existence of the stated embedding implicitly through reliance on a duality argument; obtaining a more constructive proof would be valuable.

Theorem 1 is in fact a special case of a stronger theorem that treats embeddings into targets that are not necessarily Hilbertian, L_p variants of the quadratic requirement (1), and other snowflakes of X, and it also obtains improved embeddings (i.e. with less snowflaking) if X satisfies an additional geometric assumption; see Theorem 12 below. It is beneficial to start by presenting the above basic version (quadratic, Hilbertian, without any assumption on the geometry of X) because it does not require the introduction of further terminology, and it has a noteworthy geometric consequence that we wish to explain first, prior to passing to the somewhat more involved setup of Section 1.3 below.

Remark 3. In the spirit of Theorem 1, it is simple to find other examples of metric spaces $(\mathcal{M}, d_\mathcal{M})$ whose quadratic average distortion into some Banach space Z is significantly smaller than their bi-Lipschitz distortion into Z. Indeed, it is straightforward to check that if $(\mathcal{M}, d_\mathcal{M})$ is an infinite equilateral space, i.e., $d_\mathcal{M}(x,y) = 1$ for all distinct $x, y \in \mathcal{M}$, then $\mathcal{M}$ embeds into $Z = \mathbb{R}$ with finite quadratic average distortion, but $\mathcal{M}$ does not admit a bi-Lipschitz embedding into $\mathbb{R}^n$ for any $n \in \mathbb{N}$. Much more substantially, any weighted planar graph (equipped with its shortest-path metric) or any $O(1)$-doubling metric space (see [61]) embeds into the real line with $O(1)$-quadratic average distortion (see [107, Section 7] for a justification of this, which adapts the reasoning in [125]), while such spaces need not even admit a bi-Lipschitz embedding into a Hilbert space [29, 77]. Also, if $2 < p < \infty$, then ℓ_p does not admit a bi-Lipschitz embedding into a Hilbert space (see [26]), but it follows from [107] that ℓ_p embeds into a Hilbert space with quadratic average distortion $O(p)$, and that this bound is optimal. More such examples will be obtained below.

1.1. **Notation, terminology, conventions.** Below, all metric spaces will be tacitly assumed to be separable. While some of the ensuing statements hold without a separability assumption, adhering to this convention avoids measurability side-issues that would otherwise obscure the main geometric content. Alternatively, one could harmlessly consider throughout only finitely supported measures.

In addition to the usual $O(\cdot), o(\cdot), \Omega(\cdot)$ notation, we will use the following (also standard) asymptotic notation. For $Q, Q' > 0$, the notations $Q \lesssim Q'$ and $Q' \gtrsim Q$ mean that $Q \leqslant \mathsf{K}Q'$ for a universal constant $\mathsf{K} > 0$. The notation $Q \asymp Q'$ stands for $(Q \lesssim Q') \wedge (Q' \lesssim Q)$. If we need to allow for dependence on parameters, we indicate this by subscripts. For example, in the presence of auxiliary parameters ψ, ξ, the notation $Q \lesssim_{\psi,\xi} Q'$ means that $Q \leqslant c(\psi,\xi)Q'$, where $c(\psi,\xi) \in (0,\infty)$ may depend only on ψ and ξ, and analogously for the notations $Q \gtrsim_{\psi,\xi} Q'$ and $Q \asymp_{\psi,\xi} Q'$.

We will use notions of Banach spaces [82, 83], metric embeddings [93, 119] and expanders [64, 6]. Any undefined term in the ensuing discussion is entirely standard and appears in the aforementioned references, but in this short subsection we recall a modicum of simple concepts.

For a normed space $(Z, \|\cdot\|_Z)$ and $p \geqslant 1$, the normed space $\ell_p(Z)$ consists of all those Z-valued sequences $x = (x_1, x_2, \ldots) \in Z^{\mathbb{N}}$ such that $\|x\|_{\ell_p(Z)}^p = \sum_{i=1}^{\infty} \|x_i\|_Z^p < \infty$. One writes $\ell_p(\mathbb{R}) = \ell_p$.

Theorem 1 tensorizes in a straightforward manner to give the same conclusion for $\ell_1(X)$. In order to facilitate later reference, it is beneficial to record this fact as the following separate statement.

Corollary 4. *For any normed space X of dimension $k \geqslant 2$, the $\frac{1}{2}$-snowflake of $\ell_1(X)$ embeds into a Hilbert space with quadratic average distortion $\mathsf{C}\sqrt{\log k}$, where $\mathsf{C} > 0$ is a universal constant.*

Proof. For every $i \in \mathbb{N}$ let $\mathfrak{c}_i : \ell_1(X) \to X$ denote the i'th coordinate projection, i.e., $\mathfrak{c}_i(x) = x_i$ for each $x = (x_1, x_2, \ldots,) \in \ell_1(X)$. Fix any Borel probability measure μ on $\ell_1(X)$. For each $i \in \mathbb{N}$, an application of Theorem 1 to the measure $(\mathfrak{c}_i)_\sharp \mu$ on X (the image of μ under $\mathfrak{c}_i$) yields $f_i : X \to H$ which is $\frac{1}{2}$-Hölder with constant $\mathsf{C}\sqrt{\log k}$, where $(H, \|\cdot\|_H)$ is a Hilbert space, that satisfies

$$\iint_{\ell_1(X)\times\ell_1(X)} \big\|f_i\big(\mathfrak{c}_i(x)\big) - f_i\big(\mathfrak{c}_i(y)\big)\big\|_H^2 \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant \iint_{\ell_1(X)\times\ell_1(X)} \|\mathfrak{c}_i(x) - \mathfrak{c}_i(y)\|_X \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y).$$

The desired embedding $f : \ell_1(X) \to \ell_2(H)$ is now defined by $f(x) = (f_1(x_1), f_2(x_2), \ldots)$. $\square$

The bi-Lipschitz distortion of a metric space $(\mathcal{M}, d_{\mathcal{M}})$ in a normed space $(Z, \|\cdot\|_Z)$ is a numerical invariant denoted $\mathsf{c}_Z(\mathcal{M})$ that is defined to be the infimum over those $D \in [1,\infty]$ for which there exists a D-Lipschitz mapping $f : \mathcal{M} \to Z$ satisfying $\|f(x) - f(y)\|_Z \geqslant d_{\mathcal{M}}(x,y)$ for all $x, y \in \mathcal{M}$.

The most natural setting to discuss average distortion of embeddings is that of metric probability spaces, namely triples $(\mathcal{M}, d_{\mathcal{M}}, \mu)$ where $(\mathcal{M}, d_{\mathcal{M}})$ is a metric space and μ is a Borel probability measure on $\mathcal{M}$. In this context, as an obvious variant of (1), for $p > 0$ and $D \geqslant 1$, say that $(\mathcal{M}, d_{\mathcal{M}}, \mu)$ embeds with p-average distortion D into a Banach space $(Z, \|\cdot\|_Z)$ if there is a D-Lipschitz mapping $f : \mathcal{M} \to Z$ such that $\iint_{\mathcal{M}\times\mathcal{M}} \|f(x) - f(y)\|_Z^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant \iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)$. If this holds with $p = 1$, one simply says that $(\mathcal{M}, d_{\mathcal{M}}, \mu)$ embeds with average distortion D into $(Z, \|\cdot\|_Z)$.

When a *finite* metric space $(\mathcal{M}, d_{\mathcal{M}})$ is said to embed with p-average distortion D into a Banach space $(Z, \|\cdot\|_Z)$ without explicitly specifying the underlying probability measure μ, it will always be understood that μ is the uniform probability measure on $\mathcal{M}$. Embeddings of finite metric spaces with controlled average distortion have several interesting applications, and their systematic investigation was initiated by Rabinovich [125]. If $(\mathcal{M}, d_{\mathcal{M}})$ is an *infinite* metric space, then when we say that it embeds with p-average distortion D into $(Z, \|\cdot\|_Z)$ we mean that for *every* probability measure μ on $\mathcal{M}$ the metric probability space $(\mathcal{M}, d_{\mathcal{M}}, \mu)$ embeds with p-average distortion D into $(Z, \|\cdot\|_Z)$. The difference between the terminology for finite and infinite spaces is natural because finite spaces carry a canonical probability (counting) measure while infinite spaces do not. We chose these conventions so as to be consistent with the terminology in the literature, which only treats finite spaces.

Using the above terminology, we record for ease of later reference the following immediate consequence of Corollary 4 (with the universal constant $\mathsf{C} \in [1,\infty)$ the same).

Corollary 5. *Suppose that $(X, \|\cdot\|_X)$ is a normed space of dimension $k \geqslant 2$ and that $(\mathcal{M}, d_{\mathcal{M}}, \mu)$ is a metric probability space that embeds into $\ell_1(X)$ with average distortion D. Then, $(\mathcal{M}, \sqrt{d_{\mathcal{M}}}, \mu)$ embeds into a Hilbert space with quadratic average distortion $\mathsf{C}\sqrt{D\log k}$.*

The following proposition demonstrates that the notion of average distortion is robust to changes of the moments of distances that one wishes to approximately preserve, as well as to snowflaking.

Proposition 6. *Fix $p, q, D \in [1, \infty)$ and $\omega \in (0, 1]$. Suppose that an infinite separable metric space $(\mathcal{M}, d_{\mathcal{M}})$ embeds with p-average distortion D into a Banach space $(Y, \|\cdot\|_Y)$. Then, the ω-snowflake of $(\mathcal{M}, d_{\mathcal{M}})$ embeds with q-average distortion $D' = D'(p, q, \omega) \geqslant 1$ into $(Y, \|\cdot\|_Y)$, where*

$$D' \lesssim_{p,q,\omega} D^{\max\left\{\frac{p}{q},\omega\right\}}. \tag{3}$$

We postpone discussion of Proposition 6 to Section 5.2 below, where it is proved and the implicit dependence on p, q, ω in (3) is specified; see (105). It suffices to say here that Proposition 6 shows that phenomena such as Theorem 1 (as well as more refined results that we will soon state), in which upon performing a certain amount of snowflaking the average distortion decreases from power-type behavior to logarithmic behavior, are independent of the choice of "p" in the notion of p-average distortion that one considers, and they persist if one performs an even greater amount of snowflaking.

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 $\{1, \ldots, n\}$. 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 $1 - \lambda_2(\mathsf{A})$ can be interpreted by straightforward linear algebra (expanding the squares and expressing in an eigenbasis of A) as the largest factor (in the left hand side) for which the following quadratic inequality holds true.

$$\forall\, x_1, \ldots, x_n \in \ell_2, \qquad \big(1 - \lambda_2(\mathsf{A})\big) \sum_{i=1}^n \sum_{j=1}^n \pi_i \pi_j \|x_i - x_j\|_{\ell_2}^2 \leqslant \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} \|x_i - x_j\|_{\ell_2}^2. \tag{4}$$

If $\mathsf{G} = (\{1, \ldots, n\}, E_{\mathsf{G}})$ is a connected graph, then the shortest-path metric that it induces is denoted $d_{\mathsf{G}} : \{1, \ldots, n\} \times \{1, \ldots, n\} \to \mathbb{N} \cup \{0\}$. If G is Δ-regular for some $\Delta \in \{2, \ldots, n\}$, then the normalized adjacency matrix of G, denoted $\mathsf{A}_{\mathsf{G}} \in \mathsf{M}_n(\mathbb{R})$, is the symmetric stochastic matrix whose entry at $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, n\}$ is equal to $\frac{1}{\Delta}\mathbf{1}_{\{i,j\} \in E_{\mathsf{G}}}$. Write $\lambda_2(\mathsf{G}) = \lambda_2(\mathsf{A}_{\mathsf{G}})$.

1.2. **A spectral gap is an obstruction to metric dimension reduction.** For every $n \in \mathbb{N}$ there is a $O(1)$-regular graph $\mathsf{G}_n = (\{1, \ldots, n\}, E_{\mathsf{G}_n})$ with $1/(1 - \lambda_2(\mathsf{G}_n)) = O(1)$. See the survey [64] for this statement and much more on such *expanders*. In particular, it is well-known that an argument of Linial, London and Rabinovich [84] gives that if the $\frac{1}{2}$-snowflake of $(\{1, \ldots, n\}, d_{\mathsf{G}_n})$ embeds with quadratic average distortion $D \geqslant 1$ into a Hilbert space, then necessarily $D \gtrsim \sqrt{\log n}$.

We will next recall why this nonembeddability statement holds, following an influential formulation of the approach of [84] due to Matoušek [92] and Gromov [56]. Before doing so, note that this establishes the aforementioned optimality of the distortion bound of Theorem 1, since the Fréchet embedding [53] yields an n-point subset S of $X = \ell_\infty^n$ that is isometric to $(\{1, \ldots, n\}, d_{\mathsf{G}_n})$, and therefore if μ is the uniform measure on S, then the quadratic average distortion of any embedding of the $\frac{1}{2}$-snowflake of $(X, \|\cdot\|_{\ell_\infty^n}, \mu)$ is at least a universal constant multiple of $\sqrt{\log n} = \sqrt{\log \dim(X)}$.

So, fix an integer $n \geqslant 4$ and $\Delta \in \{3, \ldots, n-1\}$. Suppose that $\mathsf{G} = (\{1, \ldots, n\}, E_{\mathsf{G}})$ is a connected Δ-regular graph. Let $(H, \|\cdot\|_H)$ be a Hilbert space and assume that $f : \{1, \ldots, n\} \to H$ satisfies

$$\sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_H^2 \geqslant \sum_{i=1}^n \sum_{j=1}^n d_{\mathsf{G}}(i, j) \qquad \text{and} \qquad \forall \{i, j\} \in E_{\mathsf{G}}, \quad \|f(i) - f(j)\|_H \leqslant D. \tag{5}$$

By a simple and standard counting argument (e.g. [92, page 193]), a positive universal constant fraction of the pairs of vertices $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, n\}$ satisfy $d_{\mathsf{G}}(i, j) \gtrsim \frac{\log n}{\log \Delta}$. Hence,

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|f(i) - f(j)\|_H^2 \overset{(5)}{\geqslant} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{\mathsf{G}}(i, j) \gtrsim \frac{\log n}{\log \Delta}. \tag{6}$$

With this observation, the average distortion D can be bounded from below through an application of the interpretation (4) of a spectral gap to the normalized adjacency matrix of G, as follows.

$$D^2 \overset{(5)}{\geqslant} \frac{1}{|E_\mathsf{G}|}\sum_{\{i,j\}\in E_\mathsf{G}} \|f(i)-f(j)\|_H^2 \overset{(4)}{\geqslant} \frac{1-\lambda_2(\mathsf{G})}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|f(i)-f(j)\|_H^2 \overset{(6)}{\gtrsim} \frac{1-\lambda_2(\mathsf{G})}{\log\Delta}\log n, \qquad (7)$$

We have thus shown (following [84, 92, 56]) that

$$D \gtrsim \frac{\sqrt{1-\lambda_2(\mathsf{G})}}{\sqrt{\log\Delta}}\sqrt{\log n}. \qquad (8)$$

So, $D \gtrsim \sqrt{\log n}$ when $\frac{1}{1-\lambda_2(\mathsf{G})} \lesssim 1$ and $\Delta \lesssim 1$, i.e., for expanders. In general, we have

Theorem 7. *Fix $D \geqslant 1$ and integers $n, \Delta \geqslant 3$ with $\Delta \leqslant n$. Let $\mathsf{G} = (\{1,\ldots,n\}, E_\mathsf{G})$ be a Δ-regular connected graph. Suppose that $(X, \|\cdot\|_X)$ is a finite-dimensional normed space such that the metric space $(\{1,\ldots,n\}, d_\mathsf{G})$ embeds with average distortion D into $\ell_1(X)$. Then necessarily*

$$\dim(X) \geqslant n^{\frac{\eta(\mathsf{G})}{D}}, \qquad \text{where} \qquad \eta(\mathsf{G}) \gtrsim \frac{1-\lambda_2(\mathsf{G})}{\log\Delta}. \qquad (9)$$

Proof. By combining (8) with Corollary 5, it follows that $\sqrt{D\log\dim(X)} \gtrsim \frac{\sqrt{1-\lambda_2(\mathsf{G})}}{\sqrt{\log\Delta}}\sqrt{\log n}$. □

Remark 8. The reasoning by which we deduced (8) from (5) did not use the entirety of the second condition of (5). Namely, in addition to the first inequality in (5), it suffices to assume that

$$\bigg(\frac{1}{|E_\mathsf{G}|}\sum_{\{i,j\}\in E_\mathsf{G}} \|f(i)-f(j)\|_H^2\bigg)^{\frac12} \leqslant D. \qquad (10)$$

Thus, we only need to have an upper bound on the discrete Sobolev $W^{1,2}$ norm of the embedding f, namely the left hand side of (10), rather than an upper bound on the Lipschitz constant of f. By using this variant in place of the average distortion assumption in Theorem 7, one deduces mutatis mutandis that the same conclusion (9) holds true if there exists $f:\{1,\ldots,n\}\to\ell_1(X)$ that satisfies the following two requirements (which can be described, respectively, as "quantitative invertibility on average," combined with a bound on the discrete Sobolev $W^{1,1}$ norm).

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|f(i)-f(j)\|_{\ell_1(X)} \geqslant \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_\mathsf{G}(i,j),$$

and

$$\frac{1}{|E_\mathsf{G}|}\sum_{\{i,j\}\in E_\mathsf{G}} \|f(i)-f(j)\|_{\ell_1(X)} \leqslant D.$$

In words, if G has a spectral gap and large average distance, and one is given $x_1,\ldots,x_n\in X$ for which the averages of the two sets of distances $\{\|x_i-x_j\|_X\}_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,n\}}$ and $\{\|x_i-x_j\|_X\}_{\{i,j\}\in E_\mathsf{G}}$ are within a fixed but potentially large constant factor from the corresponding averages of distances in G, then this crude geometric information about two specific "distance statistics" of the finite point configuration $\{x_1,\ldots,x_n\}\subset X$ forces the continuous ambient space X to be high-dimensional.

The natural variant of Theorem 7 for ℓ_1 products of k-dimensional normed spaces $\{(X_i,\|\cdot\|_{X_i})\}_{i=1}^\infty$ that need not all be the same space $(X,\|\cdot\|_X)$ holds mutatis mutandis by the same reasoning. Also, a version with $\ell_1(X)$ replaced by $\ell_p(X)$ for any $p\geqslant 1$ appears in Section 1.2.1 below, where it is explained that this seemingly more general setting is, in fact, a formal consequence of its counterpart for $\ell_1(X)$ that we deduced above from Theorem 1. Section 1.4 is a description of the history of the questions that Theorem 7 resolves, as well as an indication of subsequent algorithmic developments that rely on it and were found since the initial posting of a preliminary version of the present work.

1.2.1. *A matrix-dimension inequality.* Let $(X,\|\cdot\|_X)$ be a normed space of dimension $k\geqslant 2$. Fix $n\in\mathbb{N}$ and $\pi\in\triangle^{n-1}$. For every n-tuple of vectors $x_1,\ldots,x_n\in X$ we can apply Theorem 1 to obtain a function $f=f_{\pi,x_1,\ldots,x_n}:X\to H$, where $(H,\|\cdot\|_H)$ is a Hilbert space, satisfying

$$\forall\, x,y\in X,\qquad \|f(x)-f(y)\|_H\leqslant \mathsf{C}\sqrt{\log k}\cdot\sqrt{\|x-y\|_X}, \tag{11}$$

as well as

$$\sum_{i=1}^n\sum_{j=1}^n \pi_j\pi_j\|f(x_i)-f(x_j)\|_H^2\geqslant \sum_{i=1}^n\sum_{j=1}^n \pi_j\pi_j\|x_i-x_j\|_X. \tag{12}$$

If $\mathsf{A}=(a_{ij})\in\mathsf{M}_n(\mathbb{R})$ is a stochastic and π-reversible matrix, then

$$\begin{aligned}\mathsf{C}^2\log k &\overset{(11)}{\geqslant} \frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|f(x_i)-f(x_j)\|_H^2}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X}\\ &\overset{(4)}{\geqslant}\big(1-\lambda_2(\mathsf{A})\big)\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|f(x_i)-f(x_j)\|_H^2}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X}\\ &\overset{(12)}{\geqslant}\big(1-\lambda_2(\mathsf{A})\big)\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X},\end{aligned} \tag{13}$$

Above, and henceforth, the ratios that appear in (13) are interpreted to be equal to 0 when their denominator $\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X$ vanishes (this "disconnected" case will never be of interest).

We have thus established that, as a consequence of Theorem 1, every finite configuration of points $x_1,\ldots,x_n\in X$ imposes the following geometric lower bound on the dimension of the ambient space.

$$\dim(X)\geqslant \exp\left(\frac{1-\lambda_2(\mathsf{A})}{\mathsf{C}^2}\cdot\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X}\right). \tag{14}$$

The ℓ_1 bound (14) formally implies the following ℓ_p counterpart for every $p\geqslant 1$.

$$\dim(X)\geqslant \exp\left(\frac{1-\lambda_2(\mathsf{A})}{\beta(p)}\left(\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X^p}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X^p}\right)^{\frac1p}\right), \tag{15}$$

where the constant $\beta(p)>0$ depends only on p. This is so because, by Proposition 6, it is a formal consequence of Theorem 1 that there is also an Hilbertian embedding of the $\frac12$-snowflake of X with $(2p)$-average distortion $\mathsf{C}(p)\sqrt{\log k}$ for every $p\geqslant 1$ (the special case of Proposition 6 that was used here is due to [107, Section 7.4]). By substituting this into the above reasoning, one arrives at (15).

While the asymptotic dependence in (the above use of) Proposition 6 that we obtain in Section 5.2 improves over what is available in the literature, it does not yield the sharp dependence of $\beta(p)$ on p as $p\to\infty$, and more care is needed in order to derive Theorem 9 below, which we will prove in Section 5.3. It obtains what we expect to be the sharp asymptotic dependence on p, though at present we do not see a proof of this; the desired optimality would follow from Conjecture 11 below.

Theorem 9. *There is a universal constant $\mathsf{K}>0$ such that for every $p\geqslant 1$ and $n\in\mathbb{N}$, if $(X,\|\cdot\|_X)$ is a normed space, $\pi\in\triangle^{n-1}$ and $\mathsf{A}=(a_{ij})\in\mathsf{M}_n(\mathbb{R})$ is a stochastic π-reversible matrix, then*

$$\forall\, x_1,\ldots,x_n\in X,\qquad \dim(X)\geqslant \exp\left(\frac{1-\lambda_2(\mathsf{A})}{\mathsf{K}p}\left(\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X^p}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X^p}\right)^{\frac1p}\right). \tag{16}$$

We thus obtain the following variant of Theorem 7 in which $\ell_1(X)$ is replaced by $\ell_p(X)$ for $p\geqslant 1$.

Corollary 10. *Fix $D,p \geqslant 1$ and integers $n,\Delta \geqslant 3$ with $\Delta \leqslant n$. Let $\mathsf{G} = (\{1,\ldots,n\},E_\mathsf{G})$ be a Δ-regular connected graph. Suppose that $(X,\|\cdot\|_X)$ is a normed space such that $(\{1,\ldots,n\},d_\mathsf{G})$ embeds with average distortion D into $\ell_p(X)$. Then, $\dim(X) \geqslant n^{\eta(\mathsf{G})/(pD)}$ for $\eta(\mathsf{G}) > 0$ as in (9).*

Proof. For $f : \{1,\ldots,n\} \to \ell_p(X)$ and $m \in \mathbb{N}$ denote the m'th entry of f by $f_m : \{1,\ldots,n\} \to X$, i.e., for every $i \in \{1,\ldots,n\}$ we have $f(i) = (f_1(i), f_2(i),\ldots) \in \ell_p(X)$. Then, by Theorem 9 applied to the finite configuration $\{f_m(1),\ldots,f_m(n)\}$ of points in X for each $m \in \mathbb{N}$ separately, we see that

$$\dim(X) \geqslant \exp\left(\frac{1-\lambda_2(\mathsf{A})}{\mathsf{K}p} \sup_{m\in\mathbb{N}} \left(\frac{\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|f_m(i)-f_m(j)\|_X^p}{\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|f_m(i)-f_m(j)\|_X^p}\right)^{\frac{1}{p}}\right).$$

It remains to observe that if $f : \{1,\ldots,n\} \to \ell_p(X)$ is D-Lipschitz (with respect to the shortest-path metric d_G), yet $\sum_{i=1}^n\sum_{j=1}^n \|f(i)-f(j)\|_{\ell_p(X)} \geqslant \sum_{i=1}^n\sum_{j=1}^n d_\mathsf{G}(i,j)$, then

$$\begin{aligned}\sup_{m\in\mathbb{N}} \left(\frac{\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|f_m(i)-f_m(j)\|_X^p}{\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|f_m(i)-f_m(j)\|_X^p}\right)^{\frac{1}{p}} &\geqslant \left(\frac{\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|f(i)-f(j)\|_{\ell_p(X)}^p}{\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|f(i)-f(j)\|_{\ell_p(X)}^p}\right)^{\frac{1}{p}}\\ &\geqslant \frac{1}{Dn^2}\sum_{i=1}^n\sum_{j=1}^n \|f(i)-f(j)\|_{\ell_p(X)} \geqslant \frac{1}{Dn^2}\sum_{i=1}^n\sum_{j=1}^n d_\mathsf{G}(i,j) \gtrsim \frac{\log n}{D\log\Delta},\end{aligned}$$

where the first step holds because $\frac{a_1+a_2+\ldots}{b_1+b_2+\ldots} \leqslant \sup_{m\in\mathbb{N}} \frac{a_m}{b_m}$ for any $\{a_m\}_{m=1}^\infty, \{b_m\}_{m=1}^\infty \subset (0,\infty)$, the second step is an application of Jensen's inequality in the numerator and the D-Lipschitz condition in the denominator, the third step is our second assumption on f, and the final step is (6). $\square$

We expect that Corollary 10 is sharp in terms of its dependence on p, as expressed in Conjecture 11 below, whose positive resolution might have algorithmic applications; this could be quite tractable by adapting available methods, specifically those of [70, 90, 91, 92, 1].

Conjecture 11. There is a universal constant $\mathsf{C} \geqslant 1$ with the following property. For every $p, D \geqslant 1$ there exists $n_0 = n_0(p,D) \in \mathbb{N}$ such that if $n \geqslant n_0$, then for every n-point metric space $(\mathcal{M}, d_\mathcal{M})$ there exists $k \in \mathbb{N}$ with $k \leqslant n^{\mathsf{C}/(pD)}$ and a k-dimensional normed space $(X,\|\cdot\|_X)$ such that $\mathcal{M}$ embeds with bi-Lipschitz distortion D into $\ell_p(X)$. Conceivably this even holds true for $X = \ell_\infty^k$.

1.3. **Uniform convexity and smoothness.** Henceforth, the closed unit ball of a normed space $(X,\|\cdot\|_X)$ will be denoted by $B_X = \{x \in X : \|x\|_X \leqslant 1\}$. The moduli [45] of uniform convexity and uniform smoothness of $(X,\|\cdot\|_X)$, commonly denoted $\delta_X : [0,2] \to [0,\infty)$ and $\rho_X : [0,\infty) \to [0,\infty)$, respectively, are the (point-wise) smallest such functions for which every $x,y \in \partial B_X$ and $\tau \in [0,\infty)$ satisfy $\|x+y\|_X \leqslant 2(1-\delta_X(\|x-y\|_X))$ and $\|x+\tau y\|_X + \|x-\tau y\|_X \leqslant 2(1+\rho_X(\tau))$.

Given $p,q \in [1,\infty)$, one says that $(X,\|\cdot\|_X)$ has moduli of uniform convexity and uniform smoothness of power type q and p, respectively, if $\delta_X(\varepsilon) \gtrsim_{X,q} \varepsilon^q$ and $\rho_X(\tau) \lesssim_{X,p} \tau^p$ for all $\varepsilon \in [0,2]$ and $\tau \in [0,\infty)$. By the parallelogram identity, a Hilbert space has moduli of uniform convexity and uniform smoothness of power type 2; conversely, Figiel and Pisier [52] proved (confirming a conjecture of Lindenstrauss [81]) that if a Banach space has this property, then it is isomorphic to a Hilbert space. In the reflexive range $p \in (1,\infty)$, the works of Clarkson [40] and Hanner [59] show that any $L_p(\mu)$ space has moduli of uniform convexity and uniform smoothness of power type $\max\{p,2\}$ and $\min\{p,2\}$, respectively.

An important theorem of Pisier [122] asserts that if $\delta_X(\varepsilon) > 0$ for all $\varepsilon \in (0,2]$, then there exists $q \in [2,\infty)$ and an equivalent norm on X with respect to which it has modulus of uniform convexity of power type q. Analogously, if $\lim_{\tau\to 0^+} \rho_X(\tau)/\tau = 0$, then there exists $p \in (1,2]$ and an equivalent norm on X with respect to which it has modulus of uniform smoothness of power type p. For this reason, we will focus below only on uniform convexity and smoothness with power-type behavior.

Theorem 12. *Fix $p,q\in[1,\infty)$ that satisfy $p\leqslant 2\leqslant q$. Let $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ be Banach spaces that have moduli of uniform smoothness and uniform convexity of power type p and q, respectively. Then, there exists $D=D(\rho_X,\delta_Y,q)\in[1,\infty)$ satisfying*

$$D\lesssim_{\rho_X,\delta_Y,p,q}\big(\log(\mathsf{c}_Y(X)+1)\big)^{\frac{1}{q}},\tag{17}$$

such that the $\frac{p}{q}$-snowflake of $(X,\|\cdot\|_X)$ embeds with q-average distortion D into $\ell_q(Y)$.

Furthermore, if $(Y,\|\cdot\|_Y)$ has modulus of uniform smoothness of power type $r>p$, then for every $\varepsilon\in[0,\frac{r-p}{q}]$ the $(\frac{p}{q}+\varepsilon)$-snowflake of $(X,\|\cdot\|_X)$ embeds with q-average distortion D into $\ell_q(Y)$, where

$$D\lesssim_{\rho_X,\delta_Y,\rho_Y,p,q}\begin{cases}\big(\log(\mathsf{c}_Y(X)+1)\big)^{\frac{1}{q}} & \text{if } 0\leqslant\varepsilon\leqslant\frac{r-p}{q\log(\mathsf{c}_Y(X)+1)},\\ \left(\frac{r-p}{\varepsilon}\right)^{\frac{1}{q}}\mathsf{c}_Y(X)^{\frac{r\varepsilon}{r-p}} & \text{if } \frac{r-p}{q\log(\mathsf{c}_Y(X)+1)}\leqslant\varepsilon\leqslant\frac{r-p}{q}.\end{cases}\tag{18}$$

Because $\ell_q(\ell_q(Y))$ is isometric to $\ell_q(Y)$ and by [51] if Y satisfies the assumption of Theorem 12, then so does $\ell_q(Y)$, Theorem 12 establishes that the worst-case bi-Lipschitz distortion into $\ell_q(Y)$ is exponentially larger than its average-case counterpart. Specifically, in the setting of Theorem 12, if one finds any Borel probability measure μ on X such that no embedding of the (p/q)-snowflake of $(X,\|\cdot\|_X,\mu)$ into $\ell_q(Y)$ has q-average distortion less than $D\geqslant 1$, then any embedding of X into $\ell_q(Y)$ incurs bi-Lipschitz distortion at least $\exp(\beta D^q)$, where $\beta>0$ depends only on ρ_X,δ_Y,p,q.

Theorem 1 is a special case of Theorem 12. Indeed, if $\dim(X)=k$ and $Y=\ell_2$, then $\mathsf{c}_{\ell_2}(X)\leqslant\sqrt{k}$ by John's theorem (the simpler Auerbach lemma [119, Lemma 2.22] suffices for this application). The assumptions of Theorem 12 hold with $p=1$ for any Banach space $(X,\|\cdot\|_X)$ and with $q=2$ when $Y=\ell_2$, in which case $\ell_2(Y)$ is still a Hilbert space, so we arrive at the conclusion of Theorem 1.

When $(X,\|\cdot\|_X)$ has modulus of uniform smoothness of power type $1<p\leqslant 2$, Theorem 12 obtains the desired embedding with Hölder regularity that improves with p, namely a lesser amount of snowflaking. In particular, if $(X,\|\cdot\|_X)$ has modulus of uniform smoothness of power type 2 and $(Y,\|\cdot\|_Y)$ has modulus of uniform convexity of power type 2, then the embedding of Theorem 12 is of the original metric on $(X,\|\cdot\|_X)$ without any snowflaking (i.e., it is Lipschitz rather than Hölder).

Returning to an examination of the special case when $(Y,\|\cdot\|_Y)$ is a Hilbert space, the first part of Theorem 12 yields the same quadratic average distortion as that of Theorem 1, but now this is achieved for the $\frac{p}{2}$-snowflake of $(X,\|\cdot\|_X)$. For $p>1$ this is better (a less dramatic deformation of the original metric on X) than the $\frac{1}{2}$-snowflake of Theorem 1, albeit under the stronger assumption that the modulus of uniform smoothness of $(X,\|\cdot\|_X)$ is of power type p. For every $1\leqslant p<2$ (thus also covering the setting of Theorem 1), this amount of snowflaking is sharp in the sense that for any fixed exponent that is strictly larger than $\frac{p}{2}$, the dependence on $\mathsf{c}_Y(X)$ in (17) must sometimes grow as $\mathsf{c}_Y(X)\to\infty$ at a rate that is at least a definite positive power of $\mathsf{c}_Y(X)$. This is the content of the following lemma, which also establishes that the exponent of $\mathsf{c}_Y(X)$ in (18), which is equal to $2\varepsilon/(2-p)$ when Y is a Hilbert space (hence $q=r=2$), cannot be improved in general.

Lemma 13. *Fix $p\in[1,2)$, $\alpha\geqslant 1$, $\beta\geqslant 2$ and $\varepsilon\in[0,1-\frac{p}{2}]$. For arbitrarily large $\mathsf{c}\geqslant 1$ there exists a normed space $(X,\|\cdot\|_X)$ that satisfies $\mathsf{c}_{\ell_2}(X)=\mathsf{c}$ and whose modulus of uniform smoothness has power type p, such that if the $(\frac{p}{2}+\varepsilon)$-snowflake of $(X,\|\cdot\|_X)$ embeds with α-average distortion $D\geqslant 1$ into $\ell_\beta(\ell_2)$, then necessarily*

$$D\gtrsim\frac{1}{\sqrt{\alpha+\beta}}\mathsf{c}^{\frac{2\varepsilon}{2-p}}.\tag{19}$$

We do not know if the analogue of Lemma 13 holds in the full range of parameters of Theorem 12, namely when either $q\neq 2$ or $r\neq 2$ (note that by the aforementioned Figiel–Pisier characterization of Hilbert space [52], if $q=r=2$, then Y must be isomorphic to a Hilbert space).

Question 14. Is the exponent of $\mathsf{c}_Y(X)$ in (18), namely $\frac{r\varepsilon}{r-p}$, optimal also when $(q,r)\neq(2,2)$?

Despite the optimality for $Y=\ell_2$ of the amount of snowflaking that is required for achieving the logarithmic behavior (17), as expressed in Lemma 13, at the endpoint case of $\frac{p}{2}$-snowflakes the potential optimality of the dependence on $\mathsf{c}_{\ell_2}(X)$ in (17) is much more mysterious. The case $p=1$ is an exception, because we have already seen (in the beginning of Section 1.2) that the distortion bound of Theorem 1 is sharp; more generally, Remark 51 below shows that (17) is sharp for $p=1$ and any $q\geqslant 2$. However, this is proved by considering $X=\ell_\infty^k$, which is not pertinent to the range $p\in(1,2]$. The (in our opinion unlikely) possibility remains that if $p\in(1,2]$, then for every Banach space X whose modulus of uniform smoothness has power type p there exists $\omega=\omega(X)\in(0,1]$ such that the ω-snowflake of X embeds with average distortion[2] $D=D(X)\in[1,\infty)$ into a Hilbert space; we do not know an obstruction to this holding even for the maximal possible exponent $\omega=\frac{p}{2}$.

A Banach space X is called *superreflexive* if it admits an equivalent uniformly smooth norm, namely a norm for which $\lim_{\tau\to 0^+}\rho_X(\tau)/\tau=0$, and this holds if an only if [48] it admits an equivalent uniformly convex norm, namely a norm for which $\delta_X(\varepsilon)>0$ for all $\varepsilon\in(0,2]$. These are not the original definitions of superreflexivity (due to James [66]), but they are equivalent to them by deep work of Enflo [48]. By the aforementioned renorming theorem of Pisier [122], superreflexivity is equivalent to admitting an equivalent norm whose modulus of uniform smoothness has power type p for some $p>1$. Therefore, the above discussion coincides with the following question.

Question 15. Does every superreflexive Banach space X admit $\omega(X)\in(0,1]$ and $D(X)\in[1,\infty)$ such that the $\omega(X)$-snowflake of X embeds with average distortion $D(X)$ into a Hilbert space?

We conjecture that the answer to Question (15) is negative; constructing an example that demonstrates this conjecture would be an important achievement. Perhaps even (17) is sharp, but at present we do not have sufficient evidence in support of this more ambitious conjecture (other than that this is so when $p=1$). Notwithstanding the above expectation, if it were the case that Question (15) had a positive answer, then this would be a truly remarkable theorem, asserting that the mere presence of uniform convexity implies "bounded distance on average" from Hilbertian geometry. In particular, a positive answer to Question (15) would resolve a central open question (see e.g. [78, 124, 99]) by demonstrating that every classical expander is a super-expander; we will explain this deduction in Remark 20 below, after the relevant concepts are recalled.

1.3.1. *Embedding a complex interpolation family into its endpoint.* Theorem 12, and therefore also its special case Theorem 1 (average John) and its corollaries Theorem 7 (impossibility of dimension reduction for expanders) and Theorem 9 (matrix-dimension inequality), are all consequences of the structural statement for complex interpolation spaces that appears in Theorem 16 below. We recall the (standard) background in Section 3 below; here we explain the idea in broad stokes.

Following Calderón [34] and Lions [87], to a pair $(X,\|\cdot\|_X),(Z,\|\cdot\|_Z)$ of complex Banach spaces that satisfies a mild compatibility assumption (which will be immediate in our setting), one associates a one-parameter family $(\theta\in[0,1])\mapsto[X,Z]_\theta$ of Banach spaces which interpolates between them, namely $[X,Z]_0=X$ and $[X,Z]_1=Z$. This provides a useful way to deform the geometry of $(X,\|\cdot\|_X)$ to that of $(Z,\|\cdot\|_Z)$, and Theorem 16 is a technical statement that quantifies the extent to which elements of this complex interpolation family differ from its $\theta=1$ endpoint.

Theorem 16. *Fix $\theta\in(0,1]$ and $p,q\in[1,\infty)$ with $1\leqslant p\leqslant 2\leqslant q$. Let $(X,\|\cdot\|_X)$ and $(Z,\|\cdot\|_Z)$ be a compatible pair of complex Banach spaces such that the moduli of uniform smoothness and convexity of $[X,Z]_\theta$ and Z are of power type p and q, respectively. Then, there exists $D\geqslant 1$ satisfying*

$$D\lesssim_{\rho_{[X,Z]_\theta},\delta_Z,p,q}\left(\frac{1}{\theta}\right)^{\frac{1}{q}},\tag{20}$$

[2]For concreteness we chose to discuss average distortion, but note that due to Proposition 6, for any $q\geqslant 1$ such a qualitative statement is equivalent to the same statement with "average distortion" replaced by "q-average distortion."

such that the $\frac{p}{q}$-snowflake of $[X,Z]_\theta$ embeds with q-average distortion D into $\ell_q(Z)$.

The implicit dependence in (20) on the data $\rho_{[Y,Z]_\theta}, \delta_Z, p, q$ is specified in Section 3 below, where Theorem 16 is proved; see specifically inequality (35). Section 3.2 below demonstrates that Theorem 16 implies Theorem 12. The basic idea is as follows. Suppose that $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are Banach spaces that satisfy the assumptions of Theorem 12. Assume for simplicity that they are a compatible pair of complex Banach spaces (this side-issue is treated in Section 3 via a standard complexification step). By estimating the relevant modulus of $[X,Y]_\theta$, contrasting the bound (20) with a bound on the distance of $[X,Y]_\theta$ to X, and then optimizing over θ, we arrive at Theorem 12. Thus, the role of complex interpolation in the proof of Theorem 12 is in essence as a Banach space-valued flow starting at X and terminating at Y, parameterized by $\theta \in [0,1]$. At positive times $\theta > 0$, this flow consists of spaces that embed on average into $\ell_q(Y)$, by Theorem 16. The desired embedding of X itself is obtained since this flow tends to X as $\theta \to 0^+$ (at a definite rate).

Remark 17. Theorem 16 (combined with Proposition 6) shows that Question 15 has a positive answer for spaces of the form $[X,H]_\theta$ where X, H is a compatible pair of complex Banach spaces with H being a Hilbert space, and $\theta \in (0,1]$ (thus, by Pisier's extrapolation theorem [123], Question 15 has a positive answer for superreflexive Banach lattices). An inspection of the ensuing proof of Theorem 16 reveals that this holds also for subspaces of quotients of the class of θ-Hilbertian Banach spaces that was introduced by Pisier in [124] (we will not recall the definition here). So, a proof of our conjectured negative answer to Question 15 would entail constructing a superreflexive Banach space which is not a subspace of a quotient of any θ-Hilbertian Banach space. Such spaces are not yet known to exist; see [124, pages 15–16] for a discussion of this intriguing open question in structural Banach space theory. As stated above and justified in Remark 20, also disproving our conjecture by answering Question 15 positively would have interesting ramifications.

1.4. **Historical discussion.** Adopting terminology of [84, Definition 2.1], given $n \in \mathbb{N}$, an n-point metric space $\mathcal{M}$ and $D \geqslant 1$, define an integer $\dim_D(\mathcal{M})$, called the (bi-Lipschitz distortion-D) *metric dimension* of $\mathcal{M}$, to be the minimum $k \in \mathbb{N}$ for which there exists a k-dimensional normed space $X_{\mathcal{M}}$ such that $\mathcal{M}$ embeds into $X_{\mathcal{M}}$ with bi-Lipschitz distortion D. By the Fréchet isometric embedding into ℓ_∞^{n-1}, we always have $\dim_D(\mathcal{M}) \leqslant \dim_1(\mathcal{M}) \leqslant n-1$.

Johnson and Lindenstrauss asked [68, Problem 3] if $\dim_D(\mathcal{M}) = O(\log n)$ for some $D = O(1)$ and every n-point metric space $\mathcal{M}$. The $O(\log n)$ bound arises naturally here, because it cannot be improved due to a standard volumetric argument when one considers embeddings of the equilateral space of size n. See Remark 19 below and mainly the survey [111] for background on this question (and, more generally, the field of *metric dimension reduction*, to which the present investigations belong), including how it initially arose in the context of the Ribe program.

Bourgain proved [28, Corollary 4] that the Johnson–Lindenstrauss question has a negative answer. Specifically, he showed that for arbitrarily large $n \in \mathbb{N}$ there is an n-point metric space $\mathcal{M}_n$ such that $\dim_D(\mathcal{M}_n) \gtrsim (\log n)^2/(D \log\log n)^2$ for every $D \geqslant 1$. This naturally led him to raise the question [28, page 48] of determining the asymptotic behavior of the maximum of $\dim_D(\mathcal{M})$ over all n-point metric spaces $\mathcal{M}$. It took over a decade for this question to be resolved.

Towards this goal, Johnson, Lindenstrauss and Schechtman [70] proved that there exists a universal constant $\alpha > 0$ such that for every $D \geqslant 1$ and $n \in \mathbb{N}$ we have $\dim_D(\mathcal{M}) \lesssim_D n^{\alpha/D}$ for any n-point metric space $\mathcal{M}$. In [90, 91], Matoušek improved this result by showing that one can actually embed any such $\mathcal{M}$ with distortion D into ℓ_∞^k for some $k \in \mathbb{N}$ satisfying $k \lesssim_D n^{\alpha/D}$, i.e., the target normed space need not depend on $\mathcal{M}$ (Matoušek's proof is also simpler than that of [70], and it yields a smaller value of the constant α; see the exposition in Chapter 15 of the monograph [93]).

For small distortions, Arias-de-Reyna and Rodríguez-Piazza proved [15] the satisfactory assertion that for arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space $\mathcal{M}_n$ such that $\dim_D(\mathcal{M}_n) \gtrsim_D n$ for every $1 \leqslant D < 2$. For larger distortions, they asked [15, page 109] if for every $D \in (2,\infty)$ and

$n \in \mathbb{N}$ we have $\dim_D(\mathcal{M}) \lesssim_D (\log n)^{O(1)}$ for any n-point metric space $\mathcal{M}$. For this distortion regime, an asymptotic improvement (as $n \to \infty$) over the aforementioned lower bound of Bourgain [28] was made by Linial, London and Rabinovich [84, Proposition 4.2], who showed that for arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space $\mathcal{M}_n$ such that $\dim_D(\mathcal{M}_n) \gtrsim (\log n)^2/D^2$ for every $D \geqslant 1$.

In [91], Matoušek answered the above questions by proving Theorem 18 below via an ingenious argument that relies on (a modification of) graphs of large girth with many edges and an existential counting argument (inspired by ideas of Alon, Frankl and Rödl [4]) that uses the classical theorem of Milnor [102] and Thom [133] from real algebraic geometry.

Theorem 18 (Matoušek [91])**.** *For every $D \geqslant 1$ and arbitrarily large $n \in \mathbb{N}$, there exists an n-point metric space $\mathcal{M}_n(D)$ such that $\dim_D\big(\mathcal{M}_n(D)\big) \gtrsim_D n^{c/D}$, where $c > 0$ is a universal constant.*

Due to the Johnson–Lindenstrauss–Schechtman upper bound [70], Theorem 18 is a complete (and unexpected) answer to the aforementioned questions of Johnson–Lindenstrauss [68], Bourgain [28] and Arias-de-Reyna–Rodríguez-Piazza [15], up to the value of the universal constant c. Theorem 7 furnishes a new resolution of these questions, via an analytic approach for deducing dimension lower bounds from rough metric information that differs markedly from Matoušek's algebraic argument.

Our solution has some novel features. It shows that the spaces $\mathcal{M}_n(D)$ of Theorem 18 can actually be taken to be independent of the distortion D, while the construction of [91] depends on D (it is based on graphs whose girth is of order D). One could alternatively achieve this by considering the disjoint union of the spaces $\{\mathcal{M}_n(2^k)\}_{k=0}^m$ for $m \asymp \log n$, which is a metric space of size $O(n \log n)$.

Rather than using an ad-hoc construction (and a non-constructive existential statement) as in [91], here we specify a natural class of metric spaces, namely the shortest-path metrics on expanders (see also Remark 55 below), for which Theorem 7 holds. The question of determining the metric dimension of expanders was first considered by Linial–London–Rabinovich [84]. Indeed, their aforementioned lower bound $\dim_D(\mathcal{M}_n) \gtrsim (\log n)^2/D^2$ was obtained when $\mathcal{M}_n$ is the shortest-path metric on an n-vertex expander $\mathsf{G} = (\{1,\ldots,n\}, E_\mathsf{G})$. This lower bound remained the best-known prior to our proof of Theorem 7 that establishes the exponential improvement $\dim_D(\{1,\ldots,n\}, d_\mathsf{G}) \geqslant n^{\eta/D}$ for some $\eta = \Omega(1)$, which is best-possible up to the value of η.

We were motivated to revisit this old question because it arose more recently in the work [13] of Andoni, Nguyen, Nikolov, Razenshteyn and Waingarten on approximate nearest neighbor search (NNS). They devised an approach for proving an impossibility result for NNS that requires the existence of an n-vertex expander that embeds with bi-Lipschitz distortion $O(1)$ into some normed space of dimension $n^{o(1)}$. By Theorem 7 no such expander exists, thus resolving (negatively) a question that Andoni–Nguyen–Nikolov–Razenshteyn–Waingarten posed in [13, Section 1.6].

Unlike Theorem 18, the lower bound $\dim(X) \geqslant n^{\Omega(1)}$ of Theorem 7 assumes (when the underlying graph G is an n-vertex expander) that the embedding has $O(1)$ average distortion rather than the worst-case control that $O(1)$ bi-Lipschitz distortion entails. In fact, we only need to assume that there is $f : \{1,\ldots,n\} \to X$ that preserves up to constant factors two specific distance sums, i.e., that $\sum_{\{i,j\}\in E_\mathsf{G}} \|f(i)-f(j)\|_X \asymp \sum_{\{i,j\}\in E_\mathsf{G}} d_\mathsf{G}(i,j)$ and $\sum_{i=1}^n \sum_{j=1}^n \|f(i)-f(j)\|_X \asymp \sum_{i=1}^n \sum_{j=1}^n d_\mathsf{G}(i,j)$. Moreover, we also deduce the lower bound $\dim(X) \geqslant n^{\Omega(1)}$ from the existence of an embedding into $\ell_1(X)$ with these properties. We do not see how the algebraic technique of [91] could address such issues, namely distance preservation being only on average and infinite dimensional targets.

Thus, the new approach that we devise here is both more robust than that of [91], in the sense that it relies on significantly less stringent assumptions, and it also provides an explicit criterion (spectral gap) for intrinsic (largest-possible) high-dimensionality. Both of these features, as well as ideas within our proof, turned out to be important for subsequent developments that occurred since a preliminary version of the present work was posted (November 2016). Specifically, in a series of collaborations with Andoni, Nikolov, Razenshteyn and Waingarten [9, 10, 12, 11], we studied the algorithmic question (NNS) that Theorem 7 resolves negatively (recall that it is a negative solution to a question that would have implied an algorithmic *impossibility result*). These works design NNS data structures for arbitrary high-dimensional norms that were previously believed to be unattainable. For this purpose, the robustness of the average-case requirements in combination with our use below of a recently developed theory of *nonlinear spectral gaps* are both crucial for uncovering new structural information about general norms (a randomized hierarchical partitioning scheme that is governed by the intrinsic geometry). We refer to [9, 10, 12, 11] and the surveys [7, 111] for more information on these more recent algorithmic developments which rely on the present work.

Remark 19. The *Ribe program* aims to uncover an explicit "dictionary" between the local theory of Banach spaces and general metric spaces, inspired by a rigidity theorem of Ribe [127] which indicates that a dictionary of this sort should exist. See [29] as well as the surveys [72, 105, 20, 111] and the monograph [119] for more on this area. While much of the more recent research on dimension reduction is driven by the need to compress data, the initial motivation of the above question of [68] arose in the Ribe program. It is simplest to include here a direct quotation of Matoušek's explanation in [91, page 334] for the origin of the investigations that led to his Theorem 18.

> *...This investigation started in the context of the local Banach space theory, where the general idea was to obtain some analogs for general metric spaces of notions and results dealing with the structure of finite dimensional subspaces of Banach spaces. The distortion of a mapping should play the role of the norm of a linear operator, and the quantity* $\log n$*, where* n *is the number of points in a metric space, would serve as an analog of the dimension of a normed space. Parts of this programme have been carried out by Bourgain, Johnson, Lindenstrauss, Milman and others...*

Despite many previous successes of the Ribe program, not all of the questions that it raised turned out to have a positive answer (e.g. [97]). Theorem 18 is among the most extreme examples of failures of natural steps in the Ribe program, with the final answer being exponentially worse than the predictions. Here we provide a different derivation (and strengthening) of this phenomenon.

It is an amusing coincidence that while Johnson and Lindenstrauss raised [68, Problem 3] as a step toward a metric version of John's theorem (see [68, Problem 4]; this was resolved by Bourgain [28], who took a completely different route than the one proposed in [68]), the present work finds another nonlinear version of John's theorem and demonstrates that in fact it serves as an obstruction to the dimension reduction phenomenon that Johnson and Lindenstrauss were hoping for.

1.5. **Roadmap.** Section 2 recalls the theory of nonlinear spectral gaps that was alluded to above. Further background on uniform convexity and smoothness, as well as background on Ball's notion of Markov type (both of which are tools for subsequent proofs) appears, respectively, in Section 2.1 and Section 2.2. The link between nonlinear spectral gaps and Theorem 1 is through a duality statement that we proved in [107]; Section 2.3 describes a convenient enhancement of this duality which is proved in (the mainly technical) Section 7. Section 3 treats complex interpolation, leading to Theorem 16. A key inequality (Theorem 25) about nonlinear spectral gaps in complex interpolation spaces appears in Section 3.1. Its proof adapts an approach of [107] where a similar inequality was derived; such an adaptation is required because [107] relies on somewhat arbitrary choices of distance exponents, due to which we do not see how to use the results of [107] to prove Theorem 1. The deduction of Theorem 12 (hence also its special case Theorem 1) from Theorem 25 appears in Section 3.2. The proof of Theorem 25, namely our main nonlinear spectral gap inequality, appears in Section 4, though it assumes Proposition 6 whose proof is postponed to Section 5 which is devoted to several auxiliary embedding results of independent interest. The case $\omega = 1$ of Proposition 6 (passing from p-average distortion to q-average distortion) was first broached in [107] where a similar statement is obtained under an additional assumption that is not needed in our context, and with much (exponentially) worse dependence on p, q than what we derive here; due to the basic nature of these facts and also because obtaining them is not merely a technical adaptation of [107], full proofs are included in Section 5.1. The more novel case $\omega \in (0,1)$ of Proposition 6 is based on elementary geometric reasoning; again, due to the fundamental nature of this fact (as well as its connection to longstanding open questions), we prove it in Section 5.2 while taking care to obtain good asymptotic dependence as $\omega \to 0^+$. The proof of Theorem 9 appears in Section 5.3.1. Section 6 is devoted to several impossibility results, including those that were discussed above, such as Lemmas 2 and 13.

Acknowledgements. I am grateful to Noga Alon, Alexandr Andoni, Emmanuel Breuillard, Ilya Razenshteyn, John Pardon, Igor Rodnianski, Gideon Schechtman, Tasos Sidiropoulos, Vijay Sridhar and Ramon van Handel for helpful discussions and feedback. I also thank the anonymous referees for their careful reading of this manuscript and their useful corrections and feedback.

Throughout the work on this project (and many others), I benefited immensely from thought-provoking, inspiring and illuminating conversations with my friend and colleague Eli Stein, who passed away as the the final revision of this work was completed. He is dearly missed.

2. Nonlinear spectral gaps and duality

Suppose that $(\mathcal{M}, d_\mathcal{M})$ is a metric space, $p>0$ and $n\in\mathbb{N}$. If $\pi \in \triangle^{n-1}$ and $\mathsf{A}=(a_{ij})\in \mathsf{M}_n(\mathbb{R})$ is a stochastic and π-reversible matrix, then in analogy to (4) one measures the magnitude of the (reciprocal of) the nonlinear spectral gap of A relative to the kernel $d_\mathcal{M}^p : \mathcal{M}\times\mathcal{M}\to[0,\infty)$ through a quantity $\gamma(\mathsf{A}, d_\mathcal{M}^p)\in[0,\infty]$ which is defined [99] as the infimum over those $\gamma\in[0,\infty]$ such that

$$\forall\, x_1,\ldots,x_n\in\mathcal{M},\qquad \sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j d_\mathcal{M}(x_i,x_j)^p \leqslant \gamma \sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij} d_\mathcal{M}(x_i,x_j)^p. \tag{21}$$

Even though (21) is analogous to (4), a nonlinear spectral gap can differ markedly from the usual (reciprocal of the) gap in the (linear) spectrum; see [99, 100] for some of the subtleties and mysteries that arise from this generalization. As explained in [99], unless $\mathcal{M}$ is a singleton, if $\gamma(\mathsf{A}, d_\mathcal{M}^p)$ is finite, then $\lambda_2(\mathsf{A})$ is bounded away from 1 by a positive quantity that depends on $\gamma(\mathsf{A}, d_\mathcal{M}^p)$. So, the property of a matrix that is being considered here (determined by its interaction with the geometry of a metric space) is more stringent than requiring that it has a spectral gap in the classical sense.

A quite substantial theory of nonlinear spectral gaps was developed in a series of works, including [92, 57, 23, 65, 78, 79, 124, 116, 76, 98, 99, 107, 100, 103, 37, 46, 8], for several geometric applications, though many fundamental questions remain open. Establishing the utility of nonlinear spectral gaps to the results presented in the Introduction is a key conceptual contribution of the present work, and this underlies the algorithmic applications that were developed in [9, 10, 12, 11].

Remark 20. Fix $\Delta\in\mathbb{N}$. A sequence of Δ-regular graphs $\{\mathsf{G}_n=(V_n,E_n)\}_{n=1}^\infty$ is an expander with respect to a metric space $(\mathcal{M}, d_\mathcal{M})$ if $\lim_{n\to\infty}|V_n|=\infty$ and $\sup_{n\in\mathbb{N}}\gamma(\mathsf{A}_{\mathsf{G}_n}, d_\mathcal{M}^2)<\infty$, where we recall that $\mathsf{A}_{\mathsf{G}_n}$ is the normalized adjacency matrix of G_n. $\{\mathsf{G}_n\}_{n=1}^\infty$ is called a *super-expander* if it is an expander with respect to *every* superreflexive Banach space. It is a major open problem if a sequence $\{\mathsf{G}_n\}_{n=1}^\infty$ of bounded-degree regular graphs is a super-expander whenever $\sup_{n\in\mathbb{N}} 1/(1-\lambda_2(\mathsf{G}_n))<\infty$, i.e., when $\{\mathsf{G}_n\}_{n=1}^\infty$ is an expander in the classical sense. If Question 15 had a positive answer, then any classical expander would be a super-expander. Indeed, let $(X,\|\cdot\|_X)$ be a superreflexive Banach space. It suffices to prove that for every regular graph $\mathsf{G}=(\{1,\ldots,n\},E_\mathsf{G})$ we have

$$\gamma\big(\mathsf{A}_\mathsf{G}, \|\cdot\|_X^2\big) \lesssim_X \left(\frac{1}{1-\lambda_2(\mathsf{G})}\right)^{\frac{1}{\omega(X)}}. \tag{22}$$

To establish (22), by the hypothesized positive answer to Question 15, there are $\omega(X)\in(0,1]$ and $D(X)\in[1,\infty)$ such that the $\omega(X)$-snowflake of X embeds with average distortion $D(X)$ into ℓ_2. Proposition 6 with parameters $\omega=\omega(X)$, $D=D(X)$, $p=1$ and $q=2/\omega(X)$ shows that there exists $D'(X)\in[1,\infty)$ such that the $\omega(X)$-snowflake of X embeds with $(2/\omega(X))$-average distortion

$D'(X)$ into ℓ_2. If $x_1,\ldots,x_n\in X$, then an application of this conclusion to the uniform measure on $\{x_1,\ldots,x_n\}$ provides and embedding $f:X\to\ell_2$ which is $\omega(X)$-Hölder with constant $D'(X)$ and

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|f(x_i)-f(x_j)\|_{\ell_2}^{\frac{2}{\omega(X)}}\geqslant\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|_X^2\,. \tag{23}$$

By [23, Lemma 5.5] (see also (147) below), there exists a universal constant $C\in(0,\infty)$ such that

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|f(x_i)-f(x_j)\|_{\ell_2}^{\frac{2}{\omega(X)}}\leqslant\left(\frac{C}{\omega(X)\sqrt{1-\lambda_2(\mathsf{G})}}\right)^{\frac{2}{\omega(X)}}\frac{1}{|E_{\mathsf{G}}|}\sum_{\{i,j\}\in E_{\mathsf{G}}}\|f(x_i)-f(x_j)\|_{\ell_2}^{\frac{2}{\omega(X)}}. \tag{24}$$

The fact that f is $\omega(X)$-Hölder with constant $D'(X)$ gives

$$\frac{1}{|E_{\mathsf{G}}|}\sum_{\{i,j\}\in E_{\mathsf{G}}}\|f(x_i)-f(x_j)\|_{\ell_2}^{\frac{2}{\omega(X)}}\leqslant\frac{D'(X)^{\frac{2}{\omega(X)}}}{|E_{\mathsf{G}}|}\sum_{\{i,j\}\in E_{\mathsf{G}}}\|x_i-x_j\|_X^2\,. \tag{25}$$

By substituting (25) into (24), and then substituting the resulting inequality into (23), we see that

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|_X^2\leqslant\left(\frac{CD'(X)}{\omega(X)\sqrt{1-\lambda_2(\mathsf{G})}}\right)^{\frac{2}{\omega(X)}}\frac{1}{|E_{\mathsf{G}}|}\sum_{\{i,j\}\in E_{\mathsf{G}}}\|x_i-x_j\|_X^2\,.$$

Recalling (21), since this holds for every $x_1,\ldots,x_n\in X$, the justification of (22) is complete.

Prior to explaining how nonlinear spectral gaps relate to the embedding results that we stated in the Introduction, were recall some terminology and notation that will be used in what follows.

2.1. **Uniform convexity and smoothness.** Let $(X,\|\cdot\|_X)$ be a normed space and fix $p,q\geqslant1$ satisfying $p\leqslant2\leqslant q$. In the Introduction we recalled the traditional definitions of when it is said that X has moduli of smoothness and convexity of power type p and q, respectively. However, it is often convenient to work with an equivalent formulation of these properties due to Ball, Carlen and Lieb [21] (inspired by contributions of Pisier [122] and Figiel [51]), which we shall now recall.

The p-smoothness constant of X, denoted $\mathcal{S}_p(X)$, is the infimum over those $\mathcal{S}\in[1,\infty]$ such that

$$\forall\,x,y\in X,\qquad\frac{\|x+y\|_X^p+\|x-y\|_X^p}{2}-\|x\|_X^p\leqslant\mathcal{S}^p\|y\|_X^p. \tag{26}$$

By the triangle inequality we always have $\mathcal{S}_1(X)=1$. The q-convexity constant of X, denoted $\mathcal{K}_q(X)$, is the infimum over those $\mathcal{K}\in[1,\infty]$ such that

$$\forall\,x,y\in X,\qquad\|y\|_X^q\leqslant\mathcal{K}^q\left(\frac{\|x+y\|_X^q+\|x-y\|_X^q}{2}-\|x\|_X^q\right).$$

As shown in [21], X has moduli of smoothness and convexity of power type p and q, respectively, if and only if $\mathcal{S}_p(X)<\infty$ and $\mathcal{K}_q(X)<\infty$, respectively. It is beneficial to work with the coefficients $\mathcal{S}_p(X),\mathcal{K}_q(X)$ rather than the aforementioned classical moduli δ_X,ρ_X because they are well-behaved with respect to basic operations, an example of which is the duality $\mathcal{K}_{p/(p-1)}(X^*)=\mathcal{S}_p(X)$, as shown in [21]. Another example that is directly relevant to the present work is their especially clean behavior under complex interpolation; see Section 3.2.1 below. Further useful properties of these parameterizations of uniform convexity and uniform smoothness can be found in [99, Section 6.2].

2.2. **Markov type.** Following Ball [19], a metric space $(\mathcal{M}, d_{\mathcal{M}})$ is said to have Markov type $p \geqslant 1$ if there exists $M \geqslant 1$ with the following property. Suppose that $n \in \mathbb{N}$ and $\pi \in \triangle^{n-1}$. Then, for every stochastic and π-reversible matrix $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$, every $x_1, \ldots, x_n \in X$ and every $s \in \mathbb{N}$,

$$\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i(\mathsf{A}^s)_{ij} d_{\mathcal{M}}(x_i, x_j)^p\Big)^{\frac{1}{p}} \leqslant M s^{\frac{1}{p}} \Big(\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i a_{ij} d_{\mathcal{M}}(x_i, x_j)^p\Big)^{\frac{1}{p}}. \tag{27}$$

The infimum over those $M \geqslant 1$ which satisfy (27) is called the Markov type p constant of $\mathcal{M}$, and is denoted $\mathbf{M}_p(\mathcal{M})$. This nomenclature arises from a natural probabilistic interpretation [19] of (27) in terms of how stationary reversible Markov chains interact with the geometry of $\mathcal{M}$. We omit this description since it will not be needed below, though it is very important for other applications.

The following theorem, due[3] to [112, Theorem 2.3], will be used in the proof of Theorem 12.

Theorem 21. *For every $p \in [1, 2]$, every Banach space $(X, \|\cdot\|_X)$ with $\mathcal{S}_p(X) < \infty$ satisfies*

$$\mathbf{M}_p(X) \lesssim \mathcal{S}_p(X). \tag{28}$$

2.3. **Duality, compactness and Hölder extension.** The connection between nonlinear spectral gaps and Theorem 16 is through Theorem 22 below. For the first part of its statement, we refer to [62] for background on ultrapowers of Banach spaces. It suffices to say here that to each Banach space $(Z, \|\cdot\|_Z)$ one associates a (huge) Banach space $Z^{\mathcal{U}}$, called an ultrapower of Z, that has valuable compactness properties. Z is canonically isometric to a subspace of $Z^{\mathcal{U}}$, and any finite-dimensional linear subspace of $Z^{\mathcal{U}}$ embeds into Z with bi-Lipschitz distortion $1 + \varepsilon$ for any $\varepsilon > 0$. Thus, Z is essentially indistinguishable from any of its ultrapowers in terms of their finitary substructures. Due to Corollary 23 below, if one does not mind losing a constant factor that depends only on the moduli of uniform convexity and uniform smoothness, then one could drop all mention of ultrapowers in the ensuing discussion, and work throughout with the classical sequence space $\ell_q(Z)$ instead.

Theorem 22. *Suppose that $p, q, \mathcal{C} \geqslant 1$ and $p \leqslant q$. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space and $(Y, \|\cdot\|_Y)$ be a Banach space such that for every $n \in \mathbb{N}$, every symmetric stochastic matrix $\mathsf{A} \in \mathsf{M}_n(\mathbb{R})$ satisfies*

$$\gamma(\mathsf{A}, d_{\mathcal{M}}^p) \leqslant \mathcal{C}\gamma(\mathsf{A}, \|\cdot\|_Y^q). \tag{29}$$

Then, the $\frac{p}{q}$-snowflake of $\mathcal{M}$ embeds into some ultrapower of $\ell_q(Y)$ with q-average distortion $2\mathcal{C}^{\frac{1}{q}}$.

Furthermore, if in addition to the above assumption $\mathcal{M}$ has Markov type p and the modulus of uniform convexity of Y has power type q, then there exists $D \geqslant 1$ satisfying

$$D \lesssim \mathbf{M}_p(\mathcal{M})^{\frac{p}{q}} \mathcal{K}_q(Y) \mathcal{C}^{\frac{1}{q}},$$

such that the $\frac{p}{q}$-snowflake of $\mathcal{M}$ embeds into $\ell_q(Y)$ with q-average distortion D.

The following corollary is a combination of the second assertion of Theorem 22 and Theorem 21.

Corollary 23. *Suppose that $p, q, \mathcal{C} \geqslant 1$ and $p \leqslant 2 \leqslant q$. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces whose moduli of uniform smoothness and uniform convexity have power type p and q, respectively. Suppose also that for every $n \in \mathbb{N}$ and every symmetric stochastic matrix $\mathsf{A} \in \mathsf{M}_n(\mathbb{R})$ we have*

$$\gamma(\mathsf{A}, \|\cdot\|_X^p) \leqslant \mathcal{C}\gamma(\mathsf{A}, \|\cdot\|_Y^q).$$

Then, the $\frac{p}{q}$-snowflake of X embeds into $\ell_q(Y)$ with q-average distortion D, where

$$D \lesssim \mathcal{S}_p(X)^{\frac{p}{q}} \mathcal{K}_q(Y) \mathcal{C}^{\frac{1}{q}}. \tag{30}$$

[3]Formally, [112, Theorem 2.3] asserts that $\mathbf{M}_p(X) \lesssim_p \mathcal{S}_p(X)$ with the implicit constant tending to ∞ as $p \to 1^+$. However, [107, Theorem 4.3] adjusts the martingale argument of [112] so as to make that implicit constant universal.

Theorem 22 is deduced in Section 7 as a formal consequence of [107, Theorem 1.3], which was proved by a duality argument and implies the first assertion of Theorem 22 for average distortion embeddings of finite subsets of X. Those who are only interested in our application to metric dimension reduction could therefore skip Section 7, since for this finitary application one could use [107, Theorem 1.3] as a "black box." Theorem 22 is needed only for the full embedding statements in the Introduction, which treat arbitrary Borel measures and require that the embeddings have a controlled Lipschitz constant on all of X rather than only on the support of the given measure.

The deduction of Theorem 22 from [107, Theorem 1.3] amounts to a somewhat tedious but quite straightforward compactness argument, combined with a deep Hölder extension theorem that we use for the second assertion of Theorem 22, namely to obtain an embedding into $\ell_q(Y)$. It remains open if a loss as in (30) of a multiplicative factor that depends on the data $p, q, \mathcal{S}_p(X), \mathcal{K}_q(Y)$ is needed if one wishes to obtain an embedding into $\ell_q(Y)$ rather than into its ultrapower.

Remark 24. A version of Theorem 22 is available in which the target Y need not be a Banach space, but for that purpose further background in metric geometry is required (see [98, 107]; the pertinent concepts are metric Markov cotype q and W_q-barycentric spaces). We omit the discussion since its treatment in full generality will lead to a needlessly lengthy digression.

3. Complex interpolation

We briefly present background on the vector-valued complex interpolation method of Calderón [34] and Lions [87]; an extensive treatment can be found in e.g. [27, Chapter 4]. A pair of Banach spaces $(X, \|\cdot\|_X), (Z, \|\cdot\|_Z)$ over the complex scalars $\mathbb{C}$ is said to be a compatible pair of Banach spaces if they are both subspaces of a complex linear space W with $X+Z=W$. The space W is a complex Banach space under the norm $\|w\|_W = \inf\{\|x\|_X + \|z\|_Z :\ (x,z) \in X \times Z \text{ and } x+z=w\}$. Let $\mathcal{F}(X,Z)$ denote the space of all bounded continuous functions $\psi : \{\zeta \in \mathbb{C} :\ 0 \leqslant \Re(\zeta) \leqslant 1\} \to W$ that are analytic on $\{\zeta \in \mathbb{C} :\ 0 < \Re(\zeta) < 1\}$, such that for all $t \in \mathbb{R}$ we have $f(ti) \in X$ and $f(1+ti) \in Z$, the mappings $t \mapsto f(ti)$ and $t \mapsto f(1+ti)$ are continuous relative to the norms $\|\cdot\|_X$ and $\|\cdot\|_Z$, respectively, and $\lim_{|t|\to\infty} \|f(ti)\|_X = \lim_{|t|\to\infty} \|f(1+ti)\|_Z = 0$. To each $\theta \in [0,1]$ one associates as follows a Banach space $[X,Z]_\theta$. The underlying vector space is $\{\psi(\theta) :\ \psi \in \mathcal{F}(X,Z)\} \subset W$, and the norm of $w \in [X,Z]_\theta$ is $\|w\|_{[X,Z]_\theta} = \inf_{\{\psi \in \mathcal{F}(X,Z):\ \psi(\theta)=w\}} \max\{\sup_{t\in\mathbb{R}} \|\psi(ti)\|_X, \sup_{t\in\mathbb{R}} \|\psi(1+ti)\|_Z\}$. This turns $[X,Z]_\theta$ into a Banach space. By [27, Theorem 4.2.1] we have $[X,X]_\theta = X$ for $\theta \in [0,1]$.

By [27, Theorem 4.2.1], if $X \cap Z$ is dense in both X and Z, then $[X,Z]_0 = X$ and $[X,Z]_1 = Z$. In what follows, whenever we say that $(X, \|\cdot\|_X), (Z, \|\cdot\|_Z)$ is a compatible pair of Banach spaces we will tacitly assume that $X \cap Z$ is dense in both X and Z, thus ensuring that $\{[X,Z]_\theta\}_{\theta\in[0,1]}$ is a one-parameter family of Banach spaces starting at X and terminating at Z.

The *reiteration theorem* [34, Section 12.3] (see also [41] and the exposition in [27, Section 4.6]) asserts that if $(X, \|\cdot\|_X), (Z, \|\cdot\|_Z)$ is a compatible pair of complex Banach spaces, then

$$\forall\, \alpha, \beta, \theta \in [0,1], \qquad \big[[X,Z]_\alpha, [X,Z]_\beta\big]_\theta = [X,Z]_{(1-\theta)\alpha+\theta\beta}. \tag{31}$$

The equality in (31) means that the corresponding spaces are linearly isometric (over $\mathbb{C}$). Going forward, this is how all the ensuing equalities between complex Banach spaces are to be interpreted.

A basic property of vector-valued complex interpolation [87, 34] is that if $(X, \|\cdot\|_X), (Z, \|\cdot\|_Z)$ and $(U, \|\cdot\|_U), (V, \|\cdot\|_V)$ are two compatible pairs of complex Banach spaces and $T : X \cap Z \to U \cap V$ is a linear operator that extends to a bounded linear operator from $(X, \|\cdot\|_X)$ to $(U, \|\cdot\|_U)$ and from $(Z, \|\cdot\|_Z)$ to $(V, \|\cdot\|_V)$, then the following operator norm bounds hold true.

$$\forall\, \theta \in [0,1], \qquad \|T\|_{[X,Z]_\theta \to [U,V]_\theta} \leqslant \|T\|_{X\to U}^{1-\theta} \|T\|_{Z\to V}^{\theta}. \tag{32}$$

Fix $p \geqslant 1$ and $n \in \mathbb{N}$. For a complex Banach space $(X, \|\cdot\|_X)$ and a weight $\xi : \{1, \ldots, n\} \to [0, \infty)$, we denote (as usual) by $L_p(\xi; X)$ the vector space X^n equipped with the norm that is given by

$$\forall (x_1, \ldots, x_n) \in X^n, \qquad \|(x_1, \ldots, x_n)\|_{L_p(\xi;X)} \stackrel{\mathrm{def}}{=} \Big(\xi(1)\|x\|_X^p + \ldots + \xi(n)\|x_n\|_X^p\Big)^{\frac{1}{p}}.$$

In particular, if $\xi(1) = \ldots = \xi(n) = 1$, then $L_p(\xi; X) = \ell_p^n(X)$. Calderón's vector-valued version of Stein's interpolation theorem [132, Theorem 2] (see part(i) of §13.6 in [34] or Theorem 5.6.3 in [27]) asserts that if $(X, \|\cdot\|_X), (Z, \|\cdot\|_Z)$ is a compatible pair of complex Banach spaces, then for every $a, b \in [1, \infty]$, $\theta \in [0, 1]$ and $\xi, \zeta : \{1, \ldots, n\} \to [0, \infty)$ we have

$$[L_a(\xi; X), L_b(\zeta; Z)]_\theta = L_{\frac{ab}{\theta a + (1-\theta) b}}\Big(\xi^{\frac{(1-\theta)b}{\theta a + (1-\theta)b}} \zeta^{\frac{\theta a}{\theta a + (1-\theta)b}}; [X, Z]_\theta\Big). \tag{33}$$

The special case $\xi = \zeta$ of (33), in combination with (32), corresponds to the vector-valued version of the classical Riesz–Thorin interpolation theorem [129, 134].

3.1. **Nonlinear spectral gaps along an interpolation family.** Our main technical result is the following theorem which (under certain geometric assumptions) controls the growth of nonlinear spectral gaps along an interpolation family $\{[X, Z]_\theta\}_{\theta \in [0,1]}$ as $\theta \to 0^+$, in terms of their value at the endpoint $\theta = 1$. The relevance to Theorem 12 is through the duality of Theorem 22.

Theorem 25. *There is a universal constant $\alpha \geqslant 1$ with the following property. Fix $\theta \in (0, 1]$ and $(p, q) \in [1, 2] \times [2, \infty)$. Let $(X, \|\cdot\|_X), (Z, \|\cdot\|_Z)$ be a compatible pair of complex Banach spaces. Then, for every $n \in \mathbb{N}$, any symmetric stochastic matrix $\mathsf{A} \in \mathsf{M}_n(\mathbb{R})$ satisfies the following inequality.*

$$\gamma\big(\mathsf{A}, \|\cdot\|_{[X,Z]_\theta}^p\big) \leqslant \big(\alpha \mathcal{K}_q(Z)\big)^q \cdot \frac{\mathbf{M}_p([X, Z]_\theta)^p}{\theta} \gamma(\mathsf{A}, \|\cdot\|_Z^q) \overset{(28)}{\lesssim} \big(\alpha \mathcal{K}_q(Z)\big)^q \cdot \frac{\mathcal{S}_p([X, Z]_\theta)^p}{\theta} \gamma(\mathsf{A}, \|\cdot\|_Z^q). \tag{34}$$

By Theorem 22, Theorem 25 directly implies Theorem 16, yielding the following version of (20) with the implicit dependence of the constant factor on the relevant parameters specified explicitly.

$$D \lesssim \mathbf{M}_p([X, Z]_\theta)^{\frac{2p}{q}} \mathcal{K}_q(Z)^2 \Big(\frac{1}{\theta}\Big)^{\frac{1}{q}} \overset{(28)}{\lesssim} \mathcal{S}_p([X, Z]_\theta)^{\frac{2p}{q}} \mathcal{K}_q(Z)^2 \Big(\frac{1}{\theta}\Big)^{\frac{1}{q}}. \tag{35}$$

It is worthwhile to note in passing that, by the first part of Theorem 22, a smaller factor is achievable in (35) if one considers embeddings into an ultrapower of $\ell_q(Z)$ rather than into $\ell_q(Z)$ itself.

3.2. **Deduction of Theorem 12 from Theorem 16.** We first derive some preparatory estimates.

3.2.1. *Tracking the coefficients $\mathcal{S}_p([X, Z]_\theta)$ and $\mathcal{K}_q([X, Z]_\theta)$ as a function of $\theta \in [0, 1]$.* Suppose that $(X, \|\cdot\|_X), (Z, \|\cdot\|_Z)$ is a compatible pair of complex Banach spaces. Cwikel and Reisner estimated [42] the moduli of uniform convexity and uniform smoothness of $\{[X, Z]_\theta\}_{\theta \in [0,1]}$ in terms of the corresponding moduli of X and Z. By combining the bounds of [42] with [21], it follows that for every $p_1, p_2, \in (1, 2]$ and $q_1, q_2 \in [2, \infty)$ we have

$$\mathcal{S}_{\frac{p_1 p_2}{\theta p_1 + (1-\theta) p_2}}([X, Z]_\theta) \lesssim_{p_1, p_2} \mathcal{S}_{p_1}(X)^{1-\theta} \mathcal{S}_{p_2}(Z)^\theta,$$

and

$$\mathcal{K}_{\frac{q_1 q_2}{\theta q_1 + (1-\theta) q_2}}([X, Z]_\theta) \lesssim_{q_1, q_2} \mathcal{K}_{q_1}(X)^{1-\theta} \mathcal{K}_{q_2}(Z)^\theta.$$

We will next adjust the approach of [42] so as to obtain these estimates without any dependence on p_1, p_2, q_1, q_2 in the implicit multiplicative factors. Namely, we will demonstrate that

$$\mathcal{S}_{\frac{p_1 p_2}{\theta p_1 + (1-\theta) p_2}}([X, Z]_\theta) \leqslant \mathcal{S}_{p_1}(X)^{1-\theta} \mathcal{S}_{p_2}(Z)^\theta, \tag{36}$$

and

$$\mathcal{K}_{\frac{q_1 q_2}{\theta q_1 + (1-\theta) q_2}}([X, Z]_\theta) \leqslant \mathcal{K}_{q_1}(X)^{1-\theta} \mathcal{K}_{q_2}(Z)^\theta. \tag{37}$$

Removing the dependence on p_1, p_2, q_1, q_2 of the constant factors in the Cwikel–Reisner estimates is important in our context, as the parameters will be optimized so as to depend on other quantities that we wish to track. The ensuing reasoning is nothing more than an adaptation of [42].

Suppose that $p_1, p_2 \in [1,2]$ and that the smoothness constants $\mathcal{S}_{p_1}(X), \mathcal{S}_{p_2}(Z)$ are finite. Denote for simplicity $\mathcal{S}_1 = \mathcal{S}_{p_1}(X)$ and $\mathcal{S}_2 = \mathcal{S}_{p_2}(Z)$. Then by (26) we have

$$\forall\, y_1, y_2 \in Y, \qquad \|y_1+y_2\|_Y^{p_1} + \|y_1-y_2\|_Y^{p_1} \leqslant 2\|y_1\|_Y^{p_1} + 2\mathcal{S}_1^{p_1}\|y_2\|_Y^{p_1}, \tag{38}$$

and

$$\forall\, z_1, z_2 \in Z, \qquad \|z_1+z_2\|_Z^{p_2} + \|z_1-z_2\|_Z^{p_2} \leqslant 2\|z_1\|_Z^{p_2} + 2\mathcal{S}_2^{p_2}\|z_2\|_Z^{p_2}. \tag{39}$$

For every $\mathcal{S} > 0$ and $p \geqslant 1$ define $\xi(\mathcal{S}, p) : \{1,2\} \to (0,\infty)$ by $\xi(\mathcal{S},p)(1) = 2$ and $\xi(\mathcal{S},p)(2) = 2\mathcal{S}^p$. Also, denote the constant function $\mathbf{1}_{\{1,2\}}$ by $\zeta : \{1,2\} \to (0,\infty)$, i.e., $\zeta(1) = \zeta(2) = 1$. With this notation, if we consider the linear operator $T : (X+Z)\times(X+Z) \to (X+Z)\times(X+Z)$ that is given by setting $T(w_1,w_2) = (w_1+w_2, w_1-w_2)$ for every $w_1, w_2 \in Y+Z$, then

$$\|T\|_{L_{p_1}(\xi(\mathcal{S}_1,p_1);X)\to L_{p_1}(\zeta;Y)} \overset{(38)}{\leqslant} 1 \qquad \text{and} \qquad \|T\|_{L_{p_2}(\xi(\mathcal{S}_2,p_2);Z)\to L_{p_2}(\zeta;Z)} \overset{(39)}{\leqslant} 1. \tag{40}$$

Denoting $r = p_1p_2(\theta p_1 + (1-\theta)p_2)^{-1}$, observe that $\xi(\mathcal{S}_1,p_1)^{\frac{(1-\theta)r}{p_1}}\xi(\mathcal{S}_2,p_2)^{\frac{\theta r}{p_2}} = \xi(\mathcal{S}_1^{1-\theta}\mathcal{S}_2^{\theta}, r)$. Hence, by (33) we have $[L_{p_1}(\xi(\mathcal{S}_1,p_1);X), L_{p_2}(\xi(\mathcal{S}S_2,p_2);Z)]_\theta = L_r(\xi(\mathcal{S}_1^{1-\theta}\mathcal{S}_2^{\theta},r);[X,Z]_\theta)$ and also $[L_{p_1}(\zeta;X); L_{p_2}(\zeta;Z)]_\theta = L_r(\zeta,[X,Z]_\theta)$. In combination with (32) and (40), this implies that the norm of T as an operator from $L_r(\xi(\mathcal{S}_1^{1-\theta}\mathcal{S}_2^{\theta},r);[X,Z]_\theta)$ to $L_r(\zeta,[X,Z]_\theta)$ is at most 1. Thus,

$$\forall\, w_1, w_2 \in [X,Z]_\theta, \qquad \|w_1+w_2\|_{[X,Z]_\theta}^r + \|w_1-w_2\|_{[X,Z]_\theta}^r \leqslant 2\|w_1\|_{[X,Z]_\theta}^r + 2\big(\mathcal{S}_1^{1-\theta}\mathcal{S}_2^{\theta}\big)^r\|w_2\|_{[X,Z]_\theta}^r.$$

This is precisely (36). The bound (37) is justified mutatis mutandis via the same reasoning (only (36) will be used below); alternatively, one could derive (37) from (36) by a duality argument.

3.2.2. *Complexification.* To make Theorem 25, which treats complex Banach spaces, relevant to Theorem 12, which treats real normed spaces, we use a standard complexification procedure. Specifically, for a real normed space $(W,\|\cdot\|_W)$ associate as follows a complex normed space $(W_\mathbb{C},\|\cdot\|_{W_\mathbb{C}})$. The underlying vector space is $W_\mathbb{C} = W\times W$, which is viewed as a vector space over $\mathbb{C}$ by setting $(\alpha+\beta i)(x,y) = (\alpha x - \beta y, \beta x + \alpha y)$ for $\alpha,\beta\in\mathbb{R}$ and $x,y\in W$. The norm on $W_\mathbb{C}$ is given by

$$\forall\,(x,y)\in W\times W, \qquad \|(x,y)\|_{W_\mathbb{C}} = \bigg(\frac{1}{\pi}\int_0^{2\pi}\big\|(\cos\theta)x - (\sin\theta)y\big\|_W^2\,\mathrm{d}\theta\bigg)^{\frac12}. \tag{41}$$

The normalization of the integral in (41) was chosen so as to ensure that $x\mapsto(x,0)$ is an isometric embedding of W into $W_\mathbb{C}$. It is straightforward to check that for every $p\in[1,\infty]$ and $(x,y)\in W_\mathbb{C}$,

$$\|(x,y)\|_{W_\mathbb{C}} \asymp \big(\|x\|_W^p + \|y\|_W^p\big)^{\frac1p} = \|(x,y)\|_{\ell_p^2(W)}. \tag{42}$$

Hence, $\gamma(\mathsf{A},\|\cdot\|_X^p) \asymp \gamma(\mathsf{A},\|\cdot\|_{X_\mathbb{C}}^p)$ for every $n\in\mathbb{N}$ and every symmetric stochastic matrix $\mathsf{A}\in\mathsf{M}_n(\mathbb{R})$. Also, $\mathcal{S}_p(W_\mathbb{C}) \asymp \mathcal{S}_p(W)$ and $\mathcal{K}_q(W_\mathbb{C}) \asymp \mathcal{K}_q(W)$ for $p\in[1,2]$ and $q\in[2,\infty]$. If one were to let the implicit constants in these equivalences to depend on p,q, then they would follow from [52, 51, 21]. The fact that the constants can be taken to be universal follows by reasoning with more care, as done in [106, 99]; see specifically Lemma 6.3 and Corollary 6.4 of [99].

3.2.3. *Proof of Theorem 12.* Suppose that we are in the setting of Theorem 12. Thus, $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ are Banach spaces that satisfy $\mathsf{c}_Y(X), \mathcal{S}_p(X), \mathcal{K}_q(Y) < \infty$ where $1\leqslant p\leqslant 2\leqslant q<\infty$.

Fix $\mathsf{c} > \mathsf{c}_Y(X)$. Since Y is uniformly convex and hence (by the Milman–Pettis theorem [101, 121]) in particular reflexive, by a classical differentiation argument of Aronszajn [16], Christensen [38] and

Mankiewicz [88] (see also [26, Chapter 7] for a thorough treatment of such reductions to the linear setting) there exists a *linear* operator $T : X \to Y$ which satisfies

$$\forall\, x \in X, \qquad \|x\|_X \leqslant \|Tx\|_Y \leqslant \mathsf{c}\|x\|_X. \tag{43}$$

We define a normed space $(Z, \|\cdot\|_Z)$ by setting $Z = X$ and $\|x\|_Z \stackrel{\mathrm{def}}{=} \|Tx\|_Y$ for every $x \in X$. Thus, X and Z coincide as linear spaces and Z is linearly isometric to a subspace of Y, via the embedding T. Let $X_{\mathbb{C}}$ and $Z_{\mathbb{C}}$ be the complexifications of X and Z, respectively. Then $X_{\mathbb{C}}, Z_{\mathbb{C}}$ is a compatible pair of Banach spaces. So, we may consider the complex interpolation family $\{[X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta\}_{\theta\in[0,1]}$.

It follows from a substitution of (43) into the definition (41) that

$$\forall (x,y) \in X \times X, \qquad \|(x,y)\|_{X_{\mathbb{C}}} \leqslant \|(x,y)\|_{Z_{\mathbb{C}}} \leqslant \mathsf{c}\|(x,y)\|_{X_{\mathbb{C}}}. \tag{44}$$

Hence, the following operator norm bounds hold true for the formal identity $\mathsf{Id}_{X\times X} : X_{\mathbb{C}} \to X \times X$.

$$\|\mathsf{Id}_{X\times X}\|_{X_{\mathbb{C}}\to X_{\mathbb{C}}} \leqslant 1, \qquad \|\mathsf{Id}_{X\times X}\|_{X_{\mathbb{C}}\to Z_{\mathbb{C}}} \leqslant \mathsf{c}, \qquad \text{and} \qquad \|\mathsf{Id}_{X\times X}\|_{Z_{\mathbb{C}}\to X_{\mathbb{C}}} \leqslant 1. \tag{45}$$

The first inequality in (45) is tautological, and the rest of (45) is a restatement of (44).

For every $\theta \in [0, 1]$ we have

$$\|\mathsf{Id}_{X\times X}\|_{[X_{\mathbb{C}},Z_{\mathbb{C}}]_\theta\to X_{\mathbb{C}}} = \|\mathsf{Id}_{X\times X}\|_{[X_{\mathbb{C}},Z_{\mathbb{C}}]_\theta\to[X_{\mathbb{C}},X_{\mathbb{C}}]_\theta} \stackrel{(32)}{\leqslant} \|\mathsf{Id}_{X\times X}\|^{1-\theta}_{X_{\mathbb{C}}\to X_{\mathbb{C}}}\|\mathsf{Id}_{X\times X}\|^{\theta}_{Z_{\mathbb{C}}\to X_{\mathbb{C}}} \stackrel{(45)}{\leqslant} 1,$$

and

$$\|\mathsf{Id}_{X\times X}\|_{X_{\mathbb{C}}\to[X_{\mathbb{C}},Z_{\mathbb{C}}]_\theta} = \|\mathsf{Id}_{X\times X}\|_{[X_{\mathbb{C}},X_{\mathbb{C}}]_\theta\to[X_{\mathbb{C}},Z_{\mathbb{C}}]_\theta} \stackrel{(32)}{\leqslant} \|\mathsf{Id}_{X\times X}\|^{1-\theta}_{X_{\mathbb{C}}\to X_{\mathbb{C}}}\|\mathsf{Id}_{X\times X}\|^{\theta}_{X_{\mathbb{C}}\to Z_{\mathbb{C}}} \stackrel{(45)}{\leqslant} \mathsf{c}^\theta.$$

In other words, this simple reasoning yields the following bounds.

$$\forall (x,y) \in X \times X, \qquad \|(x,y)\|_{X_{\mathbb{C}}} \leqslant \|(x,y)\|_{[X_{\mathbb{C}},Z_{\mathbb{C}}]_\theta} \leqslant \mathsf{c}^\theta\|(x,y)\|_{X_{\mathbb{C}}}. \tag{46}$$

Fix $\sigma \geqslant 1$. By Theorem 16, applied with the value of D in (35), for any Borel probability measure μ on X (viewed as a subset of $X_{\mathbb{C}}$) there is $f : [X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta \to \ell_q(Z_{\mathbb{C}})$ satisfying

$$\forall\, x, y \in X, \qquad \|f(x) - f(y)\|_{\ell_q(Z_{\mathbb{C}})} \lesssim \frac{\mathbf{M}_\sigma([X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta)^{\frac{2\sigma}{q}}\mathcal{K}_q(Z_{\mathbb{C}})^2}{\theta^{\frac{1}{q}}}\|x-y\|^{\frac{\sigma}{q}}_{[X_{\mathbb{C}},Z_{\mathbb{C}}]_\theta}, \tag{47}$$

and

$$\iint_{X\times X} \|f(x) - f(y)\|^q_{\ell_q(Z_{\mathbb{C}})}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant \iint_{X\times X} \|x-y\|^\sigma_X\;\mu(x)\,\mathrm{d}\mu(y). \tag{48}$$

By (42) and the definition of Z, there is a linear map $S : \ell_q(Z_{\mathbb{C}}) \to \ell_q(Y)$ with $\|Sw\|_{\ell_q(Y)} \asymp \|w\|_{\ell_q(Z_{\mathbb{C}})}$ for all $w \in \ell_q(X \times X)$. Indeed, if we write $w = ((x_1, y_1), (x_2, y_2), \ldots)$ for $\{x_i\}_{i=1}^\infty, \{y_i\}_{i=1}^\infty \subset X$, then simply take $Sw = (Tx_1, Ty_1, Tx_2, Ty_2, \ldots)$. So, by considering $\phi = S \circ f : X \to \ell_q(Y)$, we get

$$\forall\, x, y \in X, \qquad \|\phi(x) - \phi(y)\|_{\ell_q(Y)} \stackrel{(46)\wedge(47)}{\lesssim} \frac{\mathsf{c}^{\frac{\sigma\theta}{q}}\mathbf{M}_\sigma([X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta)^{\frac{2\sigma}{q}}}{\theta^{\frac{1}{q}}}\mathcal{K}_q(Y)^2\|x-y\|^{\frac{\sigma}{q}}_X, \tag{49}$$

and

$$\iint_{X\times X} \|\phi(x) - \phi(y)\|^q_{\ell_q(Y)}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \stackrel{(48)}{\gtrsim} \iint_{X\times X} \|x-y\|^\sigma_X\;\mu(x)\,\mathrm{d}\mu(y). \tag{50}$$

The first part of Theorem 12 makes no assumption on the uniform smoothness of Y. So, apply (49) with $\sigma = p$ while noting that $\mathbf{M}_p([X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta) \leqslant \mathsf{c}^\theta\mathbf{M}_p(X_{\mathbb{C}}) \lesssim \mathsf{c}^\theta\mathcal{S}_p(X_{\mathbb{C}}) \lesssim \mathsf{c}^\theta\mathcal{S}_p(X)$, where the first step holds due to (46) and the second step is Theorem 21. We thus arrive at the Hölder condition

$$\forall\, x, y \in X, \qquad \|\phi(x) - \phi(y)\|_{\ell_q(Y)} \lesssim \frac{\mathsf{c}^{\frac{3p\theta}{q}}}{\theta^{\frac{1}{q}}}\mathcal{K}_q(Y)^2\mathcal{S}_p(X)^{\frac{2p}{q}}\|x-y\|^{\frac{p}{q}}_X. \tag{51}$$

The optimal choice of θ in (51) satisfies $\theta \asymp \frac{1}{\log(\mathsf{c}+1)}$, yielding the following explicit version of (17).

$$D \lesssim \mathcal{S}_p(X)^{\frac{2p}{q}} \mathcal{K}_q(Y)^2 \big(\log(\mathsf{c}_Y(X)+1)\big)^{\frac{1}{q}}. \tag{52}$$

We note in passing that with more care one obtains (51) with the term $\mathsf{c}^{\frac{3p\theta}{q}}$ replaced by $\mathsf{c}^{\frac{p\theta}{q}}$. But, upon choosing the optimal θ as we do here, this only influences the universal constant factor in (52).

For the second part of Theorem 12, we are now assuming that Y is more uniformly smooth than X, namely that $\mathcal{S}_r(Y) < \infty$ for some $r \in (p, 2]$. Under this stronger assumption, we fix $\varepsilon \in [0, \frac{r-p}{q}]$ and the aim is now to obtain an embedding of X into $\ell_q(Y)$ with higher regularity than in the first part of Theorem 12, namely an embedding of the $(\frac{p}{q}+\varepsilon)$-snowflake of X rather than of its $\frac{p}{q}$-snowflake. To this end, we apply (49) and (50) with $\sigma = p+\varepsilon q$ and $\theta \in [0,1]$ satisfying

$$\frac{\varepsilon q r}{(r-p)(p+\varepsilon q)} \leqslant \theta \leqslant 1. \tag{53}$$

Note that the range of possible values of θ in (53) is nonempty due to the assumption $\varepsilon \leqslant \frac{r-p}{q}$. The lower bound on θ in (53) is equivalent to $p+\varepsilon q \leqslant \frac{pr}{\theta p+(1-\theta)r}$, and therefore

$$\begin{aligned}\mathbf{M}_{p+\varepsilon q}([X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta) &\overset{(28)}{\lesssim} \mathcal{S}_{p+\varepsilon q}([X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta) \leqslant \mathcal{S}_{\frac{pr}{\theta p+(1-\theta)r}}([X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta)\\ &\overset{(36)}{\leqslant} \mathcal{S}_p(X_{\mathbb{C}})^{1-\theta}\mathcal{S}_r(Z_{\mathbb{C}})^{\theta} \lesssim \mathcal{S}_p(X)^{1-\theta}\mathcal{S}_r(Y)^{\theta},\end{aligned} \tag{54}$$

where the second step of (54) uses the fact that $p \mapsto \mathcal{S}_p(\cdot)$ is increasing (see [21] or [99, Section 6.2]) and the last step of (54) holds as Z is isometric to a subspace of Y. A substitution of (54) into (49) shows that the $(\frac{p}{q}+\varepsilon)$-snowflake of X embeds with q-average distortion D into $\ell_q(Y)$, where

$$D \lesssim \mathcal{S}_p(X)^{2\left(\frac{p}{q}+\varepsilon\right)}\mathcal{K}_q(Y)^2 \cdot \frac{\left(\frac{\mathcal{S}_r(Y)^2}{\mathcal{S}_p(X)^2}\mathsf{c}_Y(X)\right)^{\left(\frac{p}{q}+\varepsilon\right)\theta}}{\theta^{\frac{1}{q}}}. \tag{55}$$

By choosing θ so as to minimise the right hand side of (55) subject to the constraint (53), this leads to the following more refined version of the desired bound (18).

$$D \lesssim \begin{cases} \mathcal{S}_p(X)^{2\left(\frac{p}{q}+\varepsilon\right)}\mathcal{K}_q(Y)^2\left(\log\left(\frac{\mathcal{S}_r(Y)^2}{\mathcal{S}_p(X)^2}\mathsf{c}_Y(X)+2\right)\right)^{\frac{1}{q}} & \text{if } 0 \leqslant \varepsilon \leqslant \frac{r-p}{qr\log\left(\frac{\mathcal{S}_r(Y)^2}{\mathcal{S}_p(X)^2}\mathsf{c}_Y(X)+2\right)},\\ \left(\frac{r-p}{\varepsilon}\right)^{\frac{1}{q}}\mathcal{K}_q(Y)^2\mathcal{S}_r(Y)^{\frac{2\varepsilon r}{r-p}}\mathcal{S}_p(X)^{2p\left(1-\frac{\varepsilon q}{r-p}\right)}\mathsf{c}_Y(X)^{\frac{\varepsilon r}{r-p}} & \text{if } \frac{r-p}{qr\log\left(\frac{\mathcal{S}_r(Y)^2}{\mathcal{S}_p(X)^2}\mathsf{c}_Y(X)+2\right)} \leqslant \varepsilon \leqslant \frac{r-p}{q}.\end{cases}$$

This completes the deduction of Theorem 12 from Theorem 16. $\square$

Remark 26. Continuing with the notation and assumptions of the above proof of Theorem 12 in the special case when $Y = H$ is a Hilbert space and $q = 2$, Corollary 4.7 of [107] asserts[4] that

$$\gamma\Big(\mathsf{A}, \|\cdot\|^2_{[X_{\mathbb{C}},Z_{\mathbb{C}}]}\Big) \lesssim \frac{\mathcal{S}_p([X_{\mathbb{C}}, Z_{\mathbb{C}}])^2}{\theta^{\frac{2}{p}}\big(1-\lambda_2(\mathsf{A})\big)^{\frac{2}{p}}}. \tag{56}$$

Since $Z_{\mathbb{C}}$ is (isometrically) a Hilbert space and therefore $\mathcal{S}_2(Z_{\mathbb{C}}) = 1$, by (36) we have

$$\mathcal{S}_{\frac{2p}{p\theta+2(1-\theta)}}([X_{\mathbb{C}}, Z_{\mathbb{C}}]) \leqslant \mathcal{S}_p(X_{\mathbb{C}})^{1-\theta} \lesssim \mathcal{S}_p(X)^{1-\theta}.$$

[4]We note that in [107] (specifically, in the statement of [107, Theorem 4.5]) we have the following misprint: (56) is stated there for the transposed interpolation space $[Z_{\mathbb{C}}, X_{\mathbb{C}}]_\theta$ rather than the correct space $[X_{\mathbb{C}}, Z_{\mathbb{C}}]_\theta$ as above.

Arguing the same as above, by substituting this bound into (56) and using (46) we get that

$$\gamma\big(\mathsf{A},\|\cdot\|_X^2\big)\lesssim \frac{\mathsf{c}_2(X)^{2\theta}\mathcal{S}_p(X)^{2(1-\theta)}}{\theta^{\frac{2}{p}}\big(1-\lambda_2(\mathsf{A})\big)^{\theta+\frac{2(1-\theta)}{p}}}. \tag{57}$$

By choosing $\theta\in[0,1]$ so as to minimize the right hand side of (57), we see that

$$\gamma\big(\mathsf{A},\|\cdot\|_X^2\big)\lesssim \frac{\mathcal{S}_p(X)^2}{\big(1-\lambda_2(\mathsf{A})\big)^{\frac{2}{p}}}\bigg(\log\bigg(\frac{\mathsf{c}_2(X)^p\big(1-\lambda_2(\mathsf{A})\big)^{1-\frac{p}{2}}}{\mathcal{S}_p(X)^p}+1\bigg)\bigg)^{\frac{2}{p}}. \tag{58}$$

In particular, if $\dim(X)=k\in\{2,3,\ldots\}$ and $p=1$, by John's theorem (58) implies that

$$\gamma\big(\mathsf{A},\|\cdot\|_X^2\big)\lesssim \bigg(\frac{\log\big(\mathsf{c}_2(X)\sqrt{1-\lambda_2(\mathsf{A})}+1\big)}{1-\lambda_2(\mathsf{A})}\bigg)^2\lesssim \frac{(\log k)^2}{\big(1-\lambda_2(\mathsf{A})\big)^2}. \tag{59}$$

Note that if one is interested only in the rightmost quantity in (59) as an upper bound on $\gamma\big(\mathsf{A},\|\cdot\|_X^2\big)$, then one simply needs to substitute $\theta\asymp 1/\log k$ into (56) and use (46) as above. This slightly weaker estimate can be rewritten as the assertion that there exists a universal constant $\mathsf{K}\geqslant 1$ such that

$$\begin{aligned}\dim(X)&\geqslant \exp\bigg(\frac{1-\lambda_2(\mathsf{A})}{\mathsf{K}}\sqrt{\gamma\big(\mathsf{A},\|\cdot\|_X^2\big)}\bigg)\\ &\overset{(21)}{=}\sup_{x_1,\ldots,x_n\in X}\exp\Bigg(\frac{1-\lambda_2(\mathsf{A})}{\mathsf{K}}\bigg(\frac{\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|_X^2}{\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|x_i-x_j\|_X^2}\bigg)^{\frac12}\Bigg).\end{aligned} \tag{60}$$

(60) corresponds to the case $p=2$ of the matrix-dimension inequality (15). As explained in the Introduction, (60) is a formal consequence of the average John theorem of Theorem 1. We do not see how to deduce Theorem 1 formally from (60); we conjecture that such a reverse implication is impossible in general but we did not devote substantial effort to obtain a counterexample.

4. Proof of Theorem 25

Suppose that $(\mathcal{M},d_{\mathcal{M}})$ is a metric space, $n\in\mathbb{N}$ and $p\in(0,\infty)$. Following [99] and in analogy to (21), the (reciprocal of) the *nonlinear absolute spectral gap* with respect to $d_{\mathcal{M}}^p$ of a symmetric stochastic matrix $\mathsf{A}=(a_{ij})\in\mathsf{M}_n(\mathbb{R})$, denoted $\gamma_+(\mathsf{A},d_{\mathcal{M}}^p)$, is the smallest $\gamma_+\in(0,\infty]$ such that

$$\forall\, x_1,\ldots,x_n,y_1,\ldots,y_n\in\mathcal{M},\qquad \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_{\mathcal{M}}(x_i,y_j)^p\leqslant\frac{\gamma_+}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_{\mathcal{M}}(x_i,y_j)^p. \tag{61}$$

The reason for this terminology is that by simple linear-algebraic considerations (e.g. [99]) one has

$$\gamma_+(\mathsf{A},d_{\mathbb{R}}^2)=\gamma_+(\mathsf{A},|\cdot-\cdot|^2:\mathbb{R}\times\mathbb{R}\to\mathbb{R})=\frac{1}{1-\max_{i\in\{2,\ldots,n\}}|\lambda_i(\mathsf{A})|}.$$

Observe that the definition directly implies that $\gamma(\mathsf{A},d_{\mathcal{M}}^p)\leqslant\gamma_+(\mathsf{A},d_{\mathcal{M}}^p)$.

We record for ease of later reference the following elementary relation [107, Lemma 2.3] between nonlinear spectral gaps and their absolute counterparts. In its formulation, as well as in the rest of what follows, for every $n\in\mathbb{N}$ the n-by-n identity matrix is denoted $\mathsf{I}_n\in\mathsf{M}_n(\mathbb{R})$.

Lemma 27. *Fix $q\in[1,\infty)$ and $n\in\mathbb{N}$. For every symmetric stochastic matrix $\mathsf{A}\in\mathsf{M}_n(\mathbb{R})$ and every metric space $(\mathcal{M},d_{\mathcal{M}})$ we have*

$$2\gamma(\mathsf{A},d_{\mathcal{M}}^q)\leqslant\gamma_+\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A},d_{\mathcal{M}}^q\Big)\leqslant 2^{2q+1}\gamma(\mathsf{A},d_{\mathcal{M}}^q). \tag{62}$$

Even though our ultimate goal here is to bound nonlinear spectral gaps, one of the advantages of considering nonlinear absolute spectral gaps is that, in the case of uniformly convex normed spaces, they have a useful connection to operator norms. Specifically, suppose that $(X,\|\cdot\|_X)$ is a normed space, $n\in\mathbb{N}$ and $\mathsf{A}\in\mathsf{M}_n(\mathbb{R})$ is symmetric and stochastic. For $p\in[1,\infty)$ let $\ell_p^n(X)_0\subset\ell_p^n(X)$ denote the subspace of those $(x_1,\ldots,x_n)\in X^n$ for which $\sum_{i=1}^n x_i=0$. If $\mathsf{Id}_X:X\to X$ denotes the formal identity operator on X, then, since A is symmetric and stochastic, the operator $\mathsf{A}\otimes\mathsf{Id}_X:X^n\to X^n$ preserves $\ell_p^n(X)_0$, where we recall that $(\mathsf{A}\otimes\mathsf{Id}_X)(x_1,\ldots,x_n)=(\sum_{j=1}^n a_{1j}x_j,\ldots,\sum_{j=1}^n a_{nj}x_j)$ for each $(x_1,\ldots,x_n)\in X^n$. One can therefore consider the operator norm $\|\mathsf{A}\otimes\mathsf{Id}_X\|_{\ell_p^n(X)_0\to\ell_p^n(X)_0}$.

The following lemma coincides with [99, Lemma 6.1].

Lemma 28. *For every $q\in[1,\infty)$, every $n\in\mathbb{N}$, every symmetric and stochastic matrix $\mathsf{A}\in\mathsf{M}_n(\mathbb{R})$, and every normed space $(X,\|\cdot\|_X)$, we have*

$$\gamma_+\big(\mathsf{A},\|\cdot\|_X^q\big)\leqslant\bigg(1+\frac{4}{1-\|\mathsf{A}\otimes\mathsf{Id}_X\|_{\ell_q^n(X)_0\to\ell_q^n(X)_0}}\bigg)^q. \tag{63}$$

Even a weak converse to Lemma 28, namely the ability to bound $\|\mathsf{A}\otimes\mathsf{Id}_X\|_{\ell_q^n(X)_0\to\ell_q^n(X)_0}$ from above away from 1 by a quantity that may depend on $\gamma_+(\mathsf{A},\|\cdot\|_X^q)$ and q but not on n, fails for a general normed space $(X,\|\cdot\|_X)$; see [99, Section 6.1]. However, if $(X,\|\cdot\|_X)$ is uniformly convex, then we have the following converse statement along these lines, due to [99, Lemma 6.6].

Lemma 29. *Suppose that $q\in[2,\infty)$ and that $(X,\|\cdot\|_X)$ is a normed space for which $\mathcal{K}_q(X)<\infty$. Then, for every $n\in\mathbb{N}$ and every symmetric and stochastic matrix $\mathsf{A}\in\mathsf{M}_n(\mathbb{R})$, we have*

$$\|\mathsf{A}\otimes\mathsf{Id}_X\|_{\ell_q^n(X)_0\to\ell_q^n(X)_0}\leqslant\bigg(1-\frac{1}{(2^{q-1}-1)\mathcal{K}_q(X)^q\gamma_+\big(\mathsf{A},\|\cdot\|_X^q\big)}\bigg)^{\frac{1}{q}}. \tag{64}$$

The proof of the following lemma is an adaptation of the proof of [107, Theorem 4.15].

Lemma 30. *Suppose that $p\in[1,\infty)$ and $q\in[2,\infty)$. Let $(X,\|\cdot\|_X),(Z,\|\cdot\|_Z)$ be a compatible pair of complex Banach spaces such that $\mathcal{K}_q(Z)<\infty$. Fix $n\in\mathbb{N}$ and a symmetric stochastic matrix $\mathsf{A}\in\mathsf{M}_n(\mathbb{R})$. If $\theta\in(0,1]$ and $s\in\mathbb{N}$ satisfy*

$$s\geqslant\big(8\mathcal{K}_q(Z)\big)^q\gamma\big(\mathsf{A},\|\cdot\|_Z^q\big)\max\left\{\frac{q}{\theta},p,\frac{p(q-1)}{p-1}\right\}, \tag{65}$$

then,

$$\gamma\bigg(\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s,\|\cdot\|_{[X,Z]_\theta}^p\bigg)\leqslant\gamma_+\bigg(\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s,\|\cdot\|_{[X,Z]_\theta}^p\bigg)\leqslant e^{O(p)}.$$

In particular, $\gamma\Big(\big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\big)^s,\|\cdot\|_{[X,Z]_\theta}^p\Big)\lesssim 1$ for some $s\asymp\frac{1}{\theta}\big(9\mathcal{K}_q(Z)\big)^q\gamma\big(\mathsf{A},\|\cdot\|_Z^q\big)$.

Proof. Suppose first that $\frac{q}{\theta+(1-\theta)q}\leqslant p\leqslant\frac{q}{\theta}$, or equivalently that $\max\left\{\frac{q}{\theta},p,\frac{p(q-1)}{p-1}\right\}=\frac{q}{\theta}$. Then,

$$\frac1p=\frac{1-\theta}{r}+\frac{\theta}{q},\qquad\text{where}\qquad r\stackrel{\mathrm{def}}{=}\frac{(1-\theta)pq}{q-\theta p}\in[1,\infty]. \tag{66}$$

For p in the above range, the assumption (65) on s is equivalent to the bound

$$\big(9\mathcal{K}_q(Z)\big)^q\gamma\big(\mathsf{A},\|\cdot\|_Z^q\big)\leqslant\frac{\theta s}{q}. \tag{67}$$

Let $\mathsf{J_n} \in \mathsf{M}_n(\mathbb{R})$ be the matrix all of whose entries equal $\frac{1}{n}$. Set $\mathsf{Q}_n \stackrel{\mathrm{def}}{=} \mathsf{I}_n - \mathsf{J}_n$. By convexity and the triangle inequality, $\|\mathsf{Q}_n \otimes \mathsf{Id}_W\|_{\ell_a(W)\to\ell_a(W)_0} \leqslant 2$ for any Banach space $(W, \|\cdot\|_W)$ and $a \geqslant 1$. So,

$$\begin{aligned}\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s\mathsf{Q}_n\otimes\mathsf{Id}_W\Big\|_{\ell_a^n(W)\to\ell_a^n(W)}&\\ &=\bigg\|\bigg(\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)\otimes\mathsf{Id}_W\bigg)^s(\mathsf{Q}_n\otimes\mathsf{Id}_W)\bigg\|_{\ell_a^n(W)\to\ell_a^n(W)_0}\\ &\leqslant\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)\otimes\mathsf{Id}_W\Big\|^s_{\ell_a^n(W)_0\to\ell_a^n(W)_0}\big\|\mathsf{Q}_n\otimes\mathsf{Id}_W\big\|_{\ell_a^n(W)\to\ell_a^n(W)_0}\\ &\leqslant 2\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)\otimes\mathsf{Id}_W\Big\|^s_{\ell_a^n(W)_0\to\ell_a^n(W)_0}.\end{aligned}\tag{68}$$

Due to (66), by (33) we have $\ell_p^n([X,Z]_\theta) = [\ell_r^n(X), \ell_q^n(Z)]_\theta$. Consequently,

$$\begin{aligned}\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s\otimes\mathsf{Id}_{[X,Z]_\theta}\Big\|_{\ell_p^n([X,Z]_\theta)_0\to\ell_p^n([X,Z]_\theta)_0}&\leqslant\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s\mathsf{Q}_n\otimes\mathsf{Id}_{[X,Z]_\theta}\Big\|_{\ell_p^n([X,Z]_\theta)\to\ell_p^n([X,Z]_\theta)}\\ &\stackrel{(32)\wedge(68)}{\leqslant}2\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)\otimes\mathsf{Id}_X\Big\|^{(1-\theta)s}_{\ell_r^n(X)_0\to\ell_r^n(X)_0}\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)\otimes\mathsf{Id}_Z\Big\|^{\theta s}_{\ell_q^n(Z)_0\to\ell_q^n(Z)_0}.\end{aligned}\tag{69}$$

We claim that

$$\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)\otimes\mathsf{Id}_X\Big\|_{\ell_r^n(X)_0\to\ell_r^n(X)_0}\leqslant 1\quad\text{and}\quad\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)\otimes\mathsf{Id}_Z\Big\|_{\ell_q^n(Z)_0\to\ell_q^n(Z)_0}\leqslant\Big(1-\frac{q}{\theta s}\Big)^{\frac1q}.\tag{70}$$

Indeed, the first inequality in (70) follows from the convexity of the $\ell_r^n(X)_0$ norm, because $\frac12\mathsf{I}_n+\frac12\mathsf{A}$ is a stochastic matrix. The second inequality in (70) is justified as follows.

$$\begin{aligned}\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)\otimes\mathsf{Id}_Z\Big\|_{\ell_q^n(Z)_0\to\ell_q^n(Z)_0}&\stackrel{(64)}{\leqslant}\bigg(1-\frac{1}{2^{q-1}\mathcal{K}_q(Z)^q\gamma_+\big(\frac12\mathsf{I}_n+\frac12\mathsf{A},\|\cdot\|_Z^q\big)}\bigg)^{\frac1q}\\ &\stackrel{(62)}{\leqslant}\bigg(1-\frac{1}{\big(8\mathcal{K}_q(Z)\big)^q\gamma\big(\mathsf{A},\|\cdot\|_Z^q\big)}\bigg)^{\frac1q}\stackrel{(67)}{\leqslant}\Big(1-\frac{q}{\theta s}\Big)^{\frac1q}.\end{aligned}$$

A substitution of (70) into (69) gives

$$\Big\|\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s\otimes\mathsf{Id}_{[X,Z]_\theta}\Big\|_{\ell_p^n([X,Z]_\theta)_0\to\ell_p^n([X,Z]_\theta)_0}\leqslant 2\Big(1-\frac{q}{\theta s}\Big)^{\frac{\theta s}{q}}\leqslant\frac2e.\tag{71}$$

Hence,

$$\gamma_+\bigg(\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s,\|\cdot\|^p_{[X,Z]_\theta}\bigg)\stackrel{(63)\wedge(71)}{\leqslant}\Big(\frac{5e-2}{e-2}\Big)^p=e^{O(p)}.$$

This proves Lemma 30 when $\frac{q}{\theta+(1-\theta)q}\leqslant p\leqslant\frac{q}{\theta}$. If $p\in[1,\infty]\setminus[\frac{q}{\theta+(1-\theta)q},\frac{q}{\theta}]$, then define

$$\tau\stackrel{\mathrm{def}}{=}\min\left\{\frac qp,\frac{q(p-1)}{p(q-1)}\right\}\qquad\text{and}\qquad\alpha\stackrel{\mathrm{def}}{=}1-(1-\theta)\max\left\{\frac{p}{p-q},\frac{p(q-1)}{q-p}\right\}.$$

One checks that $\theta=(1-\tau)\alpha+\tau$. Also, the assumption on p ensures that $\alpha,\tau\in[0,1]$ and

$$\max\left\{\frac q\tau,p,\frac{p(q-1)}{p-1}\right\}=\frac q\tau=\max\left\{p,\frac{p(q-1)}{p-1}\right\}.$$

By the reiteration theorem (31), we have $[X,Z]_\theta=\big[[X,Z]_\alpha,[X,Z]_1\big]_\tau=\big[[X,Z]_\alpha,Z\big]_\tau$. So, the rest of Lemma 30 becomes the case that we already proved upon replacing X by $[X,Z]_\alpha$ and θ by τ. $\square$

Completion of the proof of Theorem 25. Continue with the notation and assumptions of Theorem 25. Fix $x_1,\ldots,x_n\in[X,Z]_\theta$. By the special case $\mathcal{M}=Y=[X,Z]_\theta$ of Proposition 6 (which is yet to be proven, but this is done in Section 5.2), we know that the $\frac{p}{2}$-snowflake of $[X,Z]_\theta$ embeds with quadratic average distortion $O(1)$ back into $[X,Z]_\theta$. An application of this conclusion to the uniform measure on $\{x_1,\ldots,x_n\}$ provides new points $y_1,\ldots,y_n\in[X,Z]_\theta$ satisfying

$$\forall\, i,j\in\{1,\ldots,n\},\qquad \|y_i-y_j\|_{[X,Z]_\theta}\lesssim \|x_i-x_j\|_{[X,Z]_\theta}^{\frac{p}{2}}, \tag{72}$$

and

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|y_i-y_j\|_{[X,Z]_\theta}^2 \geqslant \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|x_i-x_j\|_{[X,Z]_\theta}^p. \tag{73}$$

Write $\mathsf{A}=(a_{ij})\in\mathsf{M}_n(\mathbb{R})$. By Lemma 30 we know that

$$\exists\, s\in\mathbb{N},\qquad s\asymp\frac{1}{\theta}\big(9\mathcal{K}_q(Z)\big)^q\gamma\big(\mathsf{A},\|\cdot\|_Z^q\big)\qquad\text{and}\qquad \gamma\bigg(\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s,\|\cdot\|_{[X,Z]_\theta}^p\bigg)\lesssim 1. \tag{74}$$

Fixing $s\in\mathbb{N}$ as in (74), we reason as follows.

$$\begin{aligned}
\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|x_i-x_j\|_{[X,Z]_\theta}^p &\overset{(73)}{\leqslant} \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|y_i-y_j\|_{[X,Z]_\theta}^2\\
&\overset{(21)}{\leqslant} \frac1n\gamma\bigg(\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s,\|\cdot\|_{[X,Z]_\theta}^p\bigg)\sum_{i=1}^n\sum_{j=1}^n\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s_{ij}\|y_i-y_j\|_{[X,Z]_\theta}^2\\
&\overset{(74)\wedge(72)}{\lesssim} \frac1n\sum_{i=1}^n\sum_{j=1}^n\Big(\frac12\mathsf{I}_n+\frac12\mathsf{A}\Big)^s_{ij}\|x_i-x_j\|_{[X,Z]_\theta}^p\\
&\overset{(27)}{\leqslant} \frac{\mathbf{M}_p([X,Z]_\theta)^p s}{n}\sum_{i=1}^n\sum_{j=1}^n\frac12 a_{ij}\|x_i-x_j\|_{[X,Z]_\theta}^p\\
&\overset{(74)}{\asymp} \frac{\mathbf{M}_p([X,Z]_\theta)^p\big(9\mathcal{K}_q(Z)\big)^q\gamma\big(\mathsf{A},\|\cdot\|_Z^q\big)}{\theta n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|x_i-x_j\|_{[X,Z]_\theta}^p.
\end{aligned}$$

Because this holds for every $x_1,\ldots,x_n\in[X,Z]_\theta$, by the definition (21) this is precisely (34). $\square$

Remark 31. In [9, Section 6] and [111, Section 5] we presented a proof of the quadratic inequality (59) while stripping away any reference to complex interpolation; it amounts to an expository repackaging of the same mechanism as our reasoning here, but is more elementary. Applying this proof to the vectors $y_1,\ldots,y_n\in X$ that satisfy (72) and (73) with $p=1$ and $\theta=0$ (namely, using the special case $\mathcal{M}=Y=X$ and $\omega=\frac12$ of Proposition 6), and then invoking duality through Theorem 22, we get an interpolation-free proof of Theorem 1. By applying Lemma 28 and Lemma 29 in place of the linear-algebraic reasoning in [9, 111], one also obtains mutatis mutandis an interpolation-free proof of the first part (17) of Theorem 12, albeit with a worse asymptotic dependence on q in the implicit factor in (17). We do not see how to derive the second part (18) of Theorem 12 without appealing to complex interpolation. Incorporation of finite-dimensional reasoning in an interpolation argument, as we do here, is also used in our subsequent works [10, 11]; if interpolation could be avoided in the context of [10, 11], then it would be worthwhile to do so, potentially (depending on the resulting proof) with algorithmic ramifications.

5. Auxiliary embedding results

Here we will prove Proposition (6) and show how Theorem 9 (matrix-dimension inequality with what we conjecture is the asymptotically optimal dependence on p) follows from Theorem 1.

Henceforth, all balls in a metric space are closed, i.e., for a metric space $(\mathcal{M}, d_\mathcal{M})$, a point $x \in \mathcal{M}$ and a radius $r \in [0, \infty]$, we write $B_\mathcal{M}(x, r) = \{y \in \mathcal{M} : \ d_\mathcal{M}(y, x) \leqslant r\}$. Given a Borel probability measure μ on $\mathcal{M}$ and $p \geqslant 1$, when in the Introduction we discussed the p-average distortion of an embedding of the metric probability space $(\mathcal{M}, d_\mathcal{M}, \mu)$ into some Banach space, we did not impose the integrability requirement $\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x, y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) < \infty$. However, it is simple to dispose of the (inconsequential) case of those Borel probability measures μ on $\mathcal{M}$ for which $d_\mathcal{M}(\cdot, \cdot) \notin L_p(\mu\times\mu)$ through the following straightforward consequence of the triangle inequality.

If $\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x, y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = \infty$, then for every $z \in \mathcal{M}$ and $r > 0$ we have

$$\infty = \bigg(\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x, y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac{1}{p}} \leqslant \bigg(\iint_{\mathcal{M}\times\mathcal{M}} \big(d_\mathcal{M}(x, z) + d_\mathcal{M}(y, z)\big)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac{1}{p}}$$
$$\leqslant 2\bigg(\int_\mathcal{M} d_\mathcal{M}(x, z)^p \,\mathrm{d}\mu(x)\bigg)^{\frac{1}{p}} \leqslant 2\bigg((2r)^p + \int_{\mathcal{M}\smallsetminus B_\mathcal{M}(z,2r)} d_\mathcal{M}(x, z)^p \,\mathrm{d}\mu(x)\bigg)^{\frac{1}{p}}.$$

So $\int_{\mathcal{M}\smallsetminus B_\mathcal{M}(z,2r)} d_\mathcal{M}(x, z)^p \,\mathrm{d}\mu(x) = \infty$. There is $r > 0$ for which $\mu(B_\mathcal{M}(z, r)) > 0$ (μ is a probability measure). Noting that $d_\mathcal{M}(x, z) - d_\mathcal{M}(y, z) \geqslant \frac{1}{2} d_\mathcal{M}(x, z)$ when $(x, y) \in \big(\mathcal{M} \smallsetminus B_\mathcal{M}(z, 2r)\big) \times B_\mathcal{M}(z, r)$,

$$\iint_{\mathcal{M}\times\mathcal{M}} \big|d_\mathcal{M}(x, z) - d_\mathcal{M}(y, z)\big|^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant \frac{1}{2}\iint_{\left(\mathcal{M}\smallsetminus B_\mathcal{M}(z,2r)\right)\times B_\mathcal{M}(z,r)} d_\mathcal{M}(x, z)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)$$
$$= \frac{1}{2}\mu\big(B_\mathcal{M}(z, r)\big) \int_{\mathcal{M}\smallsetminus B_\mathcal{M}(z,2r)} d_\mathcal{M}(x, z)^p \,\mathrm{d}\mu(x) = \infty = \iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x, y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y).$$

Thus the 1-Lipschitz function $x \mapsto d_\mathcal{M}(x, z) \in \mathbb{R}$ is an embedding of $(\mathcal{M}, d_\mathcal{M}, \mu)$ into the real line with μ average distortion 1. Due to this (trivial) observation, we will be allowed to assume that we have $\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x, y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) < \infty$ whenever needed in the ensuing discussion.

5.1. Section 7.4 of [107] revisited.

The special case $\omega = 1$ of Proposition (6) was essentially proved in [107, Section 7.4]. Here we will derive this case of Proposition (6) while obtaining asymptotically better bounds than those of [107] and also removing an additional hypothesis (on Lipschitz extendability) that arose in the context of [107] but is not needed for Proposition (6) as stated here.

Let $(\mathcal{M}, d_\mathcal{M})$ be a metric space and fix [22] an arbitrary isometric embedding $j : \mathcal{M} \to C[0, 1]$ of $\mathcal{M}$ into the space of continuous functions on the interval $[0, 1]$, equipped (as usual) with the supremum norm $\|\cdot\|_{C[0,1]}$; it is more convenient (but not crucial) to work below with such an embedding rather than the Fréchet embedding into ℓ_∞ due to the separability of the target space. Suppose that $q \geqslant 1$ and let μ be a Borel probability measure on $\mathcal{M}$ such that

$$\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x, y)^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) < \infty. \tag{75}$$

By (75) and Fubini's theorem, $\int_\mathcal{M} d_\mathcal{M}(u, x)^q \,\mathrm{d}\mu(x) < \infty$ for some $u \in \mathcal{M}$. Since

$$\forall\, x \in \mathcal{M}, \qquad \|j(x)\|_{C[0,1]} \leqslant \|j(u)\|_{C[0,1]} + \|j(x) - j(u)\|_{C[0,1]} = \|j(u)\|_{C[0,1]} + d_\mathcal{M}(x, u),$$

we have $\int_\mathcal{M} \|j(x)\|^q_{C[0,1]} \,\mathrm{d}\mu(x) < \infty$. Hence $\int_\mathcal{M} \|j(x)\|_{C[0,1]} \,\mathrm{d}\mu(x) < \infty$, because $q \geqslant 1$ and μ is a probability measure. By Bochner's integrability criterion (see e.g. [26, Chapter 5]), this implies that the Bochner integral $\int_{\mathcal{M}} \jmath(x)\,d\mu(x)$ is a well-defined element of $C[0,1]$. We can therefore denote

$$\mathbf{I}_q = \mathbf{I}_q(\mu,\jmath) \stackrel{\mathrm{def}}{=} \left(\int_{\mathcal{M}} \left\|\jmath(x) - \int_{\mathcal{M}} \jmath(w)\,d\mu(w)\right\|_{C[0,1]}^q d\mu(x)\right)^{\frac{1}{q}}. \tag{76}$$

Observe that, using the fact that $\jmath$ is isometry and the triangle inequality in $L_q(\mu\times\mu)$, we have

$$\left(\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,d\mu(x)\,d\mu(y)\right)^{\frac{1}{q}} = \left(\iint_{\mathcal{M}\times\mathcal{M}} \|\jmath(x)-\jmath(y)\|_{C[0,1]}^q\,d\mu(x)\,d\mu(y)\right)^{\frac{1}{q}} \leqslant 2\mathbf{I}_q. \tag{77}$$

Since $q\geqslant 1$ and μ is a probability measure, by Jensen's inequality and the fact that $\jmath$ is isometry,

$$\mathbf{I}_q \leqslant \left(\iint_{\mathcal{M}\times\mathcal{M}} \|\jmath(x)-\jmath(w)\|_{C[0,1]}^q\,d\mu(x)\,d\mu(w)\right)^{\frac{1}{q}} = \left(\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,d\mu(x)\,d\mu(y)\right)^{\frac{1}{q}}. \tag{78}$$

In what follows, for $\tau\geqslant 1$ we will also consider a subset $A_\tau = A_\tau(\mu,\jmath,q)\subset\mathcal{M}$ that is defined by

$$\begin{aligned} A_\tau &\stackrel{\mathrm{def}}{=} \left\{x\in\mathcal{M}:\ \left\|\jmath(x)-\int_{\mathcal{M}}\jmath(w)\,d\mu(w)\right\|_{C[0,1]} \leqslant \tau\mathbf{I}_q\right\} \\ &= \jmath^{-1}\left(B_{C[0,1]}\left(\int_{\mathcal{M}}\jmath(w)\,d\mu(w),\tau\mathbf{I}_q\right)\right). \end{aligned} \tag{79}$$

Note that by Markov's inequality we have

$$\forall\,\tau\geqslant 1,\qquad \mu(\mathcal{M}\smallsetminus A_\tau)\leqslant\frac{1}{\tau^q}. \tag{80}$$

The following lemma provides a convenient upper bound on the q-average distortion of the metric probability space $(\mathcal{M},d_{\mathcal{M}},\mu)$ into the real line; in essence, its role in what follows is to treat the "trivial case" in which the random variable $\jmath(x)$, where $x\in\mathcal{M}$ is distributed according to μ, is not well-concentrated around its mean in a certain quantitative sense which is made precise below.

Lemma 32. *$(\mathcal{M},d_{\mathcal{M}},\mu)$ embeds with q-average distortion $D_{\mathbb{R}}\geqslant 1$ into $\mathbb{R}$, where*

$$D_{\mathbb{R}} = D_{\mathbb{R}}(\mu,\jmath,q) \stackrel{\mathrm{def}}{=} \inf_{\tau>e^{e^{-q}}} \left(\int_{\mathcal{M}\smallsetminus A_\tau}\left\|\jmath(x)-\int_{\mathcal{M}}\jmath(w)\,d\mu(w)\right\|_{C[0,1]}^q d\mu(x)\right)^{-\frac{1}{q}} \frac{6\tau\mathbf{I}_q}{\tau-e^{e^{-q}}}. \tag{81}$$

Proof. Define $f:\mathcal{M}\to\mathbb{R}$ by setting

$$\forall\,x\in\mathcal{M},\qquad f(x)\stackrel{\mathrm{def}}{=}\left\|\jmath(x)-\int_{\mathcal{M}}\jmath(w)\,d\mu(w)\right\|_{C[0,1]}.$$

Then f is 1-Lipschitz, because $\jmath$ is an isometry.

Suppose that $\tau>e^{e^{-q}}$ and observe that for every $x\in\mathcal{M}\smallsetminus A_\tau$ and $y\in A_{e^{e^{-q}}}$ we have

$$f(x)-f(y)\geqslant\left\|\jmath(x)-\int_{\mathcal{M}}\jmath(w)\,d\mu(w)\right\|_{C[0,1]} - e^{e^{-q}}\mathbf{I}_q \geqslant \left(1-\frac{e^{e^{-q}}}{\tau}\right)\left\|\jmath(x)-\int_{\mathcal{M}}\jmath(w)\,d\mu(w)\right\|_{C[0,1]}.$$

Consequently,

$$\begin{aligned} &\iint_{\mathcal{M}\times\mathcal{M}}\left|f(x)-f(y)\right|^q d\mu(x)\,d\mu(y) \\ &\geqslant 2\left(1-\frac{e^{e^{-q}}}{\tau}\right)^q\mu\left(A_{e^{e^{-q}}}\right)\int_{\mathcal{M}\smallsetminus A_\tau}\left\|\jmath(x)-\int_{\mathcal{M}}\jmath(w)\,d\mu(w)\right\|_{C[0,1]}^q d\mu(x) \\ &\geqslant \frac{\left(\tau-e^{e^{-q}}\right)^q\int_{\mathcal{M}\smallsetminus A_\tau}\left\|\jmath(x)-\int_{\mathcal{M}}\jmath(w)\,d\mu(w)\right\|_{C[0,1]}^q\,d\mu(x)}{(6\tau\mathbf{I}_q)^q}\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,d\mu(x)\,d\mu(y), \end{aligned}$$

where the final step uses (77) and the bound (80) which gives $\mu\left(A_{e^{e^{-q}}}\right)\geqslant 1-e^{-qe^{-q}}\geqslant 3^{-q}$. $\square$

Lemma 33. *Fix $p, q, D \geqslant 1$ with $q \geqslant p$. Define $\Delta \geqslant 1$ by*

$$\Delta \stackrel{\mathrm{def}}{=} D + \frac{q}{p\log\left(e+\frac{q}{pD}\right)} \asymp \begin{cases} D & \text{if } D \geqslant \frac{q}{p}, \\ \frac{q}{p\log\left(e+\frac{q}{pD}\right)} & \text{if } 1 \leqslant D \leqslant \frac{q}{p}. \end{cases} \tag{82}$$

Suppose that a metric probability space $(\mathcal{M}, d_\mathcal{M}, \mu)$ embeds with p-average distortion less than D into a Banach space $(X, \|\cdot\|_X)$. Then $(\mathcal{M}, d_\mathcal{M}, \mu)$ embeds with q-average distortion $O(\Delta)$ into $(X, \|\cdot\|_X)$.

Proof. Let $D' \geqslant 1$ denote the infimum over those $K \geqslant 1$ for which $(\mathcal{M}, d_\mathcal{M}, \mu)$ embeds with q-average distortion K into $(X, \|\cdot\|_X)$. Our task is to bound D' from above. To this end, define $\delta > 0$ by

$$\delta \stackrel{\mathrm{def}}{=} \frac{\mathbf{I}_p}{\mathbf{I}_q} \stackrel{(77)\wedge(78)}{\leqslant} 2\frac{\left(\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x,y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{p}}}{\left(\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x,y)^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{q}}}. \tag{83}$$

Since $p < q$, by Jensen's inequality we have $\delta \in [0,1]$. The premise of Lemma 33 is that there exists a D-Lipschitz mapping $f : \mathcal{M} \to X$ that satisies

$$\begin{aligned}\left(\iint_{\mathcal{M}\times\mathcal{M}} \|f(x)-f(y)\|_X^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{q}} &\geqslant \left(\iint_{\mathcal{M}\times\mathcal{M}} \|f(x)-f(y)\|_X^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{p}} \\ &\geqslant \left(\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x,y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{p}} \stackrel{(83)}{\geqslant} \frac{\delta}{2}\left(\iint_{\mathcal{M}\times\mathcal{M}} d_\mathcal{M}(x,y)^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{q}},\end{aligned}$$

where the first step is Jensen's inequality. So, the normalized mapping $\frac{2}{\delta}f : \mathcal{M} \to X$ exhibits that

$$D' \leqslant \frac{2}{\delta}D. \tag{84}$$

Note that Lemma 33 already follows from (84) if $\delta \geqslant \frac{1}{2}$, so we may assume from now on that $\delta \leqslant \frac{1}{2}$.

Next, fix $\tau > 1$ satisfying

$$e^{e^{-q}} < \tau < \left(\frac{1}{\delta}\right)^{\frac{p}{q-p}}. \tag{85}$$

The value of τ will be specified later so as to optimize the ensuing reasoning; see (87). Observe that

$$\begin{aligned}&\int_{\mathcal{M}\smallsetminus A_\tau} \left\|j(x) - \int_\mathcal{M} j(w)\,\mathrm{d}\mu(w)\right\|_{C[0,1]}^q \mathrm{d}\mu(x) \stackrel{(76)}{=} \mathbf{I}_q^q - \int_{A_\tau} \left\|j(x) - \int_\mathcal{M} j(w)\,\mathrm{d}\mu(w)\right\|_{C[0,1]}^q \mathrm{d}\mu(x) \\ &\stackrel{(79)}{\geqslant} \mathbf{I}_q^q - (\tau\mathbf{I}_q)^{q-p}\int_{A_\tau} \left\|j(x) - \int_\mathcal{M} j(w)\,\mathrm{d}\mu(w)\right\|_{C[0,1]}^p \mathrm{d}\mu(x) \stackrel{(76)}{\geqslant} \mathbf{I}_q^q - (\tau\mathbf{I}_q)^{q-p}\mathbf{I}_p^p \stackrel{(83)}{=} (1-\tau^{q-p}\delta^p)\mathbf{I}_q^q.\end{aligned}$$

In combination with Lemma 32, this implies (even for an embedding into $\mathbb{R} \subset X$) that

$$D' \leqslant \frac{6\tau}{(\tau - e^{e^{-q}})(1-\tau^{q-p}\delta^p)^{\frac{1}{q}}}. \tag{86}$$

It is obviously in our interest to choose τ so as to minimize the right had side of (86) subject to the constraints (85). While the optimal τ here does not have a closed-form expression, a straightforward (albeit somewhat tedious) inspection of (86) reveals that up to a possible loss of a universal constant factor in the final conclusion (82) of Lemma 33, one cannot do better than the following choice.

$$\tau = \left(1 - \frac{1}{e^q}\right)^{\frac{1}{q-p}}\left(\frac{1}{\delta}\right)^{\frac{p}{q-p}}. \tag{87}$$

For this value of τ one readily checks that (85) holds (recall that $0 \leqslant \delta \leqslant \frac{1}{2}$). So, (86) implies that

$$D' \lesssim 1 + \frac{q}{p\log\left(\frac{1}{\delta}\right)}.$$

In combination with (84) we therefore have

$$D' \lesssim \max_{0\leqslant\delta\leqslant\frac12} \min\left\{\frac{D}{\delta}, 1+\frac{q}{p\log\left(\frac{1}{\delta}\right)}\right\} \asymp D + \frac{q}{p\log\left(e+\frac{q}{pD}\right)}, \qquad \square$$

Lemma 34. *Fix $p,q,D \geqslant 1$ with $p\geqslant q$. Suppose that an infinite metric space $(\mathcal{M}, d_{\mathcal{M}})$ embeds with p-average distortion less than D into a Banach space $(X,\|\cdot\|_X)$. Then, $(\mathcal{M}, d_{\mathcal{M}})$ embeds with q-average distortion $D' = D'(p,q,D)\geqslant 1$ into $(X,\|\cdot\|_X)$, where for some universal constant $\kappa>1$,*

$$D' \leqslant (\kappa D)^{\frac{p}{q}}. \tag{88}$$

Proof. Fix a Borel probability measure μ on $\mathcal{M}$ that satisfies $\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) < \infty$. For the purpose of proving Lemma 34, it will suffice to consider the subset $A_\tau \subset \mathcal{M}$ in (79) only for $\tau = 8$. Note that due to Lemma 32, it suffices to prove Lemma 34 under the additional assumption

$$\bigg(\int_{\mathcal{M}\smallsetminus A_8} \Big\| j(x) - \int_{\mathcal{M}} j(w)\,\mathrm{d}\mu(w)\Big\|_{C[0,1]}^q \,\mathrm{d}\mu(x)\bigg)^{\frac1q} \leqslant \frac{\mathbf{I}_q}{8}. \tag{89}$$

Since j is an isometry, we have the following point-wise bound for every $x,y\in\mathcal{M}$.

$$d_{\mathcal{M}}(x,y) \leqslant \|j(x)-j(y)\|_{C[0,1]} \leqslant \Big\|x - \int_{\mathcal{M}} j(w)\,\mathrm{d}\mu(w)\Big\|_{C[0,1]} + \Big\|y - \int_{\mathcal{M}} j(w)\,\mathrm{d}\mu(w)\Big\|_{C[0,1]}. \tag{90}$$

Using the triangle inequality in $L_q(\mu\times\mu)$, this implies that

$$\begin{aligned}
&\bigg(\iint_{(\mathcal{M}\smallsetminus A_8)\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac1q}\\
&\qquad \overset{(90)\wedge(76)}{\leqslant} \bigg(\int_{\mathcal{M}\smallsetminus A_8} \Big\|x - \int_{\mathcal{M}} j(w)\,\mathrm{d}\mu(w)\Big\|_{C[0,1]}^q \,\mathrm{d}\mu(x)\bigg)^{\frac1q} + \mu(\mathcal{M}\smallsetminus A_8)^{\frac1q}\mathbf{I}_q \qquad (91)\\
&\qquad \overset{(89)\wedge(80)}{\leqslant} \frac14\bigg(\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac1q}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\\
&\qquad \overset{(76)}{=} \iint_{A_8\times A_8} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) + 2\iint_{(\mathcal{M}\smallsetminus A_8)\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\\
&\qquad \overset{(91)}{\leqslant} \iint_{A_8\times A_8} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) + \frac12\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y).
\end{aligned}$$

This simplifies to give

$$\iint_{A_8\times A_8} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant \frac12\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y). \tag{92}$$

An application of the assumption of Lemma (34) to the restriction of μ to A_8 (recall that by (80) we have $\mu(A_8)\geqslant 1-8^{-q}\asymp 1$) yields a D-Lipschitz mapping $f:\mathcal{M}\to X$ that satisfies

$$\begin{aligned}
&\bigg(\frac{1}{\mu(A_8)^2}\iint_{A_8\times A_8} \|f(x)-f(y)\|_X^p\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac1p} \geqslant \bigg(\frac{1}{\mu(A_8)^2}\iint_{A_8\times A_8} d_{\mathcal{M}}(x,y)^p\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac1p}\\
&\qquad \geqslant \bigg(\frac{1}{\mu(A_8)^2}\iint_{A_8\times A_8} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac1q} \overset{(92)}{\geqslant} \bigg(\frac{1}{2\mu(A_8)^2}\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac1q},
\end{aligned}$$

where the penultimate step is an application of Jensen's inequality, since $q < p$. Hence,

$$\left(\iint_{A_8\times A_8} \|f(x)-f(y)\|_X^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{p}} \gtrsim \left(\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{q}}. \qquad (93)$$

Next, for every $x,y\in A_8$ we have

$$d_{\mathcal{M}}(x,y) = \|\jmath(x)-\jmath(y)\|_{C[0,1]} \leqslant \Big\|x-\int_{\mathcal{M}} \jmath(w)\,\mathrm{d}\mu(w)\Big\|_{C[0,1]} + \Big\|y-\int_{\mathcal{M}} \jmath(w)\,\mathrm{d}\mu(w)\Big\|_{C[0,1]} \overset{(79)}{\leqslant} 16\mathbf{I}_q.$$

Therefore, because f is D-Lipschitz, the following point-wise inequality holds true.

$$\forall\, x,y\in A_8,\qquad \|f(x)-f(y)\|_X^p \leqslant \big(Dd_{\mathcal{M}}(x,y)\big)^{p-q}\|f(x)-f(y)\|_X^q \leqslant (16D\mathbf{I}_q)^{p-q}\|f(x)-f(y)\|_X^q.$$

Consequently,

$$\begin{aligned}\iint_{A_8\times A_8} \|f(x)-f(y)\|_X^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) &\leqslant (16D\mathbf{I}_q)^{p-q}\iint_{A_8\times A_8} \|f(x)-f(y)\|_X^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\\ &\leqslant (16D\mathbf{I}_q)^{p-q}\iint_{\mathcal{M}\times\mathcal{M}} \|f(x)-f(y)\|_X^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\\ &\overset{(78)}{\leqslant} (16D)^{p-q}\left(\iint_{\mathcal{M}\times\mathcal{M}} \|f(x)-f(y)\|_X^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{p}{q}}.\end{aligned}$$

A substitution of this bound into (93) gives

$$(16D)^{\frac{p}{q}-1}\left(\iint_{\mathcal{M}\times\mathcal{M}} \|f(x)-f(y)\|_X^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{q}} \gtrsim \left(\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\right)^{\frac{1}{q}}.$$

Thus, for an appropriate universal constant $C>0$, the rescaled function $C(16D)^{\frac{p}{q}-1}f:\mathcal{M}\to X$ exhibits the existence of an embedding with the stated bound on its q-average distortion. $\square$

Remark 35. We did not investigate the optimality of Lemma 33 and Lemma 34. Specifically, we do not know the extent to which the additive term in (82) and the power p/q in (88) are necessary. It would be worthwhile (and probably tractable) to clarify these basic matters in future investigations.

5.2. **Average embedding of a snowflake of a Banach space into itself.** Note that Lemma 33 and Lemma 34 imply the special case $\omega=1$ of Proposition (6), with quite good dependence on p,q in (3); in particular, $D'\lesssim D$ (essentially no loss is incured) when $q\geqslant p$ and $D\geqslant q/p$ (and $\omega=1$). We will next treat Proposition (6) for general $\omega\in(0,1]$ in the special case $\mathcal{M}=Y$ and $D=1$.

For $\omega\in(0,1]$ and $p\in[1,2]$, the ω-snowflake of $L_p(\mathbb{R})$ embeds isometrically into $L_p(\mathbb{R})$. The Hilbertian case $p=2$ of this statement is a classical theorem of Schoenberg [131], and this statement was proven for general $p\in[1,2]$ by Bretagnolle, Dacunha-Castelle and Krivine [31]; see also the monograph [135] for an extensive treatment of this and related matters. Understanding the analogous situation when $p\in(2,\infty)$ remains a longstanding open question. Specifically, it is unknown whether or not there exists $\omega\in(0,1)$ and $p\in(2,\infty)$ such that the ω-snowflake of $L_p(\mathbb{R})$ admits a bi-Lipschitz embedding into $L_p(\mathbb{R})$; see [95, 2, 25, 115, 109, 50] for results along these lines, but an answer to this seemingly simple question remains stubbornly elusive despite substantial efforts. To the best of our knowledge, even the following more general question remains unknown.

Question 36. Does there exist $\omega\in(0,1)$ and an infinite dimensional Banach space $(X,\|\cdot\|_X)$ whose ω-snowflake does not admit a bi-Lipschitz embedding into $(X,\|\cdot\|_X)$?

Proposition 37 below treats the easier variant of Question (36) in the setting of average distortion.

Proposition 37. *Fix $p \in (0,\infty)$ and $\omega \in (0,1)$. Suppose that $D \in \mathbb{R}$ satisfies*

$$D > \frac{2^{(1-\omega)\left(1+\frac{1}{p\omega}\right)}}{\eta(p,\omega)}, \tag{94}$$

where $\eta(p,\omega) \in [0,1]$ is defined by

$$\eta(p,\omega) \stackrel{\mathrm{def}}{=} \inf_{\sigma \in [0,1)} \frac{1-\sigma^{\omega}}{1-\sigma}(1+\sigma^{p\omega})^{\frac{1-\omega}{p\omega}}. \tag{95}$$

Then, for any Banach space X, the ω-snowflake of X embeds with p-average distortion D into X.

Below we will provide estimates on the quantity $\eta(p,\omega)$ in (95), but the main significance of Proposition 37 is that the p-average distortion D in (94) can be taken to be a finite quantity that depends only on p and ω. With this at hand, we will now complete the proof of Proposition (6).

Proof of Proposition 6 assuming Proposition 37. Denote

$$\beta \stackrel{\mathrm{def}}{=} \max\{q\omega, 1\}. \tag{96}$$

The assumption of Proposition (6) is that an infinite metric space $(\mathcal{M}, d_{\mathcal{M}})$ embeds with p-average distortion $D \geqslant 1$ into a Banach space $(Y, \|\cdot\|_Y)$. By Lemma 33 and Lemma 34 there is a universal constant $\alpha > 1$ such that $(\mathcal{M}, d_{\mathcal{M}})$ embeds with β-average less than $D_1 > 1$ in $(Y, \|\cdot\|_Y)$, where

$$D_1 \stackrel{\mathrm{def}}{=} \left(\alpha D + \frac{\alpha\beta}{p\log\left(e+\frac{\beta}{pD}\right)}\right)^{\max\left\{\frac{p}{\beta},1\right\}}. \tag{97}$$

Thus, for every Borel probability measure μ on $\mathcal{M}$ there exists a mapping $f : \mathcal{M} \to Y$ that satisfies

$$\forall\, x,y \in \mathcal{M}, \qquad \|f(x)-f(y)\|_Y \leqslant D_1 d_{\mathcal{M}}(x,y), \tag{98}$$

and

$$\iint_{\mathcal{M}\times\mathcal{M}} \|f(x)-f(y)\|_Y^{\beta}\, d\mu(x)\, d\mu(y) \geqslant \iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^{\beta}\, d\mu(x)\, d\mu(y). \tag{99}$$

An application of Proposition (37) to the probability measure $f_\sharp\mu$ on Y (the push-forward of μ under f) yields a mapping $g : Y \to Y$ that satisfies

$$\forall\, u,v \in Y, \qquad \|g(u)-g(v)\|_Y \leqslant D_2\|u-v\|_Y^{\omega}, \qquad \text{where} \qquad D_2 \stackrel{\mathrm{def}}{=} \frac{2^{1+\frac{1-\omega}{\beta\omega}}}{\eta\left(\frac{\beta}{\omega},\omega\right)}, \tag{100}$$

and

$$\iint_{\mathcal{M}\times\mathcal{M}} \|g\circ f(x)-g\circ(y)\|_Y^{\frac{\beta}{\omega}}\, d\mu(x)\, d\mu(y) \geqslant \iint_{\mathcal{M}\times\mathcal{M}} \|f(x)-f(y)\|_Y^{\beta}\, d\mu(x)\, d\mu(y). \tag{101}$$

Therefore,

$$\forall\, x,y \in \mathcal{M}, \qquad \|g\circ f(x)-g\circ f(y)\|_Y \stackrel{(98)\wedge(100)}{\leqslant} D_2 D_1^{\omega} d_{\mathcal{M}}(x,y)^{\omega},$$

and

$$\iint_{\mathcal{M}\times\mathcal{M}} \|g\circ f(x)-g\circ(y)\|_Y^{\frac{\beta}{\omega}}\, d\mu(x)\, d\mu(y) \stackrel{(99)\wedge(101)}{\geqslant} \iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^{\beta}\, d\mu(x)\, d\mu(y)$$

This is the same as saying that the metric space $(\mathcal{M}, d_{\mathcal{M}}^{\omega})$ embeds with (β/ω)-average distortion $D_2 D_1^{\omega}$ into $(Y, \|\cdot\|_Y)$. Recalling the definition (96) of β, we have $\beta/\omega \geqslant q$. Hence, by Lemma 34 the ω-snowflake of $(\mathcal{M}, d_{\mathcal{M}})$ embeds with q-average distortion $D' \geqslant 1$ into $(Y, \|\cdot\|_Y)$, where

$$D' \leqslant \left(\kappa D_2 D_1^{\omega}\right)^{\frac{\beta}{q\omega}}. \qquad \square$$

Having noted the validity of Proposition 37, the following open question arises naturally.

Question 38. Does there exist a universal constant D with the property that for every $p \geqslant 1$ and $\omega \in (0,1)$, the ω-snowflake of any Banach space X embeds with p-average distortion D into X? We do not know if this is so even in the special cases of greatest interest $p \in \{1,2\}$ and $\omega \to 0^+$.

Prior to proving Proposition 37, we will collect some elementary estimates on the quantity $\eta(p,\omega)$ that is defined in (95). While explicit bounds on $\eta(p,\omega)$ are not crucial for the main geometric consequences of the present work, it is worthwhile to record such estimates here due to the intrinsic geometric interest of such embeddings; we also expect that explicit bounds will be needed for future applications. Lemma 39 below reflects the fact that the function of $\sigma \in [0,1)$ in (95) whose infimum defines $\eta(p,\omega)$ is decreasing when $p\omega \geqslant 1$ and increasing when $p \leqslant 1$. In the remaining range $p \in (1, \frac{1}{\omega})$, the function in question can behave in a more complicated manner and in particular in parts of this range it attains its global minimum in the interior of the interval $[0,1]$.

Lemma 39. *Fix $\omega \in (0,1)$. If $p \geqslant \frac{1}{\omega}$, then $\eta(p,\omega) = \omega 2^{\frac{1-\omega}{p\omega}}$. If $0 < p \leqslant 1$, then $\eta(p,\omega) = 1$.*

Proof. Define $h : [0,1] \times (0,\infty) \times (0,1) \to \mathbb{R}$ by setting for every $(\sigma, p, \omega) \in [0,1] \times (0,\infty) \times (0,1)$,

$$h(\sigma,p,\omega) \stackrel{\mathrm{def}}{=} \log\left(\frac{1-\sigma^\omega}{1-\sigma}(1+\sigma^{p\omega})^{\frac{1-\omega}{p\omega}}\right) = \log(1-\sigma^\omega) - \log(1-\sigma) + \frac{1-\omega}{p\omega}\log\left(1+\sigma^{p\omega}\right),$$

with the endpoint convention $h(1,p,\omega) \stackrel{\mathrm{def}}{=} \lim_{\sigma\to 1} h(\sigma,p,\omega) = \log\left(\omega 2^{\frac{1-\omega}{p\omega}}\right)$. Then,

$$\frac{\partial h}{\partial \sigma}(\sigma,p,\omega) = \frac{1}{1-\sigma} - \frac{\omega}{\sigma^{1-\omega}(1-\sigma^\omega)} + \frac{1-\omega}{\sigma}\cdot\frac{\sigma^{p\omega}}{1+\sigma^{p\omega}}. \tag{102}$$

Since the mapping $u \mapsto \frac{u}{1+u}$ is increasing on $[0,\infty)$, for every fixed $\sigma, \omega \in (0,1)$ the right hand side of (102) is a decreasing function of p. Consequently, for every $\sigma, \omega \in (0,1)$ and $p \geqslant \frac{1}{\omega}$,

$$\begin{aligned}\frac{\partial h}{\partial \sigma}(\sigma,p,\omega) \leqslant \frac{\partial h}{\partial \sigma}\Big(\sigma, \frac{1}{\omega}, \omega\Big) &\stackrel{(102)}{=} \frac{\omega\sigma^{2-\omega} - (2-\omega)\sigma + (2-\omega)\sigma^{1-\omega} - \omega}{(1-\sigma^2)(1-\sigma^\omega)\sigma^{1-\omega}} \\ &= -\frac{\omega(1-\omega)(2-\omega)}{(1-\sigma^2)(1-\sigma^\omega)\sigma^{1-\omega}} \int_\sigma^1 \frac{(1-s)(s-\sigma)}{s^{\omega+1}}\,\mathrm{d}s < 0,\end{aligned} \tag{103}$$

where the penultimate step of (103) is a straightforward evaluation of the definite integral. Therefore $h(\sigma,p,\omega)$ is decreasing in σ if $p\omega \geqslant 1$. In particular, in this range we have $h(\sigma,p,\omega) \geqslant h(1,p,\omega)$, i.e., the maximum defining $\eta(p,\omega)$ in (95) is attained at $\sigma = 1$, as required. Also, if $p \in (0,1]$, then

$$\begin{aligned}\frac{\partial h}{\partial \sigma}(\sigma,p,\omega) \geqslant \frac{\partial h}{\partial \sigma}\Big(\sigma, 1, \omega\Big) &\stackrel{(102)}{=} \frac{\sigma^{1-\omega} - \sigma^\omega + (2\omega-1)\sigma + 1 - 2\omega}{(1-\sigma)(1-\sigma^{2\omega})\sigma^{1-\omega}} \\ &= \frac{\omega}{(1-\sigma)(1-\sigma^{2\omega})\sigma^{1-\omega}} \int_\sigma^1 s^\omega(s-\sigma)\Big(2(1-\omega) + \frac{1}{s^{1+\omega}} - 1\Big)\,\mathrm{d}s > 0,\end{aligned} \tag{104}$$

where the penultimate step of (104) is a straightforward evaluation of the definite integral and the final step of (104) holds because the integrand is point-wise positive. Hence, $h(\sigma,p,\omega)$ is increasing in σ if $p \in (0,1]$. In particular, $h(\sigma,p,\omega) \geqslant h(0,p,\omega) = 0$ when $p \in (0,1]$, i.e., in this range the maximum defining $\eta(p,\omega)$ in (95) is attained at $\sigma = 0$, as required. $\square$

The following corollary is nothing more than a substitution of Lemma 39 into Proposition 37.

Corollary 40. *For $p > 0$ and $\omega \in (0,1]$, let $\mathsf{Av}(p,\omega)$ be the infimum over those $D \geqslant 1$ such that for any Banach space X the ω-snowflake of X embeds with p-average distortion D into X. Then*

$$\mathsf{Av}(1,\omega) \leqslant 2^{\frac{1-\omega^2}{\omega}} \qquad \text{and} \qquad p\omega \geqslant 1 \implies \mathsf{Av}(p,\omega) \leqslant \frac{2^{1-\omega}}{\omega}.$$

In particular, both $\mathsf{Av}\big(1,\frac{1}{2}\big)$ and $\mathsf{Av}\big(2,\frac{1}{2}\big)$ are at most $2\sqrt{2}$ and $\mathsf{Av}\big(p,\frac{1}{p}\big) \leqslant 2p$ for all $p \geqslant 1$.

Remark 41. We proved above that the distortion D' of Proposition (6) satisfies $D' \leqslant \left(\kappa D_2 D_1^{\omega}\right)^{\frac{\beta}{q\omega}}$. Here $\kappa > 1$ is a universal constant and β, D_1, D_2 are given in (96), (97), (100), respectively. So, using Lemma 39, we have the following version of (3), in which $\mathsf{K} > 1$ is a universal constant.

$$D' \lesssim \frac{\mathsf{K}^{\frac{p}{q}+\frac{1}{q\omega}}}{\omega^{\max\left\{1,\frac{1}{q\omega}\right\}}}\left(D + \frac{q\omega}{p\log\left(e+\frac{q\omega}{pD}\right)}\right)^{\max\left\{\frac{p}{q},\omega\right\}}. \tag{105}$$

We have no reason to suspect that (105) is sharp; it would be worthwhile to find the optimal bound.

If one uses (105) in the setting of Theorem 1, in which $\mathcal{M}$ is the $\frac{1}{2}$-snowflake of a k-dimensional normed space X and $D \lesssim \sqrt{\log k}$, one see that for every $\omega \in (0, \frac{1}{2}]$ the ω-snowflake of X embeds with quadratic average distortion at most $\mathsf{C}_{\omega}\sqrt{\log k}$ into a Hilbert space, where $\mathsf{C}_{\omega} > 0$ depends only on ω. We suspect that the power of the logarithm may not be sharp here, thus leading to the following conjecture whose investigation we postpone to future research; we will see in Section 6 that its positive resolution would yield an asymptotically sharp bound (for ω fixed and $k \to \infty$).

Conjecture 42. For every $\omega \in (0, \frac{1}{2})$ there is $\mathsf{C}_{\omega} > 0$ such that for $k \in \{2, 3, \ldots\}$ the ω-snowflake of any k-dimensional normed space embeds with quadratic average distortion $C_{\omega}(\log k)^{\omega}$ into ℓ_2.

If in the formulation of Conjecture 42 quadratic average distortion is replaced by q-average distortion for $q > 2$, then by (105) the asymptotics of the distortion decreases to $\mathsf{C}_{\omega,q}(\log k)^{\max\{2/q,\omega\}}$. So, the analogue of Conjecture 42 for $(2/\omega)$-average distortion has a positive answer. For example, if, say, one considers 4-average distortion of the $\frac{1}{4}$-snowflake, then the bound becomes of order $\sqrt[4]{\log k}$.

Note also that if in Theorem 1 one considers average distortion (i.e., 1-average distortion) in place of its quadratic counterpart, then by (105) we get that the $\frac{1}{2}$-snowflake of any k-dimensional norm embeds into a Hilbert space with average distortion $O(\log k)$. It is conceivable that this bound could be reduced to $O(\sqrt{\log k})$, but we did not investigate this matter yet.

The proof of Proposition (37) relies on a natural "fractional normalization map" for which sharp bounds are contained in Lemma (43) below; cruder estimates on the modulus of continuity of such maps appear in several places, but we could not locate their optimal from in the literature.

Lemma 43. *Fix $\omega \in (0,1)$. For a Banach space $(X, \|\cdot\|_X)$, define $f_{\omega} = f_{\omega}^X : X \to X$ by setting*

$$\forall\, x \in X \smallsetminus \{0\}, \qquad f_{\omega}(x) \stackrel{\mathrm{def}}{=} \frac{1}{\|x\|_X^{1-\omega}}x, \tag{106}$$

and $f_{\omega}(0) \stackrel{\mathrm{def}}{=} 0$. Then, for every $p \in (0, \infty)$ we have

$$\forall\, x, y \in X, \qquad \frac{\eta(p,\omega)\|x-y\|_X}{\left(\|x\|_X^{p\omega} + \|y\|_X^{p\omega}\right)^{\frac{1-\omega}{p\omega}}} \leqslant \|f_{\omega}(x) - f_{\omega}(y)\|_X \leqslant 2^{1-\omega}\|x-y\|_X^{\omega}. \tag{107}$$

Both of the constants $\eta(p,\omega)$ and $2^{1-\omega}$ in the two inequalities appearing in (107) *cannot be improved.*

Note that if $p \geqslant 1$ and X is an $L_p(\mu)$ space, and we apply the mapping in (106) point-wise, then we get the mapping $(\phi \in L_p(\mu)) \mapsto \mathrm{sign}(\phi)|\phi|^{\omega}$, which is the classical Mazur map [94] from $L_p(\mu)$ to $L_q(\mu)$ for $q = p/\omega$. However, $f_{\omega}(\phi) = \|\phi\|_{L_p(\mu)}^{\omega-1}\phi$, so f_{ω} itself is different from the Mazur map.

Proof of Lemma 43. The optimality of the first inequality in (107) is seen by considering $x = \sigma y$ for every $\sigma \in [0,1]$, and the optimality of the second inequality in (107) is seen by considering $x = -y$.

Suppose that $x, y \in X \setminus \{0\}$ satisfy $\|x\|_X < \|y\|_X$. Then,

$$
\begin{aligned}
\|f_\omega(y) - f_\omega(x)\|_X &\overset{(106)}{=} \left\| \frac{1}{\|y\|_X^{1-\omega}}(y-x) - \left(\frac{1}{\|x\|_X^{1-\omega}} - \frac{1}{\|y\|_X^{1-\omega}} \right) x \right\|_X \\
&\geqslant \frac{\|y-x\|_X}{\|y\|_X^{1-\omega}} - \|x\|_X^\omega + \frac{\|x\|_X}{\|y\|_X^{1-\omega}} = \frac{1}{\|y\|_X^{1-\omega}} \left(1 - \frac{\|x\|_X^\omega \|y\|_X^{1-\omega} - \|x\|_X}{\|x-y\|_X} \right) \|x-y\|_X.
\end{aligned}
\tag{108}
$$

Observe that $(\|x\|_X^\omega \|y\|_X^{1-\omega} - \|x\|_X)/\|x-y\|_X \leqslant (\|x\|_X^\omega \|y\|_X^{1-\omega} - \|x\|_X)/(\|y\|_X - \|x\|_X)$, because $\|x\|_X^\omega \|y\|_X^{1-\omega} - \|x\|_X > 0$ and $\|x-y\|_X \geqslant \|y\|_X - \|x\|_X$. By substituting this into (108), we see that

$$
\begin{aligned}
\|f_\omega(y) - f_\omega(x)\|_X &\geqslant \frac{1}{\|y\|_X^{1-\omega}} \left(1 - \frac{\|x\|_X^\omega \|y\|_X^{1-\omega} - \|x\|_X}{\|y\|_X - \|x\|_X} \right) \|x-y\|_X \\
&= \frac{1-\sigma^\omega}{1-\sigma} (1+\sigma^{p\omega})^{\frac{1-\omega}{p\omega}} \cdot \frac{\|x-y\|_X}{\left(\|x\|_X^{p\omega} + \|y\|_X^{p\omega}\right)^{\frac{1-\omega}{p\omega}}} \overset{(95)}{\geqslant} \frac{\eta(p,\omega)\|x-y\|_X}{\left(\|x\|_X^{p\omega} + \|y\|_X^{p\omega}\right)^{\frac{1-\omega}{p\omega}}},
\end{aligned}
$$

where in the penultimate step we write $\sigma \overset{\mathrm{def}}{=} \frac{\|x\|_X}{\|y\|_X} \leqslant 1$. This justifies the first inequality in (107).

For the second inequality in (107), note that

$$
\begin{aligned}
\|f_\omega(x) - f_\omega(y)\|_X &= \left\| \left(\frac{1}{\|x\|_X^{1-\omega}} - \frac{1}{\|y\|_X^{1-\omega}} \right) x + \frac{1}{\|y\|_X^{1-\omega}}(x-y) \right\|_X \\
&\leqslant \left(\frac{1}{\|x\|_X^{1-\omega}} - \frac{1}{\|y\|_X^{1-\omega}} \right) \|x\|_X + \frac{\|x-y\|_X}{\|y\|_X^{1-\omega}} = \frac{\|x-y\|_X - \|x\|_X}{\|y\|_X^{1-\omega}} + \|x\|_X^\omega\ .
\end{aligned}
\tag{109}
$$

The quantity $\phi(\|y\|_X) \overset{\mathrm{def}}{=} (\|x-y\|_X - \|x\|_X)/\|y\|_X^{1-\omega}$ in (109) decreases with $\|y\|_X$ if $\|x\|_X < \|x-y\|_X$. Since $\|y\|_X \geqslant \|x-y\|_X - \|x\|_X$ is a better lower bound on $\|y\|_X$ than our assumption $\|y\|_X \geqslant \|x\|_X$ when $\|x\|_X \leqslant \frac{1}{2}\|x-y\|_X$, it follows that $\phi(\|y\|_X) \leqslant \phi(\|x-y\|_X - \|x\|_X)$ if $\|x\|_X \leqslant \frac{1}{2}\|x-y\|_X$ and $\phi(\|y\|_X) \leqslant \phi(\|x\|_X)$ if $\frac{1}{2}\|x-y\|_X \leqslant \|x\|_X \leqslant \|x-y\|_X$. Next, $\phi(\|y\|_X)$ is nondecreasing in $\|y\|_X$ when $\|x\|_X \geqslant \|x-y\|_X$, so due to the a priori upper bound $\|y\|_X \leqslant \|x-y\|_X + \|x\|_X$ we have $\phi(\|y\|_X) \leqslant \phi(\|x-y\|_X + \|x\|_X)$ in the remaining range $\|x\|_X \geqslant \|x-y\|_X$. These observations give

$$
\|f_\omega(x) - f_\omega(y)\|_X \overset{(109)}{\leqslant} \psi_\omega\left(\frac{\|x\|_X}{\|x-y\|_X} \right) \|x-y\|_X^\omega\ , \tag{110}
$$

where $\psi_\omega : [0,\infty) \to [0,\infty)$ is defined by

$$
\forall\, \rho \in [0,\infty), \qquad \psi_\omega(\rho) \overset{\mathrm{def}}{=} \begin{cases} (1-\rho)^\omega + \rho^\omega & \text{if } 0 \leqslant \rho \leqslant \frac{1}{2}, \\ \frac{1}{\rho^{1-\omega}} & \text{if } \frac{1}{2} \leqslant \rho \leqslant 1, \\ \rho^\omega - \frac{\rho-1}{(1+\rho)^{1-\omega}} & \text{if } \rho \geqslant 1. \end{cases} \tag{111}
$$

For $\rho \in (0, \frac{1}{2})$ we have $\psi_\omega'(\rho) = \omega(1/\rho^{1-\omega} - 1/(1-\rho)^{1-\omega}) > 0$. Also, $\psi'(\rho) = -(1-\omega)/\rho^{2-\omega} < 0$ for $\rho \in (\frac{1}{2}, 1)$. Finally, we claim that $\psi_\omega'(\rho) < 0$ if $\rho > 1$. Indeed, for every $\rho > 1$,

$$
\begin{aligned}
\psi_\omega'(\rho) &\overset{(111)}{=} \frac{\omega(1+\rho)^{2-\omega} - \omega\rho^{2-\omega} - (2-\omega)\rho^{1-\omega}}{\rho^{1-\omega}(1+\rho)^{2-\omega}} = \frac{(1-\omega)(2-\omega)}{(1+\rho)^{2-\omega}} \int_\rho^{1+\rho} \left(\omega \int_1^{\frac{r}{\rho}} \frac{\mathrm{d}s}{s^\omega} - 1 \right) \mathrm{d}r \\
&\leqslant \frac{(1-\omega)(2-\omega)}{(1+\rho)^{2-\omega}} \int_\rho^{1+\rho} \left(\omega\left(\frac{r}{\rho} - 1 \right) - 1 \right) \mathrm{d}r = -\frac{(1-\omega)(2-\omega)}{(1+\rho)^{2-\omega}} \left(1 - \frac{\omega}{2\rho} \right) < 0,
\end{aligned}
\tag{112}
$$

where the identities in the second and penultimate steps of (112) are straightforward evaluations of the respective definite integrals, the third step of (112) uses the fact that $s \geqslant 1$ in the internal integrand, and the final step of (112) holds because $0 < \omega < 1 < \rho$. We have thus established that ψ_ω is increasing on $[0,\frac12]$ and decreasing on $[\frac12,\infty)$. Hence, ψ_ω attains its global maximum at $\rho=\frac12$, where its value is $2^{1-\omega}$. Due to (110), this justifies the second inequality in (107). $\square$

Proof of Proposition 37. Fix a Borel probability measure μ on X. The assumption (94) implies that

$$\beta \stackrel{\mathrm{def}}{=} \frac12\left(\frac{D\eta(p,\omega)}{2^{1-\omega}}\right)^{\frac{p\omega}{1-\omega}} > 1. \tag{113}$$

Hence, there exists $u\in X$ such that

$$\int_X \|x-u\|_X^{p\omega}\,\mathrm{d}\mu(x) \leqslant \beta \inf_{v\in X}\int_X \|x-v\|_X^{p\omega}\,\mathrm{d}\mu(x) \leqslant \beta \iint_{X\times X}\|x-y\|_X^{p\omega}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y). \tag{114}$$

Define $\phi: X\to X$ by setting

$$\forall\, z\in X,\qquad \phi(z) \stackrel{\mathrm{def}}{=} \left(\frac{\iint_{X\times X}\|x-y\|_X^{p\omega}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}{\iint_{X\times X}\|f_\omega(x-u)-f_\omega(y-u)\|_X^{p}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}\right)^{\frac1p} f_\omega(z-u), \tag{115}$$

where f_ω is the normalization map that is given in (106). Then, by design we have

$$\iint_{X\times X}\|\phi(x)-\phi(y)\|_X^{p}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = \iint_{X\times X}\|x-y\|_X^{p\omega}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y).$$

Proposition 37 will therefore be proven if we demonstrate that the ω-Hölder constant of $\phi: X\to X$ satisfies $\|\phi\|_{\mathrm{Lip}_\omega(X,X)}\leqslant D$. Indeed,

$$\begin{aligned}
\|\phi\|_{\mathrm{Lip}_\omega(X,X)} &\stackrel{(115)\wedge(107)}{=} 2^{1-\omega}\left(\frac{\iint_{X\times X}\|x-y\|_X^{p\omega}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}{\iint_{X\times X}\|f_\omega(x-u)-f_\omega(y-u)\|_X^{p}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}\right)^{\frac1p} && (116)\\
&\stackrel{(107)}{\leqslant} \frac{2^{1-\omega}}{\eta(p,\omega)^\omega}\left(\frac{\iint_{X\times X}\left(\|x-u\|_X^{p\omega}+\|y-u\|_X^{p\omega}\right)^{1-\omega}\|f_\omega(x-u)-f_\omega(y-u)\|_X^{p\omega}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}{\iint_{X\times X}\|f_\omega(x-u)-f_\omega(y-u)\|_X^{p}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}\right)^{\frac1p}\\
&\leqslant \frac{2^{1-\omega}}{\eta(p,\omega)^\omega}\left(\frac{\iint_{X\times X}\left(\|x-u\|_X^{p\omega}+\|y-u\|_X^{p\omega}\right)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}{\iint_{X\times X}\|f_\omega(x-u)-f_\omega(y-u)\|_X^{p}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}\right)^{\frac{1-\omega}{p}} && (117)\\
&= \frac{2^{1-\omega}}{\eta(p,\omega)^\omega}\left(\frac{2\int_X\|x-u\|_X^{p\omega}\,\mathrm{d}\mu(x)}{\iint_{X\times X}\|f_\omega(x-u)-f_\omega(y-u)\|_X^{p}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}\right)^{\frac{1-\omega}{p}}\\
&\stackrel{(114)}{\leqslant} \frac{2^{1-\omega}(2\beta)^{\frac{1-\omega}{p}}}{\eta(p,\omega)^\omega}\left(\frac{\iint_{X\times X}\|x-y\|_X^{p\omega}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}{\iint_{X\times X}\|f_\omega(x-u)-f_\omega(y-u)\|_X^{p}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}\right)^{\frac{1-\omega}{p}}\\
&\stackrel{(116)}{=} \frac{2^{1-\omega}(2\beta)^{\frac{1-\omega}{p}}}{\eta(p,\omega)^\omega}\left(\frac{\|\phi\|_{\mathrm{Lip}_\omega(X,X)}}{2^{1-\omega}}\right)^{1-\omega}, && (118)
\end{aligned}$$

where (117) is an application of Jensen's inequality for the probability measure on $X\times X$ whose Radon–Nikodym derivative with respect to $\mu\times\mu$ is proportional to $(x,y)\mapsto\|f_\omega(x-u)-f_\omega(y-u)\|_X^p$. Now, the bound (118) simplifies to give the desired estimate

$$\|\phi\|_{\mathrm{Lip}_\omega(X,X)} \leqslant \frac{2^{1-\omega}(2\beta)^{\frac{1-\omega}{p\omega}}}{\eta(p,\omega)} \stackrel{(113)}{=} D. \qquad \square$$

5.3. **Deduction of Theorem 9 from** (14)**.** As we stated in the Introduction, the matrix-dimension inequality (16) of Theorem 9 is a formal consequence of the ℓ_1 matrix-dimension inequality (14) that we deduced there from Theorem 1. We also explained in the Introduction that if one settles for a matrix-dimension inequality as in (15) with a worse asymptotic dependence on p as $p \to \infty$ than that of (16), which we expect to be sharp (recall Conjecture 11), then this could be done using reductions from [107, Section 7.4] between notions of q-average distortion as q varies over $[1,\infty)$, or using the better bounds of Proposition 6. However, it seems that neither the literature nor Proposition 6 suffice for deducing Theorem 9 from (14). We rectify this here using the elementary bounds that we derived in Section 5.2 and basic input from topological degree theory. Those who are not concerned with obtaining the conjecturally sharp dependence on p can therefore skip the present section and instead mimic the argument of the Introduction that led to (14).

5.3.1. *A nonlinear Rayleigh quotient inequality.* The following lemma relates quantities that are naturally viewed as nonlinear versions of classical Rayleigh quotients. The need for estimates of this type first arose due to algorithmic concerns in [9, 10]; see also the survey [111, Section 5.1.1].

Lemma 44. *Fix $n \in \mathbb{N}$ and $p,q \in [1,\infty)$ with $p \leqslant q$. Suppose that $\pi = (\pi_1,\ldots,\pi_n) \in \triangle^{n-1}$ and that $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$ is a stochastic and π-reversible matrix. Let $(X,\|\cdot\|_X)$ be a Banach space. For any $x_1,\ldots,x_n \in X$ there are $y_1 = y_1(\frac{p}{q},\pi,x_1,\ldots,x_n),\ldots,y_n = y_n(\frac{p}{q},\pi,x_1,\ldots,x_n) \in X$ with*

$$\bigg(\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|y_i-y_j\|_X^p}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|y_i-y_j\|_X^p}\bigg)^{\frac1p} \geqslant \frac{p}{2q}\bigg(\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X^q}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X^q}\bigg)^{\frac1q}. \tag{119}$$

Prior to proving Lemma 44, we will proceed to see how, in combination with (14), it quickly implies Theorem 9. Indeed, let $(X,\|\cdot\|_X)$ be a finite-dimensional normed space, $n \in \mathbb{N}$ and $x_1,\ldots,x_n \in X$. Fix $\pi \in \triangle^{n-1}$, a stochastic π-reversible matrix $\mathsf{A} = (a_{ij}) \in \mathsf{M}_n(\mathbb{R})$ and $p \geqslant 1$. Apply Lemma 44 to get new vectors $y_1,\ldots,y_n \in X$ that satisfy the inequality.

$$\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|y_i-y_j\|_X}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|y_i-y_j\|_X} \geqslant \frac{1}{2p}\bigg(\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X^p}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X^p}\bigg)^{\frac1p}. \tag{120}$$

An application of (14) to the new configuration $\{y_1,\ldots,y_n\} \subset X$ of points in X yields the following lower bound on the dimension of X, in which $\mathsf{C} \geqslant 1$ is the universal constant of Theorem 1.

$$\dim(X) \geqslant \exp\bigg(\frac{1-\lambda_2(\mathsf{A})}{\mathsf{C}^2}\cdot\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|y_i-y_j\|_X}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|y_i-y_j\|_X}\bigg), \tag{121}$$

Upon substitution of (120) into (121) we get the desired bound (16) with $\mathsf{K} = 2\mathsf{C}^2$. $\square$

Towards the proof of Lemma 44, a variant of the following preparatory lemma also played a key role in [9, 10], for similar purposes. Its short proof relies on considerations from algebraic topology.

Lemma 45. *Suppose that $(X,\|\cdot\|_X)$ is a finite-dimensional normed space and that $f: X \to X$ is a continuous function that satisfies*

$$\lim_{R\to\infty}\ \inf_{\substack{x\in X\\ \|x\|_X\geqslant R}} \big(\|x\|_X - \|x-f(x)\|_X\big) = \infty. \tag{122}$$

Then f is surjective.

Proof. Write $\dim(X) = k$. Fix an arbitrary point $z \in \mathbb{S}^k$ in the Euclidean sphere $\mathbb{S}^k$ of $\mathbb{R}^{k+1}$, and fix also any homeomorphism $h: \mathbb{S}^k \smallsetminus \{z\} \to X$ between the punctured sphere $\mathbb{S}^k \smallsetminus \{z\}$ and X. Define $g: \mathbb{S}^k \to \mathbb{S}^k$ by $g(w) = h^{-1}\circ f\circ h(w)$ for $w \in \mathbb{S}^k \smallsetminus \{z\}$, and $g(z) = z$. We claim that g is continuous at z, and hence it is continuous on all of $\mathbb{S}^k$. Indeed, if $\{w_n\}_{n=1}^\infty \subset \mathbb{S}^k$ and $\lim_{n\to\infty} w_n = z$, then $\lim_{n\to\infty} \|h(w_n)\|_X = \infty$. Consequently,

$$\big\|f\big(h(w_n)\big)\big\|_X \geqslant \|h(w_n)\|_X - \big\|h(w_n) - f\big(h(w_n)\big)\big\|_X \xrightarrow[n\to\infty]{} \infty,$$

where we used (122). Therefore $\lim_{n\to\infty} g(w_n) = z$, as required.

We next claim that g is homotopic to the identity mapping $\mathsf{Id}_{\mathbb{S}^k} : \mathbb{S}^k \to \mathbb{S}^k$. Indeed, denote

$$\forall (t,w) \in [0,1]\times(\mathbb{S}^k \smallsetminus \{z\}), \qquad H(t,w) \stackrel{\mathrm{def}}{=} h^{-1}\Big(th(w) + (1-t)f\big(h(w)\big)\Big),$$

and $H(t,z) = z$ for all $t \in [0,1]$. If we will check that $H : [0,1]\times\mathbb{S}^k \to \mathbb{S}^k$ is continuous at every point of $[0,1]\times\{z\}$, then it would follow that it is continuous on all of $[0,1]\times\mathbb{S}^k$, thus yielding the desired homotopy. To see this, take any $\{t_n\}_{n=1}^\infty \subset [0,1]$ such that $\lim_{n\to\infty} t_n = t$ exists, and any $\{w_n\}_{n=1}^\infty \subset \mathbb{S}^k$ with $\lim_{n\to\infty} w_n = z$. We then have $\lim_{n\to\infty} \|h(w_n)\|_X = \infty$, and therefore

$$\begin{aligned}\big\|t_n h(w_n) + (1-t_n) f\big(h(w_n)\big)\big\|_X &\geqslant \|h(w_n)\|_X - (1-t_n)\big\|h(w_n) - f\big(h(w_n)\big)\big\|_X \\ &\geqslant \|h(w_n)\|_X - \big\|h(w_n) - f\big(h(w_n)\big)\big\|_X \xrightarrow[n\to\infty]{} \infty,\end{aligned}$$

where we used (122) once more. Hence $\lim_{n\to\infty} H(t_n, w_n) = z = H(t,z)$, as required.

Because we showed that g is homotopic to the identity on $\mathbb{S}^k$, it has degree 1, and therefore it is surjective; see e.g. [60, page 134]. Hence $h^{-1}\circ f\circ h(\mathbb{S}^n \smallsetminus \{z\}) = \mathbb{S}^n \smallsetminus \{z\}$, i.e., $f(X) = X$. □

Lemma 46. *Fix $n \in \mathbb{N}$, $\pi \in \triangle^{n-1}$, $\omega \in (0,1)$, a Banach space $(X, \|\cdot\|_X)$, and $x_1,\ldots,x_n \in X$. Then, there exist new vectors $y_1 = y_1(\omega,\pi,x_1,\ldots,x_n),\ldots,y_n = y_n(\omega,\pi,x_1,\ldots,x_n) \in X$ that satisfy $\sum_{i=1}^n \pi_i y_i = 0$, and for every $q \in (0,\infty)$ we have*

$$\forall\, i,j \in \{1,\ldots,n\}, \qquad 2^{1-\frac{1}{\omega}}\|x_i - x_j\|_X^{\frac{1}{\omega}} \leqslant \|y_i - y_j\|_X \leqslant \frac{(\|y_i\|_X^{q\omega} + \|y_j\|_X^{q\omega})^{\frac{1-\omega}{q\omega}}}{\eta(q,\omega)}\|x_i - x_j\|_X. \quad (123)$$

Proof. Our eventual goal is to apply Lemma 45 to the mapping

$$f = f_{\omega,\pi,x_1,\ldots,x_n} : \mathrm{span}(\{x_1,\ldots,x_n\}) \to \mathrm{span}(\{x_1,\ldots,x_n\})$$

that is defined by setting for every $x \in \mathrm{span}(\{x_1,\ldots,x_n\})$,

$$f(x) \stackrel{\mathrm{def}}{=} \sum_{i=1}^n \pi_i f_\omega^{-1}(f_\omega(x) - x_i) \stackrel{(106)}{=} \frac{1}{\|x\|_X^{\frac{1}{\omega}-1}} \sum_{i=1}^n \pi_i \big\|x - \|x\|_X^{1-\omega} x_i\big\|_X^{\frac{1}{\omega}-1}\big(x - \|x\|_X^{1-\omega}x_i\big). \quad (124)$$

Suppose for the moment that we checked that f satisfies the assumption (122) of Lemma 45. It would follow that f is surjective, and in particular there exists $x = x(\omega,\pi,x_1,\ldots,x_n) \in \mathrm{span}(\{x_1,\ldots,x_n\})$ such that $f(x) = 0$. Thus, if we choose $y_i = f_\omega^{-1}(f_\omega(x) - x_i)$ for $i \in \{1,\ldots,n\}$, then $\sum_{i=1}^n \pi_i y_i = 0$. Because $f_\omega(y_1) + x_1 = \ldots = f_\omega(y_n) + x_n = f_\omega(x)$, we have $\|f_\omega(y_i) - f_\omega(y_j)\|_X = \|x_i - x_j\|_X$ for all $i,j \in \{1,\ldots,n\}$. The desired bounds (123) now follows from Lemma 43.

Both f_ω and f_ω^{-1} are continuous, so f is also continuous. Write $\xi_z(x) = f_\omega^{-1}(f_\omega(x) - z)$ for every $x,z \in X$. Then $f(x) = \sum_{i=}^n \pi_i \xi_{x_i}(x)$, and therefore by the convexity of $\|\cdot\|_X : X \to X$ we have

$$\|x\|_X - \|x - f(x)\|_X = \|x\|_X - \Big\|\sum_{i=1}^n \pi_i(\xi_{x_i}(x) - x)\Big\|_X \geqslant \sum_{i=1}^n \pi_i\big(\|x\|_X - \|x - \xi_{x_i}(x)\|_X\big).$$

This shows that the assumption (122) of Lemma 45 would hold true if ξ_z satisfied it for every fixed $z \in X$. Since $f_\omega(x) - f_\omega(\xi_z(x)) = z$ by the definition of ξ_z, the case $p = \frac{1}{\omega}$ of Lemma 43 gives

$$\|z\|_X \geqslant \frac{2^{1-\omega}\omega}{(\|x\|_X + \|\xi_z(x)\|_X)^{1-\omega}}\|x - \xi_z(x)\|_X, \quad (125)$$

where we also used the fact that $\eta(\frac{1}{\omega},\omega)=\omega 2^{1-\omega}$, by Lemma 39. Note that

$$\|\xi_z(x)\|_X=\|f_\omega(x)-z\|_X^{\frac{1}{\omega}}\leqslant\big(\|f_\omega(x)\|_X+\|z\|_X\big)^{\frac{1}{\omega}}=\big(\|x\|_X^{\omega}\ +\|z\|_X\big)^{\frac{1}{\omega}}. \tag{126}$$

By combining (125) and (126) we conclude that

$$\|x\|_X-\|x-\xi_z(x)\|_X\geqslant\|x\|_X-\frac{\|z\|_X}{\omega}\big(\|x\|_X^{\omega}+\|z\|_X\big)^{\frac{1-\omega}{\omega}}\xrightarrow[\|x\|_X\to\infty]{}\infty. \qquad \square$$

Completion of the proof of Lemma 44. The ensuing reasoning is inspired by an idea of Matoušek [92]. Apply Lemma 46 with $\omega=\frac{p}{q}$ to get $y_1,\ldots,y_n\in X$ (depending on $\frac{p}{q},\pi,x_1,\ldots,x_n$) such that

$$\sum_{i=1}^n\pi_iy_i=0, \tag{127}$$

and for every $i,j\in\{1,\ldots,n\}$,

$$2^{1-\frac{q}{p}}\|x_i-x_j\|_X^{\frac{q}{p}}\leqslant\|y_i-y_j\|_X\leqslant\frac{q}{p}\|x_i-x_j\|_X\left(\frac{\|y_i\|_X^p+\|y_j\|_X^p}{2}\right)^{\frac1p-\frac1q}, \tag{128}$$

where we also used the fact that $\eta(q,\frac{p}{q})=\frac{p}{q}2^{\frac1p-\frac1q}$, by Lemma 39. Note that

$$\begin{aligned}\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\frac{\|y_i\|_X^p+\|y_j\|_X^p}{2}&=\frac12\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|y_i\|_X^p+\frac12\sum_{j=1}^n\sum_{i=1}^n\pi_ja_{ji}\|y_j\|_X^p\\&=\sum_{i=1}^n\pi_i\Big\|y_i-\sum_{s=1}^n\pi_sy_s\Big\|_X^p\leqslant\sum_{i=1}^n\sum_{j=1}^n\pi_i\pi_j\|y_i-y_j\|_X^p,\end{aligned} \tag{129}$$

where the first step uses π-reversibility, the second step uses stochasticity and the centering condition (127), and the final step follows from Jensen's inequality (since $p\geqslant1$). Hence,

$$\begin{aligned}\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|y_i-y_j\|_X^p&\leqslant\Big(\frac{q}{p}\Big)^p\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|x_i-x_j\|_X^p\left(\frac{\|y_i\|_X^p+\|y_j\|_X^p}{2}\right)^{1-\frac{p}{q}}\\&\leqslant\Big(\frac{q}{p}\Big)^p\bigg(\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|x_i-x_j\|_X^q\bigg)^{\frac{p}{q}}\bigg(\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\frac{\|y_i\|_X^p+\|y_j\|_X^p}{2}\bigg)^{1-\frac{p}{q}}\\&\leqslant\Big(\frac{q}{p}\Big)^p\bigg(\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|x_i-x_j\|_X^q\bigg)^{\frac{p}{q}}\bigg(\sum_{i=1}^n\sum_{j=1}^n\pi_i\pi_j\|y_i-y_j\|_X^p\bigg)^{1-\frac{p}{q}},\end{aligned}$$

where the first step is the second inequality in (128), the second step is Hölder's inequality, and the final step is (129). This simplifies to give

$$\begin{aligned}\left(\frac{\sum_{i=1}^n\sum_{j=1}^n\pi_i\pi_j\|y_i-y_j\|_X^p}{\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|y_i-y_j\|_X^p}\right)^{\frac1p}&\geqslant\frac{q}{p}\left(\frac{\sum_{i=1}^n\sum_{j=1}^n\pi_i\pi_j\|y_i-y_j\|_X^p}{\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|x_i-x_j\|_X^q}\right)^{\frac1q}\\&\geqslant\frac{p}{2^{1-\frac{p}{q}}q}\left(\frac{\sum_{i=1}^n\sum_{j=1}^n\pi_i\pi_j\|x_i-x_j\|_X^q}{\sum_{i=1}^n\sum_{j=1}^n\pi_ia_{ij}\|x_i-x_j\|_X^q}\right)^{\frac1q},\end{aligned}$$

where the final step is the first inequality in (128). $\square$

Remark 47. In [46, Proposition 3.9] de Laat and de la Salle proved that for every Banach space $(X,\|\cdot\|_X)$, every $n\in\mathbb{N}$, every $\pi\in\triangle^{n-1}$ and every π-reversible stochastic matrix $\mathsf{A}=(a_{ij})\in\mathsf{M}_n(\mathbb{R})$,

$$\forall\,1\leqslant p\leqslant q<\infty,\qquad\gamma(\mathsf{A},\|\cdot\|_X^q)^{\frac{p}{q}}\lesssim_{p,q}\gamma(\mathsf{A},\|\cdot\|_X^p)\lesssim_{p,q}\gamma(\mathsf{A},\|\cdot\|_X^q). \tag{130}$$

This is a Banach space-valued generalization of a useful extrapolation result for Poincaré inequalities that Matoušek proved in [92] for real-valued functions (see also [23, Lemma 5.5] or [116, Lemma 4.4]). Direct precursors of (130) are those of [103, 37], but they treat the case of graphs (relying on their representation as Schreier coset graphs due to [58], as well as ideas of [18]) with the resulting bound depending on the maximum degree; as such, these earlier versions are not suitable for applications that use arbitrary stochastic matrices (e.g. when using duality as we do here).

Using Theorem 22, it follows from the rightmost inequality in (130) that for every $\omega \in (0,1]$ the ω-snowflake of X embeds with $(1/\omega)$-average distortion $D_\omega \geqslant 1$ into an ultrapower of $\ell_{1/\omega}(X)$, where D_ω may depend only on ω (for this, we are considering (130) with $p=1$ and $q=1/\omega$. More generally, (130) and Theorem 22 yield an embedding of the ω-snowflake of X into an ultrapower of $\ell_q(X)$ with q-average distortion $D_{\omega,q}$). Proposition 37 shows that this is so even for embeddings into X itself. The ingredients of Proposition 37 and [46, Proposition 3.9] are similar, as [46] considers an $L_p(X)$-valued version of the normalization map that is given in (106) as a generalization of the classical Mazur map [94] (see also [118, 35, 43, 126, 128] for earlier variants in special cases, as well as the subsequent development in [10]). We will next show that by incorporating the reasoning of the present section, we obtain the following version of (130) with an explicit dependence on p,q.

$$\forall\, 1 \leqslant p \leqslant q < \infty, \qquad \Big(\frac{p}{2q}\Big)^p \gamma(\mathsf{A}, \|\cdot\|_X^q)^{\frac{p}{q}} \leqslant \gamma(\mathsf{A}, \|\cdot\|_X^p) \leqslant \Big(\frac{2q}{p}\Big)^q \gamma(\mathsf{A}, \|\cdot\|_X^q). \tag{131}$$

An inspection of the proof in [46] reveals that the dependence on p,q that it yields is much (exponentially) weaker asymptotically than that of (131), and we believe that this is inherent to the reasoning of [46]. The first inequality in (131) is sharp, as already shown in [92] for real-valued functions. We do not know if the second inequality in (131) is sharp, and conceivably the rightmost factor $(2q/p)^q$ in (131) could be replaced by $e^{O(q)}$. If this were indeed possible, then it would be a worthwhile result because it would yield a fully analogous vector-valued generalization of Cheeger's inequality [36, 39] and Matoušek's extrapolation phenomenon [92].

To deduce the first inequality in (131), take $x_1,\ldots,x_n \in X$ and use Lemma 44 to obtain new vectors $y_1,\ldots,y_n \in X$ such that

$$\Big(\frac{p}{2q}\Big)^p \cdot \bigg(\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X^q}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X^q}\bigg)^{\frac{p}{q}} \leqslant \frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|y_i-y_j\|_X^p}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|y_i-y_j\|_X^p} \leqslant \gamma(\mathsf{A}, \|\cdot\|_X^p), \tag{132}$$

where the last step of (132) is the definition of $\gamma(\mathsf{A}, \|\cdot\|_X^p)$ applied to the new configuration of vectors $\{y_1,\ldots,y_n\} \subset X$. It remains to note that by the definition of $\gamma(\mathsf{A}, \|\cdot\|_X^q)$, the supremum of the left hand side of (132) over all possible $x_1,\ldots,x_n \in X$ equals the left hand side of (131).

To deduce the second inequality in (131), use Proposition 37 and Corollary 40 with $\omega = p/q$ and $(\mu(x_1),\ldots,\mu(x_n)) = \pi$ to get new vectors $z_1,\ldots,z_n \in X$ satisfying $\|z_i-z_j\|_X^q \leqslant (p/2q)^q\|x_i-x_j\|_X^p$ for all $i,j \in \{1,\ldots,n\}$, and also $\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|z_i-z_j\|_X^q \geqslant \sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X^p$. Thus,

$$\frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_X^p}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_X^p} \leqslant \Big(\frac{2q}{p}\Big)^q \cdot \frac{\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|z_i-z_j\|_X^q}{\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|z_i-z_j\|_X^q} \leqslant \Big(\frac{2q}{p}\Big)^q \gamma(\mathsf{A}, \|\cdot\|_X^q).$$

6. Impossibility results

The main purpose of this section is to prove Lemma 2, Lemma 13 and to refine the lower bound (8) on the Hilbertian average distortion of snowflakes of regular graphs with a spectral gap.

The assertion of Lemma 2 that the ω-snowflake of any k-dimensional normed space $(X,\|\cdot\|_X)$ embeds into a Hilbert space with bi-Lipschitz distortion $k^{\frac{\omega}{2}}$ is a quick consequence of John's theorem [67], combined with Schoenberg's result [131] that the ω-snowflake of an infinite dimensional Hilbert space $(H,\|\cdot\|_H)$ embeds isometrically into H. Indeed, by John's theorem there exists a mapping $f: X \to H$ such that $\|x-y\|_X \leqslant \|f(x)-f(y)\|_H \leqslant \sqrt{k}\|x-y\|_X$ for all $x,y \in X$. By

Schoenberg's theorem there exists a mapping $g : H \to H$ such that $\|g(u) - g(v)\|_H = \|u - v\|_H^{\omega}$ for all $u, v \in H$. Hence, $\|x - y\|_X^{\omega} \leqslant \|g \circ f(x) - g \circ f(y)\|_H \leqslant k^{\frac{\omega}{2}} \|x - y\|_X^{\omega}$ for all $x, y \in X$.

The more substantial novelty of Lemma 2 is the assertion that the above composition of Schoenberg's embedding and John's embedding yields the correct worst-case asymptotic behavior (up to universal constant factors) of the Hilbertian bi-Lipschitz distortion of ω-snowflakes of k-dimensional normed spaces. The proof is a quick application of the invariant *metric cotype* 2 *with sharp scaling parameter* that was introduced in [96], but this has not been previously noted in the literature. Note that the endpoint case $\omega = 1$ here is classical, by a reduction to the linear theory through differentiation, but this approach is inherently unsuitable for treating Hölder functions.

Proof of Lemma 2. Fix $k \in \mathbb{N}$ and consider the normed space $\ell_\infty^{k^2}(\mathbb{C}) \cong \ell_\infty^{k^2}(\ell_2^2)$ whose dimension over $\mathbb{R}$ equals $2k^2$. Suppose that the ω-snowflake of $\ell_\infty^{k^2}(\mathbb{C})$ embeds into a Hilbert space $(H, \|\cdot\|_H)$ with bi-Lipschitz distortion less than D. Thus, there exists an embedding $f : \ell_\infty^{k^2}(\mathbb{C}) \to H$ such that

$$\forall\, x, y \in \ell_\infty^{k^2}(\mathbb{C}), \qquad \|x - y\|^{\omega}_{\ell_\infty^{k^2}(\mathbb{C})} \leqslant \|f(x) - f(y)\|_H \leqslant D\|x - y\|^{\omega}_{\ell_\infty^{k^2}(\mathbb{C})}. \tag{133}$$

By [96, Section 3], the following inequality holds true for any $f : \{1, \ldots, 4k\}^{k^2} \to H$.

$$\begin{aligned}\frac{1}{(4k)^{k^2}} \sum_{j=1}^{k^2} \sum_{x \in \{1,\ldots,4k\}^{k^2}} &\bigg\| f\Big(- e^{\frac{\pi i}{2k} x_j} e_j + \sum_{r \in \{1,\ldots,k^2\} \smallsetminus \{j\}} e^{\frac{\pi i}{2k} x_r} e_r \Big) - f\Big(\sum_{r=1}^{k^2} e^{\frac{\pi i}{2k} x_r} e_r \Big) \bigg\|_H^2 \\ &\lesssim \frac{k^2}{(12k)^{k^2}} \sum_{\varepsilon \in \{-1,0,1\}^{k^2}} \sum_{x \in \{1,\ldots,4k\}^{k^2}} \bigg\| f\Big(\sum_{r=1}^{k^2} e^{\frac{\pi i}{2k}(x_r + \varepsilon_r)} e_r \Big) - f\Big(\sum_{r=1}^{k^2} e^{\frac{\pi i}{2k} x_r} e_r \Big) \bigg\|_H^2,\end{aligned} \tag{134}$$

where $e_1, \ldots, e_{k^2}$ is the standard basis of $\mathbb{C}^{k^2}$. By combining (133) with (134), we conclude that

$$2^{2\omega} k^2 \lesssim k^2 D^2 \Big| e^{\frac{\pi i}{2k}} - 1 \Big|^{2\omega} \asymp k^{2(1-\omega)} D^2 \iff D \gtrsim k^{\omega} \asymp \dim\big(\ell_\infty^{k^2}(\mathbb{C})\big)^{\frac{\omega}{2}}. \qquad \square$$

The above proof of Lemma 2 works also for embeddings of ω-snowflakes of k-dimensional normed spaces into L_p when $p \in [1, 2]$, yielding the same conclusion. However, for $p > 2$ the upper and lower bounds that it yields (using sharp metric cotype p) do not match. We therefore ask

Question 48. Suppose that $\omega \in (0, 1)$ and $p \in (2, \infty)$. What is the infimum over those $\beta \in (0, 1]$ for which there exists $\alpha_\beta \in (0, \infty)$ such that the ω-snowflake of any finite-dimensional normed space X embeds into L_p with bi-Lipschitz distortion at most $\alpha_\beta \dim(X)^{\beta}$?

Next, the optimality of Theorem 12 for Hilbertian targets in the regime of higher Hölder regularity, as exhibited by Lemma (13), is a quick application of the classical invariant *Enflo type* [49, 30].

Proof of Lemma 13. Fix $k \in \mathbb{N}$ and denote $\mathsf{c} \stackrel{\mathrm{def}}{=} k^{\frac{1}{p} - \frac{1}{2}}$, so that $\mathsf{c}_{\ell_2}(\ell_p^k) = \mathsf{c}$; see e.g. [69, Section 8]. Choose $X = \ell_p^k$, so that it has modulus of uniform smoothness of power type p. Suppose that $(Z, \|\cdot\|_Z)$ is a normed space whose modulus of uniform smoothness has power type 2. We will show that if the $(\frac{p}{2} + \varepsilon)$-snowflake of ℓ_p^k embeds into Z with α-average distortion D, then necessarily

$$D \gtrsim \frac{\mathsf{c}^{\frac{2\varepsilon}{2-p}}}{\sqrt{\alpha} + \mathcal{S}_2(Z)}. \tag{135}$$

The desired lower bound (19) in Lemma (13) is the special case of (135) corresponding to $Z = \ell_\beta(\ell_2)$, because $\ell_\beta(\ell_2)$ is isometric to a subspace of L_β, and $\mathcal{S}_2(L_\beta) \asymp \sqrt{\beta}$ by [51, 21].

Let $\mathbb{F}_2$ be the field of two elements. Identify $\mathbb{F}_2^k$ with the 2^k vertices of the hypercube $\{0,1\}^k \subset \ell_p^k$. We also let $e_1, \ldots, e_k$ denote the standard coordinate basis and μ denote the uniform probability measure on $\mathbb{F}_2^k \subset \ell_p^k$, respectively. By [107, equation (6.32)], any $f : \mathbb{F}_2^k \to Z$ satisfies the following bound, in which additions that occur in the argument of f are in $\mathbb{F}_2^k$, i.e., modulo 2 coordinate-wise.

$$\begin{aligned}\bigg(\iint_{\mathbb{F}_2^k\times\mathbb{F}_2^k} \|f(x)-f(y)\|_Z^{\alpha}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac{1}{\alpha}} \\ \lesssim \big(\mathcal{S}_2(Z)+\sqrt{\alpha}\big)\sqrt{k}\bigg(\frac{1}{k}\sum_{i=1}^{k}\int_{\mathbb{F}_2^k}\|f(x+e_i)-f(x)\|_Z^{\alpha}\,\mathrm{d}\mu(x)\bigg)^{\frac{1}{\alpha}}.\end{aligned} \tag{136}$$

Suppose that $\|f(x+e_i)-f(x)\|_Z \leqslant D$ for every $x \in \mathbb{F}_2^k$ and $i \in \{1,\ldots,k\}$. This would follow if f were $(\frac{p}{2}+\varepsilon)$-Hölder with constant D, which is what is relevant in the present context, but we are in fact assuming significantly less here, namely that f is D-Lipschitz in the metric that is induced by ℓ_1^k on $\mathbb{F}_2^k$. Under this assumption, the right hand side of (136) is at most $D\big(\mathcal{S}_2(Z)+\sqrt{\alpha}\big)\sqrt{k}$.

If we also have

$$\bigg(\iint_{\mathbb{F}_2^k\times\mathbb{F}_2^k} \|f(x)-f(y)\|_Z^{\alpha}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac{1}{\alpha}} \geqslant \bigg(\iint_{\mathbb{F}_2^k\times\mathbb{F}_2^k} \|x-y\|_{\ell_p^k}^{\alpha(\frac{p}{2}+\varepsilon)}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\bigg)^{\frac{1}{\alpha}} \gtrsim k^{\frac12+\frac{\varepsilon}{p}},$$

where the last step holds because $\|x-y\|_{\ell_p^k} \gtrsim k^{\frac{1}{p}}$ for a constant fraction of $(x,y) \in \mathbb{F}_2^k \times \mathbb{F}_2^k$, then by contrasting this with (136) we conclude that

$$D \gtrsim \frac{k^{\frac{\varepsilon}{p}}}{\sqrt{\alpha}+\mathcal{S}_2(Z)} = \frac{\mathsf{C}^{\frac{2\varepsilon}{2-p}}}{\sqrt{\alpha}+\mathcal{S}_2(Z)}. \qquad \square$$

Note that the case $p=1$ of the above proof of Lemma 13 gives that if the $(\frac12+\varepsilon)$-snowflake of ℓ_1^k embeds with quadratic average distortion $D \geqslant 1$ into a Hilbert space, then necessarily $D \gtrsim k^{\varepsilon}$. A slightly more careful examination (using Enflo's original "diagonal versus edge" inequality in [47] in place of (136)) of the constant factors in this special case reveals that we actually get the sharp bound $D \geqslant k^{\varepsilon}$ (for the uniform measure on $\{0,1\}^k$), as stated in the Introduction.

We will next revisit the estimate (8) that was derived in the Introduction. Recalling the relevant setting, we are given $n \in \mathbb{N}$ and a connected regular graph $\mathsf{G} = (\{1,\ldots,n\}, E_{\mathsf{G}})$. In the course of the deduction of (8), see specifically the penultimate step in (7), we actually showed that if the $\frac12$-snowflake of the shortest-path metric $(\{1,\ldots,n\}, d_{\mathsf{G}})$ embeds into a Hilbert space $(H,\|\cdot\|_H)$ with quadratic average distortion $D \geqslant 1$, then necessarily

$$D \geqslant \sqrt{1-\lambda_2(\mathsf{G})}\bigg(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{\mathsf{G}}(i,j)\bigg)^{\frac12}.$$

Replacing the Hölder exponent $\frac12$ by an arbitrary $\omega \in (0,1]$, the same reasoning shows mutatis mutandis that if the ω-snowflake of the shortest-path metric $(\{1,\ldots,n\}, d_{\mathsf{G}})$ embeds into a Hilbert space $(H,\|\cdot\|_H)$ with quadratic average distortion $D \geqslant 1$, then necessarily

$$D \geqslant \sqrt{1-\lambda_2(\mathsf{G})}\bigg(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{\mathsf{G}}(i,j)^{2\omega}\bigg)^{\frac12}. \tag{137}$$

In particular, this implies that if G is an expander, then $D \gtrsim (\log n)^{\omega}$. Therefore, by considering the uniform measure on an isometric copy of $(\{1,\ldots,n\}, d_{\mathsf{G}})$ in ℓ_∞^n we see that Conjecture 42 asks for the optimal asymptotic dependence on the dimension for fixed $\omega \in (0,\frac12)$.

As we explained in the Introduction, the above proof of the "vanilla" spectral bound (137) goes back to [84, 92, 56]; further examples of implementations of this (by now standard) useful idea can be found in [85, 86, 120, 23, 114, 74, 78, 117, 124, 116, 55, 99, 107, 71, 113, 111]. An especially important special case of (137) is when G is a vertex-transitive graph, e.g. when it is the Cayley graph of a group of order n. In this case, by equation (4.24) in [107] (see also [117, 63]) we have

$$\forall\, p \geqslant 1,\ \forall\, \omega \in (0,1], \qquad \bigg(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{\mathsf{G}}(i,j)^{p\omega}\bigg)^{\frac{1}{p}} \asymp \mathrm{diam}(\mathsf{G})^{\omega}, \tag{138}$$

where $\mathrm{diam}(\mathsf{G})$ is the diameter of $(\{1,\ldots,n\},d_{\mathsf{G}})$. So, (137) for a vertex transitive graph becomes

$$D \gtrsim \sqrt{1-\lambda_2(\mathsf{G})}\,\mathrm{diam}(\mathsf{G})^{\omega}. \tag{139}$$

The following lemma improves the dependence on the spectral gap in (137) and (139) for small ω.

Lemma 49. *Fix $n \in \mathbb{N}$, $\omega \in (0,1]$ and $p,q,D \in [1,\infty)$. Let $\mathsf{G} = (\{1,\ldots,n\},E_{\mathsf{G}})$ be a connected regular graph such that the ω-snowflake of the metric space $(\{1,\ldots,n\},d_{\mathsf{G}})$ embeds with q-average distortion less than D into ℓ_p. Then necessarily*

$$D \gtrsim \frac{\big(1-\lambda_2(\mathsf{G})\big)^{\min\left\{\omega,\frac{1}{\min\{p,2\}}\right\}}}{(p^2+q^2)^{\frac{1}{\min\{p,2\}}}}\bigg(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{\mathsf{G}}(i,j)^{q\omega}\bigg)^{\frac{1}{q}}. \tag{140}$$

In particular, if G is a vertex-transitive graph, then

$$D \gtrsim \frac{\big(1-\lambda_2(\mathsf{G})\big)^{\min\left\{\omega,\frac{1}{\min\{p,2\}}\right\}}}{(p^2+q^2)^{\frac{1}{\min\{p,2\}}}}\,\mathrm{diam}(\mathsf{G})^{\omega}. \tag{141}$$

Prior to proving Lemma 49, we will discuss some of its consequences.

Example 50. Contrast Lemma 49 with (139) in the following illustrative classical example. Fix $\mathfrak{q},k \in \mathbb{N}$ such that $\mathfrak{q}$ is a power of a prime. Consider the Cayley graph of $\mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}})$ that is induced by the symmetric generating set $\{\mathsf{I}_k \pm \mathsf{E}_k(i,j) : (i,j) \in \{1,\ldots,k\}^2 \ \wedge\ i \neq j\}$. Here, $\mathbb{F}_{\mathfrak{q}}$ is the field of size $\mathfrak{q}$ and $\mathsf{E}_k(i,j) \in \mathsf{M}_k(\mathbb{F}_{\mathfrak{q}})$ is the elementary matrix whose (i,j)-entry equals 1 and the rest of its entries vanish. In what follows, $\mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}})$ will always be assumed to be equipped with the word metric that corresponds to this (standard) generating set. We then have

$$1-\lambda_2\big(\mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}})\big) \asymp \frac{1}{k} \qquad \text{and} \qquad \mathrm{diam}\,\big(\mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}})\big) \asymp \frac{k^2\log\mathfrak{q}}{\log k}. \tag{142}$$

The first assertion in (142) is due to Kassabov [73]. The second assertion in (142) was obtained by Alon [3] who extended[5] a similar algorithm of Andrén, Hellström and Markström [14] that proves it for $\mathfrak{q} = O(1)$; see also [130] for prior diameter bounds. If $\omega \in (0,1]$ and the ω-snowflake of $\mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}})$ embeds with quadratic average distortion $D \geqslant 1$ into a Hilbert space, then by (142) and (139),

$$D \gtrsim \frac{k^{2\omega-\frac{1}{2}}(\log\mathfrak{q})^{\omega}}{(\log k)^{\omega}}.$$

This bound is vacuous if $\omega < \frac{1}{4}$. However, if we use Lemma 49 in place of (139) we get the following lower bound on D which tends to ∞ as $k \to \infty$ in the entire range $\omega \in (0,1]$.

$$D \gtrsim \frac{k^{\max\left\{2\omega-\frac{1}{2},\omega\right\}}(\log\mathfrak{q})^{\omega}}{(\log k)^{\omega}}.$$

Remark 51. Lemma 49 in the case when G is an expander, namely it is both $O(1)$-regular and $1/(1-\lambda_2(\mathsf{G})) = O(1)$, shows that the case $p=1$ of the first assertion (17) of Theorem 12 is sharp for every $q \geqslant 2$, up to a multiplicative factor which may depend on only q. Indeed, take $Y = \ell_q$ and $X = \ell_\infty^n$. Then, the modulus of uniform convexity of Y has power type q and $\mathsf{c}_Y(X) = n^{1/q}$. By

[5] Alon obtained the second assertion in (142) independently, before we learned of the earlier work [14].

considering the uniform distribution over the image of an isometric embedding of $(\{1,\ldots,n\},d_{\mathsf{G}})$ into X, we see from Lemma 49 that if X embeds with q-average distortion D into Y, then necessarily

$$D \gtrsim \frac{1}{q}(\log n)^{\frac{1}{q}} \asymp \frac{1}{q}\big(\log \mathsf{c}_Y(X)\big)^{\frac{1}{q}}. \tag{143}$$

Our proof of Lemma 49 uses the following lemma, the case $p>q$ of which is a mixed-exponent variant of Matoušek's extrapolation phenomenon for Poincaré inequalities [92, 23, 116]. One could avoid treating mixed exponents and obtain a "vanilla" extrapolation inequality by using [46] (recall also Remark 47), but this leads to an asymptotically worse dependence on the spectral gap even when, say, $p=2$ and $1\leqslant q<2$, which is an inherent deficiency: If one considers the variant of (144) below with $p=2$ and the q'th moment on both sides for some $q\in[1,2)$, then the power of $1/(1-\lambda_2(\mathsf{A}))$ becomes $1/q$ rather than the stated $1/2<1/q$, and this is sharp, for example, when A is the transition matrix of the standard random walk on the k-dimensional Hamming cube $\{0,1\}^k$.

Lemma 52. *Fix $p,q\geqslant 1$, $n\in\mathbb{N}$ and $\pi=(\pi_1,\ldots,\pi_n)\in\triangle^{n-1}$. Suppose that $\mathsf{A}=(a_{ij})\in\mathsf{M}_n(\mathbb{R})$ is a stochastic and π-reversible matrix. Then, every $x_1,\ldots,x_n\in\ell_p$ satisfy the inequality*

$$\bigg(\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_{\ell_p}^q\bigg)^{\frac{1}{q}} \lesssim \bigg(\frac{p^2+q^2}{1-\lambda_2(\mathsf{A})}\bigg)^{\frac{1}{\min\{p,2\}}}\bigg(\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_{\ell_p}^{\max\{p,q\}}\bigg)^{\frac{1}{\max\{p,q\}}}. \tag{144}$$

Proof. By the case $X=\mathbb{R}$ (and $p=2$) of the first inequality in (131), for every $\beta\geqslant 2$ we have

$$\forall\, s_1,\ldots,s_n\in\mathbb{R},\qquad \sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j|s_i-s_j|^\beta \leqslant \bigg(\frac{\beta}{\sqrt{1-\lambda_2(\mathsf{A})}}\bigg)^\beta \sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}|s_i-s_j|^\beta. \tag{145}$$

The scalar inequality (145) with slightly weaker constant factor appears in [116, Lemma 4.4], as a natural quadratic variant (via a similar proof) of Matoušek's extrapolation lemma for Poincaré inequalities [92], which is the analogous ℓ_1 statement, namely with "spectral gap" replaced by "Cheeger constant." By a point-wise application of (145) followed by integration, we see that that

$$\forall\, f_1,\ldots,f_n\in L_\beta,\qquad \sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|f_i-f_j\|_{L_\beta}^\beta \leqslant \bigg(\frac{\beta}{\sqrt{1-\lambda_2(\mathsf{A})}}\bigg)^\beta \sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|f_i-f_j\|_{L_\beta}^\beta. \tag{146}$$

Since L_2 is isometric to a subset of L_β, it follows from (146) that also

$$\forall\, f_1,\ldots,f_n\in L_2,\qquad \sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|f_i-f_j\|_{L_2}^\beta \leqslant \bigg(\frac{\beta}{\sqrt{1-\lambda_2(\mathsf{A})}}\bigg)^\beta \sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|f_i-f_j\|_{L_2}^\beta. \tag{147}$$

Suppose first that $q\geqslant p\geqslant 2$. Then, (146) with $\beta=p$ is the same as the estimate

$$\gamma\big(\mathsf{A},\|\cdot\|_{L_p}^p\big) \leqslant \bigg(\frac{p}{\sqrt{1-\lambda_2(\mathsf{A})}}\bigg)^p. \tag{148}$$

We therefore obtain the following bound which implies the desired inequality (144) when $q\geqslant p\geqslant 2$.

$$\gamma\big(\mathsf{A},\|\cdot\|_{L_p}^q\big) \overset{(131)}{\leqslant} \Big(\frac{2q}{p}\Big)^q \gamma\big(\mathsf{A},\|\cdot\|_{L_p}^p\big)^{\frac{q}{p}} \overset{(148)}{\leqslant} \bigg(\frac{2q}{\sqrt{1-\lambda_2(\mathsf{A})}}\bigg)^q.$$

If $1\leqslant p\leqslant 2$ and $q\geqslant p$, then by [131] there exist $f_1,\ldots,f_n\in L_2$ such that

$$\forall\, i,j\in\{1,\ldots,n\},\qquad \|f_i-f_j\|_{L_2} = \|x_i-x_j\|_{\ell_p}^{\frac{p}{2}}. \tag{149}$$

An application of (147) with $\beta = \frac{2q}{p} \geqslant 2$ now shows that

$$\begin{aligned}\Bigg(\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_{\ell_p}^q\Bigg)^{\frac1q} &\overset{(149)}{=} \Bigg(\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|f_i-f_j\|_{L_2}^{\frac{2q}{p}}\Bigg)^{\frac1q}\\
&\overset{(147)}{\leqslant} \left(\frac{2q}{p\sqrt{1-\lambda_2(\mathsf{A})}}\right)^{\frac2p}\Bigg(\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|f_i-f_j\|_{L_2}^{\frac{2q}{p}}\Bigg)^{\frac1q}\\
&\overset{(149)}{\leqslant} \left(\frac{2q}{\sqrt{1-\lambda_2(\mathsf{A})}}\right)^{\frac2p}\Bigg(\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_{\ell_p}^{q}\Bigg)^{\frac1q}.\end{aligned}$$

This completes the proof of (144) in the entire range $q \geqslant p \geqslant 1$.

Suppose next that $p \geqslant q \geqslant 1$ and $p \geqslant 2$. Writing $x_i = (x_{i1}, x_{i2}, \ldots) \in \ell_p$ for each $i \in \{1,\ldots,n\}$,

$$\begin{aligned}\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\|x_i-x_j\|_{\ell_p}^q &= \sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j\Bigg(\sum_{k=1}^\infty |x_{ik}-x_{jk}|^p\Bigg)^{\frac qp}\\
&\leqslant \Bigg(\sum_{k=1}^\infty\sum_{i=1}^n\sum_{j=1}^n \pi_i\pi_j|x_{ik}-x_{jk}|^p\Bigg)^{\frac qp} && (150)\\
&\leqslant \Bigg(\sum_{k=1}^\infty\left(\frac{p}{\sqrt{1-\lambda_2(\mathsf{A})}}\right)^p\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}|x_{ik}-x_{jk}|^p\Bigg)^{\frac qp} && (151)\\
&= \left(\frac{p}{\sqrt{1-\lambda_2(\mathsf{A})}}\right)^q\Bigg(\sum_{i=1}^n\sum_{j=1}^n \pi_i a_{ij}\|x_i-x_j\|_{\ell_p}^p\Bigg)^{\frac qp}, && (152)\end{aligned}$$

where (150) is a consequence of the concavity (since $q \leqslant p$) of the function $(u \geqslant 0) \mapsto u^{\frac qp}$ and (151) is a coordinate-wise application (with $\beta = p \geqslant 2$) of the scalar inequality (145). This establishes (144) when $p \geqslant q \geqslant 1$ and $p \geqslant 2$. If $1 \leqslant q \leqslant p \leqslant 2$, then (144) follows by using (152) with $x_1,\ldots,x_n$ replaced by $f_1,\ldots,f_n$ that satisfy (149), with p replaced by 2 and with q replaced by $\frac{2q}{p} \leqslant 2$. □

Proof of Lemma 49. By assumption, there exists an embedding $f:\{1,\ldots,n\}\to\ell_p$ such that

$$\forall\, i,j\in\{1,\ldots,n\},\qquad \|f(i)-f(j)\|_{\ell_p} \leqslant D d_{\mathsf{G}}(i,j)^{\omega}. \tag{153}$$

and

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|f(i)-f(j)\|_{\ell_p}^q \geqslant \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_{\mathsf{G}}(i,j)^{q\omega} \tag{154}$$

Our task is to bound D from below. Using Lemma 52 with $\mathsf{A} = \mathsf{A}_{\mathsf{G}} \in \mathsf{M}_n(\mathbb{R})$ the adjacency matrix of G, $\pi = (\frac1n,\ldots,\frac1n) \in \triangle^{n-1}$ (G is a regular graph) and $x_i = f(i)$ for all $i \in \{1,\ldots,n\}$, we see that

$$\begin{aligned}\Bigg(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_{\mathsf{G}}(i,j)^{q\omega}\Bigg)^{\frac1q} &\overset{(154)}{\leqslant} \Bigg(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|f(i)-f(j)\|_{\ell_p}^q\Bigg)^{\frac1q}\\
&\overset{(144)}{\lesssim} \left(\frac{p^2+q^2}{1-\lambda_2(\mathsf{G})}\right)^{\frac{1}{\min\{p,2\}}}\Bigg(\frac{1}{|E_{\mathsf{G}}|}\sum_{\{i,j\}\in E_{\mathsf{G}}}^n \|f(i)-f(j)\|_{\ell_p}^{\max\{p,q\}}\Bigg)^{\frac{1}{\max\{p,q\}}}\\
&\overset{(153)}{\leqslant} \left(\frac{p^2+q^2}{1-\lambda_2(\mathsf{G})}\right)^{\frac{1}{\min\{p,2\}}} D.\end{aligned}$$

This is the desired bound (140) when $\omega \geqslant 1/\min\{p,2\}$. Note that thus far we used (153) only for those $i,j \in \{1,\ldots,n\}$ such that $\{i,j\} \in E_{\mathsf{G}}$. In other words, we derived (140) under the assumption that f is D-Lipschitz rather than that f is ω-Hölder with constant D, which is a more stringent requirement as d_{G} takes values in $[1,\infty)\cup\{0\}$. To prove (140) when $\omega \leqslant 1/\min\{p,2\}$ we will probe larger distances in $(\{1,\ldots,n\},d_{\mathsf{G}})$ for which the full Hölder condition (153) gives more information.

Denote

$$s \stackrel{\mathrm{def}}{=} \left\lceil \frac{1}{1-\lambda_2(\mathsf{G})} \right\rceil. \tag{155}$$

The function $t \mapsto \left(\frac{1+t}{2}\right)^{\frac{1}{1-t}}$ is increasing on $(-1,1)$ and tends to $\frac{1}{\sqrt{e}}$ as $t \to 1^-$. Hence,

$$\left(\frac{1+\lambda_2(\mathsf{G})}{2}\right)^{s} \leqslant \left(\frac{1+\lambda_2(\mathsf{G})}{2}\right)^{\frac{1}{1-\lambda_2(\mathsf{G})}} \leqslant \frac{1}{\sqrt{e}}. \tag{156}$$

Using Lemma 52 with $\mathsf{A} = \left(\frac12 \mathsf{I}_n + \frac12 \mathsf{A}_{\mathsf{G}}\right)^s$, we therefore see that

$$\begin{aligned}
&\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_{\mathsf{G}}(i,j)^{q\omega}\right)^{\frac1q}\\
&\stackrel{(154)}{\leqslant} \left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|f(i)-f(j)\|_{\ell_p}^q\right)^{\frac1q}\\
&\stackrel{(144)}{\lesssim} \left(\frac{p^2+q^2}{1-\lambda_2\left(\left(\frac12 \mathsf{I}_n+\frac12\mathsf{A}_{\mathsf{G}}\right)^s\right)}\right)^{\frac{1}{\min\{p,2\}}} \left(\frac1n \sum_{i=1}^n\sum_{j=1}^n \left(\frac12 \mathsf{I}_n + \frac12 \mathsf{A}_{\mathsf{G}}\right)^s_{ij} \|f(i)-f(j)\|_{\ell_p}^{\max\{p,q\}}\right)^{\frac{1}{\max\{p,q\}}}\\
&\stackrel{(153)\wedge(156)}{\lesssim} \left(p^2+q^2\right)^{\frac{1}{\min\{p,2\}}} s^{\omega} D \stackrel{(155)}{\asymp} \frac{\left(p^2+q^2\right)^{\frac{1}{\min\{p,2\}}}}{\left(1-\lambda_2(\mathsf{G})\right)^{\omega}} D,
\end{aligned}$$

where the penultimate step uses that if for some $i,j\in\{1,\ldots,n\}$ the (i,j)-entry of $(\frac12 \mathsf{I}_n + \frac12\mathsf{A}_{\mathsf{G}})^s$ is nonzero, then there is a walk in G of length at most s from i to j, hence $d_{\mathsf{G}}(i,j) \leqslant s$. $\square$

We end this section with a few further remarks and open questions.

Proposition 53. *Fix $\mathfrak{q},k \in \mathbb{N}$ such that $\mathfrak{q}$ is a power of a prime. Let $\mathsf{Av}(k,\mathfrak{q})$ denote the smallest $D \geqslant 1$ such that $\mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}})$ embeds into a Hilbert space with average distortion D. Then*

$$\mathsf{Av}(k,\mathfrak{q}) \asymp (\log \mathfrak{q})\frac{k^{\frac32}}{\log k} \asymp \sqrt[4]{\log \mathfrak{q}} \cdot \frac{(\log|\mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}})|)^{\frac34}}{\log\log|\mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}})| - \log\log \mathfrak{q}}.$$

Proof. A substitution of (142) into (141) when $\omega = q = 1$ and $p=2$ gives $\mathsf{Av}(k,\mathfrak{q}) \gtrsim \frac{k^{\frac32}\log \mathfrak{q}}{\log k}$. To prove the matching upper bound, suppose that $\mathfrak{q} = \mathfrak{p}^m$ for some prime $\mathfrak{p}$ and $m \in \mathbb{N}$. Let $v_1,\ldots,v_m$ be a basis of $\mathbb{F}_{\mathfrak{q}}$ over $\mathbb{F}_{\mathfrak{p}}$. Thus, for every $x \in \mathbb{F}_{\mathfrak{q}}$ there are unique $\chi_1(x),\ldots,\chi_m(x) \in \mathbb{Z}/\mathfrak{p}\mathbb{Z}$ such that $x = \chi_1(x)v_1 + \ldots + \chi_m(x)v_m$. Define an embedding

$$f : \mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}}) \to \underbrace{\mathsf{M}_k(\mathbb{C}) \oplus \ldots \oplus \mathsf{M}_k(\mathbb{C})}_{m \text{ times}} \cong \ell_2^{2mk^2}$$

by setting for some $C>0$,

$$\forall\, \mathsf{X} = (x_{jk}) \in \mathsf{SL}_k(\mathbb{F}_{\mathfrak{q}}), \qquad f(\mathsf{X}) \stackrel{\mathrm{def}}{=} \bigoplus_{s=1}^m \frac{Ck\log q}{\sqrt{m}\log k}\left(e^{\frac{2\pi \mathrm{i}}{\mathfrak{p}}\chi_s(x_{jk})}\right)_{(j,k)\in\{1,\ldots,k\}^2},$$

We claim that if C is a sufficiently large universal constant, then f exhibits that $\mathsf{Av}(k,\mathfrak{q}) \lesssim \frac{k^{\frac32}\log \mathfrak{q}}{\log k}$.

Fix distinct indices $\alpha, \beta \in \{1, \ldots, k\}$. Then for every $\mathsf{X} = (x_{jk}) \in \mathsf{SL}_k(\mathbb{F}_\mathfrak{q})$ we have

$$f\Big(\mathsf{X}\big(\mathsf{I}_k \pm \mathsf{E}_k(\alpha, \beta)\big)\Big) - f(\mathsf{X}) = \bigoplus_{s=1}^{m} \frac{Ck \log q}{\sqrt{m} \log k} \Big(\delta_{s\beta} e^{\frac{2\pi \mathsf{i}}{\mathfrak{p}}\chi_s(x_{j\beta})}\Big(e^{\pm \frac{2\pi \mathsf{i}}{\mathfrak{p}}\chi_s(x_{j\alpha})} - 1\Big)\Big)_{(j,k) \in \{1,\ldots,k\}^2},$$

where $\delta_{s\beta}$ is the Kronecker delta. Thus,

$$\Big\| f\Big(\mathsf{X}\big(\mathsf{I}_k \pm \mathsf{E}_k(\alpha, \beta)\big)\Big) - f(\mathsf{X}) \Big\|_{\ell_2^{2mk^2}} \leqslant \frac{Ck \log q}{\sqrt{m} \log k} \cdot 2\sqrt{km} = \frac{2k^{\frac{3}{2}} \log \mathfrak{q}}{\log k}.$$

By the definition of the (word) metric on $\mathsf{SL}_k(\mathbb{F}_\mathfrak{q})$, this means that f is $\frac{2k^{\frac{3}{2}} \log \mathfrak{q}}{\log k}$-Lipschitz.

In the reverse direction, if $\mathsf{X} = (x_{jk}), \mathsf{Y} = (y_{jk})$ are independent and chosen uniformly at random from $\mathsf{SL}_k(\mathbb{F}_\mathfrak{q})$, then with probability that is bounded below by a positive universal constant, we have $|\exp(2\pi \mathsf{i}\chi_s(x_{jk}))/\mathfrak{p} - \exp(2\pi \mathsf{i}\chi_s(y_{jk})/\mathfrak{p})| \gtrsim 1$ for a universal constant fraction of the mk^2 triples $(i, j, s) \in \{1, \ldots, k\} \times \{1, \ldots, k\} \times \{1, \ldots, m\}$. Therefore,

$$\begin{aligned}\frac{1}{|\mathsf{SL}_k(\mathbb{F}_\mathfrak{q})|^2} \sum_{\mathsf{X},\mathsf{Y} \in \mathsf{SL}_k(\mathbb{F}_\mathfrak{q})} \|f(\mathsf{X}) - f(\mathsf{Y})\|_{\ell_2^{2mk^2}} &\gtrsim \frac{Ck \log q}{\sqrt{m} \log k} \cdot \sqrt{mk^2} = \frac{Ck^2 \log \mathfrak{q}}{\log k} \\ &\overset{(142)}{\gtrsim} C \operatorname{diam}\big(\mathsf{SL}_k(\mathbb{F}_\mathfrak{q})\big) \geqslant \frac{C}{|\mathsf{SL}_k(\mathbb{F}_\mathfrak{q})|^2} \sum_{\mathsf{X},\mathsf{Y} \in \mathsf{SL}_k(\mathbb{F}_\mathfrak{q})} d_{\mathsf{SL}_k(\mathbb{F}_\mathfrak{q})}(\mathsf{X}, \mathsf{Y}). \qquad \square\end{aligned}$$

The following conjecture asserts that (at least for fixed $\mathfrak{q}$) the curious-looking but nonetheless sharp asymptotic behavior of Proposition 53 holds also for bi-Lipschitz embeddings; we suspect that its resolution is tractable, perhaps via the representation-theoretic approach of [17].

Conjecture 54. For every $k \in \mathbb{N}$ and prime power $\mathfrak{q}$ we have $\mathsf{c}_{\ell_2}\big(\mathsf{SL}_k(\mathbb{F}_\mathfrak{q})\big) \asymp_\mathfrak{q} \frac{k^{\frac{3}{2}}}{\log k}$.

Remark 55. By (14) and (138) we see that if $\mathsf{G} = (\{1, \ldots, n\}, E_\mathsf{G})$ is a vertex-transitive graph, then

$$\forall D \geqslant 1, \qquad \dim_D(\mathsf{G}) \geqslant e^{\frac{c}{D}(1-\lambda_2(\mathsf{G}))\operatorname{diam}(\mathsf{G})}, \tag{157}$$

where $c > 0$ is a universal constant, and the notation $\dim_D(\cdot)$ of [84, Definition 2.1] was recalled in Section 1.4. In fact, this reasoning shows that (Theorem 1 implies that) if $(\{1, \ldots, n\}, d_\mathsf{G})$ embeds into a normed space X with average distortion D (rather than the stronger bi-Lipschitz distortion D to which (157) alludes), then necessarily $\dim(X) \geqslant \exp(c(1 - \lambda_2(\mathsf{G}))\operatorname{diam}(\mathsf{G})/D)$. It follows in particular from (157) that if $(\{1, \ldots, n\}, d_\mathsf{G})$ embeds with average distortion $O(1)$ into some normed space of dimension $(\log n)^{O(1)}$, then necessarily $(1 - \lambda_2(\mathsf{G}))\operatorname{diam}(\mathsf{G}) \lesssim \log\log n$.

There are many examples of Cayley graphs $\mathsf{G} = (\{1, \ldots, n\}, E_\mathsf{G})$ for which $\lambda_2(\mathsf{G}) = 1 - \Omega(1)$ and $\operatorname{diam}(\mathsf{G}) \gtrsim \log n$ (see e.g. [5, 117]). In all such examples, (157) asserts that $\dim_D(\mathsf{G}) \gtrsim n^{c/D}$ for some universal constant $c > 0$. The Cayley graph that was studied in [74] (a quotient of the Hamming cube by a good code) now shows that there exist arbitrarily large n-point metric spaces $\mathcal{M}_n$ with $\dim_1(\mathcal{M}_n) \lesssim \log n$ (indeed, $\mathcal{M}_n$ embeds isometrically into ℓ_1^k for some $k \lesssim \log n$), yet $\mathcal{M}_n$ has a $O(1)$-Lipschitz quotient (see [24] for the relevant definition) that does not embed with distortion $O(1)$ into any normed space of dimension $n^{o(1)}$. To the best of our knowledge, it wasn't previously known that the metric dimension $\dim_D(\cdot)$ can become asymptotically larger (and even increase exponentially) under Lipschitz quotients, which is yet another major departure from the linear theory, in contrast to what one would normally predict in the context of the Ribe program.

Remark 56. Let $\mathsf{G} = (\{1, \ldots, n\}, E_\mathsf{G})$ be a Cayley graph of a finite group with $\lambda_2(\mathsf{G}) = 1 - \Omega(1)$. The metric space $(\{1, \ldots, n\}, d_\mathsf{G})$ embeds with bi-Lipschitz distortion $\operatorname{diam}(\mathsf{G})$ into ℓ_2^{n-1} by considering any bijection between $\{1, \ldots, n\}$ and the vertices of the n-simplex. There is therefore no a priori reason why it wouldn't be possible to embed $(\{1,\ldots,n\},d_{\mathsf{G}})$ with bi-Lipschitz distortion $O(1)$ into some normed space $(X,\|\cdot\|_X)$ whose bi-Lipschitz distortion from a Hilbert space is at least a sufficiently large constant multiple of $\mathrm{diam}(\mathsf{G})$. But this is not so if $\mathrm{diam}(\mathsf{G})$ is large. Indeed, for every $\mathsf{c}>\mathsf{c}_{\ell_2}(X)$ and $D>\mathsf{c}_X(\{1,\ldots,n\},d_{\mathsf{G}})$ by Theorem 12 the $\frac12$-snowflake of $(\{1,\ldots,n\},d_{\mathsf{G}})$ embeds into ℓ_2 with quadratic average distortion that is at most a universal constant multiple of $\sqrt{D\log(\mathsf{c}+1)}$. By contrasting this with the case $\omega=\frac12$ of (139), it follows that

$$\mathsf{c}_X(\{1,\ldots,n\},d_{\mathsf{G}})\gtrsim\frac{\mathrm{diam}(\mathsf{G})}{\log(\mathsf{c}_{\ell_2}(X)+1)}.$$

Thus, even if we allow $\mathsf{c}_{\ell_2}(X)$ to be as large as $\mathrm{diam}(\mathsf{G})^{O(1)}$, then any embedding of $(\{1,\ldots,n\},d_{\mathsf{G}})$ into X incurs distortion that is at least a positive universal constant multiple of $\mathrm{diam}(\mathsf{G})/\log\mathrm{diam}(\mathsf{G})$.

Substituting (142) into (157) gives the following noteworthy corollary. It shows that even though elements of $\mathsf{SL}_k(\mathbb{F}_\mathsf{q})$ have a representation using k^2 coordinates (over $\mathbb{F}_\mathsf{q}$, thus using $\mathsf{q}k^2$ bits), if one wishes to realize its geometry with bounded (average) distortion as a subset of the "commutative" geometry of a normed space, then the dimension of that space must be exponentially large.

Corollary 57. *Fix $\mathsf{q},k\in\mathbb{N}$ such that q is a power of a prime. For $D\geqslant1$ let $\dim_D(k,\mathsf{q})$ denote the smallest $d\in\mathbb{N}$ such that $\mathsf{SL}_k(\mathbb{F}_\mathsf{q})$ embeds into some d-dimensional normed space $X_{k,\mathsf{q}}$ with bi-Lipschitz distortion D. Then, for some universal constant $c>0$ we have*

$$\dim_D(k,\mathsf{q})\geqslant\mathsf{q}^{\frac{ck}{D\log k}}.$$

This holds even if we only require that the low-dimensional embedding has average distortion D.

The following conjecture asserts that (for fixed q) Corollary 57 is sharp. Given the lower bound that we obtained here, it remains to construct a $O(1)$-distortion embedding of $\mathsf{SL}_k(\mathbb{F}_\mathsf{q})$ into some low-dimensional normed space X. Here, "low-dimensional" means that the dimension of X grows exponentially in $k/\log k$ rather than exponentially in k^2 as in Fréchet's embedding. We suspect that, beyond its intrinsic interest, such a low-dimensional realization of $\mathsf{SL}_k(\mathbb{F}_\mathsf{q})$ will be useful elsewhere.

Conjecture 58 (dimension reduction for $\mathsf{SL}_k(\mathbb{F}_\mathsf{q})$). For every prime power q there exist $D=D(\mathsf{q})\geqslant1$ and $c=c(\mathsf{q}),C=C(\mathsf{q})>0$ such that for every integer $k\geqslant2$ we have

$$e^{\frac{ck}{\log k}}\leqslant\dim_D(k,\mathsf{q})\leqslant e^{\frac{Ck}{\log k}}.$$

7. Proof of Theorem 22

For a metric space $(\mathcal{M},d_\mathcal{M})$, a Banach space $(Y,\|\cdot\|_Y)$ and $\omega\in(0,1]$, following the notation of [89, 33, 80, 113] we consider a quantity $\mathsf{e}^\omega(\mathcal{M},Y)$, called the ω-Hölder extension modulus of the pair $(\mathcal{M},Y)$, which is defined as the infimum over those $L\in[1,\infty]$ such that for every subset $S\subset\mathcal{M}$ and every mapping $\phi:S\to Y$ which is ω-Hölder with constant 1, i.e., $\|f(x)-f(y)\|_Y\leqslant d_\mathcal{M}(x,y)^\omega$ for all $x,y\in\mathcal{M}$, there exists $\Phi:\mathcal{M}\to Y$ that extends ϕ, i.e., $\Phi(s)=\phi(s)$ for all $s\in S$, and Φ is ω-Hölder with constant L. When $\omega=1$ one uses the simpler notation $\mathsf{e}^1(\mathcal{M},T)=\mathsf{e}(\mathcal{M},Y)$. Note that $\mathsf{e}^\omega(\mathcal{M},Y)=\mathsf{e}(\mathcal{M}^\omega,Y)$, where henceforth $\mathcal{M}^\omega$ denotes the ω-snowflake of $(\mathcal{M},d_\mathcal{M})$. Thus, one could work throughout (both in the present context and elsewhere) with the more classical Lipschitz extension modulus $\mathsf{e}(\cdot,\cdot)$, but it is beneficial to use the above notation for ω-Hölder extension. Such extension moduli have been studied extensively in the literature; see e.g. [80, 32, 113] and the references therein for an indication of the large amount of work that has been done on this topic.

The following powerful extension theorem is a combination of known results. Its special case $p=q=2$ is a combination of Ball's deep work [19] on Lipschitz extension and our solution [112] in collaboration with Peres, Schramm and Sheffield of Ball's Markov type 2 problem [19]. Also, its special case when $p=1$ and Y is a Hilbert space is a theorem of Minty [104], which relies on

Kirszbraun's important theorem [75]. Its statement in full generality follows from the generalization of the above results that appears in our work with Mendel [98]. A special case of Theorem 59 was discussed in [108]; for ease of later use (below and elsewhere), it is worthwhile to formulate the full statement here and explain its quick derivation from results in the literature.

Theorem 59 (Ball's extension phenomenon for Hölder functions)**.** *Fix $p,q>0$ with $q\geqslant \max\{p,2\}$. Write $\omega=p/q$. Let $(\mathcal{M},d_{\mathcal{M}})$ be a metric space that has Markov type p and let $(Y,\|\cdot\|_Y)$ be a Banach space whose modulus of uniform convexity has power type q. Then, $\mathsf{e}^{\omega}(X,Y)\lesssim \mathbf{M}_p(\mathcal{M})^{\omega}\mathcal{K}_q(Y)$.*

Proof. As Y is uniformly convex, it is reflexive. Hence, $\mathsf{e}(\mathcal{N},Y)\lesssim \mathbf{M}_q(\mathcal{N})\mathcal{K}_q(Y)$ for any metric space $(\mathcal{N},d_{\mathcal{N}})$, by combining [99, Theorem 6.10] and [98, Theorem 1.11] (see the discussion in Section 1.5 of [98]). By definition, we have $\mathsf{e}^{\omega}(\mathcal{M},Y)=\mathsf{e}(\mathcal{M}^{\omega},Y)$ and $\mathbf{M}_q(\mathcal{M}^{\omega})\leqslant \mathbf{M}_p(\mathcal{M})^{\omega}$. Consequently,

$$\mathsf{e}^{\omega}(\mathcal{M},Y)\lesssim \mathbf{M}_q(\mathcal{M}^{\omega})\mathcal{K}_q(Y)\leqslant \mathbf{M}_p(\mathcal{M})^{\omega}\mathcal{K}_q(Y). \qquad \square$$

Note that due to Theorem 21, if we are in the setting of Theorem 59 and $(X,\|\cdot\|_X)$ is a Banach space whose modulus of uniform smoothness has power type p, then $\mathsf{e}^{\omega}(X,Y)\lesssim \mathcal{S}_p(X)^{\omega}\mathcal{K}_q(Y)$.

The proof of Lemma 60 below is a natural (a bit tedious) discretization/dominated convergence argument; we include it for the sake of completeness, but it could be skipped and left as a technical exercise. In what follows, it is convenient to use the (ad hoc) terminology that a measure μ on a set Ω is rational if it is finitely supported and $\mu(\{x\})\in\mathbb{Q}$ for every $x\in\Omega$.

Lemma 60 (compactness)**.** *Fix $p\geqslant 1$ and $D,\alpha>1$. Let $(\mathcal{M},d_{\mathcal{M}})$ be a separable infinite metric space and let $(Y,\|\cdot\|_Y)$ be a Banach space. Suppose that for any rational probability measure ρ on $\mathcal{M}$ the metric probability space $(\mathrm{supp}(\rho),d_{\mathcal{M}},\rho)$ embeds with p-average distortion D into Z. Then, $(\mathcal{M},d_{\mathcal{M}})$ embeds with p-average distortion αD into the ultrapower $Y^{\mathcal{U}}$ for any non-principal ultrafilter $\mathcal{U}$ on $\mathbb{N}$. Also, $(\mathcal{M},d_{\mathcal{M}})$ embeds with p-average distortion $\alpha\mathsf{e}(X,Y)D$ into Y.*

The proof of Theorem 22 is a direct application of Lemma 60 to Theorem 61 below, which is (the nontrivial direction of a) the duality result of [107, Theorem 1.3], while using Theorem 59 to justify the second assertion of Theorem 22. The term "duality" here indicates that the existence of the embedding that Theorem 61 asserts is proved in [107] by a separation (Hahn–Banach) argument.

Theorem 61. *Let $(\mathcal{M},d_{\mathcal{M}})$ be a metric space and $(Y,\|\cdot\|_Y)$ a Banach space. Suppose that there exist $p,K\geqslant 1$ such that $\gamma(\mathsf{A},d_{\mathcal{M}}^p)\leqslant K\gamma(\mathsf{A},\|\cdot\|_Y^p)$ for every $n\in\mathbb{N}$ and any symmetric stochastic matrix $\mathsf{A}\in\mathsf{M}_n(\mathbb{R})$. Then, for every $D>K$ and every rational probability measure ρ on $\mathcal{M}$, the (finite) metric measure space $(\mathrm{supp}(\rho),d_{\mathcal{M}},\rho)$ embeds with p-average distortion D into $\ell_p(Y)$.*

Proof of Lemma 60. For each $\delta>0$ fix an arbitrary δ-net $\{x_i^{\delta}\}_{i=1}^{\infty}$ of $(\mathcal{M},d_{\mathcal{M}})$, i.e., $d_{\mathcal{M}}(x_i^{\delta},x_j^{\delta})\geqslant\delta$ for all distinct $i,j\in\mathbb{N}$, and also $\bigcup_{i=1}^{\infty}B_{\mathcal{M}}(x_i^{\delta},\delta)=\mathcal{M}$. Define inductively $V_1^{\delta}=B_{\mathcal{M}}(x_1^{\delta},\delta)$ and $V_{j+1}^{\delta}=B_{\mathcal{M}}(x_{j+1}^{\delta},\delta)\smallsetminus\bigcup_{i=1}^{j}B_{\mathcal{M}}(x_j^{\delta},\delta)$ for $j\in\mathbb{N}$, namely $\{V_j^{\delta}\}_{j=1}^{\infty}$ is the disjoint Voronoi tessellation of $\mathcal{M}$ that is induced by the (ordered) δ-net $\{x_i^{\delta}\}_{i=1}^{\infty}$.

Fix from now on a Borel probability measure μ on $\mathcal{M}$. Let $(W,\|\cdot\|_W)$ be any Banach space (we will eventually take W to be either $Y^{\mathcal{U}}$ or Y). Suppose that there is some $\lambda>0$ such that for every $\delta>0$ and $n\in\mathbb{N}$ there exists a λ-Lipschitz mapping $f_n^{\delta}:\mathcal{M}\to W$ that satisfies

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\|f_n^{\delta}(x_i^{\delta})-f_n^{\delta}(x_j^{\delta})\|_W^p\mu(V_i^{\delta})\mu(V_j^{\delta})\geqslant\sum_{i=1}^{n}\sum_{j=1}^{n}d_{\mathcal{M}}(x_i^{\delta},x_j^{\delta})^p\mu(V_i^{\delta})\mu(V_j^{\delta}). \tag{158}$$

We will next show that this assumption formally implies that for any $\Lambda>\lambda$, the metric probability space $(\mathcal{M},d_{\mathcal{M}},\mu)$ embeds with p-average distortion Λ into $(W,\|\cdot\|_W)$.

As justified in the beginning of Section 5, we may assume that $\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) < \infty$. If we could prove that

$$\sup_{\substack{\delta>0\\ n\in\mathbb{N}}} \iint_{\mathcal{M}\times\mathcal{M}} \|f_n^\delta(x) - f_n^\delta(y)\|_W^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant \iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y), \tag{159}$$

then for some $\delta > 0$ and $n \in \mathbb{N}$ the normalized mapping $g_n^\delta \stackrel{\mathrm{def}}{=} \frac{\Lambda}{\lambda} f_n^\delta : \mathcal{M} \to W$ would be Λ-Lipschitz and satisfy $\iint_{\mathcal{M}\times\mathcal{M}} \|g_n^\delta(x) - g_n^\delta(y)\|_W^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant \iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)$, as required.

To prove (159), note that because $\{x_s^\delta\}_{s=1}^\infty$ is δ-dense in $\mathcal{M}$ we have

$$\sup_{\substack{i,j\in\mathbb{N}\\ (x,y)\in V_i^\delta\times V_j^\delta}} |d_{\mathcal{M}}(x_i^\delta, x_j^\delta) - d_{\mathcal{M}}(x,y)| \leqslant \sup_{\substack{i,j\in\mathbb{N}\\ (x,y)\in V_i^\delta\times V_j^\delta}} \big(d_{\mathcal{M}}(x_i^\delta, x) + d_{\mathcal{M}}(x_j^\delta, y)\big) \leqslant 2\delta. \tag{160}$$

Hence, the $\mathbb{R}$-valued function on $\mathcal{M}\times\mathcal{M}$ that is equal to $d_{\mathcal{M}}(x_i^\delta, x_j^\delta)^p$ on $V_i\times V_j$ for each $i,j\in\mathbb{N}$ tends point-wise to $d_{\mathcal{M}}^p : \mathcal{M}\times\mathcal{M} \to \mathbb{R}$ as $\delta \to 0$, and it is bounded from above by the $(\mu\times\mu)$-integrable function $(x,y) \mapsto (d_{\mathcal{M}}(x,y) + 2\delta)^p$. By the dominated convergence theorem we therefore have

$$\begin{aligned}
\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x,y)^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) &= \lim_{\delta\to 0} \sum_{i=1}^\infty \sum_{j=1}^\infty d_{\mathcal{M}}(x_i^\delta, x_j^\delta)^p \mu(V_i^\delta)\mu(V_j^\delta) \\
&\stackrel{(158)}{\leqslant} \limsup_{\delta\to 0} \limsup_{n\to\infty} \sum_{i=1}^n \sum_{j=1}^n \|f_n^\delta(x_i^\delta) - f_n^\delta(x_j^\delta)\|_W^p \mu(V_i^\delta)\mu(V_j^\delta) \\
&\leqslant \sup_{\substack{\delta>0\\ n\in\mathbb{N}}} \sum_{i=1}^\infty \sum_{j=1}^\infty \|f_n^\delta(x_i^\delta) - f_n^\delta(x_j^\delta)\|_W^p \mu(V_i^\delta)\mu(V_j^\delta).
\end{aligned}$$

The desired statement (159) would therefore follow if we could show that for every fixed $n \in \mathbb{N}$,

$$\lim_{\delta\to 0}\bigg(\iint_{\mathcal{M}\times\mathcal{M}} \|f_n^\delta(x) - f_n^\delta(y)\|_W^p \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) - \sum_{i=1}^\infty \sum_{j=1}^\infty \|f_n^\delta(x_i^\delta) - f_n^\delta(x_j^\delta)\|_W^p \mu(V_i^\delta)\mu(V_j^\delta)\bigg) = 0. \tag{161}$$

To justify (161), for $n\in\mathbb{N}$ and $\delta > 0$ consider the function $h_n^\delta : \mathcal{M}\times\mathcal{M} \to \mathbb{R}$ that is defined by

$$\forall\, i,j \in \mathbb{N},\ \forall (x,y) \in V_i^\delta \times V_j^\delta, \qquad h_n^\delta(x,y) \stackrel{\mathrm{def}}{=} \|f_n^\delta(x) - f_n^\delta(y)\|_W^p - \|f_n^\delta(x_i^\delta) - f_n^\delta(x_j^\delta)\|_W^p.$$

Under this notation, the assertion of (161) is the same as $\lim_{\delta\to 0} \iint_{\mathcal{M}\times\mathcal{M}} h_n^\delta(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = 0$. Since f_n^δ is assumed to be λ-Lipschitz, if $(x,y) \in V_i^\delta \times V_j^\delta$ for some $i,j\in\mathbb{N}$, then we have

$$\begin{aligned}
|h_n^\delta(x,y)| &\leqslant \|f_n^\delta(x) - f_n^\delta(y)\|_W^p + \|f_n^\delta(x_i^\delta) - f_n^\delta(x_j^\delta)\|_W^p \\
&\leqslant \lambda^p\big(d_{\mathcal{M}}(x,y)^p + d_{\mathcal{M}}(x_i^\delta, x_j^\delta)^p\big) \stackrel{(160)}{\leqslant} \lambda^p\big(d_{\mathcal{M}}(x,y)^p + (d_{\mathcal{M}}(x,y) + 2\delta)^p\big).
\end{aligned}$$

By the dominated convergence theorem, it therefore suffices to show that $\lim_{\delta\to 0} h_n^\delta = 0$ point-wise. This is so since if $\delta \leqslant 1$ and $(x,y) \in V_i^\delta \times V_j^\delta$ for $i,j\in\mathbb{N}$, then $\|f_n^\delta(x) - f_n^\delta(y)\|_W \leqslant \lambda d_{\mathcal{M}}(x,y)$ and

$$\big|\|f_n^\delta(x) - f_n^\delta(y)\|_W - \|f_n^\delta(x_i^\delta) - f_n^\delta(x_j^\delta)\|_W\big| \leqslant \|f_n^\delta(x) - f_n^\delta(x_i^\delta)\|_W + \|f_n^\delta(y) - f_n^\delta(x_j^\delta)\|_W \leqslant 2\lambda\delta \leqslant 2\lambda.$$

Hence, both of the numbers $\|f_n^\delta(x) - f_n^\delta(y)\|_W$ and $\|f_n^\delta(x_i^\delta) - f_n^\delta(x_j^\delta)\|_W$ belong to the bounded interval $[0, \lambda d_{\mathcal{M}}(x,y) + 2\lambda]$ and are within $2\lambda\delta$ of each other. By the uniform continuity of $t \mapsto t^p$ on $[0, \lambda d_{\mathcal{M}}(x,y) + 2\lambda]$, it follows that $\lim_{\delta\to 0}(\|f_n^\delta(x) - f_n^\delta(y)\|_W^p - \|f_n^\delta(x_i^\delta) - f_n^\delta(x_j^\delta)\|_W^p) = 0$, as required.

By the above considerations, it remains to establish, for fixed $\delta > 0$ and $n \in \mathbb{N}$, the existence of f_n^δ when $W = Y^{\mathcal{U}}$ or $W = Y$, and λ is less than αD or $\mathsf{e}(X,Y)\alpha D$, respectively. The case $W = Y$ is a direct consequence of the definition of $\mathsf{e}(X,Y)$ and the assumption of Lemma 60. Indeed, recalling that $\alpha > 1$, for each $i \in \{1,\ldots,n\}$ choose $\rho_i \in \mathbb{Q}$ satisfying $\mu(V_i^\delta) \leqslant \rho_i \leqslant \alpha^{\frac{p}{4}}\mu(V_i^\delta)$. By assumption, there is a D-Lipschitz mapping $\phi : \{x_i^\delta : \ i \in \{1,\ldots,n\} \ \wedge \ \rho_i > 0\} \to Y$ that satisfies $\sum_{i=1}^n\sum_{j=1}^n \|\phi(x_i^\delta)-\phi(x_j^\delta)\|_Y^p\rho_i\rho_j \geqslant \sum_{i=1}^n\sum_{j=1}^n d_{\mathcal{M}}(x_i^\delta,x_j^\delta)^p\rho_i\rho_j$. Therefore,

$$\alpha^{\frac{p}{2}}\sum_{i=1}^n\sum_{j=1}^n \|\phi(x_i^\delta)-\phi(x_j^\delta)\|_Y^p\mu(V_i^\delta)\mu(V_j^\delta) \geqslant \sum_{i=1}^n\sum_{j=1}^n d_{\mathcal{M}}(x_i^\delta,x_j^\delta)^p\mu(V_i^\delta)\mu(V_j^\delta). \tag{162}$$

Extend ϕ to a function $\Phi : \mathcal{M} \to Y$ which is $\mathsf{e}(X,Y)\sqrt[4]{\alpha}D$-Lipschitz. Then, $f_n^\delta = \sqrt{\alpha}\Phi$ has Lipschitz constant less than $\mathsf{e}(X,Y)\alpha D$ and, by virtue of (162), it satisfies the desired estimate (158).

For the remaining case, namely when $W = Y^{\mathcal{U}}$ and $\lambda < \alpha D$, fix $n \in \mathbb{N}$ and $\{y_j\}_{j=1}^\infty \subset \mathcal{M} \smallsetminus \{x_i^\delta\}_{i=1}^n$ such that $\{x_i^\delta\}_{i=1}^n \cup \{y_j\}_{j=1}^\infty$ is dense in $\mathcal{M}$. Fix also $k \in \mathbb{N}$ and $\eta \in (0,1)\cap\mathbb{Q}$, and define a measure ν on $\mathcal{F} = \{x_i^\delta\}_{i=1}^n \cup \{y_j\}_{j=1}^k$ by setting $\nu(x_i^\delta) \in \mathbb{Q}$ to be any rational number satisfying $(1-\eta)\mu(V_i^\delta) \leqslant \nu(x_i^\delta) < \mu(V_i^\delta)$ if $\mu(V_i^\delta) > 0$, and ν assigns mass η to all the other points in $\mathcal{F}$. By assumption, there is a D-Lipschitz mapping $\psi : \mathcal{F} \to Y$ such that

$$\sum_{u\in\mathcal{F}}\sum_{v\in\mathcal{F}} \|\psi(u)-\psi(v)\|_Y^p\nu(u)\nu(v) \geqslant \sum_{u\in\mathcal{F}}\sum_{v\in\mathcal{F}} d_{\mathcal{M}}(u,v)^p\nu(u)\nu(v).$$

Then, since ψ is D-Lipschitz and $\nu(w) = \eta$ for all $w \in \mathcal{F} \smallsetminus \{x_i^\delta\}_{i=1}^n = \{y_j\}_{j=1}^k$, we have

$$\sum_{i=1}^n\sum_{j=1}^n \|\psi(x_i^\delta)-\psi(x_j^\delta)\|_Y^p\mu(V_i^\delta)\mu(V_j^\delta)$$
$$\geqslant (1-\eta)^2\sum_{i=1}^n\sum_{j=1}^n d_{\mathcal{M}}(x_i^\delta,x_j^\delta)^p\mu(V_i^\delta)\mu(V_j^\delta) - 2\eta D^p\sum_{u\in\mathcal{F}}\sum_{j=1}^k d_{\mathcal{M}}(u,y_j)^p.$$

Since $\alpha > 1$, by choosing small enough η it follows from this that for each $k \in \mathbb{N}$ there exists a D-Lipschitz mapping $\psi_k : \{x_i^\delta\}_{i=1}^n \cup \{y_j\}_{j=1}^k \to Y$ that satisfies

$$\sum_{i=1}^n\sum_{j=1}^n \|\psi_k(x_i^\delta)-\psi_k(x_j^\delta)\|_Y^p\mu(V_i^\delta)\mu(V_j^\delta) \geqslant \alpha^{-\frac{p}{2}}\sum_{i=1}^n\sum_{j=1}^n d_{\mathcal{M}}(x_i^\delta,x_j^\delta)^p\mu(V_i^\delta)\mu(V_j^\delta). \tag{163}$$

We extend ψ_k to all of $\mathcal{M}$ by setting it to be identically equal to $\psi_k(x_1^\delta)$ on $\mathcal{M} \smallsetminus (\{x_i^\delta\}_{i=1}^n \cup \{y_j\}_{j=1}^k)$.

Since ψ_k is D-Lipschitz on $\{x_i^\delta\}_{i=1}^n \cup \{y_j\}_{j=1}^k$, for every $z \in \mathcal{M}$ the sequence $\{\psi_k(z)-\psi_k(x_1^\delta)\}_{k=1}^\infty$ is bounded. We can therefore define $f_n^\delta : \mathcal{M} \to Y^{\mathcal{U}}$ by setting $f_n^\delta(z) = \sqrt{\alpha}(\psi_k(z)-\psi_k(x_1^\delta))_{k=1}^\infty/\mathcal{U}$ for all $z \in \mathcal{M}$. It follows directly from the definitions that f_n^δ has Lipschitz constant $\sqrt{\alpha}D < \alpha D$ on the dense subset $\{x_i^\delta\}_{i=1}^n \cup \{y_j\}_{j=1}^\infty$, so its Lipschitz constant is less than αD on all of $\mathcal{M}$. Also, the desired bound (158) follows by passing to the ultralimit of (163) as $k \to \infty$, since the number of pairwise distances that appear in the left hand side of (163) is independent of k. $\square$

References

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

[2] F. Albiac and F. Baudier. Embeddability of snowflaked metrics with applications to the nonlinear geometry of the spaces L_p and ℓ_p for $0 < p < \infty$. *J. Geom. Anal.*, 25(1):1–24, 2015.

[3] N. Alon. The diameter of $\mathsf{SL}_n(\mathbb{F}_q)$. Preprint, 2019.

[4] N. Alon, P. Frankl, and V. Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science*, pages 277–280. IEEE Computer Society, 1985.

[5] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures Algorithms*, 5(2):271–284, 1994.

[6] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.

[7] A. Andoni, P. Indyk, and I. Razenshteyn. Approximate nearest neighbor search in high dimensions. To appear in *Proceedings of the 2018 International Congress of Mathematicians*, preprint available at `https://arxiv.org/abs/1806.09823`, 2018.

[8] A. Andoni, A. Naor, and O. Neiman. Snowflake universality of Wasserstein spaces. *Ann. Sci. Éc. Norm. Supér. (4)*, 51(3):657–700, 2018.

[9] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Data-dependent hashing via nonlinear spectral gaps. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 787–800, 2018.

[10] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Hölder homeomorphisms and approximate nearest neighbors. In *59th IEEE Annual Symposium on Foundations of Computer Science*, pages 159–169, 2018.

[11] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Complex interpolation, Hölder homeomorphisms, and algorithmic applications. Forthcoming manuscript, 2019.

[12] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Spectral partitioning of metric spaces. Forthcoming manuscript, 2019.

[13] A. Andoni, H. Nguyen, A. Nikolov, I. Razenshteyn, and E. Waingarten. Approximate near neighbors for general symmetric norms. Extended abstract in STOC'17, available at `https://arxiv.org/abs/1611.06222`, 2016.

[14] D. Andrén, L. Hellström, and K. Markström. On the complexity of matrix reduction over finite fields. *Adv. in Appl. Math.*, 39(4):428–452, 2007.

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

[16] N. Aronszajn. Differentiability of Lipschitzian mappings between Banach spaces. *Studia Math.*, 57(2):147–190, 1976.

[17] T. Austin, A. Naor, and A. Valette. The Euclidean distortion of the lamplighter group. *Discrete Comput. Geom.*, 44(1):55–74, 2010.

[18] U. Bader, A. Furman, T. Gelander, and N. Monod. Property (T) and rigidity for actions on Banach spaces. *Acta Math.*, 198(1):57–105, 2007.

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

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

[21] K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.

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

[23] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005.

[24] S. Bates, W. B. Johnson, J. Lindenstrauss, D. Preiss, and G. Schechtman. Affine approximation of Lipschitz functions and nonlinear quotients. *Geom. Funct. Anal.*, 9(6):1092–1127, 1999.

[25] F. P. Baudier. Quantitative nonlinear embeddings into Lebesgue sequence spaces. *J. Topol. Anal.*, 8(1):117–150, 2016.

[26] Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

[27] J. Bergh and J. Löfström. *Interpolation spaces. An introduction*. Springer-Verlag, Berlin-New York, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

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

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

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

[31] J. Bretagnolle, D. Dacunha-Castelle, and J.-L. Krivine. Fonctions de type positif sur les espaces L^p. *C. R. Acad. Sci. Paris*, 261:2153–2156, 1965.

[32] 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.

[33] Y. Brudnyi and P. Shvartsman. Stability of the Lipschitz extension property under metric transforms. *Geom. Funct. Anal.*, 12(1):73–79, 2002.

[34] A.-P. Calderón. Intermediate spaces and interpolation, the complex method. *Studia Math.*, 24:113–190, 1964.

[35] F. Chaatit. On uniform homeomorphisms of the unit spheres of certain Banach lattices. *Pacific J. Math.*, 168(1):11–31, 1995.

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

[37] Q. Cheng. Sphere equivalence, property H, and Banach expanders. *Studia Math.*, 233(1):67–83, 2016.

[38] J. P. R. Christensen. Measure theoretic zero sets in infinite dimensional spaces and applications to differentiability of Lipschitz mappings. *Publ. Dép. Math. (Lyon)*, 10(2):29–39, 1973. Actes du Deuxième Colloque d'Analyse Fonctionnelle de Bordeaux (Univ. Bordeaux, 1973), I, pp. 29–39.

[39] F. R. K. Chung. Laplacians of graphs and Cheeger's inequalities. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 157–172. János Bolyai Math. Soc., Budapest, 1996.

[40] J. A. Clarkson. Uniformly convex spaces. *Trans. Amer. Math. Soc.*, 40(3):396–414, 1936.

[41] M. Cwikel. Complex interpolation spaces, a discrete definition and reiteration. *Indiana Univ. Math. J.*, 27(6):1005–1009, 1978.

[42] M. Cwikel and S. Reisner. Interpolation of uniformly convex Banach spaces. *Proc. Amer. Math. Soc.*, 84(4):555–559, 1982.

[43] M. Daher. Homéomorphismes uniformes entre les sphères unité des espaces d'interpolation. *Canad. Math. Bull.*, 38(3):286–294, 1995.

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

[45] M. M. Day. Uniform convexity in factor and conjugate spaces. *Ann. of Math. (2)*, 45:375–385, 1944.

[46] T. de Laat and M. de la Salle. Banach space actions and L^2-spectral gap. Preprint, available at `https://arxiv.org/abs/1705.03296`, 2017.

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

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

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

[50] A. Eskenazis and A. Naor. On coarse and uniform embeddings into L_p. Forthcoming manuscript, 2018.

[51] T. Figiel. On the moduli of convexity and smoothness. *Studia Math.*, 56(2):121–155, 1976.

[52] T. Figiel and G. Pisier. Séries aléatoires dans les espaces uniformément convexes ou uniformément lisses. *C. R. Acad. Sci. Paris Sér. A*, 279:611–614, 1974.

[53] M. Fréchet. Sur quelques points du calcul fonctionnel. *Rend. Circ. Mat. Palermo*, 22:1–74, 1906.

[54] I. Gelfand. Abstrakte Funktionen und lineare Operatoren. *Matematicheskiĭ Sbornik*, 4(2):235–286, 1938.

[55] R. I. Grigorchuk and P. W. Nowak. Diameters, distortion, and eigenvalues. *European J. Combin.*, 33(7):1574–1587, 2012.

[56] M. Gromov. Spaces and questions. *Geom. Funct. Anal.*, (Special Volume, Part I):118–161, 2000. GAFA 2000 (Tel Aviv, 1999).

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

[58] J. L. Gross. Every connected regular graph of even degree is a Schreier coset graph. *J. Combinatorial Theory Ser. B*, 22(3):227–232, 1977.

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

[60] A. Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[61] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001.

[62] S. Heinrich. Ultraproducts in Banach space theory. *J. Reine Angew. Math.*, 313:72–104, 1980.

[63] M. Herman and J. Pakianathan. On the distribution of distances in homogeneous compact metric spaces. *Topology Appl.*, 193:97–99, 2015.

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

[65] H. Izeki and S. Nayatani. Combinatorial harmonic maps and discrete-group actions on Hadamard spaces. *Geom. Dedicata*, 114:147–188, 2005.

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

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

[68] 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.
[69] W. B. Johnson and J. Lindenstrauss. Basic concepts in the geometry of Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 1–84. North-Holland, Amsterdam, 2001.
[70] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. On Lipschitz embedding of finite metric spaces in low-dimensional normed spaces. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 177–184. Springer, Berlin, 1987.
[71] P.-N. Jolissaint and A. Valette. L^p-distortion and p-spectral gap of finite graphs. *Bull. Lond. Math. Soc.*, 46(2):329–341, 2014.
[72] N. J. Kalton. The nonlinear geometry of Banach spaces. *Rev. Mat. Complut.*, 21(1):7–60, 2008.
[73] M. Kassabov. Kazhdan constants for $\mathrm{SL}_n(\mathbb{Z})$. *Internat. J. Algebra Comput.*, 15(5-6):971–995, 2005.
[74] S. Khot and A. Naor. Nonembeddability theorems via Fourier analysis. *Math. Ann.*, 334(4):821–852, 2006.
[75] M. D. Kirszbraun. Über die zusammenziehenden und Lipschitzchen Transformationen. *Fundam. Math.*, 22:77–108, 1934.
[76] T. Kondo. CAT(0) spaces and expanders. *Math. Z.*, 271(1-2):343–355, 2012.
[77] T. J. Laakso. Plane with A_∞-weighted metric not bi-Lipschitz embeddable to $\mathbb{R}^N$. *Bull. London Math. Soc.*, 34(6):667–676, 2002.
[78] V. Lafforgue. Un renforcement de la propriété (T). *Duke Math. J.*, 143(3):559–602, 2008.
[79] V. Lafforgue. Propriété (T) renforcée banachique et transformation de Fourier rapide. *J. Topol. Anal.*, 1(3):191–206, 2009.
[80] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.
[81] J. Lindenstrauss. On the modulus of smoothness and divergent series in Banach spaces. *Michigan Math. J.*, 10:241–252, 1963.
[82] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. I.* Springer-Verlag, Berlin-New York, 1977. Sequence spaces, Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 92.
[83] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin-New York, 1979. Function spaces.
[84] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
[85] N. Linial and A. Magen. Least-distortion Euclidean embeddings of graphs: products of cycles and expanders. *J. Combin. Theory Ser. B*, 79(2):157–171, 2000.
[86] N. Linial, A. Magen, and A. Naor. Girth and Euclidean distortion. *Geom. Funct. Anal.*, 12(2):380–394, 2002.
[87] J.-L. Lions. Une construction d'espaces d'interpolation. *C. R. Acad. Sci. Paris*, 251:1853–1855, 1960.
[88] P. Mankiewicz. On Lipschitz mappings between Fréchet spaces. *Studia Math.*, 41:225–241, 1972.
[89] J. Matoušek. Extension of Lipschitz mappings on metric trees. *Comment. Math. Univ. Carolin.*, 31(1):99–104, 1990.
[90] J. Matoušek. Note on bi-Lipschitz embeddings into normed spaces. *Comment. Math. Univ. Carolin.*, 33(1):51–55, 1992.
[91] J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.
[92] J. Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.
[93] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.
[94] S. Mazur. Une remarque sur l'homéomorphie des champs fonctionels. *Studia Math.*, 1:83–85, 1929.
[95] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Adv. Math.*, 189(2):451–494, 2004.
[96] M. Mendel and A. Naor. Metric cotype. *Ann. of Math. (2)*, 168(1):247–298, 2008.
[97] M. Mendel and A. Naor. Markov convexity and local rigidity of distorted metrics. *J. Eur. Math. Soc. (JEMS)*, 15(1):287–337, 2013.
[98] M. Mendel and A. Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. *Anal. Geom. Metr. Spaces*, 1:163–199, 2013.
[99] M. Mendel and A. Naor. Nonlinear spectral calculus and super-expanders. *Publ. Math. Inst. Hautes Études Sci.*, 119:1–95, 2014.
[100] M. Mendel and A. Naor. Expanders with respect to Hadamard spaces and random graphs. *Duke Math. J.*, 164(8):1471–1548, 2015.

[101] D. Milman. On some criteria for the regularity of spaces of the type (B). *C. R. (Dokl.) Acad. Sci. URSS, n. Ser.*, 20:243–246, 1938.
[102] J. Milnor. On the Betti numbers of real varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.
[103] M. Mimura. Sphere equivalence, Banach expanders, and extrapolation. *Int. Math. Res. Not. IMRN*, (12):4372–4391, 2015.
[104] G. J. Minty. On the extension of Lipschitz, Lipschitz-Hölder continuous, and monotone functions. *Bull. Amer. Math. Soc.*, 76:334–339, 1970.
[105] A. Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.
[106] A. Naor. On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon-Roichman graphs. *Combin. Probab. Comput.*, 21(4):623–634, 2012.
[107] A. Naor. Comparison of metric spectral gaps. *Anal. Geom. Metr. Spaces*, 2:1–52, 2014.
[108] A. Naor. Uniform nonextendability from nets. *C. R. Math. Acad. Sci. Paris*, 353(11):991–994, 2015.
[109] A. Naor. Discrete Riesz transforms and sharp metric X_p inequalities. *Ann. of Math. (2)*, 184(3):991–1016, 2016.
[110] A. Naor. A spectral gap precludes low-dimensional embeddings. In *33rd International Symposium on Computational Geometry*, volume 77 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 50, 16. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.
[111] 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.
[112] A. Naor, Y. Peres, O. Schramm, and S. Sheffield. Markov chains in smooth Banach spaces and Gromov-hyperbolic metric spaces. *Duke Math. J.*, 134(1):165–197, 2006.
[113] A. Naor and Y. Rabani. On Lipschitz extension from finite subsets. *Israel J. Math.*, 219(1):115–161, 2017.
[114] A. Naor, Y. Rabani, and A. Sinclair. Quasisymmetric embeddings, the observable diameter, and expansion properties of graphs. *J. Funct. Anal.*, 227(2):273–303, 2005.
[115] A. Naor and G. Schechtman. Metric X_p inequalities. *Forum Math. Pi*, 4:e3, 81 pp., 2016.
[116] A. Naor and L. Silberman. Poincaré inequalities, embeddings, and wild groups. *Compos. Math.*, 147(5):1546–1572, 2011.
[117] I. Newman and Y. Rabinovich. Hard metrics from Cayley graphs of abelian groups. *Theory Comput.*, 5:125–134, 2009.
[118] E. Odell and T. Schlumprecht. The distortion problem. *Acta Math.*, 173(2):259–281, 1994.
[119] M. I. Ostrovskii. *Metric embeddings. Bilipschitz and coarse embeddings into Banach spaces*, volume 49 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2013.
[120] N. Ozawa. A note on non-amenability of $\mathcal{B}(l_p)$ for $p = 1, 2$. *Internat. J. Math.*, 15(6):557–565, 2004.
[121] B. J. Pettis. A proof that every uniformly convex space is reflexive. *Duke Math. J.*, 5(2):249–253, 1939.
[122] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.
[123] G. Pisier. Some applications of the complex interpolation method to Banach lattices. *J. Analyse Math.*, 35:264–281, 1979.
[124] G. Pisier. Complex interpolation between Hilbert, Banach and operator spaces. *Mem. Amer. Math. Soc.*, 208(978):vi+78, 2010.
[125] Y. Rabinovich. On average distortion of embedding metrics into the line. *Discrete Comput. Geom.*, 39(4):720–733, 2008.
[126] Y. Raynaud. On ultrapowers of non commutative L_p spaces. *J. Operator Theory*, 48(1):41–68, 2002.
[127] M. Ribe. On uniformly homeomorphic normed spaces. *Ark. Mat.*, 14(2):237–244, 1976.
[128] E. Ricard. Hölder estimates for the noncommutative Mazur maps. *Arch. Math. (Basel)*, 104(1):37–45, 2015.
[129] M. Riesz. Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires. *Acta Math.*, 49(3-4):465–497, 1927.
[130] T. R. Riley. Navigating in the Cayley graphs of $\mathrm{SL}_N(\mathbb{Z})$ and $\mathrm{SL}_N(\mathbb{F}_p)$. *Geom. Dedicata*, 113:215–229, 2005.
[131] I. J. Schoenberg. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, 44(3):522–536, 1938.
[132] E. M. Stein. Interpolation of linear operators. *Trans. Amer. Math. Soc.*, 83:482–492, 1956.
[133] R. Thom. Sur l'homologie des variétés algébriques réelles. In *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)*, pages 255–265. Princeton Univ. Press, Princeton, N.J., 1965.
[134] G. O. Thorin. Convexity theorems generalizing those of M. Riesz and Hadamard with some applications. *Comm. Sem. Math. Univ. Lund [Medd. Lunds Univ. Mat. Sem.]*, 9:1–58, 1948.
[135] J. H. Wells and L. R. Williams. *Embeddings and extensions in analysis*. Springer-Verlag, New York, 1975. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 84.

Mathematics Department, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA
E-mail address: naor@math.princeton.edu