COMPARISON OF METRIC SPECTRAL GAPS

ASSAF NAOR

Abstract. Let $A = (a_{ij}) \in M_n(\mathbb{R})$ be an n by n symmetric stochastic matrix. For $p \in [1,\infty)$ and a metric space (X, d_X), let $\gamma(A, d_X^p)$ be the infimum over those $\gamma \in (0,\infty]$ for which every $x_1,\ldots,x_n \in X$ satisfy

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p \leqslant \frac{\gamma}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^p.$$

Thus $\gamma(A, d_X^p)$ measures the magnitude of the *nonlinear spectral gap* of the matrix A with respect to the kernel $d_X^p : X\times X \to [0,\infty)$. We study pairs of metric spaces (X, d_X) and (Y, d_Y) for which there exists $\Psi : (0,\infty) \to (0,\infty)$ such that $\gamma(A, d_X^p) \leqslant \Psi(\gamma(A, d_Y^p))$ for every symmetric stochastic $A \in M_n(\mathbb{R})$ with $\gamma(A, d_Y^p) < \infty$. When Ψ is linear a complete geometric characterization is obtained.

Our estimates on nonlinear spectral gaps yield new embeddability results as well as new nonembeddability results. For example, it is shown that if $n \in \mathbb{N}$ and $p \in (2,\infty)$ then for every $f_1,\ldots,f_n \in L_p$ there exist $x_1,\ldots,x_n \in L_2$ such that

$$\forall\, i,j \in \{1,\ldots,n\}, \quad \|x_i - x_j\|_2 \lesssim p\|f_i - f_j\|_p, \tag{1}$$

and

$$\sum_{i=1}^n\sum_{j=1}^n \|x_i - x_j\|_2^2 = \sum_{i=1}^n\sum_{j=1}^n \|f_i - f_j\|_p^2.$$

This statement is impossible for $p \in [1,2)$, and the asymptotic dependence on p in (1) is sharp. We also obtain the best known lower bound on the L_p distortion of Ramanujan graphs, improving over the work of Matoušek. Links to Bourgain–Milman–Wolfson type and a conjectural nonlinear Maurey–Pisier theorem are studied.

Contents

2010 *Mathematics Subject Classification.* 51F99,05C12,05C50,46B85.

Key words and phrases. Metric embeddings, nonlinear spectral gaps, expanders, nonlinear type.

Supported by NSF grant CCF-0832795, BSF grant 2010021, the Packard Foundation and the Simons Foundation.

1. Introduction

The decreasing rearrangement of the eigenvalues of an n by n symmetric stochastic matrix $A = (a_{ij}) \in M_n(\mathbb{R})$ is denoted below by

$$1 = \lambda_1(A) \geqslant \lambda_2(A) \geqslant \ldots \geqslant \lambda_n(A).$$

We also set

$$\lambda(A) \stackrel{\mathrm{def}}{=} \max_{i \in \{2,\ldots,n\}} |\lambda_i(A)| = \max\left\{\lambda_2(A), -\lambda_n(A)\right\}. \tag{2}$$

For $p \in [1,\infty)$ and a metric space (X, d_X), we denote by $\gamma(A, d_X^p)$ the infimum over those $\gamma \in (0,\infty]$ for which every $x_1, \ldots, x_n \in X$ satisfy

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i,x_j)^p \leqslant \frac{\gamma}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} d_X(x_i,x_j)^p. \tag{3}$$

Thus for $p = 2$ and $X = (\mathbb{R}, d_{\mathbb{R}})$, where we denote $d_{\mathbb{R}}(x,y) \stackrel{\mathrm{def}}{=} |x-y|$ for every $x, y \in \mathbb{R}$, we have

$$\gamma(A, d_{\mathbb{R}}^2) = \frac{1}{1-\lambda_2(A)}. \tag{4}$$

For this reason one thinks of $\gamma(A, d_X^p)$ as measuring the magnitude of the *nonlinear spectral gap* of the matrix A with respect to the kernel $d_X^p : X \times X \to [0,\infty)$. See [MN12] for more information on the topic of nonlinear spectral gaps.

Suppose that (X,d_X) is a metric space with $|X|\geqslant 2$ and fix distinct points $a,b\in X$. Fix also $p\in[1,\infty)$, $n\in\mathbb{N}$, and an n by n symmetric stochastic matrix $A=(a_{ij})$. By considering $x_1,\ldots,x_n\in\{a,b\}$ in (3), we see that

$$\gamma(A,d_X^p)\geqslant \max_{\emptyset\neq S\subsetneq\{1,\ldots,n\}} \frac{|S|(n-|S|)}{n\sum_{(i,j)\in S\times(\{1,\ldots,n\}\smallsetminus S)} a_{ij}}.$$

It therefore follows from Cheeger's inequality [Che70] (in our context, see [JS88] and [LS88]) that

$$\gamma(A,d_X^p)\gtrsim \frac{1}{\sqrt{1-\lambda_2(A)}}. \tag{5}$$

Consequently, any finite upper bound on $\gamma(A,d_X^p)$ immediately implies a spectral gap estimate for the matrix A. The ensuing discussion always tacitly assumes that metric spaces contain at least two points.

In (5), and in what follows, the notations $U\lesssim V$ and $V\gtrsim U$ mean that $U\leqslant CV$ for some universal constant $C\in(0,\infty)$. If we need to allow C to depend on parameters, we indicate this by subscripts, thus e.g. $U\lesssim_\beta V$ means that $U\leqslant C(\beta)V$ for some $C(\beta)\in(0,\infty)$ which is allowed to depend only on the parameter β. The notation $U\asymp V$ stands for $(U\lesssim V)\wedge(V\lesssim U)$, and correspondingly the notation $U\asymp_\beta V$ stands for $(U\lesssim_\beta V)\wedge(V\lesssim_\beta U)$.

A simple application of the triangle inequality (see [MN12, Lem. 2.1]) shows that $\gamma(A,d_X^p)$ is finite if and only if $\lambda_2(A)<1$ (equivalently, the matrix A is ergodic, or the graph on $\{1,\ldots,n\}$ whose edges are the pais $\{i,j\}$ for which $a_{ij}>0$ is connected).

It is often quite difficult to obtain good estimates on nonlinear spectral gaps. This difficulty is exemplified by several problems in metric geometry that can be cast as estimates on nonlinear spectral gaps: see [Laf08, Laf09, MN12, MN13a, Lia13, dlS13] for some specific examples, as well as the ensuing discussion on nonlinear type. In this general direction, here we investigate the following basic "meta-problem."

Question 1.1 (Comparison of nonlinear spectral gaps). Given $p\in[1,\infty)$, characterize those pairs of metric spaces (X,d_X) and (Y,d_Y) for which there exists an increasing function $\Psi=\Psi_{X,Y}:(0,\infty)\to(0,\infty)$ such that for every $n\in\mathbb{N}$ and every ergodic symmetric stochastic $A\in M_n(\mathbb{R})$ we have

$$\gamma(A,d_X^p)\leqslant\Psi\left(\gamma(A,d_Y^p)\right). \tag{6}$$

The case $p=2$ and $(Y,d_Y)=(\mathbb{R},d_{\mathbb{R}})$ of Question 1.1 is especially important, so we explicitly single it out as follows. See also the closely related question that Pisier posed as Problem 3.1 in [Pis10].

Question 1.2 (Bounding nonlinear gaps by linear gaps). Characterize those metric spaces (X, d_X) for which there exists an increasing function $\Psi = \Psi_X : (0,\infty) \to (0,\infty)$ such that for every $n \in \mathbb{N}$ and every ergodic symmetric stochastic $A \in M_n(\mathbb{R})$ we have

$$\gamma(A, d_X^p) \leqslant \Psi\left(\frac{1}{1-\lambda_2(A)}\right). \tag{7}$$

Question 1.1 and Question 1.2 seem to be difficult, and they might not have a useful simple-to-state answer. As an indication of this, in [MN13a] it is shown that there exists a $CAT(0)$ metric space (X, d_X), and for each $k \in \mathbb{N}$ there exist $n_k \in \mathbb{N}$ with $\lim_{k\to\infty} n_k = \infty$ such that there are symmetric stochastic matrices $A_k, B_k \in M_{n_k}(\mathbb{R})$ with

$$\sup_{k\in\mathbb{N}} \lambda_2(A_k) < 1 \qquad \text{and} \qquad \sup_{k\in\mathbb{N}} \lambda_2(B_k) < 1,$$

yet

$$\sup_{k\in\mathbb{N}} \gamma(A_k, d_X^2) < \infty \qquad \text{and} \qquad \sup_{k\in\mathbb{N}} \gamma(B_k, d_X^2) = \infty.$$

Such questions are difficult even in the setting of Banach spaces: a well-known open question (see e.g. [Pis10]) asks whether (7) holds true when X is a uniformly convex Banach space. If true, this would yield a remarkable "linear to nonlinear transference principle for spectral gaps," establishing in particular the existence of super-expanders (see [MN12]) with logarithmic girth, a result which could then be used in conjunction with Gromov's random group construction [Gro03] to rule out the success of an approach to the Novikov conjecture that was discovered by Kasparov and Yu [KY06]. There is little evidence, however, that every uniformly convex Banach space admits an inequality such as (7), and we suspect that (7) fails for some uniformly convex Banach spaces.

When the function Ψ appearing in (6) can be taken to be linear, i.e., $\Psi(t) = Kt$ for some $K \in (0,\infty)$, Theorem 1.3 below provides the following geometric answer to Question 1.1. Given $p \in [1,\infty)$ and a metric space (Y, d_Y), for every $m \in \mathbb{N}$ we denote by $\ell_p^m(Y)$ the metric space Y^m equipped with the metric

$$\forall\, x, y \in Y^m, \qquad d_{\ell_p^m(Y)}(x,y) \stackrel{\mathrm{def}}{=} \left(\sum_{i=1}^m d_Y(x_i, y_i)^p\right)^{\frac{1}{p}}.$$

Recall also the standard notation $\ell_p^m = \ell_p^m(\mathbb{R})$. A coordinate-wise application of (3) shows that every symmetric stochastic A satisfies

$$\forall\, m \in \mathbb{N}, \qquad \gamma(A, d_Y^p) = \gamma\left(A, d_{\ell_p^m(Y)}^p\right). \tag{8}$$

Fixing $D \in [1,\infty)$, suppose that (X,d_X) is a metric space such that for every $n \in \mathbb{N}$ and $x_1,\ldots,x_n \in X$ there exists $m \in \mathbb{N}$ and a non-constant mapping $f:\{x_1,\ldots,x_n\} \to \ell_p^m(Y)$ such that

$$\sum_{i=1}^n\sum_{j=1}^n d_{\ell_p^m(Y)}(f(x_i),f(x_j))^p \geqslant \frac{\|f\|_{\mathrm{Lip}}^p}{D}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p, \tag{9}$$

where $\|f\|_{\mathrm{Lip}}$ denotes the Lipschitz constant of f, i.e.,

$$\|f\|_{\mathrm{Lip}} \stackrel{\mathrm{def}}{=} \max_{\substack{x,y\in\{x_1,\ldots,x_n\}\\ x\neq y}} \frac{d_{\ell_p^m(Y)}(f(x),f(y))}{d_X(x,y)}.$$

Then for every symmetric stochastic matrix $A \in M_n(\mathbb{R})$,

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p &\stackrel{(9)}{\leqslant} \frac{D}{n^2\|f\|_{\mathrm{Lip}}^p}\sum_{i=1}^n\sum_{j=1}^n d_{\ell_p^m(Y)}(f(x_i),f(x_j))^p\\ &\stackrel{(8)}{\leqslant} \frac{D\gamma(A,d_Y^p)}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\frac{d_{\ell_p^m(Y)}(f(x_i),f(x_j))^p}{\|f\|_{\mathrm{Lip}}^p}\\ &\leqslant \frac{D\gamma(A,d_Y^p)}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^p. \end{aligned}\tag{10}$$

Consequently, $\gamma(A,d_X^p) \leqslant D\gamma(A,d_Y^p)$.

The above simple argument, combined with standard metric embedding methods, already implies that a variety of metric spaces satisfy the spectral inequality (7) with Ψ linear. This holds in particular when (X,d_X) belongs to one of the following classes of metric spaces: doubling metric spaces, compact Riemannian surfaces, Gromov hyperbolic spaces of bounded local geometry, Euclidean buildings, symmetric spaces, homogeneous Hadamard manifolds, and forbidden-minor (edge-weighted) graph families; this topic is treated in Section 7. We will also see below that the same holds true for certain Banach spaces, including $L_p(\mu)$ spaces for $p \in [2,\infty)$.

The following theorem asserts that the above reasoning using metric embeddings is the only way to obtain an inequality such as (6) with Ψ linear. Its proof is a duality argument that is inspired by the proof of a lemma of K. Ball [Bal92, Lem. 1.1] (see also [MN13b, Lem 5.2]).

Theorem 1.3. *Fix $n \in \mathbb{N}$ and $p, K \in [1,\infty)$. Given two metric spaces (X,d_X) and (Y,d_Y), the following assertions are equivalent.*

(1) For every symmetric stochastic n by n matrix A we have

$$\gamma(A,d_X^p) \leqslant K\gamma(A,d_Y^p). \tag{11}$$

(2) *For every* $D \in (K,\infty)$ *and every* $x_1,\ldots,x_n \in X$ *there exists* $m \in \mathbb{N}$ *and a nonconstant mapping* $f:\{x_1,\ldots,x_n\}\to \ell_p^m(Y)$ *that satisfies*

$$\sum_{i=1}^n\sum_{j=1}^n d_{\ell_p^m(Y)}(f(x_i),f(x_j))^p \geqslant \frac{\|f\|_{\mathrm{Lip}}^p}{D}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p.$$

It is worthwhile to single out the special case of Theorem 1.3 that corresponds to Question 1.2, in which case the embeddings are into ℓ_2.

Corollary 1.4. *Fix* $n \in \mathbb{N}$ *and* $K \in [1,\infty)$. *Given a metric space* (X,d_X), *the following assertions are equivalent.*

(1) *For every symmetric stochastic matrix* $A \in M_n(\mathbb{R})$ *we have*

$$\gamma(A,d_X^2) \leqslant \frac{K}{1-\lambda_2(A)}.$$

(2) *For every* $D \in (K,\infty)$ *and every* $x_1,\ldots,x_n \in X$ *there exists a nonconstant mapping* $f:\{x_1,\ldots,x_n\}\to \ell_2$ *that satisfies*

$$\sum_{i=1}^n\sum_{j=1}^n \|f(x_i)-f(x_j)\|_{\ell_2}^2 \geqslant \frac{\|f\|_{\mathrm{Lip}}^2}{D}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^2.$$

In Section 4 below we prove the following theorem.

Theorem 1.5. *For every* $p \in [2,\infty)$, *every* $n \in \mathbb{N}$ *and every symmetric stochastic matrix* $A \in M_n(\mathbb{R})$ *we have*

$$\gamma\Big(A,\|\cdot\|_{\ell_p}^2\Big) \lesssim \frac{p^2}{1-\lambda_2(A)}. \tag{12}$$

Inequality (12) is proved in Section 4 via a direct argument using analytic and probabilistic techniques, but once this inequality is established one can use duality through Corollary 1.4 to deduce the following new Hilbertian embedding result for arbitrary finite subsets of ℓ_p.

Corollary 1.6. *If* $n \in \mathbb{N}$ *and* $p \in (2,\infty)$ *then for every* $x_1,\ldots,x_n \in \ell_p$ *there exist* $y_1,\ldots,y_n \in \ell_2$ *such that*

$$\forall\, i,j \in \{1,\ldots,n\},\qquad \|y_i-y_j\|_{\ell_2} \lesssim p\|x_i-x_j\|_{\ell_p}, \tag{13}$$

and

$$\sum_{i=1}^n\sum_{j=1}^n \|y_i-y_j\|_{\ell_2}^2 = \sum_{i=1}^n\sum_{j=1}^n \|x_i-x_j\|_{\ell_p}^2.$$

The dependence on p in (13) is sharp up to the implicit universal constant, and the conclusion of Corollary 1.6 is false for $p \in [1,2)$ even if one allows any dependence on p in (13) (provided it is independent of

n): for the former statement see Lemma 1.11 below and for the latter statement see Lemma 1.12 below. It would be interesting to find a constructive proof of Corollary 1.6, i.e., a direct proof that does not rely on duality to show that the desired embedding exists.

Remark 1.7. Questions in the spirit of Theorem1.5 have been previously studied by Matoušek [Mat97], who proved that there exists a universal constant $C \in (0,\infty)$ such that for every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix $A \in M_n(\mathbb{R})$,

$$p \in [2,\infty) \implies \gamma\Big(A, \|\cdot\|_{\ell_p}^p\Big) \leqslant \frac{(Cp)^p}{(1-\lambda_2(A))^{p/2}}. \tag{14}$$

See [BLMN05, Lem. 5.5] and [NS11, Lem. 4.4] for the formulation and proof of this fact (which is known as Matoušek's extrapolation lemma for Poincaré inequalities) in the form stated in (14). In order to obtain an embedding result such as Corollary 1.6 one needs to bound $\gamma(A, \|\cdot\|_{\ell_p}^2)$ rather than $\gamma(A, \|\cdot\|_{\ell_p}^p)$ by a quantity that grows linearly with $1/(1-\lambda_2(A))$. We do not see how to use Matoušek's approach in [Mat97] to obtain such an estimate, and we therefore use an entirely different method (specifically, complex interpolation and Markov type) to prove Theorem 1.5.

As stated above, when $p \in [1,2)$ the analogue of Theorem 1.5 can hold true only if we allow the right hand side of (12) to depend on n. Specifically, we ask the following question.

Question 1.8. Is it true that for every $p \in [1,2]$, every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix A we have

$$\gamma(A, \|\cdot\|_{\ell_p}^2) \lesssim \frac{(\log n)^{\frac{2}{p}-1}}{1-\lambda_2(A)} \ ?$$

Due to Theorem 1.3, an affirmative answer to Question 1.8 is equivalent to the assertion that if $p \in [1,2]$ then for every $x_1,\ldots,x_n \in \ell_p$ there exist $y_1,\ldots,y_n \in \ell_2$ that satisfy

$$\sum_{i=1}^n\sum_{j=1}^n \|y_i-y_j\|_2^2 = \sum_{i=1}^n\sum_{j=1}^n \|x_i-x_j\|_p^2,$$

and

$$\forall\, i,j \in \{1,\ldots,n\}, \qquad \|y_i-y_j\|_2 \lesssim (\log n)^{\frac{1}{p}-\frac{1}{2}}\|x_i-x_j\|_p.$$

We conjecture that the answer to Question 1.8 is positive. As partial motivation for this conjecture, we note that by an important theorem of Arora, Rao and Vazirani [ARV09] the answer is positive when $p=1$.

A possible approach towards proving this conjecture for $p \in (1,2)$ is to investigate whether the quantity $\gamma(A, \|\cdot\|_X^2)$ behaves well under interpolation. In our case one would write $\frac{1}{p} = \frac{\theta}{2} + \frac{1-\theta}{1}$ for an appropriate $\theta \in (0,1)$ and ask whether or not it is true that for every $n \in \mathbb{N}$ every n by n symmetric stochastic matrix A satisfies

$$\gamma(A, \|\cdot\|_p^2) \lesssim \gamma(A, \|\cdot\|_2^2)^\theta \cdot \gamma(A, \|\cdot\|_1^2)^{1-\theta}. \tag{15}$$

Investigating the possible validity such interpolation inequalities for nonlinear spectral gaps is interesting in its own right; in Section 4.3 we derive a weaker interpolation inequality in this spirit.

1.1. **Applications to bi-Lipschitz embeddings.** The (bi-Lipschitz) distortion of a metric space (X, d_X) in a metric space (Y, d_Y), denoted $c_Y(X)$, is define to be the infimum over those $D \in [1,\infty)$ for which there exists $s \in (0,\infty)$ and a mapping $f : X \to Y$ that satisfies

$$\forall\, x, y \in X, \qquad s d_X(x,y) \leqslant d_Y(f(x), f(y)) \leqslant D s d_X(x,y).$$

Set $c_Y(x) = \infty$ if no such D exists. When $Y = \ell_p$ for some $p \in [1,\infty]$ we use the simpler notation $c_p(X) = c_{\ell_p}(X)$. The parameter $c_2(X)$ is known in the literature as the Euclidean distortion of X.

The Fréchet–Kuratowski embedding (see e.g. [Hei01]) shows that $c_\infty(X) = 1$ for every separable metric space X. We therefore define $p(X)$ to be the infimum over those $p \in [2,\infty]$ such that $c_p(X) < 10$. The choice of the number 10 here is arbitrary, and one can equally consider any fixed number bigger than 1 in place of 10 to define the parameter $p(X)$; we made this arbitrary choice rather than adding an additional parameter only for the sake of notational simplicity.

For $n \in \mathbb{N}$ and $d \in \{3, \ldots, n-1\}$ let $p(n,d)$ be the expectation of $p(G)$ when G is distributed uniformly at random over all connected n-vertex d-regular graphs, equipped with the shortest-path metric. Thus, if $p = p(n,d)$ then in expectation a connected n-vertex d-regular graph G satisfies $c_p(G) \leqslant 10$. In [Mat97] Matoušek evaluated the asymptotic dependence of the largest possible distortion of an n-point metric space in ℓ_p, yielding the estimate $p(n,d) \gtrsim \log_d n$, which is an asymptotically sharp bound if $d = O(1)$. As a consequence of our proof of Theorem 1.5, it turns out that Matoušek's bound is not sharp as a function of the degree d. Specifically, in Section 4.1 we prove the following result.

Proposition 1.9. *For every $n \in \mathbb{N}$ and $d \in \{3, \ldots, n-1\}$ we have*

$$d \leqslant e^{(\log n)^{2/3}} \implies p(n,d) \gtrsim \frac{\log n}{\sqrt{\log d}}, \tag{16}$$

and

$$d \geqslant e^{(\log n)^{2/3}} \implies p(n,d) \gtrsim (\log_d n)^2. \tag{17}$$

The significance of the quantity $\log_d n$ appearing in Matoušek's bound is that up to universal constant factors it is the typical diameter of a uniformly random connected n-vertex d-regular graph [BFdlV82]. Thus, in (16) there must be some restriction on the size of d relative to n since when d is at least a constant power of n the typical diameter of G is $O(1)$, and therefore $c_p(G) \leqslant c_2(G) = O(1)$. Both (16) and (17) assert that $p(n,d)$ tends to ∞ faster than the typical diameter of G when $n^{o(1)} = d \to \infty$ (note also that (17) is consistent with the fact that $p(n,d)$ must become bounded when d is large enough). While we initially expected Matoušek's bound to be sharp, Proposition (1.9) indicates that the parameter $p(n,d)$, and more generally the parameter $p(X)$ for a finite metric space (X,d_X), deserves further investigation. In particular, we make no claim that Proposition (1.9) is sharp.

The link between Theorem 1.5 and Proposition 1.9 is that our proof of Theorem 1.5 yields the following bound, which holds for every $n \in \mathbb{N}$, every n by n symmetric stochastic matrix A, and every $p \in [2,\infty)$.

$$\gamma\Big(A, \|\cdot\|_{\ell_p}^2\Big) \lesssim \frac{p}{1-\lambda(A)^{2/p}}, \tag{18}$$

where we recall that $\lambda(A)$ was defined in (2). Proposition 1.9 is deduced in Section 4.1 from (18) through a classical argument of Linial, London and Rabinovich [LLR95]. The bounds appearing in Proposition 1.9 hold true with $p(n,d)$ replaced by $p(G)$ when G is an n-vertex d-regular Ramanujan graph [LPS88, Mar88], and it is an independently interesting open question to evaluate $c_p(G)$ up to universal constant factors when G is Ramanujan. Some estimates in this direction are obtained in Section 4.1, where a similar question is also studied for Abelian Alon–Roichman graphs [AR94].

1.2. **Ozawa's localization argument for Poincaré inequalities.** Theorem 1.10 below provides a partial answer to Question 1.1 when X and Y are certain Banach spaces. Its proof builds on an elegant idea of Ozawa [Oza04] that was used in [Oza04] to rule out coarse embeddings of expanders into certain Banach spaces. While Ozawa's original argument did not yield a nonlinear spectral gap inequality in the form that we require here, it can be modified so as to yield Theorem 1.10; this is carried out in Section 5.

Throughout this article the unit ball of a Banach space $(X,\|\cdot\|_X)$ is denoted by

$$B_X \stackrel{\mathrm{def}}{=} \{x \in X:\ \|x\|_X \leqslant 1\}.$$

Theorem 1.10. *Let $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ be Banach spaces. Suppose that $\alpha,\beta:[0,\infty)\to[0,\infty)$ are increasing functions, β is concave, and*

$$\lim_{t\to 0}\alpha(t)=\lim_{t\to 0}\beta(t)=0.$$

Suppose also that there exists a mapping $\phi: B_X\to Y$ that satisfies

$$\forall\, x,y\in B_X,\quad \alpha\left(\|x-y\|_X\right)\leqslant \|\phi(x)-\phi(y)\|_Y\leqslant \beta\left(\|x-y\|_X\right). \tag{19}$$

Then for every $q\in[1,\infty)$, every $n\in\mathbb{N}$, and every symmetric stochastic matrix $A\in M_n(\mathbb{R})$ we have

$$\gamma(A,\|\cdot\|_X^q)\leqslant 8^{q+1}\gamma(A,\|\cdot\|_Y^q)+\frac{8^q}{\beta^{-1}\left(\frac{\alpha(1/4)}{8\gamma\left(A,\|\cdot\|_Y^q\right)^{1/q}}\right)^q}. \tag{20}$$

The key point of Theorem 1.10 is that one can use local information such as (19) in order to deduce a Poincaré-type inequality such as (3).

Certain classes of Banach spaces X are known to satisfy the assumptions of Theorem 1.10 when Y is a Hilbert space. These include: $L_p(\mu)$ spaces for $p\in[1,\infty)$, as shown by Mazur [Maz29] (see [BL00, Ch. 9, Sec. 1]); Banach spaces of finite cotype with an unconditional basis, as shown by Odell and Schlumprecht [OS94]; more generally, Banach lattices of finite cotype, as shown by Chaatit [Cha95] (see [BL00, Ch. 9, Sec. 2]); Schatten classes of finite cotype (and more general noncommutative L_p spaces), as shown by Raynaud [Ray02]. For these classes of Banach spaces Theorem 1.10 furnishes a positive answer to Question 1.2 with Ψ given by the right hand side of (20).

Mazur proved [Maz29] that for every $p\in[1,\infty)$, if $X=\ell_p$ and $Y=\ell_2$ then there exists a mapping $\phi: X\to Y$ that satisfies (19) with $\alpha(t)=2(t/2)^{p/2}$ and $\beta(t)=pt$ if $p\in[2,\infty)$ and $\alpha(t)=t/3$ and $\beta(t)=2t^{p/2}$ if $p\in[1,2]$ (these estimates are recalled in Section 3). Therefore, it follows from Theorem 1.10 that for every $p\in[1,\infty)$, every $n\in\mathbb{N}$ and every symmetric stochastic matrix $A\in M_n(\mathbb{R})$ we have

$$1\leqslant p\leqslant 2\implies \gamma\Big(A,\|\cdot\|_{\ell_p}^2\Big)\lesssim\frac{1}{(1-\lambda_2(A))^{2/p}}, \tag{21}$$

and

$$2\leqslant p<\infty\implies \gamma\Big(A,\|\cdot\|_{\ell_p}^2\Big)\lesssim\frac{p^2 8^p}{1-\lambda_2(A)}. \tag{22}$$

As explained in Section 1.3.1, both (21) and (22) are sharp in terms of the asymptotic dependence on $1-\lambda_2(A)$, but in terms of the dependence on p the bound (22) is exponentially worse than the (sharp) bound (12). In Section 5.1 we show that this exponential loss is inherent to Ozawa's method, in the sense that for $p\in(2,\infty)$, if $\phi:\ell_p\to\ell_2$ satisfies (19) with $\beta(t) = Kt$ for some $K \in (0,\infty)$ and every $t \in (0,\infty)$ (this is to ensure that (20) yields an upper bound on $\gamma(A, \|\cdot\|_{\ell_p}^2)$ that grows linearly with $(1-\lambda_2(A))^{-1}$), then necessarily $K/\alpha(1/4) \gtrsim 2^{3p/2}$.

Note that (22) suffices via duality (i.e., Corollary 1.4) to obtain an embedding result for arbitrary subsets of ℓ_p as in Corollary 1.6, with exponentially worse dependence on p. Yet another proof of such an embedding statement appears in Section 7, though it also yields a bound in terms of p that is exponentially worse than Corollary 1.6. We do not know how to prove the sharp statement of Corollary 1.6 other than through Theorem 1.5, whose proof is not as elementary as the above mentioned proofs that yield an exponential dependence on p.

1.3. **Average distortion embeddings and nonlinear type.** For $p,q \in (0,\infty)$ the (p,q)-*average distortion* of (X,d_X) into (Y,d_Y), denoted $Av_Y^{(p,q)}(X) \in [1,\infty)$, is the infimum over those $D \in [1,\infty]$ such that for every $n \in \mathbb{N}$ and every $x_1,\ldots,x_n \in X$ there exists a nonconstant Lipschitz function $f : \{x_1,\ldots,x_n\} \to Y$ that satisfies

$$\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_Y(f(x_i),f(x_j))^p\right)^{\frac{1}{p}} \geqslant \frac{\|f\|_{\mathrm{Lip}}}{D}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^q\right)^{\frac{1}{q}}.$$

When $p=q$ we use the simpler notation $Av_Y^{(p)}(X) \stackrel{\mathrm{def}}{=} Av_Y^{(p,p)}(X)$.

The notion of average distortion and its relevance to approximation algorithms was brought to the fore in the influential work [Rab08] of Rabinovich. Parts of Section 7 below are inspired by Rabinovich's ideas in [Rab08]. Earlier applications of this notion include the work of Alon, Boppana and Spencer [ABS98] that related average distortion to asymptotically sharp isoperimetric theorems on products spaces; see Remark 7.9 below. In the linear theory of Banach spaces average distortion embeddings have been studied in several contexts; see e.g. the work on random sign embeddings in [Elt83, FJS88].

With the above terminology, Theorem 1.3 asserts that the linear dependence (11) holds true if and only if for every finite subset $S \subseteq X$ there exists $m \in \mathbb{N}$ such that

$$Av^{(p)}_{\ell_p^m(Y)}(S,d_X) \leqslant K^{1/p}.$$

By Corollary 1.6, if $p \in [2,\infty)$ then

$$Av^{(2)}_{\ell_2}(\ell_p) \lesssim p. \tag{23}$$

As stated earlier, the estimate (23) cannot be improved (up to the implicit constant factor): this is a special case of the following lemma.

Lemma 1.11. *For every $p,q,r,s\in[1,\infty)$ with $2\leqslant q\leqslant p$ and every $n\in\mathbb{N}$ there exist $x_1,\ldots,x_n\in\ell_p$ such that if $y_1,\ldots,y_n\in\ell_q$ satisfy*

$$\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|y_i-y_j\|_{\ell_q}^r\right)^{\frac{1}{r}}=\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|_{\ell_p}^s\right)^{\frac{1}{s}},\qquad(24)$$

then there exist $i,j\in\{1,\ldots,n\}$ such that

$$\|y_i-y_j\|_{\ell_q}\gtrsim\frac{p}{q+r}\cdot\|x_i-x_j\|_{\ell_p}>0.$$

The proof of Lemma 1.11 is given in Section 4.3. It suffices to say at this juncture that the points $x_1,\ldots,x_n\in\ell_p$ are the images of the vertices of a bounded degree expanding n-vertex regular graph under Matoušek's ℓ_p-variant [Mat97] of Bourgain's embedding [Bou85].

We also stated earlier that Corollary 1.6 fails for $p\in[1,2)$: this is a special case of the following lemma.

Lemma 1.12. *Fix $p,q,r,s\in[1,\infty)$ with $p\in[1,2)$ and $q\in(p,\infty)$. For arbitrarily large $n\in\mathbb{N}$ there exist $x_1,\ldots,x_n\in\ell_p$ such that for every $y_1,\ldots,y_n\in\ell_q$ with*

$$\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|y_i-y_j\|_{\ell_q}^r\right)^{\frac{1}{r}}=\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|_{\ell_p}^s\right)^{\frac{1}{s}},$$

there exist $i,j\in\{1,\ldots,n\}$ such that

$$\|y_i-y_j\|_{\ell_q}\gtrsim\frac{(\log n)^{\frac{1}{p}-\frac{1}{\min\{q,2\}}}}{\sqrt{q+r}}\cdot\|x_i-x_j\|_{\ell_p}>0.\qquad(25)$$

1.3.1. *Bourgain–Milman–Wolfson type.* The reason for the validity of the lower bound (25) is best explained in the context of nonlinear type: a metric invariant that furnishes an obstruction to the existence of average distortion embeddings. Let $\mathbb{F}_2$ be the field of cardinality 2 and for $n\in\mathbb{N}$ let $e_1,\ldots,e_n$ be the standard basis of $\mathbb{F}_2^n$. We also write $e=e_1+\ldots+e_n$. Following Bourgain, Milman and Wolfson [BMW86], given $p,T\in(0,\infty)$, a metric space (X,d_X) is said to have BMW type p with constant T if for every $n\in\mathbb{N}$ every $f:\mathbb{F}_2^n\to X$ satisfies

$$\sum_{x\in\mathbb{F}_2^n}d_Y(f(x),f(x+e))^2\leqslant T^2n^{\frac{2}{p}-1}\sum_{i=1}^n\sum_{x\in\mathbb{F}_2^n}d_Y\left(f(x),f(x+e_i)\right)^2.\quad(26)$$

(X,d_X) has BMW type p if it has BMW type p with constant T for some $T\in(0,\infty)$; in this case the infimum over those $T\in(0,\infty)$ for which (26) holds true is denoted $\mathrm{BMW}_p(X)$. For background on this notion we refer to [BMW86], as well as [Pis86, NS02, Nao12a]. These references also contain a description of the closely related important notion of Enflo type [Enf76], a notion whose definition is recalled below but will not be further investigated here.

The simple proof of the following lemma appears in Section 6.

Lemma 1.13. *Fix $p \in (0,\infty)$. For every two metric spaces (X, d_X) and (Y, d_Y) we have*

$$\mathrm{BMW}_p(X) \leqslant 2Av_Y^{(2)}(X) \cdot \mathrm{BMW}_p(Y).$$

The case $r = s = 2$ of Lemma 1.12 follows from Lemma 1.13 and the computations of BMW type that appear in the literature. The remaining cases of Lemma 1.12 are proved using similar ideas.

The Hamming cube is the Cayely graph on $\mathbb{F}_2^n$ corresponding to the set of generators $\{e_1, \dots, e_n\}$. The shortest path metric on this graph coincides with the ℓ_1^n metric under the identification of $\mathbb{F}_2^n$ with $\{0,1\}^n \subseteq \mathbb{R}^n$. Let H_n be the 2^n by 2^n symmetric stochastic matrix which is the normalized adjacency matrix of the Hamming cube. Thus for $x, y \in \mathbb{F}_2^n$ the (x, y)-entry of H_n equals 0 unless $x - y \in \{e_1, \dots, e_n\}$, in which case it equals $1/n$. It is well known (and easy to check) that $\lambda_2(H_n) = 1 - 2/n$, so $\gamma(H_n, d_{\mathbb{R}}^2) \asymp n$. A simple argument (that is explained in Section 6) shows that the definition (26) is equivalent to the requirement that $\gamma(H_n, d_X^2) \lesssim_X n^{2/p}$ for every $n \in \mathbb{N}$. The notion of Enflo type p that was mentioned above is equivalent to the requirement that $\gamma(H_n, d_X^p) \lesssim_X n$ for every $n \in \mathbb{N}$.

For $p \in [1, 2]$, by considering the identity mapping of $\mathbb{F}_2^n$ into ℓ_p^n (observe that $\|x - y\|_p^p = \|x - y\|_1$ for every $x, y \in \mathbb{F}_2^n$) one sees that $\gamma(H_n, \|\cdot\|_p^2) \gtrsim n^{2/p} \asymp 1/(1 - \lambda_2(H_n))^{2/p}$. Thus (21) is sharp.

1.3.2. *Towards a nonlinear Maurey–Pisier theorem.* Every metric space has BMW type 1 and no metric space has BMW type greater than 2; see Remark 6.4 below. Thus, for a metric space (X, d_X) define

$$p_X \stackrel{\mathrm{def}}{=} \sup\left\{p \in [1, 2] : \ \mathrm{BMW}_p(X) < \infty\right\}. \tag{27}$$

Maurey and Pisier [MP76] associate a quantity p_X to every Banach space X, which is defined analogously to (27) but with BMW type replaced by Rademacher type. The (linear) notion of Rademacher type is recalled in Section 6.2 below; at this juncture we just want to state, for the sake of readers who are accustomed to the standard Banach space terminology, that despite the apparent conflict of notation between (27) and [MP76], a beautiful theorem of Bourgain, Milman and Wolfson [BMW86] asserts that actually the two quantities coincide.

The following theorem is due to Bourgain, Milman and Wolfson.

Theorem 1.14 ([BMW86]). *Suppose that (X, d_X) is a metric space with $p_X = 1$. Then $c_X(\mathbb{F}_2^n, \|\cdot\|_1) = 1$ for every $n \in \mathbb{N}$.*

Thus, the only possible obstruction to a metric space (X, d_X) having BMW type p for some $p > 1$ is the presence of arbitrarily large Hamming cubes. This is a metric analogue of a classical theorem of Pisier [Pis73] asserting that the only obstruction to a Banach space having nontrivial Rademacher type is the presence of ℓ_1^n for every $n \in \mathbb{N}$.

In light of the Maurey–Pisier theorem [MP76] for Rademacher type, it is natural to ask if a similar result holds true for a general metric space X even when $p_X > 1$: Is it true that for every metric space (X, d_X) we have

$$\sup_{n\in\mathbb{N}} c_X(\mathbb{F}_2^n, \|\cdot\|_{p_X}) < \infty?$$

The answer to this question is negative if $p_X = 2$. Indeed, we have $p_\mathbb{R} = 2$ yet $c_\mathbb{R}(\mathbb{F}_2^n, \|\cdot\|_2)$ tends to infinity exponentially fast as $n \to \infty$ because there is an exponentially large subset S of $\mathbb{F}_2^n$ with the property that $\|x - y\|_2 \asymp \sqrt{n}$ for every distinct $x, y \in S$. If, however, $p_X \in (1, 2)$ then the above question, called the *Maurey–Pisier problem for BMW type*, remains open.

The Maurey–Pisier problem for BMW type was posed by Bourgain, Milman and Wolfson in [BMW86], where they obtained a partial result about it: they gave a condition on a metric space (X, d_X) that involves its BMW type as well as an additional geometric restriction that ensures that $\sup_{n\in\mathbb{N}} c_X(\mathbb{F}_2^n, \|\cdot\|_{p_X}) < \infty$; see Section 4 of [BMW86].

In Section 6.1 we prove the following theorem.

Theorem 1.15. *For every metric space (X, d_X) and every $d \in \mathbb{N}$ there exists $N = N(d, X) \in \mathbb{N}$ such that*

$$c_{\ell_2^N(X)}\left(\mathbb{F}_2^d, \|\cdot\|_{p_X}\right) \leqslant \mathrm{BMW}_{p_X}(X)^2.$$

Thus, if $\mathrm{BMW}_{p_X}(X) < \infty$, i.e., the supremum defining p_X in (27) is attained, then for every $d \in \mathbb{N}$ one can embed $(\mathbb{F}_2^d, \|\cdot\|_{p_X})$ into $\ell_2^N(X)$ for sufficiently large $N \in \mathbb{N}$. Note that it follows immediately from (27) that $\mathrm{BMW}_p(\ell_2^N(X)) = \mathrm{BMW}_p(X)$ for every $N \in \mathbb{N}$, so Theorem 1.15 is a complete metric characterization of the parameter p_X when the supremum defining p_X in (27) is attained. Note also that by passing to $\ell_2^N(X)$ we overcome the issue that was described above if $p_X = 2$, since trivially $(\mathbb{F}_2^n, \|\cdot\|_2)$ is isometric to a subset of $\ell_2^n(\mathbb{R})$. This indicates why Theorem 1.15 is easier than the actual Maurey–Pisier problem for BMW type, whose positive solution would require using the assumption $p_X < 2$. We therefore ask whether or not it is true that for every metric space (X, d_X) and every $p \in (1, 2)$, if there is $K \in (0, \infty)$ such that for every $d \in \mathbb{N}$ there exists $N \in \mathbb{N}$ for which $(\mathbb{F}_2^d, \|\cdot\|_p)$ embeds with distortion K into $\ell_2^N(X)$, then $\sup_{d\in\mathbb{N}} c_X(\mathbb{F}_2^d, \|\cdot\|_p) < \infty$? For $p = 1$ the answer to this question is positive due to Theorem 1.14, and by virtue of Theorem 1.15 a positive answer to this question would imply an affirmative solution of the Maurey–Pisier problem for BMW type.

Recalling Lemma 1.13, given the relation between BMW type and average distortion embeddings, it is natural to study the following weaker version of the Maurey–Pisier problem for BMW type: Is it true that for every metric space (X, d_X) we have

$$p_X < 2 \implies \sup_{n\in\mathbb{N}} Av_X^{(2)}(\mathbb{F}_2^n, \|\cdot\|_{p_X}) < \infty? \tag{28}$$

In (28) we restrict to $p_X < 2$ because one can show that

$$Av_{\mathbb{R}}^{(2)}(\mathbb{F}_2^n, \|\cdot\|_2) \asymp \sqrt[4]{n}. \tag{29}$$

See Remark 6.9 below for the proof of (29).

By Theorem 1.14 and Lemma 1.13, for every metric space (X, d_X),

$$\sup_{n\in\mathbb{N}} Av_X^{(2)}(\mathbb{F}_2^n, \|\cdot\|_1) < \infty \implies \sup_{n\in\mathbb{N}} c_X(\mathbb{F}_2^n, \|\cdot\|_1) = 1. \tag{30}$$

Given $p \in (1, \infty)$, it is therefore natural to ask whether or not for every metric space (X, d_X) we have

$$\sup_{n\in\mathbb{N}} Av_X^{(2)}(\mathbb{F}_2^n, \|\cdot\|_p) < \infty \implies \sup_{n\in\mathbb{N}} c_X(\mathbb{F}_2^n, \|\cdot\|_p) < \infty? \tag{31}$$

If (31) were true for every $p \in (1, 2)$ then a positive answer to the question that appears in (28) would imply a positive solution to the Maurey–Pisier problem for BMW type.

More generally, in light of the availability of results such as (31), it would be of interest to relate average distortion embeddings to bi-Lipschitz embeddings. For example, is it true that if a metric space (X, d_X) satisfies $Av_{\ell_2}^{(2)}(X) < \infty$ then for every finite subset $S \subseteq X$ we have $c_2(S) = o_X(\log|S|)$? If the answer to this question is positive then Corollary 1.6 would imply that for $p \in (2, \infty)$ any n-point subset of ℓ_p embeds into Hilbert space with distortion $o_p(\log n)$. No such improvement over Bourgain's embedding theorem [Bou85] is known for finite subsets of ℓ_p if $p \in (2, \infty)$; for $p \in [1, 2)$ see [CGR08, ALN08].

Acknowledgements. I thank Noga Alon, Manor Mendel, and especially Yuval Rabani for very helpful discussions. I am also grateful for the hospitality of Université Pierre et Marie Curie, Paris, France, where part of this work was completed.

2. Absolute spectral gaps

Fix $n \in \mathbb{N}$ and an n by n symmetric stochastic matrix $A = (a_{ij})$. Following [MN12], for $p \in [1, \infty)$ and a metric space (X, d_X), denote by $\gamma_+(A, d_X^p)$ the infimum over those $\gamma_+ \in (0, \infty]$ for which every $x_1, \ldots, x_n, y_1, \ldots, y_n \in X$ satisfy

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i, y_j)^p \leqslant \frac{\gamma_+}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} d_X(x_i, y_j)^p. \tag{32}$$

Note that by definition $\gamma(A, d_X^p) \leqslant \gamma_+(A, d_X^p)$. Recalling (2), we have

$$\gamma_+(A, d_{\mathbb{R}}^2) = \frac{1}{1 - \lambda(A)}.$$

For this reason one thinks of $\gamma_+(A, d_X^p)$ as measuring the magnitude of the nonlinear *absolute* spectral gap of the matrix A with respect to the kernel $d_X^p : X \times X \to [0, \infty)$.

The parameter $\gamma_+(A, d_X^p)$ is useful in various contexts (see [MN12]), and in particular it will be used in some of the ensuing arguments. It is natural to ask for the analogue of Question 1.1 with $\gamma(\cdot, \cdot)$ replaced by $\gamma_+(\cdot, \cdot)$. However, it turns out that this question is essentially the same as Question 1.1, as explained in Proposition 2.1 below.

Proposition 2.1. *Fix $p \in [1, \infty)$ and metric spaces $(X, d_X), (Y, d_Y)$. Suppose that there exists an increasing function $\Psi : (0, \infty) \to (0, \infty)$ such that for every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix A we have*

$$\gamma(A, d_X^p) \leqslant \Psi\left(\gamma(A, d_Y^p)\right). \tag{33}$$

Then for every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix A we also have

$$\gamma_+(A, d_X^p) \leqslant 2\Psi\left(2^p \gamma_+(A, d_Y^p)\right). \tag{34}$$

Conversely, suppose that $\Phi : (0, \infty) \to (0, \infty)$ is an increasing function such that for every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix A we have

$$\gamma_+(A, d_X^p) \leqslant \Phi\left(\gamma_+(A, d_Y^p)\right). \tag{35}$$

Then for every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix A we also have

$$\gamma(A, d_X^p) \leqslant \frac{1}{2}\Phi\left(2^{2p+1}\gamma(A, d_Y^p)\right). \tag{36}$$

Before passing to the (simple) proof of Proposition 2.1, we record for future use some basic facts about nonlinear spectral gaps.

Lemma 2.2. *Fix $p \in [1,\infty)$, $n \in \mathbb{N}$ and a metric space (X,d_X). Then every symmetric stochastic matrix $A=(a_{ij}) \in M_n(\mathbb{R})$ satisfies*

$$\gamma(A,d_X^p) \geqslant 1-\frac{1}{n} \qquad \text{and} \qquad \gamma_+(A,d_X^p) \geqslant 1. \tag{37}$$

Proof. Fix distinct $a,b \in X$ and let $x_1,\ldots,x_n \in X$ be i.i.d. points, each of which is chosen uniformly at random from $\{a,b\}$. Then every symmetric stochastic $A \in M_n(\mathbb{R})$ satisfies

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^p\right] = \frac{d_X(a,b)^p}{2n}\sum_{\substack{i,j\in\{1,\ldots,n\}\\ i\neq j}} a_{ij} \leqslant \frac{d_X(a,b)^p}{2},$$

while

$$\mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p\right] = \frac{n(n-1)}{2n^2}d_X(a,b)^p.$$

the desired conclusion now follows from the definition (3). The rightmost inequality in (37) follows by substituting $x_1=\ldots=x_n=a$ and $y_1=\ldots=y_n=b$ into (32). $\square$

Lemma 2.3. *Fix $p \in [1,\infty)$ and a metric space (X,d_X). Then for every integer $n \geqslant 2$, every n by n symmetric stochastic matrix $A=(a_{ij})$ satisfies*

$$2\gamma(A,d_X^p) \leqslant \gamma_+\left(\frac{I+A}{2},d_X^p\right) \leqslant 2^{2p+1}\gamma(A,d_X^p). \tag{38}$$

Proof. Since the diagonal entries of A play no role in the definition of $\gamma(A,d_X^2)$, it follows immediately from (3) that

$$\gamma\left(\frac{I+A}{2},d_X^p\right) = 2\gamma(A,d_X^p). \tag{39}$$

Because $\gamma_+\left(\frac{I+A}{2},d_X^p\right) \geqslant \gamma\left(\frac{I+A}{2},d_X^p\right)$, this implies the leftmost inequality in (38).

Next, fix $x_1,\ldots,x_n,y_1,\ldots,y_n \in X$. By the triangle inequality and the convexity of $t \mapsto |t|^p$, for every $i,j \in \{1,\ldots,n\}$ we have

$$d_X(x_i,y_j)^p \leqslant 2^{p-1}\left(d_X(x_i,x_j)^p + d_X(x_j,y_j)^p\right), \tag{40}$$

and

$$d_X(x_i,y_j)^p \leqslant 2^{p-1}\left(d_X(x_i,y_i)^p + d_X(y_i,y_j)^p\right). \tag{41}$$

By averaging (40) and (41), and then averaging the resulting inequality over $i, j \in \{1, \ldots, n\}$, we see that

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i, y_j)^p \leqslant \frac{2^p}{2n}\sum_{i=1}^{n} d_X(x_i, y_i)^p + \frac{2^p}{4n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(d_X(x_i, x_j)^p + d_X(y_i, y_j)^p\right). \quad (42)$$

Now, by the definition of $\gamma(A, d_X^p)$ we have

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(d_X(x_i, x_j)^p + d_X(y_i, y_j)^p\right) \leqslant \frac{\gamma(A, d_X^p)}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(d_X(x_i, x_j)^p + d_X(y_i, y_j)^p\right). \quad (43)$$

Next, for every $i, j \in \{1, \ldots, n\}$ we have

$$d_X(x_i, x_j)^p \leqslant 2^{p-1}\left(d_X(x_i, y_j)^p + d_X(y_j, x_j)^p\right), \quad (44)$$

$$d_X(x_i, x_j)^p \leqslant 2^{p-1}\left(d_X(x_i, y_i)^p + d_X(y_i, x_j)^p\right), \quad (45)$$

$$d_X(y_i, y_j)^p \leqslant 2^{p-1}\left(d_X(y_i, x_j)^p + d_X(x_j, y_j)^p\right), \quad (46)$$

and

$$d_X(y_i, y_j)^p \leqslant 2^{p-1}\left(d_X(y_i, x_i)^p + d_X(x_i, y_j)^p\right). \quad (47)$$

By averaging (44), (45), (46) and (47) we see that

$$d_X(x_i, x_j)^p + d_X(y_i, y_j)^p \leqslant 2^{p-1}\left(d_X(x_i, y_j)^p + d_X(y_i, x_j)^p + d_X(x_i, y_i)^p + d_X(x_j, y_j)^p\right). \quad (48)$$

By multiplying (48) by a_{ij}/n and summing over $i, j \in \{1, \ldots, n\}$ while using the fact that A is symmetric and stochastic we conclude that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(d_X(x_i, x_j)^p + d_X(y_i, y_j)^p\right) \leqslant \frac{2^p}{n}\sum_{i=1}^{n} d_X(x_i, y_i)^p + \frac{2^p}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} d_X(x_i, y_j)^p. \quad (49)$$

A substitution of (49) into (43), and then a substitution of the resulting inequality into (42), yields the following estimate.

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i,y_j)^p\\
&\leqslant \frac{2^{p-1}+2^{2p-2}\gamma(A,d_X^p)}{n}\sum_{i=1}^{n} d_X(x_i,y_i)^p\\
&\quad+\frac{2^{2p-2}\gamma(A,d_X^p)}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}d_X(x_i,y_j)^p\\
&\leqslant \frac{2^{p}+2^{2p-1}\gamma(A,d_X^p)}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{I+A}{2}\right)_{ij} d_X(x_i,y_j)^p. \qquad (50)
\end{aligned}
$$

Since (50) holds true for every $x_1,\ldots,x_n,y_1,\ldots,y_n\in X$, we deduce from the definition of $\gamma_+(\cdot,\cdot)$ in (32) that

$$
\gamma_+\left(\frac{I+A}{2},d_X^p\right)\leqslant 2^p+2^{2p-1}\gamma(A,d_X^p). \qquad (51)
$$

The rightmost inequality in (38) is a consequence of (51) since by Lemma 2.2 we have $\gamma(A,d_X^p)\geqslant\frac12$. □

Proof of Proposition 2.1. Fix $n\in\mathbb{N}$ and an n by n symmetric stochastic matrix A. Assuming the validity of (35) for the matrix $(I+A)/2$, we deduce (36). Next, by [MN12, Lem. 2.3] for every metric space (W,d_W) we have

$$
\frac{2}{2^p+1}\gamma\left(\left(\begin{smallmatrix}0 & A\\ A & 0\end{smallmatrix}\right),d_W^p\right)\leqslant\gamma_+(A,d_W^p)\leqslant 2\gamma\left(\left(\begin{smallmatrix}0 & A\\ A & 0\end{smallmatrix}\right),d_W^p\right). \qquad (52)
$$

Therefore, assuming the validity of (33) for the matrix $\left(\begin{smallmatrix}0 & A\\ A & 0\end{smallmatrix}\right)\in M_{2n}(\mathbb{R})$, the desired estimate (34) follows from (52). □

3. Duality

The implication (2) $\implies$ (1) of Theorem 1.3 was already proved in (10). Below we prove the more substantial implication (1) $\implies$ (2).

Proof of Theorem 1.3. Fix $D\in(K,\infty)$ and let $\varepsilon\in(0,\infty)$ be given by $(1+\varepsilon)(K+2\varepsilon)=D$. Fixing also $x_1,\ldots,x_n\in X$, let $C\subseteq M_n(\mathbb{R})$ be the set of n by n symmetric matrices (c_{ij}) for which there exists $y_1,\ldots,y_n\in Y$ with $|\{y_1,\ldots,y_n\}|\geqslant 2$ and

$$
\forall\, i,j\in\{1,\ldots,n\},\qquad c_{ij}=\frac{\sum_{r=1}^n\sum_{s=1}^n d_X(x_r,x_s)^p}{\sum_{r=1}^n\sum_{s=1}^n d_Y(y_r,y_s)^p}\cdot d_Y(y_i,y_j)^p.
$$

Letting $P \subseteq M_n(\mathbb{R})$ be the set of all symmetric n by n matrices with nonnegative entries and vanishing diagonal, denote

$$M \stackrel{\mathrm{def}}{=} \mathrm{conv}\,(C+P)\,.$$

For every $i,j\in\{1,\ldots,n\}$ write $t_{ij}\stackrel{\mathrm{def}}{=}(K+2\varepsilon)d_X(x_i,x_j)^p$. We first claim that the matrix $T=(t_{ij})\in M_n(\mathbb{R})$ belongs to M. Indeed, if this were not the case then by the separation theorem (Hahn–Banach) there would exist a symmetric matrix $H=(h_{ij})\in M_n(\mathbb{R})$ which has at least one nonzero off-diagonal entry and whose diagonal vanishes, satisfying

$$\inf_{B=(b_{ij})\in M}\sum_{i=1}^n\sum_{j=1}^n h_{ij}b_{ij} \geqslant (K+2\varepsilon)\sum_{i=1}^n\sum_{j=1}^n h_{ij}d_X(x_i,x_j)^p. \tag{53}$$

Since $P\subseteq M$, the fact that the left hand side of (53) is bounded from below implies that $h_{ij}\geqslant 0$ for all $i,j\in\{1,\ldots,n\}$. Fixing $\delta\in(0,1)$, if we define

$$\sigma \stackrel{\mathrm{def}}{=} \max_{i\in\{1,\ldots,n\}}\sum_{j\in\{1,\ldots,n\}\smallsetminus\{i\}}(h_{ij}+\delta),$$

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

$$a_{ij}\stackrel{\mathrm{def}}{=}\frac{1}{2\sigma}\begin{cases}2\sigma-\sum_{r\in\{1,\ldots,n\}\smallsetminus\{i\}}(h_{ir}+\delta) & \text{if } i=j,\\ h_{ij}+\delta & \text{if } i\neq j,\end{cases}$$

then, provided $\delta\in(0,1)$ is small enough, $A=(a_{ij})\in M_n(\mathbb{R})$ is a symmetric stochastic matrix all of whose entries are positive such that

$$\begin{aligned}\inf_{B=(b_{ij})\in M}\sum_{i=1}^n\sum_{j=1}^n a_{ij}b_{ij} &\stackrel{(53)}{\geqslant} (K+\varepsilon)\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^p\\ &\stackrel{(11)}{\geqslant} \frac{K+\varepsilon}{n\gamma(A,d_X^p)}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p. \end{aligned}\tag{54}$$

By the definition of $C\subseteq M_n(\mathbb{R})$ and $\gamma(A,d_Y^p)$ we have

$$\begin{aligned}&\inf_{B=(b_{ij})\in C}\sum_{i=1}^n\sum_{j=1}^n a_{ij}b_{ij}\\ &=\inf_{\substack{y_1,\ldots,y_n\in Y\\ |\{y_1,\ldots,y_n\}|\geqslant 2}}\frac{\sum_{r=1}^n\sum_{s=1}^n d_X(x_r,x_s)^p}{\sum_{r=1}^n\sum_{s=1}^n d_Y(y_r,y_s)^p}\cdot\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_Y(y_i,y_j)^p\\ &=\frac{1}{n\gamma(A,d_Y^p)}\sum_{r=1}^n\sum_{s=1}^n d_X(x_r,x_s)^p. \end{aligned}\tag{55}$$

Because $M \supseteq C$ it follows from (54) and (55) that

$$\frac{1}{n\gamma(A,d_Y^p)}\sum_{r=1}^n\sum_{s=1}^n d_X(x_r,x_s)^p \geqslant \frac{K+\varepsilon}{n\gamma(A,d_X^p)}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p$$
$$\geqslant \frac{K+\varepsilon}{Kn\gamma(A,d_Y^p)}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p, \quad (56)$$

Since all the entries of A are positive, $\gamma(A,d_Y^p)\in(0,\infty)$, and therefore (56) furnishes the desired contradiction.

Having proved that $T\in M$, it follows that there exist $N\in\mathbb{N}$ and $\mu_1,\ldots,\mu_N\in(0,1]$ with $\sum_{k=1}^N\mu_k=1$, and for every $k\in\{1,\ldots,N\}$ there are $y_1^k,\ldots,y_n^k\in Y$ with $|\{y_1^k,\ldots,y_n^k\}|\geqslant 2$, such that for every $i,j\in\{1,\ldots,n\}$ we have

$$(K+2\varepsilon)d_X(x_i,x_j)^p \geqslant \sum_{k=1}^N \mu_k\frac{\sum_{r=1}^n\sum_{s=1}^n d_X(x_r,x_s)^p}{\sum_{r=1}^n\sum_{s=1}^n d_Y(y_r^k,y_s^k)^p}\cdot d_Y(y_i^k,y_j^k)^p. \quad (57)$$

There are integers $q_1,\ldots,q_N,Q\in\mathbb{N}$ such that

$$\forall\, k\in\{1,\ldots,N\},\quad \frac{q_k}{Q}\leqslant \mu_k\frac{\sum_{r=1}^n\sum_{s=1}^n d_X(x_r,x_s)^p}{\sum_{r=1}^n\sum_{s=1}^n d_Y(y_r^k,y_s^k)^p}\leqslant(1+\varepsilon)\frac{q_k}{Q}. \quad (58)$$

Setting $q_0\stackrel{\mathrm{def}}{=}0$ and $m\stackrel{\mathrm{def}}{=}\sum_{k=1}^N q_k$, define $f:\{x_1,\ldots,x_n\}\to\ell_p^m(Y)$ by

$$\forall\, k\in\{1,\ldots,N\},\ \forall\, u\in\left[1+\sum_{j=0}^{k-1}q_j,\sum_{j=0}^{k}q_j\right]\cap\mathbb{N},\quad f(x_i)_u=y_i^k. \quad (59)$$

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

$$d_{\ell_p^m(Y)}(f(x_i),f(x_j)) \stackrel{(59)}{=} \left(\sum_{k=1}^N q_k d_Y(y_i^k,y_j^k)^p\right)^{\frac1p} \stackrel{(57)\wedge(58)}{\leqslant} Q^{1/p}(K+2\varepsilon)^{1/p}d_X(x_i,x_j).$$

Consequently,

$$\|f\|_{\mathrm{Lip}}\leqslant Q^{1/p}(K+2\varepsilon)^{1/p}. \quad (60)$$

Hence,

$$\sum_{i=1}^n\sum_{j=1}^n d_Y(f(x_i),f(x_j))^p \stackrel{(59)}{=} \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^N q_k d_Y(y_i^k,y_j^k)^p$$
$$\stackrel{(58)}{\geqslant}\frac{Q}{1+\varepsilon}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p \stackrel{(60)}{\geqslant}\frac{\|f\|_{\mathrm{Lip}}^p}{(1+\varepsilon)(K+2\varepsilon)}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p.$$

Since $(1+\varepsilon)(K+2\varepsilon)=D$, the proof of Theorem 1.3 is complete. $\square$

4. Interpolation and Markov type

Fix $p\in[1,\infty)$ and $m\in\mathbb{N}$. Following K. Ball [Bal92], given a metric space (X,d_X) define its *Markov type p constant at time m*, denoted $M_p(X;m)$, to be the infimum over those $M\in(0,\infty)$ such that for every $n\in\mathbb{N}$, every $x_1,\ldots,x_n\in X$ and every n by n symmetric stochastic matrix $A=(a_{ij})\in M_n(\mathbb{R})$ we have

$$\sum_{i=1}^n\sum_{j=1}^n (A^m)_{ij}d_X(x_i,x_j)^p \leqslant M^p m\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^p. \tag{61}$$

X is said to have Markov type p if $M_p(X)\stackrel{\mathrm{def}}{=}\sup_{m\in\mathbb{N}}M_p(X;m)<\infty$. Note that it follows from the triangle inequality that

$$\forall\, m\in\mathbb{N},\quad M_p(X;m)\leqslant m^{1-\frac{1}{p}}.$$

Remark 4.1. In Section 1.3 we recalled the notions of BMW type and Enflo type. The link between these notions and Ball's notion of Markov type is that Markov type p implies Enflo type p (see [NS02]). One can also define natural variants of Markov type that imply BMW type (see the inequalities appearing in Theorem 4.4 of [NPSS06]). Recently Kondo proved [Kon11] that there exists a Hadamard space (see e.g. [BH99]) that fails to have Markov type p for any $p>1$, answering a question posed in [NPSS06]. Since Hadamard spaces have Enflo type 2 (see [Oht09a]), this yields the only known example of a metric space that has Enflo type 2 but fails to have nontrivial Markov type (observe that the notions of Enflo type 2 and BMW type 2 coincide).

Lemma 4.2. *Fix $p\in[1,\infty)$ and $m,n\in\mathbb{N}$. Let (X,d_X) be a metric space and $A=(a_{ij})\in M_n(\mathbb{R})$ be a symmetric stochastic matrix. Then*

$$\gamma(A,d_X^p)\leqslant M_p(X;m)^p m\gamma(A^m,d_X^p)\,.$$

Proof. This is immediate from the definitions: for every $x_1,\ldots,x_n\in X$ we have

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p &\stackrel{(3)}{\leqslant} \frac{\gamma(A^m,d_X^p)}{n}\sum_{i=1}^n\sum_{j=1}^n (A^m)_{ij}d_X(x_i,x_j)^p\\ &\stackrel{(61)}{\leqslant} \frac{M_p(X;m)^p m\gamma(A^m,d_X^p)}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^p.\end{aligned}$$

$\square$

The modulus of uniform smoothness of a Banach space $(X, \|\cdot\|_X)$ is defined for $\tau \in (0,\infty)$ as

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

X is said to be uniformly smooth if $\lim_{\tau\to 0} \rho_X(\tau)/\tau = 0$. Furthermore, X is said to have modulus of smoothness of power type $q \in (0,\infty)$ if there exists a constant $C \in (0,\infty)$ such that $\rho_X(\tau) \leqslant C\tau^q$ for all $\tau \in (0,\infty)$. It is straightforward to check that in this case necessarily $q \in [1,2]$. It is shown in [BCL94] that X has modulus of smoothness of power type q if and only if there exists a constant $S \in [1,\infty)$ such that for every $x, y \in X$

$$\frac{\|x+y\|_X^q + \|x-y\|_X^q}{2} \leqslant \|x\|_X^q + S^q\|y\|_X^q. \tag{62}$$

The infimum over those $S \in [1,\infty)$ for which (62) holds true is called the q-smoothness constant of X, and is denoted $S_q(X)$. Observe that every Banach space satisfies $S_1(X) = 1$.

The following theorem is due to [NPSS06].

Theorem 4.3. *Fix $q \in [1,2]$ and $p \in [q,\infty)$. Suppose that $(X, \|\cdot\|_X)$ is a Banach space whose modulus of smoothness has power type q. Then*

$$\forall\, m \in \mathbb{N}, \qquad M_p(X;m) \lesssim \left(p^{\frac{1}{q}} + S_q(X)\right) m^{\frac{1}{q}-\frac{1}{p}}. \tag{63}$$

The statement corresponding to Theorem 4.3 in [NPSS06] (specifically, see Theorem 4.4 there), allows for a multiplicative constant with unspecified dependence on p and q, while in (63) we stated an explicit dependence on these parameters that will serve us later on several occasions. We shall therefore proceed to sketch the proof of Theorem 4.3 so as to explain why the dependence on p and q in (63) is indeed valid.

Proof of Theorem 4.3 (sketch). For every measure space (Ω,μ) we have

$$S_q\left(L_p(\mu, X)\right) \lesssim p^{\frac{1}{q}} + S_q(X). \tag{64}$$

The case $q = 2$ of (64) appears in [Nao12b], and the proof for general $q \in [1,2]$ follows mutatis mutandis from the proof in [Nao12b]. This has been carried out explicitly in Lemma 6.3 of [MN12], whose statement asserts the weaker bound $S_q\left(L_p(\mu, X)\right) \lesssim p^{1/q} S_q(X)$, but the proof of [MN12, Lem. 6.3] without any change whatsoever yields (64).

As explained in the proof of Theorem 4.4 in [NPSS06], by a result of [Lin63] (see also [Pis75, Prop. 2.2]) it follows from (64) that every

X-valued martingale $\{M_k\}_{k=0}^m$ satisfies

$$\begin{aligned}&\mathbb{E}\left[\|M_m-M_0\|_X^p\right]\\ &\leqslant K^p\left(p^{\frac{p}{q}}+S_q(X)^p\right)\left(\sum_{k=1}^m\left(\mathbb{E}\left[\|M_k-M_{k-1}\|_X^p\right]\right)^{\frac{q}{p}}\right)^{\frac{p}{q}}\\ &\leqslant K^p\left(p^{\frac{p}{q}}+S_q(X)^p\right)m^{\frac{p}{q}-1}\sum_{k=1}^m\mathbb{E}\left[\|M_k-M_{k-1}\|_X^p\right],\end{aligned}\tag{65}$$

where $K\in(0,\infty)$ is a universal constant. Now, substituting the martingale inequality (65) into the proof of Theorem 2.3 in [NPSS06], in place of the use of Pisier's martingale inequality [Pis75], yields (63). □

We record for future use the following corollary, which is an immediate consequence of Lemma 4.2 and Theorem 4.3.

Corollary 4.4. *Fix $q\in(1,2]$ and $p\in[q,\infty)$. Suppose that $(X,\|\cdot\|_X)$ is a Banach space whose modulus of smoothness has power type q. Then for every $m,n\in\mathbb{N}$ and every symmetric stochastic $A=(a_{ij})\in M_n(\mathbb{R})$,*

$$\gamma(A,\|\cdot\|_X^p)\lesssim C^p\left(p^{\frac{p}{q}}+S_q(X)^p\right)m^{\frac{p}{q}}\gamma(A^m,\|\cdot\|_X^p),\tag{66}$$

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

We refer to [Cal64] for the background on complex interpolation that is used below. We also recall the definition of $\lambda(A)$ in (2). The following theorem is the main result of this section.

Theorem 4.5. *Let (H,Z) be a compatible pair of Banach spaces with H being a Hilbert space. Suppose that $\theta\in[0,1]$ and consider the complex interpolation space $X=[H,Z]_\theta$. Fix $q\in[1,2]$ and suppose that X has modulus of smoothness of power type q. Then for every $n\in\mathbb{N}$ and every n by n symmetric stochastic matrix $A\in M_n(\mathbb{R})$ we have*

$$\gamma\big(A,\|\cdot\|_X^2\big)\lesssim\frac{S_q(X)^2}{\left(1-\lambda(A)^\theta\right)^{2/q}}.\tag{67}$$

Before proving Theorem 4.5 we present some of its immediate corollaries. First, since in the setting of Theorem 4.5 we always have $S_q(X)\lesssim 1$ for $q=2/(1+\theta)$ (see [Pis79, CR82]), the following corollary is a special case of Theorem 4.5.

Corollary 4.6. *Under the assumptions of Theorem 4.5 we have*

$$\gamma\big(A,\|\cdot\|_X^2\big)\lesssim\frac{1}{\left(1-\lambda(A)^\theta\right)^{1+\theta}}.$$

In order for the above results to fit into the framework of Question 1.2, we need to bound $\gamma(A, \|\cdot\|_X^2)$ in terms of $\lambda_2(A)$ rather than $\lambda(A)$. This is the content of the next corollary.

Corollary 4.7. *Under the assumptions of Theorem 4.5 we have*

$$\gamma\big(A, \|\cdot\|_X^2\big) \lesssim \frac{S_q(X)^2}{\theta^{2/q}} \cdot \frac{1}{(1-\lambda_2(A))^{2/q}}. \tag{68}$$

Proof. Since A is symmetric and stochastic, all of its eigenvalues are in the interval $[-1,1]$. Consequently, all the eigenvalues of the symmetric stochastic matrix $(I+A)/2$ are nonnegative, and hence

$$\lambda\left(\frac{I+A}{2}\right) = \frac{1+\lambda_2(A)}{2}. \tag{69}$$

An application of Theorem 4.5 to the matrix $(I+A)/2$ while taking into account the identities (69) and (39) implies that

$$\gamma(A, \|\cdot\|_X^2) \lesssim \frac{S_q(X)^2}{\left(1-\left(\frac{1+\lambda_2(A)}{2}\right)^{\theta}\right)^{2/q}}.$$

This yields the desired estimate (68) due to the elementary inequality

$$\forall\, x \in [-1,1], \qquad 1-\left(\frac{1+x}{2}\right)^{\theta} \geqslant \frac{\theta(1-x)}{2}. \qquad \square$$

We can now complete the proof of Theorem 1.5, and consequently also its corresponding dual statement Corollary 1.6.

Proof of Theorem 1.5. Fix $p \in [2,\infty)$ and apply Corollary 4.7 when $X = \ell_p$. Then $X = [\ell_2, \ell_\infty]_\theta$ for $\theta = 2/p$. Moreover, by [Fig76, BCL94] we have $S_2(\ell_p) = \sqrt{p-1}$, and therefore the desired estimate (12) follows from (68) (with $q=2$). $\square$

As in the above proof of Theorem 1.5, by specializing Theorem 4.5 to $X = \ell_p$ for $p \in [2,\infty)$, we obtain the following corollary, which was stated in the Introduction as inequality (18).

Corollary 4.8. *For every $p \in [2,\infty)$, every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix A we have*

$$\gamma\Big(A, \|\cdot\|_{\ell_p}^2\Big) \lesssim \frac{p}{1-\lambda(A)^{2/p}}.$$

We now proceed to prove Theorem 4.5.

Proof of Theorem 4.5. In what follows, given a Banach space $(Y, \|\cdot\|_Y)$ and $n \in \mathbb{N}$ we let $L_2^n(Y)$ denote the Banach space whose underlying vector space is Y^n, equipped with the norm

$$\forall\, y = (y_1, \ldots, y_n) \in Y^n, \quad \|y\|_{L_2^n(Y)} \stackrel{\mathrm{def}}{=} \left(\frac{1}{n}\sum_{i=1}^n \|y_i\|_Y^2\right)^{\frac{1}{2}}.$$

If H is a Hilbert space with scalar product $\langle \cdot, \cdot \rangle_H$, then $L_2^n(H)$ is a Hilbert space whose scalar product is always understood to be given by $\langle x, y\rangle_{L_2^n(H)} = \frac{1}{n}\sum_{i=1}^n \langle x_i, y_i\rangle_H$ for every $x, y \in H$.

Let $e_1(A), \ldots, e_n(A) \in L_2^n(\mathbb{R})$ be an orthonormal eigenbasis of A with $e_1(A) = (1, \ldots, 1)$ and $Ae_i(A) = \lambda_i(A)e_i(A)$ for all $i \in \{1, \ldots, n\}$. Define an operator $T : L_2^n(\mathbb{R}) \to L_2^n(\mathbb{R})$ by setting for every $x \in \mathbb{R}^n$ and $i \in \{1, \ldots, n\}$,

$$(Tx)_i \stackrel{\mathrm{def}}{=} \sum_{j=1}^n a_{ij}\left(x_j - \frac{1}{n}\sum_{k=1}^n x_k\right).$$

Equivalently,

$$Tx = \sum_{i=2}^n \lambda_i(A)\langle x, e_i(A)\rangle_{L_2^n(\mathbb{R})} e_i(A). \tag{70}$$

The operator $T \otimes I_Y : L_2^n(Y) \to L_2^n(Y)$, where I_Y denotes the identity on Y, is then given by

$$\forall\, i \in \{1, \ldots, n\}, \qquad ((T \otimes I_Y)y)_i \stackrel{\mathrm{def}}{=} \sum_{j=1}^n a_{ij}\left(y_j - \frac{1}{n}\sum_{k=1}^n y_k\right). \tag{71}$$

Recalling (2) and (70), we have the following operator norm bounds.

$$\|T \otimes I_H\|_{L_2^n(H)} = \|T\|_{L_2^n(\mathbb{R}) \to L_2^n(\mathbb{R})} = \lambda(A). \tag{72}$$

The norm of $T \otimes I_Z : L_2^n(Z) \to L_2^n(Z)$ can be bounded crudely by using the fact that A is a symmetric stochastic matrix. Indeed, for every $z \in L_2^n(Z)$ we have

$$\begin{aligned}
\|Tz\|_{L_2^n(Z)}^2 &= \frac{1}{n}\sum_{i=1}^n \left\|\sum_{j=1}^n a_{ij}\left(z_j - \frac{1}{n}\sum_{k=1}^n z_k\right)\right\|_Z^2 \\
&\leqslant \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n a_{ij}\|z_j - z_k\|_Z^2 \\
&\leqslant \frac{2}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n a_{ij}\left(\|z_j\|_Z^2 + \|z_k\|_Z^2\right) = 4\|z\|_{L_2^n(Z)}^2.
\end{aligned} \tag{73}$$

An interpolation of (72) and (73) (see [Cal64]) shows that

$$\|T\otimes I_X\|_{L_2^n(X)\to L_2^n(X)}\leqslant 2^{1-\theta}\lambda(A)^\theta. \tag{74}$$

For every $x\in L_2^n(X)$ let $\overline{x}\in L_2^n(X)$ be the vector whose ith coordinate equals $x_i-S(x)$, where $S(x)=\frac{1}{n}\sum_{k=1}^n x_k$. Then

$$\begin{aligned}\left\|(I_{L_2^n(X)}-T\otimes I_X)\overline{x}\right\|_{L_2^n(X)}^2 &\stackrel{(71)}{=} \frac{1}{n}\sum_{i=1}^n\left\|\sum_{j=1}^n a_{ij}(x_i-x_j)\right\|_X^2\\ &\leqslant \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|x_i-x_j\|_X^2. \end{aligned}\tag{75}$$

At the same time,

$$\begin{aligned}\left\|(I_{L_2^n(X)}-T\otimes I_X)\overline{x}\right\|_{L_2^n(X)} &\geqslant \|\overline{x}\|_{L_2^n(X)}-\|(T\otimes I_X)\overline{x}\|_{L_2^n(X)}\\ &\stackrel{(74)}{\geqslant} \left(1-2^{1-\theta}\lambda(A)^\theta\right)\|\overline{x}\|_{L_2^n(X)}. \end{aligned}\tag{76}$$

Hence, if we suppose that

$$2^{1-\theta}\lambda(A)^\theta<1 \iff \lambda(A)<\frac{1}{2^{(1-\theta)/\theta}}, \tag{77}$$

then

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|_X^2 &\leqslant \frac{2}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\|x_i-S(x)\|_X^2+\|x_j-S(x)\|_X^2\right)\\ =4\|\overline{x}\|_{L_2^n(X)}^2 &\stackrel{(75)\wedge(76)}{\leqslant} \frac{4}{\left(1-2^{1-\theta}\lambda(A)^\theta\right)^2}\cdot\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|x_i-x_j\|_X^2.\end{aligned}$$

Equivalently,

$$\gamma\big(A,\|\cdot\|_X^2\big)\leqslant\frac{4}{\left(1-2^{1-\theta}\lambda(A)^\theta\right)^2}. \tag{78}$$

We shall now apply a trick that was used by Pisier in [Pis10], where it is attributed to V. Lafforgue: we can ensure that the condition (77) holds true if we work with a large enough power of A. We will then be able to return back to an inequality that involves A rather than its power by using Markov type through Corollary 4.4. Specifically, define

$$m\stackrel{\mathrm{def}}{=}\left\lceil\frac{(2-\theta)\log 2}{\theta\log(1/\lambda(A))}\right\rceil. \tag{79}$$

This choice of m ensures that

$$2^{1-\theta}\lambda(A^m)^\theta=2^{1-\theta}\lambda(A)^{m\theta}\leqslant\frac{1}{2},$$

so we may apply (78) with A replaced by A^m to get the estimate

$$\gamma(A^m,\|\cdot\|_X^2)\leqslant 16. \tag{80}$$

An application of Corollary 4.4 with $p=2$ now implies that

$$\begin{aligned}\gamma(A,\|\cdot\|_X^2) &\overset{(66)\wedge(80)}{\lesssim} m^{\frac{2}{q}}S_q(X)^2\\ &\overset{(79)}{\lesssim} S_q(X)^2\left(1+\frac{1}{\theta\log(1/\lambda(A))}\right)^{\frac{2}{q}}\lesssim\frac{S_q(X)^2}{(1-\lambda(A)^\theta)^{2/q}},\end{aligned}\tag{81}$$

where in (81) we used the elementary inequality

$$\forall\, x\in[0,1],\qquad 1+\frac{1}{\log(1/x)}\leqslant\frac{2}{1-x},$$

which holds true because $\exp\left(-\frac{1-x}{1+x}\right)\geqslant 1-\frac{1-x}{1+x}=\frac{2x}{1+x}\geqslant x$. $\square$

4.1. **Ramanujan graphs and Alon–Roichman graphs.** Given a connected n-vertex graph $G=(\{1,\ldots,n\},E_G)$, let $d_G(\cdot,\cdot)$ denote the shortest path metric that G induces on $\{1,\ldots,n\}$. The diameter of the metric space $(\{1,\ldots,n\},d_G)$ will be denoted below by $\mathrm{diam}(G)$. Suppose that $d\in\{3,\ldots,n-1\}$ and that G is d-regular, i.e., for every $i\in\{1,\ldots,n\}$ the number of $j\in\{1,\ldots,n\}$ such that $\{i,j\}\in E_G$ equals d. The normalized adjacency matrix of G will be denoted A_G, i.e., $(A_G)_{ij}=\frac{1}{d}\mathbf{1}_{\{i,j\}\in E_G}$ for every $i,j\in\{1,\ldots,n\}$. Thus A_G is an n by n symmetric stochastic matrix. We denote $\lambda_i(G)=\lambda_i(A_G)$ for every $i\in\{1,\ldots,n\}$, and we correspondingly set $\lambda(G)=\lambda(A_G)$ and $\gamma(G,d_X^q)=\gamma(A_G,d_X^q)$ for every metric space (X,d_X) and $q\in[1,\infty)$.

Setting $c_X(G)\stackrel{\mathrm{def}}{=}c_X(\{1,\ldots,n\},d_G)$, an important idea of Linial, London and Rabinovich [LLR95] relates $\gamma(G,d_X^q)$ to a lower bound on $c_X(G)$ as follows. Let $f:\{1,\ldots,n\}\to X$ be a nonconstant function and apply (3) with $A=A_G$ and $x_i=f(i)$ for every $i\in\{1,\ldots,n\}$, thus obtaining the estimate

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(f(i),f(j))^q &\leqslant \frac{\gamma(G,d_X^q)}{dn}\sum_{\substack{(i,j)\in\{1,\ldots,n\}^2\\ \{i,j\}\in E_G}} d_X(f(i),f(j))^q\\ &\leqslant \gamma(G,d_X^q)\|f\|_{\mathrm{Lip}}^q.\end{aligned}\tag{82}$$

Denoting $Av_X^{(q)}(G)\stackrel{\mathrm{def}}{=}Av_X^{(q)}(\{1,\ldots,n\},d_G)$, it follows from (82) that

$$c_X(G)\geqslant Av_X^{(q)}(G)\geqslant\frac{1}{\gamma(G,d_X^q)^{1/q}}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_G(i,j)^q\right)^{\frac{1}{q}}. \tag{83}$$

For notational simplicity we will assume from now on that G is a vertex-transitive graph, since in this case we have

$$\frac{\operatorname{diam}(G)}{2^{1+1/q}} \leqslant \left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_G(i,j)^q\right)^{\frac{1}{q}} \leqslant \operatorname{diam}(G). \tag{84}$$

One verifies the validity of (84) by arguing as in the proof of Proposition 3.4 of [NR09]: for every $i \in \{1,\ldots,n\}$ and $r \in (0,\infty)$ let $B_G(i,r) = \{j \in \{1,\ldots,n\}:\ d_G(i,j) \leqslant r\}$ be the closed ball of radius r centered at i in the metric d_G. Since G is vertex-transitive, the cardinality of $B_G(i,r)$ is independent of i. Hence, if we let r_* be the minimum $r \in \mathbb{N}$ such that $|B_G(i,r)| > n/2$ for every $i \in \{1,\ldots,n\}$, then for every $i \in \{1,\ldots,n\}$ we have $|\{1,\ldots,n\}\smallsetminus B_G(i,r_*-1)| \geqslant n/2$. In other words, for every $i \in \{1,\ldots,n\}$ there are at least $n/2$ vertices $j \in \{1,\ldots,n\}$ with $d_G(i,j) \geqslant r_*$. Hence $\sum_{i=1}^n\sum_{j=1}^n d_G(i,j)^q \geqslant n^2r_*^q/2$. At the same time, by the definition of r_* we have $B_G(i,r_*)\cap B_G(j,r_*) \neq \emptyset$ for every $i,j \in \{1,\ldots,n\}$, and therefore $\operatorname{diam}(G) \leqslant 2r_*$. This proves the leftmost inequality in (84) (the remaining inequality in (84) is trivial).

By combining (83) and (84), Matoušek's argument in [Mat97] deduces from his bound (14) that if $G = (\{1,\ldots,n\},E_G)$ is a vertex-transitive graph such that $\lambda_2(G)$ is bounded away from 1 by a universal constant then for every $p \in [2,\infty)$ we have

$$c_p(G) \gtrsim \frac{\operatorname{diam}(G)}{p}. \tag{85}$$

Denote $p(G) \stackrel{\mathrm{def}}{=} p(\{1,\ldots,n\},d_G)$, where we recall that in Section 1.1 we defined for a separable metric space (X,d_X) the quantity $p(X,d_X)$ (or simply $p(X)$ if the metric is clear from the context) to be the infimum over those $p \in [2,\infty]$ for which $c_p(X) \leqslant 10$. It follows from (85) that $p(G) \gtrsim \operatorname{diam}(G)$ (still under the assumption that $\lambda_2(G)$ is bounded away from 1). Using Corollary 4.8, we now show that it is possible to improve over this estimate.

Proposition 4.9. *Fix $p \in [2,\infty)$ and let $G = (\{1,\ldots,n\},E_G)$ be a vertex-transitive graph. Then*

$$p \leqslant \log\left(\frac{1}{\lambda(G)}\right) \implies c_p(G) \gtrsim \frac{\operatorname{diam}(G)}{\sqrt{p}},$$

and

$$p \geqslant \log\left(\frac{1}{\lambda(G)}\right) \implies c_p(G) \gtrsim \frac{\operatorname{diam}(G)}{p}\sqrt{\log\left(\frac{1}{\lambda(G)}\right)}.$$

Proof. By combining (83) and (84) (for $q = 2$) with Corollary 4.8 we see that
$$c_p(G) \geqslant Av_X^{(2)}(G) \gtrsim \frac{\sqrt{1-\lambda(G)^{2/p}}}{\sqrt{p}}\,\mathrm{diam}(G). \qquad \square$$

By the definition of $p(G)$, the following corollary is a formal consequence of Proposition 4.9.

Corollary 4.10. *Let $G = (\{1,\ldots,n\}, E_G)$ be a vertex-transitive graph. Then*
$$\lambda(G) \leqslant e^{-\mathrm{diam}(G)^2} \implies p(G) \gtrsim \mathrm{diam}(G)^2,$$
and
$$\lambda(G) \geqslant e^{-\mathrm{diam}(G)^2} \implies p(G) \gtrsim \mathrm{diam}(G)\sqrt{\log\left(\frac{1}{\lambda(G)}\right)}.$$

For every $n \in \mathbb{N}$ and $d \in \{3,\ldots,n-1\}$, if $G = (\{1,\ldots,n\}, E_G)$ is a d-regular graph then by [Chu89] we have
$$\frac{\log n}{\log d} \lesssim \mathrm{diam}(G) \leqslant 1 + \frac{\log n}{\log(1/\lambda(G))}.$$
Consequently, if $1/\lambda(G)$ is at least d^c for some universal constant $c > 0$ then $\mathrm{diam}(G) \asymp \log_d n$. This happens in particular when G is a Ramanujan graph, i.e., $\lambda(G) \leqslant 2\sqrt{d-1}/d$. Such graphs have been constructed in [LPS88, Mar88]. It is natural to ask for the asymptotic evaluation of $c_p(G)$ when G is an n-vertex d-regular Ramanujan graph. While this question remains open, Proposition 4.11 below contains a lower bound on $c_p(G)$ that improves over Matoušek's bound.

Proposition 4.11. *Fix $n \in \mathbb{N}$ and $d \in \{3,\ldots,n-1\}$. Suppose that $G = (\{1,\ldots,n\}, E_G)$ is a Ramanujan graph. Then*
$$2 \leqslant p \leqslant \log d \implies c_p(G) \gtrsim \frac{\log n}{\sqrt{p}\log d}. \tag{86}$$
$$p \geqslant \log d \implies c_p(G) \gtrsim \frac{\log n}{p\sqrt{\log d}}. \tag{87}$$

Proof. By combining (83) and (84) with Corollary 4.8 we see that
$$\begin{aligned} c_p(G) &\gtrsim \frac{\sqrt{1-\lambda(G)^{2/p}}}{\sqrt{p}}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_G(i,j)^2\right)^{\frac12} \\ &\gtrsim \frac{\sqrt{1-\left(2/\sqrt{d}\right)^{2/p}}}{\sqrt{p}}\log_d n, \end{aligned}$$

where we used the fact that $\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_G(i,j)^2 \gtrsim (\log_d n)^2$ and that $\lambda(G) \leqslant 2/\sqrt{d}$. The latter bound uses the fact that G is a Ramanujan graph (in fact, weaker bounds on $\lambda(G)$ suffice for our purposes), and the former bound holds true for any connected n-vertex d-regular graph (see [Mat97] for a simple proof of this). □

If G is a uniformly random n-vertex d-regular graph then by [BS87] $\lambda(G) \leqslant 2/\sqrt[4]{d}$ with high probability (for the best known bound on $\lambda(G)$ when G is a random d-regular graph, see [Fri08]). By arguing identically to the proof of Proposition 4.11, we see that with high probability (86) and (87) hold true for such G, implying Proposition 1.9.

Corollary 4.8 also implies new distortion bounds for Abelian Alon–Roichman graphs [AR94]. These are graphs that are obtained from the following random construction. Let Γ be a finite Abelian group, and think of Γ as the set $\{1,\dots,n\}$, equipped with an Abelian group operation. Fix $\varepsilon \in (0,1/2)$ and set $k = \left\lceil \frac{3}{\varepsilon^2}\log n\right\rceil$. Let $g_1,\dots,g_k \in \Gamma$ be chosen independently and uniformly at random. This induces a random Cayley graph G whose generating multi-set is $\{g_1, g_1^{-1},\dots,g_k,g_k^{-1}\}$. As explained in [AR94], with probability that tends to 1 as $n\to\infty$ the graph G is connected. Note that since G is a Cayley graph it is vertex-transitive. It follows from [CM08] that provided n is large enough we have $\lambda(G) \leqslant \varepsilon$ with probability at least $\frac12$. Moreover, by [NR09, Prop. 3.5] we have $\operatorname{diam}(G) \gtrsim (\log n)/(\log 1/\varepsilon)$. A substitution of these estimates into Proposition 4.9 shows that

$$2 \leqslant p \leqslant \log(1/\varepsilon) \implies c_p(G) \gtrsim \frac{\log n}{\sqrt{p}\log(1/\varepsilon)},$$

and

$$p \geqslant \log(1/\varepsilon) \implies c_p(G) \gtrsim \frac{\log n}{p\sqrt{\log(1/\varepsilon)}}.$$

Remark 4.12. We warn that there is some subtlety in the definition of the parameter $p(X)$ for a separable metric space (X,d_X). Given that X is isometric to a subset of ℓ_∞, it is indeed natural to ask for the smallest $p\in[2,\infty]$ such that X embeds with bounded distortion, say, distortion 10, into ℓ_p. As an example of an application that was shown to us by Yuval Rabani, one can use the methods of [KOR00] to prove that subsets of ℓ_p admit an efficient approximate nearest neighbor data structure with approximation guarantee $e^{O(p)}$, so the parameter $p(X)$ relates to approximate nearest neighbor search in X (it would be very interesting to determine the correct asymptotic dependence on p here). But, understanding the set of $p\in[2,\infty)$ for which X admits a bi-Lipschitz embedding into ℓ_p can be subtle. In particular, it is not true that if X embeds into ℓ_p then for every $q>p$ it also embeds into ℓ_q. In fact, we have the following estimates for every $n\in\mathbb{N}$ and $p,q\in(2,\infty)$.

$$2<q<p \implies c_q\left(\ell_p^n\right)\asymp n^{\frac{1}{q}-\frac{1}{p}}, \tag{88}$$

and

$$2<p<q \implies c_q\left(\ell_p^n\right)\asymp_{p,q} n^{\frac{(q-p)(p-2)}{(q-2)p^2}}. \tag{89}$$

The asymptotic identity (88) is a standard consequence of the fact that L_q has Rademacher cotype q (see e.g. [Woj91]). The remarkable asymptotic identity (89) is due to [FJS88] (using a computation of [GPP80]). The implicit dependence on p,q in (89) is unknown, and it would be of interest to evaluate it up to a universal constant factor. Observe that the exponent of n in (89) tends to $(p-2)/p^2>0$ as $q\to\infty$, and therefore the implicit constant in (89) must tend to 0 as $q\to\infty$.

4.2. **Curved Banach spaces in the sense of Pisier.** Motivated by his work on nonlinear spectral gaps [Laf08], V. Lafforgue associated the following modulus to a Banach space $(X,\|\cdot\|_X)$, a modulus has been investigated extensively by Pisier in [Pis10]. Given $\varepsilon\in(0,\infty)$ let $\Delta_X(\varepsilon)$ denote the infimum over those $\Delta\in(0,\infty)$ such that for every $n\in\mathbb{N}$, every matrix $T=(t_{ij})\in M_n(\mathbb{R})$ with

$$\|T\|_{L_2^n(\mathbb{R})\to L_2^n(\mathbb{R})}\leqslant\varepsilon \qquad\text{and}\qquad \|\mathrm{abs}(T)\|_{L_2^n(\mathbb{R})\to L_2^n(\mathbb{R})}\leqslant 1,$$

where $\mathrm{abs}(T)\stackrel{\mathrm{def}}{=}(|t_{ij}|)$ is the entry-wise absolute value of T, satisfies

$$\|T\otimes I_X\|_{L_2^n(X)\to L_2^n(X)}\leqslant\Delta.$$

Pisier introduced the following terminology in [Pis10]: X is said to be *curved* if $\Delta_X(\varepsilon)<1$ for some $\varepsilon\in(0,1)$. X is said to be *fully curved* if $\Delta_X(\varepsilon)<1$ for all $\varepsilon\in(0,1)$, and X is said to be *uniformly curved* if $\lim_{\varepsilon\to0}\Delta_X(\varepsilon)=0$. It is shown in [Pis10] that if X is either fully curved or uniformly curved then it admits an equivalent uniformly convex norm. A remarkable characterization of Pisier [Pis10] shows that $\Delta_X(\varepsilon)\lesssim\varepsilon^\alpha$ for some $\alpha\in(0,\infty)$ if and only if X arises from complex interpolation with Hilbert space: formally, this happens if and only if X is isomorphic to a quotient of a subspace of an ultraproduct of θ-Hilbertian Banach spaces for some $\theta\in(0,1)$; we refer to [Pis10] for the definition of these notions. A more complicated structural characterization of uniformly curved spaces (based on real interpolation) is also obtained in [Pis10].

One can use the above notions to give a generalized abstract treatment of results in the spirit of Theorem 4.5. Fix $\varepsilon\in(0,1)$ and suppose that $\Delta_X(\varepsilon)<\frac12$. Let $A\in M_n(\mathbb{R})$ be symmetric and stochastic and let $T=(t_{ij})\in M_n(\mathbb{R})$ be given as in (70). By (72) we have $\|T\|_{L_2^n(\mathbb{R})\to L_2^n(\mathbb{R})} = \lambda(A)$. Moreover, since $\mathrm{abs}(T) = (|a_{ij} - 1/n|)$ and A is symmetric and stochastic, it is immediate to check that $\|\mathrm{abs}(T)\|_{L_2^n(\mathbb{R})\to L_2^n(\mathbb{R})} \leqslant 2$. By the definition of the modulus $\Delta_X(\cdot)$ we therefore have $\|T\otimes I_X\|_{L_2^n(X)\to L_2^n(X)} \leqslant 2\Delta_X(\lambda(A)/2)$. Define

$$m \stackrel{\mathrm{def}}{=} \left\lceil \frac{\log(1/(2\varepsilon))}{\log(1/\lambda(A))} \right\rceil, \tag{90}$$

so that $\lambda(A^m)/2 = \lambda(A)^m/2 \leqslant \varepsilon$, and apply the above reasoning with A replaced by A^m. Arguing as in (75) and (76), we obtain the estimate

$$\gamma\big(A^m, \|\cdot\|_X^2\big) \leqslant \frac{4}{\left(1-2\Delta_X(\varepsilon)\right)^2}.$$

We can now use the notion of Markov type through Lemma 4.2 to deduce the following statement.

Theorem 4.13. *Fix $\varepsilon \in (0,1)$ and let $(X, \|\cdot\|_X)$ be a Banach that satisfies $\Delta_X(\varepsilon) < \frac{1}{2}$. Then for every $n \in \mathbb{N}$ and every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ we have*

$$\gamma\big(A, \|\cdot\|_X^2\big) \leqslant \frac{4}{\left(1-2\Delta_X(\varepsilon)\right)^2} \cdot M_2\left(X; \left\lceil \frac{\log(1/(2\varepsilon))}{\log(1/\lambda(A))} \right\rceil\right)^2 \cdot \left\lceil \frac{\log(1/(2\varepsilon))}{\log(1/\lambda(A))} \right\rceil.$$

In particular, using Theorem 4.3 and arguing as in the proof of Corollary 4.7, if X has modulus of smoothness of power type 2 then every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ satisfies

$$\gamma\big(A, \|\cdot\|_X^2\big) \lesssim_X \frac{1}{1-\lambda_2(A)}.$$

In conjunction with Theorem 1.3 we deduce the following geometric embedding result for uniformly curved Banach space that can be renormed so as to have modulus of smoothness of power type 2.

Corollary 4.14. *Suppose that $(X, \|\cdot\|_X)$ is a uniformly curved Banach space that admits an equivalent norm whose modulus of uniform smoothness has power type 2. Then for every $x_1, \ldots, x_n \in X$ there exist $y_1, \ldots, y_n \in \ell_2$ such that*

$$\sum_{i=1}^n \sum_{j=1}^n \|x_i - x_j\|_X^2 = \sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|_2^2,$$

and

$$\forall\, i,j \in \{1,\ldots,n\}, \qquad \|y_i - y_j\|_2 \lesssim_X \|x_i - x_j\|_X.$$

4.3. **An interpolation inequality for nonlinear spectral gaps.** The *modulus of uniform convexity* of a Banach space $(X, \|\cdot\|_X)$ is defined for $\varepsilon \in [0,2]$ as

$$\delta_X(\varepsilon) \stackrel{\mathrm{def}}{=} \inf\left\{1-\frac{\|x+y\|_X}{2}:\ x,y\in B_X\ \wedge\ \|x-y\|_X=\varepsilon\right\}.$$

X is said to be *uniformly convex* if $\delta_X(\varepsilon) > 0$ for all $\varepsilon \in (0,2]$. Furthermore, X is said to have modulus of convexity of power type p if there exists a constant $c \in (0,\infty)$ such that $\delta_X(\varepsilon) \geqslant c\,\varepsilon^p$ for all $\varepsilon \in [0,2]$. It is straightforward to check that in this case necessarily $p \geqslant 2$. By Proposition 7 in [BCL94] (see also [Fig76]), X has modulus of convexity of power type p if and only if there exists a constant $K \in [1,\infty)$ such that for every $x, y \in X$

$$\left(\|x\|_X^p+\frac{1}{K^p}\|y\|_X^p\right)^{\frac1p}\leqslant\left(\frac{\|x+y\|_X^p+\|x-y\|_X^p}{2}\right)^{\frac1p}. \tag{91}$$

The infimum over those K for which (91) holds true is called the p-convexity constant of X, and is denoted $K_p(X)$. Note that every Banach space satisfies $K_\infty(X) = 1$. Below we shall use the convention $K_p(X) = \infty$ if $p \in [1,2)$. For $1 \leqslant q \leqslant 2 \leqslant p$, the p-convexity constant of X is related to the q-smoothness constant of X (recall (62)) via the following duality relation [BCL94, Lem. 5].

$$\frac1p+\frac1q=1 \implies K_p(X)=S_q(X^*).$$

Theorem 4.15. *Let (X, Y) be a compatible pair of Banach spaces. Fix $\theta \in [0,1]$ and consider the complex interpolation space $Z = [X, Y]_\theta$. Fix also $p, q \in [1,\infty]$ and $r \in [1,2]$. Then for every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix A we have*

$$\frac{\gamma(A,\|\cdot\|_Z^s)}{c^s\left(s^{\frac sr}+S_r(Z)^s\right)} \leqslant\left(\min\left\{\frac{(9K_p(X))^p\gamma(A,\|\cdot\|_X^p)}{\theta},\frac{(9K_q(Y))^q\gamma(A,\|\cdot\|_Y^q)}{1-\theta}\right\}\right)^{\frac sr},$$

where $c \in (0,\infty)$ is a universal constant and $s \in [2,\infty]$ is given by

$$\frac1s=\frac\theta p+\frac{1-\theta}{q}.$$

Observe that for every $a, b \in (0,\infty)$ and every $\theta \in (0,1)$ we have

$$\min\left\{\frac a\theta,\frac{b}{1-\theta}\right\}\leqslant\frac{2}{\frac\theta a+\frac{1-\theta}{b}}\leqslant a^\theta b^{1-\theta}.$$

Consequently, the conclusion of Theorem 4.15 implies that

$$\gamma(A,\|\cdot\|_Z^s)\lesssim_{X,Y,Z,s}\gamma(A,\|\cdot\|_X^p)^{\theta s/r}\gamma(A,\|\cdot\|_Y^q)^{(1-\theta)s/r}.$$

Such an estimate is in the spirit of the interpolation inequality (15), but it is insufficient for the purpose of addressing Question 1.8. Theorem 4.15 does suffice to prove Lemma 1.11, so we assume the validity of Theorem 4.15 for the moment and proceed now to prove Lemma 1.11.

Proof of Lemma 1.11. Matoušek proved in [Mat97] that if (X,d_X) is an n-point metric space then for $p\in[2,\infty)$ we have

$$c_p(X)\lesssim 1+\frac{\log n}{p}. \tag{92}$$

The case $p=2$ of (92) is Bourgain's embedding theorem [Bou85]. Now, for every $n\in\mathbb{N}$ let $G_n=(\{1,\ldots,n\},E_{G_n})$ be a 4-regular graph with $\sup_{n\in\mathbb{N}}\lambda_2(G_n)<1$, i.e., $\{G_n\}_{n=1}^\infty$ forms an expander sequence. Fixing $n\geqslant e^p$, by (92) we know that there exist $x_1,\ldots,x_n\in\ell_p$ such that

$$\forall\, i,j\in\{1,\ldots,n\},\qquad d_{G_n}(i,j)\leqslant\|x_i-x_j\|_{\ell_p}\lesssim\frac{\log n}{p}d_{G_n}(i,j). \tag{93}$$

Suppose that $y_1,\ldots,y_n\in\ell_q$ satisfy (24), i.e.,

$$\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|y_i-y_j\|_{\ell_q}^r\right)^{\frac1r}=\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|_{\ell_p}^s\right)^{\frac1s}, \tag{94}$$

If $\|y_i-y_j\|_{\ell_q}\leqslant D\|x_i-x_j\|_{\ell_p}$ for every $i,j\in\{1,\ldots,n\}$ then we need to show that $D\gtrsim p/(q+r)$. Note that since G_n is 4-regular, a constant fraction of the pairs $(i,j)\in\{1,\ldots,n\}^2$ satisfy $d_{G_n}(i,j)\gtrsim\log n$ (the standard argument showing this appears in e.g. [Mat97]). Hence, due to the leftmost inequality in (93), it follows from (94) that

$$\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|y_i-y_j\|_{\ell_q}^r\right)^{\frac1r}\gtrsim\log n. \tag{95}$$

Case 1. $r\leqslant q$. In this case we have

$$\begin{aligned}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|y_i-y_j\|_{\ell_q}^r\right)^{\frac1r}&\leqslant\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|y_i-y_j\|_{\ell_q}^q\right)^{\frac1q}\\&\leqslant\gamma(G_n,\|\cdot\|_{\ell_q}^q)^{\frac1q}\left(\frac{1}{4n}\sum_{\substack{(i,j)\in\{1,\ldots,n\}^2\\ \{i,j\}\in E_{G_n}}}\|y_i-y_j\|_{\ell_q}^q\right)^{\frac1q}\lesssim\frac{Dq\log n}{p},\end{aligned} \tag{96}$$

where in (96) we bounded $\gamma(G_n, \|\cdot\|_{\ell_q}^q)$ using (14), and we used the fact that if $\{i,j\} \in E_{G_n}$ then $\|y_i - y_j\|_{\ell_q} \leqslant D\|x_i - x_j\|_{\ell_p} \lesssim D(\log n)/p$, due to (93). By contrasting (96) with (95) we have $D \gtrsim p/q$, as required.

Case 2. $r > q$. Write $\theta \stackrel{\mathrm{def}}{=} 1/(r-1)$ and $t \stackrel{\mathrm{def}}{=} 2q(r-2)/(2r-q-2)$. Recalling that we are assuming in Lemma 1.11 that $q \geqslant 2$, it follows that $\theta \in (0,1)$ and $t \in [q, 2r]$. Consequently, $K_{2r}(\ell_t) \leqslant K_t(\ell_t) \leqslant 1$ (see [BCL94]). Note also that $1/q = \theta/2 + (1-\theta)/t$, so $\ell_q = [\ell_2, \ell_t]_\theta$. Since $1/r = \theta/2 + (1-\theta)/(2r)$ and $S_2(\ell_q) = \sqrt{q-1}$ (see [BCL94]), it follows from Theorem 4.15 that for every $n \in \mathbb{N}$ we have

$$\gamma(G_n, \|\cdot\|_{\ell_q}^r)^{1/r} \lesssim \sqrt{r \cdot \min\left\{r\gamma(G_n, \|\cdot\|_{\ell_2}^2), \gamma(G_n, \|\cdot\|_{\ell_t}^{2r})\right\}} \lesssim r.$$

Hence, if we argue as in (96) we see that $D \gtrsim p/r$, as required. $\square$

We now prove Theorem 4.15.

Proof of Theorem 4.15. We may assume that A is ergodic, implying that $\gamma(A, \|\cdot\|_X^p), \gamma(A, \|\cdot\|_Y^q) < \infty$, since otherwise the conclusion of Theorem 4.15 is vacuous. So, by Lemma 2.3 we have

$$\gamma_+\left(\frac{I+A}{2}, \|\cdot\|_X^p\right) \leqslant 2^{2p+1}\gamma(A, \|\cdot\|_X^p) < \infty, \tag{97}$$

and

$$\gamma_+\left(\frac{I+A}{2}, \|\cdot\|_Y^q\right) \leqslant 2^{2q+1}\gamma(A, \|\cdot\|_X^q) < \infty, \tag{98}$$

For a Banach space $(W, \|\cdot\|_W)$ and $t \in [1,\infty]$ let $L_t^n(W)_0$ be the subspace of $L_t^n(W)$ consisting of mean-zero vectors, i.e.,

$$L_t^n(W)_0 \stackrel{\mathrm{def}}{=} \left\{(w_1, \ldots, w_n) \in L_t^n(W) : \sum_{i=1}^n w_i = 0\right\}.$$

Let $Q : L_t^n(\mathbb{R}) \to L_t^n(\mathbb{R})_0$ be the canonical projection, i.e, for every $v \in L_t^n(\mathbb{R})$ and $i \in \{1, \ldots, n\}$,

$$(Qv)_i \stackrel{\mathrm{def}}{=} v_i - \frac{1}{n}\sum_{j=1}^n v_j.$$

Then by the triangle inequality, $\|Q \otimes I_W\|_{L_t^n(W) \to L_t^n(W)_0} \leqslant 2$, and consequently for every $B \in M_n(\mathbb{R})$ such that $B(L_t^n(\mathbb{R})_0) \subseteq L_t^n(\mathbb{R})_0$,

$$\|(BQ) \otimes I_W\|_{L_t^n(W) \to L_t^n(W)} \leqslant 2\|B \otimes I_W\|_{L_t^n(\mathbb{R})_0 \to L_t^n(\mathbb{R})_0}. \tag{99}$$

Note that if B is symmetric and stochastic then $B(L_t^n(\mathbb{R})_0) \subseteq L_t^n(\mathbb{R})_0$, and by [MN12, Lem. 6.6] for every $t \in [1,\infty]$ we have

$$\|B \otimes I_W\|_{L_t^n(\mathbb{R})_0 \to L_t^n(\mathbb{R})_0} \leqslant e^{-2/(t(2K_t(W))^t \gamma_+(B, \|\cdot\|_W^t))}. \tag{100}$$

(Observe that we always have $\|B\otimes I_W\|_{L_t^n(\mathbb{R})_0\to L_t^n(\mathbb{R})_0}\leqslant 1$ because B is symmetric and stochastic, so (100) is most meaningful when $t\in[2,\infty)$ and $K_t(W)<\infty$.) Consequently, for every $m\in\mathbb{N}$ we have

$$\begin{aligned}\|B^m\otimes I_W\|_{L_t^n(\mathbb{R})_0\to L_t^n(\mathbb{R})_0} &\leqslant \left(\|B\otimes I_W\|_{L_t^n(\mathbb{R})_0\to L_t^n(\mathbb{R})_0}\right)^m\\ &\leqslant e^{-2m/(t(2K_t(W))^t\gamma_+(B,\|\cdot\|_W^t))}. \qquad (101)\end{aligned}$$

Define $T\in M_n(\mathbb{R})$ by

$$T\stackrel{\mathrm{def}}{=}\left(\frac{I+A}{2}\right)^m Q.$$

An application of (101) and (99) with $W=X$ and $t=p$ while using (97) shows that

$$\|T\otimes I_X\|_{L_p^n(X)\to L_p^n(X)}\leqslant 2e^{-m/(4(9K_p(X))^p\gamma(A,\|\cdot\|_X^p))}. \qquad (102)$$

The same reasoning applied to $W=Y$ and $t=q$ while using (98) shows that

$$\|T\otimes I_Y\|_{L_q^n(Y)\to L_q^n(Y)}\leqslant 2e^{-m/(4(9K_q(Y))^q\gamma(A,\|\cdot\|_Y^q))}. \qquad (103)$$

Interpolation of (102) and (103) yields the following estimate.

$$\begin{aligned}\left\|\left(\frac{I+A}{2}\right)^m\otimes I_Z\right\|_{L_s^n(Z)_0\to L_s^n(Z)_0} &\leqslant \|T\otimes I_Y\|_{L_s^n(Z)\to L_s^n(Z)}\\ &\leqslant 2e^{-m\theta/(4(9K_p(X))^p\gamma(A,\|\cdot\|_X^p))-m(1-\theta)/(4(9K_q(Y))^q\gamma(A,\|\cdot\|_Y^q))}. \qquad (104)\end{aligned}$$

Let m be given by

$$m\stackrel{\mathrm{def}}{=}\left\lceil 7\min\left\{\frac{(9K_p(X))^p\gamma(A,\|\cdot\|_X^p)}{\theta},\frac{(9K_q(Y))^q\gamma(A,\|\cdot\|_Y^q)}{1-\theta}\right\}\right\rceil.$$

Then by (104) we have

$$\left\|\left(\frac{I+A}{2}\right)^m\otimes I_Z\right\|_{L_s^n(Z)_0\to L_s^n(Z)_0}\leqslant\frac12.$$

By [MN12, Lem. 6.1] this implies that

$$\gamma\left(\left(\frac{I+A}{2}\right)^m,\|\cdot\|_Z^s\right)\leqslant\gamma_+\left(\left(\frac{I+A}{2}\right)^m,\|\cdot\|_Z^s\right)\leqslant 9^s. \qquad (105)$$

Hence, using Corollary 4.4 we deduce that

$$
\begin{aligned}
2\gamma(A,\|\cdot\|_Z^s) &\overset{(39)}{=} \gamma\left(\frac{I+A}{2},\|\cdot\|_Z^s\right)\\
&\overset{(66)}{\lesssim} C^s\left(s^{\frac{s}{r}}+S_r(Z)^s\right)m^{\frac{s}{r}}\gamma\left(\left(\frac{I+A}{2}\right)^m,\|\cdot\|_Z^s\right)\\
&\overset{(105)}{\leqslant} (9C)^s\left(s^{\frac{s}{r}}+S_r(Z)^s\right)m^{\frac{s}{r}},
\end{aligned}
$$

where C is the universal constant from Corollary 4.4. Recalling our choice of m, this completes the proof of Theorem 4.15. □

5. Proof of Theorem 1.10

Let $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ be Banach spaces. We first note that the assumptions of Theorem 1.10 are equivalent to the assertion that B_X is uniformly homeomorphic to a subset of Y, i.e, that there is an injection $\phi : B_X \to Y$ such that both ϕ and ϕ^{-1} are uniformly continuous. Indeed, since B_X is metrically convex, the modulus of continuity of ϕ, namely the mapping $\omega_\phi : [0,\infty)\to[0,\infty)$ given by

$$
\omega_\phi(t)\overset{\mathrm{def}}{=}\sup\{\|\phi(x)-\phi(y)\|_Y : x,y\in B_X\ \wedge\ \|x-y\|_X\leqslant t\},
$$

is sub-additive (see for example [BL00, Ch. 1, Sec. 1]). Consequently, as explained in [BL00, Ch. 1, Sec. 2], there exists an increasing concave function $\beta : [0,\infty)\to[0,\infty)$ such that $\omega_\phi\leqslant\beta\leqslant 2\omega_\phi$.

Since $\beta : [0,\infty)\to[0,\infty)$ is concave, increasing, and $\beta(0)=0$, for every $q\in[1,\infty)$ the mapping $t\mapsto\beta(t^{1/q})^q$ is concave. Indeed, it suffices to verify this when β is differentiable. Denote $f(t)=\beta(t^{1/q})^q$. Then

$$
f'(t)=\left(\frac{\beta(t^{1/q})}{t^{1/q}}\right)^{q-1}\beta'(t^{1/q}). \tag{106}
$$

By our assumptions β' is nonnegative and decreasing, and $s\mapsto\beta(s)/s$ is decreasing on $(0,\infty)$. It therefore follows from (106) that f' is decreasing on $(0,\infty)$, as required.

The canonical radial retraction of X onto B_X is denoted below by $\rho : X\to B_X$, i.e.,

$$
\rho(x)\overset{\mathrm{def}}{=}\begin{cases} x & \text{if } \|x\|_X\leqslant 1,\\ \frac{x}{\|x\|_X} & \text{if } \|x\|_X\geqslant 1.\end{cases} \tag{107}
$$

It is straightforward to check that ρ is 2-Lipschitz.

Lemma 5.1. *Under the assumptions of Theorem 1.10, fix $n \in \mathbb{N}$, $q \in [1,\infty)$ and $x_1,\ldots,x_n \in X$ with*
$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\|x_i-x_j\|_X^q = 1. \tag{108}$$
For every $i \in \{1,\ldots,n\}$ let $r_i \in (0,\infty)$ be the smallest $r>0$ such that
$$\big|\{j\in\{1,\ldots,n\}:\ \|x_j-x_i\|_X \leqslant r\}\big| \geqslant \frac{n}{2}. \tag{109}$$
Then for every n by n symmetric stochastic matrix $A=(a_{ij})\in M_n(\mathbb{R})$,
$$\min_{i\in\{1,\ldots,n\}} r_i \leqslant \max\left\{2(8\gamma(A,\|\cdot\|_Y^q))^{1/q}, \frac{2}{\beta^{-1}\left(\frac{\alpha(1/4)}{8\gamma\left(A,\|\cdot\|_Y^q\right)^{1/q}}\right)}\right\}. \tag{110}$$

Proof. Note that the fact that the function $x\mapsto \beta(x^{1/q})^q$ is concave on $[0,\infty)$ implies that for every $\lambda\in(0,\infty)$ we have
$$\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\beta\left(\lambda\|x_i-x_j\|_X\right)^q\right)^{\frac{1}{q}} \leqslant \beta\left(\lambda\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\|x_i-x_j\|_X^q\right)^{\frac{1}{q}}\right) \stackrel{(108)}{=} \beta(\lambda). \tag{111}$$
By relabeling the points if necessary we may also assume without loss of generality that
$$r_1 = \min_{i\in\{1,\ldots,n\}} r_i. \tag{112}$$
Denote
$$B \stackrel{\mathrm{def}}{=} \{j\in\{1,\ldots,n\}:\ \|x_j-x_1\|_X\leqslant r_1\}. \tag{113}$$
So, by definition,
$$|B| \geqslant \frac{n}{2}. \tag{114}$$
For the sake of simplicity we denote below
$$\gamma \stackrel{\mathrm{def}}{=} \gamma(A,\|\cdot\|_Y^q)\,. \tag{115}$$
For $i\in\{1,\ldots,n\}$ define
$$y_i \stackrel{\mathrm{def}}{=} x_1 + r_1\rho\left(\frac{x_i-x_1}{r_1}\right),$$
where $\rho: X\to B_X$ is given in (107). Since ρ is 2-Lipschitz,
$$\forall\, i,j\in\{1,\ldots,n\},\qquad \|y_i-y_j\|_X \leqslant 2\|x_i-x_j\|_X. \tag{116}$$

By definition $y_i \in x_1 + r_1 B_X$, so we may consider the vectors
$$z_i \stackrel{\mathrm{def}}{=} \phi\left(\frac{y_i - x_1}{r_1}\right) \in Y. \tag{117}$$
Now,
$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|z_i - z_j\|_Y^q &\stackrel{(115)}{\leqslant} \frac{\gamma}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|z_i - z_j\|_Y^q \\ &\stackrel{(19)\wedge(117)\wedge(116)}{\leqslant} \frac{\gamma}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\beta\left(\frac{2\|x_i - x_j\|_X}{r_1}\right)^q \stackrel{(111)}{\leqslant} \gamma\beta\left(\frac{2}{r_1}\right)^q. \end{aligned}\tag{118}$$

Denoting
$$w \stackrel{\mathrm{def}}{=} \frac{1}{n}\sum_{j=1}^n z_j \in Y,$$
by the convexity of $\|\cdot\|_Y^q$ we have
$$\frac{1}{n}\sum_{i=1}^n \|z_j - w\|_Y^q \leqslant \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|z_i - z_j\|_Y^q \stackrel{(118)}{\leqslant} \gamma\beta\left(\frac{2}{r_1}\right)^q. \tag{119}$$
It follows from (119) and Markov's inequality that if we set
$$C \stackrel{\mathrm{def}}{=} \left\{i \in \{1,\ldots,n\}:\ \|z_i - w\|_Y \leqslant (4\gamma)^{1/q}\beta\left(\frac{2}{r_1}\right)\right\}, \tag{120}$$
then $|C| \geqslant 3n/4$. By (114) it follows that
$$|B \cap C| \geqslant \frac{n}{4}. \tag{121}$$

Recalling the definitions (113), (120) and (117), for every $i \in B \cap C$ we have
$$z_i = \phi\left(\frac{x_i - x_1}{r_1}\right) \qquad \text{and} \qquad \|z_i - w\|_Y \leqslant (4\gamma)^{1/q}\beta\left(\frac{2}{r_1}\right). \tag{122}$$
Hence, for every $i, j \in B \cap C$,
$$\begin{aligned}\alpha\left(\frac{\|x_i - x_j\|_X}{r_1}\right) &\stackrel{(19)}{\leqslant} \left\|\phi\left(\frac{x_i - x_1}{r_1}\right) - \phi\left(\frac{x_j - x_1}{r_1}\right)\right\|_Y \\ &\stackrel{(122)}{=} \|z_i - z_j\|_Y \leqslant \|z_i - w\|_Y + \|z_j - w\|_Y \stackrel{(122)}{\leqslant} 2(4\gamma)^{1/q}\beta\left(\frac{2}{r_1}\right).\end{aligned}$$
Consequently,
$$i, j \in B \cap C \implies \|x_i - x_j\|_X \leqslant r \stackrel{\mathrm{def}}{=} r_1\alpha^{-1}\left(2(4\gamma)^{1/q}\beta\left(\frac{2}{r_1}\right)\right). \tag{123}$$

Fix an arbitrary index $k \in B \cap C$ and define

$$S \stackrel{\mathrm{def}}{=} \{j \in \{1, \ldots, n\} : \ \|x_k - x_j\|_X \leqslant r\}. \tag{124}$$

It follows from (123) that $S \supseteq B \cap C$, and therefore by (121) we have

$$|S| \geqslant \frac{n}{4}. \tag{125}$$

Moreover, by the definition (124),

$$\frac{1}{n} \sum_{i \in S} \|x_i - x_k\|_X^q \leqslant \frac{|S| r^q}{n}. \tag{126}$$

Now, define for every $i \in \{1, \ldots, n\}$,

$$v_i \stackrel{\mathrm{def}}{=} \max\{0, \|x_i - x_k\|_X - r\} \in [0, \infty). \tag{127}$$

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

$$\begin{aligned} \|x_i - x_k\|_X^q &\leqslant 2^{q-1} \left(\|x_i - x_k\|_X - r\right)^q + 2^{q-1} r^q \\ &\stackrel{(127)}{=} 2^{q-1} |v_i - v_j|^q + 2^{q-1} r^q. \end{aligned} \tag{128}$$

Hence,

$$\begin{aligned} &\frac{1}{n} \sum_{i \in \{1, \ldots, n\} \smallsetminus S} \|x_i - x_k\|_X^q \\ &\stackrel{(128)}{\leqslant} \frac{2^{q-1}}{n|S|} \sum_{i=1}^n \sum_{j=1}^n |v_i - v_j|^q + \frac{2^{q-1}(n - |S|) r^q}{n} \\ &\stackrel{(125)}{\leqslant} \frac{2^{q+1} \gamma(A, d_{\mathbb{R}}^q)}{n} \sum_{i=1}^n \sum_{j=1}^n a_{ij} |v_i - v_j|^q + \frac{2^{q-1}(n - |S|) r^q}{n} \\ &\leqslant 2^{q+1} \gamma + \frac{2^{q-1}(n - |S|) r^q}{n}, \end{aligned} \tag{129}$$

where the last step of (125) uses the trivial fact $\gamma(A, d_{\mathbb{R}}^q) \leqslant \gamma$ (since Y contains an isometric copy of $\mathbb{R}$) and

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_{ij} |v_i - v_j|^q \stackrel{(127)}{\leqslant} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \|x_i - x_j\|_X^q \stackrel{(108)}{=} 1.$$

A combination of (126) and (129) yields the estimate

$$\frac{1}{n} \sum_{i=1}^n \|x_i - x_k\|_X^q \leqslant 2^{q+1} \gamma + 2^{q-1} r^q.$$

Consequently, an application of Markov's inequality shows that

$$\left| \left\{ i \in \{1, \ldots, n\} : \ \|x_i - x_k\|_X \leqslant 2^{1/q} \left(2^{q+1} \gamma + 2^{q-1} r^q\right)^{1/q} \right\} \right| \geqslant \frac{n}{2}.$$

Recalling the definition of r_k (see (109)), it follows that

$$r_1 \overset{(112)}{\leqslant} r_k \leqslant 2^{1/q}\left(2^{q+1}\gamma + 2^{q-1}r^q\right)^{1/q}. \tag{130}$$

If $(2r)^q \leqslant r_1^q/2$ then it follows from (130) that $r_1^q \leqslant 2^{q+3}\gamma$, implying the desired estimate (110). So, suppose that $(2r)^q > r_1^q/2$, which, by recalling the definition of r in (123), is the same as

$$\frac{r_1^q}{2} < \left(2r_1\alpha^{-1}\left(2(4\gamma)^{1/q}\beta\left(\frac{2}{r_1}\right)\right)\right)^q,$$

or

$$\alpha\left(\frac{1}{4}\right) \leqslant \alpha\left(\frac{1}{2^{1+1/q}}\right) \leqslant 2(4\gamma)^{1/q}\beta\left(\frac{2}{r_1}\right) \leqslant 8\gamma^{1/q}\beta\left(\frac{2}{r_1}\right),$$

implying the validity of (110) in this case as well. $\square$

Proof of Theorem 1.10. We continue to use the notation that was introduced in the statement and the proof of Lemma 5.1. In particular, the assumptions (108) and (112) are (without loss of generality) still in force.

Recalling the definition of B in (113), we have

$$\frac{1}{n}\sum_{i\in B}\|x_i - x_1\|_X^q \leqslant \frac{r_1^q|B|}{n}. \tag{131}$$

For $i \in \{1,\ldots,n\}$ define

$$u_i \stackrel{\mathrm{def}}{=} \max\{0, \|x_i - x_1\|_X - r_1\} \in [0,\infty). \tag{132}$$

Then for every $i \in \{1,\ldots,n\} \smallsetminus B$ and $j \in B$ we have

$$\begin{aligned}\|x_i - x_1\|_X^q &\leqslant 2^{q-1}\left(\|x_i - x_1\|_X - r_1\right)^q + 2^{q-1}r_1^q\\ &\overset{(132)}{=} 2^{q-1}|u_i - u_j|^q + 2^{q-1}r_1^q.\end{aligned}$$

Arguing exactly as in (129) (with the use of (125) replaced by the use of (114)), it follows that

$$\frac{1}{n}\sum_{i\in\{1,\ldots,n\}\smallsetminus B}\|x_i - x_1\|_X^q \leqslant 2^q\gamma + \frac{2^{q-1}(n-|B|)r_1^q}{n}. \tag{133}$$

Consequently,

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i - x_j\|_X^q &\leqslant \frac{2^{q-1}}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\|x_i - x_1\|_X^q + \|x_1 - x_j\|_X^q\right)\\ &= \frac{2^q}{n}\sum_{i=1}^n\|x_i - x_1\|_X^q \overset{(131)\wedge(133)}{\leqslant} 4^q\gamma + \frac{4^q r_1^q}{2}. \end{aligned}\tag{134}$$

The desired estimate (20) now follows substituting (110) into (134). $\square$

5.1. **Limitations of Ozawa's method.** In the discussion immediately preceding the estimates (21) and (22) we stated that for every $p \in [1,\infty)$ there exists a mapping $\phi : \ell_p \to \ell_2$ such that for every $x, y \in \ell_p$ we have

$$p \in (2,\infty) \implies \frac{\|x-y\|_{\ell_p}^{p/2}}{2^{(p-2)/2}} \leqslant \|\phi(x)-\phi(y)\|_{\ell_2} \leqslant p\|x-y\|_{\ell_p}, \quad (135)$$

and

$$p \in [1,2) \implies \frac{\|x-y\|_{\ell_p}}{3} \leqslant \|\phi(x)-\phi(y)\|_{\ell_2} \leqslant 2\|x-y\|_{\ell_p}^{p/2}. \quad (136)$$

The estimates (135) and (136) are a special case of the following bounds on the modulus of uniform continuity of the Mazur map [Maz29] (see also [BL00, Ch.9]). Let (Ω,μ) be a measure space and fix $p,q \in [1,\infty)$. Define $M_{p,q} : L_p(\mu) \to L_q(\mu)$ by

$$\forall\, f \in L_p(\mu), \qquad M_{p,q}(f) \stackrel{\mathrm{def}}{=} |f|^{p/q}\mathrm{sign}(f).$$

If $p \geqslant q$ then for every $f,g \in L_p(\mu)$ with $\|f\|_{L_p(\mu)}, \|g\|_{L_p(\mu)} \leqslant 1$ we have

$$\frac{\|f-g\|_{L_p(\mu)}^{p/q}}{2^{(p-q)/q}} \leqslant \|M_{p,q}(f)-M_{p,q}(g)\|_{L_q(\mu)} \leqslant \frac{2^{\frac{1}{q}-\frac{1}{p}}p\|f-g\|_{L_p(\mu)}}{q}. \quad (137)$$

Note that (135) is a special case of (137), and (136) is also a consequence of (137) because $M_{p,q}^{-1} = M_{q,p}$. While the bounds appearing in (137) are entirely standard, they seem to have been always stated in the literature while either using implicit multiplicative constant factors, or with suboptimal constant factors. These constants play a role in our context, so we briefly include the proof of (137), following the lines of the proof of [BL00, Prop. 9.2]. The elementary inequality

$$\big||u|^\theta\mathrm{sign}(u)-|v|^\theta\mathrm{sign}(v)\big| \geqslant \frac{|u-v|^\theta}{2^{\theta-1}},$$

which holds for every $u,v \in \mathbb{R}$ and $\theta \in [1,\infty)$, immediately implies (with $\theta = p/q$) the leftmost inequality in (137). To prove the rightmost inequality of (137), note the following elementary inequality, which also holds for every $u,v \in \mathbb{R}$ and $\theta \in [1,\infty)$.

$$\big||u|^\theta\mathrm{sign}(u)-|v|^\theta\mathrm{sign}(v)\big| \leqslant \theta|u-v|\max\left\{|u|^{\theta-1},|v|^{\theta-1}\right\}.$$

Consequently,

$$\begin{aligned}
&\|M_{p,q}(f)-M_{p,q}(g)\|_{L_q(\mu)}^q\\
&\leqslant \frac{p^q}{q^q}\int_\Omega |f-g|^q \max\{|f|,|g|\}^{p-q}\,d\mu\\
&\leqslant \frac{p^q}{q^q}\|f-g\|_{L_p(\mu)}^q\cdot\|\max\{|f|,|g|\}\|_{L_p(\mu)}^{p-q} \qquad (138)\\
&\leqslant \frac{p^q\cdot 2^{(p-q)/p}}{q^q}\|f-g\|_{L_p(\mu)}^q, \qquad (139)
\end{aligned}$$

where (138) follows from an application of Hölder's inequality with exponents p/q and $p/(p-q)$ and (139) holds true because we have $\|\max\{|f|,|g|\}\|_{L_p(\mu)}^p \leqslant \|f\|_{L_p(\mu)}^p+\|g\|_{L_p(\mu)}^p\leqslant 2$.

Returning to Theorem 1.10 (in particular using the notation and assumptions that were introduced in the statement of Theorem 1.10), if one wants the bound (20) to be compatible with the assumption of Theorem 1.3 one needs (20) to yield an upper bound on $\gamma(A,\|\cdot\|_X^q)$ that grows linearly with $\gamma(A,\|\cdot\|_Y^q)$. This is equivalent to the requirement

$$\beta^{-1}\left(\frac{\alpha(1/4)}{8\gamma(A,\|\cdot\|_Y^q)^{1/q}}\right)^q \gtrsim_{X,Y} \frac{1}{\gamma(A,\|\cdot\|_Y^q)}. \qquad (140)$$

Since (140) is supposed to hold for every $n\in\mathbb{N}$ and every n by n symmetric stochastic matrix A, (140) is the same as requiring that $\beta(t)\lesssim_{X,Y} t$ for every $t\in(0,\infty]$.

Specializing the above discussion to $Y=\ell_2$ and $q=2$, if $\beta(t)\leqslant Kt$ for some $K\in(0,\infty)$ and every $t\in(0,\infty)$ then (20) yields the estimate

$$\gamma(A,\|\cdot\|_X^q)\lesssim \frac{(K/\alpha(1/4))^2}{1-\lambda_2(A)}.$$

By Corollary 1.4, this implies that

$$Av_{\ell_2}^{(2)}(X)\lesssim \frac{K}{\alpha(1/4)}.$$

In particular, if $p\in(2,\infty)$ then due to (135) we get the estimate

$$Av_{\ell_2}^{(2)}(\ell_p)\lesssim p2^{3p/2},$$

which is exponentially worse than (23). The following lemma shows that this exponential loss is inherent to the use of Theorem 1.10 for the purpose of obtaining average distortion embedding of finite subsets of ℓ_p into ℓ_2, i.e., that $K/\alpha(1/4)$ must grow exponentially in p as $p\to\infty$.

Lemma 5.2. *Suppose that $p \in (2,\infty)$ and that $\phi : B_X \to \ell_2$ satisfies*

$$\alpha\left(\|x-y\|_{\ell_p}\right) \leqslant \|\phi(x)-\phi(y)\|_{\ell_2} \leqslant K\|x-y\|_{\ell_p}, \tag{141}$$

for every $x,y \in B_{\ell_p}$, where $\alpha : (0,2] \to (0,\infty)$ is increasing. Then

$$\forall\, \varepsilon \in (0,2), \qquad \frac{K}{\alpha(2-\varepsilon)} \gtrsim \frac{1}{(1-\varepsilon/2)^{p/2}}. \tag{142}$$

In particular, for $\varepsilon = 7/4$ we have $K/\alpha(1/4) \gtrsim 2^{3p/2}$.

Proof. Fix $m,n \in \mathbb{N}$ and $s \in (0,1]$. For every $x \in \mathbb{Z}^n$ define $\psi(x) \in \ell_p$ to be the vector whose jth coordinate equals

$$\frac{s}{n^{1/p}} e^{2\pi x_j i/m}$$

if $j \in \{1,\ldots,n\}$, and whose remaining coordinates vanish. Then we have $\|\psi(x)\|_{\ell_p} \leqslant s \leqslant 1$ for every $x \in \mathbb{Z}^n$.

By the results of Section 3 of [MN08], if m is divisible by 4 and $m \geqslant \frac{2}{3}\pi\sqrt{n}$ then we have

$$\begin{aligned}\frac{1}{m^n}\sum_{j=1}^{n}\sum_{x\in\{0,\ldots,m-1\}^n} &\left\|\phi\left(\psi\left(x+\frac{m}{2}e_j\right)\right)-\phi(\psi(x))\right\|_{\ell_2}^2\\ &\lesssim \frac{m^2}{(3m)^n}\sum_{\omega\in\{-1,0,1\}^n}\sum_{x\in\{0,\ldots,m-1\}^n}\left\|\phi\left(\psi(x+\omega)\right)-\phi(\psi(x))\right\|_{\ell_2}^2. \end{aligned} \tag{143}$$

Now, by the leftmost inequality in (141) for every $x \in \{0,\ldots,m-1\}^n$ and $j \in \{1,\ldots,n\}$ we have

$$\begin{aligned}\left\|\phi\left(\psi\left(x+\frac{m}{2}e_j\right)\right)-\phi(\psi(x))\right\|_{\ell_2} &\geqslant \alpha\left(\frac{s\left|e^{2\pi(x_j+m/2)i/m}-e^{2\pi x_j i/m}\right|}{n^{1/p}}\right)\\ &= \alpha\left(\frac{2s}{n^{1/p}}\right). \end{aligned} \tag{144}$$

Also, by the rightmost inequality in (141), for every $x \in \{0,\ldots,m-1\}^n$ and $\omega \in \{-1,0,1\}^n$ we have

$$\begin{aligned}\left\|\phi\left(\psi(x+\omega)\right)-\phi(\psi(x))\right\|_{\ell_2} &\leqslant \frac{Ks}{n^{1/p}}\left(\sum_{j=1}^{n}\left|e^{2\pi(x_j+\omega_j)i/m}-e^{2\pi x_j i/m}\right|^p\right)^{\frac{1}{p}}\\ &\lesssim \frac{Ks}{m}. \end{aligned} \tag{145}$$

Choose $n \stackrel{\mathrm{def}}{=} \lfloor 1/(1-\varepsilon/2)^p\rfloor$ and $s \stackrel{\mathrm{def}}{=} (1-\varepsilon/2)n^{1/p} \in (0,1]$. Then, if m is the smallest integer that is divisible by 4 and satisfies $m \geqslant \frac{2}{3}\pi\sqrt{n}$, by substituting (145) and (144) into (143) we see that

$$n\alpha\left(2-\varepsilon\right)^2 \lesssim m^2\cdot\frac{K^2s^2}{m^2} \lesssim n\frac{K^2(1-\varepsilon/2)^2}{n^{1-2/p}} \lesssim nK^2(1-\varepsilon/2)^p,$$

which simplifies to give the desired estimate (142). □

6. Bourgain–Milman–Wolfson type

Here we study aspects of nonlinear type in the sense of Bourgain, Milman and Wolfson [BMW86], proving in particular Lemma 1.12, Lemma 1.13 and Theorem 1.15 that were stated in the Introduction.

6.1. **On the Maurey–Pisier problem for BMW type.** In what follows, for every finite set Ω and $p\in[1,\infty)$, the space $L_p(\Omega)$ consists of all mappings $f:\Omega\to\mathbb{R}$, equipped with the norm

$$\|f\|_{L_p(\Omega)} \stackrel{\mathrm{def}}{=} \left(\frac{1}{|\Omega|}\sum_{\omega\in\Omega}|f(\omega)|^p\right)^{\frac{1}{p}}.$$

For every $k\in\{1,\ldots,n\}$ let $\Omega_k^n\subseteq\mathbb{F}_2^n\times 2^{\{1,\ldots,n\}}$ be defined by

$$\Omega_k^n \stackrel{\mathrm{def}}{=} \mathbb{F}_2^n\times\binom{\{1,\ldots,n\}}{k} = \mathbb{F}_2^n\times\left\{I\subseteq\{1,\ldots,n\}:\ |I|=k\right\}.$$

Thus $|\Omega_k^n|=2^n\binom{n}{k}$.

Fixing a metric space (X,d_X) and $f:\mathbb{F}_2^n\to X$, define

$$\forall (z,I)\in\mathbb{F}_2^n\times 2^{\mathbb{F}_2^n},\qquad \mathcal{D}_f(z,I)\stackrel{\mathrm{def}}{=} d_X\left(f\left(z+e_I\right),f(z)\right), \tag{146}$$

where for $I\subseteq\{1,\ldots,n\}$ we set

$$e_I\stackrel{\mathrm{def}}{=}\sum_{i\in I}e_i\in\mathbb{F}_2^n.$$

Thus, using the notation of the Introduction, we have $e_{\{1,\ldots,n\}}=e$.

For $q\in(0,\infty)$ define $E_k^{(q)}(f)\in[0,\infty)$ by

$$\begin{aligned}E_k^{(q)}(f)^q &\stackrel{\mathrm{def}}{=} \|\mathcal{D}_f\|_{L_q(\Omega_k^n)}^q\\ &= \frac{1}{2^n\binom{n}{k}}\sum_{\substack{I\subseteq\{1,\ldots,n\}\\ |I|=k}}\sum_{z\in\mathbb{F}_2^n} d_X\left(f\left(z+e_I\right),f(z)\right)^q.\end{aligned} \tag{147}$$

Note that since for every $I,J\subseteq\{1,\ldots,n\}$ with $|I|=|J|$ the number of permutations $\sigma\in S_n$ satisfying $\sigma(I)=J$ equals $|I|!(n-|I|)!$,

$$\forall\, I\subseteq\{1,\ldots,n\},\qquad E_{|I|}^{(q)}(f)^q=\frac{1}{2^nn!}\sum_{x\in\mathbb{F}_2^n}\sum_{\sigma\in S_n}\mathcal{D}_f(x,\sigma(I))^q. \tag{148}$$

We also record for future use the following simple consequence of the triangle inequality in (X,d_X).

Lemma 6.1. *Fix $n\in\mathbb{N}$ and $q\in[1,\infty)$. Suppose that (X,d_X) is a metric space and $f:\mathbb{F}_2^n\to X$. Then for every $k,m\in\{1,\ldots,n\}$ with $k+m\leqslant n$ we have*

$$E_{k+m}^{(q)}(f)\leqslant E_k^{(q)}(f)+E_m^{(q)}(f). \tag{149}$$

Proof. Fix $I\subseteq\{1,\ldots,n\}$ with $|I|=k+m$. Recalling (146), for every $J\subseteq I$ with $|J|=k$ and every $z\in\mathbb{F}_2^n$ we have

$$\mathcal{D}_f(z,I)\leqslant \mathcal{D}_f(z,I\smallsetminus J)+\mathcal{D}_f(z+e_{I\smallsetminus J},J).$$

Consequently,

$$\forall (z,I)\in\Omega_{k+m}^n,\qquad \mathcal{D}_f(z,I)\leqslant U_f(z,I)+V_f(z,I), \tag{150}$$

where

$$U_f(z,I)\stackrel{\mathrm{def}}{=}\frac{1}{\binom{k+m}{k}}\sum_{\substack{J\subseteq I\\ |J|=k}}\mathcal{D}_f(z,I\smallsetminus J).$$

and

$$V_f(z,I)\stackrel{\mathrm{def}}{=}\frac{1}{\binom{k+m}{k}}\sum_{\substack{J\subseteq I\\ |J|=k}}\mathcal{D}_f(z+e_{I\smallsetminus J},J).$$

The point-wise estimate (150) combined with the triangle inequality in $L_q(\Omega_{k+m}^n)$ implies that

$$E_{k+m}^{(q)}(f)=\|\mathcal{D}_f\|_{L_q(\Omega_{k+m}^n)}\leqslant \|U_f\|_{L_q(\Omega_{k+m}^n)}+\|V_f\|_{L_q(\Omega_{k+m}^n)}. \tag{151}$$

By the convexity of $t\mapsto |t|^q$ we have

$$\begin{aligned}\|U_f\|_{L_q(\Omega_{k+m}^n)}^q&\leqslant \frac{1}{2^n\binom{n}{k+m}}\sum_{\substack{I\subseteq\{1,\ldots,n\}\\ |I|=k+m}}\sum_{z\in\mathbb{F}_2^n}\frac{1}{\binom{k+m}{k}}\sum_{\substack{J\subseteq I\\ |J|=k}}\mathcal{D}_f(z,I\smallsetminus J)^q\\ &=\frac{1}{2^n\binom{n}{k+m}\binom{k+m}{k}}\sum_{z\in\mathbb{F}_2^n}\sum_{\substack{S\subseteq\{1,\ldots,n\}\\ |S|=m}}\binom{n-m}{k}\mathcal{D}_f(z,S)^q=E_m^{(q)}(f)^q,\end{aligned} \tag{152}$$

and

$$\begin{aligned}\|V_f\|_{L_q(\Omega_{k+m}^n)}^q&\leqslant \frac{1}{2^n\binom{n}{k+m}}\sum_{\substack{I\subseteq\{1,\ldots,n\}\\ |I|=k+m}}\sum_{z\in\mathbb{F}_2^n}\frac{1}{\binom{k+m}{k}}\sum_{\substack{J\subseteq I\\ |J|=k}}\mathcal{D}_f(z+e_{I\smallsetminus J},J)^q\\ &=\frac{1}{2^n\binom{n}{k+m}\binom{k+m}{k}}\sum_{\substack{I\subseteq\{1,\ldots,n\}\\ |I|=k+m}}\sum_{\substack{J\subseteq I\\ |J|=k}}\sum_{w\in\mathbb{F}_2^n}\mathcal{D}_f(w,J)^q=E_k^{(q)}(f)^q.\end{aligned} \tag{153}$$

The desired estimate (149) now follows from a substitution of (152) and (153) into (151). $\square$

For $m,n\in\mathbb{N}$ with $n\geqslant m$, for every $f:\mathbb{F}_2^m\to X$ denote its natural lifting to $\mathbb{F}_2^n$ by $f^{\uparrow n}:\mathbb{F}_2^n\to X$, that is,

$$\forall\, z\in\mathbb{F}_2^n,\qquad f^{\uparrow n}(z)\stackrel{\mathrm{def}}{=}f(z_1,\ldots,z_m).$$

We then have the following identity for every $k\in\{1,\ldots,n\}$.

$$E_k^{(q)}\left(f^{\uparrow n}\right)^q=\frac{1}{2^n\binom{n}{k}}\sum_{\substack{I\subseteq\{1,\ldots,n\}\\|I|=k}}2^{n-m}\sum_{w\in\mathbb{F}_2^m}\mathcal{D}_f(w,I\cap\{1,\ldots,m\})^q \tag{154}$$

$$=\frac{1}{\binom{n}{k}}\sum_{\ell=\max\{1,m+k-n\}}^{\min\{k,m\}}\binom{n-m}{k-\ell}\binom{m}{\ell}E_\ell^{(q)}(f)^q, \tag{155}$$

where (154) uses (146) and (147), and the identity (155) follows by observing that if $I\subseteq\{1,\ldots,n\}$ satisfies $|I\cap\{1,\ldots,m\}|=\ell$ then necessarily $\ell\geqslant m+k-n$ and for each $J\subseteq\{1,\ldots,m\}$ with $|J|=\ell$ the number of subsets $I\subseteq\{1,\ldots,n\}$ with $I\cap\{1,\ldots,m\}=J$ is $\binom{n-m}{k-\ell}$. Note in particular the following two special cases of (155).

$$E_n^{(q)}\left(f^{\uparrow n}\right)=E_n^{(q)}(f)\qquad\text{and}\qquad E_1^{(q)}\left(f^{\uparrow n}\right)=\left(\frac{m}{n}\right)^{1/q}E_1^{(q)}(f). \tag{156}$$

When $q=2$ in (147) we write $E_k(f)\stackrel{\mathrm{def}}{=}E_k^{(2)}(f)$. With this notation, a metric space (X,d_X) has BMW type $p\in(0,\infty)$ if and only if there exists $T\in(0,\infty)$ such that for every $n\in\mathbb{N}$ and every $f:\mathbb{F}_2^n\to X$,

$$E_n(f)\leqslant Tn^{1/p}E_1(f). \tag{157}$$

Let $\mathrm{BMW}_p^n(X)$ denote the infimum over those $T\in(0,\infty)$ for which (157) holds true for every $f:\mathbb{F}_2^n\to X$. Thus

$$\mathrm{BMW}_p(X)=\sup_{n\in\mathbb{N}}\mathrm{BMW}_p^n(X).$$

Remark 6.2. By definition we have $\mathrm{BMW}_p^n(X)n^{1/p}=\mathrm{BMW}_q^n(X)n^{1/q}$ for every $p,q\in(0,\infty)$. Also, unless $|X|=1$ we have $\mathrm{BMW}_p^1(X)=1$.

Remark 6.3. By Lemma 6.1 we have $E_n(f)\leqslant nE_1(f)$, so

$$\forall\, n\in\mathbb{N},\qquad \mathrm{BMW}_p^n(X)\leqslant n^{1-\frac{1}{p}}.$$

Moreover, given $m,n\in\mathbb{N}$ with $n\geqslant m$ and $f:\mathbb{F}_2^m\to X$ with $E_1(f)>0$, by (156) we have

$$\frac{E_m(f)}{E_1(f)}=\frac{E_n\left(f^{\uparrow n}\right)}{E_1\left(f^{\uparrow n}\right)}\cdot\sqrt{\frac{n}{m}}.$$

Consequently,

$$\forall\, m,n\in\mathbb{N},\quad m\leqslant n \implies \mathrm{BMW}_p^m(X)\sqrt{m}\leqslant \mathrm{BMW}_p^n(X)\sqrt{n}. \tag{158}$$

Remark 6.4. Let (X,d_X) be a metric space and $p\in(1,\infty)$ be such that X has BMW type p. Then necessarily $p\leqslant 2$. Indeed, choose distinct $x_0,x_1\in X$ and for every $n\in\mathbb{N}$ define $f_n:\mathbb{F}_2^n\to X$ by $f_n(z)=x_{z_1}$. Then $E_1(f_n)=d_X(x_0,x_1)/\sqrt{n}$ and $E_n(f_n)=d_X(x_0,x_1)$. This means that $n^{1/p}\mathrm{BMW}_p(X)\geqslant\sqrt{n}$ for all $n\in\mathbb{N}$, which implies that $p\leqslant 2$. A straightforward application of the triangle inequality (see [BMW86]) implies that every metric space has BMW type 1 with $\mathrm{BMW}_1(X)=1$.

Recalling the definition of $p_X\in[1,2]$ in (27), we have the following lemma that relies on a sub-multiplicativity argument that was introduced by Pisier [Pis73] in the context of Rademacher type of normed spaces, and has been implemented in the context of nonlinear type by Bourgain, Milman and Wolfson [BMW86] (see also [Pis86]).

Lemma 6.5. *For every metric space (X,d_X) we have*

$$\forall\, n\in\mathbb{N},\qquad \mathrm{BMW}_{p_X}^n(X)\geqslant 1.$$

Proof. Write $p=p_X$ and suppose for the sake of obtaining a contradiction that there exists $m\in\mathbb{N}$ and $\varepsilon\in(0,1)$ such that $\mathrm{BMW}_p^m(X)<\varepsilon$. We may also assume without loss of generality that $\varepsilon>1/\sqrt{m}$. Since $\mathrm{BMW}_p^1(X)=1$, we have $m\geqslant 2$. If we define

$$q\stackrel{\mathrm{def}}{=}\frac{1}{\frac{1}{p}-\frac{\log(1/\varepsilon)}{\log m}} \tag{159}$$

then $q\in(p,\infty)$ because $p\geqslant 1/2$ (recall Remark 6.4) and $\varepsilon>1/\sqrt{m}$. By [BMW86, Lem. 2.3] (see also [Pis86, Lem.7.2]), for every $k,n\in\mathbb{N}$,

$$\mathrm{BMW}_p^{kn}(X)\leqslant \mathrm{BMW}_p^k(X)\cdot\mathrm{BMW}_p^n(X).$$

Consequently, for every $i\in\mathbb{N}$ we have

$$\mathrm{BMW}_p^{m^i}(X)\leqslant \mathrm{BMW}_p^m(X)^i<\varepsilon^i\stackrel{(159)}{=}m^{-i\left(\frac{1}{p}-\frac{1}{q}\right)}. \tag{160}$$

For every $n\in\mathbb{N}$ choose $i\in\mathbb{N}$ such that $m^{i-1}\leqslant n<m^i$. Since, by (158), $\mathrm{BMW}_p^n(X)\sqrt{n}$ increases with n, it follows from (160) that

$$\begin{aligned}\mathrm{BMW}_q^n(X)&=n^{\frac{1}{p}-\frac{1}{q}-\frac{1}{2}}\cdot\mathrm{BMW}_p^n(X)\sqrt{n}\leqslant n^{\frac{1}{p}-\frac{1}{q}-\frac{1}{2}}\cdot\mathrm{BMW}_p^{m^i}(X)m^{\frac{i}{2}}\\&\stackrel{(160)}{<}n^{\frac{1}{p}-\frac{1}{q}-\frac{1}{2}}\cdot m^{-i\left(\frac{1}{p}-\frac{1}{q}-\frac{1}{2}\right)}\leqslant m^{\frac{1}{2}-\frac{1}{p}+\frac{1}{q}}\stackrel{(159)}{=}\varepsilon\sqrt{m},\end{aligned}$$

where we used the fact that $1/2-1/p+1/q>0$ (by (159) and the assumption $\varepsilon>1/\sqrt{m}$). Consequently $\mathrm{BMW}_q(X)<\infty$, i.e., X has BMW type q. Since $q>p$, this contradicts the definition of $p_X=p$. $\square$

Lemma 6.6. *Fix $p \in [1,2]$, $n \in \mathbb{N}$, and $k_1,\ldots,k_m \in \{1,\ldots,n\}$ such that $k_1+\ldots+k_m \leqslant n$. Then for every metric space (X,d_X) and every $f:\mathbb{F}_2^n \to X$ we have*

$$E_{k_1+\ldots+k_m}(f) \leqslant \mathrm{BMW}_p(X) m^{\frac{1}{p}-\frac{1}{2}} \left(\sum_{j=1}^m E_{k_j}(f)^2\right)^{\frac{1}{2}}. \tag{161}$$

Proof. Write $k_0 = 0$ and $I = \{1,\ldots,k_1+\ldots+k_m\} \subseteq \{1,\ldots,n\}$. For every $j \in \{1,\ldots,m\}$ set $I_j = \{k_1+\ldots+k_{j-1}+1,\ldots,k_1+\ldots+k_j\} \subseteq I$. Fixing $x \in \mathbb{F}_2^n$ and a permutation $\sigma \in S_n$, define $\phi_x^\sigma : \mathbb{F}_2^m \to X$ by

$$\phi_x^\sigma(z) \stackrel{\mathrm{def}}{=} f\left(x + \sum_{s=1}^m z_s e_{\sigma(I_s)}\right).$$

An application of the definition of $\mathrm{BMW}_p(X)$ to ϕ_x^σ yields the inequality

$$\begin{aligned}\frac{1}{2^m}\sum_{z\in\mathbb{F}_2^m} &\mathcal{D}_f\left(x+\sum_{s=1}^m z_s e_{\sigma(I_s)},\sigma(I)\right)^2\\ &\leqslant \frac{\mathrm{BMW}_p(X)^2 m^{\frac{2}{p}-1}}{2^m}\sum_{z\in\mathbb{F}_2^m}\sum_{j=1}^m \mathcal{D}_f\left(x+\sum_{s=1}^m z_s e_{\sigma(I_s)},\sigma(I_j)\right)^2. \end{aligned}\tag{162}$$

Recalling (148), by averaging (162) over $x \in \mathbb{F}_2^n$ and $\sigma \in S_n$ we obtain

$$\begin{aligned}E_{k_1+\ldots+k_m}(f)^2 &= \frac{1}{2^n n!}\sum_{y\in\mathbb{F}_2^n}\sum_{\sigma\in S_n}\mathcal{D}_f\left(y,\sigma(I)\right)^2\\ &\leqslant \frac{\mathrm{BMW}_p(X)^2 m^{\frac{2}{p}-1}}{2^n n!}\sum_{j=1}^m\sum_{y\in\mathbb{F}_2^n}\sum_{\sigma\in S_n}\mathcal{D}_f(y,\sigma(I_j))^2\\ &= \mathrm{BMW}_p(X)^2 m^{\frac{2}{p}-1}\sum_{j=1}^m E_{k_j}(f)^2. \qquad \square\end{aligned}$$

Proof of Theorem 1.15. Denote $p_X = p$. Fix $\varepsilon \in (0,1/3)$ and define

$$n \stackrel{\mathrm{def}}{=} \left\lceil \frac{\mathrm{BMW}_p(X)^4}{\varepsilon} d\right\rceil. \tag{163}$$

Since $n \geqslant d$, we can consider $\mathbb{F}_p^d$ as a subset of $\mathbb{F}_p^n$ (say, canonically embedded as the first d coordinates).

By Lemma 6.5 we have $\mathrm{BMW}_p^n(X) \geqslant 1$, and therefore by the definition of $\mathrm{BMW}_p^n(X)$ there exists $f:\mathbb{F}_p^n \to X$ such that

$$E_n(f) \geqslant (1-\varepsilon)n^{1/p}E_1(f) > 0. \tag{164}$$

Write

$$Y \stackrel{\mathrm{def}}{=} \ell_2^{\mathbb{F}_2^n\times S_n}(X) \cong \ell_2^{2^n n!}(X).$$

Define $F:\mathbb{F}_2^n\to Y$ by setting

$$\forall (x,z,\pi)\in \mathbb{F}_2^n\times\mathbb{F}_2^n\times S_n,\qquad F(x)_{(z,\pi)} \stackrel{\mathrm{def}}{=} f(\pi(x)+z),$$

where $\pi(x)\stackrel{\mathrm{def}}{=}\sum_{i=1}^n x_{\pi(i)}e_i$. Recalling (147), every $x,y\in\mathbb{F}_2^n$ satisfy

$$d_Y(F(x),F(y)) = \sqrt{2^n n!}\cdot E_{\|x-y\|_1}(f), \tag{165}$$

By Lemma 6.6 with $m=\|x-y\|_1$ and $k_1=\ldots=k_m=1$, for every $x,y\in\mathbb{F}_p^n$ we have

$$\begin{aligned}\frac{d_Y(F(x),F(y))}{\sqrt{2^n n!}} &\stackrel{(161)\wedge(165)}{\leqslant} \mathrm{BMW}_p(X)\|x-y\|_1^{\frac1p-\frac12}\cdot\sqrt{\|x-y\|_1}E_1(f)\\ &= \mathrm{BMW}_p(X)E_1(f)\|x-y\|_p. \end{aligned}\tag{166}$$

Fixing $x,y\in\mathbb{F}_p^d\subseteq\mathbb{F}_p^n$, write $n=a\|x-y\|_1+b$ for appropriate integers a and $b\in[0,\|x-y\|_1)$. By Lemma 6.6 with $m=b$ and $k_1=\ldots=k_m=1$,

$$E_b(f)\leqslant \mathrm{BMW}_p(X)b^{1/p}E_1(f)\leqslant \mathrm{BMW}_p(X)\|x-y\|_1^{1/p}E_1(f). \tag{167}$$

Using Lemma 6.6 once more, this time with $m=a+1$, $k_1=b$ and $k_2=\ldots=k_{a+1}=\|x-y\|_1$, and noting that since $\|x-y\|_1\leqslant d\leqslant \varepsilon n$ we have $m\leqslant(1+\varepsilon)n/\|x-y\|_1$, we conclude that

$$E_n(f)\leqslant \mathrm{BMW}_p(X)\left(\frac{(1+\varepsilon)n}{\|x-y\|_1}\right)^{\frac1p-\frac12}\sqrt{\frac{n}{\|x-y\|_1}E_{\|x-y\|_1}(f)^2+E_b(f)^2}.$$

In combination with (167) and our assumption (164), this implies

$$\begin{aligned}(1-\varepsilon)^2n^{2/p}E_1(f)^2\leqslant (1+\varepsilon)\mathrm{BMW}_p(X)^2\left(\frac{n}{\|x-y\|_1}\right)^{2/p}E_{\|x-y\|_1}(f)^2\\ +(1+\varepsilon)n^{2/p}\mathrm{BMW}_p(X)^4\frac{\|x-y\|_1}{n}E_1(f)^2.\end{aligned}\tag{168}$$

Recalling (163), we have $\|x-y\|_1/n\leqslant d/n\leqslant\varepsilon\mathrm{BMW}_p(X)^4$ (since $x,y\in\mathbb{F}_2^d$), and it therefore follows from (168) that

$$\frac{d_Y(F(x),F(y))}{\sqrt{2^n n!}}\stackrel{(165)}{=}E_{\|x-y\|_1}(f)\geqslant\frac{\|x-y\|_pE_1(f)}{\mathrm{BMW}_p(X)}\sqrt{\frac{1-3\varepsilon}{1+\varepsilon}}. \tag{169}$$

Since $\varepsilon\in(0,1/3)$ can be taken to be arbitrarily small, by combining (166) and (169) we conclude that

$$c_Y(\mathbb{F}_2^d,\|x-y\|_p)\leqslant\mathrm{BMW}_p(X)^2. \qquad\square$$

6.2. **Obstructions to average distortion embeddings of cubes.** We start by proving Lemma 1.13, whose proof is very simple.

Proof of Lemma 1.13. Fix $D > Av_Y^{(2)}(X)$ and $n \in \mathbb{N}$. If $f : \mathbb{F}_2^n \to X$ then there exists a nonconstant mapping $g : f(\mathbb{F}_2^n) \to Y$ such that

$$\sum_{(x,y)\in\mathbb{F}_2^n\times\mathbb{F}_2^n} d_Y(g(f(x)),g(f(y)))^2 \geqslant \frac{\|g\|_{\mathrm{Lip}}^2}{D^2}\sum_{(x,y)\in\mathbb{F}_2^n\times\mathbb{F}_2^n} d_X(f(x),f(y))^2.$$

Consequently,

$$\begin{aligned}
&\frac{1}{2^n}\sum_{x\in\mathbb{F}_2^n} d_X(f(x),f(x+e))^2\\
&\leqslant \frac{2}{4^n}\sum_{(x,y)\in\mathbb{F}_2^n\times\mathbb{F}_2^n}\left(d_X(f(x),f(y))^2+d_X(f(y),f(x+e))^2\right)\\
&=\frac{1}{4^{n-1}}\sum_{(x,y)\in\mathbb{F}_2^n\times\mathbb{F}_2^n} d_X(f(x),f(y))^2\\
&\leqslant \frac{D^2}{4^{n-1}\|g\|_{\mathrm{Lip}}^2}\sum_{(x,y)\in\mathbb{F}_2^n\times\mathbb{F}_2^n} d_Y(g(f(x)),g(f(y)))^2. \qquad (170)
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{1}{4^n}\sum_{(x,y)\in\mathbb{F}_2^n\times\mathbb{F}_2^n} d_Y(g(f(x)),g(f(y)))^2 &\overset{(147)}{=} \frac{1}{2^n}\sum_{k=1}^n\binom{n}{k}E_k(g\circ f)^2\\
&\overset{(161)}{\leqslant} \frac{1}{2^n}\sum_{k=1}^n\binom{n}{k}\mathrm{BMW}_p(Y)^2k^{2/p}E_1(g\circ f)^2. \quad (171)
\end{aligned}$$

Since $E_1(g\circ f)\leqslant \|g\|_{\mathrm{Lip}}E_1(f)$, it follows from (170) and (171) that $\mathrm{BMW}_p(X)\leqslant 2Av_Y^{(2)}(X)\mathrm{BMW}_p(Y)$. $\square$

Recall that (see e.g. [Mau03]) a Banach space $(X,\|\cdot\|_X)$ is said to have Rademacher type p constant $T\in(0,\infty)$ if for every $m\in\mathbb{N}$ and every $x_1,\ldots,x_m\in X$ we have

$$\left(\mathbb{E}_\varepsilon\left[\left\|\sum_{i=1}^m \varepsilon_i x_i\right\|_X^2\right]\right)^{\frac12}\leqslant T\left(\sum_{i=1}^m\|x_i\|_X^p\right)^{\frac1p}, \qquad (172)$$

where $\mathbb{E}_\varepsilon[\cdot]$ is the expectation with respect to i.i.d. ±1 Bernoulli random variables $\varepsilon_1,\ldots,\varepsilon_n$. The infimum over those $T\in(0,\infty)$ for which X has Rademacher type p constant T is denoted $T_p(X)$. If no such T exists then we write $T_p(X)=\infty$.

Lemma 6.7. *Assume that $p \in [1,2)$ and fix $q \in (p,2]$ and $r,s \in [1,\infty)$. Suppose that $(Y,\|\cdot\|_Y)$ is a Banach space with $T_q(Y) < \infty$ and that $f : \mathbb{F}_2^m \to Y$ satisfies*

$$\left(\frac{1}{4^m}\sum_{(x,y)\in\mathbb{F}_2^m\times\mathbb{F}_2^m}\|f(x)-f(y)\|_Y^r\right)^{\frac{1}{r}} = \left(\frac{1}{4^m}\sum_{(x,y)\in\mathbb{F}_2^m\times\mathbb{F}_2^m}\|x-y\|_{\ell_p}^s\right)^{\frac{1}{s}} \asymp m^{1/p}. \tag{173}$$

Then there exists $x \in \mathbb{F}_2^m$ and $i \in \{1,\ldots,n\}$ such that

$$\|f(x)-f(x+e_i)\|_Y \gtrsim \frac{1}{\sqrt{r}T_q(Y)}\cdot\frac{m^{\frac{1}{p}-\frac{1}{q}}}{\log m}. \tag{174}$$

Proof. By Pisier's inequality [Pis86] we have

$$\begin{aligned} m^{1/p} &\overset{(173)}{\lesssim} \left(\frac{1}{4^m}\sum_{(x,y)\in\mathbb{F}_2^m\times\mathbb{F}_2^m}\|f(x)-f(y)\|_Y^r\right)^{\frac{1}{r}} \\ &\lesssim \log m\left(\frac{1}{2^m}\sum_{x\in\mathbb{F}_2^n}\mathbb{E}_\varepsilon\left[\left\|\sum_{i=1}^m\varepsilon_i(f(x+e_i)-f(x))\right\|_Y^r\right]\right)^{\frac{1}{r}}. \end{aligned}\tag{175}$$

For every fixed $x \in \mathbb{F}_2^m$ it follows from Kahane's inequality (with asymptotically optimal dependence on r; see e.g. [Tal88]) that

$$\begin{aligned} &\left(\mathbb{E}_\varepsilon\left[\left\|\sum_{i=1}^m\varepsilon_i(f(x+e_i)-f(x))\right\|_Y^r\right]\right)^{\frac{1}{r}} \\ &\lesssim \sqrt{r}\left(\mathbb{E}_\varepsilon\left[\left\|\sum_{i=1}^m\varepsilon_i(f(x+e_i)-f(x))\right\|_Y^2\right]\right)^{\frac{1}{2}} \\ &\leqslant \sqrt{r}T_q(Y)\left(\sum_{i=1}^m\|f(x+e_i)-f(x)\|_Y^q\right)^{\frac{1}{q}}. \end{aligned}$$

Combined with (175), this implies that

$$\max_{\substack{x\in\mathbb{F}_2^m\\ i\in\{1,\ldots,m\}}}\|f(x+e_i)-f(x)\|_Y \gtrsim \frac{1}{\sqrt{r}T_q(Y)}\cdot\frac{m^{\frac{1}{p}-\frac{1}{q}}}{\log m}. \qquad \square$$

There are classes of Banach spaces Y, including Banach lattices of nontrivial type and UMD spaces, for which it is known that Pisier's inequality (175) holds true with the $\log m$ factor replaced by a constant that may depend on Y and r but not on m; see [NS02, HN13]. For such spaces we therefore obtain (174) without the $\log m$ term.

Lemma 6.8. *Assume that $p \in [1,2)$ and fix $q \in (p,2]$ and $r,s \in [1,\infty)$. Suppose that $(Y,\|\cdot\|_Y)$ is a Banach space with $S_q(Y) < \infty$, i.e., Y has modulus of uniform smoothness of power type q. If $f : \mathbb{F}_2^m \to Y$ satisfies (173) then there exists $x \in \mathbb{F}_2^m$ and $i \in \{1,\ldots,n\}$ such that*

$$\|f(x) - f(x+e_i)\|_Y \gtrsim \frac{m^{\frac{1}{p}-\frac{1}{q}}}{r^{1/q} + S_q(Y)}. \tag{176}$$

Proof. Due to (173), in order to prove (176) it suffices to show that for every $h : \mathbb{F}_2^m \to Y$,

$$\left(\frac{1}{4^m}\sum_{(x,y)\in\mathbb{F}_2^m\times\mathbb{F}_2^m}\|h(x)-h(y)\|_Y^r\right)^{\frac{1}{r}}$$
$$\lesssim \left(r^{\frac{1}{q}} + S_q(Y)\right) m^{\frac{1}{q}} \left(\frac{1}{m2^m}\sum_{i=1}^m\sum_{x\in\mathbb{F}_2^m}\|h(x+e_i)-h(x)\|_Y^r\right)^{\frac{1}{r}}. \tag{177}$$

Note that it suffices to prove (177) when $r \geqslant 2$, since otherwise we could replace q by r and use the fact that $S_r(Y) \leqslant S_q(Y)$.

By considering the standard random walk on the Hamming cube $\mathbb{F}_2^n$ and arguing mutatis mutandis as in [NS02, Sec. 5], (177) is a formal consequence of the Markov type estimate of Theorem 4.3. Alternatively, once can deduce (177) directly via the martingale argument in [KN06, Sec. 5], the only difference being the use of the martingale inequality (65) in place of Pisier's inequality [Pis75]. $\square$

Proof of Lemma 1.12. Since for $q \in [2,\infty)$ we have $S_2(\ell_q) \leqslant \sqrt{q-1}$, Lemma 1.12 is a special case of Lemma 6.8 with $Y = \ell_q$, $n = 2^m$ and $x_1,\ldots,x_{2^m}$ being an arbitrary enumeration of $\mathbb{F}_2^m \subseteq Y$. $\square$

Remark 6.9. As promised in the Introduction, here we justify (29). The fact that $Av_{\mathbb{R}}(\mathbb{F}_2^n,\|\cdot\|_2) \lesssim \sqrt[4]{n}$ is simple: consider the mapping $\phi : \mathbb{F}_2^n \to \mathbb{R}$ given by $\phi(x) = \sqrt{\max\{\|x\|_1 - n/2, 0\}}$. Then ϕ is 1-Lipschitz with respect to the metric induced on $\mathbb{F}_2^n$ by the Euclidean norm $\|\cdot\|_2$. By the central limit theorem the average of $|\phi(x)-\phi(y)|^2$ over $(x,y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ is of order $\sqrt{n}$.

The corresponding lower bound $Av_{\mathbb{R}}(\mathbb{F}_2^n, \|\cdot\|_2) \gtrsim \sqrt[4]{n}$ is an example of a lower bound on the average distortion of the cube $\mathbb{F}_2^n$ that is not proved through the use on nonlinear type. Suppose that $f:\mathbb{F}_2^n \to \mathbb{R}$ satisfies $|f(x)-f(y)| \leqslant \|x-y\|_2$ for every $x,y\in\mathbb{F}_2^n$. Suppose also that $S\subseteq \mathbb{F}_2^n$ satisfies $|S|\geqslant 2^{n-1}$. Then by Harper's inequality [Har66] (see also [Led01, Thm. 2.11]), for every $t\in(0,\infty)$ we have

$$\frac{\left|\{x\in\mathbb{F}_2^n:\ \forall\, y\in S,\ \|x-y\|_2\geqslant t\}\right|}{2^n} = \frac{\left|\{x\in\mathbb{F}_2^n:\ \forall\, y\in S,\ \|x-y\|_1\geqslant t^2\}\right|}{2^n} \leqslant e^{-2t^4/n}.$$

Consequently, if $M_f\in\mathbb{R}$ is a median of f then by [Led01, Prop.1.3] we have $\left|\{x\in\mathbb{F}_2^n:\ |f(x)-M_f|\geqslant t\}\right|/2^n \leqslant 2e^{-2t^4/n}$. Hence,

$$\left(\frac{1}{4^n}\sum_{(x,y)\in\mathbb{F}_2^n\times\mathbb{F}_2^n}|f(x)-f(y)|^2\right)^{\frac12} \leqslant 2\left(\frac{1}{2^n}\sum_{(x,y)\in\mathbb{F}_2^n\times\mathbb{F}_2^n}|f(x)-M_f|^2\right)^{\frac12} \leqslant \left(\int_0^\infty 4te^{-2t^4/n}dt\right)^{\frac12} \lesssim \sqrt[4]{n}.$$

Remark 6.10. Additional obstructions to average distortion embeddings that do not fall into the framework described in this section have been obtained in the context of integrality gap lower bounds for the Goemans–Linial semidefinite relaxation for the Uniform Sparsest Cut Problem. The best known result in this direction is due to [KM13] (improving over the works [DKSV06, KR09]), where it is shown that for arbitrarily large $n\in\mathbb{N}$ there exists an n-point metric space (X,d_X) such that the metric space $(X,\sqrt{d_X})$ emebds isometrically into ℓ_2, yet $Av_{\ell_1}^{(1)}(X)\geqslant \exp\left(c\sqrt{\log\log n}\right)$, where $c\in(0,\infty)$ is a universal constant. Finding the correct asymptotic dependence here remains open.

7. Existence of average distortion embeddings

The main purpose of this section is to state criteria for the existence of average distortion embeddings. In what follows we often discuss probability distributions over random subsets or random partitions of metric spaces. To avoid measurability issues we focus our discussion on finite metric spaces. Such topics can be treated for infinite spaces as well, as done in [LN05].

7.1. **Random zero sets.** Fix $\Delta,\zeta\in(0,\infty)$ and $\delta\in(0,1)$. Following [ALN08], a finite metric space (X,d_X) is said to admit a *random* *zero set* at scale Δ which is ζ-spreading with probability δ if there exists a probability distribution μ over 2^X such that every $x, y \in X$ satisfy

$$d_X(x,y) \geqslant \Delta \implies \mu\left(\left\{Z \in 2^X : \ x \in Z \ \wedge \ d_X(y,Z) \geqslant \frac{\Delta}{\zeta}\right\}\right) \geqslant \delta. \tag{178}$$

We denote by $\zeta(X;\delta)$ the infimum over those $\zeta \in (0,\infty)$ such that for every scale $\Delta \in (0,\infty)$ the finite metric space (X,d_X) admits a random zero set at scale Δ which is ζ-spreading with probability δ. If (X,d_X) is an infinite metric space then we write

$$\zeta(X;\delta) \stackrel{\mathrm{def}}{=} \sup_{\substack{S\subseteq X\\ |S|<\infty}} \zeta(X;\delta). \tag{179}$$

The following proposition asserts that random zero sets can be used to obtain embeddings into the real line $\mathbb{R}$ with low average distortion.

Proposition 7.1. *Fix $n \in \mathbb{N}$ and $\delta \in (0,1)$. Suppose that (X,d_X) is a metric space with $\zeta(X;\delta) < \infty$. Then for every $p \in [1,\infty)$ and every $x_1,\ldots,x_n \in X$ there exists a 1-Lipschitz function $f : X \to \mathbb{R}$ such that*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |f(x_i)-f(x_j)|^p \geqslant \frac{\delta}{2^{11p}\zeta(X;\delta)^p}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i,x_j)^p.$$

Thus, using the notation of Section 1.3, $Av_{\mathbb{R}}^{(p)}(X) \lesssim \zeta(X;\delta)/\delta^{1/p}$.

Proposition 7.1 will be proven in Section 7.4 below. We will now explain how Proposition 7.1 can be applied to a variety of metric spaces. Due to the discussion preceding Theorem 1.3, such spaces will satisfy the spectral inequality (7) with Ψ linear.

7.1.1. *Random partitions.* Many spaces are known to admit good random zero sets. Such examples often (though not always) arise from metric spaces for which one can construct *random padded partitions*. If (X,d_X) is a finite metric space let $\mathscr{P}(X)$ denote the set of all partitions of X. For $P \in \mathscr{P}(X)$ and $x \in X$, the unique element of P to which x belongs is denoted $P(x) \subseteq X$. Given $\varepsilon, \delta \in (0,1)$, the metric space (X,d_X) is said to admit an ε-padded random partition with probability δ if for every $\Delta \in (0,\infty)$ there exists a probability distribution μ_Δ over partitions of X with the following properties.

- $\forall\, P \in \mathscr{P}(X), \quad \mu_\Delta(P) > 0 \implies \max_{x\in X} \operatorname{diam} P(x) \leqslant \Delta.$
- For every $x \in X$ we have
$$\mu_\Delta\left(\{P \in \mathscr{P}(X) : \ B_X(x,\varepsilon\Delta) \subseteq P(x)\}\right) \geqslant \delta,$$
where $B_X(x,r) \stackrel{\mathrm{def}}{=} \{y \in X : \ d_X(x,y) \leqslant r\}$ for every $r \in [0,\infty)$.

Let $\varepsilon(X;\delta)$ denote the supremum over those $\varepsilon \in (0,1)$ for which (X,d_X) admits an ε-padded random partition with probability δ. As in (179), we extend this definition to infinite metric spaces (X,d_X) by setting

$$\varepsilon(X;\delta) \stackrel{\mathrm{def}}{=} \inf_{\substack{S\subseteq X\\ |S|<\infty}} \varepsilon(X;\delta).$$

Fact 3.4 in [ALN08] (which itself uses an idea of [Rao99]) asserts that for every $\delta \in (0,1)$, if (X,d_X) is a finite metric space then

$$\varepsilon(X;\delta)\cdot \zeta\left(X;\frac{\delta}{4}\right) \leqslant 1. \tag{180}$$

([ALN08, Fact 3.4] states this for the arbitrary choice $\delta = \frac{1}{2}$, but its proof does not use this specific value of δ in any way.) One should interpret (180) as asserting that a lower bound on $\varepsilon(X;\delta)$ implies an upper bound on $\zeta(X,\delta/4)$. The following classes of metric spaces (X,d_X) are known to satisfy $\varepsilon(X;\delta) > 0$ for some $\delta \in (0,1)$: doubling metric spaces, compact Riemannian surfaces, Gromov hyperbolic spaces of bounded local geometry, Euclidean buildings, symmetric spaces, homogeneous Hadamard manifolds, and forbidden-minor (edge-weighted) graph families. The case of doubling spaces goes back to [Ass83], with subsequent improved bounds on $\varepsilon(X;\delta) > 0$ obtained in [GKL03]. The case of forbidden-minor graph families is due to [KPR93], with subsequent improved bounds on $\varepsilon(X;\delta) > 0$ obtained in [FT03]. The case of compact Riemannian surfaces is due to [LN05], with subsequent improved bounds on $\varepsilon(X;\delta) > 0$ obtained in [LS10]. The remaining cases follow from the general fact [NS11] that if (X,d_X) has bounded Nagata dimension then $\varepsilon(X;\delta) > 0$ for some $\delta \in (0,1)$ (see [LS05] for more information on Nagata dimension of metric spaces). We single out the following two consequences of Proposition 7.1 and (the easy direction of) Theorem 1.3, with explicit quantitative bounds arising from the estimates on $\varepsilon(X;\delta)$ obtained in [GKL03, LS10].

Corollary 7.2. *Suppose that (X,d_X) is a metric space that is doubling with constant $K \in [2,\infty)$. Then for every $n \in \mathbb{N}$ and every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ we have*

$$\gamma\big(A,d_X^2\big) \lesssim \frac{(\log K)^2}{1-\lambda_2(A)}.$$

Corollary 7.3. *Suppose that (X,d_X) is a two dimensional Riemannian manifold of genus $g \in \mathbb{N}\cup\{0\}$. Then for every $n \in \mathbb{N}$ and every* *symmetric stochastic matrix $A \in M_n(\mathbb{R})$ we have*

$$\gamma\big(A, d_X^2\big) \lesssim \frac{(\log(g+1))^2}{1-\lambda_2(A)}.$$

The fact that the conclusion of Proposition 7.1 holds true under the assumption that $\varepsilon(X;\delta) > 0$ for every $\delta \in (0,\infty)$ (as follows by combining Proposition 7.1 with (180)) was proved by Rabinovich in [Rab08] in the case $p = 1$. It has long been well known to experts (and stated explicitly in [BLR10]), that the original proof of Rabinovich extends mutatis mutandis to every $p \in [1,\infty)$. The (simple) proof of Proposition 7.1 below builds on the ideas of Rabinovich in [Rab08].

An example of a class of metric spaces that admits good random zero sets for reasons other than the existence of random padded partitions is the class of spaces that admit a quasisymmetric embedding into Hilbert space. We refer to [Hei01] and the references therein for more information on quasisymmetric embeddings; it suffices to say here that $L_1(\mu)$ spaces provide such examples (see [DL97, Ch. 6]). It follows from [ALN08] (using in part ideas of [ARV09, NRS05, Lee05, CGR08]) that if (X, d_X) is a metric space that admits a quasisymmetric embedding into Hilbert space then there exist $\varepsilon, \delta \in (0,1)$ (depending only on the modulus of quasisymmetry of the implicit embedding) such that for every $n \in \mathbb{N}$, any n-point subset $S \subseteq X$ satisfies $\varepsilon(S;\delta) \geqslant \varepsilon/\sqrt{\log n}$. Consequently we have the following statement.

Corollary 7.4. *Suppose that (X, d_X) is a metric space that admits a quasisymmetric embedding into a Hilbert space. Then there exists a constant $C \in (0,\infty)$ (depending only on the modulus of quasisymmetry of the implicit embedding) such that for every $n \in \mathbb{N}$ and every symmetric stochastic matrix $A \in M_n(\mathbb{R})$ we have*

$$\gamma\big(A, d_X^2\big) \leqslant \frac{C\log n}{1-\lambda_2(A)}. \tag{181}$$

Note that Bourgain's embedding theorem [Bou85] implies that (181) holds true for *every* metric space (X, d_X) if one replaces the term $\log n$ by $(\log n)^2$ (in which case C can be taken to be a universal constant).

7.2. **Localized weakly bi-Lipschitz embeddings.** Following the terminology of [NPSS06], for $D \in [1,\infty)$ say that a metric space (X, d_X) admits a weakly bi-Lipschitz embedding with distortion D into a metric space (Y, d_Y) if for every $\Delta \in (0,\infty)$ there exists a nonconstant Lipschitz mapping $f_\Delta : X \to Y$ such that for every $x, y \in X$,

$$d_X(x,y) \geqslant \Delta \implies d_Y\left(f_\Delta(x), f_\Delta(y)\right) \geqslant \frac{\|f_\Delta\|_{\mathrm{Lip}}}{D}\Delta. \tag{182}$$

The origin of this terminology is that such embeddings preserve (by design) *weak* (p,q) metric Poincaré inequalities. Specifically, a standard way by which one rules out the existence of bi-Lipschitz embeddings is via generalized Poincaré-type inequalities as follows. Suppose that $n\in\mathbb{N}$ and $p,q,K\in(0,\infty)$, and there exist two measures μ,ν on $\{1,\ldots,n\}^2$ such that every $y_1,\ldots,y_n\in Y$ satisfy

$$\left(\sum_{i=1}^n\sum_{j=1}^n d_Y(y_i,y_j)^p\mu(i,j)\right)^{\frac1p}\leqslant\left(\sum_{i=1}^n\sum_{j=1}^n d_Y(y_i,y_j)^q\nu(i,j)\right)^{\frac1q}. \tag{183}$$

Clearly if $f:X\to Y$ is a bi-Lipschitz embedding then the inequality (183) holds for (X,d_X) as well, with the right hand side of (183) multiplied by $\|f\|_{\mathrm{Lip}}\|f^{-1}\|_{\mathrm{Lip}}$. Thus a *strong* (p,q) inequality such as (183) are bi-Lipschitz invariants that can be used to show that certain spaces (X,d_X) must incur large distortion in any bi-Lipschitz embedding into (Y,d_Y). The obvious *weak* (p,q) variant of (183) is the assertion that for every $u\in(0,\infty)$ and every $y_1,\ldots,y_n\in Y$ we have

$$\mu\left(\left\{(i,j)\in\{1,\ldots,n\}^2:\ d_Y(y_i,y_j)\geqslant u\right\}\right)^{\frac1p}\leqslant\frac1u\left(\sum_{i=1}^n\sum_{j=1}^n d_Y(y_i,y_j)^q\nu(i,j)\right)^{\frac1q}. \tag{184}$$

By definition, if a metric space (X,d_X) admits a weakly bi-Lipschitz embedding with distortion D into a metric space (Y,d_Y) satisfying (184) then for every $n\in\mathbb{N}$, any $x_1,\ldots,x_n\in X$ satisfy

$$\mu\left(\left\{(i,j)\in\{1,\ldots,n\}^2:\ d_X(x_i,x_j)\geqslant u\right\}\right)^{\frac1p}\leqslant\frac{D}{u}\left(\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^q\nu(i,j)\right)^{\frac1q}.$$

We will see below that one can prove nonlinear spectral gap inequality such as (6) with Ψ linear by showing that (X,d_X) admits a weakly bi-Lipschitz embedding into (Y,d_Y). To this end it suffices to localize the condition (182) to balls of proportional scale, as follows. For $D\in[1,\infty)$ say that a metric space (X,d_X) admits a *localized weakly bi-Lipschitz embedding* with distortion D into a metric space (Y,d_Y) if for every $z\in X$ and $\Delta\in(0,\infty)$ there exists a non-constant Lipschitz mapping $f_\Delta^z:X\to Y$ such that for every $x,y\in B_X(z,32\Delta)$ we have

$$d_X(x,y)\geqslant\Delta\implies d_Y\left(f_\Delta^z(x),f_\Delta^z(y)\right)\geqslant\frac{\|f_\Delta^z\|_{\mathrm{Lip}}}{D}\Delta. \tag{185}$$

The factor 32 here was chosen to be convenient for the ensuing arguments, but it is otherwise arbitrary.

Proposition 7.5. *Fix $n \in \mathbb{N}$ and $p, D \in [1,\infty)$. Suppose that (X,d_X) is a metric space that admits a localized weakly bi-Lipschitz embedding with distortion D into a metric space (Y,d_Y). Then for every $x_1,\ldots,x_n \in X$ there is a nonconstant mapping $f:\{x_1,\ldots,x_n\}\to Z$, for at least one of the spaces $Z\in\{Y,\mathbb{R}\}$, such that*

$$\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_Z(f(x_i),f(x_j))^p\right)^{\frac{1}{p}} \gtrsim \frac{\|f\|_{\mathrm{Lip}}}{D}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p\right)^{\frac{1}{p}}.$$

Observe that if (Y,d_Y) contains an isometric copy of an interval $[a,b]\subseteq\mathbb{R}$ (in particular if Y is a Banach space) then the conclusion of Proposition 7.5 can be taken to be $Av_Y^{(p)}(X)\lesssim D$.

Lemma 3.5 in [ALN08] asserts that for every $\delta\in(0,1)$, every finite metric space (X,d_X) admits a weakly bi-Lipschitz embedding into ℓ_2 with distortion $\zeta(X;\delta)/\sqrt{\delta}$. Consequently, all the examples that arise from random padded partitions as described in Section 7.1.1 fall into the framework of Proposition 7.5, with the only difference being that an application of Proposition 7.5 rather than Proposition 7.1 yields an embedding into Hilbert space rather than into the real line. This difference is discussed further in Section 7.3 below. The following lemma shows that Proposition 7.5 has wider applicability than Proposition 7.1: in combination with Proposition 7.5 it yields a different proof of the case $p\in(2,\infty)$ of (22) that avoids the use of Theorem 1.3.

Lemma 7.6. *Suppose that $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ are Banach spaces that satisfy the assumptions of Theorem 1.10. Suppose furthermore that there exists $K\in(0,\infty)$ such that $\beta(t)=Kt$ for all $t\in[0,\infty)$. Then $(X,\|\cdot\|_X)$ admits a localized weakly bi-Lipschitz embedding with distortion $2K/\alpha(1/32)$ into $(Y,\|\cdot\|_Y)$.*

Proof. Fix $\Delta\in(0,\infty)$ and a mapping $f:B_X\to Y$ that satisfies (19). For $z\in X$ define $f_\Delta^z:X\to Y$ by

$$f_\Delta^z(x)\stackrel{\mathrm{def}}{=}32\Delta f\left(\rho\left(\frac{x-z}{32\Delta}\right)\right),$$

where ρ is given as in (107). Since ρ is 2-Lipschitz, $\|f_\Delta^z\|_{\mathrm{Lip}}\leqslant 2K$. If $x,y\in z+32\Delta B_X$ satisfy $\|x-y\|_X\geqslant\Delta$ then

$$\|f_\Delta^z(x)-f_\Delta^z(y)\|_Y\geqslant\Delta\alpha\left(\frac{\|x-y\|_X}{32\Delta}\right)\geqslant\frac{\alpha\left(\frac{1}{32}\right)\|f_\Delta^z\|_{\mathrm{Lip}}}{2K}\Delta. \qquad\square$$

For $p \in [1,2)$, due to Lemma 1.12 and Proposition 7.5, ℓ_p does not admit a localized weakly bi-Lipschitz embedding into Hilbert space (this can also be proved directly via a shorter argument). Since finite subsets of ℓ_p embed isometrically into ℓ_1 (see e.g. [DL97]), it follows from [ARV09] that every n-point subset of ℓ_p admits a weakly bi-Lipschitz embedding into ℓ_2 with distortion $O(\sqrt{\log n})$. By [Lee05], it is also true that every n-point subset of ℓ_p admits a weakly bi-Lipschitz embedding into ℓ_2 with distortion $O((\log n)^{(2-p)/p^2})$, which is better than the $O(\sqrt{\log n})$ bound of [ARV09] if $\sqrt{5}-1 < p \leqslant 2$. Therefore for every $n \in \mathbb{N}$ and every n by n symmetric stochastic matrix A,

$$p \in \left[1, \sqrt{5}-1\right] \implies \gamma(A, \|\cdot\|_{\ell_p}^2) \lesssim \frac{\log n}{1-\lambda_2(A)},$$

and

$$p \in \left[\sqrt{5}-1, 2\right] \implies \gamma(A, \|\cdot\|_{\ell_p}^2) \lesssim \frac{(\log n)^{\frac{2(2-p)}{p^2}}}{1-\lambda_2(A)}.$$

These are the currently best known bounds towards Question 1.8.

7.3. **Dimension reduction.** As discussed in Section 7.2, Proposition 7.1 yields an average distortion embedding into the real line, while Proposition 7.5, when applied in the context of spaces with random zero sets, yields an average distortion embedding into Hilbert space. Here we briefly compare these two notions. The following lemma is a simple application of the classical Johnson-Lindenstrauss dimension reduction lemma [JL84].

Lemma 7.7. *If (X, d_X) is an n-point metric space then*

$$\frac{Av_{\mathbb{R}}^{(2)}(X)}{\sqrt{\log n}} \leqslant Av_{\ell_2}^{(2)}(X) \lesssim Av_{\mathbb{R}}^{(2)}(X). \tag{186}$$

Proof. The rightmost inequality in (186) is trivial. Write $D = Av_{\ell_2}^{(2)}(X)$ and take $x_1, \ldots, x_m \in X$. By the Johnson-Lindenstrauss lemma [JL84] there exists $k \in \mathbb{N}$ such that $k \lesssim \log n$ and there exists a 1-Lipschitz function $f = (f_1, \ldots, f_k) : \{x_1, \ldots, x_m\} \to \mathbb{R}^k$ such that

$$\sum_{i=1}^m \sum_{j=1}^m \|f(x_i) - f(x_j)\|_2^2 \geqslant \frac{1}{2D^2} \sum_{i=1}^m \sum_{j=1}^m d_X(x_i, x_j)^2.$$

Therefore there exists $s \in \{1, \ldots, k\}$ such that

$$\sum_{i=1}^m \sum_{j=1}^m |f_s(x_i) - f_s(x_j)|^2 \geqslant \frac{1}{2kD^2} \sum_{i=1}^m \sum_{j=1}^m d_X(x_i, x_j)^2.$$

Since $f_s : \{x_1, \ldots, x_m\} \to \mathbb{R}$ is also 1-Lipschitz, we conclude that $Av_{\mathbb{R}}^{(2)}(X) \leqslant \sqrt{2k}D \lesssim \sqrt{\log n}D$. □

The following lemma shows that Lemma 7.7 is almost asymptotically sharp.

Lemma 7.8. *For arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space X_n such that*

$$Av_{\mathbb{R}}^{(2)}(X_n) \gtrsim Av_{\ell_2}^{(2)}(X_n) \cdot \sqrt{\frac{\log n}{\log\log n}}. \tag{187}$$

Proof. Fix $\varepsilon \in (0,1)$ and an integer $m \geqslant 2$. Let σ denote the normalized surface measure on the unit sphere $S^{m-1} \subseteq \ell_2^m$. By Lemma 21 in [FS02] there exists a partition $\{C_1, \ldots, C_n\}$ of S^{m-1} into nonempty measurable sets such that $\sigma(C_i) = 1/n$ and $\mathrm{diam}(C_i) \leqslant \varepsilon$ for all $i \in \{1, \ldots, n\}$, and $n \leqslant (\kappa/\varepsilon)^m$ for some universal constant $\kappa \in (0, \infty)$.

Choose an arbitrary point $x_i \in C_i$ and set $X_n = \{x_1, \ldots, x_n\} \subseteq \ell_2^{m-1}$. Since X_n is isometric to a subset of Hilbert space, $Av_{\ell_2}^{(2)}(X_n) = 1$. Suppose that $f : \{x_1, \ldots, x_n\} \to \mathbb{R}$ is a 1-Lipschitz function. By the nonlinear Hahn-Banach theorem (see [BL00]) we can think of f as the restriction to X_n of a 1-Lipschitz function defined on all of S^{m-1}. The Poincaré inequality on the sphere S^{m-1} (see e.g. [Cha84, Led01]) asserts that

$$\int_{S^{m-1}} \int_{S^{m-1}} |f(x) - f(y)|^2 d\sigma(x) d\sigma(y) \leqslant \frac{2}{m-1} \int_{S^{m-1}} \|\nabla f(x)\|_2^2 \, d\sigma(x) \leqslant \frac{4}{m}. \tag{188}$$

For every $i, j \in \{1, \ldots, n\}$ and every $(x, y) \in C_i \times C_j$ we have

$$\begin{aligned}\frac{|f(x_i) - f(x_j)|^2}{3} &\leqslant |f(x_i) - f(x)|^2 + |f(x) - f(y)|^2 + |f(y) - f(x_j)|^2 \\ &\leqslant |f(x) - f(y)|^2 + \mathrm{diam}(C_i)^2 + \mathrm{diam}(C_j)^2 \\ &\leqslant |f(x) - f(y)|^2 + 2\varepsilon^2, \end{aligned} \tag{189}$$

and similarly,

$$\|x - y\|_2^2 \leqslant 3\|x_i - x_j\|_2^2 + 6\varepsilon^2. \tag{190}$$

Consequently,

$$\frac{|f(x_i) - f(x_j)|^2 - 6\varepsilon^2}{n^2} = \sigma(C_i)\sigma(C_j)\left(|f(x_i) - f(x_j)|^2 - 6\varepsilon^2\right) \stackrel{(189)}{\leqslant} 3 \int_{C_i} \int_{C_j} |f(x) - f(y)|^2 d\sigma(x) d\sigma(y). \tag{191}$$

and

$$\frac{\|x_i - x_j\|_2^2}{n^2} \geqslant \frac{1}{3}\int_{C_i}\int_{C_j} \|x-y\|_2^2 d\sigma(x)d\sigma(y) - \frac{2\varepsilon^2}{n^2}. \tag{192}$$

Hence,

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |f(x_i)-f(x_j)|^2\\ &\quad\overset{(191)}{\leqslant} 3\int_{S^{m-1}}\int_{S^{m-1}} |f(x)-f(y)|^2 d\sigma(x)d\sigma(y) + 6\varepsilon^2 \overset{(188)}{\leqslant} \frac{12}{m}+6\varepsilon^2,\end{aligned}$$

and

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|x_i-x_j\|_2^2 &\overset{(192)}{\geqslant} \frac{1}{3}\int_{S^{m-1}}\int_{S^{m-1}} \|x-y\|_2^2 d\sigma(x)d\sigma(y) - 2\varepsilon^2\\ &= \frac{2}{3}-2\varepsilon^2.\end{aligned}$$

By choosing $\varepsilon = 1/(2\sqrt{m})$, we have shown that every 1-Lipschitz function $f : X_n \to \mathbb{R}$ satisfies

$$\sum_{i=1}^n\sum_{j=1}^n |f(x_i)-f(x_j)|^2 \lesssim \frac{1}{m}\sum_{i=1}^n\sum_{j=1}^n \|x_i-x_j\|_2^2.$$

Recalling that $n \leqslant (\kappa/\varepsilon)^m = (2\kappa\sqrt{m})^m$, or $m \gtrsim \sqrt{\log n/\log\log n}$, the proof of (187) is complete. $\square$

Remark 7.9. Given an n-point metric space (X, d_X) let $\mathfrak{S}(X)$ denote the maximum of $\frac{1}{2n^2}\sum_{(x,y)\in X\times X}|f(x)-f(y)|^2$ over all 1-Lipschitz functions $f : X \to \mathbb{R}$. The quantity $\mathfrak{S}(X)$ was introduced by Alon, Boppana and Spencer in [ABS98], where they called it the *spread constant of* X. They proved that the spread constant of X governs the asymptotic isoperimetric behavior of $\ell_1^n(X)$ as $n \to \infty$. They also state that "The spread constant appears to be new and may well be of independent interest." We agree with this assertion. In particular, it would be worthwhile to investigate the computational complexity of the problem that takes as input an n-point metric space (X, d_X) and is supposed to output in polynomial time a number that is guaranteed to be a good approximation of its spread constant. We are not aware of hardness of approximation results for this question. Let $\mathfrak{S}_{\ell_2}(X)$ denote the maximum of $\frac{1}{2n^2}\sum_{(x,y)\in X\times X}\|f(x)-f(y)\|_2^2$ over all 1-Lipschitz functions $f : X \to \ell_2$. The quantity $\mathfrak{S}_{\ell_2}(X)$ can be computed in polynomial time with arbitrarily good precision, since (by definition) it can be cast as a semidefinite program (see [GLS93]). The proof of Lemma 7.7

can be viewed as a simple approximation algorithm to the spread constant, achieving an approximation guarantee of $O(\log n)$. Lemma 7.8 can be viewed as yielding an almost matching integrality gap lower bound for the semidefinite program. Note that the parameter $\mathfrak{S}_{\ell_2}(X)$ itself has also been studied in the literature in the context of the problem of finding the fastest mixing Markov process on a given graph; see [SBXD06]. See also the works [Fie89, GHW08, GHR12] that study this quantity in the context of the absolute algebraic connectivity of a graph. Clearly $(Av_{\mathbb{R}}^{(2)}(X))^2$ is closely related to $\mathfrak{S}(X)$: it amounts to finding the (multi)subset of X with largest spread constant. The same can be said for the relation between $(Av_{\ell_2}^{(2)}(X))^2$ and $\mathfrak{S}_{\ell_2}(X)$.

7.4. **Proofs of Proposition 7.1 and Proposition 7.5.** We start by recording the following very simple lemma, whose proof is a straightforward application of the triangle inequality.

Lemma 7.10. *Fix $p \in [1,\infty)$ and $n \in \mathbb{N}$. Let (X,d_X) be a metric space and $x_1,\ldots,x_n, \in X$. Define*

$$r \stackrel{\mathrm{def}}{=} \min_{i\in\{1,\ldots,n\}} \left(\frac{1}{n}\sum_{j=1}^n d_X(x_i,x_j)^p\right)^{\frac{1}{p}}. \tag{193}$$

Then

$$r \leqslant \left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p\right)^{\frac{1}{p}} \leqslant 2r. \tag{194}$$

Proof. Choose $k \in \{1,\ldots,n\}$ such that

$$r = \left(\frac{1}{n}\sum_{j=1}^n d_X(x_k,x_j)^p\right)^{\frac{1}{p}}. \tag{195}$$

The rightmost inequality in (194) follows from averaging the estimate $d_X(x_i,x_j)^p \leqslant 2^{p-1}d_X(x_i,x_k)^p + 2^{p-1}d_X(x_k,x_j)^p$ over $i,j \in \{1,\ldots,n\}$. The leftmost inequality in (194) follows from the definition of r. $\square$

Lemma 7.11. *Continuing with the notation of the statement and proof of Lemma 7.10, in particular choosing $k \in \{1,\ldots,n\}$ so as to satisfy* (195)*, write*

$$B \stackrel{\mathrm{def}}{=} \{i \in \{1,\ldots,n\}:\ d_X(x_i,x_k) \leqslant 4r\}, \tag{196}$$

and

$$M \stackrel{\mathrm{def}}{=} \left\{(i,j) \in B\times B:\ d_X(x_i,x_j) \geqslant \frac{r}{8}\right\}. \tag{197}$$

Then,

$$|M| \leqslant \frac{n^2}{2^{7p}} \implies \left(\frac{1}{n}\sum_{j\in\{1,\ldots,n\}\smallsetminus B} d_X(x_j,x_k)^p\right)^{\frac{1}{p}} \geqslant \frac{r}{8}. \tag{198}$$

Proof. Suppose that

$$|M| \leqslant \frac{n^2}{2^{7p}}. \tag{199}$$

Recalling (195) and (196), it follows from Markov's inequality that

$$\frac{n-|B|}{n} \leqslant \frac{1}{4^p}. \tag{200}$$

Hence,

$$\begin{aligned}
&\frac{2}{n^2}\sum_{i=1}^n \sum_{j\in\{1,\ldots,n\}\smallsetminus B} d_X(x_i,x_j)^p\\
&\quad\leqslant \frac{2^p}{n^2}\sum_{i=1}^n \sum_{j\in\{1,\ldots,n\}\smallsetminus B} \left(d_X(x_i,x_k)^p + d_X(x_k,x_j)^p\right)\\
&\quad\overset{(195)}{=} \frac{2^p(n-|B|)}{n}r^p + \frac{2^p}{n}\sum_{j\in\{1,\ldots,n\}\smallsetminus B} d_X(x_k,x_j)^p\\
&\quad\overset{(200)}{\leqslant} \frac{r^p}{2^p} + \frac{2^p}{n}\sum_{j\in\{1,\ldots,n\}\smallsetminus B} d_X(x_k,x_j)^p.
\end{aligned} \tag{201}$$

Since every $i,j\in B$ satisfy $d_X(x_i,x_j)\leqslant d_X(x_i,x_k)+d_X(x_j,x_k)\leqslant 8r$,

$$\begin{aligned}
\frac{1}{n^2}\sum_{(i,j)\in B\times B} d_X(x_i,x_j)^p &\overset{(196)\wedge(197)}{\leqslant} \frac{8^p r^p|M|}{n^2} + \frac{|B|^2-|M|}{n^2}\cdot\frac{r^p}{8^p}\\
&\overset{(199)}{\leqslant} \frac{r^p}{2^{4p}} + \frac{r^p}{2^{3p}}.
\end{aligned} \tag{202}$$

It follows that

$$\begin{aligned}
r^p &\overset{(194)}{\leqslant} \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p\\
&\overset{(201)\wedge(202)}{\leqslant} \frac{r^p}{2^{4p}} + \frac{r^p}{2^{3p}} + \frac{r^p}{2^p} + \frac{2^p}{n}\sum_{j\in\{1,\ldots,n\}\smallsetminus B} d_X(x_k,x_j)^p,
\end{aligned}$$

which (since $p\geqslant 1$) implies that

$$\frac{1}{n}\sum_{j\in\{1,\ldots,n\}\smallsetminus B} d_X(x_k,x_j)^p \geqslant \frac{r^p}{8^p}. \qquad \square$$

Lemma 7.12. *Continuing with the notation of the statements and proofs of Lemma 7.10 and Lemma 7.11, define*

$$\forall\, i\in\{1,\ldots,n\},\qquad s_i \stackrel{\mathrm{def}}{=} \max\left\{0,d_X(x_i,x_k)-2r\right\}.$$

If $|M|\leqslant n^2/2^{7p}$ *then*

$$\sum_{i=1}^n\sum_{j=1}^n |s_i-s_j|^p \geqslant \frac{1}{2^{5p}}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_k)^p.$$

Proof. Due to Lemma 7.11 we know that

$$\frac{1}{n}\sum_{j\in\{1,\ldots,n\}\smallsetminus B} d_X(x_k,x_j)^p \geqslant \frac{r^p}{8^p} \tag{203}$$

Define $B'\stackrel{\mathrm{def}}{=}\{i\in\{1,\ldots,n\}:\ d_X(x_i,x_k)\leqslant 2r\}$. If $i\in B'$ then $s_i=0$, and if $j\in\{1,\ldots,n\}\smallsetminus B$ then $s_j=d_X(x_j,x_k)-2r\geqslant \frac12 d_X(x_j,x_k)$. Also, recalling (195), it follows from Markov's inequality that

$$\frac{n-|B'|}{n}\leqslant \frac{1}{2^p}. \tag{204}$$

Consequently,

$$\begin{aligned}\sum_{i=1}^n\sum_{j=1}^n |s_i-s_j|^p &\geqslant 2|B'|\sum_{j\in\{1,\ldots,n\}\smallsetminus B}\frac{d_X(x_j,x_k)^p}{2^p}\\ &\stackrel{(203)\wedge(204)}{\geqslant} \frac{n\left(1-\frac{1}{2^p}\right)}{2^{p-1}}\cdot\frac{nr^p}{8^p}\stackrel{(194)}{\geqslant}\frac{1}{2^{5p}}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p. \qquad\square\end{aligned}$$

Proof of Proposition 7.1. Fix $n\in\mathbb{N}$. Suppose that $x_1,\ldots,x_n\in X$ and write $S=\{x_1,\ldots,x_n\}\subseteq X$. Define $r\in(0,\infty)$ as in (193) and let μ be a probability distribution over 2^X that satisfies (178) with $\Delta=r/8$ and $\zeta=\zeta(S;\delta)$. Let M be defined as in (197) and suppose that $|M|>n^2/2^{7p}$. Then

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\int_{2^X}|d_X(x_i,Z)-d_X(x_j,Z)|^p\,d\mu(Z)\\ &\geqslant \frac{1}{n^2}\sum_{(i,j)\in M}\left(\frac{r}{8\zeta}\right)^p\mu\left(\left\{Z\in 2^X:\ x_i\in Z\ \wedge\ d_X(x_j,Z)\geqslant\frac{r}{8\zeta}\right\}\right)\\ &\geqslant \frac{|M|}{n^2}\cdot\frac{\delta r^p}{8^p\zeta^p} &(205)\\ &\geqslant \frac{\delta}{2^{11p}\zeta^p}\cdot\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p, &(206)\end{aligned}$$

where (205) uses (178) and (197), and (206) uses (194) and the assumption $|M| > n^2/2^{7p}$. It follows that there exists $Z \subseteq X$ such that the 1-Lipschitz function $f : X \to \mathbb{R}$ given by $f(x) = d_X(x, Z)$ satisfies

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |f(x_i) - f(x_j)|^p \geqslant \frac{\delta}{2^{11p}\zeta^p}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i, x_j)^p.$$

If, on the hand, $|M| \leqslant n^2/2^{7p}$ then the desired estimate follows by choosing $f(x) = \max\{0, d_X(x, x_k) - 2r\}$ and applying Lemma 7.12. $\square$

Proof of Proposition 7.5. Fix $n \in \mathbb{N}$ and choose $x_1, \ldots, x_n \in X$. Define $r \in (0, \infty)$ as in (193) and $k \in \{1, \ldots, n\}$ as in (195). Let M be defined as in (197) and suppose that $|M| > n^2/2^{7p}$. An application of (185) with $\Delta = r/8$ and $z = x_k$ shows that

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} d_Y\left(f_\Delta^{x_k}(x_i), f_\Delta^{x_k}(x_j)\right)^p &\geqslant \sum_{(i,j)\in M} d_Y\left(f_\Delta^{x_k}(x_i), f_\Delta^{x_k}(x_j)\right)^p \\ &\geqslant |M|\frac{\|f\|_{\mathrm{Lip}}^p}{D^p}\cdot\frac{r^p}{8^p} \geqslant \frac{n^2 r^p\|f\|_{\mathrm{Lip}}^p}{2^{10p}D^p} \geqslant \frac{\|f\|_{\mathrm{Lip}}^p}{2^{11p}D^p}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i, x_j)^p.\end{aligned}$$

This yields the desired average distortion embedding into Y. If, on the hand, $|M| \leqslant n^2/2^{7p}$ then the existence of the desired embedding into $\mathbb{R}$ follows by choosing $f(x) = \max\{0, d_X(x, x_k) - 2r\}$ and applying Lemma 7.12. $\square$

It is natural to ask how the quantities $Av_Y^{(p)}(X)$ and $Av_Y^{(q)}(X)$ are related to each other for distinct $p, q \in [1, \infty)$ and two metric space (X, d_X) and (Y, d_Y). We shall now briefly address this matter.

Suppose that $D > Av_Y^{(q)}(X)$ and fix $x_1, \ldots, x_n \in X$. Then there exists a nonconstant mapping $f : \{x_1, \ldots, x_n\} \to Y$ such that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} d_Y(f(x_i), f(x_j))^q \geqslant \frac{\|f\|_{\mathrm{Lip}}^q}{D^q}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i, x_j)^q. \tag{207}$$

Suppose first that $q < p$, and continue using the notation of Lemma 7.10 and Lemma 7.11. If $|M| > n^2/2^{7p}$ then

$$\begin{aligned}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i, x_j)^q\right)^{\frac{1}{q}} &\geqslant \left(\frac{1}{n^2}\sum_{(i,j)\in M} d_X(x_i, x_j)^q\right)^{\frac{1}{q}} \\ &\overset{(197)}{\geqslant} \left(\frac{1}{2^{7p}}\cdot\frac{r^q}{8^q}\right)^{\frac{1}{q}} \overset{(194)}{\gtrsim} \frac{1}{2^{7p/q}}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i, x_j)^p\right)^{\frac{1}{p}}. \end{aligned} \tag{208}$$

Consequently,

$$\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_Y(f(x_i),f(x_j))^p\right)^{\frac{1}{p}} \geqslant \left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_Y(f(x_i),f(x_j))^q\right)^{\frac{1}{q}}$$
$$\overset{(207)\wedge(208)}{\gtrsim} \frac{\|f\|_{\mathrm{Lip}}}{2^{7p/q}D}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_X(x_i,x_j)^p\right)^{\frac{1}{p}}. \quad (209)$$

By using Lemma 7.12 if $|M| \leqslant n^2/2^{7p}$, we deduce from (209) that

$$q < p \implies Av^{(p)}_{Y\times\mathbb{R}}(X) \lesssim 2^{7p/q} Av^{(q)}_{Y}(X), \quad (210)$$

where $Y\times\mathbb{R}$ is understood to be equipped with, say, the ℓ_1-sum metric, i.e., $d_{Y\times\mathbb{R}}((x,s),(y,t)) = d(x,y)+|s-t|$ for every $(x,s),(y,t)\in Y\times\mathbb{R}$ (as in Proposition 7.5, we can conclude here that there exists an average distortion embedding into either Y or $\mathbb{R}$, but we choose to work with $Y\times\mathbb{R}$ for notational simplicity).

If $p<q$ then the following argument establishes an estimate analogous to (210), but under an additional assumption. The Lipschitz extension constant for the pair of metric spaces (X,Y), denoted $e(X,Y)$, is the infimum over those $K\in[1,\infty)$ such that for every $S\subseteq X$ every Lipschitz function $f:S\to Y$ admits an extension $F:X\to Y$ with $\|F\|_{\mathrm{Lip}}\leqslant K\|f\|_{\mathrm{Lip}}$. If no such K exists then set $e(X,Y)=\infty$. Suppose that $p<q$ and that $|M|>n^2/2^{7p}$. By the definition of D there exists a nonconstant mapping $\phi:B\to Y$ such that

$$\sum_{(i,j)\in B\times B} d_Y(\phi(x_i),\phi(x_j))^q \geqslant \frac{\|\phi\|_{\mathrm{Lip}}^q}{D^q}\sum_{(i,j)\in B\times B} d_X(x_i,x_j)^q. \quad (211)$$

Note that since for every $i,j\in B$ we have $d_X(x_i,x_j)\leqslant 8r$,

$$\sum_{(i,j)\in B\times B} d_Y(\phi(x_i),\phi(x_j))^q$$
$$\leqslant (8r\|\phi\|_{\mathrm{Lip}})^{q-p}\sum_{(i,j)\in B\times B} d_Y(\phi(x_i),\phi(x_j))^p. \quad (212)$$

Also, arguing as in (208) we have

$$\frac{1}{n^2}\sum_{(i,j)\in B\times B} d_X(x_i,x_j)^q \gtrsim (cr)^q, \quad (213)$$

where $c \in (0,\infty)$ is a universal constant. By substituting (212) and (213) into (211) we therefore have

$$\left(\frac{1}{n^2}\sum_{(i,j)\in B\times B} d_Y(\phi(x_i),\phi(x_j))^p\right)^{\frac1p} \geqslant \frac{(c/8)^{q/p}\|\phi\|_{\mathrm{Lip}} r}{D^{q/p}}$$
$$\stackrel{(194)}{\geqslant} \frac{(c/8)^{q/p}\|\phi\|_{\mathrm{Lip}}}{D^{q/p}}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p\right)^{\frac1p}. \quad (214)$$

By extending ϕ to a mapping $\Phi : X \to Y$ with $\|\Phi\|_{\mathrm{Lip}} \lesssim e(X,Y)\|\phi\|_{\mathrm{Lip}}$ we see that

$$\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_Y(\Phi(x_i),\Phi(x_j))^p\right)^{\frac1p} \geqslant \left(\frac{1}{n^2}\sum_{(i,j)\in B\times B} d_Y(\phi(x_i),\phi(x_j))^p\right)^{\frac1p}$$
$$\stackrel{(214)}{\geqslant} \frac{\|\Phi\|_{\mathrm{Lip}}}{(CD)^{q/p}e(X,Y)}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_X(x_i,x_j)^p\right)^{\frac1p}, \quad (215)$$

where $C \in (0,\infty)$ is a universal constant. By using Lemma 7.12 if $|M| \leqslant n^2/2^{7p}$, we deduce from (215) that

$$p<q \implies Av^{(p)}_{Y\times\mathbb{R}}(X) \lesssim e(X,Y)\left(CAv^{(q)}_Y(X)\right)^{q/p}. \quad (216)$$

By [Bal92, NPSS06] for $p \in [2,\infty)$ we have $e(\ell_p,\ell_2) \lesssim \sqrt{p}$. It therefore follows from (210) and (216) combined with Corollary 1.6 that for every $p\in[2,\infty)$ and $q\in[1,\infty)$ we have

$$Av^{(q)}_{\ell_2}(\ell_p) \lesssim \begin{cases} 2^{4q}p & \text{if } q \geqslant 2,\\ p^{\frac2q+\frac12} & \text{if } q<2.\end{cases} \quad (217)$$

It seems unlikely that (217) is sharp.

Appendix: a refinement of Markov type

Below is an application of Theorem 1.5 that I found in collaboration with Yuval Peres. I thank him for agreeing to include it here.

Fix $n \in \mathbb{N}$ and an n by n symmetric stochastic matrix $A = (a_{ij})$. Then for every $m \in \mathbb{N}$ and $x_1,\ldots,x_n \in \mathbb{R}$ we have

$$\frac{\sum_{i=1}^n\sum_{j=1}^n (A^m)_{ij}(x_i-x_j)^2}{\sum_{i=1}^n\sum_{j=1}^n a_{ij}(x_i-x_j)^2} \leqslant \sum_{t=0}^{m-1}\lambda_2(A)^t. \quad (218)$$

(218) becomes evident when one expresses the vector $(x_1,\ldots,x_n)\in\mathbb{R}^n$ in an orthonormal eigenbasis of A, showing also that the multiplicative

factor $1 + \lambda_2(A) + \ldots + \lambda_2(A)^{m-1}$ is sharp. By using the estimate $|\lambda_2(A)| \leqslant 1$, it follows from (218) that Hilbert space has Markov type 2 with $M_2(\ell_2) = 1$; this was Ball's original proof of this fact in [Bal92].

Suppose that $p \in (2, \infty)$. In [NPSS06] it was shown that ℓ_p has Markov type 2, and in fact that $M_2(\ell_p) \lesssim \sqrt{p}$. This is the same as asserting the following inequality, which holds true for every n by n symmetric stochastic matrix $A = (a_{ij})$ and every $x_1, \ldots, x_n \in \ell_p$.

$$\forall\, m \in \mathbb{N}, \qquad \frac{\sum_{i=1}^n \sum_{j=1}^n (A^m)_{ij} \|x_i - x_j\|_{\ell_p}^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} \|x_i - x_j\|_{\ell_p}^2} \lesssim pm. \tag{219}$$

It is natural to ask whether or not one can refine inequality (219) in the spirit of (218) so as to yield an estimate in terms of $\lambda_2(A)$ that becomes (219) if one uses the a priori bound $|\lambda_2(A)| \leqslant 1$. Below we will show how a combination of [NPSS06] and Theorem 1.5 yields the following estimate, thus answering this question positively.

$$\frac{\sum_{i=1}^n \sum_{j=1}^n (A^m)_{ij} \|x_i - x_j\|_{\ell_p}^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} \|x_i - x_j\|_{\ell_p}^2} \lesssim p \sum_{t=0}^{m-1} \left(1 - \frac{2}{p}\left(1 - \lambda_2(A)\right)\right)^t. \tag{220}$$

To explain the similarity of (220) to (218), note that if $m \geqslant p/2$ then (220) has the following equivalent form.

$$\frac{\sum_{i=1}^n \sum_{j=1}^n (A^m)_{ij} \|x_i - x_j\|_{\ell_p}^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} \|x_i - x_j\|_{\ell_p}^2} \lesssim p^2 \sum_{t=1}^{\lceil 2m/p \rceil} \lambda_2(A)^t. \tag{221}$$

Also, if $\lambda_2(A)$ is positive and bounded away from 0, say, if $\lambda_2(A) \geqslant 1/2$, then inequality (220) has the following equivalent form.

$$\frac{\sum_{i=1}^n \sum_{j=1}^n (A^m)_{ij} \|x_i - x_j\|_{\ell_p}^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} \|x_i - x_j\|_{\ell_p}^2} \lesssim p \sum_{t=0}^{m-1} \lambda_2(A)^{2t/p}. \tag{222}$$

The proof of Lemma 7.13 below is a simple variant of an unpublished argument of Mark Braverman (2009), who proved the same statement for $p = 1$; see Exercise 13.10 in [LP13]. The factor 2^p in inequality (223) below is asymptotically sharp as $n \to \infty$: this follows mutatis mutandis from an unpublished argument of Oded Schramm (2007), who proved the same statement when $p = 1$; see Exercise 13.10 in [LP13].

Lemma 7.13. *Fix $p \in [1, \infty)$ and a metric space (X, d_X). Fix also $n \in \mathbb{N}$ and an n by n symmetric stochastic matrix $A = (a_{ij})$. Then for every $s, t \in \mathbb{N}$ with $t \geqslant s$ and every $x_1, \ldots, x_n \in X$ we have*

$$\sum_{i=1}^n \sum_{j=1}^n (A^{2s})_{ij} d_X(x_i, x_j)^p \leqslant 2^p \sum_{i=1}^n \sum_{j=1}^n (A^t)_{ij} d_X(x_i, x_j)^p. \tag{223}$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(A^{2s+1})_{ij}d_X(x_i,x_j)^p \leqslant 3^p\sum_{i=1}^{n}\sum_{j=1}^{n}(A^{t})_{ij}d_X(x_i,x_j)^p. \qquad (224)$$

Proof. We have $d_X(x_i,x_j)^p \leqslant 2^{p-1}d_X(x_i,x_k)^p + 2^{p-1}d_X(x_j,x_k)^p$ for every $i,j,k \in \{1,\ldots,n\}$. For every $\ell \in \{1,\ldots,n\}$ multiply this inequality by $(A^s)_{i\ell}(A^s)_{j\ell}(A^{t-s})_{k\ell}$ and sum the resulting inequality over $i,j,k,\ell \in \{1,\ldots,n\}$, thus obtaining the following estimate.

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n}(A^s)_{i\ell}(A^s)_{j\ell}(A^{t-s})_{k\ell}d_X(x_i,x_j)^p\\ &\leqslant 2^{p-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n}(A^s)_{i\ell}(A^s)_{j\ell}(A^{t-s})_{k\ell}d_X(x_i,x_k)^p\\ &\quad+2^{p-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n}(A^s)_{i\ell}(A^s)_{j\ell}(A^{t-s})_{k\ell}d_X(x_j,x_k)^p. \qquad (225)\end{aligned}$$

Since A is symmetric and stochastic, (225) is the same as (223).

To deduce (224), observe that the convexity of $u \mapsto |u|^p$ implies that for every $i,j,k \in \{1,\ldots,n\}$ we have

$$d_X(x_i,x_j)^p \leqslant \left(\frac{3}{2}\right)^{p-1} d_X(x_i,x_k)^p + 3^{p-1}d_X(x_j,x_k)^p. \qquad (226)$$

(224) now follows by multiplying (226) by $(A^{2s})_{ik}a_{kj}$, summing the resulting inequality over $i,j,k \in \{1,\ldots,n\}$, and using (223). $\square$

Corollary 7.14. *Fix $p \in [1,\infty)$ and $m,n \in \mathbb{N}$. Suppose that $A=(a_{ij})$ is an n by n symmetric stochastic matrix. Then for every metric space (X,d_X) and every $x_1,\ldots,x_n \in X$ we have*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(A^{m})_{ij}d_X(x_i,x_j)^p \leqslant 3^p\gamma(A,d_X^p)\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}d_X(x_i,x_j)^p.$$

Proof. We may assume without loss of generality that $\gamma(A,d_X^p)<\infty$, i.e., that A is ergodic. In this case we have $\lim_{t\to\infty}(A^t)_{ij}=1/n$ for every $i,j \in \{1,\ldots,n\}$. Therefore by Lemma 7.13 (with $t \to \infty$),

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}(A^{m})_{ij}d_X(x_i,x_j)^p &\leqslant \frac{3^p}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}d_X(x_i,x_j)^p\\ &\overset{(3)}{\leqslant} 3^p\gamma(A,d_X^p)\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}d_X(x_i,x_j)^p. \qquad \square\end{aligned}$$

The following corollary is an immediate consequence of Corollary 7.14 and the definition of the Markov type p constant $M_p(X;m)$.

Corollary 7.15. *Fix $p \in [1,\infty)$ and let (X,d_X) be a metric space with Markov type p, i.e., $M_p(X)<\infty$. Then for every $m,n\in\mathbb{N}$, every n by n symmetric stochastic matrix $A=(a_{ij})$ and every $x_1,\ldots,x_n\in X$,*

$$\frac{\sum_{i=1}^n\sum_{j=1}^n (A^m)_{ij}d_X(x_i,x_j)^p}{\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^p} \leqslant \min\left\{M_p(X;m)^p m, 3^p\gamma(A,d_X^p)\right\}.$$

Assume from now on that $n\geqslant 3$, so that $\lambda_2(A)\geqslant -1/2$ (recall Lemma 2.2). Since $1+(1-u)+\ldots+(1-u)^{m-1}\asymp \min\{m,1/u\}$ for every $u\in(0,3/2]$ and $m\in\mathbb{N}$, inequality (220) is a consequence of Corollary 7.15 (with $X=\ell_p$ and $p=2$), Theorem 1.5, and (219).

To state two additional examples of consequences of this type, fix $K\in[2,\infty)$ and let (X,d_X) be a metric space that is doubling with constant K. In [DLP13] it is shown that $M_2(X)\lesssim \log K$, so in combination with Corollary 7.2 we deduce from Corollary 7.15 that

$$\frac{\sum_{i=1}^n\sum_{j=1}^n (A^m)_{ij}d_X(x_i,x_j)^2}{\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^2} \lesssim (\log K)^2\sum_{t=0}^{m-1}\lambda_2(A)^t. \tag{227}$$

Similarly, it was shown that if G is a connected planar graph then $M_2(G,d_G)\lesssim 1$, so in combination with Corollary 7.2 we deduce from Corollary 7.15 that

$$\frac{\sum_{i=1}^n\sum_{j=1}^n (A^m)_{ij}d_X(x_i,x_j)^2}{\sum_{i=1}^n\sum_{j=1}^n a_{ij}d_X(x_i,x_j)^2} \lesssim \sum_{t=0}^{m-1}\lambda_2(A)^t. \tag{228}$$

Despite the validity of satisfactory spectral estimates such as (220), (221), (222), (227) and (228), they do not follow automatically only from the fact that the metric space in question has Markov type 2. Specifically, there exists a metric space (X,d_X) that has Markov type 2, yet it is not true that $\gamma(A,d_X^2)\lesssim_X 1/(1-\lambda_2(A))$ for every symmetric stochastic matrix A. To see this, by Theorem 1.3 it suffices to prove the following result.

Theorem 7.16. *There is a metric space (X,d_X) with $Av_{\ell_2}^{(2)}(X)=\infty$, yet $M_2(X)<\infty$, i.e., X has Markov type 2.*

Proof. If $\Lambda\subseteq\mathbb{R}^n$ is a lattice of rank n then denote the length of the shortest nonzero vector in Λ by $N(\Lambda)$. Also, let $r(\Lambda)$ denote the infimum over those $r\in(0,\infty)$ such that Euclidean balls of radius r centered at Λ cover $\mathbb{R}^n$. The dual lattice of Λ is denoted Λ^*; thus Λ^* is the set of all $x\in\mathbb{R}^n$ such that $\sum_{i=1}^n x_iy_i$ is an integer for every $y\in\Lambda$.

For every $n \in \mathbb{N}$ choose an arbitrary rank n lattice $\Lambda_n \subseteq \mathbb{R}^n$ that satisfies $r(\Lambda_n) \lesssim N(\Lambda_n)$. See [Rog50] for the existence of such lattices. Let $\mathbb{R}^n/\Lambda_n^*$ be the corresponding flat torus, equipped with the natural Riemannian quotient metric $d_{\mathbb{R}^n/\Lambda_n^*}(\cdot,\cdot)$. Also, let μ_n denote the normalized Riemannian volume measure on $\mathbb{R}^n/\Lambda_n^*$.

Consider the ℓ_2 product

$$X \stackrel{\mathrm{def}}{=} \left(\bigoplus_{n=1}^{\infty} (\mathbb{R}^n/\Lambda_n^*)\right)_2 .$$

Thus X consists of all the sequences $x = (x_n)_{n=1}^{\infty} \in \prod_{n=1}^{\infty} (\mathbb{R}^n/\Lambda_n^*)$ such that $\sum_{n=1}^{\infty} d_{\mathbb{R}^n/\Lambda_n^*}(x_n, 0)^2 < \infty$, equipped with the metric given by $d_X(x,y)^2 = \sum_{n=1}^{\infty} d_{\mathbb{R}^n/\Lambda_n^*}(x_n, y_n)^2$ for every $x, y \in X$. Since $\mathbb{R}^n/\Lambda_n^*$ has vanishing sectional curvature, it is an Aleksandrov space of nonnegative curvature (see [Oht09b, Sec. 3]), and therefore by a theorem of Ohta [Oht09b] it has Markov type 2 constant at most $1+\sqrt{2}$. Being an ℓ_2 product of spaces with uniformly bounded Markov type 2 constant, X also has Markov type 2.

Fix $\varepsilon \in (0,1)$. By covering the fundamental parallelepiped of Λ_n^* by homothetic copies of itself, we see that there exists a finite measurable partition $\{U_1, \ldots, U_k\}$ of the torus $\mathbb{R}^n/\Lambda_n^*$ into sets of diameter at most ε and $\mu_n(U_i) = 1/k$ for every $i \in \{1, \ldots, k\}$.

Fix an arbitrary point $x_i \in U_i$ and suppose that $f : \{x_1, \ldots, x_k\} \to \ell_2$ is 1-Lipschitz. Since $M_2(\mathbb{R}^n/\Lambda_n^*) \leqslant 1 + \sqrt{2} \leqslant 3$, by Ball's extension theorem [Bal92] there exists $F : \mathbb{R}^n/\Lambda_n^* \to \ell_2$ that extends f and satisfies $\|F\|_{\mathrm{Lip}} \leqslant 3$. By arguing as in the proof of Lemma 7.8, we have

$$\frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} \|f(x_i) - f(x_j)\|_{\ell_2}^2 \leqslant 3 \int_{\mathbb{R}^n/\Lambda_n^*}\int_{\mathbb{R}^n/\Lambda_n^*} \|F(x) - F(y)\|_{\ell_2}^2 d\mu_n(x) d\mu_n(y) + 54\varepsilon^2, \quad (229)$$

and

$$\frac{1}{k^2}\sum_{i=1}^{k} d_{\mathbb{R}^n/\Lambda_n^*}(x_i, x_j) \geqslant \frac{1}{3}\int_{\mathbb{R}^n/\Lambda_n^*}\int_{\mathbb{R}^n/\Lambda_n^*} d_{\mathbb{R}^n/\Lambda_n^*}(x,y)^2 d\mu_n(x) d\mu_n(y) - 2\varepsilon^2, \quad (230)$$

By [KN06, Lem. 11], for every Lipschitz mapping $g : \mathbb{R}^n/\Lambda_n^* \to \ell_2$,

$$\int_{\mathbb{R}^n/\Lambda_n^*}\int_{\mathbb{R}^n/\Lambda_n^*} \|g(x) - g(y)\|_{\ell_2}^2 d\mu_n(x) d\mu_n(y) \lesssim \frac{n\|g\|_{\mathrm{Lip}}^2}{N(\Lambda_n^*)^2}. \quad (231)$$

Also, by [KN06, Lem. 10] we have

$$\int_{\mathbb{R}^n/\Lambda_n^*}\int_{\mathbb{R}^n/\Lambda_n^*} d_{\mathbb{R}^n/\Lambda_n^*}(x,y)^2 d\mu_n(x)d\mu_n(y) \gtrsim \frac{n^2}{r(\Lambda_n^*)^2}. \tag{232}$$

Hence, by letting $\varepsilon \to 0$ in (229) and (230), it follows from (231) and (232) that

$$Av_{\ell_2}^{(2)}\left(\mathbb{R}^n/\Lambda_n^*\right) \gtrsim \frac{N(\Lambda_n^*)}{r(\Lambda_n^*)}\sqrt{n} \gtrsim \sqrt{n}, \tag{233}$$

where we used the assumption $r(\Lambda_n) \lesssim N(\Lambda_n)$. Now $Av_{\ell_2}^{(2)}(X) = \infty$ follows from (233) and the definition of X. $\square$

Remark 7.17. The use of Ball's extension theorem in the proof of Theorem 7.16 can be replaced by the use of Kirszbraun's extension theorem [Kir34] combined with the fact [KN06, Thm. 6] (see also [HR13]) that $c_2(\mathbb{R}^n/\Lambda_n^*) \lesssim_n 1$; in this case the factor 54 in (229) would be replaced by a factor that depends on n, but since we let $\varepsilon \to 0$ this does not affect the rest of the proof (however, if one desires to bound k as a function of n, this approach would yield an inferior estimate).

References

[ABS98] N. Alon, R. Boppana, and J. Spencer. An asymptotic isoperimetric inequality. *Geom. Funct. Anal.*, 8(3):411–436, 1998.

[ALN08] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21(1):1–21 (electronic), 2008.

[AR94] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures Algorithms*, 5(2):271–284, 1994.

[ARV09] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2):Art. 5, 37, 2009.

[Ass83] P. Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[Bal92] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[BCL94] 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.

[BFdlV82] B. Bollobás and W. Fernandez de la Vega. The diameter of random regular graphs. *Combinatorica*, 2(2):125–134, 1982.

[BH99] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[BL00] Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

[BLMN05] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005.

[BLR10] P. Biswal, J. R. Lee, and S. Rao. Eigenvalue bounds, spectral partitioning, and metrical deformations via flows. *J. ACM*, 57(3):Art. 13, 23, 2010.

[BMW86] J. Bourgain, V. Milman, and H. Wolfson. On type of metric spaces. *Trans. Amer. Math. Soc.*, 294(1):295–317, 1986.

[Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[BS87] A. Z. Broder and E. Shamir. On the second eigenvalue of random regular graphs. In *28th Annual Symposium on Foundations of Computer Science*, pages 286–294. 1987.

[Cal64] A.-P. Calderón. Intermediate spaces and interpolation, the complex method. *Studia Math.*, 24:113–190, 1964.

[CGR08] S. Chawla, A. Gupta, and H. Räcke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. *ACM Trans. Algorithms*, 4(2):Art. 22, 18, 2008.

[Cha84] I. Chavel. *Eigenvalues in Riemannian geometry*, volume 115 of *Pure and Applied Mathematics*. Academic Press Inc., Orlando, FL, 1984. Including a chapter by Burton Randol, With an appendix by Jozef Dodziuk.

[Cha95] F. Chaatit. On uniform homeomorphisms of the unit spheres of certain Banach lattices. *Pacific J. Math.*, 168(1):11–31, 1995.

[Che70] 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.

[Chu89] F. Chung. Diameters and eigenvalues. *J. Amer. Math. Soc.*, 2(2):187–195, 1989.

[CM08] D. Christofides and K. Markström. Expansion properties of random Cayley graphs and vertex transitive graphs via matrix martingales. *Random Structures Algorithms*, 32(1):88–100, 2008.

[CR82] M. Cwikel and S. Reisner. Interpolation of uniformly convex Banach spaces. *Proc. Amer. Math. Soc.*, 84(4):555–559, 1982.

[DKSV06] N. R. Devanur, S. A. Khot, R. Saket, and N. K. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 537–546. ACM, New York, 2006.

[DL97] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997.

[DLP13] J. Ding, J. R. Lee, and Y. Peres. Markov type and threshold embeddings. *Geom. Funct. Anal.*, 23(4):1207–1229, 2013.

[dlS13] M. de la Salle. Towards Banach space strong property (T) for $\mathrm{SL}(3,\mathbb{R})$, 2013. Preprint available at `http://arxiv.org/abs/1307.2475`.

[Elt83] J. Elton. Sign-embeddings of l_1^n. *Trans. Amer. Math. Soc.*, 279(1):113–124, 1983.

[Enf76] 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.

[Fie89] M. Fiedler. Laplacian of graphs and algebraic connectivity. In *Combinatorics and graph theory (Warsaw, 1987)*, volume 25 of *Banach Center Publ.*, pages 57–70. PWN, Warsaw, 1989.

[Fig76] T. Figiel. On the moduli of convexity and smoothness. *Studia Math.*, 56:121–155, 1976.

[FJS88] T. Figiel, W. B. Johnson, and G. Schechtman. Random sign embeddings from l_r^n, $2 < r < \infty$. *Proc. Amer. Math. Soc.*, 102(1):102–106, 1988.

[Fri08] J. Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.

[FS02] U. Feige and G. Schechtman. On the optimality of the random hyperplane rounding technique for MAX CUT. *Random Structures Algorithms*, 20(3):403–440, 2002. Probabilistic methods in combinatorial optimization.

[FT03] J. Fakcharoenphol and K. Talwar. An improved decomposition theorem for graphs excluding a fixed minor. In *Approximation, randomization, and combinatorial optimization*, volume 2764 of *Lecture Notes in Comput. Sci.*, pages 36–46. Springer, Berlin, 2003.

[GHR12] F. Göring, C. Helmberg, and S. Reiss. Graph realizations associated with minimizing the maximum eigenvalue of the Laplacian. *Math. Program.*, 131(1-2, Ser. A):95–111, 2012.

[GHW08] F. Göring, C. Helmberg, and M. Wappler. Embedded in the shadow of the separator. *SIAM J. Optim.*, 19(1):472–501, 2008.

[GKL03] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science*, pages 534–543. 2003.

[GLS93] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 1993.

[GPP80] E. D. Gluskin, A. Pietsch, and J. Puhl. A generalization of Khintchine's inequality and its application in the theory of operator ideals. *Studia Math.*, 67(2):149–155, 1980.

[Gro03] M. Gromov. Random walk in random groups. *Geom. Funct. Anal.*, 13(1):73–146, 2003.

[Har66] L. H. Harper. Optimal numberings and isoperimetric problems on graphs. *J. Combinatorial Theory*, 1:385–393, 1966.

[Hei01] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001.

[HN13] T. Hytönen and A. Naor. Pisier's inequality revisited. *Studia Math.*, 215(3):221–235, 2013.

[HR13] I. Haviv and O. Regev. The Euclidean distortion of flat tori. *J. Topol. Anal.*, 5(2):205–223, 2013.

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[JS88] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for markov chains: the approximation of the permanent resolved (preliminary version). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, pages 235–244. 1988.

[Kir34] M. D. Kirszbraun. Über die zusammenziehenden und Lipschitzchen Transformationen. *Fundam. Math.*, 22:77–108, 1934.

[KM13] D. M. Kane and R. Meka. A PRG for Lipschitz functions of polynomials with applications to sparsest cut. In *STOC'13: Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 1–10. 2013.

[KN06] S. Khot and A. Naor. Nonembeddability theorems via Fourier analysis. *Math. Ann.*, 334(4):821–852, 2006.

[Kon11] T. Kondo. CAT(0) spaces and expanders. *Mathematische Zeitschrift*, pages 1–13, 2011.

[KOR00] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Comput.*, 30(2):457–474 (electronic), 2000.

[KPR93] P. N. Klein, S. A. Plotkin, and S. Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, pages 682–690. 1993.

[KR09] R. Krauthgamer and Y. Rabani. Improved lower bounds for embeddings into L_1. *SIAM J. Comput.*, 38(6):2487–2498, 2009.

[KY06] G. Kasparov and G. Yu. The coarse geometric Novikov conjecture and uniform convexity. *Adv. Math.*, 206(1):1–56, 2006.

[Laf08] V. Lafforgue. Un renforcement de la propriété (T). *Duke Math. J.*, 143(3):559–602, 2008.

[Laf09] V. Lafforgue. Propriété (T) renforcée Banachique et transformation de Fourier rapide. *J. Topol. Anal.*, 1(3):191–206, 2009.

[Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[Lee05] J. R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 92–101 (electronic). ACM, New York, 2005.

[Lia13] B. Liao. Strong Banach property (T) for simple algebraic groups of higher rank, 2013. Preprint available at `http://arxiv.org/abs/1301.1861`.

[Lin63] J. Lindenstrauss. On the modulus of smoothness and divergent series in Banach spaces. *Michigan Math. J.*, 10:241–252, 1963.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[LN05] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[LP13] R. Lyons and Y. Peres. *Probability on Trees and Networks*, 2013. Forthcoming book, available at `http://mypage.iu.edu/~rdlyons/prbtree/book.pdf`.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[LS88] G. F. Lawler and A. D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309(2):557–580, 1988.

[LS05] U. Lang and T. Schlichenmaier. Nagata dimension, quasisymmetric embeddings, and Lipschitz extensions. *Int. Math. Res. Not.*, (58):3625–3655, 2005.

[LS10] J. R. Lee and A. Sidiropoulos. Genus and the geometry of the cut graph. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 193–201. 2010.

[Mar88] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988.

[Mat97] J. Matoušek. On embedding expanders into ℓ_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[Mau03] B. Maurey. Type, cotype and K-convexity. In *Handbook of the geometry of Banach spaces, Vol. 2*, pages 1299–1332. North-Holland, Amsterdam, 2003.

[Maz29] S. Mazur. Une remarque sur l'homéomorphie des champs fonctionels. *Studia Math.*, 1:83–85, 1929.

[MN08] M. Mendel and A. Naor. Metric cotype. *Ann. of Math. (2)*, 168(1):247–298, 2008. Preliminary version in SODA '06.

[MN12] M. Mendel and A. Naor. Nonlinear spectral calculus and super-expanders, 2012. To appear in Inst. Hautes Études Sci. Publ. Math., available at `http://arxiv.org/abs/1207.4705`.

[MN13a] M. Mendel and A. Naor. Expanders with respect to Hadamard spaces and random graphs, 2013. Preprint available at `http://arxiv.org/abs/1306.5434`.

[MN13b] M. Mendel and A. Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. *Anal. Geom. Metr. Spaces*, 1:163–199, 2013.

[MP76] B. Maurey and G. Pisier. Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. *Studia Math.*, 58(1):45–90, 1976.

[Nao12a] A. Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.

[Nao12b] A. Naor. On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon-Roichman graphs. *Comb. Probab. Comput.*, 21(4):623–634, 2012.

[NPSS06] 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.

[NR09] I. Newman and Y. Rabinovich. Hard metrics from Cayley graphs of abelian groups. *Theory Comput.*, 5:125–134, 2009.

[NRS05] 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.

[NS02] A. Naor and G. Schechtman. Remarks on non linear type and Pisier's inequality. *J. Reine Angew. Math.*, 552:213–236, 2002.

[NS11] A. Naor and L. Silberman. Poincaré inequalities, embeddings, and wild groups. *Compos. Math.*, 147(5):1546–1572, 2011.

[Oht09a] S.-i. Ohta. Extending Lipschitz and Hölder maps between metric spaces. *Positivity*, 13(2):407–425, 2009.

[Oht09b] S.-i. Ohta. Markov type of Alexandrov spaces of non-negative curvature. *Mathematika*, 55(1-2):177–189, 2009.

[OS94] E. Odell and T. Schlumprecht. The distortion problem. *Acta Math.*, 173(2):259–281, 1994.

[Oza04] N. Ozawa. A note on non-amenability of $\mathscr{B}(l_p)$ for $p = 1, 2$. *Internat. J. Math.*, 15(6):557–565, 2004.

[Pis73] G. Pisier. Sur les espaces de Banach qui ne contiennent pas uniformément de l_n^1. *C. R. Acad. Sci. Paris Sér. A-B*, 277:A991–A994, 1973.

[Pis75] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.

[Pis79] G. Pisier. La méthode d'interpolation complexe: applications aux treillis de Banach. In *Séminaire d'Analyse Fonctionnelle (1978–1979)*, pages Exp. No. 17, 18. École Polytech., Palaiseau, 1979.

[Pis86] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.

[Pis10] G. Pisier. Complex interpolation between Hilbert, Banach and operator spaces. *Mem. Amer. Math. Soc.*, 208(978):vi+78, 2010.

[Rab08] Y. Rabinovich. On average distortion of embedding metrics into the line. *Discrete Comput. Geom.*, 39(4):720–733, 2008.

[Rao99] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry (Miami Beach, FL, 1999)*, pages 300–306 (electronic). ACM, New York, 1999.

[Ray02] Y. Raynaud. On ultrapowers of non commutative L_p spaces. *J. Operator Theory*, 48(1):41–68, 2002.

[Rog50] C. A. Rogers. A note on coverings and packings. *J. London Math. Soc.*, 25:327–331, 1950.

[SBXD06] J. Sun, S. Boyd, L. Xiao, and P. Diaconis. The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM Rev.*, 48(4):681–699, 2006.

[Tal88] M. Talagrand. An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities. *Proc. Amer. Math. Soc.*, 104(3):905–909, 1988.

[Woj91] P. Wojtaszczyk. *Banach spaces for analysts*, volume 25 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1991.

COURANT INSTITUTE, NEW YORK UNIVERSITY
E-mail address: naor@cims.nyu.edu