HAL open science

Average distortion embeddings, nonlinear spectral gaps, and a metric John theorem

Alexandros Eskenazis

▶ To cite this version:

Alexandros Eskenazis. Average distortion embeddings, nonlinear spectral gaps, and a metric John theorem. Asterisque, 2022, 10.24033/ast.1189 . hal-03912757

HAL Id: hal-03912757
https://hal.science/hal-03912757

Submitted on 25 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Séminaire BOURBAKI Janvier 2022
74e année, 2021–2022, nº 1188

AVERAGE DISTORTION EMBEDDINGS, NONLINEAR SPECTRAL GAPS, AND A METRIC JOHN THEOREM [after Assaf Naor]

by **Alexandros Eskenazis**

1. INTRODUCTION

Preamble. The main purpose of this survey is to present a concise exposition of some applications of the theory of nonlinear spectral gaps which can serve as a roadmap for newcomers in the field and experts alike. Having as our main focus a result (Theorem 1.1) of Naor (2021), we shall highlight some ideas which have played a pivotal role in recent developments and mention connections with classical geometric and algorithmic questions. The material of this paper is a mere expository repackaging of a selection of such developments and any difference in presentation is solely cosmetic.

Let $(\mathcal{M}, d_{\mathcal{M}})$, $(\mathcal{N}, d_{\mathcal{N}})$ be two metric spaces and $D \in [1, \infty)$. We say that $(\mathcal{M}, d_{\mathcal{M}})$ embeds into $(\mathcal{N}, d_{\mathcal{N}})$ with bi-Lipschitz distortion at most D if there exists a scaling factor $\sigma \in (0, \infty)$ and a map $f : \mathcal{M} \to \mathcal{N}$ such that

$$\forall\, x, y \in \mathcal{M}, \qquad \sigma d_{\mathcal{M}}(x, y) \le d_{\mathcal{N}}\Big(f(x), f(y)\Big) \le \sigma D d_{\mathcal{M}}(x, y). \tag{1}$$

Following Naor (2021), we say that an *infinite*[(1)] metric space $(\mathcal{M}, d_{\mathcal{M}})$ embeds into $(\mathcal{N}, d_{\mathcal{N}})$ with q-average distortion D, where $q > 0$, if for every Borel probability measure μ on $\mathcal{M}$, there exists $\sigma = \sigma_\mu \in (0, \infty)$ and a σD-Lipschitz map $f = f_\mu : \mathcal{M} \to \mathcal{N}$ with

$$\iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{N}}\Big(f(x), f(y)\Big)^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \ge \sigma^q \iint_{\mathcal{M}\times\mathcal{M}} d_{\mathcal{M}}(x, y)^q \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y). \tag{2}$$

If the target space $\mathcal{N}$ is a normed space, the parameter σ_μ can be omitted by rescaling.

The θ-snowflake of a metric space $(\mathcal{M}, d_{\mathcal{M}})$ is the metric space $(\mathcal{M}, d_{\mathcal{M}}^\theta)$, $\theta \in (0, 1]$. The primary goal of this survey is to present a self-contained proof of the following deep embedding theorem of Naor (2021) in which asymptotically optimal bounds for the quadratic average distortion (i.e. corresponding to exponent $q = 2$ in equation (2) above) of $\frac{1}{2}$-snowflakes of finite-dimensional normed spaces into the separable Hilbert space ℓ_2 are established. The, so called, *average John theorem* reads as follows.

The author was supported by a Junior Research Fellowship from Trinity College, Cambridge.

(1) The study of average distortion embeddings for *finite* metric spaces goes back at least to the work of Rabinovich (2003) (see also Abraham, Bartal, and Neiman, 2011 for various related notions).

Theorem 1.1 (Average John). — *There exists a universal constant $C \in (0,\infty)$ such that the $\frac{1}{2}$-snowflake of any finite-dimensional normed space $(X, \|\cdot\|_X)$ admits an embedding into ℓ_2 with quadratic average distortion at most $C\sqrt{\log(\dim(X)+1)}$.*

Theorem 1.1 is a metric counterpart of a classical theorem of John (1948), asserting that any finite-dimensional normed space embeds into ℓ_2 with bi-Lipschitz distortion at most $\sqrt{\dim(X)}$. This statement is famously optimal, e.g. for $X = \ell_1^d$ or $X = \ell_\infty^d$, yet Naor's theorem shows that an exponential improvement of the relevant distortion is possible if one relaxes the pointwise lower bound of the bi-Lipschitz condition (1) to the averaged requirement (2) and replaces the normed space $(X, \|\cdot\|_X)$ by its $\frac{1}{2}$-snowflake. Before explaining the ideas that come into the proof of Theorem 1.1, it is worth pointing out that both of these modifications of John's theorem are necessary in order to deduce bounds for the distortion which are subpolynomial on $\dim(X)$. In fact, the average John theorem is optimal in three distinct ways.

• If one is interested in bi-Lipschitz embeddings of snowflakes of normed spaces X into ℓ_2 in lieu of average distortion embeddings, then the relevant distortion has to depend polynomially on $\dim(X)$. Indeed, in Naor (2021, Lemma 2), it is shown that the bi-Lipschitz distortion required to embed the θ-snowflake of ℓ_∞^d into ℓ_2 is at least a constant multiple of $d^{\theta/2}$. The proof relies on metric cotype.

• The exponent $\frac{1}{2}$ is the least amount of snowflaking that one needs to perform in order to obtain embeddings whose quadratic average distortion depends subpolynomially on $\dim(X)$. More specifically, in Naor (2021, Lemma 13) it is shown that for any $\varepsilon \in (0, \frac{1}{2}]$, the quadratic average distortion required to embed the $(\frac{1}{2}+\varepsilon)$-snowflake of ℓ_1^d into ℓ_2 is at least a constant multiple of d^ε. The proof relies on Enflo type.

• Finally, $\sqrt{\log\dim(X)}$ is the asymptotically optimal bound for the quadratic average distortion required to embed the $\frac{1}{2}$-snowflake of an arbitrary finite-dimensional space X into ℓ_2. This will be further explained (for $X = \ell_\infty^d$) in Remark 6.2 below.

In the rest of the introduction, we shall describe the strategy of the proof of the average John theorem and introduce the necessary background.

1.1. Nonlinear spectral gaps

Let $\triangle^{n-1} = \{(\pi_1,\ldots,\pi_n) \in [0,1]^n : \sum_{i=1}^n \pi_i = 1\}$ be the n-dimensional standard simplex. Consider a (row)-stochastic matrix $A = (a_{ij})_{i,j=1}^n \in M_n(\mathbb{R})$, that is, a matrix for which $(a_{i1},\ldots,a_{in}) \in \triangle^{n-1}$ for every $i \in \{1,\ldots,n\}$. Given a vector $\pi = (\pi_1,\ldots,\pi_n) \in \triangle^{n-1}$, we say that the matrix A is π-reversible if $\pi_i a_{ij} = \pi_j a_{ji}$ for every $i,j \in \{1,\ldots,n\}$. These objects admit a classical probabilistic interpretation. Consider the discrete-time homogeneous Markov chain $(X_t)_{t\geq 0}$ on the state space $\{1,\ldots,n\}$ with transition probabilities given by

$$\forall\, i,j \in \{1,\ldots,n\}, \qquad \mathbb{P}\{X_{t+1} = j \mid X_t = i\} = a_{ij}, \tag{3}$$

where $t \geq 0$. If the transition matrix A is π-reversible, then π is also a stationary distribution for the process $(X_t)_{t\geq 0}$, that is, if X_0 is distributed according to π then so is X_t for any $t \geq 1$. This is expressed algebraically by the matrix identity $\pi A = \pi$, where π is thought of as a row-vector. In the probabilistic framework above, reversibility simply means that the Markov process is invariant under time reversal in the sense that $(X_0, X_1, \ldots, X_T)$ has the same joint distribution as $(X_T, X_{T-1}, \ldots, X_0)$ for any $T \in \mathbb{N}$.

Consider the Hilbert space $L_2(\pi) = (\mathbb{R}^n, \|\cdot\|_{L_2(\pi)})$ whose (semi-)norm is given by

$$\forall\, x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \qquad \|x\|_{L_2(\pi)} = \Big(\sum_{i=1}^n \pi_i x_i^2\Big)^{\frac{1}{2}}. \tag{4}$$

Analytically, the stochastic matrix A is π-reversible if and only if it defines a self-adjoint contraction on $L_2(\pi)$ with real eigenvalues which we shall denote by $1 = \lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A) \geq -1$. The spectral gap of A is the algebraic quantity $1 - \lambda_2(A)$ which is known to encode important combinatorial properties of the matrix. It is a simple linear algebra exercise to show that the reciprocal $\gamma(A) \stackrel{\text{def}}{=} (1 - \lambda_2(A))^{-1}$ of the spectral gap is the least constant $\gamma \in (0, \infty]$ for which the inequality

$$\forall\, x_1, \ldots, x_n \in \ell_2, \qquad \sum_{i,j=1}^n \pi_i \pi_j \|x_i - x_j\|_{\ell_2}^2 \leq \gamma \sum_{i,j=1}^n \pi_i a_{ij} \|x_i - x_j\|_{\ell_2}^2 \tag{5}$$

holds true. It is a well-known consequence of Cheeger's inequality (see, e.g., Davidoff, Sarnak, and Valette (2003)) that upper bounds on $\gamma(A)$ are equivalent to good expansion properties of the underlying weighted graph defined by A.

The above analytic characterization of a spectral gap as an optimal constant in a functional inequality was the starting point for the theory of *nonlinear spectral gaps*, of which Theorem 1.1 is the latest application. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space and $p \in (0, \infty)$. If $\pi \in \triangle^{n-1}$ and A is a π-reversible stochastic matrix, the spectral gap of A with respect to $d_{\mathcal{M}}^p$, denoted by $\gamma(A, d_{\mathcal{M}}^p)$, is the least $\gamma \in (0, \infty]$ such that

$$\forall\, x_1, \ldots, x_n \in \mathcal{M}, \qquad \sum_{i,j=1}^n \pi_i \pi_j d_{\mathcal{M}}(x_i, x_j)^p \leq \gamma \sum_{i,j=1}^n \pi_i a_{ij} d_{\mathcal{M}}(x_i, x_j)^p. \tag{6}$$

If the metric $d_{\mathcal{M}}$ is inherited by a norm $\|\cdot\|$, we will denote $\gamma(A, d_{\mathcal{M}}^p)$ by $\gamma(A, \|\cdot\|^p)$. As explained in Mendel and Naor (2014), unless $\mathcal{M}$ is a singleton, if $\gamma(A, d_{\mathcal{M}}^p)$ is finite then $\lambda_2(A)$ is bounded away from 1 by a positive quantity depending only on $\gamma(A, d_{\mathcal{M}}^p)$. On the other hand, obtaining sensible upper bounds for $\gamma(A, d_{\mathcal{M}}^p)$ in terms of the usual spectral gap $1 - \lambda_2(A)$ is a notoriously hard task even for very structured metric spaces $(\mathcal{M}, d_{\mathcal{M}})$. This difficulty reflects the fact that nonlinear spectral gap inequalities (6) capture delicate interactions of spectral properties of the matrix A and geometric characteristics of the underlying metric space $(\mathcal{M}, d_{\mathcal{M}})$.

The study of nonlinear spectral gap inequalities (6) has led to very fruitful investigations which have been impactful in various areas of mathematics and theoretical computer science such as metric geometry, geometric group theory,

operator algebras, Alexandrov geometry and approximation algorithms. We refer, for instance, to the works of Matoušek (1997), Gromov (2003), Lafforgue (2008, 2009), Pisier (2010), Naor and Silberman (2011), Kondo (2012), Mendel and Naor (2013, 2014, 2015), Mimura (2015), Andoni, Naor, Nikolov, Razenshteyn, and Waingarten (2018a,b) and Naor (2014, 2017, 2021) (see also Section 6 below for a high-level exposition of some of those). The pertinence of nonlinear spectral gaps to the study of average distortion embeddings into normed spaces and Theorem 1.1 stems from an important duality principle which was discovered by Naor (2014) and which we shall now describe.

1.2. Duality

Fix $\pi \in \triangle^{n-1}$ and a π-reversible stochastic matrix $A \in M_n(\mathbb{R})$. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space, $(Y, \|\cdot\|_Y)$ be a normed space and assume that the θ-snowflake of $\mathcal{M}$ embeds into Y with q-average distortion $D \in [1,\infty)$. Then, for $x_1,\ldots,x_n \in \mathcal{M}$, there exist $y_1,\ldots,y_n \in Y$ such that $\|y_i - y_j\|_Y \le D d_{\mathcal{M}}(x_i,x_j)^\theta$ for every $i,j \in \{1,\ldots,n\}$ and

$$\sum_{i,j=1}^n \pi_i\pi_j \|y_i - y_j\|_Y^q \ge \sum_{i,j=1}^n \pi_i\pi_j d_{\mathcal{M}}(x_i,x_j)^{\theta q}. \tag{7}$$

Therefore, we have

$$\sum_{i,j=1}^n \pi_i\pi_j d_{\mathcal{M}}(x_i,x_j)^{\theta q} \overset{(7)}{\le} \gamma(A,\|\cdot\|_Y^q) \sum_{i,j=1}^n \pi_i a_{ij} \|y_i - y_j\|_Y^q \le D^q \gamma(A,\|\cdot\|_Y^q) \sum_{i,j=1}^n \pi_i a_{ij} d_{\mathcal{M}}(x_i,x_j)^{\theta q}$$

which implies that $\gamma(A, d_{\mathcal{M}}^{\theta q}) \le D^q \gamma(A, \|\cdot\|_Y^q)$. Moreover[(2)], as tensorization gives the identity $\gamma(A,\|\cdot\|_Y^q) = \gamma(A,\|\cdot\|_{\ell_q(Y)}^q)$ and $\gamma(A,\|\cdot\|_W^q)$ is only determined by the finite-dimensional structure of W, the above simple argument shows that if the θ-snowflake of $\mathcal{M}$ embeds with q-average distortion $D \in [1,\infty)$ into any Banach space Z which is finitely representable in $\ell_q(Y)$, then $\gamma(A, d_{\mathcal{M}}^{\theta q}) \le D^q \gamma(A,\|\cdot\|_Y^q)$ for any π-reversible stochastic matrix $A \in M_n(\mathbb{R})$. The first important step towards Theorem 1.1 is the following striking converse to this implication, proven by Naor (2014, Theorem 1.3).

Theorem 1.2 (Naor's duality principle). — *Suppose that $q, D \in [1,\infty)$ and $\theta \in (0,1]$. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space and $(Y, \|\cdot\|_Y)$ be a Banach space such that for every $n \in \mathbb{N}$ and $\pi \in \triangle^{n-1}$, every π-reversible stochastic matrix $A \in M_n(\mathbb{R})$ satisfies*

$$\gamma(A, d_{\mathcal{M}}^{\theta q}) \le D^q \gamma(A, \|\cdot\|_Y^q). \tag{8}$$

Then, for any $\varepsilon > 0$ the θ-snowflake of $\mathcal{M}$ embeds into some ultrapower[(3)] of $\ell_q(Y)$ with q-average distortion at most $D + \varepsilon$.

---

(2) As usual, we denote by $\ell_q(Y) = \big\{ y = (y_n)_{n\ge1} \in Y^{\mathbb{N}} : \|y\|_{\ell_q(Y)} \overset{\text{def}}{=} \big(\sum_{n\ge1} \|y_n\|_Y^q\big)^{1/q} < \infty \big\}$.

(3) We refer to Heinrich (1980) for background on ultraproducts of Banach spaces. For the purposes of this discussion it suffices to say that an ultrapower $Z^{\mathcal{U}}$ of a Banach space Z is a Banach space containing Z with various compactness properties such that any finite-dimensional subspace of $Z^{\mathcal{U}}$ embeds into Z with distortion $1+\varepsilon$ for any $\varepsilon > 0$.

We emphasize that Theorem 1.2 is an *existential* result whose proof does not shed any light on any additional properties of the average distortion embeddings at hand. Its proof consists of an elegant Hahn–Banach separation argument which we shall present in Section 2. In the setting of the average John theorem, the metric space $\mathcal{M}$ is a finite-dimensional normed space $(X, \|\cdot\|_X)$, Y is the Hilbert space ℓ_2, $q = 2$ and $\theta = \frac{1}{2}$. As any ultrapower of ℓ_2 is itself a Hilbert space (see Heinrich, 1980), Naor's duality theorem shows that the embedding statement of Theorem 1.1 is equivalent to the following comparison estimate for nonlinear spectral gaps.

Theorem 1.3. — *Let $(X, \|\cdot\|_X)$ be a finite-dimensional normed space. Then, for every $n \in \mathbb{N}$ and $\pi \in \triangle^{n-1}$, every π-reversible stochastic matrix $A \in M_n(\mathbb{R})$ satisfies*

$$\gamma(A, \|\cdot\|_X) \leq \frac{C \log(\dim(X) + 1)}{1 - \lambda_2(A)}, \tag{9}$$

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

Theorem 1.3 has implicitly appeared as a special case of a much more general result concerning nonlinear spectral gaps of complex interpolation spaces (Naor, 2021, Theorem 25). This family of substantially stronger nonlinear spectral gap inequalities can be used to prove (via Theorem 1.2) the existence of refined average distortion embeddings of snowflakes of Banach spaces which are not captured by Theorem 1.1. This task is undertaken in great detail in Naor (2021), yet most of these results go beyond the scope of the present survey. In Section 4, we shall present a self-contained proof of Theorem 1.3 which completely avoids the complex interpolation machinery of Naor (2021) and is a modification of an argument which appeared in Naor (2018, Section 5). In Section 5, we shall present some extensions and refinements of Theorems 1.1 and 1.3 and highlight some key ideas from their proofs in Naor (2021).

1.3. Extrapolation

As explained above, the forthcoming proof of Theorem 1.3 does not rely on any sophisticated analytic machinery beyond elementary spectral properties of matrices. We will however use the following extrapolation principle for Poincaré inequalities.

Proposition 1.4. — *For every $1 \leq p \leq q < \infty$ there exist $c(p,q), C(p,q) \in (0, \infty)$ such that the following conclusion holds. For every normed space $(X, \|\cdot\|_X)$, every $n \in \mathbb{N}$, $\pi \in \triangle^{n-1}$ and every π-reversible stochastic matrix $B \in M_n(\mathbb{R})$, we have*

$$c(p,q) \cdot \gamma(B, \|\cdot\|_X^q)^{\frac{p}{q}} \leq \gamma(B, \|\cdot\|_X^p) \leq C(p,q) \cdot \gamma(B, \|\cdot\|_X^q). \tag{10}$$

Proposition 1.4 is the vector-valued version (due to Cheng (2016) and de Laat and de la Salle (2021)) of the extrapolation principle for Poincaré inequalities (Matoušek, 1997). In Section 3, we shall also discuss a strengthening of Proposition 1.4 and its relation to a long-standing problem in the nonlinear theory of Banach spaces.

1.4. Historical discussion

Motivated by a classical theorem of Ribe (1976) and kickstarted by Bourgain (1986), the *Ribe program* is a vast research program in metric geometry which aims to uncover deep structural analogies between the local theory of normed spaces and (nonlinear) metric spaces. In the nearly four decades that lapsed since Bourgain's formalization of its objectives, the Ribe program has been an extraordinary source of surprising phenomena which arise when one studies metric spaces through the lens of Banach space theory and, vice versa, when one considers normed spaces as objects in the metric category. Numerous such key insights obtained in the last two decades originate in works of Naor and his collaborators. We refer to the surveys of Kalton (2008), Naor (2012, 2018), Ball (2013), Baudier and Johnson (2016) and Godefroy (2017) and to the monograph of Ostrovskii (2013) for a snapshot of some of these advances and their applications to other areas of mathematics and theoretical computer science.

Theorem 1.1 is a prime example of a result conceptually belonging in the Ribe program for multiple reasons. Firstly, the statement of the theorem contains a highly nonlinear operation (snowflaking) performed on a norm and the desired embedding itself is not realized by a linear operator despite the fact that both the source and the target space are linear. Moreover, as already mentioned, the proof of Theorem 1.1 relies on the theory of nonlinear spectral gaps, a large part of which has been developed in the context of the Ribe program (see the discussion on expanders with respect to Banach spaces in Section 6.3 below). Finally, as discussed in Naor (2021, Section 1.4), Naor's initial interest in this research direction stemmed from a question regarding the embeddability of expanders into low-dimensional normed spaces raised by Andoni, Nguyen, Nikolov, Razenshteyn, and Waingarten (2017) in the context of the approximate nearest neighbor search problem. A negative answer to this question (see Theorem 6.1 below) by Naor (2017, 2021) which follows easily from Theorem 1.1 shall be explained in detail in Section 6.1. Theorem 6.1 also provides a new negative answer to an old question of Johnson and Lindenstrauss (1984) who asked whether every n-point metric space admits a bi-Lipschitz embedding with constant distortion into a d-dimensional normed space, where $d = O(\log n)$. This question had previously been answered negatively by Arias-de-Reyna and Rodríguez-Piazza (1992) for small distortions and Matoušek (1996) in general. Naor's works provide a novel and more robust approach to this problem as they highlight a specific criterion (spectral gap) which implies the intrinsic high-dimensionality of the metric space at hand. Johnson and Lindenstrauss raised this question as a step towards finding a metric version of the aforementioned classical theorem of John (1948). A deep and impactful nonlinear John theorem was discovered via a completely different route in the influential work of Bourgain (1985). Quite surprisingly, Theorem 1.1, which answers negatively the question of Johnson and Lindenstrauss, is itself a metric version of John's theorem.

Structure of the paper. In Sections 2 and 3 we present the proofs of Theorem 1.2 and Proposition 1.4 respectively. In Section 4 we use Proposition 1.4 to prove Theorem 1.3

which, combined with Theorem 1.2, completes the proof of Theorem 1.1. In Section 5 we present some refinements of Theorems 1.1 and 1.3 from Naor (2021) and highlight key ideas used in their proofs. Finally, Section 6 contains a high-level account of further geometric and algorithmic applications of the theory of nonlinear spectral gaps.

Asymptotic notation. In what follows we use the convention that for $a, b \in [0, \infty]$ the notation $a \gtrsim b$ (respectively $a \lesssim b$) means that there exists a universal constant $c \in (0, \infty)$ such that $a \geq cb$ (respectively $a \leq cb$). The notations $\lesssim_\xi$ and $\gtrsim_\chi$ mean that the implicit constant c depends on ξ and χ respectively.

Acknowledgements. I am very grateful to Florent Baudier, Manor Mendel and Assaf Naor for helpful discussions and constructive feedback.

2. DUALITY AND AVERAGE DISTORTION

In this section we present the proof of Naor's duality Theorem 1.2. Despite the fact that the theorem is stated for an *arbitrary* metric space $(\mathcal{M}, d_{\mathcal{M}})$, the crux of the argument is the following special case in which $\mathcal{M}$ is assumed to be *finite*. The general case follows by a (standard yet lengthy) discretization and compactness argument which can be found in Naor (2021, Section 7). The finitary version stated below was proven in the case that π is the normalized counting measure in Naor (2014, Theorem 1.3), where it is said that the argument is inspired by the proof of Ball (1992, Lemma 1.1).

Theorem 2.1 (Naor's duality – finitary version). — *Suppose that $q, D \in [1, \infty)$, $\theta \in (0, 1]$, $n \in \mathbb{N}$ and fix $\pi \in \mathrm{int}(\triangle^{n-1})$. Let $\mathcal{M} = (\{x_1, \ldots, x_n\}, d_{\mathcal{M}})$ be a metric space and $(Y, \|\cdot\|_Y)$ be a Banach space such that every π-reversible stochastic matrix $A \in M_n(\mathbb{R})$*

$$\gamma(A, d_{\mathcal{M}}^{\theta q}) \leq D^q \gamma(A, \|\cdot\|_Y^q). \tag{11}$$

Then, for any $\varepsilon > 0$ there exists $m \in \mathbb{N}$ and a function $f = f_\varepsilon : (\mathcal{M}, d_{\mathcal{M}}^\theta) \to \ell_q^m(Y)$ which is $(D + \varepsilon)$-Lipschitz satisfying the condition

$$\sum_{i,j=1}^n \pi_i \pi_j \|f(x_i) - f(x_j)\|_{\ell_q^m(Y)}^q \geq \sum_{i,j=1}^n \pi_i \pi_j d_{\mathcal{M}}(x_i, x_j)^{\theta q}. \tag{12}$$

Proof. — It clearly suffices to assume that $\theta = 1$ as otherwise we can simply apply the same result to the θ-snowflake of $\mathcal{M}$. Let $\mathcal{C} \subseteq M_n(\mathbb{R})$ be the class of all symmetric $n \times n$ matrices (c_{ij}) for which there exist $y_1, \ldots, y_n \in Y$, not all of which are equal, with

$$\forall\, i, j \in \{1, \ldots, n\}, \qquad c_{ij} = \frac{\sum_{r,s=1}^n \pi_r \pi_s d_{\mathcal{M}}(x_r, x_s)^q}{\sum_{r,s=1}^n \pi_r \pi_s \|y_r - y_s\|_Y^q} \cdot \|y_i - y_j\|_Y^q. \tag{13}$$

Moreover, let $\mathcal{P} \subseteq M_n(\mathbb{R})$ be the class of all symmetric $n \times n$ matrices with nonnegative entries and vanishing diagonal and consider the convex hull $\mathcal{Q} \stackrel{\text{def}}{=} \mathrm{conv}(\mathcal{C} + \mathcal{P})$.

Let $T = (t_{ij})_{i,j=1}^n$ be the $n \times n$ matrix with entries given by $t_{ij} \stackrel{\text{def}}{=} (D + \varepsilon)^q d_{\mathcal{M}}(x_i, x_j)^q$ for $i, j \in \{1, \ldots, n\}$. We shall prove that $T \in \mathcal{Q}$. Suppose that this is not the case.

Then, by the Hahn–Banach separation theorem, there exists a nonzero symmetric $n \times n$ matrix $H = (h_{ij})_{i,j=1}^n$ with vanishing diagonal such that

$$\inf_{(b_{ij})_{i,j=1}^n \in \mathcal{Q}} \sum_{i,j=1}^n h_{ij} b_{ij} \geq (D+\varepsilon)^q \sum_{i,j=1}^n h_{ij} d_M(x_i, x_j)^q. \tag{14}$$

Since $\mathcal{Q}$ contains a translate of $\mathcal{P}$, choosing $(b_{ij})_{i,j=1}^n \in \mathcal{P}$ whose only nonzero entries are those indexed by (k,ℓ) and (ℓ,k), where $k \neq \ell$, we deduce that $h_{ij} \geq 0$ for every $i,j \in \{1,\ldots,n\}$. Moreover, as $\pi_i \neq 0$ for every $i \in \{1,\ldots,n\}$, we can define the parameter

$$\sigma \stackrel{\text{def}}{=} \max_{i \in \{1,\ldots,n\}} \frac{1}{\pi_i} \sum_{r \neq i} h_{ir} \in (0,\infty) \tag{15}$$

and consider the matrix $A = (a_{ij})_{i,j=1}^n$ whose entries are given by

$$\forall\, i,j \in \{1,\ldots,n\}, \qquad a_{ij} \stackrel{\text{def}}{=} \begin{cases} \frac{h_{ij}}{\sigma \pi_i}, & \text{if } i \neq j \\ 1 - \frac{1}{\sigma \pi_i} \sum_{r \neq i} h_{ir}, & \text{if } i = j \end{cases}. \tag{16}$$

By construction, A is a π-reversible stochastic matrix as the choice of σ guarantees that its entries are nonnegative. Moreover, inequality (14) can be equivalently rewritten as

$$\inf_{(b_{ij})_{i,j=1}^n \in \mathcal{Q}} \sum_{i,j=1}^n \pi_i a_{ij} b_{ij} \geq (D+\varepsilon)^q \sum_{i,j=1}^n \pi_i a_{ij} d_M(x_i, x_j)^q. \tag{17}$$

Combining (17) with the definition (6) of nonlinear spectral gaps, we deduce that

$$\inf_{(b_{ij})_{i,j=1}^n \in \mathcal{Q}} \sum_{i,j=1}^n \pi_i a_{ij} b_{ij} \geq \frac{(D+\varepsilon)^q}{\gamma(A, d_M^q)} \sum_{i,j=1}^n \pi_i \pi_j d_M(x_i, x_j)^q \tag{18}$$

On the other hand, since $\mathcal{C} \subseteq \mathcal{Q}$, we have

$$\begin{aligned} \inf_{(b_{ij})_{i,j=1}^n \in \mathcal{Q}} \sum_{i,j=1}^n \pi_i a_{ij} b_{ij} &\leq \inf_{(b_{ij})_{i,j=1}^n \in \mathcal{C}} \sum_{i,j=1}^n \pi_i a_{ij} b_{ij} \\ &= \inf_{y_1,\ldots,y_n \in Y} \frac{\sum_{r,s=1}^n \pi_r \pi_s d_M(x_r, x_s)^q}{\sum_{r,s=1}^n \pi_r \pi_s \|y_r - y_s\|_Y^q} \sum_{i,j=1}^n \pi_i a_{ij} \|y_i - y_j\|_Y^q \\ &= \frac{1}{\gamma(A, \|\cdot\|_Y^q)} \cdot \sum_{r,s=1}^n \pi_r \pi_s d_M(x_r, x_s)^q. \end{aligned} \tag{19}$$

Combining (18), (19) and rearranging, we deduce that $\gamma(A, d_M^q) \geq (D+\varepsilon)^q \gamma(A, \|\cdot\|_Y^q)$ which contradicts the assumption (11), thus proving that $T \in \mathcal{Q}$.

Since $T \in \mathcal{Q}$ and all matrices in $\mathcal{P}$ have nonnegative entries, we deduce that there exists $m \in \mathbb{N}$, $(\mu_1,\ldots,\mu_m) \in \triangle^{m-1}$ and n-tuples of points $\{y_1(k),\ldots,y_n(k)\} \subset Y$ not all of which are equal for each $k \in \{1,\ldots,m\}$ such that

$$(D+\varepsilon)^q d_M(x_i, x_j)^q \geq \sum_{k=1}^m \underbrace{\mu_k \frac{\sum_{r,s=1}^n \pi_r \pi_s d_M(x_r, x_s)^q}{\sum_{r,s=1}^n \pi_r \pi_s \|y_r(k) - y_s(k)\|_Y^q}}_{w_k} \cdot \|y_i(k) - y_j(k)\|_Y^q, \tag{20}$$

for every $i,j \in \{1,\ldots,n\}$. Consider the mapping $f : \mathcal{M} \to \ell_q^m(Y)$ given by

$$\forall\, i \in \{1,\ldots,n\}, \qquad f(x_i) = \Big(w_1^{1/q} y_i(1), \ldots, w_m^{1/q} y_i(m)\Big). \tag{21}$$

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

$$\|f(x_i) - f(x_j)\|_{\ell_q^m(Y)} = \bigg(\sum_{k=1}^m w_k \|y_i(k) - y_j(k)\|_Y^q\bigg)^{\frac{1}{q}} \overset{(20)}{\leq} (D+\varepsilon) d_{\mathcal{M}}(x_i,x_j), \tag{22}$$

which is equivalent to $\|f\|_{\mathrm{Lip}} \leq D + \varepsilon$. Finally,

$$\begin{aligned}\sum_{i,j=1}^n \pi_i \pi_j \|f(x_i) - f(x_j)\|_{\ell_q^m(Y)}^q &= \sum_{i,j=1}^n \pi_i \pi_j \sum_{k=1}^n w_k \|y_i(k) - y_j(k)\|_Y^q \\ &\overset{(20)}{=} \sum_{i,j=1}^n \pi_i \pi_j \sum_{k=1}^m \mu_k \frac{\sum_{r,s=1}^n \pi_r \pi_s d_{\mathcal{M}}(x_r,x_s)^q}{\sum_{r,s=1}^n \pi_r \pi_s \|y_r(k) - y_s(k)\|_Y^q} \cdot \|y_i(k) - y_j(k)\|_Y^q, \\ &= \sum_{r,s=1}^n \pi_r \pi_s d_{\mathcal{M}}(x_r,x_s)^q\end{aligned} \tag{23}$$

which proves the average lower bound and completes the proof. □

Remark 2.2. — It is worth emphasizing that Naor's Theorem 2.1 is an important addition to a long list of results in which the existence of a map with favorable metric properties is proven using duality or by exploiting the cone structure of ℓ_p-distance matrices. We refer, for instance, to the works of Schoenberg (1938), Bretagnolle, Dacunha-Castelle, and Krivine (1965/1966), Krivine (1965), Wells and Williams (1975) on isometric embeddings, Maurey (1974) on factorization theory, Matoušek (2002, Proposition 15.5.2) on bi-Lipschitz embeddings, Ball (1990), Eskenazis (2021) on metric dimension reduction and Ball (1992), Mendel and Naor (2013) on extensions of Lipschitz mappings.

3. EXTRAPOLATION AND SNOWFLAKE EMBEDDINGS

In this section we present the proof of Proposition 1.4. The argument relies on some elementary properties of the vector-valued Mazur map (Mazur, 1929). If (Ω,μ) is a measure space, $(X, \|\cdot\|_X)$ is a normed space and $p,q \in [1,\infty)$, consider the map $\mathsf{M}_{p,q} : L_p(\mu;X) \to L_q(\mu;X)$ whose action on $f \in L_p(\mu;X)$ is given by

$$\forall\, \omega \in \Omega, \qquad (\mathsf{M}_{p,q} f)(\omega) = \frac{f(\omega)}{\|f(\omega)\|_X^{1-\frac{p}{q}}} \tag{24}$$

when $f(\omega) \neq 0$ and $(\mathsf{M}_{p,q} f)(\omega) = 0$ when $f(\omega) = 0$. We will use the following lemma.

Lemma 3.1. — *Let $p,q \in [1,\infty)$. For any normed space $(X, \|\cdot\|_X)$ and any functions $f,g : \Omega \to X$ with $\max\{\|f\|_{L_p(\mu;X)}, \|g\|_{L_p(\mu;X)}\} \leq 1$, we have*

$$\Big\|\mathsf{M}_{p,q} f - \mathsf{M}_{p,q} g\Big\|_{L_q(\mu;X)} \lesssim_{p,q} \Big\|f - g\Big\|_{L_p(\mu;X)}^{\min\{\frac{p}{q},1\}}. \tag{25}$$

Proof. — The scalar-valued case $X = \mathbb{R}$ of the proposition is classical[(4)] and can be found in Benyamini and Lindenstrauss (2000, Section 9.1). Consider two functions $\theta, \phi : \Omega \to \{x \in X : \ \|x\|_X = 1\}$ such that $f = \|f\|_X \theta$ and $g = \|g\|_X \phi$. Then, we have $\mathsf{M}_{p,q} f = \|f\|_X^{p/q}\theta$ and $\mathsf{M}_{p,q} g = \|g\|_X^{p/q}\phi$ which imply the inequality

$$\begin{aligned} \|\mathsf{M}_{p,q} f - \mathsf{M}_{p,q} g\|_{L_q(\mu;X)} &= \Big\| \|f\|_X^{\frac{p}{q}}\theta - \|g\|_X^{\frac{p}{q}}\phi \Big\|_{L_q(\mu;X)} \\ &\le \underbrace{\Big\| \|f\|_X^{\frac{p}{q}} - \|g\|_X^{\frac{p}{q}} \Big\|_{L_q(\mu;\mathbb{R})}}_{①} + \underbrace{\Big\| \|g\|_X^{\frac{p}{q}}(\theta - \phi) \Big\|_{L_q(\mu;X)}}_{②}. \end{aligned} \tag{26}$$

By the scalar-valued version of (25) applied to $\|f\|_X$ and $\|g\|_X$, we have

$$① \lesssim_{p,q} \Big\| \|f\|_X - \|g\|_X \Big\|_{L_p(\mu;\mathbb{R})}^{\min\{\frac{p}{q},1\}} \le \Big\| f - g \Big\|_{L_p(\mu;X)}^{\min\{\frac{p}{q},1\}}. \tag{27}$$

Moreover, if $\mathrm{d}\nu = \|g\|_X^p \, \mathrm{d}\mu$, then $\nu(\Omega) \le 1$ and thus for $q \le p$ we have

$$\begin{aligned} ② &= \|\theta - \phi\|_{L_q(\nu;X)} \le \|\theta - \phi\|_{L_p(\nu;X)} = \Big\| \|g\|_X \theta - \|g\|_X \phi \Big\|_{L_p(\mu;X)} \\ &\le \Big\| \|g\|_X - \|f\|_X \Big\|_{L_p(\mu;\mathbb{R})} + \Big\| \|f\|_X \theta - \|g\|_X \phi \Big\|_{L_p(\mu;X)} \le 2\|f - g\|_{L_p(\mu;X)}. \end{aligned} \tag{28}$$

On the other hand, if $p \le q$, Hölder's inequality gives

$$② = \|\theta - \phi\|_{L_q(\nu;X)} \le \|\theta - \phi\|_{L_p(\nu;X)}^{\frac{p}{q}} \cdot \|\theta - \phi\|_{L_\infty(\nu;X)}^{1-\frac{p}{q}} \overset{(28)}{\le} 2\|f - g\|_{L_p(\mu;X)}^{\frac{p}{q}}, \tag{29}$$

which completes the proof of the lemma. □

Equipped with the estimates of Lemma 3.1, we now ready to prove Proposition 1.4.

Proof of Proposition 1.4. — We shall first prove the rightmost inequality of (10). Let $x_1, \dots, x_n \in X$ and consider the function $f : \{1, \dots, n\} \to X$ given by $f(i) = x_i$. Unless all the vectors $x_1, \dots, x_n$ are equal, we can rescale so that the constraint

$$\sum_{i,j=1}^{n} \pi_i \pi_j \|x_i - x_j\|_X^p = 1 \tag{30}$$

is satisfied, which in particular, by Jensen's inequality, implies that $\|f - \mathbb{E}_\pi f\|_{L_p(\pi;X)} \le 1$, where $\mathbb{E}_\pi f \overset{\text{def}}{=} \sum_{i=1}^n \pi_i f(i)$. Consider the function $g = \mathsf{M}_{p,q}(f - \mathbb{E}_\pi f) : \{1, \dots, n\} \to X$.

(4) Since $M_{p,q} h = |h|^{p/q}\mathrm{sign}(h)$, the scalar case is a consequence of the pointwise inequalities

$$|\mathrm{sign}(\alpha)|\alpha|^\omega - \mathrm{sign}(\beta)|\beta|^\omega| \le c_1(\omega)|\alpha - \beta|^\omega$$

and

$$|\mathrm{sign}(\alpha)|\alpha|^{\frac{1}{\omega}} - \mathrm{sign}(\beta)|\beta|^{\frac{1}{\omega}}| \le c_2(\omega) \max\{|\alpha|, |\beta|\}^{\frac{1}{\omega}-1}|\alpha - \beta|$$

which are valid for every $\alpha, \beta \in \mathbb{R}$ and $\omega \in (0, 1]$.

Then, we have

$$1 = \sum_{i,j=1}^{n} \pi_i \pi_j \|f(i) - f(j)\|_X^p = \sum_{i,j=1}^{n} \pi_i \pi_j \|\mathsf{M}_{q,p} g(i) - \mathsf{M}_{q,p} g(j)\|_X^p$$

$$\overset{(25)}{\lesssim}_{p,q} \Big(\sum_{i,j=1}^{n} \pi_i \pi_j \|g(i) - g(j)\|_X^q \Big)^{\frac{p}{q}} \leq \gamma(B, \|\cdot\|_X^q)^{\frac{p}{q}} \Big(\sum_{i,j=1}^{n} \pi_i b_{ij} \|g(i) - g(j)\|_X^q \Big)^{\frac{p}{q}}$$

$$= \gamma(B, \|\cdot\|_X^q)^{\frac{p}{q}} \Big(\sum_{i,j=1}^{n} \pi_i b_{ij} \|\mathsf{M}_{p,q}(f - \mathbb{E}_\pi f)(i) - \mathsf{M}_{p,q}(f - \mathbb{E}_\pi f)(j)\|_X^q \Big)^{\frac{p}{q}}$$

$$\overset{(25)}{\lesssim}_{p,q} \gamma(B, \|\cdot\|_X^q)^{\frac{p}{q}} \Big(\sum_{i,j=1}^{n} \pi_i b_{ij} \|f(i) - f(j)\|_X^p \Big)^{\frac{p}{q}} = \gamma(B, \|\cdot\|_X^q)^{\frac{p}{q}} \Big(\sum_{i,j=1}^{n} \pi_i b_{ij} \|x_i - x_j\|_X^p \Big)^{\frac{p}{q}},$$

where in both inequalities we used that $\max\{\|f - \mathbb{E}_\pi f\|_{L_p(\pi;X)}, \|g\|_{L_q(\pi;X)}\} \leq 1$. Finally, taking an infimum of the right-hand side over all $x_1, \ldots, x_n \in X$ satisfying (30), we deduce that $1 \lesssim_{p,q} \gamma(B, \|\cdot\|_X^q)^{p/q} \gamma(B, \|\cdot\|_X^p)^{-p/q}$ which concludes the proof.

The proof of the leftmost inequality is almost identical yet we repeat it for completeness. Let $x_1, \ldots, x_n \in X$ and consider the function $\phi : \{1, \ldots, n\} \to X$ given by $\phi(i) = x_i$. Without loss of generality, we can again assume that the constraint

$$\sum_{i,j=1}^{n} \pi_i \pi_j \|x_i - x_j\|_X^q = 1, \tag{31}$$

is satisfied, which implies that $\|\phi - \mathbb{E}_\pi \phi\|_{L_q(\pi;X)} \leq 1$ by Jensen's inequality. Consider the function $\psi = \mathsf{M}_{q,p}(\phi - \mathbb{E}_\pi \phi) : \{1, \ldots, n\} \to X$. Then, we have

$$1 = \sum_{i,j=1}^{n} \pi_i \pi_j \|\phi(i) - \phi(j)\|_X^q = \sum_{i,j=1}^{n} \pi_i \pi_j \|\mathsf{M}_{p,q} \psi(i) - \mathsf{M}_{p,q} \psi(j)\|_X^q$$

$$\overset{(25)}{\lesssim}_{p,q} \sum_{i,j=1}^{n} \pi_i \pi_j \|\psi(i) - \psi(j)\|_X^p \leq \gamma(B, \|\cdot\|_X^p) \sum_{i,j=1}^{n} \pi_i b_{ij} \|\psi(i) - \psi(j)\|_X^p$$

$$= \gamma(B, \|\cdot\|_X^p) \sum_{i,j=1}^{n} \pi_i b_{ij} \|\mathsf{M}_{q,p}(\phi - \mathbb{E}_\pi \phi)(i) - \mathsf{M}_{q,p}(\phi - \mathbb{E}_\pi \phi)(j)\|_X^p$$

$$\overset{(25)}{\lesssim}_{p,q} \gamma(B, \|\cdot\|_X^p) \Big(\sum_{i,j=1}^{n} \pi_i b_{ij} \|\phi(i) - \phi(j)\|_X^q \Big)^{\frac{p}{q}} = \gamma(B, \|\cdot\|_X^p) \Big(\sum_{i,j=1}^{n} \pi_i b_{ij} \|x_i - x_j\|_X^q \Big)^{\frac{p}{q}}$$

where in both inequalities we used that $\max\{\|\phi - \mathbb{E}_\pi \phi\|_{L_q(\pi;X)}, \|\psi\|_{L_p(\pi;X)}\} \leq 1$. Finally, taking an infimum of the right-hand side over all $x_1, \ldots, x_n \in X$ satisfying (31), we deduce that $1 \lesssim_{p,q} \gamma(B, \|\cdot\|_X^p) \gamma(B, \|\cdot\|_X^q)^{-p/q}$ which concludes the proof. □

Even though the vector-valued version of Matoušek's extrapolation theorem (Proposition 1.4) suffices for the proof of Theorem 1.1 which will be presented in the next section, we digress to mention the following stronger result of Naor (2021).

Proposition 3.2. — *Fix $p \in (0, \infty)$ and $\theta \in (0, 1)$. There exists $D_0 = D_0(p, \theta) \in (1, \infty)$ such that the θ-snowflake of any normed space $(X, \|\cdot\|_X)$ embeds with p-average distortion D_0 into X.*

The discussion preceeding the statement of Theorem 1.2 shows that Proposition 1.4 is a formal consequence of Proposition 3.2, whose proof also relies on properties of the vector-valued Mazur map (24). In fact, as explained in Naor (2021, Remark 47), Proposition 3.2 implies improved bounds for the parameters $c(p,q), C(p,q)$ appearing in Proposition 1.4. Proposition 3.2 is the average distortion analogue of the following classical open problem in the nonlinear theory of Banach spaces.

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

A classical result of Schoenberg (1938) implies that for every $\theta \in (0,1)$, the θ-snowflake of L_2 admits an isometric embedding into L_2. Schoenberg's theorem was later extended by Bretagnolle, Dacunha-Castelle, and Krivine (1965/1966), who showed that for every $p \in (0,2]$ and $\theta \in (0,1)$, the θ-snowflake of L_p admits an isometric embedding into L_p. Despite decades of attention, Question 1 remains stubbornly open even for the spaces $X = L_p$, where $p \in (2,\infty)$. In the forthcoming work of Eskenazis and Naor (2021), it is proven that arbitrarily small logarithmic perturbations of this question have a negative answer. More precisely, it is shown that for every $\theta \in (0,1)$, $\eta \in (0, \min\{\theta, 1-\theta\})$ and $p \in (2,\infty)$, the metric transforms $(L_p, \omega_{\theta,\eta} \circ d_{L_p})$ of L_p do not admit a bi-Lipschitz embedding into L_p, where

$$\omega_{\theta,\eta}(t) = t^\theta \log^\eta(1+t) \qquad \text{or} \qquad \omega_{\theta,\eta}(t) = \frac{t^\theta}{1+\log^\eta(1+t)}. \tag{32}$$

We refer to Eskenazis (2019, Chapter 3) for further results in this direction.

4. PROOF OF THE AVERAGE JOHN THEOREM

Having established the duality principle of Theorem 1.2 and the extrapolation inequalities of Proposition 1.4, we are well equipped to proceed to the proof of Theorem 1.1 via Theorem 1.3. We start with some preliminary properties of *nonlinear Rayleigh quotients* which will help us analyze nonlinear spectral gaps. The following simplification of the original proof of Theorem 1.1 was sketched in Naor (2021, Remark 31).

4.1. Nonlinear Rayleigh quotients

Fix $p \geq 1$, a metric space $(\mathcal{M}, d_{\mathcal{M}})$ and a probability measure $\pi \in \triangle^{n-1}$. Let $L_p(\pi;\mathcal{M})$ be the metric space $(\mathcal{M}^n, d_{L_p(\pi,\mathcal{M})})$ whose metric is given by

$$d_{L_p(\pi;\mathcal{M})}(\boldsymbol{x},\boldsymbol{y}) = \Big(\sum_{i=1}^n \pi_i d_{\mathcal{M}}(x_i,y_i)^p\Big)^{\frac{1}{p}}, \tag{33}$$

where $\boldsymbol{x} = (x_1,\ldots,x_n), \boldsymbol{y} = (y_1,\ldots,y_n) \in \mathcal{M}^n$. Moreover, we shall use the ad hoc notation $L_p(\pi;\mathcal{M})^\dagger$ for the subset $L_p(\pi;M) \setminus \{(x,\ldots,x) : x \in \mathcal{M}\}$ of $L_p(\pi;\mathcal{M})$.

Let $A \in M_n(\mathbb{R})$ be a row-stochastic matrix that is π-stationary (in the sense that $\pi A = \pi$) and $\boldsymbol{x} = (x_1, \ldots, x_n) \in L_p(\pi; \mathcal{M})^\dagger$. Following Naor (2018, Section 5.1), we consider the corresponding nonlinear Rayleigh quotient given by

$$\mathcal{R}(\boldsymbol{x}; A, d_{\mathcal{M}}^p) \stackrel{\text{def}}{=} \frac{\sum_{i,j=1}^n \pi_i a_{ij} d_{\mathcal{M}}(x_i, x_j)^p}{\sum_{i,j=1}^n \pi_i \pi_j d_{\mathcal{M}}(x_i, x_j)^p}. \tag{34}$$

By definition, if A is π-reversible, the nonlinear spectral gap (6) satisfies

$$\gamma(A, d_{\mathcal{M}}^p) = \sup_{\boldsymbol{x} \in L_p(\pi; \mathcal{M})^\dagger} \frac{1}{\mathcal{R}(\boldsymbol{x}; A, d_{\mathcal{M}}^p)}. \tag{35}$$

We will need the following properties of nonlinear Rayleigh quotients.

Lemma 4.1. — *Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space, $p \geq 1$, $\lambda \in [0,1]$ and $\pi \in \triangle^{n-1}$. If $A, B \in M_n(\mathbb{R})$ are π-stationary stochastic matrices and $\boldsymbol{x} \in L_p(\pi; \mathcal{M})^\dagger$, then we have*

(i) $\mathcal{R}(\boldsymbol{x}; \lambda A + (1-\lambda)B, d_{\mathcal{M}}^p) = \lambda \mathcal{R}(\boldsymbol{x}; A, d_{\mathcal{M}}^p) + (1-\lambda)\mathcal{R}(\boldsymbol{x}; B, d_{\mathcal{M}}^p)$.
(ii) $\mathcal{R}(\boldsymbol{x}; \lambda A + (1-\lambda)\mathsf{Id}_n, d_{\mathcal{M}}^p) = \lambda \mathcal{R}(\boldsymbol{x}; A, d_{\mathcal{M}}^p)$, *where Id_n is the identity matrix.*
(iii) $\mathcal{R}(\boldsymbol{x}; AB, d_{\mathcal{M}}^p)^{\frac{1}{p}} \leq \mathcal{R}(\boldsymbol{x}; A, d_{\mathcal{M}}^p)^{\frac{1}{p}} + \mathcal{R}(\boldsymbol{x}; B, d_{\mathcal{M}}^p)^{\frac{1}{p}}$.
(iv) $\mathcal{R}(\boldsymbol{x}; B^t, d_{\mathcal{M}}^p) \leq t^p \mathcal{R}(\boldsymbol{x}; B, d_{\mathcal{M}}^p)$ *for every $t \in \mathbb{N}$.*

Proof. — The first property is evident from the definition (34) and the second follows from (i) since $\mathcal{R}(\boldsymbol{x}; \mathsf{Id}_n, d_{\mathcal{M}}^p) = 0$. Moreover, (iv) follows by iterating (iii) so we are left to prove that. Notice that AB is π-stationary and the triangle inequality gives

$$\begin{aligned}
\Big(\sum_{i,j=1}^n \pi_i (AB)_{ij} d_{\mathcal{M}}(x_i, x_j)^p\Big)^{\frac{1}{p}} &\leq \Big(\sum_{i,j=1}^n \pi_i \sum_{k=1}^n a_{ik} b_{kj} \big(d_{\mathcal{M}}(x_i, x_k) + d_{\mathcal{M}}(x_k, x_j)\big)^p\Big)^{\frac{1}{p}} \\
&\leq \Big(\sum_{i,j,k=1}^n \pi_i a_{ik} b_{kj} d_{\mathcal{M}}(x_i, x_k)^p\Big)^{\frac{1}{p}} + \Big(\sum_{i,j,k=1}^n \pi_i a_{ik} b_{kj} d_{\mathcal{M}}(x_k, x_j)^p\Big)^{\frac{1}{p}} \\
&= \Big(\sum_{i,k=1}^n \pi_i a_{ik} d_{\mathcal{M}}(x_i, x_k)^p\Big)^{\frac{1}{p}} + \Big(\sum_{k,j=1}^n \pi_k b_{kj} d_{\mathcal{M}}(x_k, x_j)^p\Big)^{\frac{1}{p}},
\end{aligned} \tag{36}$$

where in the last equality we used the stationarity of A in the form $\pi_k = \sum_{i=1}^n \pi_i a_{ik}$. The desired inequality (iii) follows from (36) after renormalizing. □

Apart from the elementary properties of Lemma 4.1, we shall also need the following standard computation of nonlinear Rayleigh quotients in Hilbert space. Recall that for every normed space $(X, \|\cdot\|_X)$, every matrix $B \in M_n(\mathbb{R})$ induces a linear operator $B \otimes \mathsf{Id}_X : L_p(\pi; X) \to L_p(\pi; X)$ that is given by $(B \otimes \mathsf{Id}_X)(x_1, \ldots, x_n) = (\sum_{j=1}^n b_{ij} x_j)_{i=1}^n$.

Lemma 4.2. — *Fix $\pi \in \triangle^{n-1}$ and let $B \in M_n(\mathbb{R})$ be a π-reversible stochastic matrix. For every Hilbert space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ and $\boldsymbol{x} \in L_2(\pi; \mathcal{H})^\dagger$ with $\sum_{i=1}^n \pi_i x_i = 0$, we have*

$$\mathcal{R}(\boldsymbol{x}; B^2, \|\cdot\|_{\mathcal{H}}^2) = 1 - \frac{\|(B \otimes \mathsf{Id}_{\mathcal{H}})\boldsymbol{x}\|_{L_2(\pi; \mathcal{H})}^2}{\|\boldsymbol{x}\|_{L_2(\pi; \mathcal{H})}^2}. \tag{37}$$

Proof. — Let $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ be the inner product of $\mathcal{H}$ and notice that

$$\sum_{i,j=1}^{n}\pi_i\pi_j\|x_i-x_j\|_{\mathcal{H}}^2 = 2\sum_{i=1}^{n}\pi_i\|x_i\|_{\mathcal{H}}^2 - 2\left\langle\sum_{i=1}^{n}\pi_i x_i,\sum_{j=1}^{n}\pi_j x_j\right\rangle_{\mathcal{H}} = 2\|\boldsymbol{x}\|_{L_2(\pi;\mathcal{H})}^2, \tag{38}$$

since $\sum_{i=1}^n \pi_i x_i = 0$. Moreover, since B^2 is π-stationary and stochastic, we have

$$\sum_{i,j=1}^{n}\pi_i(B^2)_{ij}\|x_i-x_j\|_{\mathcal{H}}^2 = \sum_{i,j=1}^{n}\pi_i(B^2)_{ij}(\|x_i\|_{\mathcal{H}}^2+\|x_j\|_{\mathcal{H}}^2) - 2\sum_{i=1}^{n}\pi_i\left\langle x_i,\sum_{j=1}^{n}(B^2)_{ij}x_j\right\rangle_{\mathcal{H}}$$
$$= 2\|\boldsymbol{x}\|_{L_2(\pi;\mathcal{H})}^2 - 2\sum_{i=1}^{n}\pi_i\Big\langle x_i,\big((B^2\otimes\mathsf{Id}_{\mathcal{H}})\boldsymbol{x}\big)_i\Big\rangle_{\mathcal{H}} = 2\|\boldsymbol{x}\|_{L_2(\pi;\mathcal{H})}^2 - 2\|(B\otimes\mathsf{Id}_{\mathcal{H}})\boldsymbol{x}\|_{L_2(\pi;\mathcal{H})}^2,$$

where in the last equality we additionally used the π-reversibility of B. The conclusion now readily follows by the definition (34) of nonlinear Rayleigh quotients. □

4.2. Proof of Theorem 1.3

In the proof of Theorem 1.3 we will use the following pointwise estimate of nonlinear Rayleigh quotients of normed spaces which are isomorphic to a Hilbert space.

Lemma 4.3. — *Let $(X,\|\cdot\|_X)$ be a normed space and $D\in[1,\infty)$. Suppose that there exists a Hilbertian norm $\|\cdot\|_{\mathcal{H}}: X\to\mathbb{R}_+$ such that*

$$\forall\, y\in X,\qquad \|y\|_{\mathcal{H}}\le\|y\|_X\le D\|y\|_{\mathcal{H}}. \tag{39}$$

Then, for every $\pi\in\triangle^{n-1}$ and every π-reversible stochastic matrix $B\in M_n(\mathbb{R})$,

$$\mathcal{R}(\boldsymbol{x};B^2,\|\cdot\|_{\mathcal{H}}^2)\ge 1-\eta^2 \quad\Longrightarrow\quad \mathcal{R}(\boldsymbol{x};B,\|\cdot\|_X^2)\ge\frac{(1-\eta D)^2}{4}, \tag{40}$$

where $\boldsymbol{x}\in L_2(\pi;X)^\dagger$ and $\eta\in(0,1/D)$.

Proof. — Without loss of generality, we can translate the components x_i of the vector $\boldsymbol{x}\in L_2(\pi;X)^\dagger$ to assume that $\sum_{i=1}^n\pi_i x_i=0$. Then, the assumption $\mathcal{R}(\boldsymbol{x};B^2,\|\cdot\|_{\mathcal{H}}^2)\ge 1-\eta^2$ can be equivalently rewritten due to Lemma 4.2 as

$$\|(B\otimes\mathsf{Id}_{\mathcal{H}})\boldsymbol{x}\|_{L_2(\pi;\mathcal{H})}\le\eta\|\boldsymbol{x}\|_{L_2(\pi;\mathcal{H})}. \tag{41}$$

Therefore,

$$\|(B\otimes\mathsf{Id}_X)\boldsymbol{x}\|_{L_2(\pi;X)}\overset{(39)}{\le}D\|(B\otimes\mathsf{Id}_{\mathcal{H}})\boldsymbol{x}\|_{L_2(\pi;\mathcal{H})}\overset{(41)}{\le}\eta D\|\boldsymbol{x}\|_{L_2(\pi;\mathcal{H})}\overset{(39)}{\le}\eta D\|\boldsymbol{x}\|_{L_2(\pi;X)}, \tag{42}$$

and thus, by the triangle inequality,

$$\|\boldsymbol{x}-(B\otimes\mathsf{Id}_X)\boldsymbol{x}\|_{L_2(\pi;X)}\ge\|\boldsymbol{x}\|_{L_2(\pi;X)}-\|(B\otimes\mathsf{Id}_X)\boldsymbol{x}\|_{L_2(\pi;X)}\overset{(42)}{\ge}(1-\eta D)\|\boldsymbol{x}\|_{L_2(\pi;X)}. \tag{43}$$

Since B is row-stochastic, Jensen's inequality for the convex function $\|\cdot\|_X^2$ gives

$$\sum_{i,j=1}^{n}\pi_i b_{ij}\|x_i-x_j\|_X^2\ge\sum_{i=1}^{n}\pi_i\Big\|x_i-\sum_{j=1}^{n}b_{ij}x_j\Big\|_X^2=\|\boldsymbol{x}-(B\otimes\mathsf{Id}_X)\boldsymbol{x}\|_{L_2(\pi;X)}^2. \tag{44}$$

On the other hand, using the triangle inequality we get

$$\sum_{i,j=1}^{n} \pi_i \pi_j \|x_i - x_j\|_X^2 \leq \sum_{i,j=1}^{n} \pi_i \pi_j (\|x_i\|_X + \|x_j\|_X)^2 \leq 4\|\boldsymbol{x}\|_{L_2(\pi;X)}^2. \tag{45}$$

Combining (34), (44), (43) and (45) we deduce that

$$\mathcal{R}(\boldsymbol{x}; B, \|\cdot\|_X^2) \geq \frac{(1-\eta D)^2}{4}, \tag{46}$$

which concludes the proof. □

Equipped with Lemma 4.3, we can complete the proof of Theorem 1.3. The main idea is to consider a Hilbertian norm which nicely approximates our given norm on $\mathbb{R}^d$ and then use the implication (40). In order to ensure that the assumption of (40) is satisfied we shall apply a trick that was used by Pisier (2010), who attributed it to V. Lafforgue: we will replace A by a large enough power of the form $\left(\frac{A+\mathsf{Id}_n}{2}\right)^t$. We will then be able to return to an inequality involving A rather than its power using Lemma 4.1.

Proof of Theorem 1.3. — Suppose that $X = (\mathbb{R}^d, \|\cdot\|_X)$ and fix $\pi \in \triangle^{n-1}$, a π-reversible stochastic matrix $A \in M_n(\mathbb{R})$ and a vector $\boldsymbol{x} \in L_1(\pi; X)^\dagger$. In view of (35), we need to prove a lower bound on $\mathcal{R}(\boldsymbol{x}; A, \|\cdot\|_X)$. Notice that, by properties (ii) and (iv) of Lemma 4.1, we have the inequality

$$\mathcal{R}(\boldsymbol{x}; A, \|\cdot\|_X) = 2\mathcal{R}\left(\boldsymbol{x}; \frac{A+\mathsf{Id}_n}{2}, \|\cdot\|_X\right) \geq \frac{2}{t}\mathcal{R}\left(\boldsymbol{x}; \left(\frac{A+\mathsf{Id}_n}{2}\right)^t, \|\cdot\|_X\right) \tag{47}$$

for every $t \in \mathbb{N}$. Moreover, by the vector-valued extrapolation inequalities of Proposition 1.4 and the expression (35) of nonlinear spectral gaps in terms of nonlinear Rayleigh quotients, we conclude that there exists a point $\boldsymbol{y} \in L_2(\pi; X)^\dagger$ satisfying

$$\mathcal{R}\left(\boldsymbol{x}; \left(\frac{A+\mathsf{Id}_n}{2}\right)^t, \|\cdot\|_X\right) \gtrsim \mathcal{R}\left(\boldsymbol{y}; \left(\frac{A+\mathsf{Id}_n}{2}\right)^t, \|\cdot\|_X^2\right). \tag{48}$$

Let $D_X \in [1, \infty)$ be the least constant for which there exists a Hilbertian norm $\|\cdot\|_{\mathcal{H}} : \mathbb{R}^d \to \mathbb{R}_+$ such that the following inequality is satisfied,

$$\forall\, y \in \mathbb{R}^d, \qquad \|y\|_{\mathcal{H}} \leq \|y\|_X \leq D_X \|y\|_{\mathcal{H}}. \tag{49}$$

As $\mathcal{H}$ is isometric to ℓ_2^d, the spectral gap of $\left(\frac{A+\mathsf{Id}_n}{2}\right)^{2t}$ with respect to $\|\cdot\|_{\mathcal{H}}^2$ satisfies

$$\gamma\left(\left(\frac{A+\mathsf{Id}_n}{2}\right)^{2t}, \|\cdot\|_{\mathcal{H}}^2\right) = \gamma\left(\left(\frac{A+\mathsf{Id}_n}{2}\right)^{2t}, |\cdot|^2\right) = \frac{1}{1-\left(\frac{1+\lambda_2(A)}{2}\right)^{2t}}. \tag{50}$$

Therefore, for the parameter

$$t^*(A) \stackrel{\text{def}}{=} \left\lceil \frac{\log(2D_X)}{\log(\frac{2}{1+\lambda_2(A)})} \right\rceil \lesssim \frac{\log(D_X+1)}{1-\lambda_2(A)} \tag{51}$$

we have the estimate

$$\gamma\left(\left(\frac{A+\mathsf{Id}_n}{2}\right)^{2t^*(A)}, \|\cdot\|_{\mathcal{H}}^2\right) \stackrel{(50)\wedge(51)}{\leq} \frac{1}{1-\frac{1}{4D_X^2}}, \tag{52}$$

which combined with (35) immediately implies that

$$\mathcal{R}\Big(\boldsymbol{y}; \Big(\frac{A+\mathsf{Id}_n}{2}\Big)^{2t^*(A)}, \|\cdot\|_{\mathcal{H}}^2\Big) \geq 1 - \frac{1}{4D_X^2}. \tag{53}$$

Therefore, in view of (49) and (53), the pointwise estimate of Lemma 4.3 applied to the matrix $B = \left(\frac{A+\mathsf{Id}_n}{2}\right)^{t^*(A)}$ and $\eta = \frac{1}{2D_X}$ implies that

$$\mathcal{R}\Big(\boldsymbol{y}; \Big(\frac{A+\mathsf{Id}_n}{2}\Big)^{t^*(A)}, \|\cdot\|_X^2\Big) \geq \frac{1}{16}. \tag{54}$$

Finally, combining (47), (48) and (54) for $t = t^*(A)$ we deduce that

$$\mathcal{R}(\boldsymbol{x}; A, \|\cdot\|_X) \gtrsim \frac{1}{t^*(A)} \overset{(51)}{\gtrsim} \frac{1-\lambda_2(A)}{\log(D_X+1)}. \tag{55}$$

By John's theorem (John, 1948), since X is d-dimensional we have $D_X \leq \sqrt{d}$ and the desired estimate (9) thus follows by rearranging (55) and using (35). □

As explained in the introduction, Theorem 1.1 is equivalent to Theorem 1.3 via the duality principle of Theorem 1.2.

Proof of Theorem 1.1. — Combining Theorems 1.3 and 1.2, we deduce that the $\frac{1}{2}$-snowflake of any d-dimensional normed space embeds into an ultrapower of ℓ_2 with quadratic average distortion at most $C\sqrt{\log d}$, where $d \geq 2$ and $C \in (0,\infty)$ is a universal constant. This immediately yields the conclusion of Theorem 1.1 since any ultrapower of ℓ_2 is itself a Hilbert space (Heinrich, 1980). □

5. BEYOND HILBERTIAN EMBEDDINGS

Theorem 1.1 is a special case of a much more general embedding theorem proven by Naor (2021). As a matter of fact, a lot of the ideas required to prove this more general statement have already been used in the Hilbertian case. A key ingredient required to go beyond Theorem 1.1 is the notion of *Markov type* introduced by Ball (1992).

Definition 5.1. — *A metric space $(\mathcal{M}, d_{\mathcal{M}})$ has Markov type $p \in (0,\infty)$ with constant $M \in (0,\infty)$ if for every $n \in \mathbb{N}$, $\pi \in \triangle^{n-1}$, every π-reversible matrix $A \in M_n(\mathbb{R})$ and every $\boldsymbol{x} \in L_p(\pi; \mathcal{M})$, we have*

$$\forall\, t \in \mathbb{N}, \qquad \mathcal{R}(\boldsymbol{x}; A^t, d_{\mathcal{M}}^p) \leq M^p t\, \mathcal{R}(\boldsymbol{x}; A, d_{\mathcal{M}}^p). \tag{56}$$

The least such constant $M \in (0,\infty)$ will be denoted by $\mathsf{M}_p(\mathcal{M})$.

In Ball (1992), it was shown that any Hilbert space $\mathcal{H}$ has $\mathsf{M}_2(\mathcal{H}) = 1$. Following Ball, Carlen, and Lieb (1994), we say that a normed space $(X, \|\cdot\|_X)$ is p-uniformly smooth, where $p \in [1,2]$, if there exists a constant $S \in (0,\infty)$ such that

$$\forall\, x, y \in X, \qquad \frac{\|x\|_X^p + \|y\|_X^p}{2} \leq \Big\|\frac{x+y}{2}\Big\|_X^p + S^p \Big\|\frac{x-y}{2}\Big\|_X^p. \tag{57}$$

The least such constant $S \in (0,\infty)$ will be denoted by $\mathsf{S}_p(X)$. A deep theorem of Naor, Peres, Schramm, and Sheffield (2006) asserts that every p-uniformly smooth normed space $(X, \|\cdot\|_X)$ has Markov type p with constant

$$\mathsf{M}_p(X) \lesssim \mathsf{S}_p(X). \tag{58}$$

An inspection of the proof of Theorem 1.3 reveals that the power of the norm in the estimate (9) can be improved for spaces of Markov type $p \in (1,2]$. Indeed, replacing (47) with (56) and using the extrapolation inequality (10), we deduce that for every $\boldsymbol{x} \in L_p(\pi; X)^\dagger$ there exists $\boldsymbol{y} \in L_2(\pi; X)^\dagger$ for which we have the estimate

$$\begin{aligned} \mathcal{R}(\boldsymbol{x}; A, \|\cdot\|_X^p) &= 2\mathcal{R}\Big(\boldsymbol{x}; \frac{A+\mathsf{Id}_n}{2}, \|\cdot\|_X^p\Big) \\ &\overset{(56)}{\underset{\mathsf{M}_p(X)}{\gtrsim}} \frac{1}{t}\mathcal{R}\Big(\boldsymbol{x}; \Big(\frac{A+\mathsf{Id}_n}{2}\Big)^t, \|\cdot\|_X^p\Big) \overset{(10)}{\gtrsim} \frac{1}{t}\mathcal{R}\Big(\boldsymbol{y}; \Big(\frac{A+\mathsf{Id}_n}{2}\Big)^t, \|\cdot\|_X^2\Big). \end{aligned} \tag{59}$$

Then, repeating the rest of the proof mutatis mutandis, we deduce the bound

$$\gamma(A, \|\cdot\|_X^p) \lesssim_{\mathsf{M}_p(X)} \frac{\log(\dim(X)+1)}{1-\lambda_2(A)}, \tag{60}$$

which, in view of (58) and Theorem 1.2, implies the following embeddability result.

Theorem 5.2. — *For every $S \in (0,\infty)$, there exists $C(S) \in (0,\infty)$ such that the following holds. If $p \in [1,2]$ and $(X, \|\cdot\|_X)$ is a finite-dimensional normed space with $\mathsf{S}_p(X) \le S$, then the $\frac{p}{2}$-snowflake of $(X, \|\cdot\|_X)$ admits an embedding into ℓ_2 with quadratic average distortion at most $C(S)\sqrt{\log(\dim(X)+1)}$.*

Theorem 1.1 is a special case of Theorem 5.2 as $\mathsf{S}_1(X) = 1$ for any $(X, \|\cdot\|_X)$. However, Theorem 5.2 is a refinement of the average John theorem in that it captures the fact that more structured normed spaces (i.e. spaces with bounded p-uniform smoothness constant) require a lesser amount of snowflaking in order to be embedded into ℓ_2 with quadratic average distortion which depends subpolynomially on the dimension. It is worth emphasizing that for 2-uniformly smooth spaces (such as $L_r(\mu)$ with $2 < r < \infty$), Theorem 5.2 shows that *no* snowflaking is necessary for such an embedding to exist.

This approach can be further exploited even for target spaces which are not Hilbertian. Following Ball, Carlen, and Lieb (1994), we say that a normed space $(X, \|\cdot\|_X)$ is q-uniformly convex, where $q \in [2,\infty)$, if there exists a constant $K \in (0,\infty)$ such that

$$\forall\, x, y \in X, \qquad \Big\|\frac{x+y}{2}\Big\|_X^q + \frac{1}{K^q}\Big\|\frac{x-y}{2}\Big\|_X^q \le \frac{\|x\|_X^q + \|y\|_X^q}{2}. \tag{61}$$

The least such constant $K \in (0,\infty)$ will be denoted by $\mathsf{K}_q(X)$. Observe that $\mathsf{K}_2(\ell_2) = 1$. Theorem 5.2 admits the following non-Hilbertian generalization. We shall denote by $\mathsf{c}_Y(\mathcal{M})$ the infimal distortion of a bi-Lipschitz embedding $f : (\mathcal{M}, d_{\mathcal{M}}) \to (Y, \|\cdot\|_Y)$.

Theorem 5.3. — *For every $S, K \in (0,\infty)$, there exists $C(S,K) \in (0,\infty)$ such that the following holds. If $1 \le p \le 2 \le q < \infty$, $(X, \|\cdot\|_X)$ is a Banach space with $\mathsf{S}_p(X) \le S$*

and $(Y, \|\cdot\|_Y)$ *is a Banach space with* $\mathsf{K}_q(Y) \le K$*, then the* $\frac{p}{q}$*-snowflake of* X *admits an embedding into* $\ell_q(Y)$ *with* q*-average distortion at most* $C(S,K)(\log(\mathsf{c}_Y(X)+1))^{1/q}$*.*

In view of the duality principle[(5)] of Theorem 1.2, Theorem 5.3 is equivalent the following nonlinear spectral gap inequality. For every $n \in \mathbb{N}$, $\pi \in \triangle^{n-1}$ and every π-reversible matrix $A \in M_n(\mathbb{R})$, we have

$$\gamma(A, \|\cdot\|_X^p) \lesssim_{\mathsf{M}_p(X), \mathsf{K}_q(Y), p, q} \log(\mathsf{c}_Y(X)+1)\gamma(A, \|\cdot\|_Y^q). \tag{62}$$

Fix $\boldsymbol{x} \in L_p(\pi; X)^\dagger$ with $\sum_{i=1}^n \pi_i x_i = 0$. Using Markov type and extrapolation as in (59), we deduce that for any $t \in \mathbb{N}$, there exists $\boldsymbol{y} \in L_q(\pi; X)^\dagger$ with $\sum_{i=1}^n \pi_i y_i = 0$ such that

$$\mathcal{R}(\boldsymbol{x}; A, \|\cdot\|_X^p) \gtrsim_{\mathsf{M}_p(X), p, q} \frac{1}{t}\mathcal{R}\Big(\boldsymbol{y}; \Big(\frac{A+\mathsf{Id}_n}{2}\Big)^t, \|\cdot\|_X^q\Big). \tag{63}$$

Moreover, if $B_t \stackrel{\text{def}}{=} (\frac{A+\mathsf{Id}_n}{2})^t$, the argument of (44) and (45) implies that

$$\begin{aligned}\mathcal{R}(\boldsymbol{y}; B_t, \|\cdot\|_X^q) &\gtrsim_q \frac{\|\boldsymbol{y} - (B_t \otimes \mathsf{Id}_X)\boldsymbol{y}\|_{L_q(\pi;X)}^q}{\|\boldsymbol{y}\|_{L_q(\pi;X)}^q} \\ &\ge \Big(1 - \|B_t \otimes \mathsf{Id}_X\|_{L_q^0(\pi;X)\to L_q^0(\pi;X)}\Big)^q,\end{aligned} \tag{64}$$

where $L_q^0(\pi; Z) = \{\boldsymbol{z} \in L_q(\pi; Z) : \sum_{i=1}^n \pi_i z_i = 0\}$. Therefore, we have

$$\gamma(A, \|\cdot\|_X^p) \stackrel{(63)\wedge(35)\wedge(64)}{\lesssim_{\mathsf{M}_p(X), p, q}} \frac{t}{(1 - \|B_t \otimes \mathsf{Id}_X\|_{L_q^0(\pi;X)\to L_q^0(\pi;X)})^q}. \tag{65}$$

Notice that by the definition of $\mathsf{c}_Y(X)$,

$$\|B_t \otimes \mathsf{Id}_X\|_{L_q^0(\pi;X)\to L_q^0(\pi;X)} \le \mathsf{c}_Y(X)\|B_t \otimes \mathsf{Id}_Y\|_{L_q^0(\pi;Y)\to L_q^0(\pi;Y)} \tag{66}$$

and moreover

$$\begin{aligned}\|B_t \otimes \mathsf{Id}_Y\|_{L_q^0(\pi;Y)\to L_q^0(\pi;Y)} &= \Big\|\Big(\frac{A+\mathsf{Id}_n}{2}\Big)^t \otimes \mathsf{Id}_Y\Big\|_{L_q^0(\pi;Y)\to L_q^0(\pi;Y)} \\ &\le \Big\|\Big(\frac{A+\mathsf{Id}_n}{2}\Big) \otimes \mathsf{Id}_Y\Big\|_{L_q^0(\pi;Y)\to L_q^0(\pi;Y)}^t.\end{aligned} \tag{67}$$

Combining (65), (66) and (67), we finally deduce that for any $t \in \mathbb{N}$,

$$\gamma(A, \|\cdot\|_X^p) \lesssim_{\mathsf{M}_p(X), p, q} t \cdot \Big(1 - \mathsf{c}_Y(X)\Big\|\Big(\frac{A+\mathsf{Id}_n}{2}\Big) \otimes \mathsf{Id}_Y\Big\|_{L_q^0(\pi;Y)\to L_q^0(\pi;Y)}^t\Big)^{-q}. \tag{68}$$

Optimizing over t we thus conclude that

$$\gamma(A, \|\cdot\|_X^p) \lesssim_{\mathsf{M}_p(X), p, q} \frac{\log(\mathsf{c}_Y(X)+1)}{\log\Big(1/\|(\frac{A+\mathsf{Id}_n}{2}) \otimes \mathsf{Id}_Y\|_{L_q^0(\pi;Y)\to L_q^0(\pi;Y)}\Big)} \tag{69}$$

(5) Observe that a direct application of Theorem 1.2 and (62) would imply that the $\frac{p}{q}$-snowflake of X admits such an embedding into an ultrapower of $\ell_q(Y)$ rather than $\ell_q(Y)$ itself. Taking an ultrapower in this statement is redundant if X is assumed to be p-uniformly smooth and Y is q-uniformly convex, as was shown in Naor (2021, Corollary 23). We shall not address this delicate issue here.

Observe that so far we have been very closely following the Hilbertian proof. Indeed, if Y is a Hilbert space and $q = 2$, then the operator norm appearing in (69) is simply $\frac{1+\lambda_2(A)}{2}$ and thus (9) follows from (69) and John's theorem which asserts that $\mathsf{c}_{\ell_2}(X) \leq \sqrt{\dim X}$. In the general (Banach space-valued) setting of Theorem 5.3, we need a more robust argument to show that the operator norm $\left\|(\frac{A+\mathsf{Id}_n}{2}) \otimes \mathsf{Id}_Y\right\|_{L_q^0(\pi;Y)\to L_q^0(\pi;Y)}$ is bounded away from 1 by a quantity which depends on the nonlinear spectral gap $\gamma(A, \|\cdot\|_Y^q)$. To do this, we will leverage the q-uniform convexity of the normed space $(Y, \|\cdot\|_Y)$.

Fix a metric space $(\mathcal{M}, d_{\mathcal{M}})$ and $q \in (0,\infty)$. If $\pi \in \triangle^{n-1}$ and $A \in M_n(\mathbb{R})$ is a π-reversible stochastic matrix, the *nonlinear absolute spectral gap* of A with respect to $d_{\mathcal{M}}^q$, denoted by $\gamma_+(A, d_{\mathcal{M}}^q)$, is the least constant $\gamma_+ \in (0,\infty]$ such that

$$\forall\, x_1,\ldots,x_n,y_1,\ldots,y_n \in \mathcal{M}, \qquad \sum_{i,j=1}^n \pi_i\pi_j d_{\mathcal{M}}(x_i,y_j)^q \leq \gamma_+ \sum_{i,j=1}^n \pi_i a_{ij} d_{\mathcal{M}}(x_i,y_j)^q. \tag{70}$$

The terminology stems from the fact that $\gamma_+(A, |\cdot|^2) = (1-\max_{i=2,\ldots,n}|\lambda_i(A)|)^{-1}$, where $1 = \lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A) \geq -1$ are the eigenvalues of A. Nonlinear spectral gaps and nonlinear absolute spectral gaps are related via the following inequalities.

Lemma 5.4. — *Fix $q \in [1,\infty)$, $n \in \mathbb{N}$ and $\pi \in \triangle^{n-1}$. For every π-reversible stochastic matrix $A \in M_n(\mathbb{R})$ and every metric space $(\mathcal{M}, d_{\mathcal{M}})$, we have*

$$2\gamma(A, d_{\mathcal{M}}^q) \leq \gamma_+\Big(\frac{A+\mathsf{Id}_n}{2}, d_{\mathcal{M}}^q\Big) \leq 2^{2q+1}\gamma(A, d_{\mathcal{M}}^q). \tag{71}$$

The elementary proof of Lemma 5.4 can be found in Naor (2014, Lemma 2.3). The pertinence of absolute spectral gaps in the ensuing discussion is that, in the case of uniformly convex spaces, they have a useful connection to vector-valued operator norms of adjacency matrices. This is manifested by the following proposition of Mendel and Naor (2014, Lemma 6.6), whose proof relies on Pisier's martingale cotype inequality for q-uniformly convex spaces (Pisier, 1975a).

Proposition 5.5. — *Fix $q \in [2,\infty)$ and let $(Y, \|\cdot\|_Y)$ be a q-uniformly convex normed space. Then, for every $n \in \mathbb{N}$, $\pi \in \triangle^{n-1}$ and every π-reversible stochastic matrix $C \in M_n(\mathbb{R})$, we have*

$$\|C \otimes \mathsf{Id}_Y\|_{L_q^0(\pi;Y)\to L_q^0(\pi;Y)} \leq \Big(1 - \frac{1}{(2^{q-1}-1)\mathsf{K}_q(Y)^q\gamma_+(C, \|\cdot\|_Y^q)}\Big)^{\frac{1}{q}}. \tag{72}$$

Proposition 5.5 is proven by Mendel and Naor (2014) for the special case that π is the uniform measure on $\{1,\ldots,n\}$ and C is a symmetric stochastic matrix. The proof of the general statement presented here is similar to this special case and we thus omit it. Plugging the bound (72) in (69) for $C = \frac{A+\mathsf{Id}_n}{2}$, we finally deduce that

$$\begin{aligned}\gamma(A, \|\cdot\|_X^p) &\overset{(69)\wedge(72)}{\lesssim}_{\mathsf{M}_p(X),\mathsf{K}_q(Y),p,q} \log(\mathsf{c}_Y(X)+1)\gamma_+\Big(\frac{A+\mathsf{Id}_n}{2}, \|\cdot\|_Y^q\Big) \\ &\overset{(71)}{\lesssim}_q \log(\mathsf{c}_Y(X)+1)\gamma(A, \|\cdot\|_Y^q). \qquad \square\end{aligned} \tag{73}$$

6. GEOMETRIC AND ALGORITHMIC APPLICATIONS

In this final section, we present a selection of geometric and algorithmic applications of nonlinear spectral gaps (mostly without proofs) and related open questions.

6.1. Nonembeddability of expanders into low-dimensional normed spaces

Let $G = (V, E)$ be a d-regular graph on the vertex set $\{1, \ldots, n\}$. We shall denote by A_G the normalized adjacency matrix of G, that is, the $n \times n$ symmetric stochastic matrix whose entries are given by $(A_G)_{ij} = \frac{\mathbf{1}_{\{i,j\}\in E}}{d}$, where $i, j \in \{1, \ldots, n\}$. A sequence $\{G_n = (V_n, E_n)\}_{n=1}^{\infty}$ of d-regular graphs with $|V_n| \to \infty$ as $n \to \infty$ is called an expander graph sequence if $\sup_{n\in\mathbb{N}} \gamma(A_{G_n}) < \infty$. The existence of regular expander graph sequences is a classical fact that can be proven via the probabilistic method (Pinsker, 1973 and Bollobás, 1988), while deterministic constructions are notoriously more involved (see, e.g., the book of Davidoff, Sarnak, and Valette, 2003). Embeddability properties of connected expanders viewed as metric spaces when equipped with the shortest path distance were first investigated by Linial, London, and Rabinovich (1995) who, among other results, showed that if an n-vertex d-regular expander embeds with quadratic average distortion D in a k-dimensional normed space, then $k \gtrsim (\log_d n)^2/D^2$. We shall now present the following (sharp) improvement of Linial, London and Rabinovich's result due to Naor (2017) as a consequence of the average John theorem.

Theorem 6.1. — *For every $q \in [1, \infty)$, there exists $c(q) \in (0, \infty)$ such that the following holds for every $\gamma, D \in [1, \infty)$. Let $G = (V, E)$ be a d-regular connected graph on n vertices with $\gamma(A_G) \le \gamma$ and let $(X, \|\cdot\|_X)$ be a normed space such that (G, d_G) admits an embedding into X with q-average distortion at most D. Then,*

$$\dim(X) \ge n^{c(q)/\gamma D \log d}. \tag{74}$$

Proof. — We shall first prove the case $q = 1$. By the assumption, there exists a D-Lipschitz map[(6)] $f : (G, d_G) \to (X, \|\cdot\|_X)$ satisfying the average lower bound

$$\frac{1}{n^2} \sum_{u,v\in V} \|f(u) - f(v)\|_X \ge \frac{1}{n^2} \sum_{u,v\in V} d_G(u, v). \tag{75}$$

Let $k = \dim(X)$. Applying Theorem 1.1 for the measure $\mu = \frac{1}{n}\sum_{u\in V} \delta_{f(u)}$ on X, we deduce that there exists a $O(\sqrt{\log k})$-Lipschitz function $h : (X, \|\cdot\|_X^{1/2}) \to \ell_2$ such that

$$\frac{1}{n^2} \sum_{u,v\in V} \|h(f(u)) - h(f(v))\|_{\ell_2}^2 \ge \frac{1}{n^2} \sum_{u,v\in V} \|f(u) - f(v)\|_X. \tag{76}$$

(6) As is common, we shall identify the graph G with its vertex set V thus writing $f : G \to X$ rather than $f : V \to X$. Moreover, we will always denote by d_G the shortest path distance on V.

Since the graph G is a regular expander, inequality (5) implies that

$$\begin{aligned}\frac{1}{n^2}\sum_{u,v\in V}\|h(f(u))-h(f(v))\|_{\ell_2}^2 &\le \frac{2\gamma}{dn}\sum_{\{a,b\}\in E}\|h(f(a))-h(f(b))\|_{\ell_2}^2\\ &\lesssim \frac{\gamma\log k}{dn}\sum_{\{a,b\}\in E}\|f(a)-f(b)\|_X \lesssim \gamma D\log k,\end{aligned}\tag{77}$$

where in the last two inequalities we used the Lipschitz conditions for h and f. On the other hand, the graph G is d-regular and therefore, for any fixed $u\in V$ there exist at least $\frac{n}{2}$ vertices $v\in V$ such that $d_G(u,v)\ge\lfloor\log_d(n/2)\rfloor$. Hence, we have

$$\frac{1}{n^2}\sum_{u,v\in V} d_G(u,v)\gtrsim \log_d n,\tag{78}$$

which, combined with (75), (76) and (77), implies that

$$\log_d n\lesssim \gamma D\log k,\tag{79}$$

thus completing the proof of (74) for $q=1$. To address the general case $q\ge 1$, we need a slight modification of this argument. It is a formal consequence of Naor (2021, Proposition 6) and Theorem 1.1, that for any $q\ge 1$, the $\frac{1}{2}$-snowflake of any finite-dimensional normed space X embeds into ℓ_2 with $(2q)$-average distortion at most $C(q)\sqrt{\log(\dim(X)+1)}$. Considering a $O(\sqrt{\log k})$-Lipschitz embedding satisfying the analogue of (76) with power $2q$ instead of the embedding h and repeating the above argument completes the proof of (74) for general $q\ge 1$. □

A few historical comments are in order. Due to the existence of regular expander graph sequences, Theorem 6.1 implies that for arbitrarily large n, there exists an n-point metric space $(\mathcal{M}_n, d_{\mathcal{M}_n})$ such that if $\mathcal{M}_n$ admits an embedding with bi-Lipschitz distortion D into a finite-dimensional normed space X, then $\dim(X)\ge n^{c/D}$ for some universal constant $c\in(0,\infty)$. Therefore, Theorem 6.1 provides a negative answer to the question of Johnson and Lindenstrauss (1984) discussed in Section 1.4. A different negative answer to this question had been given in important work of Matoušek (1996), who devised an ingenious random family of metric spaces and showed that they satisfy this property using input from real algebraic geometry. It is worth mentioning that a precursor of Theorem 6.1 is a result of Lee, Mendel, and Naor (2005, Proposition 4.1), who showed that if an n-vertex regular expander embeds in ℓ_∞^d with bi-Lipschitz distortion at most D, then $d\ge n^{c/D}$ for some universal constant $c\in(0,\infty)$.

Quantitatively, Theorem 6.1 provides a sharp relation between the dimension of the target space X, the number of vertices of G and the distortion D. Indeed, a classical theorem of Johnson, Lindenstrauss, and Schechtman (1987) asserts that for every $n\in\mathbb{N}$ and $D\ge 1$, any n-point metric space $\mathcal{M}$ admits a bi-Lipschitz embedding with distortion at most D into some d-dimensional normed space X, where $d\lesssim_D n^{C/D}$ for some universal constant $C\in(0,\infty)$. This result was later refined by Matoušek (1992), who showed that one can always take $X=\ell_\infty^d$ as a target space in this statement.

Remark 6.2. — The optimality of Theorem 6.1 which follows from the works of Johnson, Lindenstrauss, and Schechtman (1987) and Matoušek (1992) immediately implies that the $O(\sqrt{\log \dim(X)})$ upper bound for the average distortion in Theorem 1.1 is sharp. Indeed, suppose that the $\frac{1}{2}$-snowflake of $X = \ell_\infty^d$ admitted an embedding into ℓ_2 with quadratic average distortion $o(\sqrt{\log d})$. Then, the proof of Theorem 6.1 would show that if an n-vertex expander embeds with bi-Lipschitz distortion D in ℓ_∞^d, then

$$\log n = o(D \log d). \tag{80}$$

However, this inequality contradicts the embedding theorem of Matoušek (1992).

Following the terminology of Naor (2018), we say that an infinite-dimensional Banach space $(X, \|\cdot\|_X)$ admits (quadratic) average dimension reduction with distortion $D \in (1, \infty)$ if for any $n \in \mathbb{N}$ there exists $k_n = k_n^D(X) \in \mathbb{N}$ satisfying

$$\lim_{n\to\infty} \frac{\log k_n}{\log n} = 0 \tag{81}$$

such that the following condition holds. For any n points $x_1, \ldots, x_n \in X$, there exists a subspace $F = F(x_1, \ldots, x_n)$ of X with $\dim F \le k_n$ and points $y_1, \ldots, y_n \in F$ satisfying $\|y_i - y_j\|_X \le D\|x_i - x_j\|_X$ for every $i, j \in \{1, \ldots, n\}$ and

$$\frac{1}{n^2} \sum_{i,j=1}^{n} \|y_i - y_j\|_X^2 \ge \frac{1}{n^2} \sum_{i,j=1}^{n} \|x_i - x_j\|_X^2. \tag{82}$$

As every finite metric space embeds isometrically in ℓ_∞, the aforementioned result of Matoušek (1996) (or Theorem 6.1) implies that ℓ_∞ does not admit average dimension reduction with any distortion $D > 1$. The following tantalizing question remains open.

Question 2. — *Does ℓ_1 admit average dimension reduction with any distortion $D > 1$?*

We note that the bi-Lipschitz analogue of Question 2 is answered by a famous theorem of Brinkman and Charikar (2005) (see also Lee and Naor (2004) for a different influential proof) who showed that for arbitrarily large n and $D > 1$ there exists an n-point subset of ℓ_1 which does not admit a bi-Lipschitz embedding into any subspace of ℓ_1 of dimension at most n^{c/D^2}, where $c \in (0, \infty)$ is a universal constant.

6.2. Average distortion embeddings of ℓ_p into ℓ_2

In Theorem 1.1, it was established that any finite-dimensional normed space $(X, \|\cdot\|_X)$ admits an embedding into ℓ_2 with quadratic average distortion $O(\sqrt{\log(\dim(X) + 1)})$ via the nonlinear spectral gap inequality (9). As explained in Remark 6.2, this estimate for the quadratic average distortion is asymptotically optimal yet, quite surprisingly, there exist many non-Hilbertian normed spaces which embed with constant quadratic average distortion in ℓ_2. The following result is the main theorem of Naor (2014).

Theorem 6.3. — *There exists $C \in (0, \infty)$ such that for any $p \in (2, \infty)$, the normed space ℓ_p admits an embedding into ℓ_2 with quadratic average distortion Cp.*

Theorem 6.3 is established in Naor (2014) via the nonlinear spectral gap inequality

$$\forall\, p > 2, \qquad \gamma(A, \|\cdot\|_{\ell_p}^2) \lesssim \frac{p^2}{1-\lambda_2(A)}, \tag{83}$$

which holds for any $\pi \in \triangle^{n-1}$ and any π-reversible stochastic matrix $A \in M_n(\mathbb{R})$, and the duality principle of Theorem 1.2. Once again, (83) is proven in Naor (2014) in the special case that π is the uniform measure on $\{1, \ldots, n\}$ and A is a symmetric stochastic matrix. The proof of the more general statement presented here (which is equivalent to Theorem 6.3) is identical. In Naor (2014), Theorem 6.3 and (83) were used to give new lower bounds for the ℓ_p-distortion of random connected d-regular graphs, Ramanujan graphs and abelian Alon–Roichman graphs, improving earlier results of Matoušek (1997). It is worth pointing out that (83) is no longer valid when $p \in [1, 2)$.

6.3. Expanders with respect to Banach spaces

Combinatorial expanders are ubiquitous geometric objects whose metric structure is notoriously incompatible with Euclidean geometry. Nonlinear spectral gaps allow us to analyze non-Euclidean analogues of these exotic metrics. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space. A sequence $\{G_n\}_{n=1}^{\infty}$ of d-regular graphs with $|G_n| \to \infty$ is called an expander graph sequence with respect to $\mathcal{M}$ if $\sup_{n\in\mathbb{N}} \gamma(A_{G_n}, d_{\mathcal{M}}^2) < \infty$. If such graphs exist, we say that $(\mathcal{M}, d_{\mathcal{M}})$ admits a sequence of d-regular expanders. The following influential observation on the embeddability of expanders is essentially due to Matoušek (1997).

Proposition 6.4. — *Let $(\mathcal{M}, d_{\mathcal{M}})$ be a metric space and fix $\gamma, q \in (0, \infty)$. Suppose that $G = (V, E)$ is a d-regular connected graph on n vertices with $\gamma(A_G, d_{\mathcal{M}}^q) \le \gamma$. If (G, d_G) embeds into $(\mathcal{M}, d_{\mathcal{M}})$ with q-average distortion at most D, then $D \gtrsim \frac{\log_d n}{\gamma^{1/q}}$.*

Proof. — By the assumption, there exists $\sigma \in (0, \infty)$ and a σD-Lipschitz map $f : (G, d_G) \to (\mathcal{M}, d_{\mathcal{M}})$ satisfying the average lower bound

$$\frac{1}{n^2} \sum_{u,v\in V} d_{\mathcal{M}}\Big(f(u), f(v)\Big)^q \ge \frac{\sigma^q}{n^2} \sum_{u,v\in V} d_G(u,v)^q \gtrsim \sigma^q (\log_d n)^q, \tag{84}$$

where the second inequality follows from (78) and Jensen's inequality. On the other hand, by the definition (6) of $\gamma(A, d_{\mathcal{M}}^q)$, we have

$$\frac{1}{n^2} \sum_{u,v\in V} d_{\mathcal{M}}\Big(f(u), f(v)\Big)^q \le \frac{2\gamma}{dn} \sum_{\{a,b\}\in E} d_{\mathcal{M}}\Big(f(a), f(b)\Big)^q \le \gamma \sigma^q D^q, \tag{85}$$

where the last inequality follows from the Lipschitz condition for f. Rearranging, we deduce the desired lower bound for the q-average distortion D. □

Deciding whether a given non-Euclidean metric space admits a sequence of expanders is a notoriously difficult problem in metric geometry, even when specified to normed spaces. By Proposition 6.4, it is clear that if there exists a sequence $\{G_n\}_{n=1}^{\infty}$ of regular expanders with respect to a normed space $(X, \|\cdot\|_X)$, then X cannot contain subspaces uniformly isomorphic to $\{\ell_\infty^m\}_{m=1}^{\infty}$ as it would then bi-Lipschitzly contain all

finite metric spaces with uniform distortion. Normed spaces which do not uniformly contain $\{\ell_\infty^m\}_{m=1}^\infty$ are said to have finite cotype in Banach space theory jargon (Maurey, 2003). Strikingly, this is the only known necessary condition for a normed space to admit an expander graph sequence and the following general question remains open.

Question 3. — *Is every combinatorial expander also an expander with respect to any normed space of finite cotype?*

Such implications, asserting that a classical spectral gap implies a nonlinear spectral gap, are currently only known for substantially smaller classes of normed spaces from works of Matoušek (1997), Ozawa (2004), Pisier (2010) and Naor and Silberman (2011). It is worth mentioning that even the following question, which is formally weaker than Question 3 in view of Proposition 6.4, remains open.

Question 4. — *Does there exist a sequence of finite metric spaces $\{(\mathcal{M}_n, d_{\mathcal{M}_n})\}_{n=1}^\infty$ with $|\mathcal{M}_n| \to \infty$ as $n \to \infty$ such that for any normed space $(X, \|\cdot\|_X)$ of finite cotype, the bi-Lipschitz distortion required to embed $\mathcal{M}_n$ into X satisfies $\mathsf{c}_X(\mathcal{M}_n) \gtrsim_X \log|\mathcal{M}_n|$?*

A positive answer to Question 4 would imply a striking dichotomy in the embeddability of finite metric spaces into infinite-dimensional normed spaces. If such a normed space X does not have finite cotype, then it bi-Lipschitzly contains every finite metric space with distortion $1+\varepsilon$ for any $\varepsilon > 0$ (Maurey, 2003). On the other hand, if X is an arbitrary infinite-dimensional space, then any finite metric space $\mathcal{M}$ admits a bi-Lipschitz embedding into X with distortion $O(\log|\mathcal{M}|)$ by the theorems of Dvoretzky (1961) and Bourgain (1985). A positive answer to Question 4 would imply that this bound is always optimal under the (necessary) assumption that X has finite cotype.

In regard to Question 3, even the *existence* of a sequence $\{G_n\}_{n=1}^\infty$ of regular graphs which are expanders with respect to any space of finite cotype remains unknown. The strongest available result in this direction is the following profound theorem of Lafforgue (2008), whose proof is an ingenious combination of algebraic and vector-valued harmonic analytic methods. We say that a normed space $(X, \|\cdot\|_X)$ has nontrivial type if X does not contain subspaces uniformly isomorphic to $\{\ell_1^m\}_{m=1}^\infty$ (Maurey, 2003). Any space of nontrivial type has finite cotype, but the converse is not true (e.g. for ℓ_1).

Theorem 6.5. — *There exists a sequence of regular graphs $\{G_n\}_{n=1}^\infty$ which is an expander graph sequence with respect to any normed space of non-trivial type.*

Lafforgue's graphs can be obtained as Cayley graphs of finite quotients of co-compact lattices in $SL_3(\mathbb{Q}_p)$, where p is a prime and $\mathbb{Q}_p$ is the field of p-adic rationals.

A completely different construction of a sequence of regular graphs which are expanders with respect to a large family of norms was presented in work of Mendel and Naor (2014). Theirs is a vector-valued adaptation of the zig-zag product construction of Reingold, Vadhan, and Wigderson (2002) and the resulting graphs are expanders with respect to any normed space which admits an equivalent uniformly convex norm. Clearly any such space has nontrivial type but the converse is not true

(Pisier, 1975b). While we will not outline the argument of Mendel and Naor (2014), it is worth pointing out that it consists of a novel construction of a base graph along with an adaptation of the zig-zag iteration of Reingold, Vadhan, and Wigderson (2002). The necessity of the uniform convexity assumption in this argument stems from this iteration procedure. On the other hand, the construction of the base graph (which was straightforward in the case of combinatorial expanders) has raised influential questions in vector-valued harmonic analysis that led to investigations of independent interest (Mendel and Naor, 2014; Eskenazis and Ivanisvili, 2020, 2021).

6.4. Expanders with respect to Alexandrov spaces

A complete geodesic metric space $(\mathcal{M}, d_{\mathcal{M}})$ is an Alexandrov space of nonpositive curvature (or a CAT(0) space) if for any quadruple of points $x, y, z, m \in \mathcal{M}$ such that m is a metric midpoint of x and y, that is, $d_{\mathcal{M}}(m, x) = d_{\mathcal{M}}(m, y) = \frac{1}{2} d_{\mathcal{M}}(x, y)$, we have

$$d_{\mathcal{M}}(z, m)^2 \leq \frac{1}{2} d_{\mathcal{M}}(z, x)^2 + \frac{1}{2} d_{\mathcal{M}}(z, y)^2 - \frac{1}{4} d_{\mathcal{M}}(x, y)^2. \tag{86}$$

If the reverse inequality holds true for any such quadruple $x, y, z, m \in \mathcal{M}$, then $\mathcal{M}$ is an Alexandrov space of nonnegative curvature. Alexandrov spaces of nonpositive (respectively nonnegative) curvature are (potentially singular) metric spaces which generalize Riemannian manifolds with nonpositive (resp. nonnegative) sectional curvature.

An argument of Wang (1998) shows that any regular combinatorial expander is also an expander with respect to any Hilbert manifold with a CAT(0) Riemannian metric (see also Naor and Silberman, 2011, Corollary 4.10). The first systematic study of expanders with respect to (non-smooth) Alexandrov spaces of nonpositive curvature was undertaken by Mendel and Naor (2015), who showed the following theorem.

Theorem 6.6. — *There exists a CAT(0) space $(\mathcal{M}, d_{\mathcal{M}})$ and a sequence $\{G_n\}_{n=1}^{\infty}$ of 3-regular graphs such that $\sup_{n \in \mathbb{N}} \gamma(A_{G_n}, d_{\mathcal{M}}^2) < \infty$, yet a random d-regular graph $\mathbf{G}$ on n vertices satisfies $\gamma(A_{\mathbf{G}}, d_{\mathcal{M}}^2) \gtrsim (\log_d n)^2$ with probability $1 - o_n(1)$ as $n \to \infty$.*

Theorem 6.6 reveals a striking difference between nonlinear spectral gaps with respect to Alexandrov spaces of nonpositive curvature and classical spectral gaps, as a random d-regular graph on n-vertices is a combinatorial expander with probability $1 - o_n(1)$ as $n \to \infty$ for any fixed $d \in \mathbb{N}$ (Bollobás, 1988). The following question remains open.

Question 5. — *Does every CAT(0) space admit a sequence of regular expanders? More ambitiously, does there exists a sequence of $O(1)$-regular graphs $\{G_n\}_{n=1}^{\infty}$ with $|G_n| \to \infty$ such that $\sup_{n \in \mathbb{N}} \gamma(A_{G_n}, d_{\mathcal{M}}^2) < \infty$ for every CAT(0) space $(\mathcal{M}, d_{\mathcal{M}})$?*

A positive answer to the stronger statement in Question 5 would imply (in view of Proposition 6.4) the existence of arbitrarily large finite metric spaces requiring logarithmic distortion to be embedded in any Alexandrov space of nonpositive curvature. The following question was asked by Eskenazis, Mendel, and Naor (2019).

Question 6. — *Does there exist a sequence of finite metric spaces* $\{(\mathcal{M}_n, d_{\mathcal{M}_n})\}_{n=1}^{\infty}$ *with* $|\mathcal{M}_n| \to \infty$ *as* $n \to \infty$ *such that for any CAT(0) space* $(\mathcal{N}, \|\cdot\|_{\mathcal{N}})$*, the bi-Lipschitz distortion required to embed* $\mathcal{M}_n$ *into* $\mathcal{N}$ *satisfies* $\mathsf{c}_{\mathcal{N}}(\mathcal{M}_n) \gtrsim_n \log|\mathcal{M}_n|$*?*

In the dual nonnegative curvature regime, the analogue of Question 5 was answered by Andoni, Naor, and Neiman (2018), who showed that there exists an Alexandrov space of nonnegative curvature which does not admit any sequence of regular expanders. Moreover, they asked the following dual to Question 6.

Question 7. — *Does there exist an Alexandrov space of nonnegative curvature* $(\mathcal{N}, d_{\mathcal{N}})$ *such that any finite metric space* $(\mathcal{M}, d_{\mathcal{M}})$ *embeds into* $\mathcal{N}$ *with bi-Lipschitz distortion* $\mathsf{c}_{\mathcal{N}}(\mathcal{M}) \lesssim \sqrt{\log|\mathcal{M}|}$*?*

In their paper, they specifically asked Question 7 for the concrete Alexandrov space $\mathcal{P}_2(\mathbb{R}^3)$, which is the space of all Borel probability measures μ on $\mathbb{R}^3$ satisfying $\int_{\mathbb{R}^3} \|x\|_{\ell_2^3}^2 \, \mathrm{d}\mu(x) < \infty$ equipped with the Wasserstein W_2-distance.

6.5. Coarse non-universality

Let $(\mathcal{M}, d_{\mathcal{M}})$ and $(\mathcal{N}, d_{\mathcal{N}})$ be two metric spaces and $\omega, \Omega : [0,\infty) \to [0,\infty)$ two moduli satisfying $\omega \le \Omega$ pointwise and $\lim_{t\to\infty} \omega(t) = \infty$. A mapping $f : \mathcal{M} \to \mathcal{N}$ is a coarse embedding with lower and upper moduli ω and Ω respectively if

$$\forall\, x, y \in \mathcal{M}, \qquad \omega\Big(d_{\mathcal{M}}(x,y)\Big) \le d_{\mathcal{N}}\Big(f(x), f(y)\Big) \le \Omega\Big(d_{\mathcal{M}}(x,y)\Big). \tag{87}$$

A family of metric spaces $\{(\mathcal{M}_\alpha, d_{\mathcal{M}_\alpha})\}_\alpha$ is said to embed equi-coarsely into a metric space $(\mathcal{N}, d_{\mathcal{N}})$ if there exist two moduli $\omega, \Omega : [0,\infty) \to [0,\infty)$ satisfying $\omega \le \Omega$ pointwise and $\lim_{t\to\infty} \omega(t) = \infty$ and a family of coarse embeddings $\{f_\alpha : \mathcal{M}_\alpha \to \mathcal{N}\}_\alpha$ with lower and upper moduli ω and Ω. The pertinence of nonlinear spectral gaps in coarse geometry stems from the following influential observation of Gromov (2000, 2003).

Proposition 6.7. — *Fix* $d \in \mathbb{N}$*,* $p \in (0,\infty)$ *and let* $(\mathcal{M}, d_{\mathcal{M}})$ *be a metric space. Suppose that* $\{G_n = (V_n, E_n)\}_{n=1}^{\infty}$ *is a sequence of connected* d*-regular graphs with* $|V_n| \to \infty$ *and* $\sup_{n\in\mathbb{N}} \gamma(A_{G_n}, d_{\mathcal{M}}^p) < \infty$*. Then, the family of graphs* $\{(G_n, d_{G_n})\}_{n=1}^{\infty}$ *equipped with their shortest path distances does not equi-coarsely embed into* $\mathcal{M}$*.*

Proof. — Let $\gamma \stackrel{\text{def}}{=} \sup_{n\in\mathbb{N}} \gamma(A_{G_n}, d_{\mathcal{M}}^p) < \infty$. Suppose that there exist two moduli $\omega, \Omega : [0,\infty) \to [0,\infty)$ with $\lim_{t\to\infty} \omega(t) = \infty$ and mappings $f_n : V_n \to \mathcal{M}$ with

$$\forall\, x, y \in V_n, \qquad \omega\Big(d_{G_n}(x,y)\Big) \le d_{\mathcal{M}}\Big(f(x), f(y)\Big) \le \Omega\Big(d_{G_n}(x,y)\Big). \tag{88}$$

By definition of nonlinear spectral gaps, we have

$$\frac{1}{|V_n|^2} \sum_{u,v\in V_n} d_{\mathcal{M}}\Big(f_n(u), f_n(v)\Big)^p \le \frac{2\gamma}{d|V_n|} \sum_{\{a,b\}\in E_n} d_{\mathcal{M}}\Big(f_n(a), f_n(b)\Big)^p. \tag{89}$$

Moreover, using the upper modulus, we get

$$\frac{2\gamma}{d|V_n|} \sum_{\{a,b\}\in E_n} d_{\mathcal{M}}\Big(f_n(a), f_n(b)\Big)^p \stackrel{(88)}{\le} \frac{2\gamma}{d|V_n|} \sum_{\{a,b\}\in E_n} \Omega(1)^p = \gamma\Omega(1)^p. \tag{90}$$

Finally, as each graph G_n is d-regular, for any $u \in V_n$ there exist at least $\frac{|V_n|}{2}$ vertices $v \in V_n$ such that $d_{G_n}(u,v) \geq \lfloor \log_d(|V_n|/2) \rfloor$. Thus, the lower modulus gives

$$\frac{1}{|V_n|^2} \sum_{u,v \in V_n} d_{\mathcal{M}}\Big(f_n(u), f_n(v)\Big)^p \overset{(88)}{\geq} \frac{1}{|V_n|^2} \sum_{u,v \in V_n} \omega\Big(d_{G_n}(u,v)\Big)^p \geq \frac{\omega(\lfloor \log_d(|V_n|/2) \rfloor)^p}{2}.$$

Combining all the above, we deduce that

$$\forall\, n \in \mathbb{N}, \qquad \omega(\lfloor \log_d(|V_n|/2) \rfloor)^p \leq 2\gamma\Omega(1)^p, \tag{91}$$

which clearly contradicts the coarse condition $\lim_{t\to\infty} \omega(t) = \infty$. □

An important consequence of Gromov's observation is that if $(\mathcal{M}, d_{\mathcal{M}})$ admits a sequence of regular expanders $\{G_n\}_{n=1}^{\infty}$, then there exists a metric space (e.g. the disjoint union $\bigsqcup_{n\geq 1}(G_n, d_{G_n})$) which does not admit a coarse embedding into $\mathcal{M}$. Consequently, the mere existence of combinatorial expanders implies that Hilbert spaces are not coarsely universal which is a well-known theorem of Dranishnikov, Gong, Lafforgue, and Yu (2002). Moreover, Lafforgue's Theorem 6.5 implies the existence of a metric space which does not admit a coarse embedding into any Banach space of non-trivial type. The coarse non-universality of this class was previously established in work of Mendel and Naor (2008) by proving that Banach spaces of non-trivial type with cotype q have *sharp* metric cotype q. Understanding whether every Banach space of cotype q has sharp metric cotype q is the central open problem in the theory of metric cotype of Banach spaces; see Giladi, Mendel, and Naor (2011) for the best known results to date. If this was the case, then the following (currently open) question on coarse embeddings would have a negative answer.

Question 8. — *Does every separable metric space embed coarsely into some Banach space of finite cotype?*

It follows from Proposition 6.7 that a negative answer to Question 8 would also be a consequence of the existence of a sequence of regular graphs $\{G_n\}_{n=1}^{\infty}$ which are expanders with respect to any normed space of finite cotype simultaneously, let alone from a positive answer to the much stronger Question 3.

Despite the fact that Question 5 on the existence of expanders with respect to Alexandrov spaces of nonpositive curvature remains open, the coarse non-universality of this class was established by Eskenazis, Mendel, and Naor (2019), thus answering a question raised by Gromov (1993). The main technical contribution of this work is the proof that every CAT(0) space $\mathcal{M}$ has sharp metric cotype 2 which formally implies that ℓ_q does not admit a coarse embedding in $\mathcal{M}$ for any $q > 2$. In contrast to this result, the very surprising fact that there exist coarsely universal Alexandrov spaces of nonnegative curvature was proven by Andoni, Naor, and Neiman (2018).

6.6. Approximate nearest neighbor search

Fix a parameter $c > 1$. The c-Approximate Nearest Neighbor Search problem is defined as follows. Given an n-point dataset $\mathcal{P}$ in some metric space $(\mathcal{M}, d_{\mathcal{M}})$, we want to build a data structure[(7)] that, given any query point $q \in \mathcal{M}$, returns a point $\widehat{p} \in \mathcal{P}$ with $d_{\mathcal{M}}(q, \widehat{p}) \leq c \min_{p \in \mathcal{P}} d_{\mathcal{M}}(q, p)$. In practice, this problem can be reduced to its "decision version" (see Har-Peled, Indyk, and Motwani, 2012), which is the c-Approximate *Near* Neighbor Search (c-ANN) problem at a *pre-fixed* distance scale $r > 0$. In the c-ANN problem at scale r, we are again given an n-point dataset $\mathcal{P}$ in some metric space $(\mathcal{M}, d_{\mathcal{M}})$ and we want to build a data structure that, given any query point $q \in \mathcal{M}$ for which there exists a point $p^* \in \mathcal{P}$ with $d_{\mathcal{M}}(q, p^*) \leq r$, returns a point $\widehat{p} \in \mathcal{P}$ with $d_{\mathcal{M}}(q, \widehat{p}) \leq cr$. The main parameters to optimize are the *space* the data structure occupies and the *time* it takes to answer a query. A majority of the research conducted on this problem has focused on d-dimensional normed spaces rather than general metric spaces and moreover most of the algorithms in the literature are *randomized* in the sense that they return a random point $\widehat{p}$ satisfying $d_{\mathcal{M}}(q, \widehat{p}) \leq cr$ with probability at least $1 - \delta$ for some pre-fixed confidence parameter $\delta \in (0, 1)$.

The first approaches to the c-ANN problem for d-dimensional norms produced data-independent data structures, in which the memory cells accessed by the algorithm do not depend on the dataset $\mathcal{P}$ but only on the query point q. In particular, building such data structures via (oblivious) metric dimension reduction has been used with great success for the Hilbert space ℓ_2^d (Indyk and Motwani, 1999; Har-Peled, Indyk, and Motwani, 2012), the hypercube $\{-1, 1\}^d$ equipped with the Hamming distance (Kushilevitz, Ostrovsky, and Rabani, 2000) and spaces which (effectively) embed in them (Andoni, Indyk, and Krauthgamer, 2009; Nguyễn, 2014). While dimension reduction techniques yield data structures with polynomial space for these norms, these results are often far from practical due to the large degree of said polynomial. To overcome this barrier, Indyk and Motwani (1999) introduced an influential technique called *Locality-Sensitive Hashing* (LSH) relying on (data-independent) randomized space partitions. Somewhat informally, a distribution $\mathcal{D}$ over a family of partitions of $\mathcal{M}$ is called *sensitive* at scale r up to error c if any two points at distance at most r are $\mathcal{D}$-likely to belong in the same cluster of the partition and any two points at distance at least cr are $\mathcal{D}$-unlikely to do so (where the implicit probabilities affect the space and time requirements of the data structure). As shown by Indyk and Motwani (1999), a (computationally efficient) sensitive distribution over random partitions can serve as a pre-filter for the dataset $\mathcal{P}$ as the query point q is very likely to be indistinguishable from its near neighbors but is unlikely to collide with points p having $d_{\mathcal{M}}(q, p) > cr$. Using LSH, they were able to improve the space requirements over the existing c-ANN algorithms to almost linear for large enough accuracy parameters $c > 1$. We refer to the

(7) For the purposes of this survey, a data structure of size M is an array $A[1 \ldots M]$ of numbers (the "memory") along with an algorithm which, given a point $q \in \mathcal{M}$, returns a point $\widehat{p} \in \mathcal{P}$.

thorough survey of Andoni, Indyk, and Razenshteyn (2018) for a detailed account of these and other contributions on the c-ANN problem and further references.

Despite these advances towards the c-ANN problem, researchers proved strong impossibility results (Motwani, Naor, and Panigrahy, 2007; O'Donnell, Wu, and Zhou, 2014) for the existence of data-independent data structures arising from LSH, thus creating the necessity for the development of efficient data-*dependent* algorithms. Historically, the first such result was proven by Indyk (2001) for ℓ_∞^d. In recent years, this approach has gained a lot of momentum, especially in view of the works of Andoni, Indyk, and Krauthgamer (2009) for the Ulam metric, Andoni, Indyk, Nguyễn, and Razenshteyn (2014) and Andoni and Razenshteyn (2015) for ℓ_2^d and Andoni, Nguyen, Nikolov, Razenshteyn, and Waingarten (2017) for 1-symmetric norms. A breakthrough in this direction was presented in the work of Andoni, Naor, Nikolov, Razenshteyn, and Waingarten (2018a) who showed the following theorem for *general* d-dimensional normed spaces. It is worth emphasizing that their result does not a priori give any bound on the running time of the algorithm, it just restricts the number of memory locations the data structure is allowed to probe.

Theorem 6.8. — *Fix $\varepsilon \in (0,1)$ and let X be a d-dimensional normed space. There exists a randomized data structure for $O\Big(\frac{\log d}{\varepsilon^2}\Big)$-ANN over X with the following properties:*

- *The space used by the data structure is $n^{1+\varepsilon} \cdot d^{O(1)}$;*
- *The query procedure probes $n^{\varepsilon} \cdot d^{O(1)}$ words in memory.*

In order to prove Theorem 6.8, the authors introduced a geometric parameter called the cutting modulus $\Xi(\mathcal{M}, \varepsilon)$ associated to a metric space $(\mathcal{M}, d_{\mathcal{M}})$ and a parameter $\varepsilon \in (0,1)$, which governs the following data-dependent partitioning scheme: every finite dataset in $\mathcal{M}$ either has a subset of proportional size (measured appropriately) which is contained in a ball of radius $\Xi(\mathcal{M}, \varepsilon)$ or admits a cut which is ε-sparse. Relying on this notion, they were able to show the following general partitioning theorem.

Theorem 6.9. — *Let $(\mathcal{M}, d_{\mathcal{M}})$ be a finite metric space and fix $n \in \mathbb{N}$, $\varepsilon \in (0, \frac{1}{4})$. There exists a collection $\mathcal{C}$ of subsets of $\mathcal{M}$ with $\log|\mathcal{C}| = O\Big(\log|\mathcal{M}| \log(\log|\mathcal{M}|/\varepsilon)\Big)$ such that for any n-point dataset $\mathcal{P}$ in $\mathcal{M}$, we have one of the following two properties:*

- *Either there exists $x_0 \in \mathcal{M}$ and $R \leq \Xi(\mathcal{M}, \varepsilon)$ such that $|\mathcal{P} \cap B_{\mathcal{M}}(x_0, R)| \geq \frac{n}{50}$, or*
- *There exists a subcollection $\{S_1, \ldots, S_m\} \subseteq \mathcal{C}$ such that*

$$\forall\, i \in \{1, \ldots, m\}, \qquad \frac{n}{50} \leq |S_i \cap \mathcal{P}| \leq \frac{49n}{50} \tag{92}$$

and for every $x, y \in \mathcal{M}$ with $d_{\mathcal{M}}(x,y) \leq 1$, we have

$$\#\Big\{i \in \{1, \ldots, m\} : \ \mathbf{1}_{S_i}(x) \neq \mathbf{1}_{S_i}(y)\Big\} \leq 50\varepsilon m. \tag{93}$$

Theorem 6.9 suggests a very natural LSH with approximation $O(\Xi(\mathcal{M}, \varepsilon))$ since, at each step of the algorithm, we either have a dense ball of radius $\Xi(\mathcal{M}, \varepsilon)$ or we have a collection of subsets with a distribution that decreases the size of the dataset and rarely

splits the query from its nearby points in the dataset. The relevance of those results with the subject of this survey stems from the fact that Theorem 1.3 implies that if X is a normed space, then the cutting modulus satisfies

$$\Xi(X,\varepsilon) \lesssim \frac{\log(\dim(X)+1)}{\varepsilon^2}. \tag{94}$$

The main idea of the proof of (94) is to apply (9) to the adjacency matrices of geometric graphs associated to finite subsets of X. If such a graph does not have a subset of proportional size (with respect to the underlying stationary measure) contained in a ball of radius $\Omega(\log(\dim(X)))$, then the nonlinear spectral gap inequality (9) implies that it also cannot have large (classical) spectral gap and thus admits a sparse cut by Cheeger's inequality. A combination of Theorem 6.9 and (94) implies Theorem 6.8. We refer to the work of Andoni, Naor, Nikolov, Razenshteyn, and Waingarten (2018a) for the precise definition of the cutting modulus and the proofs of these results.

In their follow-up work, Andoni, Naor, Nikolov, Razenshteyn, and Waingarten (2018b) proved the existence of data structures for ANN over d-dimensional normed spaces with slightly worse (but still subpolynomial) approximation and reasonable bounds for the running time. Some of their results still rely on elements of the theory of nonlinear spectral gaps, whereas others use the existence of a remarkable uniform homeomorphism between spheres of Banach spaces which originates in the resolution of the distortion problem by Odell and Schlumprecht (1994).

REFERENCES

Ittai Abraham, Yair Bartal, and Ofer Neiman (2011). "Advances in metric embedding theory", *Adv. Math.* **228** (6), pp. 3026–3126.

Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer (2009). "Overcoming the ℓ_1 non-embeddability barrier: algorithms for product metrics". In: *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, Philadelphia, PA, pp. 865–874.

Alexandr Andoni, Piotr Indyk, Huy L. Nguyễn, and Ilya Razenshteyn (2014). "Beyond locality-sensitive hashing". In: *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, New York, pp. 1018–1028.

Alexandr Andoni, Piotr Indyk, and Ilya Razenshteyn (2018). "Approximate nearest neighbor search in high dimensions". In: *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited lectures*. World Sci. Publ., Hackensack, NJ, pp. 3287–3318.

Alexandr Andoni, Assaf Naor, and Ofer Neiman (2018). "Snowflake universality of Wasserstein spaces", *Ann. Sci. Éc. Norm. Supér. (4)* **51** (3), pp. 657–700.

Alexandr Andoni, Assaf Naor, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten (2018a). "Data-dependent hashing via nonlinear spectral gaps". In: *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, pp. 787–800.

——— (2018b). "Hölder homeomorphisms and approximate nearest neighbors", in: *59th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2018*. IEEE Computer Soc., Los Alamitos, CA, pp. 159–169.

Alexandr Andoni, Huy L. Nguyen, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten (2017). "Approximate near neighbors for general symmetric norms". In: *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, pp. 902–913.

Alexandr Andoni and Ilya Razenshteyn (2015). "Optimal data-dependent hashing for approximate near neighbors". In: *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*. ACM, New York, pp. 793–801.

Juan Arias-de-Reyna and Luis Rodríguez-Piazza (1992). "Finite metric spaces needing high dimension for Lipschitz embeddings in Banach spaces", *Israel J. Math.* **79** (1), pp. 103–111.

Keith Ball (1990). "Isometric embedding in l_p-spaces", *European J. Combin.* **11** (4), pp. 305–311.

——— (1992). "Markov chains, Riesz transforms and Lipschitz maps", *Geom. Funct. Anal.* **2** (2), pp. 137–172.

——— (2013). "The Ribe programme", *Astérisque* (352). Séminaire Bourbaki. Vol. 2011/2012. Exposés 1043–1058, Exp. No. 1047, viii, 147–159.

Keith Ball, Eric A. Carlen, and Elliott H. Lieb (1994). "Sharp uniform convexity and smoothness inequalities for trace norms", *Invent. Math.* **115** (3), pp. 463–482.

Florent P. Baudier and William B. Johnson (2016). "*Metric embeddings: bilipschitz and coarse embeddings into Banach spaces* [book review of MR3114782]", *Bull. Amer. Math. Soc. (N.S.)* **53** (3), pp. 495–506.

Yoav Benyamini and Joram Lindenstrauss (2000). *Geometric nonlinear functional analysis. Vol. 1.* Vol. 48. American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, pp. xii+488.

Béla Bollobás (1988). "The isoperimetric number of random regular graphs", *European J. Combin.* **9** (3), pp. 241–244.

Jean Bourgain (1985). "On Lipschitz embedding of finite metric spaces in Hilbert space", *Israel J. Math.* **52** (1-2), pp. 46–52.

——— (1986). "The metrical interpretation of superreflexivity in Banach spaces", *Israel J. Math.* **56** (2), pp. 222–230.

Jean Bretagnolle, Didier Dacunha-Castelle, and Jean-Louis Krivine (1965/1966). "Lois stables et espaces L^p", *Ann. Inst. H. Poincaré Sect. B (N.S.)* **2**, pp. 231–259.

Bo Brinkman and Moses Charikar (2005). "On the impossibility of dimension reduction in l_1", *J. ACM* **52** (5), pp. 766–788.

Qingjin Cheng (2016). "Sphere equivalence, property H, and Banach expanders", *Studia Math.* **233** (1), pp. 67–83.

Giuliana Davidoff, Peter Sarnak, and Alain Valette (2003). *Elementary number theory, group theory, and Ramanujan graphs*. Vol. 55. London Mathematical Society Student Texts. Cambridge University Press, Cambridge, pp. x+144.

Tim de Laat and Mikael de la Salle (2021). "Banach space actions and L^2-spectral gap", *Anal. PDE* **14** (1), pp. 45–76.

Alexander N. Dranishnikov, Guihua Gong, Vincent Lafforgue, and Guoliang Yu (2002). "Uniform embeddings into Hilbert space and a question of Gromov", *Canad. Math. Bull.* **45** (1), pp. 60–70.

Aryeh Dvoretzky (1961). "Some results on convex bodies and Banach spaces", in: *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, pp. 123–160.

Alexandros Eskenazis (2019). *Geometric Inequalities and Advances in the Ribe Program*. Thesis (Ph.D.)–Princeton University. ProQuest LLC, Ann Arbor, MI, p. 399.

——— (2021). "ε-isometric dimension reduction for incompressible subsets of ℓ_p". Preprint available at https://arxiv.org/abs/2109.06602.

Alexandros Eskenazis and Paata Ivanisvili (2020). "Polynomial inequalities on the Hamming cube", *Probab. Theory Related Fields* **178** (1-2), pp. 235–287.

——— (2021). "Sharp growth of the Ornstein–Uhlenbeck operator on Gaussian tail spaces". To appear in *Israel J. Math.*. Preprint available at https://arxiv.org/abs/2011.01359.

Alexandros Eskenazis, Manor Mendel, and Assaf Naor (2019). "Nonpositive curvature is not coarsely universal", *Invent. Math.* **217** (3), pp. 833–886.

Alexandros Eskenazis and Assaf Naor (2021). "On coarse and uniform embeddings into L_p". Preprint.

Ohad Giladi, Manor Mendel, and Assaf Naor (2011). "Improved bounds in the metric cotype inequality for Banach spaces", *J. Funct. Anal.* **260** (1), pp. 164–194.

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

Mikhail Gromov (1993). "Asymptotic invariants of infinite groups", in: *Geometric group theory, Vol. 2 (Sussex, 1991)*. Vol. 182. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 1–295.

——— (2000). "Spaces and questions", *Geom. Funct. Anal.* (Special Volume, Part I). GAFA 2000 (Tel Aviv, 1999), pp. 118–161.

——— (2003). "Random walk in random groups", *Geom. Funct. Anal.* **13** (1), pp. 73–146.

Sariel Har-Peled, Piotr Indyk, and Rajeev Motwani (2012). "Approximate nearest neighbor: towards removing the curse of dimensionality", *Theory Comput.* **8**, pp. 321–350.

Stefan Heinrich (1980). "Ultraproducts in Banach space theory", *J. Reine Angew. Math.* **313**, pp. 72–104.

Piotr Indyk (2001). "On approximate nearest neighbors under l_∞ norm", in: vol. 63. 4. Special issue on FOCS 98 (Palo Alto, CA), pp. 627–638.

Piotr Indyk and Rajeev Motwani (1999). "Approximate nearest neighbors: towards removing the curse of dimensionality", in: *STOC '98 (Dallas, TX)*. ACM, New York, pp. 604–613.

Fritz John (1948). "Extremum problems with inequalities as subsidiary conditions", in: *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*. Interscience Publishers, Inc., New York, N. Y., pp. 187–204.

William B. Johnson and Joram Lindenstrauss (1984). "Extensions of Lipschitz mappings into a Hilbert space", in: *Conference in modern analysis and probability (New Haven, Conn., 1982)*. Vol. 26. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 189–206.

William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman (1987). "On Lipschitz embedding of finite metric spaces in low-dimensional normed spaces", in: *Geometrical aspects of functional analysis (1985/86)*. Vol. 1267. Lecture Notes in Math. Springer, Berlin, pp. 177–184.

Nigel J. Kalton (2008). "The nonlinear geometry of Banach spaces", *Rev. Mat. Complut.* **21** (1), pp. 7–60.

Takefumi Kondo (2012). "CAT(0) spaces and expanders", *Math. Z.* **271** (1-2), pp. 343–355.

Jean-Louis Krivine (1965). "Plongement des espaces normés dans les L^p pour $p > 2$", *C. R. Acad. Sci. Paris* **261**, pp. 4307–4310.

Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani (2000). "Efficient search for approximate nearest neighbor in high dimensional spaces", *SIAM J. Comput.* **30** (2), pp. 457–474.

Vincent Lafforgue (2008). "Un renforcement de la propriété (T)", *Duke Math. J.* **143** (3), pp. 559–602.

——— (2009). "Propriété (T) renforcée banachique et transformation de Fourier rapide", *J. Topol. Anal.* **1** (3), pp. 191–206.

James R. Lee, Manor Mendel, and Assaf Naor (2005). "Metric structures in L_1: dimension, snowflakes, and average distortion", *European J. Combin.* **26** (8), pp. 1180–1190.

James R. Lee and Assaf Naor (2004). "Embedding the diamond graph in L_p and dimension reduction in L_1", *Geom. Funct. Anal.* **14** (4), pp. 745–747.

Nathan Linial, Eran London, and Yuri Rabinovich (1995). "The geometry of graphs and some of its algorithmic applications", *Combinatorica* **15** (2), pp. 215–245.

Jiří Matoušek (1992). "Note on bi-Lipschitz embeddings into normed spaces", *Comment. Math. Univ. Carolin.* **33** (1), pp. 51–55.

Jiří Matoušek (1996). "On the distortion required for embedding finite metric spaces into normed spaces", *Israel J. Math.* **93**, pp. 333–344.

——— (1997). "On embedding expanders into l_p spaces", *Israel J. Math.* **102**, pp. 189–197.

——— (2002). *Lectures on discrete geometry.* Vol. 212. Graduate Texts in Mathematics. Springer-Verlag, New York, pp. xvi+481.

Bernard Maurey (1974). *Théorèmes de factorisation pour les opérateurs linéaires à valeurs dans les espaces L^p.* Astérisque, No. 11. With an English summary. Société Mathématique de France, Paris, pp. ii+163.

——— (2003). "Type, cotype and K-convexity", in: *Handbook of the geometry of Banach spaces, Vol. 2.* North-Holland, Amsterdam, pp. 1299–1332.

Stanisław Mazur (1929). "Une remarque sur l'homéomorphie des champs fonctionnels", fre. *Studia Math.* **1** (1), pp. 83–85.

Manor Mendel and Assaf Naor (2008). "Metric cotype", *Ann. of Math. (2)* **168** (1), pp. 247–298.

——— (2013). "Spectral calculus and Lipschitz extension for barycentric metric spaces", *Anal. Geom. Metr. Spaces* **1**, pp. 163–199.

——— (2014). "Nonlinear spectral calculus and super-expanders", *Publ. Math. Inst. Hautes Études Sci.* **119**, pp. 1–95.

——— (2015). "Expanders with respect to Hadamard spaces and random graphs", *Duke Math. J.* **164** (8), pp. 1471–1548.

Masato Mimura (2015). "Sphere equivalence, Banach expanders, and extrapolation", *Int. Math. Res. Not. IMRN* (12), pp. 4372–4391.

Rajeev Motwani, Assaf Naor, and Rina Panigrahy (2007). "Lower bounds on locality sensitive hashing", *SIAM J. Discrete Math.* **21** (4), pp. 930–935.

Assaf Naor (2012). "An introduction to the Ribe program", *Jpn. J. Math.* **7** (2), pp. 167–233.

——— (2014). "Comparison of metric spectral gaps", *Anal. Geom. Metr. Spaces* **2** (1), pp. 1–52.

——— (2017). "A spectral gap precludes low-dimensional embeddings", in: *33rd International Symposium on Computational Geometry.* Vol. 77. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, Art. No. 50, 16.

——— (2018). "Metric dimension reduction: a snapshot of the Ribe program". In: *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. I. Plenary lectures.* World Sci. Publ., Hackensack, NJ, pp. 759–837.

——— (2021). "An average John theorem", *Geom. Topol.* **25** (4), pp. 1631–1717.

Assaf Naor, Yuval Peres, Oded Schramm, and Scott Sheffield (2006). "Markov chains in smooth Banach spaces and Gromov-hyperbolic metric spaces", *Duke Math. J.* **134** (1), pp. 165–197.

Assaf Naor and Lior Silberman (2011). "Poincaré inequalities, embeddings, and wild groups", *Compos. Math.* **147** (5), pp. 1546–1572.

Huy L. Nguyễn (2014). *Algorithms for High Dimensional Data.* Thesis (Ph.D.)–Princeton University.

Edward Odell and Thomas Schlumprecht (1994). "The distortion problem", *Acta Math.* **173** (2), pp. 259–281.

Ryan O'Donnell, Yi Wu, and Yuan Zhou (2014). "Optimal lower bounds for locality-sensitive hashing (except when q is tiny)", *ACM Trans. Comput. Theory* **6** (1), Art. 5, 13.

Mikhail I. Ostrovskii (2013). *Metric embeddings.* Vol. 49. De Gruyter Studies in Mathematics. Bilipschitz and coarse embeddings into Banach spaces. De Gruyter, Berlin, pp. xii+372.

Narutaka Ozawa (2004). "A note on non-amenability of $\mathcal{B}(l_p)$ for $p = 1, 2$", *Internat. J. Math.* **15** (6), pp. 557–565.

Mark S. Pinsker (1973). "On the complexity of a concentrator". In: *7th International Teletraffic Conference.*

Gilles Pisier (1975a). "Martingales with values in uniformly convex spaces", *Israel J. Math.* **20** (3-4), pp. 326–350.

——— (1975b). "Un exemple concernant la super-réflexivité", in: *Séminaire Maurey-Schwartz 1974–1975: Espaces L^p applications radonifiantes et géométrie des espaces de Banach, Annexe No. 2*, p. 12.

——— (2010). "Complex interpolation between Hilbert, Banach and operator spaces", *Mem. Amer. Math. Soc.* **208** (978), pp. vi+78.

Yuri Rabinovich (2003). "On average distortion of embedding metrics into the line and into l_1". In: *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing.* ACM, New York, pp. 456–462.

Omer Reingold, Salil Vadhan, and Avi Wigderson (2002). "Entropy waves, the zig-zag graph product, and new constant-degree expanders", *Ann. of Math. (2)* **155** (1), pp. 157–187.

Martin Ribe (1976). "On uniformly homeomorphic normed spaces", *Ark. Mat.* **14** (2), pp. 237–244.

Isaac J. Schoenberg (1938). "Metric spaces and positive definite functions", *Trans. Amer. Math. Soc.* **44** (3), pp. 522–536.

Mu-Tao Wang (1998). "A fixed point theorem of discrete group actions on Riemannian manifolds", *J. Differential Geom.* **50** (2), pp. 249–267.

James H. Wells and Lynn R. Williams (1975). *Embeddings and extensions in analysis.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 84. Springer-Verlag, New York-Heidelberg, pp. vii+108.

Alexandros Eskenazis

Trinity College
University of Cambridge
Cambridge, CB2 1TQ, United Kingdom

Institut de Mathématiques de Jussieu
CNRS & Sorbonne Université
Paris, 75252, France
E-mail : ae466@cam.ac.uk