Séminaire BOURBAKI Janvier 2022
73e année, 2021–2022, nº 1188, p. 295 à 333
doi: 10.24033/ast.1189

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

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

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

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

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

The 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).

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.

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 $(X_t)_{t\geqslant 0}$ 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 $t \geqslant 0$. If the transition matrix A is π-reversible, then π is also a stationary distribution for the process $(X_t)_{t\geqslant 0}$, that is, if X_0 is distributed according to π then so is X_t for any $t \geqslant 1$. 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 $1 = \lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_n(A) \geqslant -1$. 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 $\gamma(A) \stackrel{\mathrm{def}}{=} (1 - \lambda_2(A))^{-1}$ of the spectral gap is the least constant $\gamma \in (0,\infty]$ for which the inequality

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

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

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

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, et al. (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 $\|y_i - y_j\|_Y \leqslant D d_{\mathcal{M}}(x_i, x_j)^\theta$ for every $i, j \in \{1, \ldots, n\}$ and

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

Therefore, we have

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

which implies that $\gamma(A, d_{\mathcal{M}}^{\theta q}) \leqslant D^q \gamma(A, \|\cdot\|_Y^q)$. 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

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

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 $\gamma(A, d_{\mathcal{M}}^{\theta q}) \leqslant D^q \gamma(A, \|\cdot\|_Y^q)$ 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*

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

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 $D + \varepsilon$.*

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*

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

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

(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 $\varepsilon > 0$.